Stochastic Calculus for Finance
Brief Lecture Notes

Gautam Iyer

Preface

The purpose of these notes is to provide a rapid introduction to the Black-Scholes formula and the mathematical techniques used to derive it. Most mathematical concepts used are explained and motivated, but the complete rigorous proofs are beyond the scope of these notes. These notes were written in 2017 when I was teaching a seven week course in the Masters in Computational Finance program at Carnegie Mellon University.

The notes are somewhat minimal and mainly include material that was covered during the lectures itself. Only two sets of problems are included. These are problems that were used as a review for the midterm and final respectively. Supplementary problems and exams can be found on the course website: http://www.math.cmu.edu/~gautam/sj/teaching/2018-19/944-scalc-finance1.

For more comprehensive references and exercises, I recommend:

(1) *Stochastic Calculus for Finance II* by Steven Shreve.
(2) *The basics of Financial Mathematics* by Rich Bass.
(3) *Introduction to Stochastic Calculus with Applications* by Fima C Klebaner.

Contents

CHAPTER 1

Introduction

The price of a stock is not a smooth function of time, and standard calculus tools can not be used to effectively model it. A commonly used technique is to model the price S as a *geometric Brownian motion*, given by the *stochastic differential equation (SDE)*

$$dS(t) = \alpha S(t)\,dt + \sigma S(t)\,dW(t)\,,$$

where α and σ are parameters, and W is a Brownian motion. If $\sigma = 0$, this is simply the ordinary differential equation

$$dS(t) = \alpha S(t)\,dt \quad \text{or} \quad \partial_t S = \alpha S(t)\,.$$

This is the price assuming it grows at a rate α. The $\sigma\,dW$ term models *noisy fluctuations* and the first goal of this course is to understand what this means. The mathematical tools required for this are Brownian motion, and Itô integrals, which we will develop and study.

An important point to note is that the above model can not be used to *predict* the price of S, because randomness is built into the model. Instead, we will use this model is to *price securities*. Consider a *European call option* for a stock S with strike prices K and maturity T (i.e. this is the right to buy the asset S at price K at time T). Given the stock price $S(t)$ at some time $t \leqslant T$, what is a fair price for this option?

Seminal work of Black and Scholes computes the fair price of this option in terms of the time to maturity $T - t$, the stock price $S(t)$, the strike price K, the model parameters α, σ and the interest rate r. For notational convenience we suppress the explicit dependence on K, α, σ and let $c(t, x)$ represent the price of the option at time t given that the stock price is x. Clearly $c(T, x) = (x - K)^+$. For $t \leqslant T$, the Black-Scholes formula gives

$$c(t, x) = xN(d_+(T - t, x)) - Ke^{-r(T-t)}N(d_-(T - t, x))$$

where

$$d_\pm(\tau, x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{\tau}}\Big(\ln\Big(\frac{x}{K}\Big) + \Big(r \pm \frac{\sigma^2}{2}\Big)\tau\Big)\,.$$

Here r is the interest rate at which you can borrow or lend money, and N is the CDF of a standard normal random variable. (It might come as a surprise to you that the formula above is independent of α, the mean return rate of the stock.)

The second goal of this course is to understand the derivation of this formula. The main idea is to find a *replicating strategy*. If you've sold the above option, you hedge your bets by investing the money received in the underlying asset, and in an interest bearing account. Let $X(t)$ be the value of your portfolio at time t, of which you have $\Delta(t)$ invested in the stock, and $X(t) - \Delta(t)$ in the interest bearing account. If we are able to choose $X(0)$ and Δ in a way that would guarantee

$X(T) = (S(T) - K)^+$ almost surely, then $X(0)$ must be the fair price of this option. In order to find this strategy we will need to understand SDEs and the Itô formula, which we will develop subsequently.

The final goal of this course is to understand *risk neutral measures*, and use them to provide an elegant derivation of the Black-Scholes formula. If time permits, we will also study the *fundamental theorems of asset pricing*, which roughly state:

(1) The *existence* of a risk neutral measure is equivalent to the market having no arbitrage (i.e. you can't make money without taking risk).
(2) *Uniqueness* of a risk neutral measure is equivalent to the market having no arbitrage, and that all derivative securities can be hedged (i.e. for every derivative security we can find a replicating portfolio).

CHAPTER 2

Brownian motion, and an Introduction to Modern Probability

1. Scaling limit of random walks.

Our first goal is to understand *Brownian motion*, which is used to model "noisy fluctuations" of stocks, and various other objects. This is named after the botanist Robert Brown, who observed that the microscopic movement of pollen grains appears random. Intuitively, Brownian motion can be thought of as a process that performs a random walk in continuous time.

We begin by describing Brownian motion as the scaling limit of discrete random walks. Let $\xi_1, \xi_2, \ldots,$ be a sequence of i.i.d. random variables which take on the values ± 1 with probability $1/2$. Define the time interpolated random walk $S(t)$ by setting $S(0) = 0$, and

$$S(t) = S(n) + (t - n)\xi_{n+1} \quad \text{when } t \in (n, n+1]\,. \tag{1.1}$$

Note $S(n) = \sum_1^n \xi_i$, and so at integer times S is simply a symmetric random walk with step size 1.

Our aim now is to rescale S so that it takes a random step at shorter and shorter time intervals, and then take the limit. In order to get a meaningful limit, we will have to compensate by also scaling the step size. Let $\varepsilon > 0$ and define

$$S_\varepsilon(t) = \alpha_\varepsilon S\Big(\frac{t}{\varepsilon}\Big)\,, \tag{1.2}$$

where α_ε will be chosen below in a manner that ensures convergence of $S_\varepsilon(t)$ as $\varepsilon \to 0$. Note that S_ε now takes a random step of size α_ε after every ε time units.

To choose α_ε, we compute the variance of S_ε. Note first

$$\operatorname{Var} S(t) = \lfloor t \rfloor + (t - \lfloor t \rfloor)^2,$$

and[1] consequently

$$\operatorname{Var} S_\varepsilon(t) = \alpha_\varepsilon^2 \Big(\Big\lfloor \frac{t}{\varepsilon} \Big\rfloor + \Big(\frac{t}{\varepsilon} - \Big\lfloor \frac{t}{\varepsilon} \Big\rfloor \Big)^2 \Big)\,.$$

In order to get a "nice limit" of S_ε as $\varepsilon \to 0$, one would at least expect that $\operatorname{Var} S_\varepsilon(t)$ converges as $\varepsilon \to 0$. From the above, we see that choosing

$$\alpha_\varepsilon = \sqrt{\varepsilon}$$

immediately implies

$$\lim_{\varepsilon \to 0} \operatorname{Var} S_\varepsilon(t) = t\,.$$

[1] Here $\lfloor x \rfloor$ denotes the greatest integer smaller than x. That is, $\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leqslant x\}$.

THEOREM 1.1. *The processes* $S_\varepsilon(t) \stackrel{\text{def}}{=} \sqrt{\varepsilon} S(t/\varepsilon)$ *"converge" as* $\varepsilon \to 0$. *The limiting process, usually denoted by* W*, is called a (standard, one dimensional)* Brownian motion.

The proof of this theorem uses many tools from the modern theory of probability, and is beyond the scope of this course. The important thing to take away from this is that Brownian motion can be well approximated by a random walk that takes steps of variance ε on a time interval of size ε.

While the above construction provides good intuition as to what Brownian motion actually is, the scaling limit it is a somewhat unwieldy object to work with. We instead introduce an intrinsic characterization of Brownian motion, and we will shortly see that is both useful and mathematically convenient.

DEFINITION 1.2. A Brownian motion is a continuous process that has stationary independent increments.

Let us briefly explain the terms appearing in the above definition.

(1) A *process* (aka stochastic process) is simply a collection of random variables $\{X(t) \mid 0 \leqslant t < \infty\}$. The index t usually represents time, and the process is often written as $\{X_t \mid 0 \leqslant t < \infty\}$ instead.
(2) A *trajectory* (aka sample path) of a process is the outcome of one particular realization each of the random variables $X(t)$ viewed as function of time.
(3) A process is called *continuous* if each of the trajectories are continuous. That is, for every $t > 0$ we have

$$\lim_{s \to t} X(s) = X(t)\,. \tag{1.3}$$

(4) An process is said to have *stationary increments* if for every $h \geqslant 0$, the distribution of $X(t+h) - X(t)$ does not depend on t.
(5) A process is said to have *independent increments* if for every finite sequence of times $0 \leqslant t_0 < t_1 \cdots < t_N$, the random variables $X(t_0)$, $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, ..., $X(t_N) - X(t_{N-1})$ are all jointly independent.

For the process S in (1.1), note that for $n \in \mathbb{N}$, $S(n+1) - S(n) = X_{n+1}$ whose distribution *does not* depend on n as the variables $\{\xi_i\}$ were chosen to be independent and identically distributed. Similarly, $S(n+k) - S(n) = \sum_{n+1}^{n+k} \xi_i$ which has the same distribution as $\sum_1^k \xi_i$ and is independent of n.

However, if $t \in \mathbb{R}$ and is not necessarily an integer, $S(t+k) - S(t)$ will in general depend on t. So the process S (and also S_ε) do not have stationary (or independent) increments.

We claim, that the limiting process W does have stationary, independent, *normally distributed* increments. Suppose for some fixed $\varepsilon > 0$, both s and t are multiples of ε. In this case

$$S_\varepsilon(t) - S_\varepsilon(s) \sim \sqrt{\varepsilon} \sum_{i=1}^{\lfloor t-s \rfloor/\varepsilon} \xi_i \xrightarrow{\varepsilon \to 0} N(0, t-s)\,,$$

by the central limit theorem. If s, t aren't multiples of ε as we will have in general, the first equality above is true up to a remainder which can easily be shown to vanish.

The above heuristic argument suggests that the limiting process W (from Theorem 1.1) satisfies $W(t) - W(s) \sim N(0, t-s)$. This certainly has independent increments since $W(t+h) - W(t) \sim N(0,h)$ which is independent of t. Moreover, this also suggests that Brownian motion can be *equivalently* characterized as follows.

DEFINITION 1.3. A Brownian motion is a *continuous process* W such that:

(1) W has independent increments, and
(2) For $s < t$, $W(t) - W(s) \sim N(0, \sigma^2(t-s))$.

REMARK 1.4. A *standard* (one dimensional) Brownian motion is one for which $W(0) = 0$ and $\sigma = 1$.

2. A brief review of probability

In modern probability we usually start with a *probability triple* $(\Omega, \mathcal{G}, \boldsymbol{P})$.

(1) Ω is a non-empty set called the *sample space.*
(2) $\mathcal{G}$ is a σ-algebra. This is a non-empty collection of events (subsets of Ω) of which the probability is known.
(3) $\boldsymbol{P}$ is a *probability measure.* For any event $A \in \mathcal{G}$, $\boldsymbol{P}(A)$ represents the probability of the event A occurring.

A subtle, but important, point in this framework is the in most case $\mathcal{G}$ is usually not the collection of all subsets of Ω, but only a collection of some subsets of Ω. In fact, in most interesting examples, it is *impossible* to define the probability of arbitrary subsets of Ω consistently (i.e. in a manner that satisfies the required properties listed below), and we thus restrict ourselves to only talking about probabilities of elements of elements of the σ-algebra $\mathcal{G}$.

In order to be a probability space, the triple $(\Omega, \mathcal{G}, \boldsymbol{P})$ is required to satisfy certain properties. First the σ-algebra $\mathcal{G}$ must satisfy the following:

(1) It must be closed under compliments. That is, if $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$.
(2) It must be closed under *countable* unions. That us, if A_1, A_2, ... are all elements of $\mathcal{G}$, then the union $\cup_1^\infty A_i$ is also an element of $\mathcal{G}$.

The precise mathematical definition of a σ-algebra is simply a non-empty collection of sets that satisfies the above two properties. Of course, if $\mathcal{G}$ satisfies the above properties then one can quickly deduce the following:

(3) The empty set $\emptyset$ and the whole space Ω are elements of $\mathcal{G}$.
(4) If A_1, A_2, ... are all elements of $\mathcal{G}$, then the intersection $\cap_1^\infty A_i$ is also an element of $\mathcal{G}$.
(5) If A, B are events in $\mathcal{G}$, then $A - B$ is also an event in $\mathcal{G}$.

The reason for requiring the above properties is that we expect $\mathcal{G}$ is the collection of events of which the probability is known (or of which the probability can be deduced by performing repeated trials of some experiment). If you can deduce the probability of an event A, you should certainly be able to deduce the probability of A^c. Similarly, if you can deduce the probabilities of A, B, you should be able to deduce the probability of $A \cup B$ and $A \cap B$. The possibly surprising point above is that we don't require that $\mathcal{G}$ be closed under finite unions, but we require it is closed under *countable* unions. The reason for this is that we would like our framework to allow us to perform repeated trials of an experiment and take limits.

Next, we turn our attention to the probability measure $\boldsymbol{P}$. We require that $\boldsymbol{P}$ satisfies the following properties:

(1) For each $A \in \mathcal{G}$, $\boldsymbol{P}(A) \in [0,1]$. Moreover, $\boldsymbol{P}(\emptyset) = 0$, and $\boldsymbol{P}(\Omega) = 1$.

(2) *(Countable additivity)* Given *pairwise disjoint* events $A_1,\ A_2, \dots \in \mathcal{G}$, we have

$$\boldsymbol{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \boldsymbol{P}(A_i)\,.$$

The above two properties are precisely the formal definition of a probability measure. Recall that in probability we require that the probability of mutually exclusive events add. The second property above is a generalization of this to countably many events.

Using the above properties, one can quickly verify that $\boldsymbol{P}$ also satisfies the following properties:

(1) $\boldsymbol{P}(A^c) = 1 - \boldsymbol{P}(A)$. More generally, if $A, B \in \mathcal{G}$ with $A \subseteq B$, then $\boldsymbol{P}(B - A) = \boldsymbol{P}(B) - \boldsymbol{P}(A)$.
(2) If $A_1 \subseteq A_2 \subseteq A_3 \cdots$ and each $A_i \in \mathcal{G}$ then $\boldsymbol{P}(\cup A_i) = \lim_{n\to\infty} \boldsymbol{P}(A_n)$.
(3) If $A_1 \supseteq A_2 \supseteq A_3 \cdots$ and each $A_i \in \mathcal{G}$ then $\boldsymbol{P}(\cap A_i) = \lim_{n\to\infty} \boldsymbol{P}(A_n)$.

We now describe *random variables* in the above context. In discrete probability, random variables are usually just real valued functions defined on the sample space. In our context, however, we need to be a bit more careful. If X is a random variable, then one should always be able to assign probabilities to questions such as *"Is X positive?"* or *"Does X belong to the interval $(0,1)$?"*.

If X is simply a function from Ω to $\mathbb{R}$, then to compute the probability that X is positive, we should define $A = \{\omega \in \Omega \mid X(\omega) > 0\}$, and then compute $\boldsymbol{P}(A)$. This, however, is only possible if $A \in \mathcal{G}$; and since $\mathcal{G}$ *is usually not* the entire power set of Ω, we should take care to ensure that all questions we might ask about the random variable X can be answered by only computing probabilities of events in $\mathcal{G}$, and not arbitrary subsets of Ω. For this reason, we define random variables as follows.

DEFINITION 2.1. A *random variable* is a *$\mathcal{G}$-measurable function* $X : \Omega \to \mathbb{R}$. That is, a random variable is a function $X : \Omega \to \mathbb{R}$ such that for every $\alpha \in \mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) \leqslant \alpha\}$ is guaranteed to be an element of $\mathcal{G}$. (Such functions are also called *$\mathcal{G}$-measurable*, *measurable with respect to $\mathcal{G}$*, or simply *measurable* if the σ-algebra in question is clear from the context.)

REMARK 2.2. The argument ω is *always* suppressed when writing random variables. That is, the event $\{\omega \in \Omega \mid X(\omega) \leqslant \alpha\}$ is simply written as $\{X \leqslant \alpha\}$.

REMARK 2.3. Note for any random variable, $\{X > \alpha\} = \{X \leqslant \alpha\}^c$ which must also belong to $\mathcal{G}$ since $\mathcal{G}$ is closed under complements. One can check that for every $\alpha < \beta \in \mathbb{R}$ the events $\{X < \alpha\}$, $\{X \geqslant \alpha\}$, $\{X > \alpha\}$, $\{X \in (\alpha,\beta)\}$, $\{X \in [\alpha,\beta)\}$, $\{X \in (\alpha,\beta]\}$ and $\{X \in (\alpha,\beta)\}$ are all also elements of $\mathcal{G}$.

Thus to (for instance) compute the chance that X lies strictly between two real numbers α and β, we consider the event $\{X \in (\alpha,\beta)\}$. By Remark 2.3 this is guaranteed to be an element of $\mathcal{G}$, and thus we can compute the probability of it using $\boldsymbol{P}$. Hence, the quantity $\boldsymbol{P}(\{X \in (\alpha,\beta)\})$ is mathematically well defined, and represents the chance that the random variable X takes values in the interval (α,β). For brevity, we almost always omit the outermost curly braces and write $\boldsymbol{P}(X \in (\alpha,\beta))$ for $\boldsymbol{P}(\{X \in (\alpha,\beta)\})$.

REMARK 2.4. One can check that if X, Y are random variables then so are $X \pm Y$, XY, X/Y (when defined), $|X|$, $X \wedge Y$ and $X \vee Y$. In fact if $f : \mathbb{R} \to \mathbb{R}$ is any reasonably nice (more precisely, a Borel measurable) function, $f(X)$ is also a random variable.

EXAMPLE 2.5. Given $A \subseteq \Omega$, define *indicator function of A* by

$$\mathbf{1}_A(\omega) \stackrel{\text{def}}{=} \begin{cases} 1 & \omega \in A\,, \\ 0 & \omega \notin A\,. \end{cases}$$

One can check that $\mathbf{1}_A$ is a ($\mathcal{G}$-measurable) random variable if and only if $A \in \mathcal{G}$.

EXAMPLE 2.6. For $M \in \mathbb{N}$, $i \in \{1, \dots, M\}$, $a_i \in \mathbb{R}$ and $A_i \in \mathcal{G}$ be such that $A_i \cap A_j = \emptyset$ for $i \neq j$, and define

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{M} a_i \mathbf{1}_{A_i}\,. \tag{2.1}$$

Then X is a ($\mathcal{G}$-measurable) random variable. (Such variables are called *simple* random variables.)

The next important concept concerning random variables is that of *expectation*, which we assume the reader is familiar with in the discrete setting. In the measure theoretic framework, the expectation of a random variable is the *Lebesgue integral*, and is denoted by[2]

$$\boldsymbol{E}X \stackrel{\text{def}}{=} \int_\Omega X\, d\boldsymbol{P}\,.$$

The precise construction of the Lebesgue integral, however, is to lengthy to be presented here, and we only present a brief summary.

If a random variable X only takes on finitely many values $a_1, \dots a_n$, then the expectation of X is given by

$$\boldsymbol{E}X \stackrel{\text{def}}{=} \sum_{i=1}^{n} a_i \boldsymbol{P}(X = a_i)\,. \tag{2.2}$$

This means that for any *simple random variable* of the form (2.1), the expectation is given by (2.2). For general random variables (i.e. random variables that are not simple), we can compute by expressing them as a limit of simple random variables. Namely, we can compute $\boldsymbol{E}X$ by

$$\boldsymbol{E}X = \lim_{n\to\infty} \boldsymbol{E}\Big(\sum_{k=-n^2}^{n^2-1} \frac{k}{n} \mathbf{1}_{\{\frac{k}{n} \leqslant X < \frac{k+1}{n}\}} \Big) = \lim_{n\to\infty} \sum_{k=-n^2}^{n^2-1} \frac{k}{n} \boldsymbol{P}\Big(\frac{k}{n} \leqslant X < \frac{k+1}{n} \Big)\,,$$

for instance.

The above description, however, is only of theoretical importance and is not used to compute in practice. Here are a few computation rules and properties of expectations that will be useful later.

[2] If $A \in \mathcal{G}$ we define

$$\int_A Y\, d\boldsymbol{P} \stackrel{\text{def}}{=} \boldsymbol{E}(\mathbf{1}_A Y)\,,$$

and when $A = \Omega$ we will often omit writing it.

(1) *(Linearity)* If $\alpha \in \mathbb{R}$ and X, Y are random variables, then $\boldsymbol{E}(X + \alpha Y) = \boldsymbol{E}X + \alpha \boldsymbol{E}Y$.
(2) *(Positivity)* If $X \geqslant 0$ almost surely,[3] then $\boldsymbol{E}X \geqslant 0$. Moreover, if $X > 0$ almost surely, $\boldsymbol{E}X > 0$. Consequently, (using linearity) if $X \leqslant Y$ almost surely then $\boldsymbol{E}X \leqslant \boldsymbol{E}Y$.
(3) *(Layer Cake formula)* If $X \geqslant 0$ almost surely, then

$$\boldsymbol{E}X = \int_0^\infty \boldsymbol{P}(X \geqslant t)\, dt\,.$$

More generally, if φ is a increasing differentiable function with $\varphi(0) = 0$ then

$$\boldsymbol{E}\varphi(X) = \int_0^\infty \varphi'(t)\, \boldsymbol{P}(X \geqslant t)\, dt\,.$$

(4) *(Unconscious Statistician Formula)* If the probability density function of X is p, and f is any (Borel measurable) function, then

$$\boldsymbol{E}f(X) = \int_{-\infty}^\infty f(x)p(x)\, dx\,. \tag{2.3}$$

The proof of these properties goes beyond the scope of these notes. We do, however, make a few remarks. It turns out that the proof of positivity in this framework is immediate, however the proof of linearity is surprisingly not as straightforward as you would expect. While it is easy to verify linearity for simple random variables, for general random variables, the proof of linearity requires an approximation argument. The full proof of this involves either the *dominated* or *monotone* convergence theorem which guarantee $\lim \boldsymbol{E}X_n = \boldsymbol{E} \lim X_n$, under modest assumptions.

The layer cake formula can be proved by drawing a graph of X with Ω on the horizontal axis. Now $\boldsymbol{E}X$ should be the "area under the curve", which is usually computed by slicing the region into vertical strips and adding up the area of each strip. If, instead, you compute the area by slicing the region into *horizontal* strips, you get exactly the layer cake formula!

Finally, unconscious statistician formula might already be familiar to you. In fact, the reason for this somewhat unusual name is that many people use this result "unconsciously" treating it as the definition, without realizing it is in fact a theorem that requires proof. To elaborate further, introductory (non-measure theoretic) probability courses usually stipulate that if a random variable X has density p_X, then

$$\boldsymbol{E}X = \int_{-\infty}^\infty x p_X(x)\, dx\,.$$

Thus if you set $Y = f(X)$ for some function f, we should have

$$\boldsymbol{E}Y = \int_{-\infty}^\infty y p_Y(y)\, dy\,.$$

If we could compute p_Y in terms of p_X and f, you could substitute it in the above formula, and obtain a formula for $\boldsymbol{E}Y$ in terms of p_X and f. Unfortunately, this isn't easy to do. Namely, if f isn't monotone, it isn't easy to write down p_Y in terms of p_X. It turns out, however, that even though we can't easily write down p_Y in terms of f and p_X, we can prove that $\boldsymbol{E}Y$ can be computed using (2.3).

[3] By $X \geqslant 0$ almost surely, we mean that $\boldsymbol{P}(X \geqslant 0) = 1$. More generally, we say an event occurs almost surely if the probability of it occurring is 1.

Since discussing these results and proofs further at this stage will will lead us too far astray, we invite the curious to look them up in any standard measure theory book. The main point of this section was to introduce you to a framework which is capable of describing and studying the objects we will need for the remainder of the course.

We conclude this section by revisiting the notion of a *continuous process* defined in the previous section. Recall, our definition so far was that a process is simply a collection of random variables $\{X(t)\}_{t \geqslant 0}$, and a continuous process is a process whose trajectories are continuous. In our context, a process can now be thought of as a function

$$X \colon \Omega \times [0, \infty) \to \mathbb{R}\,.$$

For every fixed t, the function $\omega \mapsto X(\omega, t)$ is required to be a random variable (i.e. measurable with respect to $\mathcal{G}$). Since the ω is usually suppressed in probability, this random variable is simply denoted by $X(t)$.

The trajectory of X is now the slice of X for a fixed ω. Namely, for any fixed $\omega \in \Omega$, the function $t \mapsto X(\omega, t)$ is the trajectory of X. Saying a process has continuous trajectories means that for every $\omega \in \Omega$, the trajectory $t \mapsto X(\omega, t)$ is continuous as a function of t. Explicitly, this means for every $t \geqslant 0$ and $\omega \in \Omega$ we have

$$\lim_{s \to t} X(\omega, s) = X(\omega, s)\,.$$

Following our convention of "never writing ω", this is exactly (1.3) as we had before.

3. Independence of random variables

Recall two events A, B are independent if $P(A \mid B) = P(A)$. This is of course immediately implies the multiplication law:

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)\,.$$

The notion of independence for random variables requires that *every* event that is observable from one is necessarily independent of *every* event that is observable from the other.

For example, suppose X and Y are two random variables. For any $a, b \in \mathbb{R}$, the event $\{X \in (a, b)\}$ can be observed using the random variable X. Similarly, any $c \in \mathbb{R}$, the event $\{Y > c\}$ can be observed using the random variable Y. If X and Y were independent, then the events $\{X \in (a, b)\}$ would necessarily be independent of the event $\{Y > c\}$. Of course, this is just an example and you can write down all sorts of other events (e.g. $X^2 - e^X < 15$, or $\sin(Y + 3) < .5$). No matter how you do it, if X and Y are independent, then any event observable from X alone must necessarily be independent of any event observable from Y alone.

Since the notion of "all events that can be observed from the random variable X" will be useful later, we denote it by $\sigma(X)$.

Definition 3.1. Let X be a random variable on $(\Omega, \mathcal{G}, \boldsymbol{P})$. We define *the σ-algebra generated by* X to be the σ algebra obtained by only using events that are observable using the random variable X.

One can mathematically prove that $\sigma(X)$ is generated by the events $\{X \leqslant \alpha\}$ for every $\alpha \in \mathbb{R}$. Namely, if a σ algebra contains the events $\{X \leqslant \alpha\}$ for every $\alpha \in \mathbb{R}$, then it must necessarily contain *all* events observable through the random variable X.

In particular, it will contain events of the form $\{X \in [\alpha, \beta)\}$, $e^{X+1} < \sin X$, or any other complicated formula that you can write down.

As mentioned above, the σ-algebra $\sigma(X)$ represents all the information one can obtain by observing X. To illustrate this, consider the following example: A card is drawn from a shuffled deck, and you win a dollar if it is red, and lose one if it is black. Now the likely hood of drawing any particular card is $1/52$. However, if you are blindfolded and only told the outcome of the game, you have no way to determine that each gard is picked with probability $1/52$. The only thing you will be able to determine is that red cards are drawn as often as black ones.

This is captured by σ-algebra as follows. Let $\Omega = \{1, \dots, 52\}$ represent a deck of cards, $\mathcal{G} = \mathcal{P}(\Omega)$, and define $\boldsymbol{P}(A) = \operatorname{card}(A)/52$. Let $R = \{1, \dots 26\}$ represent the red cards, and $B = R^c$ represent the black cards. The outcome of the above game is now the random variable $X = \mathbf{1}_R - \mathbf{1}_B$, and you should check that $\sigma(X)$ is exactly $\{\emptyset, R, B, \Omega\}$.

With this, we can now revisit the notion of two random variables being independent.

Definition 3.2. We say the random variables $X_1, \dots, X_N$ are independent if for every $i \in \{1 \dots N\}$ and every $A_i \in \sigma(X_i)$ the events $A_1, \dots, A_N$ are independent.

Remark 3.3. Recall, A collection of events $A_1, \dots, A_N$ is said to be independent if *any* sub collection $\{A_{i_1}, \dots, A_{i_k}\}$ satisfies the multiplication law

$$\boldsymbol{P}\Big(\bigcap_{i=1}^{k} A_{i_k}\Big) = \prod_{i=1}^{d} \boldsymbol{P}(A_i)\,.$$

Note that this is a *stronger* condition than simply requiring

$$\boldsymbol{P}\big(A_1 \cap A_2 \cap \dots \cap A_N\big) = \boldsymbol{P}(A_1)\,\boldsymbol{P}(A_2) \cdots \boldsymbol{P}(A_N)\,.$$

In practice, one never tests independence of random variables using the above multiplication law.

Proposition 3.4. *Let $X_1, \dots, X_N$ be N random variables. The following are equivalent:*

(1) The random variables $X_1, \dots, X_N$ are independent.

(2) For every $\alpha_1, \dots, \alpha_N \in \mathbb{R}$, we have

$$\boldsymbol{P}\Big(\bigcap_{j=1}^{N} \{X_j \leqslant \alpha_j\}\Big) = \prod_{j=1}^{N} \boldsymbol{P}(X_j \leqslant \alpha_j)$$

(3) For every collection of bounded continuous functions $f_1, \dots, f_N$ we have

$$\boldsymbol{E}\Big[\prod_{j=1}^{N} f_j(X_j)\Big] = \prod_{j=1}^{N} \boldsymbol{E} f_j(X_j)\,.$$

(4) For every $\xi_1, \dots, \xi_N \in \mathbb{R}$ we have

$$\boldsymbol{E}\exp\Big(i\sum_{j=1}^{N} \xi_j X_j\Big) = \prod_{j=1}^{N} \boldsymbol{E}\exp(i\xi_j X_j)\,, \qquad \text{where } i = \sqrt{-1}\,.$$

Remark 3.5. It is instructive to explicitly check each of these implications when $N = 2$ and X_1, X_2 are simple random variables.

REMARK 3.6. The intuition behind the above result is as follows: Since the events $\{X_j \leqslant \alpha_j\}$ generate $\sigma(X_j)$, we expect the first two properties to be equivalent. Since $\mathbf{1}_{(-\infty,\alpha_j]}$ can be well approximated by continuous functions, we expect equivalence of the second and third properties. The last property is a bit more subtle: Since $\exp(a+b) = \exp(a)\exp(b)$, the third clearly implies the last property. The converse holds because of "completeness of the complex exponentials" or Fourier inversion, and again a through discussion of this will lead us too far astray.

REMARK 3.7. The third implication above implies that independent random variables are uncorrelated. Namely, if X, Y are independent random variables, then

$$\boldsymbol{E}(XY) = (\boldsymbol{E}X)(\boldsymbol{E}Y)\,. \tag{3.1}$$

The converse, is of course false. Namely if (3.1) holds, there is no reason we should have

$$\boldsymbol{E}f(X)g(Y) = \boldsymbol{E}f(X)\boldsymbol{E}g(Y)\,,$$

for *every* bounded continuous pair of functions f, g as required by the third part in Proposition 3.4. However, if (X, Y) is *jointly normal* and X, Y are uncorrelated, then the *normal correlation theorem* guarantees that X, Y are independent.

REMARK 3.8. If moment generating functions of the random variables are defined in an interval around 0, then one can test independence using real exponentials instead of the complex exponentials used in the last condition in Proposition 3.4. Explicitly, in this case $X_1, \ldots, X_N$ are independent if and only if for every $t_1, \ldots, t_N$ in some small interval containing 0 we have

$$\boldsymbol{E}\exp\Big(\sum_{j=1}^N t_j X_j\Big) = \prod_{j=1}^N \boldsymbol{E}\exp(t_j X_j)\,.$$

EXAMPLE 3.9 (Covariance of Brownian motion). The independence of increments allows us to compute covariances of Brownian motion easily. Suppose W is a standard Brownian motion, and $s < t$. Then we know $W_s \sim N(0, s)$, $W_t - W_s \sim N(0, t-s)$ and is independent of W_s. Consequently $(W_s, W_t - W_s)$ is jointly normal with mean 0 and covariance matrix $\left(\begin{smallmatrix} s & 0 \\ 0 & t-s \end{smallmatrix}\right)$. This implies that (W_s, W_t) is a jointly normal random variable. Moreover we can compute the covariance by

$$\boldsymbol{E}W_s W_t = \boldsymbol{E}W_s(W_t - W_s) + \boldsymbol{E}W_s^2 = s\,.$$

In general if you don't assume $s < t$, the above immediately implies $\boldsymbol{E}W_s W_t = s \wedge t$.

4. Conditional probability

Our next goal is to understand *conditional probability*, and we do it directly here to help understanding. In the next section we will construct *conditional expectations* independently, and the reader may choose to skip this section.

Suppose you have an incomplete deck of cards which has 10 red cards, and 20 black cards. Suppose 5 of the red cards are *high cards* (i.e. ace, king, queen, jack or 10), and only 4 of the black cards are high. If a card is chosen at random, the *conditional probability* of it being high given that it is red is 1/2, and the *conditional probability* of it being high given that it is black is 1/5. Our aim is to encode both these facts into a single entity.

We do this as follows. Let R, B denote the set of all red and black cards respectively, and H denote the set of all high cards. A σ-algebra encompassing all the above information is exactly

$$\mathcal{G} \stackrel{\text{def}}{=} \{\emptyset, R, B, H, H^c, R \cap H, B \cap H, R \cap H^c, B \cap H^c, \\ (R \cap H) \cup (B \cap H^c), (R \cap H^c) \cup (B \cap H), \Omega\}$$

and you can explicitly compute the probabilities of each of the above events. A σ-algebra encompassing only the color of cards is exactly

$$\mathcal{G} \stackrel{\text{def}}{=} \{\emptyset, R, B, \Omega\}\,.$$

Now we define the *conditional probability* of a card being high given the color to be the **random variable**

$$\boldsymbol{P}(H \mid \mathcal{C}) \stackrel{\text{def}}{=} \boldsymbol{P}(H \mid R)\mathbf{1}_R + \boldsymbol{P}(H \mid B)\mathbf{1}_B = \frac{1}{2}\mathbf{1}_R + \frac{1}{5}\mathbf{1}_B\,.$$

To emphasize:

(1) What is given is the σ-algebra $\mathcal{C}$, and not just an event.
(2) The conditional probability is now a $\mathcal{C}$*-measurable random variable* and not a number.

To see how this relates to $\boldsymbol{P}(H \mid R)$ and $\boldsymbol{P}(H \mid B)$ we observe

$$\int_R \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} \stackrel{\text{def}}{=} \boldsymbol{E}\big(\mathbf{1}_R \boldsymbol{P}(H \mid \mathcal{C})\big) = \boldsymbol{P}(H \mid R)\, \boldsymbol{P}(R)\,.$$

The same calculation also works for B, and so we have

$$\boldsymbol{P}(H \mid R) = \frac{1}{\boldsymbol{P}(R)} \int_R \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} \quad \text{and} \quad \boldsymbol{P}(H \mid B) = \frac{1}{\boldsymbol{P}(B)} \int_B \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P}\,.$$

Our aim is now to generalize this to a non-discrete scenario. The problem with the above identities is that if either R or B had probability 0, then the above would become meaningless. However, clearing out denominators yields

$$\int_R \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} = \boldsymbol{P}(H \cap R) \quad \text{and} \quad \int_B \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} = \boldsymbol{P}(H \cap B)\,.$$

This suggests that the defining property of $\boldsymbol{P}(H \mid \mathcal{C})$ should be the identity

$$\int_C \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} = \boldsymbol{P}(H \cap C) \tag{4.1}$$

for every event $C \in \mathcal{C}$. Note $\mathcal{C} = \{\emptyset, R, B, \Omega\}$ and we have only checked (4.1) for $C = R$ and $C = B$. However, for $C = \emptyset$ and $C = \Omega$, (4.1) is immediate.

Definition 4.1. Let $(\Omega, \mathcal{G}, \boldsymbol{P})$ be a probability space, and $\mathcal{F} \subseteq \mathcal{G}$ be a σ-algebra. Given $A \in \mathcal{G}$, we define the conditional probability of A given $\mathcal{F}$, denoted by $\boldsymbol{P}(A|\mathcal{F})$ to be an $\mathcal{F}$-measurable random variable that satisfies

$$\int_F \boldsymbol{P}(H \mid \mathcal{F})\, d\boldsymbol{P} = \boldsymbol{P}(H \cap F) \quad \text{for every } F \in \mathcal{F}. \tag{4.2}$$

Remark 4.2. Showing existence (and uniqueness) of the conditional probability isn't easy, and relies on the *Radon-Nikodym theorem*, which is beyond the scope of this course.

REMARK 4.3. It is crucial to require that $\boldsymbol{P}(H \,|\, \mathcal{F})$ is measurable with respect to $\mathcal{F}$. Without this requirement we could simply choose $\boldsymbol{P}(H|\mathcal{F}) = \mathbf{1}_H$ and (4.2) would be satisfied. However, note that if $H \in \mathcal{F}$, then the function $\mathbf{1}_F$ is $\mathcal{F}$-measurable, and in this case $\boldsymbol{P}(H \mid \mathcal{F}) = \mathbf{1}_F$.

REMARK 4.4. In general we can only expect (4.2) to hold for all events in $\mathcal{F}$, and it need not hold for events in $\mathcal{G}$! Indeed, in the example above we see that

$$\int_H \boldsymbol{P}(H \mid \mathcal{C})\, d\boldsymbol{P} = \frac{1}{2}\boldsymbol{P}(R \cap H) + \frac{1}{5}\boldsymbol{P}(B \cap H) = \frac{1}{2} \cdot \frac{5}{30} + \frac{1}{5} \cdot \frac{4}{30} = \frac{11}{100}$$

but

$$\boldsymbol{P}(H \cap H) = \boldsymbol{P}(H) = \frac{3}{10} \neq \frac{11}{100}\,.$$

REMARK 4.5. One situation where you can compute $\boldsymbol{P}(A \,|\, \mathcal{F})$ explicitly is when $\mathcal{F} = \sigma(\{F_i\})$ where $\{F_i\}$ is a pairwise disjoint collection of events whose union is all of Ω and $\boldsymbol{P}(F_i) > 0$ for all i. In this case

$$\boldsymbol{P}(A \mid \mathcal{F}) = \sum_i \frac{\boldsymbol{P}(A \cap F_i)}{\boldsymbol{P}(F_i)} \mathbf{1}_{F_i}\,.$$

5. Conditional expectation.

Conditional expectation arises when you have a random variable X, and want to *best approximate* it using only a (strict) subset of events. Precisely, suppose $\mathcal{F} \subseteq \mathcal{G}$ is a *σ-sub-algebra* of $\mathcal{G}$. That is, $\mathcal{F}$ is a σ-algebra, and every event in $\mathcal{F}$ is also an event in $\mathcal{G}$. Now to best approximate a ($\mathcal{G}$-measurable) random variable X using only events in $\mathcal{F}$, one would like to find an $\mathcal{F}$ measurable random variable Z that minimizes

$$\boldsymbol{E}|X - Z|^2\,.$$

The minimizer is known as the *conditional expectation of X given $\mathcal{F}$*, and denoted by $\boldsymbol{E}(X \mid \mathcal{F})$. That is,

$$\boldsymbol{E}(X \mid \mathcal{F}) \stackrel{\text{def}}{=} \arg\min\bigl\{\boldsymbol{E}|X - Z|^2 \bigm| Z \text{ is a } \mathcal{G}\text{-measurable random variable}\bigr\}\,. \tag{5.1}$$

While the above provides good intuition to the notion of conditional expectation, it is not as convenient to work with mathematically. For instance, the above requires $\boldsymbol{E}X^2 < \infty$, and we will often require conditional expectations of random variables that do not have this property.

To motivate the other definition of conditional expectation, we use the following example. Consider an incomplete deck of cards which has 10 red cards, of which 5 are high, and 20 black cards, of which 4 are high. Let X be the outcome of a game played through a dealer who pays you \$1 when a high card is drawn, and charges you \$1 otherwise. However, you are standing too far away from the dealer to tell whether the card drawn was high or not. You can only tell *the color*, and *whether or not you won.*

After playing this game often the only information you can deduce is that your expected return is 0 when a red card is drawn and $-3/5$ when a black card is drawn. That is, you approximate the game outcome X by the random variable

$$Y \stackrel{\text{def}}{=} 0\mathbf{1}_R - \frac{3}{5}\mathbf{1}_B\,,$$

where, as before R, B denote the set of all red and black cards respectively.

Note that the events you can deduce information about by playing this game (through the dealer) are exactly elements of the σ-algebra $\mathcal{C} = \{\emptyset, R, B, \Omega\}$. By construction, that your approximation Y is $\mathcal{C}$-measurable, and it is easy to verify that

$$Y = \arg\min\{\boldsymbol{E}(X - Z)^2 \mid Z \text{ is a } \mathcal{C}\text{-measurable random variable}\}\,. \tag{5.2}$$

That is $Y = \boldsymbol{E}(X \mid C)$ according to the definition (5.1). In this case, we can also verify that Y has the same averages as X on all elements of $\mathcal{C}$. That is, for every $C \in \mathcal{C}$, we have[4]

$$\int_C Y\,d\boldsymbol{P} = \int_C X\,d\boldsymbol{P}\,. \tag{5.3}$$

It turns out that in general, one can show abstractly that any $\mathcal{C}$ measurable random variable that satisfies (5.3), must in fact also be the minimizer in (5.2). We will thus use (5.3) to define conditional expectation.

Definition 5.1. Let X be a $\mathcal{G}$-measurable random variable, and $\mathcal{F} \subseteq \mathcal{G}$ be a σ-sub-algebra. We define $\boldsymbol{E}(X \mid \mathcal{F})$, the *conditional expectation of X given $\mathcal{F}$* to be a *random variable* such that:

(1) $\boldsymbol{E}(X \mid \mathcal{F})$ is $\mathcal{F}$-measurable.
(2) For every $F \in \mathcal{F}$, we have the *partial averaging* identity:

$$\int_F \boldsymbol{E}(X \mid \mathcal{F})\,d\boldsymbol{P} = \int_F X\,d\boldsymbol{P}. \tag{5.4}$$

Remark 5.2. We can only expect (5.4) to hold for all events $F \in \mathcal{F}$. In general (5.4) *will not* hold for events $G \in \mathcal{G} - \mathcal{F}$.

Remark 5.3. An equivalent way of phrasing (5.4) is to require

$$\boldsymbol{E}(XY) = \boldsymbol{E}\big(\boldsymbol{E}(X \bigm| \mathcal{F})Y\big) \tag{5.5}$$

for every $\mathcal{F}$ measurable random variable Y. As before, we can only expect (5.5) to hold when Y is $\mathcal{F}$ measurable. In general (5.5) *will not* hold when Y is not $\mathcal{F}$ measurable.

Remark 5.4. Choosing $F = \Omega$ we see $\boldsymbol{E}\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$.

Remark 5.5. More concretely, suppose Y is another random variable and $\mathcal{F} = \sigma(Y)$. Then it turns out that one can find a special (non-random) function g such that $\boldsymbol{E}(X \mid \mathcal{F}) = g(Y)$. Moreover, the function g is characterized by the property that

$$\boldsymbol{E}\big(f(Y)X\big) = \boldsymbol{E}\big(f(Y)g(Y)\big)\,.$$

for *every* bounded continuous function f.

Remark 5.6. Under mild integrability assumptions one can show that conditional expectations exist. This requires the *Radon-Nikodym* theorem and goes beyond the scope of this course. If, however, $\mathcal{F} = \sigma(\{F_i\})$ where $\{F_i\}$ is a pairwise disjoint collection of events whose union is all of Ω and $\boldsymbol{P}(F_i) > 0$ for all i, then

$$\boldsymbol{E}(X \mid \mathcal{F}) = \sum_{i=1}^{\infty} \frac{\boldsymbol{1}_{F_i}}{\boldsymbol{P}(F_i)} \int_{F_i} X\,d\boldsymbol{P}\,.$$

[4] Recall $\int_C Y\,d\boldsymbol{P}$ is simply $\boldsymbol{E}(\boldsymbol{1}_C Y)$. That is $\int_C Y\,d\boldsymbol{P}$ is the expectation of the random variable which is Y on the event C, and 0 otherwise.

REMARK 5.7. Once existence is established it is easy to see that conditional expectations are unique. Namely, if Y is any $\mathcal{F}$-measurable random variable that satisfies

$$\int_F Y\,d\boldsymbol{P} = \int_F X\,d\boldsymbol{P} \quad \text{for every } F \in \mathcal{F}\,,$$

then $Y = \boldsymbol{E}(X \mid F)$. Often, when computing the conditional expectation, we will "guess" what it is, and verify our guess by checking measurablity and the above partial averaging identity.

PROPOSITION 5.8. *If X is $\mathcal{F}$-measurable, then $\boldsymbol{E}(X \mid \mathcal{F}) = X$. On the other hand, if X is independent*[5] *of $\mathcal{F}$ then $\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$.*

PROOF. If X is $\mathcal{F}$-measurable, then clearly the random variable X is both $\mathcal{F}$-measurable and satisfies the partial averaging identity. Thus by uniqueness, we must have $\boldsymbol{E}(X \mid \mathcal{F}) = X$.

Now consider the case when X is independent of $\mathcal{F}$. Suppose first $X = \sum a_i \mathbf{1}_{A_i}$ for finitely many sets $A_i \in \mathcal{G}$. Then for any $F \in \mathcal{F}$,

$$\int_F X\,d\boldsymbol{P} = \sum a_i \boldsymbol{P}(A_i \cap F) = \boldsymbol{P}(F) \sum a_i \boldsymbol{P}(A_i) = \boldsymbol{P}(F)\boldsymbol{E}X = \int_F \boldsymbol{E}X\,d\boldsymbol{P}\,.$$

Thus the constant random variable $\boldsymbol{E}X$ is clearly $\mathcal{F}$-measurable and satisfies the partial averaging identity. This forces $\boldsymbol{E}(X \mid \mathcal{F}) = \boldsymbol{E}X$. The general case when X is not simple follows by approximation. □

The above fact has a generalization that is tremendously useful when computing conditional expectations. Intuitively, the general principle is to *average* quantities that are independent of $\mathcal{F}$, and *leave unchanged* quantities that are $\mathcal{F}$ measurable. This is known as the independence lemma.

LEMMA 5.9 (Independence Lemma). *Suppose X, Y are two random variables such that X is independent of $\mathcal{F}$ and Y is $\mathcal{F}$-measurable. Then if $f = f(x, y)$ is any function of two variables we have*

$$\boldsymbol{E}\big(f(X, Y) \bigm| \mathcal{F}\big) = g(Y)\,,$$

where $g = g(y)$ is the function[6] *defined by*

$$g(y) \stackrel{\text{def}}{=} \boldsymbol{E}f(X, y)\,.$$

REMARK. If p_X is the probability density function of X, then the above says

$$\boldsymbol{E}\big(f(X, Y) \bigm| \mathcal{F}\big) = \int_{\mathbb{R}} f(x, Y)\, p_X(x)\,dx\,.$$

Indicating the ω dependence explicitly for clarity, the above says

$$\boldsymbol{E}\big(f(X, Y) \bigm| \mathcal{F}\big)(\omega) = \int_{\mathbb{R}} f(x, Y(\omega))\, p_X(x)\,dx\,.$$

[5] We say a random variable X is independent of σ-algebra $\mathcal{F}$ if for every $A \in \sigma(X)$ and $B \in \mathcal{F}$ the events A and B are independent.

[6] To clarify, we are defining a *non-random* function $g = g(y)$ here when $y \in \mathbb{R}$ is any real number. Then, once we compute g, we substitute in $y = Y (= Y(\omega))$, where Y is the given random variable.

REMARK 5.10. Note we defined and motivated conditional expectations and conditional probabilities independently. They are however intrinsically related: Indeed, $\boldsymbol{E}(\mathbf{1}_A \mid \mathcal{F}) = \boldsymbol{P}(A \mid \mathcal{F})$, and this can be checked directly from the definition.

As we will see shortly, computing conditional expectations will be a very important part of pricing securities. Most of the time, all that is required to compute conditional expectations are the following properties.

PROPOSITION 5.11. *Conditional expectations satisfy the following properties.*

(1) (Linearity) *If X, Y are random variables, and $\alpha \in \mathbb{R}$ then*

$$\boldsymbol{E}(X + \alpha Y \mid \mathcal{F}) = \boldsymbol{E}(X \mid \mathcal{F}) + \alpha \boldsymbol{E}(Y \mid \mathcal{F}).$$

(2) (Positivity) *If $X \leqslant Y$, then $\boldsymbol{E}(X \mid \mathcal{F}) \leqslant \boldsymbol{E}(Y \mid \mathcal{F})$ (almost surely).*

(3) If X is $\mathcal{F}$ measurable and Y is an arbitrary (not necessarily $\mathcal{F}$-measurable) random variable then (almost surely)

$$\boldsymbol{E}(XY \mid \mathcal{F}) = X \boldsymbol{E}(Y \mid \mathcal{F}).$$

(4) (Tower property) *If $\mathcal{E} \subseteq \mathcal{F} \subseteq \mathcal{G}$ are σ-algebras, then (almost surely)*

$$\boldsymbol{E}(X \mid \mathcal{E}) = \boldsymbol{E}\Big(\boldsymbol{E}(X \mid \mathcal{F}) \Bigm| \mathcal{E}\Big).$$

PROOF. The first property follows immediately from linearity. For the second property, set $Z = Y - X$ and observe

$$\int_{\boldsymbol{E}(Z|\mathcal{F})} \boldsymbol{E}(Z \mid \mathcal{F})\, d\boldsymbol{P} = \int_{\boldsymbol{E}(Z|\mathcal{F})} Z\, d\boldsymbol{P} \geqslant 0,$$

which can only happen if $\boldsymbol{P}(\boldsymbol{E}(Z \mid \mathcal{F}) < 0) = 0$. The third property is easily checked for simple random variables, and follows in general by approximating. The tower property follows immediately from the definition. ☐

As an illustration, of how the above properties come in handy, we show how they can be used to deduce (5.5).

PROPOSITION 5.12. *If $\mathcal{F} \subseteq G$ is a σ-sub-algebra, X is a $\mathcal{G}$-measurable random variable, and Y is an $\mathcal{F}$-measurable random variable, then*

$$\boldsymbol{E}(XY) = \boldsymbol{E}\big(\boldsymbol{E}(X \bigm| \mathcal{F})Y\big)$$

PROOF. Using Remark 5.4 and then Proposition 5.11 part (3), we see

$$\boldsymbol{E}(XY) = \boldsymbol{E}\big(\boldsymbol{E}(XY \mid \mathcal{F})\big) = \boldsymbol{E}\big(Y \boldsymbol{E}(X \mid \mathcal{F})\big),$$

as desired. ☐

Finally we conclude this section by showing that the conditional expectation of a random variable according to Definition 5.1 is precisely the minimizer as in (5.1).

PROPOSITION 5.13. *Let X be a square integrable $\mathcal{G}$-measurable random variable, and $\mathcal{F} \subseteq \mathcal{G}$ be a σ-sub-algebra of $\mathcal{G}$. Then amongst all $\mathcal{F}$-measurable random variables Z, the one that minimizes*

$$\boldsymbol{E}(X - Z)^2$$

is precisely $Z = \boldsymbol{E}(X \mid \mathcal{F})$.

PROOF. Since $\boldsymbol{E}(X \mid \mathcal{F})$ is known to be an $\mathcal{F}$-measurable random variable, we only need to show that for any (other) $\mathcal{F}$-measurable random variable $\mathbb{Z}$ we have

$$\boldsymbol{E}(X - Z)^2 \geqslant \boldsymbol{E}\big((X - \boldsymbol{E}(X \mid \mathcal{F}))^2\big).$$

To see this, note

$$\begin{aligned}
\boldsymbol{E}(X - Z)^2 &= \boldsymbol{E}(X - \boldsymbol{E}(X \mid \mathcal{F}) + \boldsymbol{E}(X \mid \mathcal{F}) - Z)^2 \\
&= \boldsymbol{E}(X - \boldsymbol{E}(X \mid \mathcal{F}))^2 + \boldsymbol{E}((X \mid \mathcal{F}) - Z)^2 \\
&\quad + 2\boldsymbol{E}\Big(\underbrace{(X - \boldsymbol{E}(X \mid \mathcal{F}))}_{I}\underbrace{(\boldsymbol{E}(X \mid \mathcal{F}) - Z)}_{II}\Big)
\end{aligned}$$

Since term II is $\mathcal{F}$ measurable, we can use (5.5) to replace X with $\boldsymbol{E}(X \mid F)$ in term I. This yields

$$\begin{aligned}
\boldsymbol{E}(X - Z)^2 &= \boldsymbol{E}(X - \boldsymbol{E}(X \mid \mathcal{F}))^2 + \boldsymbol{E}((X \mid \mathcal{F}) - Z)^2 \\
&\quad + 2\boldsymbol{E}\Big((\boldsymbol{E}(X \mid \mathcal{F}) - \boldsymbol{E}(X \mid \mathcal{F}))(\boldsymbol{E}(X \mid \mathcal{F}) - Z)\Big) \\
&= \boldsymbol{E}(X - \boldsymbol{E}(X \mid \mathcal{F}))^2 + \boldsymbol{E}((X \mid \mathcal{F}) - Z)^2 \geqslant \boldsymbol{E}(X - \boldsymbol{E}(X \mid \mathcal{F}))^2,
\end{aligned}$$

as desired. □

6. The Martingale Property

A martingale is "fair game". Suppose you are playing a game and $M(t)$ is your cash stockpile at time t. As time progresses, you learn more and more information about the game. For instance, in blackjack getting a high card benefits the player more than the dealer, and a common card counting strategy is to have a "spotter" betting the minimum while counting the high cards. When the odds of getting a high card are favorable enough, the player will signal a "big player" who joins the table and makes large bets, as long as the high card count is favorable. Variants of this strategy have been shown to give the player up to a 2% edge over the house.

If a game is a martingale, then this extra information you have acquired *can not* help you going forward. That is, if you signal your "big player" at any point, you will not affect your expected return.

Mathematically this translates to saying that the *conditional expectation* of your stockpile at a later time given your present accumulated knowledge, is exactly the present value of your stockpile. Our aim in this section is to make this precise.

6.1. Adapted processes and filtrations. Let X be any stochastic process (for example Brownian motion). For any $t > 0$, we've seen before that $\sigma(X(t))$ represents the information you obtain by observing $X(t)$. Accumulating this over time gives us the *filtration*. To introduce this concept, we first need the notion of a σ algebra generated by a family of sets.

DEFINITION 6.1. Given a collection of sets A_α, where α belongs to some (possibly infinite) index set $\mathcal{A}$, we define $\sigma(\{A_\alpha\})$ to be the *smallest* σ-algebra that contains each of the sets A_α.

That is, if $\mathcal{G} = \sigma(\{A_\alpha\})$, then we must have each $A_\alpha \in \mathcal{G}$. Since $\mathcal{G}$ is a σ-algebra, all sets you can obtain from these by taking complements, countable unions and countable intersections intersections must also belong to $\mathcal{G}$.[7] The fact that $\mathcal{G}$ is the smallest σ-algebra containing each A_α also means that if $\mathcal{G}'$ is any other σ-algebra that contains each A_α, then $\mathcal{G} \subseteq \mathcal{G}'$.

REMARK 6.2. The smallest σ-algebra under which X is a random variable (under which X is measurable) is exactly $\sigma(X)$. It turns out that $\sigma(X) = X^{-1}(\mathcal{B}) = \{X \in B \mid B \in \mathcal{B}\}$, where $\mathcal{B}$ is the *Borel* σ-algebra on $\mathbb{R}$. Here $\mathcal{B}$ is the *Borel* σ-algebra, defined to be the σ-algebra on $\mathbb{R}$ generated by all open intervals.

DEFINITION 6.3. Given a stochastic process X, the *filtration generated by* X is the family of σ-algebras $\{\mathcal{F}_t^X \mid t \geqslant 0\}$ where

$$\mathcal{F}_t^X \stackrel{\text{def}}{=} \sigma\Big(\bigcup_{s \leqslant t} \sigma(X_s)\Big).$$

That is, $\mathcal{F}_t^X$ is all events that can be observed using only the random variables X_s when $s \leqslant t$. Clearly each $\mathcal{F}_t^X$ is a σ-algebra, and if $s \leqslant t$, $\mathcal{F}_s^X \subseteq \mathcal{F}_t^X$. A family of σ-algebras with this property is called a *filtration*.

DEFINITION 6.4. A *filtration* is a family of σ-algebras $\{\mathcal{F}_t \mid t \geqslant 0\}$ such that whenever $s \leqslant t$, we have $\mathcal{F}_s \subseteq \mathcal{F}_t$.

In our case, the filtration we work with will most often be the *Brownian filtration*, i.e. the filtration generated by Brownian motion. However, one can (and often needs to) consider more general filtrations. In this case the intuition we use is that the σ-algebra $\mathcal{F}_t$ represents the information accumulated up to time t (i.e. all events whose probabilities can be deduced up to time t). When given a filtration, it is important that all stochastic processes we construct respect the flow of information, and do not look into the future. This is of course natural: trading / pricing strategies can not rely on the price at a later time, and gambling strategies do not know the outcome of the next hand. Mathematically this property is called *adapted*, and is defined as follows.

DEFINITION 6.5. A stochastic process X is said to be *adapted* to a filtration $\{\mathcal{F}_t \mid t \geqslant 0\}$ if for every t the random variable $X(t)$ is $\mathcal{F}_t$ measurable (i.e. $\{X(t) \leqslant \alpha\} \in \mathcal{F}_t$ for every $\alpha \in \mathbb{R}$, $t \geqslant 0$).

Clearly a process X is adapted with respect to the filtration it generates $\{\mathcal{F}_t^X\}$.

6.2. Martingales. Recall, a martingale is a "fair game". Using conditional expectations, we can now define this precisely.

DEFINITION 6.6. A stochastic process M is a martingale with respect to a filtration $\{\mathcal{F}_t\}$ if:

(1) M is adapted to the filtration $\{\mathcal{F}_t\}$.
(2) For any $s < t$ we have $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) = M(s)$, almost surely.

---

[7] Usually $\mathcal{G}$ contains *much more* than all countable unions, intersections and complements of the A_α's. You might think you could keep including all sets you generate using countable unions and complements and arrive at all of $\mathcal{G}$. It turns out that to make this work, you will usually have to do this *uncountably* many times!

This won't be too important within the scope of these notes. However, if you read a rigorous treatment and find the authors using some fancy trick (using Dynkin systems or monotone classes) instead of a naive countable unions argument, then the above is the reason why.

REMARK 6.7. A *sub-martingale* is an adapted process M for which we have $\boldsymbol{E}(M(t)\,|\,\mathcal{F}_s) \geqslant M(s)$, and a *super-martingale* if $\boldsymbol{E}(M(t)\,|\,\mathcal{F}_s) \leqslant M(s)$. Thus $\boldsymbol{E}M(t)$ is an increasing function of time if M is a sub-martingale, constant in time if M is a martingale, and a decreasing function of time if M is a super-martingale.

REMARK 6.8. It is crucial to specify the filtration when talking about martingales, as it is certainly possible that a process is a martingale with respect to one filtration but not with respected to another. For our purposes the filtration will almost always be the *Brownian filtration* (i.e. the filtration generated by Brownian motion).

EXAMPLE 6.9. Let $\{\mathcal{F}_t\}$ be a filtration, $\mathcal{F}_\infty = \sigma(\cup_{t\geqslant 0}\mathcal{F}_t)$, and X be any $\mathcal{F}_\infty$-measurable random variable. The process $M(t) \stackrel{\text{def}}{=} \boldsymbol{E}(X_\infty \mid \mathcal{F}_t)$ is a martingale with respect to the filtration $\{\mathcal{F}_t\}$.

6.3. The martingale property of Brownian motion. In discrete time a random walk is a martingale, so it is natural to expect that in continuous time Brownian motion is a martingale as well.

THEOREM 6.10. *Let W be a Brownian motion, $\mathcal{F}_t = \mathcal{F}_t^W$ be the Brownian filtration. Brownian motion is a martingale with respect to this filtration.*

PROOF. By independence of increments, $W(t) - W(s)$ is certainly independent of $W(r)$ for any $r \leqslant s$. Since $\mathcal{F}_s = \sigma(\cup_{r\leqslant s}\sigma(W(r)))$ we expect that $W(t) - W(s)$ is independent of $\mathcal{F}_s$. Consequently

$$\boldsymbol{E}(W(t) \mid \mathcal{F}_s) = \boldsymbol{E}(W(t) - W(s) \mid \mathcal{F}_s) + \boldsymbol{E}(W(s) \mid \mathcal{F}_s) = 0 + W(s) = W(s)\,. \qquad \square$$

THEOREM 6.11. *Let W be a standard Brownian motion (i.e. a Brownian motion normalized so that $W(0) = 0$ and $\operatorname{Var}(W(t)) = t$). For any $C_b^{1,2}$ function[8] $f = f(t,x)$ the process*

$$M(t) \stackrel{\text{def}}{=} f(t, W(t)) - \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\Big)\,ds$$

is a martingale (with respect to the Brownian filtration).

PROOF. This is an extremely useful fact about Brownian motion follows quickly from the Itô formula, which we will discuss later. However, at this stage, we can provide a simple, elegant and instructive proof as follows.

Adaptedness of M is easily checked. To compute $\boldsymbol{E}(M(t) \mid \mathcal{F}_r)$ we first observe

$$\boldsymbol{E}\big(f(t, W(t)) \,\big|\, \mathcal{F}_r\big) = \boldsymbol{E}\big(f(t, [W(t) - W(r)] + W(r)) \,\big|\, \mathcal{F}_r\big).$$

Since $W(t) - W(r) \sim N(0, t-r)$ and is independent of $\mathcal{F}_r$, the above conditional expectation can be computed by

$$\boldsymbol{E}\big(f(t, [W(t) - W(r)] + W(r)) \,\big|\, \mathcal{F}_r\big) = \int_{\mathbb{R}} f(t, y + W(r))G(t-r, y)\,dy\,,$$

where

$$G(\tau, y) = \frac{1}{\sqrt{2\pi\tau}}\exp\Big(\frac{-y^2}{2\tau}\Big)$$

[8] Recall a function $f = f(t,x)$ is said to be $C^{1,2}$ if it is C^1 in t (i.e. differentiable with respect to t and $\partial_t f$ is continuous), and C^2 in x (i.e. twice differentiable with respect to x and $\partial_x f$, $\partial_x^2 f$ are both continuous). The space $C_b^{1,2}$ refers to all $C^{1,2}$ functions f for which and f, $\partial_t f$, $\partial_x f$, $\partial_x^2 f$ are all bounded functions.

is the density of $W(t) - W(r)$.

Similarly

$$\begin{aligned}\boldsymbol{E}\Big(\int_0^t \big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\big)\, ds \,\Big|\, \mathcal{F}_r\Big) \\ = \int_0^r \big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\big)\, ds \\ + \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + W(r)) + \frac{1}{2}\partial_x^2 f(s, y + W(r))\big) G(s - r, y)\, dy\, ds\end{aligned}$$

Hence

$$\begin{aligned}\boldsymbol{E}(M(t) \mid \mathcal{F}_r) - M(r) = \int_{\mathbb{R}} f(t, y + W(r)) G(t - r, y)\, dy \\ - \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + W(r)) + \frac{1}{2}\partial_x^2 f(s, y + W(r))\big) G(s - r, y)\, ds \\ - f(r, W(r))\,.\end{aligned}$$

We claim that the right hand side of the above vanishes. In fact, we claim the (deterministic) identity

$$\begin{aligned}f(r, x) = \int_{\mathbb{R}} f(t, y + x) G(t - r, y)\, dy \\ - \int_r^t \int_{\mathbb{R}} \big(\partial_t f(s, y + x) + \frac{1}{2}\partial_x^2 f(s, y + x))\big) G(s - r, y)\, dy\, ds\end{aligned}$$

holds for any function f and $x \in \mathbb{R}$. For those readers who are familiar with PDEs, this is simply the Duhamel's principle for the heat equation. If you're unfamiliar with this, the above identity can be easily checked using the fact that $\partial_\tau G = \frac{1}{2}\partial_y^2 G$ and integrating the first integral by parts. We leave this calculation to the reader. □

6.4. Stopping Times. For this section we assume that a filtration $\{\mathcal{F}_t\}$ is given to us, and fixed. When we refer to process being adapted (or martingales), we implicitly mean they are adapted (or martingales) with respect to this filtration.

Consider a game (played in continuous time) where you have the option to walk away at any time. Let τ be the *random* time you decide to stop playing and walk away. In order to respect the flow of information, you need to be able to decide weather you have stopped using only information up to the present. At time t, event $\{\tau \leqslant t\}$ is exactly when you have stopped and walked away. Thus, to respect the flow of information, we need to ensure $\{\tau \leqslant t\} \in \mathcal{F}_t$.

DEFINITION 6.12. A stopping time is a function $\tau \colon \Omega \to [0, \infty)$ such that for every $t \geqslant 0$ the event $\{\tau \leqslant t\} \in \mathcal{F}_t$.

A standard example of a stopping time is *hitting times*. Say you decide to liquidate your position once the value of your portfolio reaches a certain threshold. The time at which you liquidate is a hitting time, and under mild assumptions on the filtration, will always be a stopping time.

PROPOSITION 6.13. *Let X be an adapted continuous process, $\alpha \in \mathbb{R}$ and τ be the first time X hits α (i.e. $\tau = \inf\{t \geqslant 0 \mid X(t) = \alpha\}$). Then τ is a stopping time (if the filtration is right continuous).*

THEOREM 6.14 (Doob's optional sampling theorem). *If M is a martingale and τ is a bounded stopping time. Then the stopped process $M^\tau(t) \stackrel{\text{def}}{=} M(\tau \wedge t)$ is also a martingale. Consequently, $\boldsymbol{E}M(\tau) = \boldsymbol{E}M(\tau \wedge t) = \boldsymbol{E}M(0) = \boldsymbol{E}M(t)$ for all $t \geqslant 0$.*

REMARK 6.15. If instead of assuming τ is bounded, we assume M^τ is bounded the above result is still true.

The proof goes beyond the scope of these notes, and can be found in any standard reference. What this means is that if you're playing a fair game, then you can not hope to improve your odds by "quitting when you're ahead". Any rule by which you decide to stop, must be a stopping time and the above result guarantees that stopping a martingale still yields a martingale.

REMARK 6.16. Let W be a standard Brownian motion, τ be the first hitting time of W to 1. Then $\boldsymbol{E}W(\tau) = 1 \neq 0 = \boldsymbol{E}W(t)$. This is one situation where the optional sampling theorem doesn't apply (in fact, $\boldsymbol{E}\tau = \infty$, and W^τ is unbounded).

This example corresponds to the gambling strategy of walking away when you make your "million". The reason it's not a sure bet is because the time taken to achieve your winnings is finite almost surely, but very long (since $\boldsymbol{E}\tau = \infty$). In the mean time you might have incurred financial ruin and expended your entire fortune.

Suppose the price of a security you're invested in fluctuates like a martingale (say for instance Brownian motion). This is of course unrealistic, since Brownian motion can also become negative; but lets use this as a first example. You decide you're going to liquidate your position and walk away when either you're bankrupt, or you make your first million. What are your expected winnings? This can be computed using the optional sampling theorem.

PROBLEM 6.1. Let $a \geqslant 0$ and M be any continuous martingale with $M(0) = x \in (0, a)$. Let τ be the first time M hits either 0 or a. Compute $\boldsymbol{P}(M(\tau) = a)$ and your expected return $\boldsymbol{E}M(\tau)$.

CHAPTER 3

Stochastic Integration

1. Motivation

Suppose $\Delta(t)$ is your position at time t on a security whose price is $S(t)$. If you only trade this security at times $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, then the change in the value of your wealth up to time T is given by

$$X(t_n) - X(0) = \sum_{i=0}^{n-1} \Delta(t_i)(S(t_{i+1}) - S(t_i))$$

If you are trading this continuously in time, you'd expect that a "simple" limiting procedure should show that your wealth is given by the *Riemann-Stieltjes* integral:

$$X(T) - X(0) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta(t_i)(S(t_{i+1}) - S(t_i)) = \int_0^T \Delta(t)\, dS(t)\,.$$

Here $P = \{0 = t_0 < \cdots < t_n = T\}$ is a partition of $[0, T]$, and $\|P\| = \max\{t_{i+1} - t_i\}$.

This has been well studied by mathematicians, and it is well known that for the above limiting procedure to "work directly", you need S to have finite *first variation.* Recall, the *first variation* of a function is defined to be

$$V_{[0,T]}(S) \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} |S(t_{i+1}) - S(t_i)|\,.$$

It turns out that almost *any* continuous martingale S *will not* have finite first variation. Thus to define integrals with respect to martingales, one has to do something 'clever'. It turns out that if X is adapted and S is an martingale, then the above limiting procedure works, and this was carried out by Itô (and independently by Doeblin).

2. The First Variation of Brownian motion

We begin by showing that the first variation of Brownian motion is infinite.

PROPOSITION 2.1. *If W is a standard Brownian motion, and $T > 0$ then*

$$\lim_{n\to\infty} \boldsymbol{E} \sum_{k=0}^{n-1} \Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \infty\,.$$

REMARK 2.2. In fact

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \infty \quad \text{almost surely,}$$

but this won't be necessary for our purposes.

PROOF. Since $W((k+1)/n) - W(k/n) \sim N(0, 1/n)$ we know

$$\boldsymbol{E}\Big|W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big)\Big| = \int_{\mathbb{R}} |x|\, G\Big(\frac{1}{n}, x\Big)\, dx = \frac{C}{\sqrt{n}}\,,$$

where

$$C = \int_{\mathbb{R}} |y| e^{-y^2/2} \frac{dy}{\sqrt{2\pi}} = \boldsymbol{E}|N(0,1)|\,.$$

Consequently

$$\sum_{k=0}^{n-1} \boldsymbol{E}\Big|W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big)\Big| = \frac{Cn}{\sqrt{n}} \xrightarrow{n\to\infty} \infty\,. \qquad \square$$

3. Quadratic Variation

It turns out that the *second* variation of any *square integrable* martingale is almost surely finite, and this is the key step in constructing the Itô integral.

DEFINITION 3.1. Let M be any process. We define the *quadratic variation* of M, denoted by $[M, M]$ by

$$[M, M](T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} (\Delta_i M)^2\,,$$

where $P = \{0 = t_1 < t_1 \cdots < t_n = T\}$ is a partition of $[0, T]$, and $\Delta_i M = M(t_{i+1}) - M(t_i)$.

PROPOSITION 3.2. *If W is a standard Brownian motion, then $[W, W](T) = T$ almost surely.*

PROOF. For simplicity, let's assume $t_i = Ti/n$. Note

$$\sum_{i=0}^{n-1} (\Delta_i W)^2 - T = \sum_{i=0}^{n-1} \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big)\Big)^2 - T = \sum_{i=0}^{n-1} \xi_i\,,$$

where

$$\xi_i \stackrel{\text{def}}{=} (\Delta_i W)^2 - \frac{T}{n} = \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big)\Big)^2 - \frac{T}{n}\,.$$

Note that ξ_i's are i.i.d. with distribution $N(0, T/n)^2 - T/n$, and hence

$$\boldsymbol{E}\xi_i = 0 \qquad \text{and} \qquad \operatorname{Var}\xi_i = \frac{T^2(\boldsymbol{E}N(0,1)^4 - 1)}{n^2}.$$

Consequently

$$\operatorname{Var}\Big(\sum_{i=0}^{n-1} \xi_i\Big) = \frac{T^2(\boldsymbol{E}N(0,1)^4 - 1)}{n} \xrightarrow{n\to\infty} 0\,,$$

which shows

$$\sum_{i=0}^{n-1} \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big)\Big)^2 - T = \sum_{i=0}^{n-1} \xi_i \xrightarrow{n\to\infty} 0\,. \qquad \square$$

COROLLARY 3.3. *The process $M(t) \stackrel{\text{def}}{=} W(t)^2 - [W, W](t)$ is a martingale.*

Proof. We see

$$\begin{aligned}\boldsymbol{E}(W(t)^2 - t \mid \mathcal{F}_s) &= \boldsymbol{E}((W(t)-W(s))^2 + 2W(s)(W(t)-W(s)) + W(s)^2 \mid \mathcal{F}_s) - t \\ &= W(s)^2 - s\end{aligned}$$

and hence $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) = M(s)$. □

The above wasn't a co-incidence. This property in fact characterizes the quadratic variation.

Theorem 3.4. *Let M be a continuous martingale with respect to a filtration $\{\mathcal{F}_t\}$. Then $\boldsymbol{E}M(t)^2 < \infty$ if and only if $\boldsymbol{E}[M,M](t) < \infty$. In this case the process $M(t)^2 - [M,M](t)$ is also a martingale with respect to the same filtration, and hence $\boldsymbol{E}M(t)^2 - \boldsymbol{E}M(0)^2 = \boldsymbol{E}[M,M](t)$.*

The above is in fact a characterization of the quadratic variation of martingales.

Theorem 3.5. *If $A(t)$ is any continuous, increasing, adapted process such that $A(0) = 0$ and $M(t)^2 - A(t)$ is a martingale, then $A = [M,M]$.*

The proof of these theorems are a bit technical and go beyond the scope of these notes. The results themselves, however, are extremely important and will be used subsequently.

Remark 3.6. The intuition to keep in mind about the first variation and the quadratic variation is the following. Divide the interval $[0,T]$ into $T/\delta t$ intervals of size δt. If X has finite first variation, then on each subinterval $(k\delta t, (k+1)\delta t)$ the increment of X should be of order δt. Thus adding $T/\delta t$ terms of order δt will yield something finite.

On the other hand if X has finite quadratic variation, on each subinterval $(k\delta t, (k+1)\delta t)$ the increment of X should be of order $\sqrt{\delta t}$, so that adding $T/\delta t$ terms of the *square* of the increment yields something finite. Doing a quick check for Brownian motion (which has finite quadratic variation), we see

$$\boldsymbol{E}|W(t+\delta t) - W(t)| = \sqrt{\delta t}\boldsymbol{E}|N(0,1)|\,,$$

which is in line with our intuition.

Remark 3.7. If a continuous process has finite first variation, its quadratic variation will necessarily be 0. On the other hand, if a continuous process has finite (and non-zero) quadratic variation, its first variation will necessary be infinite.

4. Construction of the Itô integral

Let W be a standard Brownian motion, $\{\mathcal{F}_t\}$ be the Brownian filtration and D be an adapted process. We think of $D(t)$ to represent our position at time t on an asset whose spot price is $W(t)$.

Lemma 4.1. *Let $P = \{0 = t_0 < t_1 < t_2 < \cdots\}$ be an increasing sequence of times, and assume D is constant on $[t_i, t_{i+1})$ (i.e. the asset is only traded at times t_0, …, t_n). Let $I_P(T)$, defined by*

$$I_P(T) = \sum_{i=0}^{n-1} D(t_i)\Delta_i W + D(t_n)(W(T) - W(t_n)) \quad \text{if } T \in [t_n, t_{n+1})\,.$$

be your cumulative winnings up to time T. As before $\Delta_i W \stackrel{\text{def}}{=} W(t_{i+1}) - W(t_i)$. Then,

$$\boldsymbol{E} I_P(T)^2 = \boldsymbol{E}\Big[\sum_{i=0}^{n} D(t_i)^2(t_{i+1}-t_i) + D(t_n)^2(T-t_n)\Big] \quad \text{if } T \in [t_n, t_{n+1})\,. \tag{4.1}$$

Moreover, I_P is a martingale and

$$[I_P, I_P](T) = \sum_{i=0}^{n-1} D(t_i)^2(t_{i+1}-t_i) + D(t_n)^2(T-t_n) \quad \text{if } T \in [t_n, t_{n+1})\,. \tag{4.2}$$

This lemma, as we will shortly see, is the key to the construction of stochastic integrals (called Itô integrals).

Proof. We first prove (4.1) with $T = t_n$ for simplicity. Note

$$\boldsymbol{E} I_P(t_n)^2 = \sum_{i=0}^{n-1} \boldsymbol{E} D(t_i)^2 (\Delta_i W)^2 + 2\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \boldsymbol{E} D(t_i) D(t_j)(\Delta_i W)(\Delta_j W) \tag{4.3}$$

By the tower property

$$\begin{aligned}\boldsymbol{E} D(t_i)^2(\Delta_i W)^2 &= \boldsymbol{E}\boldsymbol{E}\big(D(t_i)^2(\Delta_i W)^2 \bigm| \mathcal{F}_{t_i}\big)\\ &= \boldsymbol{E} D(t_i)^2 \boldsymbol{E}\big((W(t_{i+1}) - W(t_i))^2 \bigm| \mathcal{F}_{t_i}\big) = \boldsymbol{E} D(t_i)^2 (t_{i+1}-t_i)\,.\end{aligned}$$

Similarly we compute

$$\begin{aligned}\boldsymbol{E} D(t_i) D(t_j)(\Delta_i W)(\Delta_j W) &= \boldsymbol{E}\boldsymbol{E}\big(D(t_i)D(t_j)(\Delta_i W)(\Delta_j W) \bigm| \mathcal{F}_{t_j}\big)\\ &= \boldsymbol{E} D(t_i) D(t_j)(\Delta_i W)\boldsymbol{E}\big((W(t_{j+1}) - W(t_j)) \bigm| \mathcal{F}_{t_j}\big) = 0\,.\end{aligned}$$

Substituting these in (4.3) immediately yields (4.1) for $t_n = T$.

The proof that I_P is an martingale uses the same "tower property" idea, and is left to the reader to check. The proof of (4.2) is also similar in spirit, but has a few more details to check. The main idea is to let $A(t)$ be the right hand side of (4.2). Observe A is clearly a continuous, increasing, adapted process. Thus, if we show $M^2 - A$ is a martingale, then using Theorem 3.5 we will have $A = [M, M]$ as desired. The proof that $M^2 - A$ is an martingale uses the same "tower property" idea, but is a little more technical and is left to the reader. □

Note that as $\|P\| \to 0$, the right hand side of (4.2) converges to the standard Riemann integral $\int_0^T D(t)^2\,dt$. Itô realised he could use this to prove that I_P itself converges, and the limit is now called the Itô integral.

Theorem 4.2. *If $\int_0^T D(t)^2\,dt < \infty$ almost surely, then as $\|P\| \to 0$, the processes I_P converge to a* continuous process *I denoted by*

$$I(T) \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} I_P(T) \stackrel{\text{def}}{=} \int_0^T D(t)\,dW(t)\,. \tag{4.4}$$

This is known as the Itô integral of D with respect to W. *If further*

$$\boldsymbol{E}\int_0^T D(t)^2\,dt < \infty\,, \tag{4.5}$$

then the process $I(T)$ *is a* martingale *and the quadratic variation* $[I, I]$ *satisfies*

$$[I, I](T) = \int_0^T D(t)^2 \, dt \quad \textit{almost surely.}$$

REMARK 4.3. For the above to work, it is *crucial* that D is adapted, and is sampled at the left endpoint of the time intervals. That is, the terms in the sum are $D(t_i)(W(t_{i+1}) - W(t_i))$, and not $D(t_{i+1})(W(t_{i+1}) - W(t_i))$ or $\frac{1}{2}(D(t_i) + D(t_{i+1}))(W(t_{i+1}) - W(t_i))$, or something else.

Usually if the process is not adapted, there is no meaningful way to make sense of the limit. However, if you sample at different points, it still works out (usually) but what you get is *different* from the Itô integral (one example is the Stratonovich integral).

REMARK 4.4. The variable t used in (4.4) is a "dummy" integration variable. Namely one can write

$$\int_0^T D(t) \, dW(t) = \int_0^T D(s) \, dW(s) = \int_0^T D(r) \, dW(r) \,,$$

or any other variable of your choice.

COROLLARY 4.5 (Itô Isometry). *If* (4.5) *holds then*

$$\boldsymbol{E}\Big(\int_0^T D(t) \, dW(t)\Big)^2 = \boldsymbol{E} \int_0^T D(t)^2 \, dt \,.$$

PROPOSITION 4.6 (Linearity). *If* D_1 *and* D_2 *are two adapted processes, and* $\alpha \in \mathbb{R}$, *then*

$$\int_0^T (D_1(t) + \alpha D_2(t)) \, dW(t) = \int_0^T D_1(t) \, dW(t) + \alpha \int_0^T D_2(t) \, dW(t) \,.$$

REMARK 4.7. Positivity, however, is not preserved by Itô integrals. Namely if $D_1 \leqslant D_2$, there is no reason to expect $\int_0^T D_1(t) \, dW(t) \leqslant \int_0^T D_2(t) \, dW(t)$. Indeed choosing $D_1 = 0$ and $D_2 = 1$ we see that we can not possibly have $0 = \int_0^T D_1(t) \, dW(t)$ to be almost surely smaller than $W(T) = \int_0^T D_2(t) \, dW(t)$.

Recall, our starting point in these notes was modelling stock prices as *geometric Brownian motions*, given by the equation

$$dS(t) = \alpha S(t) \, dt + \sigma S(t) \, dW(t) \,.$$

After constructing Itô integrals, we are now in a position to describe what this means. The above is simply shorthand for saying S is a process that satisfies

$$S(T) - S(0) = \int_0^T \alpha S(t) \, dt + \int_0^T \sigma S(t) \, dW(t) \,.$$

The first integral on the right is a standard Riemann integral. The second integral, representing the noisy fluctuations, is the Itô integral we just constructed.

Note that the above is a little more complicated than the Itô integrals we will study first, since the process S (that we're trying to define) also appears as an integrand on the right hand side. In general, such equations are called *Stochastic differential equations*, and are extremely useful in many contexts.

5. The Itô formula

Using the abstract "limit" definition of the Itô integral, it is hard to compute examples. For instance, what is

$$\int_0^T W(s)\,dW(s)\,?$$

This, as we will shortly, can be computed easily using the Itô formula (also called the Itô-Doeblin formula).

Suppose b and σ are adapted processes. (In particular, they could but need not, be random). Consider a process X defined by

$$X(T) = X(0) + \int_0^T b(t)\,dt + \int_0^T \sigma(t)\,dW(t)\,. \tag{5.1}$$

Note the first integral $\int_0^T b(t)\,dt$ is a regular Riemann integral that can be done directly. The second integral the Itô integral we constructed in the previous section.

DEFINITION 5.1. The process X is called an Itô process if $X(0)$ is deterministic (not random) and for all $T \geqslant 0$,

$$\boldsymbol{E}\int_0^T \sigma(t)^2\,dt < \infty \qquad \text{and} \qquad \int_0^T b(t)\,dt < \infty\,.$$

REMARK 5.2. The shorthand notation for (5.1) is to write

$$dX(t) = b(t)\,dt + \sigma(t)\,dW(t)\,. \tag{5.1'}$$

PROPOSITION 5.3. *The quadratic variation of X is*

$$[X,X](T) = \int_0^T \sigma(t)^2\,dt \quad \textit{almost surely.} \tag{5.2}$$

PROOF. Define B and M by

$$B(T) = \int_0^T b(t)\,dt \qquad \text{and} \qquad M(T) = \int_0^T \sigma(t)\,dW(t)\,,$$

and let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$, and $\|P\| = \max_i t_{i+1} - t_i$. Observe

$$\sum_{i=0}^{n-1} (\Delta_i X)^2 = \sum_{i=0}^{n-1} (\Delta_i M)^2 + \sum_{i=0}^{n-1} (\Delta_i B)^2 + 2\sum_{i=0}^{n-1} (\Delta_i B)(\Delta_i M)\,.$$

The first sum on the right converges (as $\|P\| \to 0$) to $[M,M](T)$, which we know is exactly $\int_0^T \sigma(t)^2\,dt$. For the second sum, observe

$$(\Delta_i B)^2 = \Big(\int_{t_i}^{t_{i+1}} b(s)\,ds\Big)^2 \leqslant \big(\max|b|^2\big)(t_{i+1}-t_i)^2 \leqslant \big(\max|b|^2\big)\|P\|(t_{i+1}-t_i)\,.$$

Hence

$$\Big|\sum_{i=0}^{n-1} (\Delta_i B)^2\Big| \leqslant \|P\|\big(\max|b|^2\big)T \xrightarrow{\|P\|\to 0} 0\,.$$

For the third term, one uses the *Cauchy-Schwartz* inequality to observe

$$\Big|\sum_{i=0}^{n-1} (\Delta_i B)(\Delta_i M)\Big| \leqslant \Big(\sum_{i=0}^{n-1} (\Delta_i B)^2\Big)^{1/2}\Big(\sum_{i=0}^{n-1} (\Delta_i M)^2\Big)^{1/2} \xrightarrow{\|P\|\to 0} 0 \cdot [M,M](T) = 0\,.$$

□

REMARK 5.4. It's common to decompose $X = B + M$ where M is a martingale and B has finite first variation. Processes that can be decomposed in this form are called *semi-martingales*, and the decomposition is unique. The process M is called the martingale part of X, and B is called the *bounded variation* part of X.

PROPOSITION 5.5. *The semi-martingale decomposition of X is unique. That is, if $X = B_1 + M_1 = B_2 + M_2$ where B_1, B_2 are continuous adapted processes with finite first variation, and M_1, M_2 are continuous (square integrable) martingales, then $B_1 = B_2$ and $M_1 = M_2$.*

PROOF. Set $M = M_1 - M_2$ and $B = B_1 - B_2$, and note that $M = -B$. Consequently, M has finite first variation and hence 0 quadratic variation. This implies $\boldsymbol{E}M(t)^2 = \boldsymbol{E}[M, M](t) = 0$ and hence $M = 0$ identically, which in turn implies $B = 0$, $B_1 = B_2$ and $M_1 = M_2$. □

Given an adapted process D, interpret X as the price of an asset, and D as our position on it. (We could either be long, or short on the asset so D could be positive or negative.)

DEFINITION 5.6. We define the integral of D with respect to X by

$$\int_0^T D(t)\,dX(t) \stackrel{\text{def}}{=} \int_0^T D(t)b(t)\,dt + \int_0^T D(t)\sigma(t)\,dW(t)\,.$$

As before, $\int_0^T D\,dX$ represents the winnings or profit obtained using the trading strategy D.

REMARK 5.7. Note that the first integral on the right $\int_0^T D(t)b(t)\,dt$ is a regular Riemann integral, and the second one is an Itô integral. Recall that Itô integrals with respect to Brownian motion (i.e. integrals of the form $\int_0^t D(s)\,dW(s)$ are martingales). Integrals with respect to a general process X are only guaranteed to be martingales if X itself is a martingale (i.e. $b = 0$), or if the integrand is 0.

REMARK 5.8. If we define I_P by

$$I_P(T) = \sum_{i=0}^{n-1} D(t_i)(\Delta_i X) + D(t_n)(X(T) - X(t_n)) \quad \text{if } T \in [t_n, t_{n+1})\,,$$

then I_P converges to the integral $\int_0^T D(t)\,dX(t)$ defined above. This works in the same way as Theorem 4.2.

Suppose now $f(t, x)$ is some function. If X is differentiable as a function of t *(which it most certainly is not)*, then the chain rule gives

$$\begin{aligned} f(T, X(T)) - f(0, X(0)) &= \int_0^T \partial_t \Big(f(t, X(t)) \Big)\,dt \\ &= \int_0^T \partial_t f(t, X(t))\,dt + \int_0^T \partial_x f(t, X(t))\,\partial_t X(t)\,dt \\ &= \int_0^T \partial_t f(t, X(t))\,dt + \int_0^T \partial_x f(t, X(t))\,dX(t)\,. \end{aligned}$$

Itô process are *almost never* differentiable as a function of time, and so the above has no chance of working. It turns out, however, that for Itô process you can make the above work by adding an *Itô correction* term. This is the celebrated Itô formula (more correctly the Itô-Doeblin[1] formula).

THEOREM 5.9 (Itô formula, aka Itô-Doeblin formula). *If $f = f(t,x)$ is $C^{1,2}$ function*[2] *then*

$$f(T,X(T)) - f(0,X(0)) = \int_0^T \partial_t f(t,X(t))\,dt + \int_0^T \partial_x f(t,X(t))\,dX(t) + \frac{1}{2}\int_0^T \partial_x^2 f(t,X(t)\,d[X,X](t)\,. \tag{5.3}$$

REMARK 5.10. To clarify notation, $\partial_t f(t,X(t))$ means the following: differentiate $f(t,x)$ with respect to t (treating x as a constant), and then substitute $x = X(t)$. Similarly $\partial_x f(t,X(t))$ means differentiate $f(t,x)$ with respect to x, and then substitute $x = X(t)$. Finally $\partial_x^2 f(t,X(t))$ means take the second derivative of the function $f(t,x)$ with respect to x, and the substitute $x = X(t)$.

REMARK 5.11. In short hand differential form, this is written as

$$df(t,X(t)) = \partial_t f(t,X(t))\,dt + \partial_x f(t,X(t))\,dX(t) + \frac{1}{2}\partial_x^2 f(t,X(t))\,d[X,X](t)\,. \tag{5.3$'$}$$

The term $\frac{1}{2}\partial_x^2 f\,d[X,X](t)$ is an "extra" term, and is often referred to as the *Itô correction term*. The Itô formula is simply a version of the *chain rule* for stochastic processes.

REMARK 5.12. Substituting what we know about X from (5.1) and (5.2) we see that (5.3) becomes

$$f(T,X(T)) - f(0,X(0)) = \int_0^T \big(\partial_t f(t,X(t)) + \partial_x f(t,X(t))b(t)\big)\,dt + \int_0^T \partial_x f(t,X(t))\sigma(t)\,dW(t) + \frac{1}{2}\int_0^T \partial_x^2 f(t,X(t))\,\sigma(t)^2\,dt\,.$$

The second integral on the right is an Itô integral (and hence a martingale). The other integrals are regular Riemann integrals which yield processes of finite variation.

While a complete rigorous proof of the Itô formula is technical, and beyond the scope of this course, we provide a quick heuristic argument that illustrates the main idea clearly.

---

[1]W. Doeblin was a French-German mathematician who was drafted for military service during the second world war. During the war he wrote down his mathematical work and sent it in a sealed envelope to the French Academy of Sciences, because he did not want it to "fall into the wrong hands". When he was about to be captured by the Germans he burnt his mathematical notes and killed himself.

The sealed envelope was opened in 2000 which revealed that he had a treatment of stochastic Calculus that was essentially equivalent to Itô's. In posthumous recognition, Itô's formula is now referred to as the Itô-Doeblin formula by many authors.

[2]Recall a function $f = f(t,x)$ is said to be $C^{1,2}$ if it is C^1 in t (i.e. differentiable with respect to t and $\partial_t f$ is continuous), and C^2 in x (i.e. twice differentiable with respect to x and $\partial_x f$, $\partial_x^2 f$ are both continuous).

Intuition behind the Itô formula. Suppose that the function f is only a function of x and doesn't depend on t, and X is a standard Brownian motion (i.e. $b = 0$ and $\sigma = 1$). In this case proving Itô's formula reduces to proving

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

Let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0, T]$. Taylor expanding f to second order gives

$$\begin{aligned} f(W(T)) - f(W(0)) &= \sum_{i=0}^{n-1} f(W(t_{i+1})) - f(W(t_i)) \\ &= \sum_{i=0}^{n-1} f'(W(t_i))(\Delta_i W) + \frac{1}{2}\sum_{i=0}^{n-1} f''(W(t_i))(\Delta_i W)^2 + \frac{1}{2}\sum_{i=0}^{n-1} o\big((\Delta_i W)^2\big)\,, \end{aligned} \tag{5.4}$$

where the last sum on the right is the remainder from the Taylor expansion.

Note the first sum on the right of (5.4) converges to the Itô integral

$$\int_0^T f'(W(t))\,dW(t)\,.$$

For the second sum on the right of (5.4), note

$$f''(W(t_i))(\Delta_i W)^2 = f''(W(t_i))(t_{i+1} - t_i) + f''(W(t_i))\big[(\Delta_i W)^2 - (t_{i+1} - t_i)\big]\,.$$

After summing over i, first term on the right converges to the Riemann integral $\int_0^T f''(W(t))\,dt$. The second term on the right is similar to what we had when computing the quadratic variation of W. The variance of $\xi_i \stackrel{\text{def}}{=} (\Delta_i W)^2 - (t_{i+1} - t_i)$ is of order $(t_{i+1} - t_i)^2$. Thus we expect that the second term above, when summed over i, converges to 0.

Finally each summand in the remainder term (the last term on the right of (5.4)) is smaller than $(\Delta_i W)^2$. (If, for instance, f is three times continuously differentiable in x, then each summand in the remainder term is of order $(\Delta_i W)^3$.) Consequently, when summed over i this should converge to 0 □

6. A few examples using Itô's formula

Technically, as soon as you know Itô's formula you can "jump right in" and derive the Black-Scholes equation. However, because of the importance of Itô's formula, we work out a few simpler examples first.

Example 6.1. Compute the quadratic variation of $W(t)^2$.

Solution. Let $f(t, x) = x^2$. Then, by Itô's formula,

$$\begin{aligned} d\big(W(t)^2\big) &= df(t, W(t)) \\ &= \partial_t f(t, W(t))\,dt + \partial_x f(t, W(t))\,dW(t) + \frac{1}{2}\partial_x^2 f(t, W(t))\,dt \\ &= 2W(t)\,dW(t) + dt\,. \end{aligned}$$

Or, in integral form,

$$W(T)^2 - W(0)^2 = W(T)^2 = 2\int_0^T W(t)\,dW(t) + T\,.$$

Now the second term on the right has finite first variation, and won't affect our computations for quadratic variation. The first term is an martingale whose quadratic variation is $\int_0^T W(t)^2\,dt$, and so

$$[W^2, W^2](T) = 4\int_0^T W(t)^2\,dt\,. \qquad \square$$

REMARK 6.2. Note the above also tells you

$$2\int_0^T W(t)\,dW(t) = W(T)^2 - T\,.$$

EXAMPLE 6.3. Let $M(t) = W(t)$ and $N(t) = W(t)^2 - t$. We know M and N are martingales. Is MN a martingale?

SOLUTION. Note $M(t)N(t) = W(t)^3 - tW(t)$. By Itô's formula,

$$d(MN) = -W(t)\,dt + (3W(t)^2 - t)\,dW(t) + 3W(t)\,dt\,.$$

Or in integral form

$$M(t)N(t) = \int_0^t 2W(s)\,ds + \int_0^t (3W(s)^2 - s)\,dW(s)\,.$$

Now the second integral on the right is a martingale, but the first integral most certainly is not. So MN can not be a martingale. $\square$

REMARK 6.4. Note, above we changed the integration variable from t to s. This is perfectly legal – the variable with which you integrate with respect to is a dummy variable (just line regular Riemann integrals) and you can replace it what your favourite (unused!) symbol.

REMARK 6.5. It's worth pointing out that the Itô integral $\int_0^t \Delta(s)\,dW(s)$ is always a martingale (under the finiteness condition (4.5)). However, the Riemann integral $\int_0^t b(s)\,ds$ is only a martingale if $b = 0$ identically.

PROPOSITION 6.6. *If $f = f(t,x)$ is $C_b^{1,2}$ then the process*

$$M(t) \stackrel{\text{def}}{=} f(t, W(t)) - \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\Big)\,ds$$

is a martingale.

REMARK 6.7. We'd seen this earlier, and the proof involved computing the conditional expectations directly and checking an algebraic identity involving the density of the normal distribution. With Itô's formula, the proof is "immediate".

PROOF. By Itô's formula (in integral form)

$$\begin{aligned} &f(t, W(t)) - f(0, W(0)) \\ &\quad= \int_0^t \partial_t f(s, W(s))\,ds + \int_0^t \partial_x f(s, W(s))\,dW(s) + \frac{1}{2}\int_0^t \partial_x^2 f(s, W(s))\,ds \\ &\quad= \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\Big)\,ds + \int_0^t \partial_x f(s, W(s))\,dW(s)\,. \end{aligned}$$

Substituting this we see

$$M(t) = f(0, W(0)) + \int_0^t \partial_x f(s, W(s))\, dW(s)\,,$$

which is a martingale. □

REMARK 6.8. Note we said $f \in C_b^{1,2}$ to "cover our bases". Recall for Itô integrals to be martingales, we need the finiteness condition (4.5) to hold. This will certainly be the case if $\partial_x f$ is bounded, which is why we made this assumption. The result above is of course true under much more general assumptions.

EXAMPLE 6.9. Let $X(t) = t\sin(W(t))$. Is $X^2 - [X, X]$ a martingale?

SOLUTION. Let $f(t, x) = t\sin(x)$. Observe $X(t) = f(t, W(t))$, $\partial_t f = \sin x$, $\partial_x f = t\cos x$, and $\partial_x^2 f = -t\sin x$. Thus by Itô's formula,

$$\begin{aligned} dX(t) &= \partial_t f(t, W(t))\, dt + \partial_x f(t, W(t))\, dW(t) + \frac{1}{2}\partial_x^2 f(t, W(t))\, d[W, W](t) \\ &= \sin(W(t))\, dt + t\cos(W(t))\, dW(t) - \frac{1}{2} t\sin(W(t))\, dt\,, \end{aligned}$$

and so

$$d[X, X](t) = t^2\cos^2(W(t))\, dt\,.$$

Now let $g(x) = x^2$ and apply Itô's formula to compute $dg(X(t))$. This gives

$$dX(t)^2 = 2X(t)\, dX(t) + d[X, X](t)$$

and so

$$\begin{aligned} &d(X(t)^2 - [X, X]) = 2X(t)\, dX(t) \\ &\quad = 2t\sin(t)\Big(\sin(W(t)) - \frac{t\sin(W(t))}{2}\Big)\, dt + 2t\sin(t)\big(t\cos(W(t))\big)\, dW(t)\,. \end{aligned}$$

Since the dt term above isn't 0, $X(t)^2 - [X, X]$ can not be a martingale. □

Recall we said earlier (Theorem 3.4) that for any martingale M, $M^2 - [M, M]$ is a martingale. In the above example X is not a martingale, and so there is no contradiction when we show that $X^2 - [X, X]$ is not a martingale. If M is a martingale, Itô's formula can be used to "prove"[3] that $M^2 - [M, M]$ is a martingale.

PROPOSITION 6.10. *Let* $M(t) = \int_0^t \sigma(s)\, dW(s)$. *Then* $M^2 - [M, M]$ *is a martingale.*

PROOF. Let $N(t) = M(t)^2 - [M, M](t)$. Observe that by Itô's formula,

$$d(M(t)^2) = 2M(t)\, dM(t) + d[M, M](t)\,.$$

Hence

$$dN = 2M(t)\, dM(t) + d[M, M](t) - d[M, M](t) = 2M(t)\sigma(t)\, dW(t)\,.$$

Since there is no "dt" term and Itô integrals are martingales, N is a martingale. □

[3] We used the fact that $M^2 - [M, M]$ is a martingale crucially in the construction of Itô integrals, and hence in proving Itô's formula. Thus proving $M^2 - [M, M]$ is a martingale using the Itô's formula is circular and not a valid proof. It is however instructive, and helps with building intuition, which is why it is presented here.

7. Review Problems

PROBLEM 7.1. If $0 \leqslant r < s < t$, compute $\boldsymbol{E}\big(W(r)W(s)W(t)\big)$.

PROBLEM 7.2. Define the processes X, Y, Z by

$$X(t) = \int_0^{W(t)} e^{-s^2}\,ds\,, \quad Y(t) = \exp\Big(\int_0^t W(s)\,ds\Big)\,, \quad Z(t) = \frac{t}{X(t)}\,.$$

Decompose each of these processes as the sum of a martingale and a process of finite first variation. What is the quadratic variation of each of these processes?

PROBLEM 7.3. Define the processes X, Y by

$$X(t) \stackrel{\text{def}}{=} \int_0^t W(s)\,ds\,, \quad Y(t) \stackrel{\text{def}}{=} \int_0^t W(s)\,dW(s)\,.$$

Given $0 \leqslant s < t$, compute the conditional expectations $\boldsymbol{E}(X(t) \mid \mathcal{F}_s)$, and $\boldsymbol{E}(Y(t) \mid \mathcal{F}_s)$.

PROBLEM 7.4. Let $M(t) = \int_0^t W(s)\,dW(s)$. Find a function f such that

$$E(t) \stackrel{\text{def}}{=} \exp\Big(M(t) - \int_0^t f(s, W(s))\,ds\Big)$$

is a martingale.

PROBLEM 7.5. Suppose $\sigma = \sigma(t)$ is a deterministic (i.e. non-random) process, and X is the Itô process defined by

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

(a) Given $\lambda, s, t \in \mathbb{R}$ with $0 \leqslant s < t$ compute $\boldsymbol{E}(e^{\lambda(X(t)-X(s))} \mid \mathcal{F}_s)$.
(b) If $r \leqslant s$ compute $\boldsymbol{E}\exp(\lambda X(r) + \mu(X(t) - X(s)))$.
(c) What is the joint distribution of $(X(r), X(t) - X(s))$?
(d) *(Lévy's criterion)* If $\sigma(u) = \pm 1$ for all u, then show that X is a standard Brownian motion.

PROBLEM 7.6. Define the process X, Y by

$$X = \int_0^t s\,dW(s)\,, \quad Y = \int_0^t W(s)\,ds\,.$$

Find a formula for $\boldsymbol{E}X(t)^n$ and $\boldsymbol{E}Y(t)^n$ for any $n \in \mathbb{N}$.

PROBLEM 7.7. Let $M(t) = \int_0^t W(s)\,dW(s)$. For $s < t$, is $M(t) - M(s)$ independent of $\mathcal{F}_s$? Justify.

PROBLEM 7.8. Determine whether the following identities are true or false, and justify your answer.

(a) $e^{2t}\sin(2W(t)) = 2\int_0^t e^{2s}\cos(2W(s))\,dW(s)$.
(b) $|W(t)| = \int_0^t \operatorname{sign}(W(s))\,dW(s)$. (Recall $\operatorname{sign}(x) = 1$ if $x > 0$, $\operatorname{sign}(x) = -1$ if $x < 0$ and $\operatorname{sign}(x) = 0$ if $x = 0$.)

8. The Black Scholes Merton equation.

The price of an asset with a constant rate of return α is given by

$$dS(t) = \alpha S(t)\, dt\,.$$

To account for noisy fluctuations we model stock prices by adding the term $\sigma S(t)\, dW(t)$ to the above:

$$dS(t) = \alpha S(t)\, dt + \sigma S(t) dW(t)\,. \tag{8.1}$$

The parameter α is called the *mean return rate* or the *percentage drift*, and the parameter σ is called the *volatility* or the *percentage volatility.*

DEFINITION 8.1. A stochastic process S satisfying (8.1) above is called a *geometric Brownian motion.*

The reason S is called a geometric Brownian motion is as follows. Set $Y = \ln S$ and observe

$$dY(t) = \frac{1}{S(t)}\, dS(t) - \frac{1}{2S(t)^2}\, d[S,S](t) = \Big(\alpha - \frac{\sigma^2}{2}\Big)\, dt + \sigma\, dW(t)\,.$$

If $\alpha = \sigma^2/2$ then $Y = \ln S$ is simply a Brownian motion.

We remark, however, that our application of Itô's formula above is not completely justified. Indeed, the function $f(x) = \ln x$ is *not* differentiable at $x = 0$, and Itô's formula requires f to be at least C^2. The reason the application of Itô's formula here is valid is because the process S *never* hits the point $x = 0$, and at all other points the function f is infinitely differentiable.

The above also gives us an explicit formula for S. Indeed,

$$\ln\Big(\frac{S(t)}{S(0)}\Big) = \Big(\alpha - \frac{\sigma^2}{2}\Big)t + \sigma W(t)\,,$$

and so

$$S(t) = S(0) \exp\Big(\Big(\alpha - \frac{\sigma^2}{2}\Big)t + \sigma W(t)\Big)\,. \tag{8.2}$$

Now consider a European call option with strike price K and maturity time T. This is a security that allows you the option (not obligation) to buy S at price K and time T. Clearly the price of this option at time T is $(S(T) - K)^+$. Our aim is to compute the *arbitrage free*[4] price of such an option at time $t < T$.

Black and Scholes realised that the price of this option at time t should only depend on the asset price $S(t)$, and the current time t (or more precisely, the time to maturity $T - t$), and of course the model parameters α, σ. In particular, the option price does not depend on the price history of S.

THEOREM 8.2. *Suppose we have an arbitrage free financial market consisting of a money market account with constant return rate r, and a risky asset whose price is given by S. Consider a European call option with strike price K and maturity T.*

---

[4] In an arbitrage free market, we say p is the arbitrage free price of a non traded security if given the opportunity to trade the security at price p, the market is still arbitrage free. (Recall a financial market is said to be *arbitrage free* if there doesn't exist a self-financing portfolio X with $X(0) = 0$ such that at some $t > 0$ we have $X(t) \geqslant 0$ and $\boldsymbol{P}(X(t) > 0) > 0$.)

(1) *If $c = c(x,t)$ is a function such that at any time $t \leqslant T$, the arbitrage free price of this option is $c(t, S(t))$, then c satisfies*

$$\partial_t c + rx\partial_x c + \frac{\sigma^2 x^2}{2}\partial_x^2 c - rc = 0 \qquad x > 0,\ t < T\,, \tag{8.3}$$

$$c(t,0) = 0 \qquad t \leqslant T\,, \tag{8.4}$$

$$c(T,x) = (x-K)^+ \qquad x \geqslant 0\,. \tag{8.5}$$

(2) *Conversely, if c satisfies (8.3)–(8.5) then $c(t, S(t))$ is the arbitrage free price of this option at any time $t \leqslant T$.*

Remark 8.3. Since α, σ and T are fixed, we suppress the explicit dependence of c on these quantities.

Remark 8.4. The above result assumes the following:

(1) The market is *frictionless* (i.e. there are no transaction costs).
(2) The asset is liquid and fractional quantities of it can be traded.
(3) The borrowing and lending rate are both r.

Remark 8.5. Even though the asset price $S(t)$ is random, the function c is a deterministic (non-random) function. The option price, however, is $c(t, S(t))$, which is certainly random.

Remark 8.6. Equation (8.3)–(8.5) are the *Black Scholes Merton PDE*. This is a *partial differential equation*, which is a differential equation involving derivatives with respect to more than one variable. Equation (8.3) governs the evolution of c for $x \in (0, \infty)$ and $t < T$. Equation (8.5) specifies the terminal condition at $t = T$, and equation (8.4) specifies a boundary condition at $x = 0$.

To be completely correct, one also needs to add a boundary condition as $x \to \infty$ to the system (8.3)–(8.5). When x is very large, the call option is deep in the money, and will very likely end in the money. In this case the replicating portfolio should be long one share of the asset and short $e^{-r(T-t)}K$, the discounted strike price, in cash. This means that when x is very large, $c(x,t) \approx x - Ke^{-r(T-t)}$, and hence a boundary condition at $x = \infty$ can be obtained by supplementing (8.4) with

$$\lim_{x\to\infty} \big(c(t,x) - (x - Ke^{-r(T-t)}) \big) = 0\,. \tag{8.4$'$}$$

Remark 8.7. The system (8.3)–(8.5) can be solved explicitly using standard calculus by substituting $y = \ln x$ and converting it into the *heat equation*, for which the solution is explicitly known. This gives the Black-Scholes-Merton formula

$$c(t,x) = xN(d_+(T-t,x)) - Ke^{-r(T-t)}N(d_-(T-t,x)) \tag{8.6}$$

where

$$d_\pm(\tau, x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{\tau}}\Big(\ln\Big(\frac{x}{K}\Big) + \Big(r \pm \frac{\sigma^2}{2}\Big)\tau\Big)\,, \tag{8.7}$$

and

$$N(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy\,, \tag{8.8}$$

is the CDF of a standard normal variable.

Even if you're unfamiliar with the techniques involved in arriving at the solution above, you can certainly check that the function c given by (8.6)–(8.7) above satisfies (8.3)–(8.5). Indeed, this is a direct calculation that only involves patience and a careful application of the chain rule. We will, however, derive (8.6)–(8.7) later using risk neutral measures.

We will prove Theorem 8.2 by using a *replicating portfolio.* This is a portfolio (consisting of cash and the risky asset) that has exactly the same cash flow at maturity as the European call option that needs to be priced. Specifically, let $X(t)$ be the value of the replicating portfolio and $\Delta(t)$ be the number of shares of the asset held. The remaining $X(t) - S(t)\Delta(t)$ will be invested in the money market account with return rate r. (It is possible that $\Delta(t)S(t) > X(t)$, in which means we borrow money from the money market account to invest in stock.) For a replicating portfolio, the trading strategy Δ should be chosen in a manner that ensures that we have the same cash flow as the European call option. That is, we must have $X(T) = (S(T) - K)^+ = c(T, S(T))$. Now the arbitrage free price is precisely the value of this portfolio.

Remark 8.8. Through the course of the proof we will see that given the function c, the number of shares of S the replicating portfolio should hold is given by the *delta hedging rule*

$$\Delta(t) = \partial_x c(t, S(t))\,. \tag{8.9}$$

Remark 8.9. Note that there is no α dependence in the system (8.3)–(8.5), and consequently the formula (8.6) does not depend on α. At first sight, this might appear surprising. (In fact, Black and Scholes had a hard time getting the original paper published because the community couldn't believe that the final formula is independent of α.) The fact that (8.6) is independent of α can be heuristically explained by the fact that the replicating portfolio also holds the same asset: thus a high mean return rate will help both an investor holding a call option and an investor holding the replicating portfolio. (Of course this isn't the entire story, as one has to actually write down the dependence and check that an investor holding the call option benefits exactly as much as an investor holding the replicating portfolio. This is done below.)

Proof of Theorem 8.2 part 1. If $c(t, S(t))$ is the arbitrage free price, then, by definition

$$c(t, S(t)) = X(t)\,, \tag{8.10}$$

where $X(t)$ is the value of a replicating portfolio. Since our portfolio holds $\Delta(t)$ shares of S and $X(t) - \Delta(t)S(t)$ in a money market account, the evolution of the value of this portfolio is given by

$$\begin{aligned} dX(t) &= \Delta(t)\,dS(t) + r\big(X(t) - \Delta(t)S(t)\big)\,dt \\ &= \big(rX(t) + (\alpha - r)\Delta(t)S(t)\big)\,dt + \sigma\Delta(t)S(t)\,dW(t)\,. \end{aligned}$$

Also, by Itô's formula we compute

$$\begin{aligned} dc(t, S(t)) &= \partial_t c(t, S(t))\,dt + \partial_x c(t, S(t))\,dS(t) + \frac{1}{2}\partial_x^2 c(t, S(t))\,d[S, S](t)\,, \\ &= \Big(\partial_t c + \alpha S \partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\,dt + \partial_x c\,\sigma S\,dW(t) \end{aligned}$$

where we suppressed the $(t, S(t))$ argument in the last line above for convenience.

Equating $dc(t, S(t)) = dX(t)$ gives

$$\big(rX(t) + (\alpha - r)\Delta(t)S(t)\big)\,dt + \sigma\Delta(t)S(t)\,dW(t) \\ = \Big(\partial_t c + \alpha S \partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\,dt + \partial_x c\,\sigma S\,dW(t)\,.$$

Using uniqueness of the semi-martingale decomposition (Proposition 5.5) we can equate the dW and the dt terms respectively. Equating the dW terms gives the delta hedging rule (8.9). Writing $S(t) = x$ for convenience, equating the dt terms and using (8.10) gives (8.3). Since the payout of the option is $(S(T) - K)^+$ at maturity, equation (8.5) is clearly satisfied.

Finally if $S(t_0) = 0$ at one particular time, then we must have $S(t) = 0$ at all times, otherwise we would have an arbitrage opportunity. (This can be checked directly from the formula (8.2) of course.) Consequently the arbitrage free price of the option when $S = 0$ is 0, giving the *boundary condition* (8.4). Hence (8.3)–(8.5) are all satisfied, finishing the proof. ☐

Proof of Theorem 8.2 part 2. For the converse, we suppose c satisfies the system (8.3)–(8.5). Choose $\Delta(t)$ by the delta hedging rule (8.9), and let X be a portfolio with initial value $X(0) = c(0, S(0))$ that holds $\Delta(t)$ shares of the asset at time t and the remaining $X(t) - \Delta(t)S(t)$ in cash. We claim that X is a replicating portfolio (i.e. $X(T) = (S(T) - K)^+$ almost surely) and $X(t) = c(t, S(t))$ for all $t \leqslant T$. Once this is established $c(t, S(t))$ is the arbitrage free price as desired.

To show X is a replicating portfolio, first claim that $X(t) = c(t, S(t))$ for all $t < T$. To see this, let $Y(t) = e^{-rt}X(t)$ be the discounted value of X. (That is, $Y(t)$ is the value of $X(t)$ converted to cash at time $t = 0$.) By Itô's formula, we compute

$$dY(t) = -rY(t)\,dt + e^{-rt}dX(t) \\ = e^{-rt}(\alpha - r)\Delta(t)S(t)\,dt + e^{-rt}\sigma\Delta(t)S(t)\,dW(t)\,.$$

Similarly, using Itô's formula, we compute

$$d\big(e^{-rt}c(t, S(t))\big) = e^{-rt}\Big(-rc + \partial_t c + \alpha S\partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\,dt + e^{-rt}\partial_x c\,\sigma S\,dW(t)\,.$$

Using (8.3) this gives

$$d\big(e^{-rt}c(t, S(t))\big) = e^{-rt}(\alpha - r)S\partial_x c\,dt + e^{-rt}\partial_x c\,\sigma S\,dW(t) = dY(t)\,,$$

since $\Delta(t) = \partial_x c(t, S(t))$ by choice. This forces

$$e^{-rt}X(t) = X(0) + \int_0^t dY(s) = X(0) + \int_0^t d\big(e^{-rs}c(s, S(s))\big) \\ = X(0) + e^{-rt}c(t, S(t)) - c(0, S(0)) = e^{-rt}c(t, S(t))\,,$$

since we chose $X(0) = c(0, S(0))$. This forces $X(t) = c(t, S(t))$ for all $t < T$, and by continuity also for $t = T$. Since $c(T, S(T)) = (S(T) - K)^+$ we have $X(T) = (S(T) - K)^+$ showing X is a replicating portfolio, concluding the proof. ☐

Remark 8.10. In order for the application of Itô's formula to be valid above, we need $c \in C^{1,2}$. This is certainly false at time T, since $c(T, x) = (x - K)^+$ which is not even differentiable, let alone twice continuously differentiable. However, if c satisfies the system (8.3)–(8.5), then it turns out that for every $t < T$ the function

c will be infinitely differentiable with respect to x. This is why our proof first shows that $c(t, S(t)) = X(t)$ for $t < T$ and not directly that $c(t, S(t)) = X(t)$ for all $t \leqslant T$.

REMARK 8.11 (Put Call Parity). The same argument can be used to compute the arbitrage free price of European put options (i.e. the option to sell at the strike price, instead of buying). However, once the price of the price of a call option is computed, the *put call parity* can be used to compute the price of a put.

Explicitly let $p = p(t, x)$ be a function such that at any time $t \leqslant T$, $p(t, S(t))$ is the arbitrage free price of a European put option with strike price K. Consider a portfolio X that is long a call and short a put (i.e. buy one call, and sell one put). The value of this portfolio at time $t < T$ is

$$X(t) = c(t, S(t)) - p(t, S(t))$$

and at maturity we have[5]

$$X(T) = (S(T) - K)^+ - (K - S(T))^+ = S(T) - K\,.$$

This payoff can be replicated using a portfolio that holds one share of the asset and borrows Ke^{-rT} in cash (with return rate r) at time 0. Thus, in an arbitrage free market, we should have

$$c(t, S(t)) - p(t, S(t)) = X(t) = S(t) - Ke^{-r(T-t)}\,.$$

Writing x for $S(t)$ this gives the *put call parity* relation

$$c(t, x) - p(t, x) = x - Ke^{-r(T-t)}\,.$$

Using this the price of a put can be computed from the price of a call.

We now turn to understanding properties of c. The partial derivatives of c with respect to t and x measure the sensitivity of the option price to changes in the time to maturity and spot price of the asset respectively. These are called "the Greeks":

(1) The *delta* is defined to be $\partial_x c$, and is given by

$$\partial_x c = N(d_+) + xN'(d_+)d'_+ - Ke^{-r\tau}N'(d_-)d'_-\,.$$

where $\tau = T - t$ is the time to maturity. Recall $d_\pm = d_\pm(\tau, x)$, and we suppressed the (τ, x) argument above for notational convenience. Using the formulae (8.6)–(8.8) one can verify

$$d'_+ = d'_- = \frac{1}{x\sigma\sqrt{\tau}} \qquad \text{and} \qquad xN'(d_+) = Ke^{-r\tau}N'(d_-)\,,$$

and hence the delta is given by

$$\partial_x c = N(d_+)\,.$$

Recall that the *delta hedging rule* (equation (8.9)) explicitly tells you that the replicating portfolio should hold precisely $\partial_x c(t, S(t))$ shares of the risky asset and the remainder in cash.

(2) The *gamma* is defined to be $\partial_x^2 c$, and is given by

$$\partial_x^2 c = N'(d_+)d'_+ = \frac{1}{x\sigma\sqrt{2\pi\tau}} \exp\Big(\frac{-d_+^2}{2}\Big)\,.$$

[5] A *forward contract* requires the holder to buy the asset at price K at maturity. The value of this contract at maturity is exactly $S(T) - K$, and so a portfolio that is long a put and short a call has exactly the same cash flow as a forward contract.

(3) Finally the *theta* is defined to be $\partial_t c$, and simplifies to

$$\partial_t c = -rKe^{-r\tau} N(d_-) - \frac{\sigma x}{2\sqrt{\tau}} N'(d_+)$$

PROPOSITION 8.12. *The function $c(t,x)$ is convex and increasing as a function of x, and is* decreasing *as a function of t.*

PROOF. This follows immediately from the fact that $\partial_x c > 0$, $\partial_x^2 c > 0$ and $\partial_t c < 0$. ☐

REMARK 8.13 (Hedging a short call). Suppose you sell a call option valued at $c(t,x)$, and want to create a replicating portfolio. The delta hedging rule calls for $x\partial_x c(t,x)$ of the portfolio to be invested in the asset, and the rest in the money market account. Consequently the value of your money market account is

$$c(t,x) - x\partial_x c = xN(d_+) - Ke^{-r\tau}N(d_-) - xN(d_+) = -Ke^{-r\tau}N(d_-) < 0\,.$$

Thus to properly hedge a short call you will have to borrow from the money market account and invest it in the asset. As $t \to T$ you will end up selling shares of the asset if $x < K$, and buying shares of it if $x > K$, so that at maturity you will hold the asset if $x > K$ and not hold it if $x < K$. To hedge a long call you do the opposite.

REMARK 8.14 (Delta neutral and Long Gamma). Suppose at some time t the price of a stock is x_0. We short $\partial_x c(t,x_0)$ shares of this stock buy the call option valued at $c(t,x_0)$. We invest the balance $M = x_0\partial_x c(t,x_0) - c(t,x_0)$ in the money market account. Now if the stock price changes to x, *and we do not change our position*, then the value of our portfolio will be

$$\begin{aligned} c(t,x) - \partial_x c(t,x_0)x + M &= c(t,x) - x\partial_x c(t,x_0) + x_0\partial_x c(t,x_0) - c(t,x_0) \\ &= c(t,x) - \big(c(t,x_0) + (x-x_0)\partial_x c(t,x_0)\big)\,. \end{aligned}$$

Note that the line $y = c(t,x_0) + (x-x_0)\partial_x c(t,x_0)$ is the equation for the tangent to the curve $y = c(t,x)$ at the point $(x_0, c(t,x_0))$. For this reason the above portfolio is called *delta neutral.*

Note that any convex function lies entirely above its tangent. Thus, under instantaneous changes of the stock price (both rises and falls), we will have

$$c(t,x) - \partial_x c(t,x_0)x + M > 0\,, \quad \text{both for } x > x_0 \text{ and } x < x_0.$$

For this reason the above portfolio is called *long gamma.*

Note, even though under instantaneous price changes the value of our portfolio always rises, this is *not* an arbitrage opportunity. The reason for this is that as time increases c decreases since $\partial_t c < 0$. The above instantaneous argument assumed c is constant in time, which it most certainly is not!

9. Multi-dimensional Itô calculus.

Finally we conclude this chapter by studying Itô calculus in higher dimensions. Let X, Y be Itô process. We typically expect X, Y will have finite and non-zero quadratic variation, and hence both the increments $X(t+\delta t) - X(t)$ and $Y(t+\delta t) - Y(t)$ should typically be of size $\sqrt{\delta}$. If we multiply these and sum over some finite interval $[0,T]$, then we would have roughly $T/\delta t$ terms each of size δt, and expect that this converges as $\delta t \to 0$. The limit is called the joint quadratic variation.

DEFINITION 9.1. Let X and Y be two Itô processes. We define the *joint quadratic variation* of X, Y, denoted by $[X,Y]$ by

$$[X,Y](T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} (X(t_{i+1}) - X(t_i))(Y(t_{i+1}) - Y(t_i)),$$

where $P = \{0 = t_1 < t_1 \cdots < t_n = T\}$ is a partition of $[0,T]$.

Using the identity

$$4ab = (a+b)^2 - (a-b)^2$$

we quickly see that

$$[X,Y] = \frac{1}{4}\big([X+Y, X+Y] - [X-Y, X-Y]\big). \tag{9.1}$$

Using this and the properties we already know about quadratic variation, we can quickly deduce the following.

PROPOSITION 9.2 (Product rule). *If X and Y are two Itô processes then*

$$d(XY) = X\,dY + Y\,dX + d[X,Y]. \tag{9.2}$$

PROOF. By Itô's formula

$$\begin{aligned} d(X+Y)^2 &= 2(X+Y)\,d(X+Y) + d[X+Y, X+Y] \\ &= 2X\,dX + 2Y\,dY + 2X\,dY + 2Y\,dX + d[X+Y, X+Y]. \end{aligned}$$

Similarly

$$d(X-Y)^2 = 2X\,dX + 2Y\,dY - 2X\,dY - 2Y\,dX + d[X-Y, X-Y].$$

Since

$$4\,d(XY) = d(X+Y)^2 - d(X-Y)^2,$$

we obtain (9.2) as desired. ☐

As with quadratic variation, processes of finite variation do not affect the joint quadratic variation.

PROPOSITION 9.3. *If X is and Itô process, and B is a continuous adapted process with finite variation, then $[X,B] = 0$.*

PROOF. Note $[X \pm B, X \pm B] = [X,X]$ and hence $[X,B] = 0$. ☐

With this, we can state the higher dimensional Itô formula. Like the one dimensional Itô formula, this is a generalization of the chain rule and has an extra correction term that involves the joint quadratic variation.

THEOREM 9.4 (Itô-Doeblin formula). *Let $X_1, \ldots, X_n$ be n Itô processes and set $X = (X_1, \ldots, X_n)$. Let $f\colon [0,\infty) \times \mathbb{R}^n$ be C^1 in the first variable, and C^2 in the remaining variables. Then*

$$\begin{aligned} f(T, X(T)) = f(0, X(0)) + \int_0^T \partial_t f(t, X(t))\,dt + \sum_{i=1}^N \int_0^T \partial_i f(t, X(t))\,dX_i(t) \\ + \frac{1}{2} \sum_{i,j=1}^N \int_0^T \partial_i \partial_j f(t, X(t))\,d[X_i, X_j](t), \end{aligned}$$

REMARK 9.5. Here we think of $f = f(t, x_1, \dots, x_n)$, often abbreviated as $f(t,x)$. The $\partial_i f$ appearing in the Itô formula above is the partial derivative of f with respect to x_i. As before, the $\partial_t f$ and $\partial_i f$ terms above are from the usual chain rule, and the last term is the extra Itô correction.

REMARK 9.6. In differential form Itô's formula says

$$d\big(f(t, X(t)\big) = \partial_t f(t, X(t))\,dt + \sum_{i=1}^n \partial_i f(t, X(t))\,dX_i(t) + \frac{1}{2}\sum_{i,j=1}^n \partial_i\partial_j f(t, X(t))\,d[X_i, X_j](t)\,.$$

For compactness, we will often omit the $(t, X(t))$ and write the above as

$$d\big(f(t, X(t)\big) = \partial_t f\,dt + \sum_{i=1}^n \partial_i f\,dX_i(t) + \frac{1}{2}\sum_{i,j=1}^n \partial_i\partial_j f\,d[X_i, X_j](t)\,.$$

REMARK 9.7. We will most often use this in two dimensions. In this case, writing X and Y for the two processes, the Itô formula reduces to

$$d\big(f(t, X(t), Y(t))\big) = \partial_t f\,dt + \partial_x f\,dX(t) + \partial_y f\,dY(t) + \frac{1}{2}\big(\partial_x^2 f\,d[X,X](t) + 2\partial_x\partial_y f\,d[X,Y](t) + \partial_y^2 f\,d[Y,Y](t)\big)\,.$$

INTUITION BEHIND THE ITÔ FORMULA. Let's assume we only have two Itô processes X, Y and $f = f(x,y)$ doesn't depend on t. Let $P = \{0 = t_0 < t_1 \cdots < t_m = T\}$ be a partition of the interval $[0,T]$ and write

$$f(X(T), Y(T)) - f(X(0), Y(0)) = \sum_{i=0}^{m-1} f(\xi_{i+1}) - f(\xi_i)\,,$$

where we write $\xi_i = (X(t_i), Y(t_i))$ for compactness. Now by Taylor's theorem,

$$f(\xi_{i+1}) - f(\xi_i) = \partial_x f(\xi_i)\,\Delta_i X + \partial_y f(\xi_i)\,\Delta_i Y + \frac{1}{2}\Big(\partial_x^2 f(\xi_i)(\Delta_i X)^2 + 2\partial_x\partial_y f(\xi_i)\,\Delta_i X\,\Delta_i Y + \partial_y^2 f(\xi_i)(\Delta_i Y)^2\Big) + \text{higher order terms.}$$

Here $\Delta_i X = X(t_{i+1}) - X(t_i)$ and $\Delta_i Y = Y(t_{i+1}) - Y(t_i)$. Summing over i, the first two terms converge to $\int_0^T \partial_x f(t)\,dX(t)$ and $\int_0^T \partial_y f(t)\,dY(t)$ respectively. The terms involving $(\Delta_i X)^2$ should to $\int_0^T \partial_x^2 f\,d[X,X]$ as we had with the one dimensional Itô formula. Similarly, the terms involving $(\Delta_i Y)^2$ should to $\int_0^T \partial_y^2 f\,d[Y,Y]$ as we had with the one dimensional Itô formula. For the cross term, we can use the identity (9.1) and quickly check that it converges to $\int_0^T \partial_x\partial_y f\,d[X,Y]$. The higher order terms are typically of size $(t_{i+1} - t_i)^{3/2}$ and will vanish as $\|P\| \to 0$. ☐

The most common use of the multi-dimensional Itô formula is when the Itô processes are specified as a combination of Itô integrals with respect to different Brownian motions. Thus our next goal is to find an effective way to to compute the joint quadratic variations in this case.

We've seen earlier (Theorems 3.4–3.5) that the quadratic variation of a martingale M is the unique increasing process that make $M^2 - [M,M]$ a martingale. A similar result holds for the joint quadratic variation.

PROPOSITION 9.8. *Suppose M, N are two continuous martingales with respect to a common filtration $\{\mathcal{F}_t\}$ such that $\boldsymbol{E}M(t)^2, \boldsymbol{E}N(t)^2 < \infty$.*

(1) The process $MN - [M,N]$ is also a martingale with respect to the same filtration.
(2) Moreover, if A is any continuous adapted process with finite first variation such that $A(0) = 0$ and $MN - A$ is a martingale with respect to $\{\mathcal{F}_t\}$, then $A = [M,N]$.

PROOF. The first part follows immediately from Theorem 3.4 and the fact that

$$4\big(MN - [M,N]\big) = (M+N)^2 - [M+N, M+N] - \big((M-N)^2 - [M-N, M-N]\big)\,.$$

The second part follows from the first part and uniqueness of the semi-martingale decomposition (Proposition 5.5). ☐

PROPOSITION 9.9 (Bi-linearity). *If X, Y, Z are three Itô processes and $\alpha \in \mathbb{R}$ is a (non-random) constant, then*

$$[X, Y + \alpha Z] = [X, Y] + \alpha[X, Z]\,. \tag{9.3}$$

PROOF. Let L, M and N be the martingale part in the Itô decomposition of X, Y and Z respectively. Clearly

$$L(M + \alpha N) - \big([L, M] + \alpha[L, N]\big) = \big(LM - [L, M]\big) + \alpha\big(LN - [L, N]\big)\,,$$

which is a martingale. Thus, since $[L, M] + \alpha[L, N]$ is also continuous adapted and increasing, by Proposition 9.8 we must have $[L, M + \alpha N] = [L, M] + \alpha[L, N]$. Since the joint quadratic variation of Itô processes can be computed in terms of their martingale parts alone, we obtain (9.3) as desired. ☐

For integrals with respect to Itô processes, we can compute the joint quadratic variation explicitly.

PROPOSITION 9.10. *Let X_1, X_2 be two Itô processes, σ_1, σ_2 be two adapted processes and let I_j be the integral defined by $I_j(t) = \int_0^t \sigma_j(s)\, dX_j(s)$ for $j \in \{1, 2\}$. Then*

$$[I_1, I_2](t) = \int_0^t \sigma_1(s)\sigma_2(s)\, d[X_1, X_2](s)\,.$$

PROOF. Let P be a partition and, as above, let $\Delta_i X = X(t_{i+1}) - X(t_i)$ denote the increment of a process X. Since

$$I_j(T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_j(t_i)\Delta_i X_j\,, \quad \text{and} \quad [X_1, X_2] = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta_i X_1\, \Delta_i X_2\,,$$

we expect that $\sigma_j(t_i)\Delta_i(X_j)$ is a good approximation for $\Delta_i I_j$, and $\Delta_i X_1\, \Delta_i X_2$ is a good approximation for $\Delta_i[X_1, X_2]$. Consequently, we expect

$$[I_i, I_j](T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta_i I_1\, \Delta_i I_2 = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_1(t_i)\Delta_i X_1\, \sigma_2(t_i)\Delta_i X_2$$

$$= \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_1(t_i)\,\sigma_2(t_i)\Delta_i[X_1, X_2] = \int_0^T \sigma_1(t)\sigma_2(t)\,d[X_1, X_2](t)\,,$$

as desired. □

PROPOSITION 9.11. *Let M, N be two continuous martingales with respect to a common filtration $\{\mathcal{F}_t\}$ such that $\boldsymbol{E}M(t)^2 < \infty$ and $\boldsymbol{E}N(t)^2 < \infty$. If M, N are independent, then $[M, N] = 0$.*

REMARK 9.12. If X and Y are independent, we know $\boldsymbol{E}XY = \boldsymbol{E}X\boldsymbol{E}Y$. However, we *need not* have $\boldsymbol{E}(XY \mid \mathcal{F}) = \boldsymbol{E}(X \mid F)\boldsymbol{E}(Y \mid F)$. So we can not prove the above result by simply saying

$$\boldsymbol{E}\big(M(t)N(t) \bigm| \mathcal{F}_s\big) = \boldsymbol{E}(M(t) \mid \mathcal{F}_s)\boldsymbol{E}(N(t) \mid \mathcal{F}_s) = M(s)N(s) \tag{9.4}$$

because M and N are independent. Thus MN is a martingale, and hence $[M, N] = 0$ by Proposition 9.8.

This reasoning is incorrect, even though the conclusion is correct. If you're not convinced, let me add that there exist martingales that are *not continuous* which are independent and have nonzero joint quadratic variation. The above argument, if correct, would certainly also work for martingales that are not continuous. The error in the argument is that the first equality in (9.4) need not hold even though M and N are independent.

PROOF. Let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0, T]$, $\Delta_i M = M(t_{i+1}) - M(t_i)$ and $\Delta_i N = N(t_{i+1}) - N(t_i)$. Observe

$$\boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 = \boldsymbol{E}\sum_{i=0}^{n-1} (\Delta_i M)^2 (\Delta_i N)^2 + 2\boldsymbol{E}\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \Delta_i M\, \Delta_i N \Delta_j M\, \Delta_j N\,. \tag{9.5}$$

We claim the cross term vanishes because of independence of M and N. Indeed,

$$\begin{aligned}\boldsymbol{E}\Delta_i M\, \Delta_i N \Delta_j M\, \Delta_j N &= \boldsymbol{E}\big(\Delta_i M\, \Delta_j M\big)\boldsymbol{E}\big(\Delta_i N\, \Delta_j N\big)\\ &= \boldsymbol{E}\big(\Delta_i M \boldsymbol{E}(\Delta_j M \mid \mathcal{F}_{t_j})\big)\boldsymbol{E}\big(\Delta_i N\, \Delta_j N\big) = 0\,.\end{aligned}$$

Thus from (9.5)

$$\boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 = \boldsymbol{E}\sum_{i=0}^{n-1} (\Delta_i M)^2 (\Delta_i N)^2 \leqslant \boldsymbol{E}\Big(\max_i (\Delta_i M)^2\Big)\Big(\sum_{i=0}^{n-1} (\Delta_i N)^2\Big)$$

As $\|P\| \to 0$, $\max_i \Delta_i M \to 0$ because M is continuous, and $\sum_i (\Delta_i N)^2 \to [N, N](T)$. Thus we expect[6]

$$\boldsymbol{E}[M, N](T)^2 = \lim_{\|P\|\to 0} \boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 = 0\,,$$

finishing the proof. □

[6]For this step we need to use $\lim_{\|P\|\to 0} \boldsymbol{E}(\cdots) = \boldsymbol{E}\lim_{\|P\|\to 0}(\cdots)$. To make this rigorous we need to apply the Lebesgue dominated convergence theorem. This is done by first assuming M and N are bounded, and then choosing a *localizing sequence* of stopping times, and a full discussion goes beyond the scope of these notes.

REMARK 9.13. The converse is false. If $[M, N] = 0$, it does *not* mean that M and N are independent. For example, if

$$M(t) = \int_0^t \mathbf{1}_{\{W(s)<0\}} \, dW(s) \,, \qquad \text{and} \qquad N(t) = \int_0^t \mathbf{1}_{\{W(s)\geqslant 0\}} \, dW(s) \,,$$

then clearly $[M, N] = 0$. However,

$$M(t) + N(t) = \int_0^t 1 \, dW(s) = W(t) \,,$$

and with a little work one can show that M and N are not independent.

DEFINITION 9.14. We say $W = (W_1, W_2, \dots, W_d)$ is a standard d-dimensional Brownian motion if:

(1) Each coordinate W_i is a standard (1-dimensional) Brownian motion.
(2) If $i \neq j$, the processes W_i and W_j are independent.

When working with a multi-dimensional Brownian motion, we usually choose the filtration to be that generated by *all* the coordinates.

DEFINITION 9.15. Let W be a d-dimensional Brownian motion. We define the filtration $\{\mathcal{F}_t^W\}$ by

$$\mathcal{F}_t^W = \sigma\Big(\bigcup_{\substack{s \leqslant t, \\ i \in \{1, \dots, d\}}} \sigma(W_i(s)) \Big)$$

With $\{\mathcal{F}_t^W\}$ defined above note that:

(1) Each coordinate W_i is a martingale with respect to $\{\mathcal{F}_t^W\}$.
(2) For every $s < t$, the increment of each coordinate $W_i(t) - W_i(s)$ is independent of $\{\mathcal{F}_s^W\}$.

REMARK 9.16. Since W_i is independent of W_j when $i \neq j$, we know $[W_i, W_j] = 0$ if $i \neq j$. When $i = j$, we know $d[W_i, W_j] = dt$. We often express this concisely as

$$d[W_i, W_j](t) = \mathbf{1}_{\{i=j\}} \, dt \,.$$

An extremely important fact about Brownian motion is that the converse of the above is also true.

THEOREM 9.17 (Lévy). *If $M = (M_1, M_2, \dots, M_d)$ is a continuous martingale such that $M(0) = 0$ and*

$$d[M_i, M_j](t) = \mathbf{1}_{\{i=j\}} \, dt \,,$$

then M is a d-dimensional Brownian motion.

PROOF. The main idea behind the proof is to compute the moment generating function (or characteristic function) of M, in the same way as in Problem 7.5. This can be used to show that $M(t) - M(s)$ is independent of $\mathcal{F}_s$ and $M(t) \sim N(0, tI)$, where I is the $d \times d$ identity matrix. □

EXAMPLE 9.18. If W is a 2-dimensional Brownian motion, then show that

$$B = \int_0^t \frac{W_1(s)}{|W(t)|} \, dW_1(s) + \int_0^t \frac{W_2(s)}{|W(t)|} \, dW_2(s) \,,$$

is also a Brownian motion.

PROOF. Since B is the sum of two Itô integrals, it is clearly a continuous martingale. Thus to show that B is a Brownian motion, it suffices to show that $[B,B](t) = t$. For this, define

$$X(t) = \int_0^t \frac{W_1(s)}{|W(t)|}\, dW_1(s) \qquad \text{and} \qquad Y(t) = \int_0^t \frac{W_2(s)}{|W(t)|}\, dW_2(s)\,,$$

and note

$$\begin{aligned} d[B,B](t) = d[X+Y, X+Y](t) &= d[X,X](t) + d[Y,Y](t) + 2d[X,Y](t) \\ &= \Big(\frac{W_1(t)^2}{|W(t)|^2} + \frac{W_2(t)^2}{|W(t)|^2}\Big)\, dt + 0 = dt\,. \end{aligned}$$

So by Lévy's criterion, B is a Brownian motion. □

EXAMPLE 9.19. Let W be a 2-dimensional Brownian motion and define

$$X = \ln(|W|^2) = \ln(W_1^2 + W_2^2)\,.$$

Compute dX. Is X a martingale?

SOLUTION. This is a bit tricky. First, if we set $f(x) = \ln|x|^2 = \ln(x_1^2 + x_2^2)$, then it is easy to check

$$\partial_i f = \frac{2x_i}{|x|^2} \qquad \text{and} \qquad \partial_1^2 f + \partial_2^2 f = 0\,.$$

Consequently,

$$dX(t) = \frac{2W_1(t)}{|W|^2}\, dW_1(t) + \frac{2W_2(t)}{|W|^2}\, dW_2(t)\,.$$

With this one would be tempted to say that since there are no dt terms above, X is a martingale. This, however, is false! Martingales have constant expectation, but

$$\begin{aligned} \boldsymbol{E}X(t) &= \frac{1}{2\pi t} \iint_{\mathbb{R}^2} \ln\big(x_1^2 + x_2^2\big) \exp\Big(-\frac{x_1^2 + x_2^2}{2t}\Big)\, dx_1\, dx_2 \\ &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \ln\big(t(y_1^2 + y_2^2)\big) \exp\Big(-\frac{y_1^2 + y_2^2}{2}\Big)\, dy_1\, dy_2 \\ &= \ln t + \frac{1}{2\pi} \iint_{\mathbb{R}^2} \ln\big(y_1^2 + y_2^2\big) \exp\Big(-\frac{y_1^2 + y_2^2}{2}\Big)\, dy_1\, dy_2 \xrightarrow{t\to\infty} \infty\,. \end{aligned}$$

Thus $\boldsymbol{E}X(t)$ is not constant in t, and so X can not be a martingale. □

REMARK 9.20. We have repeatedly used the fact that Itô integrals are martingales. The example above obtains X as an Itô integral, but can not be a martingale. The reason this doesn't contradict Theorem 4.2 is that in order for Itô integral $\int_0^t \sigma(s)\, dW(s)$ to be defined, we only need the finiteness condition $\int_0^t \sigma(s)^2\, ds < \infty$ almost surely. However, for an Itô integral to be a martingale, we need the stronger condition $\boldsymbol{E}\int_0^t \sigma(s)^2\, ds < \infty$ (given in (4.5)) to hold. This is precisely what fails in the previous example. The process X above is an example of a *local martingale* that is not a martingale, and we will encounter a similar situation when we study exponential martingales and risk neutral measures.

EXAMPLE 9.21. Let $f = f(t, x_1, \dots, x_d) \in C^{1,2}$ and W be a d-dimensional Brownian motion. Then Itô's formula gives

$$d\big(f(t, W(t))\big) = \Big(\partial_t f(t, W(t)) + \frac{1}{2}\Delta f(t, W(t))\Big)\,dt + \sum_{i=1}^{d} \partial_i f(t, W(t))\,dW_i(t)\,.$$

Here $\Delta f = \sum_1^d \partial_i^2 f$ is the Laplacian of f.

EXAMPLE 9.22. Consider a d-dimensional Brownian motion W, and n Itô processes $X_1, \dots, X_n$ which we write (in differential form) as

$$dX_i(t) = b_i(t)\,dt + \sum_{k=1}^{d} \sigma_{i,k}(t)\,dW_k(t)\,,$$

where each b_i and $\sigma_{i,j}$ are adapted processes. For brevity, we will often write b for the vector process $(b_1, \dots, b_n)$, σ for the *matrix* process $(\sigma_{i,j})$ and X for the n-dimensional Itô process $(X_1, \dots, X_n)$.

Now to compute $[X_i, X_j]$ we observe that $d[W_i, W_j] = dt$ if $i = j$ and 0 otherwise. Consequently,

$$d[X_i, X_j](t) = \sum_{k,l=1}^{d} \sigma_{i,k}\sigma_{j,l}\mathbf{1}_{\{k=l\}}\,dt = \sum_{k=1}^{d} \sigma_{i,k}(t)\sigma_{j,k}(t)\,dt\,.$$

Thus if f is any $C^{1,2}$ function, Itô formula gives

$$d\Big(f(t, X(t))\Big) = \Big(\partial_t f + \sum_{i=1}^{n} b_i \partial_i f\Big)\,dt + \sum_{i=1}^{n} \sigma_i \partial_i f\,dW_i(t) + \frac{1}{2}\sum_{i,j=1}^{n} a_{i,j}\partial_i\partial_j f\,dt$$

where

$$a_{i,j}(t) = \sum_{k=1}^{N} \sigma_{i,k}(t)\sigma_{j,k}(t)\,.$$

In matrix notation, the matrix $a = \sigma\sigma^T$, where σ^T is the transpose of the matrix σ.

CHAPTER 4

Risk Neutral Measures

Our aim in this section is to show how risk neutral measures can be used to price derivative securities. The key advantage is that under a risk neutral measure the discounted hedging portfolio becomes a martingale. Thus the price of any derivative security can be computed by conditioning the payoff at maturity. We will use this to provide an elegant derivation of the Black-Scholes formula, and discuss the fundamental theorems of asset pricing.

1. The Girsanov Theorem.

DEFINITION 1.1. Two probability measures $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ are said to be equivalent if for every event A, $\boldsymbol{P}(A) = 0$ if and only if $\tilde{\boldsymbol{P}}(A) = 0$.

EXAMPLE 1.2. Let Z be a random variable such that $\boldsymbol{E}Z = 1$ and $Z > 0$. Define a new measure $\tilde{\boldsymbol{P}}$ by

$$\tilde{\boldsymbol{P}}(A) = \boldsymbol{E}Z\mathbf{1}_A = \int_A Z\,d\boldsymbol{P}\,. \tag{1.1}$$

for every event A. Then $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ are equivalent.

REMARK 1.3. The assumption $\boldsymbol{E}Z = 1$ above is required to guarantee $\tilde{\boldsymbol{P}}(\Omega) = 1$.

DEFINITION 1.4. When $\tilde{\boldsymbol{P}}$ is defined by (1.1), we say

$$d\,\tilde{\boldsymbol{P}} = Z\,d\boldsymbol{P} \qquad \text{or} \qquad Z = \frac{d\tilde{\boldsymbol{P}}}{d\boldsymbol{P}}\,,$$

and Z is called the *density* of $\tilde{\boldsymbol{P}}$ with respect to $\boldsymbol{P}$.

THEOREM 1.5 (Radon-Nikodym). *Two measures $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ are equivalent if and only if there exists a random variable Z such that $\boldsymbol{E}Z = 1$, $Z > 0$ and $\tilde{\boldsymbol{P}}$ is given by* (1.1).

The proof of this requires a fair amount of machinery from Lebesgue integration and goes beyond the scope of these notes. (This is exactly the result that is used to show that conditional expectations exist.) However, when it comes to risk neutral measures, it isn't essential since in most of our applications the density will be explicitly chosen.

Suppose now $T > 0$ is fixed, and Z is a martingale. Define a new measure $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}_T$ by

$$d\tilde{\boldsymbol{P}} = d\tilde{\boldsymbol{P}}_T = Z(T)d\boldsymbol{P}\,.$$

We will denote expectations and conditional expectations with respect to the new measure by $\tilde{\boldsymbol{E}}$. That is, given a random variable X,

$$\tilde{\boldsymbol{E}}X = \boldsymbol{E}Z(T)X = \int Z(T)X\,d\boldsymbol{P}\,.$$

Also, given a σ-algebra $\mathcal{F}$, $\tilde{\boldsymbol{E}}(X \mid \mathcal{F})$ is the unique $\mathcal{F}$-measurable random variable such that

$$\int_F \tilde{\boldsymbol{E}}(X \mid \mathcal{F})\, d\tilde{\boldsymbol{P}} = \int_F X\, d\tilde{\boldsymbol{P}}\,, \tag{1.2}$$

holds for all $\mathcal{F}$ measurable events F. Of course, equation (1.2) is equivalent to requiring

$$\int_F Z(T)\tilde{\boldsymbol{E}}(X \mid \mathcal{F})\, d\boldsymbol{P} = \int_F Z(T) X\, d\boldsymbol{P}\,, \tag{1.2$'$}$$

for all $\mathcal{F}$ measurable events F.

The main goal of this section is to prove the Girsanov theorem.

THEOREM 1.6 (Cameron, Martin, Girsanov). *Let $b(t) = (b_1(t), b_2(t), \dots, b_d(t))$ be a d-dimensional adapted process, W be a d-dimensional Brownian motion, and define*

$$\tilde{W}(t) = W(t) + \int_0^t b(s)\, ds\,.$$

Let Z be the process defined by

$$Z(t) = \exp\Big(-\int_0^t b(s) \cdot dW(s) - \frac{1}{2}\int_0^t |b(s)|^2\, ds\Big)\,,$$

and define a new measure $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}_T$ by $d\tilde{\boldsymbol{P}} = Z(T)\, d\boldsymbol{P}$. If Z is a martingale then $\tilde{W}$ is a Brownian motion under the measure $\tilde{\boldsymbol{P}}$ up to time T.

REMARK 1.7. Above

$$b(s) \cdot dW(s) \stackrel{\text{def}}{=} \sum_{i=1}^d b_i(s)\, dW_i(s) \qquad \text{and} \qquad |b(s)|^2 = \sum_{i=1}^d b_i(s)^2\,.$$

REMARK 1.8. Note $Z(0) = 1$, and if Z is a martingale then $\boldsymbol{E}Z(T) = 1$ ensuring $\tilde{\boldsymbol{P}}$ is a probability measure. You might, however, be puzzled at need for the assumption that Z is a martingale. Indeed, let $M(t) = \int_0^t b(s) \cdot dW(s)$, and $f(t,x) = \exp(-x - \frac{1}{2}\int_0^t b(s)^2\, ds)$. Then, by Itô's formula,

$$\begin{aligned} dZ(t) = d\big(f(t, M(t))\big) &= \partial_t f\, dt + \partial_x f\, dM(t) + \frac{1}{2}\partial_x^2 f\, d[M,M](t) \\ &= -\frac{1}{2}Z(t)|b(t)|^2\, dt - Z(t)b(t) \cdot dW(t) + \frac{1}{2}Z(t)|b(t)|^2\, dt\,, \end{aligned}$$

and hence

$$dZ(t) = -Z(t)b(t) \cdot dW(t)\,. \tag{1.3}$$

Thus you might be tempted to say that Z is always a martingale, assuming it explicitly is unnecessary. However, we recall from Chapter 3, Theorem 4.2 in that Itô integrals are only guaranteed to be martingales under the square integrability condition

$$\boldsymbol{E}\int_0^T |Z(s)b(s)|^2\, ds < \infty\,. \tag{1.4}$$

Without this finiteness condition, Itô integrals are only *local martingales*, whose expectation need not be constant, and so $\boldsymbol{E}Z(T) = 1$ is not guaranteed. Indeed, there are many examples of processes b where the finiteness condition (1.4) does *not* hold and we have $\boldsymbol{E}Z(T) < 1$ for some $T > 0$.

REMARK 1.9. In general the process Z above is always a super-martingale, and hence $\boldsymbol{E}Z(T) \leqslant 1$. Two conditions that guarantee Z is a martingale are the Novikov and Kazamaki conditions: If either

$$\boldsymbol{E}\exp\Big(\frac{1}{2}\int_0^t |b(s)|^2\,ds\Big) < \infty \quad \text{or} \quad \boldsymbol{E}\exp\Big(\frac{1}{2}\int_0^t b(s)\cdot dW(s)\Big) < \infty\,,$$

then Z is a martingale and hence $\boldsymbol{E}Z(T) = 1$ for all $T > 0$. Unfortunately, in many practical situations these conditions do not apply, and you have to show Z is a martingale by hand.

REMARK 1.10. The components $b_1, \ldots, b_d$ of the process b are not required to be independent. Yet, under the new measure, the process $\tilde{W}$ is a Brownian motion and hence has *independent* components.

The main idea behind the proof of the Girsanov theorem is the following: Clearly $[\tilde{W}_i, \tilde{W}_j] = [W_i, W_j] = \mathbf{1}_{i=j}t$. Thus if we can show that $\tilde{W}$ is a martingale *with respect to the new measure* $\tilde{\boldsymbol{P}}$, then Lévy's criterion will guarantee $\tilde{W}$ is a Brownian motion. We now develop the tools required to check when processes are martingales under the new measure.

LEMMA 1.11. *Let $0 \leqslant s \leqslant t \leqslant T$. If X is a $\mathcal{F}_t$-measurable random variable then*

$$\tilde{\boldsymbol{E}}\big(X \mid \mathcal{F}_s\big) = \frac{1}{Z(s)}\boldsymbol{E}\big(Z(t)X \mid \mathcal{F}_s\big) \tag{1.5}$$

PROOF. Let $A \in \mathcal{F}_s$ and observe that

$$\begin{aligned}\int_A \tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s)\,d\tilde{\boldsymbol{P}} &= \int_A Z(T)\tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s)\,d\boldsymbol{P}\\ &= \int_A \boldsymbol{E}\big(Z(T)\tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s) \mid \mathcal{F}_s\big)\,d\boldsymbol{P} = \int_A Z(s)\tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s)\,d\boldsymbol{P}\,.\end{aligned}$$

Also,

$$\begin{aligned}\int_A \tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s)\,d\tilde{\boldsymbol{P}} = \int_A X\,d\tilde{\boldsymbol{P}} = \int_A XZ(T)\,d\boldsymbol{P} &= \int_A \boldsymbol{E}\big(XZ(T) \mid \mathcal{F}_t\big)\,d\boldsymbol{P}\\ &= \int_A Z(t)X\,d\boldsymbol{P} = \int_A \boldsymbol{E}\big(Z(t)X \mid \mathcal{F}_s\big)\,d\boldsymbol{P}\end{aligned}$$

Thus

$$\int_A Z(s)\tilde{\boldsymbol{E}}(X \mid \mathcal{F}_s)\,d\boldsymbol{P} = \int_A \boldsymbol{E}\big(Z(t)X \mid \mathcal{F}_s\big)\,d\boldsymbol{P}\,,$$

for every $\mathcal{F}_s$ measurable event A. Since the integrands are both $\mathcal{F}_s$ measurable this forces them to be equal, giving (1.5) as desired. ∎

LEMMA 1.12. *An adapted process M is a martingale under $\tilde{\boldsymbol{P}}$ if and only if MZ is a martingale under $\boldsymbol{P}$.*

PROOF. Suppose first MZ is a martingale with respect to $\boldsymbol{P}$. Then

$$\tilde{\boldsymbol{E}}(M(t) \mid \mathcal{F}_s) = \frac{1}{Z(s)}\boldsymbol{E}(Z(t)M(t) \mid \mathcal{F}_s) = \frac{1}{Z(s)}Z(s)M(s) = M(s)\,,$$

showing M is a martingale with respect to $\boldsymbol{P}$.

Conversely, suppose M is a martingale with respect to $\tilde{\boldsymbol{P}}$. Then

$$\boldsymbol{E}\big(M(t)Z(t) \bigm| \mathcal{F}_s\big) = Z(s)\tilde{\boldsymbol{E}}(M(t) \mid \mathcal{F}_s) = Z(s)M(s)\,,$$

and hence ZM is a martingale with respect to $\boldsymbol{P}$. □

Proof of Theorem 1.6. Clearly $\tilde{W}$ is continuous and

$$d[\tilde{W}_i, \tilde{W}_j](t) = d[W_i, W_j](t) = \mathbf{1}_{i=j}\,dt\,.$$

Thus if we show that each $\tilde{W}_i$ is a martingale (under $\tilde{\boldsymbol{P}}$), then by Lévy's criterion, $\tilde{W}$ will be a Brownian motion under $\tilde{\boldsymbol{P}}$.

We now show that each $\tilde{W}_i$ is a martingale under $\tilde{\boldsymbol{P}}$. By Lemma 1.12, $\tilde{W}_i$ is a martingale under $\tilde{\boldsymbol{P}}$ if and only if $Z\tilde{W}_i$ is a martingale under $\boldsymbol{P}$. To show $Z\tilde{W}_i$ is a martingale under $\boldsymbol{P}$, we use the product rule and (1.3) to compute

$$\begin{aligned} d\big(Z\tilde{W}_i\big) &= Z\,d\tilde{W}_i + \tilde{W}_i\,dZ + d[Z, \tilde{W}_i] \\ &= Z\,dW_i + Zb_i\,dt - \tilde{W}_i Zb \cdot dW - b_i Z\,dt = Z\,dW_i - \tilde{W}_i Zb \cdot dW\,. \end{aligned}$$

Thus $Z\tilde{W}_i$ is a martingale[1] under $\boldsymbol{P}$, and by Lemma 1.12, $\tilde{W}_i$ is a martingale under $\tilde{\boldsymbol{P}}$. This finishes the proof. □

2. Risk Neutral Pricing

Consider a stock whose price is modelled by a generalized geometric Brownian motion

$$dS(t) = \alpha(t)S(t)\,dt + \sigma(t)S(t)\,dW(t)\,,$$

where $\alpha(t)$, $\sigma(t)$ are the (time dependent) mean return rate and volatility respectively. Here α and σ are no longer constant, but allowed to be *adapted processes*. We will, however, assume $\sigma(t) > 0$.

Suppose an investor has access to a money market account with variable interest rate $R(t)$. Again, the interest rate R need not be constant, and is allowed to be any adapted process. Define the *discount process* D by

$$D(t) = \exp\Big(-\int_0^t R(s)\,ds\Big)\,,$$

and observe

$$dD(t) = -D(t)R(t)\,dt\,.$$

Since the price of one share of the money market account at time t is $1/D(t)$ times the price of one share at time 0, it is natural to consider the *discounted stock price* DS.

Definition 2.1. A *risk neutral measure* is a measure $\tilde{\boldsymbol{P}}$ that is equivalent to $\boldsymbol{P}$ under which the discounted stock price process $D(t)S(t)$ is a martingale.

Remark 2.2. It turns out that the *existence* of a risk neutral measure is equivalent to there being no arbitrage opportunity in the market. Moreover, *uniqueness* of a risk neutral measure is equivalent to both the absence of an arbitrage opportunity, and that every derivative security can be hedged. These are the *fundamental theorems of asset pricing*.

[1] Technically, we have to check the square integrability condition to ensure that $Z\tilde{W}_i$ is a martingale, and not a *local martingale*. This, however, follows quickly from the Cauchy-Schwartz inequality and our assumption.

Using the Girsanov theorem, we can compute the risk neutral measure explicitly. Observe

$$\begin{aligned} d\big(D(t)S(t)\big) &= -RDS\,dt + D\,dS = (\alpha - R)DS\,dt + DS\sigma\,dW(t) \\ &= \sigma(t)D(t)S(t)\Big(\theta(t)\,dt + dW(t)\Big) \end{aligned}$$

where

$$\theta(t) \stackrel{\text{def}}{=} \frac{\alpha(t) - R(t)}{\sigma(t)}$$

is known as the *market price of risk.*

Define a new process $\tilde{W}$ by

$$d\tilde{W}(t) = \theta(t)\,dt + dW(t)\,,$$

and observe

$$d\big(D(t)S(t)\big)\,dt = \sigma(t)D(t)S(t)\,d\tilde{W}(t)\,. \tag{2.1}$$

PROPOSITION 2.3. *If Z is the process defined by*

$$Z(t) = \exp\Big(-\int_0^t \theta(s)\,dW(s) - \frac{1}{2}\int_0^t \theta(s)^2\,ds\Big)\,,$$

then the measure $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}_T$ defined by $d\tilde{\boldsymbol{P}} = Z(T)\,d\boldsymbol{P}$ is a risk neutral measure.

PROOF. By the Girsanov theorem 1.6 we know $\tilde{W}$ is a Brownian motion under $\tilde{\boldsymbol{P}}$. Thus using (2.1) we immediately see that the discounted stock price is a martingale. □

Our next aim is to develop *risk neutral pricing formula.*

THEOREM 2.4 (Risk Neutral Pricing formula). *Let $V(T)$ be a $\mathcal{F}_T$-measurable random variable that represents the payoff of a derivative security, and let $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}_T$ be the risk neutral measure above. The arbitrage free price at time t of a derivative security with payoff $V(T)$ and maturity T is given by*

$$V(t) = \tilde{\boldsymbol{E}}\Big(\exp\Big(-\int_t^T R(s)\,ds\Big)V(T)\,\Big|\,\mathcal{F}_t\Big)\,. \tag{2.2}$$

REMARK 2.5. It is important to note that the price $V(t)$ above is the actual arbitrage free price of the security, and there is no alternate "risk neutral world" which you need to teleport to in order to apply this formula. The risk neutral measure is simply a tool that is used in the above formula, which gives the arbitrage free price under the *standard* measure.

As we will see shortly, the reason for this formula is that under the risk neutral measure, the discounted replicating portfolio becomes a martingale. To understand why this happens we note

$$dS(t) = \alpha(t)S(t)\,dt + \sigma(t)S(t)\,dW(t) = R(t)S(t)\,dt + \sigma(t)S(t)\,d\tilde{W}\,. \tag{2.3}$$

Under the standard measure $\boldsymbol{P}$ this isn't much use, since $\tilde{W}$ isn't a martingale. However, under the risk neutral measure, the process $\tilde{W}$ is a Brownian motion and hence certainly a martingale. Moreover, S becomes a geometric Brownian motion under $\tilde{\boldsymbol{P}}$ with mean return rate of S exactly the same as that of the money market account. The fact that S and the money market account have exactly the same mean return rate (under $\tilde{\boldsymbol{P}}$) is precisely what makes the replicating portfolio (or any self-financing portfolio for that matter) a martingale (under $\tilde{\boldsymbol{P}}$).

LEMMA 2.6. *Let Δ be any adapted process, and $X(t)$ be the wealth of an investor with that holds $\Delta(t)$ shares of the stock and the rest of his wealth in the money market account. If there is no external cash flow (i.e. the portfolio is* self financing*), then the discounted portfolio $D(t)X(t)$ is a martingale under $\tilde{\boldsymbol{P}}$.*

PROOF. We know

$$dX(t) = \Delta(t)\,dS(t) + R(t)(X(t) - \Delta(t)S(t))\,dt\,.$$

Using (2.3) this becomes

$$\begin{aligned} dX(t) &= \Delta RS\,dt + \Delta\sigma S\,d\tilde{W} + RX\,dt - R\Delta S\,dt \\ &= RX\,dt + \Delta\sigma S\,d\tilde{W}\,. \end{aligned}$$

Thus, by the product rule,

$$\begin{aligned} d(DX) &= D\,dX + X\,dD + d[D,X] = -RDX\,dt + DRXdt + D\Delta\sigma S\,d\tilde{W} \\ &= D\Delta\sigma S\,d\tilde{W}\,. \end{aligned}$$

Since $\tilde{W}$ is a martingale under $\tilde{\boldsymbol{P}}$, DX must be a martingale under $\tilde{\boldsymbol{P}}$. □

PROOF OF THEOREM 2.4. Suppose $X(t)$ is the wealth of a replicating portfolio at time t. Then by definition we know $V(t) = X(t)$, and by the previous lemma we know DX is a martingale under $\tilde{\boldsymbol{P}}$. Thus

$$V(t) = X(t) = \frac{1}{D(t)} D(t)X(t) = \frac{1}{D(t)} \tilde{\boldsymbol{E}}\big(D(T)X(T) \mid \mathcal{F}_t\big) = \tilde{\boldsymbol{E}}\Big(\frac{D(T)V(T)}{D(t)} \Bigm| \mathcal{F}_t\Big)\,,$$

which is precisely (2.2). □

REMARK 2.7. Our proof assumes that a security with payoff $V(T)$ *has* a replicating portfolio. This is true in general because of the *martingale representation theorem*, which guarantees any martingale (with respect to the Brownian filtration) can be expressed as an Itô integral with respect to Brownian motion. Recall, we already know that Itô integrals are martingales. The martingale representation theorem is a partial converse.

Now clearly the process Y defined by

$$Y(t) = \tilde{\boldsymbol{E}}\Big(\exp\Big(-\int_0^T R(s)\,ds\Big)V(T) \Bigm| \mathcal{F}_t\Big)\,,$$

is a martingale. Thus, the *martingale representation theorem* can be used to express this as an Itô integral (with respect to $\tilde{W}$). With a little algebraic manipulation one can show that $D(t)^{-1}Y(t)$ is the wealth of a self financing portfolio. Since the terminal wealth is clearly $V(T)$, this must be a replicating portfolio.

REMARK 2.8. If $V(T) = f(S(T))$ for some function f and R is not random, then the *Markov Property* guarantees $V(t) = c(t, S(t))$ for some function c. Equating $c(t, S(t)) = X(t)$, the wealth of a replicating portfolio and using Itô's formula, we immediately obtain the *Delta hedging rule*

$$\Delta(t) = \partial_x c(t, S(t))\,. \tag{2.4}$$

If, however, that if V is not of the form $f(S(T))$ for some function f, then the option price will in general depend on the *entire history* of the stock price, and not only the spot price $S(t)$. In this case we will not (in general) have the delta hedging rule (2.4).

3. The Black-Scholes formula

Recall our first derivation of the Black-Scholes formula only obtained a PDE. The Black-Scholes formula is the solution to this PDE, which we simply wrote down without motivation. Use the risk neutral pricing formula can be used to derive the Black-Scholes formula quickly, and independently of our previous derivation. We carry out his calculation in this section.

Suppose σ and R are deterministic constants, and for notational consistency, set $r = R$. The risk neutral pricing formula says that the price of a European call is

$$c(t, S(t)) = \tilde{\boldsymbol{E}}\Big(e^{-r(T-t)}(S(T) - K)^+ \Bigm| \mathcal{F}_t\Big),$$

where K is the strike price. Since

$$S(t) = S(0)\exp\Big(\Big(r - \frac{\sigma^2}{2}\Big)t + \sigma\tilde{W}(t)\Big),$$

and $\tilde{W}$ is a Brownian motion under $\tilde{\boldsymbol{P}}$, we see

$$\begin{aligned}
c(t, S(t)) &= e^{-r\tau}\tilde{\boldsymbol{E}}\Big(\Big[S(0)\exp\Big(\Big(r - \frac{\sigma^2}{2}\Big)T + \sigma\tilde{W}(T)\Big) - K\Big]^+ \Bigm| \mathcal{F}_t\Big) \\
&= e^{-r\tau}\tilde{\boldsymbol{E}}\Big(\Big[S(t)\exp\Big(\Big(r - \frac{\sigma^2}{2}\Big)\tau + \sigma(\tilde{W}(T) - \tilde{W}(t))\Big) - K\Big]^+ \Bigm| \mathcal{F}_t\Big) \\
&= \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{\mathbb{R}}\Big[S(t)\exp\Big(\Big(r - \frac{\sigma^2}{2}\Big)\tau + \sigma\sqrt{\tau}y\Big) - K\Big]^+ e^{-y^2/2}\,dy,
\end{aligned}$$

by the independence lemma. Here $\tau = T - t$.

Now set $S(t) = x$,

$$d_\pm(\tau, x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{\tau}}\Big(\ln\Big(\frac{x}{K}\Big) + \Big(r \pm \frac{\sigma^2}{2}\Big)\tau\Big),$$

and

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy = \frac{1}{\sqrt{2\pi}}\int_{-x}^{\infty} e^{-y^2/2}\,dy.$$

Observe

$$\begin{aligned}
c(t, x) &= \frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty} x\exp\Big(\frac{-\sigma^2\tau}{2} + \sigma\sqrt{\tau}y - \frac{y^2}{2}\Big)\,dy - e^{-r\tau}KN(d_-) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty} x\exp\Big(\frac{-(y - \sigma\sqrt{\tau})^2}{2}\Big)\,dy - e^{-r\tau}KN(d_-) \\
&= xN(d_+) - e^{-r\tau}KN(d_-),
\end{aligned}$$

which is precisely the Black-Scholes formula.

4. Review Problems

PROBLEM 4.1. Let f be a deterministic function, and define

$$X(t) \stackrel{\text{def}}{=} \int_0^t f(s)W(s)\,ds\,.$$

Find the distribution of X.

PROBLEM 4.2. Suppose σ, τ, ρ are two deterministic functions and M and N are two martingales with respect to a common filtration $\{\mathcal{F}_t\}$ such that $M(0) = N(0) = 0$, and

$$d[M,M](t) = \sigma(t)\,dt\,, \quad d[N,N](t) = \tau(t)\,dt\,, \qquad \text{and} \qquad d[M,N](t) = \rho\,dt\,.$$

(a) Compute the joint moment generating function $\boldsymbol{E}\exp(\lambda M(t) + \mu N(t))$.

(b) *(Lévy's criterion)* If $\sigma = \tau = 1$ and $\rho = 0$, show that (M, N) is a two dimensional Brownian motion.

PROBLEM 4.3. Consider a financial market consisting of a risky asset and a money market account. Suppose the return rate on the money market account is r, and the price of the risky asset, denoted by S, is a geometric Brownian motion with mean return rate α and volatility σ. Here r, α and σ are all deterministic constants. Compute the arbitrage free price of derivative security that pays

$$V(T) = \frac{1}{T}\int_0^T S(t)\,dt$$

at maturity T. Also compute the trading strategy in the replicating portfolio.

PROBLEM 4.4. Let $X \sim N(0,1)$, and $a, \alpha, \beta \in \mathbb{R}$. Define a new measure $\tilde{\boldsymbol{P}}$ by

$$d\tilde{\boldsymbol{P}} = \exp\bigl(\alpha X + \beta\bigr)\,d\boldsymbol{P}\,.$$

Find α, β such that $X + a \sim N(0,1)$ under $\tilde{\boldsymbol{P}}$.

PROBLEM 4.5. Let $x_0, \mu, \theta, \sigma \in \mathbb{R}$, and suppose X is an Itô process that satisfies

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

with $X(0) = x_0$.

(a) Find functions $f = f(t)$ and $g = g(s,t)$ such that

$$X(t) = f(t) + \int_0^t g(s,t)\,dW(s)\,.$$

The functions f, g may depend on the parameters x_0, θ, μ and σ, but should not depend on X.

(b) Compute $\boldsymbol{E}X(t)$ and $\operatorname{cov}(X(s), X(t))$ explicitly.

PROBLEM 4.6. Let M be a martingale, and φ be a convex function. Must the process $\varphi(M)$ be a martingale, sub-martingale, or a super-martingale? If yes, explain why. If no, find a counter example.

PROBLEM 4.7. Let $\theta \in \mathbb{R}$ and define

$$Z(t) = \exp\Bigl(\theta W(t) - \frac{\theta^2 t}{2}\Bigr)\,.$$

Given $0 \leqslant s < t$, and a function f, find a function such that

$$\boldsymbol{E}\big(f(Z(t)) \bigm| \mathcal{F}_s\big) = g(Z(s))\,.$$

Your formula for the function g can involve f, s, t and integrals, but not the process Z or expectations.

Problem 4.8. Let W be a Brownian motion, and define

$$B(t) = \int_0^t \operatorname{sign}(W(s))\,dW(s)\,.$$

(a) Show that B is a Brownian motion.
(b) Is there an adapted process σ such that

$$W(t) = \int_0^t \sigma(s)\,dB(s)\,?$$

If yes, find it. If no, explain why.
(c) Compute the joint quadratic variation $[B, W]$.
(d) Are B and W uncorrelated? Are they independent? Justify.

Problem 4.9. Let W be a Brownian motion. Does there exist an equivalent measure $\tilde{\boldsymbol{P}}$ under which the process $tW(t)$ is a Brownian motion? Prove it.

Problem 4.10. Suppose M is a continuous process such that both M and M^2 are martingales. Must M be constant in time? Prove it, or find a counter example.