HKU STAT3911 Study Notes

Chiu Ka Long (Leo)*

Last Updated: 2026-01-17

Contents

*email: `leockl@connect.hku.hk`; personal website: `https://leochiukl.github.io`

1 Probability Theory

1.0.1 Many topics in modern financial economics are built on *stochastic calculus*, which is touched upon in Section 4. As suggested by its name, it is a new type of calculus "with randomness"; for instance, rather than differentiating with respect to a deterministic time variable, we differentiate with respect to a *random variable* (you will learn how to make sense of this in STAT3911).

Due to the probabilistic nature of stochastic calculus, having a solid foundation in probability theory is critical for studying stochastic calculus properly. Unfortunately, the probability knowledge from your first course in probability is *not* enough, so we will cover the basics of some more advanced probability theory (known as the *measure-theoretic* probability theory) in Section 1 and Section 2. Next, armed with these knowledge, we will start diving into the major sections in STAT3911, which are more related to financial economics and hopefully more interesting 😊!

1.1 Set Theory Preliminaries

1.1.1 **Set terminologies.** In the following, we will review some terminologies about sets, which should be very familiar to you. Here, we let Ω be a nonempty *universal* set (meaning that every set below is a *subset* of it).

Name	Notation	Definition
A is a subset of Ω, or Ω is a superset of A	$A \subseteq \Omega$, or $\Omega \supseteq A$	For all $\omega \in A$, we have $\omega \in \Omega$.
(Absolute) complement	A^c	$\{\omega \in \Omega : \omega \notin A\}$
Set difference	$A \setminus B$	$A \cap B^c$
Intersection (I: index set)	$\bigcap_{i\in I} A_i$	$\{\omega \in \Omega : \omega \in A_i \text{ for all } i \in I\}$
Union	$\bigcup_{i\in I} A_i$	$\{\omega \in \Omega : \omega \in A_i \text{ for some } i \in I\}$
Disjoint union	$\biguplus_{i\in I} A_i$	meaning the same as $\bigcup_{i\in I} A_i$, with the emphasis on the pairwise disjointness, i.e., $A_i \cap A_j = \varnothing$ for all $i, j \in I$ with $i \neq j$.

1.1.2 **Basic set properties.** You should be very familiar to the following properties of sets (the sets below are arbitrary):

- *Associativity of union and intersection:* $(A\cup B)\cup C = A\cup(B\cup C)$ and $(A\cap B)\cap C = A\cap(B\cap C)$.
- *Commutativity of union and intersection:* $A \cup B = B \cup A$ and $A \cap B = B \cap A$.
- *De Morgan's laws:* $(\bigcup_{i\in I} A_i)^c = \bigcap_{i\in I} A_i^c$ and $(\bigcap_{i\in I} A_i)^c = \bigcup_{i\in I} A_i^c$.

Make sure you are able to *prove* these! (An useful way to prove a set equality $S = T$ is to prove (i) $S \subseteq T$ and (ii) $T \subseteq S$.)

1.1.3 **Further set terminologies.** For the following set theoretic terminologies, you may not have encountered them before, so perhaps you should pay more attention here:

Name	Notation	Definition
Infimum (set)	$\inf_{k\geq n} A_k$	$\bigcap_{k\geq n} A_k$
Supremum (set)	$\sup_{k\geq n} A_k$	$\bigcup_{k\geq n} A_k$
Limit inferior (set)	$\liminf_{n\to\infty} A_n$	$\bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k = \sup_{n\geq 1} \inf_{k\geq n} A_k$
Limit superior (set)	$\limsup_{n\to\infty} A_n$	$\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \inf_{n\geq 1} \sup_{k\geq n} A_k$
Limit of A_n	$A = \lim_{n\to\infty} A_n$ or $A_n \to A$	$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A$
$\{A_n\}_{n\in\mathbb{N}}$ is **increasing**	$A_n \nearrow$	$A_1 \subseteq A_2 \subseteq \cdots$
$\{A_n\}_{n\in\mathbb{N}}$ is **decreasing**	$A_n \searrow$	$A_1 \supseteq A_2 \supseteq \cdots$

1.1.4 **Further set properties.** Here we will discuss some results about the less familiar set terminologies from [1.1.3]:

(a) *(Interpreting limit inferior and limit superior)*

i. $\liminf_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for } \underline{\text{a}}\text{ll } \underline{\text{b}}\text{ut } \underline{\text{f}}\text{initely } \underline{\text{m}}\text{any } n\} =: \{\omega \in A_n \text{ abfm}\}$.

ii. $\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\} = \{\omega \in \Omega : \omega \in A_n \text{ } \underline{\text{i}}\text{nfinitely } \underline{\text{o}}\text{ften}\} =: \{\omega \in A_n \text{ io}\}$.

Proof.

i. Note that

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k = \{\omega \in \Omega : \exists n \in \mathbb{N} \text{ s.t. } \omega \in A_k \ \forall k \geq n\}$$

and "$\exists n \in \mathbb{N}$ s.t. $\omega \in A_k\ \forall k \geq n$" just means "$\omega \in A_n$ for all but finitely many n" in words.

ii. Similar to above, and we can interpret "$\forall n \in \mathbb{N}\ \exists k \geq n$ s.t. $\omega \in A_k$" as "$\omega \in A_n$ for infinitely many n".

□

(b) *(Relating limit inferior and limit superior)*

i. $\liminf_{n\to\infty} A_n \subseteq \liminf_{n\to\infty} A_n$.

ii. $(\liminf_{n\to\infty} A_n)^c = \limsup_{n\to\infty} A_n^c$.

iii. $(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c$.

Proof.

i. Because "$\omega \in A_n$ for all but finitely many n" is just a special case of "$\omega \in A_n$ for infinitely many n".

ii. Apply De Morgan's laws twice:

$$\left(\liminf_{n\to\infty} A_n\right)^c = \left(\bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k\right)^c \overset{\text{DM}}{=} \bigcap_{n=1}^{\infty} \left(\bigcap_{k\geq n} A_k\right)^c \overset{\text{DM}}{=} \bigcap_{n=1}^{\infty} \left(\bigcup_{k\geq n} A_k^c\right) = \limsup_{n\to\infty} A_n^c.$$

iii. Again apply De Morgan's laws twice:

$$\left(\limsup_{n\to\infty} A_n\right)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right)^c \overset{\text{DM}}{=} \bigcup_{n=1}^{\infty} \left(\bigcup_{k\geq n} A_k\right)^c \overset{\text{DM}}{=} \bigcup_{n=1}^{\infty} \left(\bigcap_{k\geq n} A_k^c\right) = \liminf_{n\to\infty} A_n^c.$$

□

(c) *(Limits of monotone sequences of sets)*

i. If $A_n \nearrow$, then $\lim_{n\to\infty} A_n$ exists and equals $\bigcup_{k=1}^{\infty} A_k$.
[Note: Setting $A_n := \bigcup_{i=1}^{n} B_i$, since $A_n \nearrow$, we can get the intuitively appealing equality $\lim_{n\to\infty} \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{\infty} B_i$, as $\bigcup_{i=1}^{\infty} B_i = \bigcup_{k=1}^{\infty}\left(\bigcup_{i=1}^{k} B_k\right)$.]

ii. If $A_n \searrow$, then $\lim_{n\to\infty} A_n$ exists and equals $\bigcap_{k=1}^{\infty} A_k$.

iii. For all collections $\{A_n\} \subseteq \mathcal{P}(\Omega)$, $\liminf_{n\to\infty} A_n = \lim_{n\to\infty}(\inf_{k\geq n} A_k)$ and $\limsup_{n\to\infty} A_n = \lim_{n\to\infty}(\sup_{k\geq n} A_k)$. [Note: This explains the rationale behind the notations "$\liminf_{n\to\infty} A_n$" and "$\limsup_{n\to\infty} A_n$".]

Proof.

i. Assuming $A_n \nearrow$, we have $A_k \subseteq \bigcap_{i=k}^{\infty} A_i$. Hence,

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \subseteq \boxed{\bigcup_{k=1}^{\infty} A_k} \subseteq \bigcup_{n=1}^{\infty} \bigcap_{i=k}^{\infty} A_i = \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n.$$

This forces $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} A_k$, as desired.

ii. Let $B_n = A_n^c$ for all n, then $B_n \nearrow$. Applying (i), we get

$$\liminf_{n\to\infty} B_n = \limsup_{n\to\infty} B_n = \bigcup_{k=1}^{\infty} B_k.$$

By [1.1.4]b, we have $\liminf_{n\to\infty} B_n = (\limsup_{n\to\infty} A_n)^c$ and $\limsup_{n\to\infty} B_n = (\liminf_{n\to\infty} A_n)^c$. It then follows that

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \left(\bigcup_{k=1}^{\infty} B_k\right)^c = \bigcap_{k=1}^{\infty} A_k^c.$$

iii. Note that $\inf_{k\geq n} A_k \nearrow$ and $\sup_{k\geq n} A_k \searrow$. Hence,

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty}\inf_{k\geq n} A_k \overset{\text{(i)}}{=} \lim_{n\to\infty}\left(\inf_{k\geq n} A_k\right)$$

and

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty}\sup_{k\geq n} A_k \overset{\text{(ii)}}{=} \lim_{n\to\infty}\left(\sup_{k\geq n} A_k\right).$$

□

1.1.5 **Indicator functions.** You should have learnt what an indicator function is in your first probability course. This function continues to be of great use here, so let us review it a bit.

- The *indicator function* of $A \subseteq \Omega$ is given by

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

- Property: $\mathbf{1}_A \leq \mathbf{1}_B$ (i.e., $\mathbf{1}_A(\omega) \leq \mathbf{1}_B(\omega)$ for all $\omega \in \Omega$) iff $A \subseteq B$.

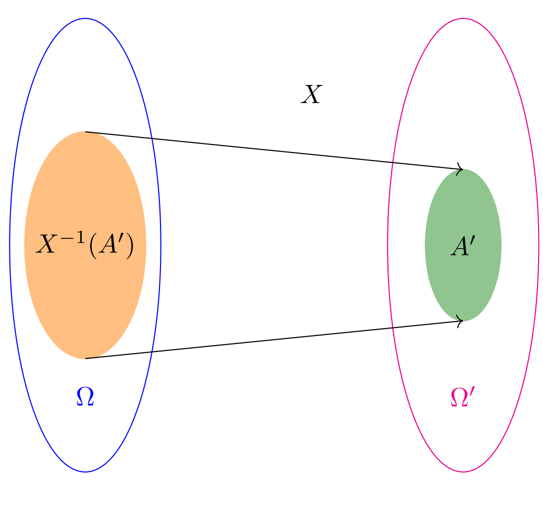
Indicator function can be applied for describing $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$:

$$\limsup_{n\to\infty} A_n = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty\right\}, \quad \liminf_{n\to\infty} A_n = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(\omega) < \infty\right\}.$$

This is because when $\omega \in A_n$ infinitely often, infinitely many $\mathbf{1}_{A_n}(\omega)$ would equal one; when $\omega \in A_n$ for all but finitely many n, finitely many $\mathbf{1}_{A_n^c}(\omega)$ would equal one.

1.1.6 **Images and preimages.** A pair of concepts that will frequently appear in our discussion of measure theoretic probability theory is *image* and *preimage*, which is covered in MATH2012.

Let Ω and Ω' be two sets, and $X : \Omega \to \Omega'$ be a function. The *image* of $A \subseteq \Omega$ under X is $X(A) := \{X(\omega) : \omega \in A\}$, and the *preimage* of $A' \subseteq \Omega'$ under is $X^{-1}(A') := \{\omega \in \Omega : X(\omega) \in A'\}$.



Let $\mathcal{A}'$ be a family/set of sets in Ω' (i.e., a subfamily/subset of $\mathcal{P}(\Omega')$. Then, the notation $X^{-1}(\mathcal{A}')$ refers to the family/set of all preimages of sets in $\mathcal{A}$, i.e., $\{X^{-1}(A') : A' \in \mathcal{A}'\}$.

1.1.7 **Properties of preimages.** Let A', B' denote any subsets of Ω', and $\{A'_i\}_{i\in I}$ be any subcollection in $\mathcal{P}(\Omega')$.

(a) *(preservation of complementation)* $\left(X^{-1}(A')\right)^c = X^{-1}(A'^c)$.

(b) *(preservation of union)* $X^{-1}(\bigcup_{i\in I} A'_i) = \bigcup_{i\in I} X^{-1}(A'_i)$.

(c) *(preservation of intersection)* $X^{-1}(\bigcap_{i\in I} A'_i) = \bigcap_{i\in I} X^{-1}(A'_i)$.

(d) *(monotonicity)* Let $\mathcal{A}', \mathcal{B}' \subseteq \mathcal{P}(\Omega')$. Then, $\mathcal{A}' \subseteq \mathcal{B}' \implies X^{-1}(\mathcal{A}') \subseteq X^{-1}(\mathcal{B}')$.

Proof. We demonstrate the proof for (b) and (d) here and leave the rest as exercises.

(b) "$\subseteq$": Fix any $\omega \in X^{-1}(\bigcup_{i\in I} A'_i)$. By definition, $X(\omega) \in \bigcup_{i\in I} A'_i$, thus there exists some $i \in I$ such that $X(\omega) \in A'_i$, or $\omega \in X^{-1}(A'_i)$. This means $\omega \in \bigcup_{i\in I} X^{-1}(A'_i)$.
"$\supseteq$": Highly similar to "$\subseteq$"; just work backward.

(d) Assume $\mathcal{A}' \subseteq \mathcal{B}'$. Now, fix any $A \in X^{-1}(\mathcal{A}')$. By definition, we have $A = X^{-1}(A')$ for some $A' \in \mathcal{A}'$. Since $\mathcal{A}' \subseteq \mathcal{B}'$, we also have $A' \in \mathcal{B}'$, meaning that $A \in \{X^{-1}(A') : A' \in \mathcal{B}'\} = X^{-1}(\mathcal{B}')$.

$\square$

1.2 Probability Spaces

1.2.1 **Systems of sets.** *Systems of sets* are utilized for constructing families of "selected" sets on which a probability measure $\mathbb{P}$ can be defined consistently without having any issue. It turns out that there are indeed some sets to which we cannot assign any reasonable probability, like *Vitali set* (see STAT7610 for more details), so in general we cannot just blindly choose the whole power set of sample space as the domain of the probability measure $\mathbb{P}$ (though it is quite tempting!), and we need to exclude those "pathological" sets by restricting the domain. [Note: However, such issues only arise when the sample space Ω is uncountable. If Ω is countable, then we can define $\mathbb{P}$ on the power set $\mathcal{P}(\Omega)$ without issues.]

There are multiple systems of sets here, but they all share a common theme, which is about constructing a family of sets that is *closed* under certain set operations, i.e., performing these operations on sets in $\mathcal{F}$ would not yield something outside $\mathcal{F}$ — being "stable" in some sense. Intuitively, we are interested in this kind of properties because they can make the families of sets "rich enough", in the sense that the families contain "sufficiently many well-behaved sets". Here, we will discuss two types of systems of sets: (i) algebra and (ii) σ-algebra (important concept for probability theory!)

1.2.2 **Definitions of algebra and σ-algebra.**

- A family $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is an **algebra** (or **field**) on Ω if
  (1) $\varnothing \in \mathcal{F}$.
  (2) *(closed under complementations)* $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$.
  (3) *(closed under unions)* $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$.
- A family $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a **σ-algebra** (or **σ-field**) on Ω if
  (1) $\varnothing \in \mathcal{F}$.
  (2) *(closed under complementations)* $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$.
  (3) *(closed under countable unions)* $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

  [Note: A σ-algebra is actually also closed under several more set operations, e.g., finite unions, finite/countable intersections, and set differences:

  - *finite intersections:* It follows from the closedness under countable unions from setting $A_{n+1} = A_{n+2} = \cdots = \varnothing$, because $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$ in this case.

– *countable intersections:* To show this, we can use De Morgan law, and the closedness under countable unions and complementations:

$$A_1, A_2, \ldots \in \mathcal{F} \implies \bigcap_{i=1}^{\infty} A_i \overset{\text{DM}}{=} \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{F}.$$

– *finite intersections:* It follows from the closedness under countable intersections from setting $A_{n+1} = A_{n+2} = \cdots = \Omega$, because $\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i$ in this case.

– *set differences:* Fix any $A, B \in \mathcal{F}$. Then $B^c \in \mathcal{F}$. Applying the closedness under finite intersections, we have $A \setminus B = A \cap B^c \in \mathcal{F}$.

]

1.2.3 **Examples of σ-algebras.**

(1) The **trivial σ-algebra** on Ω is $\mathcal{F} = \{\varnothing, \Omega\}$. [Note: It is the *smallest* σ-algebra on Ω, i.e., every σ-algebra on Ω is a superset of the trivial σ-algebra. This is because containing $\varnothing$ and being closed under complementations would force a σ-algebra to at least contain $\varnothing$ and Ω.]

(2) The power set $\mathcal{F} = \mathcal{P}(\Omega)$ is a σ-algebra on Ω.

Remarks:

- It is a σ-algebra on Ω because complement of subset of Ω is still a subset of Ω, and countable union of subsets of Ω is still a subset of Ω.
- It is the *largest* σ-algebra on Ω, i.e., every σ-algebra on Ω is a subset of $\mathcal{P}(\Omega)$ (which follows from definition).

1.2.4 **σ-algebra generated by a family of sets.** Let us motivate this concept by considering some examples of constructing σ-algebras from sets:

- *Constructing from one set:* Start with a set $A \subseteq \Omega$. Then consider the following two sets that partition Ω:

$$A \qquad A^c$$

By putting zero/one/two of them into an union, we can get 4 combinations:

– *zero:* $\varnothing$

– *one:* A, A^c

– *two:* $A \cup A^c = \Omega$

These four sets form a σ-algebra: $\{\varnothing, A, A^c, \Omega\}$.

- *Constructing from two sets:* Here we start with two sets $A, B \subseteq \Omega$. Then consider the following $2^2 = 4$ sets that partition Ω: $A \cap B$, $A^c \cap B$, $A \cap B^c$, and $A^c \cap B^c$.

B^c	$A \cap B^c$	$A^c \cap B^c$
B	$A \cap B$	$A^c \cap B$
	A	A^c

By putting zero/one/two/three/four of them into an union, we can get $\binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = (1+1)^4 = 16$ combinations:

– *zero:* $\varnothing$

- *one:* $A \cap B$, $A^c \cap B$, $A \cap B^c$, $A^c \cap B^c$
- *two:*
  * $B = (A \cap B) \cup (A^c \cap B)$
  * $A = (A \cap B) \cup (A \cap B^c)$
  * $(A \cap B) \cup (A^c \cap B^c)$
  * $(A^c \cap B) \cup (A \cap B^c)$
  * $A^c = (A^c \cap B) \cup (A^c \cap B^c)$
  * $B^c = (A \cap B^c) \cup (A^c \cap B^c)$
- *three:* (just complement of each of the four sets in the partition indeed)
  * $A^c \cup B^c = (A \cap B)^c$
  * $A \cup B^c = (A^c \cap B)^c$
  * $A^c \cup B = (A \cap B^c)^c$
  * $A \cup B = (A^c \cap B^c)^c$
- *four:* Ω

These 16 sets form a σ-algebra.

The σ-algebras above are indeed examples of σ-algebra *generated by a family of sets*. The one-set example is the σ-algebra generated by the family $\mathcal{A} = \{A\}$, and the two-set example is the σ-algebra generated by the family $\mathcal{A} = \{A, B\}$.

In general, let $\mathcal{A}$ be a family of sets in Ω. Then the **σ-algebra generated by $\mathcal{A}$**, denoted by $\sigma(\mathcal{A})$, is the *smallest* σ-algebra that contains every set in the family $\mathcal{A}$, i.e., (i) $\sigma(\mathcal{A})$ is a σ-algebra on Ω that contains $\mathcal{A}$, and (ii) for every σ-algebra $\mathcal{F}'$ on Ω with $\mathcal{A} \subseteq \mathcal{F}'$, we have $\sigma(\mathcal{A}) \subseteq \mathcal{F}'$. We consider the smallest one here so that all "irrelevant" sets are excluded in the σ-algebra.

1.2.5 **Borel σ-algebra.** Next, we shall consider a commonly used σ-algebra that contains most sets (subsets of $\mathbb{R}$) we can imagine, called *Borel σ-algebra* (it is difficult to think of a set that is not in this σ-algebra!).

The **Borel σ-algebra on $\mathbb{R}$**, denoted by $\mathcal{B}(\mathbb{R})$ or $\mathcal{B}$, is the σ-algebra generated by the family $\mathcal{A} = \{(a, b] : a < b\}$ of left-open and right-closed intervals, i.e., $\mathcal{B} = \sigma(\mathcal{A})$. Every set in $\mathcal{B}$ is said to be a **Borel set**.

To see why $\mathcal{B}$ includes most subsets of $\mathbb{R}$ we can imagine, let us explore what types of subsets of $\mathbb{R}$ are included in $\mathcal{B}$:

- *singletons:* $\bigcap_{i=1}^{\infty}(a - 1/i, a] = \{a\} \in \mathcal{B}$
- *closed intervals:* $(a, b] \cup \{a\} = [a, b] \in \mathcal{B}$
- *open intervals:* $(a, b] \setminus \{a\} = (a, b) \in \mathcal{B}$
- *left-closed and right-open intervals:* $[a, b] \setminus \{b\} = [a, b) \in \mathcal{B}$
- any finite/countable union(s)/intersection(s) of (complement) of sets in the types above belongs to $\mathcal{B}$

As you can see, it is rather difficult (but not impossible!) to think of a subset of $\mathbb{R}$ that is in none of the types above.

1.2.6 **σ-algebra on subset of $\mathbb{R}$.** It turns out that Borel σ-algebra can be defined in a more general fashion, and the Borel σ-algebra on $\mathbb{R}$ is just a special case; see STAT7610 for more details. Generally, a Borel σ-algebra on $A \subseteq \mathbb{R}$ can be obtained by $\mathcal{B}(A) = \sigma(\{(a, b] \cap A : a < b\})$. For example, the Borel σ-algebra on $[0, 1]$ is given by $\mathcal{B}([0, 1]) = \sigma(\{(a, b] : 0 \le a < b \le 1\})$.

1.2.7 **Probability measures.** With the knowledge of σ-algebra, we can define probability measure formally (the definition of probability measure you have seen previously is perhaps informal). Let Ω be a sample space and $\mathcal{F}$ be a σ-algebra on Ω. Then, a **probability measure** $\mathbb{P}$ on $\mathcal{F}$ (or on $(\Omega, \mathcal{F})$, which is known as a **measurable space**) is a real-valued function on $\mathcal{F}$ satisfying:

(1) *(nonnegativity)* $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.

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

(3) *(countable additivity)* $A_1, A_2, \ldots \in \mathcal{F}$ pairwise disjoint $\implies \mathbb{P}(\biguplus_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Remarks:

- Informally, countable additivity allows us to "pull $\biguplus$ out of μ and make it $\sum$" (explaining why there is a "+" in the notation "$\biguplus$").
- *(probability of empty set)* By taking $A_1 = \Omega$ and $A_2 = A_3 = \cdots = \varnothing$, from countable additivity we know $1 = \mathbb{P}(\Omega) = \mathbb{P}(\biguplus_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(\Omega) + \sum_{i=2}^{\infty} \mathbb{P}(\varnothing)$, which implies $\sum_{i=2}^{\infty} \mathbb{P}(\varnothing) = 0$. This forces $\mathbb{P}(\varnothing) = 0$, as expected.
- *(finite additivity)* By taking $A_{n+1} = A_{n+2} = \cdots = \varnothing$, we can show also the *finite additivity*: $A_1, \ldots, A_n \in \mathcal{F}$ pairwise disjoint $\implies \mathbb{P}(\biguplus_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

These three defining properties of a probability measure are sometimes known as the **probability axioms**. Every set in $\mathcal{F}$ is called an **event**. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**. [Note: If we replace $\mathbb{P}$ by a *measure* in general[1] (e.g., Lebesgue measure, which corresponds to usual notions of "length", "area", "volume", etc.), then the triple is called a **measure space**.]

1.2.8 **Properties of probability measures.** Based on the three probability axioms, we can deduce many properties of probability which may be familiar to you. In the following, let A, B, and A_i's be sets in $\mathcal{F}$.

(a) *(subtractivity)* If $A \subseteq B$, then $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

(b) *(monotonicity)* $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$.

(c) *(complement)* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(d) *(union + intersection = individual sum)* $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$.

(e) *(interchanging limit and probability)* If $A_n \nearrow$ or $A_n \searrow$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n\to\infty} A_n)$. [Note: We have $A_1, A_2, \ldots \in \mathcal{F} \implies \lim_{n\to\infty} A_n \in \mathcal{F}$ as we can write $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k$, which is a *countable* union of *countable* intersections of sets in $\mathcal{F}$, thus it is in $\mathcal{F}$.]

(f) *(countable subadditivity)* $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$. [Note: By setting $A_{n+1} = A_{n+2} = \cdots = \varnothing$, we can obtain *finite subadditivity*: $\mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$.]

Proof.

(a) It follows from the finite additivity since A and $B \setminus A$ are pairwise disjoint.

(b) By subtractivity, we have $\mathbb{P}(A) = \mathbb{P}(B) - \underbrace{\mathbb{P}(B \setminus A)}_{\geq 0} \leq \mathbb{P}(B)$.

(c) It follows from setting $B = \Omega$ in the set difference property.

(d) The trick is to write $A \cup B = A \uplus (B \setminus A) = A \uplus (B \setminus (A \cap B))$, and thus by finite additivity and subtractivity we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Rearranging this gives the desired result.

(e) First consider the case with $A_n \nearrow$. Let $B_1 := A_1$ and $B_n := A_n \setminus A_{n-1}$ for every integer $n \geq 2$. By construction, B_n's are pairwise disjoint, and $A_n = \bigcup_{i=1}^{n} A_i = \biguplus_{i=1}^{n} B_i$ for every $n \in \mathbb{N}$. We

[1] As we are primarily focusing on probability measures, we omit the definition of a general measure here; see STAT7610 for more details.

then have $\bigcup_{i=1}^{\infty} A_i = \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i = \lim_{n\to\infty} \biguplus_{i=1}^{n} B_i = \biguplus_{i=1}^{\infty} B_i$, and thus

$$\mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) \overset{[1.1.4]c}{=} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\biguplus_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i)$$
$$= \lim_{n\to\infty} \mathbb{P}\left(\biguplus_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

For the other case with $A_n \searrow$, observe that $A_n^c \nearrow$, and we have $\lim_{n\to\infty} A_n^c = (\lim_{n\to\infty} A_n)^c$ by [1.1.4]b. Therefore,

$$\mathbb{P}\Big(\lim_{n\to\infty} A_n\Big) = 1 - \mathbb{P}\Big(\Big(\lim_{n\to\infty} A_n\Big)^c\Big) = 1 - \mathbb{P}\Big(\lim_{n\to\infty} A_n^c\Big) \overset{\text{proven}}{=} 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c)$$
$$= \lim_{n\to\infty} (1 - \mathbb{P}(A_n^c)) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(f) Let $B_1 := A_1$ and $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$ for every integer $n \geq 2$. Then by construction, (i) B_n's are pairwise disjoint, (ii) $B_n \subseteq A_n$ for every $n \in \mathbb{N}$, and (iii) $\bigcup_{i=1}^{n} A_i = \biguplus_{i=1}^{n} B_i$ for every $n \in \mathbb{N}$. By (iii), we have $\bigcup_{i=1}^{\infty} A_i = \lim_{N\to\infty} \bigcup_{i=1}^{N} A_i = \lim_{N\to\infty} \biguplus_{i=1}^{N} B_i = \biguplus_{i=1}^{\infty} B_i$. Thus,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\biguplus_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) \overset{\text{(ii), monotonicity}}{\leq} \sum_{i=1}^{\infty} \mu(A_i).$$

□

1.3 Random Variables

1.3.1 In your first course in probability, you should have learnt that a random variable is a function from the sample space Ω to $\mathbb{R}$: It quantifies each sample point ω in the sample space. However, this "definition" is indeed an informal one, and actually an extra condition is needed, namely *measurability*.

A function $X : \Omega \to \mathbb{R}$ is a **($\mathcal{F}$-)random variable** if $X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\} \subseteq \mathcal{F}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. [Note: In some contexts, we may implicitly adapt the definition a bit by replacing $\mathbb{R}$ by $\overline{\mathbb{R}} = [-\infty, \infty]$, the set of extended real numbers. In respond to this change, we would change $\mathcal{B}$ to the Borel σ-algebra on $\overline{\mathbb{R}}$, which is given by $\{B \cup E : B \in \mathcal{B}, E \in \{-\infty, \infty\}\}$.]

Recall that the probability $\mathbb{P}(X \in B)$ is indeed given by $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathbb{P}\big(X^{-1}(B)\big)$. So the idea of this extra condition is to ensure that for all sets of practical interest (the Borel sets), it is possible to compute $\mathbb{P}(X \in B)$. In other words, we need to require X to be sufficiently well-behaved such that for all Borel sets B, $X^{-1}(B)$ would not fall outside the σ-algebra $\mathcal{F}$, avoiding the probability $\mathbb{P}(X \in B)$ to be undefined.

We sometimes call such extra condition as being **$\mathcal{F}$-measurable**. So we can say that a random variable X is a $\mathcal{F}$-measurable function. [Note: This is sometimes denoted by the shorthand "$X \in \mathcal{F}$".] For a generic function $f : \mathbb{R} \to \mathbb{R}$, we can define measurability in a similar fashion: It is **(Borel-)measurable** if $f^{-1}(\mathcal{B}) \subseteq \mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. Here we do not have "$\mathcal{F}$" as the default σ-algebra on $\mathbb{R}$ is always chosen to be $\mathcal{B}$.

1.3.2 **Distribution measure.** Indeed, a random variable X would *induce* a probability measure on $\mathbb{R}$ (equipped with the Borel σ-algebra $\mathcal{B}$), called the **distribution (measure)** of X, denoted by μ_X or $\mathbb{P}_X$, which is given by

$$\mu_X(B) := \mathbb{P}(X \in B) = \mathbb{P}\big(X^{-1}(B)\big) \text{ for all } B \in \mathcal{B}.$$

We can verify that μ_X is a probability measure as follows:

(1) The nonnegativity of μ_X follows from the nonnegativity of $\mathbb{P}$. ✔

(2) $\mu_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$. ✔

(3) Fix any pairwise disjoint events $A_1, A_2, \ldots \in \mathcal{B}$. Then, $\mu_X(\bigcup_{i=1}^{\infty} A_i) = \mathbb{P}\big(X^{-1}(\biguplus_{i=1}^{\infty} A_i)\big) \overset{[1.1.7]}{=} \mathbb{P}\big(\biguplus_{i=1}^{\infty} X^{-1}(A_i)\big) = \sum_{i=1}^{\infty} \mathbb{P}\big(X^{-1}(A_i)\big) = \sum_{i=1}^{\infty} \mu_X(A_i)$. ✔

1.3.3 **Properties of random variables.** With the formal definition of random variable, we can discuss the following theoretical properties of random variables.

(a) *(equivalent definition when X takes finitely many values)* Suppose a function X takes finitely many values $x_1, \ldots, x_n$ only. Then X is a random variable iff $X^{-1}(\{x_i\}) \in \mathcal{F}$ for all $i = 1, \ldots, n$.
Proof. "$\Rightarrow$": Because $\{x_i\}$ is a Borel set for all $i = 1, \ldots, n$.
"$\Leftarrow$": Fix any $B \in \mathcal{B}$. For the case where B contains none of $x_1, \ldots, x_n$, we have $X^{-1}(B) = \varnothing \in \mathcal{F}$. So it remains to prove for the case where B contains some of $x_1, \ldots, x_n$, say $x_{i_1}, \ldots, x_{i_j}$. Then, we have $X^{-1}(B) = X^{-1}(\{x_{i_1}\} \cup \cdots \cup \{x_{i_j}\}) = X^{-1}(\{x_{i_1}\}) \cup \cdots \cup X^{-1}(\{x_{i_j}\}) \in \mathcal{F}$ due to the closedness under unions for $\mathcal{F}$. ☐

(b) *(measurable function of random variable is random variable)* If X is a $\mathcal{F}$-random variable and f is a *measurable* function, then $f \circ X = f(X)$ is also a $\mathcal{F}$-random variable.
Proof. Fix any $B \in \mathcal{B}$. Then, we have

$$\begin{aligned}(f \circ X)^{-1}(B) &= \{\omega \in \Omega : f(X(\omega)) \in B\} = \{\omega \in \Omega : X(\omega) \in f^{-1}(B)\} \\ &= \{\omega \in \Omega : \omega \in X^{-1}(f^{-1}(B))\} = X^{-1}(f^{-1}(B)).\end{aligned}$$

Since f is measurable, we have $f^{-1}(B) \in \mathcal{B}$. With X being a $\mathcal{F}$-random variable, we have $(f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F}$, as desired. ☐

1.4 Expectations

1.4.1 Given a random variable, often we would like to compute its expectation. In your first probability course, you should have learnt these formulas for computing expectations:

$$\mathbb{E}[X] = \begin{cases} \sum_k x_k f(x_k) & \text{if } X \text{ is discrete, taking countably many values } x_1, x_2, \ldots, \\ \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

where f denotes the mass function (for discrete case) or the density function (for continuous case). However, we will face difficulties when X is *neither discrete nor continuous* (this is possible; the definition for X to be continuous is not "X is not discrete"⚠). These formulas are not flexible enough, and are not enough for achieving an in-depth understanding of the theories behind financial economics. Hence, we will introduce a more general (yet more abstract) definition of expectation, which involves the notion of *Lebesgue integral*.

1.4.2 **Motivation.** The above formulas of expectations are both capturing the idea of "weighted average"; the expectation is considered to be an average weighted according to the probabilities. This idea is preserved in the general and abstract definition of expectation.

To motivate the general definition, we first consider the special case where the sample space Ω is countable, which forces all random variables defined on Ω to be discrete. Hence, we can express the expectation as follows:

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

which is an alternative expression of the discrete expectation formula above. Through this expression, we can clearly see that the expectation is an average weighted according to probabilities.

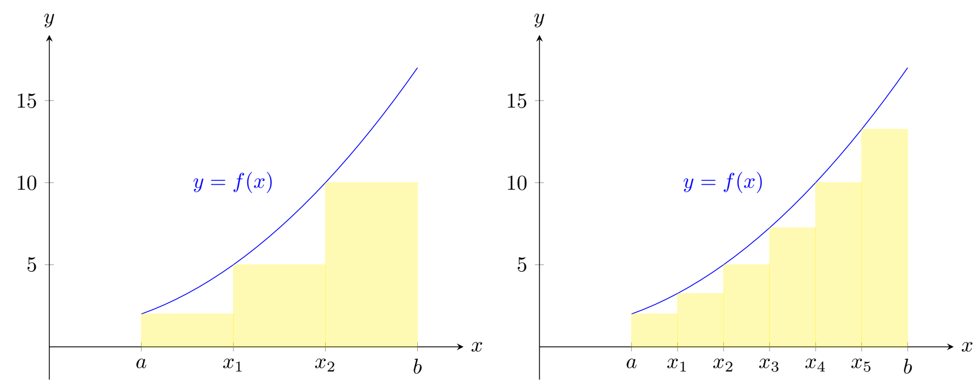
To generalize the notion above to the case where Ω is uncountable, we need to utilize the notion of *integral*, which may be seen as a mathematical way to "sum over uncountably many values".

Intuitively, one may consider using *Riemann integral* for the definition (the usual integral you learn in your calculus course). However, it only makes sense for functions whose domain is a set of *real* *numbers*. In our case here, the function of interest is the *random variable* X, whose domain is the sample space Ω, which may *not* contain numbers in general! We only know that the *values taken* by X (on the "y-axis") are numeric.

Recall that the Riemann integral is defined by partitioning the *x-axis* into finer and finer pieces. In our case, knowing only the y-axis is numeric, it seems natural to consider an integral defined by partitioning the *y-axis* instead; such integral is known as the *Lebesgue integral*.

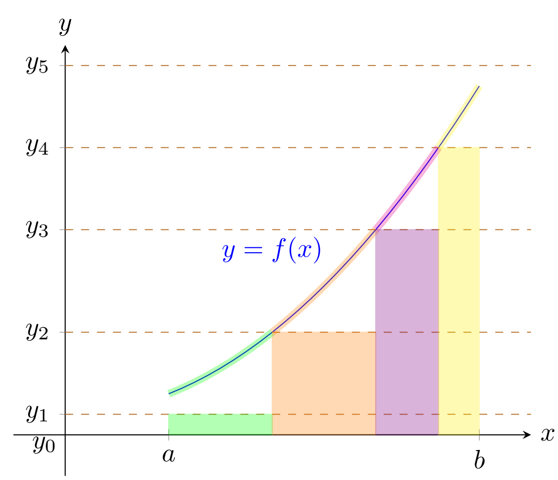
1.4.3 **Comparing Riemann and Lebesgue integrals.** Both Riemann and Lebesgue integrals are trying to find out the *area under curve*, but in different ways. To illustrate the idea, consider the following.

- *Riemann integral:* We partition the x-axis into finer and finer pieces to approach to the area under curve.



- *Lebesgue integral:* We partition the y-axis into finer and finer pieces to approach to the area under curve.

  More details: First assume that the function f is nonnegative on $[a, b]$. Let $\Pi = \{y_0, y_1, y_2, \dots\}$ be a partition of the y-axis, where $0 = y_0 < y_1 < y_2 < \cdots$. For each subinterval $[y_k, y_{k+1}]$, let $A_k := \{x \in [a, b] : y_k \le f(x) < y_{k+1}\}$. Then the **lower Lebesgue sum** is defined to be $\mathrm{LS}_\Pi^-(f) = \sum_{k=1}^{\infty} y_k \mu(A_k)$, where μ is a measure to be specified, e.g., Lebesgue measure. With a finer and finer partition, the lower Lebesgue sum (sum of areas of rectangles below) would approach to the area under curve (when the curve is "sufficiently nice").



1.4.4 **Definition of Lebesgue integral.** As we are primarily discussing expectations here, we will focus on defining Lebesgue integral of a *random variable* X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (although it can be similarly defined for measurable real-valued functions in general). The definition takes two steps:

(1) *(definition for nonnegative random variables)* For now, assume that $0 \le X(\omega) < \infty$ for all $\omega \in \Omega$. Let $\Pi = \{y_0, y_1, y_2, \dots\}$ be a partition where $0 = y_0 < y_1 < y_2 < \cdots$. For each subinterval

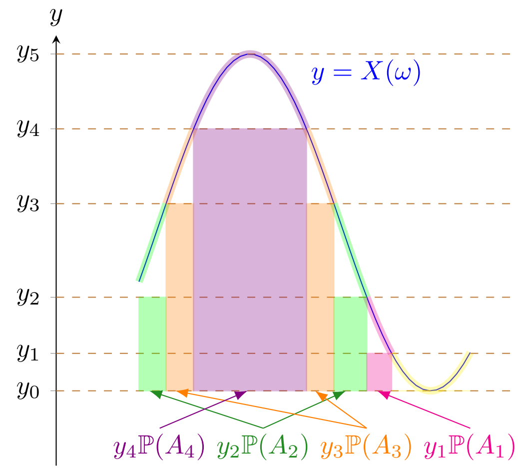
$[y_k, y_{k+1}]$, let $A_k = \{\omega \in \Omega : y_k \leq X(\omega) < y_{k+1}\} = X^{-1}([y_k, y_{k+1})) \in \mathcal{F}$ (as X is a random variable). Consider the lower Lebesgue sum $\mathrm{LS}_\Pi^-(X) = \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k)$.

Let $\|\Pi\|$ be the maximal distance between the neighbouring partition points in Π. Then, the **Lebesgue integral** is defined to be the limit of the lower Lebesgue sum $\mathrm{LS}_\Pi^-(X)$ as $\|\Pi\| \to 0$:

$$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) := \lim_{\|\Pi\|\to 0} \mathrm{LS}_\Pi^-(X)^2,$$

which is always nonnegative and could be ∞. [Note: This notation is "inspired" by the previous expectation formula $\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$ in the case with countable Ω.]

An alternative notation for the Lebesgue integral is $\int_\Omega X\,\mathrm{d}\mathbb{P}$.



Some extensions: (assuming X is allowed to take the value of ∞)

- If $\mathbb{P}(\{\omega \in \Omega : X(\omega) = \infty\}) = 0$ or $\mathbb{P}(\{\omega \in \Omega : X(\omega) < 0\}) = 0$, then the Lebesgue integral turns out to remain unchanged. [Note: This suggests that we indeed only require $0 \leq X < \infty$ to hold with probability 1.]
- If $\mathbb{P}(\{\omega \in \Omega : X(\omega) = \infty\}) > 0$, then we *define* $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) := \infty$.

(2) *(definition for general random variables)* Having the definition for nonnegative random variables, we can define the Lebesgue integral for general random variables through splitting them into *positive parts* and *negative parts*. Given any random variable X,

- the **positive part** of X is given by $X^+(\omega) = \max\{X(\omega), 0\}$ for all $\omega \in \Omega$;
- the **negative part** of X is given by $X^-(\omega) = \max\{-X(\omega), 0\}$ for all $\omega \in \Omega$.

Note that both X^+ and X^- are nonnegative random variables. Also we have $X = X^+ - X^-$ and $|X| = X^+ + X^-$.

In this general setting, the **Lebesgue integral** $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ is defined based on the positive and negative parts.

- *Case 1:* $\int_\Omega X^+(\omega)\,\mathrm{d}\mathbb{P}(\omega) < \infty$ *and* $\int_\Omega X^-(\omega)\,\mathrm{d}\mathbb{P}(\omega) < \infty$. We define

$$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) := \int_\Omega X^+(\omega)\,\mathrm{d}\mathbb{P}(\omega) - \int_\Omega X^-(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

  In this case, we also say that X is **integrable**.

- *Case 2:* $\int_\Omega X^+(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \infty$ *and* $\int_\Omega X^-(\omega)\,\mathrm{d}\mathbb{P}(\omega) < \infty$. We define

$$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) := \infty.$$

[2] Actually some technical details are omitted here (e.g., what does "$\lim_{\|\Pi\|\to 0}$" mean mathematically?), and the formal treatment of Lebesgue integral is more involved (see STAT7610 for more details). But for our purpose, it is enough to have basic understanding on Lebesgue integral above and have some geometric intuition about it, e.g., the Lebesgue integral $\int_\Omega c\,\mathrm{d}\mathbb{P}(\omega)$ can be geometrically viewed as the area of rectangle with length $\mathbb{P}(\Omega) = 1$ and height c, which is c.

- *Case 3:* $\int_\Omega X^+(\omega)\,\mathrm{d}\mathbb{P}(\omega) < \infty$ *and* $\int_\Omega X^-(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \infty$. We define
$$\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) := -\infty.$$
- *Case 4:* $\int_\Omega X^+(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \infty$ *and* $\int_\Omega X^-(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \infty$. We leave $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ *undefined*.

Apart from the Lebesgue integral over the whole sample space Ω discussed above, we can also define the **Lebesgue integral over a set** $A \in \mathcal{F}$ (where $\mathcal{F}$ is a σ-algebra on Ω) as follows:
$$\int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\Omega \mathbf{1}_A(\omega)X(\omega)\,\mathrm{d}\mathbb{P}(\omega),$$
where $\mathbf{1}_A$ denotes the indicator function. [Note: In a similar manner, the Lebesgue integral can be defined for measures other than $\mathbb{P}$, e.g., the Lebesgue measure λ, and similar notations apply; see STAT7610 for more details.]

1.4.5 **Relationship between Riemann and Lebesgue integral.** The Lebesgue integral can indeed be viewed as a generalization to the Riemann integral, as suggested by the following result.

Proposition 1.4.a (Comparison of Riemann and Lebesgue integrals)**.** Let f be a bounded function on $[a,b]$ with $a < b$. Then:

(a) The Riemann integral $\int_a^b f(x)\,\mathrm{d}x$ is defined (i.e., f is *Riemann integrable*) iff the set $\{x \in [a,b] : f \text{ is not continuous at } x\}$ has Lebesgue measure zero (or in short: f is continuous *almost everywhere*; see [1.5.1]).

(b) If the Riemann integral $\int_a^b f(x)\,\mathrm{d}x$ is defined, then f is Borel measurable and the Lebesgue integral $\int_{[a,b]} f(x)\,\mathrm{d}\lambda(x)$ is also defined. Furthermore, the Riemann and Lebesgue integrals agree. [Note: In view of this, sometimes we write $\int_{[a,b]} f(x)\,\mathrm{d}x$ to denote the common value of the Riemann and Lebesgue integrals.]

Proof. Omitted. □

1.4.6 **Properties of Lebesgue integral.** The Lebesgue integral turns out to possess similar properties as the Riemann integral. Let X and Y be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) If X takes only finitely many values $y_0, \dots, y_n$, then $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \sum_{k=0}^n y_k \mathbb{P}(X = y_k)$.

(b) *(integrability)* X is integrable iff $\int_\Omega |X(\omega)|\,\mathrm{d}\mathbb{P}(\omega) < \infty$.

(c) *(monotonicity)* If $\mathbb{P}(X \le Y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \le Y(\omega)\}) = 1$, and both $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ are defined, then $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \le \int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$.

(d) If $\mathbb{P}(X = Y) = 1$ and one of the integrals $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ is defined, then the other is also defined and $\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$

(e) *(linearity)* If (i) α and β are real constants and X and Y are both integrable, or (ii) α and β are *nonnegative* constants and X and Y are both *nonnegative*, then $\int_\Omega (\alpha X(\omega) + \beta Y(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \alpha\int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) + \beta\int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)$.

Proof. Due to the lack of formal definition of Lebesgue integral here, we shall omit the proof; see STAT7610 for more details. □

1.4.7 **Definition and properties of expectation.** With the knowledge of Lebesgue integral, it is straightforward to define the concept of expectation. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the **expectation** of X is defined by
$$\mathbb{E}[X] := \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

It is well-defined if (i) X is integrable (equivalently: $\mathbb{E}[|X|] = \int_\Omega |X(\omega)|\, d\mathbb{P}(\omega) < \infty$), or (ii) $\mathbb{P}(X \geq 0) = 1$. For (i), we always have $\mathbb{E}[X] < \infty$. However, for (ii) we may have $\mathbb{E}[X] = \infty$.

Since the expectation is essentially a Lebesgue integral, the properties for Lebesgue integral in [1.4.6] still apply (just using the notation "$\mathbb{E}[\cdot]$" instead of "$\int$"). But apart from them, we also have some extra.

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

(a) If X takes only finitely many values $y_0, \ldots, y_n$, then $\mathbb{E}[X] = \sum_{k=0}^n y_k \mathbb{P}(X = y_k)$.

(b) *(integrability)* X is integrable iff $\mathbb{E}[|X|] < \infty$.

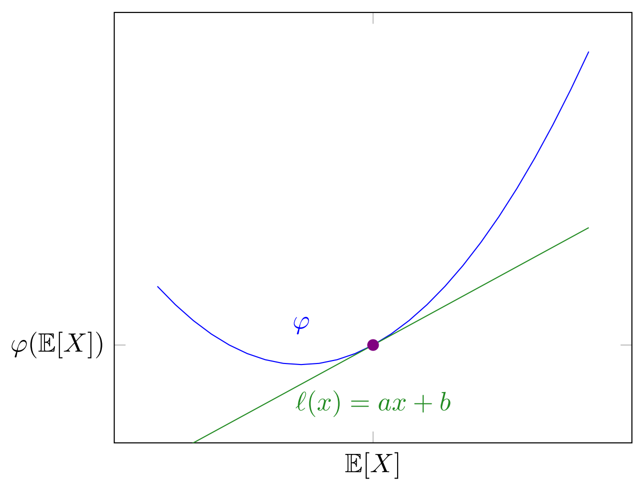
(c) *(monotonicity)* If $\mathbb{P}(X \leq Y) = 1$ and both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are defined, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(d) If $\mathbb{P}(X = Y) = 1$ and one of the expectations $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ is defined, then the other is also defined and $\mathbb{E}[X] = \mathbb{E}[Y]$.

(e) *(linearity)* If (i) α and β are real constants and X and Y are both integrable, or (ii) α and β are *nonnegative* constants and X and Y are both *nonnegative*, then $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$.

(f) *(Jensen's inequality)* If φ is a convex and real-valued function on $\mathbb{R}$, and $\mathbb{E}[|X|] < \infty$, then $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$.

Proof. Here we only prove the Jensen's inequality (based on the monotonicity and linearity).



Let $\ell(x) = ax + b$ denote a *support line* of the convex function φ at $x = \mathbb{E}[X]$, i.e., a line that lies under the graph of φ always and passes through the point $(\mathbb{E}[X], \varphi(\mathbb{E}[X]))$.[3]

Then, we have $aX(\omega) + b \leq \varphi(X(\omega))$ for all $\omega \in \Omega$. By monotonicity, we have $\mathbb{E}[aX + b] \leq \mathbb{E}[\varphi(X)]$. Applying the linearity of the expectation on the LHS (with $\alpha = a, \beta = b, Y \equiv 1$) gives $\ell(\mathbb{E}[X]) = a\mathbb{E}[X] + b \leq \mathbb{E}[\varphi(X)]$. Since the line ℓ passes through $(\mathbb{E}[X], \varphi(\mathbb{E}[X]))$, we have $\ell(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$, so we get $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ as desired. ☐

1.5 Convergence Theorems

1.5.1 **Preliminary terms.** In Section 1.5, we are going to discuss several *convergence theorems*, which give us conditions under which we can "interchange limit and integral". We first introduce some terminologies for describing the types of convergence to be investigated here.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space (e.g., μ may be the Lebesgue measure λ or a probability measure $\mathbb{P}$). Every set $N \in \mathcal{F}$ with $\mu(N) = 0$ is called a **(μ-)null set**. If a statement holds for all $\omega \in \Omega \setminus N$ where N is a null set, then it is said to hold **(μ-)almost everywhere** (a.e.), or **(μ-)almost surely** (a.s.) when μ is a probability measure.

[3] The existence of support line is guaranteed by the convexity of φ. But here we will not delve into the mathematical details about it.

1.5.2 **Almost everywhere/surely convergence.** Let $\{f_n\}$ be a sequence of real-valued (Borel-)measurable functions on $\mathbb{R}$, and let f be another real-valued (Borel-)measurable function on $\mathbb{R}$. Then we say that the sequence $\{f_n\}$ **converges to f (λ-)almost everywhere**, denoted by "$\lim_{n\to\infty} f_n = f$ (λ-)a.e.", if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in \mathbb{R} \setminus N$ where $N \in \mathcal{B}$ is a (λ-)null set (λ denotes the Lebesgue measure). [Note: With the value x fixed, we can view $\{f_n(x)\}$ as a real-valued sequence with index n, and "$\lim_{n\to\infty} f_n(x) = f(x)$" denote the usual convergence of the sequence $\{f_n(x)\}$ (to the fixed value $f(x)$), i.e., the one you have seen in your previous calculus/analysis course.]

In a similar way, we can define almost surely convergence for a sequence of random variables. Let $\{X_n\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let X be another random variable on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then we say that the sequence $\{X_n\}$ **converges to X ($\mathbb{P}$-)almost surely**, denoted by "$\lim_{n\to\infty} f_n = f$ ($\mathbb{P}$-)a.s.", if $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega \setminus N$ where $N \in \mathcal{F}$ is a ($\mathbb{P}$-)null set.

1.5.3 **Monotone convergence theorem.** Now we are ready to state the convergence theorems (without proofs as they are quite technical). For each of them, we will state a version for real-valued measurable functions on $\mathbb{R}$ and a version for random variables (like what we did above).

The first convergence theorem is known as the *monotone convergence theorem* (MCT).

Theorem 1.5.a (Monotone convergence theorem)**.**

(a) *(for functions on $\mathbb{R}$)* Let $\{f_n\}$ be a sequence of measurable real-valued functions on $\mathbb{R}$ converging almost everywhere to another measurable real-valued function f on $\mathbb{R}$. If we have $0 \le f_1 \le f_2 \le \cdots$ almost everywhere, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x.$$

(b) *(for random variables)* Let $\{X_n\}$ be a sequence of random variables converging almost surely to another random variable X. If we have $0 \le X_1 \le X_2 \le \cdots$ almost surely, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\Big[\lim_{n\to\infty} X_n\Big] = \mathbb{E}[X],$$

or more explicitly,

$$\lim_{n\to\infty} \int_{\Omega} X_n(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_{\Omega} \lim_{n\to\infty} X_n(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_{\Omega} X(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

While the conditions involved in the MCT are "$0 \le f_1 \le f_2 \le \cdots$ a.e." or "$0 \le X_1 \le X_2 \le \cdots$ a.s.", we can actually still apply the MCT to more general sequences, by making the following observations:

- If $c \le X_1 \le X_2 \le \cdots$ a.s., then we have $0 \le X_1 - c \le X_2 - c \le \cdots$ a.s., so we can apply the MCT to the sequence $\{X_n - c\}$.
- If $c \ge X_1 \ge X_2 \ge \cdots$ a.s., then we have $0 \le c - X_1 \le c - X_2 \le \cdots$ a.s., so we can apply the MCT to the sequence $\{c - X_n\}$.

[Note: It is similar for the case with functions on $\mathbb{R}$.]

1.5.4 **Dominated convergence theorem.** Another convergence theorem to be discussed here is the *dominated convergence theorem* (DCT). Instead of requiring *monotonicity* as in MCT, here we require the functions/random variables in the sequence to be bounded ("<u>dominated</u>") by a function/random variable.

Theorem 1.5.b (Dominated convergence theorem)**.**

(a) *(for functions on $\mathbb{R}$)* Let $\{f_n\}$ be a sequence of measurable real-valued functions on $\mathbb{R}$ converging almost everywhere to another measurable real-valued function f on $\mathbb{R}$. If we have $|f_n| \le g$ almost everywhere for every $n \in \mathbb{N}$, where g is a function such that $\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x < \infty$, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x.$$

(b) *(for random variables)* Let $\{X_n\}$ be a sequence of random variables converging almost surely to another random variable X. If we have $|X_n| \leq Y$ almost surely for every $n \in \mathbb{N}$, where Y is a random variable (defined on the same probability space as others) with $\mathbb{E}[Y] < \infty$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\Big[\lim_{n\to\infty} X_n\Big] = \mathbb{E}[X].$$

1.6 Computation of Expectations

1.6.1 So far we have dealt with the concept of expectations theoretically. But in practice, we are also interested in how we can *compute* expectations. Of course, you have already learnt computational formulas in your first course in probability, namely:

$$\mathbb{E}[g(X)] = \begin{cases} \sum_k g(x_k)f(x_k) & \text{if } X \text{ is discrete, taking countably many values } x_1, x_2, \ldots, \\ \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

Here we will investigate how our general and abstract definition of expectation would get reduced to these formulas in these special cases. We start by stating the following result.

Proposition 1.6.a. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a measurable function on $\mathbb{R}$. Then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x),$$

provided that $\mathbb{E}[g(X)] < \infty$.

In the case where X is discrete, taking countably many values $x_1, x_2, \ldots$, the distribution measure μ_X would then be given by $\mu_X(\{x_k\}) = \mathbb{P}(X = x_k) = f(x_k)$ for all $k = 1, 2, \ldots$, where f is the mass function. This allows us to reduce the expression $\int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x)$ to $\sum_k g(x_k)f(x_k)$, which is more familiar to us. For the case where X is continuous, there is a density function f such that $\mu_X(B) = \mathbb{P}(X \in B) = \int_B f(x)\,\mathrm{d}x$ for all $B \in \mathcal{B}$. Based on this, it can be shown that $\int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x$ (details omitted).

1.7 Change of Measure

1.7.1 We have now reached the last subsection in Section 1, which is perhaps the most important one here also, due to its foundational role for justifying the technique of *risk-neutral pricing* that is heavily used in financial economics (there we are changing the probability measure to the *risk-neutral measure*).

1.7.2 **Changing measure through a random variable Z.** The basic idea of *change of measure* is to introduce a random variable Z for "distorting" the probability measure. For illustration purpose, consider the case where Ω is countable with $\mathbb{P}(\{\omega\}) > 0$ for every $\omega \in \Omega$. Suppose we would like to change the probability measure $\mathbb{P}$ to a new probability measure $\widetilde{\mathbb{P}}$ via an introduction of a random variable Z. One natural way to do that is to define Z by $Z(\omega) = \widetilde{\mathbb{P}}(\{\omega\})/\mathbb{P}(\{\omega\})$ for all $\omega \in \Omega$, such that multiplying $\mathbb{P}(\{\omega\})$ by $Z(\omega)$ gives us the "distorted" probability $\widetilde{\mathbb{P}}(\{\omega\})$, i.e., $Z(\omega)\mathbb{P}(\{\omega\}) = \widetilde{\mathbb{P}}(\{\omega\})$, for all $\omega \in \Omega$.

However, difficult arises when Ω is uncountable and $\mathbb{P}(\{\omega\}) = \widetilde{\mathbb{P}}(\{\omega\}) = 0$ for all $\omega \in \Omega$. In such case, *any* random variable Z would satisfy $Z(\omega)\mathbb{P}(\{\omega\}) = \widetilde{\mathbb{P}}(\{\omega\})$ for all $\omega \in \Omega$, since the equation is just saying "$0 = 0$", which is not meaningful and not what we want for describing how the probability "distortion" should take place. Hence, instead of describing the "distorting" on a "per sample point ω" basis, we would describe it on a "per *event*" basis, as suggested by the following result.

Proposition 1.7.a. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $\mathbb{E}[Z] = 1$. Consider a function $\widetilde{\mathbb{P}}$ defined by

$$\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{for every } A \in \mathcal{F}. \tag{1}$$

Then $\widetilde{\mathbb{P}}$ is a probability measure on $\mathcal{F}$.

Proof.

(1) For all $A \in \mathcal{F}$, since we have $\mathbf{1}_A Z \geq 0$ almost surely, by [1.4.6] we have $\widetilde{\mathbb{P}}(A) \geq 0$. ✔

(2) We have $\widetilde{\mathbb{P}}(\Omega) = \int_\Omega Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}[Z] = 1$. ✔

(3) Fix any pairwise disjoint events $A_1, A_2, \ldots \in \mathcal{F}$. Let $B_n := \biguplus_{k=1}^n A_k$ for every $n \in \mathbb{N}$ and $B_\infty := \biguplus_{k=1}^\infty A_k$. Then $B_n \nearrow$ and so we have $0 \leq \mathbf{1}_{B_1} \leq \mathbf{1}_{B_2} \leq \cdots$, and hence $0 \leq \mathbf{1}_{B_1} Z \leq \mathbf{1}_{B_2} Z \leq \cdots$ almost surely. Using the monotone convergence theorem and the properties that for all $\omega \in \Omega$, $\lim_{n\to\infty} \mathbf{1}_{B_n}(\omega) = \mathbf{1}_{B_\infty}(\omega)$ and $\mathbf{1}_{B_n}(\omega) = \sum_{k=1}^n \mathbf{1}_{A_k}(\omega)$, we have

$$\begin{aligned}
\widetilde{\mathbb{P}}(B_\infty) &= \int_\Omega \mathbf{1}_{B_\infty}(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\Omega \lim_{n\to\infty} (\mathbf{1}_{B_n}(\omega) Z(\omega))\,\mathrm{d}\mathbb{P}(\omega) \\
&\overset{\text{(MCT)}}{=} \lim_{n\to\infty} \int_\Omega \mathbf{1}_{B_n}(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \lim_{n\to\infty} \int_\Omega \mathbf{1}_{B_n}(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= \lim_{n\to\infty} \int_\Omega \sum_{k=1}^n \mathbf{1}_{A_k}(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) \overset{\text{(linearity)}}{=} \lim_{n\to\infty} \sum_{k=1}^n \int_\Omega \mathbf{1}_{A_k}(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= \lim_{n\to\infty} \sum_{k=1}^n \widetilde{\mathbb{P}}(A_k) = \sum_{k=1}^\infty \widetilde{\mathbb{P}}(A_k). \checkmark
\end{aligned}$$

□

1.7.3 **Expectations after change of measure.** In *risk-neutral pricing*, after changing the probability measure to the *risk-neutral measure*, we would like to compute expectations under such probability measure. In view of this, here we also study the effect of changing probability measure on the expectations.

Proposition 1.7.b. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $\mathbb{E}[Z] = 1$. Let $\widetilde{\mathbb{P}}$ be defined as in Equation (1).

(a) If X is a nonnegative random variable, then $\widetilde{\mathbb{E}}[X] = \mathbb{E}[XZ]$.

(b) If $Z > 0$ almost surely, then $\mathbb{E}[Y] = \widetilde{\mathbb{E}}[Y/Z]$ for every nonnegative random variable Y.

[Note: Here $\widetilde{\mathbb{E}}[\cdot]$ denotes the expectation under the probability measure $\widetilde{\mathbb{P}}$, i.e., $\widetilde{\mathbb{E}}[X] = \int_\Omega X(\omega)\,\mathrm{d}\widetilde{\mathbb{P}}(\omega)$.]

Proof. First, note that (b) follows from (a) since in the case where $Z > 0$ and $Y \geq 0$ almost surely, the expression Y/Z can be defined (ignoring those ω's where $Z(\omega) = 0$ would not affect the calculation of expectation) and is almost surely nonnegative, so (b) follows by setting $X = Y/Z$ in (a). So we will only prove (a) in the following.

We will utilize a proof strategy known as the *standard machine*, which is a standard argument for proving results in measure theory. The **standard machine** takes four steps, which show the desired result holds for larger and larger classes of functions: (1) indicator functions ➔ (2) **simple functions** (i.e., linear combinations of finitely many indicator functions of pairwise disjoint sets) ➔ (3) nonnegative random variables ➔ (4) general random variables. As the results here are for nonnegative random variables, we only need to perform the first three steps here. (We can also perform the step 4 to extend the results to general random variables; see the remark afterwards.)

Step 1: Indicator functions. For all $A \in \mathcal{F}$, we have

$$\widetilde{\mathbb{E}}[\mathbf{1}_A] = \widetilde{\mathbb{P}}(A) = \int_\Omega \mathbf{1}_A(\omega) Z(\omega)\,\mathrm{d}\omega = \mathbb{E}[\mathbf{1}_A Z].$$

Step 2: Simple functions. Fix any pairwise disjoint events $A_1, \ldots, A_n \in \mathcal{F}$, and any constants $c_1, \ldots, c_n \in \mathbb{R}$. Then

$$\widetilde{\mathbb{E}}\left[\sum_{i=1}^n c_i \mathbf{1}_{A_i}\right] = \sum_{i=1}^n c_i \widetilde{\mathbb{E}}[\mathbf{1}_{A_i}] \overset{\text{(step 1)}}{=} \sum_{i=1}^n c_i \mathbb{E}[\mathbf{1}_{A_i} Z] = \mathbb{E}\left[\left(\sum_{i=1}^n c_i \mathbf{1}_{A_i}\right) Z\right].$$

Step 3: Nonnegative random variables. We will use the fact that, given any nonnegative random variable X, there exists a sequence of nonnegative simple functions $\{X_n\}$ such that $0 \le X_1 \le X_2 \le \cdots \le X$ a.s. and $\lim_{n\to\infty} X_n = X$ a.s. (loosely speaking, they are approaching to X "from below"). By monotone convergence theorem, we have

$$\widetilde{\mathbb{E}}[X] = \widetilde{\mathbb{E}}\Big[\lim_{n\to\infty} X_n\Big] = \lim_{n\to\infty} \widetilde{\mathbb{E}}[X_n] \overset{\text{(step 2)}}{=} \lim_{n\to\infty} \mathbb{E}[X_n Z] = \mathbb{E}\Big[\lim_{n\to\infty} (X_n Z)\Big] = \mathbb{E}[XZ].$$

□

Remarks:

- Given any general random variable X that may not be nonnegative, we can write $X = X^+ - X^-$ and apply (a) by $\widetilde{\mathbb{E}}[X] = \widetilde{\mathbb{E}}[X^+] - \widetilde{\mathbb{E}}[X^-] \overset{\text{(a)}}{=} \mathbb{E}[X^+ Z] - \mathbb{E}[X^- Z]$, as long as it does not result in "$\infty - \infty$".
- Similarly for (b), we can apply it for every general random variable Y through writing $Y = Y^+ - Y^-$: $\mathbb{E}[Y] = \widetilde{\mathbb{E}}[Y^+/Z] - \widetilde{\mathbb{E}}[Y^-/Z]$, again as long as it does not result in "$\infty - \infty$".

1.7.4 **Equivalent probability measures.** Let Ω be nonempty and $\mathcal{F}$ be a σ-algebra on Ω. Then two probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are called **equivalent** if we have $\mathbb{P}(A) = 0$ iff $\widetilde{\mathbb{P}}(A) = 0$, i.e., they assign zero probability to *exactly* the same collection of events in $\mathcal{F}$. [Note: While equivalent probability measures agree on which events in $\mathcal{F}$ have zero probability, they can still differ a lot on the probability assignments for events with positive probability.]

The following result provides a sufficient condition for two probability measures to be equivalent.

Proposition 1.7.c. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely *positive* random variable with $\mathbb{E}[Z] = 1$. Let $\widetilde{\mathbb{P}}$ be defined as in Equation (1). Then, the two probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent.

Proof. Fix any $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$. Then we have $\mathbf{1}_A Z = 0$ $\mathbb{P}$-almost surely. Hence, $\widetilde{\mathbb{P}}(A) = \int_\Omega \mathbf{1}_A(\omega) Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\Omega 0\,\mathrm{d}\mathbb{P}(\omega) = 0$.

Conversely, fix any $B \in \mathcal{F}$ with $\widetilde{\mathbb{P}}(B) = 0$. Then we have $\mathbf{1}_B/Z = 0$ $\widetilde{\mathbb{P}}$-almost surely. Hence, $\mathbb{P}(B) = \mathbb{E}[\mathbf{1}_B] \overset{\text{(Proposition 1.7.b)}}{=} \widetilde{\mathbb{E}}[\mathbf{1}_B/Z] = 0$. □

1.7.5 **Radon-Nikodym theorem and Radon-Nikodym derivatives.** The *Radon-Nikodym theorem* is a theoretical result that guarantees the existence of a random variable Z satisfying Equation (1) given two *equivalent* probability measures.

Theorem 1.7.d (Radon-Nikodym)**.** Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be equivalent probability measures on a measurable space $(\Omega, \mathcal{F})$. Then there exists an almost surely *positive* random variable Z with $\mathbb{E}[Z] = 1$ such that

$$\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{for every } A \in \mathcal{F}.$$

Proof. Omitted. □

Inspired by $\widetilde{\mathbb{P}}(A) = \int_\Omega \mathbf{1}_A\,\mathrm{d}\widetilde{\mathbb{P}}(\omega) = \int_A 1\,\mathrm{d}\widetilde{\mathbb{P}}(\omega)$ and the Radon-Nikodym theorem, we may write $Z(\omega) =: \mathrm{d}\widetilde{\mathbb{P}}(\omega)/\mathrm{d}\mathbb{P}(\omega)$ for all $\omega \in \Omega$, as a mnemonic device. (This looks somewhat similar to the equation "$Z(\omega) = \widetilde{\mathbb{P}}(\{\omega\})/\mathbb{P}(\{\omega\})$" in [1.7.2].) Indeed, such random variable Z in the Radon-Nikodym theorem is known as the **Radon-Nikodym derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$**, which is often denoted by $\mathrm{d}\widetilde{\mathbb{P}}/\mathrm{d}\mathbb{P}$.

2 Information and Conditioning

2.0.1 In financial economics, often we are dealing with *stochastic processes*, which describe how different economic variables change over time; we are handling some "dynamic" random variables. An important topic to be studied here is *hedging*, which is about adjusting our portfolio positions in response to the changes in the economic variables over time. Of course, when we are constructing our portfolio position at time t, we would only have access to the *information* available as of time t; we cannot look into the future! In view of this, it is crucial to investigate the *accrual of information* over time, to ensure that our hedging strategy would not "accidentally" use some information that cannot be accessed in reality.

Therefore, in Section 2, we will establish a rigorous framework for studying the concepts of *information* and *conditioning* ("incorporating the information on hand"), which are fundamental for our later discussions on the stochastic processes appearing in financial economics.

2.1 Information and σ-algebras

2.1.1 **Coin tossing example.** To illustrate the concept of *information*, we will consider the following example of tossing a coin three times independently, whose sample space can be expressed as

$$\begin{aligned}\Omega &= \{H,T\}^3 = \{\omega = (\omega_1,\omega_2,\omega_3) : \omega_1,\omega_2,\omega_3 \in \{H,T\}\} \\ &= \{(H,H,H),(T,H,H),(H,T,H),(H,H,T),(T,T,H),(T,H,T),(H,T,T),(T,T,T)\}\end{aligned}$$

(H: heads, T: tails). As Ω is finite, we can set the σ-algebra on Ω as $\mathcal{F} = \mathcal{P}(\Omega)$ without issues. For every $n = 0,1,2,3$, let $\mathcal{F}_n$ denote the family of all events whose occurrence can be decided (or events *resolved*) after the first n tosses, i.e.,

$$\begin{aligned}\mathcal{F}_0 &= \{\varnothing,\Omega\}, \\ \mathcal{F}_1 &= \{\{\omega\in\Omega : \omega_1 \in A\} : A \subseteq \{H,T\}\} \\ &= \{\{\omega\in\Omega : \omega_1 \in \varnothing\}, \{\omega\in\Omega : \omega_1 = H\}, \{\omega\in\Omega : \omega_1 = T\}, \{\omega\in\Omega : \omega_1 \in \{H,T\}\}\} \\ &= \{\varnothing, \{(H,H,H),(H,T,H),(H,H,T),(H,T,T)\}, \{(T,H,H),(T,T,H),(T,H,T),(T,T,T)\}, \Omega\}, \\ \mathcal{F}_2 &= \{\{\omega\in\Omega : (\omega_1,\omega_2) \in A\} : A \subseteq \{H,T\}^2\}, \\ \mathcal{F}_3 &= \mathcal{P}(\Omega).\end{aligned}$$

The families $\mathcal{F}_0,\mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3$ are all σ-algebras on Ω (check!). It is instructive to interpret these σ-algebras as representations of the available information about the true outcome ω, which is identified by what events are resolved. When there are more events resolved, more information is available and we can specify the true ω more precisely. Examples:

- The set $\mathcal{F}_0 = \{\varnothing,\Omega\}$ represents the information available initially before having any tosses. Initially we know nothing more than just "ω belongs to Ω (and does not belong to $\varnothing$)", so only $\varnothing$ and Ω appear in $\mathcal{F}_0$.
- The set $\mathcal{F}_1$ represents the information available after the first toss. Intuitively, we know the information gained by having the first toss is the outcome of the first toss; this piece of information is represented mathematically by the two sets: $\{(H,H,H),(H,T,H),(H,H,T),(H,T,T)\}$ and $\{(T,H,H),(T,T,H),(T,H,T),(T,T,T)\}$. More precisely, with this piece of information, we can determine whether the true ω lies in each of these sets: If the first toss is heads, then the true ω lies in the former set; otherwise, the true ω lies in the latter set. This allows us to specify the true ω more precisely than the mere "the true ω is in Ω".
- The set $\mathcal{F}_3 = \mathcal{P}(\Omega)$ represents the information available after all three tosses. Since we would be able to know exactly what the true ω is after the three tosses, we can determine for every subset of Ω whether the true ω lies in that set (including singletons); "full" information is available, and we can specify the true ω *exactly* by identifying the singleton to which the true ω belongs.

The feature that more information about the true ω would be gained when we toss the coin more number of times is reflected by the increasing number of events resolved upon more number of tosses:

$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$; these form a *filtration*, which is a concept used for describing the information accrual over time.

2.1.2 **Filtration.** The coin tossing example in [2.1.1] illustrates (i) how *information* can be represented mathematically (namely through the events appearing in σ-algebras, which can elucidate the true ω), and (ii) a mathematical way to represent the information accrual over time (namely through *filtration*).

Now, let us introduce some terminologies for describing information. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{F}_t \subseteq \mathcal{F}$ be a σ-algebra on Ω (such σ-algebra $\mathcal{F}_t$ is sometimes said to be a **sub-σ-algebra** of $\mathcal{F}$) for all $t \in I$, where $I \subseteq [0, \infty)$ is an index set for the time points in consideration (e.g., $I = [0, T]$ where $T > 0$ is a fixed constant). If we have $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$, then the collection $\{\mathcal{F}_t\}$ (or $\{\mathcal{F}_t\}_{t\in I}$ to be more specific) is said to be a **filtration**. In such case, the quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is called a **filtered probability space**.

At time $t \in I$, we would know for each set in $\mathcal{F}_t$ whether the true ω lies in that set, so $\mathcal{F}_t$ effectively "filters" out impossible candidates for the true ω. For instance, in the coin tossing example from [2.1.1], assuming that the first toss in heads, $\mathcal{F}_1$ would "filter out" the outcomes in $\{(T, H, H), (T, T, H), (T, H, T), (T, T, T)\}$, as we know the true ω must be in the set $\{(H, H, H), (H, T, H), (H, H, T), (H, T, T)\}$.

2.1.3 **σ-algebras generated by random variables.** Let X be a random variable. Then the **σ-algebra generated by X**, denoted by $\sigma(X)$, is given by $\sigma(X) := X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\}$. We can verify that it is indeed a σ-algebra as follows:

Proof.

(1) We have $\varnothing = X^{-1}(\underbrace{\varnothing}_{\in \mathcal{B}}) \in \sigma(X)$. ✔

(2) Fix any $A \in \sigma(X)$. Then $A = X^{-1}(B)$ for some $B \in \mathcal{B}$. Hence,

$$A^c = (X^{-1}(B))^c = X^{-1}(\underbrace{B^c}_{\in \mathcal{B}}) \in \sigma(X). ✔$$

(3) Fix any $A_1, A_2, \ldots \in \sigma(X)$. Then for all $i \in \mathbb{N}$, $A_i = X^{-1}(B_i)$ for some $B_i \in \mathcal{B}$. Thus,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} X^{-1}(B_i) = X^{-1}\Bigg(\underbrace{\bigcup_{i=1}^{\infty} B_i}_{\in \mathcal{B}}\Bigg) \in \sigma(X). ✔$$

□

2.1.4 **Understanding $\sigma(X)$ from the information perspective.** From the information perspective, $\sigma(X)$ is a special kind of σ-algebra that represents the information about the true ω available from the random variable X, or more precisely, from knowing the value $X(\omega)$ where ω is the true outcome (but not the value of the true ω itself). To understand this better, we will revisit the coin tossing example in [2.1.1].

To relate with financial economics more closely, we can associate the three coin tosses in [2.1.1] with a *three-period binomial tree*, where the outcomes in Ω represents the paths taken in the three periods, say H represents an up move and T represents a down move. Let us fix the parameters as follows:

- *up and down factors:* $u = 1.2$ and $d = 0.8$,
- *initial stock price:* $S_0 = 100$.

Now, let us consider the time-2 stock price S_2 for example. We may treat S_2 as a random variable (a function from Ω to $\mathbb{R}$), defined by:

$$\begin{aligned}
S_2(H, H, H) = S_2(H, H, T) &= 120 \ (uu \text{ node}),\\
S_2(H, T, H) = S_2(H, T, T) = S_2(T, H, H) = S_2(T, H, T) &= 96 \ (ud \text{ node}),\\
S_2(T, T, H) = S_2(T, T, T) &= 80 \ (dd \text{ node}).
\end{aligned}$$

(Note that the value taken by S_2 depends only on the first two entries of the outcome ω.)

Let $A_{HH}=\{(H,H,H),(H,H,T)\}$, $A_{HT}=\{(H,T,H),(H,T,T)\}$, $A_{TH}=\{(T,H,H),(T,H,T)\}$, and $A_{TT}=\{(T,T,H),(T,T,T)\}$. Then, note that:

$$S_2^{-1}(\{120\})=A_{HH},\quad S_2^{-1}(\{96\})=A_{HT}\cup A_{TH},\quad S_2^{-1}(\{80\})=A_{TT}.$$

Intuitively, this means that:

- if we observe $S_2(\omega)=120$, then we know $\omega\in A_{HH}$ (both heads in the first two tosses);
- if we observe $S_2(\omega)=96$, then we know $\omega\in A_{HT}\cup A_{TH}$ (one heads and one tails in the first two tosses);
- if we observe $S_2(\omega)=80$, then we know $\omega\in A_{TT}$ (both tails in the first two tosses)

(ω is the true outcome). Thus, by knowing the value of $S_2(\omega)$, we can determine whether the true ω lies in each of the sets A_{HH}, $A_{HT}\cup A_{TH}$, and A_{TT}, i.e., these sets are resolved. Of course $\varnothing$ and Ω are also resolved, so do all unions and complements of these sets. These resolved sets altogether would then form the σ-algebra $\sigma(S_2)$:

$$\sigma(S_2)=\{\varnothing,A_{HH},A_{HT}\cup A_{TH},A_{TT},A_{HH}\cup A_{HT}\cup A_{TH},A_{HT}\cup A_{TH}\cup A_{TT},A_{HH}\cup A_{TT},\Omega\}$$

[Note: We have $S_2^{-1}(\varnothing)=\varnothing$, $S_2^{-1}(\{96,120\})=A_{HH}\cup A_{HT}\cup A_{TH}$, $S_2^{-1}(\{80,96\})=A_{HT}\cup A_{TH}\cup A_{TT}$, $S_2^{-1}(\{80,120\})=A_{HT}\cup A_{TT})$, and $S_2^{-1}(\{80,96,120\})=\Omega$.]

2.1.5 **Understanding measurability from the information perspective.** Continuing the example above, note that every set in $\sigma(S_2)$ is also in $\mathcal{F}_2$ in the coin tossing example, because the occurrence of each set in $\sigma(S_2)$ can be decided (and thus also the value taken by S_2) after the first two coin tosses (two periods in three-period binomial tree). This suggests that $S_2^{-1}(\mathcal{B})=\sigma(S_2)\subseteq\mathcal{F}_2$ which, by our definition of measurability in Section 1.3, means that S_2 is $\mathcal{F}_2$-measurable. Also, it is clear by definition that S_2 is always $\sigma(S_2)$-measurable and $\sigma(S_2)$ is the smallest σ-algebra under which S_2 is measurable (this applies for random variable in general).

Identifying $\sigma(S_2)=S_2^{-1}(\mathcal{B})$ as the representation of the information available from the random variable S_2, if S_2 is $\mathcal{G}$-measurable (i.e., $\sigma(S_2)=S_2^{-1}(\mathcal{B})\subseteq\mathcal{G}$), then it means that $\mathcal{G}$ contains all information from $\sigma(S_2)$. Noting that $\sigma(S_2)$ is the smallest σ-algebra that contains all the sets needed to be resolved in order to know exactly the value taken by S_2[4], having the information from $\mathcal{G}$ would be sufficient for determining the value $S_2(\omega)$.

In general, if a random variable X is $\mathcal{G}$-measurable, it means that the information represented by $\mathcal{G}$ is sufficient for determining the value $X(\omega)$ where $\omega\in\Omega$ is the true outcome.

On the other hand, if X is *not* $\mathcal{G}$-measurable (i.e., $\sigma(X)=X^{-1}(\mathcal{B})\not\subseteq\mathcal{G}$), then it means that some information needed for determining $X(\omega)$ is missing from the information represented by $\mathcal{G}$ and knowing the information from $\mathcal{G}$ alone is *not* enough for determining $X(\omega)$ exactly (but the information from $\mathcal{G}$ may help *inferring* $X(\omega)$; see Section 2.3). For instance, in the example above, S_2 is not $\mathcal{F}_1$-measurable as $\mathcal{F}_1=\{\varnothing,A_H,A_T,\Omega\}\not\subseteq\sigma(S_2)$, where $A_H=\{(H,H,H),(H,T,H),(H,H,T),(H,T,T)\}$ and $A_T=\{(T,H,H),(T,T,H),(T,H,T),(T,T,T)\}$. More intuitively, knowing only the outcome of the first toss (whether we have an "up move" or a "down move" in the first period) is not enough for determining exactly the value of $S_2(\omega)$, which is at the end of the *second* period. [Note: However, assuming the first toss is heads, we know either A_{HH} or A_{HT} would occur, so $S_2(\omega)=96$ or 120; we have narrowed down the possible values of $S_2(\omega)$. This illustrates how the information from $\mathcal{F}_1$ could help inferring the value $S_2(\omega)$, although it is not sufficient for exactly determining the value.]

2.1.6 **Adapted stochastic processes.** In the discussion of hedging, we will work in a filtered probability space where the filtration $\{\mathcal{F}_t\}_{t\in I}$ would be treated as a model of the accrual of public information available in the market. For our time-t hedging strategy to be sensible, the portfolio position taken

[4] We must be able to determine whether the true ω lies in each of A_{HH}, $A_{HT}\cup A_{TH}$, and A_{TT}. The smallest σ-algebra containing them is indeed $\sigma(S_2)$.

$\Delta_t(\omega)$ at time t should be determinable from the information available at time t, i.e., $\sigma(\Delta_t) \subseteq \mathcal{F}_t$. It means that Δ_t should be $\mathcal{F}_t$-measurable for all time $t \in I$. This gives rise to the definition of *adapted stochastic process*.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, \mathbb{P})$ be a filtered probability space. Then a stochastic process $\{X_t\}_{t\in I}$ is said to be **adapted (to the filtration $\{\mathcal{F}_t\}_{t\in I}$)** if X_t is $\mathcal{F}_t$-measurable for all time $t \in I$. [Note: $\{X_t\}_{t\in I}$ is a **stochastic process** if X_t is a $\mathcal{F}$-random variable for all $t \in I$.]

2.2 Independence

2.2.1 Recall that if a random variable X is $\mathcal{G}$-measurable, then the information represented by $\mathcal{G}$ is sufficient for determining exactly the value $X(\omega)$ where $\omega \in \Omega$ is the true outcome. Another extreme would be the case where the information represented by $\mathcal{G}$ is of *no use* for determining the value $X(\omega)$; it does not even marginally help estimating $X(\omega)$. This case corresponds to the concept of *independence*. While the concept of measurability has nothing to do with probability measure (we have made no reference to $\mathbb{P}$ in the definition), the probability measure $\mathbb{P}$ is crucial in the definition of independence, and so the choice of $\mathbb{P}$ would affect whether we have independence or not.

Intuitively, probability measure plays no role for measurability since this concept is about *exactness*: whether we can determine the value $X(\omega)$ *exactly*; it carries no probabilistic meaning. On the other hand, independence requires the information to be of *no use* for inferring the value $X(\omega)$, so we need to ensure that the information would not "indirectly" help inferring the value $X(\omega)$ by making some impacts on the *probability assessments* about X.

2.2.2 **Independence of σ-algebras.** In your first probability course, you should have learnt the concept of independence for *events*: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the events $A, B \in \mathcal{F}$ are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Intuitively, this means that knowing the true outcome ω is in A does not affect our assessment on the probability that ω is in B, and vice versa.

We would like to extend this idea to random variables also. Intuitively, if two random variables X and Y are independent, then having knowledge about X should not affect our assessment on the probabilities about Y, and vice versa. (You may have also learnt the independence of random variables in your first probability course.)

It turns out that, by utilizing the concepts of *σ-algebras*, we can neatly unify these two kinds of independence together, and we can focus on discussing about the independence of σ-algebras in general. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ be σ-algebras on Ω. Then the σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are called **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathcal{G}, B \in \mathcal{H}$.

To illustrate how this independence unifies the notions of independence of events and independence of random variables, consider the following.

- *(independence of events)* Given two events $A, B \in \mathcal{F}$, we note that A and B are independent iff the σ-algebras $\sigma(\{A\}) = \{\varnothing, A, A^c, \Omega\}$ and $\sigma(\{B\}) = \{\varnothing, B, B^c, \Omega\}$ are independent (they are *σ-algebras generated by families of sets in* Ω).
- *(independence of random variables)* Let X and Y be random variables. They are said to be independent if $\mathbb{P}(X \in C \cap Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D)$ for all *Borel subsets* C and D of $\mathbb{R}$[5]. Then, X and Y are independent iff the σ-algebras $\sigma(X) = X^{-1}(\mathcal{B})$ and $\sigma(Y) = Y^{-1}(\mathcal{B})$ are independent.

2.2.3 Here we give an example about the independence of random variables. Consider the previous three-period binomial tree example. We let S_3 denote the time-3 stock price, which can be treated as a

[5] The condition that the sets C and D involved should be *Borel subsets* of $\mathbb{R}$ may not be emphasized in your first probability course. It ensures that we can assign probabilities to $\{X \in C\}$ and $\{X \in D\}$, and also their intersection.

random variable defined by:

$$\begin{aligned}
S_3(H,H,H) &= 144 \ (uuu \text{ node}),\\
S_3(H,H,T) &= S_3(H,T,H) = S_3(T,H,H) = 115.2 \ (uud \text{ node}),\\
S_3(H,T,T) &= S_3(T,H,T) = S_3(T,T,H) = 76.8 \ (udd \text{ node}),\\
S_3(T,T,T) &= 51.2 \ (ddd \text{ node}).
\end{aligned}$$

We would then like to investigate whether S_2 and S_3 are independent. Intuition tells us that they should not be independent: For instance, if we know that $S_2 = 120$, then S_3 cannot possibly take the values of 76.8 or 51.2 anymore. So, knowledge about S_2 *should* affect our assessment on the probabilities about S_3. This is indeed the case and we can show it by considering the events $\{(H,H,H),(H,H,T)\} = \{S_2 = 120\} \in \sigma(S_2)$ and $\{(H,H,H)\} = \{S_3 = 144\} \in \sigma(S_3)$. We have $\mathbb{P}(\{S_2 = 120\} \cap \{S_3 = 144\}) = \mathbb{P}(\{(H,H,H)\}) = p^3$, while $\mathbb{P}(S_2 = 120)\mathbb{P}(S_3 = 144) = \mathbb{P}(\{(H,H,H),(H,H,T)\})\mathbb{P}(\{(H,H,H)\}) = p^2 \cdot p^3 = p^5$ (where p is the probability of getting heads in a coin toss, assumed to be positive and the same for all three *independent* tosses). To understand this more intuitively, the probability for $S_3 = 144$ when we know $S_2 = 120$ is the *conditional probability*

$$\mathbb{P}(\{S_3 = 144\}|\{S_2 = 120\}) = \frac{\mathbb{P}(\{S_2 = 120\} \cap \{S_3 = 144\})}{\mathbb{P}(\{S_2 = 120\})} = \frac{p^3}{p^2} = p,$$

but without this piece of knowledge, the probability for $S_3 = 144$ is instead $\mathbb{P}(\{S_3 = 144\}) = p^3$. In other words, having this piece of knowledge raises the probability for $S_3 = 144$ from p^3 to p.

On the other hand, S_2 and S_3/S_2 are independent. This can be verified by considering the σ-algebras:

$$\sigma(S_2) = \{\varnothing, A_{HH}, A_{HT} \cup A_{TH}, A_{TT}, A_{HH} \cup A_{HT} \cup A_{TH}, A_{HT} \cup A_{TH} \cup A_{TT}, A_{HH} \cup A_{TT}, \Omega\},$$

and

$$\sigma(S_3/S_2) = \{\varnothing, (S_3/S_2)^{-1}(\{1.2\}), (S_3/S_2)^{-1}(\{0.8\}), \Omega\} = \{\varnothing, A_{\bullet\bullet H}, A_{\bullet\bullet T}, \Omega\},$$

where $A_{\bullet\bullet H} = \{(H,H,H),(H,T,H),(T,H,H),(T,T,H)\}$ and $A_{\bullet\bullet T} = \{(H,H,T),(H,T,T),(T,H,T),(T,T,T)\}$.

2.2.4 **More general definitions of independence.** We can extend the definition for independence of two σ-algebras in a natural way. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{G}_1, \mathcal{G}_2, \ldots \subseteq \mathcal{F}$ be σ-algebras on Ω. Then the σ-algebras $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are called **independent** if $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\cdots\mathbb{P}(A_n)$ for all $A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$. Also, the sequence of σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ is called **independent** if the n σ-algebras $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent for every $n \in \mathbb{N}$.

Let $X_1, X_2, \ldots$ be random variables. The random variables $X_1, \ldots, X_n$ are said to be independent if $\mathbb{P}(X_1 \in B_1 \cap \cdots \cap X_n \in B_n) = \mathbb{P}(X_1 \in B_1)\cdots\mathbb{P}(X_n \in B_n)$ for all *Borel subsets* $B_1, \ldots, B_n$ of $\mathbb{R}$. Also, the sequence of random variable $X_1, X_2, \ldots$ is called independent if the n random variables $X_1, \ldots, X_n$ are independent for every $n \in \mathbb{N}$. It can then be shown that $X_1, \ldots, X_n$ are independent iff $\sigma(X_1), \ldots, \sigma(X_n)$ are independent, and the sequence $X_1, X_2, \ldots$ is independent iff the sequence $\sigma(X_1), \sigma(X_2), \ldots$ is independent.

Sometimes we also talk about independence of a random variable and a σ-algebra, which is defined as follows. A random variable X and a σ-algebra $\mathcal{G}$ are called **independent** if $\sigma(X)$ and $\mathcal{G}$ are independent.

2.2.5 **Some results about independence.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(a) *(measurable functions of random variables are independent)* Let X and Y be independent random variables, and f and g be measurable real-valued functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.

[Note: More generally, if $X_1, X_2, \ldots$ are independent random variables and $f_1, f_2, \ldots$ are measurable real-valued functions on $\mathbb{R}$, then for all $1 \le r < s < t < \cdots$, $f_1(X_1, \ldots, X_r)$, $f_2(X_{r+1}, \ldots, X_s)$, $f_3(X_{s+1}, \ldots, X_t), \ldots$ are independent random variables.]

Proof. From [1.3.3] we know that $f(X)$ and $g(Y)$ are random variables, so it suffices to show that they are independent, or $\sigma(f(X))$ and $\sigma(f(Y))$ are independent.

Fix any sets $A \in \sigma(f(X))$ and $B \in \sigma(g(Y))$. Since $A \in \sigma(f(X))$, there is $C \in \mathcal{B}$ such that $A = \{\omega \in \Omega : f(X(\omega)) \in C\}$. Letting $D = \{x \in \mathbb{R} : f(x) \in C\} \in \mathcal{B}$ (as f is measurable), we have

$$A = \{\omega \in \Omega : f(X(\omega)) \in C\} = \{\omega \in \Omega : X(\omega) \in D\}$$

(since $f(X(\omega)) \in C \iff X(\omega) \in D$ by construction of D). Hence $A \in \sigma(X)$. Similarly, we can show that $B \in \sigma(Y)$. Now, since X and Y are independent, we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, as desired. ☐

(b) *(expectation of product is product of expectations)* If X and Y are independent and integrable random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(c) *(equivalent characterization of independence of random variables)* Let X and Y be random variables. The **joint distribution measure** of (X, Y) is defined by

$$\mu_{X,Y}(C) = \mathbb{P}(\{(X, Y) \in C\}) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in C\})$$

for all $C \in \mathcal{B}(\mathbb{R}^2)$, where $\mathcal{B}(\mathbb{R}^2)$ denotes the Borel σ-algebra on $\mathbb{R}^2$[6]. Then X and Y are independent iff $\mu_{X,Y}(A \times B) = \mu_X(A)\mu_Y(B)$ for all $A, B \in \mathcal{B}(\mathbb{R})$.

2.3 Conditional Expectations

2.3.1 After studying the concept of *information*, the next important topic to be studied in Section 2 is *conditioning*, which is about incorporating information available in probabilistic calculations. Based on our previous discussion, we know that:

- If a random variable X is $\mathcal{G}$-measurable, then the information from the σ-algebra $\mathcal{G}$ is sufficient for determining the value of $X(\omega)$.
- If a random variable X is independent of the σ-algebra $\mathcal{G}$, then the information from $\mathcal{G}$ should be of no use for inferring the value of $X(\omega)$ (more precisely, we have $\mathbb{P}(\{X \in B\} \cap C) = \mathbb{P}(X \in B)\mathbb{P}(C)$ for all $B \in \mathcal{B}$ and $C \in \mathcal{G}$, as $\sigma(X)$ and $\mathcal{G}$ are independent).

In Section 2.3, we are going to investigate the "middle case", where the information from the σ-algebra $\mathcal{G}$ may not be sufficient for determining the value of $X(\omega)$, but may help inferring it. An estimation of the value $X(\omega)$ based on the information from $\mathcal{G}$ (as a function of $\omega \in \Omega$, and indeed a random variable) is known as the *conditional expectation of X given $\mathcal{G}$*, denoted by $\mathbb{E}[X|\mathcal{G}]$.

2.3.2 **Motivation.** To motivate the definition of conditional expectation, we again consider the three-period binomial tree example. Here we would like to investigate how the information from $\mathcal{F}_2$ (information available after two tosses, or two periods) can be incorporated in calculating the expectation of S_3; more colloquially, we would like to compute the expectation of S_3 at time 2.

Intuitively, the conditional expectation $\mathbb{E}[S_3|\mathcal{F}_2]$ should be given by

$$\mathbb{E}[S_3|\mathcal{F}_2](\omega) = \begin{cases} pS_3(H,H,H) + (1-p)S_3(H,H,T) & \text{if } \omega \in A_{HH} = \{(H,H,H),(H,H,T)\}, \\ pS_3(H,T,H) + (1-p)S_3(H,T,T) & \text{if } \omega \in A_{HT} = \{(H,T,H),(H,T,T)\}, \\ pS_3(T,H,H) + (1-p)S_3(T,H,T) & \text{if } \omega \in A_{TH} = \{(T,H,H),(T,H,T)\}, \\ pS_3(T,T,H) + (1-p)S_3(T,T,T) & \text{if } \omega \in A_{TT} = \{(T,T,H),(T,T,T)\}. \end{cases}$$

The intuitive idea is that, based on the information from $\mathcal{F}_2$, we know that one of these cases must hold (we can determine if the true ω lies in each of these sets). Then, the formula above estimates the value of $S_3(\omega)$ through taking average over all the possible candidates of ω (i.e., the outcomes in the set containing ω) weighted by the conditional probabilities. For example, if we know $\omega \in$

[6] We shall omit the definition of $\mathcal{B}(\mathbb{R}^2)$ here, but we can treat $\mathcal{B}(\mathbb{R}^2)$ to contain all "normal" subsets of $\mathbb{R}^2$. Particularly, if we have $A, B \in \mathcal{B}$, then we always have $A \times B \in \mathcal{B}(\mathbb{R}^2)$.

A_{HH}, then the conditional probability of having the outcome (H,H,H) is $\mathbb{P}(\{(H,H,H)\}|A_{HH}) = \mathbb{P}(\{(H,H,H)\})/\mathbb{P}(A_{HH}) = p^3/p^2 = p$, and similarly the conditional probability of having (H,H,T) is $1-p$. Hence, the conditional expectation would take the value $\mathbb{E}[S_3|\mathcal{F}_2](\omega) = pS_3(H,H,H) + (1-p)S_3(H,H,T)$.

Multiplying the probability of each of the events $A_{HH}, A_{HT}, A_{TH}, A_{TT}$ in the corresponding case from the formula above gives

$$
\begin{aligned}
\mathbb{E}[S_3|\mathcal{F}_2](\omega)\mathbb{P}(A_{HH}) &= S_3(H,H,H)\mathbb{P}(\{(H,H,H)\}) + (1-p)S_3(H,H,T)\mathbb{P}(\{H,H,T\}) && \text{if } \omega \in A_{HH}, \\
\mathbb{E}[S_3|\mathcal{F}_2](\omega)\mathbb{P}(A_{HT}) &= S_3(H,T,H)\mathbb{P}(\{(H,T,H)\}) + (1-p)S_3(H,T,T)\mathbb{P}(\{H,T,T\}) && \text{if } \omega \in A_{HT}, \\
\mathbb{E}[S_3|\mathcal{F}_2](\omega)\mathbb{P}(A_{TH}) &= S_3(T,H,H)\mathbb{P}(\{(T,H,H)\}) + (1-p)S_3(T,H,T)\mathbb{P}(\{T,H,T\}) && \text{if } \omega \in A_{TH}, \\
\mathbb{E}[S_3|\mathcal{F}_2](\omega)\mathbb{P}(A_{TT}) &= S_3(T,T,H)\mathbb{P}(\{(T,T,H)\}) + (1-p)S_3(T,T,T)\mathbb{P}(\{T,T,T\}) && \text{if } \omega \in A_{TT}.
\end{aligned}
$$

To generalize the idea, we represent these equations more abstractly through *Lebesgue integrals*:

$$
\begin{aligned}
\int_{A_{HH}} \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{A_{HH}} S_3(\omega)\,\mathrm{d}\mathbb{P}(\omega), \\
\int_{A_{HT}} \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{A_{HT}} S_3(\omega)\,\mathrm{d}\mathbb{P}(\omega), \\
\int_{A_{TH}} \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{A_{TH}} S_3(\omega)\,\mathrm{d}\mathbb{P}(\omega), \\
\int_{A_{TT}} \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{A_{TT}} S_3(\omega)\,\mathrm{d}\mathbb{P}(\omega).
\end{aligned}
$$

[Note: Since $\mathbb{E}[S_3|\mathcal{F}_2](\omega)$ takes constant value for all $\omega \in A_{HH}$, we have $\int_{A_{HH}} \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}[S_3|\mathcal{F}_2](\omega)\mathbb{P}(A_{HH})$, and similarly for others.]

With these equations, we can show that

$$\int_A \mathbb{E}[S_3|\mathcal{F}_2](\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_A S_3(\omega)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F}_2,$$

which is known as the *partial averaging property*, since, as we see above, it originates from computing the conditional expectation via averaging over "parts" of Ω, weighted by probabilities.

2.3.3 **Definition of conditional expectation.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω, and X be a random variable that integrable. The **conditional expectation of X given $\mathcal{G}$**, denoted by $\mathbb{E}[X|\mathcal{G}]$, is *any* random variable satisfying:

(1) *(measurability)* $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable.

(2) *(partial averaging)*

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{G}.$$

Remarks:

- We require $\mathbb{E}[X|\mathcal{G}]$ to be $\mathcal{G}$-measurable since the information from $\mathcal{G}$ should be sufficient for determining/computing the value of $\mathbb{E}[X|\mathcal{G}](\omega)$, or in other words, the computation of such conditional expectation must not utilize information that is *unavailable* based on $\mathcal{G}$.
- If we have $\mathcal{G} = \sigma(W)$ for some random variable W, then we usually write $\mathbb{E}[X|W]$ in place of $\mathbb{E}[X|\sigma(W)]$.

2.3.4 **Properties of conditional expectations.** Based on the abstract definition of conditional expectation above, we can prove the following properties of conditional expectations; some of which are generalizations to the properties of conditional expectation you have seen in your first probability course.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω. Capital roman letters like X, Y, and Z always denote random variables below.

(a) *(existence and uniqueness)* Given any integrable random variable X, there always exists a conditional expectation $\mathbb{E}[X|\mathcal{G}]$ (i.e., a random variable satisfying both the *measurability* and *partial averaging* requirements). Moreover, such random variable is unique in *almost sure* sense, i.e., if Y and Z both qualify to be a conditional expectation of X given $\mathcal{G}$, then $Y \overset{\text{a.s.}}{=} Z$.

[Note: Due to the uniqueness in almost sure sense, formulas for conditional expectations below are always understood in the almost sure sense: "$=$" refers to "$\overset{\text{a.s.}}{=}$".]

(b) *(linearity)* If X and Y are integrable random variables, and c_1 and c_2 are constants, then $\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$.

(c) *(monotonicity)* If $X \le Y$ a.s. with X and Y being integrable, then $\mathbb{E}[X|\mathcal{G}] \le \mathbb{E}[Y|\mathcal{G}]$ a.s.

(d) *(taking out what is known (TOWIK))* If Y and XY are integrable, and X is $\mathcal{G}$-measurable, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$. [Note: Particularly, we have $\mathbb{E}[X|\mathcal{G}] = X$ by letting $Y \equiv 1$.]

(e) *(independence)* If X is integrable and is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

(f) *(tower property)* If $\mathcal{H} \subseteq \mathcal{G}$ is a σ-algebra on Ω and X is integrable, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$. [Note: In fact, we also have $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$, but this is just a simple corollary of the TOWIK property: Since $\sigma(\mathbb{E}[X|\mathcal{H}]) \subseteq \mathcal{H} \subseteq \mathcal{G}$, $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{G}$-measurable and thus $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$.]

(g) *(iterated conditioning)* If X is integrable, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

[Note: Both $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]]$ and $\mathbb{E}[X]$ are non-random values, so the equality is exact rather than in the almost sure sense.]

(h) *(conditional Jensen's inequality)* If φ is a convex and real-valued function on $\mathbb{R}$ and $X, \varphi(X)$ are integrable, then $\varphi(\mathbb{E}[X|\mathcal{G}]) \le \mathbb{E}[\varphi(X)|\mathcal{G}]$ a.s.

(i) *(independence lemma)* Suppose $X_1, \dots, X_k$ are $\mathcal{G}$-measurable and each of $Y_1, \dots, Y_\ell$ is independent of $\mathcal{G}$. Let $f(x_1, \dots, x_k, y_1, \dots, y_\ell)$ be a measurable[7] function and define $g(x_1, \dots, x_k) := \mathbb{E}[f(x_1, \dots, x_k, Y_1, \dots, Y_\ell)]$. Suppose that $f(X_1, \dots, X_k, Y_1, \dots, Y_\ell)$ is integrable. Then,

$$\mathbb{E}[f(X_1, \dots, X_k, Y_1, \dots, Y_\ell)|\mathcal{G}] = g(X_1, \dots, X_k).$$

[Intuition 💡: With $X_1, \dots, X_k$ being $\mathcal{G}$-measurable, they may be treated as "constants" given $\mathcal{G}$, so they remain in the inputs of g (and are held fixed in the evaluation of the expectation from the function g). On the other hand, since $Y_1, \dots, Y_\ell$ are independent of $\mathcal{G}$, they are "averaged out" in the (ordinary) expectation $\mathbb{E}[\cdot]$ (see the definition of g).]

Proof.

(a) Omitted.

(b) With $\mathbb{E}[X|\mathcal{G}]$ and $\mathbb{E}[Y|\mathcal{G}]$ being $\mathcal{G}$-measurable, $c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$ is also $\mathcal{G}$-measurable. Next, for all $A \in \mathcal{G}$, we have

$$\begin{aligned}\int_A (c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}])\,\mathrm{d}\mathbb{P} &= c_1\int_A \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} + c_2\int_A \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \\ &\overset{\text{(partial averaging)}}{=} c_1\int_A X\,\mathrm{d}\mathbb{P} + c_2\int_A Y\,\mathrm{d}\mathbb{P} \\ &\overset{\text{(linearity)}}{=} \int_A (c_1X + c_2Y)\,\mathrm{d}\mathbb{P},\end{aligned}$$

thus $c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$ also satisfies the partial averaging property. This then equals the conditional expectation $\mathbb{E}[c_1X + c_2Y|\mathcal{G}]$ almost surely, by the uniqueness of conditional expectation.

(c) Consider the set $A = \{\omega \in \Omega : \mathbb{E}[Y|\mathcal{G}](\omega) - \mathbb{E}[X|\mathcal{G}](\omega) < 0\} \in \mathcal{G}$. By the partial averaging property and linearity, we have

$$\int_A (\mathbb{E}[Y|\mathcal{G}](\omega) - \mathbb{E}[X|\mathcal{G}](\omega))\,\mathrm{d}\mathbb{P}(\omega) = \int_A Y(\omega)\,\mathrm{d}\mathbb{P}(\omega) - \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \overset{(X\le Y\text{ a.s.})}{\ge} 0.$$

[7]Here, "measurable" refers to $\mathcal{B}(\mathbb{R}^{k+\ell})$-measurable, where $\mathcal{B}(\mathbb{R}^{k+\ell})$ is the Borel σ-algebra on $\mathbb{R}^{k+\ell}$; here we will delve into the details about it.

By construction of A, the integrand is always positive. Hence, we have $\mathbb{P}(A) = \mathbb{P}(\mathbb{E}[X|\mathcal{G}] > \mathbb{E}[Y|\mathcal{G}]) = 0$, and thus $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$ a.s.

(d) As X and $\mathbb{E}[Y|\mathcal{G}]$ are $\mathcal{G}$-measurable, $X\mathbb{E}[Y|\mathcal{G}]$ is also $\mathcal{G}$-measurable. To show the partial averaging property, we utilize the 4-step *standard machine*.

Step 1: Indicator functions. Fix any $B \in \mathcal{G}$, and consider the indicator function $\mathbf{1}_B$. For all $A \in \mathcal{G}$, we have

$$\begin{aligned}\int_A \mathbf{1}_B \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &= \int_\Omega \mathbf{1}_A \mathbf{1}_B \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_\Omega \mathbf{1}_{A\cap B} \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \\ &= \int_{A\cap B} \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \overset{\text{(partial averaging)}}{=} \int_{A\cap B} Y\,\mathrm{d}\mathbb{P} \\ &= \int_A \mathbf{1}_B Y\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Step 2: Simple random variables. Fix any pairwise disjoint $B_1, \dots, B_n \in \mathcal{G}$, and any constants $c_1, \dots, c_n \in \mathbb{R}$. For all $A \in \mathcal{G}$, we have

$$\begin{aligned}\int_A \left(\sum_{i=1}^n c_i \mathbf{1}_{B_i}\right) \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &\overset{\text{(linearity)}}{=} \sum_{i=1}^n c_i \int_A \mathbf{1}_{B_i} \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \overset{\text{(step 1)}}{=} \sum_{i=1}^n c_i \int_A \mathbf{1}_{B_i} Y\,\mathrm{d}\mathbb{P} \\ &\overset{\text{(linearity)}}{=} \int_A \sum_{i=1}^n c_i \mathbf{1}_{B_i} Y\,\mathrm{d}\mathbb{P}\end{aligned}$$

Step 3: Nonnegative random variables. Let X be any nonnegative random variable, and $\{X_n\}$ be a sequence of nonnegative simple functions such that $0 \leq X_1 \leq X_2 \leq \cdots \leq X$ a.s. and $\lim_{n\to\infty} X_n = X$ a.s.

First consider the special case where Y is nonnegative, which implies by monotonicity that $\mathbb{E}[Y|\mathcal{G}] \geq 0$ a.s., to ensure the applicability of monotone convergence theorem. In this case, for all $A \in \mathcal{G}$, we have

$$\begin{aligned}\int_A X\mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &= \int_A \lim_{n\to\infty} X_n \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \overset{\text{(MCT)}}{=} \lim_{n\to\infty} \int_A X_n \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \\ &\overset{\text{(partial averaging)}}{=} \lim_{n\to\infty} \int_A X_n Y\,\mathrm{d}\mathbb{P} \overset{\text{(MCT)}}{=} \int_A \left(\lim_{n\to\infty} X_n\right) Y\,\mathrm{d}\mathbb{P} = \int_A XY\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Next, consider the general case where Y is integrable. In this case we write $Y = Y^+ - Y^-$. Of course, the positive and negative parts themselves are integrable random variables also, so applying the linearity of conditional expectation gives $\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[Y^+|\mathcal{G}] - \mathbb{E}[Y^-|\mathcal{G}]$.

With XY being integrable, for all $A \in \mathcal{G}$ we have

$$\begin{aligned}\int_A X\mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &= \int_A X\mathbb{E}\left[Y^+|\mathcal{G}\right]\mathrm{d}\mathbb{P} - \int_A X\mathbb{E}\left[Y^-|\mathcal{G}\right]\mathrm{d}\mathbb{P} \overset{\text{(above)}}{=} \int_A XY^+\,\mathrm{d}\mathbb{P} - \int_A XY^-\,\mathrm{d}\mathbb{P} \\ &\overset{\text{(linearity)}}{=} \int_A X(Y^+ - Y^-)\,\mathrm{d}\mathbb{P} = \int_A XY\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Step 4: General random variables. Let X be an integrable random variable in general. Then we write $X = X^+ - X^-$. After that, for all $A \in \mathcal{G}$ we have

$$\begin{aligned}\int_A X\mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} &= \int_A X^+\mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} - \int_A X^-\mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P} \overset{\text{(step 3)}}{=} \int_A X^+Y\,\mathrm{d}\mathbb{P} - \int_A X^-Y\,\mathrm{d}\mathbb{P} \\ &= \int_A (X^+ - X^-)Y\,\mathrm{d}\mathbb{P} = \int_A XY\,\mathrm{d}\mathbb{P}.\end{aligned}$$

This completes the standard machine and we have established the partial averaging property. The result then follows from the uniqueness of conditional expectation.

(e) First, $\mathbb{E}[X]$ is $\mathcal{G}$-measurable, as a deterministic constant. Next, for all $A \in \mathcal{G}$, we have

$$\int_A \mathbb{E}[X]\,d\mathbb{P} = \mathbb{E}[X]\mathbb{P}(A) = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A] \overset{(X \perp\!\!\!\perp \mathcal{G} \implies X \perp\!\!\!\perp \mathbf{1}_A)}{=} \mathbb{E}[X\mathbf{1}_A] = \int_\Omega X\mathbf{1}_A\,d\mathbb{P} = \int_A X\,d\mathbb{P}.$$

Thus $\mathbb{E}[X|\mathcal{G}] \overset{\text{a.s.}}{=} \mathbb{E}[X]$.

(f) First, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable, as it is a conditional expectation given $\mathcal{H}$. Next, for all $A \in \mathcal{H} \subseteq \mathcal{G}$,

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\,d\mathbb{P} \overset{\text{(partial averaging)}}{=} \int_A \mathbb{E}[X|\mathcal{G}]\,d\mathbb{P} \overset{\text{(partial averaging)}}{=} \int_A X\,d\mathbb{P}.$$

Hence, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] \overset{\text{a.s.}}{=} \mathbb{E}[X|\mathcal{H}]$.

(g) Take $A = \Omega \in \mathcal{G}$. Then,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \int_A \mathbb{E}[X|\mathcal{G}]\,d\mathbb{P} \overset{\text{(partial averaging)}}{=} \int_A X\,d\mathbb{P} = \mathbb{E}[X].$$

(h) Omitted.

(i) Omitted.

□

2.3.5 **Martingales and Markov processes.** We close Section 2 by introducing some terminologies that relate *conditioning* and *stochastic processes*, which will be revisited in later sections.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, \mathbb{P})$ be a filtered probability space. A stochastic process $\{X_t\}_{t\in I}$ is said to be a **($\{\mathcal{F}_t\}$-)martingale** if:

(1) *(adaptivity)* $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$.

(2) *(integrability)* X_t is integrable for all $t \in I$.

(3) *(no tendency to rise or fall)* $X_s = \mathbb{E}[X_t|\mathcal{F}_s]$ for all $s \le t$.

- If we replace (3) by "$X_s \le \mathbb{E}[X_t|\mathcal{F}_s]$ for all $s \le t$" *(no tendency to fall)*, then $\{X_t\}$ is called a **($\{\mathcal{F}_t\}$-)submartingale**.
- If we replace (3) by "$X_s \ge \mathbb{E}[X_t|\mathcal{F}_s]$ for all $s \le t$" *(no tendency to rise)*, then $\{X_t\}$ is called a **($\{\mathcal{F}_t\}$-)supermartingale**.
- If we replace (3) by "For all $s \le t$ and all nonnegative measurable functions f such that $f(X_t)$ is integrable, there exists a measurable function g such that $\mathbb{E}[f(X_t)|\mathcal{F}_s] = g(X_s)$" ***(Markov property)***, then $\{X_t\}$ is called a **($\{\mathcal{F}_t\}$-)Markov process**.

  [Intuition 💡: This suggests that the conditional expectation $\mathbb{E}[f(X_t)|\mathcal{F}_s]$ about a "future" quantity $f(X_t)$ given the "current information" $\mathcal{F}_s$ can always be evaluated based on *only* the "current" process value (as a function of X_s) and not the past. This indeed generalizes the concept with the same name from STAT3903.]

  The independence lemma from [2.3.4]i is often helpful for showing that a stochastic process is a Markov process; see, e.g., Proposition 3.4.a.

2.3.6 **Doob martingale.** A notable example of martingale is called *Doob martingale*, which is the process of conditional expectations of a fixed random variable. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, \mathbb{P})$ be a filtered probability space and X be an integrable random variable. Then the **Doob martingale** is a stochastic process $\{X_t\}_{t\in I}$ defined by $X_t = \mathbb{E}[X|\mathcal{F}_t]$ for all $t \in I$. Let us verify that $\{X_t\}$ is indeed a $\{\mathcal{F}_t\}$-martingale:

Proof.

(1) By the definition of conditional expectation, $X_t = \mathbb{E}[X|\mathcal{F}_t]$ is $\mathcal{F}_t$-measurable for all $t \in I$. Thus $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$.

(2) For all $t \in I$, we have

$$\mathbb{E}[|X_t|] = \mathbb{E}\big[|\mathbb{E}[X|\mathcal{F}_t]|\big] \overset{\text{(conditional Jensen's inequality)}}{\leq} \mathbb{E}\big[\mathbb{E}\big[|X|\big|\mathcal{F}_t\big]\big] \overset{\text{(iterated conditioning)}}{=} \mathbb{E}[|X|] \overset{(X\text{ integrable})}{<} \infty,$$

so X_t is integrable.

(3) For all $s \leq t$, we have $\mathcal{F}_s \subseteq \mathcal{F}_t$, thus

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_t]|\mathcal{F}_s] \overset{\text{(tower property)}}{=} \mathbb{E}[X|\mathcal{F}_s] = X_s.$$

□

2.3.7 **Linear combinations of martingales is a martingale.** Due to the linearity of conditional expectations, linear combination of martingales remains as a martingale.

Proposition 2.3.a. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, \mathbb{P})$ be a filtered probability space. If $\{X_t\}$ and $\{Y_t\}$ are $\{\mathcal{F}_t\}$-martingales, then $\{\alpha X_t + \beta Y_t\}$ is also a $\{\mathcal{F}_t\}$-martingale for all $\alpha, \beta \in \mathbb{R}$.

Proof.

(1) Since $\{X_t\}$ and $\{Y_t\}$ are both adapted to $\{\mathcal{F}_t\}$, X_t and Y_t are $\mathcal{F}_t$-measurable for all $t \in I$. Hence, $\alpha X_t + \beta Y_t$ is $\mathcal{F}_t$-measurable for all $t \in I$, meaning that $\{\alpha X_t + \beta Y_t\}$ is adapted to $\{\mathcal{F}_t\}$.

(2) With X_t and Y_t being integrable for all $t \in I$, by triangle inequality we have for all $t \in I$,

$$\mathbb{E}[|\alpha X_t + \beta Y_t|] \leq |\alpha| \underbrace{\mathbb{E}[|X_t|]}_{<\infty} + |\beta| \underbrace{\mathbb{E}[|Y_t|]}_{<\infty} < \infty,$$

thus $\alpha X_t + \beta Y_t$ is integrable.

(3) By the linearity of conditional expectation, for all $s \leq t$ we have

$$\alpha X_s + \beta Y_s \overset{(\{X_t\},\ \{Y_t\}\text{ martingales})}{=} \alpha\mathbb{E}[X_t|\mathcal{F}_s] + \beta\mathbb{E}[Y_t|\mathcal{F}_s] = \mathbb{E}[\alpha X_t + \beta Y_t|\mathcal{F}_s].$$

□

3 Brownian Motions

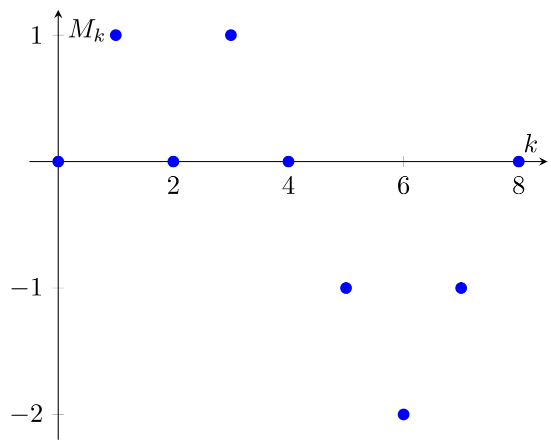
3.0.1 After learning the fundamentals of probabilistic concepts in Sections 1 and 2, we are ready to study an important kind of stochastic process, known as *Brownian motion*, which lies in the heart of many popular models for economic variables (e.g., stock prices) in financial economics. It turns out that Brownian motion can be obtained as a *limit* of a more elementary (and understandable ☺) stochastic process known as *random walk*. Thus, to motivate the development of Brownian motion and gain some intuition about it, we start with discussing random walks.

3.1 Random Walks

3.1.1 **Definition of random walk.** The random walk can be obtained from the experiment of tossing a fair coin repetitively and independently, once per unit time, with 1 and -1 being labels of heads and tails respectively:

$$X_j = \begin{cases} 1 & \text{if } j\text{th toss is heads (with probability } 1/2), \\ -1 & \text{if } j\text{th toss is tails (with probability } 1/2), \end{cases}$$

for all $j = 1, 2, \ldots$, with $X_1, X_2, \ldots$ being i.i.d. Then, we consider the random variable M_k that sums up the X_j's for all $j \le k$: $M_0 = 0$ and $M_k = \sum_{j=1}^{k} X_j$ for all $k = 1, 2, \ldots$. The stochastic process $\{M_k\}$ is then called a **symmetric random walk**.



3.1.2 **Properties of random walk.**

(a) *(independent increments)* For all $0 = k_0 < k_1 < \cdots < k_m$, the random variables (**increments**)

$$M_{k_1} = M_{k_1} - M_{k_0}, \quad M_{k_2} - M_{k_1}, \quad \ldots \quad , \quad M_{k_m} - M_{k_{m-1}}$$

are independent.

(b) *(mean and variance of increments)* For all $0 \le k_i < k_{i+1}$, we have $\mathbb{E}\big[M_{k_{i+1}} - M_{k_i}\big] = 0$ and $\mathrm{Var}\big(M_{k_{i+1}} - M_{k_i}\big) = k_{i+1} - k_i$.

(c) *(martingale)* Let $\mathcal{F}_k$ denote the family of all events whose occurrence can be decided after the first k tosses (you have seen this in [2.1.1]). Then $\{M_k\}$ is a $\{\mathcal{F}_k\}$-martingale.

(d) *(quadratic variation)* The **quadratic variation** up to time k of $\{M_k\}$ is defined by $[M, M]_k := \sum_{j=1}^{k}(M_j - M_{j-1})^2$. Then, we have $[M, M]_k = k$. [Note: Each summand involves a quadratic term that describes the variation in each time step, hence the name "quadratic variation".]

Proof.

(a) Since the increments are sums over *disjoint* sets of X_j's, the increments are independent by [2.2.5].

(b) Note that for all $0 \le k_i < k_{i+1}$ we have $M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$. Since $\mathbb{E}[X_j] = 1(1/2) + (-1)(1/2) = 0$ and $\mathrm{Var}(X_j) = 1^2(1/2)+(-1)^2(1/2) = 1$ for all $j = 1, 2, \ldots$, we have $\mathbb{E}\big[M_{k_{i+1}} - M_{k_i}\big] = 0$ and $\mathrm{Var}\big(M_{k_{i+1}} - M_{k_i}\big) = k_{i+1} - k_i$ for all $0 \le k_i < k_{i+1}$.

(c) (1) By construction of the filtration $\{\mathcal{F}_k\}$, $\{M_k\}$ is adapted to $\{\mathcal{F}_k\}$.

(2) For all $k = 0, 1, 2, \ldots$, M_k is clearly integrable as it is a finite sum of zero-mean random variables.

(3) Fix any $k \le \ell$. If $k = \ell$, then we have $\mathbb{E}[M_\ell|\mathcal{F}_k] = \mathbb{E}[M_k|\mathcal{F}_k] = M_k$ as M_k is a sum of $\mathcal{F}_k$-measurable random variables, thus is $\mathcal{F}_k$-measurable also. Now consider the case with $k < \ell$. Since $M_\ell - M_k$ ("future" increment) is independent of $\mathcal{F}_k$, we have

$$\mathbb{E}[M_\ell|\mathcal{F}_k] = \mathbb{E}[(M_\ell - M_k) + M_k|\mathcal{F}_k] = \mathbb{E}[M_\ell - M_k|\mathcal{F}_k] + \mathbb{E}[M_k|\mathcal{F}_k] = \underbrace{\mathbb{E}[M_\ell - M_k]}_{0} + M_k = M_k.$$

(d) Since each $M_j - M_{j-1}$ is either 1 or -1, we have $(M_j - M_{j-1})^2 = 1$ for all $j = 1, 2, \ldots$. Thus, $[M, M]_k = k$.

□

3.2 Brownian Motions

3.2.1 As we "speed up" the coin tossing process discussed previously, the corresponding random walk process turns out to converge to the *Brownian motion* (see Shreve (2004) for more details). As a result, Brownian motion does share some properties with the random walk.

3.2.2 **Definition of Brownian motion.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a stochastic process $\{W_t\}_{t\ge 0}$ is called a **(standard) Brownian motion** or **Wiener process** (on $(\Omega, \mathcal{F}, \mathbb{P})$) if:

(1) *(starting at 0)* $W_0 = 0$.

(2) *(continuity)* For every fixed $\omega \in \Omega$, the function $t \mapsto W_t(\omega)$ on $[0, \infty)$ is continuous.

(3) *(independent and normal increments)* For all $0 = t_0 < t_1 < \cdots < t_m$, the increments

$$W_{t_1} = W_{t_1} - W_{t_0}, \quad W_{t_2} - W_{t_1}, \quad \ldots \quad , \quad W_{t_m} - W_{t_{m-1}}$$

are independent, and $W_{t_{i+1}} - W_{t_i} \sim \mathrm{N}(0, t_{i+1} - t_i)$ for all $i = 0, 1, \ldots, m-1$.

3.2.3 **Properties of Brownian motion.**

(a) *(martingale)* Let $\{\mathcal{F}_t\}_{t\ge 0}$ be a filtration such that (i) *(adaptivity)* $\{W_t\}$ is adapted to $\{\mathcal{F}_t\}$ and (ii) *(independent of future increments)* $W_u - W_t$ is independent of $\mathcal{F}_t$ for all $0 \le t < u$. [Note: Such filtration is known as a **filtration for Brownian motion** $\{W_t\}$.]

Then, the Brownian motion $\{W_t\}$ is a $\{\mathcal{F}_t\}$-martingale.

(b) *(multivariate normality)* For all $0 \le t_1 < t_2 < \cdots < t_m$, $(W_{t_1}, \ldots, W_{t_m})$ follows a multivariate normal distribution with mean vector being the zero vector and covariance matrix being

$$\begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{bmatrix}$$

whose (i, j)th entry is $\min\{i, j\}$ for all $i, j = 1, \ldots, m$.

(c) *(joint moment generating function)* For all $0 \le t_1 < t_2 < \cdots < t_m$, the joint moment generating function of $(W_{t_1}, \ldots, W_{t_m})$ is

$$M(u_1, \ldots, u_m) = \exp\left[\frac{1}{2}\bigg((u_1 + \cdots + u_m)^2 t_1 + (u_2 + \cdots + u_m)^2 (t_2 - t_1) + \cdots \right.$$
$$\left. + (u_{m-1} + u_m)^2 (t_{m-1} - t_{m-2}) + u_m^2 (t_m - t_{m-1})\bigg)\right]$$

Proof.

(a) (1) By the definition of $\{\mathcal{F}_t\}$, $\{W_t\}$ is adapted to $\mathcal{F}_t$.

(2) By the normal increment property, we know $W_t \sim \mathrm{N}(0,t)$ for all $t \geq 0$. Thus W_t is integrable for all $t \geq 0$.

(3) Fix any $0 \leq s \leq t$. If $s = t$, then we have $\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_s|\mathcal{F}_s] = W_s$ as W_s is $\mathcal{F}_s$-measurable, due to the adaptivity. Now consider the case with $s < t$. Since $W_t - W_s$ is independent of $\mathcal{F}_s$ by assumption, we have

$$\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s|\mathcal{F}_s] + \mathbb{E}[W_s|\mathcal{F}_s] = \mathbb{E}[W_t - W_s] + W_s = 0 + W_s = W_s.$$

(b) By the independent and normal increments property, the vector of increments $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}})$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\Sigma = \mathrm{diag}(t_1, t_2 - t_1, \ldots, t_m - t_{m-1})$, i.e., the matrix

$$\Sigma = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 - t_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_m - t_{m-1} \end{bmatrix}.$$

Since we can write

$$\begin{bmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_m} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{A} \begin{bmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ \vdots \\ W_{t_m} - W_{t_{m-1}} \end{bmatrix},$$

$(W_{t_1}, \ldots, W_{t_m})$ follows a multivariate normal distribution with mean vector $A\mathbf{0} = \mathbf{0}$ and covariance matrix

$$A\Sigma A^T = \begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{bmatrix}.$$

(c) The joint moment generating function of $(W_{t_1}, \ldots, W_{t_m})$ is

$$\begin{aligned} M(u_1, \ldots, u_m) &= \mathbb{E}\left[e^{u_m W_{t_m} + u_{m-1} W_{t_{m-1}} + \cdots + u_1 W_{t_1}}\right] \\ &= \mathbb{E}\left[e^{u_m (W_{t_m} - W_{t_{m-1}}) + (u_{m-1} + u_m)(W_{t_{m-1}} - W_{t_{m-2}}) + \cdots + (u_1 + \cdots + u_m) W_{t_1}}\right] \\ &\overset{\text{(independent increments)}}{=} \mathbb{E}\left[e^{u_m (W_{t_m} - W_{t_{m-1}})}\right] \mathbb{E}\left[e^{(u_{m-1} + u_m)(W_{t_{m-1}} - W_{t_{m-2}})}\right] \cdots \mathbb{E}\left[e^{(u_1 + \cdots + u_m) W_{t_1}}\right] \\ &\overset{(W_{t_{i+1}} - W_{t_i} \sim \mathrm{N}(0, t_{i+1} - t_i))}{=} \exp\left[\frac{1}{2}(t_m - t_{m-1}) u_m^2\right] \exp\left[\frac{1}{2}(t_{m-1} - t_{m-2})(u_{m-1} + u_m)^2\right] \\ &\qquad \cdots \exp\left[\frac{1}{2} t_1 (u_1 + \cdots + u_m)^2\right] \\ &= \exp\left[\frac{1}{2}\Big((u_1 + \cdots + u_m)^2 t_1 + (u_2 + \cdots + u_m)^2 (t_2 - t_1) + \cdots \right. \\ &\qquad \left. + (u_{m-1} + u_m)^2 (t_{m-1} - t_{m-2}) + u_m^2 (t_m - t_{m-1})\Big)\right]. \end{aligned}$$

□

3.2.4 **Characterizations of Brownian motion.** As it turns out, the multivariate normality in [3.2.3]b and the joint moment generating function in [3.2.3]c can be used to *characterize* Brownian motion, or more precisely, serve as alternatives to the *independent and normal increments* property in the definition:

Proposition 3.2.a. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{W_t\}_{t\geq 0}$ be a stochastic process. Suppose that $W_0 = 0$ and for every fixed $\omega \in \Omega$, the function $t \mapsto W_t(\omega)$ on $[0,\infty)$ is continuous. Then the following are equivalent.

(a) $\{W_t\}$ satisfies the *independent and normal increments* property in the definition of Brownian motion.

(b) $\{W_t\}$ satisfies the *multivariate normality* as in [3.2.3]b.

(c) $\{W_t\}$ possesses the *joint moment generating functions* as in [3.2.3]b.

Proof. Omitted. □

3.3 First-Order and Quadratic Variations

3.3.1 Like random walk, we can also define *quadratic variation* for a *Brownian motion*, but the definition is more sophisticated and involves taking limit, unlike the more elementary one for random walk in [3.1.2]d. For a symmetric random walk $\{M_k\}$, its quadratic variation up to time k is $[M,M]_k = k$, and we may say that it "accumulates" k units of quadratic variation between times 0 and k. Brownian motion turns out to share this property also: it "accumulates" quadratic variation at rate one per unit time.

Here, apart from quadratic variation, we will also discuss another kind of variation known as the *first-order variation*, in contrast to the quadratic variation which may be seen as a *second-order* variation. It turns out that Brownian motion has an *infinite* first-order variation, explaining why one often focuses on quadratic variations in the study of Brownian motions.

3.3.2 **First-order variations.** Let f be a function on $[0,\infty)$ and $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a *partition* of $[0,T]$ (set of time points falling in $[0,T]$), where $0 = t_0 < t_1 < \cdots < t_n = T$. We denote the maximal time step in the partition (**norm** of Π) by $\|\Pi\| := \max_{j=0,\ldots,n-1}(t_{j+1} - t_j)$. Then the **first-order variation** of f up to time T is

$$\mathrm{FV}_T(f) := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|.$$

[Note: It can be shown that $\mathrm{FV}_T(f)$ is either a nonnegative constant or ∞.]

Here each summand only involves a "first-order" term that captures the variation. Also, "$\lim_{\|\Pi\|\to 0}$" refers to a limit taken with respect to a certain sequence of partitions Π whose norms converging to zero. In many cases, the choice of such sequence does not matter. However sometimes it would matter (e.g., in Proposition 3.3.b) and we shall implicitly assume the sequence is "suitably" chosen such that the target result holds. (We will not discuss the technical details here.)

[Note: Assuming f is differentiable, by mean value theorem we have $f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j)$ for some $t_j^* \in [t_j, t_{j+1}]$. Therefore in this case we can express the first-order variation as

$$\mathrm{FV}_T(f) = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|(t_{j+1} - t_j) \overset{\text{(definition of Riemann integral)}}{=} \int_0^T |f'(t)|\,\mathrm{d}t.$$

]

3.3.3 **Quadratic variations.** To obtain the quadratic variation, we replace each first-order term $|f(t_{j+1}) - f(t_j)|$ in the first-order variation by the quadratic term $(f(t_{j+1}) - f(t_j))^2$: The **quadratic variation** of f up to time T is

$$[f,f]_T := \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2.$$

[Note: It can be shown that $[f,f]_T$ is either a nonnegative constant or ∞.]

Intuitively, quadratic variation provides us a measure of "roughness" of the function. As the following result demonstrates, as long as the function f is "sufficiently smooth", then its quadratic variation is always zero.

Proposition 3.3.a. If $\mathrm{FV}_T(f) < \infty$ and f is continuous, then $[f,f]_T = 0$ for all $T > 0$.

[Note: The condition is particularly satisfied when f is *continuously differentiable*, i.e., its derivative f' is continuous. This is because in such case, f' is (Riemann-)integrable and thus $\mathrm{FV}_T(f) = \int_0^T |f'(t)|\,\mathrm{d}t < \infty$.
]

Proof. Note that

$$\sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \le \left(\max_{0\le j\le n-1} |f(t_{j+1}) - f(t_j)|\right) \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|.$$

Then, taking the limit $\lim_{\|\Pi\|\to 0}$ on both sides gives

$$\begin{aligned}
[f,f]_T &\le \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1} |f(t_{j+1}) - f(t_j)|\right) \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)| \\
&= \underbrace{\left(\lim_{\|\Pi\|\to 0}\max_{0\le j\le n-1} |f(t_{j+1}) - f(t_j)|\right)}_{=0 \text{ since } f \text{ is continuous}} \left(\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|\right) \\
&= 0 \cdot \mathrm{FV}_T(f) \\
&= 0.
\end{aligned}$$

On the other hand, we have $\sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \ge 0$ always since each summand is nonnegative, so $[f,f]_T \ge 0$. Thus we have $[f,f]_T = 0$. $\square$

3.3.4 **Quadratic variation of Brownian motion.** Now we are ready to show that Brownian motion does accumulate T units of quadratic variation between times 0 and T:

Proposition 3.3.b. Let $\{W_t\}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, and let W denote the function $t \mapsto W_t$. Then, we have $[W,W]_T \overset{\text{a.s.}}{=} T$ for all $T > 0$.[8]

Proof. Let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0,T]$ and write $Q_\Pi := \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})^2$. Then,

$$\mathbb{E}[Q_\Pi] = \sum_{j=0}^{n-1} \underbrace{\mathbb{E}\big[(W_{t_{j+1}} - W_{t_j})^2\big]}_{\mathrm{Var}(W_{t_{j+1}} - W_{t_j})} = \sum_{j=0}^{n-1} (t_{j+1} - t_j) = T.$$

Noting that

$$\begin{aligned}
\mathrm{Var}\big((W_{t_{j+1}} - W_{t_j})^2\big) &= \mathbb{E}\big[(W_{t_{j+1}} - W_{t_j})^4\big] - \big(\mathbb{E}\big[(W_{t_{j+1}} - W_{t_j})^2\big]\big)^2 \\
&\overset{(X\sim \mathrm{N}(0,\sigma^2) \Rightarrow \mathbb{E}[X^4] = 3(\sigma^2)^2)}{=} 3(t_{j+1} - t_j)^2 - (t_{j+1} - t_j)^2 \\
&= 2(t_{j+1} - t_j)^2,
\end{aligned}$$

we have

$$\mathrm{Var}(Q_\Pi) = \sum_{j=0}^{n-1} \mathrm{Var}\big((W_{t_{j+1}} - W_{t_j})^2\big) = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \le 2\|\Pi\| \sum_{j=0}^{n-1} (t_{j+1} - t_j) = 2\|\Pi\| T.$$

[8] To have the almost sure equality, the limit is supposed to be taken with respect to a sequence of partitions whose norms converge to zero "sufficiently fast". Here we shall not delve into these technical details and we will assume throughout that this is done, so that we have the almost sure equality always.

Thus, $\lim_{\|\Pi\|\to 0} \mathrm{Var}(Q_\Pi) = 0$. So, we have

$$\lim_{\|\Pi\|\to 0} \mathbb{E}\Big[(Q_\Pi - T)^2\Big] = \lim_{\|\Pi\|\to 0} \mathbb{E}\Big[(Q_\Pi - \mathbb{E}[Q_\Pi])^2\Big] = \lim_{\|\Pi\|\to 0} \mathrm{Var}(Q_\Pi) = 0.$$

With the sequence of partitions Π suitably chosen, we can then have $[W,W]_T = \lim_{\|\Pi\|\to 0} Q_\Pi \overset{\text{a.s.}}{=} \mathbb{E}[Q_\Pi] = T$ for all $T > 0$. □

3.3.5 **First-order variation of Brownian motion.** Using Proposition 3.3.b, we can show that the first-order variation of Brownian motion is indeed infinite.

Proposition 3.3.c. Let $\{W_t\}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, and consider the function $t \mapsto W_t(\omega)$ for each fixed $\omega \in \Omega$ (call it W). Then we have $\mathrm{FV}_T(W) \overset{\text{a.s.}}{=} \infty$ for all $T > 0$.

Proof. Consider

$$\sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})^2 = \sum_{j=0}^{n-1} |W_{t_{j+1}} - W_{t_j}| \cdot |W_{t_{j+1}} - W_{t_j}| \le \left(\max_{0\le k\le n-1} |W_{t_{k+1}} - W_{t_k}| \right) \sum_{j=0}^{n-1} |W_{t_{j+1}} - W_{t_j}|.$$

By Proposition 3.3.b, we have $\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} (W_{t_{j+1}}(\omega) - W_{t_j}(\omega))^2 = T > 0$ for all $\omega \in \Omega \setminus N$ where N is a null set. Furthermore, by continuity we have $\lim_{\|\Pi\|\to 0} \big(\max_{0\le k\le n-1} |W_{t_{k+1}}(\omega) - W_{t_k}(\omega)|\big) = 0$ for all $\omega \in \Omega$.

Now fix any $\omega \in \Omega \setminus N$. If we had $\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |W_{t_{j+1}}(\omega) - W_{t_j}(\omega)| = L$ for some $L < \infty$, then the inequality above would imply $T \le 0 \cdot L = 0$, contradiction. Thus, we must have $\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} |W_{t_{j+1}}(\omega) - W_{t_j}(\omega)| = \infty$, for all $\omega \in \Omega \setminus N$. □

3.3.6 **Differential rules for Brownian motions.** While the function $t \mapsto W_t(\omega)$ with $\omega \in \Omega$ fixed is generally not differentiable, there are still some *informal* differential rules available for the Brownian motion, which will be very helpful for handling *stochastic calculus* (see Section 4):

(1) $\boxed{\mathrm{d}W_t\, \mathrm{d}W_t = \mathrm{d}t}$.

Mathematical meaning: $\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \underbrace{(W_{t_{j+1}} - W_{t_j})}_{\text{“}\mathrm{d}W_t\text{”}} \underbrace{(W_{t_{j+1}} - W_{t_j})}_{\text{“}\mathrm{d}W_t\text{”}} = [W,W]_T = T = \int_0^T \mathrm{d}t$.

(2) $\boxed{\mathrm{d}W_t\, \mathrm{d}t = 0}$.

Mathematical meaning: $\lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \underbrace{(W_{t_{j+1}} - W_{t_j})}_{\text{“}\mathrm{d}W_t\text{”}} \underbrace{(t_{j+1} - t_j)}_{\text{“}\mathrm{d}t\text{”}} = 0 = \int_0^T 0 \,\mathrm{d}t$.

Proof. Consider

$$\begin{aligned}
\lim_{\|\Pi\|\to 0} \left| \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})(t_{j+1} - t_j) \right| &\le \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \big|(W_{t_{j+1}} - W_{t_j})(t_{j+1} - t_j)\big| \\
&\le \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} \max_{j=0,\dots,n-1} |W_{t_{j+1}} - W_{t_j}|(t_{j+1} - t_j) \\
&= \lim_{\|\Pi\|\to 0} \max_{j=0,\dots,n-1} |W_{t_{j+1}} - W_{t_j}| \sum_{j=0}^{n-1} (t_{j+1} - t_j) \\
&= T \underbrace{\lim_{\|\Pi\|\to 0} \max_{j=0,\dots,n-1} |W_{t_{j+1}} - W_{t_j}|}_{=0 \text{ due to continuity}} \\
&= 0.
\end{aligned}$$

□

(3) $\boxed{\mathrm{d}t\,\mathrm{d}t = 0}$.

Mathematical meaning: $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\underbrace{(t_{j+1}-t_j)}_{\text{“d}t\text{”}}\underbrace{(t_{j+1}-t_j)}_{\text{“d}t\text{”}} = 0 = \int_0^T 0\,\mathrm{d}t$.

Proof. Note that

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

□

[Mnemonic: Whenever the product of differentials involves *at least three* “$\mathrm{d}W_t$”s, the result would be zero. (Here we regard “$\mathrm{d}t$” as the product of *two* “$\mathrm{d}W_t$”s, so “$\mathrm{d}W_t\,\mathrm{d}t$” would involve three “$\mathrm{d}W_t$”s, for example.)]

3.4 Markov Property

3.4.1 **Brownian motion is a Markov process.** Recall the definition of *Markov process* in [2.3.5]. Here, we will show that the Brownian motion is indeed a Markov process, with respect to a filtration *for Brownian motion* (i.e., a filtration that satisfies also the *adaptivity* and *independence of future increments*).

Proposition 3.4.a. Let $\{W_t\}$ be a Brownian motion and $\{\mathcal{F}_t\}$ be a filtration for the Brownian motion $\{W_t\}$. Then $\{W_t\}$ is a $\{\mathcal{F}_t\}$-Markov process.

Proof. It suffices to verify the *Markov property* as the adaptivity and integrability of $\{W_t\}$ have already been shown previously (when showing that it is a $\{\mathcal{F}_t\}$-martingale).

Fix any $s \le t$ and any nonnegative measurable function f. First, we write

$$\mathbb{E}[f(W_t)|\mathcal{F}_s] = \mathbb{E}\big[f\big(W_s + (W_t - W_s)\big)|\mathcal{F}_s\big].$$

[Note: We can view f as a measurable function with inputs W_s and $(W_t - W_s)$.]

Noting that W_s is $\mathcal{F}_s$-measurable and $W_t - W_s$ is independent of $\mathcal{F}_s$, we can apply the independence lemma from [2.3.4]i with $X_1 := W_s$, $Y_1 := W_t - W_s$, and $\mathcal{G} := \mathcal{F}_s$: Defining $g(x) := \mathbb{E}[f(x + W_t - W_s)]$, we have $\mathbb{E}[f(W_t)|\mathcal{F}_s] = \mathbb{E}\big[f\big(W_s + (W_t - W_s)\big)|\mathcal{F}_s\big] = g(W_s)$. It then remains to show that g is measurable.

Since $W_t - W_s \sim \mathrm{N}(0, t-s)$, we have

$$g(x) = \mathbb{E}[f(x + W_t - W_s)] = \int_{-\infty}^{\infty} f(x+w)\cdot\frac{1}{\sqrt{2\pi(t-s)}}e^{-\frac{w^2}{2(t-s)}}\,\mathrm{d}w \tag{2}$$

which can be shown to be a continuous function of x. This would then imply the (Borel-)measurability of g, using the property that the preimage of an open set under a continuous function is also open (details omitted). □

3.4.2 **Transition density representation of Markov property.** After some changes of variables in Equation (2), we can interpret the Markov property for the Brownian motion more intuitively. We let $\tau = t - s$ and $y = w + x$ to get

$$g(x) = \int_{-\infty}^{\infty} f(y)\cdot\frac{1}{\sqrt{2\pi\tau}}e^{-\frac{(y-x)^2}{2\tau}}\,\mathrm{d}y.$$

Defining the **transition density** for the Brownian motion to be $p(\tau, x, y) := \frac{1}{\sqrt{2\pi\tau}}e^{-\frac{(y-x)^2}{2\tau}}$, we have

$$g(x) = \int_{-\infty}^{\infty} f(y)p(\tau, x, y)\,\mathrm{d}y,$$

thus the Markov property would suggest

$$\mathbb{E}[f(W_t)|\mathcal{F}_s] = g(W_s) = \int_{-\infty}^{\infty} f(y)p(\tau, W_s, y)\,\mathrm{d}y.$$

Interpretation: Based on the information from $\mathcal{F}_s$, the "conditional density" of W_t is $p(\tau, W_s, y)$ (normal density with mean W_s and variance $\tau = t - s$), in the variable y. This "conditional density" depends on the information from $\mathcal{F}_s$ only through W_s (the past information has no impact on $p(\tau, W_s, y)$), capturing the essential idea of Markov process that only the "current" process value is relevant for the evaluation of the conditional expectation.

3.5 First Passage Time Distribution

3.5.1 In STAT3910, we have learnt about a kind of exotic option known as the *barrier option*: The option would *come into existence* or *cease to exist* upon hitting/passing a specified barrier (price of the underlying asset). Therefore, we would like to study how the *first passage time* (first time of hitting the barrier) behaves, in order to analyze such kind of options.

3.5.2 **Exponential martingale.** To study the first passage time distribution, a key concept is the *exponential martingale* corresponding to a constant σ, which contains a Brownian motion in the exponential function.

Proposition 3.5.a. Let $\{W_t\}_{t\geq 0}$ be a Brownian motion, $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration for the Brownian motion $\{W_t\}$, and σ be a constant. Then the stochastic process $\{Z_t\}_{t\geq 0} := \left\{e^{\sigma W_t - \sigma^2 t/2}\right\}_{t\geq 0}$ is a $\{\mathcal{F}_t\}$-martingale. [Note: Such stochastic process $\{Z_t\}$ is called the **exponential martingale** corresponding to σ.]

Proof.

(1) Since $\{W_t\}$ is adapted to $\{\mathcal{F}_t\}$, $\{Z_t\} = \left\{e^{\sigma W_t - \sigma^2 t/2}\right\}$ is adapted to $\{\mathcal{F}_t\}$ also.

(2) Since $W_t \sim \mathrm{N}(0, t)$ for all $t \geq 0$, we have

$$\mathbb{E}[|Z_t|] = \mathbb{E}\left[e^{\sigma W_t - \sigma^2 t/2}\right] = e^{-\sigma^2 t/2}\mathbb{E}\left[e^{\sigma W_t}\right] = e^{-\sigma^2 t/2}e^{\sigma^2 t/2} = 1 < \infty$$

for all $t \geq 0$, thus Z_t is integrable for all $t \geq 0$.

(3) Fix any $0 \leq s \leq t$. If $s = t$, then we have $\mathbb{E}[Z_t|\mathcal{F}_s] = \mathbb{E}[Z_s|\mathcal{F}_s] = Z_s$ due to the adaptivity. So henceforth we suppose $s < t$. In this case we have

$$\begin{aligned}\mathbb{E}[Z_t|\mathcal{F}_s] &= \mathbb{E}\left[e^{\sigma W_t - \sigma^2 t/2}\middle|\mathcal{F}_s\right] = \mathbb{E}\left[e^{\sigma(W_t - W_s)}e^{\sigma W_s - \sigma^2 t/2}\middle|\mathcal{F}_s\right] \\ &\overset{\text{(TOWIK)}}{=} e^{\sigma W_s - \sigma^2 t/2}\mathbb{E}\left[e^{\sigma(W_t - W_s)}\middle|\mathcal{F}_s\right] \overset{\text{(independent of future increments)}}{=} e^{\sigma W_s - \sigma^2 t/2}\mathbb{E}\left[e^{\sigma(W_t - W_s)}\right] \\ &\overset{(W_t - W_s \sim \mathrm{N}(0, t-s))}{=} e^{\sigma W_s - \sigma^2 t/2}e^{\frac{1}{2}(t-s)\sigma^2} = e^{\sigma W_s - \sigma^2 s/2} = Z_s.\end{aligned}$$

□

3.5.3 **First passage time, stopping time, and maximum/minimum of Brownian motions.** Let m be a real number and $\{W_t\}$ be a Brownian motion. Then the **first passage time of Brownian motion to level m** is $\tau_m := \inf\{t \geq 0 : W_t = m\}$. Particularly, if $W_t \neq m$ for all $t \geq 0$, then the first passage time would be $\tau_m = \inf \varnothing = \infty$.

Remarks:

- As long as $\{t \geq 0 : W_t = m\}$ is nonempty, we have $\inf\{t \geq 0 : W_t = m\} = \min\{t \geq 0 : W_t = m\}$ due to the continuity of $t \mapsto W_t$.

- The first passage time τ_m is a random variable.
- Often we consider the case with $m \neq 0$, since we always have $\tau_0 = 0$ due to the property that $W_0 = 0$.

A random variable τ is called a **stopping time** if $\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for all $t > 0$. [Intuition 💡: Based on the information about the process at time t ($\mathcal{F}_t$), we should be able to determine whether the process has been "stopped", i.e., whether the stopping time τ is less than or equal to t.]

The first passage time is indeed a kind of stopping time. To see this, consider:

- *($m > 0$)* $\{\omega \in \Omega : \tau_m(\omega) > t\} = \{\omega \in \Omega : W_s < m \ \forall s \leq t\} \overset{(\text{adaptivity of } \{W_t\})}{\in} \mathcal{F}_t$, so $\{\omega \in \Omega : \tau(\omega) \leq t\} = \{\omega \in \Omega : \tau_m(\omega) > t\}^c \in \mathcal{F}_t$ for all $t > 0$.
- *($m < 0$)* $\{\omega \in \Omega : \tau_m(\omega) > t\} = \{\omega \in \Omega : W_s > m \ \forall s \leq t\} \overset{(\text{adaptivity of } \{W_t\})}{\in} \mathcal{F}_t$, so $\{\omega \in \Omega : \tau(\omega) \leq t\} = \{\omega \in \Omega : \tau_m(\omega) > t\}^c \in \mathcal{F}_t$ for all $t > 0$.
- *($m = 0$)* $\{\omega \in \Omega : \tau_0(\omega) \leq t\} = \Omega \in \mathcal{F}_t$ for all $t > 0$.

The argument here also leads to an useful relationship between the first passage time and the maximum/minimum of Brownian motions, namely

$$\boxed{\tau_m \leq t \iff \begin{cases} M_t \geq m & \text{if } m > 0, \\ m_t \leq m & \text{if } m < 0, \end{cases}}$$

where $M_t = \max_{0 \leq s \leq t} W_s$ and $m_t = \min_{0 \leq s \leq t} W_s$, for all $t > 0$. This allows us to deduce the distribution of maximum/minimum of Brownian motions through the one for the first passage time (which will be investigated in Propositions 3.5.c and 3.6.b).

3.5.4 **Optional stopping theorem.** A remarkable result that is related to both the concepts of *martingale* and *stopping time* is the *optional stopping theorem*, which suggests that a martingale that is "stopped" ⏸ or "frozen" ❄ at a stopping time (based on your "option") remains a martingale.

Theorem 3.5.b (Optional stopping theorem)**.** Let $\{X_t\}_{t \in I}$ be a $\{\mathcal{F}_t\}$-martingale and τ be a stopping time. Then, the *stopped process* $\{X_{t \wedge \tau}\}_{t \in I}$[9] is also a $\{\mathcal{F}_t\}$-martingale. Particularly, if X_0 is a deterministic constant, then $\mathbb{E}[X_{t \wedge \tau}] = X_0$ for all $t \in I$.

Proof. We will omit the part of proving that the stopped process is also a martingale. Once we have this result, we know that for all $t \in I$, $\mathbb{E}[X_{t \wedge \tau}] = \mathbb{E}[\mathbb{E}[X_{t \wedge \tau} | \mathcal{F}_0]] = \mathbb{E}[X_{0 \wedge \tau}] = \mathbb{E}[X_0] = X_0$. □

For a more intuitive understanding of this result, let us interpret $\{X_t\}$ as a stock price process and the stopping time τ as the time of selling the stock. Suppose that you purchase a share of stock at the (nonrandom) price X_0 now (time 0). Then, the optional stopping theorem tells you that, regardless of the "rule" for deciding the selling time based on the *available* information (corresponding to the stopping time τ), the expected selling price is always the same as your purchase price X_0, meaning that you cannot "beat the market" in this way.

3.5.5 **First passage time distribution.** Now we have enough tools to derive the first passage time distribution. Here, the distribution is characterized indirectly (but uniquely) via the *Laplace transform*.

Proposition 3.5.c. Let m be a real number and $\{W_t\}$ be a Brownian motion. For the first passage time τ_m to level m, we have:

(a) *(Almost sure finiteness)* $\mathbb{P}(\tau_m < \infty) = 1$.

[9] The notation $t \wedge \tau$ denotes $\min\{t, \tau\}$, and we have

$$X_{t \wedge \tau} = \begin{cases} X_t & \text{if } t < \tau, \\ X_\tau & \text{if } t \geq \tau. \end{cases}.$$

(b) Its Laplace transform is given by $\mathbb{E}[e^{-\alpha\tau_m}] = e^{-|m|\sqrt{2\alpha}}$ for every $\alpha > 0$.

(c) *(Infinite mean)* $\mathbb{E}[\tau_m] = \infty$ if $m \neq 0$.

Proof.

(a) Let $\sigma > 0$ and $m > 0$. Applying Theorem 3.5.b to the *exponential martingale* $\{Z_t\} = \left\{e^{\sigma W_t - \sigma^2 t/2}\right\}$ with the stopping time being the first passage time τ_m gives

$$1 = Z_0 = \mathbb{E}[Z_{t\wedge\tau_m}] = \mathbb{E}\left[e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2}\right]$$

Noting that the Brownian motion $\{W_t\}$ is always at or below the level m for every $t \leq \tau_m$, we have $0 \leq e^{\sigma W_{t\wedge\tau_m}} \leq e^{\sigma m}$. Now, we consider the limiting behaviour of the term $e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2}$ as $t \to \infty$.

- *Case 1: $\tau_m < \infty$.* Since we have $e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2} = e^{\sigma W_{\tau_m} - \sigma^2 \tau_m/2} = e^{\sigma m - \sigma^2\tau_m/2}$ for sufficiently large t, the term would converge to $e^{\sigma W_{\tau_m} - \sigma^2\tau_m/2}$ as $t \to \infty$.
- *Case 2: $\tau_m = \infty$.* In this case, we know that

$$0 \leq e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2} = e^{\sigma W_t - \sigma^2 t/2} \leq e^{\sigma m} e^{-\sigma^2 t/2},$$

  and hence the term would converge to 0 as $t \to \infty$.

In short, we can summarize these two cases as:

$$\lim_{t\to\infty} e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2} = e^{\sigma m - \sigma^2\tau_m/2}\mathbf{1}_{\{\tau_m < \infty\}}.$$

Using DCT, we can get

$$1 = \lim_{t\to\infty} \mathbb{E}\left[e^{\sigma W_{t\wedge\tau_m} - \sigma^2 (t\wedge\tau_m)/2}\right] = \mathbb{E}\left[e^{\sigma m - \sigma^2\tau_m/2}\mathbf{1}_{\{\tau_m < \infty\}}\right],$$

which implies that $\mathbb{E}\left[e^{-\sigma^2\tau_m/2}\mathbf{1}_{\{\tau_m < \infty\}}\right] = e^{-\sigma m}$. Taking the limit $\sigma \to 0^+$ on both sides then yields $\mathbb{E}\left[\mathbf{1}_{\{\tau_m < \infty\}}\right] = \mathbb{P}(\tau_m < \infty) = 1$, by the MCT.

(b) First of all, in case $m = 0$, we would have $\tau_m = 0$ and both sides of the equality would be 1, so the result holds immediately.
Then, with $\mathbb{P}(\tau_m < \infty) = 1$, we know that

$$\mathbb{E}\left[e^{-\sigma^2\tau_m/2}\right] = \mathbb{E}\left[e^{-\sigma^2\tau_m/2}\mathbf{1}_{\{\tau_m < \infty\}}\right]\mathbb{P}(\tau_m < \infty) + 0 = e^{-\sigma m}.$$

Now, consider first the case where $m > 0$. Fix any $\alpha > 0$, and then set $\sigma = \sqrt{2\alpha}$. After that, we have $\sigma^2/2 = \alpha$, and thus the equation above implies that $\mathbb{E}[e^{-\alpha\tau_m}] = e^{-m\sqrt{2\alpha}} = e^{-|m|\sqrt{2\alpha}}$. Next, for the case where $m < 0$, it can be shown that τ_m and $\tau_{|m|} = \tau_{-m}$ have the same distribution by the symmetry of the Brownian motion. Therefore, we have $\mathbb{E}[e^{-\alpha\tau_m}] = \mathbb{E}[e^{-\alpha\tau_{-m}}] \overset{(-m>0)}{=} e^{-|-m|\sqrt{2\alpha}} = e^{-|m|\sqrt{2\alpha}}$ in this case also.

(c) Differentiating the Laplace transform from (b) with respect to α gives $\mathbb{E}[\tau_m e^{-\alpha\tau_m}] = \frac{|m|}{\sqrt{2\alpha}}e^{-|m|\sqrt{2\alpha}}$ for every $\alpha > 0$. Letting $\alpha \to 0^+$ then yields $\mathbb{E}[\tau_m] = \infty$, provided that $m \neq 0$.

$\square$

[Note: This result suggests that, while the first passage time is *finite almost surely*, its mean is still *infinite* (quite counterintuitive)!]

3.6 Reflection Principle

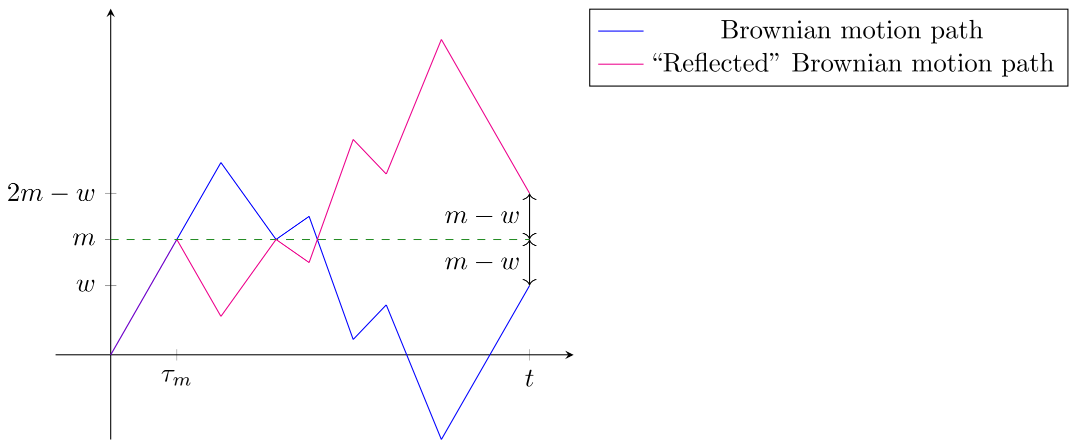
3.6.1 **Reflection principle.** To derive further distributions about Brownian motion, the *reflection principle* serves as an useful tool. The idea of reflection principle is as follows. First fix $m > 0$ and $t > 0$. Our task here is to count the number of Brownian motion paths that reach level m at or before time t, or in other words, those that lead to $\tau_m \leq t$. Those paths can be classified into two types:

(1) reaching level m before time t but locating at a level $w < m$ at time t,

(2) exceeding level m at time t.

[Note: Technically, there can be paths that are located *exactly* at level m at time t, but it is with probability zero to have such Brownian path realized. So, henceforth we shall ignore these without loss of generality.]

The reflection principle is best understood through a picture:



For every Brownian motion path of type 1 (i.e., reaching level m before time t but locating at a level $w < m$ at time t), there would be a corresponding "reflected" path (due to the symmetry of Brownian motion) that is at the level $2m - w$ at time t, which coincides with the original path at or before time τ_m, and is a mirror image of the original path about the line $y = m$ after time τ_m.

Applying this argument repetitively to numerous paths, the "number" of paths that reach level m before time t and locate at a level $\leq w$ at time t would "equal" the "number" of paths that are at or above the level $2m - w$ at time t. Therefore, intuitively, it is of equal probability to have a path realized from any one of these types. This leads to the *reflection equality*.

Theorem 3.6.a (Reflection equality)**.** Let $m > 0$, $\{W_t\}$ be a Brownian motion, and τ_m be the first passage time to level m. Then, for every $w \leq m$, we have

$$\mathbb{P}(\tau_m \leq t, W_t \leq w) = \mathbb{P}(W_t \geq 2m - w).$$

Proof. Omitted. □

3.6.2 **Distribution and density functions of first passage time.** Using the reflection equality, we can deduce the *distribution function* and *density function* of the first passage time τ_m (with $m \neq 0$), serving as tools for characterizing its distribution uniquely.

Proposition 3.6.b. Let $m \neq 0$ and τ_m be the first passage time. Then:

(a) The distribution function of τ_m is

$$\mathbb{P}(\tau_m \leq t) = \frac{2}{\sqrt{2\pi}} \int_{|m|/\sqrt{t}}^{\infty} e^{-y^2/2}\, dy, \quad t \geq 0.$$

(b) The density function of τ_m is

$$f_{\tau_m}(t) = \frac{|m|}{t\sqrt{2\pi t}} e^{-m^2/(2t)}, \quad t > 0.$$

Proof.

(a) Consider first the case where $m > 0$. With $w = m$ in Theorem 3.6.a, we get $\mathbb{P}(\tau_m \le t, W_t < m) = \mathbb{P}(\tau_m \le t, W_t \le m) = \mathbb{P}(W_t \ge m)$. Also, noting that $W_t \ge m \implies \tau_m \le t$, we have $\mathbb{P}(\tau_m \le t, W_t \ge m) = \mathbb{P}(W_t \ge m)$. Adding the two equations together then yields

$$\begin{aligned}\mathbb{P}(\tau_m \le t) = 2\mathbb{P}(W_t \ge m) &= \frac{2}{\sqrt{2\pi t}} \int_m^\infty \underbrace{e^{-x^2/(2t)}}_{\text{pdf of N}(0,t)} \, \mathrm{d}t \\ &\overset{(\text{substitution: } y = x/\sqrt{t})}{=} \frac{2}{\sqrt{2\pi}} \int_{m/\sqrt{t}}^\infty e^{-y^2/2} \, \mathrm{d}y = \frac{2}{\sqrt{2\pi}} \int_{|m|/\sqrt{t}}^\infty e^{-y^2/2} \, \mathrm{d}y.\end{aligned}$$

Next, for the case where $m < 0$, since τ_m and $\tau_{|m|} = \tau_{-m}$ have the same distribution, the distribution function of τ_m is

$$\mathbb{P}(\tau_m \le t) = \mathbb{P}(\tau_{-m} \le t) \overset{(-m>0)}{=} \frac{2}{\sqrt{2\pi}} \int_{|-m|/\sqrt{t}}^\infty e^{-y^2/2} \, \mathrm{d}y.$$

(b) Differentiating the distribution function from (a) with respect to t gives

$$\begin{aligned}f_{\tau_m}(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{2}{\sqrt{2\pi}} \int_{|m|/\sqrt{t}}^\infty e^{-y^2/2} \, \mathrm{d}y \underset{(\text{chain rule})}{\overset{(\text{Fundamental theorem of calculus})}{=}} -\frac{2}{\sqrt{2\pi}} e^{-(|m|/\sqrt{t})^2/2} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \frac{|m|}{\sqrt{t}} \\ &= \frac{|m|}{t\sqrt{2\pi t}} e^{-m^2/(2t)}, \quad t > 0.\end{aligned}$$

□

3.6.3 **Distributions of Brownian motion and its maximum.** Let $\{W_t\}$ be a Brownian motion. For every $t \ge 0$, the **maximum to date** (time t) for the Brownian motion is defined by $M_t = \max_{0 \le s \le t} W_s$. Such random variable is helpful in analyzing barrier options, and hence we would like to study its distributional behaviours. The following result suggests the joint distribution and conditional distribution of Brownian motion W_t and the maximum to date M_t:

Proposition 3.6.c. Let $\{W_t\}$ be a Brownian motion and M_t be its maximum to date (time t) for every $t \ge 0$. Then, for every $t > 0$:

(a) The joint density function of (M_t, W_t) is

$$f_{M_t,W_t}(m, w) = \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-(2m-w)^2/(2t)}, \quad m > 0, \; w < m.$$

(b) The conditional density function of M_t given $W_t = w$ is

$$f_{M_t|W_t}(m|w) = \frac{2(2m - w)}{t} e^{2m(m-w)/t}, \quad m > 0, \; w < m.$$

Proof.

(a) Fix any $m > 0$ and $w \le m$. By [3.5.3], we have $M_t \ge m$ iff $\tau_m \le t$. Hence,

$$\mathbb{P}(M_t \ge m, W_t \le w) = \mathbb{P}(\tau_m \le t, W_t \le w) \overset{(\text{Theorem 3.6.a})}{=} \mathbb{P}(W_t \ge 2m - w) = \frac{1}{\sqrt{2\pi t}} \int_{2m-w}^\infty e^{-z^2/(2t)} \, \mathrm{d}z.$$

Also, we have

$$\mathbb{P}(M_t \geq m, W_t \leq w) = \int_m^\infty \int_{-\infty}^w f_{M_t,W_t}(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

These imply that

$$\int_m^\infty \int_{-\infty}^w f_{M_t,W_t}(x,y)\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{\sqrt{2\pi t}} \int_{2m-w}^\infty e^{-z^2/(2t)}\,\mathrm{d}z.$$

Differentiating both sides with respect to m gives

$$-\int_{-\infty}^w f_{M_t,W_t}(m,y)\,\mathrm{d}y = -\frac{2}{\sqrt{2\pi t}} e^{-(2m-w)^2/(2t)}.$$

Next, differentiating both sides with respect to w gives

$$f_{M_t,W_t}(m,w) = \frac{2(2m-w)}{t\sqrt{2\pi t}} e^{-(2m-w)^2/(2t)}, \quad m > 0,\ w < m.$$

(b) By definition, we have

$$\begin{aligned} f_{M_t|W_t}(m|w) = \frac{f_{M_t,W_t}(m,w)}{f_{W_t}(w)} &= \frac{2(2m-w)e^{-(2m-w)^2/(2t)}/(t\sqrt{2\pi t})}{\underbrace{e^{w^2/2t}/\sqrt{2\pi t}}_{\text{pdf of N}(0,t)}} \\ &= \frac{2(2m-w)}{t} e^{-2m(m-w)/t}, \quad m > 0,\ w < m. \end{aligned}$$

□

4 Stochastic Calculus

4.0.1 Equipped with the knowledge from Sections 1 to 3, we are now ready to study the first major topic of this course, namely *stochastic calculus*. Here, we will be developing a "new kind of calculus", with "rules" that you should have never seen before in the previous calculus classes. In modern financial economics, stochastic calculus plays a fundamental role and it is involved in many results. The major notion within stochastic calculus is the *Itô's integral* (another type of new integral, apart from the Lebesgue integral in Section 1), so let us start by constructing such integral and studying its properties.

4.1 Construction of Itô integral

4.1.1 **Idea of Itô integral.** Like the construction of Lebesgue integral (not covered here; see STAT7610 if interested), the construction of Itô integral starts with defining the integral for *simple* integrands only, and then extending the definition for *general integrands* through taking limits on some "approximating sequences of simple integrands" that would get "closer and closer" to the desired general integrands.

Before going into the construction, it is instructive to have some intuitive idea about the Itô integral in mind. Our main goal here is to develop a more definition of integral looking like $\int_0^T \Delta_t \,\mathrm{d}W_t$ with a fixed $T > 0$. Here, we have a Brownian motion $\{W_t\}$ and we suppose that $\{\Delta_t\}$ is a stochastic process adapted to a filtration $\{\mathcal{F}_t\}$ for the Brownian motion. To have a better intuition, we may interpret Δ_t as the position taken for an asset at time t, and interpret W_t as the time-t "price" of that asset (not so properly as it can be negative). Then, loosely speaking, the integral $\int_0^T \Delta_t \,\mathrm{d}W_t$ is the "sum" of profits (positive/negative) earned over the period $[0, T]$ based on the "continuous evolution" of our positions taken and "continuous trading" in such period: For small time increment $h > 0$, the profit gained over $[t, t+h]$ would be (approximately) $\Delta_t \times \underbrace{(W_{t+h} - W_t)}_{\text{loosely: } \mathrm{d}W_t}$. We are going to formalize this idea in the following.

4.1.2 **Itô integral for simple integrands.** Let $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0, T]$, with $0 = t_0 \le t_1 \le \cdots \le t_n = T$. A process $\{\Delta_t\}$ is called a **simple process** if Δ_t is constant in t on each subinterval $[t_j, t_{j+1})$. It is not hard to see that every path of a simple process (with ω fixed) is a simple function in t, so in short, a simple process is a process with simple paths.

Now, let $\{W_t\}$ be a Brownian motion and $\{\mathcal{F}_t\}$ be a filtration for the Brownian motion. Here, we shall consider a simple process $\{\Delta_t\}$ that is adapted to $\{\mathcal{F}_t\}$. For every $k = 0, 1, \ldots, n-1$ and every $t \in [t_k, t_{k+1}]$, define

$$I(t) := \sum_{j=0}^{k-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_k}(W_t - W_{t_k}),$$

which gives the total profit from taking the positions as instructed in the simple process (until time t) and selling all the positions at time t. [Note: For example, take $t_1 = 1$, $t_2 = 3$, and $t_3 = 6$, and suppose that $\Delta_0 = 5$, $\Delta_1 = 7$, $\Delta_3 = 2$, $W_1 = 5$, $W_3 = 1$, and $W_4 = 7$. Then, $I(4) = \Delta_0(W_1 - W_0) + \Delta_1(W_3 - W_1) + \Delta_3(W_4 - W_3) = 5(5-0) + 7(1-5) + 2(7-1) = 9$.]

It is precisely the **Itô integral of the simple process $\{\Delta_t\}$**, and the notation for the Itô integral is:

$$I_t = \int_0^t \Delta_u \,\mathrm{d}W_u.$$

Differential form. In stochastic calculus, it is often handy to express formulas involving Itô integral in the *differential form* (as a shorthand), which involves some "differentials". For example, we would write $\mathrm{d}I_t = \Delta_t \,\mathrm{d}W_t$ (or $\mathrm{d}I_u = \Delta_u \,\mathrm{d}W_u$) as a shorthand for $\int_0^t \mathrm{d}I_u \overset{\text{(notation)}}{=} I_t - I_0 = \int_0^t \Delta_u \,\mathrm{d}W_u$. So, here the definition of Itô integral can be expressed in differential form as $\mathrm{d}I_t = \Delta_t \,\mathrm{d}W_t$ with $I_0 = 0$. This kind of differential equation that has "differential of stochastic process" is known as **stochastic differential equation** (SDE).

While the SDE $\mathrm{d}I_t = \Delta_t \,\mathrm{d}W_t$ may be intuitively interpreted as after moving "a little" forward in time from time t, the change in the Itô integral I would be approximately Δ_t times the change in the

Brownian motion W, such interpretation is imprecise since the terms "a little" and "approximately" do not have clear meanings. Mathematically, the SDE should always be understood to carry the (well-defined) meaning from the integral form (the corresponding formula with Itô integrals). [Note: To "derive" such intuitive interpretation, we may take a small $h > 0$, and then we would have $I_{t+h} - I_t = \int_0^{t+h} \Delta_u \, \mathrm{d}W_u - \int_0^t \Delta_u \, \mathrm{d}W_u \overset{\text{(consider definition)}}{\approx} \Delta_t(W_{t+h} - W_t)$.]

4.1.3 **Properties of Itô integral for simple integrands.** The Itô integral for simple integrands carries fairly nice properties, and they are helpful for establishing analogous properties for the later Itô integral for *general* integrands.

Let $\{\Delta_t\}$ be a simple process adapted to a filtration $\{\mathcal{F}_t\}$ for a Brownian motion $\{W_t\}$. Let $I_t = \int_0^t \Delta_u \, \mathrm{d}W_u$ denote the Itô integral.

(a) *(Martingale)* The process $\{I_t\}$ is a $\{\mathcal{F}_t\}$-martingale. Particularly, we have $\mathbb{E}[I_t] = 0$ for all t.

(b) *(Itô isometry)* $\mathbb{E}\big[I_t^2\big] = \mathbb{E}\Big[\int_0^t \Delta_u^2 \, \mathrm{d}u\Big]$.

(c) *(Quadratic variation)* The quadratic variation of I (as a function of time) up to time t is $[I, I]_t = \int_0^t \Delta_u^2 \, \mathrm{d}u$.

[Note: Recall the definition of quadratic variation: $[I, I]_t = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} (I_{t_{j+1}} - I_{t_j})^2$. Here, the quadratic variation is still a random variable. With $\omega \in \Omega$ realized, the value of quadratic variation can be computed by using the Itô integrals evaluated at ω in the expression: $[I, I]_t(\omega) = \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} (I_{t_{j+1}}(\omega) - I_{t_j}(\omega))^2$.]

Proof.

(a) We first show that $\{I_t\}$ is a $\{\mathcal{F}_t\}$-martingale:

(1) For every $t \in [t_k, t_{k+1}]$, we have $I(t) = \sum_{j=0}^{k-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_k}(W_t - W_{t_k})$. With $t_0 \le t_1 \le \cdots \le t_k \le t$, all Δ's and W's here are $\mathcal{F}_t$-measurable. Hence, $\{I_t\}$ is adapted to $\{\mathcal{F}_t\}$.

(2) Since the simple process $\{\Delta_t\}$ can only take finitely many different values, we know that there is a constant $M > 0$ such that $|\Delta_t| \le M$ for all t. Hence, for every $t \in [t_k, t_{k+1}]$, we have

$$\begin{aligned}
\mathbb{E}[|I_t|] &\overset{\text{(triangle)}}{\le} \sum_{j=0}^{k-1} \mathbb{E}\big[|\Delta_{t_j}| \cdot |W_{t_{j+1}} - W_{t_j}|\big] + \mathbb{E}[|\Delta_{t_k}| \cdot |W_t - W_{t_k}|] \\
&\le \sum_{j=0}^{k-1} M\mathbb{E}\big[|W_{t_{j+1}} - W_{t_j}|\big] + M\mathbb{E}[|W_t - W_{t_k}|] \\
&\overset{(|x|\le 1+x^2)}{\le} \sum_{j=0}^{k-1} M\big(1 + \mathbb{E}\big[(W_{t_{j+1}} - W_{t_j})^2\big]\big) + M\big(1 + \mathbb{E}\big[(W_t - W_{t_k})^2\big]\big) < \infty.
\end{aligned}$$

(3) Fix any $0 \le s \le t \le T$. We know that $s \in [t_\ell, t_\ell + 1)$ and $t \in [t_k, t_k + 1)$ for some $\ell \le k$ with $t_\ell \le t_k$. We then write

$$I_t = \sum_{j=0}^{\ell-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_\ell}(W_{t_{\ell+1}} - W_{t_\ell}) + \sum_{j=\ell+1}^{k-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_k}(W_t - W_{t_k}).$$

Consider first the case where $t_\ell = t_k$. We can then simplify the expression above to $I_t =$

$\sum_{j=0}^{\ell-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_\ell}(W_t - W_{t_\ell})$, and hence

$$\begin{aligned}
\mathbb{E}[I_t|\mathcal{F}_s] &= \sum_{j=0}^{\ell-1} \mathbb{E}\big[\Delta_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_s\big] + \mathbb{E}[\Delta_{t_\ell}(W_t - W_{t_\ell})|\mathcal{F}_s] \\
&\overset{\text{(TOWIK)}}{=} \sum_{j=0}^{\ell-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_\ell}(\mathbb{E}[W_t|\mathcal{F}_s] - W_{t_\ell}) \\
&\overset{(\{W_t\}\text{ is martingale})}{=} \sum_{j=0}^{\ell-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_\ell}(W_s - W_{t_\ell}) = I_s.
\end{aligned}$$

Next, consider the case where $t_\ell < t_k$. From above, we know that

$$\mathbb{E}\left[\sum_{j=0}^{\ell-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j}) + \Delta_{t_\ell}(W_{t_{\ell+1}} - W_{t_\ell})\middle|\mathcal{F}_s\right] = I_s.$$

Now, consider the rest of the terms in I_t:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=\ell+1}^{k-1} \Delta_{t_j}(W_{t_{j+1}} - W_{t_j})\middle|\mathcal{F}_s\right] &= \sum_{j=\ell+1}^{k-1} \mathbb{E}\big[\Delta_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_s\big] \\
&\overset{\text{(tower)}}{=} \sum_{j=\ell+1}^{k-1} \mathbb{E}\big[\mathbb{E}\big[\Delta_{t_j}(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}\big]|\mathcal{F}_s\big] \\
&\overset{\text{(TOWIK)}}{=} \sum_{j=\ell+1}^{k-1} \mathbb{E}\big[\Delta_{t_j}\mathbb{E}\big[(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}\big]|\mathcal{F}_s\big] \\
&\overset{\text{(independence)}}{=} \sum_{j=\ell+1}^{k-1} \mathbb{E}\big[\Delta_{t_j}\mathbb{E}\big[W_{t_{j+1}} - W_{t_j}\big]|\mathcal{F}_s\big] = \sum_{j=\ell+1}^{k-1} \mathbb{E}\big[\Delta_{t_j}\cdot 0|\mathcal{F}_s\big] = 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\Delta_{t_k}(W_t - W_{t_k})|\mathcal{F}_s] &\overset{\text{(tower)}}{=} \mathbb{E}[\mathbb{E}[\Delta_{t_k}(W_t - W_{t_k})|\mathcal{F}_{t_k}]|\mathcal{F}_s] \overset{\text{(TOWIK)}}{=} \mathbb{E}[\Delta_{t_k}\mathbb{E}[W_t - W_{t_k}|\mathcal{F}_{t_k}]|\mathcal{F}_s] \\
&\overset{\text{(independence)}}{=} \mathbb{E}[\Delta_{t_k}\mathbb{E}[W_t - W_{t_k}]|\mathcal{F}_s] = 0.
\end{aligned}$$

Therefore, we have $\mathbb{E}[I_t|\mathcal{F}_s] = I_s$ in this case also.

Knowing that $\{I_t\}$ is a $\{\mathcal{F}_t\}$-martingale, we have

$$\mathbb{E}[I_t] = \mathbb{E}[\mathbb{E}[I_t|\mathcal{F}_0]] \overset{\text{(martingale)}}{=} \mathbb{E}[I_0] \overset{\text{(definition)}}{=} \mathbb{E}[0] = 0$$

for every t.

(b) Let $D_j := W_{t_{j+1}} - W_{t_j}$ for every $j = 0, 1, \ldots, k-1$ and $D_k := W_t - W_{t_k}$. With these notations, we can simplify the Itô integral to $I_t = \sum_{j=0}^{k} \Delta_{t_j} D_j$. Then, direct expansion gives $I_t^2 = \sum_{j=0}^{k} \Delta_{t_j}^2 D_j^2 + 2\sum_{0\le i<j\le k} \Delta_{t_i}\Delta_{t_j} D_i D_j$. We consider these two terms one by one.
First, for every $j = 0, 1, \ldots, k$, we know that $\Delta_{t_j}^2$ is $\mathcal{F}_{t_j}$-measurable while D_j^2 is independent of $\mathcal{F}_{t_j}$ by the independent increments property. Hence, $\Delta_{t_j}^2$ and D_j^2 are independent, thus

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=0}^{k} \Delta_{t_j}^2 D_j^2\right] &= \sum_{j=0}^{k} \mathbb{E}\Big[\Delta_{t_j}^2\Big]\mathbb{E}\big[D_j^2\big] = \sum_{j=0}^{k-1} \mathbb{E}\Big[\Delta_{t_j}^2\Big](t_{j+1} - t_j) + \mathbb{E}\big[\Delta_{t_k}^2\big](t - t_k) \\
&\overset{\text{(simple process)}}{=} \sum_{j=0}^{k-1} \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \Delta_u^2 \,\mathrm{d}u\right] + \mathbb{E}\left[\int_{t_k}^{t} \Delta_u^2 \,\mathrm{d}u\right] = \mathbb{E}\left[\int_0^t \Delta_u^2 \,\mathrm{d}u\right].
\end{aligned}$$

Second, for every $i < j$, $\Delta_{t_i}\Delta_{t_j}D_i$ is $\mathcal{F}_{t_j}$-measurable, while D_j is independent of $\mathcal{F}_j$. So, $\Delta_{t_i}\Delta_{t_j}D_i$ and D_j are independent. Therefore,

$$\mathbb{E}\left[2\sum_{0\le i<j\le k}\Delta_{t_i}\Delta_{t_j}D_iD_j\right] = 2\sum_{0\le i<j\le k}\mathbb{E}\big[\Delta_{t_i}\Delta_{t_j}D_i\big]\underbrace{\mathbb{E}[D_j]}_{0} = 0.$$

It follows that $\mathbb{E}\big[I_t^2\big] = \mathbb{E}\Big[\int_0^t \Delta_u^2\,\mathrm{d}u\Big]$, as desired.

(c) With $t \in [t_k, t_{k+1}]$, fix any subinterval $[t_j, t_{j+1}] \subseteq [0,t]$ on which Δ_u is constant (so $j \in \{0,1,\dots,k-1\}$). Let $\Pi_j = \{s_0, s_1, \dots, s_m\}$ be a partition of $[t_j, t_{j+1}]$, with $t_j = s_0 < s_1 < \dots < s_m = t_{j+1}$. Consider

$$\sum_{i=0}^{m-1}(I_{s_{i+1}} - I_{s_i})^2 = \sum_{i=0}^{m-1}(\Delta_{t_j}(W_{s_{i+1}} - W_{s_i}))^2 = \Delta_{t_j}^2\sum_{i=0}^{m-1}(W_{s_{i+1}} - W_{s_i})^2.$$

From this, we know that

$$\begin{aligned}\lim_{\|\Pi_j\|\to 0}\sum_{i=0}^{m-1}(I_{s_{i+1}} - I_{s_i})^2 &= \Delta_{t_j}^2\lim_{\|\Pi_j\|\to 0}\sum_{i=0}^{m-1}(W_{s_{i+1}} - W_{s_i})^2 = \Delta_{t_j}^2([W,W]_{t_{j+1}} - [W,W]_{t_j}) \\ &\overset{\text{(Proposition 3.3.b)}}{=} \Delta_{t_j}^2(t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}}\Delta_u^2\,\mathrm{d}u\end{aligned}$$

In words, this means that the quadratic variation accumulated by the Itô integral between times t_j and t_{j+1} is $\int_{t_j}^{t_{j+1}}\Delta_u^2\,\mathrm{d}u$. In a similar way, one can show that the quadratic variation accumulated between times t_k and t is $\int_{t_k}^{t}\Delta_u^2\,\mathrm{d}u$. Hence, the quadratic variation of I up to time t is

$$[I,I]_t = \sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\Delta_u^2\,\mathrm{d}u + \int_{t_k}^{t}\Delta_u^2\,\mathrm{d}u = \int_0^t\Delta_u^2\,\mathrm{d}u$$

(when the norm of partition of the whole interval $[0,t]$ goes to zero, the norms of partitions of subintervals must go to zero as well).

□

One can also informally derive the quadratic variation of Itô integral through algebraic manipulations of differentials and differential rules:

$$\mathrm{d}I_t\,\mathrm{d}I_t \overset{\text{(substitution)}}{=} (\Delta_t\,\mathrm{d}W_t)(\Delta_t\,\mathrm{d}W_t) = \Delta_t^2\,\mathrm{d}W_t\,\mathrm{d}W_t \overset{\text{(differential rule)}}{=} \Delta_t^2\,\mathrm{d}t.$$

This method is frequently utilized in stochastic calculus for quick and efficient derivations, but it should be understood that such manipulations on the differentials are mathematically justified like the proof above (handling some kinds of limits). Later we will study more results about stochastic calculus, which provide us more differential rules to work with. In STAT3911, we put higher emphasis on such manipulations of differentials rather than the mathematical justifications behind them.

4.1.4 **Itô integral for general integrands.** Let $\{\Delta_t\}_{t\in[0,T]}$ be a stochastic process that is adapted to a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ for a Brownian motion $\{W_t\}_{t\in[0,T]}$, and is also **square-integrable**, i.e., $\mathbb{E}\Big[\int_0^T\Delta_t^2\,\mathrm{d}t\Big] < \infty$. The Itô integral of such general process $\{\Delta_t\}$ is then defined through approximating $\{\Delta_t\}$ by a sequence of simple processes, for which the Itô integral has been defined already. More precisely, the following result asserts the existence of such approximating sequence:

Proposition 4.1.a. For such adapted and square-integrable stochastic process $\{\Delta_t\}_{t\in[0,T]}$, there exists a sequence $\{\{\Delta_{t,n}\}_{t\in[0,T]}\}_{n\in\mathbb{N}}$ of simple processes such that

$$\lim_{n\to\infty}\mathbb{E}\left[\int_0^T|\Delta_{t,n} - \Delta_t|^2\,\mathrm{d}t\right] = 0.$$

Proof. Omitted. $\square$

With such approximating sequence, we can define the **Itô integral of the adapted and square-integrable process $\{\Delta_t\}$** by

$$I_t := \int_0^t \Delta_u \, \mathrm{d}W_u := \lim_{n\to\infty} \int_0^t \Delta_{u,n} \, \mathrm{d}W_u$$

for every $t \in [0,T]$. [Note: It can be shown that such limit always exists and so the Itô integral is well-defined. But we shall omit the technical details here.]

4.1.5 **Properties of Itô integral for general integrands.** Since the Itô integral for general integrands is defined as a limit of Itô integrals for simple integrands, one may naturally expect it to carry similar properties as the Itô integral for simple integrand. This is indeed the case, and its properties are collected below:

(a) *(Continuity)* For every fixed $\omega \in \Omega$, the function (path) $t \mapsto I_t(\omega)$ is continuous.

(b) *(Adaptivity)* $\{I_t\}$ is adapted to $\{\mathcal{F}_t\}$.

(c) *(Linearity)* If $I_t = \int_0^t \Delta_u \, \mathrm{d}W_u$ and $J_t = \int_0^t \Gamma_u \, \mathrm{d}W_u$, then $I_t \pm J_t = \int_0^t (\Delta_u \pm \Gamma_u) \, \mathrm{d}W_u$. Also, for every constant c, we have $cI_t = \int_0^t c\Delta_u \, \mathrm{d}W_u$.

(d) *(Martingale)* $\{I_t\}$ is a $\{\mathcal{F}_t\}$-martingale. Particularly, we have $\mathbb{E}[I_t] = 0$ for all t.

(e) *(Itô isometry)* $\mathbb{E}\big[I_t^2\big] = \mathbb{E}\Big[\int_0^t \Delta_u^2 \, \mathrm{d}u\Big]$.

(f) *(Quadratic variation)* $[I,I]_t = \int_0^t \Delta_u^2 \, \mathrm{d}u$.

4.2 Itô Formula and Itô Processes

4.2.1 **Itô formula for Brownian motion.** To work with stochastic calculus *efficiently*, the *Itô formula* (or *Itô lemma*) is an indispensable tool and is perhaps the most frequently used formula in STAT3911 (so make sure that you are familiar with it)! Intuitively, the Itô formula serves as a "stochastic version" of chain rule in multivariable calculus — Due to the extra "roughness" from the stochastic processes we are dealing with, there are some extra terms compared with the chain rule formula.

There are several versions of Itô formula, depending on the level of generality. We first consider the least general one (but still very important), which is for *Brownian motion*.

Theorem 4.2.a (Itô formula for Brownian motion)**.** Let $f(t,x)$ be a function of class C^2 (i.e., with continuous partial derivatives up to order 2) and let $\{W_t\}$ be a Brownian motion. For every $T \geq 0$, we have

$$f(T, W_T) = f(0, W_0) + \int_0^T f_t(t, W_t) \, \mathrm{d}t + \int_0^T f_x(t, W_t) \, \mathrm{d}W_t + \frac{1}{2} \int_0^T f_{xx}(t, W_t) \, \mathrm{d}t.$$

Differential form:

$$\mathrm{d}f(t, W_t) = f_t(t, W_t) \, \mathrm{d}t + f_x(t, W_t) \, \mathrm{d}W_t + \frac{1}{2} f_{xx}(t, W_t) \, \mathrm{d}t.$$

It can be informally derived as follows. By writing it as

$$\frac{\mathrm{d}f(t, W_t)}{\mathrm{d}t} = f_t(t, W_t) \frac{\mathrm{d}t}{\mathrm{d}t} + f_x(t, W_t) \frac{\mathrm{d}W_t}{\mathrm{d}t} + \frac{1}{2} f_{xx}(t, W_t) \frac{\mathrm{d}t}{\mathrm{d}t},$$

we can see that it takes a similar form as the chain rule in multivariable calculus, with the exception that there is an extra "$\frac{1}{2} f_{xx}(t, W_t) \frac{\mathrm{d}t}{\mathrm{d}t}$". Intuitively, such extra term appears since we have the differential rule $\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t$. Consequently, by performing "Taylor expansion on the differential" we would get

$$\begin{aligned}\mathrm{d}f(t,W_t) &= f_t(t,W_t)\,\mathrm{d}t + f_x(t,W_t)\,\mathrm{d}W_t + \frac{1}{2}f_{xx}(t,W_t)\underbrace{\mathrm{d}W_t\,\mathrm{d}W_t}_{\mathrm{d}t} \\ &\quad + f_{xt}(t,W_t)\underbrace{\mathrm{d}t\,\mathrm{d}W_t}_{0} + \frac{1}{2}f_{tt}(t,W_t)\underbrace{\mathrm{d}t\,\mathrm{d}t}_{0} + \underbrace{\text{higher order term}}_{0} \\ &= f_t(t,W_t)\,\mathrm{d}t + f_x(t,W_t)\,\mathrm{d}W_t + \frac{1}{2}f_{xx}(t,W_t)\,\mathrm{d}t.\end{aligned}$$

The much more involved mathematical proof is provided below for reference.

Proof. (If you are interested) Consider any partition $\Pi = \{t_0, t_1, \dots, t_n\}$ of $[0,T]$, with $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$. The main idea in the proof is to express the difference $f(T,W_T) - f(0,W_0)$ in terms of *Taylor series* of approximated differences within subintervals that involve partial derivatives. By Taylor's theorem, we have

$$\begin{aligned}f(T,W_T) - f(0,W_0) &= \sum_{j=0}^{n-1}\big(f(t_{j+1},W_{t_{j+1}}) - f(t_j,W_{t_j})\big) \\ &\overset{\text{(Taylor)}}{=} \sum_{j=0}^{n-1} f_t(t_j,W_{t_j})(t_{j+1}-t_j) + \sum_{j=0}^{n-1} f_x(t_j,W_{t_j})(W_{t_{j+1}} - W_{t_j}) \\ &\quad + \frac{1}{2}\sum_{j=0}^{n-1} f_{xx}(t_j,W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 + \sum_{j=0}^{n-1} f_{tx}(t_j,W_{t_j})(t_{j+1}-t_j)(W_{t_{j+1}} - W_{t_j}) \\ &\quad + \frac{1}{2}\sum_{j=0}^{n-1} f_{tt}(t_j,W_{t_j})(t_{j+1}-t_j)^2 + \text{higher order remainder term}.\end{aligned}$$

We now consider the limit as $\|\Pi\| \to 0$ and consider the six terms on the right-hand side one by one:

- $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_t(t_j,W_{t_j})(t_{j+1}-t_j) = \int_0^T f_t(t,W_t)\,\mathrm{d}t$ (by definition).
- $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_x(t_j,W_{t_j})(W_{t_{j+1}} - W_{t_j}) = \int_0^T f_x(t_j,W_{t_j})\,\mathrm{d}W_t$ (by definition).
- **Claim:** $\lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}(t_j,W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 = \frac{1}{2}\int_0^T f_{xx}(t,W_t)\,\mathrm{d}t$.
  Proof. Here we use a similar technique as in the proof of Proposition 3.3.b. For notational convenience, we let $\Delta W_j := W_{t_{j+1}} - W_{t_j}$ and $\Delta t_j = t_{j+1} - t_j$ for every $j = 0, 1, \dots, n-1$. First,

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\big(t_j,W_{t_j}\big)(\Delta W_j^2 - \Delta t_j)\right] \\ &= \frac{1}{2}\sum_{j=0}^{n-1}\mathbb{E}\big[\mathbb{E}\big[f_{xx}\big(t_j,W_{t_j}\big)(\Delta W_j^2 - \Delta t_j)\big|\mathcal{F}_{t_j}\big]\big] \overset{\text{(TOWIK)}}{=} \frac{1}{2}\sum_{j=0}^{n-1}\mathbb{E}\big[f_{xx}\big(t_j,W_{t_j}\big)\mathbb{E}\big[(\Delta W_j^2 - \Delta t_j)\big|\mathcal{F}_{t_j}\big]\big] \\ &\overset{\text{(independence)}}{=} \frac{1}{2}\sum_{j=0}^{n-1}\mathbb{E}\big[f_{xx}\big(t_j,W_{t_j}\big)\mathbb{E}\big[(\Delta W_j^2 - \Delta t_j)\big]\big] = \frac{1}{2}\sum_{j=0}^{n-1}\mathbb{E}\big[f_{xx}\big(t_j,W_{t_j}\big)\cdot 0\big] = 0.\end{aligned}$$

Second,

$$
\begin{aligned}
&\operatorname{Var}\left(\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})(\Delta W_j^2 - \Delta t_j)\right) \overset{\text{(zero mean)}}{=} \mathbb{E}\left[\left(\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})(\Delta W_j^2 - \Delta t_j)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})^2(\Delta W_j^2 - \Delta t_j)^2\right] \\
&\quad + 2\mathbb{E}\left[\sum_{0\le i<j\le n-1} f_{xx}(t_i, W_{t_i}) f_{xx}(t_j, W_{t_j})(\Delta W_i^2 - \Delta t_i)(\Delta W_j^2 - \Delta t_j)\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}\big[\mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2(\Delta W_j^2 - \Delta t_j)^2 | \mathcal{F}_{t_j}\big]\big] \\
&\quad + 2\sum_{0\le i<j\le n-1} \mathbb{E}\big[\mathbb{E}\big[f_{xx}(t_i, W_{t_i}) f_{xx}(t_j, W_{t_j})(\Delta W_i^2 - \Delta t_i)(\Delta W_j^2 - \Delta t_j) | \mathcal{F}_{t_j}\big]\big] \\
&\overset{\text{(TOWIK)}}{=} \sum_{j=0}^{n-1} \mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2 \mathbb{E}\big[(\Delta W_j^2 - \Delta t_j)^2 | \mathcal{F}_{t_j}\big]\big] \\
&\quad + 2\sum_{0\le i<j\le n-1} \mathbb{E}\big[f_{xx}(t_i, W_{t_i}) f_{xx}(t_j, W_{t_j})(\Delta W_i^2 - \Delta t_i)\mathbb{E}\big[\Delta W_j^2 - \Delta t_j | \mathcal{F}_{t_j}\big]\big] \\
&\overset{\text{(independence)}}{=} \sum_{j=0}^{n-1} \mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2 \mathbb{E}\big[(\Delta W_j^2 - \Delta t_j)^2\big]\big] \\
&\quad + 2\sum_{0\le i<j\le n-1} \mathbb{E}\big[f_{xx}(t_i, W_{t_i}) f_{xx}(t_j, W_{t_j})(\Delta W_i^2 - \Delta t_i)\mathbb{E}\big[\Delta W_j^2 - \Delta t_j\big]\big] \\
&\overset{(\mathbb{E}[\Delta W_j^2 - \Delta t_j]=0)}{=} \sum_{j=0}^{n-1} \mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2\big] \operatorname{Var}\big(\Delta W_j^2 - \Delta t_j\big) = \sum_{j=0}^{n-1} \mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2\big] \operatorname{Var}\big(\Delta W_j^2\big) \\
&\overset{(X\sim \mathrm{N}(0,\sigma^2)\Rightarrow \mathbb{E}[X^4]=3(\sigma^2)^2)}{=} \sum_{j=0}^{n-1} \mathbb{E}\big[f_{xx}(t_j, W_{t_j})^2\big](3\Delta t_j^2 - \Delta t_j^2) \le 2\|\Pi\|\mathbb{E}\left[\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})^2 \Delta t_j\right].
\end{aligned}
$$

Since $\lim_{\|\Pi\|\to 0} 2\|\Pi\|\mathbb{E}\Big[\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})^2 \Delta t_j\Big] = 0$[10], we know that

$$
\lim_{\|\Pi\|\to 0} \operatorname{Var}\left(\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})(\Delta W_j^2 - \Delta t_j)\right) = 0.
$$

Therefore, we have

$$
\begin{aligned}
&\lim_{\|\Pi\|\to 0} \frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\big(t_j, W_{t_j}\big)(\Delta W_j^2 - \Delta t_j) = 0 \\
\implies &\lim_{\|\Pi\|\to 0} \frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\big(t_j, W_{t_j}\big)(W_{t_{j+1}} - W_{t_j})^2 = \frac{1}{2}\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_{xx}\big(t_j, W_{t_j}\big)\Delta t_j = \frac{1}{2}\int_0^T f_{xx}(t, W_t)\,\mathrm{d}t.
\end{aligned}
$$

$\square$

[10]Some work is actually needed for establishing this (particularly, we need to ensure that $\mathbb{E}\Big[\sum_{j=0}^{n-1} f_{xx}(t_j, W_{t_j})^2 \Delta t_j\Big]$ would not "explode" as $\|\Pi\| \to 0$); see Etheridge (2002, Theorem 4.3.1) for more details.

- **Claim:** $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_{tx}(t_j, W_{t_j})(t_{j+1}-t_j)(W_{t_{j+1}}-W_{t_j}) = 0$.

  Proof. Note that

$$
\begin{aligned}
&\lim_{\|\Pi\|\to 0}\left|\sum_{j=0}^{n-1} f_{tx}(t_j, W_{t_j})(t_{j+1}-t_j)(W_{t_{j+1}}-W_{t_j})\right| \\
&\overset{\text{(triangle)}}{\le} \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}|f_{tx}(t_j, W_{t_j})|(t_{j+1}-t_j)|W_{t_{j+1}}-W_{t_j}| \\
&\le \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|W_{t_{j+1}}-W_{t_j}|\right)\sum_{j=0}^{n-1}|f_{tx}(t_j, W_{t_j})|(t_{j+1}-t_j) \overset{\text{(continuity)}}{=} 0\cdot\underbrace{\int_0^T |f_{tx}(t, W_t)|\,\mathrm{d}t}_{<\infty} = 0.
\end{aligned}
$$

  □

- **Claim:** $\lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1} f_{tt}(t_j, W_{t_j})(t_{j+1}-t_j)^2 = 0$.

  Proof. Note that

$$
\begin{aligned}
&\lim_{\|\Pi\|\to 0}\left|\frac{1}{2}\sum_{j=0}^{n-1} f_{tx}(t_j, W_{t_j})(t_{j+1}-t_j)^2\right| \\
&\overset{\text{(triangle)}}{\le} \lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1}|f_{tt}(t_j, W_{t_j})|(t_{j+1}-t_j)^2 \\
&\le \frac{1}{2}\lim_{\|\Pi\|\to 0}\|\Pi\|\sum_{j=0}^{n-1}|f_{tx}(t_j, W_{t_j})|(t_{j+1}-t_j) = \frac{1}{2}\cdot 0\cdot\underbrace{\int_0^T |f_{tt}(t, W_t)|\,\mathrm{d}t}_{<\infty} = 0.
\end{aligned}
$$

  □

- **Claim:** The higher order remainder term also converges to 0 as $\|\Pi\|\to 0$.

  Proof. Similar to the two proofs above. □

With all these (tedious) arguments, we can finally establish that

$$
f(T, W_T)-f(0, W_0) = \lim_{\|\Pi\|\to 0} f(T, W_T)-f(0, W_0) = \int_0^T f_t(t, W_t)\,\mathrm{d}t + \int_0^T f_x(t, W_t)\,\mathrm{d}W_t + \frac{1}{2}\int_0^T f_{xx}(t, W_t)\,\mathrm{d}t,
$$

as desired. □

4.2.2 **Examples of applications of Itô formula.**

(a) Let $f(t,x) = x^2$. Then by Itô formula we have

$$
\mathrm{d}W_t^2 = \underbrace{f_t(t, W_t)}_{0}\,\mathrm{d}t + \underbrace{f_x(t, W_t)}_{2x|_{x=W_t}=2W_t}\,\mathrm{d}W_t + \frac{1}{2}\underbrace{f_{xx}(t, W_t)}_{2}\,\mathrm{d}t = 2W_t\,\mathrm{d}W_t + \mathrm{d}t.
$$

Expressing this in integral form, we have

$$
W_T^2 - \underbrace{W_0^2}_{0} = \int_0^T 2W_t\,\mathrm{d}W_t + \int_0^T \mathrm{d}t,
$$

which implies that

$$
\int_0^T W_t\,\mathrm{d}W_t = \frac{1}{2}W_T^2 - \frac{T}{2}.
$$

[Note: For deterministic function $g(t)$ with $g(0) = 0$, we would have $\int_0^T g(t)\,\mathrm{d}g(t) = \frac{1}{2}g(T)^2$. The extra term $-T/2$ appearing above comes from the "roughness" of the Brownian motion $\{W_t\}$.]

(b) *(Geometric Brownian motion)* Let $f(t,x) = S_0 e^{(\mu-\sigma^2/2)t+\sigma x}$, and let $S_t := f(t, W_t) = S_0 e^{(\mu-\sigma^2/2)t+\sigma W_t}$, with nonrandom $S_0 > 0$ and parameters $\mu \in \mathbb{R}, \sigma > 0$. By Itô formula, we get

$$\mathrm{d}S_t = \underbrace{f_t(t,W_t)}_{S_t(\mu-\sigma^2/2)}\,\mathrm{d}t + \underbrace{f_x(t,W_t)}_{S_t\sigma}\,\mathrm{d}W_t + \frac{1}{2}\underbrace{f_{xx}(t,W_t)}_{S_t\sigma^2}\,\mathrm{d}t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t.$$

Here the process $\{S_t\}$ follows the **geometric Brownian motion** (the one studied in STAT3910!), and this SDE representation is a common way to describe this process.

The coefficient of $\mathrm{d}t$ (namely μS_t here) is sometimes known as the **drift term**, and the coefficient of $\mathrm{d}W_t$ (namely σS_t here) is sometimes known as the **volatility term**. Intuitively, the SDE suggests that after moving "a little" forward in time from time t, the movement in S (e.g., stock price) would be approximately the sum of (i) the drift term (μS_t here) times the length of time elapsed and (ii) the volatility term (σS_t here) times the change in the Brownian motion W. The geometric Brownian motion is often seen as a nice and "simple" model for capturing the dynamics of stock price movements since historically stock price generally exhibits an "exponential growth" (corresponding to the *drift term*) but with volatilities in between (corresponding to the *volatility term*).

4.2.3 **Itô processes.** Previously we have defined the Itô integral $I_t = \int_0^t \Delta_u\,\mathrm{d}W_u$ for adapted and square-integrable process $\{\Delta_t\}$. The process $\{I_t\}$ of such Itô integrals only captures a limited amount of stochastic processes we are interested in. To generalize this, we will develop the concept *Itô process*, which includes almost all stochastic processes of interest, except those with jumps.

Let $\{W_t\}$ be a Brownian motion and $\{\mathcal{F}_t\}$ be a filtration for the Brownian motion. An **Itô process** is a $\{\mathcal{F}_t\}$-adapted stochastic process $\{X_t\}$ given by

$$X_t = X_0 + \int_0^t \Delta_u\,\mathrm{d}W_u + \int_0^t \Theta_u\,\mathrm{d}u \quad \text{for every } t,$$

where X_0 is nonrandom, and both $\{\Delta_u\}$ and $\{\Theta_u\}$ are adapted to $\{\mathcal{F}_t\}$ with integrability conditions satisfied: (i) $\mathbb{E}\left[\int_0^t \Delta_u^2\,\mathrm{d}u\right] < \infty$ and (ii) $\int_0^t |\Theta_u|\,\mathrm{d}u < \infty$, for every t.

[Note: Whenever we work with expressions like this, it should be understood that these integrability conditions are always assumed (sometimes implicitly), so that the expressions are defined and the discussion becomes meaningful.]

Differential form: $\mathrm{d}X_t = \Delta_t\,\mathrm{d}W_t + \Theta_t\,\mathrm{d}t$.

[Note: By taking $\Theta_u \equiv 0$, $\Delta_u \equiv 1$, and $X_0 = W_0$, the Itô process would be reduced to simply a Brownian motion (as $\int_0^t 1\,\mathrm{d}W_u = W_t - W_0$ by considering the definition).]

4.2.4 **Quadratic variation of Itô process.** While an Itô process $\{X_t\}$ is much more general than a process of Itô integrals $\{I_t\}$, they turn out to accumulate quadratic variation at the same rate. Intuitively, this happens because the only "source" of quadratic variation for X_t is the I_t term: $\int_0^t \Delta_u\,\mathrm{d}W_u$. It can be more formally proven as follows.

Proposition 4.2.b. Let $\{X_t\}$ be an Itô process. Then, its quadratic variation is $[X,X]_t = \int_0^t \Delta_u^2\,\mathrm{d}u$ for every t.

Differential form: $\mathrm{d}X_t\,\mathrm{d}X_t = \Delta_t^2\,\mathrm{d}t$.

This result can be easily derived by manipulating the differentials as follows:

$$\mathrm{d}X_t\,\mathrm{d}X_t = (\Delta_t\,\mathrm{d}W_t + \Theta_t\,\mathrm{d}t)(\Delta_t\,\mathrm{d}W_t + \Theta_t\,\mathrm{d}t) \overset{\text{(expansion)}}{=} \Delta_t^2 \underbrace{\mathrm{d}W_t\,\mathrm{d}W_t}_{\mathrm{d}t} + 2\Delta_t\Theta_t \underbrace{\mathrm{d}W_t\,\mathrm{d}t}_{0} + \Theta_t^2 \underbrace{\mathrm{d}t\,\mathrm{d}t}_{0} = \Delta_t^2\,\mathrm{d}t.$$

The mathematical proof is more involved, and is provided below for reference.

Proof. *(If you are interested)* Let $I_t = \int_0^t \Delta_u \,\mathrm{d}W_u$ and $R_t = \int_0^t \Theta_u \,\mathrm{d}u$, and $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0,t]$ with $0 = t_0 \le t_1 \le \cdots \le t_n = t$. By [4.1.5] and the fundamental theorem of calculus, both I_t and R_t are continuous in t. Next, write

$$\sum_{j=0}^{n-1}(X_{t_{j+1}} - X_{t_j})^2 = \sum_{j=0}^{n-1}(I_{t_{j+1}} - I_{t_j})^2 + \sum_{j=0}^{n-1}(R_{t_{j+1}} - R_{t_j})^2 + 2\sum_{j=0}^{n-1}(I_{t_{j+1}} - I_{t_j})(R_{t_{j+1}} - R_{t_j}).$$

Now, consider the limit as $\|\Pi\| \to 0$ and consider the three terms on the right-hand side one by one:

- $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(I_{t_{j+1}} - I_{t_j})^2 \overset{\text{(definition)}}{=} [I,I]_t \overset{[4.1.5]}{=} \int_0^t \Delta_u^2 \,\mathrm{d}u.$
- **Claim:** $\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(R_{t_{j+1}} - R_{t_j})^2 = 0.$
  Proof. Note that

$$\begin{aligned}
\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(R_{t_{j+1}} - R_{t_j})^2 &\le \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|R_{t_{j+1}} - R_{t_j}|\right)\sum_{j=0}^{n-1}|R_{t_{j+1}} - R_{t_j}| \\
&= \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|R_{t_{j+1}} - R_{t_j}|\right)\sum_{j=0}^{n-1}\left|\int_{t_j}^{t_{j+1}}\Theta_u \,\mathrm{d}u\right| \\
&\overset{\text{(triangle)}}{\le} \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|R_{t_{j+1}} - R_{t_j}|\right)\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|\Theta_u| \,\mathrm{d}u \\
&= \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|R_{t_{j+1}} - R_{t_j}|\right)\underbrace{\int_0^t |\Theta_u| \,\mathrm{d}u}_{<\infty} \\
&\overset{\text{(continuity)}}{=} 0.
\end{aligned}$$

$\square$

- **Claim:** $\lim_{\|\Pi\|\to 0} 2\sum_{j=0}^{n-1}(I_{t_{j+1}} - I_{t_j})(R_{t_{j+1}} - R_{t_j}) = 0.$
  Proof. Note that

$$\begin{aligned}
\lim_{\|\Pi\|\to 0} 2\sum_{j=0}^{n-1}(I_{t_{j+1}} - I_{t_j})(R_{t_{j+1}} - R_{t_j}) &\le 2\lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|I_{t_{j+1}} - I_{t_j}|\right)\sum_{j=0}^{n-1}|R_{t_{j+1}} - R_{t_j}| \\
&\le \lim_{\|\Pi\|\to 0}\left(\max_{0\le j\le n-1}|I_{t_{j+1}} - I_{t_j}|\right)\underbrace{\int_0^t |\Theta_u| \,\mathrm{d}u}_{<\infty} \\
&\overset{\text{(continuity)}}{=} 0.
\end{aligned}$$

$\square$

Therefore,

$$[X,X]_t = \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(X_{t_{j+1}} - X_{t_j})^2 = \int_0^t \Delta_u^2 \,\mathrm{d}u.$$

$\square$

4.2.5 **Itô integral with respect to Itô process.** Previously, we have discussed the Itô integral with respect to a Brownian motion: $\int_0^t \Delta_u \,\mathrm{d}W_u$. We can generalize it to allow taking integral with respect to an *Itô process*. Let $\{X_t\}$ be an Itô process associated with a filtration $\{\mathcal{F}_t\}$, and $\{\Gamma_t\}$ be a $\{\mathcal{F}_t\}$-adapted process. Then, we define the **Itô integral with respect to an Itô process** by

$$\int_0^t \Gamma_u \,\mathrm{d}X_u := \int_0^t \Gamma_u\Delta_u \,\mathrm{d}W_u + \int_0^t \Gamma_u\Theta_u \,\mathrm{d}u \quad \text{for every } t,$$

provided that the integrability conditions are satisfied: (i) $\mathbb{E}\left[\int_0^t \Gamma_u^2 \Delta_u^2 \,\mathrm{d}u\right] < \infty$ and (ii) $\int_0^t |\Gamma_u \Theta_u| \,\mathrm{d}u < \infty$, for every t.

In differential form, we can write $\Gamma_t \,\mathrm{d}X_t = \Gamma_t \Delta_t \,\mathrm{d}W_t + \Gamma_t \Theta_t \,\mathrm{d}t$. So, this definition basically permits us to "distribute" Γ_t in the differential form: $\Gamma_t \,\mathrm{d}X_t = \Gamma_t(\Delta_t \,\mathrm{d}W_t + \Theta_t \,\mathrm{d}t) = \Gamma_t \Delta_t \,\mathrm{d}W_t + \Gamma_t \Theta_t \,\mathrm{d}t$.

4.2.6 **Itô formula for Itô process.** With the Itô integral with respect to Itô process defined, we can also generalize the Itô formula before, to the case where Itô process is considered rather than Brownian motion.

Theorem 4.2.c (Itô formula for Itô process)**.** Let $\{X_t\}$ be an Itô process, and let $f(t,x)$ be a function of class C^2. Then, for every $T \geq 0$, we have

$$f(T, X_T) = f(0, X_0) + \int_0^T f_t(t, X_t) \,\mathrm{d}t + \int_0^T f_x(t, X_t) \,\mathrm{d}X_t + \frac{1}{2} \int_0^T f_{xx}(t, X_t) \Delta_t^2 \,\mathrm{d}t.$$

[Note: Sometimes, one may abuse notations slightly and write the last term as "$\frac{1}{2}\int_0^T f_{xx}(t, X_t) \,\mathrm{d}X_t \,\mathrm{d}X_t$", which is perhaps more intuitive but it should be understood to carry the meaning above.]

Differential form:

$$\mathrm{d}f(t, X_t) = f_t(t, X_t) \,\mathrm{d}t + f_x(t, X_t) \,\mathrm{d}X_t + \frac{1}{2} f_{xx}(t, X_t) \Delta_t^2 \,\mathrm{d}t.$$

It again has a similar form as the chain rule in multivariable calculus, and can be informally derived through the following manipulations of differentials:

$$\begin{aligned}
\mathrm{d}f(t, X_t) &= f_t(t, X_t) \,\mathrm{d}t + f_x(t, X_t) \,\mathrm{d}X_t + \frac{1}{2} f_{xx}(t, X_t) \underbrace{\mathrm{d}X_t \,\mathrm{d}X_t}_{\Delta_t^2 \,\mathrm{d}t} \\
&\quad + f_{xt}(t, X_t) \underbrace{\mathrm{d}t \,\mathrm{d}X_t}_{\mathrm{d}t(\Delta_t \,\mathrm{d}W_t + \Theta_t \,\mathrm{d}t) = \Delta_t \,\mathrm{d}t \,\mathrm{d}W_t + \Theta_t \,\mathrm{d}t \,\mathrm{d}t = 0} + \frac{1}{2} f_{tt}(t, X_t) \underbrace{\mathrm{d}t \,\mathrm{d}t}_{0} + \underbrace{\text{higher order term}}_{0} \\
&= f_t(t, X_t) \,\mathrm{d}t + f_x(t, X_t) \,\mathrm{d}X_t + \frac{1}{2} f_{xx}(t, X_t) \Delta_t^2 \,\mathrm{d}t.
\end{aligned}$$

Proof. Omitted. □

4.2.7 **Generalized geometric Brownian motion.** To illustrate the usage of the more general Itô formula from Theorem 4.2.c, we consider the *generalized geometric Brownian motion*. Let $\{W_t\}$ be a Brownian motion and $\{\mathcal{F}_t\}$ be a filtration for the Brownian motion. Let $\{\mu_t\}$ and $\{\sigma_t\}$ be $\{\mathcal{F}_t\}$-adapted processes (such that the relevant integrability conditions are satisfied for the following Itô process), with $\sigma_t > 0$ for every t. Consider an Itô process $\{X_t\}$ defined by

$$X_t = \int_0^t \sigma_u \,\mathrm{d}W_u + \int_0^t \left(\mu_u - \frac{1}{2}\sigma_u^2\right) \mathrm{d}u \quad \text{for every } t.$$

In differential form, we write $\mathrm{d}X_t = \sigma_t \,\mathrm{d}W_t + (\mu_t - \sigma_t^2/2) \,\mathrm{d}t$. Also, by Proposition 4.2.b, we know that $\mathrm{d}X_t \,\mathrm{d}X_t = \sigma_t^2 \,\mathrm{d}t$. The **generalized geometric Brownian motion** is the process $\{S_t\}$ defined by

$$S_t = S_0 e^{X_t} = S_0 e^{\int_0^t \sigma_u \,\mathrm{d}W_u + \int_0^t \left(\mu_u - \frac{1}{2}\sigma_u^2\right) \mathrm{d}u} \quad \text{for every } t,$$

where $S_0 > 0$ is nonrandom. Letting $f(t,x) = S_0 e^x$, we have $S_t = f(t, X_t)$. Hence, by Itô formula, we have

$$\mathrm{d}S_t = \underbrace{f_t(t, X_t)}_{0} \,\mathrm{d}t + \underbrace{f_x(t, X_t)}_{S_0 e^x|_{x = X_t} = S_t} \,\mathrm{d}X_t + \frac{1}{2} \underbrace{f_{xx}(t, X_t)}_{S_0 e^x|_{x = X_t} = S_t} \sigma_t^2 \,\mathrm{d}t = S_t \underbrace{\mathrm{d}X_t}_{\sigma_t \,\mathrm{d}W_t + (\mu_t - \sigma_t^2/2) \,\mathrm{d}t} + \frac{1}{2} S_t \sigma_t^2 \,\mathrm{d}t = \mu_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t.$$

From this, we can see that the generalized geometric Brownian motion is indeed a generalization to the geometric Brownian motion, with the parameters μ and σ allowed to be time-varying.

4.2.8 **Zero drift term leads to a martingale.** For the geometric Brownian motion, in case the drift term (coefficient of $\mathrm{d}t$) is always zero, the SDE above would be reduced to simply $\mathrm{d}S_t = \sigma_t S_t \,\mathrm{d}W_t$. Expressing this in integral form, we have

$$S_t = S_0 + \int_0^t \sigma_u S_u \,\mathrm{d}W_u =: S_0 + I_t$$

By [4.1.5], we know that the process of Itô integrals $\{I_t\}$ is a martingale. Furthermore, the constant process $\{S_0\}$ is clearly a martingale. Hence, by Proposition 2.3.a, $\{S_t\}$ is also a martingale. This gives rise to a very useful technique for showing that a stochastic process $\{M_t\}$ is martingale:

(1) Write $M_t = f(t, W_t)$ where $\{W_t\}$ is a Brownian motion, or $M_t = f(t, X_t)$ where $\{X_t\}$ is an Itô process.

(2) Apply the Itô formula for Brownian motion/Itô process on M_t to get

$$\mathrm{d}M_t = Y \,\mathrm{d}t + Z \,\mathrm{d}W_t$$

for some (possibly random) Y and Z.

(3) Show that the drift term Y equals 0 always, and hence conclude that $\{M_t\}$ is a martingale.

[Note: It can be shown that the converse of this result also holds, i.e., being a martingale implies that the drift term is zero.]

4.2.9 **Normality of Itô integral for a nonrandom integrand.** As an application of the technique in [4.2.8], we will prove a result that asserts the normality of Itô integral for a nonrandom integrand.

Proposition 4.2.d. Let $\{W_t\}$ be a Brownian motion and Δ_u be a nonrandom and square-integrable function of u. Then, we have

$$I_t = \int_0^t \Delta_u \,\mathrm{d}W_u \sim \mathrm{N}\left(0, \int_0^t \Delta_u^2 \,\mathrm{d}u\right).$$

Proof. We will show the normality through the moment generating function. Fix any $s \in \mathbb{R}$ and consider

$$\mathbb{E}\left[e^{sI_t - \frac{1}{2}s^2 \int_0^t \Delta_u \,\mathrm{d}u}\right] = \mathbb{E}\left[e^{\int_0^t s\Delta_u \,\mathrm{d}u + \int_0^t -\frac{1}{2}(s\Delta_u)^2 \,\mathrm{d}u}\right] =: \mathbb{E}[M_t].$$

Note that $\{M_t\}$ follows the geometric Brownian motion with $S_0 = 1$, $\mu_t \equiv 0$ and $\sigma_t = s\Delta_t$. As the drift term is always zero, we conclude by [4.2.8] that $\{M_t\}$ is a martingale, and hence $\mathbb{E}[M_t] = \mathbb{E}[\mathbb{E}[M_t|\mathcal{F}_0]] = \mathbb{E}[M_0] = e^0 = 1$. Since Δ_u is nonrandom, we have $e^{-\frac{1}{2}s^2 \int_0^t \Delta_u \,\mathrm{d}u}\mathbb{E}\left[e^{sI_t}\right] = 1$, and thus

$$\mathbb{E}\left[e^{sI_t}\right] = e^{\frac{1}{2}s^2 \int_0^t \Delta_u \,\mathrm{d}u},$$

which is the moment generating function of $\mathrm{N}\left(0, \int_0^t \Delta_u^2 \,\mathrm{d}u\right)$. □

4.2.10 **More examples of applications of Itô formula: Two interest rate models.** To become more familiar with the very important Itô formula, we will analyze two well-known interest rate models via the Itô formula in the following, namely the *Vasicek interest rate model* and the *Cox-Ingersoll-Ross (CIR) interest rate model*.

(a) *(Vasicek)* Let $\{W_t\}$ be a Brownian motion. The **Vasicek model** for an instantaneous interest rate (or *short rate*) process $\{R_t\}$ is described by the SDE

$$\mathrm{d}R_t = (\alpha - \beta R_t) \,\mathrm{d}t + \sigma \,\mathrm{d}W_t$$

where α, β, and σ are positive parameters, and R_0 is nonrandom. A closed-form solution (as an Itô process) is available for the SDE above, and can be derived by the Itô formula.

(1) *Idea of solving SDE.* Before working with the mathematics, let us provide some ideas on how to solve a SDE in general. Unfortunately, like the case for the ordinary differential equation (ODE), there is not a general "rule" for solving *any* SDE. However, a typical approach would be collecting the terms and "guessing" the potential solutions (like the case for ODE). To illustrate this, here we rewrite the SDE above by collecting terms involving "R_t" together:

$$\mathrm{d}R_t + \beta R_t \,\mathrm{d}t = \alpha \,\mathrm{d}t + \sigma \,\mathrm{d}W_t.$$

For the expression "$\mathrm{d}R_t + \beta R_t \,\mathrm{d}t$", we *guess* that the solution would be in exponential form as this kind of expression arises from "differentiating exponentials": More specifically, applying Itô formula (for Itô process) with $f(t,x) = e^{\beta t}x$ gives

$$\mathrm{d}(e^{\beta t}R_t) = \mathrm{d}f(t,R_t) = \underbrace{f_t(t,R_t)}_{\beta e^{\beta t}R_t}\,\mathrm{d}t + \underbrace{f_x(t,R_t)}_{e^{\beta t}}\,\mathrm{d}R_t + 0 = e^{\beta t}(\beta R_t \,\mathrm{d}t + \mathrm{d}R_t),$$

where the expression "$\beta R_t \,\mathrm{d}t + \mathrm{d}R_t$" appears. This observation then inspires us to "complete the exponential" as follows.

(2) *Completing the exponential.* Multiplying both sides of the rewritten SDE by $e^{\beta t}$, we get

$$\mathrm{d}(e^{\beta t}R_t) \overset{\text{(previous)}}{=} e^{\beta t}(\mathrm{d}R_t + \beta R_t \,\mathrm{d}t) = \alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t}\,\mathrm{d}W_t.$$

Expressing this in integral form, we have

$$e^{\beta t}R_t - e^0 R_0 = \int_0^t \alpha e^{\beta u}\,\mathrm{d}u + \int_0^t \sigma e^{\beta u}\,\mathrm{d}W_u,$$

which implies that

$$R_t = e^{-\beta t}R_0 + e^{-\beta t}\left[\frac{\alpha}{\beta}(e^{\beta t}-1)\right] + \sigma e^{-\beta t}\int_0^t e^{\beta u}\,\mathrm{d}W_u,$$

which forms an Itô process. This establishes the "forward implication" (i.e., if the SDE holds, then R_t can only be in the form above), which shows that such R_t is the only *potential* solution to the SDE. Mathematically speaking, to solve the SDE completely, we also need to show the "backward implication" (i.e., if R_t takes the form above, then the SDE with such R_t plugged in would hold), which can ensure that such R_t serves as the actual (and only) solution to the SDE.

(3) *Verifying the "backward implication".* To establish the "backward implication", let $f(t,x) = e^{-\beta t}R_0 + \frac{\alpha}{\beta}(1-e^{-\beta t}) + \sigma e^{-\beta t}x$. Compute:
- $f_t(t,x) = -\beta e^{-\beta t}R_0 + \alpha e^{-\beta t} - \sigma\beta e^{-\beta t}x = \alpha - \beta f(t,x)$.
- $f_x(t,x) = \sigma e^{-\beta t}$.
- $f_{xx}(t,x) = 0$.

Applying the Itô formula with $X_t = \int_0^t e^{\beta u}\,\mathrm{d}W_u$ (in differential form: $\mathrm{d}X_t = e^{\beta t}\,\mathrm{d}W_t$, with $X_0 = 0$), we have

$$\mathrm{d}f(t,X_t) = f_t(t,X_t)\,\mathrm{d}t + f_x(t,X_t)\underbrace{\mathrm{d}X_t}_{e^{\beta t}\,\mathrm{d}W_t} + 0 = (\alpha - \beta f(t,X_t))\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Since $f(t,X_t) = e^{-\beta t}R_0 + e^{-\beta t}\left[\frac{\alpha}{\beta}(e^{\beta t}-1)\right] + \sigma e^{-\beta t}\int_0^t e^{\beta u}\,\mathrm{d}W_u$, the SDE indeed holds with such form of R_t plugged in.

Therefore, the closed-form solution $\{R_t\}$ to the Vasicek SDE is given by

$$R_t = e^{-\beta t}R_0 + e^{-\beta t}\left[\frac{\alpha}{\beta}(e^{\beta t}-1)\right] + \sigma e^{-\beta t}\int_0^t e^{\beta u}\,\mathrm{d}W_u.$$

Properties of the Vasicek model.

- *(normality)* By Proposition 4.2.d, we know $\int_0^t e^{\beta u}\,\mathrm{d}W_u \sim \mathrm{N}\Big(0, \int_0^t e^{2\beta u}\,\mathrm{d}u\Big) = \mathrm{N}\Big(0, \frac{1}{2\beta}(e^{2\beta t}-1)\Big)$. Hence,

$$R_t \sim \mathrm{N}\Bigg(e^{-\beta t}R_0 + e^{-\beta t}\left[\frac{\alpha}{\beta}(e^{\beta t}-1)\right], \frac{\sigma^2}{2\beta}(1-e^{-2\beta t}).\Bigg)$$

  [Note: The normality reveals a potential issue ⚠ of the Vasicek model, namely that R_t can take negative values. While it has recently been observed that interest rates can indeed get negative (e.g., in Japan), it is seldom the case. So, one should be careful in assessing whether the potential negativity of interest rates implied by the Vasicek model is sensible.]
- *(mean-reverting)*
  - When $R_t = \alpha/\beta$, the drift term is zero, and so there is not a "tendency" for the interest rate movement.
  - When $R_t > \alpha/\beta$, the drift term is negative, and so there is a *downward* "tendency" for the interest rate movement (pushing it back towards α/β).
  - When $R_t < \alpha/\beta$, the drift term is positive, and so there is an *upward* "tendency" for the interest rate movement (pushing it back towards α/β).

  Furthermore, we note that if $R_0 = \alpha/\beta$, then we have $\mathbb{E}[R_t] = \alpha/\beta$ for every t. Even if $R_0 \neq \alpha/\beta$, we also have $\lim_{t\to\infty}\mathbb{E}[R_t] = \alpha/\beta$ in this case. These suggest that the Vasicek model exhibits a *mean-reverting* behaviour, in the sense that the interest rate level R_t would always tend to "revert" back to its (long-term) mean level α/β. The larger the deviation from its (long-term) mean level, the "stronger" the reverting power.

(b) *(CIR)* Let $\{W_t\}$ be a Brownian motion. The **Cox-Ingersoll-Ross (CIR) model** for an instantaneous interest rate process $\{R_t\}$ is described by the SDE

$$\mathrm{d}R_t = (\alpha - \beta R_t)\,\mathrm{d}t + \sigma\sqrt{R_t}\,\mathrm{d}W_t$$

where α, β, and σ are positive constants, and $R_0 \geq 0$ is nonrandom. While this SDE does not admit a closed-form solution, it avoids the potential issue of negative interest rate existing in the Vasicek model (but may not necessarily be a good thing as negative interest rate is indeed possible). Intuitively, negative interest rate is avoided since as $R_t \to 0^+$, the volatility term would go to 0 while the drift term would go to $\alpha > 0$, forcing R_t to rise and take positive values, thereby eliminating the possibility for R_t to be negative.

Moreover, since the drift term in the CIR model is the same as the one for Vasicek model, the CIR model has the mean-reverting property also.

Due to the lack of closed-form solution, here we will just derive the expectation and variance of R_t. We start by rearranging the SDE as

$$\mathrm{d}R_t + \beta R_t\,\mathrm{d}t = \alpha\,\mathrm{d}t + \sigma\sqrt{R_t}\,\mathrm{d}W_t.$$

Then, we "complete the exponential" by multiplying both sides by $e^{\beta t}$ like before:

$$\mathrm{d}e^{\beta t}R_t = e^{\beta t}(\mathrm{d}R_t + \beta R_t\,\mathrm{d}t) = \alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t}\sqrt{R_t}\,\mathrm{d}W_t.$$

Expressing this in integral form, we have

$$e^{\beta t}R_t = R_0 + \int_0^t \alpha e^{\beta u}\,\mathrm{d}u + \int_0^t \sigma e^{\beta u}\sqrt{R_u}\,\mathrm{d}W_u = R_0 + \frac{\alpha}{\beta}(e^{\beta t}-1) + \sigma\int_0^t e^{\beta u}\sqrt{R_u}\,\mathrm{d}W_u,$$

which implies that $\mathbb{E}\big[e^{\beta t}R_t\big] = R_0 + \frac{\alpha}{\beta}(e^{\beta t}-1)$ (as the Itô integral $\int_0^t e^{\beta u}\sqrt{R_u}\,\mathrm{d}W_u$ has zero mean). Thus,

$$\mathbb{E}[R_t] = e^{-\beta t}R_0 + \frac{\alpha}{\beta}(1-e^{-\beta t}).$$

To derive the variance, we let $X_t = e^{\beta t}R_t$. Then, we have

- $\mathrm{d}X_t = \alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t}\sqrt{R_t}\,\mathrm{d}W_t = \alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t/2}\sqrt{X_t}\,\mathrm{d}W_t$.

- $\mathbb{E}[X_t] = R_0 + (\alpha/\beta)(e^{\beta t} - 1)$.

Now, applying the Itô formula for Itô process $\{X_t\}$ with $f(t,x) = x^2$ gives

$$\begin{aligned}
\mathrm{d}X_t^2 &= \underbrace{f_t(t, X_t)}_{0}\,\mathrm{d}t + \underbrace{f_x(t, X_t)}_{2x|_{x=X_t}=2X_t}\,\mathrm{d}X_t + \frac{1}{2}\underbrace{f_{xx}(t, X_t)}_{2}\,\mathrm{d}X_t\,\mathrm{d}X_t \\
&= 2X_t\,\mathrm{d}X_t + (\sigma e^{\beta t}\sqrt{R_t})^2\,\mathrm{d}t = 2X_t\,\mathrm{d}X_t + \sigma^2 e^{2\beta t} R_t\,\mathrm{d}t = 2X_t\,\mathrm{d}X_t + \sigma^2 e^{\beta t} X_t\,\mathrm{d}t \\
&= 2X_t(\alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t}\sqrt{R_t}\,\mathrm{d}W_t) + \sigma^2 e^{\beta t} X_t\,\mathrm{d}t = 2X_t(\alpha e^{\beta t}\,\mathrm{d}t + \sigma e^{\beta t/2}\sqrt{X_t}\,\mathrm{d}W_t) + \sigma^2 e^{\beta t} X_t\,\mathrm{d}t \\
&= (2\alpha + \sigma^2)e^{\beta t} X_t\,\mathrm{d}t + 2\sigma e^{\beta t/2} X_t^{3/2}\,\mathrm{d}W_t.
\end{aligned}$$

Expressing this in integral form, we get

$$X_t^2 = X_0^2 + (2\alpha + \sigma^2)\int_0^t e^{\beta u} X_u\,\mathrm{d}u + 2\sigma\int_0^t e^{\beta u/2} X_u^{3/2}\,\mathrm{d}W_u.$$

Since the mean of Itô integral is zero, taking expectation then yields

$$\begin{aligned}
\mathbb{E}\big[X_t^2\big] &= X_0^2 + (2\alpha + \sigma^2)\mathbb{E}\left[\int_0^t e^{\beta u} X_u\,\mathrm{d}u\right] \overset{\text{(Tonelli)}}{=} X_0^2 + (2\alpha + \sigma^2)\int_0^t e^{\beta u}\mathbb{E}[X_u]\,\mathrm{d}u \\
&= X_0^2 + (2\alpha + \sigma^2)\int_0^t e^{\beta u}\left(R_0 + \frac{\alpha}{\beta}(e^{\beta u} - 1)\right)\mathrm{d}u \\
&= X_0^2 + (2\alpha + \sigma^2)\int_0^t \left(R_0 - \frac{\alpha}{\beta}\right)e^{\beta u} + \frac{\alpha}{\beta}e^{2\beta u}\,\mathrm{d}u \\
&= R_0^2 + \frac{2\alpha + \sigma^2}{\beta}\left(R_0 - \frac{\alpha}{\beta}\right)(e^{\beta t} - 1) + \frac{2\alpha + \sigma^2}{2\beta}\frac{\alpha}{\beta}(e^{2\beta t} - 1).
\end{aligned}$$

Hence,

$$\mathbb{E}\big[R_t^2\big] \overset{(X_t^2 = e^{2\beta t} R_t^2)}{=} e^{-2\beta t}\mathbb{E}\big[X_t^2\big] = e^{-2\beta t} R_0^2 + \frac{2\alpha + \sigma^2}{\beta}\left(R_0 - \frac{\alpha}{\beta}\right)(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha(2\alpha + \sigma^2)}{2\beta^2}(1 - e^{-2\beta t}).$$

Thus,

$$\begin{aligned}
\mathrm{Var}(R_t) &= \mathbb{E}\big[R_t^2\big] - \mathbb{E}[R_t]^2 \\
&= e^{-2\beta t} R_0^2 + \frac{2\alpha + \sigma^2}{\beta}\left(R_0 - \frac{\alpha}{\beta}\right)(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha(2\alpha + \sigma^2)}{2\beta^2}(1 - e^{-2\beta t}) - \left(e^{-\beta t} R_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t})\right)^2 \\
&= \frac{\sigma^2}{\beta} R_0(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta^2}(1 - 2e^{-\beta t} + e^{-2\beta t}).
\end{aligned}$$

Particularly, we have $\lim_{t\to\infty}\mathrm{Var}(R_t) = \alpha\sigma^2/(2\beta^2)$.

4.3 Black-Scholes Equation

4.3.1 In STAT3905/STAT3910, we have studied option pricing under the Black-Scholes model, via the *Black-Scholes formula*. Here, we will explore the mathematical details underlying the Black-Scholes model and derive an important partial differential equation that describes the dynamics of option prices under the Black-Scholes model. It can be derived based on a replication argument.

4.3.2 **Black-Scholes model.** Under the **Black-Scholes model**, we assume that the market is arbitrage-free, with a stock and a (risk-free) bond which can be freely bought or (short) sold in any amount without transaction cost. We suppose that the bond earns a continuously compounded risk-free rate r, and the stock price process $\{S_t\}$ follows a *geometric Brownian motion*:

$$\mathrm{d}S_t = \alpha S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$$

where $\alpha \in \mathbb{R}$ and $\sigma > 0$ are constants.

4.3.3 **Constructing a self-financing portfolio.** In the derivation of Black-Scholes equation, a *self-financing* portfolio is utilized for replicating the option payoffs. So, we will first develop such portfolio here as a preparation. At each time t, let X_t denote the investor's portfolio value and Δ_t denote the number of shares of stock held by the investor. Here, we suppose that $\{\Delta_t\}$ is adapted to a filtration $\{\mathcal{F}_t\}$ for the Brownian motion $\{W_t\}$, and at each time t, the non-stock component of the portfolio, with value $X_t - \Delta_t S_t$, is all invested in the bond.

The **self-financing** property of the portfolio then suggests that the "change" $\mathrm{d}X_t$ for the investor's portfolio at time t would source from the following two factors: (i) the gain $\Delta_t \,\mathrm{d}S_t$ on the stock position and (ii) the interest earning $r(X_t - \Delta_t S_t)\,\mathrm{d}t$ on the bond position. Hence, we can write

$$\begin{aligned}\mathrm{d}X_t &= \Delta_t \,\mathrm{d}S_t + r(X_t - \Delta_t S_t)\,\mathrm{d}t = \Delta_t(\alpha S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t) + r(X_t - \Delta_t S_t)\,\mathrm{d}t \\ &= rX_t \,\mathrm{d}t + (\alpha - r)\Delta_t S_t \,\mathrm{d}t + \sigma \Delta_t S_t \,\mathrm{d}W_t.\end{aligned}$$

With the final expression, we can identify three sources for the changes in portfolio values, namely: (i) an average underlying rate of return r on the portfolio ($rX_t \,\mathrm{d}t$), (ii) a risk premium $\alpha - r$ for investing in the (risky) stock ($(\alpha - r)\Delta_t S_t \,\mathrm{d}t$), and (iii) a volatility term proportional to the size of stock investment ($\sigma \Delta_t S_t \,\mathrm{d}W_t$).

For the purpose of pricing, we often need to consider the *discounted* stock price $\{e^{-rt}S_t\}$ and the *discounted* portfolio value $\{e^{-rt}X_t\}$. Hence, we are interested in deriving SDEs that involve these discounted processes. This can be done by using the Itô formula (for Itô process) with $f(t,x) = e^{-rt}x$:

- *Discounted stock price:*

$$\begin{aligned}\mathrm{d}(e^{-rt}S_t) = \mathrm{d}f(t\,\mathrm{d}S_t) &= \underbrace{f_t(t,S_t)}_{-re^{-rt}S_t}\mathrm{d}t + \underbrace{f_x(t,S_t)}_{e^{-rt}}\underbrace{\mathrm{d}S_t}_{\alpha S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t} + \frac{1}{2}\underbrace{f_{xx}(t,S_t)}_{0}\mathrm{d}S_t \,\mathrm{d}S_t \\ &= (\alpha - r)e^{-rt}S_t \,\mathrm{d}t + \sigma e^{-rt}S_t \,\mathrm{d}W_t.\end{aligned}$$

- *Discounted portfolio value:*

$$\begin{aligned}\mathrm{d}(e^{-rt}X_t) = \mathrm{d}f(t,X_t) &= \underbrace{f_t(t,X_t)}_{-re^{-rt}X_t}\mathrm{d}t + \underbrace{f_x(t,X_t)}_{e^{-rt}}\underbrace{\mathrm{d}X_t}_{rX_t \,\mathrm{d}t + (\alpha-r)\Delta_t S_t \,\mathrm{d}t + \sigma\Delta_t S_t \,\mathrm{d}W_t} + \frac{1}{2}\underbrace{f_{xx}(t,X_t)}_{0}\mathrm{d}X_t \,\mathrm{d}X_t \\ &= \Delta_t(\alpha - r)e^{-rt}S_t \,\mathrm{d}t + \Delta_t \sigma e^{-rt}S_t \,\mathrm{d}W_t = \Delta_t \,\mathrm{d}(e^{-rt}S_t).\end{aligned}$$

4.3.4 **Deriving the Black-Scholes equation through a replication argument.** One way to derive the famous *Black-Scholes equation* is to utilize a replication argument. The basic idea is that, if the European call (or more generally, derivative whose payoff depends only on the terminal stock price) in consideration can be replicated by a self-financing portfolio (more on this condition will be discussed in [5.2.2]), then the call prices must obey a certain dynamic described by the *Black-Scholes equation*, and the solution to that equation gives us the call price formula (known as *Black-Scholes formula*).

Theorem 4.3.a (Black-Scholes equation)**.** Let $V(t,x) = V(t,S_t)$ denote the time-t value of a T-year derivative, with the time-t stock price being S_t, for all $0 \le t \le T$. Suppose that the payoff of the derivative depends only on the terminal stock price S_T, the function $V(t,S_t)$ is of class C^2, and the derivative can be **replicated** (or **hedged**), i.e., there is a self-financing portfolio, having a price process $\{X_t\}$, such that $X_T = V(T,S_T)$ (known as a **replicating portfolio** or **hedging portfolio**). Under the Black-Scholes model, the function $V(t,S_t)$ satisfies the following *Black-Scholes partial differential equation (PDE)*:

$$V_t(t,S_t) + rS_t V_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t) = rV(t,S_t).$$

Proof. By the Itô formula for the Itô process $\{S_t\}$, we have

$$\begin{aligned}\mathrm{d}V(t,S_t) &= V_t(t,S_t)\,\mathrm{d}t + V_x(t,S_t)\underbrace{\mathrm{d}S_t}_{\alpha S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t} + \frac{1}{2}V_{xx}(t,S_t)(\sigma S_t)^2 \,\mathrm{d}t \\ &= \left[V_t(t,S_t) + \alpha S_t V_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t)\right]\mathrm{d}t + \sigma S_t V_x(t,S_t)\,\mathrm{d}W_t.\end{aligned}$$

Next, applying the Itô formula for the Itô process $\{V(t,x)\}_{t\in[0,T]}$ with $f(t,x)=e^{-rt}x$, we get

$$\begin{aligned}
\mathrm{d}(e^{-rt}V(t,S_t)) &= \underbrace{f_t(t,V(t,S_t))}_{-re^{-rt}V(t,S_t)}\,\mathrm{d}t + \underbrace{f_x(t,V(t,S_t))}_{e^{-rt}}\,\mathrm{d}V(t,S_t) + \frac{1}{2}\underbrace{f_{xx}(t,V(t,S_t))}_{0}\,\mathrm{d}V(t,S_t)\,\mathrm{d}V(t,S_t) \\
&= e^{-rt}\left[-rV(t,S_t)+V_t(t,S_t)+\alpha S_t V_x(t,S_t)+\frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t)\right]\mathrm{d}t + e^{-rt}\sigma S_t V_x(t,S_t)\,\mathrm{d}W_t.
\end{aligned}$$

On the other hand, we have the following SDE for discounted portfolio value from [4.3.3]:

$$\mathrm{d}(e^{-rt}X_t)=\Delta_t(\alpha-r)e^{-rt}S_t\,\mathrm{d}t+\Delta_t\sigma e^{-rt}S_t\,\mathrm{d}W_t.$$

From the replication nature of the self-financing portfolio, we have $X_T=V(T,S_T)$. Since the portfolio is self-financing and there is no arbitrage, this implies that $X_t=V(t,S_t)$ for all $0\le t<T$ also. Therefore, we must have $\mathrm{d}(e^{-rt}V(t,S_t))=\mathrm{d}(e^{-rt}X_t)$, i.e.,

$$\begin{aligned}
&\Delta_t(\alpha-r)e^{-rt}S_t\,\mathrm{d}t+\Delta_t\sigma e^{-rt}S_t\,\mathrm{d}W_t \\
&= e^{-rt}\left[-rV(t,S_t)+V_t(t,S_t)+\alpha S_t V_x(t,S_t)+\frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t)\right]\mathrm{d}t + e^{-rt}\sigma S_t V_x(t,S_t)\,\mathrm{d}W_t.
\end{aligned}$$

This implies that the drift and volatility terms from both sides are equal. Equating the volatility terms gives $\Delta_t=V_x(t,S_t)$, which is known as the **delta-hedging rule** (recall that the partial derivative $V_x(t,S_t)$ is known as the **delta** of the call). Next, equating the drift terms with $\Delta_t=V_x(t,S_t)$ yields

$$\begin{aligned}
V_x(t,S_t)(\alpha-r)S_t &= -rV(t,S_t)+V_t(t,S_t)+\alpha S_t V_x(t,S_t)+\frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t) \\
\implies rV(t,S_t) &= V_t(t,S_t)+rS_tV_x(t,S_t)+\frac{1}{2}\sigma^2 S_t^2 V_{xx}(t,S_t),
\end{aligned}$$

which gives us the Black-Scholes partial differential equation. □

4.3.5 **Solving the Black-Scholes equation.** Suppose that the derivative in consideration is a $T-$-year K-strike European call. After solving the Black-Scholes equation with the *terminal condition* $V(T,S_T)=(S_T-K)_+$ and the *boundary conditions* $V(t,0)=0\ \forall t\in[0,T]$ and $\lim_{x\to\infty}[V(t,x)-(x-Ke^{-r(T-t)})]=0\ \forall t\in[0,T]$[11], we can get the *Black-Scholes formula* for European call, namely

$$V(t,S_t)=S_t\Phi(d_+(T-t,S_t))-Ke^{-r(T-t)}\Phi(d_-(T-t,S_t))$$

for every $t\in[0,T)$, where

$$d_\pm(T-t,S_t)=\frac{1}{\sigma\sqrt{T-t}}\left[\ln\frac{S_t}{K}+\left(r\pm\frac{\sigma^2}{2}\right)(T-t)\right],$$

and Φ is the standard normal distribution function. [Note: The function $c(t,x)$ here is indeed of class C^2.]

4.4 Multivariable Stochastic Calculus

4.4.1 Previously, we have studied stochastic calculus with the randomness sourcing from a *single* Brownian motion. However, in practice there are often multiple sources of randomness driving the dynamics of the stochastic process in consideration. In view of this, the *univariate* stochastic calculus (involving only a single Brownian motion) discussed before would not be enough, and *multivariable* stochastic calculus is needed. Its development starts from the concept of higher-dimensional Brownian motion.

[11] Intuitively, as the stock price goes to infinity, the call would almost certainly expire in-the-money, and hence would worth almost the same as the time-t value of a (synthetic) long forward on a stock deliverable at the price K and time T, namely $S_t-Ke^{-r(T-t)}=x-Ke^{-r(T-t)}$.

4.4.2 **d-dimensional Brownian motion.** A **d-dimensional Brownian motion** is a stochastic process $\{\boldsymbol{W}_t\}$ given by $\boldsymbol{W}_t = (W_t^{(1)}, \dots, W_t^{(d)})$, satisfying:

(1) For every $i = 1, \dots, d$, the process $\{W_t^{(i)}\}$ is a Brownian motion.

(2) For all $i \neq j$, the processes $\{W_t^{(i)}\}$ and $\{W_t^{(j)}\}$ are independent.[12]

(3) There is a filtration $\{\mathcal{F}_t\}$ such that:

i. *(Adaptivity)* $\boldsymbol{W}_t$ is $\mathcal{F}_t$-measurable for all t.[13]

ii. *(Independence of future increments)* For all $t < u$, $\boldsymbol{W}_u - \boldsymbol{W}_t$ is independent of $\{\mathcal{F}_t\}$.[14]

[Note: Such $\{\mathcal{F}_t\}$ is sometimes called a **filtration for the d-dimensional Brownian motion $\{\boldsymbol{W}_t\}$**.]

4.4.3 **Zero cross-variation for d-dimensional Brownian motion.** By Proposition 3.3.b, we know that the quadratic variation of an individual Brownian motion $\{W_t^{(i)}\}$ is given by

$$[W^{(i)}, W^{(i)}]_T = \lim_{\|\Pi\| \to 0} \sum_{k=0}^{n-1} (W_{t_{k+1},i} - W_{t_k,i})^2 = T$$

for all $T > 0$. With the presence of multiple Brownian motions here, we would also like to investigate the **cross variation** of the two processes $\{W_t^{(i)}\}$ and $\{W_t^{(j)}\}$, given by

$$[W^{(i)}, W^{(j)}]_T = \lim_{\|\Pi\| \to 0} \sum_{k=0}^{n-1} (W_{t_{k+1},i} - W_{t_j,i})(W_{t_{k+1},j} - W_{t_k,j})$$

for all $T > 0$. Intuitively, due to the independence of $\{W_t^{(i)}\}$ and $\{W_t^{(j)}\}$, one may expect such cross variation to be zero always (somewhat "similar" to the result that independence implies uncorrelatedness). This is indeed the case, and we will prove it in the following.

Proposition 4.4.a. Let $\{\boldsymbol{W}_t\}$ be a d-dimensional Brownian motion, with $\boldsymbol{W}_t = (W_t^{(1)}, \dots, W_{t,d})$ for all t. Fix any $i \neq j$ and let W_i and W_j denote the functions $t \mapsto W_t^{(i)}$ and $t \mapsto W_t^{(j)}$ respectively. Then, we have $[W_i, W_j]_T \overset{\text{a.s.}}{=} 0$ for all $T > 0$.

Proof. Let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0, T]$ and write $C_\Pi := \sum_{k=0}^{n-1} (W_{t_{k+1},i} - W_{t_k,i})(W_{t_{k+1},j} - W_{t_k,j})$. Then,

$$\begin{aligned}
\mathbb{E}[C_\Pi] &= \sum_{k=0}^{n-1} \mathbb{E}\big[(W_{t_{k+1},i} - W_{t_k,i})(W_{t_{k+1},j} - W_{t_k,j})\big] \\
&\overset{\text{(independence)}}{=} \sum_{k=0}^{n-1} \mathbb{E}\big[W_{t_{k+1},i} - W_{t_k,i}\big]\mathbb{E}\big[W_{t_{k+1},j} - W_{t_k,j}\big] = 0.
\end{aligned}$$

Next, consider

$$\begin{aligned}
C_\Pi^2 &= \sum_{k=0}^{n-1} (W_{t_{k+1},i} - W_{t_k,i})^2 (W_{t_{k+1},j} - W_{t_k,j})^2 \\
&\quad + 2 \sum_{0 \le \ell < k \le n-1} \big[(W_{t_{\ell+1},i} - W_{t_\ell,i})(W_{t_{\ell+1},j} - W_{t_\ell,j})\big]\big[(W_{t_{k+1},i} - W_{t_k,i})(W_{t_{k+1},j} - W_{t_k,j})\big].
\end{aligned}$$

[12]This independence means that for all $n \in \mathbb{N}$ and $t_1, \dots, t_n$, the random vectors $\boldsymbol{W}^{(i)} = (W_{t_1}^{(i)}, \dots, W_{t_n}^{(i)})$ and $\boldsymbol{W}^{(j)} = (W_{t_1}^{(j)}, \dots, W_{t_n}^{(j)})$ are independent, which in turn means that the σ-algebras generated by them, $(\boldsymbol{W}^{(i)})^{-1}(\mathcal{B}(\mathbb{R}^n))$ and $(\boldsymbol{W}^{(j)})^{-1}(\mathcal{B}(\mathbb{R}^n))$, are independent.

[13]This means that the σ-algebra generated by $\boldsymbol{W}_t$, namely $\boldsymbol{W}_t^{-1}(\mathcal{B}(\mathbb{R}^d))$, is a subset of $\mathcal{F}_t$, for all t.

[14]This means that the σ-algebra generated by the increment $\boldsymbol{W}_u - \boldsymbol{W}_t$, which is $(\boldsymbol{W}_u - \boldsymbol{W}_t)^{-1}(\mathcal{B}(\mathbb{R}^d))$, and $\{\mathcal{F}_t\}$ are independent.

For all $0 \le \ell < k \le n+1$, we have

$$\begin{aligned}
&\mathbb{E}\big[(W_{t_{\ell+1},i} - W_{t_\ell,i})(W_{t_{\ell+1},j} - W_{t_\ell,j})(W_{t_{k+1},i} - W_{t_k,i})(W_{t_{k+1},j} - W_{t_k,j})\big] \\
&\overset{\text{(independent Brownian motions)}}{=} \mathbb{E}\big[(W_{t_{\ell+1},i} - W_{t_\ell,i})(W_{t_{k+1},i} - W_{t_k,i})\big]\mathbb{E}\big[(W_{t_{\ell+1},j} - W_{t_\ell,j})(W_{t_{k+1},j} - W_{t_k,j})\big] \\
&\overset{\text{(independent increments)}}{=} \mathbb{E}\big[W_{t_{\ell+1},i} - W_{t_\ell,i}\big]\mathbb{E}\big[W_{t_{k+1},i} - W_{t_k,i}\big]\mathbb{E}\big[W_{t_{\ell+1},j} - W_{t_\ell,j}\big]\mathbb{E}\big[W_{t_{k+1},j} - W_{t_k,j}\big] = 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(C_\Pi) = \mathbb{E}\big[C_\Pi^2\big] &= \mathbb{E}\left[\sum_{k=0}^{n-1}(W_{t_{k+1},i} - W_{t_k,i})^2(W_{t_{k+1},j} - W_{t_k,j})^2\right] \\
&= \sum_{k=0}^{n-1}\mathbb{E}\big[(W_{t_{k+1},i} - W_{t_k,i})^2(W_{t_{k+1},j} - W_{t_k,j})^2\big] \\
&\overset{\text{(independence)}}{=} \sum_{k=0}^{n-1}\mathbb{E}\big[(W_{t_{k+1},i} - W_{t_k,i})^2\big]\mathbb{E}\big[(W_{t_{k+1},j} - W_{t_k,j})^2\big] \\
&= \sum_{k=0}^{n-1}(t_{k+1} - t_k)^2 \le \|\Pi\|\sum_{k=0}^{n-1}(t_{k+1} - t_k) = \|\Pi\|T,
\end{aligned}$$

and so $\lim_{\|\Pi\|\to 0}\mathrm{Var}(C_\Pi) = 0$. Like the proof of Proposition 3.3.b, the result then follows. □

Differential rule: Informally, we can express the result in Proposition 4.4.a as $\mathrm{d}W_t^{(i)}\,\mathrm{d}W_t^{(j)} = 0$, which gives rise to another differential rule. This differential rule is often helpful when dealing with differentials about higher-dimensional Brownian motions.

4.4.4 **Itô processes with two Brownian motions.** With the higher-dimensional Brownian motion, we can generalize the previous notion of Itô process to incorporate more sources of randomness. To keep things simple, here we shall only focus on the case with two-dimensional Brownian motions.

Let $\{\boldsymbol{W}_t\}$ be a two-dimensional Brownian motion with $\boldsymbol{W}_t = (W_t^{(1)}, W_t^{(2)})$ for all t, and $\{\mathcal{F}_t\}$ be a filtration for $\{\boldsymbol{W}_t\}$. A **(two-dimensional) Itô process** is a $\{\mathcal{F}_t\}$-adapted process $\{X_t\}$ given by

$$X_t = X_0 + \int_0^t \Theta_u\,\mathrm{d}u + \int_0^t \sigma_u^{(1)}\,\mathrm{d}W_u^{(1)} + \int_0^t \sigma_u^{(2)}\,\mathrm{d}W_u^{(2)}$$

where $\{\Theta_u\}$, $\{\sigma_u^{(1)}\}$, and $\{\sigma_u^{(2)}\}$ are adapted to $\{\mathcal{F}_t\}$.

Differential form: $\mathrm{d}X_t = \Theta_t\,\mathrm{d}t + \sigma_u^{(1)}\,\mathrm{d}W_t^{(1)} + \sigma_u^{(2)}\,\mathrm{d}W_t^{(2)}$.

[Note: By taking $\sigma_u^{(1)} \equiv 0$ or $\sigma_u^{(2)} \equiv 0$, the two-dimensional Itô process would be reduced to the Itô process before.]

4.4.5 **A pair of two-dimensional Itô processes.** In the two-dimensional case of multivariable stochastic calculus, often we are dealing with two Itô processes that are two-dimensional each: $\{X_t\}$ and $\{Y_t\}$ with differential form

$$\begin{aligned}
\mathrm{d}X_t &= \Theta_t^{(1)}\,\mathrm{d}t + \sigma_u^{(11)}\,\mathrm{d}W_t^{(1)} + \sigma_u^{(12)}\,\mathrm{d}W_t^{(2)}, \\
\mathrm{d}Y_t &= \Theta_t^{(2)}\,\mathrm{d}t + \sigma_u^{(21)}\,\mathrm{d}W_t^{(1)} + \sigma_u^{(22)}\,\mathrm{d}W_t^{(2)}.
\end{aligned}$$

For example, $\{X_t\}$ and $\{Y_t\}$ may represent two stock price processes and the Brownian motions $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ may correspond to the "risk factors" underlying the two stocks.

With these SDEs, we can derive numerous properties informally by applying differential rules, such as:

$$\begin{aligned}
\mathrm{d}X_t\,\mathrm{d}X_t &= ((\sigma_t^{(11)})^2 + (\sigma_t^{(12)})^2)\,\mathrm{d}t, \\
\mathrm{d}Y_t\,\mathrm{d}Y_t &= ((\sigma_t^{(21)})^2 + (\sigma_t^{(22)})^2)\,\mathrm{d}t, \\
\mathrm{d}X_t\,\mathrm{d}Y_t &= (\sigma_t^{(11)}\sigma_t^{(21)} + \sigma_t^{(12)}\sigma_t^{(22)})\,\mathrm{d}t.
\end{aligned}$$

These can be more precisely expressed as:

$$[X, X]_T = \int_0^T ((\sigma_t^{(11)})^2 + (\sigma_t^{(12)})^2)\,\mathrm{d}t,$$
$$[Y, Y]_T = \int_0^T ((\sigma_t^{(21)})^2 + (\sigma_t^{(22)})^2)\,\mathrm{d}t,$$
$$[X, Y]_T = \int_0^T (\sigma_t^{(11)}\sigma_t^{(21)} + \sigma_t^{(12)}\sigma_t^{(22)})\,\mathrm{d}t.$$

Their proofs are somewhat similar to the one for Proposition 4.2.b, so we will omit them here.

4.4.6 **Two-dimensional Itô formula.** The Itô formula for an Itô process studied in Theorem 4.2.c can be further generalized to the one for a pair of two-dimensional Itô processes as follows.

Theorem 4.4.b (Two-dimensional Itô formula)**.** Let $\{X_t\}$ and $\{Y_t\}$ be a pair of two-dimensional Itô processes, and let $f(t, x, y)$ be a function of class C^2. Then, for every $T \geq 0$, we have

$$\begin{aligned} f(T, X_T, Y_T) &= f(0, X_0, Y_0) + \int_0^T f_t(t, X_t, Y_t)\,\mathrm{d}t + \int_0^T f_x(t, X_t, Y_t)\,\mathrm{d}X_t + \int_0^T f_x(t, X_t, Y_t)\,\mathrm{d}Y_t \\ &\quad + \frac{1}{2}\int_0^T f_{xx}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}X_t + \int_0^T f_{xy}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}Y_t + \frac{1}{2}\int_0^T f_{xx}(t, X_t, Y_t)\,\mathrm{d}Y_t\,\mathrm{d}Y_t \end{aligned}$$

where the "product" of differentials appearing here are to be interpreted as in the SDEs from [4.4.5], e.g., "$\int_0^T f_{xy}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}Y_t$" should be interpreted as "$\int_0^T f_{xy}(t, X_t, Y_t)((\sigma_t^{(21)})^2 + (\sigma_t^{(22)})^2)\,\mathrm{d}t$".

Differential form:

$$\begin{aligned} \mathrm{d}f(t, X_t, Y_t) &= f_t(t, X_t, Y_t)\,\mathrm{d}t + f_x(t, X_t, Y_t)\,\mathrm{d}X_t + f_y(t, X_t, Y_t)\,\mathrm{d}Y_t \\ &\quad + \frac{1}{2} f_{xx}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}X_t + f_{xy}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}Y_t + \frac{1}{2} f_{yy}(t, X_t, Y_t)\,\mathrm{d}Y_t\,\mathrm{d}Y_t. \end{aligned}$$

It can be informally derived as follows, like the Itô formula for Itô process:

$$\begin{aligned} \mathrm{d}f(t, X_t, Y_t) &= f_t(t, X_t, Y_t)\,\mathrm{d}t + f_x(t, X_t, Y_t)\,\mathrm{d}X_t + f_y(t, X_t, Y_t)\,\mathrm{d}Y_t \\ &\quad + \frac{1}{2} f_{xx}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}X_t + f_{xy}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}Y_t + \frac{1}{2} f_{yy}(t, X_t, Y_t)\,\mathrm{d}Y_t\,\mathrm{d}Y_t \\ &\quad + \frac{1}{2} f_{tt}(t, X_t, Y_t)\underbrace{\mathrm{d}t\,\mathrm{d}t}_{0} + f_{xt}(t, X_t, Y_t)\underbrace{\mathrm{d}X_t\,\mathrm{d}t}_{0} + f_{yt}(t, X_t, Y_t)\underbrace{\mathrm{d}Y_t\,\mathrm{d}t}_{0} + \underbrace{\text{higher order term}}_{0} \\ &= f_t(t, X_t, Y_t)\,\mathrm{d}t + f_x(t, X_t, Y_t)\,\mathrm{d}X_t + f_y(t, X_t, Y_t)\,\mathrm{d}Y_t \\ &\quad + \frac{1}{2} f_{xx}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}X_t + f_{xy}(t, X_t, Y_t)\,\mathrm{d}X_t\,\mathrm{d}Y_t + \frac{1}{2} f_{yy}(t, X_t, Y_t)\,\mathrm{d}Y_t\,\mathrm{d}Y_t. \end{aligned}$$

Proof. Omitted. □

A corollary of the Itô formula here is the *Itô product rule*, which is a stochastic version of the product rule for calculus. Just like how Itô formula has extra terms compared with the ordinary chain rule in calculus, the Itô product rule also carries an extra term compared with the ordinary product rule in calculus.

Corollary 4.4.c (Itô product rule)**.** Let $\{X_t\}$ and $\{Y_t\}$ be a pair of two-dimensional Itô processes. Then, $\mathrm{d}(X_t Y_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \mathrm{d}X_t\,\mathrm{d}Y_t$.

Proof. Take $f(t, x, y) = xy$ in Theorem 4.4.b (with $f_t = 0$, $f_x = y$, $f_y = x$, $f_{xx} = 0$, $f_{xy} = 1$, and $f_{yy} = 0$). □

4.4.7 **Lévy's characterization theorem.** In stochastic calculus, we need to work with Brownian motions frequently, and so the first step is often to identify whether the processes in consideration are Brownian motions. In Proposition 3.2.a, we have provided several characterizations of Brownian motion. Here, we will study one more characterization, given by the *Lévy's characterization theorem*.

Theorem 4.4.d (Lévy's characterization theorem)**.** Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration and $\{M_t\}_{t\geq 0}$ be a $\{\mathcal{F}_t\}$-martingale. The process $\{M_t\}$ is a Brownian motion iff

(1) *(Starting at zero)* $M_0 = 0$.

(2) *(Continuous paths)* For every fixed $\omega \in \Omega$, $M_t(\omega)$ is continuous in t.

(3) *(Unit quadratic variations)* $[M, M]_t = t$ for all $t \geq 0$.

Proof. (Sketch) The "$\Rightarrow$" direction follows by the definition of Brownian motion and Proposition 3.3.b. So it remains to prove the "$\Leftarrow$" direction. By the definition of Brownian motion, it suffices to prove that $M_t \sim \mathrm{N}(0, t)$ for all $t \geq 0$ with the three conditions assumed.

While $\{M_t\}$ is just a martingale (we have not yet shown that it is a Brownian motion for this direction), we can indeed apply the Itô formula for Brownian motion from Theorem 4.2.a with $\{W_t\}$ replaced by $\{M_t\}$ here. To see why, observe that the proof of Theorem 4.2.a only utilizes the properties of continuous path and accumulation of quadratic variation at unit rate, which are satisfied by $\{M_t\}$ by assumption. Furthermore, one can define the "Itô integral with respect to the martingale $\{M_t\}$" in an analogous fashion as the one for Brownian motion. Then, it can be shown that such Itô integral with respect to the martingale $\{M_t\}$, $\int_0^t \Delta_u \, \mathrm{d}M_u$, still forms a martingale.[15]

Now, let $f(t, x) = e^{ux - u^2 t/2}$ and fix any $t \geq 0$. Applying the Itô formula gives

$$f(t, M_t) = \underbrace{f(0, M_0)}_{1} + \int_0^t \underbrace{f_t(u, M_u)}_{-\frac{u^2}{2} e^{uM_u - u^2 t/2}} + \frac{1}{2} \underbrace{f_{xx}(u, M_u)}_{u^2 e^{uM_u - u^2 t/2}} \, \mathrm{d}u + \int_0^t f_x(u, M_u) \, \mathrm{d}M_u = 1 + \int_0^t f_x(u, M_u) \, \mathrm{d}M_u.$$

From this, we know that $\{f(t, M_t)\}$ is a martingale. Hence, we have $\mathbb{E}\left[e^{uM_t - u^2 t/2}\right] = \mathbb{E}[f(t, M_t)] = \mathbb{E}[\mathbb{E}[f(t, M_t)]|\mathcal{F}_0] = \mathbb{E}[f(0, M_0)] = 1$. This implies that the moment generating function of M_t is $\mathbb{E}\left[e^{uM_t}\right] = e^{u^2 t/2}$, and so $M_t \sim \mathrm{N}(0, t)$. $\square$

4.4.8 **Two-dimensional Lévy's characterization theorem.** When working with two-dimensional stochastic calculus, we need to deal with two-dimensional Brownian motions that are formed by pairs of independent Brownian motions. It turns out that there is a two-dimensional version of the Lévy's characterization theorem for providing a characterization of such independent Brownian motions.

Theorem 4.4.e (Two-dimensional Lévy's characterization theorem)**.** Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration, and $\{M_t^{(1)}\}_{t\geq 0}$ and $\{M_t^{(2)}\}_{t\geq 0}$ be $\{\mathcal{F}_t\}$-martingales. The processes $\{M_t^{(1)}\}$ and $\{M_t^{(2)}\}$ are *independent* Brownian motions iff

(1) *(Starting at zero)* $M_0^{(1)} = M_0^{(2)} = 0$.

(2) *(Continuous paths)* For every fixed $\omega \in \Omega$, the functions $t \mapsto M_t^{(1)}(\omega)$ and $t \mapsto M_t^{(2)}(\omega)$ are continuous.

(3) *(Unit quadratic variations and zero cross variations)* $[M^{(1)}, M^{(1)}]_t = [M^{(2)}, M^{(2)}]_t = t$ and $[M^{(1)}, M^{(2)}]_t = 0$ for all $t \geq 0$.

Proof. (Sketch) "$\Rightarrow$": Like before, it follows by definition, Proposition 3.3.b, and Proposition 4.4.a (also applicable for independent Brownian motions).

[15] The rough idea is to first note that it is a martingale for simple integrand Δ_t (by Proposition 2.3.a), and then show that taking limit preserves the martingale properties.

"$\Leftarrow$": Under such assumptions, by Theorem 4.4.d we know that $\{M_t^{(1)}\}$ and $\{M_t^{(2)}\}$ are Brownian motions individually. It then suffices to show that they are independent. We establish this by considering the moment generating function. Fix any $u_1, u_2 \in \mathbb{R}$, and any $t \geq 0$. By the two-dimensional Itô formula (Theorem 4.4.b) with $f(t,x,y) = e^{u_1 x + u_2 y - (u_1^2+u_2^2)t/2}$, we have

$$\begin{aligned}
\mathrm{d}f(t, M_t^{(1)}, M_t^{(2)}) &= f_t(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}t + f_x(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(1)} + f_y(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(2)} \\
&\quad + \frac{1}{2} f_{xx}(t, M_t^{(1)}, M_t^{(2)}) \underbrace{\mathrm{d}M_t^{(1)}\,\mathrm{d}M_t^{(1)}}_{\mathrm{d}t\text{ by assumption}} + f_{xy}(t, M_t^{(1)}, M_t^{(2)}) \underbrace{\mathrm{d}M_t^{(1)}\,\mathrm{d}M_t^{(2)}}_{0\text{ by assumption}} \\
&\quad + \frac{1}{2} f_{yy}(t, M_t^{(1)}, M_t^{(2)}) \underbrace{\mathrm{d}M_t^{(2)}\,\mathrm{d}M_t^{(2)}}_{\mathrm{d}t\text{ by assumption}} \\
&= \Bigg(\underbrace{f_t(t, M_t^{(1)}, M_t^{(2)})}_{-\frac{(u_1^2+u_2^2)}{2} f(t, M_t^{(1)}, M_t^{(2)})} + \frac{1}{2} \underbrace{f_{xx}(t, M_t^{(1)}, M_t^{(2)})}_{u_1^2 f(t, M_t^{(1)}, M_t^{(2)})} + \frac{1}{2} \underbrace{f_{yy}(t, M_t^{(1)}, M_t^{(2)})}_{u_2^2 f(t, M_t^{(1)}, M_t^{(2)})} \Bigg)\,\mathrm{d}t \\
&\quad + f_x(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(1)} + f_y(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(2)} \\
&= f_x(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(1)} + f_y(t, M_t^{(1)}, M_t^{(2)})\,\mathrm{d}M_t^{(2)}.
\end{aligned}$$

Hence,

$$f(t, M_t^{(1)}, M_t^{(2)}) = \underbrace{f(0, M_0^{(1)}, M_0^{(2)})}_{1} + \int_0^t f_x(u, M_u^{(1)}, M_u^{(2)})\,\mathrm{d}M_u^{(1)} + \int_0^t f_y(u, M_u^{(1)}, M_u^{(2)})\,\mathrm{d}M_u^{(2)}.$$

Since both Itô integrals here are martingales, by Proposition 2.3.a we know that $\{f(t, M_t^{(1)}, M_t^{(2)})\}$ is a martingale. Thus,

$$\mathbb{E}\left[e^{u_1 M_t^{(1)} + u_2 M_t^{(2)} - \frac{1}{2}(u_1^2+u_2^2)t}\right] = \mathbb{E}\left[f(t, M_t^{(1)}, M_t^{(2)})\right] = 1,$$

which implies that $\mathbb{E}\left[e^{u_1 M_t^{(1)} + u_2 M_t^{(2)}}\right] = e^{\frac{1}{2}u_1^2 t} e^{\frac{1}{2}u_2^2 t}$, meaning that the joint moment generating function of $M_t^{(1)}$ and $M_t^{(2)}$ can be written as a product of the two respective marginal moment generating functions. Hence, the random variables $M_t^{(1)}$ and $M_t^{(2)}$ are independent for all t.

Using higher-dimensional Itô formulas (not covered here), one can similarly show that the random vectors $(M_{t_1}^{(1)}, \dots, M_{t_n}^{(1)})$ and $(M_{t_1}^{(2)}, \dots, M_{t_n}^{(2)})$ are independent. □

4.4.9 **Constructing correlated stock price processes from independent Brownian motions.** From the preceding discussion on multivariable stochastic calculus, we have worked with *independent* Brownian motions. Despite the independence, we can actually construct some *correlated* stock price processes from them. To see this, consider the following example.

Let $\{S_t^{(1)}\}$ and $\{S_t^{(2)}\}$ be stock price processes (two-dimensional Itô processes) with dynamics given by the following SDEs:

$$\begin{aligned}
\mathrm{d}S_t^{(1)} &= \alpha_1 S_t^{(1)}\,\mathrm{d}t + \sigma_1\,\mathrm{d}W_t^{(1)} \\
\mathrm{d}S_t^{(2)} &= \alpha_2 S_t^{(2)}\,\mathrm{d}t + \sigma_2 S_t^{(2)} \left[\rho\,\mathrm{d}W_t^{(1)} + \sqrt{1-\rho^2}\,\mathrm{d}W_t^{(2)}\right]
\end{aligned}$$

where $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ are independent Brownian motions, $\alpha_1, \alpha_2 \in \mathbb{R}$, $\sigma_1 > 0, \sigma_2 > 0$, and $-1 \leq \rho \leq 1$. Define $W_t^{(3)} = \rho W_t^{(1)} + \sqrt{1-\rho^2} W_t^{(2)}$ for all t. Then, $\{W_t^{(3)}\}$ is a martingale with continuous paths, which satisfies that (i) $W_0^{(3)} = 0$, and (ii)

$$\mathrm{d}W_t^{(3)}\,\mathrm{d}W_t^{(3)} = \rho^2\,\mathrm{d}W_t^{(1)}\,\mathrm{d}W_t^{(1)} + 2\rho\sqrt{1-\rho^2}\,\mathrm{d}W_t^{(1)}\,\mathrm{d}W_t^{(2)} + (1-\rho)^2\,\mathrm{d}W_t^{(2)}\,\mathrm{d}W_t^{(2)} = \rho^2\,\mathrm{d}t + (1-\rho^2)\,\mathrm{d}t = \mathrm{d}t,$$

meaning that $[W_3, W_3]_t = t$, for every $t \geq 0$. Thus, by Theorem 4.4.d, we know that $\{W_t^{(3)}\}$ is a Brownian motion. Writing $\mathrm{d}S_t^{(2)} = \alpha_2 S_t^{(2)}\,\mathrm{d}t + \sigma_2 S_t^{(2)} W_t^{(3)}$, we can readily observe that $\{S_t^{(2)}\}$ follows a geometric Brownian motion, like $\{S_t^{(1)}\}$.

After establishing that both processes are geometric Brownian motions under such SDEs, we will then show that they are indeed correlated. To start with, applying Corollary 4.4.c gives

$$\begin{aligned}
\mathrm{d}W_t^{(1)} W_t^{(3)} &= W_t^{(1)}\,\mathrm{d}W_t^{(3)} + W_t^{(3)}\,\mathrm{d}W_t^{(1)} + \mathrm{d}W_t^{(1)}\,\mathrm{d}W_t^{(3)} \\
&= W_t^{(1)}\,\mathrm{d}W_t^{(3)} + W_t^{(3)}\,\mathrm{d}W_t^{(1)} + \mathrm{d}W_t^{(1)}(\rho W_t^{(1)} + \sqrt{1-\rho^2} W_t^{(2)}) \\
&= W_t^{(1)}\,\mathrm{d}W_t^{(3)} + W_t^{(3)}\,\mathrm{d}W_t^{(1)} + \rho\,\mathrm{d}t.
\end{aligned}$$

Expressing this in integral form, we have

$$W_t^{(1)} W_t^{(3)} = \underbrace{W_0^{(1)} W_0^{(3)}}_{0} + \int_0^t W_u^{(1)}\,\mathrm{d}W_u^{(3)} + \int_0^t W_u^{(3)}\,\mathrm{d}W_u^{(1)} + \underbrace{\int_0^t \rho\,\mathrm{d}t}_{\rho t}.$$

Since Itô integral has zero mean, taking expectation yields $\mathbb{E}\Big[W_t^{(1)} W_t^{(3)}\Big] = \rho t$. Hence, the correlation of $W_t^{(1)}$ and $W_t^{(3)}$ is

$$\operatorname{Corr}\Big(W_t^{(1)}, W_t^{(3)}\Big) = \frac{\operatorname{Cov}\Big(W_t^{(1)}, W_t^{(3)}\Big)}{\sqrt{\operatorname{Var}\Big(W_t^{(1)}\Big)}\sqrt{\operatorname{Var}\Big(W_t^{(3)}\Big)}} = \frac{\mathbb{E}\Big[W_t^{(1)} W_t^{(3)}\Big]}{\sqrt{t}\sqrt{t}} = \rho,$$

for every t. This means that the Brownian motions that drive the dynamics of the stock price processes (geometric Brownian motions) $\{S_t^{(1)}\}$ and $\{S_t^{(2)}\}$ are correlated with coefficient ρ.

5 Risk-Neutral Pricing

5.0.1 Equipped with the tools developed in Section 4, we are now able to study the mathematical details of an extremely powerful method for pricing derivatives, known as *risk-neutral pricing*, which lies in the heart of modern financial economics. In STAT3905/STAT3910, we have briefly investigated the risk-neutral pricing and carried out some numerical computations about those risk-neutral pricing formulas; the mathematical details about the risk-neutral pricing have not been fully discussed. In Section 5, we will delve into the mathematical details and justify rigorously why the risk-neutral pricing works, using tools from previous sections.

5.0.2 **Basic idea of risk-neutral pricing.** Mathematically, the method of risk-neutral pricing is closely related to the notion of *martingale* (hence it is sometimes called *martingale pricing*). The idea is to change the probability measure (as suggested in Section 1.7) such that a martingale is formed from the stock prices discounted by the risk-free rate, $e^{-rt}S_t$. With $\{e^{-rt}S_t\}$ being a $\{\mathcal{F}_t\}$-martingale, we know that $\mathbb{E}[e^{-rt}S_t] = \mathbb{E}[\mathbb{E}[e^{-rt}S_t|\mathcal{F}_0]] = \mathbb{E}\big[e^{-0}S_0\big] = S_0$ (assuming S_0 is nonrandom), meaning that the future discounted stock price $e^{-rt}S_t$ can be "passed" to the present price S_0 upon taking expectation.

Rearranging this equation yields $\mathbb{E}[S_t] = S_0e^{rt}$, which tells us that the mean return rate of the (risky) stock is still the risk-free rate. As such mean return rate would arise under the *risk-neutral* preference, we call this pricing method as <u>risk-neutral</u> pricing.

[Note: In the real market (with actual probability measure), the mean return rate of risky stock is usually higher than the risk-free rate (the difference is known as the *risk premium*)), which corresponds to the case of *risk-averse* preference, under which extra return is demanded for investing in the risky stock.]

5.1 Risk-Neutral Measure

5.1.1 **Absorbing drift terms through changing to risk-neutral measure.** Intuitively, changing the probability measure to a *risk-neutral measure* can be interpreted as "absorbing" the drift term in the stock price process.

To understand this better, consider a stock price process $\{S_t\}$ following a geometric Brownian motion: $\mathrm{d}S_t = \alpha S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$, where $\{W_t\}$ is a Brownian motion, $\alpha \in \mathbb{R}$, and $\sigma > 0$. Let r be the risk-free rate and $D_t = e^{-rt}$ be the time-t discount factor for every t. Then, by Itô formula with $f(t,x) = e^{-rt}x = D_tx$, we have

$$\begin{aligned}\mathrm{d}(D_tS_t) &= f_t(t,S_t)\,\mathrm{d}t + f_x(t,S_t)\,\mathrm{d}S_t + 0 = -rD_tS_t\,\mathrm{d}t + D_t\,\mathrm{d}S_t\\ &= -rD_tS_t\,\mathrm{d}t + \alpha D_tS_t\,\mathrm{d}t + \sigma D_tS_t\,\mathrm{d}W_t = D_tS_t(\alpha-r)\,\mathrm{d}t + \sigma D_tS_t\,\mathrm{d}W_t\\ &= \sigma D_tS_t\left(\frac{\alpha-r}{\sigma}\,\mathrm{d}t + \mathrm{d}W_t\right) \overset{(\widetilde{W}_t=\int_0^t \frac{\alpha-r}{\sigma}\,\mathrm{d}u+\int_0^t 1\,\mathrm{d}W_u=\frac{\alpha-r}{\sigma}t+W_t)}{=} \sigma D_tS_t\,\mathrm{d}\widetilde{W}_t.\end{aligned}$$

Under the original probability measure, $\{\widetilde{W}_t\}$ would generally *not* be a Brownian motion, and so one cannot just apply [4.2.8] to conclude that $\{D_tS_t\}$ is martingale. However, after changing the probability measure to a *risk-neutral* one, the process $\{\widetilde{W}_t\}$ would indeed become a Brownian motion. Therefore, in such case, changing the probability measure to the risk-neutral measure does "absorb" the original drift term $D_tS_t(\alpha-r)$, without influencing the volatility term (still σD_tS_t). As the drift term becomes zero, we can then apply [4.2.8] to conclude that $\{D_tS_t\}$ is a martingale *under the risk-neutral measure*.

5.1.2 **Preliminaries.** The idea put forward in [5.1.1] can be mathematically supported through the *Girsanov theorem*. Here we will lay some groundwork that will be helpful in proving the Girsanov theorem. In the following, we are going to establish a "conditional version" of Proposition 1.7.b, which tells us the effect on *conditional* expectation after change of measure.

Lemma 5.1.a. Let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration, Z be an almost surely positive random variable satisfying $\mathbb{E}[Z] = 1$. Let $\widetilde{\mathbb{P}}$ be defined as in Equation (1), i.e., $\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ for every $A \in \mathcal{F}$. Define a process $\{Z_t\}_{t\in[0,T]}$ by $Z_t = \mathbb{E}[Z|\mathcal{F}_t]$ for every $t \in [0,T]$ (which is a Doob martingale).

(a) For all $t \in [0,T]$, $\widetilde{\mathbb{E}}[Y] = \mathbb{E}[YZ_t]$, where Y is $\mathcal{F}_t$-measurable and YZ_t is integrable.

(b) For all $0 \le s \le t \le T$, $\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \mathbb{E}[YZ_t|\mathcal{F}_s]/Z_s$, where Y is $\mathcal{F}_t$-measurable and YZ_t is integrable.

Proof.

(a) Note that

$$\widetilde{\mathbb{E}}[Y] \overset{\text{(Proposition 1.7.b)}}{=} \mathbb{E}[YZ] = \mathbb{E}[\mathbb{E}[YZ|\mathcal{F}_t]] \overset{\text{(TOWIK)}}{=} \mathbb{E}[Y\mathbb{E}[Z|\mathcal{F}_t]] = \mathbb{E}[YZ_t].$$

(b) With the assumption that $Z > 0$ almost surely, it can be shown that $Z_s = \mathbb{E}[Z|\mathcal{F}_s] > 0$ almost surely, so the (almost sure) equality is well-defined. Next, we verify the two properties of conditional expectations:

(1) Since $\mathbb{E}[YZ_t|\mathcal{F}_s]$ and Z_s are both $\mathcal{F}_s$-measurable, $\mathbb{E}[YZ_t|\mathcal{F}_s]/Z_s$ is $\mathcal{F}_s$-measurable.

(2) Fix any $A \in \mathcal{F}_s$. Then,

$$\begin{aligned}\int_A \frac{\mathbb{E}[YZ_t|\mathcal{F}_s]}{Z_s}\,\mathrm{d}\widetilde{\mathbb{P}} &= \widetilde{\mathbb{E}}\left[\mathbf{1}_A \frac{1}{Z_s}\mathbb{E}[YZ_t|\mathcal{F}_s]\right] \underset{(\times Z_s)}{\overset{\text{(a)}}{=}} = \mathbb{E}[\mathbf{1}_A\mathbb{E}[YZ_t|\mathcal{F}_s]] \overset{\text{(TOWIK)}}{=} \mathbb{E}[\mathbb{E}[\mathbf{1}_A YZ_t|\mathcal{F}_s]] \\ &\overset{\text{(iterated conditioning)}}{=} \mathbb{E}[\mathbf{1}_A YZ_t] \overset{\text{(a)}}{=} \widetilde{\mathbb{E}}[\mathbf{1}_A Y] = \int_A Y\,\mathrm{d}\widetilde{\mathbb{P}}.\end{aligned}$$

□

[Note: Lemma 5.1.a is typically applied in the following context. Let $\{Z_t\}$ be a $\{\mathcal{F}_t\}$-martingale, and $Z = Z_T$. Assume that $Z = Z_T$ is an almost surely positive random variable satisfying $\mathbb{E}[Z] = 1$. Noting that $\mathbb{E}[Z|\mathcal{F}_t] = \mathbb{E}[Z_T|\mathcal{F}_t] = Z_t$ for all $t \in [0,T]$ in such case, we can then use Lemma 5.1.a to conclude that $\widetilde{\mathbb{E}}[Y] = \mathbb{E}[YZ_t]$ and $\widetilde{\mathbb{E}}[Y|\mathcal{F}_s] = \mathbb{E}[YZ_t|\mathcal{F}_s]/Z_s$, where Y is $\mathcal{F}_t$-measurable, for all $0 \le s \le t \le T$.]

5.1.3 **Girsanov theorem.** Now, we are ready to study the *Girsanov theorem*, which justifies the idea of "absorbing drift term" from [5.1.1].

Theorem 5.1.b (Girsanov)**.** Let $\{W_t\}_{t\in[0,T]}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $T > 0$ is a constant, $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration for the Brownian motion, and $\{\Theta_t\}_{t\in[0,T]}$ be a $\{\mathcal{F}_t\}$-adapted process. For all $t \in [0,T]$, define

$$\begin{aligned} Z_t &= \exp\left[-\int_0^t \Theta_u\,\mathrm{d}W_u - \frac{1}{2}\int_0^t \Theta_u^2\,\mathrm{d}u\right], \\ \widetilde{W}_t &= W_t + \int_0^t \Theta_u\,\mathrm{d}u. \end{aligned}$$

Suppose the square-integrability condition $\mathbb{E}\left[\int_0^T \Theta_u^2 Z_u^2\,\mathrm{d}u\right] < \infty$ is satisfied. Let $Z = Z_T > 0$ and define $\widetilde{\mathbb{P}}$ as in Equation (1), i.e., $\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ for every $A \in \mathcal{F}$. Then, $\{\widetilde{W}_t\}_{t\in[0,T]}$ is a Brownian motion on the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.

Proof. **Showing that $\{\widetilde{W}_t\}$ accumulates quadratic variations at unit rate.** Note that we can write $\widetilde{W}_t = \int_0^t 1\,\mathrm{d}W_u + \int_0^t \Theta_u\,\mathrm{d}u$ for all t, and hence $\{\widetilde{W}_t\}$ is an Itô process (under $\mathbb{P}$). Thus, by Proposition 4.2.b, we know that $\mathrm{d}\widetilde{W}_t\,\mathrm{d}\widetilde{W}_t = 1^2\,\mathrm{d}t = \mathrm{d}t$.

Showing that $\{\widetilde{W}_t Z_t\}$ is a martingale under $\mathbb{P}$. Let $X_t = -\int_0^t \Theta_u\,\mathrm{d}W_u - \frac{1}{2}\int_0^t \Theta_u^2\,\mathrm{d}u$. Then, $\{X_t\}$ is an Itô process and we have $\mathrm{d}X_t = -\frac{1}{2}\Theta_t^2\,\mathrm{d}t - \Theta_t\,\mathrm{d}W_t$. Writing $Z_t = e^{X_t}$, by the Itô formula for Itô process we have

$$\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t + \frac{1}{2}Z_t\,\mathrm{d}X_t\,\mathrm{d}X_t = -\frac{1}{2}\Theta_t^2 Z_t\,\mathrm{d}t - \Theta_t Z_t\,\mathrm{d}W_t + \frac{1}{2}Z_t\Theta_t^2\,\mathrm{d}t = -\Theta_t Z_t\,\mathrm{d}W_t.$$

Therefore, applying the Itô product rule gives

$$\begin{aligned}
\mathrm{d}\widetilde{W}_t Z_t &= \widetilde{W}_t \,\mathrm{d}Z_t + Z_t \,\mathrm{d}\widetilde{W}_t + \mathrm{d}\widetilde{W}_t \,\mathrm{d}Z_t \\
&= \widetilde{W}_t(-\Theta_t Z_t \,\mathrm{d}W_t) + Z_t(\mathrm{d}W_t + \Theta_t \,\mathrm{d}t) + (\mathrm{d}W_t + \Theta_t \,\mathrm{d}t)(-\Theta_t Z_t \,\mathrm{d}W_t) \\
&= (-\widetilde{W}_t \Theta_t Z_t + Z_t) \,\mathrm{d}W_t.
\end{aligned}$$

Since the drift term is zero, by [4.2.8] we conclude that $\{\widetilde{W}_t Z_t\}$ is a martingale under $\mathbb{P}$.

Showing that $\{\widetilde{W}_t\}$ is a martingale under $\widetilde{\mathbb{P}}$. It is straightforward to see that $\widetilde{W}_t$ is both integrable and $\mathcal{F}_t$-measurable for all $t \in [0,T]$. Thus, it remains to show that $\widetilde{\mathbb{E}}\big[\widetilde{W}_t|\mathcal{F}_s\big] = \widetilde{W}_s$ for all $0 \le s \le t \le T$. Fix any $0 \le s \le t \le T$. Since the drift term is zero in the SDE $\mathrm{d}Z_t = -\Theta_t Z_t \,\mathrm{d}W_t$, we know by [4.2.8] that $\{Z_t\}$ is a martingale under $\mathbb{P}$, and so $\mathbb{E}[Z] = \mathbb{E}[Z_T] = \mathbb{E}[\mathbb{E}[Z_T|\mathcal{F}_0]] = \mathbb{E}[Z_0] = 1$. Hence, we can apply Lemma 5.1.a to get

$$\widetilde{\mathbb{E}}\Big[\widetilde{W}_t|\mathcal{F}_s\Big] = \frac{1}{Z_s}\mathbb{E}\Big[\widetilde{W}_t Z_t|\mathcal{F}_s\Big] = \frac{1}{Z_s}\widetilde{W}_s Z_s = \widetilde{W}_s.$$

Applying Lévy's characterization theorem. Observing that $\{\widetilde{W}_t\}$ starts at zero and has continuous paths, by Theorem 4.4.d we can then conclude that $\{\widetilde{W}_t\}$ is a Brownian motion on the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$. □

5.1.4 **Risk-neutral measure.** The probability measure $\widetilde{\mathbb{P}}$ from the Girsanov theorem is indeed a *risk-neutral measure*. Like [5.1.1], it can be interpreted as being designed for absorbing the drift term. But here, rather than working with only the geometric Brownian motion, we are considering a more general setting with *generalized* geometric Brownian motion.

Let $\{W_t\}_{t\in[0,T]}$ be a Brownian motion and $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration for the Brownian motion, where $T > 0$ is fixed. Suppose that the stock price process $\{S_t\}_{t\in[0,T]}$ follows a generalized geometric Brownian motion:

$$\mathrm{d}S_t = \alpha_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t$$

where $\{\alpha_t\}$ and $\{\sigma_t\}$ are $\{\mathcal{F}_t\}$-adapted processes, with $\sigma_t > 0$ for all $t \in [0,T]$. In integral form, we can express it as

$$S_t = S_0 \exp\left[\int_0^t \sigma_u \,\mathrm{d}W_u + \int_0^t \left(\alpha_u - \frac{1}{2}\sigma_u^2\right)\mathrm{d}u\right] \qquad \text{for all } t \in [0,T].$$

Furthermore, suppose that the risk-free rate process $\{r_t\}$ is $\{\mathcal{F}_t\}$-adapted. For every $t \in [0,T]$, let $D_t = e^{-\int_0^t r_u \,\mathrm{d}u}$ be the time-t discount factor. Then, we have

$$D_t S_t = S_0 \exp\left[\int_0^t \sigma_u \,\mathrm{d}W_u + \int_0^t \left(\alpha_u - r_u - \frac{1}{2}\sigma_u^2\right)\mathrm{d}u\right].$$

Now, applying Itô formula with $X_t = \int_0^t \sigma_u \,\mathrm{d}W_u + \int_0^t \big(\alpha_u - r_u - \frac{1}{2}\sigma_u^2\big)\,\mathrm{d}u$ and $f(t,x) = S_0 e^x$ gives

$$\begin{aligned}
\mathrm{d}(D_t S_t) &= S_0 e^{X_t}\,\mathrm{d}X_t + \frac{1}{2}S_0 e^{X_t}\sigma_t^2 \,\mathrm{d}t = D_t S_t\big[(\alpha_t - r_t - \sigma_t^2/2)\,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t\big] + \frac{1}{2}D_t S_t \sigma_t^2 \,\mathrm{d}t \\
&= \sigma_t D_t S_t\left(\frac{\alpha_t - r_t}{\sigma_t}\,\mathrm{d}t + \mathrm{d}W_t\right) = \sigma_t D_t S_t(\Theta_t \,\mathrm{d}t + \mathrm{d}W_t) \overset{(\widetilde{W}_t = W_t + \int_0^t \Theta_u \,\mathrm{d}u)}{=} \sigma_t D_t S_t \,\mathrm{d}\widetilde{W}_t,
\end{aligned}$$

where $\Theta_t = (\alpha_t - r_t)/\sigma_t$ is known as the **market price of risk** (which satisfies the square-integrability condition mentioned in the Girsanov theorem).

Applying Girsanov theorem, we know that $\{\widetilde{W}_t\}$ is a Brownian motion under the probability measure $\widetilde{\mathbb{P}}$, and so the drift term of $\{D_t S_t\}$ is "absorbed" after changing the probability measure to $\widetilde{\mathbb{P}}$, and $\{D_t S_t\}$ is a martingale under the probability measure $\widetilde{\mathbb{P}}$. In addition, by Proposition 1.7.c, we know that $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent.

Mathematically, a probability measure $\widetilde{\mathbb{P}}$ is said to be a **risk-neutral measure** (or **(equivalent) martingale measure**) if

(1) $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent.

(2) Under $\widetilde{\mathbb{P}}$, the discounted stock price process $\{D_t S_t\}$ is a $\{\mathcal{F}_t\}$-martingale.

[Note: In [5.3.5], we will revisit the concept of risk-neutral measure in a more general setting.]

5.1.5 **Dynamics under risk-neutral measure.** After studying the theoretical foundation of risk-neutral measure, we are going to analyze the behaviour of the prices of different assets under the risk-neutral measure, to better understand the impacts on the market dynamics from changing the probability measure to the risk-neural one.

(a) *(Stock)* Combining the SDE $\mathrm{d}S_t = \alpha_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t$ for the generalized geometric Brownian motion and the SDE $\mathrm{d}\widetilde{W}_t = \Theta_t \,\mathrm{d}t + \mathrm{d}W_t$, we get

$$\mathrm{d}S_t = \alpha_t S_t \,\mathrm{d}t + \sigma_t S_t(\mathrm{d}\widetilde{W}_t - \Theta_t \,\mathrm{d}t) = r_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}\widetilde{W}_t.$$

This suggests that under the risk-neutral measure $\widetilde{\mathbb{P}}$, the stock price process $\{S_t\}$ still follows a generalized geometric Brownian motion with the drift term changed to $r_t S_t$ (mean return being risk-free rate) and the volatility term unchanged.

(b) *(Self-financing portfolio)* Consider a self-financing portfolio analogous to the one in [4.3.3], but for the case of generalized geometric Brownian motion here.

Like [4.3.3], at each time t, let X_t denote the investor's portfolio value and Δ_t denote the number of shares of stock held by the investor. Suppose that $\{\Delta_t\}$ is adapted to a filtration $\{\mathcal{F}_t\}$ for the Brownian motion $\{W_t\}$, and at each time t, the non-stock component of the portfolio, with value $X_t - \Delta_t S_t$, is all invested in the bond.

Then, the "change" $\mathrm{d}X_t$ for investor's portfolio at time t sources from (i) the gain $\Delta_t \,\mathrm{d}S_t$ on the stock position and (ii) the interest earning $r_t(X_t - \Delta_t S_t)\,\mathrm{d}t$ on the bond position. Hence, we can write

$$\begin{aligned}\mathrm{d}X_t &= \Delta_t \underbrace{\mathrm{d}S_t}_{\alpha_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t} + r_t(X_t - \Delta_t S_t)\,\mathrm{d}t = r_t X_t \,\mathrm{d}t + \Delta_t(\alpha_t - r_t)S_t \,\mathrm{d}t + \Delta_t \sigma_t S_t \,\mathrm{d}W_t \\ &= r_t X_t \,\mathrm{d}t + \Delta_t \sigma_t S_t(\Theta_t \,\mathrm{d}t + \mathrm{d}W_t) \overset{(\widetilde{W}_t = W_t + \int_0^t \Theta_u \,\mathrm{d}u)}{=} r_t X_t \,\mathrm{d}t + \Delta_t \sigma_t S_t \,\mathrm{d}\widetilde{W}_t\end{aligned}$$

where $\Theta_t = (\alpha_t - r_t)/\sigma_t$. This suggests that under the risk-neutral measure, the (undiscounted) portfolio values also follow a generalized geometric Brownian motion, where the mean return is still risk-free rate like the stock, but the volatility term differs and depends on the stock position Δ_t.

Further insights can be gained by studying the behaviour of *discounted* portfolio values. Again, let $D_t = e^{-\int_0^t r_u \,\mathrm{d}u}$ be the time-t discount factor for every t. Applying Itô formula with $X_t = \int_0^t r_u \,\mathrm{d}u$ and $f(t, x) = e^{-x}$ gives

$$\mathrm{d}D_t = -\underbrace{e^{-X_t}}_{D_t} \underbrace{\mathrm{d}X_t}_{r_t \,\mathrm{d}t} + 0 = -r_t D_t \,\mathrm{d}t.$$

Hence, by Itô product rule, we have

$$\begin{aligned}\mathrm{d}(D_t X_t) &= D_t \,\mathrm{d}X_t + X_t \,\mathrm{d}D_t + \mathrm{d}D_t \,\mathrm{d}X_t = D_t(r_t X_t \,\mathrm{d}t + \Delta_t \sigma_t S_t \,\mathrm{d}\widetilde{W}_t) + X_t(-r_t D_t \,\mathrm{d}t) + 0 \\ &= \Delta_t \underbrace{\sigma_t D_t S_t \,\mathrm{d}\widetilde{W}_t}_{\mathrm{d}D_t S_t} = \Delta_t \,\mathrm{d}(D_t S_t).\end{aligned}$$

This shows that $\{D_t X_t\}$ is a martingale under the risk-neutral measure, like the stock prices. Furthermore, the SDE above indicates that the "change" in discounted portfolio value $\mathrm{d}(D_t X_t)$ indeed only sources from the "discounted gain" $\Delta_t \,\mathrm{d}(D_t S_t)$ on the stock position.

(c) *(Derivative)* In the proof of Theorem 4.3.a, we have utilized a replication argument for an European call option. Here, we extend this idea to *any* derivative that can be replicated; we shall still impose the usual assumption that the market is arbitrage-free, with a stock and a (risk-free) bond which can be freely bought or (short) sold in any amount without transaction cost.

[Note: We will have more discussions on the existence of replicating portfolio in [5.2.2].]

Consider a T-year derivative with time-T payoff V_T. Under the assumption that it can be replicated, there is a self-financing portfolio with price process $\{X_t\}$ such that $X_T = V_T$. Under the no-arbitrage condition, this implies that $X_t = V_t$ for all $0 \leq t < T$ also, where V_t denotes the time-t price of the derivative.

After having such self-financing portfolio constructed, by [5.1.5]b we know that $\{D_t X_t\}$ is a martingale under the risk-neutral measure $\widetilde{\mathbb{P}}$. Thus, for every $t \in [0, T]$, we have

$$D_t V_t \overset{(X_t = V_t)}{=} D_t X_t = \widetilde{\mathbb{E}}[D_T X_T | \mathcal{F}_t] \overset{(X_T = V_T)}{=} \widetilde{\mathbb{E}}[D_T V_T | \mathcal{F}_t].$$

Since $D_T / D_t = e^{-\int_0^T r_u \,\mathrm{d}u} / e^{-\int_0^t r_u \,\mathrm{d}u} = e^{-\int_t^T r_u \,\mathrm{d}u}$, rearranging the equation gives

$$V_t = \widetilde{\mathbb{E}}\left[e^{-\int_t^T r_u \,\mathrm{d}u} V_T \middle| \mathcal{F}_t \right] \quad \text{for all } t \in [0, T],$$

which is the *risk-neutral pricing formula*.

5.1.6 **Deriving Black-Scholes equation and Black-Scholes formula through risk-neutral measure.** In Section 4.3, we have derived the Black-Scholes equation and formula purely by stochastic calculus. However, this is not the only way to do so, and the notion of *risk-neutral measure* gives us an alternative route for deriving them. Let $V(t, S_t)$ denote the time-t value of a T-year derivative satisfying the assumptions specified in Theorem 4.3.a. Assume that $r_t = r$ and $\sigma_t = \sigma$ for all $t \in [0, T]$, where r and $\sigma > 0$ are constants.

(a) *(Deriving Black-Scholes formula)* Previously the Black-Scholes formula is derived as the solution to the Black-Scholes equation, for European call. A more direct way of deriving the formula is to utilize the *risk-neutral pricing formula* from [5.1.5]c; such approach is indeed used in STAT3905/STAT3910, where the Black-Scholes formula is derived through suitable algebraic manipulations. The key idea is as follows. Suppose the derivative in consideration is a T-year K-strike European call.

$$V(t, S_t) \overset{[5.1.5]\text{c}}{=} \mathbb{E}\left[e^{-r(T-t)} (S_T - K)_+ | \mathcal{F}_t \right] \overset{\text{(algebra)}}{=} S_t \Phi(d_+(T-t, S_t)) - K e^{-r(T-t)} \Phi(d_-(T-t, S_t)).$$

(b) *(Deriving Black-Scholes equation)* The idea of risk-neutral measure also leads to an alternative method for deriving the Black-Scholes equation, by working with the differentials directly and applying the converse of [4.2.8].

Consider the process $\{V(t, S_t)\}$ of derivative prices, and its corresponding replicating portfolio $\{X_t\}$, satisfying that $X_t = V(t, S_t)$ for all $t \in [0, T]$. From the proof of Theorem 4.3.a, we know

$$\begin{aligned}
\mathrm{d}V(t, S_t) &= \left[V_t(t, S_t) + \alpha S_t V_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t, S_t) \right] \mathrm{d}t + \sigma S_t V_x(t, S_t) \,\mathrm{d}W_t \\
&\overset{(\widetilde{W}_t = \frac{\alpha - r}{\sigma} t + W_t)}{=} \left[V_t(t, S_t) + \alpha S_t V_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t, S_t) \right] \mathrm{d}t + \sigma S_t V_x(t, S_t) \left(\mathrm{d}\widetilde{W}_t - \frac{\alpha - r}{\sigma} \,\mathrm{d}t \right) \\
&= \left[V_t(t, S_t) + r S_t V_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t, S_t) \right] \mathrm{d}t + \sigma S_t V_x(t, S_t) \,\mathrm{d}\widetilde{W}_t.
\end{aligned}$$

Hence, letting $D_t = e^{-rt}$ denote the time-t discount factor and applying Itô product rule, we have

$$\begin{aligned}
\mathrm{d}(D_t V(t, S_t)) &= V(t, S_t) \underbrace{\mathrm{d}D_t}_{-r D_t \,\mathrm{d}t} + D_t \,\mathrm{d}V(t, S_t) + \underbrace{\mathrm{d}D_t \,\mathrm{d}V(t, S_t)}_{0} \\
&= -r D_t V(t, S_t) \,\mathrm{d}t + D_t \left\{ \left[V_t(t, S_t) + r S_t V_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t, S_t) \right] \mathrm{d}t + \sigma S_t V_x(t, S_t) \,\mathrm{d}\widetilde{W}_t \right\} \\
&= D_t \left[V_t(t, S_t) + r S_t V_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 V_{xx}(t, S_t) - r V(t, S_t) \right] \mathrm{d}t + \sigma D_t S_t V_x(t, S_t) \,\mathrm{d}\widetilde{W}_t.
\end{aligned}$$

By [5.1.5]b, we know that $\{D_tX_t\} = \{D_tV(t,S_t)\}$ is a martingale. Therefore, applying the converse of [4.2.8] gives

$$V_t(t,S_t) + rS_tV_x(t,S_t) + \frac{1}{2}\sigma^2S_t^2V_{xx}(t,S_t) - rV(t,S_t) = 0.$$

Rearranging this yields the Black-Scholes equation.

5.2 Martingale Representation Theorem

5.2.1 **Martingale representation theorem.** Recall from [4.1.5] that Itô integral forms a martingale. It turns out that we can also work backwards, and establish that a martingale can be represented in terms of an Itô integral, as suggested by the *martingale representation theorem*.

Theorem 5.2.a (Martingale representation)**.** Let $\{W_t\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t\}$ be the filtration generated by $\{W_t\}$[16]. If $\{M_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\mathbb{P}$, then there is a $\{\mathcal{F}_t\}$-adapted process $\{\Gamma_u\}$ such that $M_t = M_0 + \int_0^t \Gamma_u \, \mathrm{d}W_u$ for all t.

Proof. Omitted. □

There is also a version of martingale representation theorem for the risk-neutral measure as follows.

Theorem 5.2.b (Martingale representation (risk-neutral))**.** Let $\{W_t\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t\}$ be the filtration generated by $\{W_t\}$. Consider the risk-neutral probability measure $\widetilde{\mathbb{P}}$ and the corresponding Brownian motion $\{\widetilde{W}_t\}$ as specified in the Girsanov theorem (assuming the conditions there hold). If $\{\widetilde{M}_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\widetilde{\mathbb{P}}$, then there is a $\{\mathcal{F}_t\}$-adapted process $\{\widetilde{\Gamma}_u\}$ such that $\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \widetilde{\Gamma}_u \, \mathrm{d}\widetilde{W}_u$ for all t.

[Note: This does not immediately follow from Theorem 5.2.a since the filtration $\{\mathcal{F}_t\}$ is generated by the Brownian motion $\{W_t\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, rather than the Brownian motion $\{\widetilde{W}_t\}$ on $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.]

Proof. Omitted. □

5.2.2 **Replication with one stock.** We are now equipped with enough tools to show that the existence of replication portfolio as specified in [5.1.5]c. It turns out that a requirement for the existence is that the underlying filtration $\{\mathcal{F}_t\}$ needs to be the one generated by the Brownian motion $\{W_t\}$, i.e., there should not be "extra" information available other than the ones coming from Brownian motion.

Proposition 5.2.c. Let $\{W_t\}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t\}$ be the filtration generated by $\{W_t\}$. Let $T > 0$ and consider any derivative with time-T payoff V_T, where V_T is integrable (with respect to the risk-neutral measure $\widetilde{\mathbb{P}}$) and $\mathcal{F}_T$-measurable[17]. Then, the derivative can be replicated, by a self-financing portfolio with $X_0 = \widetilde{\mathbb{E}}[D_TV_T|\mathcal{F}_0]$ and $\Delta_t = \widetilde{\Gamma}_t/(\sigma_tD_tS_t)$ for all $t \in [0,T]$, where $\widetilde{\Gamma}_t$ comes from Theorem 5.2.b, and the other notations carry their usual meanings.

Proof. Let $\widetilde{M}_t = \widetilde{\mathbb{E}}[D_TV_T|\mathcal{F}_t]$ for all $t \in [0,T]$. Note that $\{\widetilde{M}_t\}$ is a Doob martingale under $\widetilde{\mathbb{P}}$. Thus, by Theorem 5.2.b we have

$$\widetilde{M}_T = \widetilde{M}_0 + \int_0^T \widetilde{\Gamma}_t \, \mathrm{d}\widetilde{W}_t$$

where $\{\widetilde{\Gamma}_t\}$ is a $\{\mathcal{F}_t\}$-adapted process.

[16] This means that for all t, $\mathcal{F}_t$ is the smallest σ-algebra for which W_s is $\mathcal{F}_t$-measurable for all $0 \le s \le t$. Intuitively, this means $\mathcal{F}_t$ contains only the information arising from the Brownian motion up to time t. Particularly, $\{\mathcal{F}_t\}$ is a special kind of filtration for Brownian motion. For a more formal definition, see STAT7610.

[17] This condition just means that the payoff can be determined based on the information at time T, which is satisfied for derivatives with T-year term.

Next, consider any self-financing portfolio with time-t value X_t and time-t stock position Δ_t (with $\{\Delta_t\}$ being $\{\mathcal{F}_t\}$-adapted and the non-stock component being invested in the risk-free bond). By [5.1.5]b, we know that $\{D_t X_t\}$ follows the SDE $\mathrm{d}D_t X_t = \Delta_t \sigma_t D_t S_t \,\mathrm{d}\widetilde{W}_t$. This implies that

$$D_T X_T = \underbrace{D_0}_{1} X_0 + \int_0^T \Delta_t \sigma_t D_t S_t \,\mathrm{d}\widetilde{W}_t = X_0 + \int_0^T \Delta_t \sigma_t D_t S_t \,\mathrm{d}\widetilde{W}_t.$$

Hence, by setting $X_0 = \widetilde{M}_0 = \mathbb{E}[D_T V_T | \mathcal{F}_0]$ and $\Delta_t = \widetilde{\Gamma}_t / (\sigma_t D_t S_t)$ for all $t \in [0, T]$, we can ensure that $D_T X_T = \widetilde{M}_T = \widetilde{\mathbb{E}}[D_T V_T | \mathcal{F}_T] \overset{\text{(TOWIK)}}{=} D_T V_T$. This implies that $X_T = V_T$, and so a replicating portfolio is constructed. □

[Note: While Proposition 5.2.c assures the *existence* of replicating portfolio under such conditions, the explicit construction of such portfolio remains unclear since Theorem 5.2.b does not give us a method for finding $\widetilde{\Gamma}_t$.]

5.3 Market With Multiple Stocks

5.3.1 So far, we have exclusively worked in an arbitrage-free market consisting of a *single* stock and a (risk-free) bond which can be freely bought or (short) sold in any amount without transaction cost, and studied the replication of a *single* derivative under such market. In Section 5.3, we will generalize the idea and allow the market to have more than one stock (which can all be transacted freely without cost still), through which *multiple* derivatives (whose payoffs may depend on more than one stock) can be replicated.

5.3.2 **Multidimensional Girsanov theorem and martingale representation theorem.** Like before, the *Girsanov theorem* and the *martingale representation theorem* serve as two important tools for justifying the replication argument here. However, since we are working in a multidimensional market (having multiple stocks), we need the *multidimensional version* of these results. They will be stated without proof below.

Theorem 5.3.a (Girsanov (multidimensional))**.** *Fix a constant $T > 0$. Let $\{\boldsymbol{W}_t\}_{t\in[0,T]}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\boldsymbol{W}_t = (W_t^{(1)}, \ldots, W_t^{(d)})$ for all t, $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration for the d-dimensional Brownian motion, and $\{\boldsymbol{\Theta}_t\}_{t\in[0,T]}$ be a d-dimensional $\{\mathcal{F}_t\}$-adapted process, with $\boldsymbol{\Theta}_t = (\Theta_t^{(1)}, \ldots, \Theta_t^{(d)})$ for all t. For all $t \in [0, T]$, define*

$$Z_t = \exp\left[-\int_0^t \boldsymbol{\Theta}_u \cdot \mathrm{d}\boldsymbol{W}_u - \frac{1}{2}\int_0^t \|\boldsymbol{\Theta}_u\|^2 \,\mathrm{d}u\right],$$
$$\widetilde{\boldsymbol{W}}_t = \boldsymbol{W}_t + \int_0^t \boldsymbol{\Theta}_u \,\mathrm{d}u,$$

where $\int_0^t \boldsymbol{\Theta}_u \cdot \mathrm{d}\boldsymbol{W}_u := \sum_{j=1}^d \int_0^t \Theta_u^{(j)} \,\mathrm{d}W_u^{(j)}$ ("dot product" notation) and $\int_0^t \boldsymbol{\Theta}_u \,\mathrm{d}u := (\int_0^t \Theta_u^{(1)} \,\mathrm{d}u, \ldots, \int_0^t \Theta_u^{(d)} \,\mathrm{d}u)$ ("vectorized" notation).

Suppose the square-integrability condition $\mathbb{E}\left[\int_0^T \|\Theta_u\|^2 Z_u^2 \,\mathrm{d}u\right] < \infty$ is satisfied. Let $Z = Z_T > 0$ and define $\widetilde{\mathbb{P}}$ as in Equation (1), i.e., $\widetilde{\mathbb{P}}(A) = \int_A Z(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ for every $A \in \mathcal{F}$. Then, $\{\widetilde{\boldsymbol{W}}_t\}_{t\in[0,T]}$ is a d-dimensional Brownian motion on the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.

Proof. Omitted. □

Theorem 5.3.b (Martingale representation (multidimensional))**.** *Let $\{\boldsymbol{W}_t\}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t\}$ be the filtration generated by $\{\boldsymbol{W}_t\}$. If $\{M_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\mathbb{P}$, then there is a $\{\mathcal{F}_t\}$-adapted and d-dimensional process $\{\boldsymbol{\Gamma}_u\}$, with $\boldsymbol{\Gamma}_u = (\Gamma_u^{(1)}, \ldots, \Gamma_u^{(d)})$ for all u, such that $M_t = M_0 + \int_0^t \boldsymbol{\Gamma}_u \cdot \mathrm{d}\boldsymbol{W}_u$ for all t.*

Consider the probability measure $\widetilde{\mathbb{P}}$ and the corresponding Brownian motion $\{\widetilde{\boldsymbol{W}}_t\}$ as specified in the multidimensional Girsanov theorem (Theorem 5.3.a) (assuming the conditions there hold). If $\{\widetilde{M}_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\widetilde{\mathbb{P}}$, then there is a $\{\mathcal{F}_t\}$-adapted and d-dimensional process $\{\widetilde{\boldsymbol{\Gamma}}_u\}$, with $\widetilde{\boldsymbol{\Gamma}}_u = (\widetilde{\Gamma}_u^{(1)}, \dots, \widetilde{\Gamma}_u^{(d)})$ for all u, such that $\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \widetilde{\boldsymbol{\Gamma}}_u \cdot \mathrm{d}\widetilde{\boldsymbol{W}}_u$ for all t.

5.3.3 **Assumptions on the market dynamics.** In addition to the usual no-arbitrage and free-transaction assumptions on the multidimensional market, we will impose the following assumptions on the market dynamics (i.e., the behaviours of the price processes).

Let $\{\boldsymbol{W}_t\}_{t\in[0,T]}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, associated with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$, where $T > 0$ is a constant. Suppose that $\mathcal{F} = \mathcal{F}_T$, and there are m stocks, whose time-t prices are $S_t^{(1)}, \dots, S_t^{(m)}$ respectively, driven by the following SDEs:

$$\mathrm{d}S_t^{(i)} = \alpha_t^{(i)} S_t^{(i)}\,\mathrm{d}t + S_t^{(i)} \sum_{j=1}^{d} \sigma_t^{(ij)}\,\mathrm{d}W_t^{(j)}, \quad i = 1, \dots, m, \tag{3}$$

where $\{(\alpha_t^{(1)}, \dots, \alpha_t^{(m)})\}_t$ and $\{[\sigma_t^{(ij)}]_{i,j=1}^{m,d}\}_t$ are $\{\mathcal{F}_t\}$-adapted[18], with $\sigma_t^{(ij)} \geq 0$ and $\sum_{j=1}^{d} \sigma_t^{(ij)} > 0$ for all t. Furthermore, we suppose the risk-free rate process $\{r_t\}$ is $\{\mathcal{F}_t\}$-adapted.

[Note: The parameters can be interpreted as follows. The value $\alpha_t^{(i)}$ refers to the mean return rate of stock i at time t, and the value $\sigma_t^{(ij)}$ refers to the volatility contribution from the jth component of Brownian motion to the stock i.]

5.3.4 **Analysis of the stock price behaviours.**

(a) *(Generalized geometric Brownian motion for each stock)* For every $i = 1, \dots, m$, define a process $\{B_t^{(i)}\}$ by

$$B_t^{(i)} = \sum_{j=1}^{d} \int_0^t \frac{\sigma_u^{(ij)}}{\sigma_u^{(i)}}\,\mathrm{d}W_u^{(j)}$$

where $\sigma_u^{(i)} := \sqrt{\sum_{j=1}^{d} (\sigma_u^{(ij)})^2} > 0$. In differential form, we can write $\mathrm{d}B_t^{(i)} = \sum_{j=1}^{d} (\sigma_t^{(ij)})/(\sigma_t^{(i)})\,\mathrm{d}W_t^{(j)}$. As a sum of Itô integrals, by [4.1.5] and Proposition 2.3.a we know that each $\{B_t^{(i)}\}$ has continuous paths and is a $\{\mathcal{F}_t\}$-martingale. Moreover, we have $B_0^{(i)} = 0$ and

$$\mathrm{d}B_t^{(i)}\,\mathrm{d}B_t^{(i)} = \sum_{j=1}^{d} \frac{(\sigma_t^{(ij)})^2}{(\sigma_t^{(i)})^2}\,\mathrm{d}t = \mathrm{d}t$$

for each $i = 1, \dots, m$. Therefore, by Lévy's characterization theorem (Theorem 4.4.d), we conclude that each $\{B_t^{(i)}\}$ is a Brownian motion. Using $B_t^{(i)}$, we can rewrite the SDEs in (3) as

$$\mathrm{d}S_t^{(i)} = \alpha_t^{(i)} S_t^{(i)}\,\mathrm{d}t + \sigma_t^{(i)} S_t^{(i)}\,\mathrm{d}B_t^{(i)}, \quad i = 1, \dots, m,$$

indicating that each stock indeed follows a generalized geometric Brownian motion.

(b) *(Relationship between the sources of randomness)* This representation of SDEs also allows us to better understand the relationships between the sources of randomness for different stocks as follows. Fix any $i \neq k$. Then we have

$$\mathrm{d}B_t^{(i)}\,\mathrm{d}B_t^{(k)} = \sum_{j=1}^{d} \frac{\sigma_t^{(ij)}\sigma_t^{(kj)}}{\sigma_t^{(i)}\sigma_t^{(k)}}\,\mathrm{d}t = \rho_t^{(ik)}\,\mathrm{d}t$$

[18]Intuitively, this means that at each time t these vectors and matrices can be determined based on the information from $\{\mathcal{F}_t\}$. We will not give a formal definition here; see STAT7610 if interested.

where $\rho_t^{(ik)} := \frac{1}{\sigma_t^{(i)}\sigma_t^{(k)}} \sum_{j=1}^{d} \sigma_t^{(ij)}\sigma_t^{(kj)}$. By Itô product rule, we have

$$\mathrm{d}B_t^{(i)}B_t^{(k)} = B_t^{(i)}\,\mathrm{d}B_t^{(k)} + B_t^{(k)}\,\mathrm{d}B_t^{(i)} + \mathrm{d}B_t^{(i)}B_t^{(k)}, = B_t^{(i)}\,\mathrm{d}B_t^{(k)} + B_t^{(k)}\,\mathrm{d}B_t^{(i)} + \rho_t^{(ik)}\,\mathrm{d}t$$

which can be expressed in integral form as

$$B_t^{(i)}B_t^{(k)} = \int_0^t B_u^{(i)}\,\mathrm{d}B_u^{(k)} + \int_0^t B_u^{(k)}\,\mathrm{d}B_u^{(i)} + \int_0^t \rho_u^{(ik)}\,\mathrm{d}u.$$

Upon taking expectations, we get the following covariance formula

$$\mathrm{Cov}\Big(B_t^{(i)}, B_t^{(k)}\Big) = \mathbb{E}\left[\int_0^t \rho_u^{(ik)}\,\mathrm{d}u\right],$$

since the expectation of Itô integral is zero.

(c) *(Dynamics of discounted stock prices)* Like before, in the replication argument we often need to deal with *discounted* prices. So, here we will briefly investigate the behaviour of the discounted stock prices. Let $D_t = e^{-\int_0^t r_u\,\mathrm{d}u}$ be the time-t discount factor, which satisfies that $\mathrm{d}D_t = -r_t\,\mathrm{d}D_t\,\mathrm{d}t$. Fix any $i = 1, \dots, m$, and consider the discounted stock price process $\{D_t S_t^{(i)}\}$. By Itô product rule, we have

$$\begin{aligned}\mathrm{d}(D_t S_t^{(i)}) &= D_t\,\mathrm{d}S_t^{(i)} + S_t^{(i)}\,\mathrm{d}D_t + \underbrace{\mathrm{d}D_t\,\mathrm{d}S_t^{(i)}}_{0} \\ &= D_t\left(\alpha_t^{(i)}S_t^{(i)}\,\mathrm{d}t + S_t^{(i)}\sum_{j=1}^{d}\sigma_t^{(ij)}\,\mathrm{d}W_t^{(j)}\right) + S_t^{(i)}(-r_t D_t\,\mathrm{d}t) \\ &= D_t S_t^{(i)}\left[(\alpha_t^{(i)} - r_t)\,\mathrm{d}t + \sum_{j=1}^{d}\sigma_t^{(ij)}\,\mathrm{d}W_t^{(j)}\right].\end{aligned}$$

5.3.5 **Risk-neutral measure.** In the multidimensional market here, we have a more general definition of risk-neutral measure. A probability measure $\widetilde{\mathbb{P}}$ is said to be a **risk-neutral measure** (or **(equivalent) martingale measure**) if

(1) $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent.

(2) Under $\widetilde{\mathbb{P}}$, the discounted stock price process $\{D_t S_t^{(i)}\}$ is a $\{\mathcal{F}_t\}$-martingale, for every $i = 1, \dots, m$.

To investigate the second condition that $\{D_t S_t^{(i)}\}$ is a $\{\mathcal{F}_t\}$-martingale for every $i = 1, \dots, m$, it is helpful to write the SDE from [5.3.4]c as follows:

$$\mathrm{d}(D_t S_t^{(i)}) = D_t S_t^{(i)} \sum_{j=1}^{d} \sigma_t^{(ij)}(\Theta_t^{(j)}\,\mathrm{d}t + \mathrm{d}W_t^{(j)}) \overset{(\widetilde{W}_t^{(j)} = W_t^{(j)} + \int_0^t \Theta_u^{(j)}\,\mathrm{d}u)}{=} D_t S_t^{(i)} \sum_{j=1}^{d} \sigma_t^{(ij)}\,\mathrm{d}\widetilde{W}_t^{(j)}$$

where the processes $\{\Theta_t^{(j)}\}$'s satisfy $\alpha_t^{(i)} - r_t = \sum_{j=1}^{d} \sigma_t^{(ij)}\Theta_t^{(j)}$ for all $i = 1, \dots, m$, known as the **market price of risk equations**.

It can be shown that those processes $\{\Theta_t^{(j)}\}$'s exist (i.e., a solution to the market price of risk equations exists) iff there is no arbitrage in the market.

[Note: The "$\Rightarrow$" direction can be established by the first fundamental theorem of asset pricing (Theorem 5.3.e).]

Hence, under the no-arbitrage assumption, we can then apply the multidimensional Girsanov theorem to obtain a probability measure $\widetilde{\mathbb{P}}$ that is equivalent to $\mathbb{P}$, under which $\{\widetilde{\boldsymbol{W}}_t\}$ is a d-dimensional Brownian motion, hence $\{D_t S_t^{(i)}\}$ is a martingale under $\widetilde{\mathbb{P}}$ for all $i = 1, \dots, m$. Therefore, such probability measure $\widetilde{\mathbb{P}}$ is a risk-neutral measure. This suggests a method for constructing a risk-neutral measure in our market.

5.3.6 **Dynamics of self-financing portfolio under risk-neutral measure.** It turns out that, like [5.1.5]b, the discounted prices of a self-financing portfolio always form a martingale under $\widetilde{\mathbb{P}}$. This result is proved below and serves as a lemma for an important result in financial economics, known as the *first fundamental theorem of asset pricing*.

Lemma 5.3.c. Let $\widetilde{\mathbb{P}}$ be a risk-neutral measure and X_t be the time-t value of a self-financing portfolio, with time-t position on stock i being $\Delta_t^{(i)}$, where $\{\Delta_t^{(i)}\}$ is $\{\mathcal{F}_t\}$-adapted for each $i = 1, \ldots, m$, and the non-stock component being invested in the risk-free bond. Then, the discounted portfolio value process $\{D_t X_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\widetilde{\mathbb{P}}$.

Proof. By the self-financing property, we have

$$\begin{aligned}
\mathrm{d}X_t &= \sum_{i=1}^m \Delta_t^{(i)}\,\mathrm{d}S_t^{(i)} + r_t \underbrace{\left(X_t - \sum_{i=1}^m \Delta_t^{(i)} S_t^{(i)}\right)}_{\text{non-stock component}}\mathrm{d}t = r_t X_t\,\mathrm{d}t + \sum_{i=1}^m \Delta_t^{(i)}\left(\mathrm{d}S_t^{(i)} - r_t S_t^{(i)}\,\mathrm{d}t\right) \\
&= r_t X_t\,\mathrm{d}t + \sum_{i=1}^m \frac{\Delta_t^{(i)}}{D_t}(D_t\,\mathrm{d}S_t^{(i)} - S_t^{(i)} r_t D_t\,\mathrm{d}t) = r_t X_t\,\mathrm{d}t + \sum_{i=1}^m \frac{\Delta_t^{(i)}}{D_t}(D_t\,\mathrm{d}S_t^{(i)} + S_t^{(i)}\,\mathrm{d}D_t) \\
&\overset{\text{(Itô product rule)}}{=} r_t X_t\,\mathrm{d}t + \sum_{i=1}^m \frac{\Delta_t^{(i)}}{D_t}\,\mathrm{d}(D_t S_t^{(i)}).
\end{aligned}$$

Therefore, for the discounted portfolio value process $\{D_t X_t\}$, by Itô product rule we have

$$\begin{aligned}
\mathrm{d}(D_t X_t) &= D_t\,\mathrm{d}X_t + X_t\,\mathrm{d}D_t + \underbrace{\mathrm{d}D_t\,\mathrm{d}X_t}_{0} = D_t(\mathrm{d}X_t - r_t X_t\,\mathrm{d}t) \\
&= \sum_{i=1}^m \Delta_t^{(i)}\,\mathrm{d}(D_t S_t^{(i)}).
\end{aligned}$$

Under the risk-neutral measure $\widetilde{\mathbb{P}}$, by definition we know that $\{D_t S_t^{(i)}\}$ is a $\{\mathcal{F}_t\}$-martingale for every $i = 1, \ldots, m$. Hence, each $\{D_t S_t^{(i)}\}$ has zero drift term (by the converse of [4.2.8]), implying that $\{D_t X_t\}$ has zero drift term also. So, $\{D_t X_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\widetilde{\mathbb{P}}$. □

5.3.7 **Arbitrage.** The *first fundamental theorem of asset pricing* relates the concepts of risk-neutral measure and arbitrage. To prove the result mathematically, first we need to give a mathematical definition of *arbitrage* (though we already have some understanding on what it means from previous courses). An **arbitrage** is a portfolio value process $\{X_t\}$ satisfying that

(1) *(No initial investment)* $X_0 = 0$.
(2) *(Potential for earning profit without risk)* For some $T > 0$, $\mathbb{P}(X_T \geq 0) = 1$ and $\mathbb{P}(X_T > 0) > 0$.

Basically, this is a mathematical way to express the idea that arbitrage is a way to potentially earn profit without initial investment and without any risk, essentially a *free lunch*.

A criterion for the existence of arbitrage is the following.

Proposition 5.3.d. An arbitrage exists iff there is a portfolio price process $\{Y_t\}$ satisfying that

(1) *(Positive initial investment)* $Y_0 > 0$.
(2) *(Earning at least risk-free rate and potentially earning more than risk-free rate)* For some $T > 0$, $\mathbb{P}(Y_T \geq Y_0/D_T) = 1$ and $\mathbb{P}(Y_T > Y_0/D_T) > 0$.

[Note: We have $Y_0/D_T = Y_0 e^{\int_0^T r_t\,\mathrm{d}t}$. So, this condition means that the portfolio earns at least the risk-free rate almost surely, and has a positive probability to earn more than risk-free rate.]

Proof. "⇒": Assume that an arbitrage $\{X_t\}$ exists. We can construct a portfolio by investing $Y_0 > 0$ in the risk-free bond and having a long position in the portfolio corresponding to the arbitrage. Then the resulting portfolio price process, denoted by $\{Y_t\}$, satisfies two conditions above.

"⇐": Assume such portfolio price process $\{Y_t\}$ exists. We can construct a portfolio by having a long position in the portfolio corresponding to such process, and borrowing $Y_0 > 0$ at risk-free rate. Then, the resulting portfolio price process, denoted by $\{X_t\}$, satisfies the two conditions for qualifying as an arbitrage. □

5.3.8 **First fundamental theorem of asset pricing.** We are now ready to prove the first fundamental theorem of asset pricing, which is about the *existence* of risk-neutral measure.

Theorem 5.3.e (First fundamental theorem of asset pricing)**.** If a risk-neutral measure *exists* in a market (having the dynamics suggested by [5.3.3]), then the market has no arbitrage.

Proof. Assume that the market has a risk-neutral measure $\widetilde{\mathbb{P}}$. By Lemma 5.3.c, every discounted portfolio value process $\{D_t X_t\}$ is a $\{\mathcal{F}_t\}$-martingale under $\widetilde{\mathbb{P}}$, which particularly implies that $\widetilde{\mathbb{E}}[D_T X_T] = D_0 X_0 = X_0$. Now, consider any portfolio value process $\{X_t\}$ with $X_0 = 0$. We then have $\widetilde{\mathbb{E}}[D_T X_T] = X_0 = 0$.

Now, assume to the contrary that $\mathbb{P}(X_T \geq 0) = 1$ and $\mathbb{P}(X_T > 0) > 0$ (so arbitrage exists). From this, we know that $\mathbb{P}(X_T < 0) = 0$. Since $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent, we have $\widetilde{\mathbb{P}}(X_T < 0) = 0$, and hence $\widetilde{\mathbb{P}}(X_T \geq 0) = 1$. As we have $\widetilde{\mathbb{E}}[D_T X_T] = 0$, this forces that $\widetilde{\mathbb{P}}(X_T = 0) = 1$ and $\widetilde{\mathbb{P}}(X_T > 0) = 0$; otherwise, we would have $\widetilde{\mathbb{E}}[D_T X_T] > 0$. Applying the equivalence of $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ again gives $\mathbb{P}(X_T > 0) = 0$, contradiction. □

The first fundamental theorem of asset pricing provides a sufficient condition for ensuring that our market model to be arbitrage-free, namely the existence of risk-neutral measure. From the discussion in [5.3.5], we know that the existence of solution to the *market price of risk equations* (existence of those $\{\Theta_t^{(j)}\}$'s) implies that a risk-neutral measure exists, which in turn implies that the market has no arbitrage, by the first fundamental theorem of asset pricing. In short, the existence of solution to the market price of risk equations is a guarantee for having no arbitrage in the market.

5.3.9 **Completeness of market.** As one may expect, there is also a *second* fundamental theorem of asset pricing. This time, it relates the *uniqueness* of risk-neutral measure and the possibility of replicating derivatives. Prior to the discussion of the theorem, we first lay some groundwork about the replication of derivatives. A market is said to be **complete** if every derivative can be replicated (or hedged). Now, let us consider what we need for the market to be complete, by utilizing an argument similar to the one in [5.2.2].

Consider a market where the underlying filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the one generated by the d-dimensional Brownian motion $\{\boldsymbol{W}_t\}_{t\in[0,T]}$. Suppose that we have found a solution to the market price of risk equations, thereby getting a risk-neutral measure $\widetilde{\mathbb{P}}$ by applying the multidimensional Girsanov theorem (and implying the market is arbitrage-free by the first fundamental theorem of asset pricing).

Consider any derivative with time-T payoff V_T, which is $\mathcal{F}_T$-measurable and integrable. Let $\widetilde{M}_t = \widetilde{\mathbb{E}}[D_T V_T|\mathcal{F}_t]$ for all $t \in [0, T]$. Since $\{\widetilde{M}_t\}$ is a Doob martingale under $\widetilde{\mathbb{P}}$, by Theorem 5.3.b we have

$$\widetilde{M}_T = \widetilde{M}_0 + \sum_{j=1}^{d} \int_0^T \widetilde{\Gamma}_t \,\mathrm{d}\widetilde{W}_t^{(j)}$$

where $\{\widetilde{\Gamma}_t^{(1)}\}, \ldots, \{\widetilde{\Gamma}_t^{(d)}\}$ are $\{\mathcal{F}_t\}$-adapted processes.

Next, consider any self-financing portfolio with time-t position on stock i being $\Delta_t^{(i)}$, where $\{\Delta_t^{(i)}\}$ is $\{\mathcal{F}_t\}$-adapted for each $i = 1, \ldots, m$, and the non-stock component being invested in the risk-free bond. Let X_t be the time-t value of the portfolio.

From the proof of Lemma 5.3.c, we have

$$\mathrm{d}(D_tX_t)=\sum_{i=1}^{m}\Delta_t^{(i)}\,\mathrm{d}(D_tS_t^{(i)})\overset{[5.3.5]}{=}\sum_{i=1}^{m}\Delta_t^{(i)}D_tS_t^{(i)}\sum_{j=1}^{d}\sigma_t^{(ij)}\,\mathrm{d}\widetilde{W}_t^{(j)}=\sum_{j=1}^{d}\sum_{i=1}^{m}\Delta_t^{(i)}D_tS_t^{(i)}\sigma_t^{(ij)}\,\mathrm{d}\widetilde{W}_t^{(j)}.$$

This implies that

$$D_TX_T=X_0+\sum_{j=1}^{d}\int_0^T\sum_{i=1}^{m}\Delta_t^{(i)}D_tS_t^{(i)}\sigma_t^{(ij)}\,\mathrm{d}\widetilde{W}_t^{(j)}.$$

Therefore, for this self-financing portfolio to replicate the derivative, we need to set $X_0=\widetilde{M}_0=\widetilde{\mathbb{E}}[D_TV_T|\mathcal{F}_0]$, and the positions $\Delta_t^{(1)},\dots,\Delta_t^{(m)}$ such that the **hedging equations**

$$\frac{\widetilde{\Gamma}_t^{(j)}}{D_t}=\sum_{i=1}^{m}\Delta_t^{(i)}S_t^{(i)}\sigma_t^{(ij)},\quad j=1,\dots,d$$

are satisfied for all $t\in[0,T]$. Note that the hedging equations form a system of d equations in m unknowns $\Delta_t^{(1)},\dots,\Delta_t^{(m)}$. Unlike Proposition 5.2.c, here in general we cannot guarantee that there *are* positions that solve the hedging equations, and so generally we cannot ensure that every derivative can be replicated, i.e., the market may not be complete. Nonetheless, the second fundamental theorem of asset pricing gives us a criterion for the completeness of the market, namely the *uniqueness of risk-neutral measure*.

5.3.10 **Second fundamental theorem of asset pricing.**

Theorem 5.3.f (Second fundamental theorem of asset pricing)**.** Consider a market (having the dynamics suggested by [5.3.3]) that has a risk-neutral measure. The market is complete iff the risk-neutral measure is unique.

Proof. (Sketch) "$\Rightarrow$": Assume the market is complete. Let $\widetilde{\mathbb{P}}_1$ and $\widetilde{\mathbb{P}}_2$ be risk-neutral measures, and fix any $A\in\mathcal{F}_T\overset{[5.3.3]}{=}\mathcal{F}$. Consider a derivative with payoff $V_T=\mathbf{1}_A/D_T$. Due to the completeness of the market, this derivative can be replicated. Let $\{X_t\}$ be the value process of the replicating portfolio. By Lemma 5.3.c, the discounted portfolio value process $\{D_tX_t\}$ is a $\{\mathcal{F}_t\}$-martingale under both $\widetilde{\mathbb{P}}_1$ and $\widetilde{\mathbb{P}}_2$. Hence, we have

$$\widetilde{\mathbb{P}}_1(A)=\widetilde{\mathbb{E}}_1[D_TV_T]\overset{(X_T=V_T)}{=}\widetilde{\mathbb{E}}_1[D_TX_T]\overset{\text{(martingale)}}{=}D_0X_0\overset{\text{(martingale)}}{=}\widetilde{\mathbb{E}}_2[D_TX_T]\overset{(X_T=V_T)}{=}\widetilde{\mathbb{E}}_2[D_TV_T]=\widetilde{\mathbb{P}}_2(A).$$

It follows that $\widetilde{\mathbb{P}}_1=\widetilde{\mathbb{P}}_2$, establishing the uniqueness.

"$\Leftarrow$": Assume the risk-neutral measure $\widetilde{\mathbb{P}}$ is unique. It can then be shown that the underlying filtration $\{\mathcal{F}_t\}$ must be the one generated by the d-dimensional Brownian motion $\{\boldsymbol{W}_t\}$. Knowing this, one can further show that the uniqueness of risk-neutral measure forces that there is only one solution to the market price of risk equations. Now, fix any $t\in[0,T]$ and $\omega\in\Omega$. The market price of risk equations can be expressed as a system of linear equations in matrix-vector form: $A\boldsymbol{x}=\boldsymbol{b}$, where $A=[\sigma_t^{(ij)}(\omega)]_{i,j=1,1}^{m,d}\in\mathbb{R}^{m\times d}$, $\boldsymbol{x}=(\Theta_t^{(1)}(\omega),\dots,\Theta_t^{(d)}(\omega))\in\mathbb{R}^d$, and $\boldsymbol{b}=(\alpha_t^{(1)}(\omega)-r_t(\omega),\dots,\alpha_t^{(m)}(\omega)-r_t(\omega))\in\mathbb{R}^m$. Then, we know that there is a unique solution $\boldsymbol{x}$ to the system.

In a similar way, we express the hedging equations as a system of linear equations in matrix-vector form: $A^T\boldsymbol{y}=\boldsymbol{c}$, where A is defined above, $\boldsymbol{y}=(y_1,\dots,y_m):=(\Delta_t^{(1)}(\omega)S_t^{(1)}(\omega),\dots,\Delta_t^{(m)}(\omega)S_t^{(m)}(\omega))\in\mathbb{R}^m$, and $\boldsymbol{c}=(\widetilde{\Gamma}_t^{(1)}(\omega)/D_t,\dots,\widetilde{\Gamma}_t^{(d)}(\omega)/D_t)\in\mathbb{R}^d$. Based on the unique solution $\boldsymbol{x}$, it can be shown that this system has a solution $\boldsymbol{y}$ for every $\boldsymbol{c}\in\mathbb{R}^d$, meaning that there always exists a solution to the hedging equations, and so every derivative can be replicated, i.e., the market is complete. □

5.3.11 **Relationship between arbitrage/completeness and the market price of risk equations.** Based on [5.3.5] and the proof sketch of Theorem 5.3.f above, we can deduce the following connection between arbitrage/completeness and the existence/uniqueness of solution to the market price of risk equations:

(a) *(Relating arbitrage and existence of solution)* The market price of risk equations has a solution iff the market has no arbitrage.

(b) *(Relating completeness and uniqueness of solution)* If the market price of risk equations have a unique solution, then the market is complete.

References

Etheridge, A. (2002). *A course in financial calculus*. Cambridge University Press.
Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*. Springer.

Concepts and Terminologies

Results

Section 1

Section 2

Section 3

- [3.1.2]: properties of symmetric random walk
- [3.2.3]: properties of Brownian motion
- Proposition 3.2.a: characterizations of Brownian motion
- Proposition 3.3.a: continuous function with finite first-order variation has a zero quadratic variation
- Proposition 3.3.b: Brownian motion accumulates quadratic variation at unit rate
- Proposition 3.3.c: Brownian motion has an infinite first-order variation
- [3.3.6]: differential rules for Brownian motions
- Proposition 3.4.a: Brownian motion is a Markov process
- Proposition 3.5.a: exponential martingale is a martingale
- [3.5.3]: relationship between first passage time and maximum/minimum of Brownian motions
- Theorem 3.5.b: optional stopping theorem
- Proposition 3.5.c: properties about first passage time
- Theorem 3.6.a: reflection equality
- Proposition 3.6.b: distribution and density functions of first passage time
- Proposition 3.6.c: joint distribution and conditional distribution functions of Brownian motion and its maximum

Section 4

- [4.1.3]: properties of Itô integral for simple integrand
- Proposition 4.1.a: approximating sequence for Itô integral
- [4.1.5]: properties of Itô integral for general (adapted and square-integrable) integrand
- Theorem 4.2.a: Itô formula for Brownian motion
- Proposition 4.2.b: quadratic variation for Itô process
- Theorem 4.2.c: Itô formula for Itô process
- [4.2.8]: zero drift term leads to a martingale (and also converse)
- Proposition 4.2.d: normality of Itô integral with nonrandom integrand
- Theorem 4.3.a: Black-Scholes equation
- Proposition 4.4.a: zero cross variation for d-dimensional Brownian motion
- Theorem 4.4.b: two-dimensional Itô formula
- Corollary 4.4.c: Itô product rule
- Theorem 4.4.d: Lévy's characterization theorem
- Theorem 4.4.e: two-dimensional Lévy's characterization theorem

Section 5

- Lemma 5.1.a: effects on conditional expectations after change of measure
- Theorem 5.1.b: Girsanov theorem
- [5.1.5]b: the discounted value process for a self-financing portfolio is a martingale
- [5.1.5]c: risk-neutral pricing formula
- Theorem 5.2.a: martingale representation theorem
- Theorem 5.2.b: martingale representation theorem (risk-neutral version)
- Proposition 5.2.c: construction of replicating portfolio with one stock
- Theorem 5.3.a: multidimensional Girsanov theorem
- Theorem 5.3.b: multidimensional martingale representation theorem
- Lemma 5.3.c: the discounted value process for a self-financing portfolio is a martingale, under market with multiple stocks
- Proposition 5.3.d: criterion for existence of arbitrage
- Theorem 5.3.e: first fundamental theorem of asset pricing
- Theorem 5.3.f: second fundamental theorem of asset pricing
- [5.3.11]a: criterion for having no arbitrage in the market based on existence of solution to the market price of risk equations
- [5.3.11]b: sufficient condition for completeness of market based on uniqueness of solution to the market price of risk equations