

MATH5975 Introduction to Stochastic Analysis

June 2022

CONTENTS

1	General Probability Theory	3
1.1	σ -algebra	3
1.2	Probability Space	3
1.3	Random Variables	4
1.4	Expectation	5
1.5	Conditional Expectation	5
1.5.1	Conditional Expectation w.r.t. \mathcal{A} a Partition	5
1.5.2	General Conditional Expectation	6
1.5.3	Properties of Conditional Expectation	7
1.5.4	Geometric Interpretation of Conditional Expectation	8
1.6	Visual Representation of Expectation	8
2	Stochastic Processes: Preliminaries	10
2.1	Stochastic Processes	10
2.2	Martingales	11
2.3	Local Martingales	11
2.4	Doob Decomposition Theorem	12
2.4.1	Discrete Time: Doob Decomposition	12
2.4.2	Continuous Time: Doob–Meyer Decomposition	12
2.5	Stopping Times	12
3	Brownian Motion	14
3.1	Brownian Motion	14
3.2	Properties of Brownian Motion	14
3.3	Quadratic Variation	15
3.3.1	Total Variation	15
3.3.2	Quadratic Variation	16
3.4	Reflection Principle	17
4	Ito’s Integral	19
4.1	Ito’s Integral for Processes from $L^2_{\mathbb{P}}(W)$	19
4.2	Ito’s Integral for Processes from $L^2_{\mathbb{P}}(W)$	20
4.3	Ito’s Lemma in 1D	20
4.4	Ito’s Lemma in Multidimensional Space	22
5	Semimartingales	24
5.1	Local Martingales	24
5.2	Semimartingales	25
5.2.1	Properties of Semimartingales	25
5.2.2	Itô’s Lemma for Continuous Semimartingales	26
5.3	Recognising a Brownian Motion	26
5.4	Martingale Representation Theorem (for the Brownian Filtration)	27
6	Stochastic Differential Equations	29
6.1	Itô’s Existence and Uniqueness Theorem	29
6.2	Linear SDE	29
6.3	Non-Linear SDE	30
6.4	Stochastic Exponential and Stochastic Logarithmic	30

7	Girsanov Theorem	32
7.1	Change of Measure	32
7.2	Radon–Nikodym Density Process	32
7.3	Abstract Bayes Formula	33
7.4	Girsanov Theorem	34
7.4.1	Linear Drift	34
7.4.2	Stochastic Drift	34
7.4.3	Continuous Semimartingales	35
A	Some Useful Results	37

1 GENERAL PROBABILITY THEORY

1.1 σ -algebra

Definition 1.1.1.

- The set of all possible outcomes of an experiment is called the **state space**, denoted by Ω .
- The family of all possible subsets of Ω , i.e. the power set is denoted by 2^Ω .

Definition 1.1.2 (σ -algebra). A family of subsets $\mathcal{A} \subset 2^\Omega$ is called a σ -algebra if it satisfies

- (i) $\Omega \in \mathcal{A}$, or $\emptyset \in \mathcal{A}$.
- (ii) Closure under complement: If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.
- (iii) Closure under countable unions and intersections: If a sequence of sets $(A_i)_{i \geq 1}$ belongs to \mathcal{A} , then $\bigcup_{i=1}^{\infty} A_i$ (or $\bigcap_{i=1}^{\infty} A_i$, by De Morgan's law¹) also belongs to \mathcal{A} .

The smallest possible σ -algebra is the trivial σ -algebra: $\{\emptyset, \Omega\}$. The largest possible σ -algebra is 2^Ω . Between the two, we have e.g. for a 6-faced dice, $\mathcal{A} = \{\{1\}, \{2, 3, 4, 5, 6\}, \emptyset, \Omega\}$.

Definition 1.1.3 (σ -algebra generated by families of sets).

- Let \mathcal{C} be an arbitrary family of subsets of Ω . We denote by $\sigma(\mathcal{C})$ the smallest σ -algebra which contains every set in \mathcal{C} (i.e. $\mathcal{C} \subseteq \sigma(\mathcal{C})$) and call this the σ -algebra generated by \mathcal{C} .
- (Borel σ -algebra) An important example is the Borel σ -algebra over any topological space Ω , denoted by $\mathcal{B}(\Omega)$, which is the σ -algebra generated by the open sets of Ω (or, equivalently, by the closed sets²). In other words, $\mathcal{B}(\Omega) := \sigma(\mathcal{O}(\Omega))$, where $\mathcal{O}(\cdot)$ denotes the collection of all open sets.
 - Recall that an open set of \mathbb{R} is a subset $E \subseteq \mathbb{R}$ such that for every $x \in E$ there exists $\epsilon > 0$ such that $B_\epsilon(x) = \{y \in \mathbb{R} : |x - y| < \epsilon\}$ is contained in E .
 - A set $F \subseteq \mathbb{R}$ is said to be closed if F^c is open.
 - \mathbb{R} and \emptyset are simultaneously both open and closed sets.
- (σ -algebra generated by a random variable) Given $X : (\Omega, \mathcal{A}) \rightarrow (\Psi, \mathcal{G})$, the σ -algebra generated by X , denoted $\sigma(X)$ is the smallest σ -algebra on Ω such that X is a random variable, that is, X is measurable with respect to $\sigma(X)$ and \mathcal{G} . Equivalently, $\sigma(X) = X^{-1}(\mathcal{G}) = \{X^{-1}(S) \mid S \in \mathcal{G}\}$ (by Definition 1.3.1 and Theorem 1.3.2).

Theorem 1.1.1. The Borel σ -algebra $\mathcal{B}(\mathbb{R})$ is generated by intervals of the form $(-\infty, a]$ where $a \in \mathbb{Q}$ is a rational number.

Proof. Let \mathcal{O} denote the collection of all open intervals. Since every open set in \mathbb{R} is an at most countable union of (disjoint) open intervals, we have $\sigma(\mathcal{O}) = \mathcal{B}(\mathbb{R})$.

Let \mathcal{D} be the collection of all intervals of the form $(-\infty, a]$, $a \in \mathbb{Q}$. For any $a < b$, $a, b \in \mathbb{Q}$,

$$(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}] = \bigcup_{n=1}^{\infty} (-\infty, b - \frac{1}{n}] \cap (-\infty, a]^c,$$

which implies that $(a, b) \in \sigma(\mathcal{D})$, i.e. $\mathcal{O} \subseteq \sigma(\mathcal{D}) \Rightarrow \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{D})$. However, every element of \mathcal{D} is a closed set (complement of an open set) so $\mathcal{D} \subseteq \mathcal{B}(\mathbb{R}) \Rightarrow \sigma(\mathcal{D}) \subseteq \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O})$, completing the proof. ■

1.2 Probability Space

Definition 1.2.1 (measurability).

- Given a state space Ω and a σ -algebra $\mathcal{A} \subseteq 2^\Omega$, the duple (Ω, \mathcal{A}) called a **measurable space**.
- A subset $A \in \Omega$ is said to be \mathcal{A} -measurable if $A \in \mathcal{A}$. We call A an **event**.

Definition 1.2.2 (probability measure). Let Ω be a nonempty set, and let \mathcal{A} be a σ -algebra of subsets of Ω . A probability measure \mathbb{P} is a function that, to every set $A \in \mathcal{A}$, assigns a number in $[0, 1]$. We require:

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

¹ $\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c$.

²To show this, we just need to show that $\sigma(\mathcal{C}) \subseteq \mathcal{B}(\Omega)$ and $\mathcal{B}(\Omega) \subseteq \sigma(\mathcal{C})$. All closed sets are complements of open sets. Since $\mathcal{B}(\Omega)$ being a σ -algebra is closed under complement, it contains all the closed sets, i.e. $\mathcal{C} \subseteq \mathcal{B}(\Omega) \Rightarrow \sigma(\mathcal{C}) \subseteq \mathcal{B}(\Omega)$. A similar argument can be used to show that $\mathcal{O} \subseteq \sigma(\mathcal{C}) \Rightarrow \mathcal{B}(\Omega) = \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{C})$.

(ii) (countable additivity) whenever A_1, A_2, \dots is a countable sequence of disjoint sets in \mathcal{A} , then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **probability space**.

Theorem 1.2.1 (probability measure \mathbb{P} is continuous along monotone sequences of events). *If $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ is a probability measure, then*

- *If $(A_i)_{i \geq 1} \in \mathcal{A}$ is increasing, i.e. $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ with $\lim_{i \rightarrow \infty} A_i = \bigcup_{i=1}^{\infty} A_i = A$, denoted $A_i \uparrow A$, then*

$$\lim_{i \rightarrow \infty} \mathbb{P}(A_i) = \mathbb{P}\left(\lim_{i \rightarrow \infty} A_i\right) = \mathbb{P}(A).$$

- *If $(A_i)_{i \geq 1} \in \mathcal{A}$ is decreasing, i.e. $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ with $\lim_{i \rightarrow \infty} A_i = \bigcap_{i=1}^{\infty} A_i = A$, denoted $A_i \downarrow A$, then*

$$\lim_{i \rightarrow \infty} \mathbb{P}(A_i) = \mathbb{P}\left(\lim_{i \rightarrow \infty} A_i\right) = \mathbb{P}(A).$$

Proof omitted (course notes pp.8–9).

1.3 Random Variables

Let (Ω, \mathcal{A}) and (Ψ, \mathcal{G}) be two measurable spaces (Definition 1.2.1).

Definition 1.3.1 (measurability of a function). A function $X : \Omega \rightarrow \Psi$ is said to be **measurable** with respect to \mathcal{A} and \mathcal{G} if the pre-image $X^{-1}(G) = \{\omega : X(\omega) \in G\} \in \mathcal{A}$ for all $G \in \mathcal{G}$, or equivalently, $X^{-1}(\mathcal{G}) \subseteq \mathcal{A}$.

Definition 1.3.2 (random variable). A **random variable** is a real-valued function $X : \Omega \rightarrow \mathbb{R}$ that is measurable with respect to \mathcal{A} and $\mathcal{B}(\mathbb{R})$. That is, the pre-image $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{A}$ for all Borel sets $B \in \mathcal{B}(\mathbb{R})$. We sometimes simply say X is \mathcal{A} -measurable.

Definition 1.3.3 (almost sure equality of two r.v.s). Two random variables X and Y are said to be **equal almost surely**, written $X \stackrel{\text{a.s.}}{=} Y$ if the set $\{\omega : X(\omega) \neq Y(\omega)\}$ is of probability zero.

The theorem below says, to check if a function is measurable w.r.t. a σ -algebra, it suffices to check the pre-image of the generating family of the σ -algebra.

Theorem 1.3.1 (measurability of a function). *Let \mathcal{C} be a class of subsets of Ψ such that $\sigma(\mathcal{C}) = \mathcal{G}$. In order for a function $X : \Omega \rightarrow \Psi$ to be measurable with respect to \mathcal{A} and \mathcal{G} , it is necessary and sufficient that $X^{-1}(\mathcal{C}) \subseteq \mathcal{A}$.*

Proof. By Definition 1.3.1, we just need to show that $X^{-1}(\mathcal{C}) \subseteq \mathcal{A}$ iff $X^{-1}(\sigma(\mathcal{C})) \subseteq \mathcal{A}$.

The backward direction is straightforward since $\mathcal{C} \subseteq \sigma(\mathcal{C})$. To show the forward direction, suppose $X^{-1}(\mathcal{C}) \subseteq \mathcal{A}$. Define

$$\mathcal{K} = \{B \in \sigma(\mathcal{C}) \mid X^{-1}(B) \in \mathcal{A}\}.$$

It is easy to verify that \mathcal{K} is a σ -algebra (by using results on pre-image of complement, union, and intersection). By assumption, $\mathcal{C} \subseteq \mathcal{K}$ and therefore $\sigma(\mathcal{C}) \subseteq \mathcal{K}$. The definition of \mathcal{K} then implies $X^{-1}(\sigma(\mathcal{C})) \subseteq \mathcal{A}$. ■

Corollary 1.3.1. (from Theorem 1.3.1 and Theorem 1.1.1) *A function $X : \Omega \rightarrow \mathbb{R}$ is a random variable if and only if $\{\omega : X(\omega) \leq a\} = X^{-1}((-\infty, a]) \in \mathcal{A}$ for all $a \in \mathbb{Q}$.*

Example to Corollary 1.3.1. Consider the probability space $([0, 1], \mathcal{B}[0, 1], \mathbb{P})$ where \mathbb{P} is the Lebesgue measure, i.e. $\mathbb{P}[a, b] = b - a$, $0 \leq a \leq b \leq 1$. Consider the identity function, i.e. $X(\omega) = \omega$. To show X is a r.v., just note that

$$X^{-1}((-\infty, a]) = \{\omega : X(\omega) \leq a\} = [0, a] \in \mathcal{B}[0, 1],$$

for all $a \in \mathbb{Q}$, $a \leq 1$.

Corollary 1.3.2. (from Theorem 1.3.1) *Two random variables X and Y are independent iff $\mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b)$ for all $a, b \in \mathbb{Q}$.*

Theorem 1.3.2 (pre-image of σ -algebra on domain is also σ -algebra). *Let $X : \Omega \rightarrow \Psi$ be a mapping. Let \mathcal{G} be a σ -algebra on Ψ . Then $X^{-1}(\mathcal{G})$ is a σ -algebra on Ω .*

Proof. We verify the axioms for a σ -algebra:

- (i) $X^{-1}(\Psi) = \Omega \Rightarrow \Omega \in X^{-1}(\mathcal{G})$ as $\Psi \in \mathcal{G}$.
- (ii) Let $S \in X^{-1}(\mathcal{G})$. Then $S = X^{-1}(G)$ for some $G \in \mathcal{G}$. We know that the pre-image of set difference equals the set difference of the respective pre-images, so

$$\Omega \setminus S = X^{-1}(\Psi) \setminus X^{-1}(G) = X^{-1}(\underbrace{\Psi \setminus G}_{\in \mathcal{G}}) \Rightarrow \Omega \setminus S \in X^{-1}(\mathcal{G}).$$

- (iii) Let $(S_i)_{i \geq 1} \in X^{-1}(\mathcal{G})$. Then $S_i = X^{-1}(G_i)$ for some $G_i \in \mathcal{G}$ for all i . It follows that

$$\bigcup_{i=1}^{\infty} S_i = \bigcup_{i=1}^{\infty} X^{-1}(G_i) = X^{-1}\left(\bigcup_{i=1}^{\infty} G_i\right) \in X^{-1}(\mathcal{G}),$$

where we used the fact that the union of pre-images is the same as the pre-image of the union. ■

1.4 Expectation

Definition 1.4.1 (characteristic function, mgf).

- The **characteristic function** of a r.v. X is given by $\Phi_X(t) = \mathbb{E}[e^{itX}]$.
- The **moment generating function** of a r.v. X is given by $M_X(t) = \mathbb{E}[e^{tX}]$.

Definition 1.4.2 (expectation). Let X be a r.v. on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The **expectation** of X is

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

When X is a simple random variable, i.e. it takes on only finitely many values, it can be written as

$$X = \sum_{i=1}^n x_i \mathbf{1}_{A_i} \iff X(\omega) = \sum_{i=1}^n x_i \mathbf{1}_{A_i}(\omega), \quad \omega \in \Omega$$

where $x_i \in \mathbb{R}$ and $A_i = \{\omega : X(\omega) = x_i\} \in \mathcal{A}$. Its expectation is given by

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(A_i) = \sum_{i=1}^n x_i \mathbb{P}\{X = x_i\}.$$

Definition 1.4.3 (L^1). We say X is **integrable** if $\mathbb{E}(X) < \infty$. The space of integrable r.v.s is denoted by $L^1(\Omega, \mathbb{P})$.

Theorem 1.4.1. The random variable X is integrable if and only if $\mathbb{E}(|X|) < \infty$.

Proof. We can write $X = X^+ - X^-$ where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Then $|X| = X^+ + X^-$. ■

Theorem 1.4.2 (Jensen's inequality). Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function and X a r.v. with $\mathbb{E}|X| < \infty$. Then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

1.5 Conditional Expectation

1.5.1 Conditional Expectation w.r.t. a Partition

Given a simple random variable X taking values x_1, \dots, x_n ,

$$\mathcal{D}(X) = \{D_1^X, \dots, D_n^X\} \quad \text{where } D_i^X = \{X = x_i\}$$

is the partition of Ω associated with X .

Definition 1.5.1 (conditional expectation w.r.t. a partition). For any partition $\mathcal{D} = \{D_1, \dots, D_n\}$ of Ω , the conditional expectation of a simple random variable X is given by

$$\mathbb{E}(X|\mathcal{D}) = \sum_{i=1}^n \mathbb{E}(X|D_i)\mathbf{1}_{D_i} = \underbrace{\sum_{i=1}^n \sum_{j=1}^m x_j \mathbb{P}(D_j^X|D_i)\mathbf{1}_{D_i}}_{\text{a simple } \sigma(\mathcal{D})\text{-measurable r.v.}},$$

where $\mathbb{E}(X|D_i) = \mathbb{E}(X\mathbf{1}_{D_i})\mathbb{P}(D_i)^{-1}$ is the usual conditional expectation, and $\mathcal{D}(X)$ partitions Ω w.r.t. X . If we have two simple random variables X and Y , then

$$\mathbb{E}(X|Y) := \mathbb{E}[X|\mathcal{D}(Y)] = g(Y)$$

for some measurable function g , which, clearly, is $\sigma(Y)$ -measurable.

Result 1.5.1.

- Given a partition \mathcal{D} of Ω , $\mathbb{P}(A|\mathcal{D}) = \mathbb{P}(A|\sigma(\mathcal{D}))$ for all $A \in \mathcal{A}$. More generally, $\mathbb{E}(X|\mathcal{D}) = \mathbb{E}(X|\sigma(\mathcal{D}))$.
- For a simple random variable X , $\sigma(X) = \sigma(\mathcal{D}(X))$.

Proof. For (1), $\mathbb{E}(X|\mathcal{D})$ is clearly $\sigma(\mathcal{D})$ -measurable. Now from the uniqueness of conditional expectation (Definition 1.5.2), it suffices to check that for all $A \in \sigma(\mathcal{D})$,

$$\mathbb{E}[\mathbb{E}(X|\mathcal{D})\mathbf{1}_A] = \mathbb{E}[\mathbf{1}_A X].$$

We can see that this indeed holds as

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A X] &= \mathbb{E}[\mathbb{E}[\mathbf{1}_A X|\mathcal{D}]] && \text{(tower property)} \\ &= \mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathcal{D}]] && \text{(taking out what is known).} \end{aligned}$$

For (2), X is $\sigma(\mathcal{D}(X))$ -measurable, so $\sigma(X) \subseteq \sigma(\mathcal{D}(X))$. On the other hand, $\mathcal{D}(X) \subseteq \sigma(X) \Rightarrow \sigma(\mathcal{D}(X)) \subseteq \sigma(X)$, hence the equality. \blacksquare

1.5.2 General Conditional Expectation

Definition 1.5.2 (conditional expectation w.r.t. a σ -algebra). Let X be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$. Given an arbitrary sub- σ -algebra $\mathcal{G} \subseteq \mathcal{A}$, the **conditional expectation** of X with respect to \mathcal{G} , denoted by $\mathbb{E}[X|\mathcal{G}]$, is the unique integrable random variable satisfying the following conditions:

- (measurability) $\mathbb{E}[X|\mathcal{G}]$ is \mathcal{G} -measurable, i.e. $\mathbb{E}[X|\mathcal{G}]^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{G}$ or $\mathbb{E}[X|\mathcal{G}]^{-1}(B) \in \mathcal{G}$ for all $A \in \mathcal{B}(\mathbb{R})$.
- (partial averaging) For any $A \in \mathcal{G}$, we have

$$\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathcal{G}]]$$

$$\text{or in integral form } \int_A \mathbb{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega).$$

For the definition to make sense, we need to prove the existence and uniqueness of conditional expectation.

Theorem 1.5.1 (existence and uniqueness of conditional expectation). *Let \mathcal{G} be a sub- σ -algebra of \mathcal{A} . Then*

1. (existence) *there exists a conditional expectation $\mathbb{E}[X|\mathcal{G}]$ for any $X \in L^1(\Omega, \mathbb{P})$.*
2. (uniqueness) *any two conditional expectations of $X \in L^1(\Omega, \mathbb{P})$ respective to \mathcal{G} are equal \mathbb{P} -a.s.*

Proof. Existence. Suppose $X \in L^1(\Omega, \mathbb{P})$ then $X^\pm \in L^1(\Omega, \mathbb{P})$. Without loss of generality, we may assume that $X \geq 0$. Define a new probability measure \mathbb{Q} on (Ω, \mathcal{A}) by setting for any $A \in \mathcal{A}$,

$$\mathbb{Q}(A) := \frac{\mathbb{E}[\mathbf{1}_A X]}{\mathbb{E}[X]}$$

The probability \mathbb{Q} is absolutely continuous with respect to \mathbb{P} on \mathcal{A} (i.e. for $A \in \mathcal{A}$, $\mathbb{P}(A) = 0 \Rightarrow \mathbb{Q}(A) = 0$). This implies that $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{G} \subseteq \mathcal{A}$ (Remark following Definition 7.1.1). Therefore from the Radon–Nikodym Theorem, there exists a positive \mathcal{G} -measurable Radon–Nikodym derivative $\eta = d\mathbb{Q}/d\mathbb{P} \in L^1(\Omega, \mathbb{P})$ such that

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A X] &= \mathbb{E}[X]\mathbb{Q}(A) \\ &= \mathbb{E}[X]\mathbb{E}[\eta\mathbf{1}_A] \\ &= \mathbb{E}[\mathbb{E}[X]\mathbb{E}[\eta\mathbf{1}_A|\mathcal{G}]] \\ &= \mathbb{E}[\eta\mathbb{E}[X]\mathbf{1}_A] \end{aligned}$$

then $\eta\mathbb{E}[X]$ is a version of the conditional expectation $\mathbb{E}[X|\mathcal{G}]$.

Uniqueness. Suppose Y and Y' are both \mathcal{G} -conditional expectation of X . Let $G = \{\omega : Y(\omega) > Y'(\omega)\}$ and we assume that $\mathbb{P}(G) > 0$. To this end, we note that

$$G := \{Y - Y' > 0\} = \bigcup_{n=1}^{\infty} \{Y - Y' > \frac{1}{n}\}$$

$$G_n := \{Y - Y' > \frac{1}{n}\} = \bigcup_{j=1}^n \{Y - Y' > \frac{1}{j}\}$$

By Theorem 1.2.1, $G_n \uparrow G \Rightarrow \mathbb{P}(G_n) \uparrow \mathbb{P}(G)$, so there exists $m > 0$ such that $\mathbb{P}(G_m) > 0$.

Since Y and Y' are both \mathcal{G} -conditional expectations, we have by (ii) of Definition 1.5.2 that for every $A \in \mathcal{G}$, in particular G , we have $\mathbb{E}[\mathbf{1}_G Y] = \mathbb{E}[\mathbf{1}_G Y'] \Rightarrow \mathbb{E}[\mathbf{1}_G (Y - Y')] = 0$. But

$$\begin{aligned} \mathbb{E}[\mathbf{1}_G (Y - Y')] &\geq \mathbb{E}[\mathbf{1}_{G_m} (Y - Y')] && (\mathbf{1}_G \geq \mathbf{1}_{G_m}) \\ &\geq \frac{1}{m} \mathbb{P}[G_m] > 0 && (\mathbf{1}_{G_m} = \mathbf{1}_{\{Y - Y' > 1/m\}}) \end{aligned}$$

This is a contradiction and hence $\mathbb{P}(G) = 0$. ■

1.5.3 Properties of Conditional Expectation

Result 1.5.2.

- Consider a constant r.v. $X = c$, i.e. $X(\omega) = c$ for all $\omega \in \Omega$. Then $\sigma(X)$ is the trivial σ -algebra $\mathcal{A}_0 = \{\emptyset, \Omega\}$.
- $\mathbb{E}[\cdot|\mathcal{A}_0] = \mathbb{E}[\cdot]$.

Proof. For (1), the trivial σ -algebra is obviously the smallest. Further, X is measurable w.r.t. \mathcal{A}_0 as

$$X^{-1}(B) = \begin{cases} \emptyset & \text{if } c \notin B \\ \Omega & \text{if } c \in B \end{cases}$$

for all $B \in \mathcal{B}(\mathbb{R})$. As $X^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{A}_0$, by definition $\sigma(X) = \mathcal{A}_0$.

For (2), let X be any random variable on $(\Omega, \mathcal{A}, \mathbb{P})$. The earlier result implies that $\mathbb{E}[X]$ is \mathcal{A}_0 -measurable. Next, we check that

$$\begin{cases} \mathbb{E}(\mathbf{1}_{\emptyset} X) = \mathbb{E}(\mathbf{1}_{\emptyset} \mathbb{E}(X)) = 0 \\ \mathbb{E}(\mathbf{1}_{\Omega} X) = \mathbb{E}(\mathbf{1}_{\Omega} \mathbb{E}(X)) = \mathbb{E}(X) \end{cases}$$

That is, $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X]]$ for all $A \in \mathcal{A}_0$. The required equality follows immediately from the uniqueness of conditional expectation. ■

Theorem 1.5.2 (conditional expectation). *Let X, Y be two integrable random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, and \mathcal{G}, \mathcal{H} be two sub σ -algebras of \mathcal{A} . Then*

1. **(taking out what is known)** *If X is \mathcal{G} -measurable (or equivalently $\sigma(X) \subseteq \mathcal{G}$), then $\mathbb{E}[X|\mathcal{G}] = X$.*
2. **(linearity)** *For $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$.*
3. **(tower property)** *If $\mathcal{H} \subseteq \mathcal{G}$ then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$

in particular, by taking $\mathcal{H} = \{\emptyset, \Omega\}$ to be the trivial σ -algebra, we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}(X).$$

4. *If X is independent of \mathcal{G} in the sense that for all $A \in \sigma(X)$ and $B \in \mathcal{G}$ we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, then*

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

5. *If X is \mathcal{G} -measurable and Y is independent of \mathcal{G} then for any Borel function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ we have*

$$\mathbb{E}[h(X, Y)|\mathcal{G}] = H(X),$$

where $H : \mathbb{R} \rightarrow \mathbb{R}$ is given by the formula $H(x) = \mathbb{E}[h(x, Y)]$. Consider, e.g. X represents the present, Y the future and \mathcal{G} the information generated by past events.

6. (**conditional Jensen's inequality**) Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function and for any σ -algebra $\mathcal{G} \subseteq \mathcal{A}$,

$$g(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[g(X)|\mathcal{G}].$$

Proof. (for 5)

$$\begin{aligned} \mathbb{E}[h(X, Y)|\mathcal{G}] &= \mathbb{E}[h(x, Y)|\mathcal{G}]|_{x=X} && \text{(taking out what is known)} \\ &= \mathbb{E}[h(x, Y)]|_{x=X} && (Y \text{ is independent of } \mathcal{G}) \\ &=: H(x)|_{x=X} \end{aligned}$$

■

Lemma 1.5.1 (Doob's measurability theorem). Let $X : \Omega \rightarrow \Psi$ be a mapping and (Ψ, \mathcal{G}) a measurable space. A function $Y : \Omega \rightarrow \mathbb{R}$ is $\sigma(X)$ -measurable if and only if there exists a \mathcal{G} -measurable function $h : \Psi \rightarrow \mathbb{R}$ s.t. $Y = h(X)$.

Remark 1.5.1. Doob's measurability theorem tells us for a integrable random variable Y , $\mathbb{E}[Y|\sigma(X)]$ which by definition is $\sigma(X)$ -measurable, must be a function of X .

1.5.4 Geometric Interpretation of Conditional Expectation

Theorem 1.5.3. Let X be a square integrable random variable, i.e. $\mathbb{E}[X^2] < \infty$. Then $\mathbb{E}[X|\mathcal{G}]$ is the orthogonal projection of X on $L^2(\Omega, \mathbb{P})$. This means that for every square integrable \mathcal{G} -measurable Z ,

$$\mathbb{E}[(X - Z)^2] \geq \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2]$$

with equality if and only if $Z = \mathbb{E}[X|\mathcal{G}]$.

Remark 1.5.2. One can show that $X - \mathbb{E}(X|\mathcal{G})$ is orthogonal to $\mathbb{E}(X|\mathcal{G})$. That is,

$$\mathbb{E}[(X - \mathbb{E}(X|\mathcal{G})) \cdot \mathbb{E}(X|\mathcal{G})] = 0.$$

Proof. (Sketch) We consider

$$\begin{aligned} \mathbb{E}[(X - Z)^2] &= \mathbb{E}[(X - \mathbb{E}(X|\mathcal{G}) + \mathbb{E}(X|\mathcal{G}) - Z)^2] \\ &= \mathbb{E}[(X - \mathbb{E}(X|\mathcal{G}))^2] + \mathbb{E}[(\mathbb{E}(X|\mathcal{G}) - Z)^2] + \underbrace{2\mathbb{E}[(X - \mathbb{E}(X|\mathcal{G}))(\mathbb{E}(X|\mathcal{G}) - Z)]}_{(*)} \end{aligned}$$

But

$$\begin{aligned} (*) &= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}(X|\mathcal{G}))(\mathbb{E}(X|\mathcal{G}) - Z)|\mathcal{G}]] && \text{(tower property)} \\ &= \mathbb{E}\left[(\mathbb{E}(X|\mathcal{G}) - Z) \underbrace{\mathbb{E}[(X - \mathbb{E}(X|\mathcal{G}))|\mathcal{G}]}_{=\mathbb{E}(X|\mathcal{G}) - \mathbb{E}(X|\mathcal{G})=0}\right] = 0 && \text{(taking out what is known)} \end{aligned}$$

and therefore $\mathbb{E}[(X - Z)^2] \geq \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2]$. ■

1.6 Visual Representation of Expectation

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

The lower Riemann sum is pictured below, with $A_k = \{\omega \in \Omega : y_k \leq X(\omega) < y_{k+1}\}$.

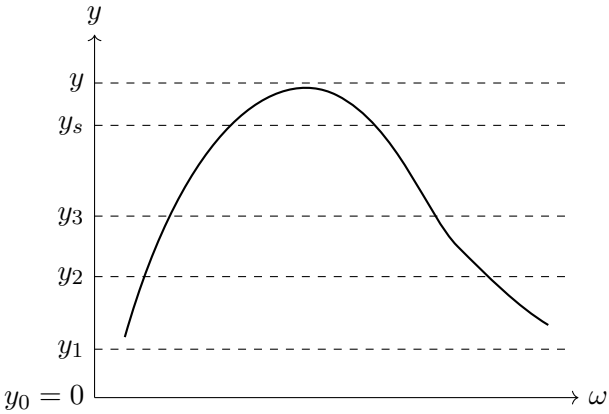


Fig. 1.3.2. Lower Lebesgue sum.

2 STOCHASTIC PROCESSES: PRELIMINARIES

2.1 Stochastic Processes

Definition 2.1.1 (stochastic process). For a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a (real-valued) **stochastic process** $(X_t)_{t \in I}$ is a collection of \mathcal{A} -measurable random variables X_t where $t \in I$ (known as the index set). It can also be written as $(X_{t,\omega})_{t \in I, \omega \in \Omega}$ to reflect that it is actually a function of two variables mapping from $I \times \Omega \rightarrow \mathbb{R}$.

Definition 2.1.2 (filtration, \mathcal{F} -adapted process, natural filtration).

- A **filtration** $\mathcal{F} = (\mathcal{F}_t)_{t \in [0, T]}$ is an increasing family of σ -algebras, that is, $\mathcal{F}_u \subseteq \mathcal{F}_t$ for any $0 \leq u \leq t \leq T$.
- A stochastic process $X = (X_t)_{t \in [0, T]}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is **\mathcal{F} -adapted** if for any $t \in [0, T]$, the random variable X_t is \mathcal{F}_t -measurable, i.e. for any $x \in \mathbb{Q}$, the event $\{X_t \leq x\} \in \mathcal{F}_t$ (Corollary 1.3.1).
- The **natural filtration** of X (or the filtration generated by X) is defined as $\mathcal{F}^X = (\mathcal{F}_t^X)_{t \in [0, T]}$ where $\mathcal{F}_t^X = \sigma(X_u, u \leq t)$. Any stochastic process X is, by definition, adapted to its natural filtration.

Remark 2.1.1. We assume that $\mathcal{A} = \mathcal{F}_T$ and hereafter we shall write $(\Omega, \mathcal{F}, \mathbb{P})$ instead of $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P})$.

Definition 2.1.3 (“equality” of stochastic processes). Two processes X and Y defined on a common probability space are said to be

- (stronger) **indistinguishable** if the event $\{X_t = Y_t, \text{ for all } t \in [0, T]\}$ has probability 1, i.e. $\mathbb{P}(\bigcap_t \{X_t = Y_t\}) = 1$.
- (weaker) **modifications** of each other if for all $t \geq 0$, $\mathbb{P}(X_t = Y_t) = 1$.

Remark 2.1.2.

1. X and Y are indistinguishable $\Rightarrow X$ and Y are modifications of each other, but not the reverse.
2. In discrete time these two concepts coincide.

Proof. We only need to show the converse. Since X and Y are modifications of each other, $\mathbb{P}(X_t = Y_t) = 1$ for all t in some countable set I . Since countable unions of null sets are again null sets³, countable intersections of sets with full measure, have again full measure. Hence, $\mathbb{P}(\bigcap_{t \in I} \{X_t = Y_t\}) = 1$. ■

3. If X and Y two càdlàg processes are modification of each other, then X and Y are indistinguishable.

Examples 2.1.1 (Counterexample of the converse). We give an example where X and Y are modifications of each other, but not distinguishable. Consider the space $([0, 1], \mathcal{B}([0, 1]), \mathbb{P})$, where \mathbb{P} is the Lebesgue measure, i.e. $\mathbb{P}([a, b]) = b - a$. Let $(Y_t)_{t \in [0, 1]}$ denote a constant process given by $Y_t = 0$ and $(X_t)_{t \in [0, 1]}$ given by

$$X_t(\omega) = \begin{cases} 1 & \text{if } t = \omega, \\ 0 & \text{if } t \neq \omega. \end{cases}$$

For a fixed ω , the trajectories $X_t(\omega)$ and $Y_t(\omega)$ differs only at the point $t = \omega$. To see that X and Y are modifications of each other, for every $t \in [0, 1]$, we have

$$\{\omega : X_t(\omega) = Y_t(\omega)\} = \{\omega : \omega \neq t\} = \Omega \setminus \{\omega : \omega = t\}$$

This shows that $\mathbb{P}(X_t = Y_t) = 1 - \mathbb{P}(\{t\}) = 1$. Here the process X is not a right continuous process (càdlàg), therefore one cannot conclude that X and Y are indistinguishable. In fact, we see that

$$\{\omega : X_t(\omega) = Y_t(\omega), \text{ for all } t \in [0, 1]\} = \bigcap_{t \in [0, 1]} \{\omega : X_t(\omega) = Y_t(\omega)\} = \emptyset$$

since the complement is given by

$$\bigcup_{t \in [0, 1]} \{\omega : X_t(\omega) \neq Y_t(\omega)\} = \bigcup_{t \in [0, 1]} \{\omega : \omega = t\} = \Omega$$

which means $(X_t)_{t \in [0, 1]}$ and $(Y_t)_{t \in [0, 1]}$ cannot be indistinguishable.

³Follows from countable additivity of probability measure (Definition 1.2.2).

2.2 Martingales

Definition 2.2.1 (martingales, submartingales, supermartingales). Consider a real-valued, \mathcal{F} -adapted process $M = (M_t)_{t \in [0, T]}$, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If

- (i) M is integrable, that is, $\mathbb{E}|M_t| < \infty$ for $t \in [0, T]$, and
- (ii) (martingale property) for any $0 \leq s \leq t \leq T$,
 - $\mathbb{E}(M_t|\mathcal{F}_s) = M_s$, then M is an \mathcal{F} -**martingale**. The expectation is a constant: $\mathbb{E}(M_t) = \mathbb{E}(M_0)$, for all $t \in [0, T]$.
 - $\mathbb{E}(M_t|\mathcal{F}_s) \geq M_s$, then M is an \mathcal{F} -**submartingale**. The expectation is increasing: $\mathbb{E}(M_t) \geq \mathbb{E}(M_0)$ for any $t \in [0, T]$.
 - $\mathbb{E}(M_t|\mathcal{F}_s) \leq M_s$, then M is an \mathcal{F} -**supermartingale**. The expectation is decreasing: $\mathbb{E}(M_t) \leq \mathbb{E}(M_0)$ for any $t \in [0, T]$.

Remark 2.2.1. (equivalent conditions for discrete time martingales) For a discrete time process, it suffices to show

- $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ for all $n = 0, 1, \dots, N-1$, or
- $\mathbb{E}(M_N|\mathcal{F}_n) = M_n$ for all $n = 0, 1, \dots, N$.

Other ways to prove a martingale in continuous case include:

- Lemma 2.2.1: A sub- or supermartingale with constant expectation.
- $\mathbb{E}(M_T|\mathcal{F}_t) = M_t$ for all $t \in [0, T]$.
- Corollary 3.3.2: $M^2 - \langle M \rangle$ is an \mathcal{F} -martingale if M is a continuous, square integrable \mathcal{F} -martingale.
- Theorem 4.1.1: $I(\gamma)$ for $\gamma \in L^2_{\mathbb{P}}(W)$ is an \mathcal{F} -martingale.
- Theorem 4.2.1: $I(\gamma)$ for $\gamma \in L_{\mathbb{P}}(W)$ is an \mathcal{F} -local martingale.
- Itô's lemma (Theorem 4.3.1) proves a continuous semimartingale when g is sufficiently smooth.

Lemma 2.2.1. Assume that M is either an \mathcal{F} -submartingale or an \mathcal{F} -supermartingale. Then M is an \mathcal{F} -martingale if and only if the expected value of M is constant, i.e. $\mathbb{E}[M_0] = \mathbb{E}[M_t]$ for all $t \in [0, T]$.

Proof. The only if part is straightforward. We prove only the converse.

Suppose M is a submartingale with constant expectation. Let $0 \leq u \leq t \leq T$. Then $\mathbb{E}[M_t|\mathcal{F}_u] - M_u \geq 0$ and

$$\mathbb{E}[\mathbb{E}[M_t|\mathcal{F}_u] - M_u] = \mathbb{E}[M_t] - \mathbb{E}[M_u] = 0$$

by assumption. A non-negative random variable X with $\mathbb{E}(X) = 0$ almost surely, as the Markov inequality implies $\mathbb{P}(X \geq 2^{-n}) \leq 2^n \mathbb{E}(X) = 0$. This implies that $\mathbb{E}[M_t|\mathcal{F}_u] = M_u$ almost surely, i.e. M is an \mathcal{F} -martingale. ■

Lemma 2.2.2 (convex transform of a martingale). Let M be an \mathcal{F} -martingale and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. If the process $X = h(M)$ is integrable, then it is an \mathcal{F} -submartingale.

Examples 2.2.1 (Conditional Jensen's inequality (Theorem 1.5.2)). Conditional Jensen's inequality can come handy when dealing with martingales. For example, if S is a martingale, then the process $(S - K)^+$ is a submartingale, as $f : x \mapsto (x - K)^+$ is convex:

$$\mathbb{E}[f(S_t)|\mathcal{F}_s] \geq f(\mathbb{E}[S_t|\mathcal{F}_s]) = f(S_s),$$

for $0 \leq s \leq t \leq T$. Similarly we can show that (S^2) is also a submartingale.

Lemma 2.2.3 (conditional expectation process is a martingale). Let X be an \mathcal{F}_T -measurable and integrable random variable. Then the process defined by $M_t = \mathbb{E}(X|\mathcal{F}_t)$ for $t \in [0, T]$ is an \mathcal{F} -martingale.

Proof. This is an immediate result of the tower property. ■

2.3 Local Martingales

Definition 2.3.1 (local martingales). A process M is an \mathcal{F} -**local martingale** if there exists an increasing sequence $(\tau_n)_{n \in \mathbb{N}}$ of stopping times such that $\tau_n \rightarrow T$ a.s. (T can be ∞), and for every n the stopped process M^{τ_n} is a uniformly integrable martingale. Any sequence $(\tau_n)_{n \in \mathbb{N}}$ with these properties is called the **reducing sequence** for a local martingale M .

2.4 Doob Decomposition Theorem

2.4.1 Discrete Time: Doob Decomposition

Theorem 2.4.1 (Doob decomposition theorem). *A discrete time submartingale $X = (X_n)_{n=0,1,\dots,N}$ can be decomposed into*

$$X_n = M_n + A_n$$

where M is a martingale and A is a predictable, increasing process with $A_0 = 0$. Similarly, a supermartingale can be decomposed into a martingale and a decreasing process. This decomposition is almost surely unique.

Proof. Existence. We prove the existence of Doob decomposition by construction. Define the processes A and M by setting, for every $n \in I = \{0, 1, \dots, N\}$,

$$A_n = \sum_{k=1}^n (\mathbb{E}[X_k | \mathcal{F}_{k-1}] - X_{k-1})$$

and

$$M_n = X_0 + \sum_{k=1}^n (X_k - \mathbb{E}[X_k | \mathcal{F}_{k-1}]).$$

It is easy to check that $X_n = M_n + A_n$ for all $n \in I$, by writing $X_n - X_0$ as a telescoping sum.

It is clear that A is increasing since $\mathbb{E}[X_k | \mathcal{F}_{k-1}] - X_{k-1} \geq 0$. To check that M is a martingale, we note that

$$\begin{aligned} \mathbb{E}[M_n | \mathcal{F}_{n-1}] &= \mathbb{E}[M_{n-1} + X_n - \mathbb{E}[X_n | \mathcal{F}_{n-1}] | \mathcal{F}_{n-1}] \\ &= M_{n-1} + \mathbb{E}[X_n | \mathcal{F}_{n-1}] - \mathbb{E}[X_n | \mathcal{F}_{n-1}] \\ &= M_{n-1} \end{aligned}$$

for all $n = 1, \dots, N$.

Uniqueness. Let $X = M' + A'$ be an additional decomposition. Then the process $Y := M - M' = A' - A$ is a martingale, implying that

$$\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = Y_{n-1},$$

and also predictable (as A is predictable), implying that

$$\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = Y_n,$$

for any $n = 1, \dots, N$. Since $Y_0 = A'_0 - A_0 = 0$ by the convention about the starting point of the predictable processes, this implies iteratively that $Y_n = 0$ almost surely for all $n = 0, 1, \dots, N$. \blacksquare

2.4.2 Continuous Time: Doob–Meyer Decomposition

Theorem 2.4.2 (Doob–Meyer decomposition theorem). *Let X be a non-negative continuous submartingale⁴. Then there exists a continuous increasing process A with $A_0 = 0$ such that the process $M = X - A$ is a continuous martingale. The processes M and A in the Doob–Meyer decomposition*

$$X = M + A$$

are unique up to indistinguishability of stochastic processes (Definition 2.1.3).

2.5 Stopping Times

Definition 2.5.1 (stopping time). An \mathcal{F} -**stopping time** τ is a random variable taking values in $[0, \infty]$ such that

$$\{\tau \leq t\} \in \mathcal{F}_t$$

for all $t \geq 0$.

⁴A more general version of the theorem applies to any càdlàg supermartingale of class D.

Lemma 2.5.1 (stopped processes in discrete time, p.26).

- Let $X = (X_n)_{n=0,\dots,N}$ be an \mathcal{F} -adapted process and τ be an \mathcal{F} -stopping time. Then the stopped process

$$X^\tau := (X_{n \wedge \tau}) = X_n \mathbf{1}_{n < \tau} + X_\tau \mathbf{1}_{n \geq \tau}$$

is also \mathcal{F} -adapted.

- Let $M = (M_n)_{n=0,\dots,N}$ be an \mathcal{F} -martingale process and τ be an \mathcal{F} -stopping time. Then the stopped process

$$M^\tau := (M_{n \wedge \tau}) = M_n \mathbf{1}_{n < \tau} + M_\tau \mathbf{1}_{n \geq \tau}$$

is also an \mathcal{F} -martingale.

Definition 2.5.2. Given a stopping time τ , the σ -algebra \mathcal{F}_τ defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_T : \text{for all } t \in [0, T], A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$$

is called the σ -algebra of events prior to τ .

Theorem 2.5.1 (Doob's optional sampling theorem). Let $\sigma \leq \tau$ be two stopping times taking values in $\{0, 1, \dots, N\}$. For any martingale, submartingale or supermartingale X , we have

- (i) The random variables X_σ and X_τ are integrable.
- (ii) $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$ (if X is a martingale), $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma$ (submartingale) and $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \leq X_\sigma$ (supermartingale) holds respectively.

Lemma 2.5.2 (a discrete time result). Let $M = (M_n)_{n=0,1,\dots,N}$ be an \mathcal{F} -adapted and integrable stochastic process. Then M is an \mathcal{F} -martingale if and only if $\mathbb{E}(M_\tau) = \mathbb{E}(M_0)$ for any stopping time τ with values in $\{0, 1, \dots, N\}$.

Proof. (\Rightarrow) This is a direct result of Lemma 2.2.1.

(\Leftarrow) By tower property, it suffices to check that $M_n = \mathbb{E}(M_N | \mathcal{F}_n)$ for every $n \in \{0, 1, \dots, N\}$. By assumption, we have $\mathbb{E}(M_\tau) = \mathbb{E}(M_N)$ for any stopping time τ with values in $\{0, 1, \dots, N\}$ (this is true since $\tau = N$ is also a stopping time). Let us fix t and consider an event $A \in \mathcal{F}_t$. Define τ_A as

$$\tau_A = \begin{cases} n, & \text{if } \omega \in A, \\ N, & \text{if } \omega \notin A, \end{cases}$$

then τ_A is an \mathcal{F} -stopping time with values in $\{0, 1, \dots, N\}$ and so $\mathbb{E}(M_{\tau_A}) = \mathbb{E}(M_N)$. This yields $\mathbb{E}(\mathbf{1}_A M_n) = \mathbb{E}(\mathbf{1}_A M_N)$. Since this equality holds for any event $A \in \mathcal{F}_t$, by definition we have $M_n = \mathbb{E}(M_N | \mathcal{F}_n)$, which in turn implies that M is an \mathcal{F} -martingale. ■

3 BROWNIAN MOTION

3.1 Brownian Motion

Definition 3.1.1 (standard Brownian motion). A real-valued process $W = (W_t)_{t \in [0, T]}$, with $W_0 = 0$, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a standard \mathcal{F} -Brownian motion if:

1. W is \mathcal{F} -adapted.
2. For any $0 \leq u \leq t \leq T$, $W_t - W_u$ is independent of \mathcal{F}_u .
3. For any $0 \leq u \leq t \leq T$, $W_t - W_u \sim \mathcal{N}(0, t - u)$.
4. W is sample-paths continuous, i.e. almost all sample paths of W are continuous.

Definition 3.1.2 (Gaussian process). A stochastic process $(X_t)_{t \geq 0}$ is said to be a Gaussian process, if for arbitrary times $0 < t_1 < t_2 < \dots < t_n$, the vector $(X_{t_1}, \dots, X_{t_n})$ follows a multivariate Gaussian distribution.

3.2 Properties of Brownian Motion

Theorem 3.2.1 (properties of Brownian motion). *Let W be an \mathcal{F} -Brownian motion. Then*

1. (**martingale property**) W is a continuous \mathcal{F} -martingale.
2. (**Markov property**) For any bounded Borel measurable function $g : \mathbb{R} \rightarrow \mathbb{R}$ and $0 \leq u \leq t \leq T$,

$$\mathbb{E}[g(W_t) | \mathcal{F}_u] = \mathbb{E}[g(W_t) | \sigma(W_u)] = \underbrace{h(W_u)}_{\text{Doob's measurability theorem}}$$

3. W is a Gaussian process.
4. $\text{Cov}(W_s, W_t) = \mathbb{E}[W_s W_t] = \min(s, t)$ for all $s, t \geq 0$ and in particular $\mathbb{E}[W_t^2] = t$.

Proof. For (1), note that W is an integrable, \mathcal{F} -adapted process with $\mathbb{E}(W_t | \mathcal{F}_u) = \mathbb{E}(W_t - W_u + W_u | \mathcal{F}_u) = W_u$ for any $0 \leq u \leq t \leq T$.

For (2),

$$\begin{aligned} \mathbb{E}[g(W_t) | \mathcal{F}_u] &= \mathbb{E}[g(W_t - W_u + W_u) | \mathcal{F}_u] \\ &= \mathbb{E}[g(W_t - W_u + x) | \mathcal{F}_u] \Big|_{x=W_u} && (W_u \text{ is } \mathcal{F}_u\text{-measurable}) \\ &= \mathbb{E}[g(W_t - W_u + x)] \Big|_{x=W_u} && (\text{independent increments}) \end{aligned}$$

which is a function of W_u and therefore $\sigma(W_u)$ -measurable (Lemma 1.5.1). It follows that

$$\begin{aligned} \mathbb{E}[\mathbb{E}[g(W_t) | \mathcal{F}_u] | \sigma(W_u)] &= \mathbb{E}[g(W_t) | \mathcal{F}_u] \\ \Rightarrow \mathbb{E}[g(W_t) | \sigma(W_u)] &= \mathbb{E}[g(W_t) | \mathcal{F}_u]. \end{aligned}$$

For (3), note that

$$\begin{pmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \underbrace{\begin{pmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{pmatrix}}_{=: y \sim \mathcal{N}}$$

We can then conclude by using the fact that any linear transform of a Gaussian vector is again Gaussian. In particular, $(W_{t_1}, \dots, W_{t_n}) \sim \mathcal{N}(0, A \Sigma A^T)$ where $\Sigma = \text{Var}(y)$.

For (4),

$$\begin{aligned} \text{Cov}(W_s, W_t) &= \mathbb{E}[W_s W_t] \\ &= \mathbb{E}[W_s(W_t - W_s)] + \mathbb{E}[W_s^2] \\ &= \mathbb{E}[W_s] \mathbb{E}[W_t - W_s] + \mathbb{E}[W_s^2] \\ &= s \end{aligned}$$

Similarly when $s \geq t$, $\text{Cov}(W_s, W_t) = t$. Therefore $\mathbb{E}[W_t W_s] = \min(s, t)$. ■

Examples 3.2.1 (Exponential Brownian motion with drift). Let $x, m \in \mathbb{R}$ and $\sigma > 0$ and we set $Y_t = xe^{mt+\sigma W_t}$. The process Y is called an exponential Brownian motion with drift m and variance σ^2 . We first notice that the process Y is adapted to the filtration \mathcal{F} . Next,

$$\mathbb{E}[Y_t] = xe^{mt}\mathbb{E}(e^{\sigma W_t}) = xe^{mt}M_{W_t}(\sigma) = xe^{(m+\sigma^2/2)t}.$$

We want to show that the process Y is an \mathcal{F} -martingale if and only if $m = -\sigma^2/2$. Indeed, for $s \leq t$,

$$\begin{aligned}\mathbb{E}[xe^{mt+\sigma W_t}|\mathcal{F}_s] &= \mathbb{E}[xe^{ms+m(t-s)+\sigma W_s+\sigma(W_t-W_s)}|\mathcal{F}_s] \\ &= Y_s\mathbb{E}[e^{m(t-s)+\sigma(W_t-W_s)}] \\ &= Y_se^{m(t-s)+\sigma^2(t-s)/2}\end{aligned}$$

and the claim follows.

3.3 Quadratic Variation

3.3.1 Total Variation

Definition 3.3.1 (total variation, aka first-order variation). Let $\pi_n = \{t_0 = 0, t_1, \dots, t_{m_n} = T\}$ be a partition of $[0, T]$ with $n(\pi) = m_n$ subintervals. The maximum step size of the partition is denoted by

$$|\pi_n| := \max_{1 \leq k \leq m_n} (t_k - t_{k-1}).$$

For a given partition π_n , we define the corresponding measure of the oscillation of a function f by

$$V_n(f) = \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})|.$$

The total variation of f is defined as

$$V(f) := \lim_{|\pi_n| \rightarrow 0} V_n(f) = \lim_{|\pi_n| \rightarrow 0} \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})|$$

Lemma 3.3.1. Any continuously differentiable function f is a function of finite variation, i.e. $V(f) < \infty$.

Proof.

$$V_n(f) = \sum_{k=1}^{m_n} \left| \int_{t_{k-1}}^{t_k} f'(t) dt \right| \leq \sum_{k=1}^{m_n} \int_{t_{k-1}}^{t_k} |f'(t)| dt = \int_0^T |f'(t)| dt < \infty$$

where we have used the fact that any continuous function on a closed and bounded interval is bounded. By taking the limit as $|\pi_n| \rightarrow 0$, we have proved that any continuously differentiable function f on $[0, T]$ is a function of finite variation. ■

Theorem 3.3.1 (Jordan's decomposition and related results).

- Any bounded increasing function f is a function of finite variation, i.e. $V(f) < \infty$.
- (Jordan's decomposition) If f is a function of finite variation, i.e. $V(f) < \infty$, then $f = f_1 - f_2$ where f_1, f_2 are non-decreasing functions.
- The difference of two bounded increasing functions is a function of finite variation.

Proof. (Sketch) The first result comes from $V(f) = |f(T) - f(0)| < \infty$.

The second result can be proved by construction (omitted).

The third result follows from (1) as well as the fact that the difference of two functions of finite variation also has finite variation (a direct result of triangle inequality). ■

3.3.2 Quadratic Variation

Definition 3.3.2 (quadratic variation). For a given partition $\pi_n = \{t_0 = 0, t_1, \dots, t_{m_n} = T\}$ of $[0, T]$, the quadratic variation of f is defined as

$$\langle f \rangle = V^{(2)}(f) := \lim_{|\pi_n| \rightarrow 0} V_n^{(2)}(f) = \lim_{|\pi_n| \rightarrow 0} \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})|^2$$

Lemma 3.3.2. Any continuously differentiable function f has zero quadratic variation, i.e. $V^{(2)}(f) = 0$.

Proof.

$$\begin{aligned} V_n^{(2)}(f) &= \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})|^2 = \sum_{k=1}^{m_n} \left(\int_{t_{k-1}}^{t_k} f'(t) dt \right)^2 \\ &\leq \sum_{k=1}^{m_n} (t_k - t_{k-1}) \int_{t_{k-1}}^{t_k} |f'(t)|^2 dt \quad (\text{by Cauchy-Schwarz, see Theorem A.0.1}) \\ &\leq |\pi_n| \int_0^T |f'(t)|^2 dt < \infty \end{aligned}$$

Taking the limit as $|\pi_n| \rightarrow 0$, we have $V^{(2)}(f) = \lim_{|\pi_n| \rightarrow 0} V_n^{(2)}(f) = 0$, completing the proof. Alternatively, this can also be shown by combining the results of Lemma 3.3.1 and 3.3.3. \blacksquare

Lemma 3.3.3 (finite variation \Rightarrow zero quadratic variation). Let f be a continuous function of finite variation on $[0, T]$. Then the quadratic variation of f is equal to zero.

Proof. Observe that for any partition $\pi_n = \{t_0 = 0, t_1, \dots, t_{m_n} = T\}$ of $[0, T]$ we have

$$\begin{aligned} V_n^{(2)}(f) &= \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})|^2 \\ &\leq \max_{1 \leq m \leq m_n} |f(t_m) - f(t_{m-1})| \sum_{k=1}^{m_n} |f(t_k) - f(t_{k-1})| \\ &= \max_{1 \leq m \leq m_n} |f(t_m) - f(t_{m-1})| \cdot V_n(f). \end{aligned}$$

By using the fact that any continuous function is uniformly continuous on a closed bounded interval, taking the limit as $|\pi_n| \rightarrow 0$ completes the proof. \blacksquare

Corollary 3.3.1. (from Lemma 3.3.3) If the quadratic variation of a continuous stochastic process is a.s. strictly positive, then the total variation of the process is almost surely infinity. In particular, sample paths of a Brownian motion are a.s. functions of infinite variation on any interval.

Theorem 3.3.2. Let W be a standard Brownian motion. W is a process of finite quadratic variation. In particular, $\langle W \rangle_t = t$ (a.s.) for all $t \geq 0$.

Proof. Let $t \geq 0$ and $\pi_n = \{t_0 = 0, t_1, \dots, t_{n(\pi)} = t\}$ be a partition of $[0, t]$. We want to show that

$$\lim_{n(\pi) \rightarrow \infty} \sum_{k=1}^{n(\pi)} (W_{t_k} - W_{t_{k-1}})^2 =: V_n^{(2)}(W) = t$$

where the convergence is in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, that is,

$$\lim_{n(\pi) \rightarrow \infty} \mathbb{E} \left[\left(\sum_{k=1}^{n(\pi)} (W_{t_k} - W_{t_{k-1}})^2 - t \right)^2 \right] = 0.$$

For brevity of notation, write $\Delta t_k = t_k - t_{k-1}$, $\Delta W_k = W_{t_k} - W_{t_{k-1}}$ and $\theta_k = (\Delta W_k)^2 - \Delta t_k$. Our goal is to prove that $I_n = \mathbb{E}[(\sum \theta_k)^2] \rightarrow 0$ as $n(\pi) \rightarrow \infty$. To show this,

$$I_n = \mathbb{E} \left[\left(\sum_{k=1}^{n(\pi)} \theta_k \right)^2 \right] = \mathbb{E} \left[\sum_{i,j} \theta_i \theta_j \right] = \sum_{i=1}^{n(\pi)} \mathbb{E}(\theta_i^2) + 2 \sum_{i \neq j} \mathbb{E}(\theta_i \theta_j)$$

We can show that the second term is 0 by using independent increments and $\mathbb{E}[(\Delta W_k)^2] = \Delta t_k$. Then

$$\begin{aligned} I_n &= \sum_{k=1}^{n(\pi)} (\mathbb{E}[(\Delta W_k)^4] - 2\Delta t_k \mathbb{E}[(\Delta W_k)^2] + (\Delta t_k)^2) \\ &= 2 \sum_{k=1}^{n(\pi)} (\Delta t_k)^2 \leq 2|\pi_n| \sum_{k=1}^{n(\pi)} \Delta t_k = 2t|\pi_n| \end{aligned}$$

where in the second equality we used the fact the 4th central moment of a $X \sim \mathcal{N}(\mu, \sigma^2)$ r.v. is $\mathbb{E}((X - \mu)^4) = 3\sigma^4$. It is now clear that the quantity I_n tends to zero as n tends to infinity, since $|\pi_n|$ tends to zero. \blacksquare

Lemma 3.3.4. *The process $M = W^2 - \langle W \rangle$ is a continuous \mathcal{F} -martingale.*

Proof. The process M is \mathcal{F} -adapted and integrable since W^2 is \mathcal{F} -adapted and integrable, while the quadratic variation $\langle W \rangle_t = t$ is clearly \mathcal{F} -adapted and integrable.

For $0 \leq u \leq t \leq T$, from the equality $\mathbb{E}(W_t W_u | \mathcal{F}_u) = W_u \mathbb{E}(W_t | \mathcal{F}_u) = W_u^2$, we deduce that

$$\begin{aligned} \mathbb{E}(M_t | \mathcal{F}_u) &= \mathbb{E}(W_t^2 - t | \mathcal{F}_u) \\ &= \mathbb{E}(W_t^2 - W_u^2 + W_u^2 - t | \mathcal{F}_u) \\ &= \mathbb{E}(W_t^2 - W_u^2 | \mathcal{F}_u) + W_u^2 - t \\ &= \mathbb{E}((W_t - W_u)^2 | \mathcal{F}_u) + W_u^2 - t \\ &= W_u^2 - u = M_u. \end{aligned}$$

\blacksquare

Theorem 3.3.3 (alternative definition of quadratic variation for martingales). *Let M be a continuous, square integrable \mathcal{F} -martingale, i.e. $\mathbb{E}(M_t^2) < \infty$ for all $t \geq 0$. Then the quadratic variation $\langle M \rangle$ is the unique continuous, increasing and \mathcal{F} -adapted process with $\langle M \rangle_0 = 0$ such that $M^2 - \langle M \rangle$ give us an \mathcal{F} -martingale.*

Remark 3.3.1. The existence of $\langle M \rangle$ in the theorem above follows from the Doob–Meyer decomposition (Theorem 2.4.2).

We now obtain a generalisation of Lemma 3.3.4:

Corollary 3.3.2. *(from Theorem 3.3.3) If M is a continuous, square integrable \mathcal{F} -martingale, then*

$$M^2 - \langle M \rangle$$

is also an \mathcal{F} -martingale. In particular, if $M_0 = 0$, then $\mathbb{E}[M_t^2 - \langle M \rangle_t] = M_0^2 - \langle M \rangle_0 = 0$ and therefore

$$\text{Var}(M_t) = \mathbb{E}(M_t^2) = \mathbb{E}(\langle M \rangle_t).$$

3.4 Reflection Principle

Let $b > 0$ and B a standard Brownian motion. Define the hitting time

$$T_b = \inf\{s : B_s = b\}.$$

Theorem 3.4.1. *The following results hold (pp.41–44):*

- **(reflection principle)** $\mathbb{P}(T_b < t) = 2\mathbb{P}(B_t > b)$. More precisely, T_b follows a Lévy distribution with parameters $(0, b^2)$ (or equivalently, an inverse Gamma distribution with shape $1/2$ and scale b^2).
- We can therefore show that $\mathbb{P}(T_b < \infty) = 1$ but $\mathbb{E}(T_b) = \infty$.

- The moment generating function of T_b is $e^{-\sqrt{2\lambda}b}$ for $b, \lambda > 0$.
- For any $t \geq 0$, the running supremum $B_t^* = \sup_{s \leq t} B_s$ and $|B_t|$ have the same distribution.
- The joint distribution of $B_t^* = \sup_{s \leq t} B_s$ and B_t is given by

$$\mathbb{P}(B_t \leq x, B_t^* > y) = \mathbb{P}(B_t > 2y - x) = \mathbb{P}(B_t < x - 2y)$$

for every $t \geq 0$, $y \geq 0$ and $x \leq y$.

- The joint density of $B_t^* = \sup_{s \leq t} B_s$ and B_t is given by

$$\frac{2(2y - x)}{t} \cdot \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(2y - x)^2}{2t}\right)$$

for $x \leq y$, $y \geq 0$.

- The drawdown $B_t^* - B_t$ has the same distribution as $|B_t|$.

4 ITO'S INTEGRAL

5

4.1 Ito's Integral for Processes from $L^2_{\mathbb{P}}(W)$

Definition 4.1.1. $L^2_{\mathbb{P}}(W)$ denotes the class of all \mathcal{F} -progressively measurable processes γ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ s.t.

$$\|\gamma\|_W^2 := \mathbb{E} \left[\int_0^T |\gamma_u|^2 du \right] < \infty.$$

Definition 4.1.2 (Ito's integral for general integrands). For any process $\gamma \in L^2_{\mathbb{P}}(W)$, the random variable $I_T(\gamma)$ is called Ito's integral of γ with respect to W over $[0, T]$. We write (notation only, not a formal definition):

$$I_T(\gamma) = \int_0^T \gamma_u \cdot dW_u$$

$$I_t(\gamma) = I_T(\gamma \mathbf{1}_{[0, t]}) = \int_0^t \gamma_u \cdot dW_u$$

Theorem 4.1.1 (properties of Ito's integrals). Let $\gamma \in L^2_{\mathbb{P}}(W)$, i.e. $\mathbb{E} \left[\int_0^T |\gamma_u|^2 du \right] < \infty$. Then:

- (i) **(martingale)** $I(\gamma)$ is a continuous, square integrable \mathcal{F} -martingale.
- (ii) **(linearity)** For constants a and b , $I_t(a\gamma + b\eta) = aI_t(\gamma) + bI_t(\eta)$.
- (iii) **(Ito isometry)** $\mathbb{E}[I_t^2(\gamma)] = \mathbb{E} \left[\int_0^t \gamma_u^2 du \right]$.
- (iv) **(quadratic variation)** The quadratic variation accumulated up to time t by the Ito's integral is

$$\langle I(\gamma) \rangle_t = \int_0^t \gamma_u^2 du.$$

Furthermore, Theorem 3.3.3 says that $I^2(\gamma) - \langle I(\gamma) \rangle$ is an \mathcal{F} -martingale. Corollary 3.3.2 implies that

$$\text{Var}(I_t(\gamma)) = \mathbb{E}[I_t^2(\gamma)] = \mathbb{E}[\langle I(\gamma) \rangle_t].$$

- (v) **(local property)** For any \mathcal{F} -stopping time τ , $I(\gamma \mathbf{1}_{[0, \tau]}) = I_\tau(\gamma)$.

Proof. For (iii), we prove the case for an elementary process $\gamma \in \mathcal{R}$. Since the expected value of the cross terms are zero, i.e. $\mathbb{E}[(W_{t_{j+1}} - W_{t_j})(W_{t_{k+1}} - W_{t_k})] = 0$ for $j \neq k$, we obtain

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=0}^{m-1} \gamma_j (W_{t_{j+1}} - W_{t_j}) \right)^2 \right] &= \mathbb{E} \left[\sum_{j=0}^{m-1} \gamma_j^2 (W_{t_{j+1}} - W_{t_j})^2 \right] \\ &= \mathbb{E} \left[\sum_{j=0}^{m-1} \gamma_j^2 \mathbb{E}[(W_{t_{j+1}} - W_{t_j})^2 | \mathcal{F}_{t_j}] \right] \\ &= \mathbb{E} \left[\sum_{j=0}^{m-1} \gamma_j^2 (t_{j+1} - t_j) \right] \\ &= \|\gamma\|_W^2 \end{aligned}$$

This means that $\hat{I}_T : (\mathcal{K}, \|\cdot\|_W) \rightarrow L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ is an isometry, i.e. a distance preserving transformation. This can be extended to the isometry $I_T : (L^2_{\mathbb{P}}(W), \|\cdot\|_W) \rightarrow L^2(\Omega, \mathcal{F}_T, \mathbb{P})$.⁶

⁵I will skip the definition of Ito's integral for elementary processes, which is required to formally define Ito's integral for more general processes as discussed here.

⁶The extension is done using the fact that the class of simple process \mathcal{K} is a dense subset in the Banach space $L^2_{\mathbb{P}}(W)$ and the space $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ is also complete (since it a Banach space or more specifically a Hilbert space).

For (iv), again we consider an elementary process γ and we can then do a similar proof to that of Theorem 3.3.2. We first compute the quadratic variation accumulated by the Ito integral on one of the subintervals $[t_j, t_{j+1}]$ on which $\gamma_u = \gamma_j$ is constant. For this, we choose partition points

$$t_j = s_0 < s_1 < \cdots < s_m = t_{j+1}$$

and consider

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

As $m \rightarrow \infty$ and the step size $\max_{0 \leq i \leq m-1} (s_{i+1} - s_i)$ approaches zero, the term $\sum_{i=0}^{m-1} (W_{s_{i+1}} - W_{s_i})^2$ converges to the quadratic variation accumulated by Brownian motion between times t_j and t_{j+1} , which is $t_{j+1} - t_j$. Therefore, the limit of the above is

$$\gamma_j^2 (t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}} \gamma_u^2 du.$$

Adding up all these pieces for each of the subintervals $[t_j, t_{j+1}]$ completes the proof. ■

Examples 4.1.1 (Ito's integral for a deterministic integrand). For a deterministic function f , we have

$$X_t = \int_0^t f(s) dW_s \sim \mathcal{N} \left(0, \int_0^t f^2(s) ds \right).$$

The computation of $\mathbb{E}(X_t)$ and $\text{Var}(X_t)$ follows directly from Theorem 4.1.1. Normality follows from the fact that X_t is (the limit of) the sum of independent normal r.v.s:

$$\int_0^t f(s) dW_s = \lim_{m \rightarrow \infty} \sum_{j=0}^{m-1} f(t_j) (W_{t_{j+1}} - W_{t_j}).$$

Examples 4.1.2 (Ito's integral). We can show that $\int_0^t 2W_s dW_s$ is a martingale, by noting that

$$\mathbb{E} \left[\int_0^T |2W_u|^2 du \right] = \int_0^T \mathbb{E}[4W_t^2] dt = \int_0^T 4t dt = 2T^2 < \infty$$

where the first equality uses Fubini's Theorem.

4.2 Ito's Integral for Processes from $L_{\mathbb{P}}(W)$

Definition 4.2.1. $L_{\mathbb{P}}(W)$ denotes the class of all \mathcal{F} -progressively measurable processes γ on $(\Omega, \mathcal{F}, \mathbb{P})$ s.t.

$$\mathbb{P} \left(\int_0^T |\gamma_u|^2 du < \infty \right) = 1.$$

By applying the optional stopping technique (aka localisation), we can extend the definition of Ito's integral to the more general class $L_{\mathbb{P}}(W)$. Details are omitted, the important conclusion is that:

Theorem 4.2.1. *If $\gamma \in L_{\mathbb{P}}(W)$, i.e. $\mathbb{P} \left(\int_0^T |\gamma_u|^2 du < \infty \right) = 1$, then $I(\gamma)$ is a continuous \mathcal{F} -local martingale.*

4.3 Ito's Lemma in 1D

Ito processes are a special class of continuous semimartingales.

Definition 4.3.1 (Ito process). An \mathcal{F} -adapted continuous process X is called an Ito process if it admits a representation

$$X_t = X_0 + \int_0^t \alpha_u du + \int_0^t \beta_u \cdot dW_u, \quad \text{for all } t \in [0, T]$$

for some \mathcal{F} -adapted processes α and β that are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and satisfy suitable integrability conditions. It is customary to write the integral formula above using differential notation as

$$dX_t = \alpha_t dt + \beta_t \cdot dW_t.$$

Lemma 4.3.1 (quadratic variation of Ito processes). *Consider an Ito process denoted by $dX_t = \alpha_t dt + \beta_t \cdot dW_t$.*

(i) *The quadratic variation of the Ito process is $\langle X \rangle_t = \int_0^t \beta_u^2 du$.*

(ii) *The quadratic covariation (aka cross-variation) of two Ito processes $X_t = \int_0^t \alpha_u dW_u$ and $Y_t = \int_0^t \beta_u dW_u$ is*

$$\langle X, Y \rangle_t = \int_0^t \alpha_u \beta_u du.$$

(iii) (**polarisation formula**) *A more general version of (ii) gives the quadratic covariation of two continuous semimartingales. If $X^i = X_0^i + M^i + A^i$ and $X^j = X_0^j + M^j + A^j$ are in $\mathcal{S}^c(\mathbb{P})$, then*

$$\langle X^i, X^j \rangle = \langle M^i, M^j \rangle = \frac{1}{2} (\langle M^i + M^j \rangle - \langle M^i \rangle - \langle M^j \rangle) = \frac{1}{4} (\langle M^i + M^j \rangle - \langle M^i - M^j \rangle).$$

Proof. (i) First note that the process $A_t = \int_0^t \alpha_u du$ is always a process of finite variation (thus zero quadratic variation). This is clear when $\alpha \geq 0$ and A_t is increasing, since the total variation on $[0, T]$ is just A_T (Theorem 3.3.1). If α is not positive, we can write $\alpha = \alpha^+ - \alpha^-$ and $A_t = A_t^+ - A_t^-$ where $\alpha^+ = \max(\alpha, 0)$, $\alpha^- = \max(-\alpha, 0)$ and

$$A_t^\pm = \int_0^t \alpha_s^\pm ds$$

and it is not difficult to see that the total variation of A is bounded by $A_T^+ + A_T^-$. The quadratic variation of the process $B_t = \int_0^t \beta_u \cdot dW_u$ is given in Theorem 4.1.1, thus the result.

(ii) We can use the polarisation formula:

$$\begin{aligned} \langle X, Y \rangle_t &= \frac{1}{2} (\langle X + Y \rangle_t - \langle X \rangle_t - \langle Y \rangle_t) \\ &= \frac{1}{2} \left(\left\langle \int_0^t (\alpha_u + \beta_u) dW_u, \int_0^t (\alpha_u + \beta_u) dW_u \right\rangle_t - \left\langle \int_0^t \alpha_u dW_u, \int_0^t \alpha_u dW_u \right\rangle_t - \left\langle \int_0^t \beta_u dW_u, \int_0^t \beta_u dW_u \right\rangle_t \right) \\ &= \frac{1}{2} \left(\int_0^t (\alpha_u + \beta_u)^2 ds - \int_0^t \alpha_u^2 ds - \int_0^t \beta_u^2 ds \right) \\ &= \int_0^t \alpha_u \beta_u ds, \end{aligned}$$

which concludes the proof. ■

Theorem 4.3.1 (Ito's Lemma, 1D). *Suppose that $g : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is a function of class $C^{2,1}([0, T] \times \mathbb{R}, \mathbb{R})$, i.e. g is twice continuously differentiable. Then for any Ito process X , the process $Y_t = g(t, X_t)$, $t \in [0, T]$, is an Ito process (and therefore also a continuous semimartingale). Moreover, its canonical decomposition is given by the Ito formula*

$$dg(t, X_t) = g_t(t, X_t) dt + g_x(t, X_t) dX_t + \frac{1}{2} g_{xx}(t, X_t) d\langle X \rangle_t.$$

or equivalently,

$$g(t, X_t) = g(0, X_0) + \int_0^t g_t(s, X_s) ds + \int_0^t g_x(s, X_s) dX_s + \frac{1}{2} \int_0^t g_{xx}(s, X_s) d\langle X \rangle_s.$$

If $dX_t = \alpha_t dt + \beta_t dW_t$, then

$$dg(t, X_t) = g_t(t, X_t) dt + g_x(t, X_t) \alpha_t dt + g_x(t, X_t) \beta_t dW_t + \frac{1}{2} g_{xx}(t, X_t) \beta_t^2 dt.$$

Examples 4.3.1 (Ito's lemma, 1D). We can use Ito formula to determine whether a process is a martingale (no drift), submartingale (drift > 0) or supermartingale (drift < 0).

1. $f(W_t) = (W_t)^2$. Then we have for all $t \in [0, T]$,

$$df(W_t) = 2W_t dW_t + \frac{1}{2} \cdot 2 dt \implies W_t^2 = W_0^2 + \int_0^t 2W_s dW_s + \int_0^t ds = \int_0^t 2W_s dW_s + t.$$

This also gives the Doob–Meyer decomposition (Theorem 2.4.2) of the positive submartingale W^2 . Also recall that $W_t^2 - t = \int_0^t 2W_s dW_s$ is a martingale (e.g. Theorem 3.3.3, or Example 4.1.2).

2. $Y_t = g(t, W_t) = e^{\mu t + W_t}$. Let $g(t, x) = e^{\mu t + x}$ and then $g_t = \mu Y_t$, $g_x = Y_t$, $g_{xx} = Y_t$ and so

$$\begin{aligned} dY_t &= \mu Y_t dt + Y_t dW_t + \frac{1}{2} \cdot Y_t d\langle W \rangle_t = \left(\mu + \frac{1}{2} \right) Y_t dt + Y_t dW_t \\ \implies Y_t &= Y_0 + \underbrace{\int_0^t \left(\mu + \frac{1}{2} \right) Y_u du}_{\text{increasing}} + \underbrace{\int_0^t Y_u dW_u}_{\text{a martingale, as } Y \in L^2_{\mathbb{P}}(W)}. \end{aligned}$$

Therefore, we conclude that Y is a submartingale.

3. (integration by parts, Corollary 4.4.1) $Z_t = \sin(W_t) \cos(W_t)$. Then

$$dZ_t = \sin W_t d(\cos W_t) + \cos W_t d(\sin W_t) + d\langle \sin W, \cos W \rangle_t.$$

By Ito's lemma,

$$\begin{aligned} d(\cos W_t) &= -\sin W_t dW_t - \frac{1}{2} \cos W_t dt, \\ d(\sin W_t) &= \cos W_t dW_t - \frac{1}{2} \sin W_t dt. \end{aligned}$$

Also $d\langle \sin W, \cos W \rangle_t = -\sin W_t \cos W_t dt$, so

$$dZ_t = -\sin^2(W_t) dW_t + \cos^2(W_t) dW_t - 2 \sin W_t \cos W_t dt$$

The sign of the drift term is not always positive or negative, and therefore Z is none of a martingale, submartingale or supermartingale.

4.4 Ito's Lemma in Multidimensional Space

Definition 4.4.1 (Ito's integral, multidimensional). Let $\gamma \in L_W(\mathbb{P})$, that is, γ is an \mathbb{R}^d -valued \mathcal{F} -progressively measurable process satisfying

$$\mathbb{P} \left(\int_0^T |\gamma_u|^2 du < \infty \right) = 1,$$

where $|\cdot|$ stands for the Euclidean norm in \mathbb{R}^d . Then the Ito stochastic integral of γ w.r.t. W is well defined and for any $t \in [0, T]$,

$$I_t(\gamma) = \int_0^t \gamma_u \cdot dW_u = \sum_{i=1}^d \int_0^t \gamma_u^i dW_u^i.$$

Definition 4.4.2 (Ito process, multidimensional). $X = (X^1, \dots, X^k)$ is a k -dimensional Ito process if

$$X_t^i = X_0^i + \int_0^t \alpha_u^i du + \int_0^t \beta_u^i \cdot dW_u$$

where α^i are real-valued processes and $\beta^i = (\beta^{i1}, \dots, \beta^{id})$ are \mathbb{R}^d -valued processes for $i = 1, \dots, k$. It is implicitly assumed in the above that the processes $\alpha^i, \beta^i, i = 1, \dots, k$ are integrable in a suitable sense.

Lemma 4.4.1 (quadratic variation of k -dimensional Ito processes). For $X = (X^1, \dots, X^k)$ defined in Definition 4.4.2, for any $t \in [0, T]$,

$$\langle X^i, X^j \rangle_t = \int_0^t \beta_u^i \cdot \beta_u^j du = \int_0^t \sum_{l=1}^d \beta_u^{il} \beta_u^{jl} du.$$

Theorem 4.4.1 (Ito's Lemma, multidimensional). Suppose that g is a function of class $C^2(\mathbb{R}^k, \mathbb{R})$, i.e. g is twice

continuously differentiable, and that X a k -dimensional Ito process. Then

$$\begin{aligned} dg(X_t) &= \sum_{i=1}^k g_{x_i}(X_t) dX_t^i + \frac{1}{2} \sum_{i,j=1}^k g_{x_i x_j}(X_t) \beta_t^i \cdot \beta_t^j dt \\ &= \sum_{i=1}^k g_{x_i}(X_t) \alpha_t^i dt + \sum_{i=1}^k g_{x_i}(X_t) \beta_t^i \cdot dW_t + \frac{1}{2} \sum_{i,j=1}^k g_{x_i x_j}(X_t) \beta_t^i \cdot \beta_t^j dt. \end{aligned}$$

Corollary 4.4.1 (integration by parts formula). *Given two continuous semimartingales X and Y , then*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_u dY_u + \int_0^t Y_u dX_u + \langle X, Y \rangle_t$$

or equivalently,

$$dX_t Y_t = X_t dY_t + Y_t dX_t + d\langle X, Y \rangle_t.$$

If at least one of X or Y is a process of finite variation, then

$$dX_t Y_t = X_t dY_t + Y_t dX_t.$$

Examples 4.4.1 (Ito's lemma, multidimensional). Let

$$R_t(d) := \sqrt{(W_t^1)^2 + (W_t^2)^2 + \dots + (W_t^d)^2}$$

where $W = (W^1, W^2, \dots, W^d)$ is a standard d -dimensional Brownian motion for $d \geq 3$. The process R is called a Bessel process. We want to show that $1/R_t(3)$ for $1 \leq t \leq T$ is a local martingale but not a martingale.

1. To show it is a local martingale, we apply Ito's lemma to compute the dynamics of $1/R_t(3)$ (and show that there is no drift term). Let

$$Y_t = R_t^2(3) = (W_t^1)^2 + (W_t^2)^2 + (W_t^3)^2 \implies dY_t = \sum_{i=1}^3 2W_t^i dW_t^i + 3 dt.$$

But $1/R_t(3) = 1/\sqrt{Y_t} = g(Y_t)$ where $g(x) = 1/\sqrt{x}$. Strictly speaking, one cannot apply directly Ito's formula as g is not differentiable at zero. But for $d \geq 3$, the d -dimensional Brownian motion never return to the origin after time 0. So (heuristically) one can apply Ito formula again to obtain that

$$\begin{aligned} d(1/R_t(3)) &= -\frac{1}{2} \frac{1}{Y_t^{3/2}} dY_t + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{Y_t^{5/2}} d\langle Y \rangle_t \\ &= -\frac{1}{2} \frac{1}{Y_t^{3/2}} \left(\sum_{i=1}^3 2W_t^i dW_t^i + 3 dt \right) + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{Y_t^{5/2}} (4Y_t dt) \\ &= \frac{1}{R_t^3} \sum_{i=1}^3 W_t^i dW_t^i = \frac{1}{R_t^3} W_t \cdot dW_t \end{aligned}$$

as $d\langle Y \rangle_t = \sum_{i=1}^3 (2W_t^i)^2 dt = 4R_t^2 dt$. This is not a true martingale as $\|1/R_t^3 \cdot W_t\|_W^2 = \infty$.

2. We can also show that

$$\mathbb{E}[1/R_t(d)] = \frac{1}{\sqrt{t}} \mathbb{E} \left[\frac{1}{\sqrt{X}} \right] = \sqrt{\frac{2}{\pi t}}$$

where $X \sim \chi_3^2$. As the expectation is decreasing in t , $1/R_t(3)$ is indeed a supermartingale (see also Theorem 5.1.1).

5 SEMIMARTINGALES

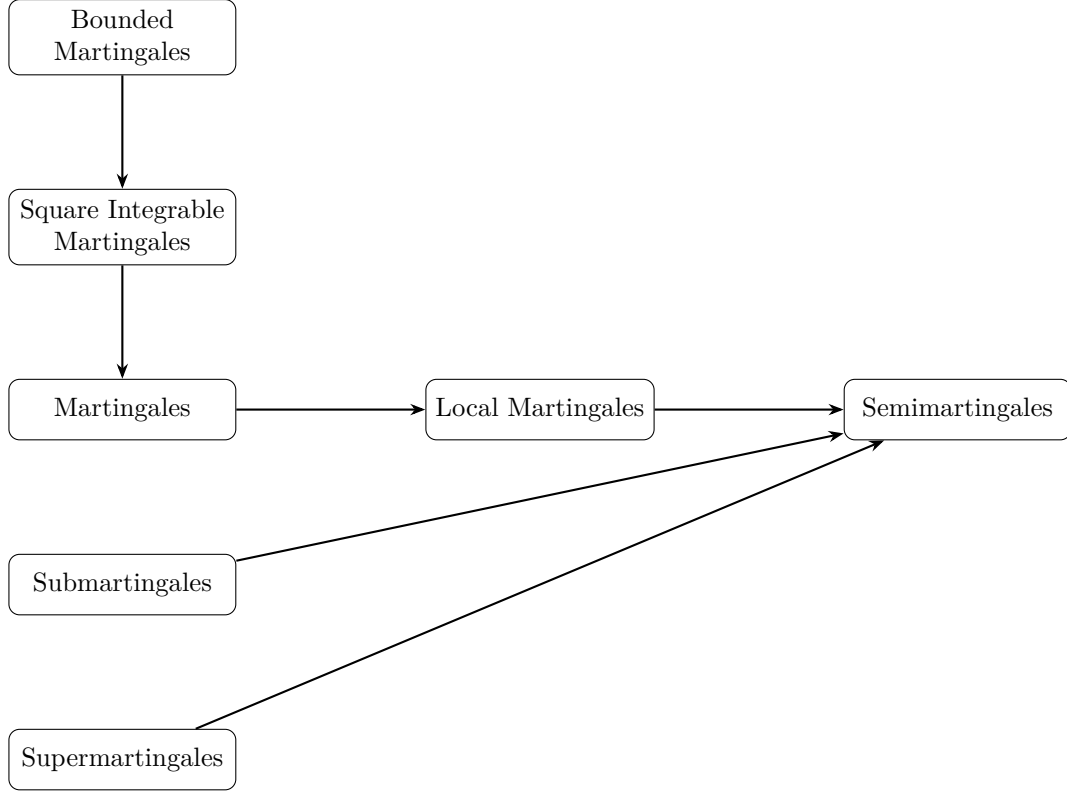


Figure 1: The family of martingales. An arrow from A to B indicates that A is subclass of B .

5.1 Local Martingales

We introduced local martingales in Definition 2.3.1 and Theorem 4.2.1. In this section, we state a few important results about local martingales.

Theorem 5.1.1 (local martingales).

- (i) Any non-negative \mathcal{F} -local martingale M is an \mathcal{F} -supermartingale⁷. If, in addition, $\mathbb{E}(M_t)$ is constant then M is an \mathcal{F} -martingale.
- (ii) (corollary of above) Let M be a non-negative \mathcal{F} -local martingale. If $M_0 = 0$ then $M_t = 0$ for every $t \in [0, T]$.
- (iii) (quadratic variation) Let M be a continuous \mathcal{F} -local martingale. Then the quadratic variation $\langle M \rangle$ is the unique continuous, increasing and \mathcal{F} -adapted process with $\langle M \rangle_0 = 0$ such that the process $M^2 - \langle M \rangle$ is a continuous \mathcal{F} -local martingale.

Remark 5.1.1. The quadratic variation is invariant with respect to an \mathcal{F}_0 -measurable shift of M ; specifically, if $N = \psi + M$ for some \mathcal{F}_0 -measurable random variable ψ is a continuous local martingale, then $\langle N \rangle = \langle M \rangle$. In particular, $\langle M \rangle = \langle M - M_0 \rangle$.

- (iv) Let M be a continuous \mathcal{F} -local martingale s.t. $\langle M \rangle_t = 0$ for $t \in [0, T]$. Then $M_t = M_0$ for every $t \in [0, T]$.
- (v) Let M be a continuous \mathcal{F} -local martingale of finite variation⁸. Then $M_t = M_0$ for every $t \in [0, T]$.
- (vi) (corollary of above) If $M_0 = 0$ and M is a continuous \mathcal{F} -local martingale of finite variation, then M vanishes (more precisely, it is indistinguishable from the null process).

Proof. (ii) Assume that $M_{t_0} \neq 0$ for some $t_0 \in [0, T]$. Then $\mathbb{E}(M_{t_0}) > 0 = \mathbb{E}(M_0)$, which contradicts the property that M is a supermartingale.

(iv) We may assume, without loss of generality, that M is a continuous bounded martingale and $M_0 = 0$. This is because one can always take $M' = M - M_0$ and stop the continuous local martingale M' such that it is bounded.

⁷The assumption that M is non-negative can be replaced by the assumption that $M \geq \eta$ for some random variable η with $\mathbb{E}(\eta) > -\infty$.

⁸An \mathcal{F} -adapted stochastic process $X = (X_t)_{t \in [0, T]}$ is said to be of finite variation if almost all sample paths of X are functions of finite variation on $[0, T]$.

Then

$$\mathbb{E}[M_t^2 - \langle M \rangle_t] = M_0^2 - \langle M \rangle_0 = 0 \implies \mathbb{E}[M_t^2] = \mathbb{E}[\langle M \rangle_t] = 0$$

for any $t \in [0, T]$ and thus $M_t = 0$ for any $t \in [0, T]$.

(v) It suffices to apply (iv) and to observe that the assumption that sample paths of M are continuous functions of finite variation implies that the quadratic variation of M vanishes. \blacksquare

Examples 5.1.1 (Quadratic variation). Recall that $d(W_t^2) = 2W_t dW_t + dt$. This implies

$$W_t^2 = W_0^2 + \int_0^t 2W_s dW_s + \int_0^t ds$$

$$W_t^2 - t = W_0^2 + \int_0^t 2W_s dW_s$$

where the RHS is an \mathcal{F} -martingale. We therefore conclude that $\langle W \rangle_t = t$.

5.2 Semimartingales

5.2.1 Properties of Semimartingales

Definition 5.2.1 (continuous semimartingale). A real-valued, continuous, \mathcal{F} -adapted process X is called a (real-valued) **continuous semimartingale** if it admits a (canonical) decomposition

$$X_t = X_0 + M_t + A_t, \quad \forall t \in [0, T],$$

where X_0 , M and A satisfy:

- (i) X_0 is an \mathcal{F}_0 -measurable random variable.
- (ii) M is a continuous local martingale with $M_0 = 0$.
- (iii) A is a continuous process whose almost all sample paths are of finite variation on the interval $[0, T]$ with $A_0 = 0$.

We denote by $\mathcal{S}^c(\mathbb{P})$ the class of all real-valued continuous semimartingales on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A continuous semimartingale is a continuous local martingale if and only if the process A in its canonical decomposition $X = X_0 + M + A$ vanishes, that is, is indistinguishable from the null process.

Examples 5.2.1 (Semimartingale). Consider the drawdown process $X_t = W_t^* - W_t$ where W_t is a standard Brownian motion and $W_t^* = \sup_{s \leq t} W_s$ is the running supremum. Then X is a semimartingale. In particular, X is a submartingale as $A_t = W_t^*$ is an increasing process.

Theorem 5.2.1 (uniqueness of semimartingale decomposition). *Let X be a continuous semimartingale with the decomposition $X = X_0 + M + A$ so that $M_0 = A_0 = 0$. If X admits also a decomposition $X = X_0 + \widetilde{M} + \widetilde{A}$ for some continuous local martingale \widetilde{M} with $\widetilde{M}_0 = 0$ and some continuous process \widetilde{A} of finite variation on $[0, T]$ with $\widetilde{A}_0 = 0$ then $M_t = \widetilde{M}_t$ and $A_t = \widetilde{A}_t$ for $t \in [0, T]$.*

Proof. It is enough to observe that a difference of two continuous local martingales is also a continuous local martingale, and a difference of two continuous processes of finite variation also follows a continuous process of finite variation. In our case, we have $M - \widetilde{M} = \widetilde{A} - A$ and $M_0 - \widetilde{M}_0 = \widetilde{A}_0 - A_0 = 0$.

Consequently, in view of (vi) of Theorem 5.1.1, any continuous local martingale of finite variation starting at 0 at time 0 is a null process. We therefore conclude that $M_t = \widetilde{M}_t$ and $A_t = \widetilde{A}_t$ for every $t \in [0, T]$. \blacksquare

Theorem 5.2.2 (quadratic variation of semimartingales). *If $X = X_0 + M + A \in \mathcal{S}^c(\mathbb{P})$, then the quadratic variation $\langle X \rangle = \langle M \rangle$. More generally, if $X^i = X_0^i + M^i + A^i$ and $X^j = X_0^j + M^j + A^j$ are in $\mathcal{S}^c(\mathbb{P})$, then*

$$\langle X^i, X^j \rangle = \langle M^i, M^j \rangle = \frac{1}{2} (\langle M^i + M^j \rangle - \langle M^i \rangle - \langle M^j \rangle) = \frac{1}{4} (\langle M^i + M^j \rangle - \langle M^i - M^j \rangle).$$

5.2.2 Itô's Lemma for Continuous Semimartingales

Theorem 5.2.3 (Itô's Lemma). *If $X = X_0 + M + A$ is a real-valued continuous semimartingale, and g is a function of class $C^{2,1}([0, T] \times \mathbb{R}, \mathbb{R})$, then the process $Y_t = g(t, X_t)$ follows a continuous semimartingale with the following canonical decomposition*

$$dY_t = g_t(t, X_t) dt + g_x(t, X_t) dX_t + \frac{1}{2} g_{xx}(t, X_t) d\langle M \rangle_t.$$

Examples 5.2.2 (Itô's lemma for semimartingales). We again consider the drawdown process $X_t = W_t^* - W_t$ where W_t is a standard Brownian motion and $W_t^* = \sup_{s \leq t} W_s$ is the running supremum. We note that X is not an Itô process, but is a continuous semimartingale with $M_t = -W_t$, $A_t = W_t^*$ in its canonical decomposition. By Itô's lemma, we can compute, e.g.

$$\begin{aligned} dX_t^2 &= 2X_t dX_t + \frac{1}{2} \cdot 2 d\langle X \rangle_t \\ &= -2X_t dW_t + 2X_t dW_t^* + d\langle -W \rangle_t \\ &= -2X_t dW_t + 2X_t dW_t^* + dt. \end{aligned}$$

5.3 Recognising a Brownian Motion

In some instances, it would be convenient to have the possibility of checking whether a given process is a standard Brownian motion by establishing its martingale property and by computing its quadratic variation. We may use the martingale characterisation of a Brownian motion, due to Paul Lévy.

Theorem 5.3.1 (Lévy, 1D). *Let M be a continuous \mathcal{F} -local martingale such that $M_0 = 0$ and $\langle M \rangle_t = t$ for every $t \in [0, T]$ (i.e. the process $M_t^2 - t$ is a continuous \mathcal{F} -local martingale). Then M is a standard \mathcal{F} -Brownian motion. Under the stronger assumption that M is a square integrable continuous \mathcal{F} -martingale, this property can be represented as follows*

$$\mathbb{E}[M_t^2 - M_u^2 \mid \mathcal{F}_u] = t - u, \quad u \leq t \leq T,$$

or equivalently (in view of the martingale property of M , see proof of Lemma 3.3.4)

$$\mathbb{E}((M_t - M_u)^2 \mid \mathcal{F}_u) = t - u, \quad u \leq t \leq T.$$

Proof. Suppose M satisfies the assumptions of Theorem 5.3.1. We want to show that M is an \mathcal{F} -Brownian motion. Let $\lambda \in \mathbb{R}$. We apply the Itô formula theorem to $F(x) = e^{i\lambda x}$. Then $F_x(\lambda, x) = i\lambda F(\lambda, x)$, $F_{xx}(\lambda, x) = -\lambda^2 F(\lambda, x)$.

$$\begin{aligned} F(\lambda, M_t) &= F(\lambda, M_s) + i\lambda \int_s^t e^{i\lambda M_u} dM_u - \frac{1}{2} \lambda^2 \int_s^t e^{i\lambda M_u} du \\ \implies e^{i\lambda(M_t - M_s)} &= 1 + i\lambda \int_s^t e^{i\lambda(M_u - M_s)} dM_u - \frac{1}{2} \lambda^2 \int_s^t e^{i\lambda(M_u - M_s)} du, \end{aligned}$$

for all $s \leq t$. By the martingale property of stochastic integrals,

$$\mathbb{E} \left[\int_s^t e^{i\lambda M_u} dM_u \mid \mathcal{F}_s \right] = 0$$

Then

$$\mathbb{E}[e^{i\lambda(M_t - M_s)} \mid \mathcal{F}_s] = 1 - \frac{1}{2} \lambda^2 \int_s^t \mathbb{E}[e^{i\lambda(M_u - M_s)} \mid \mathcal{F}_s] du.$$

If we write $g(t) = \mathbb{E}[e^{i\lambda(M_t - M_s)} \mid \mathcal{F}_s]$, then the above translates to the following differential equation (in integral form)

$$g(t) = 1 - \frac{1}{2} \lambda^2 \int_s^t g(u) du$$

to which $g(t) = e^{-\frac{\lambda^2}{2}(t-s)}$ is the unique solution, i.e.

$$\mathbb{E}[e^{i\lambda(M_t - M_s)} \mid \mathcal{F}_s] = e^{-\frac{\lambda^2}{2}(t-s)}$$

and therefore the random variable $M_t - M_s$ is independent of \mathcal{F}_s and has the normal $\mathcal{N}(0, t - s)$ distribution (characteristic function, Definition 1.4.1). Hence, M is an \mathcal{F} -Brownian motion. ■

Lemma 5.3.1. *A real-valued continuous \mathcal{F} -adapted process X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is a standard \mathcal{F} -Brownian motion if and only if for any $\lambda \in \mathbb{R}$ the process*

$$M_t^\lambda = \exp\left(\lambda X_t - \frac{1}{2}\lambda^2 t\right), \quad \forall t \in [0, T],$$

is a \mathcal{F} -local martingale and $M_0^\lambda = 1$.

Theorem 5.3.2 (an oversimplified version of Dubins–Schwarz theorem). *If M is a continuous \mathcal{F} -local martingale such that $\langle M \rangle$ is strictly increasing and $\langle M \rangle_\infty = \infty$, then M is a time changed Brownian motion.*

Proof. $\langle M \rangle$ is continuous and strictly increasing therefore the inverse of $\langle M \rangle$ is given by $C_t := \inf\{s : \langle M \rangle_s > t\}$. We consider the process $W_t := M_{C_t}$ for $t \geq 0$, which is adapted to the filtration $\mathcal{G} = (\mathcal{F}_{C_t})_{t \geq 0}$. We now assume, without loss of generality, M is a bounded martingale. Then from the Doob Optional Sampling Theorem (2.5.1) and the fact that C_t is \mathcal{F} -stopping time for all t , we obtain

$$\mathbb{E}[W_t \mid \mathcal{G}_s] = \mathbb{E}[M_{C_t} \mid \mathcal{F}_{C_s}] = M_{C_s} = W_s.$$

Therefore W is a \mathcal{G} -local martingale. Furthermore, since C is the (right) inverse of M then $\langle W \rangle_t = \langle M \rangle_{C_t} = t$. Hence, by Lévy characterisation theorem, W is a Brownian motion in the filtration \mathcal{G} . Finally, since $\langle M \rangle$ is strictly increasing, C is the (left) inverse of $\langle M \rangle$ and so $C_{\langle M \rangle_t} = t$. This shows that

$$M_t = M_{C_{\langle M \rangle_t}} = W_{\langle M \rangle_t},$$

where W is a Brownian motion in the filtration \mathcal{G} . ■

5.4 Martingale Representation Theorem (for the Brownian Filtration)

Itô representation theorem states that any square integrable and \mathcal{F}_T^W -measurable random variable admits a representation as the Itô integral of some stochastic process.

Theorem 5.4.1 (Itô representation theorem). *For any random variable $X \in L^2(\Omega, \mathcal{F}_T^W, \mathbb{P})$, there exists a unique \mathcal{F} -predictable process γ from the class $L_{\mathbb{P}}^2(W)$ such that the following equality is valid*

$$X = \mathbb{E}^{\mathbb{P}} X + \int_0^T \gamma_u \cdot dW_u.$$

The condition that γ belongs to $L_{\mathbb{P}}^2(W)$ implies that the Itô integral $I(\gamma)$ is a square integrable \mathcal{F} -martingale and thus we also have that, for every $t \in [0, T]$,

$$\mathbb{E}^{\mathbb{P}}(X \mid \mathcal{F}_t^W) = \mathbb{E}^{\mathbb{P}} X + \int_0^t \gamma_u \cdot dW_u.$$

Remark 5.4.1. We omit the definition of an \mathcal{F} -predictable process, but we note any left-continuous and \mathcal{F} -adapted process is \mathcal{F} -predictable. Also, a \mathcal{F} -predictable process is also \mathcal{F} -progressively measurable, but the converse does not hold.

Theorem 5.4.2 (martingale representation theorem). *Let a process $M = (M_t)_{t \in [0, T]}$ be an \mathcal{F}^W -martingale s.t. $\mathbb{E}^{\mathbb{P}}[M_T^2] < \infty$. Then there exists a unique \mathcal{F} -predictable process γ from the class $L_{\mathbb{P}}^2(W)$ such that, for any $t \in [0, T]$,*

$$M_t = M_0 + \int_0^t \gamma_u \cdot dW_u.$$

Proof. A straightforward consequence of Itô representation theorem with $X = M_T$. ■

Remark 5.4.2. The theorem suggests the existence of γ , but in general it is very difficult to find it.

Examples 5.4.1 (Martingale representation theorem). Consider $X = W_T^3$. We want to find γ such that

$$W_T^3 = \mathbb{E}(W_T^3) + \int_0^T \gamma_u dW_u.$$

Note if we directly apply Itô's lemma to X , then $dW_t^3 = 3W_t^2 dW_t + \frac{1}{2} \cdot 6W_t dt$ and therefore,

$$W_T^3 = \int_0^T 3W_s^2 dW_s + \int_0^T 3W_s ds$$

which is not the form we want. We consider the martingale $M_t = \mathbb{E}(W_T^3 | \mathcal{F}_t)$ with $M_T = W_T^3$ [why? The MRT implies that a martingale can be written as a stochastic integral w.r.t. the Brownian motion]. For $t \in [0, T]$,

$$\begin{aligned} M_t &= \mathbb{E}(W_T^3 | \mathcal{F}_t) \\ &= \mathbb{E}((W_T - W_t + W_t)^3 | \mathcal{F}_t) \\ &= \mathbb{E}((W_T - W_t + x)^3 | \mathcal{F}_t) \big|_{x=W_t} && (W_t \text{ is } \mathcal{F}_t\text{-measurable}) \\ &= \mathbb{E}[(W_T - W_t + x)^3] \big|_{x=W_t} && (\text{independent increments}) \\ &= [x^3 + 3x(T-t)] \big|_{x=W_t} && (W_T - W_t \sim \mathcal{N}(0, T-t)) \\ &= W_t^3 + 3W_t(T-t) \\ &= f(t, W_t) \end{aligned}$$

where $f(t, x) = x^3 + 3x(T-t)$. Applying Itô's formula to $M_t = f(t, W_t)$ gives

$$\begin{aligned} dM_t &= -3W_t dt + 3W_t^2 dW_t + 3(T-t) dW_t + \frac{1}{2} \cdot 6W_t d\langle W \rangle_t \\ &= [3W_t^2 + 3(T-t)] dW_t. \end{aligned}$$

Therefore we can take $\gamma_u = 3W_u^2 + 3(T-u)$ for $u \in [0, T]$.

Corollary 5.4.1. Any \mathcal{F}^W -local martingale is necessarily a continuous process.

6 STOCHASTIC DIFFERENTIAL EQUATIONS

By a solution of the stochastic differential equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) \cdot dW_t$$

with the initial condition X_0 given as an \mathcal{F}_0 -measurable r.v., we mean an \mathbb{R}^k -valued, \mathcal{F} -adapted stochastic process X defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and such that, for every $t \in [0, T]$,

$$X_t = X_0 + \int_0^t \mu(u, X_u) du + \int_0^t \sigma(u, X_u) \cdot dW_u.$$

Note that any solution X to the SDE is an Itô process.

Definition 6.0.1 (pathwise uniqueness). We say that the **pathwise uniqueness** of solutions to the above SDE holds if for any filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, any d -dimensional standard Brownian motions W and \widetilde{W} defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and any two solutions X and \widetilde{X} driven by W and \widetilde{W} respectively, the following implication is true

$$\mathbb{P}\{W_t = \widetilde{W}_t \mid \forall t \in [0, T]\} = 1 \implies \mathbb{P}\{X_t = \widetilde{X}_t \mid \forall t \in [0, T]\} = 1.$$

Remark 6.0.1. The pathwise uniqueness of solutions is sometimes referred to as the strong uniqueness. This is due to the fact that under pathwise uniqueness any solution to the SDE is strong, meaning that it is adapted to the filtration generated by the driving Brownian motion W , \mathcal{F}^W .

6.1 Itô's Existence and Uniqueness Theorem

Theorem 6.1.1 (Itô's theorem, sufficient but not necessary conditions for the uniqueness of SDE solution). *Let $\mu : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ satisfy the following conditions:*

- (i) (Lipschitz continuity) μ and σ are Lipschitz continuous with respect to the variable x , that is, there exist constants $K_1, K_2 > 0$ such that, for any $x, y \in \mathbb{R}$ and $t \in \mathbb{R}_+$,

$$\begin{aligned} |\mu(t, x) - \mu(t, y)| &\leq K_1 |x - y| \\ |\sigma(t, x) - \sigma(t, y)| &\leq K_2 |x - y| \end{aligned}$$

- (ii) (linear growth condition) μ and σ satisfy the linear growth condition, i.e. there exist constants $C_1, C_2 > 0$ such that, for any $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$,

$$\begin{aligned} |\mu(t, x)| &\leq C_1(1 + |x|) \\ |\sigma(t, x)| &\leq C_2(1 + |x|) \end{aligned}$$

Then the SDE has a unique solution X .

Remark 6.1.1. To check Lipschitz continuity, it is sufficient to check that μ and σ have bounded derivative. By Mean Value Theorem,

$$\mu(t, x) - \mu(t, y) = \int_y^x \mu'(t, z) dz$$

so if $|\mu'| \leq K$, then

$$|\mu(t, x) - \mu(t, y)| \leq K \left| \int_y^x dz \right| = K|x - y|.$$

6.2 Linear SDE

Proposition 6.2.1. *The unique solution of the SDE*

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) \cdot dW_t$$

with

$$\begin{aligned} \mu(t, X_t) &= A_t X_t + a_t, \\ \sigma(t, X_t) &= [B_t^1 X_t + b_t^1, \dots, B_t^d X_t + b_t^d], \end{aligned}$$

where A, a, B^i, b^i ($i = 1, \dots, d$) are all \mathcal{F} -adapted bounded processes, is given by the formula

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_u^{-1} [a_u - \mathbf{B}_u \cdot \mathbf{b}_u] du + \int_0^t \Phi_u^{-1} \mathbf{b}_u \cdot d\mathbf{W}_u \right)$$

where

$$\Phi_t = \exp \left(\int_0^t \left(A_u - \frac{\mathbf{B}_u \cdot \mathbf{B}_u}{2} \right) du + \int_0^t \mathbf{B}_u \cdot d\mathbf{W}_u \right).$$

Remark 6.2.1. Analogous to ODE, we can consider the integration factor given by

$$Y_t = \Phi_t^{-1} = \exp \left(- \int_0^t A_u du - \int_0^t \mathbf{B}_u \cdot d\mathbf{W}_u + \frac{1}{2} \int_0^t \mathbf{B}_u \cdot \mathbf{B}_u du \right)$$

Proof. To check that X is a solution to the SDE, it suffices to differentiate the RHS using Itô's formula (easier: compute $d(X_t Y_t)$). The uniqueness of solution can be deduced from Itô's Theorem (6.1.1). ■

6.3 Non-Linear SDE

Solving non-linear SDEs requires us to apply some suitable transform to the original equation.

Examples 6.3.1 (Non-linear SDE: CIR process). We solve the SDE

$$dX_t = \left(\frac{1}{4} - bX_t \right) dt + \sqrt{X_t} dW_t, \quad X_0 = 1$$

by considering the $Y_t := \sqrt{X_t}$. We apply Itô's lemma to $Y_t := \sqrt{X_t}$ (with the omission of some technicalities).

$$\begin{aligned} dY_t &= \frac{1}{2\sqrt{X_t}} dX_t + \frac{1}{2} \cdot \left(-\frac{1}{4} \frac{1}{X_t^{3/2}} \right) d\langle X \rangle_t \\ &= \frac{1}{2\sqrt{X_t}} \left[\left(\frac{1}{4} - bX_t \right) dt + \sqrt{X_t} dW_t \right] - \frac{1}{8} \frac{1}{X_t^{3/2}} X_t dt \\ &= -\frac{b\sqrt{X_t}}{2} dt + \frac{1}{2} dW_t =: -\frac{b}{2} Y_t dt + \frac{1}{2} dW_t \end{aligned}$$

...and we know how to solve linear SDEs! By applying an integration factor of $e^{bt/2}$ or otherwise, we can show

$$Y_t = e^{-bt/2} + \frac{1}{2} \int_0^t e^{-\frac{1}{2}b(t-s)} dW_s$$

and therefore the solution to the SDE is given by

$$X_t = \left(e^{-bt/2} + \frac{1}{2} \int_0^t e^{-\frac{1}{2}b(t-s)} dW_s \right)^2.$$

6.4 Stochastic Exponential and Stochastic Logarithmic

Recall the definition of an Itô's integral: Let W be a d -dimensional standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For an \mathbb{R}^d -valued process $\gamma \in \mathcal{L}_{\mathbb{P}}(W)$, we define the real-valued \mathcal{F} -adapted process U by setting

$$U_t = I_t(\gamma) = \int_0^t \gamma_u \cdot d\mathbf{W}_u, \quad t \in [0, T]$$

The process U defined in this way is, of course, a continuous \mathcal{F} -local martingale.

Definition 6.4.1 (stochastic exponential). The **stochastic exponential** of U is given by the formula

$$\mathcal{E}_t(U) = \mathcal{E}_t \left(\int_0^t \gamma_u \cdot d\mathbf{W}_u \right) = \exp \left(\int_0^t \gamma_u \cdot d\mathbf{W}_u - \frac{1}{2} \int_0^t |\gamma_u|^2 du \right), \quad t \in [0, T]$$

that is, $\mathcal{E}_t(U) = \exp(U_t - \langle U \rangle_t/2)$. More generally, for any continuous local martingale M , we define

$$\mathcal{E}_t(M) = \exp\left(M_t - \frac{1}{2}\langle M \rangle_t\right), \quad t \in [0, T].$$

Lemma 6.4.1. *The stochastic exponential of U is the unique solution X of the stochastic differential equation*

$$dX_t = X_t \gamma_t \cdot d\mathbf{W}_t = X_t dU_t$$

with the initial condition $X_0 = 1$. More generally, for any continuous local martingale M , the stochastic exponential $\mathcal{E}(M)$ is the unique solution X to $dX_t = X_t dM_t$ with the initial condition $X_0 = 1$.

Remark 6.4.1.

1. It follows immediately that $d\mathcal{E}_t(U) = \mathcal{E}_t(U) \gamma_t \cdot d\mathbf{W}_t = \mathcal{E}_t(U) dU_t$.
2. Note that $\mathcal{E}(U)$ is a strictly positive continuous local martingale under \mathbb{P} and thus, it follows a supermartingale with respect to \mathcal{F} . If, in addition,

$$\mathbb{E}^\mathbb{P}(\mathcal{E}_T(U)) = \mathbb{E}^\mathbb{P}(\mathcal{E}_0(U)) = 1$$

so that $\mathbb{E}^\mathbb{P}(\mathcal{E}_t(U))$ is constant on $t \in [0, T]$ ⁹, then the process $\mathcal{E}(U)$ is a continuous \mathcal{F} -martingale.

Definition 6.4.2 (stochastic logarithm). Given a continuous, strictly positive process U , the **stochastic logarithm** of U , denoted by $\mathcal{L}_t(U)$ is the (unique) solution to

$$d\mathcal{L}_t(U) = \frac{1}{U_t} dU_t$$

with $\mathcal{L}_0(U) = 0$.

Remark 6.4.2.

1. One can check that

$$\mathcal{L}_t(U) = \int_0^t \frac{1}{U_s} dU_s = \log U_t + \frac{1}{2} \int_0^t \frac{1}{U_s^2} d\langle U \rangle_s$$

is one (and “the”) solution, by noting that $d \log U_t = \frac{1}{U_t} dU_t - \frac{1}{2} \frac{1}{U_t^2} d\langle U \rangle_t$.

2. Stochastic logarithm is an inverse operation to stochastic exponential: $\mathcal{E}(\mathcal{L}(U)) = \mathcal{L}(\mathcal{E}(U)) = U$. Check

$$\begin{aligned} d\mathcal{L}_t(\mathcal{E}(U)) &= \frac{1}{\mathcal{E}_t(U)} d\mathcal{E}_t(U) = \frac{1}{\mathcal{E}_t(U)} \mathcal{E}_t(U) dU_t = dU_t, \\ d\mathcal{E}_t(\mathcal{L}(U)) &= \mathcal{E}_t(\mathcal{L}(U)) d\mathcal{L}_t(U) = U_t \frac{1}{U_t} dU_t = dU_t. \end{aligned}$$

⁹For a supermartingale, we have $\mathbb{E}(M_T) \leq \mathbb{E}(M_t) \leq \mathbb{E}(M_0)$ for all $t \in [0, T]$, hence the “so that” conclusion.

7 GIRSANOV THEOREM

7.1 Change of Measure

Definition 7.1.1. Consider two probability measures \mathbb{P} and \mathbb{Q} defined on (Ω, \mathcal{A}) .

- (equivalence) \mathbb{P} and \mathbb{Q} are said to be **equivalent** if

$$\mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0$$

for all $A \in \mathcal{A}$; i.e. they have the same set of null events in the σ -algebra \mathcal{A} .

- (absolute continuity) \mathbb{Q} is said to be **absolutely continuous** with respect to \mathbb{P} , written $\mathbb{Q} \ll \mathbb{P}$, if

$$\mathbb{P}(A) = 0 \implies \mathbb{Q}(A) = 0$$

for all $A \in \mathcal{A}$.

Remark 7.1.1. If \mathbb{P} and \mathbb{Q} are equivalent on \mathcal{A} , then they also enjoy this property on any σ -algebra $\mathcal{G} \subseteq \mathcal{A}$. In particular, if $\mathbb{P} \sim \mathbb{Q}$ on \mathcal{F}_T then $\mathbb{P} \sim \mathbb{Q}$ on \mathcal{F}_t for every $t \in [0, T]$.

Theorem 7.1.1 (Radon–Nikodym Theorem). *Let \mathbb{P} and \mathbb{Q} be two probability measures on (Ω, \mathcal{A}) such that $\mathbb{Q} \ll \mathbb{P}$, then there exist a unique positive \mathbb{P} -integrable random variable η (the **Radon–Nikodym density**) such that for all $A \in \mathcal{A}$,*

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(\eta \mathbf{1}_A).$$

We also write

$$\eta = \frac{d\mathbb{Q}}{d\mathbb{P}}.$$

Examples 7.1.1 (Radon–Nikodym Density).

1. Coin flip with $\Omega = \{T, H\}$. Consider \mathbb{P} given by $\mathbb{P}\{H\} = \mathbb{P}\{T\} = 1/2$ and \mathbb{Q} by $\mathbb{Q}\{H\} = 1/3$, $\mathbb{Q}\{T\} = 2/3$. Clearly \mathbb{P} and \mathbb{Q} are equivalent; further,

$$\begin{aligned} \mathbb{Q}\{T\} &= \mathbb{E}^{\mathbb{P}}[\eta \mathbf{1}_{\{T\}}] = \eta\{T\} \mathbb{P}\{T\} \\ \mathbb{Q}\{H\} &= \mathbb{E}^{\mathbb{P}}[\eta \mathbf{1}_{\{H\}}] = \eta\{H\} \mathbb{P}\{H\} \end{aligned}$$

$$\implies \eta(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}, \omega \in \{T, H\}.$$

2. Consider $\Omega = [0, \infty)$ and $\mathcal{A} = \mathcal{B}([0, \infty))$. One can define two equivalent probability measure on (Ω, \mathcal{A}) by setting, for every $A \in \mathcal{A}$,

$$\begin{aligned} \mathbb{P}(A) &= \int_{\Omega} \mathbf{1}_A(x) \mathbf{1}_{[0,1]}(x) dx = \int_0^{\infty} \mathbf{1}_A(x) d\mathbb{P}(x) \\ \mathbb{Q}(A) &= \int_{\Omega} \mathbf{1}_A(x) e^{-x} dx = \int_0^{\infty} \mathbf{1}_A(x) d\mathbb{Q}(x) \end{aligned}$$

The probability measure \mathbb{P} is the uniform (Lebesgue) measure on $[0, 1]$ and \mathbb{Q} is the exponential measure. It is not difficult to see that \mathbb{P} is absolutely continuous with respect to \mathbb{Q} but not the reverse, as $\mathbb{P}(A) \leq e\mathbb{Q}(A)$.

To find the Radon–Nikodym density of \mathbb{Q} with respect to \mathbb{P} we note that for any $A \in \mathcal{A}$,

$$\mathbb{P}(A) = \int_{\Omega} \mathbf{1}_A(x) \mathbf{1}_{[0,1]}(x) dx = \int_{\Omega} \mathbf{1}_A(x) \frac{\mathbf{1}_{[0,1]}(x)}{e^{-x}} e^{-x} dx = \int_{\Omega} \mathbf{1}_A(x) \frac{\mathbf{1}_{[0,1]}(x)}{e^{-x}} d\mathbb{Q}(x)$$

Therefore the Radon–Nikodym density of \mathbb{P} w.r.t. \mathbb{Q} is given by $\eta(x) = \mathbf{1}_{[0,1]}(x)/e^{-x}$. Further, if we let $X(\omega) = \omega$, then we can show that $X \sim U[0, 1]$ under \mathbb{P} and $X \sim \text{Exp}(1)$ under \mathbb{Q} by computing the CDFs.

7.2 Radon–Nikodym Density Process

For Theorem 7.1.1, in the special case of $\mathcal{A} = \mathcal{F}_T$, we usually write η_T to denote the (unique) \mathcal{F}_T -measurable r.v. such that for every $A \in \mathcal{F}_T$,

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(\eta_T \mathbf{1}_A) = \int_A \eta_T d\mathbb{P}.$$

Remark 7.2.1.

(i) For any \mathbb{Q} -integrable r.v. ψ , we have

$$\mathbb{E}^{\mathbb{Q}}(\psi) = \mathbb{E}^{\mathbb{P}}(\eta_T \psi).$$

ψ is \mathbb{Q} -integrable if and only if $\eta_T \psi$ is \mathbb{P} -integrable.

(ii) $\mathbb{P}\{\eta_T > 0\} = 1$.

(iii) $\mathbb{E}^{\mathbb{P}}\eta_T = \mathbb{Q}(\Omega) = 1$.

Definition 7.2.1 (Radon–Nikodym density process). The **Radon–Nikodym density process** $\eta = (\eta_t)_{t \in [0, T]}$ of \mathbb{Q} with respect to \mathbb{P} and a given filtration \mathcal{F} is defined by setting

$$\eta_t := \mathbb{E}^{\mathbb{P}}(\eta_T \mid \mathcal{F}_t), \quad \forall t \in [0, T]$$

Remark 7.2.2.

(i) The tower property implies that Radon–Nikodym density process η is a strictly positive martingale under \mathbb{P} .

(ii) The random variable η_t is the Radon–Nikodym density of \mathbb{Q} with respect to \mathbb{P} on (Ω, \mathcal{F}_t) . That is,

$$\eta_t = \left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t}$$

7.3 Abstract Bayes Formula

Lemma 7.3.1 (abstract Bayes formula). *Let \mathcal{G} be a sub- σ -algebra of \mathcal{F}_T , and let ψ be a \mathbb{Q} -integrable random variable. Then*

$$\mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G}) = \frac{\mathbb{E}^{\mathbb{P}}(\eta \psi \mid \mathcal{G})}{\mathbb{E}^{\mathbb{P}}(\eta \mid \mathcal{G})}.$$

Proof. It can be easily checked that $\mathbb{E}^{\mathbb{P}}(\eta \mid \mathcal{G})$ is strictly positive \mathbb{P} -a.s. so that the RHS is well-defined. By our assumption, the random variable $\eta \psi$ is \mathbb{P} -integrable, it is therefore enough to show that

$$\mathbb{E}^{\mathbb{P}}(\eta \psi \mid \mathcal{G}) = \mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G}) \mathbb{E}^{\mathbb{P}}(\eta \mid \mathcal{G}).$$

We want to verify that, for all $A \in \mathcal{G}$,

$$\mathbb{E}^{\mathbb{P}}(\eta \psi \mathbf{1}_A) = \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G}) \mathbb{E}^{\mathbb{P}}(\eta \mid \mathcal{G}) \mathbf{1}_A \right].$$

Write $Y = \mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G})$. Then

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G}) \mathbb{E}^{\mathbb{P}}(\eta \mid \mathcal{G}) \mathbf{1}_A \right] &= \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}}(\eta Y \mid \mathcal{G}) \mathbf{1}_A \right] && \text{(since } Y \text{ is } \mathcal{G}\text{-measurable)} \\ &= \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}}(\eta Y \mathbf{1}_A \mid \mathcal{G}) \right] && \text{(since } \mathbf{1}_A \text{ is } \mathcal{G}\text{-measurable)} \\ &= \mathbb{E}^{\mathbb{P}}[\eta Y \mathbf{1}_A] && \text{(tower property)} \\ &= \mathbb{E}^{\mathbb{Q}}[Y \mathbf{1}_A] = \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}}(\psi \mid \mathcal{G}) \mathbf{1}_A \right] \\ &= \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}}(\psi \mathbf{1}_A \mid \mathcal{G}) \right] \\ &= \mathbb{E}^{\mathbb{Q}}[\psi \mathbf{1}_A] \\ &= \mathbb{E}^{\mathbb{P}}(\eta \psi \mathbf{1}_A) \end{aligned}$$

The definition of conditional expectation then gives the desired result. ■

Lemma 7.3.2. *A stochastic process X is an \mathcal{F} -martingale under \mathbb{Q} if and only if the product ηX is an \mathcal{F} -martingale under \mathbb{P} .*

Proof. This is a direct application of the abstract Bayes formula. Assume first that ηX is an \mathcal{F} -martingale under \mathbb{P} so that $\mathbb{E}^{\mathbb{P}}(\eta_t X_t \mid \mathcal{F}_u) = \eta_u X_u$ for any $0 \leq u \leq t \leq T$. Then the Bayes formula yields,

$$\mathbb{E}^{\mathbb{Q}}(X_t \mid \mathcal{F}_u) = \frac{\mathbb{E}^{\mathbb{P}}(\eta_T X_t \mid \mathcal{F}_u)}{\mathbb{E}^{\mathbb{P}}(\eta_T \mid \mathcal{F}_u)} = \frac{\mathbb{E}^{\mathbb{P}}(X_t \mathbb{E}^{\mathbb{P}}(\eta_T \mid \mathcal{F}_t) \mid \mathcal{F}_u)}{\mathbb{E}^{\mathbb{P}}(\eta_T \mid \mathcal{F}_u)} = \frac{\mathbb{E}^{\mathbb{P}}(X_t \eta_t \mid \mathcal{F}_u)}{\eta_u} = \frac{X_u \eta_u}{\eta_u} = X_u$$

for any $0 \leq u \leq t \leq T$. We conclude that X is an \mathcal{F} -martingale under \mathbb{Q} . The proof of the converse implication goes along the same lines. ■

7.4 Girsanov Theorem

7.4.1 Linear Drift

Proposition 7.4.1. *Let W be a one-dimensional standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For a real number $\gamma \in \mathbb{R}$, we define the process \widetilde{W} by setting $\widetilde{W}_t = W_t - \gamma t$ for $t \in [0, T]$. Let the probability measure $\widetilde{\mathbb{P}}$, equivalent to \mathbb{P} on (Ω, \mathcal{F}_T) , be defined through the formula (see Definition 6.4.1 for \mathcal{E})*

$$\eta_T = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \exp\left(\gamma W_T - \frac{1}{2}\gamma^2 T\right) = \mathcal{E}_T(\gamma W).$$

Then \widetilde{W} is a standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.

Proof. To establish the proposition, we make use of the abstract Bayes formula (Lemma 7.3.1) and Lévy's characterisation theorem (Lemma 5.3.1). In view of the latter, it suffices to show that for any $\lambda \in \mathbb{R}$ the process

$$M_t^\lambda = \exp\left(\lambda \widetilde{W}_t - \frac{1}{2}\lambda^2 t\right), \quad \forall t \in [0, T],$$

is an \mathcal{F} -martingale under $\widetilde{\mathbb{P}}$. By Lemma 7.3.2, we can equivalently check that ηM^λ is an \mathcal{F} -martingale under \mathbb{P} . But

$$\eta_t M_t^\lambda = \exp\left(\gamma W_t - \frac{1}{2}\gamma^2 t\right) \exp\left(\lambda(W_t - \gamma t) - \frac{1}{2}\lambda^2 t\right) = \exp\left(\alpha W_t - \frac{1}{2}\alpha^2 t\right)$$

where $\alpha = \lambda + \gamma$. Clearly $\mathbb{E}^\mathbb{P}(\eta_t M_t^\lambda) = 1$ for all $t \in [0, T]$ and thus this process follows an \mathcal{F} -martingale under \mathbb{P} . From the Lévy characterisation theorem we conclude that \widetilde{W} is a standard Brownian motion on $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$. ■

7.4.2 Stochastic Drift

Let W be a d -dimensional standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For an \mathbb{R}^d -valued process $\gamma \in \mathcal{L}_\mathbb{P}(W)$, we define the real-valued \mathcal{F} -adapted process U by setting

$$U_t = I_t(\gamma) = \int_0^t \gamma_u \cdot d\mathbf{W}_u, \quad t \in [0, T].$$

Proposition 7.4.2. *Suppose that γ is an \mathbb{R}^d -valued \mathcal{F} -progressively measurable process such that $\mathbb{E}^\mathbb{P}[\mathcal{E}_T(U)] = 1$. Define a probability measure $\widetilde{\mathbb{P}}$ on (Ω, \mathcal{F}_T) equivalent to \mathbb{P} by means of the Radon–Nikodym derivative*

$$\eta_T = \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \mathcal{E}_T\left(\int_0^\cdot \gamma_u \cdot d\mathbf{W}_u\right) = \mathcal{E}_T(U).$$

Then the process \widetilde{W} given by the formula

$$\widetilde{W}_t = W_t - \int_0^t \gamma_u du, \quad \forall t \in [0, T]$$

follows a standard d -dimensional Brownian motion on the space $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.

The condition $\mathbb{E}^\mathbb{P}[\mathcal{E}_T(U)] = 1$ is required in Proposition 7.4.2 to ensure that $\eta = \mathcal{E}_T(U)$ is a true \mathcal{F} -martingale under \mathbb{P} : $\eta = \mathcal{E}_T(U)$ is in general a non-negative \mathcal{F} -local martingale and therefore an \mathcal{F} -supermartingale. Having $\mathbb{E}^\mathbb{P}[\mathcal{E}_T(U)] = \mathbb{E}^\mathbb{P}[\mathcal{E}_0(U)] = 1$ guarantees η is a true \mathcal{F} -martingale.

However, it is difficult to check this condition in general. We present two sufficient conditions below.

Proposition 7.4.3.

(i) (Novikov's condition) If

$$\mathbb{E}^\mathbb{P}\left[\exp\left(\frac{1}{2}\int_0^T |\gamma_u|^2 du\right)\right] < \infty,$$

then $\mathbb{E}^\mathbb{P}(\mathcal{E}_T(U)) = 1$. Consequently, the process $\eta = \mathcal{E}(U)$ is a strictly positive continuous \mathcal{F} -martingale. In particular, if the process γ is uniformly bounded, that is, there exists a constant K such that $|\gamma_t| \leq K$ for $t \in [0, T]$, then Novikov's condition is satisfied and thus $\mathbb{E}^\mathbb{P}(\mathcal{E}_T(U)) = 1$.

(ii) (Kazamaki's condition) A weaker but also sufficient condition is the Kazamaki condition:

$$\mathbb{E}^{\mathbb{P}} \left[\exp \left(\frac{1}{2} \int_0^t \gamma_u \cdot d\mathbf{W}_u \right) \right] < \infty, \quad \forall t \in [0, T].$$

The next result shows that, if the underlying filtration is generated by a d -dimensional Brownian motion, the Radon–Nikodym density process of any probability measure equivalent to \mathbb{P} has necessarily the form of the stochastic exponential for some process γ .

Proposition 7.4.4. *Assume $\mathcal{F} = \mathcal{F}^W$. Then for any probability measure $\tilde{\mathbb{P}}$ on (Ω, \mathcal{F}_T) equivalent to \mathbb{P} , there exists an \mathcal{F}^W -progressively measurable, \mathbb{R}^d -valued process γ such that*

$$\eta_T = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \mathcal{E}_T \left(\int_0^\cdot \gamma_u \cdot d\mathbf{W}_u \right).$$

Proof. Let η_T be the Radon–Nikodym density of $\tilde{\mathbb{P}}$ with respect to \mathbb{P} on (Ω, \mathcal{F}_T) (whose existence is guaranteed by Radon–Nikodym theorem). Clearly, the process $\eta_t = \mathbb{E}^{\tilde{\mathbb{P}}}(\eta_T \mid \mathcal{F}_t)$ is an \mathcal{F} -martingale and $\eta_0 = 1$. Further, since the underlying filtration $\mathcal{F} = \mathcal{F}^W$ is a Brownian filtration, the martingale representation theorem (Theorem 5.4.2) implies the existence of a process $\tilde{\gamma} \in \mathcal{L}_{\mathbb{P}}(W)$ such that

$$\eta_t = 1 + \int_0^t \tilde{\gamma}_u \cdot d\mathbf{W}_u, \quad \forall t \in [0, T].$$

Since $\mathbb{P}\{\eta_T > 0\} = 1$, we also have that $\mathbb{P}\{\eta_t > 0\} = 1$ for any $t \in [0, T]$, and further, in view of continuity of η (which is apparent from the representation above) we obtain $\mathbb{P}\{\eta_t > 0, \forall t \in [0, T]\} = 1$. Therefore, the process $\gamma_t = \tilde{\gamma}_t \eta_t^{-1}$ is well defined and we have

$$\eta_t = 1 + \int_0^t \tilde{\gamma}_u \cdot d\mathbf{W}_u = 1 + \int_0^t \eta_u \gamma_u \cdot d\mathbf{W}_u.$$

We conclude that the Radon–Nikodym density process is the unique solution to the SDE

$$d\eta_t = \eta_t \gamma_t \cdot d\mathbf{W}_t,$$

and thus, in view of the form of the stochastic exponential, it satisfies, for any $t \in [0, T]$,

$$\eta_t = \mathcal{E}_t \left(\int_0^\cdot \gamma_u \cdot d\mathbf{W}_u \right)$$

This completes the proof of the proposition. ■

7.4.3 Continuous Semimartingales

Theorem 7.4.1 (Girsanov theorem for continuous semimartingales). *Let $\tilde{\mathbb{P}} \sim \mathbb{P}$ be two equivalent probability measures on (Ω, \mathcal{F}_T) with Radon–Nikodym density η_T . We assume, in addition, that the Radon–Nikodym density process $\eta_t := \mathbb{E}^{\tilde{\mathbb{P}}}(\eta_T \mid \mathcal{F}_t)$, $t \in [0, T]$ is continuous¹⁰. Then any continuous real-valued \mathbb{P} -semimartingale X is a continuous $\tilde{\mathbb{P}}$ -semimartingale. If the canonical decomposition of X under \mathbb{P} is $X = X_0 + M + A$ then its canonical decomposition under $\tilde{\mathbb{P}}$ is $X = X_0 + \tilde{M} + \tilde{A}$ where*

$$\tilde{M}_t = M_t - \int_0^t \frac{1}{\eta_u} d\langle \eta, M \rangle_u = M_t - \langle \mathcal{L}(\eta), M \rangle_t$$

and

$$\tilde{A}_t = A_t + \int_0^t \frac{1}{\eta_u} d\langle \eta, M \rangle_u = A_t + \langle \mathcal{L}(\eta), M \rangle_t$$

In particular, X follows a local martingale under $\tilde{\mathbb{P}}$ if and only if the process $\tilde{A} = A + \langle \mathcal{L}(\eta), M \rangle$ vanishes identically (Definition 5.2.1), that is, $A_t + \langle \mathcal{L}(\eta), M \rangle_t = 0$ for every $t \in [0, T]$.

The transform $\varphi : M \mapsto M - \langle \mathcal{L}(\eta), M \rangle$ is called **Girsanov transform**.

¹⁰We know already that this holds if $\mathcal{F} = \mathcal{F}^W$ for some Brownian motion W , see Proof of Proposition 7.4.4.

Proof. We need to prove that

(i) \widetilde{M} is a local martingale under $\widetilde{\mathbb{P}}$.

(ii) \widetilde{A} is a process of finite variation with $\widetilde{A}_0 = 0$ (i.e. almost all sample paths are of finite variation on $[0, T]$).

To show (ii), recall the polarisation formula

$$\langle X, Y \rangle = \frac{1}{4} (\langle X + Y \rangle - \langle X - Y \rangle),$$

which is the difference of two increasing processes and therefore must have finite variation (Theorem 3.3.1). Hence, $\widetilde{A} = A + \langle \mathcal{L}(\eta), M \rangle$ being the sum of two processes of finite variation, must be a process of finite variation itself.

To show (i), we check its equivalence: $\eta \widetilde{M}$ is a local martingale under \mathbb{P} . Consider the dynamics of $\eta \widetilde{M}$:

$$\begin{aligned} d\eta_t \widetilde{M}_t &= \eta_t d\widetilde{M}_t + \widetilde{M}_t d\eta_t + d\langle \eta, \widetilde{M} \rangle_t && \text{(integration by parts)} \\ &= \eta_t d[M_t - \langle \mathcal{L}(\eta), M \rangle_t] + \widetilde{M}_t d\eta_t + d\langle \eta, M - \langle \mathcal{L}(\eta), M \rangle \rangle_t \\ &= \eta_t dM_t - \eta_t d\langle \mathcal{L}(\eta), M \rangle_t + \widetilde{M}_t d\eta_t + d\langle \eta, M \rangle_t && (\langle \mathcal{L}(\eta), M \rangle \text{ has finite variation}) \\ &= \eta_t dM_t + \widetilde{M}_t d\eta_t && \text{(observe } d\langle \eta, M \rangle_t = \eta_t d\langle \mathcal{L}(\eta), M \rangle_t) \end{aligned}$$

where M and η are both \mathbb{P} -local martingales. Hence, $\eta \widetilde{M}$ is also a \mathbb{P} -local martingale, completing the proof. ■

A SOME USEFUL RESULTS¹¹

Theorem A.0.1 (Cauchy–Schwarz inequality).

1. (definite integrals) Let f and g be real functions which are continuous on the closed interval $[a, b]$. Then:

$$\left(\int_a^b f(t)g(t) dt \right)^2 \leq \int_a^b f^2(t) dt \int_a^b g^2(t) dt.$$

As a corollary, we have

$$\left(\int_a^b f(t) dt \right)^2 \leq (b - a) \int_a^b f^2(t) dt.$$

2. (expectations) For any two random variables X and Y ,

$$[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2),$$

or equivalently,

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Theorem A.0.2 (Fubini–Tonelli theorem). Let X be a stochastic process such that $\int_0^T \mathbb{E}(|X_s|) ds < \infty$, then

$$\mathbb{E} \left[\int_0^T X_s ds \right] = \int_0^T \mathbb{E}[X_s] ds.$$

¹¹I may omit some technical conditions in this section.