

Lecture Notes for Stochastic Analysis

JYUNYI LIAO

Contents

1	Basic Measure Theory	3
1.1	Measurable Spaces, Sierpiński-Dynkin π - λ System and Monotone Classes	3
1.2	Measures, Pre-measures and Carathéodory's Extension	6
1.3	Measurable Functions and Simple Function Approximation	12
1.4	Lebesgue Integration	16
1.5	Signed Measures, Jordan Decomposition and Radon-Nikodym Theorem	22
1.6	Convergence of Measurable Functions and Measures	26
2	Random Variables	32
2.1	Random Variables and Independence	32
2.2	Expectation	36
2.3	Conditional Expectation and Distribution	38
2.4	Stochastic Convergence	44
2.5	Characteristic Functions	49
2.6	Continuity Theorem and Central Limit Theorems	54
3	Martingales	60
3.1	Processes, Filtrations and Stopping Times	60
3.2	Discrete-time Martingales	67
3.3	Continuous-time Martingales	76
3.4	Local Martingales, Quadratic Variation and Bracket	83
4	Brownian Motions	94
4.1	Pre-Brownian Motions and Brownian Motions	94
4.2	Canonical Construction and Wiener's Construction	99
4.3	Sample Paths of Brownian Motion	105
4.4	Strong Markov Property and Reflection Principle	109
5	Stochastic Integration	113
5.1	Construction of Stochastic Integrals	113
5.2	Itô's Formula and its Consequences	122
5.3	The Representation of Martingales as Stochastic Integral	127
5.4	Stochastic Differential Equations	130
5.5	Girsanov's Theorem and Cameron-Martin Formula	136

6	Markov Processes	142
6.1	Transition Semigroups and Feller Semigroups	142
6.2	Markov Processes and Feller Processes	148
6.3	The Generator and Feynman-Kac Formula	153
6.4	Diffusion Processes	158

1 Basic Measure Theory

1.1 Measurable Spaces, Sierpiński-Dynkin π - λ System and Monotone Classes

Let Ω be a nonempty set. Denote by 2^Ω the set of all subsets of Ω , namely, $2^\Omega = \{A : A \subset \Omega\}$. Any subset $\mathcal{A} \subset 2^\Omega$ is called a collection of subsets of Ω .

Definition 1.1 (σ -algebra and measurable space). Let \mathcal{F} be a collection of subsets of Ω . Then \mathcal{F} is said to be a σ -algebra (or σ -field) if the following conditions are satisfied:

- (i) $\Omega \in \mathcal{F}$;
- (ii) For all $A \in \mathcal{F}$, the complement $A^c := \Omega \setminus A \in \mathcal{F}$;
- (iii) For all sequences $(A_n)_{n=1}^\infty$ in \mathcal{F} , $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

A pair (Ω, \mathcal{F}) consisting of a set Ω and a σ -algebra of subsets of Ω is called a *measurable space*. A set $A \in \mathcal{F}$ is said to be \mathcal{F} -measurable.

Remark. Clearly, $\{\emptyset, \Omega\}$ and 2^Ω are two trivial σ -algebras of subsets of Ω . Furthermore, we can show that a σ -algebra \mathcal{F} also satisfies the following:

- (i) $\emptyset \in \mathcal{F}$;
- (ii) For all $A, B \in \mathcal{F}$, $A \setminus B, A \cup B, A \cap B \in \mathcal{F}$;
- (iii) For all sequences $(A_n)_{n=1}^\infty$ in \mathcal{F} , $\bigcap_{n=1}^\infty A_n \in \mathcal{F}$.

Moreover, the intersection of any collection of σ -algebras is again a σ -algebra.

Definition 1.2. Let \mathcal{A} be a collection of subsets of Ω . The σ -algebra generated by \mathcal{A} is the minimal σ -algebra of subsets of Ω that contains \mathcal{A} :

$$\sigma(\mathcal{A}) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra and } \mathcal{F} \supset \mathcal{A} \}.$$

Remark. We can generate the minimal σ -algebra \mathcal{F} from \mathcal{A} as follows: (i) Complete \mathcal{A} : $\mathcal{F} \leftarrow \mathcal{A} \cup \{\Omega, \emptyset\}$;

(ii) For all $A \in \mathcal{A}$, add A^c to \mathcal{F} if necessary; (iii) For all sequences of sets in \mathcal{F} , include their union in \mathcal{F} .

Definition 1.3. Let (X, \mathcal{T}) be a topological space. The *Borel σ -algebra* on X is defined as the σ -algebra generated by all open sets in X . We write $\mathcal{B}(X) = \sigma(\mathcal{T})$.

Remark. One of the most commonly used σ -algebra is the Borel σ -algebra on \mathbb{R} . By definition, $\mathcal{B}(\mathbb{R})$ contains all open subsets, closed subsets, finite subsets and countable subsets of \mathbb{R} . Also, $\mathcal{B}(\mathbb{R})$ contains all G_δ -sets (a countable intersection of open sets) and all F_σ -sets (a countable union of closed sets) in \mathbb{R} .

Now we introduce the Sierpiński-Dynkin π - λ theorem, which is a useful tool in analysis.

Definition 1.4 (π -system). Let \mathcal{P} be a collection of subsets of Ω . If $A \cap B \in \mathcal{P}$ for all $A, B \in \mathcal{P}$, then \mathcal{P} is said to be a π -system.

Definition 1.5 (λ -system). Let \mathcal{L} be a collection of subsets of Ω . Then \mathcal{L} is said to be a λ -system (or *Dynkin system*) if it satisfies the following conditions:

- (i) $\Omega \in \mathcal{L}$;
- (ii) For all $A, B \in \mathcal{L}$ such that $A \subset B$, it holds $B \setminus A \in \mathcal{L}$;
- (iii) For all increasing sequences $A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots$ in \mathcal{L} , it holds $\bigcup_{n=1}^\infty A_n \in \mathcal{L}$.

Remark. Another equivalent formulation of λ -system is stated below:

- (i) $\Omega \in \mathcal{L}$;
- (ii) For all $A \in \mathcal{L}$, it holds $A^c \in \mathcal{L}$;
- (iii) For all sequences $(A_n)_{n=1}^\infty$ of disjoint sets in \mathcal{L} , it holds $\coprod_{n=1}^\infty A_n \in \mathcal{L}$.

We also observe that the intersection of any collection of λ -systems is again a λ -system. Therefore, similar to Definition 1.2, we can define the minimal λ -system generated by a collection \mathcal{A} of subsets of Ω :

$$\lambda(\mathcal{A}) = \bigcap \{ \mathcal{L} : \mathcal{L} \text{ is a } \lambda\text{-system and } \mathcal{L} \supset \mathcal{A} \}.$$

In general, a λ -system is not a σ -algebra, since it is not always closed under countable unions unless they are disjoint. For instance, let $\Omega = \{0, 1, 2, 3\}$, and consider $\mathcal{L} = \{\Omega, \{0, 1\}, \{2, 3\}, \{0, 2\}, \{1, 3\}, \{1, 2\}, \{0, 3\}, \emptyset\}$.

Lemma 1.6. \mathcal{F} is a σ -algebra if and only if \mathcal{F} is a π -system and λ -system.

Proof. By definition, a σ -algebra is a π -system and λ -system. Conversely, if \mathcal{F} is a π -system and λ -system, we only need to verify Definition 1.1 (iii). Let $(A_n)_{n=1}^\infty$ be a sequence in \mathcal{F} . Define

$$B_n = \bigcup_{j=1}^n A_j, \quad \forall n \in \mathbb{N}.$$

Clearly, $B_1 \in \mathcal{F}$. Moreover, if $B_{n-1} \in \mathcal{F}$, we have $C_n = B_{n-1} \cap A_n \in \mathcal{F}$ since \mathcal{F} is a π -system, and

$$B_n = A_n \cup B_{n-1} = \underbrace{(A_n \setminus C_n)}_{\in \mathcal{F}} \cup \underbrace{(B_{n-1} \setminus C_n)}_{\in \mathcal{F}} \cup C_n. \quad (1.1)$$

Note that (1.1) is a union of disjoint sets in \mathcal{F} , which is a λ -system. Hence B_n is an increasing sequence in \mathcal{F} , which implies $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n \in \mathcal{F}$. \square

Theorem 1.7 (Sierpiński-Dynkin π - λ theorem). Let \mathcal{P} and \mathcal{L} be two collections of subsets of Ω such that $\mathcal{P} \subset \mathcal{L}$. If \mathcal{P} is a π -system, and \mathcal{L} is a λ -system, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof. We first claim that $\lambda(\mathcal{P})$ is a σ -algebra. By Lemma 1.6, it suffices to show that $\lambda(\mathcal{P})$ is a π -system.

Step I: We show that for all $A \in \lambda(\mathcal{P})$, the collection

$$\lambda_A := \{B \subset \Omega : A \cap B \in \lambda(\mathcal{P})\}$$

is a λ -system. Clearly, λ_A contains Ω and is closed under countable disjoint unions. For any $B \in \lambda_A$,

$$A \cap B^c = A \setminus (A \cap B) \in \lambda(\mathcal{P}),$$

because it is the proper difference of sets in $\lambda(\mathcal{P})$. Hence λ_A is a λ -system.

Step II: We show that $A \cap B \in \lambda(\mathcal{P})$ for all $A \in \mathcal{P}$ and all $B \in \lambda(\mathcal{P})$. Fix $A \in \mathcal{P}$. Since \mathcal{P} is a π -system and $\mathcal{P} \subset \lambda(\mathcal{P})$, we have $\mathcal{P} \subset \lambda_A$. Note that $\lambda(\mathcal{P})$ is the minimal λ -system generated by \mathcal{P} , we have $\lambda(\mathcal{P}) \subset \lambda_A$.

Step III: We show that $\lambda(\mathcal{P})$ is a σ -algebra. Let $B \in \lambda(\mathcal{P})$. By Step I, λ_B is a λ -system. If $E \in \mathcal{P}$, using Step II, we have $E \cap B \in \lambda(\mathcal{P})$, which implies $E \in \lambda_B$. Then $\mathcal{P} \in \lambda_B$, and $\lambda(\mathcal{P}) \subset \lambda_B$. Hence for all $A \in \lambda(\mathcal{P}) \subset \lambda_B$, $A \cap B \in \lambda(\mathcal{P})$. As a result, $\lambda(\mathcal{P})$ is a π -system.

Since $\lambda(\mathcal{P})$ is a σ -algebra, we have $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P}) \subset \mathcal{L}$. In fact, we can prove that $\sigma(\mathcal{P}) = \lambda(\mathcal{P})$: The other direction holds because $\sigma(\mathcal{P})$ is a λ -system, which implies $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. \square

Now we introduce the monotone class theorem.

Definition 1.8 (Monotone class). A collection \mathcal{M} of subsets of Ω is said to be a *monotone class* if the following hold: (i) For all increasing sequence $(A_n)_{n=1}^\infty$ in \mathcal{M} , it holds $\bigcup_{n=1}^\infty A_n \in \mathcal{M}$; (ii) For all decreasing sequence $(B_n)_{n=1}^\infty$ in \mathcal{M} , it holds $\bigcap_{n=1}^\infty B_n \in \mathcal{M}$.

Remark. Note that the intersection of a collection of monotone classes is also a monotone class. Again, we can define the monotone class generated by a collection \mathcal{A} of subsets of Ω :

$$m(\mathcal{A}) = \bigcap \{ \mathcal{M} : \mathcal{M} \text{ is a monotone class and } \mathcal{M} \supset \mathcal{A} \}.$$

Definition 1.9 (Algebra). Let \mathcal{A} be a collection of subsets of Ω . Then \mathcal{A} is said to be an *algebra* if the following hold: (i) $\Omega \in \mathcal{A}$; (ii) $A^c \in \mathcal{A}$ for all $A \in \mathcal{A}$; (iii) $A \cup B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$.

Remark. By (ii) and (iii), we can show that an algebra \mathcal{A} is closed under finite unions and finite intersections. In fact, another formulation of algebra uses the ring.

Definition 1.9*. (Ring and algebra). A *ring* (or *pre-algebra*) is a collection \mathcal{R} of subsets of Ω such that $A \setminus B, A \cap B, A \cup B \in \mathcal{R}$ for all $A, B \in \mathcal{R}$. Following this, an *algebra* is a ring that contain Ω .

Lemma 1.10. If \mathcal{F} is an algebra that is also a monotone class, then \mathcal{F} is a σ -algebra.

Proof. It suffices to check Definition 1.1 (iii). For any sequence $(A_n)_{n=1}^\infty$ in algebra \mathcal{F} , the partial unions $B_n := \bigcup_{k=1}^n A_k$ form an increasing sequence in \mathcal{F} . Since \mathcal{F} is a monotone class, $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n \in \mathcal{F}$. \square

Theorem 1.11 (Monotone class theorem). Let \mathcal{A} be a algebra of subsets of Ω . Then the monotone class generated by \mathcal{A} coincides with the σ -algebra generated by \mathcal{A} .

Proof. Clearly, a σ -algebra is a monotone class. If we can show that the monotone class $m(\mathcal{A})$ generated by \mathcal{A} is a σ -algebra, then $m(\mathcal{A}) = \sigma(\mathcal{A})$. Following Lemma 1.10, it suffices to show that $m(\mathcal{A})$ is an algebra.

For any $E \in m(\mathcal{A})$, define

$$\mathcal{M}_E = \{ F \in m(\mathcal{A}) : E \setminus F, F \setminus E, E \cup F \in m(\mathcal{A}) \}$$

We claim that $\mathcal{M}_E = m(\mathcal{A})$ for all $E \in m(\mathcal{A})$. For any increasing sequence F_n in \mathcal{M}_E , the sequence $E \setminus F_n$ is decreasing in $m(\mathcal{A})$, and the sequences $F_n \setminus E$ and $E \cup F_n$ are increasing in $m(\mathcal{A})$. Then

$$E \setminus \left(\bigcup_{n=1}^\infty F_n \right) = \bigcap_{n=1}^\infty (E \setminus F_n), \quad \left(\bigcup_{n=1}^\infty F_n \right) \setminus E = \bigcup_{n=1}^\infty (F_n \setminus E), \quad E \cup \left(\bigcup_{n=1}^\infty F_n \right) = \bigcup_{n=1}^\infty (E \cup F_n)$$

are all contained in $m(\mathcal{A})$, and $\bigcup_{n=1}^\infty F_n \in \mathcal{M}_E$. A similar statement holds for decreasing sequences in \mathcal{M}_E . Hence \mathcal{M}_E is a monotone class.

Assume $E \in \mathcal{A}$. Since \mathcal{A} is an algebra, we have $\mathcal{A} \subset \mathcal{M}_E$, which implies $m(\mathcal{A}) \subset \mathcal{M}_E$. Then for all $E \in \mathcal{A}$ and all $F \in m(\mathcal{A})$, we have $F \in \mathcal{M}_E$, which holds if and only if $E \in \mathcal{M}_F$. As a result, we have $\mathcal{A} \subset \mathcal{M}_F$ for all $F \in m(\mathcal{A})$, which again implies $m(\mathcal{A}) \subset \mathcal{M}_F$. Hence for all $E, F \in m(\mathcal{A})$, we have $E \setminus F, F \setminus E, E \cup F \in m(\mathcal{A})$. Clearly, $\Omega \in m(\mathcal{A})$. Hence $m(\mathcal{A})$ is an algebra, as desired. \square

Remark. There is an equivalent statement of Theorem 1.11: If \mathcal{A} is an algebra, and \mathcal{M} is a monotone class such that $\mathcal{M} \supset \mathcal{A}$, then $\sigma(\mathcal{A}) \subset \mathcal{M}$.

1.2 Measures, Pre-measures and Carathéodory's Extension

Definition 1.12 (Measure). Let (Ω, \mathcal{F}) be a measurable space. A (*nonnegative*) *measure* μ on (Ω, \mathcal{F}) is a function $\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}_+ := [0, \infty]$ that satisfies the following:

- (i) $\mu(\emptyset) = 0$;
- (ii) (Countable additivity). If $(A_n)_{n=1}^\infty$ is a sequence of pairwise disjoint sets in \mathcal{F} , then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*. Furthermore, (i) μ is called a *finite measure* if $\mu(\Omega) < \infty$;

(ii) μ is called a *σ -finite measure* if there exists $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{F}$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and $\mu(\Omega_n) < \infty$, $\forall n$;

(iii) μ is called *probability measure* if $\mu(\Omega) = 1$, and $(\Omega, \mathcal{F}, \mu)$ is called a *probability space*.

Remark. A measure μ also has the following properties:

- For all $A, B \in \mathcal{F}$ such that $A \subset B$, it holds $\mu(B \setminus A) = \mu(B) - \mu(A) \leq \mu(B)$.
- For all $A, B \in \mathcal{F}$, it holds $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.
- Using the following Lemma 1.13, we obtain countable subadditivity of μ :

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=1}^n A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(A_k) = \sum_{n=1}^{\infty} \mu(A_n), \quad \forall \text{ sequences } (A_n)_{n=1}^\infty \text{ in } \mathcal{F}.$$

Lemma 1.13. Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space.

- (i) If $(A_n)_{n=1}^\infty$ is an increasing sequence in \mathcal{F} , then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n);$$

- (ii) If $(A_n)_{n=1}^\infty$ is a decreasing sequence in \mathcal{F} such that $\mu(A_1) < \infty$, then

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n);$$

Proof. (i) Define $B_1 = A_1$, and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then $(B_n)_{n=1}^\infty$ is a disjoint sequence of sets in \mathcal{F} . By countable additivity of μ , we have $\mu(A_n) = \sum_{k=1}^n \mu(B_k)$. Hence

$$\lim_{n \rightarrow \infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(B_n) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right).$$

- (ii) Choose an increasing sequence $C_n = A_1 \setminus A_n$. By (i), we have

$$\lim_{n \rightarrow \infty} \mu(C_n) = \mu\left(\bigcup_{n=1}^{\infty} C_n\right). \tag{1.2}$$

Since $\mu(A_1) < \infty$, and since $\bigcup_{n=1}^\infty C_n = A_1 \setminus (\bigcap_{n=1}^\infty A_n)$, the left-hand side of (1.2) is $\mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n)$, and the right-hand side is $\mu(A_1) - \mu(\bigcap_{n=1}^\infty A_n)$. The result follows. \square

Remark. The condition $\mu(A_1) < \infty$ in (ii) cannot be removed. For example, let $A_n = [n, \infty)$, and let μ be the Lebesgue measure. Then $\mu(E_n) = \infty$ for all $n \in \mathbb{N}$, but $\mu(\bigcup_{n=1}^\infty A_n) = \mu(\emptyset) = 0$.

Theorem 1.14 (Borel-Cantelli Lemma). Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If $(A_n)_{n=1}^\infty$ is a sequence of sets in \mathcal{F} such that $\sum_{n=1}^\infty \mu(A_n) < \infty$, then

$$\mu \left(\bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k \right) = 0.$$

In other words, almost all $\omega \in \Omega$ belongs to at most finitely many A_k 's.

Proof. Let $B_n = \bigcup_{k=n}^\infty A_k$ for all n . Then B_n is a decreasing sequence in \mathcal{F} , and $\mu(B_1) \leq \sum_{k=1}^\infty \mu(A_k) < \infty$. By Lemma 1.13,

$$0 \leq \mu \left(\bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k \right) = \mu \left(\bigcup_{n=1}^\infty B_n \right) = \lim_{n \rightarrow \infty} \mu(B_n) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^\infty \mu(A_k) = 0. \quad \square$$

We are going to construct a measure from a ring of subsets of Ω .

Definition 1.15 (Pre-measure). Let \mathcal{A} be a collection of subsets of Ω such that $\emptyset \in \mathcal{A}$. A *pre-measure* on \mathcal{A} is a function $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ satisfying the following:

- (i) $\mu(\emptyset) = 0$;
- (ii) (Countable additivity). If $(A_n)_{n=1}^\infty$ is a sequence of pairwise disjoint sets in \mathcal{A} with $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$, then

$$\mu \left(\bigcup_{n=1}^\infty A_n \right) = \sum_{n=1}^\infty \mu(A_n).$$

Definition 1.16 (Outer measure). An *outer measure* on Ω is a set function $\mu^* : 2^\Omega \rightarrow \overline{\mathbb{R}}_+$ satisfying the following properties:

- (i) $\mu^*(\emptyset) = 0$;
- (ii) (Monotonicity). If $A \subset B$ then $\mu^*(A) \leq \mu^*(B)$;
- (iii) (Countable subadditivity). For any sequence $(A_n)_{n=1}^\infty$ of subsets of Ω , we have

$$\mu^* \left(\bigcup_{n=1}^\infty A_n \right) \leq \sum_{n=1}^\infty \mu^*(A_n).$$

Lemma 1.17 (Induced outer measure). Let \mathcal{R} be a ring of subsets of Ω , and let $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ be a pre-measure on \mathcal{R} . Define $\mu^* : 2^\Omega \rightarrow \overline{\mathbb{R}}_+$ by

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^\infty \mu(A_n) : \{A_n\}_{n=1}^\infty \subset \mathcal{R}, E \subset \bigcup_{n=1}^\infty A_n \right\}, \quad \forall E \in 2^\Omega.$$

Define $\inf \emptyset = \infty$. Then μ^* is an outer measure on Ω , and $\mu^*|_{\mathcal{R}} = \mu$.

Proof. It is easy to check that μ^* is an outer measure. For all $E \in \mathcal{R}$, take $A_1 = E$ and $A_2 = A_3 = \dots = \emptyset$. Then we know $\mu^*(E) \leq \mu(E)$. Hence it remains to show $\mu^*(E) \geq \mu(E)$.

For an arbitrary sequence $(A_n)_{n=1}^\infty$ such that $E \subset \bigcup_{n=1}^\infty A_n$, take $B_1 = A_1$ and $B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right)$ for $n \geq 2$. Then $(B_n)_{n=1}^\infty$ is a disjoint sequence in \mathcal{R} , and we have

$$E \subset \bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n = \bigcup_{n=1}^\infty (E \cap B_n) \Rightarrow \mu(E) = \sum_{n=1}^\infty \mu(E \cap B_n) \leq \sum_{n=1}^\infty \mu(B_n) \leq \sum_{n=1}^\infty \mu(A_n).$$

Hence $\mu^*(E) \geq \mu(E)$. \square

Definition 1.18 (Carathéodory condition). Let \mathcal{R} be a ring of subsets of Ω , and let $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ be a pre-measure on \mathcal{R} . Let μ^* be the outer measure induced by μ . A subset $E \subset \Omega$ is said to be μ^* -measurable if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E), \quad \forall A \subset \Omega. \quad (1.3)$$

Denote by \mathcal{R}^* the collection of all μ^* -measurable sets on Ω .

Remark. To check (1.3), it suffices to check $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \setminus E)$, since the opposite holds by definition. Moreover, for all $E \subset X$ with $\mu^*(E) = 0$, the Carathéodory condition is automatically satisfied.

Proposition 1.19. The collection \mathcal{R}^* given in Definition 1.18 is a σ -algebra.

Proof. It is clear that $\Omega, \emptyset \in \mathcal{R}^*$ and that $E^c \in \mathcal{R}^*$ for all $E \in \mathcal{R}^*$.

Step I: We claim that \mathcal{R}^* is an algebra. Let $E, F \in \mathcal{R}^*$. Then for each $A \subset \Omega$,

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap E) + \mu^*(A \cap E^c) \\ &= \mu^*(A \cap E \cap F) + \mu^*(A \cap E \cap F^c) + \mu^*(A \cap E^c) \\ &\geq \mu^*(A \cap E \cap F) + \mu^*\left(\underbrace{(A \cap E \cap F^c) \cup (A \cap E^c)}_{= A \cap (E^c \cup F^c)}\right) \quad (\text{By subadditivity of } \mu^*) \\ &= \mu^*(A \cap E \cap F) + \mu^*(A \cap (E \cap F)^c) \Rightarrow E \cap F \in \mathcal{R}^*. \end{aligned}$$

Hence \mathcal{R}^* is closed under finite intersections. Note that \mathcal{R}^* is closed under complements, it is also closed under finite unions. Thus \mathcal{R}^* is an algebra, as desired.

Step II: Following Lemma 1.10, it remains to show \mathcal{R}^* is a monotone class. Let $(E_n)_{n=1}^\infty$ be an increasing sequence in \mathcal{R}^* . We want to show that $G := \bigcup_{n=1}^\infty E_n \in \mathcal{R}^*$.

Take $F_1 = E_1$ and $F_n = E_n \setminus E_{n-1}$ for $n \geq 2$. Then $(F_n)_{n=1}^\infty$ is a disjoint sequence in \mathcal{R}^* , and $G = \bigcup_{n=1}^\infty F_n$. For all $A \subset X$ and all $n \in \mathbb{N}$,

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap E_n) + \mu^*(A \cap E_n^c) \geq \mu^*(A \cap E_n) + \mu^*(A \cap G^c) \\ &= \mu^*(A \cap E_{n-1}) + \mu^*(A \cap F_n) + \mu^*(A \cap G^c) \quad (\text{by } F_n \in \mathcal{R}^*) \\ &= \dots = \sum_{k=1}^n \mu^*(A \cap F_k) + \mu^*(A \cap G^c). \quad (\text{by } F_{n-1}, \dots, F_2 \in \mathcal{R}^*) \end{aligned}$$

Therefore $\mu^*(A) \geq \sum_{n=1}^\infty \mu^*(A \cap F_k) + \mu^*(A \cap G^c) = \mu^*(A \cap G) + \mu^*(A \cap G^c)$, and $G \in \mathcal{R}^*$, as desired. (To show that decreasing sequences in \mathcal{R}^* have their limits in \mathcal{R}^* , take the complement.) \square

Proposition 1.20. μ^* is a measure on (Ω, \mathcal{R}^*) .

Proof. It suffices to show countable additivity. Let $(A_n)_{n=1}^\infty$ be a sequence of disjoint sets in \mathcal{R}^* , and let $B_n = \bigcup_{k=1}^n A_k$. Then for all $n \geq 1$,

$$\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \geq \mu^*(B_n) \stackrel{A_n \in \mathcal{R}^*}{=} \mu^*(A_n) + \mu^*(B_{n-1}) \stackrel{A_{n-1} \in \mathcal{R}^*}{=} \dots \stackrel{A_2 \in \mathcal{R}^*}{=} \sum_{k=1}^n \mu^*(A_k).$$

Hence $\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \geq \sum_{n=1}^\infty \mu^*(A_n)$. Since the opposite inequality holds by countable subadditivity of outer measure μ^* , the equality of countable additivity follows. \square

Now we introduce the Carathéodory's extension theorem.

Theorem 1.21 (Carathéodory's extension theorem). Let \mathcal{R} be a ring of subsets of Ω , and let $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ be a pre-measure on \mathcal{R} . Let μ^* and \mathcal{R}^* be given as in Definition 1.18. Then $(\Omega, \mathcal{R}^*, \mu^*)$ is a measure space, and $\mu^*|_{\mathcal{R}} = \mu$. Furthermore, $\mathcal{R} \subset \mathcal{F} := \sigma(\mathcal{R}) \subset \mathcal{R}^*$. As a result, $\mu^*|_{\mathcal{F}}$ is an extension of μ , which is called the *Carathéodory's extension*.

Proof. It suffices to show $\mathcal{R} \subset \mathcal{R}^*$. Fix $E \in \mathcal{R}$, we want to show that

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^c), \quad \forall A \subset \Omega.$$

We can certainly assume $\mu^*(A) < \infty$. Then for all $\epsilon > 0$, there exists a sequence $(F_n)_{n=1}^\infty$ in \mathcal{R} such that $A \subset \bigcup_{n=1}^\infty F_n$ and that

$$\sum_{n=1}^\infty \mu(F_n) \leq \mu^*(A) + \epsilon.$$

Take a disjoint sequence $(G_n)_{n=1}^\infty$ of sets in \mathcal{R} such that $G_1 = F_1$ and $G_n = F_n \setminus \left(\bigcup_{k=1}^{n-1} F_k\right)$ for all $n \geq 2$. Then $\bigcup_{n=1}^\infty G_n = \bigcup_{n=1}^\infty F_n \supset A$, and

$$\begin{aligned} \mu^*(A) + \epsilon &\geq \sum_{n=1}^\infty \mu(G_n) = \sum_{n=1}^\infty \mu(G_n \cap E) + \sum_{n=1}^\infty \mu(G_n \setminus E) \\ &\geq \mu(A \cap E) + \mu(A \setminus E). \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, the result follows. \square

Remark. $(\Omega, \mathcal{R}^*, \mu^*)$ is a complete measure space, since all $E \subset X$ such that $\mu^*(E) = 0$ is contained in \mathcal{R}^* .

We have proved existence of an extension of the pre-measure on a ring. Now we discuss uniqueness.

Lemma 1.22. Let μ and ν be two measures on a measurable space (Ω, \mathcal{F}) . Let $\mathcal{P} \subset \mathcal{F}$ be a π -system such that $\sigma(\mathcal{P}) = \mathcal{F}$ and that $\mu|_{\mathcal{P}} = \nu|_{\mathcal{P}}$.

- (i) If $\mu(\Omega) = \nu(\Omega)$, then $\mu = \nu$;
- (ii) If there exists an increasing sequence $(\Omega_n)_{n=1}^\infty$ of sets in \mathcal{P} such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and $\mu(\Omega_n) = \nu(\Omega_n)$ for all $n \in \mathbb{N}$, then $\mu = \nu$.

Proof. (i) Let $\mathcal{L} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}$. Then \mathcal{L} is a λ -system that contains π -system \mathcal{P} . By Theorem 1.7, $\mathcal{F} \subset \sigma(\mathcal{P}) \subset \mathcal{L} \subset \mathcal{F}$. Hence $\mathcal{L} = \mathcal{F}$, as desired.

(ii) Denote $\mu_n = \mu|_{\Omega_n}$. Using (i), we have $\mu_n = \nu_n$ for all $n \in \mathbb{N}$. Then

$$\mu(A) = \lim_{n \rightarrow \infty} \mu_n(A \cap \Omega_n) = \lim_{n \rightarrow \infty} \nu_n(A \cap \Omega_n) = \nu(A), \quad \forall A \in \mathcal{F}. \quad \square$$

Theorem 1.23 (Uniqueness of extension). Let \mathcal{R} be a ring of subsets of Ω , and let $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ be a pre-measure on \mathcal{R} . If μ is σ -finite, then its extension on $\mathcal{F} = \sigma(\mathcal{R})$ is unique.

Proof. Any ring \mathcal{R} of subsets of Ω is a π -system. Apply Lemma 1.22. \square

Remark. Define the collection \mathcal{A} of subsets of \mathbb{R} which are finite unions of intervals of the following forms: $(-\infty, b]$, $(a, b]$, (a, ∞) , $(-\infty, \infty)$, where $a < b$. Then \mathcal{A} is an algebra. For each $A \in \mathcal{A}$, define $\ell(A)$ to be the length of A . Then $(\mathbb{R}, \mathcal{A}, \ell)$ is a σ -finite pre-measure space. Indeed, the *Lebesgue measure* on \mathbb{R} is obtained by the extension procedure described above.

Definition 1.24 (Semi-ring). A *semi-ring* is a π -system \mathcal{S} of subsets of Ω such that for all $A, B \in \mathcal{S}$, there exists finite collection $\{A_k\}_{k=1}^n \subset \mathcal{S}$ of pairwise disjoint sets such that $A \setminus B = \bigsqcup_{k=1}^n A_k$.

Remark. We can expand a semi-ring \mathcal{S} to a ring by including all finite disjoint unions of sets in \mathcal{S} :

$$\mathcal{R} := \left\{ \bigsqcup_{k=1}^n A_k : n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{S} \text{ are pairwise disjoint} \right\}.$$

Clearly, \mathcal{R} is closed under finite disjoint unions. For all $A, B \in \mathcal{S}$, we have $A \setminus B, B \setminus A \in \mathcal{R}$, and their union $A \cup B = (A \setminus B) \sqcup (A \cap B) \sqcup (B \setminus A) \in \mathcal{R}$. Suppose the union of any $n-1$ sets in \mathcal{S} lies in \mathcal{R} . Then for all $A_1, \dots, A_n \in \mathcal{S}$,

$$\begin{aligned} A_1 \cup \dots \cup A_n &= ((A_1 \cup \dots \cup A_{n-1}) \cup A_n) \sqcup ((A_1 \cup \dots \cup A_{n-1}) \setminus A_n) \sqcup (A_n \setminus (A_1 \cup \dots \cup A_{n-1})) \\ &= \underbrace{\left(\bigcup_{k=1}^{n-1} (A_k \cap A_n) \right)}_{(a)} \sqcup \underbrace{\left(\bigcup_{k=1}^{n-1} (A_k \setminus A_n) \right)}_{(b)} \sqcup \underbrace{\left(\bigcap_{k=1}^{n-1} (A_n \setminus A_k) \right)}_{(c)}. \end{aligned}$$

Note that both (a) and (b) are $(n-1)$ -unions of sets in \mathcal{S} , that (c) is finite intersection of sets in \mathcal{S} , and (a), (b), (c) are disjoint sets, we have $\bigcup_{k=1}^n A_k \in \mathcal{R}$. By induction, any finite union of sets in \mathcal{S} is in \mathcal{R} . Hence \mathcal{R} is closed under finite unions. To show that \mathcal{R} is a ring, it remains to show that \mathcal{R} is closed under finite intersections and differences:

$$\begin{aligned} \left(\bigsqcup_{k=1}^n A_k \right) \cap \left(\bigsqcup_{l=1}^m B_l \right) &= \bigcup_{k=1}^n \bigcup_{l=1}^m \underbrace{(A_k \cap B_l)}_{\in \mathcal{S}} \in \mathcal{R}, \quad \forall \text{ disjoint } \{A_k\}_{k=1}^n, \{B_l\}_{l=1}^m \subset \mathcal{S}, \\ \left(\bigsqcup_{k=1}^n A_k \right) \setminus \left(\bigsqcup_{l=1}^m B_l \right) &= \bigcup_{k=1}^n \bigcap_{l=1}^m \underbrace{(A_k \setminus B_l)}_{\in \mathcal{R}} \in \mathcal{R}, \quad \forall \text{ disjoint } \{A_k\}_{k=1}^n, \{B_l\}_{l=1}^m \subset \mathcal{S}. \end{aligned}$$

Hence \mathcal{R} is a ring of subsets of Ω . Furthermore, we can extend a pre-measure $\mu : \mathcal{S} \rightarrow \overline{\mathbb{R}}_+$ to \mathcal{R} by defining

$$\mu^* \left(\bigcup_{k=1}^n A_k \right) = \sum_{k=1}^n \mu(A_k), \quad \forall \text{ pairwise disjoint } A_1, \dots, A_n \in \mathcal{S}.$$

Then μ^* is a pre-measure on \mathcal{R} , and $\mu^*|_{\mathcal{S}} = \mu$. Applying Carathéodory's extension procedure discussed above, we can also extend a pre-measure space $(\Omega, \mathcal{S}, \mu)$ on a semi-ring to a complete measure space $(\Omega, \sigma(\mathcal{S}), \mu^*)$.

We are going to discuss product measures.

Theorem 1.25 (Product measure). Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two σ -finite measure spaces.

$$\mathcal{F}_1 \times \mathcal{F}_2 := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$

is a collection of measurable rectangles in $\Omega_1 \times \Omega_2$. Define $\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$, which is a σ -algebra of subsets of $\Omega_1 \times \Omega_2$. Then there exists a unique measure μ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \quad \forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2.$$

The measure $\mu_1 \otimes \mu_2 := \mu$ is called the *product measure* on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Moreover, the triple $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$ forms a product measure space.

Proof. We use the Carathéodory's extension theorem to prove this. We check that (i) $\mathcal{F}_1 \times \mathcal{F}_2$ is a semi-ring; and (ii) $\mu_1 \times \mu_2 : A_1 \times A_2 \mapsto \mu_1(A_1)\mu_2(A_2)$ is a pre-measure on $\mathcal{F}_1 \times \mathcal{F}_2$. If (i) and (ii) are satisfied, the existence of an extension on $\mathcal{F}_1 \otimes \mathcal{F}_2$ is ensured.

(i) Let $A = A_1 \times A_2, B = B_1 \times B_2 \in \mathcal{F}_1 \times \mathcal{F}_2$. Then $A \cap B = (A_1 \cap B_1) \times (A_2 \cap B_2) \in \mathcal{F}_1 \times \mathcal{F}_2$, and $\mathcal{F}_1 \times \mathcal{F}_2$ is a π -system. Moreover, $(B_1 \times B_2)^c = (B_1^c \times \Omega_2) \cup (B_1 \times B_2^c)$, and

$$A \setminus B = \underbrace{((A_1 \cap B_1^c) \times A_2)}_{\in \mathcal{F}_1 \times \mathcal{F}_2} \cup \underbrace{((A_1 \cap B_1) \times (A_2 \cap B_2^c))}_{\in \mathcal{F}_1 \times \mathcal{F}_2}$$

Hence $\mathcal{F}_1 \times \mathcal{F}_2$ is a semi-ring.

(ii) Clearly, $(\mu_1 \times \mu_2)(\emptyset) = 0$. Then we need to verify the countable additivity of μ . Let $E \times F \in \mathcal{F}_1 \times \mathcal{F}_2$, and assume there exists disjoint sets $\{E_n \times F_n\}_{n=1}^\infty$ such that $E \times F = \coprod_{n=1}^\infty (E_n \times F_n)$. In other words,

$$\chi_E(x)\chi_F(y) = \sum_{n=1}^\infty \chi_{E_n}(x)\chi_{F_n}(y), \quad \forall x \in \Omega_1, y \in \Omega_2. \quad (1.4)$$

Fix $y \in \Omega_2$. By monotone convergence theorem (MCT, Theorem 1.40), we integrate both sides of (1.4) with respect to x on Ω_1 . Then we obtain $\mu_1(E)\chi_F(y) = \sum_{n=1}^\infty \mu_1(E_n)\chi_{F_n}(y)$. Again by MCT, we have $\mu_1(E)\mu_2(F) = \sum_{n=1}^\infty \mu_1(E_n)\mu_2(F_n)$. Hence $\mu_1 \times \mu_2$ is a pre-measure on $\mathcal{F}_1 \times \mathcal{F}_2$.

Now we show that $\mu_1 \times \mu_2$ is σ -finite, so uniqueness of extension then follows from Theorem 1.23. By σ -finiteness of μ_1 and μ_2 , there exist $\{A_n\}_{n=1}^\infty \subset \mathcal{F}_1$ and $\{B_n\}_{n=1}^\infty \subset \mathcal{F}_2$ such that $\bigcup_{n=1}^\infty A_n = \Omega_1, \bigcup_{n=1}^\infty B_n = \Omega_2$ and $\mu_1(A_n), \mu_2(B_n) < \infty$ for all n . Clearly, $\Omega_1 \times \Omega_2 = \bigcup_{(j,k) \in \mathbb{N}^2} (A_j \times B_k)$, and $(\mu_1 \times \mu_2)(A_j \times B_k)$ is finite for all $(j,k) \in \mathbb{N}^2$. Since \mathbb{N}^2 is countable, $\mu_1 \times \mu_2$ is σ -finite. \square

Remark I. In general, the set of measurable rectangles $\mathcal{F}_1 \times \mathcal{F}_2$ is not a σ -algebra, since it is possibly not closed under complements countable intersections. For example, consider $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$, where $\mathcal{B}(\mathbb{R})^2 = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. The union of $(0,1) \times (0,1)$ and $(-1,0) \times (-1,0)$ is not in $\mathcal{B}(\mathbb{R})^2$.

Remark II. Let (X_1, \mathcal{T}_1) and (X_2, \mathcal{T}_2) are two second-countable topological spaces. The *product topology* $\mathcal{T}_1 \otimes \mathcal{T}_2$ is the topology generated by all open rectangles $\mathcal{T}_1 \times \mathcal{T}_2$. Let $\mathcal{B}_1 = \sigma(\mathcal{T}_1)$ and $\mathcal{B}_2 = \sigma(\mathcal{T}_2)$ be the Borel σ -algebras generated by \mathcal{T}_1 and \mathcal{T}_2 , respectively. Then the σ -algebras generated by the product topology $\mathcal{T}_1 \otimes \mathcal{T}_2$ and by Borel rectangles $\mathcal{B}_1 \times \mathcal{B}_2$ coincide. In a nutshell, $\sigma(\mathcal{T}_1) \otimes \sigma(\mathcal{T}_2) = \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$.

Proof. Given $A \in \mathcal{B}_1$, let \mathcal{V}_A be the collection of all $B \subset X_2$ such that $A \times B \in \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$. Clearly, \mathcal{V}_A is a σ -algebra of subsets of X_2 , and it contains all open sets in X_2 . Hence $\mathcal{B}_2 \subset \mathcal{V}_A$. Similarly, for $B \in \mathcal{B}_2$, the collection \mathcal{U}_B of all $A \subset X_1$ such that $A \times B \in \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$ is a σ -algebra containing \mathcal{B}_1 . As a result, $\sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$ contains all Borel rectangles $\mathcal{B}_1 \times \mathcal{B}_2$, hence contains $\sigma(\mathcal{B}_1 \times \mathcal{B}_2)$.

In the other direction, let $\{U_m\}_{m \in \mathbb{N}}$ be a topological basis for X_1 , and $\{V_n\}_{n \in \mathbb{N}}$ a topological basis for X_2 . Then the collection $\mathcal{A} = \{U_m \times V_n\}_{m,n \in \mathbb{N}}$ is a topological basis for the product space $X_1 \times X_2$. Furthermore, any open set in $X_1 \times X_2$ is a union of these basis elements, which must be countable. Hence the σ -algebra generated by \mathcal{A} contains all open sets in $X_1 \times X_2$, and $\sigma(\mathcal{A}) \supset \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$. On the other hand, note that $\mathcal{A} \subset \mathcal{T}_1 \times \mathcal{T}_2$, which is the set of all open rectangles in $X_1 \times X_2$, we have $\sigma(\mathcal{A}) \subset \sigma(\mathcal{T}_1 \times \mathcal{T}_2) \subset \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$. Furthermore, since $\mathcal{A} \subset \mathcal{B}_1 \times \mathcal{B}_2$, we have $\sigma(\mathcal{A}) \subset \sigma(\mathcal{B}_1 \times \mathcal{B}_2)$.

To summarize, $\sigma(\mathcal{T}_1) \otimes \sigma(\mathcal{T}_2) = \sigma(\mathcal{B}_1 \times \mathcal{B}_2) = \sigma(\mathcal{A}) = \sigma(\mathcal{T}_1 \otimes \mathcal{T}_2)$. \square

Since the real line \mathbb{R} given the standard topology is second-countable, we have $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

1.3 Measurable Functions and Simple Function Approximation

Definition 1.26. Given a function $T : \Omega_1 \rightarrow \Omega_2$ and a subset $A \subset \Omega_2$, define

$$T^{-1}A = \{\omega \in \Omega_1 : T\omega \in A\}$$

to be the *inverse image* of A . If \mathcal{A} is a collection of subsets of Ω_2 , define $T^{-1}\mathcal{A} = \{T^{-1}A : A \in \mathcal{A}\}$.

Proposition 1.27. Let $T : \Omega_1 \rightarrow \Omega_2$. It is easy to verify the following properties of T .

- (i) $T^{-1}\Omega_2 = \Omega_1$, $T^{-1}\emptyset = \emptyset$;
- (ii) For all $A \subset \Omega_2$, $T^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus T^{-1}A$.
- (iii) If $\{A_\alpha\}_{\alpha \in J}$ is a collection of subsets of Ω_2 , then $T^{-1}(\bigcup_{\alpha \in J} A_\alpha) = \bigcup_{\alpha \in J} T^{-1}A_\alpha$.
- (iv) If \mathcal{F} is a σ -algebra of subsets of Ω_2 , then $T^{-1}\mathcal{F}$ is again a σ -algebra.

Definition 1.28 (Measurable functions). Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. A function $T : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ is said to be a *measurable function* if $T^{-1}\mathcal{F}_2 \subset \mathcal{F}_1$. In other words, the inverse image of every \mathcal{F}_2 -measurable set in Ω_2 is \mathcal{F}_1 -measurable.

Remark. By definition, we can immediately verify that the composition $T \circ S$ of two measurable functions $(\Omega_1, \mathcal{F}_1) \xrightarrow{S} (\Omega_2, \mathcal{F}_2) \xrightarrow{T} (\Omega_3, \mathcal{F}_3)$ is measurable.

Lemma 1.29 (Pushforward measure). Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. If $\mu : \mathcal{F}_1 \rightarrow \overline{\mathbb{R}}_+$ is a measure on $(\Omega_1, \mathcal{F}_1)$, and $T : \Omega_1 \rightarrow \Omega_2$ is a measurable function, then $T_*\mu : \mathcal{F}_2 \rightarrow \overline{\mathbb{R}}_+$, $A \mapsto \mu(T^{-1}A)$ is a measure on $(\Omega_2, \mathcal{F}_2)$, called the *pushforward* of μ .

Proof. This lemma immediately follows from Proposition 1.27 (i) and (iii). □

Remark. A function $T : (\Omega_1, \mathcal{F}_1, \mu_1) \rightarrow (\Omega_2, \mathcal{F}_2, \mu_2)$ is said to be *measure preserving* if $\mu_2 = T_*\mu_1$. In other words, the measure of any measurable set $A \in \mathcal{F}_2$ does not change after inverse transformation.

Definition 1.30 (Measurable real-valued functions). Let (Ω, \mathcal{F}) be a measurable space.

- (i) A real-valued function $f : \Omega \rightarrow \mathbb{R}$ is said to be *measurable* if $f^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$. In other words, the function $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable.
- (ii) An extended real-valued function $f : \Omega \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is said to be *measurable* if the sets $\{\omega : f(\omega) = -\infty\}$ and $\{\omega : f(\omega) = \infty\}$ are measurable, and the following real-valued function is measurable:

$$f_1(\omega) = \begin{cases} f(\omega), & \text{if } f(\omega) \in \mathbb{R}; \\ 0, & \text{otherwise} \end{cases}$$

Remark. We can generalize (i) to any topological space (X, \mathcal{T}) , where a Borel σ -algebra can be defined.

Proposition 1.31. Let (Ω, \mathcal{F}) be a measurable space, and $f : \Omega \rightarrow \mathbb{R}$. The following are equivalent:

- (i) $\{\omega : f(\omega) > \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$;
- (ii) $\{\omega : f(\omega) \geq \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$;
- (iii) $\{\omega : f(\omega) < \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$;
- (iv) $\{\omega : f(\omega) \leq \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$;
- (v) f is a measurable function.

Proof. Clearly (i) and (iii) are equivalent. It is easy to see that (i) and (ii) are equivalent, since

$$\{\omega : f(\omega) > \alpha\} = \bigcap_{n=1}^{\infty} \left\{ \omega : f(\omega) \geq \alpha + \frac{1}{n} \right\}, \quad \{\omega : f(\omega) \geq \alpha\} = \bigcap_{n=1}^{\infty} \left\{ \omega : f(\omega) > \alpha - \frac{1}{n} \right\}.$$

Similarly (iii) and (iv) are equivalent. Then it remains to show (i)-(iv) \Rightarrow (v).

Let $\mathcal{A} = \{A \subset \mathbb{R} : f^{-1}(A) \in \mathcal{F}\}$. Clearly, \mathcal{A} is a σ -algebra. Then it suffices to show that \mathcal{A} contains all open intervals: for all $\alpha < \beta$, $f^{-1}((\alpha, \beta)) = \{\omega : f(\omega) < \beta\} \cap \{\omega : f(\omega) > \alpha\} \in \mathcal{F}$. \square

Remark. By definition, all constant functions, indicator functions, continuous functions (the inverse images of open sets remain open) and monotone functions on \mathbb{R} are measurable. Furthermore, this proposition remains true for extended real-valued functions $f : \Omega \rightarrow \overline{\mathbb{R}}$.

Definition 1.32. Given a function $f : \Omega \rightarrow \mathbb{R}$, define $f^+ = \max\{f, 0\}$ to be the *positive part* of f , and define $f^- = \max\{-f, 0\}$ to be the *negative part* of f . Then we have

$$f = f^+ - f^-, \quad |f| = f^+ + f^-.$$

Proposition 1.33. Let (Ω, \mathcal{F}) be a measurable space. Let f and g be two real-valued measurable functions. Let $\alpha \in \mathbb{R}$. The following functions are measurable: f^+ , f^- , $|f|$, αf , $f + g$, fg .

Proof. Clearly, f^+ , f^- , $|f|$, $|f|^2$ and αf are measurable. To show $f + g$ is measurable, note that

$$\{\omega : f(\omega) + g(\omega) > \alpha\} = \bigcup_{r_n \in \mathbb{Q}} \{\omega : f(\omega) > r_n\} \cap \{\omega : g(\omega) > \alpha - r_n\} \in \mathcal{B}(\mathbb{R}).$$

To show fg is measurable, note that $(f + g)^2 - |f|^2 - |g|^2 = 2fg$ is measurable. \square

Remark. The proposition also holds for extended real-valued f and g . (Note $f + g$ should be well-defined.)

Proposition 1.34. Given a measurable space (Ω, \mathcal{F}) and a sequence of measurable functions $f_n : \Omega \rightarrow \overline{\mathbb{R}}$, $n \in \mathbb{N}$, then following functions are also measurable:

$$g_1(\omega) = \sup_{n \geq 1} f_n(\omega), \quad g(\omega) = \limsup_{n \rightarrow \infty} f_n(\omega), \quad h_1(\omega) = \inf_{n \geq 1} f_n(\omega), \quad h(\omega) = \liminf_{n \rightarrow \infty} f_n(\omega).$$

Proof. Define $g_k(\omega) = \sup_{n \geq k} f_n(\omega)$, $k \in \mathbb{N}$. Then $(g_k)_{k=1}^{\infty}$ is a decreasing sequence. For all $\alpha \in \mathbb{R}$,

$$\{\omega : g_k(\omega) \geq \alpha\} = \bigcap_{n=k}^{\infty} \{\omega : f_n(\omega) \geq \alpha\} \in \mathcal{F}.$$

Hence g_k is measurable. Similarly, $h_k(\omega) = \inf_{n \geq k} f_n(\omega)$ is an increasing sequence of measurable functions. Furthermore,

$$g(\omega) = \lim_{k \rightarrow \infty} g_k(\omega) = \inf_{k \geq 1} g_k(\omega), \quad h(\omega) = \lim_{k \rightarrow \infty} h_k(\omega) = \sup_{k \geq 1} h_k(\omega)$$

are also measurable. \square

Remark. Following the result above, If $\{f_n : \Omega \rightarrow \overline{\mathbb{R}}, n \in \mathbb{N}\}$ is a sequence of measurable functions that converges pointwise to a function $f : \Omega \rightarrow \overline{\mathbb{R}}$, then f is also measurable.

Sometimes we are also interested in the measurability of vector-valued functions.

Theorem 1.35. Let (Ω, \mathcal{F}) be a measurable space. Let X and Y be two second-countable measurable spaces. A vector-valued function $f = (f_X, f_Y) : \Omega \rightarrow X \times Y$ is measurable if and only if its two components f_X and f_Y are both measurable.

Proof. If $f = (f_X, f_Y)$ is measurable, consider the projection map $\pi_X : X \times Y \rightarrow X$, $(x, y) \mapsto x$. Clearly, π_X is continuous, hence is measurable. Then $f_X = \pi_X \circ f$ is measurable. The same holds for f_Y .

Conversely, let $\{U_m\}_{m=1}^\infty$ be a topological basis for X , and $\{V_n\}_{n=1}^\infty$ a topological basis for Y . For an open set W in $X \times Y$, it can be written as a countable union of some basis elements:

$$W = \bigcup_{k=1}^\infty U_{m_k} \times V_{n_k} \Rightarrow f^{-1}(W) = \bigcup_{k=1}^\infty (f_X^{-1}(U_{m_k}) \cap f_Y^{-1}(V_{n_k})).$$

Since f_X and f_Y are measurable, $f^{-1}(W) \in \mathcal{F}$ for all open set $W \subset X \times Y$. Since f preserves set operations (intersection, union and complement), we have $f^{-1}(W) \in \mathcal{F}$ for all Borel set W in $X \times Y$. \square

Remark. By induction, a real-vector-valued function $f = (f_1, \dots, f_n)$ is measurable if and only if each of its components f_k is measurable.

Now we introduce simple function approximation.

Theorem 1.36 (Simple function approximation). Let (Ω, \mathcal{F}) be a measurable space. A (measurable) simple function φ is a finite linear combination of indicator functions of measurable sets. That is, there exists $A_1, \dots, A_n \in \mathcal{F}$ and $c_1, \dots, c_n \in \mathbb{R}$ such that

$$\varphi = \sum_{k=1}^n c_k \chi_{A_k}. \quad (1.5)$$

Let (Ω, \mathcal{F}) be a measurable space, and let $f : \Omega \rightarrow \mathbb{R}$ be a nonnegative measurable function. Then there exists an increasing sequence $(\varphi_n)_{n=1}^\infty$ of measurable functions such that $f(\omega) = \lim_{n \rightarrow \infty} \varphi_n(\omega)$ for all $\omega \in \Omega$. Namely, φ_n converges pointwise to f . Furthermore, if there exists $M > 0$ such that $f(\omega) \leq M$ for all $\omega \in \Omega$, then we are able to choose φ_n that converges uniformly to f .

Proof. For each $n \in \mathbb{N}$ and $0 \leq k < 4^n$, define

$$E_{n,k} = \{\omega : 2^{-n}k \leq f(\omega) < 2^{-n}(k+1)\}, \quad E_{n,4^n} = \{\omega : f(\omega) \geq 2^{-n}\}.$$

Then choose a nonnegative measurable simple function φ_n as follows:

$$\varphi_n = \sum_{k=0}^{4^n} \frac{k}{2^n} \chi_{E_{n,k}} \Rightarrow \varphi_n(\omega) = \max_{k \in \{0,1,\dots,4^n\}} \{2^{-n}k : 2^{-n}k \leq f(\omega)\}, \quad \forall \omega \in \Omega.$$

Clearly, φ_n is increasing, and $\varphi_n(\omega) \rightarrow f(\omega)$ for all $\omega \in \Omega$. If there exists $M > 0$ such that $f(\omega) \leq M$ for all $\omega \in \Omega$, then $E_{n,4^n} = \emptyset$ once $2^n > N$, and $|f(\omega) - \varphi_n(\omega)| < 2^{-n}$ for all $\omega \in \Omega$. \square

Remark. If f is a measurable function, we can extract its positive part $f^+ = \max\{f, 0\}$ and negative part $f^- = \max\{-f, 0\}$. By approaching f^+ and f^- respectively, we obtain a simple function approximation for f .

In proper conditions, a pointwise convergent sequence of measurable functions is nearly uniformly convergent.

Theorem 1.37 (Egoroff). Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. Let $f_n : \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}$ be a sequence of measurable functions that converges μ -almost everywhere to $f : \Omega \rightarrow \mathbb{R}$. Then for all $\epsilon > 0$, there exists $E \subset \mathcal{F}$ such that $\mu(\Omega \setminus E) < \epsilon$ and that f_n converges to f uniformly on E .

Proof. Choose $\Omega_0 \in \mathcal{F}$ such that $\mu(\Omega \setminus \Omega_0) = 0$ and $f_n(x) \rightarrow f(x)$ everywhere on Ω_0 . For all $n, k \in \mathbb{N}$, define

$$A_{k,n} := \left\{ \omega : |f_n(\omega) - f(\omega)| \geq \frac{1}{k} \right\}, \quad B_{k,n} := \bigcup_{j=n}^{\infty} A_{k,j}, \quad A_k := \bigcap_{n=1}^{\infty} B_{k,n}.$$

If $\omega_0 \in \Omega_0$, there exists $N > 0$ such that $|f_n(\omega_0) - f(\omega_0)| < k^{-1}$ for all $n \geq N$. Then $\omega_0 \notin B_{k,N}$, and $\omega_0 \notin A_k$ for all $k \in \mathbb{N}$. This implies $\bigcup_{k=1}^{\infty} A_k \subset \Omega \setminus \Omega_0$. Since μ is finite, we have

$$\lim_{n \rightarrow \infty} \mu(B_{k,n}) = \mu\left(\bigcap_{n=1}^{\infty} B_{k,n}\right) = \mu(A_k) = 0 \Rightarrow \exists N_k > 0 \text{ such that } \mu(B_{k,N_k}) < 2^{-k}\epsilon.$$

Let $E = \Omega \setminus (\bigcup_{k=1}^{\infty} B_{k,N_k}) \in \mathcal{F}$. Then $\mu(\Omega \setminus E) < \epsilon$. Furthermore, for all $\omega \in E$, $\omega \notin B_{k,N_k}$ for all $k \in \mathbb{N}$. In other words, given any $k \in \mathbb{N}$, we have $|f_n(\omega) - f(\omega)| < k^{-1}$ for all $n \geq N_k$ and all $\omega \in E$. Hence E is the desired set on which f_n converges uniformly to f . \square

1.4 Lebesgue Integration

If not specifically indicated, our discussion is based on the measure space $(\Omega, \mathcal{F}, \mu)$.

Definition 1.38. (Lebesgue integral for nonnegative measurable functions). A simple function $\varphi : \Omega \rightarrow \mathbb{R}$ takes only finitely many values $a_1, \dots, a_n \in \mathbb{R}$. Hence it has the unique *standard expression*:

$$\varphi = \sum_{k=1}^n a_k \chi_{A_k}, \text{ where } A_k = \{\omega : \varphi(\omega) = a_k\}. \quad (1.6)$$

For any nonnegative simple function $\varphi : \Omega \rightarrow \mathbb{R}_+$ with standard expression (1.6), we define

$$S(\varphi) = \sum_{k=1}^n a_k \mu(A_k).$$

Given a nonnegative measurable function $f : \Omega \rightarrow \overline{\mathbb{R}}_+$, define its *Lebesgue integral* on Ω as follows:

$$\int f \, d\mu = \sup \{S(\varphi) : 0 \leq \varphi \leq f, \varphi \text{ is a measurable simple function}\}.$$

In addition, given $A \in \mathcal{F}$, define

$$\int_A f \, d\mu = \int f \chi_A \, d\mu.$$

Remark. Recall the convention $\infty \cdot 0 = 0$. These integrals are well defined but may take value ∞ . Moreover, we can immediately check that $\int \varphi \, d\mu = S(\varphi)$ for all measurable simple functions φ .

Proposition 1.39. If f and g are nonnegative measurable functions such that $f \leq g$, then $\int f \, d\mu \leq \int g \, d\mu$. It is to check by definition of Lebesgue integral.

We have the following important theorem, which ensures the interchangeability of limit and integration.

Theorem 1.40 (Monotone convergence theorem/Levi's theorem). Let $(f_n)_{n=1}^\infty$ be a monotone increasing sequence of nonnegative measurable functions, and let $f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ for all $\omega \in \Omega$. Then

$$\int f \, d\mu = \lim_{n \rightarrow \infty} \int f_n \, d\mu.$$

Proof. By Proposition 1.39, $\int f_n \, d\mu$ is monotone increasing, and

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \sup_{n \geq 1} \int f_n \, d\mu \leq \int f \, d\mu. \quad (1.7)$$

Now we prove the opposite. Let $0 < \alpha < 1$, and let φ be any simple function such that $0 \leq \varphi \leq f$. Take $A_n := \{\omega : f_n(\omega) \geq \alpha \varphi(\omega)\}$, which is an increasing sequence in \mathcal{F} such that $\Omega = \bigcup_{n=1}^\infty A_n$. Note that f_n is nonnegative, and φ is simple. Then

$$\int f_n \, d\mu \geq \int_{A_n} f_n \, d\mu \geq \alpha \int_{A_n} \varphi \, d\mu \xrightarrow{n \rightarrow \infty} \lim_{n \rightarrow \infty} \int f_n \, d\mu \geq \alpha \int \varphi \, d\mu \xrightarrow{\alpha \rightarrow 1} \lim_{n \rightarrow \infty} \int f_n \, d\mu \geq \int \varphi \, d\mu.$$

Since the simple function $0 \leq \varphi \leq f$ is arbitrary, by definition of Lebesgue integral, we complete the proof of the opposite of (1.7). \square

Proposition 1.41. For all nonnegative measurable functions f and g and all $\alpha, \beta \in \mathbb{R}_+$,

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Proof. The equality is clear when f and g is simple. In general case, use simple function approximation and monotone convergence theorem. \square

Proposition 1.42. Let f and g be nonnegative measurable functions. Then

$$\int f d\mu = 0 \Leftrightarrow f = 0 \text{ a.e..}$$

Furthermore, if $f = g$ a.e., then

$$\int f d\mu = \int g d\mu.$$

Proof. Let $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ be a simple function such that $0 \leq \varphi \leq f$. If $f = 0$ a.e., then either $\mu(A_k) = 0$ or $a_k = 0$ for each k , which implies $S(\varphi) = 0$. By definition of Lebesgue integral, $\int f d\mu = 0$.

Now assume $\int f d\mu = 0$. Take $E = \{\omega : f(\omega) > 0\}$ and $E_n = \{\omega : f(\omega) > n^{-1}\}$ for all $n \in \mathbb{N}$. Then we have $E = \bigcup_{n=1}^{\infty} E_n$, and

$$0 \leq \mu(E) = \lim_{n \rightarrow \infty} \mu(E_n) \leq \lim_{n \rightarrow \infty} n \int_{E_n} f d\mu \leq \lim_{n \rightarrow \infty} n \int f d\mu = 0 \Rightarrow \mu(E) = 0, f = 0 \text{ a.e..}$$

Finally assume $f = g$ a.e.. Take $h = \max\{f, g\}$, then $h - f$ is a nonnegative measurable function, and $h - f = 0$ a.e.. As a result, $\int h d\mu = \int f d\mu$. Similarly, we have $\int h d\mu = \int g d\mu$. \square

The Fatou's lemma is useful when we do not know whether limit and integration are interchangeable.

Theorem 1.43 (Fatou's lemma). Let $(f_n)_{n=1}^{\infty}$ be a sequence of nonnegative measurable functions, and let $f(\omega) = \liminf_{n \rightarrow \infty} f_n(\omega)$ for all $\omega \in \Omega$. Then

$$\int f d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu. \quad (1.8)$$

Proof. Let $g_n(\omega) = \inf_{k \geq n} f_k(\omega)$. Then g is measurable, and

$$\int g_n d\mu \leq \int f_k d\mu, \forall k \geq n \Rightarrow \int g_n d\mu \leq \inf_{k \geq n} \int f_k d\mu.$$

Furthermore, $(g_n)_{n=1}^{\infty}$ is a monotone increasing sequence converging to f . By monotone convergence theorem,

$$\int f d\mu = \lim_{n \rightarrow \infty} \int g_n d\mu \leq \int \lim_{n \rightarrow \infty} \inf_{k \geq n} f_k d\mu.$$

This is indeed the inequality (1.8). \square

Remark. Even though $f = \lim_{n \rightarrow \infty} f_n$ (pointwise), limit and integration are not interchangeable in general. For example, let $f_n = n \chi_{[0, n^{-1}]}$. Then $f = \lim_{n \rightarrow \infty} f_n = \infty \chi_{\{0\}}$, but

$$0 = \int f d\mu < \lim_{n \rightarrow \infty} \int f_n d\mu = 1.$$

Definition 1.44 (Lebesgue integrable functions). A measurable function f is said to be *integrable* if

$$\int f^+ d\mu < \infty \quad \text{and} \quad \int f^- d\mu < \infty.$$

We denote by $L^1(\Omega, \mathcal{F}, \mu)$ the set of all integrable functions. The *Lebesgue integral* of f is defined as

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \in \mathbb{R}.$$

In addition, given $A \in \mathcal{F}$, define

$$\int_A f d\mu = \int f \chi_A d\mu.$$

Remark. A measurable function f is said to be *quasi-integrable* if at least one of f^+ and f^- is integrable. In this case, the Lebesgue integral of f takes value in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Proposition 1.45 (Linearity of Lebesgue integral). For all $f, g \in L^1(\Omega, \mathcal{F}, \mu)$ and all $\alpha, \beta \in \mathbb{R}$,

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Proof. By Proposition 1.41,

$$\int \alpha f d\mu = \int (\alpha f)^+ d\mu - \int (\alpha f)^- d\mu = \alpha \int f^+ d\mu - \alpha \int f^- d\mu = \alpha \int f d\mu.$$

Let $A = \{f \geq 0, g \geq 0\}$, $B = \{f < 0, g < 0\}$, $P_1 = \{f \geq 0, g < 0, f + g \geq 0\}$, $P_2 = \{f < 0, g \geq 0, f + g \geq 0\}$, $N_1 = \{f < 0, g \geq 0, f + g < 0\}$, $N_2 = \{f \geq 0, g < 0, f + g < 0\}$. Then $\Omega = A \cup B \cup P_1 \cup P_2 \cup N_1 \cup N_2$, and

$$\begin{aligned} \int (f + g) d\mu &= \int (f + g)^+ d\mu - \int (f + g)^- d\mu \\ &= \int_A (f^+ + g^+) d\mu + \int_{P_1} (f^+ - g^-) d\mu + \int_{P_2} (g^+ - f^-) d\mu \\ &\quad - \int_B (f^- + g^-) d\mu - \int_{N_1} (f^- - g^+) d\mu - \int_{N_2} (g^- - f^+) d\mu \\ &= \int_A f^+ d\mu + \int_{P_1} f^+ d\mu + \int_{N_2} f^+ d\mu - \int_B f^- d\mu - \int_{P_2} f^- d\mu - \int_{N_1} f^- d\mu \\ &\quad + \int_A g^+ d\mu + \int_{P_2} g^+ d\mu + \int_{N_1} g^+ d\mu - \int_B g^- d\mu - \int_{P_1} g^- d\mu - \int_{N_2} g^- d\mu \\ &= \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu = \int f d\mu + \int g d\mu. \end{aligned}$$

Hence the linearity follows. \square

Proposition 1.46. Let f be a measurable function. Then $f \in L^1(\Omega, \mathcal{F}, \mu)$ if and only if $|f| \in L^1(\Omega, \mathcal{F}, \mu)$. In that case,

$$\left| \int f d\mu \right| \leq \int |f| d\mu. \quad (1.9)$$

Proof. Note that $f = f^+ - f^-$, $|f| = f^+ + f^-$. Then f and $|f|$ is integrable if and only if f^+ and f^- is integrable. Moreover, (1.9) follows from the triangle inequality. \square

The Lebesgue's dominated convergence theorem ensures interchangeability of limit and integration.

Theorem 1.47. (Lebesgue's dominated convergence theorem). Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions such that $f_n \rightarrow f$ a.e., where f is also a measurable function. If there exists $g \in L^1(\Omega, \mathcal{F}, \mu)$ such that $|f_n| \leq g$ for all $n \in \mathbb{N}$, then all functions f_n and f are integrable, and

$$\int f \, d\mu = \lim_{n \rightarrow \infty} \int f_n \, d\mu.$$

Proof. We may assume $f_n \rightarrow f$ pointwise by redefining f and f_n on a set of measure zero. For all $n \in \mathbb{N}$, we have $|f_n| \leq g$, then $|f| \leq g$. By Proposition 1.46, all functions f_n and f are integrable.

Since $g + f_n \geq 0$, and $g - f_n \geq 0$, apply Fatou's lemma (Theorem 1.43):

$$\begin{aligned} \int (g + f) \, d\mu &\leq \liminf_{n \rightarrow \infty} \int (g + f_n) \, d\mu = \int g \, d\mu + \liminf_{n \rightarrow \infty} \int f_n \, d\mu, \\ \int (g - f) \, d\mu &\leq \liminf_{n \rightarrow \infty} \int (g - f_n) \, d\mu = \int g \, d\mu - \limsup_{n \rightarrow \infty} \int f_n \, d\mu. \end{aligned}$$

Hence we have

$$\limsup_{n \rightarrow \infty} \int f_n \, d\mu \leq \int f \, d\mu \leq \liminf_{n \rightarrow \infty} \int f_n \, d\mu,$$

and the result follows. \square

Remark. If $(\Omega, \mathcal{F}, \mu)$ is a complete measure space, then f is automatically measurable. Inspired by this proof, we summarize another commonly used version of Fatou's lemma as follows.

Corollary 1.48. (Fatou's lemma). Let $(f_n)_{n=1}^{\infty}$ be a sequence of integrable functions.

(i) If there exists an integrable function g such that $f_n \geq g$ for all $n \in \mathbb{N}$, then

$$\int \liminf_{n \rightarrow \infty} f_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int f_n \, d\mu$$

(ii) If there exists an integrable function g such that $f_n \leq g$ for all $n \in \mathbb{N}$, then

$$\limsup_{n \rightarrow \infty} \int f_n \, d\mu \leq \int \limsup_{n \rightarrow \infty} f_n \, d\mu.$$

Now we discuss integral transform among different measure spaces.

Theorem 1.49 (Integral transform). Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two measure spaces. If function $T : (\Omega_1, \mathcal{F}_1, \mu_1) \rightarrow (\Omega_2, \mathcal{F}_2, \mu_2)$ is measure-preserving, i.e. $\mu_2 = T_*\mu_1$ is the pushforward of μ_1 , then

$$\int f \circ T \, d\mu_1 = \int f \, d\mu_2, \quad \forall f \in L^1(\Omega_2, \mathcal{F}_2, \mu_2). \quad (1.10)$$

Proof. For all $A \in \mathcal{F}_2$, we have

$$\int \chi_A \circ T \, d\mu_1 = \mu_1(T^{-1}A) = \mu_2(A) = \int \chi_A \, d\mu_2.$$

Then (1.10) holds for all nonnegative measurable simple functions f . Similar to the procedure of defining Lebesgue integral (Definition 1.38 and Definition 1.44), it holds for all $f \in L^1(\Omega_2, \mathcal{F}_2, \mu_2)$. \square

Now we discuss Lebesgue integral on product spaces. We first present the conclusion in Theorem 1.50.

Theorem 1.50 (Fubini's theorem). Let $f \in L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$.

- (i) For all $\omega_1 \in \Omega_1$, the function $\omega_2 \mapsto f(\omega_1, \omega_2)$ is integrable.
- (ii) The function $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2)$ is integrable.
- (iii) For all $\omega_2 \in \Omega_2$, the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is integrable.
- (iv) The function $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$ is integrable.
- (v) The following integrals are equivalent:

$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).$$

The proof of Fubini's theorem uses Tonelli theorem. We first prove the following proposition.

Proposition 1.51. Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ -finite measure spaces, and denote their product space by $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$. For all $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$, we define the *slices*

$$A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\} \text{ and } A^{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}.$$

Then the following hold for all $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$:

- (i) $A_{\omega_1} \in \mathcal{F}_2$ for all $\omega_1 \in \Omega_1$;
- (ii) The function $f_A : \omega_1 \mapsto \mu_2(A_{\omega_1})$ is measurable;
- (iii) $(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_1} \mu_2(A_{\omega_1}) d\mu_1(\omega_1) := \int_{\Omega_1} f_A d\mu_1$.

Symmetric statements also holds for slices A^{ω_2} .

Proof. Denote by \mathcal{M} the collection of all subsets of $\Omega_1 \times \Omega_2$ which satisfy (i), (ii) and (iii). We prove that $\mathcal{M} \supset \mathcal{F}_1 \otimes \mathcal{F}_2$. Clearly, all measurable rectangles in $\Omega_1 \times \Omega_2$ satisfy (i), (ii) and (iii). Hence $\mathcal{M} \supset \mathcal{F}_1 \times \mathcal{F}_2$.

Step I: We prove that for any increasing sequence $(A_n)_{n=1}^\infty$ of sets in \mathcal{M} , it holds $A := \bigcup_{n=1}^\infty A_n \in \mathcal{M}$.

- (i) For all $\omega_1 \in \Omega_1$, we have

$$A_{\omega_1} = \left(\bigcup_{n=1}^\infty A_n \right)_{\omega_1} = \bigcup_{n=1}^\infty (A_n)_{\omega_1} \in \mathcal{F}_2;$$

- (ii) Note that $(A_n)_{\omega_1}$ is an increasing sequence, we have

$$\mu_2(A_{\omega_1}) = \lim_{n \rightarrow \infty} \mu_2((A_n)_{\omega_1}) \Rightarrow f_A = \lim_{n \rightarrow \infty} f_n \text{ is measurable.}$$

- (iii) Note that f_{A_n} is monotone increasing, by monotone convergence theorem,

$$(\mu_1 \otimes \mu_2)(A) = \lim_{n \rightarrow \infty} (\mu_1 \otimes \mu_2)(A_n) = \lim_{n \rightarrow \infty} \int_{\Omega_1} f_{A_n} d\mu_1 = \int_{\Omega_1} f d\mu_1.$$

Step II: Similar to Step I, we can prove that for any decreasing sequence $(B_n)_{n=1}^\infty$ of sets in \mathcal{M} such that $\mu_1((B_1)_{\omega_1}) < \infty$ for all $\omega_1 \in \Omega_1$ and $(\mu_1 \otimes \mu_2)(B_1) < \infty$, it holds $B := \bigcap_{n=1}^\infty B_n \in \mathcal{M}$.

Step III: We prove that for any sequence $(E_n)_{n=1}^\infty$ of disjoint sets in \mathcal{M} , it holds $\bigcup_{n=1}^\infty E_n \in \mathcal{M}$. Clearly, if E, F are disjoint sets in \mathcal{M} , we have $E \cup F \in \mathcal{M}$. Then our result immediately follows from Step I by choosing increasing sequence $A_n := \bigcup_{k=1}^n E_k$ in \mathcal{M} .

Step IV: Denote by \mathcal{A} the collection of all finite unions of measurable rectangles in $\Omega_1 \times \Omega_2$. Then \mathcal{A} is an algebra. By σ -finiteness of Ω_1 , choose an increasing sequence $(X_n)_{n=1}^\infty$ such that $\mu_1(X_n) < \infty$ for all $n \in \mathbb{N}$ and $\Omega_1 = \bigcup_{n=1}^\infty X_n$. Similarly choose an increasing sequence $(Y_n)_{n=1}^\infty$ for Ω_2 .

Let $\mathcal{A}_n = \{A \in \mathcal{A} : A \subset X_n \times Y_n\}$, and $\mathcal{M}_n = \{M \in \mathcal{M} : M \subset X_n \times Y_n\}$. Then $\mathcal{A}_n \subset \mathcal{M}_n$. Clearly, \mathcal{A}_n is an algebra, and \mathcal{M}_n is a monotone class by Steps I and II. By monotone class theorem, $\sigma(\mathcal{A}_n) \subset \mathcal{M}_n$. Since $\sigma(\mathcal{A}_n)$ contains all measurable subsets of $X_n \times Y_n$, so does \mathcal{M} . As a result, $\mathcal{M} \supset \mathcal{M}_n$ contains all measurable subsets of $X_n \times Y_n$ for all $n \in \mathbb{N}$. Since \mathcal{M} is closed under countable unions, $\mathcal{M} \supset \mathcal{F}_1 \otimes \mathcal{F}_2$. \square

The Tonelli's theorem gives the integral of nonnegative functions on product measure spaces.

Theorem 1.52 (Tonelli's theorem). Let $f : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2) \rightarrow \overline{\mathbb{R}}_+$ be a measurable function.

- (i) For all $\omega_1 \in \Omega_1$, the function $\omega_2 \mapsto f(\omega_1, \omega_2)$ is measurable.
- (ii) The function $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2)$ is measurable.
- (iii) For all $\omega_2 \in \Omega_2$, the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is measurable.
- (iv) The function $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$ is measurable.
- (v) The following integrals are equivalent:

$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).$$

Proof. By Theorem 1.51, the theorem holds for all indicator functions χ_A , where $A \in \mathcal{F}_1 \times \mathcal{F}_2$. Consequently, it holds for all nonnegative simple functions φ . For a general nonnegative measurable function f , choose a monotone increasing sequence φ_n of nonnegative simple functions such that $f = \lim_{n \rightarrow \infty} \varphi_n$. Applying monotone convergence theorem, we know that the theorem holds for f . \square

Proof of Fubini's theorem. Since $f = f^+ - f^-$, using Tonelli's theorem to f^+ and f^- completes the proof. \square

Remark. If $f \notin L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$, we cannot change the order of integration. For example, consider the function on $f : [0, 1] \times [0, 1] \rightarrow \overline{\mathbb{R}}$:

$$f(x, y) = \frac{x^2 - y^2}{(x^2 + y^2)^2} = -\frac{\partial^2}{\partial x \partial y} \arctan\left(\frac{y}{x}\right).$$

Then

$$\begin{aligned} \int_{[0,1]} \left(\int_{[0,1]} \frac{x^2 - y^2}{(x^2 + y^2)^2} dy \right) dx &= \int_{[0,1]} \frac{1}{1 + x^2} dx = \frac{\pi}{4}; \\ \int_{[0,1]} \left(\int_{[0,1]} \frac{x^2 - y^2}{(x^2 + y^2)^2} dx \right) dy &= \int_{[0,1]} \frac{-1}{1 + y^2} dy = -\frac{\pi}{4}; \end{aligned}$$

and

$$\int_{[0,1]} \int_{[0,1]} \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dx dy = \infty.$$

1.5 Signed Measures, Jordan Decomposition and Radon-Nikodym Theorem

Definition 1.53 (Signed measure). Let (Ω, \mathcal{F}) be a measurable space. A *signed measure* μ on (Ω, \mathcal{F}) is a set function $\mu : \mathcal{F} \rightarrow \mathbb{R}$ that satisfies the following:

- (i) $\mu(\emptyset) = 0$;
- (ii) (Countable additivity). If $(A_n)_{n=1}^\infty$ is a sequence of disjoint sets in \mathcal{F} , then

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n). \quad (1.11)$$

When the left-hand side of (1.11) is finite, the right-hand side converges absolutely.

Remark. One immediate consequence of (ii) is that a signed measure μ may take ∞ or $-\infty$ as a value, but it cannot take both, because the expression $\infty - \infty$ is undefined.

A signed measure μ is said to be *finite* if it only takes values in \mathbb{R} .

A signed measure μ is said to be *σ -finite* if there exists $\{\Omega_n\}_{n=1}^\infty$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$, and $-\infty < \mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$.

Theorem 1.54 (Hahn decomposition theorem). Let μ be a signed measure on a measurable space (Ω, \mathcal{F}) . Then there exist measurable sets $P, N \in \mathcal{F}$ that satisfy the following:

- (i) $P \amalg N$ is a division of Ω . (In other words, $P \cup N = \Omega$ and $P \cap N = \emptyset$.)
- (ii) For all $A \in \mathcal{F}$ with $A \subset P$, $\mu(A) \geq 0$. (In other words, P is a positive set.)
- (iii) For all $B \in \mathcal{F}$ with $B \subset N$, $\mu(B) \leq 0$. (In other words, N is a negative set.)

Proof. We may assume that μ does not take ∞ as a value. Otherwise apply the following proof on $-\mu$.

Denote by \mathcal{P} the collection of all positive sets in \mathcal{F} , then $\emptyset \in \mathcal{P}$. Let $M = \sup_{A \in \mathcal{P}} \mu(A)$, and choose $\{A_n\}_{n=1}^\infty \subset \mathcal{P}$ such that $\mu(A_n) \rightarrow M$. Clearly, $P = \bigcup_{n=1}^\infty A_n$ is a positive set, and $\mu(P) = M$.

We prove that $N := \Omega \setminus P$ is a negative set. If not, there exists a measurable set $E \subset N$ with $\mu(E) > 0$. Clearly, E is not a positive set. (Otherwise, $P \cup E \in \mathcal{P}$, but $\mu(P \cup E) = \mu(P) + \mu(E) > M = \sup_{A \in \mathcal{P}} \mu(A)$, a contradiction!) Hence there exists $B \subset E$ with $\mu(B) < 0$. We choose the smallest positive integer k_1 such that there exists $B_1 \subset B$ with $\mu(B_1) < -k_1^{-1}$. Since k_1 is the smallest, once $k_1 > 1$, any measurable subset A of E satisfies $\mu(A) \geq -(k_1 - 1)^{-1}$.

Again, $E \setminus B_1$ is not positive. Then we choose the smallest $k_2 \in \mathbb{N}$ such that there exists $B_2 \subset E \setminus B_1$ with $\mu(B_2) < -k_2^{-1}$. Repeat this procedure, we obtain a sequence $k_n \in \mathbb{N}$ and $B_n \subset \mathcal{F}$ such that

- $B_n \subset E \setminus \left(\bigcup_{k=1}^{n-1} B_k\right)$ and $\mu(B_n) < -k_n^{-1}$, and
- Once $k_n > 1$, any measurable subset A of $E \setminus \left(\bigcup_{k=1}^{n-1} B_k\right)$ satisfies $\mu(A) \geq -(k_n - 1)^{-1}$.

Take $C = E \setminus \left(\bigcup_{n=1}^\infty B_n\right)$. By assumption that μ does not take ∞ ,

$$\mu(C) = \mu(E) - \sum_{n=1}^\infty \mu(B_n) = \mu(E) + \sum_{n=1}^\infty \frac{1}{k_n} < \infty \Rightarrow k_n \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Since any measurable subset A of C satisfies $\mu(A) \geq -\lim_{n \rightarrow \infty} (k_n - 1)^{-1} = 0$, C is a positive set disjoint from P . However $\mu(P \cup C) = \mu(P) + \mu(E) + \sum_{n=1}^\infty k_n^{-1} > M = \sup_{A \in \mathcal{P}} \mu(A)$, again a contradiction! \square

Remark. We called a set $E \in \mathcal{F}$ a μ -null set if $\mu(A) = 0$ for any measurable subset A of E . Following this proof, the Hahn decomposition $P \amalg N$ is unique up to adding to/subtracting μ -null sets from P and N :

Given a Hahn decomposition $P' \amalg N'$, the set $P \cap N'$ is a positive set and also a negative set. The same applies to $N \cap P'$. Then $P \Delta P' = N \Delta N' = (P \cap N') \cup (N \cap P')$ is a μ -null set.

Corollary 1.55 (Jordan decomposition). Given a signed measure μ on a measurable space (Ω, \mathcal{F}) , take the Hahn decomposition $\Omega = P \amalg N$ on μ . Define

$$\mu^+(A) = \mu(A \cap P), \quad \mu^-(A) = -\mu(A \cap N), \quad \forall A \in \mathcal{F}.$$

Then μ^+ and μ^- are two (finite) measures on (Ω, \mathcal{F}) , and we have the *Jordan decomposition* $\mu = \mu^+ - \mu^-$. Since the Hahn decomposition is unique up to the difference of a μ -null set, the Jordan decomposition is unique. The measure $|\mu| := \mu^+ + \mu^-$ is called the *variation* of μ . Its maximum value $\|\mu\| = |\mu|(\Omega)$ is called the *total variation* of μ .

Now we discuss the relationship between signed measures and Lebesgue integration. We first introduce the absolute continuity and singularity of signed measures.

Definition 1.56. Let μ be a measure on a measurable space (Ω, \mathcal{F}) .

- (i) (Absolute continuity). A signed measure ν is said to be *absolutely continuous with respect to μ* , denoted by $\nu \ll \mu$, if $\nu(A) = 0$ for all $A \in \mathcal{F}$ such that $\mu(A) = 0$.
- (ii) (Singularity). A signed measure ν is said to be *singular with respect to μ* , denoted by $\nu \perp \mu$, if there exists $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(\Omega \setminus A) = 0$.

The following theorem tells that every measurable function f is associated with a signed measure.

Theorem 1.57. Let $f \in L^1(\Omega, \mathcal{F}, \mu)$. Define $\nu : \mathcal{F} \rightarrow \mathbb{R}$ by

$$\nu(A) = \int_A f \, d\mu, \quad \forall A \in \mathcal{F}.$$

Then ν is a (finite) signed measure on (Ω, \mathcal{F}) . Furthermore, for any $\epsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mu(A) < \delta$, we have $\nu(A) < \epsilon$. In particular, ν is absolutely continuous with respect to μ .

Proof. We may assume $f \geq 0$, and the result follows from $f = f^+ - f^-$.

Clearly $\nu(\emptyset) = 0$. Let $(A_n)_{n=1}^\infty$ be a sequence of disjoint sets in \mathcal{F} . Then $\sum_{k=1}^n f \chi_{A_k}$ is a monotone increasing sequence of nonnegative measurable functions that converges pointwise to $f \chi_A$, where $A = \bigcup_{n=1}^\infty A_n$. By monotone convergence theorem,

$$\nu(A) = \int f \chi_A \, d\mu = \lim_{n \rightarrow \infty} \sum_{k=1}^n \int f \chi_{A_k} \, d\mu = \sum_{n=1}^\infty \nu(A_n).$$

Thus ν is a signed measure on (Ω, \mathcal{F}) . Now fix $\epsilon > 0$. We define $E_n = \{\omega : f(\omega) > n\}$ for all $n \in \mathbb{N}$. Since f is integrable, $\mu(E_n) \rightarrow 0$. Again by monotone convergence theorem,

$$\int f \, d\mu = \lim_{n \rightarrow \infty} \int f \chi_{\Omega \setminus E_n} \, d\mu \Rightarrow \exists N > 0 \text{ such that } \int_{E_N} f \, d\mu < \frac{\epsilon}{2}.$$

Then for all $A \in \mathcal{F}$ with $\mu(A) < \epsilon/2N$, we have

$$\int_A f \, d\mu = \int_A f \chi_{\Omega \setminus E_N} \, d\mu + \int_A f \chi_{E_N} \, d\mu \leq N\mu(A) + \int_{E_N} f \, d\mu < \epsilon.$$

For the last statement, note that if $\nu(A) > 0$, then $\nu(A) \geq \epsilon$ for some $\epsilon > 0$, and there exists $\delta > 0$ such that $\mu(A) \geq \delta > 0$. Hence $\mu(A) = 0$ implies $\nu(A) = 0$. \square

In fact, the converse of Theorem 1.57 also holds true. It is the generalization of the fundamental theorem of calculus on measures, known as Radon-Nikodym theorem.

Theorem 1.58 (Radon-Nikodym theorem). Let μ and ν be two σ -finite measures defined on a measurable space (Ω, \mathcal{F}) . If $\nu \ll \mu$, then there exists a nonnegative measurable function $f : \Omega \rightarrow \mathbb{R}_+$ such that

$$\nu(A) = \int_A f \, d\mu, \quad \forall A \in \mathcal{F}.$$

The function $\frac{d\nu}{d\mu} := f$, called the *Radon-Nikodym derivative of ν with respect to μ* , is uniquely determined up to a μ -null set.

Proof. Step I: We first assume that both μ and ν are finite. Denote by \mathcal{F} the collection of all measurable functions $f : \Omega \rightarrow \overline{\mathbb{R}}$ such that (i) $f \geq 0$ a.e., and (ii) $\int_A f \, d\mu \leq \nu(A)$ for all $A \in \mathcal{F}$. Then $f \equiv 0$ is in \mathcal{F} , and \mathcal{F} is closed under finite maximum:

$$\int_A \max\{f_1, f_2\} \, d\mu = \int_{A \cap \{f_1 \geq f_2\}} f_1 \, d\mu + \int_{A \cap \{f_1 < f_2\}} f_2 \, d\mu \leq \nu(A \cap \{f_1 \geq f_2\}) + \nu(A \cap \{f_1 < f_2\}) = \nu(A).$$

Define $M = \sup_{f \in \mathcal{F}} \int f \, d\mu$, we prove that $M = \nu(\Omega)$. Let f_n be a sequence in \mathcal{F} such that $\int f_n \, d\mu \rightarrow M$. We choose an increasing sequence $g_n = \max\{f_1, \dots, f_n\} \in \mathcal{F}$. By monotone convergence theorem, the function $g = \lim_{n \rightarrow \infty} g_n$ lies in \mathcal{F} and satisfies

$$\int g \, d\mu = \lim_{n \rightarrow \infty} \int g_n \, d\mu \geq \lim_{n \rightarrow \infty} \int f_n \, d\mu = M \stackrel{g \in \mathcal{F}}{\Rightarrow} \int g \, d\mu = M.$$

Argue by contradiction. If $M < \nu(\Omega)$, choose $0 < \epsilon < \frac{\nu(\Omega) - M}{\mu(\Omega)}$ and define $\nu'(A) = \int_A g \, d\mu + \epsilon\mu(A)$. Then

$$\delta := \sup_{A \in \mathcal{F}} (\nu - \nu')(A) \geq \nu(\Omega) - \nu'(\Omega) = \nu(\Omega) - M - \epsilon\mu(\Omega) > 0.$$

Using Hahn decomposition theorem on signed measure $\nu - \nu'$, there exists a positive set $P \in \mathcal{F}$ such that $\nu(P) - \nu'(P) = \delta > 0$. Since $\nu'(A) \leq \nu(A) + \epsilon\mu(A)$ and $\nu \ll \mu$, we have $\nu' \ll \mu$, and $\mu(P) > 0$. By maximal property of P , we have $\nu'(A) \leq \nu(A)$ for all $A \subset P$. (Otherwise $\nu(P \setminus A) - \nu'(P \setminus A) > \delta$.) Then

$$\int_A (g + \epsilon\chi_P) \, d\mu = \int_A g \, d\mu + \epsilon\mu(A \cap P) = \int_{A \setminus P} g \, d\mu + \nu'(A \cap P) \leq \nu(A \setminus P) + \nu(A \cap P) = \nu(A), \quad \forall A \in \mathcal{F}.$$

Hence $g + \epsilon\chi_P \in \mathcal{F}$. However, $\int (g + \epsilon\chi_P) \, d\mu = \int g \, d\mu + \epsilon\mu(P) > M$, a contradiction! As a result, we have $\int g \, d\mu = M = \nu(\Omega)$. Since $g \in \mathcal{F}$, it holds

$$0 \leq \nu(A) - \int_A g \, d\mu = \int_{\Omega \setminus A} g \, d\mu - \nu(\Omega \setminus A) \leq 0, \quad \forall A \in \mathcal{F}.$$

Note that $g : \Omega \rightarrow \overline{\mathbb{R}}$ is integrable. The set $E = \{\omega : g(\omega) = \infty\}$ is μ -null. Choose $f = g\chi_{\Omega \setminus E}$, then f is the desired real-valued function.

Step II: If μ and ν are σ -finite, take a disjoint sequence $(\Omega_n)_{n=1}^\infty$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and $\mu(\Omega_n), \nu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$. For each n , by the finite case, there exists a measurable function $f_n : \Omega_n \rightarrow \mathbb{R}_+$ such that

$$\nu(A) = \int_A f_n \, d\mu, \quad \forall A \in \mathcal{F} \text{ with } A \subset \Omega_n.$$

Let $f = \sum_{n=1}^n f_n$. Apply monotone convergence theorem to $(\sum_{k=1}^n f_k)_{n=1}^\infty$:

$$\nu(A) = \sum_{n=1}^{\infty} \nu(\Omega_n \cap A) = \sum_{n=1}^{\infty} \int_A f_n d\mu = \int_A f d\mu, \quad \forall A \in \mathcal{F}.$$

Step III: Finally we show that f is uniquely determined up to a μ -null set. Let $h : \Omega \rightarrow \mathbb{R}_+$ be another function satisfying the desired property. Then

$$\int_A (f - h) d\mu = 0, \quad \forall A \in \mathcal{F}.$$

Take $A = \{\omega : f(\omega) > h(\omega)\}$, we have $\int_X (f - h)^+ d\mu = 0$, and $(f - h)^+ = 0$ a.e.. Similarly $(f - h)^- = 0$ a.e.. Hence $f = h$ a.e., as desired. \square

Corollary 1.59 (Radon-Nikodym). Let μ (resp. ν) be a σ -finite measure (resp. finite signed measure) on a measurable space (Ω, \mathcal{F}) . If $\nu \ll \mu$, then there exists $f \in L^1(\Omega, \mathcal{F}, \mu)$ such that

$$\nu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{F}.$$

The *Radon-Nikodym derivative* $\frac{d\nu}{d\mu} := f$ is uniquely determined up to a μ -null set.

Proof. Use the Jordan decomposition of signed measure $\nu = \nu^+ - \nu^-$. Then there exist measurable functions $g, h : \Omega \rightarrow \mathbb{R}_+$ such that

$$\nu^+(A) = \int_A g d\mu, \quad \nu^-(A) = \int_A h d\mu, \quad \forall A \in \mathcal{F}.$$

Since ν is finite, g and h are integrable. Then $f = g - h$ is the desired integrable function. \square

Theorem 1.60 (Lebesgue decomposition theorem). Let μ and ν be two σ -finite measures on a measurable space (Ω, \mathcal{F}) . Then there exist unique measures $\nu_0 \ll \mu$ and $\nu_1 \perp \mu$ such that $\nu = \nu_0 + \nu_1$.

Proof. Let $\lambda = \mu + \nu$, then $\mu, \nu \ll \lambda$, and λ is σ -finite. By Radon-Nikodym theorem, there exists measurable functions $f, g : \Omega \rightarrow \mathbb{R}_+$ such that

$$\mu(A) = \int_A f d\lambda, \quad \nu(A) = \int_A g d\lambda, \quad \forall A \in \mathcal{F}.$$

Let $E = \{\omega : f(\omega) = 0\}$, and define $\nu_1(A) = \nu(A \cap E)$, $\nu_0 = \nu(A \cap E^c)$ for all $A \in \mathcal{F}$. Clearly, $\nu_1 \perp \mu$, since $\nu_1(X \setminus E) = \mu(E) = 0$. It remains to show $\nu_0 \ll \mu$. If $\mu(A) = 0$, let $B := A \cap E^c = \{\omega \in A : f(\omega) \neq 0\}$. Note that $\mu(A) = \int_A f d\lambda = \int_B f d\lambda$, we have $\lambda(B) = 0$. Hence $0 \leq \nu_0(A) = \nu(B) \leq \lambda(B) = 0$.

Finally we prove uniqueness. If $\nu = \nu'_0 + \nu'_1$ with $\nu'_0 \ll \mu$ and $\nu'_1 \perp \mu$, there exists $E' \in \mathcal{F}$ such that $\nu'_1(X \setminus E') = \mu(E') = 0$. Then for all measurable $A \subset X \setminus (E \cup E')$, we have $\nu_0(A) = \mu(A) = \nu'_0(A)$. Moreover, for all measurable $A \subset E \cup E'$, since $\nu_0, \nu'_0 \ll \mu$, we have $\nu_0(A) = \nu'_0(A) = \mu(A) = 0$. Therefore $\nu'_0 = \nu_0$. \square

We can apply Theorem 1.60 to the Jordan decomposition of a signed measure $\nu = \nu^+ - \nu^-$.

Corollary 1.61 (Lebesgue). Let μ (resp. ν) be a σ -finite measure (resp. σ -finite signed measure) on (Ω, \mathcal{F}) . Then there exist unique signed measures $\nu_0 \ll \mu$ and $\nu_1 \perp \mu$ such that $\nu = \nu_0 + \nu_1$.

Remark. When ν is not absolutely continuous with respect to μ , we can apply Radon-Nikodym theorem to the pair $\nu_0 \ll \mu$.

1.6 Convergence of Measurable Functions and Measures

Definition 1.62 (Cauchy sequence in measure). Let $(f_n)_{n=1}^\infty$ be a sequence of measurable functions on $(\Omega, \mathcal{F}, \mu)$. If there exists a function f such that for all $\epsilon > 0$ and all $\eta > 0$, there exists N such that $\mu(|f_n - f_m| \geq \eta) < \epsilon$ for all $n, m \geq N$, then f_n is said to be a *Cauchy sequence in measure*.

Lemma 1.63. If $(f_n)_{n=1}^\infty$ is a Cauchy sequence in measure, there exists a subsequence $(f_{n_k})_{k=1}^\infty$ that converges a.e. to a measurable function f .

Proof. Since f_n is a Cauchy sequence, we can choose a subsequence f_{n_k} such that

$$\mu(E_k) < \frac{1}{2^k}, \text{ where } E_k = \left\{ |f_{n_{k+1}} - f_{n_k}| \geq \frac{1}{2^k} \right\}.$$

Let $F_N = \bigcup_{k=N}^\infty E_k$, and $E = \bigcap_{N=1}^\infty F_N$. Then $\mu(F_N) < 2^{-N+1}$, and $\mu(E) = \lim_{N \rightarrow \infty} \mu(F_N) = 0$. For each $\omega \in \Omega \setminus E$, we have $\omega \notin F_N$ for some $N \in \mathbb{N}$, which implies $|f_{n_{k+1}}(\omega) - f_{n_k}(\omega)| < 2^{-k}$ for all $k \geq N$. Hence $f_{n_k}(\omega)$ is a Cauchy sequence, which converges to some $f(\omega) \in \mathbb{R}$. For $\omega \in E$, define $f(\omega) = 0$. As a result, $f_n \rightarrow f$ a.e., which is measurable. \square

Definition 1.64 (Convergence in measure). Let $(f_n)_{n=1}^\infty$ be a sequence of measurable functions on $(\Omega, \mathcal{F}, \mu)$. If there exists a function f such that for all $\eta > 0$,

$$\lim_{n \rightarrow \infty} \mu(|f_n - f| \geq \eta) = 0,$$

then f is said to *converges to f in measure*, and we write $f_n \xrightarrow{\mu} f$.

Theorem 1.65. A function sequence $(f_n)_{n=1}^\infty$ converges in measure if and only if it is a Cauchy sequence.

Proof. Given $\epsilon > 0$ and $\eta > 0$. If $f_n \xrightarrow{\mu} f$, there exists N such that $\mu(|f_n - f| \geq \eta/2) < \epsilon/2$ for all $n \geq N$. Then for all $m, n \geq N$, we have

$$\mu(|f_n - f_m| \geq \eta) \leq \mu\left(\left\{\omega : |f_n(\omega) - f(\omega)| \geq \frac{\eta}{2}\right\} \cup \left\{\omega : |f_m(\omega) - f(\omega)| \geq \frac{\eta}{2}\right\}\right) < \epsilon.$$

Conversely, if f_n is a Cauchy sequence in measure, by Lemma 1.63, one of its subsequence f_{n_k} converges a.e. to a measurable function f . Furthermore, if we choose F_k in Lemma 1.63, for all $k \geq N$, we have

$$|f_{n_k}(\omega) - f(\omega)| \leq \sum_{l=k}^\infty |f_l(\omega) - f_{l+1}(\omega)| \leq \frac{1}{2^{k-1}}, \forall \omega \in \Omega \setminus F_k \Rightarrow \mu\left(|f_{n_k} - f| > \frac{1}{2^{k-1}}\right) \leq \mu(F_k) < \frac{1}{2^{k-1}}.$$

Hence $f_{n_k} \xrightarrow{\mu} f$. Now given $\epsilon, \eta > 0$, we choose $k > 0$ such that $\mu(|f_{n_k} - f| > \eta/2) < \epsilon/2$, and choose $N > 0$ such that $\mu(|f_n - f_m| > \eta/2) < \epsilon/2$ for all $n \geq N$. Then

$$\mu(|f_n - f| \geq \eta) \leq \mu\left(\left\{|f_n - f_{n_k}| \geq \frac{\eta}{2}\right\} \cup \left\{|f_{n_k} - f| \geq \frac{\eta}{2}\right\}\right) < \epsilon, \forall n \geq \max\{n_k, N\}.$$

Therefore $f_n \xrightarrow{\mu} f$. \square

Theorem 1.66 (Egoroff). Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. Then $f_n \rightarrow f$ μ -a.e. $\Rightarrow f_n \xrightarrow{\mu} f$.

Proof. Let $\epsilon > 0$ and $\eta > 0$. By Theorem 1.37, choose $E \in \mathcal{F}$ such that $\mu(E) < \epsilon$ and f_n converges f uniformly on $\Omega \setminus E$. Then there exists N such that $|f_n - f| < \eta$ for all $n \geq N$ and all $\omega \in \Omega \setminus E$. Hence $\mu(|f_n - f| \geq \eta) < \mu(E) < \epsilon$ for all $n \geq N$. Since $\epsilon > 0$ and $\eta > 0$ are arbitrary, $f_n \xrightarrow{\mu} f$. \square

Now we introduce L^p -spaces.

Definition 1.67 (L^p -spaces). Let $(\Omega, \mathcal{F}, \mu)$ be a measurable space. For $1 \leq p < \infty$, define $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ to be the set of all measurable functions f such that $|f|^p$ is integrable, i.e. $\int_X |f|^p d\mu < \infty$. We define

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{1/p}, \quad f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu).$$

By Minkowski's inequality, $\|\cdot\|_p$ is a seminorm on $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$. Let $f \sim g \stackrel{\text{def}}{\iff} f = g \text{ a.e.}$ be an equivalence relation on $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$. We define L^p -space as the quotient space

$$L^p(\Omega, \mathcal{F}, \mu) = \mathcal{L}^p(\Omega, \mathcal{F}, \mu) / \sim,$$

and maintain the norm $\|[f]\|_p = \|f\|_p$. This is a well-defined norm, since $\|f\|_p = \|g\|_p$ if $f \sim g$. For simplicity, we drop the brackets and use f to denote its corresponding equivalence class $[f]$ in $L^p(\Omega, \mathcal{F}, \mu)$. Then the space $(L^p(\Omega, \mathcal{F}, \mu), \|\cdot\|_p)$ is a normed space.

Theorem 1.68 (Chebyshev inequality). Let $1 \leq p < \infty$, and $f \in L^p(\Omega, \mathcal{F}, \mu)$. Then

$$\mu(|f| \geq \eta) \leq \frac{1}{\eta^p} \|f\|_p^p, \quad \forall \eta > 0.$$

Proof. Let $E = \{\omega : |f(\omega)| \geq \eta\}$. Then

$$\|f\|_p^p = \int |f|^p d\mu \geq \int_E |f|^p d\mu \geq \eta^p \mu(E), \quad \forall \eta > 0. \quad \square$$

Remark. As a result of Theorem 1.68, convergence in L^p -norm implies convergence in measure.

The following theorem tells us that L^p -spaces are complete.

Theorem 1.69 (Riesz-Fisher). $L^p(\Omega, \mathcal{F}, \mu)$ is a Banach space. In other words, if $f_n \in L^p(\Omega, \mathcal{F}, \mu)$ is a Cauchy sequence in L^p -norm, i.e. for every $\epsilon > 0$ there exists $N \geq 0$ such that $\|f_n - f_m\|_p < \epsilon$ for all $n, m \geq N$, then f_n converges in $L^p(\Omega, \mathcal{F}, \mu)$.

Proof. By Chebyshev's inequality, f_n is also a Cauchy sequence in measure, and there exists a subsequence f_{n_k} that converges *a.e.* to some measurable f . Given $\epsilon > 0$, we choose N such that $\|f_n - f_m\|_p < \epsilon$ for all $n, m \geq N$. By Fatou's lemma,

$$\int |f - f_m|^p d\mu = \int \lim_{k \rightarrow \infty} |f_{n_k} - f_m|^p d\mu \leq \liminf_{k \rightarrow \infty} \int |f_{n_k} - f_m|^p d\mu \leq \epsilon^p, \quad \forall m \geq N.$$

Hence $f - f_m \in L^p(\Omega, \mathcal{F}, \mu)$, $f = f_m + (f - f_m) \in L^p(\Omega, \mathcal{F}, \mu)$. Since $\epsilon > 0$ is arbitrary, $\|f - f_m\|_p \rightarrow 0$. \square

Now we introduce uniform integrability of function classes.

Definition 1.70 (Uniform Integrability). Let $L_+^1(\Omega, \mathcal{F}, \mu) = \{g \in L^1(\Omega, \mathcal{F}, \mu) : g \geq 0\}$. A collection of integrable functions $\mathcal{F} \subset L^1(\Omega, \mathcal{F}, \mu)$ is said to be *uniformly integrable*

$$\inf_{g \in L_+^1(\Omega, \mathcal{F}, \mu)} \sup_{f \in \mathcal{F}} \int_{\{|f| > g\}} |f| d\mu = 0. \quad (1.12)$$

Clearly, a collection of finitely many integrable functions is uniformly integrable. So we are interested in the case of infinitely many integrable functions. We have the following basic fact.

Lemma 1.71. If $\mathcal{F} \subset L^1(\Omega, \mathcal{F}, \mu)$ is uniformly integrable, then

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \int_{\{|f| > n\}} |f| d\mu = 0. \quad (1.13)$$

Proof. If (2.10) holds, we fix $\epsilon > 0$ and choose $g \in L^1_+(\Omega, \mathcal{F}, \mu)$ and $n > 0$ such that

$$\sup_{f \in \mathcal{F}} \int_{\{|f| > g\}} |f| d\mu < \frac{\epsilon}{2}, \quad \int_{\{g > n\}} g d\mu < \frac{\epsilon}{2}.$$

Then for all $f \in \mathcal{F}$, we have

$$\int_{\{|f| > n\}} |f| d\mu \leq \int_{\{|f| > g\}} |f| d\mu + \int_{\{|g| \geq f > n\}} |f| d\mu \leq \int_{\{|f| > g\}} |f| d\mu + \int_{\{|g| > n\}} g d\mu < \epsilon.$$

Hence (1.11) implies (1.13). \square

Remark. In addition, if $(\Omega, \mathcal{F}, \mu)$ is a finite measure space, (1.13) can be used as an equivalent definition of uniform integrability: Since constant functions are always integrable in finite measure spaces, (1.13) \Rightarrow (1.12).

The following theorem gives a depiction of uniformly integrable function classes in finite measure spaces, which has a similar form to the Arzelà-Ascoli theorem in functional analysis.

Theorem 1.72. Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. A collection of functions $\mathcal{F} \subset L^1(\Omega, \mathcal{F}, \mu)$ is uniformly integrable if and only if it satisfies the following:

- (i) (Uniform L^1 -boundedness). $\sup_{f \in \mathcal{F}} \|f\|_1 < \infty$.
- (ii) (Uniform absolute continuity). For all $\epsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mu(A) < \delta$,

$$\int_A |f| d\mu < \epsilon, \quad \forall f \in \mathcal{F}.$$

Proof. Let \mathcal{F} be uniform integrable. For all $A \in \mathcal{F}$ and $n > 0$, we have

$$\int_A |f| d\mu \leq \int_{A \cap \{|f| > n\}} |f| d\mu + \int_{A \cap \{|f| \leq n\}} |f| d\mu \leq \int_{\{|f| > n\}} |f| d\mu + n\mu(A), \quad \forall f \in \mathcal{F}.$$

- (i) Choose $A = \Omega$. Since $\mu(\Omega) < \infty$, and \mathcal{F} be uniform integrable, both terms are uniformly bounded.
- (ii) Given $\epsilon > 0$, we choose N such that $\sup_{f \in \mathcal{F}} \int_{\{|f| > N\}} |f| d\mu < \epsilon/2$ and $\delta = \frac{\epsilon}{2N}$.

Conversely, if \mathcal{F} satisfies (i) and (ii), by Chebyshev inequality,

$$\sup_{f \in \mathcal{F}} \mu(|f| \geq n) \leq \frac{1}{Nn} \sup_{f \in \mathcal{F}} \|f\|_1 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Given $\epsilon > 0$, we choose the δ mentioned in (ii), and choose N such that $\mu(|f| \geq n) < \delta$ for all $n \geq N$ and all $f \in \mathcal{F}$. By uniform absolute continuity of \mathcal{F} , we have $\sup_{f \in \mathcal{F}} \int_{\{|f| > n\}} |f| d\mu < \epsilon$ for all $n \geq N$. Since ϵ is arbitrary, \mathcal{F} is uniformly integrable. \square

With uniform integrability, we can deduce L^1 -convergence using convergence in measure.

Theorem 1.73. Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space, and $f_n \in L^1(\Omega, \mathcal{F}, \mu)$. Then $\|f_n - f\|_1 \rightarrow 0$ if and only if $f_n \xrightarrow{\mu} f$ and $\{f_n\}_{n \in \mathbb{N}}$ is uniformly integrable.

Proof. Assume $\|f_n - f\|_1 \rightarrow 0$, and fix $\epsilon > 0$. By Chebyshev inequality, $f_n \xrightarrow{\mu} f$. Furthermore, by choosing N such that $\|f_n - f\|_1 < \frac{\epsilon}{2\mu(\Omega)}$ for all $n \geq N$, we have

$$\sup_{n \in \mathbb{N}} \|f_n\| \leq \max \left\{ \|f_1\|_1, \dots, \|f_{N-1}\|_1, \|f\| + \frac{\epsilon}{2\mu(\Omega)} \right\} < \infty.$$

Hence $\{f_n\}_{n \in \mathbb{N}}$ is uniformly L^1 -bounded. Furthermore, for all $A \in \mathcal{F}$, we have

$$\sup_{n \in \mathbb{N}} \int_A |f_n| d\mu \leq \max \left\{ \int_A |f_1| d\mu, \dots, \int_A |f_{N-1}| d\mu, \int_A |f_N| d\mu + \frac{\epsilon}{2} \right\}.$$

By Theorem 1.57, for each $j \in \{1, \dots, N\}$, we choose $\delta_j > 0$ such that for all $A \in \mathcal{F}$ with $\mu(A) < \delta_j$, $\int_A |f_j| d\mu < \frac{\epsilon}{2}$. Choose $\delta := \min_{j \in \{1, \dots, N\}} \delta_j$, we have $\sup_{n \in \mathbb{N}} \int_A |f_n| d\mu < \epsilon$ for all $\mu(A) < \delta$. Hence $\{f_n\}_{n \in \mathbb{N}}$ is uniformly absolutely continuous. By Theorem 1.72, it is uniformly integrable.

Conversely, assume $f_n \xrightarrow{\mu} f$ and $\{f_n\}_{n \in \mathbb{N}}$ is uniformly integrable. Given $\epsilon > 0$, we have

$$\begin{aligned} \|f_n - f\|_1 &\leq \epsilon\mu(\Omega) + \int_{\{|f_n - f| > \epsilon\}} |f_n - f| d\mu \\ &\leq \epsilon\mu(\Omega) + \int_{\{|f_n - f| > \epsilon\}} |f_n| d\mu + \int_{\{|f_n - f| > \epsilon\}} |f| d\mu \end{aligned}$$

Since f is integrable, $\{f_n\}_{n=1}^\infty$ is uniformly integrable, we choose $\delta > 0$ such that $\sup_{n \in \mathbb{N}} \int_A |f_n| d\mu < \epsilon$ and $\int_A |f| d\mu < \epsilon$ for all $\mu(A) < \delta$. Since $f_n \xrightarrow{\mu} f$, we also choose N such that $\mu(\{|f_n - f| > \epsilon\}) < \delta$ for all $n \geq N$. Then $\|f_n - f\|_1 \leq \epsilon(\mu(\Omega) + 2)$ for all $n \geq N$. Note that $\epsilon > 0$ is arbitrary, we have $\|f_n - f\|_1 \rightarrow 0$. \square

Remark. By Theorem 1.73, if $(\Omega, \mathcal{F}, \mu)$ is a finite measure space, $\{f_n\}_{n \in \mathbb{N}} \subset L^1(\Omega, \mathcal{F}, \mu)$ is a uniformly integrable sequence, and $f_n \xrightarrow{\mu} f$, then

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

This is a weaker condition than Lebesgue dominated convergence theorem.

Definition 1.74. (Weak convergence). Let Ω be a metric space with its Borel σ -algebra \mathcal{B} . Let $C_b(\Omega)$ be the set of all bounded continuous functions on Ω . Let μ_n be a sequence of probability measures on (Ω, \mathcal{B}) . If there exists a probability measure μ on (Ω, \mathcal{B}) such that

$$\int f d\mu_n \rightarrow \int f d\mu, \quad \forall f \in C_b(\Omega),$$

then μ_n is said to *converge weakly* to μ , and we write $\mu_n \xrightarrow{w} \mu$.

Review: Semi-continuity. Recall that a function $f : \Omega \rightarrow \overline{\mathbb{R}}$ is *upper semi-continuous* at ω_0 if for any real $y > f(\omega_0)$ there exists a neighborhood U of ω_0 such that $f(x) < y$ for all $x \in U$. In a nutshell, f does not take a much larger value than $f(\omega_0)$ at a point closed to ω_0 .

Similarly, a function f is said to be *lower semi-continuous* at ω_0 if for any real $y < f(\omega_0)$ there exists a neighborhood U of ω_0 such that $f(x) > y$ for all $x \in U$. In addition, If f is upper (resp. lower) semi-continuous at each $\omega \in \Omega$, we say f is *upper (resp. lower) semi-continuous*.

Lemma 1.75. Let Ω be a metric space. For every nonnegative lower semi-continuous function $f : \Omega \rightarrow \overline{\mathbb{R}}_+$, there exists a sequence of nonnegative bounded Lipschitz continuous functions f_n such that $f_n \nearrow f$ pointwise.

Proof. For every $n \in \mathbb{N}$, define $g_n(x) = \inf_{y \in \Omega} \{f(y) + nd(x, y)\}$. Clearly, we have $0 \leq f_n \leq f_{n+1} \leq f$. Furthermore, for all $x, y \in \Omega$,

$$g_n(x) - g_n(y) = \inf_{z \in \Omega} \{f(z) + nd(x, z)\} - \inf_{z \in \Omega} \{f(z) + nd(y, z)\} \leq \inf_{z \in \Omega} \{f(z) + nd(x, y) + nd(y, z)\} - \inf_{z \in \Omega} \{f(z) + nd(y, z)\} = nd(x, y).$$

Symmetrically $g_n(y) - g_n(x) \leq nd(x, y)$. Hence g_n is n -Lipschitz. It remains to show $g_n \nearrow f$ pointwise.

Fix $x \in \Omega$, and choose $0 < \epsilon < f(x)$. Since f is lower semi-continuous, there exists $\delta > 0$ such that $f(y) > f(x) - \epsilon$ for all $y \in O(x, \delta)$. Choose $N > f(x)/\delta$. If $n \geq N$, we have $f(y) + nd(x, y) \geq N\delta > f(x)$ for all $y \notin O(x, \delta)$, and $f(y) + nd(x, y) \geq f(y) > f(x) - \epsilon$ for all $y \in O(x, \delta)$. Hence $f(x) - \epsilon < g_n(x) \leq f(x)$ for all $n \geq N$. Since $0 < \epsilon < f(x)$ is arbitrary, $g_n(x) \nearrow f(x)$. Then $f_n = \min\{g_n, n\}$ is the desired sequence. \square

The following lemma states that the converging point of a weakly convergent sequence is unique.

Lemma 1.76. Let Ω be a metric space with its Borel σ -algebra \mathcal{B} . Let μ_n be a sequence of probability measures on (Ω, \mathcal{B}) . If $\mu_n \xrightarrow{w} \mu$ and $\mu_n \xrightarrow{w} \mu'$, then $\mu = \mu'$.

Proof. By definition of weak convergence, $\int f_n d\mu = \int f_n d\mu'$ for all $f \in C_b(\Omega)$.

Let G be a closed set, then χ_G is lower semi-continuous. By Lemma 1.75, we choose a sequence of bounded Lipschitz continuous functions $f_n \nearrow \chi_G$. By monotone convergence theorem,

$$\mu(G) = \lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu' = \mu'(G).$$

Let \mathcal{T} be the topology on Ω , i.e. \mathcal{T} is the collection of all open subsets of Ω . Then $\mu|_{\mathcal{T}} = \mu'|_{\mathcal{T}}$. Since \mathcal{T} is a π -system, and $\sigma(\mathcal{T}) = \mathcal{B}$, by Lemma 1.22, $\mu = \mu'$. \square

Portmanteau lemma gives multiple equivalent definitions of weak convergence.

Theorem 1.77 (Portmanteau lemma). Let Ω be a metric space with its Borel σ -algebra \mathcal{B} . Let μ_n be a sequence of probability measures on (Ω, \mathcal{B}) . The following are equivalent:

- (i) $\int f d\mu_n \rightarrow \int f d\mu$ for all bounded continuous functions f . In other words, $\mu_n \xrightarrow{w} \mu$;
- (ii) $\int f d\mu_n \rightarrow \int f d\mu$ for all bounded Lipschitz continuous functions f ;
- (iii) $\liminf_{n \rightarrow \infty} \int f d\mu_n \geq \int f d\mu$ for all lower semi-continuous function f bounded from below;
- (iv) $\limsup_{n \rightarrow \infty} \int f d\mu_n \leq \int f d\mu$ for all upper semi-continuous function f bounded from above;
- (v) $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ for every open sets G ;
- (vi) $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ for every closed sets F ;
- (vii) $\lim_{n \rightarrow \infty} \mu_n(B) \rightarrow \mu(B)$ for all Borel sets B with $\mu(\partial B) = 0$, where $\partial B = \overline{B} \setminus \overset{\circ}{B}$ is the boundary of B .

Remark. A Borel set B is said to be a μ -continuity set if $\mu(\partial B) = 0$. Conversely, if a Borel set B is not a μ -continuity set, it is said to be a μ -discontinuity set.

Proof. (i) \Rightarrow (ii) is clear. (iii) \Leftrightarrow (iv) follows by taking negation. (v) \Leftrightarrow (vi) follows by taking complements.

(ii) \Rightarrow (iii): Without loss of generality, assume $f \geq 0$ is lower semi-continuous. By Lemma 1.75, choose a sequence f_k of nonnegative bounded Lipschitz continuous functions such that $f_k \nearrow f$ pointwise.

Since f_k is Lipschitz and $f_k \leq f$, by (ii) and monotone convergence theorem, we have

$$\liminf_{n \rightarrow \infty} \int f \, d\mu_n \geq \liminf_{n \rightarrow \infty} \int f_k \, d\mu_n = \int f_k \, d\mu, \quad \forall k \in \mathbb{N} \Rightarrow \liminf_{n \rightarrow \infty} \int f \, d\mu_n \geq \lim_{k \rightarrow \infty} \int f_k \, d\mu = \int f \, d\mu.$$

(iii) + (iv) \Rightarrow (i): Let f be a bounded continuous function. By (iii) and (iv). Then

$$\liminf_{n \rightarrow \infty} \int f \, d\mu_n \geq \int f \, d\mu \geq \limsup_{n \rightarrow \infty} \int f \, d\mu_n \Rightarrow \lim_{n \rightarrow \infty} \int f \, d\mu_n = \int f \, d\mu.$$

(iii) \Rightarrow (v): If G is an open set, χ_G is bounded and lower semi-continuous. Take $f = \chi_G$ in (iii).

(v) \Rightarrow (i): Pick $f \in C_n(\Omega)$, and without loss of generality assume $0 < f < 1$. Then the function $\chi_{\{(\omega, t): f(\omega) > t\}} = \chi_{(0, \infty)}(f(\omega) - t)$ is measurable. By Fubini's theorem,

$$\int_0^1 \mu(f > t) \, dt = \int_0^1 \left(\int_{\Omega} \chi_{\{(\omega, t): f(\omega) > t\}} \, d\mu \right) dt = \int_{\Omega} \left(\int_0^1 \chi_{\{(\omega, t): t < f(\omega)\}} \, dt \right) d\mu = \int f \, d\mu.$$

Since f is continuous, $\{\omega : f(\omega) > t\} = f^{-1}((t, \infty))$ is open. By (v), $\liminf_{n \rightarrow \infty} \mu_n(f > t) \geq \mu(f > t)$. Using Fatou's lemma, we have

$$\liminf_{n \rightarrow \infty} \int f \, d\mu_n = \liminf_{n \rightarrow \infty} \int_0^1 \mu_n(f > t) \, dt \geq \int_0^1 \liminf_{n \rightarrow \infty} \mu_n(f > t) \, dt \geq \int \mu(f > t) \, dt = \int f \, d\mu.$$

By repeating the same procedure on $-f$, we have $\limsup_{n \rightarrow \infty} \int f \, d\mu_n \leq \int f \, d\mu$. Then (i) follows.

(v) + (vi) \Rightarrow (vii): Let $B \in \mathcal{B}$. Then

$$\mu(\overset{\circ}{B}) \stackrel{(v)}{\leq} \liminf_{n \rightarrow \infty} \mu_n(\overset{\circ}{B}) \leq \liminf_{n \rightarrow \infty} \mu_n(B) \leq \limsup_{n \rightarrow \infty} \mu_n(B) \leq \limsup_{n \rightarrow \infty} \mu_n(\overline{B}) \stackrel{(vi)}{\leq} \mu(\overline{B}).$$

If $\mu(\delta B) = 0$, all above inequalities become equalities, and they equal $\mu(B)$.

(vii) \Rightarrow (vi): Fix a closed set $F \subset \Omega$, and define the collection of sets $\{B_F(r) : r \geq 0\}$, where

$$B_F(r) := \{\omega \in \Omega : d(\omega, F) \leq r\}.$$

Claim. There exists a countable subset C of $[0, \infty)$ such that $B_F(r)$ is a μ -continuity set for all $r \in [0, \infty) \setminus C$.

Proof of claim. Given $r \geq 0$, let $D_F(r) = \{\omega \in \Omega : d(\omega, F) = r\}$. Then $\{D_F(r) : r \geq 0\}$ is a partition of Ω .

By continuity of $d(\cdot, F)$, $B_F(r)$ is a closed set. Furthermore, if $\omega \in \partial B_F(r) = \overline{\Omega \setminus B_F(r)} \cap B_F(r)$, choose a sequence $\omega_n \in \Omega \setminus B_F(r)$ such that $\omega_n \rightarrow \omega$. Again by continuity of $d(\cdot, F)$, $d(\omega, F) = \lim_{n \rightarrow \infty} d(\omega_n, F) \geq r$. Hence $d(\omega, F) = r$, which implies $\partial B_F(r) \subset D_F(r)$.

Let $C = \{r \geq 0 : \mu(D_F(r)) > 0\}$ and $C_n = \{r \geq 0 : \mu(D_F(r)) > 1/n\}$. Then $|C_n| < n$, and $C = \bigcup_{n=1}^{\infty} C_n$ is at most countable. Furthermore, for each $r \in [0, \infty) \setminus C$, it holds

$$0 \leq \mu(\partial B_F(r)) \leq \mu(D_F(r)) = 0.$$

Therefore $B_F(r)$ is a μ -continuity set, and C is the desired countable set. Q.E.D.

Now we choose a sequence $r_k \searrow 0$ in $[0, \infty) \setminus C$. Then $B_F(r_k) \searrow F$. By (vii),

$$\mu(B_F(r_k)) = \lim_{n \rightarrow \infty} \mu_n(B_F(r_k)) \geq \limsup_{n \rightarrow \infty} \mu_n(F), \quad \forall k \in \mathbb{N} \Rightarrow \mu(F) = \lim_{k \rightarrow \infty} \mu(B_F(r_k)) \geq \limsup_{n \rightarrow \infty} \mu_n(F).$$

Hence (vi) follows. □

2 Random Variables

In this section, if not specifically indicated, our discussion is based on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1 Random Variables and Independence

Definition 2.1 (Random variables and distribution). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A (*real-valued*) *random variable* is a measurable real-valued function X on (Ω, \mathcal{F}) . In other words, a real-valued function $X : \Omega \rightarrow \mathbb{R}$ is a random variable if

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}, \forall x \in \mathbb{R} \Leftrightarrow X^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{F}.$$

The collection $X^{-1}(\mathcal{B}(\mathbb{R}))$ is said to be the σ -algebra generated by X . The function $F : \mathbb{R} \rightarrow [0, 1]$,

$$F(x) = \mathbb{P}(X \leq x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}), \forall x \in \mathbb{R} \quad (2.1)$$

is said to be the *cumulative distribution function* (*c.d.f.*) of X , written $X \sim F$.

Remark. Generally, a measurable extended real-valued function $X : \Omega \rightarrow \overline{\mathbb{R}}$ is also called a random variable, if we have $\mathbb{P}(|X| = \infty) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \{-\infty, \infty\}\}) = 0$.

Proposition 2.2. Let F be the c.d.f. of a random variable X . Then F satisfies the following:

- (i) F is monotone increasing on \mathbb{R} ;
- (ii) F is *right-continuous*, i.e. $F(x) = \lim_{\epsilon \rightarrow 0^+} F(x + \epsilon)$ for all $x \in \mathbb{R}$;
- (iii) $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$, and $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$.

In fact, any function $F : \mathbb{R} \rightarrow [0, 1]$ satisfying the properties (i)-(iii) is called a *c.d.f.*.

Proof. Clearly F is monotone increasing, and its left and right-hand limits exist everywhere. Then

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} F(x + \epsilon) &= \lim_{n \rightarrow \infty} F\left(x + \frac{1}{n}\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\{\omega : X(\omega) \leq x + \frac{1}{n}\right\}\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \left\{\omega : X(\omega) \leq x + \frac{1}{n}\right\}\right) \\ &= \mathbb{P}(\{\omega : X(\omega) \leq x\}) = F(x). \end{aligned}$$

Then (ii) holds, and (iii) follows from a similar procedure. □

Remark. Inspired by this proof, we can also associated $\mathbb{P}(X < x)$ with F by the following formula:

$$\mathbb{P}(X < x) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \left\{\omega : X(\omega) \leq x - \frac{1}{n}\right\}\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(X \leq x - \frac{1}{n}\right) = \lim_{\epsilon \rightarrow 0^+} F(x - \epsilon).$$

Since $\mathbb{P}(X = x) = F(x) - \lim_{\epsilon \rightarrow 0^+} F(x - \epsilon)$, F is continuous at a point $x \in \mathbb{R}$ if and only if $\mathbb{P}(X = x) = 0$.

Definition 2.3 (Distribution measure). A random variable $X \sim F$ on $(\Omega, \mathcal{F}, \mathbb{P})$ determines a pushforward measure $\mu_F = \mathbb{P} \circ X^{-1}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\mu_F(B) = \mathbb{P}(X^{-1}B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\}), \forall B \in \mathcal{B}(\mathbb{R}).$$

The pushforward μ_F is said to be the *distribution measure* of X , written $X \sim \mu_F$. It is easy to check that

$$\mu_F((-\infty, b]) = F(b), \mu_F((a, \infty)) = 1 - F(a), \mu_F((a, b]) = F(b) - F(a), \forall a < b. \quad (2.2)$$

Remark. Let \mathcal{S} be the collection of all finite unions of intervals of the following forms:

$$(-\infty, b], (a, \infty), (a, b].$$

Then \mathcal{S} is a semiring of subsets of \mathbb{R} , and $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$. Given a c.d.f. $F : \mathbb{R} \rightarrow [0, 1]$, we define a pre-measure μ_F on \mathcal{S} by equation (2.2). Using Carathéodory's extension theorem, μ_F can be uniquely extended to a probability measure on $\mathcal{B}(\mathbb{R})$. Thus we find a one-to-one correspondence between a c.d.f. $F : \mathbb{R} \rightarrow [0, 1]$ and Borel probability measure μ_F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In later discussion, we may not distinguish them.

Let's see what the c.d.f. of a random variable may look like.

Definition 2.4. Let μ be a Borel measure on \mathbb{R} . A point $x \in \mathbb{R}$ is said to be an *atom* of μ if $\mu(\{x\}) > 0$.

(i) (Discrete measure). μ is said to be *discrete*, if there exists a countable subset C of \mathbb{R} such that

$$\mu(A) = \sum_{x \in C \cap A} \mu(\{x\}), \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

The atoms of μ is $D := \{x \in \mathbb{R} : \mu(\{x\}) > 0\} \subset C$.

- (ii) (Continuous measure). μ is said to be (*absolutely*) *continuous* if $\mu \ll m$, where m is the Lebesgue measure on \mathbb{R} . The Radon-Nikodym derivative $\rho := \frac{d\mu}{dm}$ is said to be the *density function* of μ . If μ is a probability measure, ρ is said to be the *probability density function* (p.d.f.) of μ .
- (iii) (Singular measure). μ is said to be *singular* (*continuous*) if μ has no atom and $\mu \perp m$, where m is the Lebesgue measure on \mathbb{R} . In other words, μ is concentrated on a Lebesgue-null set E , where μ takes zero at each point of E .

Remark. The discrete and continuous measures are common. For example, the Poisson distribution is a discrete measure on \mathbb{N} , and the Gaussian distribution is a continuous measure on \mathbb{R} . Here we give an example of singular measures.

Ternary Cantor sets and devil's stair. A ternary Cantor set K is obtained by repeatedly removing the open middle third from a collection of line segments, starting from the unit interval $[0, 1]$:

$$K_1 = [0, 1] \rightarrow K_2 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] \rightarrow K_3 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \rightarrow \dots$$

Indeed, the ternary Cantor set $K = \bigcup_{n=1}^{\infty} K_n$ is the set of numbers in $[0, 1]$ with a ternary expansion omitting the digit 1:

$$K = \left\{ \sum_{n=1}^{\infty} \frac{c_n}{3^n} : c_n \in \{0, 2\} \right\},$$

which is an uncountable Lebesgue-null set in \mathbb{R} . We define a function f on K by the following formula:

$$f\left(\sum_{n=1}^{\infty} \frac{c_n}{3^n}\right) = \sum_{n=1}^{\infty} \frac{c_n}{2^{n+1}}. \quad (\text{Devil's stair})$$

It is clear that function $f : K \rightarrow [0, 1]$ is monotone increasing and onto. Furthermore, for each connected component (a, b) of $[0, 1] \setminus K$, one have $f(a) = f(b)$. Hence we can extend f to a continuous monotone increasing function on $F : [0, 1] \rightarrow [0, 1]$ that is constant on each connected component (a, b) of $[0, 1] \setminus K$. Clearly, F is a c.d.f., and the associated distribution measure μ_F is a singular probability measure.

Theorem 2.5 (Decomposition of Borel measures). If μ is a σ -finite Borel measure on \mathbb{R} , there exist uniquely discrete, continuous and singular σ -finite measures μ_d , μ_c and μ_s such that $\mu = \mu_d + \mu_c + \mu_s$.

Proof. Let $D = \{x \in \mathbb{R} : \mu(\{x\}) > 0\}$. Since μ is σ -finite, D is at most countable. Then $\mu_d(A) = \mu(A \cap D)$ is a discrete measure. ($\mu_d \equiv 0$ if $D = \emptyset$.) Furthermore, μ_d is unique and supported on all atoms of μ .

By Theorem 1.60, the measure $\mu - \mu_d$ has a unique Lebesgue decomposition $\mu - \mu_d = \mu_c + \mu_s$, where $\mu_c \ll m$ and $\mu_s \perp m$. Since $\mu - \mu_d$ has no atom, the result follows. \square

Remark. Likewise, a c.d.f. F admits a unique convex combination $F = \alpha F_d + \beta F_c + (1 - \alpha - \beta)F_s$, where the associated distribution measures of F_d , F_c and F_s are discrete, continuous and singular, respectively.

Definition 2.6 (Random Vectors). When X_1, X_2, \dots, X_n are all random variables, the function

$$X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$$

is said to be a *random vector*. The function

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

is called the *joint distribution (function)* of (X_1, \dots, X_n) . For each k , the function

$$F_k(x) = \mathbb{P}(X_k = x) = F(\infty, \dots, \infty, \underset{k\text{-th}}{x}, \infty, \dots, \infty)$$

is called the *marginal distribution (function)* of X_k .

Remark. By Theorem 1.35, a random vector is also a measurable function $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Furthermore, the c.d.f. $F : \mathbb{R}^n \rightarrow [0, 1]$ of X satisfies the following:

- (i) F is monotone increasing with respect to each variable on \mathbb{R} ;
- (ii) F is *right-continuous* on each variable;
- (iii) For all $x_1, \dots, x_n \in \mathbb{R}$, $\lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_n) = 0$, and $\lim_{x_1, \dots, x_n \rightarrow \infty} F(x_1, \dots, x_n) = 1$.

Lemma 2.7. Assume that two random variables X and Y has the same distribution measure. We say they are *identically distributed* and write $X \stackrel{d}{=} Y$. For any measurable function φ such that $\varphi(X)$ and $\varphi(Y)$ are well-defined, we have $\varphi(X) \stackrel{d}{=} \varphi(Y)$.

Proof. Denote by μ the distribution measures of X and Y , respectively. For all $b \in \mathbb{R}$, we have

$$\mathbb{P}(\varphi(X) \leq b) = \mathbb{P}(X \in \varphi^{-1}((-\infty, b])) = \mu(\varphi^{-1}((-\infty, b])) = \mathbb{P}(Y \in \varphi^{-1}((-\infty, b])) = \mathbb{P}(\varphi(Y) \leq b).$$

Then $\varphi(X)$ and $\varphi(Y)$ has the same c.d.f., hence the same distribution measure. \square

Now we introduce independence of random variables.

Definition 2.8 (Independence). Two events $A, B \in \mathcal{F}$ are said to be *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Similarly, given n events $A_1, \dots, A_n \in \mathcal{F}$, they are said to be *mutually independent* if for all index sets $I \subset \{1, \dots, n\}$, it holds

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Remark. A group of pairwise independent events are not always mutually independent. For example, consider the discrete measure on $\Omega = \{1, 2, 3, 4\}$ with a probability mass of $1/4$ at each atom. Let $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$. Then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = 1/4$, and so do event pairs (B, C) and (C, A) . Nevertheless, A, B and C are not mutually independent, since $\mathbb{P}(A \cap B \cap C) = 0$.

Definition 2.9 (Independence of random variables). Two random variables X and Y are said to be *independent* if their joint distribution is the product of marginal distributions:

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y), \quad \forall x, y \in \mathbb{R}.$$

Similarly, given n random variables X_1, \dots, X_n , they are said to be (*mutually*) *independent* if

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i), \quad \forall x_1, \dots, x_n \in \mathbb{R}. \quad (2.3)$$

Clearly, if (2.3) is satisfied, then for all index sets $I \subset \{1, \dots, n\}$, we have

$$\mathbb{P}\left(\bigcup_{i \in I} \{X_i \leq x_i\}\right) = \prod_{i \in I} \mathbb{P}(X_i \leq x_i), \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

Theorem 2.10. Two random variables X and Y are independent if and only if for all $A, B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \quad (2.4)$$

In other words, if $X \sim \mu_X$ and $Y \sim \mu_Y$, then X and Y are independent if and only if $(X, Y) \sim \mu_X \otimes \mu_Y$.

Proof. We only prove that (2.4) holds when X and Y are independent. Given $A = (-\infty, x]$, let

$$\mathcal{M}_A = \{B \in \mathcal{B}(\mathbb{R}) : \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\}.$$

Clearly, \mathcal{M}_A is a λ -system. By independence of X and Y , we have $\mathcal{M}_0 := \{(-\infty, y] : y \in \mathbb{R}\} \subset \mathcal{M}_A$. Since \mathcal{M}_0 is a π -system, and since $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{M}_0)$, by Sierpiński-Dynkin π - λ system, we have $\mathcal{M}_A = \mathcal{B}(\mathbb{R})$.

Given any $B \in \mathcal{B}(\mathbb{R})$, let

$$\mathcal{M}^B = \{A \in \mathcal{B}(\mathbb{R}) : \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\}.$$

We can also verify that \mathcal{M}^B is a λ -system, and $\mathcal{M}_0 \subset \mathcal{M}^B$. Again by Sierpiński-Dynkin π - λ system, we have $\mathcal{M}^B = \mathcal{B}(\mathbb{R})$. Hence (2.4) holds for all $A, B \in \mathcal{B}(\mathbb{R})$. \square

Remark. Let \mathcal{F}_1 and \mathcal{F}_2 be two sub σ -algebras of \mathcal{F} . Then \mathcal{F}_1 and \mathcal{F}_2 are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathcal{F}_1$ and all $B \in \mathcal{F}_2$. By Theorem 2.10, two random variables X and Y are independent indeed implies that the σ -algebras $\sigma(X)$ and $\sigma(Y)$ generated by X and Y are independent.

Corollary 2.11. If X and Y are two independent random variables, and $\psi, \varphi : \mathbb{R} \rightarrow \mathbb{R}$ are Lebesgue-measurable functions, then $\psi(X)$ and $\varphi(Y)$ are independent.

Proof. Let A and B be two Borel sets on \mathbb{R} . Then

$$\begin{aligned} \mathbb{P}(\psi(X) \in A, \varphi(Y) \in B) &= \mathbb{P}(X \in \psi^{-1}(A), Y \in \varphi^{-1}(B)) \\ &= \mathbb{P}(X \in \psi^{-1}(A))\mathbb{P}(Y \in \varphi^{-1}(B)) = \mathbb{P}(\psi(X) \in A)\mathbb{P}(\varphi(Y) \in B). \end{aligned} \quad \square$$

2.2 Expectation

Definition 2.12 (Expectation). Let $X \sim F$ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If X is integrable, the *expectation* of X is defined as the Lebesgue integral

$$\mathbb{E}X = \int X \, d\mathbb{P}.$$

Similarly, for any Lebesgue-measurable function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, if $\varphi \circ X$ is integrable, define

$$\mathbb{E}[\varphi(X)] = \int \varphi \circ X \, d\mathbb{P}.$$

Theorem 2.13 (Integral transform). Using the integral transform formula in Theorem 1.49, we immediately know that the expectation of $\varphi(X)$ equals the Lebesgue-Stieltjes integral

$$\mathbb{E}[\varphi(X)] = \int \varphi \circ X \, d\mathbb{P} = \int \varphi \, d\mu_F =: \int \varphi(x) \, dF(x).$$

Particularly,

$$\mathbb{E}X = \int x \, d\mu_F(x) = \int x \, dF(x).$$

If $X \sim \mu_F$ is a discrete random variable, let \mathcal{A} be the set of all atoms of μ_F . Then we have

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \mathcal{A}} \varphi(x) \mu_F(\{x\}).$$

If $X \sim \mu_F$ is a continuous random variable with density ρ , i.e. μ_F is continuous and $\frac{d\mu_F}{dm} = \rho$, then

$$\mathbb{E}[\mathbb{1}_A(X)] = \int \mathbb{1}_A \, d\mu_F = \mu_F(A) = \int \mathbb{1}_A \rho \, dm.$$

By simple function approximation, for all measurable φ with $\varphi(X)$ integrable, we have

$$\mathbb{E}[\varphi(X)] = \int \varphi \rho \, dm = \int \varphi(x) \rho(x) \, dx.$$

Another useful formula for calculating expectation follows from Fubini's theorem.

Theorem 2.14. Let X be a nonnegative random variable. Then

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > x) \, dx.$$

Proof. By Fubini's theorem,

$$\int_0^\infty \mathbb{P}(X > x) \, dx = \int_0^\infty \mathbb{E}[\mathbb{1}_{\{X > x\}}] \, dx = \mathbb{E} \left[\int_0^\infty \mathbb{1}_{\{X > x\}} \, dx \right] = \mathbb{E} \left[\int_0^X \, dx \right] = \mathbb{E}X.$$

Note that the function

$$\mathbb{1}_{\{X > x\}} = \mathbb{1}_{\{X(\omega) > x\}} = \mathbb{1}_{(0, \infty)}(X(\omega) - x)$$

is defined on $\Omega \times \mathbb{R}$. Since X is measurable, so is $\mathbb{1}_{\{X > x\}}$. □

Proposition 2.15. Let X and Y be two random variables. The following properties of expectation follows from Lebesgue integral:

- (i) If $X \geq 0$ a.s., i.e. $\mathbb{P}(X \geq 0) = 1$, then $\mathbb{E}X \geq 0$. Additionally, if $\mathbb{E}X = 0$, then $X = 0$ a.s..
- (ii) For all $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}X + \beta \mathbb{E}Y$.
- (iii) For $1 \leq p < \infty$, denote $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$.
 - (Hölder's inequality). If $p, q > 1$, $r \geq 1$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, then $\|XY\|_r \leq \|X\|_p \|Y\|_q$;
 - (Moment inequality). If $1 \leq p < q < \infty$, then $\|X\|_p \leq \|X\|_q$.
 - (Minkowski's inequality). If $1 \leq p \leq \infty$, then $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.
- (iv) (Jensen's inequality). If $g : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function, and both $\mathbb{E}X$ and $\mathbb{E}[g(X)]$ are well-defined, then $\mathbb{E}[g(X)] \geq g(\mathbb{E}X)$. In addition, if g is strongly convex and $\mathbb{E}[g(X)] = g(\mathbb{E}X)$, then $X = \mathbb{E}X$ a.s..

Proof. (iii) We first assume $r = 1$. The convexity of $x \mapsto -\ln x$ implies Young's inequality:

$$\frac{1}{p} \ln a^p + \frac{1}{q} \ln b^q \geq \ln \left(\frac{a^p}{p} + \frac{b^q}{q} \right) \Rightarrow \frac{a^p}{p} + \frac{b^q}{q} \leq ab, \quad \forall a, b > 0, \frac{1}{p} + \frac{1}{q} = 1.$$

Then we have

$$\frac{1}{p} \frac{|X|^p}{\|X\|_p^p} + \frac{1}{q} \frac{|Y|^q}{\|Y\|_q^q} \leq \frac{|XY|}{\|X\|_p \|Y\|_q}, \quad \frac{1}{p} + \frac{1}{q} = 1. \quad (2.5)$$

Taking expectation on both sides of (2.5) concludes. For the case $r > 1$, we have

$$\|XY\|_r^r = \| |XY|^r \|_1 \leq \| |X|^r \|_{p/r} \| |Y|^r \|_{q/r} = \|X\|_p^r \|Y\|_q^r, \quad \frac{1}{p} + \frac{1}{q} = \frac{1}{r}.$$

Hence we have Hölder's inequality. By taking $Y = 1$ in Hölder's inequality, we have the moment equality. To obtain Minkowski's inequality ($p > 1$), take $1/q = 1 - 1/p$. Then

$$\begin{aligned} \|X + Y\|_p^p &\leq \mathbb{E}[|X + Y|^{p-1}|X|] + \mathbb{E}[|X + Y|^{p-1}|Y|] \\ &\leq \|X\|_p \left(\mathbb{E}[|X + Y|^{(p-1)q}] \right)^{1/q} + \|Y\|_p \left(\mathbb{E}[|X + Y|^{(p-1)q}] \right)^{1/q} \quad (\text{By Hölder's inequality}) \\ &= (\|X\|_p + \|Y\|_p) \|X + Y\|_p^{p/q} = (\|X\|_p + \|Y\|_p) \|X + Y\|_p^{p-1}. \end{aligned}$$

(iv) Since g is a convex function defined on an open set \mathbb{R} , its subgradient set $\partial_x g$ at x is nonempty for every $x \in \mathbb{R}$. By taking $x_0 = \mathbb{E}X$ and $\alpha \in \partial_{x_0} g$, we have

$$g(x) \geq g(x_0) + \alpha(x - x_0), \quad \forall x \in \mathbb{R}. \quad (2.6)$$

Taking expectation on both sides of (2.6) immediately yields Jensen's inequality.

Now assume g is strongly convex, then the inequality (2.6) becomes strict when $x \neq x_0$. Let

$$\varphi(x) := g(x) - g(x_0) - \alpha(x - x_0),$$

then $\varphi(x) = 0$ implies $x = x_0$, and $\varphi(X)$ is a nonnegative random variable. If $\mathbb{E}[g(X)] = g(x_0)$, we have $\mathbb{E}[\varphi(X)] = 0$. By (i), we have $\varphi(X) = 0$ a.s., and $X = x_0$ a.s.. \square

2.3 Conditional Expectation and Distribution

Now we introduce conditional expectation. In contrast to expectation, which takes real number, the conditional expectation is a random variable.

Definition 2.16 (Conditional expectation). Let \mathcal{G} be a sub σ -algebra of \mathcal{F} , i.e. $\mathcal{G} \subset \mathcal{F}$ and \mathcal{G} is a σ -algebra of subsets of Ω . Let Y be a random variable. If $\mathbb{E}|Y| < \infty$, the *conditional expectation of Y with respect to \mathcal{G}* is defined as any random variable ξ satisfying the following:

- (i) ξ is \mathcal{G} -measurable, i.e. $\xi^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{G}$;
- (ii) $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[\xi\mathbf{1}_A]$ for all $A \in \mathcal{G}$.

Remark. We define a finite signed measure $\mu : \mathcal{G} \rightarrow \overline{\mathbb{R}}$ by assigning $\mu(A) := \mathbb{E}[Y\mathbf{1}_A]$ for all $A \in \mathcal{G}$. Then $\mathbb{P}(A) = 0$ implies $\mu(A) = 0$, and $\mu \ll \mathbb{P}|_{\mathcal{G}}$. By Radon-Nikodym theorem, we take

$$\rho = \frac{d\mu}{d\mathbb{P}|_{\mathcal{G}}} \Rightarrow \mu(A) = \int_A \rho d\mathbb{P}|_{\mathcal{G}} = \mathbb{E}[\rho\mathbf{1}_A], \forall A \in \mathcal{G}.$$

Then ρ is the desired \mathcal{G} -measurable function.

Furthermore, if ξ is a conditional expectation of Y with respect to \mathcal{G} , we define $A = \{\omega : \rho(\omega) \geq \xi(\omega)\}$. Then A is \mathcal{G} -measurable, and

$$\mathbb{E}[(\rho - \xi)\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A] - \mathbb{E}[Y\mathbf{1}_A] = 0 \Rightarrow \rho \leq \xi \text{ a.e..}$$

By symmetry, we also have $\xi \leq \rho$ a.e.. Hence $\xi = \rho$ a.e., and we see that the conditional expectation is uniquely determined almost everywhere, and we write $\mathbb{E}[Y|\mathcal{G}]$ for it.

In fact, the expectation $\mathbb{E}[Y]$ can be viewed as the conditional expectation $\mathbb{E}[Y|\mathcal{F}_0]$, where $\mathcal{F}_0 = \{\Omega, \emptyset\}$ is the smallest σ -algebra on Ω .

Proposition 2.17 (Properties of conditional expectation). Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra of subsets of Ω . Let X and Y be two integrable random variables, that is, $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

- (i) (Total expectation formula). $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}X$. In addition, if \mathcal{H} is another sub σ -algebra of \mathcal{F} such that $\mathcal{H} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$ a.s..
- (ii) (Monotonicity). If $X \geq 0$ a.s., then $\mathbb{E}[X|\mathcal{G}] \geq 0$ a.s.. Hence $X \leq Y$ a.s. implies $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$ a.s.. In particular, $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ a.s..
- (iii) (\mathcal{G} -linearity). For any \mathcal{G} -measurable random variable ξ and η , it holds $\mathbb{E}[\xi X + \eta Y] = \xi \mathbb{E}X + \eta \mathbb{E}Y$ a.s..
- (iv) $\sigma(X)$ is independent of \mathcal{G} if and only if $\mathbb{E}[\varphi(X)|\mathcal{G}] = \mathbb{E}[\varphi(X)]$ a.s. for all measurable $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.
- (v) (Conditional Jensen's inequality). If $g : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function such that $g(X)$ is integrable, then $\mathbb{E}[g(X)|\mathcal{G}] \geq g(\mathbb{E}[X|\mathcal{G}])$ a.s..
- (vi) The following inequalities almost surely hold:

- (Conditional Hölder's inequality). If $p, q > 1$, $r \geq 1$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, then

$$\mathbb{E}[|XY|^r|\mathcal{G}]^{1/r} \leq \mathbb{E}[|X|^p|\mathcal{G}]^{1/p} \mathbb{E}[|Y|^q|\mathcal{G}]^{1/q} \text{ a.s..}$$

- (Conditional moment inequality). If $1 \leq p < q$, then $\mathbb{E}[|X|^p|\mathcal{G}]^{1/p} \leq \mathbb{E}[|X|^q|\mathcal{G}]^{1/q}$ a.s..
- (Conditional Minkowski inequality). If $p \geq 1$, then

$$\mathbb{E}[|X + Y|^p|\mathcal{G}]^{1/p} \leq \mathbb{E}[|X|^p|\mathcal{G}]^{1/p} + \mathbb{E}[|Y|^p|\mathcal{G}]^{1/p} \text{ a.s..}$$

Proof. (i) Let $\xi = \mathbb{E}[X|\mathcal{G}]$, then $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\xi\mathbf{1}_A]$ for all $A \in \mathcal{G}$. Choose $A = \Omega$, so we have $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}X$.

If \mathcal{H} is another sub σ -algebra of \mathcal{F} such that $\mathcal{H} \subset \mathcal{G}$, then a \mathcal{G} -measurable function is also \mathcal{H} -measurable, and $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$ a.s.. Let $\xi = \mathbb{E}[X|\mathcal{G}]$, and $\eta = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$. Then for all $A \in \mathcal{H} \subset \mathcal{G}$, we have $\mathbb{E}[\eta\mathbf{1}_A] = \mathbb{E}[\xi\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$, which implies $\mathbb{E}[X|\mathcal{H}] = \eta$ a.s..

(ii) Let $\xi = \mathbb{E}[X|\mathcal{G}]$, and define $A_n = \{\omega : \xi(\omega) \leq -n^{-1}\} \subset \mathcal{G}$. Then

$$-n^{-1}\mathbb{P}(A_n) \geq \mathbb{E}[\xi\mathbf{1}_{A_n}] = \mathbb{E}[X\mathbf{1}_{A_n}] \geq 0 \Rightarrow \mathbb{P}(A_n) = 0, \forall n \in \mathbb{N}.$$

Let $A = \{\omega : \xi(\omega) < 0\} = \bigcup_{n=1}^{\infty} A_n$. Then $\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$, which implies $\xi \geq 0$ a.s..

Since $|X| - X^+$ and $|X| - X^-$ are a.s. nonnegative, we have $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ a.s..

(iii) The \mathbb{R} -linearity of $\mathbb{E}[\cdot|\mathcal{G}]$ follows from linear operator $\mathbb{E} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$. Now we prove that $\mathbb{E}[\cdot|\mathcal{G}]$ is \mathcal{G} -linear. For all $A \in \mathcal{G}$, we have

$$\mathbb{E}[X\mathbf{1}_A|\mathcal{G}] = \mathbf{1}_A \mathbb{E}[X|\mathcal{G}] \Rightarrow \mathbb{E}[X\mathbf{1}_A\mathbf{1}_B] = \mathbf{1}_A \mathbb{E}[X\mathbf{1}_B], \forall B \in \mathcal{G}.$$

By simple function approximation, for a \mathcal{G} -measurable function ξ such that $\xi X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, we have

$$\mathbb{E}[\xi X\mathbf{1}_B] = \xi \mathbb{E}[X\mathbf{1}_B], \forall B \in \mathcal{G} \Rightarrow \mathbb{E}[\xi X|\mathcal{G}] = \xi \mathbb{E}[X|\mathcal{G}].$$

Hence $\mathbb{E}[\cdot|\mathcal{G}]$ is a \mathcal{G} -linear operator.

(iv) If $\sigma(X)$ and \mathcal{G} are independent, we have

$$\mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B] = \mathbb{P}(\{X \in A\} \cap B) = \mathbb{P}(X \in A) \mathbb{P}(B) = \mathbb{E}[\mathbf{1}_A(X)] \mathbb{E}[\mathbf{1}_B], \forall A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{G}.$$

Since A is arbitrary, by simple function approximation, for any measurable φ such that $\mathbb{E}|\varphi(X)| < \infty$,

$$\mathbb{E}[\varphi(X)\mathbf{1}_B] = \mathbb{E}[\varphi(X)] \mathbb{E}[\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[\varphi(X)]\mathbf{1}_B], \forall B \in \mathcal{G} \Rightarrow \mathbb{E}[\varphi(X)|\mathcal{G}] = \mathbb{E}[\varphi(X)] \text{ a.s..}$$

Conversely, if $\mathbb{E}[X|\mathcal{G}]$, then for all $A \in \mathcal{B}(\mathbb{R})$ and all $B \in \mathcal{G}$, we have

$$\mathbb{P}(\{X \in A\} \cap B) = \mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B|\mathcal{G}]] = \mathbb{E}[\mathbf{1}_B] \mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}(B) \mathbb{P}(X \in A).$$

(v) Since g is a convex function, there exists a countable set $\mathcal{C} \subset \mathbb{R}^2$ such that $g(x) = \sup_{(a,b) \in \mathcal{S}} (a+bx)$. That is, g is the supremum of a countable collection of affine functions. Then $a + bX \leq g(X)$ for all $(a,b) \in \mathcal{S}$. By monotonicity and linearity of conditional expectation, we have $a + b \mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[g(X)|\mathcal{G}]$ a.s. for all $a, b \in \mathcal{S}$. Since \mathcal{S} is countable, we have

$$\begin{aligned} \mathbb{P}\left(\sup_{(a_n, b_n) \in \mathcal{S}} (a_n + b_n \mathbb{E}[X|\mathcal{G}]) > \mathbb{E}[g(X)|\mathcal{G}]\right) &= \mathbb{P}\left(\bigcap_{(a_n, b_n) \in \mathcal{S}} \{a_n + b_n \mathbb{E}[X|\mathcal{G}] > \mathbb{E}[g(X)|\mathcal{G}]\}\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P}(a_k + b_k \mathbb{E}[X|\mathcal{G}] > \mathbb{E}[g(X)|\mathcal{G}]) = 0. \end{aligned}$$

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

(vi) The conditional Hölder's inequality follows from Young's inequality and monotonicity of conditional expectation. The remaining part of this proof is totally parallel to Proposition 2.15 (iii). \square

Remark. Given a random variable $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, where $1 \leq p < \infty$. By conditional Jensen's inequality,

$$(\mathbb{E}[X|\mathcal{G}])^p \leq \mathbb{E}[|X|^p|\mathcal{G}] \text{ a.s.} \Rightarrow \|\mathbb{E}[X|\mathcal{G}]\|_p^p \leq \mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]] = \|X\|_p^p \Rightarrow \frac{\|\mathbb{E}[X|\mathcal{G}]\|_p}{\|X\|_p} \leq 1.$$

Hence $\mathbb{E}[\cdot|\mathcal{G}] : L^p(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a bounded linear operator, and $\|\mathbb{E}[\cdot|\mathcal{G}]\| \leq 1$. Particularly, it can be viewed as a projection operator on the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ of square-integrable variables.

The Lebesgue dominated convergence theorem can be extended to conditional expectation.

Theorem 2.18. (Dominated convergence theorem for conditional expectation). Let \mathcal{G} be a sub σ -algebra of \mathcal{F} . Let $(X_n)_{n=1}^\infty$ be a sequence of random variables such that $X_n \rightarrow X$ a.s.. If there exists a nonnegative integrable random variable $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X_n| \leq Y$ a.s. for all $n \in \mathbb{N}$, then

$$\mathbb{E}[X|\mathcal{G}] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n|\mathcal{G}] \text{ a.s..}$$

Proof. Define $Z_n = \sup_{k \geq n} |X_k - X|$. Then $0 \leq Z_n \leq 2Y$ and $Z_n \rightarrow 0$ a.s.. By Lebesgue dominated convergence theorem, $\mathbb{E}Z_n \rightarrow 0$. Then

$$\mathbb{E}[|\mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]|] \leq \mathbb{E}[\mathbb{E}[|X_n - X||\mathcal{G}]] = \mathbb{E}|X_n - X| \leq \mathbb{E}Z_n \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Hence $\mathbb{E}[X_n|\mathcal{G}]$ converges to $\mathbb{E}[X|\mathcal{G}]$ in L^1 -norm. Furthermore, since Z_n is a lower-bounded decreasing sequence, let $Z = \lim_{n \rightarrow \infty} Z_n$. By Fatou's lemma, we have $\mathbb{E}Z \leq \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{E}[Z_n|\mathcal{G}]] = \lim_{n \rightarrow \infty} \mathbb{E}Z_n = 0$. Hence $Z = 0$ a.s., and $|\mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[Z_n|\mathcal{F}] \rightarrow 0$ a.s.. \square

Theorem 2.19. Let $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ be the space of square-integrable random variables. Define

$$\langle X, Y \rangle = \mathbb{E}[XY], \quad \forall X, Y \in \mathcal{H}.$$

Then $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a Hilbert space. Given a sub σ -algebra $\mathcal{G} \subset \mathcal{F}$, we also define $\mathcal{H}_{\mathcal{G}} = L^2(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$, then $\mathcal{H}_{\mathcal{G}}$ is a closed subspace of \mathcal{H} , and the conditional expectation operator $\mathbb{E}[\cdot|\mathcal{G}]$ is the projection onto $\mathcal{H}_{\mathcal{G}}$, i.e.

$$\mathbb{E}[Y|\mathcal{G}] = \operatorname{argmin}_{X \in \mathcal{H}_{\mathcal{G}}} \mathbb{E}[(Y - X)^2], \quad \forall Y \in \mathcal{H}. \quad (2.8)$$

Proof. The construction of Hilbert space \mathcal{H} and $\mathcal{H}_{\mathcal{G}}$ follows from completeness of L^p -spaces. To prove that $\mathbb{E}[\cdot|\mathcal{G}]$ is the projection onto $\mathcal{H}_{\mathcal{G}}$, it suffices to show $\xi := \mathbb{E}[Y|\mathcal{G}]$ is orthogonal to $Y - \xi$:

$$\langle \xi, Y - \xi \rangle = \mathbb{E}[\xi(Y - \xi)] = \mathbb{E}[\mathbb{E}[\xi(Y - \xi)|\mathcal{G}]] = \mathbb{E}[\xi \mathbb{E}[Y - \xi|\mathcal{G}]] = 0.$$

The equation (2.8) follows from the definition of projection. \square

Similar to Theorem 2.13, we also have the integral transform formula for conditional expectation.

Theorem 2.20 (Conditional integral transform). Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces, and let $T : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ be a measure-preserving transform, i.e. $\mathbb{P}_2 = \mathbb{P}_1 \circ T^{-1}$. If $\varphi \in L^1(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, and $\mathcal{G}_2 \subset \mathcal{F}_2$ is a sub σ -algebra, then

$$\mathbb{E}_1[\varphi \circ T | T^{-1}\mathcal{G}_1] = \mathbb{E}_2[\varphi | \mathcal{G}_2] \circ T, \quad (2.7)$$

where \mathbb{E}_1 and \mathbb{E}_2 are expectation operators on $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, respectively.

Proof. Let $\xi_2 = \mathbb{E}_2[\varphi|\mathcal{G}_2]$. Then $\xi_1 := \xi_2 \circ T$ is a $T^{-1}\mathcal{G}_2$ -measurable function on $(\Omega_1, \mathcal{F}_1)$. By Theorem 1.49, we have $\mathbb{E}_1[\psi \circ T] = \mathbb{E}_2[\psi]$ for all $\psi \in L^1(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. For any $A_1 \in T^{-1}\mathcal{G}_2$, we have $A_2 := TA_1 \in \mathcal{G}_2$, and

$$\begin{aligned}\mathbb{E}_1[\xi_1 \cdot \mathbb{1}_{A_1}] &= \mathbb{E}_1[(\xi_2 \circ T) \cdot \mathbb{1}_{A_1}] = \mathbb{E}_1[(\xi_2 \circ T) \cdot (\mathbb{1}_{A_2} \circ T)] = \mathbb{E}_1[(\xi_2 \cdot \mathbb{1}_{A_2}) \circ T] = \mathbb{E}_2[\xi_2 \cdot \mathbb{1}_{A_2}] \\ &= \mathbb{E}_2[\varphi \cdot \mathbb{1}_{A_2}] = \mathbb{E}_1[(\varphi \cdot \mathbb{1}_{A_2}) \circ T] = \mathbb{E}_1[(\varphi \circ T) \cdot (\mathbb{1}_{A_2} \circ T)] = \mathbb{E}_1[(\varphi \circ T) \cdot \mathbb{1}_{A_1}]\end{aligned}$$

Since $A_1 \in T^{-1}\mathcal{G}_2$ is arbitrary, we have $\mathbb{E}_1[\varphi \circ T | T^{-1}\mathcal{G}_1] = \xi_2 \circ T$ a.s., which is (2.7). \square

Now we introduce the conditional expectation with respect to a random variable.

Definition 2.21 (Conditional expectation). Let X and Y be two random variables. If $\mathbb{E}|Y| < \infty$, the *conditional expectation of Y with respect to X* is defined as

$$\mathbb{E}[Y|X] = \mathbb{E}[Y|\sigma(X)],$$

where $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$ is the σ -algebra generated by X .

Remark. In fact, if the conditional expectation $\mathbb{E}[Y|\sigma(X)]$ exists, it can be written as the composition of a measurable function φ and the random variable X .

Theorem 2.22 (Doob-Dynkin). Let (Ω, \mathcal{F}) and (Γ, \mathcal{G}) be two measurable spaces. Given a measurable function $T : \Omega \rightarrow \Gamma$, let $\sigma(T) := T^{-1}\mathcal{G} \subset \mathcal{F}$ be the σ -algebra generated by T . Then an extended real-valued function $g : \Omega \rightarrow \overline{\mathbb{R}}$ is $\sigma(T)$ -measurable if and only if there exists a \mathcal{G} -measurable function $\varphi : \Gamma \rightarrow \overline{\mathbb{R}}$ such that $g = \varphi \circ T$.

Proof. The sufficiency is clear, so we only prove the necessity. Assume g is nonnegative and $\sigma(T)$ -measurable. By simple function approximation, there exists $\{A_n\}_{n=1}^\infty \subset \sigma(T)$ and nonnegative numbers $\{\alpha_n\}_{n=1}^\infty$ such that

$$g = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n}.$$

Since $A_n \in \sigma(T)$, there exists $B_n \in \mathcal{G}$ such that $A_n = T^{-1}B_n$, namely, $\mathbb{1}_{A_n} = \mathbb{1}_{B_n} \circ T$. Hence we define

$$\varphi = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{B_n} \Rightarrow \varphi \circ T = g.$$

For a general $\sigma(T)$ -measurable function g , there exists \mathcal{G} -measurable function $\varphi^+, \varphi^- : \Gamma \rightarrow \overline{\mathbb{R}}$ such that $g^+ = \varphi^+ \circ T$ and $g^- = \varphi^- \circ T$. Let $E = \{\psi \in \Gamma : \varphi^+(\psi) = \varphi^-(\psi) = \infty\} \in \mathcal{G}$. Then $T^{-1}E = \emptyset$, since $\infty - \infty$ is undefined. Then $\varphi = \mathbb{1}_{\Gamma \setminus E}(\varphi^+ - \varphi^-)$ is the desired \mathcal{G} -measurable function. \square

Remark. By Theorem 2.22, since $\mathbb{E}[Y|\sigma(X)] : (\Sigma, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a $\sigma(X)$ -measurable function, there exists a Borel-measurable function $\varphi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbb{E}[Y|\sigma(X)] = \varphi(X)$.

Theorem 2.23. Let $X \sim \mu_X$ and $Y \sim \mu_Y$ be two independent random variables. If $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ is measurable and $\varphi(X, Y)$ is integrable, then

$$\mathbb{E}[\varphi(X, Y)|Y] = \mathbb{E}[\varphi(X, y)]|_{y=Y} := \int \varphi(x, Y) d\mu_X(x). \quad (2.9)$$

Proof. Define $\psi(y) = \int \varphi(x, y) d\mu_X(x)$, $y \in \mathbb{R}$. For all $A \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned} \mathbb{E}[\psi(Y)\mathbb{1}_A(Y)] &= \int \psi(y)\mathbb{1}_A(y) d\mu_Y(y) \\ &= \int \left(\int \varphi(x, y)\mathbb{1}_A(y) d\mu_X(x) \right) d\mu_Y(y) \\ &= \int \varphi(x, y)\mathbb{1}_A(y) d(\mu_X \otimes \mu_Y)(x, y) = \mathbb{E}[\varphi(X, Y)\mathbb{1}_A(Y)]. \end{aligned} \quad (\text{By Fubini's theorem})$$

Hence we have $\mathbb{E}[\varphi(X, Y)|Y] = \psi(Y)$. \square

Remark. In fact, the equation (2.9) has a more direct form:

$$\mathbb{E}[\varphi(X, Y)|Y = y] = \mathbb{E}[\varphi(X, y)|Y = y] \stackrel{X \perp\!\!\!\perp Y}{=} \mathbb{E}[\varphi(X, y)].$$

Given a sub σ -algebra \mathcal{G} of \mathcal{F} , we can define conditional probability $\mathbb{P}(\cdot|\mathcal{G})$ by conditional expectation: $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$. This automatically induces a probability measure $\mathbb{P}(\cdot|\mathcal{G})(\omega)$ for each $\omega \in \Omega$.

Definition 2.24 (Regular conditional probability). Let \mathcal{G} be a sub σ -algebra of \mathcal{F} . A *regular conditional probability* is a function $\mathbb{P}(\cdot|\mathcal{G})(\cdot) : \mathcal{F} \times \Omega \rightarrow [0, 1]$ satisfying the following:

(i) For \mathbb{P} -a.e. $\omega \in \Omega$,

$$\mathbb{P}(B|\mathcal{G})(\omega) = \mathbb{E}[\mathbb{1}_B|\mathcal{G}](\omega), \quad \forall B \in \mathcal{B}(\mathbb{R});$$

(ii) For \mathbb{P} -a.e. $\omega \in \Omega$, $\mathbb{P}(\cdot|\mathcal{G})(\omega)$ is a probability measure on (Ω, \mathcal{F}) .

Definition 2.25 (Regular conditional distribution). Let \mathcal{G} be a sub σ -algebra of \mathcal{F} . Let $X \sim \mu_X$ be a random variable, where μ_X is a distribution measure. We define $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G}) : \mathcal{B}(\mathbb{R}) \times \Omega \rightarrow [0, 1]$ as follows:

$$\mu_{X|\mathcal{G}}(B|\mathcal{G})(\omega) = \mathbb{P}(X^{-1}(B)|\mathcal{G})(\omega), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

The function $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G}) : \mathcal{B}(\mathbb{R}) \times \Omega \rightarrow [0, 1]$ is called a *regular conditional distribution of Y given \mathcal{G}* if (i) $\omega \mapsto \mu_{X|\mathcal{G}}(B|\mathcal{G})(\omega)$ is \mathcal{F} -measurable for all $B \in \mathcal{B}(\mathbb{R})$, and (ii) $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G})(\omega)$ is a probability measure on $\mathcal{B}(\mathbb{R})$ for \mathbb{P} -a.e. $\omega \in \Omega$. Moreover, for all $A \in \mathcal{G}$ and $B \in \mathcal{B}(\mathbb{R})$, we have

$$\mathbb{P}(\omega \in A, X \in B) = \int_A \mu_{X|\mathcal{G}}(B|\mathcal{G}) d\mathbb{P}.$$

Theorem 2.26. (Existence of regular conditional distribution). Let \mathcal{G} be a sub σ -algebra of \mathcal{F} . Then every random variable X has a regular conditional distribution $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G})$ given \mathcal{G} .

Proof. For all rationals $r \in \mathbb{Q}$, define

$$F(r|\cdot) = \mathbb{P}(X \in (-\infty, r]|\mathcal{G}) := \mathbb{E}[\mathbb{1}_{\{\omega \in \Omega : X(\omega) \in (-\infty, r]\}}|\mathcal{G}].$$

Clearly, for $r \leq s$, we have $\mathbb{1}_{\{X \in (-\infty, r]\}} \leq \mathbb{1}_{\{X \in (-\infty, s]\}}$. By monotonicity of conditional expectation, we have $F(r|\cdot) \leq F(s|\cdot)$ for \mathbb{P} -a.e. $\omega \in \Omega$. Denote $A_{r,s} = \{\omega \in \Omega : F(r|\omega) > F(s|\omega)\}$, so $\mathbb{P}(A_{r,s}) = 0$. Moreover, by dominated convergence theorem (Theorem 2.18), there exist null sets $\{B_r\}_{r \in \mathbb{Q}} \subset \mathcal{F}$ and $C \in \mathcal{F}$ such that

$$\lim_{n \rightarrow \infty} F\left(r + \frac{1}{n} \middle| \omega\right) = F(r|\omega), \quad \forall \omega \in \Omega \setminus B_r$$

as well as

$$\inf_{r \in \mathbb{Q}} F(r|\omega) = 0 \quad \text{and} \quad \sup_{r \in \mathbb{Q}} F(r|\omega) = 1, \quad \forall \omega \in \Omega \setminus C.$$

Let $E = \left(\bigcup_{r,s \in \mathbb{Q}: r < s} A_{r,s} \right) \cap \left(\bigcup_{r \in \mathbb{Q}} B_r \right) \cup C$. Then $\mu(N) = 0$. For $\omega \in \Omega \setminus E$, define

$$\tilde{F}(x|\omega) := \inf_{r \in \mathbb{Q}: r > x} F(r|\omega), \quad \forall x \in \mathbb{R}.$$

Since $F(\cdot|\omega)$ is monotone increasing on \mathbb{Q} , $\tilde{F}(\omega)|_{\mathbb{Q}} = F(\cdot|\omega)$. By construction, $F(\cdot|\omega) : \mathbb{R} \rightarrow [0, 1]$ is a c.d.f., and we can extend this to a unique probability measure $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G})(\omega)$ on $\mathcal{B}(\mathbb{R})$ by Carathéodory's extension theorem. Hence $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G})(\omega)$ is a Borel probability measure for \mathbb{P} -a.e. $\omega \in \Omega$.

For $\omega \in E$, define $\tilde{F}(\cdot|\omega) = F_0$, where F_0 is an arbitrary but fixed c.d.f.. Then for $r \in \mathbb{Q}$ and $B = (-\infty, r]$,

$$\omega \mapsto \mu_{X|\mathcal{G}}(B|\mathcal{G})(\omega) = \mathbf{1}_E(\omega)F_0(r) + \mathbf{1}_{\Omega \setminus E}(\omega)F(r|\omega)$$

is \mathcal{F} -measurable, since $F(r|\omega)$ is \mathcal{G} -measurable by definition. We define

$$\mathcal{D} = \{B \in \mathcal{B}(\mathbb{R}) : \omega \mapsto \mu_{X|\mathcal{G}}(B|\mathcal{G})(\omega) \text{ is } \mathcal{F}\text{-measurable}\}.$$

This is a λ -system, because

- (i) $\mu_{X|\mathcal{G}}(\Omega|\mathcal{G})(\omega) \equiv 1$, which implies $\Omega \in \mathcal{D}$;
- (ii) For $E \subset F \in \mathcal{D}$, $\mu_{X|\mathcal{G}}(F \setminus E|\mathcal{G})(\omega) = \mu_{X|\mathcal{G}}(F|\mathcal{G})(\omega) - \mu_{X|\mathcal{G}}(E|\mathcal{G})(\omega)$, hence $F \setminus E \in \mathcal{D}$;
- (iii) For increasing sequence $B_n \in \mathcal{D}$, $\mu_{X|\mathcal{G}}(B|\mathcal{G})(\omega) = \lim_{n \rightarrow \infty} \mu_{X|\mathcal{G}}(B_n|\mathcal{G})(\omega)$, hence $B := \bigcup_{n=1}^{\infty} B_n \in \mathcal{D}$.

Note that $\{(-\infty, r], r \in \mathbb{Q}\}$ is a π -system generating $\mathcal{B}(\mathbb{R})$. By Sierpinski-Dynkin π - λ theorem, we have $\mathcal{D} = \mathcal{B}(\mathbb{R})$. Then $\mu_{X|\mathcal{G}}(B|\mathcal{G})(\cdot)$ is \mathcal{F} -measurable for all $B \in \mathcal{B}(\mathbb{R})$. Furthermore, for all $A \in \mathcal{G}$ and $B \in \mathcal{B}(\mathbb{R})$,

$$\int_A \mu_{X|\mathcal{G}}(B|\mathcal{G}) \, d\mathbb{P} = \mathbb{E} [\mathbf{1}_A \mathbb{E}[\mathbf{1}_{\{X \in B\}}|\mathcal{G}]] = \mathbb{E}[\mathbf{1}_{A \cap \{X \in B\}}] = \mathbb{P}(\omega \in A, X \in B).$$

Thus we complete the full proof. □

Remark. If X is independent of \mathcal{G} , the conditional distribution $\mu_{X|\mathcal{G}}(\cdot|\mathcal{G})(\omega)$ is the same as the unconditional distribution μ_X for \mathbb{P} -a.e. $\omega \in \Omega$.

Definition 2.27 (Conditional distribution). Let $X \sim \mu_X$ and $Y \sim \mu_Y$ be two random variables, where μ_X and μ_Y are distribution measures. If there exists a collection of probability measures $\{\mu_{Y|X}(\cdot|x)\}_{x \in \mathbb{R}}$ such that for μ_X -a.e. x and all Borel sets A and B ,

$$\mathbb{P}(X \in A, Y \in B) = \int_A \mu_Y(B|x) \, d\mu_X(x),$$

then $\{\mu_Y(\cdot|x)\}_{x \in \mathbb{R}}$ is called the *conditional distribution of Y with respect to X* , and we write $Y|X \sim \mu_{Y|X}$. Furthermore, if there exists two-variable measurable function $(x, y) \mapsto F_{Y|X}(y|x)$ such that $F(\cdot|x)$ is a c.d.f. for any fixed x , and

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{(-\infty, x]} F_{Y|X}(y|t) \, d\mu_X(t),$$

then $\{F_{Y|X}(\cdot|x)\}_{x \in \mathbb{R}}$ is called the *family of conditional distribution functions of Y with respect to X* .

2.4 Stochastic Convergence

Review: Convergence of Random Variables. Here are several categories of convergence of measurable functions we covered in *Section 1.6*. We summarized them in random variable version. Let $X_n : (\Omega, \mathcal{F}) \rightarrow (\mathcal{B}(\mathbb{R}))$ be a sequence of random variables. Let X be a random variable.

(i) (Almost sure convergence). X_n is said to converge almost surely to X if

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

(ii) (Convergence in Probability). X_n is said to converge in probability to X if for all $\eta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \eta) = 0.$$

(iii) (Convergence in L^p -norm). Let $1 \leq p < \infty$. X_n is said to converge to X in L^p -norm if

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

All these modes of convergence can be generalized to the case of random vectors by giving \mathbb{R}^p a proper metric, e.g. the Euclidean distance and L^p -distance.

Since a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a finite measure space, almost sure convergence implies convergence in probability. Also, by Chebyshev inequality, convergence in L^p -norm implies convergence in probability. Moreover, if X_n is a uniformly integrable sequence that converges to X in probability, it also converges to X in L^1 -norm. Now we introduce another convergence of random variables.

Definition 2.28 (Convergence in distribution). A sequence of random variables $X_n \sim F_n$ is said to *converges in distribution* to a random variable $X \sim F$ if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

for each points x of continuity of F , and we write $X \xrightarrow{d} X_n$. We also say that the sequence of cumulative distribution functions F_n *converges weakly* to F , and write $F_n \xrightarrow{w} F$.

Remark. In fact, a sequence of random variables converges in distribution if and only if their distribution measures converges weakly. This is used an alternative definition of convergence in distribution.

Theorem 2.29 (Portmanteau lemma). Let $X_n \sim F_n$ be a sequence of random variables, and $X \sim F$. Then $X_n \xrightarrow{d} X$ if and only if the following equivalent conditions hold:

- (i) $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ for all bounded continuous functions f ;
- (ii) $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ for all bounded Lipschitz continuous functions f ;
- (iii) $\liminf_{n \rightarrow \infty} \mathbb{E}[f(X_n)] \geq \mathbb{E}[f(X)]$ for all lower semi-continuous function f bounded from below;
- (iv) $\limsup_{n \rightarrow \infty} \mathbb{E}[f(X_n)] \leq \mathbb{E}[f(X)]$ for all upper semi-continuous function f bounded from above;
- (v) $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ for every open sets G ;
- (vi) $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ for every closed sets F ;
- (vii) $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in B) \rightarrow \mathbb{P}(X \in B)$ for all continuity sets B , i.e. $\mathbb{P}(X \in \partial B) = 0$.

Proof. \Rightarrow (i): Without loss of generality, we let bounded continuous function f take values in $[-1, 1]$. Assume that F is continuous. By $X_n \xrightarrow{d} X$, we have $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in I) = \mathbb{P}(X \in I)$ for all closed intervals I on \mathbb{R} . Given $\epsilon > 0$, choose a sufficiently large I so that $\mathbb{P}(X \notin I) < \epsilon/5$. Since f is uniformly continuous on the

compact set I , we choose a partition $I = \bigcup_{j=1}^k I_j$ such that f varies at most $\epsilon/5$ on each I_j . Take a point x_j from each I_j , and define $\varphi = f(x_j)\mathbb{1}_{I_j}$, then φ is a simple function, and

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[\varphi(X_n)]| = \frac{\epsilon}{5} + \mathbb{P}(X_n \notin I), \quad |\mathbb{E}[f(X)] - \mathbb{E}[\varphi(X)]| = \frac{\epsilon}{5} + \mathbb{P}(X \notin I) \leq \frac{2\epsilon}{5}. \quad (2.10)$$

Since $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in I) = \mathbb{P}(X \in I)$ and $\mathbb{P}(X \notin I) < \epsilon/5$, there exists N_0 such that $\mathbb{P}(X_n \notin I) < \epsilon/5$ for all $n \geq N_0$. Note that

$$|\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leq \sum_{j=1}^k |\mathbb{P}(X_n \in I_j) - \mathbb{P}(X \in I_j)| |f(x_j)|. \quad (2.11)$$

We also choose N_1 such that $|\mathbb{P}(X_n \in I_j) - \mathbb{P}(X \in I_j)| < \epsilon/(5k)$ for all I_j and all $n \geq N_1$. Combine (2.10) with (2.11) and use triangle inequality, then $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| < \epsilon$ for all $n \geq \max\{N_0, N_1\}$. Since $\epsilon > 0$ is arbitrary, (ii) holds for all continuous c.d.f. F .

If $F : \mathbb{R} \rightarrow [0, 1]$ is not continuous everywhere, we use rarity of discontinuity sets. The collection of sets $\{(-\infty, \alpha] : \alpha \in \mathbb{R}\}$ has disjoint boundaries, and at most countably many of them are discontinuity sets, say $\mathbb{P}(X = \alpha) > 0$. As a result, there exists a dense subset $D \subset \mathbb{R}$ such that F is continuous at each $\alpha \in D$. We choose closed intervals I with boundaries on D .

(vii) \Rightarrow : For each point x of continuity of F , choose $B = (-\infty, x]$. \square

Remark. This theorem can be easily extended to the case of random vectors.

Following is an extremely useful theorem in analysis. It asserts that a continuous mapping preserves several modes of stochastic convergence of random variable sequences. The case of convergence in distribution applies Portmanteau lemma.

Theorem 2.30 (Continuous mapping). Let X be a random variable. If $g : \mathbb{R} \rightarrow \mathbb{R}$ is continuous everywhere on a set C such that $\mathbb{P}(X \in C) = 1$, then g preserves the following modes of convergence:

- (i) If $X_n \xrightarrow{a.s.} X$, then $g(X_n) \xrightarrow{a.s.} g(X)$;
- (ii) If $X_n \xrightarrow{\mathbb{P}} X$, then $g(X_n) \xrightarrow{\mathbb{P}} g(X)$;
- (iii) If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.

Proof. (i) is trivial. (ii) Given $\eta > 0$, define

$$E_k = \left\{ x \in C : \exists y \in \mathbb{R} \text{ such that } |y - x| < \frac{1}{k} \text{ and } |g(y) - g(x)| \geq \eta \right\}, \quad k \in \mathbb{N}.$$

Since g is continuous everywhere on C , the sequence $E_k \searrow \emptyset$. Then

$$\mathbb{P}(|g(X_n) - g(X)| \geq \eta) \leq \mathbb{P}(X \in E_k) + \mathbb{P}\left(|X_n - X| \geq \frac{1}{k}\right) \quad (2.12)$$

Given $\epsilon > 0$, we first choose K such that $\mathbb{P}(X \in E_K) \leq \epsilon/2$, then choose N such that $\mathbb{P}(|X_n - X| \geq 1/K) < \epsilon/2$ for all $n \geq N$. Hence (2.12) is controlled by arbitrarily small $\epsilon > 0$.

(iii) Let $F \subset \mathbb{R}$ be a closed set. If $x \in \overline{g^{-1}(F)}$, there exists sequence $x_k \in g^{-1}(F)$ such that $x_k \rightarrow x$ and $g(x_k) \in F$. Since F is closed and g is continuous on C , if $x \in C$, we have $g(x) \in F$. Hence the following inclusions hold for all closed set F :

$$g^{-1}(F) \subset \overline{g^{-1}(F)} \subset g^{-1}(F) \cup C^c.$$

Using the Portmanteau lemma, we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(g(X_n) \in F) \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(X_n \in \overline{g^{-1}(F)}\right) \leq \mathbb{P}\left(X \in \overline{g^{-1}(F)}\right) \leq P(g(X) \in F),$$

where the last inequality holds because $\mathbb{P}(X \in C) = 0$. Again by Portmanteau lemma, $g(X_n) \xrightarrow{d} g(X)$. \square

Another important property associated with weak convergence is uniform tightness. Any random variable X is *tight* or $O_p(1)$, i.e. for each $\epsilon > 0$, there exists $M > 0$ such that $\mathbb{P}(|X| > M) < \epsilon$. This is a consequence of the properties of c.d.f. $F_X(x) = \mathbb{P}(X \leq x)$.

Definition 2.31. (Uniform tightness). A collection of random variables $\{X_\alpha, \alpha \in J\}$ is said to be *uniformly tight*, if for every $\epsilon > 0$, there exists a constant $M > 0$ such that

$$\sup_{\alpha \in J} \mathbb{P}(|X_\alpha| > M) < \epsilon.$$

Clearly, any finite collection of random variable is uniformly tight.

The following Prohorov's theorem is a generalization of Heine-Borel theorem.

Theorem 2.32 (Prohorov). Let X_n be a sequence of random variables.

- (i) If $X_n \xrightarrow{d} X$ for some random variable X , then $\{X_n : n \in \mathbb{N}\}$ is uniformly tight;
- (ii) If $\{X_n : n \in \mathbb{N}\}$ is uniformly tight, then there exists a subsequence X_{n_k} that converges in distribution to some random variable X .

Proof of Theorem 2.32 (i). Given $\epsilon > 0$, we choose M_0 such that $\mathbb{P}(|X| > M_0) < \epsilon/2$. By Portmanteau's theorem, we have $\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n| > M_0) \leq \mathbb{P}(|X| > M_0)$. Hence we can choose N such that

$$\sup_{n \geq N} \mathbb{P}(|X_n| > M_0) < \mathbb{P}(|X| > M_0) + \frac{\epsilon}{2} < \epsilon.$$

Note that any finite collection of random variables is uniformly tight. Then we choose M_1 such that $\mathbb{P}(|X_j| > M_1) < \epsilon$ for all $j = 1, \dots, N-1$. Let $M = \max\{M_0, M_1\}$, then $\sup_{n \in \mathbb{N}} \mathbb{P}(|X_n| > M) < \epsilon$. \square

The proof of Theorem 2.32 (ii) uses Helly's selection theorem.

Theorem 2.33 (Helly's selection theorem). Let $f_n : \mathbb{R} \rightarrow [-M, M]$ be a uniformly bounded sequence of monotone increasing functions. Then there exists a subsequence $(f_{n_k})_{k=1}^\infty$ that converges pointwise to an monotone increasing function $f : \mathbb{R} \rightarrow [-M, M]$.

Proof. Choose a countable dense subset $\mathbb{Q} = \{r_k, k \in \mathbb{N}\}$ of \mathbb{R} . Then $f_n(r_1)$ is a bounded sequence. By Bolzano-Weierstrass theorem, choose a convergent subsequence $f_{1n}(r_1) \rightarrow f(r_1)$. Then $f_{1n}(r_2)$ is a bounded sequence, and again we choose one of its convergent subsequence $f_{2n}(r_2) \rightarrow f(r_2)$.

"diagonal trick": Repeat this procedure, so for each $k \in \mathbb{N}$, we choose a subsequence f_{kn} such that $f_{kn}(r_j) \rightarrow f(r_j)$ for all indices $j \leq k$. Since $(f_{kn})_{n=1}^\infty$ is a subsequence of its predecessor $(f_{k-1,n})_{n=1}^\infty$, $(f_{nn}(r_k))_{n=1}^\infty$ is a subsequence of $(f_{kn}(r_k))_{n=1}^\infty$ from $n = k$ on, and we have $\lim_{n \rightarrow \infty} f_{nn}(r_k) = f(r_k)$ for all $k \in \mathbb{N}$. Hence we obtain a subsequence f_{nn} that converges to f pointwise on \mathbb{Q} . Clearly, $f : \mathbb{Q} \rightarrow [-M, M]$ is increasing.

For all irrationals $x \in \mathbb{R} \setminus \mathbb{Q}$, choose a increasing rational sequence $r_{k_j} \rightarrow x$, and let $f(x) = \lim_{j \rightarrow \infty} f(r_{k_j})$. Note this limit exists because $f(r_{k_j})$ is a bounded increasing sequence. Clearly, f is increasing on \mathbb{R} and bounded by M , and $r_{k_i} < x < r_{k_j}$ implies $f_{nn}(r_{k_i}) - f(r_{k_j}) < f_{nn}(x) - f(x) < f_{nn}(r_{k_j}) - f(r_{k_i})$ for all $n \in \mathbb{N}$.

Finally we prove $f_{nn} \rightarrow f$ pointwise on \mathbb{R} . Given $\epsilon > 0$. If $x \in \mathbb{R} \setminus \mathbb{Q}$ is a point of continuity of f , we choose rationals $r < x < r'$ with $|f(r_{k_i}) - f(r_{k_j})| < \epsilon$. Then

$$-\epsilon \leq \liminf_{n \rightarrow \infty} (f_{nn}(r) - f(r')) \leq f_{nn}(x) - f(x) \leq \limsup_{n \rightarrow \infty} (f_{nn}(r') - f(r)) < \epsilon.$$

Hence f_{nn} converges pointwise to f , except possibly at points of discontinuity of f . Being monotone increasing, f has at most countably points of discontinuity. Since f_{nn} is uniformly bounded by M , we repeat the “diagonal” trick to obtain a subsequence of f_n that converges everywhere on \mathbb{R} . \square

Corollary 2.34 (Helly’s selection theorem). Let $F_n : \mathbb{R} \rightarrow [0, 1]$ be a sequence of cumulative distribution functions. Then there exists a subsequence F_{n_k} such that $F_{n_k}(x) \rightarrow F(x)$ at each point x of continuity of a possibly defective distribution function F , i.e. F only satisfies properties (i) and (ii) in Proposition 2.2.

Proof. By Theorem 2.33, we choose a subsequence F_{n_k} of F_n that converges pointwise to an increasing function $G : \mathbb{R} \rightarrow [0, 1]$. Define $F(x) = \lim_{\epsilon \rightarrow 0+} G(x + \epsilon)$ for all $x \in \mathbb{R}$. Then F is right-continuous on \mathbb{R} , and F_{n_k} converges to F at all points of continuity of F . \square

Proof of Theorem 2.32 (ii). Let $X_n \sim F_n$. By Helly’s selection theorem, there exists a subsequence F_{n_k} of the c.d.f. sequence F_n that converges to a possibly defective distribution function F . It suffices to show that F is proper. That is, $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.

Given $\epsilon > 0$, by uniform tightness of $\{X_n, n \in \mathbb{N}\}$, we choose $M > 0$ such that $F(M) > 1 - \epsilon$. Since the points of discontinuity of F are rare, we slide M slightly larger so that M is a point of continuity of F . Then $F_{n_k}(M) \rightarrow F(M) > 1 - \epsilon$. Since $\epsilon > 0$ is arbitrary, $F(x) \rightarrow 1$ as $x \rightarrow \infty$. The case $x \rightarrow -\infty$ is similar. \square

Now we discuss the relationship between convergence in probability and convergence in distribution.

Theorem 2.35. Let X_n, X, Y_n and Y be random variables. Let $c \in \mathbb{R}$ be a constant. Then

- (i) If $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$, then $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$;
- (ii) If $X_n \xrightarrow{d} X$ and $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$, then $Y_n \xrightarrow{d} X$;
- (iii) If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$;
- (iv) $X_n \xrightarrow{\mathbb{P}} c$ if and only if $X_n \xrightarrow{d} c$;
- (v) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $(X_n, Y_n) \xrightarrow{d} (X, c)$;

Proof. (i) The result follows since $\rho((x_n, y_n), (x, y)) := \sqrt{|x - x_n|^2 + |y - y_n|^2} \leq |x - x_n| + |y - y_n|$.

(ii) For every bounded 1-Lipschitz continuous function $f : \mathbb{R} \rightarrow [0, 1]$, we have

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| \leq \epsilon \mathbb{E}[\mathbb{1}_{\{|X_n - Y_n| \leq \epsilon\}}] + \mathbb{E}[\mathbb{1}_{\{|X_n - Y_n| > \epsilon\}}] \leq \epsilon + \mathbb{P}(|X_n - Y_n| > \epsilon), \quad \forall \epsilon > 0.$$

Since $\epsilon > 0$ is arbitrary, and $\mathbb{P}(|X_n - Y_n| > \epsilon)$ converges to zero, we have $\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)] \rightarrow 0$. By Portmanteau lemma, $\mathbb{E}[f(Y_n)] \rightarrow \mathbb{E}[f(X)]$, and $Y_n \xrightarrow{d} X$.

(iii) Since $X \xrightarrow{d} X$ trivially, this is a special case of (ii).

(iv) The “only if” case is a special case of (iii). For the converse, given any $\epsilon > 0$, by Portmanteau lemma, $X_n \xrightarrow{d} c$ implies $X_n \xrightarrow{\mathbb{P}} c$:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n - c| \geq \epsilon) = \limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in \mathbb{R} \setminus (c - \epsilon, c + \epsilon)) \leq \mathbb{P}(c \in \mathbb{R} \setminus (c - \epsilon, c + \epsilon)) = 0.$$

(v) Since $\rho((X_n, Y_n), (X_n, c)) = |Y_n - c| \xrightarrow{\mathbb{P}} 0$, by (ii), it suffices to show that $(X_n, c) \xrightarrow{d} (X, c)$. For every $f \in C_b(\mathbb{R}^2)$, the mapping $x \mapsto f(x, c)$ is also bounded and continuous. By Portmanteau lemma, we have $\mathbb{E}[f(X_n, c)] \rightarrow \mathbb{E}[f(X, c)]$. Thus $(X_n, c) \xrightarrow{d} (X, c)$, and the result follows. \square

We have the following useful corollary.

Lemma 2.36 (Slutsky). Let X_n , X and Y_n be random variables. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c \in \mathbb{R}$, then

- (i) $X_n + Y_n \xrightarrow{d} X + c$;
- (ii) $X_n Y_n \xrightarrow{d} cX$;
- (iii) If $c \neq 0$, then $Y_n^{-1} X_n \xrightarrow{d} c^{-1} X$.

Proof. By Theorem 2.35 (v) and continuous mapping theorem [Theorem 2.30 (iii)]. □

Finally we introduce small o and big O symbols.

Definition 2.37 (Stochastic o and O symbols). The notation $o_{\mathbb{P}}(1)$ denotes a sequence of random variables that converges to 0 in probability. The notation $O_{\mathbb{P}}(1)$ denotes a sequence of random variables that is uniformly tight. More generally, given a sequence of random variables R_n ,

$$\begin{aligned} X_n = o_{\mathbb{P}}(R_n) &\Leftrightarrow X_n = Y_n R_n \text{ and } Y_n \xrightarrow{\mathbb{P}} 0; \\ X_n = O_{\mathbb{P}}(R_n) &\Leftrightarrow X_n = Y_n R_n \text{ and } Y_n = O_{\mathbb{P}}(1). \end{aligned}$$

Theorem 2.38 (Calculus with o and O symbols).

- (i) $o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$;
- (ii) $o_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) = O_{\mathbb{P}}(1)$;
- (iii) $O_{\mathbb{P}}(1) o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$;
- (iv) $(1 + o_{\mathbb{P}}(1))^{-1} = O_{\mathbb{P}}(1)$;
- (v) $o_{\mathbb{P}}(R_n) = R_n o_{\mathbb{P}}(1)$, $O_{\mathbb{P}}(R_n) = R_n O_{\mathbb{P}}(1)$;
- (vi) $o_{\mathbb{P}}(O_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)$;

Proof. (i), (v) follows from definition.

(ii) Let $X_n = o_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$. Given $\epsilon > 0$, choose M such that $\mathbb{P}(|Y_n| > M/2) < \epsilon/2$ for all $n \in \mathbb{N}$, and choose N such that $\mathbb{P}(|X_n| > M/2) < \epsilon/2$ for all $n \geq N$. Then $\mathbb{P}(|X_n + Y_n| > M) < \epsilon$ for all $n \geq N$. Since $(X_n + Y_n)_{n=N}^{\infty}$ is uniformly tight, so is $(X_n + Y_n)_{n=1}^{\infty}$.

(iii) Let $X_n = o_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$. Given $\epsilon > 0$, choose M such that $\mathbb{P}(|Y_n| > M) < \epsilon/2$. Given $\eta > 0$, choose N such that $\mathbb{P}(|X_n| > \eta/M) < \epsilon/2$ for all $n \geq N$. Then $\mathbb{P}(|X_n Y_n| > \eta) < \epsilon$ for all $n \geq N$.

(iv) Let $X_n = o_{\mathbb{P}}(1)$. For any $\epsilon > 0$, there exists $0 < \eta < 1$ and $N > 0$ such that $\mathbb{P}(|X_n| > \eta) < \epsilon$. Then $\mathbb{P}((1 + X_n)^{-1} > \frac{1}{1-\eta}) < \epsilon$. As a result, $((1 + X_n)^{-1})_{n=1}^{\infty}$ is uniformly tight, and so is $((1 + X_n)^{-1})_{n=N}^{\infty}$.

(vi) Follows from (iii) and (v). □

Theorem 2.39. Let $R : \mathbb{R} \rightarrow \mathbb{R}$ be a function such that $R(0) = 0$. Let $X_n = o_{\mathbb{P}}(1)$. Then for every $p > 0$,

- (i) If $R(h) = o(|h|^p)$ as $h \rightarrow 0$, then $R(X_n) = o_{\mathbb{P}}(|X_n|^p)$;
- (ii) If $R(h) = O(|h|^p)$ as $h \rightarrow 0$, then $R(X_n) = O_{\mathbb{P}}(|X_n|^p)$.

Proof. Define g as $g(h) = |h|^{-p} R(h)$ for $h \neq 0$ and $g(0) = 0$. Then $R(X_n) = |X_n|^p g(X_n)$.

(i) By assumption, g is continuous at 0. Then $g(X_n) \xrightarrow{\mathbb{P}} g(0) = 0$ by continuous mapping theorem.

(ii) By assumption, there exists $\delta > 0$ and $M > 0$ such that $|g(h)| \leq M$ for all $|h| \leq \delta$. Then we have $\mathbb{P}(|g(X_n)| > M) \leq \mathbb{P}(|X_n| > \delta) \rightarrow 0$, and the sequence $g(X_n)$ is uniformly tight. □

2.5 Characteristic Functions

Definition 2.40. Let $X \sim \mu$ be a (real-valued) random variable, where μ is a distribution measure. The *characteristic function of X* is defined as

$$\phi_X : \mathbb{R} \rightarrow \mathbb{C}, \quad \phi_X(\lambda) = \mathbb{E} [e^{i\lambda X}] = \int_{\mathbb{R}} e^{i\lambda x} d\mu(x), \quad i^2 = -1.$$

Proposition 2.41 (Properties of characteristic functions). If ϕ_X is the characteristic function of a random variable $X \sim \mu$, the following are true:

- (i) $\phi_X(0) = 1$;
- (ii) $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ is bounded and uniformly continuous.
- (iii) If $\mathbb{E}[|X|^n] < \infty$ for some $n \in \mathbb{N}$, then ϕ_X is n -differentiable, and its k -th derivative is

$$\phi_X^{(k)}(\lambda) = \mathbb{E} [(iX)^k e^{i\lambda X}], \quad k = 1, \dots, n. \quad (2.13)$$

Furthermore, all these derivatives are uniformly continuous. Particularly, we have $\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$.

- (iv) ϕ_X is twice differentiable if and only if $\mathbb{E}[X^2] < \infty$. Generally, for each $k \in \mathbb{N}$, ϕ_X is $2k$ -differentiable if and only if $\mathbb{E}[X^{2k}] < \infty$.
- (v) If X is a continuous variable, then $\lim_{\lambda \rightarrow \pm\infty} \phi_X(\lambda) = 0$.

Proof. (i) is clear by definition. To prove (ii), note that $|\phi_X(\lambda)| < 1$ for all $\lambda \in \mathbb{R}$. For uniform continuity, we use the following inequality:

$$e^{i\theta} - 1 = 2ie^{\frac{i\theta}{2}} \sin \frac{\theta}{2} \Rightarrow |e^{i\theta} - 1| \leq 2 \left| \sin \frac{\theta}{2} \right| \leq |\theta|.$$

Then for all $\lambda_1, \lambda_2 \in \mathbb{R}$, we have

$$|\phi_X(\lambda_1) - \phi_X(\lambda_2)| = \left| \mathbb{E} \left[\left(e^{i(\lambda_1 - \lambda_2)X} - 1 \right) e^{i\lambda_2 X} \right] \right| \leq \mathbb{E} \left| e^{i(\lambda_1 - \lambda_2)X} - 1 \right| \leq 2 \int_{\mathbb{R}} \left| \sin \frac{(\lambda_1 - \lambda_2)x}{2} \right| d\mu(x). \quad (2.14)$$

Given $\epsilon > 0$, we choose $[-R, R]$ such that $\mu([-R, R]) > 1 - \epsilon/4$. Then whenever $|\lambda_1 - \lambda_2| < \frac{\epsilon}{2R}$, we have

$$2 \int_{\mathbb{R}} \left| \sin \frac{(\lambda_1 - \lambda_2)x}{2} \right| d\mu(x) \leq 2 \int_{-R}^R \left| \sin \frac{(\lambda_1 - \lambda_2)x}{2} \right| d\mu(x) + \frac{\epsilon}{2} < 2 \int_{[-R, R]} \frac{\epsilon}{4} d\mu + \frac{\epsilon}{2} = \epsilon. \quad (2.15)$$

Hence ϕ_X is uniformly continuous.

- (iii) Assume $\mathbb{E}|X| < \infty$. By Lebesgue dominated convergence theorem, since $\frac{1}{\epsilon} |\exp(i\epsilon X) - 1| \leq |X|$,

$$\lim_{\epsilon \rightarrow 0} \frac{\phi_X(\lambda + \epsilon) - \phi_X(\lambda)}{\epsilon} = \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[\frac{e^{i\epsilon X} - 1}{\epsilon} e^{i\lambda X} \right] = \mathbb{E} \left[\lim_{\epsilon \rightarrow 0} \frac{e^{i\epsilon X} - 1}{\epsilon} e^{i\lambda X} \right] = \mathbb{E} [iX e^{i\lambda X}].$$

Hence ϕ_X is differentiable. Furthermore, by monotone convergence theorem, we choose $R_1 > 0$ such that $\mathbb{E}[|X| \mathbf{1}_{\{|X| > R_1\}}] < \epsilon/2$. Alike (2.14) and (2.15), whenever $|\lambda_1 - \lambda_2| < \epsilon/(2R_1^2)$, we have

$$|\phi_X'(\lambda_1) - \phi_X'(\lambda_2)| = \mathbb{E} [iX (e^{i\lambda_1 X} - e^{i\lambda_2 X})] < 2 \int_{-R_1}^{R_1} \left| x \sin \frac{(\lambda_1 - \lambda_2)x}{2} \right| d\mu(x) + \frac{\epsilon}{2} < \epsilon.$$

Therefore the derivative ϕ_X' is uniformly continuous. Now assume (2.13) holds for $k - 1$. If $\mathbb{E}[|X|^k] < \infty$,

$$\lim_{\epsilon \rightarrow 0} \frac{\phi_X^{(k-1)}(\lambda + \epsilon) - \phi_X^{(k-1)}(\lambda)}{\epsilon} = \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[\frac{e^{i\epsilon X} - 1}{\epsilon} (iX)^{k-1} e^{i\lambda X} \right] = \mathbb{E} \left[\lim_{\epsilon \rightarrow 0} \frac{e^{i\epsilon X} - 1}{\epsilon} (iX)^{k-1} e^{i\lambda X} \right] = \mathbb{E} [(iX)^k e^{i\lambda X}].$$

Therefore ϕ_X is k -differentiable. Again by monotone convergence theorem, we choose $R_k > 0$ such that $\mathbb{E}[|X|^k \mathbf{1}_{\{|X| > R_k\}}] < \epsilon/2$. Whenever $|\lambda_1 - \lambda_2| < \epsilon/(2R_k^{k+1})$, we have

$$\left| \phi_X^{(k)}(\lambda_1) - \phi_X^{(k)}(\lambda_2) \right| = \mathbb{E}[(iX)^k (e^{i\lambda_1 X} - e^{i\lambda_2 X})] < 2 \int_{-R_k}^{R_k} \left| x^k \sin \frac{(\lambda_1 - \lambda_2)x}{2} \right| d\mu(x) + \frac{\epsilon}{2} < \epsilon.$$

Hence $\phi_X^{(k)}$ is uniformly continuous. By induction we finish the proof of (iii).

(iv) We only prove necessity, since (iii) implies sufficiency. By definition, if $\phi_X''(0)$ exists, we have

$$\phi_X''(0) = \lim_{h \rightarrow 0} \frac{\phi_X(h) + \phi_X(-h) - 2\phi(0)}{h^2} = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{2\cos(hx) - 2}{h^2} d\mu(x) = -2 \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{1 - \cos(hx)}{h^2} d\mu(x).$$

By Fatou's lemma, we have

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 d\mu(x) = 2 \int_{\mathbb{R}} \lim_{h \rightarrow 0} \frac{1 - \cos(hx)}{h^2} d\mu(x) \leq 2 \liminf_{h \rightarrow 0} \int_{\mathbb{R}} \frac{1 - \cos(hx)}{h^2} d\mu(x) = -\phi_X''(0) < \infty.$$

Generally, if $\phi_X^{(2k-2)}(0)$ exists, then $\mathbb{E}[X^{2k-2}] < \infty$, and by (iii), $\phi_X^{(2k-2)}(\lambda) = \mathbb{E}[(iX)^{2k-2} e^{i\lambda X}]$. Then

$$\begin{aligned} \phi_X^{(2k)}(0) &= \lim_{h \rightarrow 0} \frac{\phi_X^{(2k-2)}(h) + \phi_X^{(2k-2)}(-h) - 2\phi(0)}{h^2} = \lim_{h \rightarrow 0} \int_{\mathbb{R}} (ix)^{2k-2} \frac{2\cos(hx) - 2}{h^2} d\mu(x) \\ &= (-1)^{k-1} 2 \lim_{h \rightarrow 0} \int_{\mathbb{R}} x^{2k-2} \frac{1 - \cos(hx)}{h^2} d\mu(x). \end{aligned}$$

By Fatou's lemma, we have

$$\begin{aligned} \mathbb{E}[X^{2k}] &= \int_{\mathbb{R}} x^{2k} d\mu(x) = 2 \int_{\mathbb{R}} \lim_{h \rightarrow 0} x^{2k-2} \frac{1 - \cos(hx)}{h^2} d\mu(x) \\ &\leq 2 \liminf_{h \rightarrow 0} \int_{\mathbb{R}} x^{2k-2} \frac{1 - \cos(hx)}{h^2} d\mu(x) = (-1)^{k-1} \phi_X^{(2k-2)}(0) < \infty. \end{aligned}$$

Then (iv) follows from induction.

(v) Since X is continuous, there exists a density function $\rho \in L^1(\mathbb{R})$ of X , and $\phi_X(\lambda) = \int_{\mathbb{R}} \rho(x) e^{i\lambda x} dx$. The conclusion immediately follows from Riemann-Lebesgue lemma. We give a complete proof here.

Firstly, suppose $\rho \in C_c(\mathbb{R})$. For $\lambda \neq 0$, the substitution $x \rightarrow x - \frac{\pi}{\lambda}$ implies

$$\phi_X(\lambda) = \int_{\mathbb{R}} \rho(x) e^{i\lambda x} dx = \int_{\mathbb{R}} \rho\left(x - \frac{\pi}{\lambda}\right) e^{i\lambda x} e^{i\pi} dx = - \int_{\mathbb{R}} \rho\left(x - \frac{\pi}{\lambda}\right) e^{i\lambda x} dx$$

Use the two formulae to compute $\phi_X(\lambda)$, we have

$$|\phi_X(\lambda)| \leq \frac{1}{2} \int_{\mathbb{R}} \left| \rho(x) - \rho\left(x - \frac{\pi}{\lambda}\right) \right| dx.$$

Since ρ is continuous, $|\rho(x) - \rho(x - \frac{\pi}{\lambda})| \rightarrow 0$ as $|\lambda| \rightarrow \infty$ for all $x \in \mathbb{R}$. By Lebesgue dominated convergence theorem, $|\phi_X(\lambda)| \rightarrow 0$ as $|\lambda| \rightarrow \infty$.

For the general case $\rho \in L^1(\mathbb{R})$, we use function approximation. Since $C_c(\mathbb{R})$ is dense in $L^1(\mathbb{R})$, we can choose $f \in C_c(\mathbb{R})$ such that $\|\rho - f\|_1 \leq \epsilon$ for any $\epsilon > 0$. Then

$$\limsup_{\lambda \rightarrow \pm\infty} |\phi_X(\lambda)| \leq \limsup_{\lambda \rightarrow \pm\infty} \left| \int (\rho(x) - f(x)) e^{i\lambda x} dx \right| + \limsup_{\lambda \rightarrow \pm\infty} \left| \int f(x) e^{i\lambda x} dx \right| \leq \epsilon + 0 = \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $|\phi_X(\lambda)| \rightarrow 0$ as $\lambda \rightarrow \pm\infty$. □

In fact, we can determine a measure distribution uniquely by its characteristic function.

Theorem 2.42 (Inversion formula). Let F be a c.d.f., and ϕ_F is the associated characteristic function.

(i) For any two points $a < b$ of continuity of F ,

$$F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dt.$$

Distinct c.d.f. F have distinct characteristic functions.

(ii) For any $\varphi \in C_c(\mathbb{R})$,

$$\int \varphi dF = \lim_{\delta \rightarrow 0^+} \frac{1}{2\pi} \int \varphi(x) \left(\int e^{-itx} \phi_F(t) e^{-\frac{\delta}{2}t^2} dt \right) dx.$$

(iii) If $\|\phi_F\|_1 := \int |\phi_F(t)| dt < \infty$, then F has density function ρ :

$$\rho(x) = \frac{1}{2\pi} \int e^{-itx} \phi_F(t) dt,$$

and $\sup_{x \in \mathbb{R}} \rho(x) \leq \frac{1}{2\pi} \|\phi_F\|_1 < \infty$.

Proof. (ii) Let $Z \sim N(0, 1)$ be independent of $X \sim F$. Then for all $\varphi \in C_c(\mathbb{R})$ and $\delta > 0$,

$$\begin{aligned} \mathbb{E} [\varphi(X + \sqrt{\delta}Z)] &= \mathbb{E} \left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) e^{-\frac{\lambda^2}{2}} d\lambda \right] \\ &= \mathbb{E} \left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) \mathbb{E} [e^{-i\lambda Z}] d\lambda \right] = \mathbb{E} \left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) e^{-i\lambda Z} d\lambda \right] \\ &= \mathbb{E} \left[\frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) e^{-i\frac{\xi - X}{\sqrt{\delta}} Z} d\xi \right] = \frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) \mathbb{E} \left[e^{-i\frac{\xi - X}{\sqrt{\delta}} Z} \right] d\xi \\ &= \frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) \mathbb{E} \left[e^{-i\frac{\xi Z}{\sqrt{\delta}}} \phi_F \left(\frac{Z}{\sqrt{\delta}} \right) \right] d\xi = \frac{1}{2\pi\sqrt{\delta}} \int \varphi(\xi) \left(\int e^{-i\frac{\xi z}{\sqrt{\delta}}} \phi_F \left(\frac{z}{\sqrt{\delta}} \right) e^{-\frac{z^2}{2}} dz \right) d\xi \end{aligned}$$

Note that φ is continuous and bounded. Let $\delta \searrow 0$, we obtain (ii) by dominated convergence theorem.

(iii) If $a < b$ are points of continuity of F , then by (i),

$$|F(b) - F(a)| \leq \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \left| \frac{e^{-iat} - e^{-ibt}}{it} \right| |\phi_F(t)| dt.$$

Since $\left| \frac{\exp(-iat) - \exp(-ibt)}{it} \right| \leq |b - a|$, we have

$$|F(b) - F(a)| \leq \frac{|b - a|}{2\pi} \|\phi_F\|_1. \quad (2.16)$$

For general $a < b$, we can find two sequence $a_n \nearrow a$ and $b_n \searrow b$ of points of continuity of F . Hence the estimate (2.16) holds for all $a < b$, and F is continuous. As a result, F has density ρ , and

$$\frac{F(b) - F(a)}{b - a} \leq \frac{1}{2\pi} \int \frac{e^{iat} - e^{ibt}}{i(b - a)t} \phi_F(t) dt.$$

Let $b \searrow a$, the equation in (iii) holds, and the estimate of upper bound follows from (2.16). \square

The proof of (i) requires some technical lemma.

Lemma 2.43 (Dirichlet). Let $\alpha \in \mathbb{R}$. Then

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin(\alpha t)}{t} dt = \pi \operatorname{sgn}(\alpha),$$

where $\operatorname{sgn}(\alpha) = \mathbb{1}_{(0, \infty)}(\alpha) - \mathbb{1}_{(-\infty, 0)}(\alpha)$.

Proof. The case $\alpha = 0$ is clear. It suffices to prove the case $\alpha = 1$. Since $(s, t) \mapsto e^{-st}$ is absolutely integrable on $[0, \infty) \times [0, \infty)$, by Fubini's theorem,

$$\int_0^\infty \frac{\sin t}{t} dt = \int_0^\infty \left(\int_0^\infty e^{-st} \sin t ds \right) dt = \int_0^\infty \left(\int_0^\infty e^{-st} \sin t dt \right) ds = \int_0^\infty \frac{1}{1+s^2} ds = \frac{\pi}{2}.$$

The result follows by changing variables. \square

Proof of Theorem 2.42 (i). Let $a < b$ be two points of continuity of F , and $X \sim F$. By Fubini's theorem,

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dt &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \mathbb{E}[e^{itX}] dt \\ &= \mathbb{E} \left[\frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} e^{itX} dt \right] \\ &= \mathbb{E} \left[\frac{1}{2\pi} \int_{-T}^T \frac{\sin(tX - at) - \sin(tX - bt)}{t} e^{itX} dt \right] \end{aligned} \quad (2.17)$$

Note that $\int_0^T \frac{\sin t}{t} dt \leq \int_0^\pi \frac{\sin t}{t} dt \leq \pi$ for all $T > 0$, the integrand in (2.17) is bounded by 2. By Lebesgue dominated convergence theorem,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dt &= \mathbb{E} \left[\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{\sin(tX - at) - \sin(tX - bt)}{t} e^{itX} dt \right] \\ &= \frac{1}{2} \mathbb{E} [\operatorname{sgn}(X - a) - \operatorname{sgn}(X - b)] \quad (\text{By Lemma 2.43}) \\ &= \frac{1}{2} (1 - 2F(a) - 1 + 2F(b)) = F(b) - F(a), \end{aligned}$$

where the last row follows from continuity of F at a and b . \square

Remark. In high-dimensional case, a similar conclusion follows: Let μ_F be a distribution measure on $\mathcal{B}(\mathbb{R}^p)$, and let $\phi_F(\lambda) = \int_{\mathbb{R}^p} \exp(i\langle x, \lambda \rangle) d\mu_F(x)$ be the characteristic function of μ_F . Let $A \subset \mathbb{R}^d$ be a cell of the form

$$A = \{(x_1, \dots, x_p) : a_j \leq x_j \leq b_j \text{ for all } j\},$$

where $a_j < b_j$ for all j and $\mu_F(\partial A) = 0$. Then

$$\mu_F(A) = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^p} \int_{[-T, T]^p} \prod_{j=1}^p \left\{ \frac{e^{ia_j t_j} - e^{ib_j t_j}}{it_j} \right\} \phi_F(t) dt, \text{ where } t = (t_1, \dots, t_p) \in \mathbb{R}^p. \quad (2.18)$$

Note that at most countably many hyperplanes perpendicular to the coordinate axes can have positive μ_F measure. As a result, the cells A with $\mu(\partial A) = 0$ form a π -system that generate $\mathcal{B}(\mathbb{R}^p)$. Thus a distribution μ_F is uniquely determined by its characteristic function ϕ_F by (2.18) and Lemma 1.22.

Corollary 2.44 (Independence). Let X and Y be two random variables. Let ϕ_X and ϕ_Y be the characteristic functions of X and Y , respectively, and let $\phi_{X,Y}$ be the characteristic function of (X, Y) . Then X and Y are independent if and only if $\phi_{X,Y}(u, v) = \phi_X(u)\phi_Y(v)$.

Proof. The necessity is clear. We prove sufficiency here. If $\phi_{X,Y}(u, v) = \phi_X(u)\phi_Y(v)$, by inversion formula, $\mu_{X,Y}([a_1, b_1] \times [a_2, b_2]) = \mu_X([a_1, b_1]) \times \mu_Y([a_2, b_2])$ for all continuity rectangles $[a_1, b_1] \times [a_2, b_2]$, which form a π -system that generates $\mathcal{B}(\mathbb{R}^2)$. Then the result follows from Lemma 1.22. \square

We also have the following useful corollary, which allows us to simplify some future proofs by doing only the 1-dimension case.

Lemma 2.45. (Cramér-Wold device). Let X and Y be two p -dimensional random vectors. Then $X \stackrel{d}{=} Y$ if and only if $\langle X, \alpha \rangle \stackrel{d}{=} \langle Y, \alpha \rangle$ for all $\alpha \in \mathbb{R}^p$.

Proof. The necessity is clear. For sufficiency, note that when $\langle X, \alpha \rangle \stackrel{d}{=} \langle Y, \alpha \rangle$ for all $\alpha \in \mathbb{R}^p$, the characteristic functions of X and Y are the same. \square

We can use characteristic functions to investigate the tail properties of distribution functions.

Proposition 2.46. Let ϕ_X be the characteristic function of a random variable X . For each $\epsilon, \delta > 0$, there exists a constant $K > 0$ depending only on δ such that

$$\begin{aligned} \mathbb{P}(|X| \geq \epsilon) &\leq K \int_0^1 \left[1 - \operatorname{Re} \left(\phi_X \left(\frac{t\delta}{\epsilon} \right) \right) \right] dt, \\ \mathbb{E} [X^2 \mathbb{1}_{\{|X| \leq \epsilon\}}] &\leq K \epsilon^2 \left[1 - \operatorname{Re} \left(\phi_X \left(\frac{t\delta}{\epsilon} \right) \right) \right]. \end{aligned}$$

Proof. We redefine that $\frac{\sin 0}{0} = 1$ and $\frac{1 - \cos 0}{0^2} = \frac{1}{2}$, so both $\frac{\sin x}{x}$ and $\frac{1 - \cos x}{x^2}$ become uniformly continuous functions on \mathbb{R} . Then for any $\delta > 0$, there exists $K > 0$ such that

$$1 - \frac{\sin x}{x} \geq \frac{1}{K}, \quad \forall |x| \geq \delta, \quad \text{and} \quad \frac{1 - \cos x}{x^2} \geq \frac{1}{K}, \quad \forall |x| \leq \delta.$$

Let $X \sim F$, and let $\eta = \delta/\epsilon$. Then

$$\begin{aligned} \mathbb{P}(|X| \geq \epsilon) &= \mathbb{E} [\mathbb{1}_{\{|\eta X| \geq \delta\}}] \leq K \mathbb{E} \left[1 - \frac{\sin(\eta X)}{\eta X} \right] \\ &= K - \mathbb{E} \left[K \int_0^1 \cos(\eta t X) dt \right] = K \int_0^1 [1 - \operatorname{Re}(\phi_X(\eta t))] dt, \end{aligned} \quad (\text{By Fubini's theorem})$$

and

$$\mathbb{E} [X^2 \mathbb{1}_{\{|X| \leq \epsilon\}}] = \frac{1}{\eta^2} \mathbb{E} [|\eta X|^2 \mathbb{1}_{\{|\eta X| \leq \delta\}}] \leq \frac{K}{\eta^2} \mathbb{E} [1 - \cos(\eta X)] = K \epsilon^2 [1 - \operatorname{Re}(\phi_X(\eta t))].$$

Thus we complete the proof. \square

2.6 Continuity Theorem and Central Limit Theorems

Definition 2.47 (Quantile). Given a c.d.f. $F : \mathbb{R} \rightarrow [0, 1]$, the *quantile* of distribution F is defined as

$$F^{-1}(p) = \inf \{ \alpha \in \mathbb{R} : F(\alpha) \geq p \}, \quad \forall p \in (0, 1).$$

Remark. The quantile $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ satisfies the following properties:

- (i) $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ is monotone increasing.
- (ii) F^{-1} has at most countably many points of discontinuity. To see this, let E be the set of these points. Then for each $x \in E$, since F^{-1} is monotone, define

$$l_x := \lim_{y \nearrow x} F^{-1}(y) < \lim_{y \searrow x} F^{-1}(y) =: r_x.$$

Since \mathbb{Q} is dense in \mathbb{R} , choose $q_x \in \mathbb{Q} \cap (l_x, r_x)$. Since F^{-1} is monotone increasing, the intervals (l_x, r_x) are pairwise disjoint. Thus we obtain a bijection $x \mapsto q_x$ from E to a subset of \mathbb{Q} . Hence E has at most countably many elements. In fact, this conclusion holds for all monotone functions on \mathbb{R} .

- (iii) F^{-1} is left-continuous. This follows from the right-continuity of F :

$$\{ \alpha \in \mathbb{R} : F(\alpha) \geq p \} = \bigcap_{n=1}^{\infty} \left\{ \alpha \in \mathbb{R} : F(\alpha) \geq p - \frac{1}{n} \right\} \Rightarrow F^{-1}(p) = \lim_{n \rightarrow \infty} F^{-1} \left(p - \frac{1}{n} \right).$$

Lemma 2.48 (Galois inequality). Let $\alpha \in \mathbb{R}$ and $p \in (0, 1)$. Then $F(\alpha) \geq p$ if and only if $F^{-1}(p) \leq \alpha$. Particulièrement, we have $F(F^{-1}(p)) \geq p$ and $F^{-1}(F(\alpha)) \leq \alpha$.

Proof. The “only if” case follows from definition. Conversely, assume $\alpha \geq F^{-1}(p) := \inf \{ z \in \mathbb{R} : F(z) \geq p \}$. Then we have $\alpha + n^{-1} \in \{ z \in \mathbb{R} : F(z) \geq p \}$ for all $n \in \mathbb{N}$. By right-continuity of F ,

$$F(\alpha) = \lim_{n \rightarrow \infty} F \left(\alpha + \frac{1}{n} \right) \geq p.$$

Thus we finish the proof. □

Corollary 2.49 (Quantile transformation). Let $U \sim \text{Unif}(0, 1)$. Then $F^{-1}(U) \sim F$.

Proof. By Galois inequality, for all $x \in \mathbb{R}$, we have $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(F(x) \geq U) = F(x)$. □

Theorem 2.50 (Weak convergence of quantiles). Let F_n be a c.d.f. sequence, and F a c.d.f.. Then $F_n \xrightarrow{w} F$ if and only if $F_n^{-1}(p) \rightarrow F^{-1}(p)$ for each point p of continuity of F^{-1} .

Proof. Assume $F_n \xrightarrow{w} F$, and let $Z \sim N(0, 1)$. Since F is discontinuous at at most countably many points, we have $F_n(Z) \xrightarrow{a.s.} F(Z)$, and $F_n(Z) \xrightarrow{d} F(Z)$. By Portmanteau lemma [Theorem 2.29 (vii)], if the function $p \mapsto \mathbb{P}(F(Z) < p)$ is continuous at $p \in (0, 1)$, we have $\mathbb{P}(F(Z) = p) = 0$, and $\mathbb{P}(F_n(Z) < p) \rightarrow \mathbb{P}(F(Z) < p)$.

Let Φ be the c.d.f. of standard Gaussian variables. By Galois inequality,

$$\Phi(F_n^{-1}(p)) = \mathbb{P}(Z < F_n^{-1}(p)) = \mathbb{P}(F_n(Z) < p) \xrightarrow{n \rightarrow \infty} \mathbb{P}(F(Z) < p) = \Phi(F^{-1}(p))$$

for each point p of continuity of $\Phi \circ F^{-1}$. By continuity of Φ , these contain all points of continuity of F^{-1} . Again, by continuity of Φ^{-1} , we have $F_n^{-1}(p) \rightarrow F^{-1}(p)$ for each point p of continuity of F^{-1} .

Conversely, assume that $F_n^{-1}(p) \rightarrow F^{-1}(p)$ for each point p of continuity of F^{-1} . Let $U \sim \text{Unif}(0, 1)$, then $F_n^{-1}(U) \xrightarrow{a.s.} F^{-1}(U)$, since F^{-1} has at most countably many points of discontinuity. Since $F_n^{-1}(U) \sim F_n$ and $F^{-1}(U) \sim F$, we have $F_n \xrightarrow{w} F$. □

Theorem 2.51 (Skorokhod's almost sure representations). Let $F_n : \mathbb{R} \rightarrow [0, 1]$ be a sequence of c.d.f.'s such that $F_n \xrightarrow{w} F$, where F is also a c.d.f.. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of random variables X_n on it such that $X_n \sim F_n$ for all $n \in \mathbb{N}$, and $X_n \xrightarrow{a.s.} X$, where $X \sim F$.

Proof. We use the quantile transformation. Let $\Omega = [0, 1]$, and let \mathbb{P} be the Lebesgue measure on $[0, 1]$. Define $X(\omega) = F^{-1}(\omega)$ and $X_n(\omega) = F_n^{-1}(\omega)$ for all $n \in \mathbb{N}$. Then $X \sim F$, and $X_n \sim F_n$. By Theorem 2.50, $X_n \rightarrow X$ on Ω except possibly at countably many points of discontinuity of F^{-1} , which form a null set. \square

Corollary 2.52 (Convergence of characteristic functions). Let F_n be a c.d.f. sequence, and let ϕ_{F_n} be the sequence of associated characteristic functions. If $F_n \xrightarrow{w} F$, where F is a c.d.f., then $\phi_{F_n} \rightarrow \phi_F$ pointwise.

Proof. By Skorokhod's representation theorem, we can choose $X_n \sim F_n$ and $X \sim F$ such that $X_n \xrightarrow{a.s.} X$. Then $e^{i\lambda X_n} \xrightarrow{a.s.} e^{i\lambda X}$ for all $\lambda \in \mathbb{R}$. By Lebesgue dominated convergence theorem, $\phi_{F_n} \rightarrow \phi_F$ pointwise. \square

Theorem 2.53 (Lévy's continuity theorem). Let X_n be a sequence of random variables, and let ϕ_n be the sequence of associated characteristic functions. If ϕ_n converges pointwise to a function $\phi : \mathbb{R} \rightarrow \mathbb{C}$, the following are equivalent:

- (i) $\{X_n\}_{n=1}^\infty$ is uniformly tight, i.e. $\lim_{M \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq M) = 0$.
- (ii) $X_n \xrightarrow{d} X$ for some random variable X .
- (iii) ϕ is the characteristic of some random variable X , i.e. $\phi(\lambda) = \mathbb{E}[e^{i\lambda X}]$;
- (iv) ϕ is continuous everywhere on \mathbb{R} ;
- (v) ϕ is continuous at 0.

Proof. (i) \Rightarrow (ii): Let F_n be the c.d.f. of X_n . By Theorem 2.32 (ii), for every subsequence of X_n , we can extract a further subsequence which converges some random variable $X \sim F$. By Corollary 2.52, $\phi_F = \phi$, hence F is uniquely determined by ϕ . We can fix X and conclude that every subsequence X_{n_k} of X_n has a further subsequence that converges in distribution to X .

It remains to show $X_n \xrightarrow{d} X$. If not, choose $f \in C_b(\mathbb{R})$ such that $\mathbb{E}[f(X_n)]$ does not converge to $\mathbb{E}[f(X)]$. Then there exists $\epsilon > 0$ such that for all $k \in \mathbb{N}$ we can find $n_k \geq k$ such that $|\mathbb{E}[f(X_{n_k})] - \mathbb{E}[f(X)]| > \epsilon$. As a result, $\mathbb{E}[f(X_{n_k})]$ has no subsequence converging to $\mathbb{E}[f(X)]$, and X_{n_k} has no subsequence converging in distribution to X , a contradiction! Hence $X_n \xrightarrow{d} X$. (This is called the subsequence trick.)

(ii) \Rightarrow (iii) follows from Corollary 2.52. (iii) \Rightarrow (iv) and (iv) \Rightarrow (v) are trivial.

(v) \Rightarrow (i): Following Proposition 2.46, we set $\delta = 2$ and $K = 5$. The following estimate holds for all $n \in \mathbb{N}$:

$$\mathbb{P}(|X_n| \geq T) \leq 5 \int_0^1 \left[1 - \operatorname{Re} \left(\phi_n \left(\frac{2t}{T} \right) \right) \right] dt.$$

By Lebesgue dominated convergence theorem,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| \geq T) \leq 5 \int_0^1 \left[1 - \operatorname{Re} \left(\phi \left(\frac{2t}{T} \right) \right) \right] dt.$$

Since $\phi(0) = 1$, and ϕ is continuous at 0, the right-hand side of the above estimate converges to 0 as $T \rightarrow \infty$. Given $\epsilon > 0$, we choose T_0 such that $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| \geq T_0) < \epsilon/2$, and choose N such that $\mathbb{P}(|X_n| \geq T_0) < \epsilon$ for all $n \geq N$. Then $\{X_n\}_{n=N}^\infty$ is uniformly tight, and so is $\{X_n\}_{n=1}^\infty$. \square

Remark. We can summarize a commonly used conclusion from Theorem 2.53, which can be viewed as a converse of Corollary 2.52:

Let F_n be a c.d.f. sequence, and let ϕ_{F_n} be the sequence of associated characteristic functions. If $\phi_{F_n} \rightarrow \phi$ pointwise, and ϕ is continuous at 0, then ϕ is the characteristic function of some c.d.f. F , and $F_n \xrightarrow{w} F$.

Theorem 2.54 (Khinchine's weak law of large numbers). Let $(X_n)_{n=1}^\infty$ be a sequence of independent and identically distributed (i.i.d.) random variables such that $\mathbb{E}[|X_1|] < \infty$. Denote $\mu := \mathbb{E}[X_1] < \infty$. Define

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Then $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$.

Proof. Without loss of generality, assume $\mu = 0$. Let $\phi : \mathbb{R} \rightarrow \mathbb{C}$ be the characteristic function of X_1 . Then the characteristic function of \bar{X}_n is

$$\phi_n(\lambda) = \mathbb{E} \left[\prod_{j=1}^n e^{i \frac{\lambda}{n} X_j} \right] = \phi \left(\frac{\lambda}{n} \right)^n.$$

Since $\mathbb{E}[|X_1|] < \infty$, ϕ is differentiable, $\phi'(0) = i \mathbb{E}X_1 = 0$, and ϕ' is uniformly continuous. Fix $\lambda \in \mathbb{R}$. Given any $\epsilon > 0$, we can choose N such that $|\phi'(t)| \leq \epsilon$ for all $|t| \leq |\lambda|/N$. Hence

$$\lim_{n \rightarrow \infty} |\phi_n(\lambda) - 1| = \lim_{n \rightarrow \infty} \left| \phi \left(\frac{\lambda}{n} \right)^n - 1 \right| \leq \lim_{n \rightarrow \infty} \left| \left(1 + \int_0^{\lambda/n} \phi'(t) dt \right)^n - 1 \right| \leq \max \left\{ e^{|\lambda|\epsilon} - 1, 1 - e^{-|\lambda|\epsilon} \right\}.$$

Since $\epsilon > 0$ is arbitrary, we have $\phi_n(\lambda) \rightarrow 1$ pointwise. By Lévy's continuity theorem, $\bar{X}_n \xrightarrow{d} 0$. By Theorem 2.35 (iv), we have $\bar{X}_n \xrightarrow{\mathbb{P}} 0$. \square

Theorem 2.55 (Lindeberg-Lévy central limit theorem). Let $(X_n)_{n=1}^\infty$ be a sequence of i.i.d. random variables such that $\mathbb{E}[|X_1|^2] < \infty$. Denote $\mu := \mathbb{E}X_1 < \infty$, and $0 < \sigma^2 := \text{Var}(X_1) < \infty$. Define

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad Z_n = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu)$$

Then $Z_n \xrightarrow{d} Z$, where $Z \sim N(0, 1)$.

Proof. Without loss of generality, assume that $\mu = 0$ and $\sigma^2 = 1$. Let ϕ be the characteristic function of X_1 . Then the characteristic function of Z_n is

$$\phi_n(\lambda) = \mathbb{E} \left[\prod_{j=1}^n e^{i \frac{\lambda}{\sqrt{n}} X_j} \right] = \phi \left(\frac{\lambda}{\sqrt{n}} \right)^n.$$

Since $\mathbb{E}[|X_1|^2] < \infty$, ϕ is twice-differentiable, $\phi'(0) = i \mathbb{E}X_1 = 0$, and $\phi''(0) = -\mathbb{E}[X_1^2] = -1$.

$$\phi \left(\frac{\lambda}{\sqrt{n}} \right) = 1 + \int_0^{\lambda/\sqrt{n}} \phi'(t) dt = 1 + \int_0^{\lambda/\sqrt{n}} \int_0^t \phi''(u) du dt = 1 - \frac{\lambda^2}{2n} + \int_0^{\lambda/\sqrt{n}} \int_0^t (1 + \phi''(u)) du dt.$$

Note that ϕ'' is uniformly continuous. Given $\epsilon > 0$, choose N such that $|1 + \phi''(u)| < \epsilon$ for all $|u| \leq |\lambda|/\sqrt{N}$. Then for all $n \geq N$, we have

$$1 - \frac{\lambda^2(1+\epsilon)}{2n} \leq \phi \left(\frac{\lambda}{\sqrt{n}} \right) \leq 1 - \frac{\lambda^2(1-\epsilon)}{2n} \Rightarrow e^{-\frac{\lambda^2(1+\epsilon)}{2}} \leq \lim_{n \rightarrow \infty} \phi_n(\lambda) \leq e^{-\frac{\lambda^2(1-\epsilon)}{2}}.$$

Since $\epsilon > 0$ is arbitrary, $\lim_{n \rightarrow \infty} \phi_n(\lambda) = e^{-\lambda^2/2}$, which is the characteristic function of $Z \sim N(0, 1)$. \square

Theorem 2.56 (Lindeberg-Feller central limit theorem). Let $(X_n)_{n=1}^\infty$ be a sequence of independent random variables such that $\mathbb{E}X_n = 0$ and $0 < \sigma_n^2 := \mathbb{E}[X_n^2] < \infty$. Define

$$s_n^2 = \sum_{k=1}^n \sigma_k^2, \quad Z_n = \frac{1}{s_n} \sum_{k=1}^n X_k.$$

Then $Z_n \xrightarrow{d} N(0, 1)$ and

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq k \leq n} \sigma_k^2}{s_n^2} = 0 \quad (2.19)$$

if and only if the following *Lindeberg's condition* is satisfied:

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}] = 0, \quad \forall \epsilon > 0. \quad (2.20)$$

Proof of Sufficiency. Suppose the Lindeberg's condition (2.20) holds. Then for $1 \leq k \leq n$ and all $\epsilon > 0$,

$$\frac{\max_{1 \leq k \leq n} \sigma_k^2}{s_n^2} = \max_{1 \leq k \leq n} \left\{ \frac{1}{s_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{\{|X_k| < \epsilon s_n\}}] + \frac{1}{s_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}] \right\} \leq \epsilon^2 + \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}].$$

Then (2.19) is true follows by letting $n \rightarrow \infty$ and $\epsilon \searrow 0$. Let ϕ_n be the characteristic function of X_n . To prove $Z_n \rightarrow N(0, 1)$, we need to show that the characteristic function of Z_n satisfies

$$\phi_{Z_n}(\lambda) = \prod_{k=1}^n \phi_k\left(\frac{\lambda}{s_n}\right) \rightarrow e^{-\lambda^2/2} \text{ as } n \rightarrow \infty. \quad (2.21)$$

The result then follows from Lévy's continuity theorem. We claim that (2.21) holds if and only if

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \left(\phi_k\left(\frac{\lambda}{s_n}\right) - 1 \right) + \frac{\lambda^2}{2} = 0, \quad (2.22)$$

We first prove the following (2.23), which together with (2.22) implies (2.21).

$$\lim_{n \rightarrow \infty} \left| \exp \left\{ \sum_{k=1}^n \left(\phi_k\left(\frac{\lambda}{s_n}\right) - 1 \right) \right\} - \prod_{k=1}^n \phi_k\left(\frac{\lambda}{s_n}\right) \right| = 0. \quad (2.23)$$

Claim I. If $\phi : \mathbb{R} \rightarrow \mathbb{C}$ is a characteristic function, so is $\lambda \mapsto e^{\phi(\lambda)-1}$.

Let $(Y_n)_{n=1}^\infty$ be a sequence of i.i.d. random variables, and let $N \sim \text{Poisson}(1)$ be a random variable independent of Y_n 's. Define $W = \sum_{k=1}^N Y_k$. Then the characteristic function of W is

$$\mathbb{E}[e^{i\lambda W}] = \mathbb{E}[\mathbb{E}[e^{i\lambda W} | N]] = \mathbb{E}[\phi(\lambda)^N] = \sum_{n=0}^{\infty} \frac{e^{-1}}{n!} \phi(\lambda)^n = e^{\phi(\lambda)-1}.$$

Claim II (Product comparison). Given $\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\} \subset \{z \in \mathbb{C} : |z| \leq 1\}$, it holds

$$\left| \prod_{j=1}^n a_j - \prod_{j=1}^n b_j \right| \leq \sum_{j=1}^n |a_j - b_j|.$$

The case $n = 1$ is clear. Then prove the case $n = 2$:

$$|a_1 b_1 - a_2 b_2| = |a_1(b_1 - b_2) - (a_1 - b_1)b_2| \leq |a_1 - b_1| + |a_2 - b_2|.$$

We apply this formula on $\prod_{j=1}^{n-1} a_j, \prod_{j=1}^{n-1} b_j, a_n, b_n$, so the general case follows from induction.

Proof of (2.23). By Claims I and II,

$$\left| \exp \left\{ \sum_{k=1}^n \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) \right\} - \prod_{k=1}^n \phi_k \left(\frac{\lambda}{s_n} \right) \right| \leq \sum_{k=1}^n \left| \exp \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) - \phi_k \left(\frac{\lambda}{s_n} \right) \right|. \quad (2.24)$$

By Taylor's theorem, we have

$$\left| \phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right| \leq 1 + \frac{\lambda^2}{2s_n^2} \sup_{t \in \mathbb{R}} \phi''(t) - 1 = \frac{\lambda^2 \sigma_k^2}{2s_n^2} \leq \frac{\lambda^2}{2} \max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2}.$$

Given any $0 < \epsilon < 1$, by (2.19), we can choose N such that $|\phi_k(\lambda/s_n) - 1| \leq \epsilon/2$ for all $n \geq N$ and all $1 \leq k \leq n$. Since $|e^z - z - 1| \leq \epsilon|z|$ for all $|z| \leq \epsilon/2$, following (2.24), we have

$$\left| \exp \left\{ \sum_{k=1}^n \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) \right\} - \prod_{k=1}^n \phi_k \left(\frac{\lambda}{s_n} \right) \right| \leq \sum_{k=1}^n \epsilon \left| \phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right| \leq \frac{\epsilon \lambda^2}{2s_n^2} \sum_{k=1}^n \sigma_k^2 = \frac{\epsilon \lambda^2}{2}, \quad \forall n \geq N.$$

Since $\epsilon > 0$ is arbitrary, the limit (2.23) holds.

Proof of (2.22). Given $\epsilon > 0$, we use the following expansion:

$$\begin{aligned} \sum_{k=1}^n \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) + \frac{\lambda^2}{2} &= \sum_{k=1}^n \mathbb{E} \left[\underbrace{e^{i\lambda X_k/s_n} - 1 - \frac{i\lambda}{s_n} X_k - \frac{(i\lambda)^2}{2s_n^2} X_k^2}_{=: A_{n,k}} \right] \\ &= \sum_{k=1}^n \mathbb{E} [A_{n,k} \mathbf{1}_{\{|X_k| < \epsilon s_n\}}] + \sum_{k=1}^n \mathbb{E} [A_{n,k} \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}] =: S_{n,\epsilon}^{(1)} + S_{n,\epsilon}^{(2)}. \end{aligned}$$

Now we bound the two terms. For the first term,

$$\begin{aligned} |S_{n,\epsilon}^{(1)}| &\leq \sum_{k=1}^n \mathbb{E} |A_{n,k} \mathbf{1}_{\{|X_k| < \epsilon s_n\}}| \leq \sum_{k=1}^n \mathbb{E} \left[\frac{1}{3!} \left| \frac{\lambda}{s_n} X_k \right|^3 \mathbf{1}_{\{|X_k| < \epsilon s_n\}} \right] \\ &= \frac{|\lambda|^3}{6s_n^3} \sum_{k=1}^n \mathbb{E} [|X_k|^3 \mathbf{1}_{\{|X_k| < \epsilon s_n\}}] \leq \frac{|\lambda|^3 \epsilon}{6s_n^2} \sum_{k=1}^n \mathbb{E} [|X_k|^2] = \frac{|\lambda|^3 \epsilon}{6}. \end{aligned} \quad (2.25)$$

By Lindeberg's condition, we can bound the second term as $n \rightarrow \infty$:

$$|S_{n,\epsilon}^{(2)}| \leq \sum_{k=1}^n \mathbb{E} |A_{n,k} \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}| \leq \sum_{k=1}^n \mathbb{E} \left[\left| \frac{\lambda}{s_n} X_k \right|^2 \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}} \right] = \frac{|\lambda|^2}{s_n^2} \sum_{k=1}^n \mathbb{E} [X_k^2 \mathbf{1}_{\{|X_k| \geq \epsilon s_n\}}] \rightarrow 0. \quad (2.26)$$

Since $\epsilon > 0$, we can bound (2.22) by arbitrarily small numbers, and the result follows. \square

Remark. In estimates (2.25) and (2.26), we used the following estimate in case $n = 2$:

$$\left| e^{i\theta} - \sum_{k=1}^n \frac{(i\theta)^k}{k!} \right| \leq \min \left\{ \frac{2\theta^n}{n!}, \frac{\theta^{n+1}}{(n+1)!} \right\}, \quad \forall \theta \in \mathbb{R}.$$

The proof of necessity is given by William Feller.

Proof of Necessity. Assume that (2.19) holds and $Z_n \xrightarrow{d} N(0, 1)$. Note the proof of (2.23) only uses (2.19). Then both (2.21) and (2.23) hold, which together imply (2.22). Let $\epsilon > 0$. If $\lambda \neq 0$, we have

$$\begin{aligned}
\frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E} [X_k^2 \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}}] &= 1 - \frac{2}{\lambda^2} \sum_{k=1}^n \mathbb{E} \left[\frac{\lambda^2 X_k^2}{2s_n^2} \mathbb{1}_{\{|X_k| < \epsilon s_n\}} \right] \\
&\leq \frac{2}{\lambda^2} \left(\frac{\lambda^2}{2} + \sum_{k=1}^n \mathbb{E} \left[\left(\cos \left(\frac{\lambda X_k}{s_n} \right) - 1 \right) \mathbb{1}_{\{|X_k| < \epsilon s_n\}} \right] \right) \quad (\text{by } 1 - \cos x \leq \frac{x^2}{2}) \\
&= \frac{2}{\lambda^2} \left(\frac{\lambda^2}{2} + \sum_{k=1}^n \operatorname{Re} \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) + \sum_{k=1}^n \mathbb{E} \left[\left(1 - \cos \left(\frac{\lambda X_k}{s_n} \right) \right) \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}} \right] \right) \\
&\leq \frac{2}{\lambda^2} \left(\frac{\lambda^2}{2} + \sum_{k=1}^n \operatorname{Re} \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) \right) + \frac{4}{\lambda^2} \sum_{k=1}^n \mathbb{P}(|X_k| \geq \epsilon s_n) \\
&\leq \frac{2}{\lambda^2} \left(\frac{\lambda^2}{2} + \sum_{k=1}^n \operatorname{Re} \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) \right) + \frac{4}{\lambda^2} \sum_{k=1}^n \frac{\sigma_k^2}{\epsilon^2 s_n^2} \quad (\text{By Chebyshev's inequality}) \\
&\leq \frac{2}{\lambda^2} \left(\frac{\lambda^2}{2} + \sum_{k=1}^n \operatorname{Re} \left(\phi_k \left(\frac{\lambda}{s_n} \right) - 1 \right) \right) + \frac{4}{\lambda^2 \epsilon^2}.
\end{aligned}$$

By (2.22), the first term converges to 0 as $n \rightarrow \infty$. Since $\lambda \neq 0$ is arbitrary, we obtain the Lindeberg's condition (2.20) by letting $\lambda^2 \rightarrow \infty$. \square

In practice, the Lindeberg's condition is not convenient to verify. In many cases, we would rather use one of its sufficient condition proposed by Lyapunov.

Theorem 2.57 (Lyapunov Condition). Let $(X_n)_{n=1}^\infty$ be a sequence of independent random variables such that $\mathbb{E}X_n = 0$ and $0 < \sigma_n^2 := \mathbb{E}[X_n^2] < \infty$. Let $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$. If there exists $\delta > 0$ satisfying the *Lyapunov condition*:

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}[|X_k|^{2+\delta}] = 0, \quad (2.27)$$

then the Lindeberg's condition (2.20) holds, hence the central limit theorem [Theorem 2.56].

Proof. If there exists $\delta > 0$ that satisfies the Lyapunov condition (2.27), then

$$\begin{aligned}
\frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E} [X_k^2 \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}}] &= \sum_{k=1}^n \mathbb{E} \left[\left| \frac{X_k}{s_n} \right|^2 \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}} \right] \\
&\leq \sum_{k=1}^n \mathbb{E} \left[\left| \frac{X_k}{s_n} \right|^2 \left| \frac{X_k}{\epsilon s_n} \right|^\delta \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}} \right] \leq \frac{1}{\epsilon^\delta s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E} [|X_k|^{2+\delta}] \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

Hence the Lindeberg's condition (2.20) is satisfied. \square

Remark. A useful special case of the Lyapunov condition is when $\delta = 1$:

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^3} \sum_{k=1}^n \mathbb{E} [|X_k|^3] = 0.$$

3 Martingales

3.1 Processes, Filtrations and Stopping Times

Definition 3.1 (Stochastic processes). Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a metric space (E, d) and a nonempty set \mathcal{T} , a *stochastic process* is a function $(t, \omega) \mapsto X_t(\omega)$ defined on the set $\Omega \times \mathcal{T}$ and taking values in E such that $X_t(\cdot)$ is measurable for each $t \in \mathcal{T}$.

Remark. We can also view a stochastic process as a collection $X = \{X_t\}_{t \in \mathcal{T}}$ of E -valued random variables indexed by elements of \mathcal{T} . If \mathcal{T} is a topological space given the discrete topology, we call $\{X_t\}_{t \in \mathcal{T}}$ a *discrete stochastic process*. Furthermore, if $\mathcal{T} = \mathbb{N}_0$, we call the process $\{X_n\}_{n=0}^\infty$ a *stochastic sequence*.

Definition 3.2 (Filtrations). Let \mathcal{T} be \mathbb{N}_0 or \mathbb{R}_+ . A *filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ indexed by elements \mathcal{T} of increasing sub σ -algebras of \mathcal{F} , i.e. $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$.

Remark. We can also define the limit of a filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ by $\mathcal{F}_\infty = \sigma(\bigcup_{t \in \mathcal{T}} \mathcal{F}_t)$. If $\mathcal{T} = \mathbb{N}$, then a filtration $\{\mathcal{F}_n\}_{n=1}^\infty$ satisfies

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots \subset \mathcal{F}_\infty \subset \mathcal{F}.$$

If $\mathcal{T} = [0, \infty)$, then a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ satisfies

$$\mathcal{F}_0 \subset \mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{F}, \quad \forall t > s \geq 0.$$

Definition 3.3 (Adaptation). Let \mathcal{T} be \mathbb{N}_0 or \mathbb{R}_+ , and let $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A stochastic process $\{X_t\}_{t \in \mathcal{T}}$ is said to be *adapted to* $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$, if X_t is \mathcal{F}_t -measurable for each $t \in \mathcal{T}$.

Remark. A stochastic process $\{X_t\}_{t \in \mathcal{T}}$ automatically induces a *canonical filtration*

$$\mathcal{F}_t = \sigma(\{X_s\}_{s \leq t}), \quad t \in \mathcal{T}.$$

It is the minimal sub σ -algebra where every X_s with $s \leq t$ is measurable. We also call this the σ -algebra generated by $\{X_s\}_{s \leq t}$. Clearly, the process $\{X_t\}_{t \in \mathcal{T}}$ is adapted to its canonical filtration.

Definition 3.4 (Stopping time). Let \mathcal{T} be \mathbb{N}_0 or \mathbb{R}_+ . A random variable $\tau : \Omega \rightarrow \bar{\mathcal{T}} := \mathcal{T} \cup \{\infty\}$ is said to be a *stopping time of the filtration* $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathcal{T}$. Without ambiguity, if the filtration is fixed, we say that τ is a *stopping time*.

Remark. If τ is a stopping time, then the set $\{\tau < t\}$ is also \mathcal{F}_t -measurable for all $t \in \mathcal{T}$, since

$$\{\tau < t\} = \bigcup_{n=1}^{\infty} \left\{ \tau \leq t - \frac{1}{n} \right\}.$$

Furthermore,

$$\{\tau = \infty\} = \left(\bigcup_{n=1}^{\infty} \{\tau \leq n\} \right)^c \in \mathcal{F}_\infty.$$

We can also change the definition of stopping time in discrete case. If $\mathcal{T} = \mathbb{N}_0$, then $\tau : \Omega \rightarrow \bar{\mathbb{N}}_0$ is a stopping time of the filtration $\{\mathcal{F}_n\}_{n=1}^\infty$ if and only if $\{\tau = n\} \subset \mathcal{F}_n$ for all $n \in \mathbb{N}_0$, since $\{\tau \leq n\} = \bigcup_{k=0}^n \{\tau = k\}$.

Definition 3.5 (σ -algebra generated by a stopping time). Let \mathcal{T} be \mathbb{N}_0 or \mathbb{R}_+ , and let τ be a stopping time of the filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$. The σ -algebra generated by τ is defined as

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t, \forall t \in \mathcal{T}\}.$$

Remark. We need to check that \mathcal{F}_τ is a σ -algebra. Clearly, $\Omega \in \mathcal{F}_\tau$. If $A \in \mathcal{F}_\tau$, then for all $t \in \mathcal{T}$, we have $A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \cap (A \cap \{\tau \leq t\})^c \in \mathcal{F}_t$, which implies $A^c \in \mathcal{F}_\tau$. Finally, given $\{A_n\}_{n=1}^\infty \subset \mathcal{F}_\tau$, we have

$$\left(\bigcup_{n=1}^\infty A_n\right) \cap \{\tau \leq t\} = \bigcup_{n=1}^\infty (A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t, \forall t \in \mathcal{T}.$$

Then $\bigcup_{n=1}^\infty A_n \in \mathcal{F}_\tau$. Therefore \mathcal{F}_τ is a σ -algebra.

Now we concentrate on the case $\mathcal{T} = \mathbb{R}_+$.

Definition 3.6 (Right-continuity). Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. For every $t \in \mathbb{R}_+$, define

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s, \quad \text{and} \quad \mathcal{F}_{\infty+} = \mathcal{F}_\infty.$$

Then \mathcal{F}_{t+} is a σ -algebra, and the collection $\{\mathcal{F}_{t+}\}_{t \geq 0}$ is also a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \in \mathbb{R}_+$, then the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is said to be *right-continuous*. By construction, the filtration $\{\mathcal{F}_{t+}\}_{t \geq 0}$ is automatically right-continuous.

Definition 3.7. (Completeness). Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$, and let \mathcal{N} be the sets of all $(\mathcal{F}_\infty, \mathbb{P})$ -negligible sets, i.e. $A \in \mathcal{N}$ if there exists $A' \in \mathcal{F}_\infty$ such that $A' \supset A$ and $\mathbb{P}(A') = 0$. The filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is said to be *complete* if $\mathcal{N} \subset \mathcal{F}_0$.

Remark. If $\{\mathcal{F}_t\}_{t \geq 0}$ is not complete, we can complete it by letting $\mathcal{F}'_t = \sigma(\mathcal{F}_t \cup \sigma(\mathcal{N}))$ for every $t \in \mathbb{R}_+$. Apply this completion procedure to the canonical filtration $\mathcal{F}_t = \sigma(\{X_s\}_{s \leq t})$ of a stochastic process $\{X_t\}_{t \geq 0}$, we obtain the *completed canonical filtration* of $\{X_t\}_{t \geq 0}$.

Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a complete filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. By definition, if two random variables $\xi \stackrel{a.s.}{=} \eta$, and ξ is \mathcal{F}_t -measurable, then η is also \mathcal{F}_t -measurable.

Definition 3.8 (Measurability and Progressiveness). A stochastic process $\{X_t\}_{t \geq 0}$ over a metric space (E, d) is said to be *measurable* if the mapping

$$(\omega, t) \mapsto X_t(\omega)$$

defined on $\Omega \times \mathbb{R}_+$ equipped with the product σ -field $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ is measurable. In addition, we fix a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$. If for each $t \geq 0$, the mapping

$$(\omega, s) \mapsto X_s(\omega)$$

defined on $\Omega \times [0, t]$ equipped with the product σ -field $\mathcal{F}_t \otimes \mathcal{B}([0, t])$ is measurable, then the process $\{X_t\}_{t \geq 0}$ is said to be *progressive*.

Remark. By definition, a progressive process $\{X_t\}_{t \geq 0}$ is both adapted and measurable. In later discussion, we fix the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as well as the state space (E, d) .

Proposition 3.9. Let $\{X_t\}_{t \geq 0}$ be an adapted stochastic process. If $\{X_t\}_{t \geq 0}$ is (sample) right-continuous, i.e. for all $\omega \in \Omega$, the mapping $t \mapsto X_t(\omega)$ is right-continuous, then $\{X_t\}_{t \geq 0}$ is progressive. The same conclusion holds if one replaces right-continuous with left-continuous.

Proof. We only prove the case of sample-right-continuity. The case of sample-left-continuity is similar. Fix $t > 0$. For each $n \in \mathbb{N}$, define

$$X_s^{(n)} = X_{\frac{kt}{n}} \text{ if } s \in \left[\frac{(k-1)t}{n}, \frac{kt}{n} \right), \quad k \in \{1, \dots, n\} \quad \text{and} \quad X_t^{(n)} = X_t.$$

The sample-right-continuity of $\{X_t\}_{t \geq 0}$ implies that for all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} X_s^{(n)}(\omega) = X_s(\omega), \quad \forall s \in [0, t].$$

Furthermore, for every Borel set $B \in \mathcal{B}(E)$,

$$\left\{ (\omega, s) \in \Omega \times [0, t] : X_s^{(n)}(\omega) \in B \right\} = \left(\bigcup_{k=1}^n \left\{ X_{\frac{kt}{n}}(\omega) \in B \right\} \times \left[\frac{(k-1)t}{n}, \frac{kt}{n} \right) \right) \cup (\{X_t \in B\} \times \{t\}).$$

This belongs to the product σ -algebra $\mathcal{F}_t \otimes \mathcal{B}([0, t])$. Hence the mapping $(\omega, s) \mapsto X_s^{(n)}(\omega)$ is measurable on $(\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t]))$, and so is the pointwise limit $(\omega, s) \mapsto X_s(\omega)$. \square

Remark. If the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ was complete, we would only require that the sample path $t \mapsto X_t(\omega)$ is left/right-continuous for \mathbb{P} -a.e. $\omega \in \Omega$.

Proposition 3.10. Write $\mathcal{G}_t = \mathcal{F}_{t+}$ for every $t \in [0, \infty]$.

- (i) A random variable $\tau : \Omega \rightarrow [0, \infty]$ is a stopping time of the filtration $\{\mathcal{G}_t\}_{t \geq 0}$ if and only if $\{\tau < t\} \in \mathcal{F}_t$ for all $t > 0$. This is equivalent to the condition that $\tau \wedge t$ is \mathcal{F}_t -measurable for all $t > 0$.
- (ii) Let τ be a stopping time of the filtration $\{\mathcal{G}_t\}_{t \geq 0}$. Then

$$\mathcal{F}_{\tau+} := \{A \in \mathcal{F}_\infty : A \cap \{\tau < t\} \in \mathcal{F}_t, \quad \forall t > 0\} = \mathcal{G}_\tau.$$

Proof. (i) Assume $\{\tau < s\} \in \mathcal{F}_s$ for all $s > 0$, and fix $t \geq 0$. Then for all $s > t$,

$$\{\tau \leq t\} = \bigcap_{n \in \mathbb{N}: t+n^{-1} < s} \left\{ \tau < t + \frac{1}{n} \right\} \in \mathcal{F}_s \Rightarrow \{\tau \leq t\} \in \bigcup_{s > t} \mathcal{F}_s = \mathcal{G}_t.$$

Conversely, if τ is a stopping time of $\{\mathcal{G}_t\}_{t \geq 0}$, then for all $t > 0$, we have

$$\{\tau < t\} = \bigcup_{n \in \mathbb{N}: t-n^{-1} > 0} \underbrace{\left\{ \tau \leq t - \frac{1}{n} \right\}}_{\in \mathcal{G}_{t-n^{-1}} \subset \mathcal{F}_t} \in \mathcal{F}_t.$$

If $\tau \wedge t$ is \mathcal{F}_t -measurable, we have $\{\tau \leq s\} \in \mathcal{F}_t$ for all $s < t$. Then we have $\{\tau < t\} \in \mathcal{F}_t$ by taking $s_n \nearrow t$ and $\{\tau < t\} = \bigcup_{n=1}^\infty \{\tau \leq s_n\}$. Conversely, if τ is a stopping time of $\{\mathcal{G}_n\}_{n \geq 0}$, we have $\{\tau \leq s\} \in \mathcal{G}_s \subset \mathcal{F}_t$ for all $s < t$, and $\tau \wedge t$ is thus \mathcal{F}_t -measurable.

- (ii) By definition, $\mathcal{G}_\tau := \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{G}_t, \quad \forall t \geq 0\}$. If $A \cap \{\tau < t\} \in \mathcal{F}_t$ for all $t > 0$, then

$$A \cap \{\tau \leq t\} = \bigcap_{n \in \mathbb{N}: t+n^{-1} < s} \left(A \cap \left\{ \tau < t + \frac{1}{n} \right\} \right) \in \mathcal{F}_s, \quad \forall s > t \geq 0 \Rightarrow A \cap \{\tau \leq t\} \in \mathcal{G}_t, \quad \forall t \geq 0.$$

Conversely, if $A \in \mathcal{G}_\tau$, we have

$$A \cap \{\tau < t\} = \bigcup_{n \in \mathbb{N}: t-n^{-1} > 0} \underbrace{\left(A \cap \left\{ \tau \leq t - \frac{1}{n} \right\} \right)}_{\in \mathcal{G}_{t-n^{-1}} \subset \mathcal{F}_t} \in \mathcal{F}_t, \forall t > 0.$$

Hence the conclusion follows. \square

Proposition 3.11 (Properties of stopping times). Let τ, σ be two stopping times of the filtration $\{\mathcal{F}_t\}_{t \geq 0}$.

- (i) τ is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$, and $\mathcal{F}_\tau \subset \mathcal{F}_{\tau+}$. If $\{\mathcal{F}_t\}_{t \geq 0}$ is right-continuous, we have $\mathcal{F}_\tau = \mathcal{F}_{\tau+}$.
- (ii) If $\tau = t$ is a constant stopping time, then $\mathcal{F}_\tau = \mathcal{F}_t$, and $\mathcal{F}_{\tau+} = \mathcal{F}_{t+}$.
- (iii) τ is \mathcal{F}_τ -measurable.
- (iv) Given $A \in \mathcal{F}_\infty$, define

$$\tau^A(\omega) = \begin{cases} \tau(\omega), & \omega \in A, \\ \infty, & \omega \notin A. \end{cases}$$

Then $\tau^A \in \mathcal{F}_\tau$ if and only if τ^A is a stopping time.

- (v) If $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$, and $\mathcal{F}_{\sigma+} \subset \mathcal{F}_{\tau+}$.
- (vi) All $\sigma \wedge \tau$, $\sigma \vee \tau$ and $\sigma + \tau$ are stopping times, and $\{\sigma \leq \tau\}, \{\sigma = \tau\} \in \mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.
- (vii) A function $\omega \mapsto Y(\omega)$ defined on $\{\tau < \infty\}$ is \mathcal{F}_τ -measurable if and only if for each $t \geq 0$, the restriction of Y to the set $\{\tau \leq t\}$ is \mathcal{F}_t -measurable.
- (viii) If $(\tau_n)_{n=1}^\infty$ is a monotone sequence of increasing stopping times, then $\tau_\infty = \lim_{n \rightarrow \infty} \tau_n$ is a stopping time.
- (ix) If $(\tau_n)_{n=1}^\infty$ is a monotone sequence of decreasing stopping times, then $\tau_\infty = \lim_{n \rightarrow \infty} \tau_n$ is a stopping time of the filtration $\{\mathcal{F}_{t+}\}_{t \geq 0}$, and

$$\mathcal{F}_{\tau_\infty+} = \bigcap_{n=1}^\infty \mathcal{F}_{\tau_n+}.$$

In addition, if $(\tau_n)_{n=1}^\infty$ is stationary, i.e. for each $\omega \in \Omega$, there exists $N_\omega \in \mathbb{N}$ such that $\tau(\omega) = \tau_n(\omega)$ for all $n \geq N_\omega$, then τ is a stopping time, and

$$\mathcal{F}_{\tau_\infty} = \bigcap_{n=1}^\infty \mathcal{F}_{\tau_n}.$$

Proof. (i) By Remark of Definition 3.4 and Proposition 3.10, τ is also a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$. The statement $\mathcal{F}_\tau \subset \mathcal{F}_{\tau+}$ follows from $\mathcal{F}_t \subset \mathcal{G}_t$. (ii) immediately follows from definition.

(iii) For all $\alpha \in \mathbb{R}$, we have

$$\{\tau \leq \alpha\} \cap \{\tau \leq t\} = \{\tau \leq \alpha \wedge t\} \in \mathcal{F}_t, \forall t \geq 0 \Rightarrow \{\tau \leq \alpha\} \in \mathcal{F}_\tau.$$

(iv) The result immediately follows from the definition of \mathcal{F}_τ , since

$$\{\tau^A \leq t\} = A \cap \{\tau \leq t\}, \forall t \geq 0.$$

(v) If $A \in \mathcal{F}_\sigma$, then $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Since $\sigma \leq \tau$, we have $\{\sigma \leq t\} \supset \{\tau \leq t\}$, and

$$A \cap \{\tau \leq t\} = (A \cap \{\sigma \leq t\}) \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

(vi) For all $t \geq 0$, we have

$$\begin{aligned} \{\sigma \wedge \tau \leq t\} &= \{\sigma \leq t\} \cup \{\tau \leq t\} \in \mathcal{F}_t, \quad \{\sigma \vee \tau \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t, \\ \text{and } \{\sigma + \tau > t\} &= \{\tau \geq t\} \cup \left(\bigcup_{q \in \mathbb{Q} \cap (0, t]} \{t - q \leq \tau < t\} \cap \{\sigma > q\} \right) \in \mathcal{F}_t. \end{aligned}$$

Hence $\sigma \wedge \tau$, $\sigma \vee \tau$ and $\sigma + \tau$ are stopping times. By (v), $\mathcal{F}_{\sigma \wedge \tau} \subset \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Conversely, if $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$,

$$A \cap \{\sigma \wedge \tau \leq t\} = \underbrace{(A \cap \{\sigma \leq t\})}_{\in \mathcal{F}_t} \cup \underbrace{(A \cap \{\tau \leq t\})}_{\in \mathcal{F}_t} \in \mathcal{F}_t, \quad \forall t \geq 0 \Rightarrow A \in \mathcal{F}_{\sigma \wedge \tau}.$$

Hence $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. By Proposition 3.10 (i), for all $t \geq 0$, both $\sigma \wedge t$ and $\tau \wedge t$ are \mathcal{F}_t -measurable, and

$$\begin{aligned} \{\sigma \leq \tau\} \cap \{\sigma \leq t\} &= \{\sigma \leq t\} \cap \{\sigma \wedge t \leq \tau \wedge t\} \in \mathcal{F}_t \Rightarrow \{\sigma \leq \tau\} \in \mathcal{F}_\sigma, \\ \{\sigma \leq \tau\} \cap \{\tau \leq t\} &= \{\sigma \leq t\} \cap \{\tau \leq t\} \cap \{\sigma \wedge t \leq \tau \wedge t\} \in \mathcal{F}_t \Rightarrow \{\sigma \leq \tau\} \in \mathcal{F}_\tau. \end{aligned}$$

Then $\{\sigma \leq \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau = \mathcal{F}_{\sigma \wedge \tau}$, and $\{\sigma = \tau\} = \{\sigma \leq \tau\} \cap \{\sigma \geq \tau\} \in \mathcal{F}_{\sigma \wedge \tau}$.

(vii) We first assume that for each $t \geq 0$, the restriction $Y|_{\{\tau \leq t\}}$ is \mathcal{F}_t -measurable. Then for every Borel set $B \in \mathcal{B}(E)$, we have $\{Y \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$. Since Y is defined on $\{\tau < \infty\}$, we have

$$\{Y \in B\} = \bigcup_{n=1}^{\infty} (\{Y \in B\} \cap \{\tau \leq n\}) \in \mathcal{F}_\infty.$$

Hence $\{Y \in B\} \in \mathcal{F}_\tau$. Conversely, if Y is \mathcal{F}_τ -measurable, then $\{Y \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

(viii) For every $t \geq 0$, $\{\tau_\infty \leq t\} = \bigcap_{n=1}^{\infty} \{\tau_n \leq t\} \in \mathcal{F}_t$.

(ix) For every $t \geq 0$, $\{\tau_\infty < t\} = \bigcup_{n=1}^{\infty} \{\tau_n < t\} \in \mathcal{F}_t$. Hence τ is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$ by Proposition 3.10 (i). By (v), we have $\mathcal{F}_{\tau+} \subset \mathcal{F}_{\tau_n+}$ for each $n \in \mathbb{N}$. Conversely, if $A \in \mathcal{F}_{\tau_n+}$ for each $n \in \mathbb{N}$,

$$A \cap \{\tau_\infty < t\} = \bigcup_{n=1}^{\infty} (A \cap \{\tau_n < t\}) \in \mathcal{F}_t, \quad \forall t > 0 \Rightarrow A \in \mathcal{F}_{\tau_\infty+}.$$

Hence $\mathcal{F}_{\tau_\infty+} = \bigcap_{n=1}^{\infty} \mathcal{F}_{\tau_n+}$. Furthermore, if τ_n is stationary, then $\{\tau_\infty \leq t\} = \bigcup_{n=1}^{\infty} \{\tau_n \leq t\} \in \mathcal{F}_t$. Thus τ_∞ is a stopping time, and $\mathcal{F}_{\tau_\infty} \subset \mathcal{F}_{\tau_n}$ for each $n \in \mathbb{N}$ by (v). Conversely, if $A \in \mathcal{F}_{\tau_n}$ for each $n \in \mathbb{N}$,

$$A \cap \{\tau_\infty \leq t\} = \bigcup_{n=1}^{\infty} (A \cap \{\tau_n \leq t\}) \in \mathcal{F}_t, \quad \forall t > 0 \Rightarrow A \in \mathcal{F}_{\tau_\infty}.$$

Hence $\bigcap_{n=1}^{\infty} \mathcal{F}_{\tau_n} = \mathcal{F}_{\tau_\infty}$. □

Proposition 3.12. Let $X = \{X_t\}_{t \geq 0}$ be a progressive process of $\{\mathcal{F}_t\}_{t \geq 0}$. If T is a stopping time of $\{\mathcal{F}_t\}_{t \geq 0}$, then the function $X_\tau : \omega \mapsto X_{\tau(\omega)}(\omega)$, defined on the set $\{\tau < \infty\}$, is \mathcal{F}_τ -measurable.

Proof. By Proposition 3.11 (vii), it suffices to show that the restriction of X_τ to $\{\tau \leq t\}$ is \mathcal{F}_t -measurable for all $t \geq 0$. The restriction $X_\tau|_{\{\tau \leq t\}}$ is a composition of two measurable mappings:

$$\begin{aligned} \tau \wedge t \text{ is } \mathcal{F}_t\text{-measurable} : (\{\tau \leq t\}, \mathcal{F}_t) &\rightarrow (\{F \leq t\} \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t])), \quad \omega \mapsto (\omega, \tau(\omega) \wedge t), \\ X \text{ is progressive} : (\{\tau \leq t\} \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t])) &\rightarrow (E, \mathcal{B}(E)), \quad (\omega, s) \mapsto X_s(\omega). \end{aligned}$$

Hence $X_\tau|_{\{\tau \leq t\}}$ is \mathcal{F}_t -measurable, and the result follows. □

Remark. We also consider the discrete case. Let $(X_n)_{n=0}^\infty$ be an adapted sequence. Then for all $B \in \mathcal{B}(E)$, we have $\{X_\tau \in B\} \cap \{\tau \leq t\} = \bigcup_{n=0}^t \{X_n \in B\} \cap \{\tau = n\} \in \mathcal{F}_t$ for every $t \in \mathbb{N}$. Then we have $\{X_\tau \in B\} \in \mathcal{F}_\tau$, and X_τ is always \mathcal{F}_τ -measurable.

Proposition 3.13 (Hitting time/début). Let $\{X_t\}_{t \geq 0}$ be a adapted process taking values in (E, d) , and let $A \subset E$ be a measurable subset of E . The *hitting time (or début) of A* is defined as

$$\tau_A = \inf \{t \geq 0 : X_t \in A\}.$$

Given a random time $\sigma : \Omega \rightarrow [0, \infty]$, the *first hitting time of A after σ* is defined as

$$\tau_A^\sigma = \inf \{t > \sigma : X_t \in A\}.$$

Note that we set $\inf \emptyset = \infty$. Then

- (i) If $\{X_t\}_{t \geq 0}$ is sample-right-continuous and $G \subset E$ is an open set, then τ_G is a stopping time of the filtration $\{\mathcal{F}_{t+}\}_{t \geq 0}$. Furthermore, if σ is a stopping time of the filtration $\{\mathcal{F}_{t+}\}_{t \geq 0}$, so is τ_G^σ .
- (ii) If $\{X_t\}_{t \geq 0}$ is sample-continuous and $F \subset E$ is a closed set, then τ_F is a stopping time (of the filtration $\{\mathcal{F}_t\}_{t \geq 0}$). Furthermore, if σ is a stopping time of the filtration $\{\mathcal{F}_{t+}\}_{t \geq 0}$, so is τ_F^σ .

Proof. (i) Fix $t > 0$. If $\tau_G(\omega) < t$, then there exists $\tau_G(\omega) < s < t$ such that $X_s(\omega) \in G$. Since G is open, and $t \mapsto X_t(\omega)$ is right-continuous, we can choose a rational $q \in (s, t)$ such that $X_q(\omega) \in G$. Hence

$$\{\tau_G < t\} = \bigcup_{q \in \mathbb{Q} \cap [0, t)} \{X_q \in G\} \in \mathcal{F}_t, \quad \forall t > 0.$$

Then by Proposition 3.10 (i), τ_G is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$. Furthermore, if σ is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$, we have

$$\begin{aligned} \{\tau_G^\sigma < t\} &= \bigcup_{q \in (0, t)} (\{\sigma < q\} \cap \{\inf\{s \geq q : X_s \in G\} < t\}) \\ &= \bigcup_{q \in (0, t)} \left(\{\sigma < q\} \cap \left(\bigcup_{r \in \mathbb{Q} \cap [q, t)} \{X_r \in G\} \right) \right) \in \mathcal{F}_t, \quad \forall t > 0. \end{aligned}$$

(ii) Fix $t \geq 0$. If $\tau_G(\omega) \leq t$, choose $s_n \searrow s := \tau_G(\omega)$ such that $X_{s_n} \in F$. Since $t \mapsto X_t(\omega)$ is continuous and F is closed, we have $X_{s_n}(\omega) \rightarrow X_s(\omega) \in F$. Hence

$$\{\tau_F \leq t\} = \bigcup_{s \in [0, t]} \{X_s \in F\} = \left\{ \inf_{q \in \mathbb{Q} \cap [0, t]} d(X_q, F) = 0 \right\} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

where the second equality holds because $d(\cdot, F)$ is continuous, and the inclusion holds because $d(\cdot, F)$ is Borel-measurable and X_q is \mathcal{F}_t -measurable for all $q \in \mathbb{Q} \cap [0, t]$, and countable infimum preserves measurability. Furthermore, if σ is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$, we have

$$\begin{aligned} \{\tau_F^\sigma < t\} &= \bigcup_{q \in (0, t)} (\{\sigma < q\} \cap \{\inf\{s \geq q : X_s \in F\} < t\}) \\ &= \bigcup_{q \in (0, t)} \left(\{\sigma < q\} \cap \left\{ \inf_{r \in \mathbb{Q} \cap [q, t]} d(X_r, F) = 0 \right\} \right) \in \mathcal{F}_t, \quad \forall t > 0. \end{aligned}$$

Therefore τ_F^σ is a stopping time of $\{\mathcal{F}_{t+}\}_{t \geq 0}$. □

Remark. Similarly, given any Borel set $A \in \mathcal{B}(\mathbb{R})$, we can define the hitting time of A associated to a discrete process $X = \{X_t\}_{t=0}^\infty$:

$$\tau_A = \min\{n \in \mathbb{N}_0 : X_n \in A\}, \quad \tau_A^\sigma = \min\{n \in \mathbb{N}_0 : n > \sigma, X_n \in A\}.$$

It is easy to show that τ_A is a stopping time for any measurable set A , since $\{\tau_A \leq n\} = \bigcup_{k=0}^n \{X_k \in A\}$. Furthermore, if σ is a stopping time, then the first hitting time τ_A^σ after σ is also a stopping time:

$$\{\tau_A^\sigma \leq n\} = \bigcup_{k=0}^{n-1} (\{\sigma = k\} \cap \{k < \tau_A^\sigma \leq n\}) = \bigcup_{k=0}^{n-1} \left(\{\sigma = k\} \cap \left(\bigcup_{j=k+1}^n \{X_j \in A\} \right) \right) \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}_0.$$

Finally we introduce a technical lemma about stopping times which resembles the form of simple function approximation.

Proposition 3.14. Let τ be a stopping time. If $\sigma : \Omega \rightarrow [0, \infty]$ is a \mathcal{F}_τ -measurable random variable such that $\sigma \geq \tau$. Then σ is also a stopping time. Furthermore,

$$\tau_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{\{k2^{-n} < \tau \leq (k+1)2^{-n}\}} + \infty \mathbb{1}_{\{\tau = \infty\}}, \quad n \in \mathbb{N}$$

is a sequence of stopping times decreasing to τ .

Proof. Since σ is \mathcal{F}_τ -measurable, we have $\{\sigma \leq t\} \in \mathcal{F}_\tau$, and $\{\sigma \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Hence σ is also a stopping time.

Note that $\tau_n(\omega) = \inf\{k2^{-n} : k2^{-n} \geq \tau(\omega), k \in \mathbb{Z}\}$. Then we have $\tau_n \searrow \tau$. Since τ_n is a measurable function of τ , it is \mathcal{F}_τ -measurable, hence a stopping time by the first assertion. \square

3.2 Discrete-time Martingales

Definition 3.15 (Discrete-time martingales). Let $(X_n)_{n=0}^\infty$ be a real-valued and L^1 process that is adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^\infty$. Here L^1 means $\mathbb{E}|X_n| < \infty$ for all $n \geq 0$. Then

- (i) $(X_n)_{n=0}^\infty$ is said to be a *martingale* if $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ for all $n > m \geq 0$;
- (ii) $(X_n)_{n=0}^\infty$ is said to be a *supermartingale* if $\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m$ for all $n > m \geq 0$;
- (iii) $(X_n)_{n=0}^\infty$ is said to be a *submartingale* if $\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m$ for all $n > m \geq 0$;

All these notations depends on the choice of the filtration $\{\mathcal{F}_n\}_{n=0}^\infty$, which is fixed in later discussion.

Remark. (i) If $(X_n)_{n=0}^\infty$ is a submartingale, then $(-X_n)_{n=0}^\infty$ is a supermartingale.

(ii) The set of all martingales in $(L^1(\Omega, \mathcal{F}, \mathbb{P}))^{\mathbb{N}_0}$ is a vector space.

Proposition 3.16. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function such that $\mathbb{E}[f(X_n)] < \infty$ for all $n \in \mathbb{N}_0$.

- (i) If $(X_n)_{n=0}^\infty$ is a martingale, then $(f(X_n))_{n=0}^\infty$ is a submartingale.
- (ii) If f is monotone increasing and $(X_n)_{n=0}^\infty$ is a submartingale, then $(f(X_n))_{n=0}^\infty$ is a submartingale.

Remark. This is an immediate corollary of conditional Jensen's inequality [Proposition 2.17 (v)]. Here is another useful corollary:

- (i) If $(X_n)_{n=0}^\infty$ is a martingale, then $(|X_n|)_{n=0}^\infty$ is a submartingale;
- (ii) If $(X_n)_{n=0}^\infty$ is a submartingale, then $(X_n^+)_{n=0}^\infty$ is a submartingale.

Definition 3.17 (Predictable processes). A discrete process $(H_n)_{n=0}^\infty$ is said to be *predictable* if H_0 is a constant and H_n is \mathcal{F}_{n-1} -measurable for all $n \geq 1$. We define the integral of $(H_n)_{n=0}^\infty$ with respect to a adapted process $(X_n)_{n=0}^\infty$ by

$$(H \cdot X)_0 = H_0 X_0, \quad (H \cdot X)_n = (H \cdot X)_{n-1} + H_n(X_n - X_{n-1}) = H_0 X_0 + \sum_{k=1}^n H_k(X_k - X_{k-1}), \quad \forall n \in \mathbb{N}.$$

Clearly, $(H \cdot X)_n$ is an adapted process. If X_n is a martingale, so is $(H \cdot X)_n$. Moreover, if X_n is a submartingale (or supermartingale) and H_n is nonnegative, then $(H \cdot X)_n$ is a submartingale (or supermartingale).

Theorem 3.18 (Doob's decomposition theorem). Let $(X_n)_{n=1}^\infty$ be a submartingale. Then there exists an increasing predictable L^1 process $(A_n)_{n=0}^\infty$ starting with $A_0 = 0$ and a martingale $(M_n)_{n=0}^\infty$ such that $X_n = M_n + A_n$ for each $n \geq 0$, and the decomposition is unique.

Proof. We first prove the existence. Define $M_0 = X_0$, $A_0 = 0$ and

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

Then $(M_n)_{n=1}^\infty$ and $(A_n)_{n=1}^\infty$ are the desired processes. To prove uniqueness, let $X_n = M'_n + A'_n$ be another decomposition satisfying the conditions given. Then $Y_n = M_n - M'_n = A_n - A'_n$ is a martingale and a predictable L^1 sequence, which implies $Y_n = \mathbb{E}[Y_n|\mathcal{F}_{n-1}] = Y_{n-1} = \dots = Y_0 = 0$. Hence $Y_n \equiv 0$, and the Doob's decomposition is unique. \square

Remark. We have a similar conclusion for supermartingales: If $(X_n)_{n=1}^\infty$ is a submartingale, then there exists an decreasing predictable L^1 process $(A_n)_{n=0}^\infty$ starting with $A_0 = 0$ and a martingale $(M_n)_{n=0}^\infty$ such that $X_n = M_n + A_n$ for each $n \geq 0$, and the decomposition is unique.

Theorem 3.19. (Doob's optional stopping theorem for discrete-time submartingales). Let $(X_n)_{n=1}^\infty$ be a submartingale, and let τ be a bounded stopping time. Then

- (i) $\mathbb{E}[X_\tau] \geq \mathbb{E}[X_0]$;
- (ii) If τ is bounded by $N \in \mathbb{N}$, then $\mathbb{E}[X_N | \mathcal{F}_\tau] \geq X_\tau$;
- (iii) If σ is another bounded stopping time and $\sigma \leq \tau$, then $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma$.

Proof. (i) By definition, we have $\{\tau \geq k\} = \{\tau \leq k-1\}^c \in \mathcal{F}_{k-1}$ for all $k \in \mathbb{N}$. Then

$$\begin{aligned} \mathbb{E}[X_\tau] &= \mathbb{E}\left[X_0 + \sum_{k=1}^N (X_k - X_{k-1}) \mathbf{1}_{\{\tau \geq k\}}\right] = \mathbb{E}[X_0] + \sum_{k=1}^N \mathbb{E}[(X_k - X_{k-1}) \mathbf{1}_{\{\tau \geq k\}}] \\ &= \mathbb{E}[X_0] + \sum_{k=1}^N \underbrace{\mathbb{E}[(X_k - X_{k-1}) | \mathcal{F}_{k-1}]}_{\geq 0} \mathbb{E}[\mathbf{1}_{\{\tau \geq k\}}] \geq \mathbb{E}[X_0]. \end{aligned}$$

(ii) If $A \in \mathcal{F}_\tau$, we have $A \cap \{\tau = n\} \in \mathcal{F}_n$ for all $N \in \mathbb{N}_0$, and

$$\begin{aligned} \mathbb{E}[X_N \mathbf{1}_A] &= \sum_{n=0}^N \mathbb{E}[X_N \mathbf{1}_{A \cap \{\tau=n\}}] = \sum_{n=0}^N \mathbb{E}[\mathbb{E}[X_N | \mathcal{F}_n] \mathbf{1}_{A \cap \{\tau=n\}}] \\ &\geq \sum_{n=0}^N \mathbb{E}[X_n \mathbf{1}_{A \cap \{\tau=n\}}] = \sum_{n=0}^N \mathbb{E}[X_\tau \mathbf{1}_{A \cap \{\tau=n\}}] = \mathbb{E}[X_\tau \mathbf{1}_A]. \end{aligned}$$

Since X_τ is \mathcal{F}_τ -measurable, we have $\mathbb{E}[X_N | \mathcal{F}_\tau] \geq X_\tau$.

Remark. Similarly, if $(X_n)_{n=1}^\infty$ is a supermartingale, then we have $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_0]$ and $X_\tau \geq \mathbb{E}[X_N | \mathcal{F}_\tau]$. Furthermore, if $(X_n)_{n=1}^\infty$ is a martingale, then we have $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ and $X_\tau = \mathbb{E}[X_N | \mathcal{F}_\tau]$.

(iii) Since $\sigma \leq \tau \leq N$, we have $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$. We use Doob's decomposition $X_t = M_t + A_t$ of submartingale, where M_t is a martingale and A_t is an increasing predictable sequence. By (ii),

$$M_\sigma = \mathbb{E}[M_N | \mathcal{F}_\sigma] = \mathbb{E}[\mathbb{E}[M_N | \mathcal{F}_\tau] | \mathcal{F}_\sigma] = \mathbb{E}[M_\tau | \mathcal{F}_\sigma].$$

Clearly, $A_\tau \geq A_\sigma$, and $\mathbb{E}[A_\tau | \mathcal{F}_\sigma] \geq \mathbb{E}[A_\sigma | \mathcal{F}_\sigma] = A_\sigma$. Hence $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = \mathbb{E}[M_\tau + A_\tau | \mathcal{F}_\sigma] \geq M_\sigma + A_\sigma = X_\sigma$. \square

The following theorem gives a necessary and sufficient condition of an adapted integrable sequence to be a martingale.

Theorem 3.20. Let $(X_n)_{n=0}^\infty$ be an adapted and L^1 sequence. Then $(X_n)_{n=0}^\infty$ is a martingale if and only if $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ for every bounded stopping time τ .

Proof. “ \Rightarrow ”: By Theorem 3.19 (i).

“ \Leftarrow ”: Assume that $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ for every bounded stopping time τ . To prove that $(X_n)_{n=0}^\infty$ is a martingale, we show that for each $n \in \mathbb{N}$,

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] = X_{n-1} \Leftrightarrow \mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}[X_{n-1} \mathbf{1}_A], \quad \forall A \in \mathcal{F}_{n-1}.$$

Let $\tau = (n-1)\mathbf{1}_A + n\mathbf{1}_{A^c}$, so τ is a bounded stopping time, and $\mathbb{E}[X_0] = \mathbb{E}[X_\tau] = \mathbb{E}[X_{n-1} \mathbf{1}_A] + \mathbb{E}[X_n \mathbf{1}_{A^c}]$. Since n is a constant stopping time, we have $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. Then $\mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}[X_{n-1} \mathbf{1}_A]$. \square

Proposition 3.21 (Maximal inequality). If $(X_n)_{n=0}^\infty$ is a submartingale, then for every $n \in \mathbb{N}$,

$$\lambda \mathbb{P} \left(\max_{0 \leq k \leq n} |X_k| > \lambda \right) \leq \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_n|], \quad \forall \lambda > 0. \quad (3.1)$$

In addition, if $(X_n)_{n=0}^\infty$ is nonnegative, then

$$\lambda \mathbb{P}(X_n^* > \lambda) \leq \mathbb{E}[X_n \mathbf{1}_{\{X_n^* > \lambda\}}] \leq \mathbb{E}[X_n], \quad \forall \lambda > 0, \quad (3.2)$$

where $X_n^* = \max_{0 \leq k \leq n} X_k$.

Proof. Let $\tau = n \wedge \min\{n \in \mathbb{N}_0 : |X_n| > \lambda\}$. Then τ is a bounded stopping time. By Markov's inequality,

$$\lambda \mathbb{P} \left(\max_{0 \leq k \leq n} |X_k| > \lambda \right) = \lambda \mathbb{P}(|X_\tau| > \lambda) \leq \mathbb{E}[|X_\tau| \mathbf{1}_{\{|X_\tau| > \lambda\}}] \leq \mathbb{E}[|X_\tau|]. \quad (3.3)$$

Note that both (X_n) and (X_n^+) are submartingales, and $|X_\tau| \leq 2X_\tau^+ - X_\tau$. By Theorem 3.19,

$$\lambda \mathbb{P} \left(\max_{0 \leq k \leq n} |X_k| > \lambda \right) \leq 2\mathbb{E}[X_\tau^+] - \mathbb{E}[X_\tau] \leq 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0] \leq \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_n|].$$

In addition, if $(X_n)_{n=0}^\infty$ is nonnegative, (3.2) immediately follows from (3.3) and Theorem 3.19. \square

Remark. Note that the inequality (3.1) holds for both submartingales and supermartingales. If $(X_n)_{n=0}^\infty$ is a martingale, then $(|X_n|)_{n=0}^\infty$ is a nonnegative submartingale, and we can use (3.2):

$$\lambda \mathbb{P} \left(\max_{0 \leq k \leq n} |X_k| > \lambda \right) \leq \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n|^* > \lambda\}}] \leq \mathbb{E}[|X_n|], \quad \forall \lambda > 0.$$

Proposition 3.22 (Doob's L^p -inequality). If $(X_n)_{n=0}^\infty$ is a nonnegative submartingale or a martingale, and $1 < p < \infty$, then

$$\mathbb{E} \left[\max_{0 \leq k \leq n} |X_k|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|X_n|^p], \quad \forall n \in \mathbb{N}. \quad (3.4)$$

Proof. We use a corollary of Fubini's theorem: for $p > 0$ and a nonnegative random variable $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}[X^p] = \int_0^\infty p\lambda^{p-1} \mathbb{P}(X > \lambda) d\lambda.$$

Without loss of generality, let $(X_n)_{n=0}^\infty$ be a nonnegative submartingale, and $Y := \max_{0 \leq k \leq n} X_k$. Then

$$\begin{aligned} \mathbb{E}[Y^p] &= \int_0^\infty p\lambda^{p-1} \mathbb{P}(Y > \lambda) d\lambda \leq \int_0^\infty p\lambda^{p-2} \mathbb{E}[X_n \mathbf{1}_{\{\lambda < Y\}}] d\lambda && \text{(By Proposition 3.21)} \\ &= \mathbb{E} \left[X_n \int_0^\infty p\lambda^{p-2} \mathbf{1}_{\{\lambda < Y\}} d\lambda \right] = \left(\frac{p}{p-1} \right) \mathbb{E}[X_n Y^{p-1}]. \end{aligned} \quad (3.5)$$

Note that $q = p/(p-1)$ is the Hölder's conjugate of p , we have

$$\mathbb{E}[X_n Y^{p-1}] \leq (\mathbb{E}[X_n^p])^{1/p} \left(\mathbb{E}[Y^{(p-1)q}] \right)^{1/q} = \|X_n\|_p (\mathbb{E}[Y^p])^{1-1/p}. \quad (3.6)$$

Combining (3.5) and (3.6), we have $\|Y\|_p \leq \frac{p}{p-1} \|X_n\|_p$, which is (3.4). \square

Now we discuss the convergence of martingales.

Definition 3.23 (Upcrossing number). Given a real sequence $(x_n)_{n=0}^\infty$ and $a < b$, the *upcrossing number* of this sequence along $[a, b]$ before time n , denoted by $U_{[a,b]}^x(n)$, is the largest integer k such that there exists a strictly increasing sequence

$$0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq n$$

of integers such that $x_{s_j} \leq a$ and $x_{t_j} \geq b$ for all $j \in \{1, \dots, k\}$. The *total upcrossing number* $U_{[a,b]}^x(\infty)$ of this sequence along $[a, b]$ is defined as the limit of monotone increasing sequence $U_{[a,b]}^x(n)$, which possibly takes ∞ .

Lemma 3.24. A real sequence $(x_n)_{n=0}^\infty$ converges if and only if $U_{[a,b]}^x(\infty) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$.

Proof. Assume $x_\infty = \lim_{n \rightarrow \infty} x_n$ (possibly infinite) exists. Then for all $a, b \in \mathbb{Q}$ with $a < b$, we have either $a < x_\infty$ or $b > x_\infty$. For the first case, there exists $N_a \in \mathbb{N}$ such that $x_n > a$ for all $n \geq N_a$, which implies $U_{[a,b]}^x(\infty) \leq N_a/2 < \infty$. The second case is similar.

Conversely, if the sequence has no limit, let $a := \liminf_{n \rightarrow \infty} x_n < b := \limsup_{n \rightarrow \infty} x_n$, and pick $p, q \in \mathbb{Q}$ such that $a < p < q < b$. Now we choose two sequences $\{s_j\}_{j=1}^\infty$ and $\{t_j\}_{j=1}^\infty$ recursively as follows. First we pick any $s_1 \in \mathbb{N}_0$ such that $x_{s_1} < p$, then pick $t_1 > s_1$ such that $x_{t_1} > q$. This is possible, because $\limsup_{n \rightarrow \infty} x_n > q$. Repeated this procedure, we choose $s_2 > t_1$ such that $x_{s_2} < p$, and $t_2 > s_2$ such that $x_{t_2} > q, \dots$. By definition of upcrossing numbers, we have $U_{[p,q]}^x(\infty) = \infty$, a contradiction! \square

Definition 3.25 (Upcrossing number). Given an adapted sequence $(X_n)_{n=0}^\infty$, the associated *upcrossing number* is define as $U_{[a,b]}^X(n) : \omega \mapsto U_{[a,b]}^{X(\omega)}(n)$, where $n \in \bar{\mathbb{N}}$. It can be depicted by a sequence of stopping times. Let $\tau_0 = -\infty$, and define

$$\sigma_j = \min \{k \in \mathbb{N}_0 : k > \tau_{j-1}, X_k \leq a\}, \quad \tau_j = \min \{k \in \mathbb{N}_0 : k > \sigma_j, X_k \geq b\}, \quad j \geq 1.$$

Then the upcrossing number of $(x_n)_{n=0}^\infty$ along $[a, b]$ before time $n \in \mathbb{N}$ is $U_{[a,b]}^X(n) = \max\{j : \tau_j \leq n\}$, and $U_{[a,b]}^X(\infty) = \lim_{n \rightarrow \infty} U_{[a,b]}^X(n) = \max\{j : \tau_j < \infty\} = \min\{j : \tau_j = \infty\}$.

Remark. The upcrossing number $U_{[a,b]}^X(n)$ is also a stopping time, since $\{U_{[a,b]}^X(n) \leq k\} = \{\tau_k \geq n\} \in \mathcal{F}_k$.

Proposition 3.26 (Doob's upcrossing inequality). If $(X_n)_{n=0}^\infty$ is a submartingale, then for all real numbers $a < b$ and all $n \in \mathbb{N}$,

$$\mathbb{E} \left[U_{[a,b]}^X(n) \right] \leq \frac{\mathbb{E}[(X_n - a)^+ - (X_0 - a)^+]}{b - a}. \quad (3.7)$$

Proof. Define stopping times $\{\sigma_j\}_{j=1}^\infty$ and $\{\tau_j\}_{j=1}^\infty$ as in Definition 3.25. Then $\{U_{[a,b]}^X(n) \geq k\} = \{\tau_k \leq n\}$, and

$$H_k = \begin{cases} 1, & \text{if } \sigma_j < k \leq \tau_j \text{ for some } j \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

defines a nonnegative predictable process, since $\{H_k = 1\} = \left(\bigcup_{j=0}^{k-1} \{X_j \leq a\} \right) \cap \{X_{k-1} < b\}$.

Define $Y_n = (X_n - a)^+$. Then Y_n is a nonnegative submartingale, and we have $U_{[a,b]}^X(n) = U_{[0,b-a]}^Y(n)$. By definition, $(H \cdot Y)_n$ satisfies

$$(H \cdot Y)_n = 0 + \sum_{k=1}^n H_n(Y_n - Y_{n-1}) \geq (b - a)U_{[0,b-a]}^Y(n) = (b - a)U_{[a,b]}^X(n)$$

Note that $((1 - H) \cdot Y)_n$ is also a submartingale. Then

$$(b - a)\mathbb{E} \left[U_{[a,b]}^X(n) \right] \leq \mathbb{E}[(H \cdot Y)_n] = \mathbb{E}[Y_n - ((1 - H) \cdot Y)_n] \leq \mathbb{E}[Y_n] - \mathbb{E}[(1 - H_0)Y_0] = \mathbb{E}[Y_n] - \mathbb{E}[Y_0].$$

This is indeed the inequality (3.7). \square

Following is the first martingale convergence theorem of Doob's.

Theorem 3.27 (Doob's convergence theorem for discrete-time submartingales). If $(X_n)_{n=0}^\infty$ is a submartingale, and $\sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+] < \infty$, then $X_\infty = \lim_{n \rightarrow \infty} X_n$ a.s. exists, and $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Let $M := \sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+] < \infty$. Then for all $a \in \mathbb{R}$,

$$(x - a)^+ \leq x^+ + a^-, \quad \forall x \in \mathbb{R} \Rightarrow \mathbb{E}[(X_n - a)^+] \leq \mathbb{E}[X_n] + a^- \leq M + a^-, \quad \forall n \in \mathbb{N}.$$

By Fatou's lemma and Proposition 3.26, the total upcrossing number $U_{[a,b]}^X(n) \nearrow U_{[a,b]}^X(\infty)$ satisfies

$$\mathbb{E} \left[U_{[a,b]}^X(\infty) \right] \leq \liminf_{n \rightarrow \infty} \mathbb{E} \left[U_{[a,b]}^X(n) \right] \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[(X_n - a)^+]}{b - a} \leq \frac{M + |a|}{b - a} < \infty, \quad \forall \mathbb{R} \ni b > a.$$

Then for all real numbers $a < b$, we have

$$\mathbb{P} \left(U_{[a,b]}^X(\infty) < \infty \right) = 1.$$

Let $\Omega_0 = \bigcap_{a,b \in \mathbb{Q}: a < b} \left\{ U_{[a,b]}^X(\infty) < \infty \right\}$, so $\mathbb{P}(C) = 1$. By Lemma 3.24, $X_n(\omega)$ converges for each $\omega \in \Omega_0$, and we denote the converging point by $X_\infty(\omega)$. For $\omega \notin \Omega_0$, define $X_\infty(\omega) = 0$. Then $X_n \rightarrow X_\infty$ a.s..

Now we prove that $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. By Fatou's lemma, $\mathbb{E}[X_\infty^+] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^+] \leq M$, and $\mathbb{E}[X_\infty^-] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^-] = \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^+ - X_n] \leq M - \mathbb{E}[X_0] < \infty$. \square

Remark. Theorem 3.27 does not imply L^1 -convergence. As a counter example, consider the random walk $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$, where $\{\xi_k\}_{k=1}^\infty$ are i.i.d. Rademacher variables. Fix \mathcal{F}_n to be the canonical filtration of $(X_n)_{n=0}^\infty$, and define $\tau = \min\{n \in \mathbb{N} : X_n = 1\}$, which is a stopping time. Then the stopped process $Y_n = X_{n \wedge \tau}$ is a submartingale. Since $\mathbb{E}[Y_n^+] \leq 1$, the sequence Y_n converges a.s.. Furthermore, $Y_n \rightarrow Y_\infty = 1$ a.s., because $Y_{n+1} = Y_n \pm 1$ once $Y_n < 1$. On the other hand, $\mathbb{E}[Y_n] = \mathbb{E}[X_{n \wedge \tau}] = \mathbb{E}[X_0] = 0$ for all $n \in \mathbb{N}_0$.

Theorem 3.28 (Doob's convergence theorem for uniformly integrable discrete-time martingales). Let $(X_n)_{n=0}^\infty$ be a martingale. The following are equivalent:

- (i) The collection $\{X_n\}_{n \in \mathbb{N}_0}$ is uniformly integrable.
- (ii) X_n converges a.s. and in L^1 -norm.
- (iii) $(X_n)_{n=0}^\infty$ is closed, i.e. there exists $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n = \mathbb{E}[Z | \mathcal{F}_n]$ for all $n \in \mathbb{N}_0$.

Proof. (i) \Rightarrow (ii): By Theorem 1.72, we have $\sup_{n \in \mathbb{N}_0} \mathbb{E}[|X_n|] < \infty$. By Doob's convergence theorem for discrete-time submartingales [Theorem 3.27], there exists $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \rightarrow X_\infty$ a.s., which implies $X_n \xrightarrow{\mathbb{P}} X_\infty$. Since X_n is uniformly integrable, by Theorem 1.73, $\|X_n - X_\infty\|_1 \rightarrow 0$.

(ii) \Rightarrow (iii): Fix $n \in \mathbb{N}_0$. Since $\mathbb{E}[\cdot | \mathcal{F}_s]$ is a bounded linear operator on $L^1(\Omega, \mathcal{F}, \mathbb{R})$, we have

$$\mathbb{E}[X_\infty | \mathcal{F}_n] = \lim_{m \rightarrow \infty} \mathbb{E}[X_m | \mathcal{F}_n] = X_n.$$

(iii) \Rightarrow (i) is an immediate corollary of the following STRONGER CLAIM. \square

Claim 3.28*. Given $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, the following collection is uniformly integrable:

$$\{\mathbb{E}[Z|\mathcal{G}] : \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

Proof. Since Z is integrable, by Theorem 1.57, for every $\epsilon > 0$, there exists $\delta > 0$ such that $\mathbb{E}[|Z|\mathbb{1}_A] < \epsilon$ for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$. Given $M > 0$, and define $X_{\mathcal{G}} = \mathbb{E}[Z|\mathcal{G}]$, $Y_{\mathcal{G}} = \mathbb{E}[|Z||\mathcal{G}]$. Then $|X_{\mathcal{G}}| \leq Y_{\mathcal{G}}$, and

$$\mathbb{E}[|X_{\mathcal{G}}|\mathbb{1}_{\{|X_{\mathcal{G}}|>M\}}] \leq \mathbb{E}[Y_{\mathcal{G}}\mathbb{1}_{\{Y_{\mathcal{G}}>M\}}] = \mathbb{E}[|Z|\mathbb{1}_{\{Y_{\mathcal{G}}>M\}}].$$

By Chebyshev inequality, $\mu(Y_{\mathcal{G}} > M) \leq \frac{1}{M} \mathbb{E}[Y_{\mathcal{G}}] = \frac{1}{M} \mathbb{E}[|Z|]$. Once $M > \mathbb{E}[|Z|]/\delta$, we have $\mathbb{E}[|Z|\mathbb{1}_{\{Y_{\mathcal{G}}>M\}}] < \epsilon$. Note the choice of M is independent of \mathcal{G} . Since $\epsilon > 0$, we have

$$0 \leq \lim_{M \rightarrow \infty} \sup_{\mathcal{G} \subset \mathcal{F}} \mathbb{E}[|X_{\mathcal{G}}|\mathbb{1}_{\{|X_{\mathcal{G}}|>M\}}] \leq \lim_{M \rightarrow \infty} \sup_{\mathcal{G} \subset \mathcal{F}} \mathbb{E}[|Z|\mathbb{1}_{\{Y_{\mathcal{G}}>M\}}] = 0.$$

Hence the given random variable collection is uniformly integrable. \square

Remark. Given $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, the martingale $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ described in Theorem 3.28 (iii) is also called a *Lévy's martingale* or *Doob's process*. In this case, we also have $X_{\infty} = \mathbb{E}[Z|\mathcal{F}_{\infty}]$ a.s..

This is easy to prove. By our proof of (ii) \Rightarrow (iii), we have $\mathbb{E}[X_{\infty}|\mathcal{F}_n] = X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for all $n \in \mathbb{N}_0$. Hence for all $A \in \bigcup_{n=0}^{\infty} \mathcal{F}_n$, we have $\mathbb{E}[X_{\infty}\mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A]$. Note that $\bigcup_{n=0}^{\infty} \mathcal{F}_n$ is a π -system that generates \mathcal{F}_{∞} . By Lemma 1.22, we have $\mathbb{E}[X_{\infty}\mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A]$ for all $A \in \mathcal{F}_{\infty}$. Also, if $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$, and Z is an integrable random variable, then

$$\mathbb{E}[Z|\mathcal{F}_n] \rightarrow \mathbb{E}[Z|\mathcal{F}_{\infty}] \text{ a.s. and in } L^1.$$

We have a stronger version of martingale convergence theorem in L^p when $p > 1$.

Theorem 3.29 (Convergence theorem for L^p -bounded discrete-time martingales). If $(X_n)_{n=0}^{\infty}$ is a martingale such that $\sup_{n \in \mathbb{N}_0} \mathbb{E}[|X_n|^p] < \infty$, where $p > 1$, then X_n converges a.s. and in L^p -norm.

Proof. Let $Y = \sup_{n \in \mathbb{N}_0} |X_n|$ and $M = \sup_{n \in \mathbb{N}_0} \mathbb{E}[|X_n|^p]$. By Theorem 3.27, there exists $X_{\infty} \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \rightarrow X_{\infty}$ a.s.. By Doob's L^p -inequality [Proposition 3.22] and monotone convergence theorem,

$$\mathbb{E}\left[\max_{1 \leq k \leq n} |X_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p] \leq \left(\frac{p}{p-1}\right)^p M \Rightarrow \mathbb{E}[Y^p] \leq \left(\frac{p}{p-1}\right)^p M < \infty.$$

Since $|X_n - X_{\infty}| \leq 2Y$, by Lebesgue dominated convergence theorem, $\|X_n - X_{\infty}\|_p \rightarrow 0$. \square

Definition 3.30 (Stopped processes). Let $(X_n)_{n=1}^{\infty}$ be an adapted process, and let τ be a stopping time. The *stopped process* $(X_n^{\tau})_{n=0}^{\infty}$ is defined by

$$X_n^{\tau}(\omega) = X_{t \wedge \tau(\omega)}(\omega), \quad \forall \omega \in \Omega.$$

Remark. Clearly, the stopped process $(X_n^{\tau})_{n=0}^{\infty}$ is also adapted, and

$$X_n^{\tau} = \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau=k\}} + X_n \mathbb{1}_{\{\tau \geq n\}} \Rightarrow \begin{cases} \mathbb{E}[|X_n^{\tau}|] \leq \sum_{k=0}^n \mathbb{E}[|X_k|], \\ \mathbb{E}[X_n^{\tau}|\mathcal{F}_{n-1}] = \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau=k\}} + \mathbb{E}[X_n|\mathcal{F}_{n-1}]\mathbb{1}_{\{\tau \geq n\}}. \end{cases}$$

It is seen that if X_n is a martingale (or submartingale or supermartingale, respectively), so is X_n^{τ} .

Theorem 3.31 (Optional stopping theorem). Let $(X_n)_{n=1}^\infty$ be a martingale, and let τ be a stopping time. If at least one of the following conditions holds:

- (i) τ is a bounded stopping time;
- (ii) $\mathbb{E}[\tau] < \infty$, and there exists $c > 0$ such that $\mathbb{E}[|X_{n+1} - X_n| | \mathcal{F}_n] \leq c$ for all $n \in \mathbb{N}_0$; or
- (iii) The stopped process X_n^τ is uniformly bounded, and $\tau < \infty$ a.s.;

Then X_τ is a.s. well-defined, and $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

Proof. Note that (i) is proved in Theorem 3.19 (i). To prove (ii) and (iii), we use the following fact:

$$\mathbb{E}[X_n^\tau] = \mathbb{E}[X_{n \wedge \tau}] = \mathbb{E}[X_0], \quad \forall n \in \mathbb{N}_0,$$

which holds because $n \wedge \tau$ is a bounded stopping time. We write the stopped process as

$$X_n^\tau = X_0 + \sum_{k=1}^{n \wedge \tau} (X_k - X_{k-1}), \quad \forall n \in \mathbb{N}_0.$$

If (ii) holds, we have $|X_n^\tau| \leq |X_0| + c\tau$ for all n , and $\mathbb{E}[|X_0|] + c\mathbb{E}[\tau] < \infty$. If (iii) holds, choose $M > 0$ such that $|X_n^\tau| \leq M$ for all n . Hence under either (ii) or (iii), the sequence X_n^τ is dominated by a integrable function, and $X_n^\tau \rightarrow X_\tau$ a.s.. By Lebesgue dominated convergence theorem, we have $\mathbb{E}[X_n^\tau] \rightarrow \mathbb{E}[X_\tau] = \mathbb{E}[X_0]$. \square

Theorem 3.32 (Doob's optional stopping theorem for uniformly integrable discrete-time martingales). Let $(X_n)_{n=1}^\infty$ be a uniformly integrable martingale, and $X_\infty = \lim_{n \rightarrow \infty} X_n$ a.s. and in L^1 . If τ is a stopping time, then $X_\tau \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and

$$X_\tau = \mathbb{E}[X_\infty | \mathcal{F}_\tau]$$

with the convention that $X_\tau = X_\infty$ on $\{\tau = \infty\}$. If $\sigma \leq \tau$ is another stopping time, then $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$.

Proof. By Theorem 3.28, we have $\mathbb{E}[X_\infty | \mathcal{F}_n] = X_n$. Then for all $A \in \mathcal{F}_\tau$, $A \cap \{\tau = n\} \in \mathcal{F}_n$, and

$$\begin{aligned} \mathbb{E}[X_\infty \mathbf{1}_A] &= \sum_{n=0}^{\infty} \mathbb{E}[X_\infty \mathbf{1}_{A \cap \{\tau=n\}}] + \mathbb{E}[X_\infty \mathbf{1}_{A \cap \{\tau=\infty\}}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[X_n \mathbf{1}_{A \cap \{\tau=n\}}] + \mathbb{E}[X_\infty \mathbf{1}_{A \cap \{\tau=\infty\}}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[X_\tau \mathbf{1}_{A \cap \{\tau=n\}}] + \mathbb{E}[X_\tau \mathbf{1}_{A \cap \{\tau=\infty\}}] = \mathbb{E}[X_\tau \mathbf{1}_A]. \end{aligned}$$

Since X_τ is \mathcal{F}_τ -measurable, we have $\mathbb{E}[X_\infty | \mathcal{F}_\tau] = X_\tau$. Furthermore, if $\sigma \leq \tau$ is another stopping time, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$, and $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_\tau] | \mathcal{F}_\sigma] = \mathbb{E}[X_\infty | \mathcal{F}_\sigma] = X_\sigma$. \square

Definition 3.33 (Backward martingales). A *backward filtration* is an increasing sequence of sub σ -algebras $\{\mathcal{F}_n\}_{n \in -\mathbb{N}_0}$ indexed by nonpositive integers, i.e. $\mathcal{F}_n \supset \mathcal{F}_{n-1}$ for all $n \in -\mathbb{N}_0$. Let $(Y_n)_{n \in -\mathbb{N}_0}$ be an adapted sequence of integrable variables indexed by nonpositive integers.

- (i) $(Y_n)_{n \in -\mathbb{N}_0}$ is said to be a *backward martingale*, if $\mathbb{E}[X_n | \mathcal{F}_m] = X_m$ for all $m < n \leq 0$.
- (ii) $(Y_n)_{n \in -\mathbb{N}_0}$ is said to be a *backward submartingale*, if $\mathbb{E}[X_n | \mathcal{F}_m] \geq X_m$ for all $m < n \leq 0$.
- (iii) $(Y_n)_{n \in -\mathbb{N}_0}$ is said to be a *backward supermartingale*, if $\mathbb{E}[X_n | \mathcal{F}_m] \leq X_m$ for all $m < n \leq 0$.

Remark. Likewise, we define the limit of the backward filtration $\{\mathcal{F}_n\}_{n \in -\mathbb{N}_0}$ by $\mathcal{F}_{-\infty} = \bigcap_{n \in -\mathbb{N}_0} \mathcal{F}_n$.

Theorem 3.34 (Doob's convergence theorem for backward discrete-time submartingales). If $(X_n)_{n \in -\mathbb{N}_0}$ is a backward submartingale, and $\lim_{n \rightarrow -\infty} \mathbb{E}[X_n] > -\infty$, then $\{X_n\}_{n \in \mathbb{N}_0}$ is uniformly integrable. Furthermore, there exists $X_{-\infty} \in L^1(\Omega, \mathcal{F}_{-\infty}, \mathbb{P})$ such that $X_{-\infty} = \lim_{n \rightarrow -\infty} X_n$ a.s. and in L^1 . Moreover, we have $\mathbb{E}[X_n | \mathcal{F}_{-\infty}] \geq X_{-\infty}$ for all $n \in -\mathbb{N}_0$.

Proof. Recall Doob's upcrossing inequality Proposition 3.26. Since the sequence $\{X_n, X_{n+1}, \dots, X_0, X_0, \dots\}$ is also a submartingale for all $n \in -\mathbb{N}_0$, we have

$$\mathbb{E}[U_{[a,b]}^X(n)] \leq \frac{\mathbb{E}[(X_0 - a)^+ - (X_n - a)^+]}{b - a} \leq \frac{\mathbb{E}[|X_0|] + |a|}{b - a}, \quad \forall a < b.$$

Since $U_{[a,b]}^X(n) \nearrow U_{[a,b]}^X(-\infty)$, by monotone convergence theorem, we have $\mathbb{P}(U_{[a,b]}^X(-\infty) < \infty) = 1$, and

$$\mathbb{P}\left(\bigcup_{a,b \in \mathbb{Q}: a < b} \{U_{[a,b]}^X(-\infty) < \infty\}\right) = 1$$

A symmetric version of Lemma 3.24 implies that X_n converges a.s. when $n \rightarrow -\infty$. Since every sequence $(X_k)_{k \leq n}$ is \mathcal{F}_n -measurable, the limit $X_{-\infty}$ is \mathcal{F}_n -measurable for each $n \in -\mathbb{N}_0$, hence $\mathcal{F}_{-\infty}$ -measurable. Furthermore, since $\mathbb{E}[X_n] \searrow L := \lim_{n \rightarrow -\infty} \mathbb{E}[X_n] > -\infty$, by Lebesgue dominated convergence theorem

$$\mathbb{E}[|X_n|] = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq 2\mathbb{E}[X_0^+] - L < \infty, \quad \forall n \in -\mathbb{N}_0 \Rightarrow \mathbb{E}[X_{-\infty}] = \lim_{n \rightarrow -\infty} \mathbb{E}[X_n] = L.$$

Given $\epsilon > 0$, we can choose $N \in -\mathbb{N}_0$ such that $\mathbb{E}[X_n] > \mathbb{E}[X_N] - \epsilon/4$ for all $n \leq N$. Then

$$\begin{aligned} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > M\}}] &= \mathbb{E}[X_n \mathbf{1}_{\{X_n > M\}}] - \mathbb{E}[X_n] + \mathbb{E}[X_n \mathbf{1}_{\{X_n \geq -M\}}] \\ &< \mathbb{E}[\mathbb{E}[X_N | \mathcal{F}_n] \mathbf{1}_{\{X_n > M\}}] - \mathbb{E}[X_N] + \frac{\epsilon}{4} + \mathbb{E}[\mathbb{E}[X_N | \mathcal{F}_n] \mathbf{1}_{\{X_n \geq -M\}}] \\ &\leq \mathbb{E}[X_N \mathbf{1}_{\{|X_n| > M\}}] + \frac{\epsilon}{4} \\ &\leq \mathbb{E}[|X_N| \mathbf{1}_{\{|X_n| > M, |X_N| > M/2\}}] + \mathbb{E}[|X_N| \mathbf{1}_{\{|X_n| > M, |X_N| \leq M/2\}}] + \frac{\epsilon}{4} \\ &\leq \mathbb{E}[|X_N| \mathbf{1}_{\{|X_N| > M/2\}}] + \frac{1}{2} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > M\}}] + \frac{\epsilon}{4}, \quad \forall n \leq N. \end{aligned}$$

Hence $\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > M\}}] < \mathbb{E}[|X_N| \mathbf{1}_{\{|X_N| > M/2\}}] + \epsilon/2$ for all $n \leq N$. Choose $M > 0$ so that the first term is less than $\epsilon/2$. Hence $\{X_n\}_{n \leq N}$ is uniformly integrable, and so is $\{X_n\}_{n \in -\mathbb{N}_0}$. Therefore $\|X_n - X_{-\infty}\|_1 \rightarrow 0$. Moreover, for all $A \in \mathcal{F}_{-\infty} \subset \bigcap_{n \in -\mathbb{N}_0} \mathcal{F}_n$,

$$\mathbb{E}[X_n \mathbf{1}_A] - \mathbb{E}[X_m \mathbf{1}_A] = \mathbb{E}[X_n \mathbf{1}_A] - \mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_m] \mathbf{1}_A] \geq \mathbb{E}[X_n \mathbf{1}_A] - \mathbb{E}[X_n \mathbf{1}_A] \geq 0, \quad \forall m < n.$$

Let $m \rightarrow -\infty$. By Lebesgue dominated convergence theorem, $\mathbb{E}[X_n \mathbf{1}_A] \geq \mathbb{E}[X_{-\infty} \mathbf{1}_A]$ for all $n \in -\mathbb{N}_0$. Hence we have $\mathbb{E}[X_n | \mathcal{F}_{-\infty}] \geq X_{-\infty}$. \square

Remark. If $(X_n)_{n \in -\mathbb{N}_0}$ is a backward martingale, the requirement $\lim_{n \rightarrow -\infty} \mathbb{E}[X_n] = \mathbb{E}[X_0] > -\infty$ is automatically satisfied. Then we have $X_n \rightarrow X_{-\infty}$ a.s. and in L^1 . Moreover, $\mathbb{E}[X_0 | \mathcal{F}_{-\infty}] = X_{-\infty}$.

We also have an immediate corollary of this theorem: If $\mathcal{F}_n \searrow \mathcal{F}_{-\infty}$ is a backward filtration, and Z is a integrable random variable, then $\mathbb{E}[Z | \mathcal{F}_n] \rightarrow \mathbb{E}[Z | \mathcal{F}_{-\infty}]$ a.s. and in L^1 .

The convergence theorem of backward martingales also gives rise to the strong law of large numbers.

Lemma 3.35 (Kolmogorov's 0-1 law). Let $(X_n)_{n=1}^\infty$ be a sequence of independent random variables. Define the tail σ -algebra $\mathcal{F}_{-\infty}$ as follows:

$$\mathcal{F}_{-n} = \sigma(X_{n+1}, X_{n+2}, \dots), \quad \forall n \in \mathbb{N}_0 \quad \text{and} \quad \mathcal{F}_{-\infty} = \bigcap_{n \in \mathbb{N}_0} \mathcal{F}_{-n}.$$

Then $\mathcal{F}_{-\infty}$ is \mathbb{P} -trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{F}_{-\infty}$.

Proof. Define $\mathcal{F}_0 = \{\Omega, \emptyset\}$, and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ for all $n \in \mathbb{N}$. Then $\mathcal{F}_n \perp \mathcal{F}_{-n}$, and $\mathcal{F}_{-\infty} \subset \mathcal{F}_{-n}$ is independent of \mathcal{F}_n for all $n \in \mathbb{N}_0$. Let $\mathcal{F}_\infty = \sigma(\bigcup_{n=0}^\infty \mathcal{F}_n)$, then $\mathcal{F}_{-\infty}$ is independent of \mathcal{F}_∞ . Meanwhile, $\mathcal{F}_{-\infty} \subset \mathcal{F}_\infty$ implies that $\mathcal{F}_{-\infty}$ is independent of itself, and the result follows. \square

Remark. We give a brief interpretation of the independence between

$$\mathcal{F}_n = \sigma\left(\bigcup_{k=0}^n \sigma(X_k)\right) \quad \text{and} \quad \mathcal{F}_{-n} = \sigma\left(\bigcup_{k=n+1}^\infty \sigma(X_k)\right).$$

Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra independent of every member of collection $\{\mathcal{F}_\alpha \subset \mathcal{F}\}_{\alpha \in J}$ of σ -algebras. Then \mathcal{G} is also independent of $\sigma(\bigcup_{\alpha \in J} \mathcal{F}_\alpha)$, which is the smallest σ -algebra containing every member of $\{\mathcal{F}_\alpha\}_{\alpha \in J}$. This is easy to prove: For every $A \in \mathcal{G}$, the class $\{B \in \mathcal{F} : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$ is a λ -system that contains $\bigcup_{\alpha \in J} \mathcal{F}_\alpha$, which is a π -system. The result then follows from Sierpiński-Dynkin π - λ theorem.

Now let $\{\mathcal{F}_\alpha\}_{\alpha \in J}$ and $\{\mathcal{G}_\beta\}_{\beta \in I}$ be two collections of σ -algebras such that $\mathcal{F}_\alpha \perp \mathcal{G}_\beta$ for all α and β . By applying the above conclusion twice, we have $\sigma(\bigcup_{\alpha \in J} \mathcal{F}_\alpha) \perp \sigma(\bigcup_{\beta \in I} \mathcal{G}_\beta)$.

Theorem 3.36 (Kolmogorov's strong law of large numbers). Let $(X_n)_{n=1}^\infty$ be a sequence of i.i.d. random variables such that $\mathbb{E}[|X_1|] < \infty$. Then

$$\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j \rightarrow \mu := \mathbb{E}[X_1] \quad a.s..$$

Proof. Let $S_n = X_1 + \dots + X_n$, and $Y_{-n} = \bar{X}_n = S_n/n$. Let $\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \dots) \searrow \mathcal{F}_{-\infty}$. Then

$$\begin{aligned} \mathbb{E}[Y_{-n} | \mathcal{F}_{-n-1}] &= \frac{1}{n} \mathbb{E}[S_{n+1} - X_{n+1} | \sigma(S_{n+1}, X_{n+2}, \dots)] \\ &= \frac{S_{n+1}}{n} - \frac{1}{n} \mathbb{E}[X_{n+1} | \sigma(S_{n+1}, X_{n+2}, \dots)] = \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = Y_{-n-1}, \end{aligned}$$

where the third equality holds because of the exchangeability of X_1, \dots, X_n conditioning on S_{n+1} . Then $(Y_{-n})_{n \in \mathbb{N}}$ is a backward martingale. By Theorem 3.34, $Y_{-n} \rightarrow Y_{-\infty}$ a.s. and in L^1 , with $Y_{-\infty} = \mathbb{E}[X_1 | \mathcal{F}_{-\infty}]$.

Let $\mathcal{G}_{-k} = (X_{k+1}, X_{k+2}, \dots)$. For all $k \in \mathbb{N}$, we have $S_k/n \rightarrow 0$ a.s. as $n \rightarrow \infty$. As a result, $(Y_{-n} - \frac{S_k}{n})_{n=k}^\infty$ is a \mathcal{G}_{-k} -measurable sequence that converges to $Y_{-\infty}$ a.s.. Then we conclude that $Y_{-\infty}$ is a \mathcal{G}_{-k} -measurable function by redefining it on a negligible set if required, and $Y_{-\infty}$ is measurable on the tail σ -algebra $\mathcal{G}_{-\infty}$. By Lemma 3.35, we have $\mathbb{P}(Y_{-\infty} > r) \in \{0, 1\}$ for all $r \in \mathbb{R}$. Let $\mu = \inf\{r : \mathbb{P}(Y_{-\infty} > r) = 0\}$. Then

$$\mathbb{P}(Y_{-\infty} > \mu) = \mathbb{P}\left(\bigcup_{n=1}^\infty \left\{Y_{-\infty} > \mu + \frac{1}{n}\right\}\right) = 0, \quad \text{and} \quad \mathbb{P}(Y_{-\infty} \geq \mu) = \mathbb{P}\left(\bigcap_{n=1}^\infty \left\{Y_{-\infty} > \mu - \frac{1}{n}\right\}\right) = 1.$$

Hence $\mathbb{P}(Y_{-\infty} = \mu) = 1$, and $\mathbb{E}[X_1] = \mathbb{E}[\mathbb{E}[X_1 | \mathcal{F}_{-\infty}]] = \mathbb{E}[Y_{-\infty}] = \mu$. \square

3.3 Continuous-time Martingales

Definition 3.37 (Continuous-time martingales). Let $(X_t)_{t \geq 0}$ be a real-valued and L^1 process that is adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$. Here L^1 means $\mathbb{E}[|X_t|] < \infty$ for all $t \in \mathbb{R}_+$. Then

- (i) $(X_t)_{t \geq 0}$ is said to be a *martingale* if $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ for all $t > s \geq 0$;
- (ii) $(X_t)_{t \geq 0}$ is said to be a *supermartingale* if $\mathbb{E}[X_t | \mathcal{F}_s] \leq X_s$ for all $t > s \geq 0$;
- (iii) $(X_t)_{t \geq 0}$ is said to be a *submartingale* if $\mathbb{E}[X_t | \mathcal{F}_s] \geq X_s$ for all $t > s \geq 0$;

All these notations depends on the choice of the filtration $\{\mathcal{F}_t\}_{t \geq 0}$, which is fixed in later discussion.

Remark I. Clearly, if $(X_t)_{t \geq 0}$ is a submartingale, then $(-X_t)_{t \geq 0}$ is a supermartingale.

Remark II. Similar to Proposition 3.16, we have an immediate corollary of conditional Jensen's inequality. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function such that $\mathbb{E}[f(X_t)] < \infty$ for all $t \in \mathbb{R}_+$.

- (i) If $(X_t)_{t \geq 0}$ is a martingale, then $(f(X_t))_{t \geq 0}$ is a submartingale. Particularly, $(|X_t|)_{t \geq 0}$ is a submartingale.
- (ii) In addition, if f is monotone increasing and $(X_t)_{t \geq 0}$ is a submartingale, $(f(X_t))_{t \geq 0}$ is a submartingale. Particularly, $(X_t^+)_{t \geq 0}$ is a submartingale.

Proposition 3.38. Let $(X_t)_{t \geq 0}$ be a submartingale. Then for all $t \geq 0$,

$$\sup_{0 \leq s \leq t} \mathbb{E}[|X_s|] < \infty.$$

Proof. Clearly, we have $\mathbb{E}[X_s] = \mathbb{E}[\mathbb{E}[X_s | \mathcal{F}_0]] \geq \mathbb{E}[X_0]$. On the other hand, since (X_t^+) is also a submartingale, we have $\mathbb{E}[X_s^+] \leq \mathbb{E}[X_t^+]$ for all $0 \leq s \leq t$. Note that $|x| = 2x^+ - x$. Hence we have

$$\mathbb{E}[|X_s|] = 2\mathbb{E}[X_s^+] - \mathbb{E}[X_s] \leq 2\mathbb{E}[X_t^+] - \mathbb{E}[X_0] < \infty, \quad \forall s \in [0, t].$$

The result immediately follows. □

Proposition 3.39. Let $(X_t)_{t \geq 0}$ be an L^2 martingale. Then for all reals $0 \leq s < t$ and all finite partitions $s = t_0 < t_1 < \dots < t_k = t$, we have

$$\mathbb{E} \left[\sum_{j=1}^k (X_{t_j} - X_{t_{j-1}})^2 \middle| \mathcal{F}_s \right] = \mathbb{E} [X_t^2 - X_s^2 | \mathcal{F}_s] = \mathbb{E} [(X_t - X_s)^2 | \mathcal{F}_s].$$

Proof. For each $j = 1, \dots, k$,

$$\mathbb{E} [(X_{t_j} - X_{t_{j-1}})^2 | \mathcal{F}_s] = \mathbb{E} [\mathbb{E} [(X_{t_j} - X_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}}] | \mathcal{F}_s] = \mathbb{E} [\mathbb{E} [X_{t_j}^2 - X_{t_{j-1}}^2 | \mathcal{F}_{t_{j-1}}] | \mathcal{F}_s]$$

Then the desired result follows by summing over j . □

Now we extend the inequalities in Proposition 3.21 and Proposition 3.22 to continuous-time martingales with continuous sample paths.

Proposition 3.40 (i) (Maximal inequality). If $(X_t)_{t \geq 0}$ is a sample-right-continuous submartingale, then for every $t > 0$,

$$\lambda \mathbb{P} \left(\sup_{0 \leq s \leq t} |X_s| > \lambda \right) \leq \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|], \quad \forall \lambda > 0.$$

In addition, if $(X_t)_{t \geq 0}$ is nonnegative, then

$$\lambda \mathbb{P}(X_t^* > \lambda) \leq \mathbb{E}[X_t], \quad \forall \lambda > 0, \text{ where } X_t^* = \sup_{0 \leq s \leq t} X_s.$$

(ii) (Doob's L^p -inequality). If $(X_t)_{t \geq 0}$ is a sample-right-continuous martingale, then for every $t > 0$,

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|X_t|^p], \quad \forall p > 1.$$

Proof. We fix $t > 0$ and take a countable dense subset $t \in D \subset \mathbb{R}$. If f is a right-continuous function, we have

$$\sup_{s \in D \cap [0, t]} f(s) = \sup_{0 \leq s \leq t} f(s)$$

Here is a brief interpretation using diagonal trick. Let $M = \sup_{0 \leq s \leq t} f(s)$, then we can find a sequence $s_n \in [0, t]$ such that $f(s_n) \nearrow M$. Since f is right continuous, and $D \ni t$ is dense in \mathbb{R} , we can find a sequence $D \cap [0, t] \ni t_{nk} \searrow s_n$ such that $f(t_{nk}) \rightarrow f(s_n)$ for every n . Then $f(t_{nn}) \rightarrow M$, and $\sup_{s \in D \cap [0, t]} f(s) = M$.

Hence, by right-continuity of $s \mapsto X_s(\omega)$ and the fact that $t \in D$, we have $\sup_{s \in D \cap [0, t]} |X_s| = \sup_{0 \leq s \leq t} |X_s|$. Furthermore, we can view $D \cap [0, t]$ as the union of an increasing sequence of partitions $D_k = \{t_0^k, t_1^k, \dots, t_k^k\}$, where $0 \leq t_0^k < t_1^k < \dots < t_k^k = t$.

(i) For each $k \in \mathbb{N}$, we can apply the maximal inequality [Proposition 3.21] of discrete form on sequence $Y_n = X_{t_{n \wedge k}^k}$, which is a submartingale of the filtration $\mathcal{G}_n = \mathcal{F}_{t_{n \wedge k}^k}$:

$$\lambda \mathbb{P} \left(\max_{s \in D_k} |X_s| > \lambda \right) \leq \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|], \quad \forall k \in \mathbb{N}, \lambda > 0.$$

Note that $\max_{s \in D_k} |X_s| \nearrow \sup_{s \in D \cap [0, t]} |X_s| = \sup_{0 \leq s \leq t} |X_s|$ as $k \rightarrow \infty$. By monotone convergence theorem,

$$\lambda \mathbb{P} \left(\max_{s \in D_k} |X_s| > \lambda \right) \nearrow \lambda \mathbb{P} \left(\sup_{s \in [0, t]} |X_s| > \lambda \right) \leq \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|], \quad \forall \lambda > 0.$$

The case of nonnegative submartingale is similar.

(ii) Same as the proof of (i), we apply Doob's inequality [Proposition 3.22] of discrete form:

$$\mathbb{E} \left[\max_{s \in D_k} |X_k|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|X_t|^p], \quad \forall p > 1.$$

Since $\max_{s \in D_k} |X_k| \nearrow \sup_{0 \leq s \leq t} |X_s|$, use monotone convergence theorem. \square

Remark. If $(X_t)_{t \geq 0}$ is a submartingale, then for any dense subset $D \subset \mathbb{R}$ and every $t > 0$,

$$\mathbb{P} \left(\sup_{s \in D \cap [0, t]} |X_s| > \lambda \right) \leq \frac{1}{\lambda} (\mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|]), \quad \forall \lambda > 0.$$

Let $\lambda \rightarrow \infty$, we obtain that $\sup_{s \in D \cap [0, t]} |X_s| < \infty$ a.s. for all $t > 0$.

Definition 3.41 (Upcrossing number). Given a function $f : E \rightarrow \mathbb{R}$ and $a < b$, where $E \subset \mathbb{R}$, the *upcrossing number* of this sequence along $[a, b]$, denoted by $U_{[a, b]}^f(E)$, is the largest $k \in \mathbb{N}$ such that there exists a finite and strictly increasing sequence $s_1 < t_1 < s_2 < t_2 < \dots < s_k < t_k$ of elements of E such that $f(s_j) \leq a$ and $f(t_j) \geq b$ for all $j \in \{1, \dots, k\}$. If there exists no such sequence, we take $U_{[a, b]}^f(E) = 0$. If such sequence exists for all $k \in \mathbb{N}$, we take $U_{[a, b]}^f(E) = \infty$,

Given a sequence $x_n \in \mathbb{R}$, we write $x_n \uparrow x$ if $x_n \nearrow x$ and $x_n < x$ for all n . Also, $x_n \downarrow x$ if $x_n \searrow x$ and $x_n > x$ for all n . Following this notation, one can define the left and right limits of a function $f : E \rightarrow \mathbb{R}$:

$$f(t-) := \lim_{s \uparrow t} f(s), \quad \text{and} \quad f(t+) := \lim_{s \downarrow t} f(s).$$

The function $f : E \rightarrow \mathbb{R}$ is said to be *càdlàg* (French: *continue à droite, limite à gauche*), if for all $t \in \overset{\circ}{E}$, the left limit $f(t-) < \infty$ exists, and the right limit $f(t+)$ exists and equals $f(t)$.

Lemma 3.42. Let D be a countable dense subset of \mathbb{R}_+ , and $f : D \rightarrow \mathbb{R}$. Assume that for every $T \in D$,

- (i) the function f is bounded on $D \cap [0, T]$, and
- (ii) $U_{[a,b]}^f(D \cap [0, T]) < \infty$ for all rationals $a < b$.

Then the right limit $f(t+) = \lim_{D \ni s \downarrow t} f(s)$ exists for all $t \in \mathbb{R}_+$, and the left limit $f(t-) = \lim_{D \ni s \uparrow t} f(s)$ exists for all $t \in \mathbb{R}_{++}$. Furthermore, the function $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ defined by $g(t) = f(t+)$ is *càdlàg*.

Proof. We first fix $t \in \overline{\mathbb{R}}_+$, and prove that $f(t+) = \lim_{D \ni s \downarrow t} f(s)$ exists. Take $D \ni T > t$. By assumption (i), there exists $M > 0$ such that $|f(t)| \leq M$ for all $t \in D \cap [0, T]$. Take a sequence in $s_n \in D \cap [0, T]$ such that $s_n \downarrow t$. By Heine-Borel theorem, every subsequence of $f(s_n)$ has a further subsequence that converges. We prove that all such subsequences converges to the same point, which implies that $f(s_n)$ converges.

Argue by contradiction. If there exists two subsequence s_n and t_n such that $f(s_n) \rightarrow a$ and $f(t_n) = b > a$, take two rationals $a < p < q < b$. Then for any $k \in \mathbb{N}$, we can find a $f(t_{n_1}) \geq q$, and $s_{n_1} < t_{n_1}$ such that $f(s_{n_1}) \leq p$, and $t_{n_2} < s_{n_1}$ such that $f(t_{n_2}) \geq q$, \dots , and $s_{n_k} < t_{n_k}$ such that $f(s_{n_k}) \leq p$. Thus we obtain an upcrossing sequence $s_{n_k} < t_{n_k} < s_{n_{k-1}} < t_{n_{k-1}} < \dots < s_{n_1} < t_{n_1}$ of elements of $D \cap [0, T]$. Therefore, $U_{[p,q]}(D \cap [0, T]) > k$ for all $k \in \mathbb{N}$, a contradiction to (ii)!

As a result, all such sequences $D \cap [0, T] \ni s_n \downarrow t$ converges. They should converge to the same point. Otherwise, we can construct a sequence not converging by interlacing two sequences that converges to distinct points. Therefore, the right limit $f(t+) = \lim_{D \ni s \downarrow t} f(s)$ exists for all $t \in \mathbb{R}_+$. Similarly, we can prove that the left limit $f(t-) = \lim_{D \ni s \uparrow t} f(s)$ exists for all $t \in \mathbb{R}_{++}$.

Now we prove $g(t) = f(t+)$ is *càdlàg*. Given $\epsilon > 0$, we take $\delta > 0$ such that $|f(s) - f(t-)| < \epsilon$ for all $s \in (t - \delta, t)$, and $|f(r) - f(t+)| < \epsilon$ for all $r \in (t, t + \delta)$. Take $r_n \downarrow r \in (t, t + \delta)$, and $s_n \downarrow s \in (t - \delta, t)$. Then

$$|g(r) - f(t+)| = \lim_{n \rightarrow \infty} |f(r_n) - f(t+)| < \epsilon, \quad \text{and} \quad |g(s) - f(t-)| = \lim_{n \rightarrow \infty} |f(s_n) - f(t-)| < \epsilon.$$

Hence $\lim_{r \downarrow t} g(r) = f(t+)$, $\lim_{s \uparrow t} g(s) = f(t-)$, and g is *càdlàg*. □

Theorem 3.43. Let $(X_t)_{t \geq 0}$ be a submartingale, and let D be a countable dense subset of \mathbb{R}_+ .

- (i) For \mathbb{P} -a.e. $\omega \in \Omega$, the restriction of $t \mapsto X_t(\omega)$ to D has a right limit

$$X_{t+}(\omega) = \lim_{D \ni s \downarrow t} X_s(\omega), \quad \forall t \in \mathbb{R}_+$$

and a left limit

$$X_{t-}(\omega) = \lim_{D \ni s \uparrow t} X_s(\omega), \quad \forall t \in \mathbb{R}_{++}.$$

- (ii) For every $t \in \mathbb{R}_+$, $X_{t+} \in L^1(\Omega, \mathcal{F}_{t+}, \mathbb{P})$, and $\mathbb{E}[X_{t+} | \mathcal{F}_t] \geq X_t$ with equality holds if the mean function $t \mapsto \mathbb{E}[X_t]$ is right-continuous (in particular, if $(X_t)_{t \geq 0}$ is a martingale). The process $(X_{t+})_{t \geq 0}$ is a submartingale with respect to the filtration $(\mathcal{F}_{t+})_{t \geq 0}$.

Remark. In (ii), if X_{t+} is undefined on a negligible set N , we can just take $X_{t+}(\omega) = 0$ for $\omega \in N$.

Proof. (i) Fix $T \in D$. Then $\sup_{s \in D \cap [0, T]} |X_s| < \infty$ a.s.. As in the proof of Proposition 3.40, we take an sequence D_k increasing to $D \cap [0, T]$. Using Doob's upcrossing inequality [Proposition 3.26] and monotone convergence theorem, for all $a < b$, we have

$$\mathbb{E} \left[U_{[a, b]}^X(D \cap [0, T]) \right] \leq \frac{\mathbb{E}[(X_T - a)^+ - (X_0 - a)^+]}{b - a} < \infty \Rightarrow U_{[a, b]}^X(D \cap [0, T]) < \infty \text{ a.s..}$$

Set the negligible set N as

$$N = \bigcup_{T \in D} \left(\left\{ \sup_{s \in D \cap [0, T]} |X_s| = \infty \right\} \cap \left(\bigcup_{a, b \in \mathbb{Q}} \left\{ U_{[a, b]}^X(D \cap [0, T]) = \infty \right\} \right) \right) \Rightarrow \mathbb{P}(N) = 0. \quad (3.8)$$

Outside N , the assumptions in Lemma 3.42 are satisfied, and the result follows.

(ii) We supplement the definition $X_{t+}(\omega) = 0$ if $\lim_{D \ni s \downarrow t} X_s(\omega)$ does not exist, which occurs negligibly. Then X_{t+} is \mathcal{F}_{t+} -measurable. Fix $t \in \mathbb{R}_+$, and choose $t_n \downarrow t$. Then we have $X_{t_n} \rightarrow X_{t+}$ a.s.. Set $Y_n = X_{t_n}$ for all $n \in -\mathbb{N}_0$. Then Y_n is a backward submartingale with respect to the backward filtration $\mathcal{G}_n = \mathcal{F}_{t_n}$:

$$\mathbb{E}[Y_n | \mathcal{G}_{n-1}] = \mathbb{E}[X_{t_n} | \mathcal{F}_{t_{-(n-1)}}] \geq X_{t_{-(n-1)}} = Y_{n-1}, \quad \forall n \in -\mathbb{N}_0.$$

By Proposition 3.38, we have $\sup_{s \in D \cap [0, T]} \mathbb{E}[|X_s|] < \infty$, which implies $\lim_{n \rightarrow -\infty} \mathbb{E}[Y_n] > -\infty$. Using Doob's convergence theorem for discrete-time backward submartingales [Theorem 3.34], we have $X_{t_n} \rightarrow X_{t+}$ in L^1 , and $X_{t+} \in L^1(\Omega, \mathcal{F}_{t+}, \mathbb{P})$. Due to convergence in L^1 , we have

$$\mathbb{E}[X_{t+} | \mathcal{F}_t] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n} | \mathcal{F}_t] \geq X_t, \quad \text{and} \quad \mathbb{E}[X_{t+}] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n}].$$

Note the first equality holds because the conditional expectation operator $\mathbb{E}[\cdot | \mathcal{F}_t]$ is a bounded linear operator on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. In addition, if the mean function $s \mapsto \mathbb{E}[X_s]$ is right-continuous, the second equality implies $\mathbb{E}[X_{t+}] = \mathbb{E}[X_t]$, which requires $\mathbb{E}[X_{t+} | \mathcal{F}_t] = X_t$.

Let $s < t$, and take $s_n \downarrow s$ such that $s_n \leq t_n$. Then $X_{s_n} \rightarrow X_{s+}$ a.s. and in L^1 . Moreover, if $A \in \mathcal{F}_{s+}$,

$$\mathbb{E}[X_{t+} \mathbf{1}_A] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{t_n} \mathbf{1}_A] \geq \lim_{n \rightarrow \infty} \mathbb{E}[X_{s_n} \mathbf{1}_A] = \mathbb{E}[X_{s+} \mathbf{1}_A].$$

Since X_{s+} is \mathcal{F}_{s+} -measurable, we have $\mathbb{E}[X_{t+} | \mathcal{F}_{s+}] \geq X_{s+}$. Therefore, $(X_{t+})_{t \geq 0}$ is a submartingale with respect to the filtration $(\mathcal{F}_{t+})_{t \geq 0}$. \square

Theorem 3.44. Let $(\mathcal{F}_t)_{t \geq 0}$ be a right-continuous and complete filtration. Let $(X_t)_{t \geq 0}$ be a submartingale such that the mean function $t \mapsto \mathbb{E}[X_t]$ is right continuous. Then $(X_t)_{t \geq 0}$ has an a.s. modification with càdlàg sample paths, which is also a submartingale with respect to $(\mathcal{F}_t)_{t \geq 0}$.

Proof. Let D be a countable subset of \mathbb{R}_+ , and let N be the negligible set defined in (3.8). We take $Y_t := X_{t+}$ with the refinement $Y_t(\omega) = 0$ for $\omega \in N$. By Lemma 3.42, the sample paths of $(Y_t)_{t \geq 0}$ are càdlàg.

Since X_{t+} is \mathcal{F}_t -measurable by right-continuity of $(\mathcal{F}_t)_{t \geq 0}$, and since the negligible set N falls in all \mathcal{F}_t by completeness of $(\mathcal{F}_t)_{t \geq 0}$, the function Y_t is \mathcal{F}_t -measurable. Furthermore,

$$X_t = \mathbb{E}[X_{t+} | \mathcal{F}_t] = X_{t+} \stackrel{\text{a.s.}}{=} Y_t, \quad \forall t \in \mathbb{R}_+.$$

Hence $(Y_t)_{t \geq 0}$ is an a.s. modification of $(X_t)_{t \geq 0}$, which is adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Furthermore, we have $\mathbb{E}[Y_t | \mathcal{F}_s] = \mathbb{E}[X_t | \mathcal{F}_s] \geq X_s = Y_s$ for all $t > s \geq 0$. Hence $(Y_t)_{t \geq 0}$ is also a submartingale. \square

Theorem 3.45 (Doob's first martingale convergence theorem). If $(X_t)_{t \geq 0}$ is a sample-right-continuous submartingale, and $\sup_{t \geq 0} \mathbb{E}[X_t^+] < \infty$, then $X_\infty = \lim_{t \rightarrow \infty} X_t$ *a.s.* exists, and $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Let D be a countable subset of \mathbb{R}_+ , and let $M := \sup_{t \geq 0} \mathbb{E}[X_t^+] < \infty$. For all $a < b$, we can follow the proof of Theorem 3.43 (i) and use monotone convergence theorem:

$$\mathbb{E} \left[U_{[a,b]}^X(D \cap [0, T]) \right] \leq \frac{\mathbb{E}[(X_T - a)^+ - (X_0 - a)^+]}{b - a} \leq \frac{M + |a|}{b - a} \Rightarrow \mathbb{E} \left[U_{[a,b]}^X(D) \right] \leq \frac{M + |a|}{b - a} < \infty.$$

Hence we have

$$U_{[a,b]}^X(D) < \infty \text{ a.s. } \forall a, b \in \mathbb{Q} \text{ with } a < b \Rightarrow X_\infty = \lim_{D \ni t \rightarrow \infty} X_t \in [-\infty, \infty] \text{ a.s. exists.}$$

We can further exclude values ∞ and $-\infty$ \mathbb{P} -a.e., because the Fatou's lemma gives

$$\mathbb{E}[X_\infty^+] \leq \liminf_{D \ni t \rightarrow \infty} \mathbb{E}[X_t^+] \leq M, \quad \mathbb{E}[X_\infty^-] \leq \liminf_{D \ni t \rightarrow \infty} \mathbb{E}[X_t^-] = \liminf_{D \ni t \rightarrow \infty} \mathbb{E}[X_t^+ - X_t] \leq M - \mathbb{E}[X_0].$$

Hence $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, since $(X_t)_{t \geq 0}$ has right-continuous sample paths, we can drop the restriction $D \ni t$ in our limit. \square

Similarly, we need uniform integrability of martingale to obtain convergence in L^1 .

Theorem 3.46 (Doob's second martingale convergence theorem). Let $(X_t)_{t \geq 0}$ be a sample-right-continuous martingale. The following are equivalent:

- (i) The collection $\{X_t\}_{t \in \mathbb{R}_+}$ is uniformly integrable.
- (ii) X_t converges *a.s.* and in L^1 -norm.
- (iii) $(X_t)_{t \geq 0}$ is closed, i.e. there exists $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_t = \mathbb{E}[Z | \mathcal{F}_t]$ for all $t \in \mathbb{R}_+$.

Proof. (i) \Rightarrow (ii): By Theorem 3.45, the limit $X_\infty = \lim_{t \rightarrow \infty} X_t$ *a.s.* exists. Since $\{X_t\}_{t \in \mathbb{R}_+}$ is uniformly integrable, convergence in L^1 also holds by Theorem 1.73.

(ii) \Rightarrow (iii) follows from the continuity of conditional expectation operator on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

(iii) \Rightarrow (i) follows from Claim 3.28*. \square

Remark. If (i)-(iii) are satisfied, the limit $X_\infty = \lim_{t \rightarrow \infty} X_t$ satisfies $\mathbb{E}[X_\infty | \mathcal{F}_t] = \lim_{s \rightarrow \infty} \mathbb{E}[X_s | \mathcal{F}_t] = X_t$.

Theorem 3.47 (Convergence theorem for L^p -bounded continuous-time martingales). Let $(X_t)_{t \geq 0}$ is a martingale such that $\sup_{t \geq 0} \mathbb{E}[|X_t|^p] < \infty$, where $p > 1$. Then $X_\infty = \lim_{t \rightarrow \infty} X_t$ *a.s.* and in L^p .

Proof. Let $Y = \sup_{t \geq 0} |X_t|$, and $M = \sup_{t \geq 0} \mathbb{E}[|X_t|^p] < \infty$. By Theorem 3.45, there exists $X_\infty \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_t \rightarrow X_\infty$ *a.s.* By Doob's L^p -inequality [Proposition 3.38 (ii)] and monotone convergence theorem,

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|X_t|^p] \leq \left(\frac{p}{p-1} \right)^p M \Rightarrow \mathbb{E}[Y^p] \leq \left(\frac{p}{p-1} \right)^p M < \infty.$$

Since $|X_t - X_\infty| \leq 2Y$, by Lebesgue dominated convergence theorem, $\|X_t - X_\infty\|_p \rightarrow 0$. \square

Given a sample-right-continuous submartingale $(X_t)_{t \geq 0}$ such that $\sup_{t \geq 0} \mathbb{E}[X_t^+] < \infty$, and a stopping time τ , we define the random variable

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega) \mathbb{1}_{\{\tau < \infty\}}(\omega) + X_\infty(\omega) \mathbb{1}_{\{\tau = \infty\}}(\omega), \text{ where } X_\infty = \lim_{t \rightarrow \infty} X_t \text{ a.s..}$$

By Proposition 3.9 and Proposition 3.12, $(X_t)_{t \geq 0}$ is progressive, and the restriction of X_τ to $\{\tau < \infty\}$ is \mathcal{F}_τ -measurable. Meanwhile, $\{X_\infty \mathbf{1}_{\{\tau=\infty\}} \leq \alpha\} \subset \{\tau \leq t\}$ if $\alpha \geq 0$, and $\{X_\infty \mathbf{1}_{\{\tau=\infty\}} \leq \alpha\} \cap \{\tau \leq t\} = \emptyset$ if $\alpha < 0$. Therefore, $\{X_\infty \mathbf{1}_{\{\tau=\infty\}} \leq \alpha\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$, and $X_\infty \mathbf{1}_{\{\tau=\infty\}}$ is \mathcal{F}_τ -measurable. As a result, the random variable X_τ is \mathcal{F}_τ -measurable.

Theorem 3.48 (Optional stopping theorem for submartingales). Let $(X_t)_{t \geq 0}$ be a sample-right-continuous submartingale. Let τ and σ be two stopping times such that $\sigma \leq \tau$. Then $X_\tau, X_\sigma \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma$, if either of the following conditions holds:

- (i) τ and σ are bounded stopping times;
- (ii) $(X_t)_{t \geq 0}$ is uniformly bounded by some $U \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$ from above, i.e. $X_t \leq U$ for all $t \geq 0$.

Proof. (i) Suppose $\tau \leq M$, where $M \in \mathbb{N}$. Akin to Proposition 3.14, we define two sequences of stopping times $\sigma_n \leq \tau_n$ that decrease to σ and τ , respectively:

$$\sigma_n = \sum_{k=0}^{M2^n-1} \frac{k+1}{2^n} \mathbf{1}_{\{k2^{-n} < \sigma \leq (k+1)2^{-n}\}}, \quad \tau_n = \sum_{k=0}^{M2^n-1} \frac{k+1}{2^n} \mathbf{1}_{\{k2^{-n} < \tau \leq (k+1)2^{-n}\}}, \quad n \in \mathbb{N}.$$

We fix $n \geq 2$. Then the sequence $(X_{k2^{-n}})_{k=0}^\infty$ is a discrete-time submartingale with respect to the filtration $(\mathcal{F}_{k2^{-n}})_{k=0}^\infty$. Furthermore, both $2^n \tau_{n-1}$ and $2^n \tau_n$ are stopping times of $(\mathcal{F}_{k2^{-n}})_{k=0}^\infty$:

$$\begin{aligned} 2^n \tau_{n-1} &= \min \{k \in \{2, 4, \dots, M2^n\} : k2^{-n} \geq \tau\} \Rightarrow \{2^n \tau_{n-1} \leq p\} = \left\{ \tau \leq \left\lfloor \frac{p}{2} \right\rfloor 2^{1-n} \right\} \in \mathcal{F}_{p2^{-n}}, \quad \forall p \in \mathbb{N}; \\ 2^n \tau_n &= \min \{k \in \{1, 2, \dots, M2^n\} : k2^{-n} \geq \tau\} \Rightarrow \{2^n \tau_n \leq p\} = \{\tau \leq p2^{-n}\} \in \mathcal{F}_{p2^{-n}}, \quad \forall p \in \mathbb{N}. \end{aligned}$$

By the optional stopping theorem for discrete-time submartingales [Theorem 3.19 (iii)], we know that $\mathbb{E}[X_{\tau_{n-1}} | \mathcal{F}_{\tau_n}] \geq X_{\tau_n}$. Hence $Y_n := X_{\tau_{n-1}}$ is a backward submartingale. Furthermore, by Theorem 3.19 (i, ii), we have $\mathbb{E}[X_{\tau_n}] \geq \mathbb{E}[X_0]$ for all $n \in \mathbb{N}$. Apply Theorem 3.34, the sequence $(X_{\tau_n})_{n=1}^\infty$ is uniformly integrable.

Since $\tau_n(\omega) \searrow \tau(\omega)$, and the sample path $t \mapsto X_t(\omega)$ is right-continuous, we have $X_{\tau_n} \rightarrow X_\tau$ a.s., and this convergence also holds in L^1 . Also, $X_{\sigma_n} \rightarrow X_\sigma$ a.s. and in L^1 . Then for all $A \in \mathcal{F}_\sigma$, we have

$$\text{(By Theorem 3.19)} \quad \mathbb{E}[X_{\tau_n} \mathbf{1}_A] \geq \mathbb{E}[X_{\sigma_n} \mathbf{1}_A], \quad \forall n \in \mathbb{N} \Rightarrow \mathbb{E}[X_\tau \mathbf{1}_A] \geq \mathbb{E}[X_\sigma \mathbf{1}_A],$$

where the \Rightarrow follows from L^1 convergence. Hence $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma$.

(ii) Apply (i) to bounded stopping times 0 and $\tau \wedge n$, we have $\mathbb{E}[X_{\tau \wedge n}] \geq \mathbb{E}[X_0] > -\infty$. Since X_t is bounded from above by U , by Fatou's lemma, $\mathbb{E}[X_\tau] \geq \limsup_{n \rightarrow \infty} \mathbb{E}[X_{\tau \wedge n}] \geq \mathbb{E}[X_0] > -\infty$. Similarly we have $\mathbb{E}[X_\sigma] \geq \mathbb{E}[X_0] > -\infty$. Hence $X_\tau, X_\sigma \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Fix $A \in \mathcal{F}_\sigma \subset \mathcal{F}_\tau$, and define $\tau^A(\omega) = \tau(\omega) \mathbf{1}_A + \infty \mathbf{1}_{A^c}$. According to Proposition 3.11 (d), both τ^A and σ^A are also stopping times. By (i), we have

$$\begin{aligned} 0 &\leq \mathbb{E}[X_{\tau^A \wedge n}] - \mathbb{E}[X_{\sigma^A \wedge n}] \\ &= \mathbb{E}[X_n \mathbf{1}_{A^c} + X_{\tau \wedge n} \mathbf{1}_{A \cap \{\sigma \leq n\}} + X_n \mathbf{1}_{A \cap \{\sigma > n\}}] - \mathbb{E}[X_n \mathbf{1}_{A^c} + X_\sigma \mathbf{1}_{A \cap \{\sigma \leq n\}} + X_n \mathbf{1}_{A \cap \{\sigma > n\}}] \\ &= \mathbb{E}[X_{\tau \wedge n} \mathbf{1}_{A \cap \{\sigma \leq n\}}] - \mathbb{E}[X_\sigma \mathbf{1}_{A \cap \{\sigma \leq n\}}] \end{aligned}$$

Note that $\mathbf{1}_{A \cap \{\sigma \leq n\}} \rightarrow \mathbf{1}_{A \cap \{\sigma < \infty\}}$ and $X_{\tau \wedge n} \rightarrow X_\tau$ as $n \rightarrow \infty$. By Lebesgue dominated convergence theorem, we have $\mathbb{E}[X_\tau \mathbf{1}_{A \cap \{\sigma < \infty\}}] \geq \mathbb{E}[X_\sigma \mathbf{1}_{A \cap \{\sigma < \infty\}}]$. Clearly, $\mathbb{E}[X_\tau \mathbf{1}_{A \cap \{\sigma = \infty\}}] = \mathbb{E}[X_\sigma \mathbf{1}_{A \cap \{\sigma = \infty\}}]$. Hence

$$\mathbb{E}[\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \mathbf{1}_A] = \mathbb{E}[X_\tau \mathbf{1}_A] \geq \mathbb{E}[X_\sigma \mathbf{1}_A], \quad \forall A \in \mathcal{F}_\sigma.$$

Since X_σ is \mathcal{F}_σ -measurable, we have $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma$. □

Theorem 3.49 (Optional stopping theorem for uniformly integrable martingales). Let $(X_t)_{t \geq 0}$ be a uniformly integrable sample-right-continuous martingale. Let τ be a stopping time. Then we have

$$\mathbb{E}[X_\infty | \mathcal{F}_\tau] = X_\tau \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Furthermore, if σ is another stopping time such that $\sigma \leq \tau$, then $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = X_\sigma$.

Proof. Using Proposition 3.14, we define a sequence of stopping times $\tau_n \searrow \tau$ as follows:

$$\tau_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{\{k2^{-n} < \tau \leq (k+1)2^{-n}\}} + \infty \mathbb{1}_{\{\tau = \infty\}}, \quad n \in \mathbb{N}.$$

Clearly, $2^n \tau_n$ is a stopping time of the filtration $\{\mathcal{F}_{(k+1)2^{-n}}\}_{k=0}^{\infty}$. Apply Theorem 3.32 to discrete-time martingale $\{X_{(k+1)2^{-n}}\}_{k=0}^{\infty}$ with respect to the filtration $\{\mathcal{F}_{(k+1)2^{-n}}\}_{k=0}^{\infty}$, which is uniformly integrable, we have $\mathbb{E}[X_\infty | \mathcal{F}_{\tau_n}] = X_{\tau_n} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Since $\tau_n(\omega) \searrow \tau(\omega)$, and the sample path $t \mapsto X_t(\omega)$ is right-continuous, we have $X_{\tau_n} \rightarrow X_\tau$ a.s.. Furthermore, $\{X_{\tau_n}\}_{n=1}^{\infty}$ is uniformly integrable, so the convergence also holds in L^1 . For all $A \in \mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$, $\mathbb{E}[X_\infty | \mathcal{F}_{\tau_n}] = X_{\tau_n}$ implies $\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[X_{\tau_n} \mathbb{1}_A]$. Then

$$\mathbb{E}[X_\tau \mathbb{1}_A] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{\tau_n} \mathbb{1}_A] = \mathbb{E}[X_\infty \mathbb{1}_A], \quad \forall A \in \mathcal{F}_\tau \quad \Rightarrow \quad \mathbb{E}[X_\infty | \mathcal{F}_\tau] = X_\tau.$$

Furthermore, if $\sigma \leq \tau$ is a stopping time, then $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] = \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_\tau] | \mathcal{F}_\sigma] = \mathbb{E}[X_\infty | \mathcal{F}_\sigma] = X_\sigma$. \square

Like in Definition 3.30, if an adapted process $(X_t)_{t \geq 0}$ is stopped by a stopping time τ , we define the stopped process as $X_t^\tau = X_{t \wedge \tau}$ for $t \geq 0$.

Corollary 3.50. Let $(X_t)_{t \geq 0}$ be a sample-right-continuous submartingale. Let τ be a stopping time.

- (i) The stopped process $(X_t^\tau)_{t \geq 0}$ is a submartingale.
- (ii) In addition, if $(X_t)_{t \geq 0}$ is a uniformly integrable martingale, so is the stopped process $(X_t^\tau)_{t \geq 0}$. Moreover,

$$X_t^\tau = \mathbb{E}[X_\tau | \mathcal{F}_t].$$

Proof. (i) Fix $t > s \geq 0$. If $A \in \mathcal{F}_s \subset \mathcal{F}_t$, we have $A \cap \{\tau > s\} \in \mathcal{F}_s$, and $A \cap \{\tau > s\} \in \mathcal{F}_\tau$ by the very definition of \mathcal{F}_τ . Hence $A \cap \{\tau > t\} \in \mathcal{F}_s \cap \mathcal{F}_\tau = \mathcal{F}_{\tau \wedge s}$, and

$$\begin{aligned} \mathbb{E}[X_t^\tau \mathbb{1}_A] - \mathbb{E}[X_s^\tau \mathbb{1}_A] &= \mathbb{E}[X_{\tau \wedge t} \mathbb{1}_{A \cap \{\tau \leq s\}}] + \mathbb{E}[X_{\tau \wedge t} \mathbb{1}_{A \cap \{\tau > s\}}] - \mathbb{E}[X_{\tau \wedge s} \mathbb{1}_A] \\ &= \mathbb{E}[X_{\tau \wedge s} \mathbb{1}_{A \cap \{\tau \leq s\}}] + \mathbb{E}[X_{\tau \wedge t} \mathbb{1}_{A \cap \{\tau > s\}}] - \mathbb{E}[X_{\tau \wedge s} \mathbb{1}_A] \\ &= \mathbb{E}[(X_{\tau \wedge t} - X_{\tau \wedge s}) \mathbb{1}_{A \cap \{\tau > s\}}] = \mathbb{E}[\mathbb{E}[X_{\tau \wedge t} - X_{\tau \wedge s} | \mathcal{F}_{\tau \wedge s}] \mathbb{1}_{A \cap \{\tau > s\}}] = 0, \end{aligned}$$

where the last inequality follows from Theorem 3.48, because $\tau \wedge s \leq \tau \wedge t$ are two bounded stopping times.

(ii) Fix $t \geq 0$. If $A \in \mathcal{F}_t$, we have $A \cap \{\tau > t\} \in \mathcal{F}_t$, and $A \cap \{\tau > t\} \in \mathcal{F}_\tau$ by the very definition of \mathcal{F}_τ . Hence $A \cap \{\tau > t\} \in \mathcal{F}_t \cap \mathcal{F}_\tau = \mathcal{F}_{\tau \wedge t}$, and

$$\begin{aligned} \mathbb{E}[X_\tau \mathbb{1}_A] - \mathbb{E}[X_t^\tau \mathbb{1}_A] &= \mathbb{E}[X_{\tau \wedge t} \mathbb{1}_{A \cap \{\tau \leq t\}}] + \mathbb{E}[X_\tau \mathbb{1}_{A \cap \{\tau > t\}}] - \mathbb{E}[X_{\tau \wedge t} \mathbb{1}_A] \\ &= \mathbb{E}[(X_\tau - X_{\tau \wedge t}) \mathbb{1}_{A \cap \{\tau > t\}}] = \mathbb{E}[\mathbb{E}[X_\tau - X_{\tau \wedge t} | \mathcal{F}_{\tau \wedge t}] \mathbb{1}_{A \cap \{\tau > t\}}] = 0, \end{aligned}$$

where the last inequality follows from Theorem 3.49, because $\tau \wedge t \leq \tau$ is a stopping time. Since $A \in \mathcal{F}_t$ is arbitrary, and $X_t^\tau = X_{\tau \wedge t}$ is \mathcal{F}_t -measurable, we have $\mathbb{E}[X_\tau | \mathcal{F}_t] = X_t^\tau$. Since $X_\tau \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, the stopped process $(X_t^\tau)_{t \geq 0}$ is uniformly integrable. \square

3.4 Local Martingales, Quadratic Variation and Bracket

Review: Functions of bounded variation. Given a function $f : [0, T] \rightarrow \mathbb{R}$, we define the *total variation* of f on interval $[0, T]$ as

$$V_0^T f = \sup \left\{ \sum_{j=1}^n |f(x_j) - f(x_{j-1})| : n \in \mathbb{N}, 0 = x_0 < x_1 < \dots < x_n = T \right\}$$

If $V_0^T f < \infty$, we say that $f : [0, T] \rightarrow \mathbb{R}$ has *bounded variation*.

We can show that every function of bounded variation is the difference of two monotone increasing functions. If $f : [0, T] \rightarrow \mathbb{R}$ is a function of bounded variation, we define $v_f(0) = 0$ and $v_f(t) = V_0^t f$ for all $t \in (0, T]$, which is the total variation of $f|_{[0, t]}$. Then we have $v_f(t) - v_f(s) = V_s^t(f) \geq |f(t) - f(s)|$ for all $0 \leq s < t \leq T$. Therefore, both $v_f + f$ and $v_f - f$ are monotone increasing functions on $[0, T]$, and $f = \frac{1}{2}(v_f + f) - \frac{1}{2}(v_f - f)$. In addition, if $f(0) = 0$, we can require both $v_f + f$ and $v_f - f$ to be nonnegative.

Furthermore, if $f : [0, T] \rightarrow \mathbb{R}$ is a càdlàg function of bounded variation such that $f(0) = 0$, there exists a finite signed measure μ such that $\mu([0, t]) = f(t)$ for all $t \in [0, T]$.

Since v_f is monotone on $[0, T]$, v_f has a left limit $v_f(s-)$ at every $s \in (0, T]$, and a right limit $v_f(t+)$ at every $t \in [0, T)$. We prove that $v_f(t+) = v_f(t)$ for all $t \in [0, T)$. Fix $\epsilon > 0$. Since f is right-continuous, choose $\delta > 0$ such that $|f(x) - f(t)| < \epsilon/2$ for all $x \in (t, t + \delta)$. We also choose a partition $t = x_0 < x_1 < \dots < x_n = T$ such that $\sum_{j=1}^n |f(x_j) - f(x_{j-1})| > V_t^T f - \frac{\epsilon}{2}$. Then for all $x < \min\{x_1, \delta\}$, we have

$$V_t^T f - \frac{\epsilon}{2} < \sum_{j=1}^n |f(x_j) - f(x_{j-1})| \leq |f(x) - f(t)| + |f(x_1) - f(x)| + \sum_{j=2}^n |f(x_j) - f(x_{j-1})| \leq \frac{\epsilon}{2} + V_x^T f.$$

Hence $v_f(x) - v_f(t) = V_t^x f = V_t^T f - V_x^T f < \epsilon$ for all $x \in \min\{x_1, \delta\}$, and v_f is right-continuous. As a result, both $v_f + f$ and $v_f - f$ are nonnegative, monotone increasing and càdlàg functions on $[0, T]$. Akin to the Carathéodory extension procedure of a c.d.f. in the Remark of Definition 2.3, there exists two Borel measures μ^+ and μ^- such that $\mu^+([0, t]) = \frac{1}{2}(v_f(t) + f(t))$ and $\mu^-([0, t]) = \frac{1}{2}(v_f(t) - f(t))$ for all $t \in [0, T]$. Then $\mu = \mu^+ - \mu^-$ is a signed measure with $\mu([0, t]) = f(t)$ for all $t \in [0, T]$.

Moreover, the total variation measure $|\mu|$ of μ satisfies $|\mu|([0, T]) = v_f(T) = V_0^T f$. For any partition $0 = x_0 < x_1 < \dots < x_n = T$, we have $\sum_{j=1}^n |f(x_j) - f(x_{j-1})| \leq |\mu|([0, T])$, hence $V_0^T f \leq |\mu|([0, T])$. To prove the opposite, we define a probability measure $\mathbb{P}(A) = \frac{|\mu|(A)}{|\mu|([0, T])}$ on $\mathcal{B}([0, T])$. Let $P \amalg N = [0, T]$ be the Hahn decomposition associated with μ , and define $Y = \mathbf{1}_P - \mathbf{1}_N$. Let $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = T$ be an increasing sequence of partitions of interval $[0, t]$ such that the mesh $\max_{1 \leq j \leq k_n} (t_j^n - t_{j-1}^n) \rightarrow 0$, and let \mathcal{B}_n be the sub σ -algebra generated by intervals $(t_j^n, t_{j-1}^n]$. Then $(\mathcal{B}_n)_{n=1}^\infty$ is a filtration with $\mathcal{B}_\infty = \mathcal{B}([0, T])$, and $X_n = \mathbb{E}[Y | \mathcal{B}_n]$ is a uniformly integrable martingale sequence. Furthermore, by properties of conditional expectation, X_n is a constant on each subinterval $(t_{j-1}^n, t_j^n]$, and

$$X_n|_{(t_{j-1}^n, t_j^n]} = \frac{\mathbb{E}[X_n \mathbf{1}_{(t_{j-1}^n, t_j^n]}]}{\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{\mathbb{E}[Y \mathbf{1}_{(t_{j-1}^n, t_j^n]}]}{\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{\mu((t_{j-1}^n, t_j^n])}{|\mu|([0, T])\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{f(t_j^n) - f(t_{j-1}^n)}{|\mu|((t_{j-1}^n, t_j^n])}, \quad \forall 1 \leq j \leq k_n.$$

Now it suffices to prove that $\sum_{j=1}^{k_n} |f(t_j^n) - f(t_{j-1}^n)| \rightarrow \mu([0, T])$. By Doob's convergence theorem for uniformly integrable martingales, we have $X_n \rightarrow Y$ a.s. and in L^1 . As a result,

$$\mathbb{E}[|X_n|] = \sum_{j=1}^{k_n} \frac{|f(t_j^n) - f(t_{j-1}^n)|}{\mu([0, T])} \rightarrow \mathbb{E}[|Y|] = 1.$$

Also, we can see that $\mu = \mu^+ - \mu^-$ is indeed the Jordan decomposition of μ .

Review: Functions of finite variation. A function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is said to have *finite variation*, if the restriction $f|_{[0,T]}$ has bounded variation on $[0, T]$ for all $T \in \mathbb{R}_{++}$. In addition, if $f(0) = 0$ and f is càdlàg, we can find a unique σ -finite signed measure μ on $\mathcal{B}(\mathbb{R}_+)$ such that $\mu([0, t]) = f(t)$ for all $t \in \mathbb{R}_+$, and $|\mu|([0, t])$ is the total variation of $f|_{[0,t]}$.

Review: Lebesgue Stieltjes integral. Let $f : [0, T] \rightarrow \mathbb{R}$ be a càdlàg function of bounded variation with $f(0) = 0$, so we can find a finite signed measure on $[0, T]$ such that $\mu([0, t]) = f(t)$ for all $t \in [0, T]$. If $\varphi : [0, T] \rightarrow \mathbb{R}$ is a measurable function such that $\int_{[0,T]} |\varphi| d|\mu| < \infty$, define the Lebesgue-Stieltjes integral

$$\int_0^T \varphi(s) df(s) = \int_{[0,T]} \varphi d\mu, \quad \int_0^T \varphi(s) |df(s)| = \int_{[0,T]} \varphi d|\mu|.$$

It is seen that the function $t \mapsto \int_0^t \varphi(s) df(s)$ is also of bounded variation on $[0, T]$. To see this, note the associated signed measure is $\nu(A) = \int_A \varphi d\mu$, and the total variation on $[0, T]$ is $|\nu|([0, T]) = \int_{[0,T]} |\varphi| d|\mu| < \infty$.

Also, if $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a càdlàg function of finite variation with $f(0) = 0$, we can define the Lebesgue-Stieltjes integral

$$\int_0^\infty \varphi(s) df(s) = \lim_{T \rightarrow \infty} \int_0^T \varphi(s) df(s)$$

for all measurable functions φ such that $\int_0^\infty |\varphi(s)| |df(s)| < \infty$.

Review: Approximation of Lebesgue Stieltjes integral. We can approximate a Lebesgue-Stieltjes integral of a continuous function by differentiating on a mesh. Let $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = T$ be a sequence of partitions whose mesh $\max_{1 \leq j \leq k_n} (t_j^n - t_{j-1}^n) \rightarrow 0$. Then

$$\int_0^T \varphi_n(s) df(s) = \sum_{j=1}^{k_n} \varphi(t_j^n) (f(t_j^n) - f(t_{j-1}^n)), \text{ where } \varphi_n(s) = \varphi(0) \mathbf{1}_{\{0\}}(s) + \sum_{j=1}^{k_n} \varphi(t_j^n) \mathbf{1}_{(t_{j-1}^n, t_j^n]}(s).$$

By continuity of φ , we have $\varphi_n \rightarrow \varphi$, and all these functions are dominated by a constant $\max_{s \in [0, T]} |\varphi(s)|$. By Lebesgue dominated convergence theorem, we have

$$\int_0^T \varphi(s) df(s) = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} \varphi(t_j^n) (f(t_j^n) - f(t_{j-1}^n)).$$

Definition 3.51 (Finite variation processes). An adapted process $(X_t)_{t \geq 0}$ is said to be a *finite variation process* if all its sample paths $t \mapsto X_t(\omega)$ are functions of finite variation on \mathbb{R}_+ . In addition, if all sample paths $t \mapsto X_t(\omega)$ are monotone increasing, the process $(X_t)_{t \geq 0}$ is said to be an *increasing process*.

Remark. If $(A_t)_{t \geq 0}$ is a finite variation process, then

$$V_t = \int_0^t |dA_s|, \quad \forall t \in \mathbb{R}_+$$

is an increasing process. Writing $A_t = \frac{1}{2}(V_t + A_t) - \frac{1}{2}(V_t - A_t)$ shows that any finite variation process can be written as the difference of two increasing processes. Note that V_t is \mathcal{F}_t -measurable, because

$$V_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} |A_{t_j^n} - A_{t_{j-1}^n}|, \text{ where } 0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t \text{ is an increasing sequence of partitions of } [0, t].$$

Proposition 3.52. Let $A = (A_t)_{t \geq 0}$ be a finite variation process, and let $H = (H_t)_{t \geq 0}$ be a progressive process such that $\int_0^t |H_s(\omega)| |dA_s(\omega)| < \infty$ for all $t \in \mathbb{R}_+$ and all $\omega \in \Omega$. Then the process $H \cdot A$ defined by

$$(H \cdot A)_t = \int_0^t H_s dA_s, \quad \forall t \in \mathbb{R}_+.$$

is also a finite variation process.

Proof. It suffices to show that $H \cdot A$ is an adapted process: If $h : \Omega \times [0, t] \rightarrow \mathbb{R}$ is measurable on $\mathcal{F}_t \times \mathcal{B}([0, t])$, and $\int_0^t |h(\omega, s)| |dA_s(\omega)| < \infty$ for all $\omega \in \Omega$, then $\omega \mapsto \int_0^t h(\omega, s) dA_s(\omega)$ is \mathcal{F}_t -measurable.

If $h(\omega, s) = \mathbf{1}_F(\omega) \mathbf{1}_{(p, q]}(s)$, where $F \in \mathcal{F}_t$ and $(p, q] \subset [0, t]$, we have

$$\int_0^t h(\omega, s) dA_s(\omega) = \mathbf{1}_F(\omega) (A_q(\omega) - A_p(\omega)).$$

Clearly, $\mathbf{1}_F(A_q - A_p)$ is \mathcal{F}_t -measurable. Now we define

$$\mathcal{L}_t = \left\{ G \in \mathcal{F}_t \otimes \mathcal{B}([0, t]) : \omega \mapsto \int_0^t \mathbf{1}_G(\omega, s) dA_s(\omega) \text{ is } \mathcal{F}_t\text{-measurable} \right\}.$$

Note this is a λ -system containing $\{F \times (p, q] : F \in \mathcal{F}_t, (p, q] \subset [0, t]\}$, which is a π -system generating $\mathcal{F}_t \otimes \mathcal{B}([0, t])$. By Sierpiński-Dynkin π - λ theorem, we have $\mathcal{L}_t = \mathcal{F}_t \otimes \mathcal{B}([0, t])$. Hence $\omega \mapsto \int_0^t h(\omega, s) dA_s(\omega)$ is \mathcal{F}_t -measurable for all simple functions h . The remaining part follows from simple function approximation and the Lebesgue dominated convergence theorem. \square

Remark. If the filtration $(\mathcal{F}_t)_{t \geq 0}$ is complete, then Proposition 3.52 holds for all progressive process $(H_t)_{t \geq 0}$ such that $\int_0^t |H_s(\omega)| |dA_s(\omega)| < \infty$ for all $t \in \mathbb{R}_+$ and \mathbb{P} -a.e. $\omega \in \Omega$. To see this, define

$$H'_t(\omega) = \begin{cases} H_t(\omega) & \text{if } \int_0^t |H_s(\omega)| |dA_s(\omega)| < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Then $(H'_t)_{t \geq 0}$ is still progressive, and we define $H \cdot A = H' \cdot A$.

Definition 3.53 (Continuous local martingales). An adapted sample-continuous process $X = (X_t)_{t \geq 0}$ with $X_0 = 0$ a.s. is said to be a *continuous local martingale* if there exists an increasing sequence $(\tau_n)_{n=1}^\infty$ of stopping times such that $\tau_n \nearrow \infty$ and the stopped process $X^{\tau_n} = (X_t^{\tau_n})_{t \geq 0}$ is a uniformly integrable martingale. In this case, the sequence $(\tau_n)_{n=1}^\infty$ of stopping times is said to *reduce* process X .

More generally, an adapted sample-continuous process $X = (X_t)_{t \geq 0}$ is said to be a *continuous local martingale* if the process $Y_t = X_t - X_0$ is a continuous local martingale. Here we do not assume X_0 is L^1 .

Remark. In the Definition 3.53, one can replace “uniformly integrable martingale” by “martingale”. In the latter case, $\tau_n \wedge n$ is a sequence of stopping times such that $X^{\tau_n \wedge n}$ is uniformly integrable, and $\tau_n \wedge n \nearrow \infty$.

The following two basic facts about continuous local martingales immediately follow from Corollary 3.50. Suppose $X = (X_t)_{t \geq 0}$ is a continuous local martingale. Then:

- (i) For any stopping time τ , the stopped process $(X_t^\tau)_{t \geq 0}$ is also a continuous local martingale.
- (ii) If $(\tau_n)_{n=1}^\infty$ is a sequence of stopping times reducing X , and $(\sigma_n)_{n=1}^\infty$ is a sequence of stopping times such that $\sigma_n \nearrow \infty$, then $(\sigma_n \wedge \tau_n)_{n=1}^\infty$ also reduces X .

We can show that all continuous local martingales form a vector space. To see this, let X and X' be two continuous local martingales reduced by stopping time sequences $(\tau_n)_{n=1}^\infty$ and $(\tau'_n)_{n=1}^\infty$, respectively. Using property (ii), we know that $(\tau_n \wedge \tau'_n)_{n=1}^\infty$ is a stopping time sequence that reduces process $X + X'$.

Proposition 3.54. Let $X = (X_t)_{t \geq 0}$ be a continuous local martingale.

- (i) If $(X_t)_{t \geq 0}$ is nonnegative and $X_0 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then $(X_t)_{t \geq 0}$ is a supermartingale.
- (ii) If there exists random variable $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X_t| \leq Z$ for all $t \geq 0$, then $(X_t)_{t \geq 0}$ is a uniformly integrable martingale.
- (iii) If $X_0 = 0$, the sequence of stopping times $\tau_n = \inf\{t \geq 0 : |X_t| \geq n\}$ reduces X .
- (iv) If W is a \mathcal{F}_0 -measurable (real) random variable, then $(WX_t)_{t \geq 0}$ is also a continuous local martingale.

Proof. (i) Write $X_t = X_0 + Y_t$. By definition, there exists a sequence $(\tau_n)_{n=1}^\infty$ of stopping times reducing Y . Whenever $t > s \geq 0$, since $X_0 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, we have

$$Y_{s \wedge \tau_n} = \mathbb{E}[Y_{t \wedge \tau_n} | \mathcal{F}_s] \Rightarrow X_{s \wedge \tau_n} = \mathbb{E}[X_{t \wedge \tau_n} | \mathcal{F}_s] \quad (3.9)$$

Since X is nonnegative, by Fatou's lemma (conditional version), we have

$$X_s = \liminf_{n \rightarrow \infty} X_{s \wedge \tau_n} = \liminf_{n \rightarrow \infty} \mathbb{E}[X_{t \wedge \tau_n} | \mathcal{F}_s] \geq \mathbb{E}\left[\lim_{n \rightarrow \infty} X_{t \wedge \tau_n} | \mathcal{F}_s\right] = \mathbb{E}[X_t | \mathcal{F}_s].$$

(ii) Following (3.9), we use Lebesgue dominated convergence theorem, because $|X_{t \wedge \tau_n}| \leq Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ for all $n \in \mathbb{N}$ and $t \geq 0$. Then $X_{t \wedge \tau_n} \rightarrow X_t$ in L^1 , and $X_s = \mathbb{E}[X_t | \mathcal{F}_s]$ for $0 \leq s < t$.

(iii) By Proposition 3.13 (ii), $\tau_n = \inf\{t \geq 0 : X_t \geq n\}$ is indeed a stopping time. By (ii), the stopped process X^{τ_n} , bounded by n , is a uniformly integrable martingale for each $n \in \mathbb{N}$, and the result follows.

(iv) It suffices to show the case $X_0 = 0$. Choose the stopping times τ_n defined in (iii). Clearly, the process $(WX_{t \wedge \tau_n})_{t \geq 0}$ is adapted, and $|WX_{t \wedge \tau_n}| \leq n|W|$ is L^1 . Furthermore, since $(\tau_n)_{n=1}^\infty$ reduces $(X_t)_{t \geq 0}$, and W is \mathcal{F}_0 -measurable, we have $\mathbb{E}[WX_{t \wedge \tau_n} | \mathcal{F}_s] = W\mathbb{E}[X_{t \wedge \tau_n} | \mathcal{F}_s] = WX_{s \wedge \tau_n}$ for all $t > s \geq 0$. Hence $(\tau_n)_{n=1}^\infty$ also reduces $(WX_t)_{t \geq 0}$, and the conclusion follows. \square

Proposition 3.55. If $X = (X_t)_{t \geq 0}$ is both a continuous local martingale and a finite variation process with $X_0 = 0$, then there exists a negligible set N such that $X_t(\omega) = 0$ for all $t \in \mathbb{R}_+$ and all $\omega \in \Omega \setminus N$.

Proof. Since X is a finite variation process, $\int_0^t |dX_s|$ is an increasing process with continuous sample paths. For every $n \in \mathbb{N}$, define the stopping time

$$\tau_n = \inf \left\{ t \geq 0 : \int_0^t |dX_s| \geq n \right\},$$

and set $Y_t = X_t^{\tau_n}$. Then $Y_t \leq \int_0^{t \wedge \tau_n} |dX_s| \leq n$. By Proposition 3.54 (ii), Y is a uniformly integrable martingale. Let $0 = t_0 < t_1 < \dots < t_k = t$ be a partition of $[0, t]$. By Proposition 3.39, we have

$$\mathbb{E}[Y_t^2] = \sum_{j=1}^p \mathbb{E}\left[(Y_{t_j} - Y_{t_{j-1}})^2\right] \leq \mathbb{E}\left[\sup_{1 \leq j \leq k} |Y_{t_j} - Y_{t_{j-1}}| \sum_{j=1}^p |Y_{t_j} - Y_{t_{j-1}}|\right] \leq n \mathbb{E}\left[\sup_{1 \leq j \leq k} |Y_{t_j} - Y_{t_{j-1}}|\right].$$

Now we take a sequence of increasing partitions $0 = t_0^m < t_1^m < \dots < t_{k_m}^m = t$ of $[0, t]$ whose mesh converges to 0. By continuity of sample paths of Y and Lebesgue dominated convergence theorem, we have

$$\sup_{1 \leq j \leq k_m} |Y_{t_j^m}(\omega) - Y_{t_{j-1}^m}(\omega)| \rightarrow 0 \text{ as } m \rightarrow \infty, \forall \omega \in \Omega \Rightarrow \mathbb{E}\left[\sup_{1 \leq j \leq k_m} |Y_{t_j^m} - Y_{t_{j-1}^m}|\right] \rightarrow 0.$$

Note we are able to use dominated convergence theorem because Y is bounded by n . Hence $\mathbb{E}[Y_t^2] = 0$, and $X_t^{\tau_n} = 0$ a.s.. Letting $n \rightarrow \infty$, we then have $X_t = 0$ a.s. for all $t \in \mathbb{R}_+$. To show that $X(\omega) \equiv 0$ for a.s. $\omega \in \Omega$, we take a countable dense subset $D \subset \mathbb{R}_+$. Then $N = \{\omega \in \Omega : \exists t \in D, X_t(\omega) \neq 0\}$ is a negligible set. By continuity of sample paths of X , we have $X_t(\omega) = 0$ for all $t \in \mathbb{R}_+$ and all $\omega \in \Omega \setminus N$. \square

Remark. Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two stochastic processes. If there exists a negligible set $N \subset \Omega$ such that $X_t(\omega) = Y_t(\omega)$ for all $t \in \mathbb{R}_+$ and all $\omega \in \Omega \setminus N$, then X and Y are said to be *indistinguishable*. Note this is a stronger condition than *a.s. modification*.

From now on we assume completeness of the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Theorem 3.56 (Quadratic variation). Let $X = (X_t)_{t \geq 0}$ be a continuous local martingale. There exists an increasing process denoted by $(\langle X, X \rangle_t)_{t \geq 0}$, which is unique up to indistinguishability, such that $X_t^2 - \langle X, X \rangle_t$ is a continuous local martingale. Furthermore, for every fixed $t \in \mathbb{R}_{++}$, if $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ is an increasing sequence of partitions of $[0, t]$ with the mesh $\max_{1 \leq j \leq k_n} |t_j^n - t_{j-1}^n| \rightarrow 0$, then

$$\langle X, X \rangle_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \quad (3.10)$$

in probability. The process $\langle X, X \rangle$ is said to be the *quadratic variation* of X .

Proof. Step I: Let Y_t and Y'_t be two processes satisfying the conditions given in the statement. Then the process $Y'_t - Y_t = (X_t^2 - Y_t) - (X_t^2 - Y'_t)$ is both a finite variation process and a continuous local martingale. According to Proposition 3.55, $Y'_t - Y_t = 0$ *a.s.*, and the statement of uniqueness follows.

Step II: Now we prove existence. We first assume that $X_0 = 0$ and X is bounded. Hence X is a uniformly integrable martingale by Proposition 3.54 (ii). We fix $T > 0$ and an increasing sequence of partitions of $[0, T]$ with the mesh $\max_{1 \leq j \leq k_n} |t_j^n - t_{j-1}^n| \rightarrow 0$. The for every $s > r \geq 0$ and every bounded \mathcal{F}_r -measurable random variable Z , the process $(Z(X_{s \wedge t} - X_{r \wedge t}))_{t \geq 0}$ is adapted and L^1 , and for all $0 \leq t' < t$,

$$\mathbb{E}[Z(X_{s \wedge t} - X_{r \wedge t}) | \mathcal{F}_{t'}] = \begin{cases} \mathbb{E}[\mathbb{E}[Z(X_{s \wedge t} - X_{r \wedge t}) | \mathcal{F}_r] | \mathcal{F}_{t'}] = 0 & \text{if } t' < r \\ Z \mathbb{E}[X_{s \wedge t} - X_r | \mathcal{F}_{t'}] = Z(X_{s \wedge t'} - X_r) & \text{if } t' \geq r \end{cases}.$$

Hence $(Z(X_{s \wedge t} - X_{r \wedge t}))_{t \geq 0}$ is a bounded martingale. Following this, the process

$$M_t^n = \sum_{j=1}^{k_n} X_{t_{j-1}^n} \left(X_{t_j^n \wedge t} - X_{t_{j-1}^n \wedge t} \right), \quad \text{satisfying } X_{t_j^n}^2 - 2M_{t_j^n}^n = \sum_{i=1}^j \left(X_{t_i^n} - X_{t_{i-1}^n} \right)^2,$$

is also a bounded martingale.

Claim. $\lim_{n, m \rightarrow \infty} \mathbb{E}[(M_T^n - M_T^m)^2] = 0$.

Proof of the Claim. We fix $m \leq n$ and evaluate the product $\mathbb{E}[M_T^n M_T^m]$:

$$\begin{aligned} \mathbb{E}[M_T^n M_T^m] &= \sum_{i=1}^{k_m} \sum_{j=1}^{k_n} \mathbb{E} \left[X_{t_{i-1}^m} \left(X_{t_i^m} - X_{t_{i-1}^m} \right) X_{t_{j-1}^n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right) \right] \\ &= \sum_{j=1}^{k_n} \sum_{i: (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E} \left[X_{t_{i-1}^m} \left(X_{t_i^m} - X_{t_{i-1}^m} \right) X_{t_{j-1}^n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right) \right] \\ &= \sum_{j=1}^{k_n} \sum_{i: (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \sum_{l: (t_{l-1}^n, t_l^n] \subset (t_{i-1}^m, t_i^m]} \mathbb{E} \left[X_{t_{i-1}^m} \left(X_{t_l^n} - X_{t_{l-1}^n} \right) X_{t_{j-1}^n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right) \right] \\ &= \sum_{j=1}^{k_n} \sum_{i: (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E} \left[X_{t_{i-1}^m} X_{t_{j-1}^n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right] \end{aligned} \quad (3.11)$$

The second equality holds because once $t_{j-1}^n \geq t_i^m$ (resp. $t_j^n \leq t_{i-1}^m$), we can take conditional expectation with respect to $\mathcal{F}_{t_{j-1}^n}$ (resp. $\mathcal{F}_{t_{i-1}^m}$) to eliminate the corresponding term in the double sum. The fourth equality because once $l < j$ (resp. $l > j$), we can take conditional expectation with respect to $\mathcal{F}_{t_{j-1}^n}$ (resp. $\mathcal{F}_{t_{l-1}^n}$) to eliminate the corresponding term in the triple sum. As a special case of (3.11), we have

$$\mathbb{E}[(M_T^n)^2] = \sum_{j=1}^{k_n} \mathbb{E} \left[X_{t_{j-1}^n}^2 \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right]. \quad (3.12)$$

And by Proposition 3.39,

$$\begin{aligned} \mathbb{E}[(M_T^m)^2] &= \sum_{i=1}^{k_m} \mathbb{E} \left[X_{t_{i-1}^m}^2 \left(X_{t_i^m} - X_{t_{i-1}^m} \right)^2 \right] = \sum_{i=1}^{k_m} \mathbb{E} \left[X_{t_{i-1}^m}^2 \mathbb{E} \left[\left(X_{t_i^m} - X_{t_{i-1}^m} \right)^2 \middle| \mathcal{F}_{t_{i-1}^m} \right] \right] \\ &= \sum_{i=1}^{k_m} \sum_{j: (t_{j-1}^n, t_j^n] \subset (t_{i-1}^m, t_i^m]} \mathbb{E} \left[X_{t_{i-1}^m}^2 \mathbb{E} \left[\left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \middle| \mathcal{F}_{t_{i-1}^m} \right] \right] \\ &= \sum_{j=1}^{k_n} \sum_{i: (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E} \left[X_{t_{i-1}^m}^2 \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right]. \end{aligned} \quad (3.13)$$

Note that for every $j \in \{1, \dots, k_n\}$, there is a unique $i \in \{1, \dots, k_m\}$ such that $(t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]$. Combining (3.11), (3.12) and (3.13), we have

$$\begin{aligned} \mathbb{E}[(M_T^n - M_T^m)^2] &= \sum_{j=1}^{k_n} \sum_{i: (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E} \left[\left(X_{t_{j-1}^n} - X_{t_{i-1}^m} \right)^2 \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right] \\ &\leq \mathbb{E} \left[\sup_{1 \leq j \leq k_n, (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \left(X_{t_{j-1}^n} - X_{t_{i-1}^m} \right)^2 \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right] \\ &\leq \mathbb{E} \left[\sup_{1 \leq j \leq k_n, (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \left(X_{t_{j-1}^n} - X_{t_{i-1}^m} \right)^4 \right]^{\frac{1}{2}} \mathbb{E} \left[\left(\sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right)^2 \right]^{\frac{1}{2}} \end{aligned} \quad (3.14)$$

By continuity of the sample paths of X and the fact that X is bounded (so we can use dominated convergence theorem), the first term in (3.14) converges to 0 as $n, m \rightarrow \infty$. Hence our result follows if we can bound the second term with a finite constant independent of n . Suppose that $|X_t| \leq K$ for all $t \geq 0$. Then

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right)^2 \right] &= \sum_{j=1}^{k_n} \mathbb{E} \left[\left(X_{t_j^n} - X_{t_{j-1}^n} \right)^4 \right] + 2 \sum_{1 \leq j < i \leq k_n} \mathbb{E} \left[\left(X_{t_i^n} - X_{t_{i-1}^n} \right)^2 \left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right] \\ &\leq 4K^2 \sum_{j=1}^{k_n} \mathbb{E} \left[\left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \right] + 2 \sum_{j=1}^{k_n-1} \mathbb{E} \left[\left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \sum_{i=j+1}^{k_n} \mathbb{E} \left[\left(X_{t_i^n} - X_{t_{i-1}^n} \right)^2 \middle| \mathcal{F}_{t_j^n} \right] \right] \\ &= 4K^2 \mathbb{E} \left[(X_T - X_0)^2 \right] + 2 \sum_{j=1}^{k_n-1} \mathbb{E} \left[\left(X_{t_j^n} - X_{t_{j-1}^n} \right)^2 \mathbb{E} \left[(X_T - X_{t_j^n})^2 \middle| \mathcal{F}_{t_j^n} \right] \right] \quad (\text{By Proposition 3.39}) \\ &\leq 12K^2 \mathbb{E} \left[(X_T - X_0)^2 \right] \leq 48K^4. \end{aligned}$$

Hence we can bound the second term in (3.14) by $4\sqrt{3}K^2$, which completes the proof. \square

Proof of Theorem 3.56 (Cont). Step III: By Doob's L^p -inequality [Proposition 3.40 (ii)] and our claim,

$$0 \leq \lim_{n,m \rightarrow \infty} \mathbb{E} \left[\sup_{0 \leq t \leq T} (M_t^n - M_t^m)^2 \right] \leq \lim_{n,m \rightarrow \infty} 4\mathbb{E} \left[(M_T^n - M_T^m)^2 \right] = 0.$$

Hence for all $t \in [0, T]$, $(M_t^n)_{n=1}^\infty$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and thus converges in L^2 . We choose a subsequence $n_k \nearrow \infty$ such that

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} (M_t^{n_{k+1}} - M_t^{n_k})^2 \right] < 2^{-k}, \quad \forall k \in \mathbb{N}.$$

As a result,

$$\mathbb{E} \left[\sum_{k=1}^\infty \sup_{0 \leq t \leq T} |M_t^{n_{k+1}} - M_t^{n_k}| \right] < \sum_{k=1}^\infty 2^{-k} < \infty \quad \Rightarrow \quad \sum_{k=1}^\infty \sup_{0 \leq t \leq T} |M_t^{n_{k+1}} - M_t^{n_k}| < \infty \text{ a.s.}$$

Therefore, except on a negligible set N where the series in the above display diverges, the function sequence $t \mapsto M_t^{n_k}(\omega)$ converges uniformly on $[0, T]$ as $k \rightarrow \infty$. Let $Y_t(\omega) = \lim_{k \rightarrow \infty} M_t^{n_k}(\omega)$ for all $t \in [0, T]$ if $\omega \notin N$, and otherwise $Y_t(\omega) = 0$ for all $t \in [0, T]$. Then $(Y_t)_{t \geq 0}$ has continuous sample paths. Also, Y_t is adapted by completeness of our filtration $(\mathcal{F}_t)_{t \geq 0}$. Moreover, since $(M_t^n)_{n=1}^\infty$ converges in L^2 , it must converge to the a.s. limit Y_t in L^2 . Also, since the conditional expectation is a bounded linear operator in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, we can pass the martingale property of M_t^n to Y_t to obtain that $\mathbb{E}[Y_t | \mathcal{F}_s] = Y_s$ for all $0 \leq s < t \leq T$. Hence $(Y_{t \wedge T})_{t \geq 0}$ is a sample-continuous martingale.

Meanwhile, the process $X_t^2 - 2M_t^n$ restricted to the finite sequence $(t_j^n)_{j=1}^{k_n}$ is increasing. Take the limit $n_k \nearrow \infty$, we have $X_t^2 - 2M_t^{n_k} \Rightarrow X_t^2 - Y_t$ on $[0, T]$ except possibly on the negligible set N . Set $V_t^T = X_t^2 - 2Y_t$ on $\Omega \setminus N$, and $V_t^T = 0$ on N . Then $V_0^T = 0$, V_t^T is \mathcal{F}_t -measurable for all $t \in [0, T]$, and V^T has increasing continuous sample paths. Also, $X_{t \wedge T}^2 - V_{t \wedge T}^T$ is a sample-continuous martingale.

For every $T \in \mathbb{N}$, by the uniqueness argument proposed in *Step I*, we have $V_{t \wedge T}^T = V_{t \wedge T}^{T+1}$ a.s.. Hence we can define an increasing process $\langle X, X \rangle_t = V_t^T$ for all $t \in [0, T]$ and all $T \in \mathbb{N}$. Clearly, $X_t^2 - \langle X, X \rangle_t$ is a sample-continuous martingale. To obtain (3.10), note that $X_{t \wedge T}^2 - V_{t \wedge T}^T$ and $X_{t \wedge T}^2 - \langle X, X \rangle_{t \wedge T}$ are martingales. Again by the uniqueness argument, we have $V_{t \wedge T}^T = \langle X, X \rangle_{t \wedge T}$ a.s., and particularly, $V_T^T = \langle X, X \rangle_T$ a.s.. Note that $M_T^n \rightarrow Y_T = \frac{1}{2}(X_T^2 - V_T^T)$ in L^2 , we have

$$X_T^2 - 2M_T^n = \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n})^2 \xrightarrow{L^2} V_T^T = \langle X, X \rangle_T \text{ a.s.}$$

Then the proof for the case where $X_0 = 0$ and X is bounded is completed.

Step IV: If $X_0 = 0$, but X is not bounded, let $\tau_n = \inf\{t \geq 0 : |X_t| \geq n\}$. By Proposition 3.54 (iii), the stopped process X^{τ_n} is a bounded martingale, and we set $V^{[n]} = \langle X^{\tau_n}, X^{\tau_n} \rangle$. Again, the uniqueness argument shows that $V_t^{[n]}$ and $V_{t \wedge \tau_n}^{[n+1]}$ are indistinguishable. Then there exists an increasing and continuous process V such that $V_{t \wedge \tau_n} = V_t^{[n]}$ a.s. for all $t \geq 0$, and $X_{t \wedge \tau_n}^2 - V_{t \wedge \tau_n}$ is a martingale for every $n \in \mathbb{N}$. As a result, $X_t^2 - V_t$ is a continuous local martingale, and taking $\langle X, X \rangle_t = V_t$ suffices.

To obtain (3.10) (in probability), note that for all $\eta > 0$,

$$\mathbb{P} \left(\left| \sum_{j=1}^{k_m} (X_{t_j} - X_{t_{j-1}})^2 - \langle X, X \rangle_t \right| \geq \eta \right) \leq \frac{1}{\eta^2} \left\| \sum_{j=1}^{k_m} (X_{t_j}^{\tau_n} - X_{t_{j-1}}^{\tau_n})^2 - \langle X, X \rangle_{t \wedge \tau_n} \right\|_2^2 + \mathbb{P}(\tau_n < t). \quad (3.15)$$

In (3.15), the first term converges to 0 as $m \rightarrow \infty$, because (3.10) holds in L^2 when we replace X and $\langle X, X \rangle_t$ by X^{τ_n} and $\langle X, X \rangle_{t \wedge \tau_n}$, respectively. Also, the second term converges to 0 as $n \rightarrow \infty$ by definition.

Step V: For the general case, we write $X_t = X_0 + Z_t$, so $X_t^2 = X_0^2 + 2X_0Z_t + Z_t^2$. By Proposition 3.54 (iv), the process X_0Z_t is also a continuous local martingale. Hence $X_t^2 - \langle Z, Z \rangle_t$ remains a continuous local martingale. Meanwhile, (3.10) does not change by adding a \mathcal{F}_0 -measurable variable X_0 . \square

Remark. According to our proof in *Step V*, the quadratic variation of a continuous local martingale $X = (X_t)_{t \geq 0}$ does not depend on the initial value X_0 , i.e. if we write $X_t = X_0 + Z_t$, then we have $\langle X, X \rangle_t = \langle Z, Z \rangle_t$.

Proposition 3.57. Let $X = (X_t)_{t \geq 0}$ be a continuous local martingale.

- (i) If τ is a stopping time, then $\langle X^\tau, X^\tau \rangle_t = \langle X, X \rangle_{t \wedge \tau}$.
- (ii) Assume $X_0 = 0$. Then $\langle X, X \rangle = 0$ if and only if $X = 0$ a.s..

Proof. (i) Since the stopped process $X_{t \wedge \tau}^2 - \langle X, X \rangle_{t \wedge \tau}$ is a continuous local martingale, the result follows.

(ii) Assume $\langle X, X \rangle_t = 0$ for all $t \geq 0$. Then $X_t^2 - 0$ is a nonnegative continuous local martingale, hence a supermartingale by Proposition 3.54. This implies $\mathbb{E}[X_t^2] \leq \mathbb{E}[X_0^2] = 0$, and $X_t = 0$ a.s.. To prove that $X = 0$ a.s., take the intersection of $\{X_t = 0, t \in D\}$ for a dense set $D \subset \mathbb{R}$, then use sample-path continuity. \square

Theorem 3.58. Let $X = (X_t)_{t \geq 0}$ be a continuous local martingale such that $X_0 \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:

- (i) X is a martingale, and $\sup_{t \geq 0} \mathbb{E}[|X_t|^2] < \infty$.
- (ii) $\mathbb{E}[\langle X, X \rangle_\infty] < \infty$.

Furthermore, if these properties hold, then $X_t^2 - \langle X, X \rangle_t$ is a uniformly integrable martingale, and in particular we have $\mathbb{E}[X_\infty^2] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X, X \rangle_\infty]$.

Proof. Without loss of generality let $X_0 = 0$.

(i) \Rightarrow (ii): By Doob's L^p -inequality [Proposition 3.40] and monotone convergence theorem, we have

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t|^2 \right] \leq 4\mathbb{E}[|X_T|^2], \quad \forall T \in \mathbb{R}_+ \quad \Rightarrow \quad \mathbb{E} \left[\sup_{t \geq 0} |X_t|^2 \right] \leq 4 \sup_{t \geq 0} \mathbb{E}[|X_t|^2] < \infty$$

Let $\sigma_n = \inf\{t \geq 0 : \langle X, X \rangle_t \geq n\} \nearrow \infty$. Then the continuous local martingale $X_{t \wedge \sigma_n}^2 - \langle X, X \rangle_{t \wedge \sigma_n}$ is dominated by the integrable variable $\sup_{t \geq 0} |X_t|^2 + n$. According to Proposition 3.54 (ii), this is a uniformly integrable martingale, and

$$\mathbb{E}[\langle X, X \rangle_{t \wedge \sigma_n}] = \mathbb{E}[X_{t \wedge \sigma_n}^2] \leq \mathbb{E} \left[\sup_{t \geq 0} |X_t|^2 \right] \leq 4 \sup_{t \geq 0} \mathbb{E}[|X_t|^2].$$

Let $n \rightarrow \infty$ and $t \rightarrow \infty$, we have $\mathbb{E}[\langle X, X \rangle_\infty] \leq 4 \sup_{t \geq 0} \mathbb{E}[|X_t|^2] < \infty$ by monotone convergence theorem.

(ii) \Rightarrow (i): Let $\tau_n = \{t \geq 0 : |X_t| \geq n\}$. Then the continuous local martingale $X_{t \wedge \tau_n}^2 - \langle X, X \rangle_{t \wedge \tau_n}$ is dominated by the integrable variable $\langle X, X \rangle_\infty + n^2$. According to Proposition 3.54 (ii), this is a uniformly integrable martingale. By Fatou's lemma, we have

$$\mathbb{E}[X_{t \wedge \tau_n}^2] = \mathbb{E}[\langle X, X \rangle_{t \wedge \tau_n}] \leq \mathbb{E}[\langle X, X \rangle_\infty] < \infty \quad \Rightarrow \quad \mathbb{E}[X_t^2] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_{t \wedge \tau_n}^2] \leq \mathbb{E}[\langle X, X \rangle_\infty] < \infty.$$

Meanwhile, the sequence $|X_{t \wedge \tau_n}| \nearrow |X_t|$ as $n \rightarrow \infty$, and $(X_{t \wedge \tau_n})_{n=1}^\infty$ is uniformly integrable:

$$\begin{aligned} \lim_{M \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}[|X_{t \wedge \tau_n}| \mathbb{1}_{\{|X_{t \wedge \tau_n}| \geq M\}}] &\leq \lim_{M \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}[X_{t \wedge \tau_n}^2] \mathbb{E}[\mathbb{1}_{\{|X_{t \wedge \tau_n}| \geq M\}}] \\ &\leq \mathbb{E}[\langle X, X \rangle_\infty] \lim_{M \rightarrow \infty} \mathbb{E}[\mathbb{1}_{\{|X_t| \geq M\}}] = 0. \end{aligned}$$

As a result, $X_{t \wedge \tau_n} \rightarrow X_t$ a.s. and in L^1 . By Proposition 3.54 (iii), $(X_{t \wedge \tau_n})_{t \geq 0}$ is a martingale, and we have $\mathbb{E}[X_{t \wedge \tau_n} | \mathcal{F}_s] = X_{s \wedge \tau_n}$ for all $t > s \geq 0$. Convergence in L^1 implies $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$, hence X is a martingale.

Finally, if (i) and (ii) hold, the continuous local martingale $X_t^2 - \langle X, X \rangle_t$ is dominated by the integrable variable $\langle X, X \rangle_\infty + \sup_{t \geq 0} |X_t|^2$. By Proposition 3.54 (ii), this is a uniformly integrable martingale. \square

The following corollary is derived by applying Theorem 3.58 on $(X_{t \wedge T})_{t \geq 0}$ for each $T \geq 0$.

Corollary 3.59. Let $X = (X_t)_{t \geq 0}$ be a continuous local martingale such that $X_0 \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:

- (i) X is a martingale, and $X_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in \mathbb{R}_+$.
- (ii) $\mathbb{E}[\langle X, X \rangle_t] < \infty$ for all $t \in \mathbb{R}_+$.

Furthermore, if these properties hold, then $X_t^2 - \langle X, X \rangle_t$ is a martingale.

Definition 3.60 (Bracket). Given $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are two continuous local martingales, their bracket $\langle X, Y \rangle$ is defined as the following finite variation process:

$$\langle X, Y \rangle_t = \frac{1}{2} (\langle X + Y, X + Y \rangle_t - \langle X, X \rangle_t - \langle Y, Y \rangle_t), \quad \forall t \geq 0.$$

Proposition 3.61. Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two continuous local martingales.

- (i) $\langle X, Y \rangle$ is the unique (up to indistinguishability) finite variation process such that $X_t Y_t - \langle X, Y \rangle_t$ is a continuous local martingale.
- (ii) The mapping $(X, Y) \mapsto \langle X, Y \rangle$ is bilinear and symmetric.
- (iii) For any increasing sequence of partitions $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ of $[0, t]$ with mesh tending to 0,

$$\langle X, Y \rangle_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n}) \quad \text{in probability.}$$

- (iv) For every stopping time τ , $\langle X^\tau, Y^\tau \rangle_t = \langle X^\tau, Y \rangle_t = \langle X, Y \rangle_{t \wedge \tau}$.
- (v) For every stopping time τ , $X^\tau(Y - Y^\tau)$ is a continuous martingale.
- (vi) If X and Y are two L^2 -bounded sample-continuous martingales, $X_t Y_t - \langle X, Y \rangle_t$ is a uniformly integrable martingale. Consequently, $\langle X, Y \rangle_\infty$ is well-defined as the a.s. limit of $\langle X, Y \rangle_t$ as $t \rightarrow \infty$, and $\mathbb{E}[X_\infty Y_\infty] = \mathbb{E}[X_0 Y_0] + \mathbb{E}[\langle X, Y \rangle_\infty]$.
- (vii) $\langle X, Y \rangle_t = 0$ a.s. for all $t \geq 0$ if and only if XY is a continuous local martingale. In this case, the two continuous local martingales X and Y are said to be *orthogonal*.

Proof. (i) Clearly, $X_t Y_t - \langle X, Y \rangle_t$ is a continuous local martingale, and the uniqueness argument is similar to Theorem 3.56. (ii) is a consequence of the uniqueness argument. (iii) follows from (3.10). (v) is a consequence of (iv), since $X_t^\tau(Y_t - Y_t^\tau) = X_t^\tau Y_t - \langle X^\tau, Y \rangle_t - (X_t^\tau Y_t^\tau - \langle X^\tau, Y^\tau \rangle_t)$. For (iv), according to (iii), we have

$$\begin{aligned} \langle X^\tau, Y^\tau \rangle_t &= \langle X^\tau, Y \rangle_t = \langle X, Y \rangle_t \quad \text{on } \{\tau \geq t\}, \\ \langle X^\tau, Y^\tau \rangle_t - \langle X^\tau, Y^\tau \rangle_\tau &= \langle X^\tau, Y \rangle_t - \langle X^\tau, Y \rangle_\tau = 0 \quad \text{on } \{\tau < t\}. \end{aligned}$$

(vi) is a consequence of Theorem 3.58.

(vii) If XY is a continuous local martingale, so is $\langle X, Y \rangle = XY - (XY - \langle X, Y \rangle)$, which is also a finite variation process. Conversely, if $\langle X, Y \rangle_t = 0$ a.s. for all $t \geq 0$, then $XY = (XY - \langle X, Y \rangle) + \langle X, Y \rangle$ is a continuous local martingale. \square

Theorem 3.62 (Kunita-Watanabe). Let X and Y be two continuous local martingales, and let H and K be two measurable processes. Then

$$\int_0^\infty |H_s| |K_s| |d\langle X, Y \rangle_s| \leq \left(\int_0^\infty H_s^2 d\langle X, X \rangle_s \right)^{1/2} \left(\int_0^\infty K_s^2 d\langle Y, Y \rangle_s \right)^{1/2}.$$

Proof. Given $t > s \geq 0$, we abuse the notation $\langle X, Y \rangle_s^t = \langle X, Y \rangle_t - \langle X, Y \rangle_s$. Let $s = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ be a increasing sequence of partitions of $[s, t]$ with the mesh tending to 0. Let $S_{XX}^n = \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n})^2$, $S_{YY}^n = \sum_{j=1}^{k_n} (Y_{t_j^n} - Y_{t_{j-1}^n})^2$, and $S_{XY}^n = \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n})$. By Cauchy-Schwarz inequality, we have $S_{XY}^n \leq \sqrt{S_{XX}^n S_{YY}^n}$. Note that $\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} \xrightarrow{\mathbb{P}} S_{XY}^n - \sqrt{S_{XX}^n S_{YY}^n}$. For all $\eta > 0$,

$$\mathbb{P} \left(\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} > \eta \right) \leq \mathbb{P} \left(\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} - S_{XY}^n + \sqrt{S_{XX}^n S_{YY}^n} > \eta \right) \rightarrow 0.$$

By taking the union of all rationals $\eta > 0$, we have $\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} \leq 0$ a.s.. Since $t > s \geq 0$ are arbitrary, $\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} \leq 0$ holds for all rationals $t > s \geq 0$ for a.s. $\omega \in \Omega$. By continuity of X and Y , we have for a.s. $\omega \in \Omega$ that $\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} \leq 0$ for all reals $t > s \geq 0$.

Now we fix $\omega \in \Omega$ with $\langle X, Y \rangle_s^t - \sqrt{\langle X, X \rangle_s^t \langle Y, Y \rangle_s^t} \leq 0$ for all reals $t > s \geq 0$. Then all remaining results are deterministic. For any subdivisions $s = t_0 < t_1 < \dots < t_k = t$, we have

$$\sum_{j=1}^k |\langle X, Y \rangle_{t_{j-1}}^{t_j}| \leq \sum_{j=1}^k \sqrt{\langle X, X \rangle_{t_{j-1}}^{t_j}} \sqrt{\langle Y, Y \rangle_{t_{j-1}}^{t_j}} \leq \sqrt{\sum_{j=1}^k \langle X, X \rangle_{t_{j-1}}^{t_j}} \sqrt{\sum_{j=1}^k \langle Y, Y \rangle_{t_{j-1}}^{t_j}} = \sqrt{\langle X, X \rangle_s^t} \sqrt{\langle Y, Y \rangle_s^t}.$$

Let the mesh of our partition tends to 0, we obtain

$$\int_s^t |d\langle X, Y \rangle_u| \leq \sqrt{\langle X, X \rangle_s^t} \sqrt{\langle Y, Y \rangle_s^t}$$

Fix $T > 0$, and let \mathcal{M}_T be the collection of all $A \in \mathcal{B}([0, T])$ such that

$$\int_A |d\langle X, Y \rangle_u| \leq \sqrt{\int_A d\langle X, X \rangle_u} \sqrt{\int_A d\langle Y, Y \rangle_u}. \quad (3.16)$$

By monotone convergence theorem, \mathcal{M}_T is a monotone class, and it contains the collection of all finite intersections of closed intervals in $[0, T]$, which is an algebra. By monotone class theorem [Theorem 1.11], we have $\mathcal{M}_T = \mathcal{B}([0, T])$. As a result, (3.16) holds for all bounded Borel sets $A \in \mathcal{B}(\mathbb{R}_+)$. Also, for all nonnegative simple functions h, k on $[0, T]$, choose finite partition A_1, \dots, A_m of $[0, T]$ such that $h = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$ and $k = \sum_{i=1}^m \beta_i \mathbb{1}_{A_i}$. Then we have

$$\begin{aligned} \int h(s)k(s) |d\langle X, Y \rangle_s| &= \sum_{i=1}^m \alpha_i \beta_i \int_{A_i} |d\langle X, Y \rangle_s| \leq \sqrt{\sum_{i=1}^m \alpha_i^2 \int_{A_i} d\langle X, X \rangle_u} \sqrt{\sum_{i=1}^m \beta_i^2 \int_{A_i} d\langle Y, Y \rangle_u} \\ &= \sqrt{\int h(s)^2 d\langle X, X \rangle_s} \sqrt{\int k(s)^2 d\langle Y, Y \rangle_s} \end{aligned}$$

Note that every nonnegative measurable function on $[0, T]$ is the limit of an increasing sequence of nonnegative simple functions $[0, T]$, and every nonnegative measurable function h on \mathbb{R}_+ is the increasing limit of $h \mathbb{1}_{[0, T]}$ as $T \rightarrow \infty$. Hence an application of monotone convergence theorem finishes the proof. \square

Definition 3.63 (Continuous semimartingales). A process $X = (X_t)_{t \geq 0}$ is said to be a *continuous semimartingale* if it can be written as

$$X_t = M_t + A_t, \quad \forall t \in \mathbb{R}_+, \quad (3.17)$$

where $M = (M_t)_{t \geq 0}$ is a continuous local martingale and $A = (A_t)_{t \geq 0}$ is a continuous finite variation process.

Remark. Thanks to Proposition 3.55, the decomposition (3.17) is unique up to indistinguishability. We call this the *canonical decomposition* of a continuous semimartingale X .

Definition 3.64 (Brackets). Given two continuous semimartingale $X = M + A$ and $Y = M' + A'$ (which are canonical decompositions), we define the bracket $\langle X, Y \rangle = \langle M, M' \rangle$, which is a finite variation process.

Proposition 3.65. Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two continuous semimartingales. Let $t > 0$. Let $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ be any increasing sequence of partitions of $[0, t]$ whose mesh tends to 0. Then

$$\langle X, Y \rangle_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n}) (Y_{t_j^n} - Y_{t_{j-1}^n}) \text{ in probability.}$$

Proof. Let $X = M + A$ and $Y = M' + A'$ be the canonical decompositions. Then

$$\begin{aligned} \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n}) (Y_{t_j^n} - Y_{t_{j-1}^n}) &= \sum_{j=1}^{k_n} (M_{t_j^n} - M_{t_{j-1}^n}) (M'_{t_j^n} - M'_{t_{j-1}^n}) + \sum_{j=1}^{k_n} (M_{t_j^n} - M_{t_{j-1}^n}) (A'_{t_j^n} - A'_{t_{j-1}^n}) \\ &\quad + \sum_{j=1}^{k_n} (A_{t_j^n} - A_{t_{j-1}^n}) (M'_{t_j^n} - M'_{t_{j-1}^n}) + \sum_{j=1}^{k_n} (A_{t_j^n} - A_{t_{j-1}^n}) (A'_{t_j^n} - A'_{t_{j-1}^n}) \end{aligned}$$

According to Proposition 3.61 (iii),

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} (M_{t_j^n} - M_{t_{j-1}^n}) (M'_{t_j^n} - M'_{t_{j-1}^n}) = \langle M, M' \rangle_t = \langle X, Y \rangle_t \text{ in probability.}$$

Also, note that

$$\begin{aligned} \left| \sum_{j=1}^{k_n} (M_{t_j^n} - M_{t_{j-1}^n}) (A'_{t_j^n} - A'_{t_{j-1}^n}) \right| &\leq \left(\int_0^t |dA'_s| \right) \sup_{1 \leq j \leq k_n} |M_{t_j^n} - M_{t_{j-1}^n}| \rightarrow 0, \text{ a.s.,} \\ \left| \sum_{j=1}^{k_n} (A_{t_j^n} - A_{t_{j-1}^n}) (M'_{t_j^n} - M'_{t_{j-1}^n}) \right| &\leq \left(\int_0^t |dA_s| \right) \sup_{1 \leq j \leq k_n} |M'_{t_j^n} - M'_{t_{j-1}^n}| \rightarrow 0, \text{ a.s.,} \\ \left| \sum_{j=1}^{k_n} (A_{t_j^n} - A_{t_{j-1}^n}) (A'_{t_j^n} - A'_{t_{j-1}^n}) \right| &\leq \left(\int_0^t |dA_s| \right) \sup_{1 \leq j \leq k_n} |A'_{t_j^n} - A'_{t_{j-1}^n}| \rightarrow 0, \text{ a.s.,} \end{aligned}$$

where the *a.s.* convergence holds by continuity of sample paths of M , M' and A' . □

4 Brownian Motions

4.1 Pre-Brownian Motions and Brownian Motions

Definition 4.1 (Gaussian spaces). A (centered) *Gaussian space* H is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ that contains only centered Gaussian variables.

Remark. To justify the closedness of a Gaussian space $H \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, we let $H \ni X_n \sim N(0, \sigma_n^2) \xrightarrow{L^2} X$. Convergence in L^2 implies $\mathbb{E}[X_n^2] = \sigma_n^2 \rightarrow \sigma^2$. Then for all $\lambda \in \mathbb{R}$, by dominated convergence theorem,

$$\mathbb{E}[e^{i\lambda X}] = \lim_{n \rightarrow \infty} \mathbb{E}[e^{i\lambda X_n}] = \lim_{n \rightarrow \infty} \exp\left(-\frac{\sigma_n^2}{2}\lambda^2\right) = \exp\left(-\frac{\sigma^2}{2}\lambda^2\right).$$

Hence $X \sim N(\mu, \sigma^2) \in H$. Furthermore, since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a complete space, so is a Gaussian subspace.

We can make H a Hilbert space by define the inner product $\langle X, Y \rangle = \mathbb{E}[XY]$ for $X, Y \in H$. In this space, orthogonality and independence are equivalent. To be specific, in the Gaussian space H , two variables X and Y are independent if and only if they are orthogonal, i.e. $\mathbb{E}[XY] = 0$. To see the “if” case, note that X, Y are jointly Gaussian. Then for all $s, t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[e^{i(sX+tY)}] &= \exp\left(-\frac{s^2\mathbb{E}[X^2] + 2st\mathbb{E}[XY] + t^2\mathbb{E}[Y^2]}{2}\right) \\ &= \exp\left(-\frac{s^2\mathbb{E}[X^2]}{2}\right) \exp\left(-\frac{t^2\mathbb{E}[Y^2]}{2}\right) = \mathbb{E}[e^{isX}] \mathbb{E}[e^{itY}]. \end{aligned}$$

By Corollary 2.44, X and Y are independent.

Likewise, assume that G, K are two subspaces of the Gaussian space H . Then $G \perp K$ if and only if the sub σ -algebras $\sigma(G)$ and $\sigma(K)$ generated by G and K are independent.

We also point out the equivalence between orthogonal projection onto a Gaussian space and conditional expectation. If H is a Gaussian space, and G is a closed subspace of H , then for all $X \in H$, the conditional expectation $\mathbb{E}[X|\sigma(G)]$ is the projection of X onto G . To see this, let ξ be the orthogonal projection of X onto G , so that $X - \xi \perp G$. As a result, $\mathbb{E}[X|\sigma(G)] = \mathbb{E}[\xi + (X - \xi)|\sigma(G)] = \mathbb{E}[\xi|\sigma(G)] = \xi$.

Definition 4.2 (Gaussian white noise). Let (E, \mathcal{E}) be a measurable space, and let μ be a σ -finite measure on (E, \mathcal{E}) . A *Gaussian white noise with intensity* μ is an isometry W from $L^2(E, \mathcal{E}, \mu)$ into a Gaussian space.

Remark I. According to the polarization identity, an isometry W also preserves inner product. Therefore, if $f, g \in L^2(E, \mathcal{E}, \mu)$, then we have

$$\mathbb{E}[W(f)W(g)] = \langle f, g \rangle = \int fg \, d\mu, \quad \text{and in particular } \mathbb{E}[W(f)^2] = \|f\|_2^2 = \int |f|^2 \, d\mu.$$

If $f = \mathbb{1}_A$ with $\mu(A) < \infty$, we write $W(A) = W(\mathbb{1}_A)$, and $W(A) \sim N(0, \mu(A))$.

Remark II. Given any σ -finite measure μ on (E, \mathcal{E}) , we can always find a Gaussian white noise with intensity μ on an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{e_\lambda, \lambda \in \Lambda\}$ be an orthonormal basis of $L^2(E, \mathcal{E}, \mu)$. According to Corollary 4.15, we define $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^\Lambda, \mathcal{B}(\mathbb{R})^{\otimes \Lambda}, \mathbb{G})$, where \mathbb{G} extends Gaussian measures

$$\mathbb{G}_{(t_1, \dots, t_n)}(A) = \frac{1}{(2\pi)^{n/2}} \int_A e^{-\frac{1}{2}(z_1^2 + \dots + z_n^2)} \, dz, \quad \forall t_1, \dots, t_n \in \mathcal{T}, \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

Then the coordinate maps $(\pi_\lambda)_{\lambda \in \Lambda}$ is a collection of independent standard Gaussian variables.

Now for every $f \in L^2(E, \mathcal{E}, \mu)$, we define

$$W(f) = \sum_{\lambda \in \Lambda} \langle f, e_\lambda \rangle \pi_\lambda.$$

This series converges in L^2 since $\{e_\lambda, \lambda \in \Lambda\}$ is an orthonormal basis of $L^2(E, \mathcal{E}, \mu)$. Hence W takes values in the Gaussian space $H = \overline{\text{span}}(\pi_\lambda)_{\lambda \in \Lambda}$. Since W maps an orthonormal basis in $L^2(E, \mathcal{E}, \mu)$ to one in H , it is an isometry. Thus we find a Gaussian white noise W with intensity μ .

Remark III. Given a measurable set A in (E, \mathcal{E}, μ) with $\mu(A) < \infty$, we can approximate $\mu(A)$ with a Gaussian white noise W with intensity μ . Let $A = A_1^n \amalg \cdots \amalg A_{k_n}^n$ be a sequence of partitions of A such that

$$\lim_{n \rightarrow \infty} \left(\max_{j \in \{1, \dots, k_n\}} \mu(A_j^n) \right) = 0.$$

Then $W(A_j^n)$, $j = 1, \dots, k_n$ are independent Gaussian variables, and $\mathbb{E}[W(A_j^n)^2] = \mu(A_j^n)$. Furthermore,

$$\mathbb{E} \left[\left(\sum_{j=1}^{k_n} W(A_j^n)^2 - \mu(A) \right)^2 \right] = \sum_{j=1}^{k_n} \mathbb{E} \left[(W(A_j^n)^2 - \mu(A_j^n))^2 \right] = 2 \sum_{j=1}^{k_n} \mu(A_j^n)^2 \leq 2\mu(A) \max_{1 \leq j \leq k_n} \mu(A_j^n) \rightarrow 0.$$

This implies

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} W(A_j^n)^2 = \mu(A) \quad \text{in } L^2.$$

Now we introduce pre-Brownian motion.

Definition 4.3 (Pre-Brownian motion). Give \mathbb{R}_+ the Borel σ -algebra $\mathcal{B}(\mathbb{R}_+)$ and the Lebesgue measure m , and let W be a Gaussian white noise on \mathbb{R}_+ with intensity m . The process $(B_t)_{t \geq 0}$ defined by

$$B_t = W(\mathbb{1}_{[0,t]}), \quad \forall t \in \mathbb{R}_+$$

is said to be a *pre-Brownian motion*.

Remark. By definition, a pre-Brownian motion $B = (B_t)_{t \geq 0}$ is a Gaussian process, i.e. the linear combination of any finitely many observations B_{t_1}, \dots, B_{t_n} is Gaussian. The covariance function of this process is given by

$$K(s, t) = \mathbb{E}[B_s B_t] = \int \mathbb{1}_{[0,s] \cap [0,t]} dm = s \wedge t, \quad \forall s, t \in \mathbb{R}_+.$$

Proposition 4.4 (Characterization of pre-Brownian motions). Let $B = (B_t)_{t \geq 0}$ be a (real-valued) stochastic process. The following are equivalent:

- (i) $(B_t)_{t \geq 0}$ is a pre-Brownian motion.
- (ii) $(B_t)_{t \geq 0}$ is a centered Gaussian process with covariance $K(s, t) = s \wedge t$.
- (iii) $B_0 = 0$ a.s., and for every $t > s \geq 0$, the random variable $B_t - B_s$ is independent of $\sigma(B_r, r \in [0, s])$ and distributed according to $N(0, t - s)$.
- (iv) $B_0 = 0$ a.s., and for every choice of $0 = t_0 < t_1 < \dots < t_n$, the variables $\{B_{t_j} - B_{t_{j-1}}, j = 1, \dots, n\}$ are independent, and for every $j = 1, \dots, p$, the variable $B_{t_j} - B_{t_{j-1}}$ is distributed according to $N(0, t_j - t_{j-1})$.

Proof. The facts that (i) \Rightarrow (ii) and that (iii) \Rightarrow (iv) are clear. To prove (ii) \Rightarrow (iii), let H be the Gaussian space spanned by $\{B_r, r \in [0, s]\}$ and X_t . Then $B_t - B_s \in H$ is a centered Gaussian variable, and

$$\begin{aligned}\mathbb{E}[(B_t - B_s)^2] &= t - 2(s \wedge t) + s = t - s, \\ \mathbb{E}[(B_t - B_s)B_r] &= t \wedge r - s \wedge r = r - r = 0, \quad \forall r \in [0, s].\end{aligned}$$

Hence $X_t - X_s \sim N(0, t - s)$, and X_t is independent of all X_r with $r \in [0, s]$.

To prove (iii) \Rightarrow (iv), it suffices to show that there exists an isometry W between $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ and a Gaussian space H . For all step functions $f = \sum_{j=1}^n \lambda_j \mathbf{1}_{(t_{j-1}, t_j]}$ in $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, define

$$W(f) = \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}).$$

If $f, g \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ are two step functions, we can find a partition $0 = t_0 < t_1 < \dots < t_n$ such that $f = \sum_{j=1}^n \lambda_j \mathbf{1}_{(t_{j-1}, t_j]}$ and $g = \sum_{j=1}^n \nu_j \mathbf{1}_{(t_{j-1}, t_j]}$. According to (iii), we have

$$\mathbb{E}[W(f)W(g)] = \mathbb{E}\left[\sum_{j=1}^n \sum_{k=1}^n \lambda_j \nu_k (B_{t_j} - B_{t_{j-1}})(B_{t_k} - B_{t_{k-1}})\right] = \sum_{j=1}^n \lambda_j \nu_j (t_j - t_{j-1}) = \int fg \, dm.$$

Therefore W is an isometry from the vector space of all step functions in $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ into the Gaussian space spanned by $\{B_t, t \in \mathbb{R}_+\}$. Since the step functions are dense in $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, we immediately extend W to an isometry between $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ and $\text{span}\{B_t, t \geq 0\}$. \square

Remark. According to our proof of (iii) \Rightarrow (iv), we can determine a Gaussian noise W with intensity m given a pre-Brownian motion $B = (B_t)_{t \geq 0}$. For all $f \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, we write the notation

$$W(f) = \int_0^\infty f(s) \, dB_s, \quad W(f \mathbf{1}_{[0, t]}) = \int_0^t f(s) \, dB_s, \quad \text{and} \quad W(f \mathbf{1}_{(s, t]}) = \int_s^t f(r) \, dB_r, \quad \forall t > s \geq 0.$$

The mapping $W : f \mapsto \int_0^\infty f(s) \, dB_s$ is called the *Wiener integral* with respect to B . Clearly, we have $W(f) \sim N(0, \int_0^\infty |f|^2 \, dm)$.

Proposition 4.5. Let $B = (B_t)_{t \geq 0}$ be a pre-Brownian motion. The following statements are true:

- (i) (Symmetry). $-B$ is also a pre-Brownian motion.
- (ii) (Scale invariance). For all $\lambda > 0$, the process $B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$ is a pre-Brownian motion.
- (iii) (Simple Markov property). For all $s \geq 0$, the process $B_t^{(s)} = B_t - B_s$ is a pre-Brownian motion that is independent of $\sigma(B_r, r \in [0, s])$.
- (iv) (Time inversion). The process \hat{B} defined by $\hat{B}_0 = 0$ and $\hat{B}_t = t B_{\frac{1}{t}}$ is (indistinguishably) a pre-Brownian motion.

Proof. The statement (i) is clear. (ii) follows from Proposition 4.4 (iv). For (iii), Proposition 4.4 (iv) implies that $B_t^{(s)}$ is a pre-Brownian motion, and the independence argument follows from Proposition 4.4 (iii). The statement (iv) follows from Proposition 4.4 (ii). \square

Before introducing Brownian motions, we first discuss continuity of sample paths of a stochastic process.

Theorem 4.6 (Kolmogorov's lemma). Let $X = (X_t)_{t \in I}$ be a stochastic process indexed by a bounded interval $I \subset \mathbb{R}$ and taking values in a complete metric space (E, d) . Assume there exist $\epsilon, q, C > 0$ such that

$$\mathbb{E}[d(X_s, X_t)^q] \leq C|t - s|^{1+\epsilon}, \quad \forall s, t \in I.$$

Then there exists an *a.s.* modification \tilde{X} of X whose sample paths are Hölder continuous of exponent α for all $\alpha \in (0, \frac{\epsilon}{q})$. That is, for each $\omega \in \Omega$ and each $\alpha \in (0, \frac{\epsilon}{q})$, there exists $C_\alpha(\omega) > 0$ such that

$$d(\tilde{X}_s(\omega), \tilde{X}_t(\omega)) \leq C_\alpha(\omega)|t - s|^\alpha, \quad \forall s, t \in I.$$

Proof. Without loss of generality, we take $I = [0, 1]$ and fix $\alpha \in (0, \frac{\epsilon}{q})$. Then for all $\eta > 0$,

$$\mathbb{P}(d(X_s, X_t) \geq \eta) \leq \eta^{-q} \mathbb{E}[d(X_s, X_t)^q] \leq C\eta^{-q}|t - s|^{1+\epsilon}, \quad \forall s, t \in I.$$

Take $s = (j-1)2^{-n}$, $t = j2^{-n}$ and $\eta = 2^{-n\alpha}$, where $n \in \mathbb{N}$ and $j \in \{1, \dots, 2^n\}$. Then

$$\mathbb{P}(d(X_{(j-1)2^{-n}}, X_{j2^{-n}}) \geq 2^{-n\alpha}) \leq C2^{-n(1+\epsilon-\alpha q)}, \quad \forall n \in \mathbb{N}, \forall j \in \{1, \dots, 2^n\}.$$

By summing over j , we obtain

$$\mathbb{P}\left(\bigcup_{j=1}^{2^n} \{d(X_{(j-1)2^{-n}}, X_{j2^{-n}}) \geq 2^{-n\alpha}\}\right) \leq C2^{-n(\epsilon-\alpha q)}, \quad \forall n \in \mathbb{N}.$$

Since $\epsilon - \alpha q > 0$, we can also sum over n and get

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\bigcup_{j=1}^{2^n} \{d(X_{(j-1)2^{-n}}, X_{j2^{-n}}) \geq 2^{-n\alpha}\}\right) < \infty.$$

According to Borel-Cantelli lemma, for *a.s.* $\omega \in \Omega$, there exists $n_0(\omega) \in \mathbb{N}$ such that

$$d(X_{(j-1)2^{-n}}(\omega), X_{j2^{-n}}(\omega)) < 2^{-n\alpha}, \quad \forall n \geq n_0(\omega), \forall j \in \{1, \dots, 2^n\}.$$

As a result, for *a.s.* $\omega \in \Omega$, we have

$$K_\alpha(\omega) = \sup_{n \in \mathbb{N}} \sup_{1 \leq j \leq 2^n} \frac{d(X_{(j-1)2^{-n}}(\omega), X_{j2^{-n}}(\omega))}{2^{-n\alpha}} < \infty.$$

Now we take $D = \{j2^{-n} : n \in \mathbb{N}, j \in \{1, \dots, 2^n\}\}$, which is a countable dense subset of $I = [0, 1]$. Then for *a.s.* $\omega \in \Omega$, we have the following conclusion:

$$d(X_s, X_t) \leq C_\alpha(\omega)|t - s|^\alpha, \quad \forall s, t \in D, \quad \text{where } C_\alpha(\omega) = \frac{2K_\alpha(\omega)}{1 - 2^{-\alpha}} < \infty.$$

This is an immediately corollary the Claim 4.6*:

Claim 4.6*. If there exists $\alpha, K > 0$ such that $f : D \rightarrow E$ satisfies

$$d(f((j-1)2^{-n}), f(j2^{-n})) \leq K2^{-n\alpha}, \quad \forall n \in \mathbb{N}, \forall j \in \{1, \dots, 2^n\},$$

then we have

$$d(f(s), f(t)) \leq 2K (1 - 2^{-\alpha})^{-1} |t - s|^\alpha.$$

Proof. We fix $t > s > 0$, and choose $p \in \mathbb{N}$ such that $t - s \leq 2^{-p}$. We also choose k such that $s \leq k2^{-p} < t$. Then we can write s and t as

$$s = k2^{-p} - \sum_{j=1}^l \delta_j 2^{-p-j}, \quad t = k2^{-p} + \sum_{j=1}^m \delta'_j 2^{-p-j}, \quad \text{where } \delta_1, \dots, \delta_l, \delta'_1, \dots, \delta'_m \in \{0, 1\}.$$

We take finite sequences $s_j \searrow s_l = s$ and $t_j \nearrow t_m = t$ defined by partial sums. Then $s_0 = t_0 = k2^{-p}$, and

$$d(f(s), f(t)) \leq \sum_{j=1}^l d(f(s_{j-1}), f(s_j)) + \sum_{j=1}^m d(f(t_{j-1}), f(t_j)) \leq \frac{2K2^{-\alpha p}}{1 - 2^{-\alpha}}.$$

Since $|t - s| \leq 2^{-p}$, the result follows. \square

Proof of Theorem 4.6 (Cont). According to our previous discussion, the mapping $t \mapsto X_t(\omega)$ is Hölder continuous of exponent α on D for all $\omega \in \Omega \setminus N$. By completeness of (E, d) , we define

$$\tilde{X}_t(\omega) = \begin{cases} \lim_{D \ni s \rightarrow t} X_s(\omega) & \text{if } K_\alpha(\omega) < \infty \\ x_0 & \text{otherwise,} \end{cases}$$

where $x_0 \in E$ is an arbitrary fixed point. Then $(\tilde{X}_t)_{t \in I}$ has Hölder continuous sample paths of exponent α .

It remains to show that \tilde{X} is an *a.s.* modification of X . Fix $t \in I = [0, 1]$. By assumption in this theorem, we have $X_{t_n} \xrightarrow{\mathbb{P}} X_t$ for $D \ni t_n \rightarrow t$, and $X_{t_n} \rightarrow \tilde{X}_t$ *a.s.* by definition of \tilde{X} . Then $X_t = \tilde{X}_t$ *a.s.* \square

Remark. If I is unbounded, for instance $I = \mathbb{R}_+$, we apply Theorem 4.6 repeatedly on $I_n = [0, n]$ for $n \in \mathbb{N}$. Then $(X_t)_{t \geq 0}$ has an *a.s.* modification $(\tilde{X}_t)_{t \geq 0}$ whose sample paths are locally Hölder continuous of exponent α for all $\alpha \in (0, \frac{\varepsilon}{q})$. We can apply Theorem 4.6 to a pre-Brownian motion.

Corollary 4.7. Let $B = (B_t)_{t \geq 0}$ be a pre-Brownian motion. Then it has an *a.s.* modification whose sample paths are locally Hölder continuous with exponent α for all $\alpha \in (0, \frac{1}{2})$.

Proof. Take $\delta > 0$. For all $s, t \geq 0$, we have

$$\mathbb{E}[|B_t - B_s|^{2+\delta}] = |t - s|^{1+\frac{\delta}{2}} \mathbb{E}[|Z|^{2+\delta}], \quad \text{where } Z \in N(0, 1).$$

By Theorem 4.6, process B has an *a.s.* modification \tilde{B} whose sample paths are locally Hölder continuous of exponent α for all $\alpha \in (0, \frac{\delta}{4+2\delta})$. If δ is great enough we can take α arbitrarily close to $\frac{1}{2}$. \square

Definition 4.8 (Brownian motion/Wiener process). If $B = (B_t)_{t \geq 0}$ is a pre-Brownian motion and B is sample-continuous, then process B is said to be a *Brownian motion/Wiener process*.

Remark. The Corollary 4.7 is in fact a justification for the existence of a Brownian motion. Furthermore, the Proposition 4.5 still holds if pre-Brownian motion is everywhere replaced by Brownian motion. In addition, we have $\lim_{t \rightarrow \infty} \frac{1}{t} B_t = 0$ by continuity of sample paths and the fact that $(\frac{1}{t} B_t)_{t > 0} \stackrel{d}{=} (B_{1/t})_{t > 0}$.

4.2 Canonical Construction and Wiener's Construction

Definition 4.9. Let $\{(\Omega_\alpha, \mathcal{F}_\alpha), \alpha \in J\}$ be a collection of measurable spaces. We define the collection of all *measurable rectangles* by

$$\prod_{\alpha \in J} \mathcal{F}_\alpha = \left\{ \prod_{\alpha \in J} A_\alpha : A_\alpha \in \mathcal{F}_\alpha \text{ for all } \alpha \in J, \text{ and } A_\alpha = \Omega_\alpha \text{ except for finitely many } \alpha \in J \right\}.$$

Akin to the proof (i) of Theorem 1.25, we can prove that $\prod_{\alpha \in J} \mathcal{F}_\alpha$ is a semi-ring. Similar to the definition of product of two measurable spaces, we define the product σ -algebra:

$$\bigotimes_{\alpha \in J} \mathcal{F}_\alpha = \sigma \left(\prod_{\alpha \in J} \mathcal{F}_\alpha \right)$$

The measurable space $(\bigcup_{\alpha \in J} \Omega_\alpha, \bigotimes_{\alpha \in J} \mathcal{F}_\alpha)$ is said to be the *product space* of $\{(\Omega_\alpha, \mathcal{F}_\alpha), \alpha \in J\}$.

Remark. Every coordinate mapping $\pi_\beta : (\omega_\alpha)_{\alpha \in J} \mapsto \omega_\beta$ is measurable when defined on $\{(\Omega_\alpha, \mathcal{F}_\alpha), \alpha \in J\}$. Furthermore, for all finite subset $I \subset J$, the projection mapping $\pi_I : (\omega_\alpha)_{\alpha \in J} \mapsto (\omega_\alpha)_{\alpha \in I}$ is measurable.

Proposition 4.10. Let \mathcal{I}_F be the collection of all finite subsets of J . For $I \in \mathcal{I}_F$, define $\bigotimes_{\alpha \in I}^J \mathcal{F}_\alpha$ to be the sub σ -algebra consisting of all *measurable cylinders* A with base in $\prod_{\alpha \in I} \Omega_\alpha$. That means, $A = B \times \prod_{\alpha \in I^c} \Omega_\alpha$ for some measurable $B \subset \prod_{\alpha \in I} \Omega_\alpha$. Then

$$\bigotimes_{\alpha \in J} \mathcal{F}_\alpha = \sigma \left(\bigcup_{I \in \mathcal{I}_F} \bigotimes_{\alpha \in I}^J \mathcal{F}_\alpha \right)$$

Proof. Clearly $\bigotimes_{\alpha \in I}^J \mathcal{F}_\alpha \subset \bigotimes_{\alpha \in J} \mathcal{F}_\alpha$ for all $I \in \mathcal{I}_F$. On the other hand, every element of $\prod_{\alpha \in J} \mathcal{F}_\alpha$ is contained by some $\bigotimes_{\alpha \in I}^J \mathcal{F}_\alpha$. \square

Proposition 4.11. Let $\{\mathbb{P}_I, I \in \mathcal{I}_F\}$ be a collection of probability measures defined on finite product spaces $\{(\prod_{\alpha \in I} \Omega_\alpha, \bigotimes_{\alpha \in I} \mathcal{F}_\alpha), I \in \mathcal{I}_F\}$. The following compatibility condition is necessary and sufficient for the existence of a finitely additive probability measure \mathbb{P} on $\bigotimes_{\alpha \in J} \mathcal{F}_\alpha$ such that the pushforward $(\pi_I)_* \mathbb{P} = \mathbb{P}_I$.

Compatibility: If $I_1 \subset I_2$ are two finite subsets of J , then $(\pi_{I_2 \rightarrow I_1})_* \mathbb{P}_{I_2} = \mathbb{P}_{I_1}$.

Proof. We only prove sufficiency, since necessity is clear. For every cylinder $A = B \times \prod_{\alpha \in I^c} \Omega_\alpha$ such that $B \in \bigotimes_{\alpha \in I} \mathcal{F}_\alpha$, define $\mathbb{P}(A) = \mathbb{P}_I(B)$. By compatibility condition, we obtain a finitely additive function \mathbb{P} on $\mathcal{A} = \bigcup_{I \in \mathcal{I}_F} \bigotimes_{\alpha \in I}^J \mathcal{F}_\alpha$, which is an algebra. We extend \mathbb{P} to $\bigotimes_{\alpha \in J} \mathcal{F}_\alpha = \sigma(\mathcal{A})$. For $A \in \bigotimes_{\alpha \in J} \mathcal{F}_\alpha$, define

$$\mathbb{P}(A) = \sup \{ \mathbb{P}(F) : F \subset A, F \in \mathcal{A} \} \in [0, 1].$$

Then for any collection of disjoint sets $A_1, \dots, A_n \in \bigotimes_{\alpha \in J} \mathcal{F}_\alpha$, we have

$$\begin{aligned} \sum_{j=1}^n \mathbb{P}(A_j) &= \sum_{j=1}^n \sup \{ \mathbb{P}(F_j) : F_j \subset A_j, F_j \in \mathcal{A} \} \\ &= \sup \left\{ \mathbb{P} \left(\bigcup_{j=1}^n F_j \right) : F_j \subset A_j, F_j \in \mathcal{A}, \forall j \in \{1, \dots, n\} \right\} = \mathbb{P} \left(\bigcup_{j=1}^n A_j \right). \end{aligned}$$

Thus we complete the proof of finite additivity of \mathbb{P} on $\bigotimes_{\alpha \in J} \mathcal{F}_\alpha$. \square

Proposition 4.12 (Compact class). A class \mathcal{C} of subsets of Ω is said to be *compact*, if all its countable subclasses with *finite intersection property* have nonempty intersection. That is, for all sequences $(C_n)_{n=1}^\infty \subset \mathcal{C}$ such that $\bigcap_{k=1}^n C_k \neq \emptyset$ for all $n \in \mathbb{N}$, it holds $\bigcap_{n=1}^\infty C_n \neq \emptyset$.

If \mathcal{C} is a compact class, so are the following: (i) The class \mathcal{C}_δ containing all countable intersections of elements of \mathcal{C} ; (ii) The class \mathcal{C}_s containing all finite unions of elements of \mathcal{C} .

Proof. (i) Since every countable intersection of elements of \mathcal{C}_δ is also a countable intersection of elements of the compact class \mathcal{C} , the result follows immediately.

(ii) We take a sequence $D_n = \bigcup_{j=1}^{m_n} C_j^n \in \mathcal{C}_s$ such that $\bigcap_{k=1}^n D_k \neq \emptyset$, and prove that $\bigcap_{n=1}^\infty D_n \neq \emptyset$. For each $n \in \mathbb{N}$, define the multi-index set $I_n = \prod_{k=1}^n \{1, \dots, m_k\}$. Then according to distributivity law, we have

$$\bigcap_{k=1}^n D_k = \bigcap_{k=1}^n \left(\bigcup_{j=1}^{m_k} C_j^k \right) = \bigcup_{\alpha \in I_n} \left(\bigcap_{k=1}^n C_{\alpha_k}^k \right) \neq \emptyset.$$

Hence for every $n \in \mathbb{N}$, there exists a multi-index $\alpha \in I_n$ such that $\bigcap_{k=1}^n C_{\alpha_k}^k \neq \emptyset$, and we define

$$J_n = \left\{ \alpha \in \prod_{k=1}^\infty \{1, \dots, m_k\} : \bigcap_{k=1}^n C_{\alpha_k}^k \neq \emptyset \right\}, \quad \forall n \in \mathbb{N}.$$

Clearly, the definition of J_n only concerns about the first n elements. Then $J_n \neq \emptyset$ for all $n \in \mathbb{N}$, and $(J_n)_{n=1}^\infty$ is monotone decreasing. Now we choose a sequence $\alpha^{[n]} \in J_n$ for each $n \in \mathbb{N}$. By induction on k , we are able to determine a sequence $\alpha_k^* \in \{1, \dots, m_k\}$ such that $\alpha_{1:k}^* = \alpha_{1:k}^{[n]}$ for infinitely many n . As a result, for each $k \in \mathbb{N}$, we can find $n \geq k$ such that $\alpha_{1:k}^* = \alpha_{1:k}^{[n]}$, which implies $\alpha^* \in J_n \subset J_k$. Hence $\alpha^* \in \bigcup_{k=1}^\infty J_k$, and by compactness of \mathcal{C} we have $\bigcap_{k=1}^\infty D_k \supset \bigcap_{k=1}^\infty C_{\alpha_k^*}^k \neq \emptyset$, completing the proof. \square

Remark. This definition is also in accordance with compactness in topology. If X is a topological space, and \mathcal{K} is the collection of all compact subspaces of X . If $(K_n)_{n=1}^\infty \subset \mathcal{K}$ is a sequence such that $\bigcap_{n=1}^\infty K_j = \emptyset$, define $L_n = \bigcap_{j=1}^n K_j$. Then the increasing sequence $(K_1 \setminus L_n)_{n=1}^\infty$ forms an open cover of K_1 . By compactness of K_1 , there is a finite subcover, and we can find $N \in \mathbb{N}$ such that $L_N = \emptyset$.

Theorem 4.13 (Daniell-Kolmogorov extension). Let $\{\mathbb{P}_I, I \in \mathcal{I}_F\}$ be a collection of probability measures defined on finite product spaces $\left\{ \left(\prod_{\alpha \in I} \Omega_\alpha, \bigotimes_{\alpha \in I} \mathcal{F}_\alpha \right), I \in \mathcal{I}_F \right\}$ that satisfies the compatibility condition in Proposition 4.11. If for each $\alpha \in J$, there exists a compact class $\mathcal{C}_\alpha \subset \mathcal{F}_\alpha$ such that

$$\mathbb{P}_\alpha(A) = \sup \{ \mathbb{P}_\alpha(C) : C \in \mathcal{C}_\alpha, C \subset A \}, \quad \forall A \in \mathcal{F}_\alpha.$$

Then there exists a unique probability measure \mathbb{P} on $(\prod_{\alpha \in J} \Omega_\alpha, \bigotimes_{\alpha \in J} \mathcal{F}_\alpha)$ that extends each \mathbb{P}_I .

Proof. Step I: Let \mathbb{P} be the finitely additive set function found in Proposition 4.11. We first prove that there exists a compact subclass \mathcal{C} of the semiring $\mathcal{S} = \prod_{\alpha \in J} \mathcal{F}_\alpha$ of all measurable rectangles such that

$$\mathbb{P}(A) = \sup \{ \mathbb{P}(C) : C \in \mathcal{C}, C \subset A \}, \quad \forall A \in \mathcal{S}. \quad (4.1)$$

Let $\mathcal{D} = \{ C \times \prod_{\alpha \neq \beta} \Omega_\alpha : \beta \in J, C \in \mathcal{C}_\beta \}$. Then every countable intersection $D = \bigcap_{n \in \mathbb{N}} (C_n \times \prod_{\alpha \neq \beta_n} \Omega_\alpha)$ of elements of \mathcal{D} has the form $\prod_{\beta \in J} B_\beta$, where $B_\beta = \bigcap_{\{n: \beta_n = \beta\}} C_n$. If the countable intersection D is empty, let $\beta \in J$ be such that $B_\beta = \emptyset$. By compactness of \mathcal{C}_β , there exists a finite subset $I_\beta \subset \{n : \beta_n = \beta\}$ such that $\bigcap_{n \in I_\beta} C_n = \emptyset$, which implies $\bigcap_{n \in I_\beta} (C_n \times \prod_{\alpha \neq \beta_n} \Omega_\alpha) = \emptyset$. Therefore \mathcal{D} is a compact subclass of \mathcal{S} . Again, the class \mathcal{C} of all finite intersections of elements of \mathcal{D} is compact.

Now we prove (4.1). Take any $\epsilon > 0$. If A is a measurable rectangle with base $\prod_{j=1}^n A_{\alpha_j} \subset \prod_{j=1}^n \Omega_{\alpha_j}$, choose $\mathcal{C}_{t_j} \ni C_j \subset A_{\alpha_j}$ such that $\mathbb{P}_{\alpha_j}(C_j) \geq \mathbb{P}_{\alpha_j}(A_{t_j}) - \frac{\epsilon}{n}$. Then $\mathcal{C} \ni C = \bigcap_{j=1}^n (C_j \times \prod_{\alpha \neq \alpha_j} \Omega_{\alpha}) \subset A$, and

$$\mathbb{P}(A \setminus C) = \mathbb{P} \left(\bigcup_{j=1}^n \left((A_{\alpha_j} \setminus C_j) \times \prod_{\alpha \neq \alpha_j} \Omega_{\alpha} \right) \right) \leq \sum_{j=1}^n \mathbb{P}_{\alpha_j}(A_{\alpha_j} \setminus C_j) = \epsilon \downarrow 0.$$

Step II: We prove the σ -additivity of \mathbb{P} on \mathcal{S} . We take the class \mathcal{C}_s consisting of all finite intersection of elements of \mathcal{C} , which is again a compact class by Proposition 4.12 (ii) and is contained in the ring \mathcal{R} generated by \mathcal{S} according to the Remark under Definition 1.24. Similar to the proof of (4.1) in Step I, we can prove that $\mathbb{P}(A) = \sup\{\mathbb{P}(C) : C \in \mathcal{C}_s, C \subset A\}$ for all $A = \prod_{k=1}^n A_k \in \mathcal{R}$ by taking $C = \prod_{k=1}^n C_k \in \mathcal{C}_s$ with $\mathbb{P}(A_k \setminus C_k) \leq \epsilon/n$ for arbitrarily small ϵ . We prove that \mathbb{P} is σ -additive on \mathcal{R} , hence on \mathcal{S} .

Given $\epsilon > 0$, we take a sequence $\mathcal{R} \ni A_n \downarrow \emptyset$, and take $\mathcal{C}_s \ni C_n \subset A_n$ with $\mathbb{P}(A_n) \leq \mathbb{P}(C_n) + \epsilon 2^{-n}$. Then $\bigcap_{n=1}^{\infty} C_n \subset \bigcap_{n=1}^{\infty} A_n = \emptyset$, and there exists $N \in \mathbb{N}$ such that $\bigcap_{n=1}^N C_n = \emptyset$ by compactness of \mathcal{C}_s . As a result,

$$\mathbb{P}(A_N) = \mathbb{P} \left(A_N \setminus \left(\bigcap_{n=1}^N C_n \right) \right) = \mathbb{P} \left(\bigcup_{n=1}^N (A_N \setminus C_n) \right) \leq \sum_{n=1}^N \mathbb{P}(A_n \setminus C_n) \leq \epsilon \downarrow 0.$$

Therefore \mathbb{P} is continuous at \emptyset . If $(B_n)_{n=1}^{\infty} \subset \mathcal{R}$ is a sequence of disjoint sets, take $A_n = \bigcup_{k=n+1}^{\infty} B_k$. Then we have $A_n \downarrow \emptyset$. Finite additivity of \mathbb{P} implies

$$\mathbb{P} \left(\bigcup_{n=1}^{\infty} B_n \right) - \sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(A_N) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Step III: According to Step II, \mathbb{P} is a finite pre-measure on the semiring \mathcal{S} which generates $\bigotimes_{\alpha \in J} \mathcal{F}_{\alpha}$. By Carathéodory's extension theorem, \mathbb{P} can be uniquely extended to a probability measure $\bigotimes_{\alpha \in J} \mathcal{F}_{\alpha}$. On the other hand, the finite additive function \mathbb{P} is uniquely defined on \mathcal{S} , which is specified by the family of measures $\{\mathbb{P}_I, I \in \mathcal{I}_F\}$ on finite-dimensional subspaces. Therefore the extension \mathbb{P} is unique. \square

Proposition 4.14. Let Ω be a Hausdorff topological space, and equip Ω with the Borel σ -algebra \mathcal{B} . Let \mathcal{C} be the collection of all closed sets in Ω , and \mathcal{K} the collection of all compact sets in Ω . Let \mathbb{P} be a probability measure on (Ω, \mathcal{B}) . We define the collections \mathcal{R}_c of *closed regular sets* and \mathcal{R}_k of *regular sets* as

$$\mathcal{R}_c = \left\{ B \in \mathcal{B} : \mathbb{P}(B) = \sup_{C \in \mathcal{C} : C \subset B} \mathbb{P}(C) \right\}, \quad \mathcal{R}_k = \left\{ B \in \mathcal{B} : \mathbb{P}(B) = \sup_{K \in \mathcal{K} : K \subset B} \mathbb{P}(K) \right\}.$$

We say \mathbb{P} is *tight* if $\Omega \in \mathcal{R}_k$. We say \mathbb{P} is *closed regular* (resp. *regular*) if $\mathcal{R}_c = \mathcal{B}$ (resp. $\mathcal{R}_k = \mathcal{B}$).

- (i) The collection $\mathcal{R}_c^* = \{B \in \mathcal{R}_c : \Omega \setminus B \in \mathcal{R}_c\}$ is a σ -algebra. In addition, if \mathbb{P} is tight, then the collection $\mathcal{R}_k^* = \{B \in \mathcal{R}_k : \Omega \setminus B \in \mathcal{R}_k\}$ is also a σ -algebra.
- (ii) If Ω is metrizable, then \mathbb{P} is closed regular. In addition, if \mathbb{P} is tight, then it is regular.
- (iii) (Ulam). If Ω is a *Polish space* (a separable completely metrizable space), then \mathbb{P} is regular.

Proof. (i) Clearly, $\Omega \in \mathcal{R}_c^*$, and $B \in \mathcal{R}_c^*$ implies $\Omega \setminus B \in \mathcal{R}_c^*$. Given any sequence $(A_n)_{n=1}^{\infty} \subset \mathcal{R}_c^*$, we prove $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{R}_c^*$. Let $\epsilon > 0$. Take closed sets $C_n \subset A_n$ such that $\mathbb{P}(C_n) \geq \mathbb{P}(A_n) - \epsilon 3^{-n}$, and $D_n \subset \Omega \setminus A_n$ such that $\mathbb{P}(D_n) \geq \mathbb{P}(\Omega \setminus A_n) - \epsilon 2^{-n}$. Then there exists $N \in \mathbb{N}$ such that $\mathbb{P}(A) > \mathbb{P}(\bigcup_{n=1}^{\infty} A_n) - \epsilon/2$. Note that $\bigcup_{n=1}^N C_n \subset A$ is closed, and

$$\mathbb{P}(A) - \mathbb{P} \left(\bigcup_{n=1}^N C_n \right) < \frac{\epsilon}{2} + \mathbb{P} \left(\bigcup_{n=1}^N A_n \right) - \mathbb{P} \left(\bigcup_{n=1}^N C_n \right) \leq \frac{\epsilon}{2} + \mathbb{P} \left(\bigcup_{n=1}^N (A_n \setminus C_n) \right) < \epsilon.$$

Meanwhile, $\bigcap_{n=1}^{\infty} D_n \subset \Omega \setminus A$ is also closed, and

$$\mathbb{P}(\Omega \setminus A) - \mathbb{P}\left(\bigcap_{n=1}^{\infty} D_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} ((\Omega \setminus A) \setminus D_n)\right) \leq \sum_{n=1}^{\infty} \mathbb{P}((\Omega \setminus A_n) \setminus D_n) < \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $A \in \mathcal{R}_c^*$. Hence \mathcal{R}_c^* is a σ -algebra.

If \mathbb{P} is tight, we have $\Omega, \emptyset \in \mathcal{R}_k^*$, and $B \in \mathcal{R}_k^*$ implies $\Omega \setminus B \in \mathcal{R}_k^*$. Similar to the above proof, since finite unions and countable intersections of compact sets are still compact, we conclude \mathcal{R}_k^* is a σ -algebra.

(ii) Let d be the metric of Ω . If $U \subset \Omega$ is an open set, take its complement $F = \Omega \setminus U$, and define $F_n = \{x \in \Omega : d(x, F) \geq 1/n\}$. Then $F_n \uparrow U$, and $U \in \mathcal{R}_c^*$ defined in (i). Since \mathcal{R}_c^* is a σ -algebra containing all open sets in Ω , we have $\mathcal{R}_c^* = \mathcal{B}$, and \mathbb{P} is closed regular.

In addition, if \mathbb{P} is tight, we can take a compact set K such that $\mathbb{P}(\Omega \setminus K) < \epsilon/2$ for every $\epsilon > 0$. For any Borel set $B \in \mathcal{B}$, take closed set $F \subset B$ with $\mu(B \setminus F) < \epsilon/2$. Then $F \cap K \subset B$ is a compact set, and $\mathbb{P}(B \setminus (K \cap F)) \leq \mathbb{P}(B \setminus F) + \mathbb{P}(\Omega \setminus K) < \epsilon$. Since $\epsilon > 0$ is arbitrary, $B \in \mathcal{R}_k$. Hence \mathbb{P} is regular.

(iii) Following (ii), it suffices to show that \mathbb{P} is tight. Let $(\omega_n)_{n=1}^{\infty}$ be a dense sequence in Ω . For any $\eta > 0$ and $\omega \in \Omega$, let $B(\omega, \eta)$ be the closed ball centered at ω of radius η . Given $\epsilon > 0$, by density of $(\omega_n)_{n=1}^{\infty}$, we are able to take $N_m \in \mathbb{N}$ such that

$$\mathbb{P}\left(\Omega \setminus \bigcup_{n=1}^{N_m} B\left(\omega_n, \frac{1}{m}\right)\right) < \frac{\epsilon}{2^m}, \quad \forall m \in \mathbb{N}.$$

Let $K = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{N_m} B\left(\omega_n, \frac{1}{m}\right)$. Then K is both closed and totally bounded. Since Ω is a complete metric space, K is compact. Furthermore, we have

$$\mathbb{P}(\Omega \setminus K) \leq \sum_{m=1}^{\infty} \frac{\epsilon}{2^m} = \epsilon \downarrow 0.$$

Hence \mathbb{P} is tight, and the result follows from (ii). □

Theorem 3.13 and Proposition 3.14 together imply the following conclusion.

Corollary 4.15. Let $\{(\Omega_\alpha, \mathcal{B}_\alpha), \alpha \in J\}$ be a family of Polish spaces equipped with their Borel σ -algebras. For any compatible family $\{\mathbb{P}_I, I \in \mathcal{I}_F\}$ of probability measures defined on $\{(\prod_{\alpha \in I} \Omega_\alpha, \bigotimes_{\alpha \in I} \mathcal{B}_\alpha), I \in \mathcal{I}_F\}$, there exists a unique probability measure \mathbb{P} on $(\prod_{\alpha \in J} \Omega_\alpha, \bigotimes_{\alpha \in J} \mathcal{B}_\alpha)$ that extends each \mathbb{P}_I .

Remark. Let E be a metric space equipped with its Borel σ -algebra \mathcal{B} . The product space $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$ is called the *canonical space*. Since every evaluation map $\pi_t : (E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}}) \rightarrow (E, \mathcal{B})$, $x \mapsto x(t)$ is measurable, we can define a process $(\pi_t)_{t \in \mathcal{T}}$ on $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$, which is called the *canonical process*. Given a probability measure μ on $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$, the sample paths of the canonical process $(\pi_t)_{t \in \mathcal{T}}$ are distributed according to μ .

Given $(X_t)_{t \in \mathcal{T}}$ is a stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ whose state space is the metric space E , the mapping $\Phi : (\Omega, \mathcal{F}) \rightarrow (E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$ defined to map ω to its sample path $t \mapsto X_t(\omega)$ is measurable. In fact, for every measurable rectangle $A \in E^{\mathcal{T}}$ with basis $\prod_{j=1}^n A_{t_j}$, its pre-image $\Phi^{-1}(A) = \bigcap_{j=1}^n X_{t_j}^{-1}(A_{t_j}) \in \mathcal{F}$. Since Φ^{-1} preserves complement and countable union operations, and $\mathcal{B}^{\otimes \mathcal{T}}$ is generated by all measurable rectangles, we obtain that Φ is measurable. As a result, the process $(X_t)_{t \in \mathcal{T}}$ determines a pushforward probability measure $\mu = \Phi_* \mathbb{P}$ on the canonical space $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$. Furthermore, the canonical process $(\pi_t)_{t \in \mathcal{T}}$ defined on $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}}, \mu)$ is identically distributed to $(X_t)_{t \in \mathcal{T}}$.

According to our previous discussion, we can construct a stochastic process indexed by \mathcal{T} and taking values in E by constructing a probability measure on the canonical space $(E^{\mathcal{T}}, \mathcal{B}^{\otimes \mathcal{T}})$.

Example: the canonical construction of Brownian motion. Let $C(\mathbb{R}_+)$ be the space of all real-valued continuous function defined on \mathbb{R}_+ . We give $C(\mathbb{R}_+)$ the metric

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\sup_{t \in [0, n]} |f(t) - g(t)|}{1 + \sup_{t \in [0, n]} |f(t) - g(t)|}, \quad \forall f, g \in C(\mathbb{R}_+).$$

This metric induces the compact convergence (c.c.) topology on $C(\mathbb{R}_+)$, because a sequence $f_n \in C(\mathbb{R}_+)$ converges uniformly on each compact set to $f \in C(\mathbb{R}_+)$ if and only if $d(f_n, f) \rightarrow 0$. Then for each $t \in \mathbb{R}_+$, the coordinate mapping $\pi_t : C(\mathbb{R}_+) \rightarrow \mathbb{R}$, $f \mapsto f(t)$ is continuous. Consequently, if we give $C(\mathbb{R}_+)$ the Borel σ -algebra \mathcal{C} generated by c.c.topology, then all coordinate mappings $\{\pi_t, t \in \mathbb{R}_+\}$ are measurable.

In fact, \mathcal{C} coincides the smallest σ -algebra \mathcal{B}_p on $C(\mathbb{R}_+)$ for which the coordinate mappings $\pi_t : f \mapsto f(t)$ are measurable for all $t \in \mathbb{R}_+$. To see this, note that $C(\mathbb{R}_+)$ is separable with respect to the c.c. topology, because we can approximate each $f \in C(\mathbb{R}_+)$ within arbitrary precision with a polynomial with rational coefficients on some $[0, n]$. As a result, every open set in $C(\mathbb{R}_+)$ with respect to the c.c.topology is a countable union of basic sets of the form $B_{[0, n]}(f_0, \epsilon) = \{f \in C(\mathbb{R}_+) : \sup_{t \in [0, n]} |f(t) - f_0(t)| < \epsilon\}$, where $f_0 \in C(\mathbb{R}_+)$ and $\epsilon > 0$. By continuity, every basic set $B_{[0, n]}(f_0, \epsilon) = \bigcap_{t \in \mathbb{Q} \cap [0, n]} \pi_t^{-1}(O(f_0(t), \epsilon)) \in \mathcal{B}_p$. Hence every open set in $C(\mathbb{R}_+)$ with respect to the c.c.topology is contained in \mathcal{B}_p . Therefore \mathcal{B}_p coincides \mathcal{C} .

For each measurable rectangle $A = \bigcap_{j=1}^n \pi_{t_j}^{-1}(A_j)$ [note every coordinate map π_{t_j} is defined on $C(\mathbb{R}_+)$], where $0 = t_0 < t_1 < \dots < t_n$ and $A_0, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, define

$$W(A) = \frac{\mathbb{1}_{A_0}(0)}{\sqrt{(2\pi)^n \prod_{j=1}^n (t_j - t_{j-1})}} \int_{A_1 \times \dots \times A_n} \exp\left(-\sum_{j=1}^n \frac{(z_j - z_{j-1})^2}{2(t_j - t_{j-1})}\right) dz_1 \dots dz_n, \quad \text{where } z_0 = 0.$$

This is a pre-measure on the semi-ring \mathcal{S} of all measurable rectangles, and we can extend it to a measure the σ -algebra \mathcal{C} generated by \mathcal{S} . This is called the *Wiener measure* on $(C(\mathbb{R}_+), \mathcal{C})$.

If we choose $(\Omega, \mathcal{F}, \mathbb{P}) = (C(\mathbb{R}_+), \mathcal{C}, W)$, then the canonical process $(\pi_t)_{t \in \mathbb{R}_+}$ is a Brownian motion. This is a consequence of Proposition 4.4 (iv) and the fact that $(\pi_t)_{t \in \mathbb{R}_+}$ has continuous sample paths. In fact, the distribution law of every Brownian motion $(B_t)_{t \geq 0}$ is determined, which is W .

Example: Wiener's construction of Brownian Process. Let $\{e_n, n \in \mathbb{N}\}$ be a countable orthonormal basis of $L^2([0, 1])$, which is a separable Hilbert space. According to Corollary 4.15, it is possible to construct a collection of independent standard Gaussian variables $(Z_n)_{n=1}^{\infty}$ on an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define a process $(B_t)_{t \in [0, 1]}$ as follows:

$$B_t = \sum_{n=1}^{\infty} \langle \mathbb{1}_{[0, t]}, e_n \rangle Z_n, \quad \forall t \in [0, 1]. \quad (4.2)$$

Since $\sum_{n=1}^{\infty} |\langle \mathbb{1}_{[0, t]}, e_n \rangle|^2 = t < \infty$, the series (4.2) converges in L^2 . Consequently, $(B_t)_{t \in [0, 1]}$ is a Gaussian process. Furthermore, for any partition $0 = t_0 < t_1 < \dots < t_p = 1$, we have

$$\begin{aligned} \mathbb{E}[(B_{t_j} - B_{t_{j-1}})(B_{t_k} - B_{t_{k-1}})] &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \langle \mathbb{1}_{(t_{j-1}, t_j]}, e_m \rangle \langle \mathbb{1}_{(t_{k-1}, t_k]}, e_n \rangle \underbrace{\langle e_m, e_n \rangle}_{=\mathbb{E}[Z_m Z_n]} \\ &= \left\langle \sum_{m=1}^{\infty} \langle \mathbb{1}_{(t_{j-1}, t_j]}, e_m \rangle e_m, \sum_{n=1}^{\infty} \langle \mathbb{1}_{(t_{k-1}, t_k]}, e_n \rangle e_n \right\rangle \\ &= \langle \mathbb{1}_{(t_{j-1}, t_j]}, \mathbb{1}_{(t_{k-1}, t_k]} \rangle = \begin{cases} t_j - t_{j-1}, & \text{if } j = k; \\ 0, & \text{if } j \neq k. \end{cases} \end{aligned}$$

Consequently, the process $(B_t)_{t \in [0,1]}$ has independent increments, and $B_{t_j} - B_{t_{j-1}} \sim N(0, t_j - t_{j-1})$ for each j . By Proposition 4.4 (iv), $(B_t)_{t \in [0,1]}$ is a pre-Brownian motion on $[0, 1]$. If we choose a particular basis for $L^2([0, 1])$: $e_0 = 1$, and $e_n(t) = \sqrt{2} \cos(n\pi t)$, $\forall n \in \mathbb{N}$, we obtain a process

$$B_t = tZ_0 + \sum_{n=1}^{\infty} \frac{\sqrt{2} \sin(n\pi t)}{n} Z_n = tZ_0 + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{\sqrt{2} \sin(k\pi t)}{k} Z_k, \quad \forall t \in [0, 1].$$

We set $S_m(t) = \sum_{k=m}^{2^{m-1}} \frac{\sqrt{2} \sin(k\pi t)}{k} Z_k$, and write $B_t = tZ_0 + \sum_{n=0}^{\infty} S_{2^n}(t)$. Let $T_m = \sup_{t \in [0,1]} |S_m(t)|$. Then

$$\begin{aligned} T_m^2 &\leq \sqrt{2} \sup_{t \in [0,1]} \left| \sum_{k=m}^{2^{m-1}} \frac{e^{ik\pi t}}{k} Z_k \right|^2 = \sqrt{2} \sup_{t \in [0,1]} \sum_{j=m}^{2^{m-1}} \sum_{k=m}^{2^{m-1}} \frac{e^{i(k-j)\pi t}}{jk} Z_j Z_k \\ &\leq \sqrt{2} \sum_{k=m}^{2^{m-1}} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sup_{t \in [0,1]} \left| \sum_{k=m}^{2^{m-1}} \sum_{l=1}^{2^{m-k-1}} e^{il\pi t} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \\ &= \sqrt{2} \sum_{k=m}^{2^{m-1}} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sup_{t \in [0,1]} \left| \sum_{l=1}^{m-1} e^{il\pi t} \sum_{k=m}^{2^{m-l-1}} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \\ &\leq \sqrt{2} \sum_{k=m}^{2^{m-1}} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sum_{l=1}^{m-1} \left| \sum_{k=m}^{2^{m-l-1}} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \end{aligned}$$

Let us bound the expectation of the second term:

$$\begin{aligned} \mathbb{E} \left[\left| \sum_{k=m}^{2^{m-l-1}} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \right] &\leq \sqrt{\mathbb{E} \left[\left(\sum_{k=m}^{2^{m-l-1}} \frac{Z_k Z_{k+l}}{k(k+l)} \right)^2 \right]} = \left(\sum_{k=m}^{2^{m-l-1}} \frac{1}{k^2(k+l)^2} \right)^{1/2} \leq \frac{1}{m^{3/2}}, \\ \Rightarrow \mathbb{E} \left[\sum_{l=1}^{m-1} \left| \sum_{k=m}^{2^{m-l-1}} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \right] &\leq \frac{1}{\sqrt{m}}. \end{aligned}$$

Since $\sum_{k=m}^{2^{m-1}} k^{-2} \leq m^{-1}$, we have $\mathbb{E}[T_m^2] < c/\sqrt{m}$ for some constant $c > 0$ not dependent on m . Consequently,

$$\mathbb{E} \left[\sum_{n=0}^{\infty} T_{2^n} \right] \leq \sum_{n=0}^{\infty} \sqrt{\mathbb{E}[T_{2^n}^2]} \leq c \sum_{n=0}^{\infty} \frac{1}{2^{n/4}} < \infty.$$

By Weierstrass M-test, with probability 1, the mapping $t \mapsto B_t(\omega)$ converges uniformly on $[0, 1]$, and the uniform limit is continuous on $[0, 1]$. By redefine the sample path of $(B_t)_{t \in [0,1]}$ on a negligible set, we obtain a Brownian motion $(B_t)_{t \in [0,1]}$ on $[0, 1]$. To construct a Brownian motion $(B_t)_{t \geq 0}$ on \mathbb{R}_+ , we concatenate Brownian processes defined on each $[n-1, n]$:

$$B_t = B_{n-1} + (t - n + 1)Z_0^{(n)} + \sum_{k=1}^{\infty} \frac{\sqrt{2} \sin(k\pi t)}{k} Z_k^{(n)}, \quad t \in [n-1, n],$$

where $(Z_k^{(n)})_{k=1}^{\infty}$ is a family of independent standard Gaussian variables.

4.3 Sample Paths of Brownian Motion

Let $(B_t)_{t \geq 0}$ be a Brownian motion. We take the canonical filtration $(\mathcal{F}_t)_{t \geq 0}$ of $(B_t)_{t \geq 0}$:

$$\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t), \quad \forall t \geq 0.$$

Then $(B_t)_{t \geq 0}$ is a martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$.

Theorem 4.16 (Blumenthal's 0-1 law). The σ -algebra $\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t$ is \mathbb{P} -trivial.

Proof. Let $0 < t_1 < \dots < t_n$, and let g be a bounded continuous function on \mathbb{R}^n . According to continuity and dominated convergence theorem, for all $A \in \mathcal{F}_{0+}$, we have

$$\mathbb{E}[\mathbb{1}_A g(B_{t_1}, \dots, B_{t_n})] = \lim_{\epsilon \downarrow 0} \mathbb{E}[\mathbb{1}_A g(B_{t_1} - B_\epsilon, \dots, B_{t_n} - B_\epsilon)].$$

By simple Markov property of Brownian motions [Proposition 4.5 (iii)], $B_{t_1} - B_\epsilon, \dots, B_{t_n} - B_\epsilon$ is independent of $\mathcal{F}_\epsilon \supset \mathcal{F}_{0+}$ whenever $0 < \epsilon < t_1$. Hence

$$\mathbb{E}[\mathbb{1}_A g(B_{t_1}, \dots, B_{t_n})] = \mathbb{P}(A) \lim_{\epsilon \downarrow 0} \mathbb{E}[g(B_{t_1} - B_\epsilon, \dots, B_{t_n} - B_\epsilon)] = \mathbb{P}(A) \mathbb{E}[g(B_{t_1}, \dots, B_{t_n})].$$

For any open set $U \in \mathcal{B}(\mathbb{R}^n)$, take a sequence $g_n(x) = d(x, U^c)/(d(x, U^c) + n^{-1})$ of bounded continuous functions such that $g_n \uparrow \mathbb{1}_U$ pointwise. Then

$$\mathbb{E}[\mathbb{1}_A \mathbb{1}_U(B_{t_1}, \dots, B_{t_n})] = \mathbb{P}(A) \mathbb{E}[\mathbb{1}_U(B_{t_1}, \dots, B_{t_n})].$$

Since $\mathcal{B}(\mathbb{R}^n)$ is generated by all open sets in \mathbb{R}^n , an argument of Sierpiński-Dynkin π - λ theorem implies that \mathcal{F}_{0+} is independent of $\sigma(B_{t_1}, \dots, B_{t_n})$. This holds for all finite marginals $0 < t_1 < \dots < t_n$, hence \mathcal{F}_{0+} is independent of $\sigma(B_t, t > 0)$. By right-continuity of $t \mapsto B_t(\omega)$, $B_0 = \lim_{t \rightarrow 0} B_t$ is measurable with respect to $\sigma(B_t, t > 0)$, and we have $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0) \supset \mathcal{F}_{0+}$. Consequently, \mathcal{F}_{0+} is independent of itself, and the result follows. \square

Proposition 4.17. Let $(B_t)_{t \geq 0}$ be a Brownian motion.

(i) Almost surely, we have

$$\sup_{0 \leq s \leq \epsilon} B_s > 0 \quad \text{and} \quad \inf_{0 \leq s \leq \epsilon} B_s < 0, \quad \forall \epsilon > 0.$$

(ii) For every $\alpha \in \mathbb{R}$, define a stopping time $\tau_\alpha = \inf\{t \geq 0 : B_t = \alpha\}$ (with respect to the canonical filtration, with the convention $\inf \emptyset = \infty$). Then we have $\tau_\alpha < \infty$ a.s.. Consequently, it holds

$$\limsup_{t \rightarrow \infty} B_t = \infty \quad \text{and} \quad \liminf_{t \rightarrow \infty} B_t = -\infty \quad \text{a.s..}$$

Proof. (i) We choose a sequence $0 < \epsilon_n \downarrow 0$, and define the monotone decreasing intersection:

$$A = \bigcap_{n=1}^{\infty} \left\{ \sup_{0 \leq s \leq \epsilon_n} B_s > 0 \right\} \Rightarrow \forall t > 0, \exists \epsilon_n < t \text{ with } \bigcap_{k=n}^{\infty} \left\{ \sup_{0 \leq s \leq \epsilon_k} B_s > 0 \right\} \in \mathcal{F}_t \Rightarrow A \in \mathcal{F}_{0+}.$$

Hence we have either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. If $\mathbb{P}(A) = 1$, the result follows. To show $\mathbb{P}(A) = 1$, note that

$$\mathbb{P}\left(\sup_{0 \leq s \leq \epsilon_n} B_s > 0\right) \geq \mathbb{P}(B_{\epsilon_n} > 0) = \frac{1}{2}, \quad \forall n \in \mathbb{N} \Rightarrow \mathbb{P}(A) \geq \frac{1}{2}.$$

(ii) By simple Markov property of Brownian motions [Proposition 4.5 (ii)], for all $\lambda > 0$, the process $B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$ is also a Brownian motion. Since all Brownian motions are identically distributed [according to the Wiener measure on $C(\mathbb{R}_+)$], we have.

$$\mathbb{P} \left(\sup_{0 \leq s \leq 1} B_s > \lambda \right) = \mathbb{P} \left(\sup_{0 \leq s \leq 1/\lambda^2} B_s^\lambda > 1 \right) = \mathbb{P} \left(\sup_{0 \leq s \leq 1/\lambda^2} B_s > 1 \right)$$

Let $\lambda \downarrow 0$. By monotone convergence theorem, we have

$$1 = \mathbb{P} \left(\sup_{0 \leq s \leq 1} B_s > 0 \right) = \mathbb{P} \left(\sup_{s \geq 0} B_s > 1 \right) = \mathbb{P} \left(\sup_{s \geq 0} B_s^\alpha > 1 \right) = \mathbb{P} \left(\sup_{s \geq 0} B_s > \alpha \right) \leq \mathbb{P}(\tau_\alpha < \infty), \quad \forall \alpha > 0.$$

Therefore $\tau_\alpha < \infty$ a.s. for all $\alpha > 0$, which holds only if $\limsup_{t \rightarrow \infty} B_t = \infty$ a.s..

The symmetric arguments of (i) and (ii) follow by replacing B_s by $-B_s$. \square

The following proposition follows immediately from Remark III under Definition 4.2, which is a property of Gaussian white noise.

Proposition 4.18. Let $(B_t)_{t \geq 0}$ be a Brownian motion. Let $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ be a sequence of partitions of $[0, t]$ with the mesh $\max_{1 \leq j \leq k_n} (t_j^n - t_{j-1}^n) \rightarrow 0$. Then

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} (B_{t_j^n} - B_{t_{j-1}^n})^2 = t \quad \text{in } L^2.$$

Corollary 4.19. Let $(B_t)_{t \geq 0}$ be a Brownian motion.

- (i) Almost surely, the sample path $t \mapsto B_t$ is not monotone on any nontrivial interval $[a, b]$.
- (ii) Almost surely, the sample path $t \mapsto B_t$ has infinite total variation on any nontrivial interval $[a, b]$.
- (iii) $(B_t)_{t \geq 0}$ has finite quadratic variation $\langle B, B \rangle_t = t$.

Proof. (i) By simple Markov property of $(B_t)_{t \geq 0}$ and Proposition 4.17, we have

$$\sup_{q \leq t \leq q+\epsilon} B_t > B_q, \quad \inf_{q \leq t \leq q+\epsilon} B_t < B_q, \quad \forall \epsilon > 0, \quad \forall q \in \mathbb{Q}, \quad \text{a.s.}$$

For any nontrivial interval $[a, b]$, we just choose $q \in \mathbb{Q}$ and $\epsilon > 0$ with $[q, q + \epsilon] \subset [a, b]$.

(ii) By simple Markov property of $(B_t)_{t \geq 0}$, it suffices to consider intervals $[0, t]$. Choose an increasing sequence of partitions of $[0, t]$ as in Proposition 4.18, we have

$$\left(\sup_{1 \leq j \leq k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \right) \sum_{j=1}^{k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \geq \sum_{j=1}^{k_n} (B_{t_j^n} - B_{t_{j-1}^n})^2 \xrightarrow{L^2} t.$$

As $n \rightarrow \infty$, we have $\sup_{1 \leq j \leq k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \rightarrow 0$ by continuity, which implies $\sum_{j=1}^{k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \rightarrow \infty$ a.s.. The result follows by taking intersection of all $0 < t \in \mathbb{Q}$. (iii) is a consequence of Theorem 3.56. \square

Proposition 4.20 (Non-differentiability). Let $(B_t)_{t \geq 0}$ be a Brownian motion. Then

$$\limsup_{t \downarrow 0} \frac{B_t}{\sqrt{t}} = \infty \quad \text{and} \quad \liminf_{t \downarrow 0} \frac{B_t}{\sqrt{t}} = -\infty \quad \text{a.s.}$$

Consequently, by simple Markov property of $(B_t)_{t \geq 0}$, for every $s > 0$, the function $t \mapsto B_t$ has a.s. no right derivative, hence is non-differentiable at s .

Proof. We prove that for all $\alpha > 0$, almost surely,

$$\sup_{0 \leq s \leq \epsilon} \frac{B_s}{\sqrt{s}} > \alpha, \quad \forall \epsilon > 0. \quad (4.3)$$

This statement holds only if $\limsup_{t \downarrow 0} \frac{B_t}{\sqrt{t}} = \infty$. Take a decreasing sequence $0 < \epsilon_n \downarrow 0$, and define the decreasing intersection

$$A = \bigcap_{n=1}^{\infty} \left\{ \sup_{0 \leq s \leq \epsilon_n} \frac{B_s}{\sqrt{s}} > \alpha \right\} \in \mathcal{F}_{0+}.$$

According to Blumenthal's 0-1 law, we have either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Note that

$$\mathbb{P} \left(\sup_{0 \leq s \leq \epsilon_n} \frac{B_s}{\sqrt{s}} > \alpha \right) \geq \mathbb{P} \left(\frac{B_{\epsilon_n}}{\sqrt{\epsilon_n}} > \alpha \right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}} dx \Rightarrow \mathbb{P}(A) \geq \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}} dx > 0.$$

Therefore $\mathbb{P}(A) = 1$, and the result (4.3) follows. A symmetric argument holds if we replace B_s by $-B_s$. \square

Proposition 4.21 (Global non-differentiability). Let $(B_t)_{t \geq 0}$ be a Brownian motion. Almost surely, the sample path $t \mapsto B_t$ is nowhere differentiable.

Proof. By continuity of sample paths of $(B_t)_{t \geq 0}$ on $[0, 2]$, we have

$$\left\{ \exists t_0 \in [0, 1] \text{ such that } \limsup_{h \downarrow 0} \frac{|B_{t_0+h} - B_{t_0}|}{h} < \infty \right\} \Rightarrow \left\{ \exists t_0 \in [0, 1] \text{ such that } \sup_{h \in (0, 1]} \frac{|B_{t_0+h} - B_{t_0}|}{h} < \infty \right\}$$

To show that $t \mapsto B_t$ is *a.s.* nowhere differentiable on $[0, 1]$, it suffices to show that for all $M > 0$,

$$\mathbb{P} \left(\exists t_0 \in [0, 1] \text{ such that } \sup_{h \in (0, 1]} \frac{|B_{t_0+h} - B_{t_0}|}{h} \leq M \right) = 0. \quad (4.4)$$

Fix $M > 0$. Given $n > 2$, choose $k \in \mathbb{Z}$ with $t_0 \in [(k-1)2^{-n}, k2^{-n}]$. Then for all $1 \leq j \leq 2^n - k$, it holds

$$|B_{(k+j)2^{-n}} - B_{(k+j-1)2^{-n}}| \leq |B_{(k+j)2^{-n}} - B_{t_0}| + |B_{t_0} - B_{(k+j-1)2^{-n}}| \leq M(2j+1)2^{-n}.$$

Define events $E_{n,k} := \{|B_{(k+j)2^{-n}} - B_{(k+j-1)2^{-n}}| \leq M(2j+1)2^{-n}, \forall j = 1, 2, 3\}$. Since the process $(B_t)_{t \geq 0}$ has independent increments, we have

$$\begin{aligned} \mathbb{P}(E_{n,k}) &\leq \prod_{j=1}^3 \mathbb{P}(|B_{(k+j)2^{-n}} - B_{(k+j-1)2^{-n}}| \leq M(2j+1)2^{-n}) \leq \mathbb{P}(|B_1| \leq 7M2^{-n/2})^3 \leq \frac{(7M)^3 2^{-3n/2}}{27} \\ &\Rightarrow \mathbb{P} \left(\bigcup_{k=1}^{2^n-3} E_{n,k} \right) \leq 2^n \frac{(7M)^3 2^{-3n/2}}{27} = \frac{(7M)^3 2^{-n/2}}{27} \Rightarrow \sum_{n=1}^{\infty} \mathbb{P} \left(\bigcup_{k=1}^{2^n-3} E_{n,k} \right) < \infty. \end{aligned}$$

By Borel-Cantelli lemma, we have

$$\mathbb{P} \left(\exists t_0 \in [0, 1] \text{ such that } \sup_{h \in (0, 1]} \frac{|B_{t_0+h} - B_{t_0}|}{h} \leq M \right) \leq \mathbb{P} \left(\bigcup_{k=1}^{2^n-3} E_{n,k} \text{ infinitely occurs} \right) = 0.$$

This completes the proof of (4.4). To show that $t \mapsto B_t$ is *a.s.* nowhere differentiable on \mathbb{R}_+ , take intersection of an analogue of (4.4) on all intervals $[n-1, n]$, $n \in \mathbb{N}$. \square

Theorem 4.22 (Law of iterated logarithm). Let $(B_t)_{t \geq 0}$ be a Brownian motion. Then

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1 \quad a.s..$$

Proof. We first give a tail bound of $B_1 \sim N(0, 1)$, which is a standard Gaussian variable:

$$\frac{\alpha e^{-\alpha^2/2}}{(1+\alpha^2)\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \frac{e^{-t^2/2}}{(1+t^2)^2} dt \leq \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-t^2/2} dt \leq \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \frac{te^{-t^2/2}}{\alpha} dt = \frac{e^{-\alpha^2/2}}{\alpha\sqrt{2\pi}}, \quad \forall \alpha > 0.$$

Let $h(t) = \sqrt{2t \log \log t}$. Using the law of $M_t := \sup_{0 \leq s \leq t} B_s \stackrel{d}{=} |B_t|$ given in Corollary 4.25, for all $r > 1$ and all $\delta > 0$, whenever $n \geq n_0 := 1 + \frac{e}{\log r}$, we have

$$\mathbb{P} \left(\frac{M_{r^n}}{h(r^{n-1})} > \sqrt{r+\delta} \right) = \mathbb{P} \left(\frac{|B_{r^n}|}{\sqrt{r^n}} > \sqrt{2 \left(1 + \frac{\delta}{r}\right) \log \log r^{n-1}} \right) \leq C(n-1)^{-1-\frac{\delta}{r}},$$

where $C = C(\delta, r)$ is some constant independent of n . Hence we have

$$\sum_{n=n_0}^{\infty} \mathbb{P} \left(\sup_{s \in [r^{n-1}, r^n]} \frac{B_s}{h(s)} > \sqrt{r+\delta} \right) \leq \sum_{n=n_0}^{\infty} \mathbb{P} \left(\frac{M_{r^n}}{h(r^{n-1})} > \sqrt{r+\delta} \right) \leq C \sum_{n=n_0-1}^{\infty} n^{-1-\frac{\delta}{r}} < \infty.$$

By Borel-Cantelli lemma, we have

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \frac{B_s}{h(s)} \leq \sqrt{r+\delta} \right) = \mathbb{P} \left(\sup_{s \in [r^{n-1}, r^n]} \frac{B_s}{h(s)} > \sqrt{r+\delta} \text{ for finitely many } n \right) = 1.$$

Let $\delta \downarrow 0$ and $r \downarrow 1$, we have $\mathbb{P} \left(\limsup_{n \rightarrow \infty} \frac{B_s}{h(s)} \leq 1 \right) = 1$.

Now we prove $\limsup_{n \rightarrow \infty} \frac{B_s}{h(s)} \geq 1$ a.s.. Given $n \geq n_0$, we have $\sqrt{2 \log(n \log r)} > 1$, and

$$\mathbb{P} \left(B_{r^n} - B_{r^{n-1}} \geq \sqrt{\frac{r-1}{r}} h(r^n) \right) \geq \mathbb{P} \left(\frac{B_{r^n} - B_{r^{n-1}}}{\sqrt{r^{n-1}(r-1)}} \geq \sqrt{2 \log(n \log r)} \right) \geq \frac{C}{n\sqrt{\log n}}$$

for some constant $C = C(r)$ independent of n . Hence

$$\sum_{n=n_0}^{\infty} \mathbb{P} \left(B_{r^n} - B_{r^{n-1}} \geq \sqrt{\frac{r-1}{r}} h(r^n) \right) \geq C \sum_{n=n_0}^{\infty} \frac{1}{n\sqrt{\log n}} = \infty.$$

Since $(B_t)_{t \geq 0}$ has independent increments, the Borel-Cantelli lemma implies that $B_{r^n} - B_{r^{n-1}} \geq \sqrt{\frac{r-1}{r}} h(r^n)$ for infinitely many n . A symmetric argument of the first part of our proof implies $\liminf_{n \rightarrow \infty} \frac{B_s}{h(s)} \geq -1$ a.s.. Hence for a.s. $\omega \in \Omega$, we can find some $N(\omega)$ such that for all $n \geq N(\omega)$,

$$B_{r^{n-1}} \geq -2h(r^{n-1}) \geq -\frac{2}{\sqrt{r}} h(r^n) \Rightarrow B_{r^n} \geq \left(1 - \frac{3}{\sqrt{r}}\right) h(r^n) \text{ infinitely occurs.}$$

Consequently, for all $r > 1$, we have $\mathbb{P} \left(\limsup_{t \rightarrow \infty} \frac{B_t}{h(t)} \geq 1 - \frac{3}{\sqrt{r}} \right) = 1$. Letting $r \rightarrow \infty$ suffices. \square

4.4 Strong Markov Property and Reflection Principle

In this section we discuss the martingale properties of a Brownian motion $(B_t)_{t \geq 0}$. Let τ be a stopping time (with respect to the canonical filtration $(\mathcal{F}_t)_{t \geq 0}$ of $(B_t)_{t \geq 0}$). We define a random variable $\mathbb{1}_{\{\tau < \infty\}} B_\tau$:

$$\mathbb{1}_{\{\tau < \infty\}} B_\tau(\omega) = \mathbb{1}_{\{\tau(\omega) < \infty\}} B_{\tau(\omega)}(\omega) + 0 \mathbb{1}_{\{\tau(\omega) = \infty\}}.$$

This is a \mathcal{F}_τ -measurable variable. To see this, note that $(B_t)_{t \geq 0}$ is an adaptive and sample-continuous process, hence is progressive [Proposition 3.9]. Then the desired result follows from Proposition 3.12.

Theorem 4.23 (Strong Markov property). Let τ be a stopping time with $\mathbb{P}(\tau < \infty) > 0$. Let

$$B_t^{(\tau)} = \mathbb{1}_{\{\tau < \infty\}} (B_{\tau+t} - B_\tau), \quad \forall t \in \mathbb{R}_+.$$

Then $(B_t^{(\tau)})_{t \geq 0}$ is a Brownian motion under the measure $\mathbb{P}(\cdot | \tau < \infty)$, and is independent of \mathcal{F}_τ .

Proof. We first deal with the case $\tau < \infty$ a.s.. Fix $A \in \mathcal{F}_\tau$ and $0 = t_0 < t_1 < \dots < t_n$. We claim that

$$\mathbb{E}[\mathbb{1}_A g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \dots, B_{t_n}^{(\tau)})] = \mathbb{P}(A) \mathbb{E}[g(B_{t_0}, B_{t_1}, \dots, B_{t_n})]$$

for all bounded continuous functions $g : \mathbb{R}^n \rightarrow \mathbb{R}$. If we take $A = \Omega$, a similar argument to the proof of Theorem 4.16 implies that $(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \dots, B_{t_n}^{(\tau)}) \stackrel{d}{=} (B_{t_0}, B_{t_1}, \dots, B_{t_n})$ for all choices $0 = t_0 < t_1 < \dots < t_n$. Since sample paths of $(B_t^{(\tau)})_{t \geq 0}$ are continuous, Proposition 4.4 implies that $(B_t^{(\tau)})_{t \geq 0}$ is also a Brownian motion. Furthermore, $(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \dots, B_{t_n}^{(\tau)})$ is independent of \mathcal{F}_τ , and $(B_t^{(\tau)})_{t \geq 0}$ is independent of \mathcal{F}_τ .

Now we prove the claim. For $p \in \mathbb{N}$, take $[t]_p = \min\{k2^{-p} : k2^{-p} \geq t, k \in \mathbb{Z}\}$ with convention $[\infty]_p = \infty$, and write $\tau_p = [\tau]_p$. By continuity we have $B_t^{(\tau_p)} \rightarrow B_t^{(\tau)}$ a.s., and dominated convergence implies

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \dots, B_{t_n}^{(\tau)})] &= \lim_{p \rightarrow \infty} \mathbb{E}[\mathbb{1}_A g(B_{t_0}^{(\tau_p)}, B_{t_1}^{(\tau_p)}, \dots, B_{t_n}^{(\tau_p)})] \\ &= \lim_{p \rightarrow \infty} \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{1}_{A \cap \{(k-1)2^{-p} < \tau \leq k2^{-p}\}} g(B_{k2^{-p}+t_0} - B_{k2^{-p}}, B_{k2^{-p}+t_1} - B_{k2^{-p}}, \dots, B_{k2^{-p}+t_n} - B_{k2^{-p}})] \\ &= \lim_{p \rightarrow \infty} \sum_{k=0}^{\infty} \mathbb{P}(A \cap \{(k-1)2^{-p} < \tau \leq k2^{-p}\}) \mathbb{E}[g(B_{t_0}, B_{t_1}, \dots, B_{t_n})] = \mathbb{P}(A) \mathbb{E}[g(B_{t_0}, B_{t_1}, \dots, B_{t_n})], \end{aligned}$$

where the last row follows from the fact that

$$A \cap \{(k-1)2^{-p} < \tau \leq k2^{-p}\} = (A \cap \{\tau \leq k2^{-p}\}) \cap \{\tau \leq (k-1)2^{-p}\}^c \in \mathcal{F}_{k2^{-p}}$$

and simple Markovian property of $(B_t)_{t \geq 0}$. Thus we completes the proof for case $\tau < \infty$ a.s.. For the general case $\mathbb{P}(\tau < \infty) > 0$, we have

$$\mathbb{E}[\mathbb{1}_{A \cap \{\tau < \infty\}} g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \dots, B_{t_n}^{(\tau)})] = \mathbb{P}(A \cap \{\tau < \infty\}) \mathbb{E}[g(B_{t_0}, B_{t_1}, \dots, B_{t_n})]$$

Then the desired result follows in a straightforward way. \square

Theorem 4.24 (Reflection principle). Given $t > 0$, let $M_t = \sup_{0 \leq s \leq t} B_s$. For any $a > 0$ and $b \in (-\infty, a]$,

$$\mathbb{P}(M_t \geq a, B_t \leq b) = \mathbb{P}(B_t \geq 2a - b). \quad (4.5)$$

In particular, we have $M_t \stackrel{d}{=} |B_t|$.

Proof. Define the hitting time $\tau_a = \inf\{s \geq 0 : B_s = a\}$. By Proposition 4.17 (ii), we have $\tau_a < \infty$ a.s.. And by Theorem 4.23, the process $B_t^{(\tau_a)} = B_{\tau_a+t} - B_{\tau_a} = B_{\tau_a+t} - a$ is a Brownian process independent of \mathcal{F}_{τ_a} . Consequently, $(\tau_a, B^{(\tau_a)}) \stackrel{d}{=} (\tau_a, -B^{(\tau_a)})$, whose distribution equals the product of the law of τ_a and the Wiener measure μ_W on $C(\mathbb{R}_+)$. Let $H = \{(s, f) \in \mathbb{R} \times C(\mathbb{R}_+) : s \leq t, f(t-s) \leq b-a\}$. Hence

$$\begin{aligned} \mathbb{P}(M_t \geq a, B_t \leq b) &= \mathbb{P}(\tau_a \leq t, B_t \leq b) = \mathbb{P}(\tau_a \leq t, B_{t-\tau_a}^{(\tau_a)} \leq b-a) = \mathbb{P}((\tau_a, B^{(\tau_a)}) \in H) \\ &= \mathbb{P}((\tau_a, -B^{(\tau_a)}) \in H) = \mathbb{P}(\tau_a \leq t, -B_{t-\tau_a}^{(\tau_a)} \leq b-a) \\ &= \mathbb{P}(\tau_a \leq t, B_t \geq 2a-b) = \mathbb{P}(B_t \geq 2a-b). \end{aligned}$$

For the last assertion, note that

$$\mathbb{P}(M_t \geq a) = \mathbb{P}(M_t \geq a, B_t \geq a) + \mathbb{P}(M_t \geq a, B_t \leq a) = \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq 2a-a) = 2\mathbb{P}(B_t \geq a).$$

Thus we complete the proof. \square

Remark. According to (4.5), we also have

$$\mathbb{P}(M_t \leq y, B_t \leq x) = \begin{cases} \mathbb{P}(B_t \leq x) - \mathbb{P}(B_t \geq 2y-x) & \text{if } y > 0, x \leq y \\ \mathbb{P}(M_t \leq y) & \text{if } y > 0, x > y. \end{cases}$$

This gives the density of (M_t, B_t) :

$$\rho_{M_t, B_t}(y, x) = \frac{\partial}{\partial y} \left(\frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{x^2}{2t}} + e^{-\frac{(2y-x)^2}{2t}} \right) \right) = \frac{2(2y-x)}{t^{3/2}\sqrt{2\pi}} e^{-\frac{(2y-x)^2}{2t}}, \quad y > 0, x \leq y.$$

Corollary 4.25 (Law of hitting times). Given $a > 0$, the hitting time $\tau_a = \inf\{s \geq 0 : B_s = a\}$ is identically distributed to $a^2 B_1^{-2}$.

Proof. By the last assertion of Theorem 4.24, we have

$$\mathbb{P}(\tau_a \leq t) = \mathbb{P}(M_t \geq a) = \mathbb{P}(|B_t| \geq a) = \mathbb{P}(B_t^2 \geq a^2) = \mathbb{P}\left(\frac{a^2}{B_1^2} \leq t\right),$$

which holds for all $t \geq 0$. \square

Remark. Since $B_1 \sim N(0, 1)$, we can derive the density of τ_a :

$$\rho_{\tau_a}(t) = \frac{2\mathbb{1}_{\{t>0\}}}{\sqrt{2\pi}} e^{-\frac{a^2}{2t}} \frac{d}{dt} \left(\frac{a}{\sqrt{t}} \right) = \frac{a}{\sqrt{2\pi}t^3} e^{-\frac{a^2}{2t}} \mathbb{1}_{\{t>0\}}.$$

This also implies $\mathbb{E}[\tau_a] = \infty$ for all $a > 0$.

Proposition 4.26. Let $(\mathcal{F}_t)_{t \geq 0}$ be the canonical filtration of a Brownian motion $(B_t)_{t \geq 0}$. All these processes are martingales with respect to $(\mathcal{F}_t)_{t \geq 0}$: (i) B_t ; (ii) $B_t^2 - t$; (iii) $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$, $\theta \in \mathbb{R}$.

Proposition 4.27 (Laplacian transform of hitting times). Given $a > 0$, the hitting time $\tau_a = \inf\{s \geq 0 : B_s = a\}$ satisfies

$$\mathbb{E}[e^{-\lambda \tau_a}] = e^{-a\sqrt{2\lambda}}, \quad \forall \lambda > 0.$$

Proof. We consider the martingale $N_t^\theta = \exp(\theta B_t - \frac{1}{2}\theta^2 t)$, where $\theta > 0$. By Corollary 3.50, $N_{t \wedge \tau_a}^\theta$ is a martingale bounded by $e^{\theta a}$ from above, hence is uniformly integrable. As a result,

$$\mathbb{E}[N_{\tau_a}^\theta] = \lim_{t \rightarrow \infty} \mathbb{E}[N_{t \wedge \tau_a}^\theta] = \mathbb{E}[N_0^\theta] = 1.$$

Since $\tau_a < \infty$ a.s., we have

$$1 = \mathbb{E}[N_{\tau_a}^\theta] = \mathbb{E}\left[\exp\left(\theta a - \frac{1}{2}\theta^2 \tau_a\right)\right] \Rightarrow \mathbb{E}[e^{-\lambda \tau_a}] = e^{-a\sqrt{2\lambda}},$$

where the implication follows by setting $\theta = \sqrt{2\lambda}$. □

Proposition 4.28 (Exit times). Given $a \in \mathbb{R}$, set the hitting time $\tau_a = \inf\{s \geq 0 : B_s = a\}$.

(i) (Law of the exit point from an interval). For every $a < 0 < b$, we have

$$\mathbb{P}(\tau_a < \tau_b) = \frac{b}{b-a}, \quad \text{and} \quad \mathbb{P}(\tau_b < \tau_a) = \frac{-a}{b-a}.$$

(ii) (First moment of exit times). For every $a < 0 < b$, the exit time $\tau = \tau_a \wedge \tau_b$ satisfies $\mathbb{E}[\tau] = -ab$.

(iii) (Laplacian transform of exit times). For every $a > 0$ and every $\lambda > 0$, the exit time $\tau = \tau_a \wedge \tau_b$ satisfies

$$\mathbb{E}[e^{-\lambda \tau_a}] = \frac{\cosh\left(\frac{b+a}{2}\sqrt{2\lambda}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right)}.$$

Proof. (i) We define a stopping time $\tau = \tau_a \wedge \tau_b$. By Corollary 3.50, we choose the stopped martingale $(B_{t \wedge \tau})_{t \geq 0}$, which satisfies $|B_{t \wedge \tau}| \leq (-a) \vee b$, hence is uniformly integrable. As a result, $\mathbb{E}[B_\tau] = \mathbb{E}[B_0] = 0$. Note that $\tau_a \neq \tau_b$ a.s., and $\mathbb{E}[B_\tau] = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_b < \tau_a)$, the result follows.

(ii) Consider the martingale $A_t = B_t^2 - t$. Then $\mathbb{E}[A_{t \wedge \tau}] = \mathbb{E}[A_0] = 0$, which gives $\mathbb{E}[B_{t \wedge \tau}^2] = \mathbb{E}[t \wedge \tau]$. On the other hand, the monotone convergence theorem implies $\mathbb{E}[t \wedge \tau] \rightarrow \mathbb{E}[\tau]$ as $t \rightarrow \infty$. On the other hand, since $B_{t \wedge \tau}^2 \leq a^2 \vee b^2$, we have $\mathbb{E}[B_{t \wedge \tau}^2] \rightarrow \mathbb{E}[B_\tau^2]$ as $t \rightarrow \infty$ by dominated convergence theorem. Note that

$$\mathbb{E}[B_\tau^2] = \mathbb{E}[a^2 \mathbb{1}_{\{B_\tau=a\}} + b^2 \mathbb{1}_{\{B_\tau=b\}}] = a^2 \mathbb{P}(\tau_a < \tau_b) + b^2 \mathbb{P}(\tau_b < \tau_a) = -ab.$$

(iii) Similar to Proposition 4.26 (iii), we take the following martingale:

$$\begin{aligned} N_t &= \frac{1}{2} \exp\left(\sqrt{2\lambda}\left(B_t - \frac{a+b}{2}\right) - \lambda t\right) + \frac{1}{2} \exp\left(-\sqrt{2\lambda}\left(B_t - \frac{a+b}{2}\right) - \lambda t\right) \\ &= e^{-\lambda t} \cosh\left(\sqrt{2\lambda}\left(B_t - \frac{a+b}{2}\right)\right), \quad t \geq 0. \end{aligned}$$

Since $0 \leq N_{t \wedge \tau} \leq \cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right)$, it is a uniformly integrable martingale. Consequently,

$$\mathbb{E}[N_\tau] = \lim_{n \rightarrow \infty} \mathbb{E}[N_{t \wedge \tau}] = \mathbb{E}[N_0] = \cosh\left(\frac{a+b}{2}\sqrt{2\lambda}\right).$$

On the other hand, since $B_\tau \in \{a, b\}$ a.s., we have

$$\mathbb{E}[N_\tau] = \mathbb{E}\left[e^{-\lambda \tau} \left(\mathbb{1}_{\{B_\tau=a\}} \cosh\left(\frac{a-b}{2}\sqrt{2\lambda}\right) + \mathbb{1}_{\{B_\tau=b\}} \cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right)\right)\right] = \mathbb{E}[e^{-\lambda \tau}] \cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right).$$

Then the desired result follows. □

Proposition 4.29 (Time reversal). Set $\tilde{B}_t = B_1 - B_{1-t}$ for every $t \in [0, 1]$. Then $(\tilde{B}_t)_{t \in [0, 1]}$ is a Brownian motion on $[0, 1]$, which has the same law as $(B_t)_{t \in [0, 1]}$.

Proof. Clearly, $\tilde{B}_0 = 0$, and $(\tilde{B}_t)_{t \in [0, 1]}$ has continuous sample paths. We show that $(B_t)_{t \in [0, 1]}$ and $(\tilde{B}_t)_{t \in [0, 1]}$ has the same finite-dimensional marginal distributions, which are extended to the Wiener measure on $C([0, 1])$. Take a partition $0 = t_0 < t_1 < \dots < t_n = 1$. Then the increments

$$(\tilde{B}_{t_1}, \tilde{B}_{t_2} - \tilde{B}_{t_1}, \dots, \tilde{B}_{t_n} - \tilde{B}_{t_{n-1}}) = (B_1 - B_{1-t_1}, B_{1-t_1} - B_{1-t_2}, \dots, B_{1-t_{n-1}})$$

are jointly Gaussian and independent, and the desired result follows. \square

Corollary 4.30. Let $M_t = \sup_{0 \leq s \leq t} B_s$. Then $M_t - B_t \stackrel{d}{=} M_t \stackrel{d}{=} |B_t|$ for every $t > 0$.

Proof. Fix $t > 0$. Akin to Proposition 4.29, we define $\tilde{B}_s = B_t - B_{t-s}$ for every $s \in [0, t]$. By symmetry and time reversal property of Brownian motions, all $(B_s)_{s \in [0, t]}$, $(\tilde{B}_s)_{s \in [0, t]}$ and $(-\tilde{B}_s)_{s \in [0, t]}$ have the same law. Consequently, $\sup_{0 \leq s \leq t} B_t \stackrel{d}{=} \sup_{0 \leq s \leq t} (-\tilde{B}_t)$, which is in fact $M_t \stackrel{d}{=} M_t - B_t$. \square

Proposition 4.31 (Arcsine law). We let $M_1 = \sup_{0 \leq s \leq 1} B_s$, and define $\tau = \inf\{t \geq 0 : B_t = M_1\}$. Then $\tau < 1$ a.s. and τ is not a stopping time. The density of τ is given by the following *arcsine law*:

$$\rho_\tau(t) = \frac{1}{\pi \sqrt{t(1-t)}} \mathbf{1}_{(0,1)}(t).$$

The results holds true if we replace τ by $\ell = \sup\{t \in [0, 1] : B_t = 0\}$.

Proof. Define $\tilde{B}_t = B_1 - B_{1-t}$ for all $t \in [0, 1]$, which is a Brownian motion on $[0, 1]$. Then for all $\epsilon > 0$,

$$B_1 - M_1 \leq B_1 - \sup_{1-\epsilon \leq s \leq 1} B_s = \inf_{0 \leq s \leq \epsilon} \tilde{B}_s < 0 \text{ a.s..}$$

Clearly, we have $0 \leq \tau \leq 1$, and $\{\tau = 1\} \subset \{B_1 = M_1\}$. Hence $\mathbb{P}(\tau < 1) = 1$.

Fix $0 < t < 1$, and let $M_t = \sup_{0 \leq s \leq t} B_s$. By simple Markov property, $(B_{t+s} - B_t)_{s \in [0, 1-t]}$ is independent of $(B_s)_{s \in [0, t]}$. Define $N_t = \sup_{0 \leq s \leq 1-t} (B_{t+s} - B_t)$. Then N_t is independent of $(M_t, B_t, M_t - B_t)$, and

$$\mathbb{P}(\tau \leq t) = \mathbb{P}(M_t \geq B_t + N_t) = \mathbb{P}(N_t \leq M_t - B_t) = \mathbb{P}(|B_1 - B_t| \leq |B_t|).$$

Let Z_1, Z_2 be $N(0, 1)$ i.i.d., and θ is uniformly distributed on $[0, 2\pi)$. By calculus,

$$\mathbb{P}(\tau \leq t) = \mathbb{P}(\sqrt{1-t}|Z_1| \leq \sqrt{t}|Z_2|) = \mathbb{P}\left(\frac{|Z_1|}{\sqrt{Z_1^2 + Z_2^2}} \leq t\right) = \mathbb{P}(|\sin \theta| \leq \sqrt{t}) = \frac{2}{\pi} \arcsin(\sqrt{t}).$$

The density of τ follows by finding the derivative of $\frac{2}{\pi} \arcsin(\sqrt{t})$. The case for ℓ is rather straightforward:

$$\begin{aligned} \mathbb{P}(\ell \leq t) &= 2 \int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \mathbb{P}\left(\sup_{s \in [0, 1-t]} (B_t - B_{t+s}) \geq y\right) dy = 2 \int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \mathbb{P}(\tau_y \geq 1-t) dy \\ &= 2 \int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \left(\int_{1-t}^\infty \frac{y}{\sqrt{2\pi z^3}} e^{-\frac{y^2}{2z}} dz\right) dy = \frac{1}{\pi} \int_{1-t}^\infty \left(\int_0^\infty \frac{y}{\sqrt{tz^3}} e^{-\frac{y^2}{2t} - \frac{y^2}{2z}} dy\right) dz \\ &= \frac{1}{\pi} \int_{1-t}^\infty \frac{1}{\sqrt{tz^3}} \frac{tz}{t+z} dz \stackrel{s=\frac{t}{t+z}}{=} \frac{1}{\pi} \int_0^t \frac{s^{3/2}}{t\sqrt{1-s}} \frac{\partial z}{\partial s} ds = \frac{1}{\pi} \int_0^t \frac{1}{\sqrt{s(1-s)}} ds = \frac{2}{\pi} \arcsin(\sqrt{t}). \end{aligned}$$

Hence we complete the proof. \square

5 Stochastic Integration

In this chapter, our discussion is based on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a complete filtration $(\mathcal{F}_t)_{t \geq 0}$. All processes we study are indexed by \mathbb{R}_+ and take real values.

5.1 Construction of Stochastic Integrals

Preliminary: Space \mathbb{H}^2 . Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, we denote by \mathcal{H}^2 the vector space of all sample-continuous martingales $M = (M_t)_{t \geq 0}$ that are bounded in L^2 [i.e. $\sup_{t \geq 0} \mathbb{E}[M_t^2] < \infty$] with $M_0 = 0$, and we write $M \sim N$ if $M, N \in \mathcal{H}^2$ are indistinguishable. Then we define $\mathbb{H}^2 = \mathcal{H}^2 / \sim$, and for brevity we write $[M] = M$ for all $M \in \mathcal{H}^2$. By Theorem 3.58, a continuous local martingale falls in \mathbb{H}^2 if and only if $M_0 = 0$ and $\mathbb{E}[\langle M, M \rangle_\infty] < \infty$. Consequently, the martingale $M = (M_t)_{t \geq 0}$ has *a.s.* and L^2 limit X_∞ such that $\mathbb{E}[M_\infty | \mathcal{F}_t] = M_t$ for all $t \in \mathbb{R}_+$.

If $M, N \in \mathbb{H}^2$, the bracket $\langle M, N \rangle = \frac{1}{2}(\langle M+N, M+N \rangle - \langle M, M \rangle - \langle N, N \rangle)$ then satisfies $\mathbb{E}[\langle M, N \rangle_\infty] < \infty$. This gives rise to a bilinear form:

$$\langle M, N \rangle_{\mathbb{H}^2} = \mathbb{E}[\langle M, N \rangle_\infty] = \mathbb{E}[M_\infty N_\infty] \quad \Rightarrow \quad \|M\|_{\mathbb{H}^2} = \mathbb{E}[\langle M, M \rangle_\infty] = \mathbb{E}[M_\infty^2].$$

One can easily show that $\langle \cdot, \cdot \rangle_{\mathbb{H}^2}$ forms an inner product on \mathbb{H}^2 , of which positive definiteness follows from Proposition 3.57. Furthermore, $(\mathbb{H}^2, \langle \cdot, \cdot \rangle_{\mathbb{H}^2})$ is a Hilbert space.

Proof of completeness. Let $M^{(n)} \in \mathbb{H}^2$ be a Cauchy sequence with respect to the norm $\|\cdot\|_{\mathbb{H}^2}$. Then

$$\lim_{n, m \rightarrow \infty} \mathbb{E} \left[\left(M_\infty^{(n)} - M_\infty^{(m)} \right)^2 \right] = \lim_{n, m \rightarrow \infty} \|M^{(n)} - M^{(m)}\|_{\mathbb{H}^2}^2 = 0.$$

Then $M_\infty^{(n)}$ is a Cauchy sequence in L^2 , and we denote $Z = \lim_{n \rightarrow \infty} M_\infty^{(n)}$ in L^2 . On the other hand, the Doob's L^2 -inequality [Proposition 3.40 (ii)] and an argument of dominated convergence theorem imply

$$\mathbb{E} \left[\sup_{t \geq 0} \left| M_t^{(n)} - M_t^{(m)} \right|^2 \right] \leq 4 \mathbb{E} \left[\left(M_\infty^{(n)} - M_\infty^{(m)} \right)^2 \right].$$

Hence $(M_t^{(n)})_{n=1}^\infty$ is a Cauchy sequence for every $t \geq 0$, which converges in L^2 . To conclude the proof, it suffices to show that the limit process is in \mathbb{H}^2 . We choose a subsequence $M^{(n_k)}$ such that

$$\mathbb{E} \left[\sup_{t \geq 0} \left| M_t^{(n_{k+1})} - M_t^{(n_k)} \right|^2 \right] < 2^{-k} \quad \Rightarrow \quad \mathbb{E} \left[\sup_{t \geq 0} \left| M_t^{(n_{k+1})} - M_t^{(n_k)} \right| \right] \leq 2^{-k/2}.$$

Consequently, we have $\sum_{k=1}^\infty \sup_{t \geq 0} |M_t^{(n_{k+1})} - M_t^{(n_k)}| < \infty$ *a.s.* By Weierstrass M-test, the limit process $M_t = \lim_{k \rightarrow \infty} M_t^{(n_k)}$ is a *a.s.* uniform limit on \mathbb{R}_+ , hence has continuous sample paths. On the zero probability set where the uniform convergence does not hold, we take $M_t = 0$ for each $t > 0$. By completeness of the filtration $(\mathcal{F}_t)_{t \geq 0}$, the process $M = (M_t)_{t \geq 0}$ is adapted. Also, the continuity of conditional expectation passes $M_t^{(n_k)} = \mathbb{E}[M_\infty^{(n_k)} | \mathcal{F}_t]$ to $M_t = \mathbb{E}[Z | \mathcal{F}_t]$ as $k \rightarrow \infty$, hence $(M_t)_{t \geq 0}$ is a uniformly integrable martingale, which converges to M_∞ *a.s.* and in L^2 . By uniform convergence, we have $M_\infty = Z$ *a.s.* Therefore

$$\lim_{n \rightarrow \infty} \|M^{(n)} - M\|_{\mathbb{H}^2}^2 = \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(M_\infty^{(n)} - M_\infty \right)^2 \right] = \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(M_\infty^{(n)} - Z \right)^2 \right] = 0.$$

Thus $M \in \mathbb{H}^2$ is indeed the limit of sequence $M^{(n)}$, completing the proof. \square

Preliminary: Progressive σ -fields. Given $(\Omega, \mathcal{F}, \mathbb{P})$, we define the progressive σ -field on $\Omega \times \mathbb{R}_+$ as

$$\mathcal{P} = \{A \subset \Omega \times \mathbb{R}_+ : A \cap (\Omega \times [0, t]) \in \mathcal{F} \otimes \mathcal{B}([0, t]), \forall t \in \mathbb{R}_+\} \subset \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+),$$

where the inclusion holds since $A = \bigcup_{n=1}^{\infty} (A \cap (\Omega \times [0, n]))$. Clearly, if $A \in \mathcal{P}$, the process $X_t(\omega) = \mathbf{1}_A(\omega, t)$ is a progressive process. Furthermore, one can verify that \mathcal{P} is indeed a σ -algebra on $\Omega \times \mathbb{R}_+$, and a process $(X_t)_{t \geq 0}$ is progressive if and only if the mapping $(\omega, t) \mapsto X_t(\omega)$ is \mathcal{P} -measurable.

Preliminary: Space $L^2(M)$. Given a process $M \in \mathbb{H}^2$, the Theorem 3.56 determines to an increasing process $(\langle M, M \rangle_s)_{s \geq 0}$, which is called the quadratic variation of M . Then for every $A \in \mathcal{P}$, one can define

$$\mu_M(A) = \mathbb{E} \left[\int_0^\infty \mathbf{1}_A(\cdot, s) d\langle M, M \rangle_s \right].$$

This is a measure on $(\Omega \times \mathbb{R}_+, \mathcal{P})$. We denote by $\mathcal{L}^2(M)$ the space of all progressive processes H such that

$$\|H\|_{L^2(M)}^2 = \mathbb{E} \left[\int_0^\infty H_s^2 d\langle M, M \rangle_s \right] < \infty,$$

and choose the quotient space $L^2(M)$ that makes $\|\cdot\|_{L^2(M)}$ a proper norm. Then $L^2(M) = L^2(\Omega \times \mathbb{R}_+, \mathcal{P}, \mu_M)$ can be viewed as an ordinary L^2 -space, and we can define the inner product

$$\langle H, K \rangle_{L^2(M)} = \mathbb{E} \left[\int_0^\infty H_s K_s d\langle M, M \rangle_s \right]$$

Remark. Recall that $X_t^\tau = X_{t \wedge \tau}$ is the stopped process associated with a stopping time τ . If $M \in \mathbb{H}^2$, then we have $\langle M^\tau, M^\tau \rangle_\infty = \langle M, M \rangle_\tau$, which implies that $M^\tau \in \mathbb{H}^2$. Furthermore, if $H \in L^2(M)$, the process $\mathbf{1}_{[0, \tau]} H$ defined by $(\mathbf{1}_{[0, \tau]} H)_s(\omega) = \mathbf{1}_{[0, \tau(\omega)]}(s) H_s(\omega)$ also belongs to $L^2(M)$. Note that $\mathbf{1}_{[0, \tau]} H$ is progressive since it has left-continuous sample paths.

Definition 5.1 (Elementary processes). An *elementary process* is a progressive process of the form

$$H_s(\omega) = \sum_{j=1}^n H_{(j)}(\omega) \mathbf{1}_{(t_{j-1}, t_j]}(s),$$

where $0 = t_0 < t_1 < \dots < t_n$, and $H_{(j)}$ is a bounded $\mathcal{F}_{t_{j-1}}$ -measurable random variable for all $j \in \{1, \dots, n\}$. Clearly, the set \mathcal{E} of all (equivalence classes of) elementary processes is a subspace of $L^2(M)$.

Proposition 5.2. For every $M \in \mathbb{H}^2$, \mathcal{E} is dense in $L^2(M)$.

Proof. Fix $M \in \mathbb{H}^2$. By elementary Hilbert space theory, it suffices to show that $L^2(M) \ni K \perp \mathcal{E}$ implies $K = 0$. Assume that $K \in L^2(M)$ is orthogonal to \mathcal{E} , and set

$$X_t = \int_0^t K_s d\langle M, M \rangle_s, \quad \forall t \geq 0.$$

According to Proposition 3.52, since

$$\mathbb{E} \left[\int_0^t |K_s| d\langle M, M \rangle_s \right] \leq \left(\mathbb{E} \left[\int_0^t K_s^2 d\langle M, M \rangle_s \right] \right)^{1/2} \left(\mathbb{E} \left[\int_0^t d\langle M, M \rangle_s \right] \right)^{1/2} \leq \|K\|_{L^2(M)} \|M\|_{\mathbb{H}^2},$$

the process $(X_t)_{t \geq 0}$ is a finite-variation process. In addition, it is bounded in L^1 .

Now we prove that $(X_t)_{t \geq 0}$ is a continuous martingale. Given $0 \leq s < t$, let $H_r = Y \mathbb{1}_{(s,t]}$, where Y is a bounded \mathcal{F}_s -measurable random variable. Then

$$0 = \langle H, K \rangle_{L^2(M)} = \mathbb{E} \left[Y \int_s^t K_u d\langle M, M \rangle_u \right] = \mathbb{E}[Y(X_t - X_s)]$$

Since $\mathbb{E}[Y(X_t - X_s)] = 0$ for all bounded \mathcal{F}_s -measurable random variable Y , we have $\mathbb{E}[X_t - X_s | \mathcal{F}_s] = 0$. Note that $X = (X_t)_{t \geq 0}$ is adapted, and by definition it has continuous sample paths. Hence X is a continuous martingale. By Proposition 3.55, we have $X = 0$ a.s., i.e.

$$\int_0^t K_s d\langle M, M \rangle_s = 0 \quad \forall t \geq 0, \quad a.s. \quad \Rightarrow \quad K_s = 0 \quad d\langle M, M \rangle_s\text{-a.e.}, \quad a.s..$$

Therefore $\|K\|_{L^2(M)} = 0$, and the result follows. \square

Theorem 5.3 (Stochastic integrals for L^2 -bounded martingales). Let $M \in \mathbb{H}^2$. For every elementary process $H \in \mathcal{E}$, we define the following formula:

$$H_s = \sum_{j=1}^n H_{(j)} \mathbb{1}_{(t_{j-1}, t_j]}(s) \quad \Rightarrow \quad (H \cdot M)_t = \sum_{j=1}^n H_{(j)} (M_{t_j \wedge t} - M_{t_{j-1} \wedge t})$$

This defines a process $H \cdot M \in \mathbb{H}^2$, and the mapping $H \mapsto H \cdot M$ extends to an isometry from $L^2(M)$ into \mathbb{H}^2 . Furthermore, $H \cdot M$ is the unique martingale of \mathbb{H}^2 that satisfies the property

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle, \quad \forall N \in \mathbb{H}^2, \quad (5.1)$$

where the quantity $H \cdot \langle M, N \rangle$ in the right-hand side is the integral with respect to a finite variation process. If τ is a stopping time, we then have

$$(\mathbb{1}_{[0, \tau]} H) \cdot M = (H \cdot M)^\tau = H \cdot M^\tau. \quad (5.2)$$

The process $H \cdot M$ is called the *stochastic integral of H with respect to M* .

Proof. Since the process $H \cdot M$ does not depend on the choice of partition when H is given, it is easy to see that $H \mapsto H \cdot M$ is a linear mapping. Then we verify that $H \mapsto H \cdot M$ is an isometry from \mathcal{E} into $L^2(M)$.

We fix the process $H = (H_s)_{s \geq 0}$ of the form given in the theorem. For every $j \in \{1, \dots, n\}$, define $M_t^{(j)} = H_{(j)}(M_{t_j \wedge t} - M_{t_{j-1} \wedge t})$ for all $t \geq 0$. Akin to our proof of Theorem 3.56 at Step II, the process $(M_t^{(j)})_{t \geq 0}$ is a continuous martingale, and it belongs to \mathbb{H}^2 . Hence $H \cdot M = \sum_{j=1}^n M^{(j)}$ also belongs to \mathbb{H}^2 . Moreover, note that

$$\begin{aligned} \left\langle M^{(j)}, M^{(j)} \right\rangle_t &= \sum_{j=1}^n H_{(j)}^2 (\langle M, M \rangle_{t_j \wedge t} - \langle M, M \rangle_{t_{j-1} \wedge t}), & (\text{By the approximation formula}) \\ \left\langle M^{(j)}, M^{(k)} \right\rangle_{\mathbb{H}^2} &= \mathbb{E} [M_\infty^{(j)} M_\infty^{(k)}] = \mathbb{E} [H_{(j)} H_{(k)} (M_{t_j} - M_{t_{j-1}})(M_{t_k} - M_{t_{k-1}})] \\ &= \mathbb{E} [\mathbb{E} [H_{(j)} H_{(k)} (M_{t_j} - M_{t_{j-1}})(M_{t_k} - M_{t_{k-1}}) | \mathcal{F}_{t_{j-1}}]] = 0, \quad \forall 1 \leq j < k \leq n. \end{aligned}$$

By orthogonality of $M^{(j)}$'s and bilinearity of quadratic variation, it holds

$$\langle H \cdot M, H \cdot M \rangle_t = \sum_{j=1}^n H_{(j)}^2 (\langle M, M \rangle_{t_j \wedge t} - \langle M, M \rangle_{t_{j-1} \wedge t}) = \int_0^t H_s^2 d\langle M, M \rangle_s.$$

Consequently, we have

$$\|H \cdot M\|_{\mathbb{H}^2}^2 = \mathbb{E}[\langle H \cdot M, H \cdot M \rangle_\infty] = \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M, M \rangle_s\right] = \|H\|_{L^2(M)}^2, \quad \forall t \geq 0.$$

By linearity, if $H = H'$ in $L^2(M)$, then $H \cdot M = H' \cdot M$ in \mathbb{H}^2 . Therefore the mapping $\mathcal{E} \rightarrow \mathbb{H}^2$ is well-defined. Since it is norm-preserving and linear, it is an isometry. By Proposition 5.2 and the fact that \mathbb{H}^2 is complete, we can uniquely extend this mapping to an isometry from $L^2(M)$ into \mathbb{H}^2 .

Now we verify the property (5.1). We fix $N \in \mathbb{H}^2$. The Kunita-Watanabe inequality [Theorem 3.62] implies

$$\mathbb{E}\left[\int_0^\infty |H_s| |d\langle M, N \rangle_s|\right] \leq \|H\|_{L^2(M)} \|N\|_{\mathbb{H}^2} < \infty, \quad \forall H \in L^2(M).$$

Then the variable $(H \cdot \langle M, N \rangle)_\infty = \int_0^\infty H_s d\langle M, N \rangle_s$ is well-defined and in L^1 . For the case where H is an elementary process, we have

$$\begin{cases} \langle H \cdot M, N \rangle = \sum_{j=1}^n \langle M^{(j)}, N \rangle \\ \langle M^{(j)}, N \rangle_t = H_{(j)}^2 (\langle M, N \rangle_{t_j \wedge t} - \langle M, N \rangle_{t_{j-1} \wedge t}), \quad \forall t \geq 0. \end{cases}$$

This gives (5.1) in the case $H \in \mathcal{E}$:

$$\langle H \cdot M, N \rangle_t = \sum_{j=1}^n H_{(j)}^2 (\langle M, N \rangle_{t_j \wedge t} - \langle M, N \rangle_{t_{j-1} \wedge t}) = \int_0^t H_s d\langle M, N \rangle_s = (H \cdot \langle M, N \rangle)_t, \quad \forall t \geq 0.$$

To prove the general case where $H \in L^2(M)$, note the continuity of the linear mapping $X \mapsto \langle X, N \rangle_\infty$ from \mathbb{H}^2 into $L^1(\Omega, \mathcal{F}, \mathbb{P})$:

$$\mathbb{E}[\|\langle X, N \rangle_\infty\|] = \mathbb{E}[\langle X, X \rangle_\infty]^{1/2} \mathbb{E}[\langle N, N \rangle_\infty]^{1/2} = \|X\|_{\mathbb{H}^2} \|N\|_{\mathbb{H}^2}.$$

Let $H^{(n)} \in \mathcal{E}$ be a sequence that converges to H in $L^2(M)$. Then $H^{(n)} \cdot M \rightarrow H \cdot M$ in \mathbb{H}^2 , and

$$\langle H \cdot M, N \rangle_\infty = \lim_{n \rightarrow \infty} \langle H^{(n)} \cdot M, N \rangle_\infty = \lim_{n \rightarrow \infty} (H^{(n)} \cdot \langle M, N \rangle)_\infty = (H \cdot \langle M, N \rangle)_\infty,$$

where the last equality holds in L^1 by Kunita-Watanabe:

$$\mathbb{E}\left[\left|\int_0^\infty (H_s^{(n)} - H_s) d\langle M, N \rangle_s\right|\right] \leq \|H_s^{(n)} - H_s\|_{L^2(M)} \|N\|_{\mathbb{H}^2}.$$

Hence we have $\langle H \cdot M, N \rangle_\infty = (H \cdot \langle M, N \rangle)_\infty$. By replacing N with the stopped martingale N^t for any $t \geq 0$, one obtain $\langle H \cdot M, N \rangle_t = (H \cdot \langle M, N \rangle)_t$. For uniqueness, let $X \in \mathbb{H}^2$ satisfy (5.1). Then $\langle H \cdot M - X, N \rangle = 0$ for all $N \in \mathbb{H}^2$, which implies $\langle H \cdot M - X, H \cdot M - X \rangle = 0$. By Proposition 3.57 (ii), we have $H \cdot M - X = 0$ a.s..

Finally it remains to show (5.2). By Proposition 3.61 (iv), for all $N \in \mathbb{H}^2$, we have

$$\langle (H \cdot M)^\tau, N \rangle_t = \langle H \cdot M, N \rangle_{t \wedge \tau} = (H \cdot \langle M, N \rangle)_{t \wedge \tau} = (\mathbf{1}_{[0, \tau]} H \cdot \langle M, N \rangle)_t,$$

which implies the first inequality. The proof for the second one is similar:

$$\langle H \cdot M^\tau, N \rangle = H \cdot \langle M^\tau, N \rangle = H \cdot \langle M, N \rangle^\tau = \mathbf{1}_{[0, \tau]} H \cdot \langle M, N \rangle.$$

Note that the property (5.1) can be used as an alternative definition of the stochastic integral $H \cdot M$. \square

Remark. We use the following notation for a stochastic integral:

$$\int_0^t H_s dM_s = (H \cdot M)_t, \quad \forall t \geq 0, \quad \text{and} \quad \int_0^\infty H_s dM_s = (H \cdot M)_\infty.$$

The property (5.1) gives commutativity of stochastic integral and bracket:

$$\left\langle \int_0^\cdot H_s dM_s, N \right\rangle_t = \int_0^t H_s d\langle M, N \rangle_s$$

The following proposition concerns about associativity.

Proposition 5.4 (Associativity). Let $K = (K_s)_{s \geq 0}$ and $H = (H_s)_{s \geq 0}$ be two progressive progresses.

- (i) Let $A = (A_s)_{s \geq 0}$ be a finite variation process, and $\int_0^\infty |H_s| |dA_s| < \infty$ a.s.. If $\int_0^\infty |K_s H_s| |dA_s| < \infty$ a.s., then $(KH) \cdot A = K \cdot (H \cdot A)$.
- (ii) Let $M \in \mathbb{H}^2$, and $H \in L^2(M)$. Then $KH \in L^2(M)$ if and only if $K \in L^2(H \cdot M)$. In this case, we have $(KH) \cdot M = K \cdot (H \cdot M)$.

Proof. The statement (i) follows from an analogous deterministic result. Using the property (5.1) twice and (i) gives $\langle H \cdot M, H \cdot M \rangle = H^2 \cdot \langle M, M \rangle$, and $K^2 \cdot \langle H \cdot M, H \cdot M \rangle = K^2 H^2 \cdot \langle M, M \rangle$. Then the first assertion of (ii) follows from a monotone convergence argument:

$$\mathbb{E} \left[(K^2 H^2 \cdot \langle M, M \rangle)_\infty \right] = \mathbb{E} \left[(K^2 \cdot \langle H \cdot M, H \cdot M \rangle)_\infty \right]$$

For the second assertion, note that

$$\langle (KH) \cdot M, N \rangle = KH \cdot \langle M, N \rangle = K \cdot (H \cdot \langle M, N \rangle) = K \cdot \langle H \cdot M, N \rangle, \quad \forall N \in \mathbb{H}^2.$$

The result immediately follows from the uniqueness argument in Theorem 5.3. □

Remark. Let $M, N \in \mathbb{H}^2$, $H \in L^2(M)$ and $K \in L^2(N)$. Using (5.1) and (i) gives a more general result:

$$\langle H \cdot M, K \cdot N \rangle_t = \left\langle \int_0^\cdot H_s dM_s, \int_0^\cdot K_s dN_s \right\rangle_t = \int_0^t H_s K_s d\langle M, N \rangle_s.$$

According to Proposition 3.61 (vi), we have

$$\mathbb{E} \left[\left(\int_0^t H_s dM_s \right) \left(\int_0^t K_s dN_s \right) \right] = \mathbb{E} \left[\int_0^t H_s K_s d\langle M, N \rangle_s \right].$$

Note that $H \cdot M = \int_0^\cdot H_s dM_s$ is a martingale of \mathbb{H}^2 . For all $0 \leq s < t \leq \infty$, we have

$$\mathbb{E} \left[\int_0^t H_s dM_s \right] = 0, \quad \text{and} \quad \mathbb{E} \left[\int_0^t H_u dM_u \middle| \mathcal{F}_s \right] = \int_0^s H_u dM_u.$$

According to Theorem 3.58, the second moment of the stochastic integral is given by

$$\mathbb{E} \left[\left(\int_0^t H_s dM_s \right)^2 \right] = \mathbb{E} \left[\int_0^t H_s^2 d\langle M, M \rangle_s \right].$$

Next we discuss stochastic integrals for local martingales.

Preliminaries: Space $L^2_{\text{loc}}(M)$. Similar to the case $M \in \mathbb{H}^2$, we can define a Hilbert space $L^2(M)$ associated with a continuous local martingale M containing all (equivalent classes of) progressive processes H such that $\mathbb{E} \left[\int_0^\infty H_s^2 d\langle M, M \rangle_s \right] < \infty$. Furthermore, we denote by $L^2_{\text{loc}}(M)$ the set of all progressive processes such that

$$\int_0^t H_s^2 d\langle M, M \rangle_s < \infty, \quad \forall t \geq 0, \quad a.s..$$

Clearly, $L^2(M)$ is a subspace of $L^2_{\text{loc}}(M)$.

Theorem 5.5 (Stochastic integrals for continuous local martingales). Let M be a continuous local martingale. For every $H \in L^2_{\text{loc}}(M)$, there exists a unique continuous local martingale starting from 0, denoted by $H \cdot M$, such that for every continuous local martingale N ,

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle. \quad (5.3)$$

If τ is a stopping time, we then have

$$(1_{[0, \tau]} H) \cdot M = (H \cdot M)^\tau = H \cdot M^\tau. \quad (5.4)$$

In addition, if $K = (K_s)_{s \geq 0}$ is a progressive process, then $KH \in L^2_{\text{loc}}(M)$ if and only if $K \in L^2_{\text{loc}}(H \cdot M)$. In this case, we have $(KH) \cdot M = K \cdot (H \cdot M)$.

Proof. Without loss of generality, we assume that $M_0 = 0$, since we can replace $(M_t)_{t \geq 0}$ by $(M_t - M_0)_{t \geq 0}$. We also assume that the property $\int_0^s H_s^2 d\langle M, M \rangle_s < \infty$ for all $t \geq 0$ holds for all $\omega \in \Omega$ by resetting $H = 0$ on a negligible set if required. For all $n \in \mathbb{N}$, we choose a sequence of stopping times τ_n increasing to ∞ as follows:

$$\tau_n = \inf \left\{ t \geq 0 : \int_0^t (1 + H_s^2) d\langle M, M \rangle_s \geq n \right\}.$$

By definition, $\langle M^{\tau_n}, M^{\tau_n} \rangle_t = \langle M, M \rangle_{t \wedge \tau_n} \leq n$, hence the stopped martingale M^{τ_n} belongs to \mathbb{H}^2 . Furthermore,

$$\int_0^\infty H_s^2 d\langle M^{\tau_n}, M^{\tau_n} \rangle_s = \int_0^{\tau_n} H_s^2 d\langle M, M \rangle_s \leq n$$

Hence $H \in L^2(M^{\tau_n})$, and the definition of $H \cdot M^{\tau_n}$ make sense by Theorem 5.3. Note that for all $n > m \geq 1$, the property (5.2) implies $(H \cdot M^{\tau_n})^{\tau_m} = H \cdot M^{\tau_m}$. Let $(H \cdot M)_t = \lim_{n \rightarrow \infty} (H \cdot M^{\tau_n})_t$ for every $t \geq 0$, where the limit exists for all $\omega \in \Omega$ (we find m with $\tau_m(\omega) \geq t$, then $(H \cdot M^{\tau_n})_t(\omega) = (H \cdot M^{\tau_m})_t(\omega)$ for all $n \geq m$). Then $H \cdot M$ is an adapted process, and $(H \cdot M)^{\tau_n} = \lim_{m \rightarrow \infty} (H \cdot M^{\tau_m})^{\tau_n} = H \cdot M^{\tau_n} \in \mathbb{H}^2$. Consequently, $H \cdot M$ has continuous sample paths, and is a continuous local martingale.

Now we verify (5.3). Let N be a continuous local martingale with $N_0 = 0$, and choose stopping times $\tau'_n = \inf\{t \geq 0 : |N_t| \geq n\}$, $\sigma_n = \tau_n \wedge \tau'_n$. Then $N^{\tau'_n} \in \mathbb{H}^2$. Note that $M^{\tau_n} \in \mathbb{H}^2$, and $H \in L^2(M^{\tau_n})$. Hence

$$\langle H \cdot M, N \rangle^{\sigma_n} = \langle (H \cdot M)^{\tau_n}, N^{\tau'_n} \rangle = \langle H \cdot M^{\tau_n}, N^{\tau'_n} \rangle = H \cdot \langle M^{\tau_n}, N^{\tau'_n} \rangle = H \cdot \langle M, N \rangle^{\sigma_n} = (H \cdot \langle M, N \rangle)^{\sigma_n}.$$

Since $\sigma_n \rightarrow \infty$ as $n \rightarrow \infty$, the equality (5.3) follows. The uniqueness of $H \cdot M$ follows from a similar argument presented in the proof of Theorem 5.3. The equality (5.4) is a consequence of (5.3), as is shown in the proof of Theorem 5.3. Finally, the associativity follows in an analogous way in Proposition 5.4. \square

Remark: Consistency of two definitions. If $M \in \mathbb{H}^2$ and $H \in L^2(M)$, the two definitions of $H \cdot M$ given in Theorems 5.3 and 5.5 coincide. To see this, note that the definition of $H \cdot M$ in Theorem 5.5 satisfies $H \cdot M \in \mathbb{H}^2$. This is a consequence of the property (5.3), which gives $\langle H \cdot M, H \cdot M \rangle = H^2 \cdot \langle M, M \rangle$.

Remark: Connection to Wiener's integral. If $B = (B_t)_{t \geq 0}$ is a Brownian motion, we fix the filtration $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$ so that B is a continuous martingale. According to the Remark of Proposition 4.4, for each $h \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, we can define the Wiener integral $\int_0^t h(s) dB_s = W(h\mathbb{1}_{[0,t]})$, where W is the Gaussian white noise associated with B . This definition coincides with the stochastic integral $(h \cdot B)_t$, where we view h as a (deterministic) progressive process. For all step functions $h = \sum_{j=1}^n \lambda_j \mathbb{1}_{(t_{j-1}, t_j]}$, we have

$$\int_0^t h(s) dB_s = W(h\mathbb{1}_{[0,t]}) = \sum_{j=1}^n \lambda_j (B_{t_j \wedge t} - B_{t_{j-1} \wedge t}).$$

Then for all continuous local martingales N , we have

$$\left\langle \int_0^\cdot h(s) dB_s, N \right\rangle_t = \sum_{j=1}^n \lambda_j (\langle B_{t_j \wedge \cdot}, N \rangle_t - \langle B_{t_{j-1} \wedge \cdot}, N \rangle_t) = \sum_{j=1}^n \lambda_j (\langle B, N \rangle_{t_j \wedge t} - \langle B, N \rangle_{t_{j-1} \wedge t}) = (h \cdot \langle B, N \rangle)_t.$$

Therefore we have $\int_0^t h(s) dB_s = (h \cdot B)_t$ for all step functions h . For the general case $h \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$, the quadratic variation $\langle B, B \rangle_t = t$ implies $h \in L^2(B)$, and the result follows from a density argument.

Remark: Moment formulae. In the setting of Theorem 5.5, we again write $\int_0^t H_s dM_s = (H \cdot M)_t$. If M, N are continuous local martingales, $H \in L^2_{\text{loc}}(M)$ and $K \in L^2_{\text{loc}}(N)$, the first formula in the Remark under Proposition 5.4 still holds true for $\langle H \cdot M, K \cdot N \rangle$:

$$\langle H \cdot M, K \cdot N \rangle_t = \int_0^t H_s K_s d\langle M, N \rangle_s, \quad \forall t \geq 0;$$

whereas the formulae for moments of $\int_0^t H_s dM_s$ may fail.

We can make an extension. For a continuous local martingale M and a progressive process $H \in L^2_{\text{loc}}(M)$, and for some fixed $t > 0$, assume the following condition holds:

$$\mathbb{E} \left[\int_0^t H_s^2 d\langle M, M \rangle_s \right] < \infty. \quad (5.5)$$

According to Theorem 3.58, the stopped process $(H \cdot M)^t$ is a martingale of \mathbb{H}^2 . As a result, we have

$$\mathbb{E} \left[\int_0^t H_s dM_s \right] = 0, \quad \mathbb{E} \left[\left(\int_0^t H_s dM_s \right)^2 \right] = \mathbb{E} \left[\int_0^t H_s^2 d\langle M, M \rangle_s \right].$$

Consequently, regardless of whether the condition (5.5) holds, we have the following bound:

$$\mathbb{E} \left[\left(\int_0^t H_s dM_s \right)^2 \right] \leq \mathbb{E} \left[\int_0^t H_s^2 d\langle M, M \rangle_s \right]. \quad (5.6)$$

This result remains true if we replace t by a stopping time τ .

Preliminary: Locally bounded processes. A progressive process $H = (H_s)_{s \geq 0}$ is said to be *locally bounded* if $\sup_{0 \leq s \leq t} |H_s| < \infty$ a.s. for all $t > 0$. In this case, for every finite variation process V , one have

$$\int_0^t |H_s| |dV_s| \leq \sup_{0 \leq s \leq t} |H_s| \left(\int_0^t |dV_s| \right) < \infty \quad \text{a.s.,} \quad \forall t > 0.$$

In particular, an adapted and sample-continuous process is locally bounded progressive process.

Theorem 5.6 (Stochastic integrals for continuous semimartingales). Let X be a continuous semimartingale with canonical decomposition $X = M + A$. If H is a locally bounded progressive process, the stochastic integral $H \cdot X$ is the continuous semimartingale with canonical decomposition $H \cdot X = H \cdot M + H \cdot A$. The following properties hold for this stochastic integral:

- (i) The mapping $(H, X) \mapsto H \cdot X$ is bilinear.
- (ii) If K is another locally bounded progressive process, then $K \cdot (H \cdot A) = (KH) \cdot A$.
- (iii) For a stopping time τ , we have $H \mathbb{1}_{[0, \tau]} \cdot X = (H \cdot X)^\tau = H \cdot X^\tau$.
- (iv) If X is a continuous local martingale, resp. if X is a finite variation process, so is $H \cdot X$;
- (v) If H is of the form $H_s(\omega) = \sum_{j=1}^n H_{(j)}(\omega) \mathbb{1}_{(t_{j-1}, t_j]}(s)$, where $0 = t_0 < t_1 < \dots < t_n$, and $H_{(j)}$ is a $\mathcal{F}_{t_{j-1}}$ -measurable random variable for every $j \in \{1, \dots, n\}$, then

$$(H \cdot X)_t = \sum_{j=1}^n H_{(j)} (X_{t_j \wedge t} - X_{t_{j-1} \wedge t}).$$

Proof. The properties (i)-(iv) follow from the results obtained when X is a continuous local martingale, resp. a finite variation process. To obtain (iv), it suffices to consider the case where $X = M$ is a continuous local martingale with $M_0 = 0$. We may even assume that $M \in \mathbb{H}^2$ by stopping it at a suitable time and using (5.2). We choose the following sequence of stopping times, with the convention $\inf \emptyset = \infty$:

$$\tau_k = \inf \{t \geq 0 : |H_s| \geq k\} = \inf \{t_{j-1} : |H_{(j)}| \geq k\}.$$

Then $\tau_k \uparrow \infty$ as $k \rightarrow \infty$. Furthermore, for every k ,

$$H_s \mathbb{1}_{[0, \tau_k]}(s) = \sum_{j=1}^n H_{(j)}^k \mathbb{1}_{(t_{j-1}, t_j]}(s), \quad \text{where } H_{(j)}^k = H_{(j)} \mathbb{1}_{\{\tau_k \geq t_{j-1}\}} \leq k.$$

Consequently, $H \mathbb{1}_{[0, \tau_k]}$ is an elementary process, and its stochastic integral with respect to $M \in \mathbb{H}^2$ is

$$(H \cdot M)_{t \wedge \tau_k} = (H \mathbb{1}_{[0, \tau_k]} \cdot M)_t = \sum_{j=1}^n H_{(j)}^k (X_{t_j \wedge t} - X_{t_{j-1} \wedge t}).$$

Then the desired result follows by letting $k \rightarrow \infty$. □

We introduce an important convergence result for stochastic integrals.

Theorem 5.7 (Dominated convergence theorem for stochastic integrals). Let $X = M + A$ be the canonical decomposition of a continuous semimartingale X , and let $t \geq 0$. Let $(H^n)_{n=1}^\infty$ be a sequence of locally bounded progressive processes such that $\lim_{n \rightarrow \infty} H_s^n = H_s$ a.s. for every $s \in [0, t]$, where H is a locally bounded progressive process. Let K be a nonnegative progressive process such that

$$\int_0^t K_s^2 d\langle M, M \rangle_s < \infty \text{ a.s.}, \quad \text{and} \quad \int_0^t K_s |dA_s| < \infty \text{ a.s.} \quad (5.7)$$

If the sequence $(H^n)_{n=1}^\infty$ is dominated by K , i.e. $|H_s^n| \leq K_s$ a.s. for every $n \in \mathbb{N}$ and every $s \in [0, t]$, then the following convergence result holds in probability:

$$\int_0^t H_s dX_s = \lim_{n \rightarrow \infty} \int_0^t H_s^n dX_s.$$

Note the property (5.7) holds if K is a locally bounded progressive process.

Proof. The convergence $\int_0^t H_s^n dA_s \rightarrow \int_0^t H_s dA_s$ a.s. holds by Lebesgue dominated convergence theorem. So we only need to deal with the case $X = M$ is a continuous local martingale. For every $p \in \mathbb{N}$, choose the following stopping time:

$$\tau_p = \inf \left\{ r \in [0, t] : \int_0^r K_s^2 d\langle M, M \rangle_s \geq p \right\} \wedge t.$$

By assumption (5.7), we have $\mathbb{P}(\tau_p = t) \rightarrow 1$ as $p \rightarrow \infty$. The bound (5.6) of stopping time version gives the following estimate of second moment:

$$\mathbb{E} \left[\left(\int_0^{\tau_p} H_s^n dM_s - \int_0^{\tau_p} H_s dM_s \right)^2 \right] = \mathbb{E} \left[\int_0^{\tau_p} (H_s^n - H_s)^2 d\langle M, M \rangle_s \right].$$

Since $\int_0^{\tau_p} K_s^2 d\langle M, M \rangle_s \leq p$, and $|H_s^n - H_s| \leq 2K_s$ a.s., and $H_s^n \rightarrow H_s$ a.s. for all $s \in [0, t]$, by Lebesgue dominated convergence theorem, the right-hand side converges to 0 as $n \rightarrow \infty$. Then

$$\mathbb{P} \left(\left| \int_0^t H_s^n dM_s - \int_0^t H_s dM_s \right| \geq \eta \right) \leq \mathbb{P}(\tau_p < t) + \frac{1}{\eta^2} \mathbb{E} \left[\left(\int_0^{\tau_p} H_s^n dM_s - \int_0^{\tau_p} H_s dM_s \right)^2 \right].$$

Since the right-hand result can be controlled by arbitrarily small quantity, the result follows. \square

For continuous integrands, we have the following useful approximation result.

Proposition 5.8. Let X be a continuous martingale, and let H be an sample-continuous adapted process. For every $t > 0$ and every sequence of partitions $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ whose mesh tends to 0, we have

$$\int_0^t H_s dX_s = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} H_{t_{j-1}^n} (X_{t_j^n} - X_{t_{j-1}^n}) \quad \text{in probability.} \quad (5.8)$$

Proof. For every $n \in \mathbb{N}$, we define $H_0^n = H_0$ and $H_s^n = \sum_{j=1}^{k_n} H_{t_{j-1}^n} \mathbb{1}_{(t_{j-1}^n, t_j^n]}$ for all $s > 0$. Then H^n is a left-continuous adapted process, hence is progressive. By Theorem 5.6 (v), we have

$$\int_0^t H_s^n dX_s = \sum_{j=1}^{k_n} H_{t_{j-1}^n} (X_{t_j^n} - X_{t_{j-1}^n}), \quad \forall n \in \mathbb{N}.$$

We take $K_s = \max_{0 \leq r \leq s} |H_r|$. This is a locally bounded progressive process dominating H^n . Since H is sample continuous, we have $H_s^n \rightarrow H_s$ a.s. for all $s \in [0, t]$. Using Theorem 5.7 concludes the proof. \square

Remark. Note in (5.8), we evaluate H at the left end of every interval $(t_{j-1}^n, t_j^n]$. The result fails if we replace $H_{t_{j-1}^n}$ by $H_{t_j^n}$. To see a counterexample, we take $H = Y$ to be another continuous martingale. Then

$$\sum_{j=1}^{k_n} Y_{t_j^n} (X_{t_j^n} - X_{t_{j-1}^n}) = \sum_{j=1}^{k_n} Y_{t_{j-1}^n} (X_{t_j^n} - X_{t_{j-1}^n}) + \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n}) (Y_{t_j^n} - Y_{t_{j-1}^n})$$

The convergence results in (5.8) and (3.10) imply $\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} Y_{t_j^n} (X_{t_j^n} - X_{t_{j-1}^n}) = \int_0^t Y_s dX_s + \langle X, Y \rangle_t$, and

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t. \quad (5.9)$$

This is known as the *formula of integration by parts*.

5.2 Itô's Formula and its Consequences

Theorem 5.9 (Itô's Lemma). Let X^1, \dots, X^p be p continuous semimartingales, and let $F \in C^2(\mathbb{R}^p, \mathbb{R})$. Then for every $t \geq 0$, we have

$$F(X_t^1, \dots, X_t^p) = F(X_0^1, \dots, X_0^p) + \sum_{j=1}^p \int_0^t \frac{\partial F(X_s^1, \dots, X_s^p)}{\partial x^j} dX_s^j + \frac{1}{2} \sum_{i,j=1}^p \int_0^t \frac{\partial^2 F(X_s^1, \dots, X_s^p)}{\partial x^i \partial x^j} d\langle X^i, X^j \rangle_s.$$

Proof. We write $(X^1, \dots, X^p) = X$ for brevity. Fix $t > 0$, and consider an increasing sequence of partitions $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = t$ whose mesh tends to 0. According to Taylor's theorem,

$$\begin{aligned} F(X_t) &= F(X_0) + \sum_{l=1}^{k_n} \left(F(X_{t_l^n}) - F(X_{t_{l-1}^n}) \right) \\ &= F(X_0) + \underbrace{\sum_{j=1}^p \sum_{l=1}^{k_n} \frac{\partial F}{\partial x^j}(X_{t_{l-1}^n})(X_{t_l^n}^j - X_{t_{l-1}^n}^j)}_{(a)} + \underbrace{\frac{1}{2} \sum_{i,j=1}^p \sum_{l=1}^{k_n} f_{n,l}^{i,j}(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j)}_{(b)}, \end{aligned}$$

where the quantity $f_{n,l}^{i,j}$ can be written as

$$f_{n,l}^{i,j} = \frac{\partial^2 F}{\partial x^i \partial x^j}((1 - \xi)X_{t_{l-1}^n}^i + \xi X_{t_l^n}^i)$$

for some random variable $\xi : \Omega \rightarrow [0, 1]$. By Proposition 5.8, the term (a) converges to $\sum_{j=1}^p \int_0^t \frac{\partial F}{\partial x^j}(X_s) dX_s^j$ in probability as $n \rightarrow \infty$. So it remains to find the limit of term (b). For brevity we write $D_{ij}F = \frac{\partial^2 F}{\partial x^i \partial x^j}$. By uniform continuity of the second derivatives of F on compact intervals, we have for all $i, j \in \{1, \dots, p\}$ that

$$\sup_{1 \leq l \leq k_n} \left| f_{n,l}^{i,j} - D_{ij}F(X_{t_{l-1}^n}) \right| \leq \sup_{1 \leq l \leq k_n} \left(\sup_{x \in [X_{t_{l-1}^n}^i \wedge X_{t_l^n}^i, X_{t_{l-1}^n}^i \vee X_{t_l^n}^i]} \left| D_{ij}F(x) - D_{ij}F(X_{t_{l-1}^n}) \right| \right) \rightarrow 0 \quad a.s..$$

By Proposition 3.65, $\sum_{l=1}^{k_n} (X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \xrightarrow{\mathbb{P}} \langle X^i, X^j \rangle_t < \infty$. This gives an estimate of (b):

$$\left| \sum_{l=1}^{k_n} D_{ij}F(X_{t_{l-1}^n})(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) - \sum_{l=1}^{k_n} f_{n,l}^{i,j}(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \right| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty.$$

According to (5.9), the process $X^i X^j = (X_s^i X_s^j)_{s \geq 0}$ is also a semimartingale. We then transform (b) as

$$\begin{aligned} &\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} D_{ij}F(X_{t_{l-1}^n})(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \\ &= \int_0^t D_{ij}F(X_s) d\langle X^i, X^j \rangle_s - \int_0^t D_{ij}F(X_s) X_s^i dX_s^j - \int_0^t D_{ij}F(X_s) X_s^j dX_s^i \quad (\text{in probability}) \\ &= \int_0^t D_{ij}F(X_s) d\langle X^i, X^j \rangle_s - \int_0^t D_{ij}F(X_s) d(X^i \cdot X^j)_s - \int_0^t D_{ij}F(X_s) d(X^j \cdot X^i)_s \quad (\text{by associativity}) \\ &= \int_0^t D_{ij}F(X_s) d\langle X^i, X^j \rangle_s. \quad (\text{by linearity and (5.9)}) \end{aligned}$$

Thus we finish the proof of Itô's formula. \square

Remark. The formula (5.9) of integration by parts is a special case of Itô's lemma.

Proposition 5.10. Take a twice continuously differentiable function $F(r, x)$ in \mathbb{R}^2 . Itô's formula implies

$$F(\langle X, X \rangle_t, X_t) = F(0, X_0) + \int_0^t \frac{\partial F}{\partial x}(\langle X, X \rangle_s, X_s) dX_s + \int_0^t \left(\frac{\partial F}{\partial r} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \right) (\langle X, X \rangle_s, X_s) d\langle M, M \rangle_s$$

For a continuous local martingale M , $F(\langle M, M \rangle_t, M_t)$ is a continuous local martingale if $\frac{\partial F}{\partial r} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} = 0$.

Remark. We take $F(r, x) = \exp(\lambda x - \frac{\lambda^2}{2} r)$, where $\lambda \in \mathbb{C}$. Then both the real and imaginary parts of $F : \mathbb{R}^2 \rightarrow \mathbb{C}$ satisfies the above condition, and $\frac{\partial F}{\partial x} = \lambda F$. We define

$$\mathcal{E}(\lambda M)_t = \exp\left(\lambda M_t - \frac{\lambda^2}{2} \langle M, M \rangle_t\right), \quad \forall t \geq 0.$$

Consequently, $\mathcal{E}(\lambda M)$ is a complex continuous local martingale (i.e. both its real and imaginary parts), and

$$\mathcal{E}(\lambda M)_t = e^{\lambda M_0} + \lambda \int_0^t \mathcal{E}(\lambda M)_s dM_s.$$

From now on, we are going to discuss the Brownian motions. A *d-dimensional Brownian motion* is a stochastic process $\{B_t = (B_t^1, \dots, B_t^d), t \geq 0\}$ with values in \mathbb{R}^d whose component processes B^1, \dots, B^d are independent Brownian motions. A Brownian motion $(B_t)_{t \geq 0}$ is called a (\mathcal{F}_t) -Brownian motion if it is adapted and has independent increments with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Theorem 5.11 (Multi-dimensional Brownian motions and Lévy's theorem). An adapted and sample-continuous process $B = (B^1, \dots, B^d)$ is a (\mathcal{F}_t) -Brownian motion if and only if its component processes B_1, \dots, B_d are continuous local martingales such that $\langle X^i, X^j \rangle_t = \delta_{ij}t$ for all $i, j \in \{1, \dots, d\}$ and all $t \geq 0$.

Proof. Necessity is clear by definition: If B is a Brownian motion, we have $\langle B^j, B^j \rangle_t = t$ for all $j \in \{1, \dots, d\}$ and all $t \in \mathbb{R}_+$, and independence of distinct components implies that $\langle B^i, B^j \rangle = 0$ for all $i \neq j$.

Now we prove sufficiency. Take $\alpha = (\alpha^1, \dots, \alpha^d) \in \mathbb{R}^d$ with $|\alpha|^2 = \sum_{j=1}^d \alpha_j^2$. Then $\alpha^\top X_t = \sum_{j=1}^d \alpha_j B_t^j$ is a continuous local martingale with quadratic variation $\langle \alpha^\top B, \alpha^\top B \rangle = |\alpha|^2 t$. By Proposition 5.10, the process $(\exp(i\alpha^\top B_t - \frac{1}{2}|\alpha|^2 t))_{t \geq 0}$ is a continuous local martingale bounded on each compact interval $[0, t]$, $t > 0$, hence is a martingale. As a result, for all $t \geq s > 0$,

$$\mathbb{E} \left[\exp \left(i\alpha^\top B_t - \frac{1}{2}|\alpha|^2 t \right) \middle| \mathcal{F}_s \right] = \exp \left(i\alpha^\top B_s - \frac{1}{2}|\alpha|^2 s \right) \Rightarrow \mathbb{E} [\exp(i\alpha^\top (B_t - B_s)) | \mathcal{F}_s] = e^{-\frac{1}{2}|\alpha|^2 (t-s)}.$$

Given any $A \in \mathcal{F}_s$, we take the measure $\mathbb{P}_A(E) = \mathbb{P}(A \cap E) / \mathbb{P}(A)$, $\forall E \in \mathcal{F}$. Comparing the characteristic functions, the law of $B_t - B_s$ does not change from \mathbb{P} to \mathbb{P}_A . Then $\mathbb{E}[f(B_t - B_s) \mathbf{1}_A] = \mathbb{P}(A) \mathbb{E}[f(B_t - B_s)]$ for all nonnegative measurable functions. Choosing f to be indicator functions do we obtain that $B_t - B_s$ is independent of \mathcal{F}_s , and $X_t - X_s \sim N(0, (t-s)I_d)$. Since B is adapted and sample-continuous and has independent Gaussian increments, it is a (\mathcal{F}_t) -Brownian motion. \square

Remark. Let $B = (B^1, \dots, B^d)$ be a *d-dimensional* (\mathcal{F}_t) -Brownian motion. By Itô's formula, for a twice continuously differentiable real-valued function $F(x_1, \dots, x_d)$ on \mathbb{R}^d ,

$$\begin{aligned} F(B_t^1, \dots, B_t^d) &= F(B_0) + \sum_{j=1}^d \int_0^t \frac{\partial F}{\partial x_j}(B_s^1, \dots, B_s^d) dB_s^j + \frac{1}{2} \sum_{j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_j^2}(B_s^1, \dots, B_s^d) ds \\ &\Rightarrow F(B_t) = F(B_0) + \int_0^t \nabla F(B_s) \cdot dB_s + \frac{1}{2} \int_0^t \Delta F(B_s) ds. \end{aligned}$$

Likewise, for a twice continuously differentiable real-valued function $G(t, x_1, \dots, x_d)$ on $\mathbb{R}_+ \times \mathbb{R}^d$, we have

$$G(t, B_t) = G(0, B_0) + \int_0^t \nabla_x G(s, B_s) \cdot dB_s + \int_0^t \left(\frac{\partial G}{\partial t} + \frac{1}{2} \Delta_x G \right) (s, B_s) ds.$$

Now we introduce a time-changed Brownian motion representation of continuous local martingales. Before presenting the general conclusion, we first prove some technical results.

Lemma 5.12. Let $M = (M_t)_{t \geq 0}$ be a continuous local martingale. Almost surely, we have $M_a = M_b$ for all $0 \leq a < b$ such that $\langle M, M \rangle_b = \langle M, M \rangle_a$.

Proof. Fix $0 \leq a < b$. Consider the continuous local martingale $N_t = M_t - M_{t \wedge a}$, whose quadratic variation is given by $\langle N, N \rangle_t = \langle M, M \rangle_t - \langle M, M \rangle_{t \wedge a}$. We choose the sequence of stopping times $\tau_n = \inf\{t \geq 0 : \langle N, N \rangle_t \geq 1/n\}$. Since $\langle N^{\tau_n}, N^{\tau_n} \rangle \leq 1/n$, we have $N^{\tau_n} \in \mathbb{H}^2$, and

$$\mathbb{E}[N_{t \wedge \tau_n}^2] = \mathbb{E}[\langle N, N \rangle_{t \wedge \tau_n}] \leq \frac{1}{n}, \quad \forall t \in [a, b].$$

On the event $A_{a,b} := \{\langle M, M \rangle_b = \langle M, M \rangle_a\} \subset \{\tau_n \geq b\}$, we have

$$\mathbb{E}[N_b^2 \mathbf{1}_{A_{a,b}}] = \mathbb{E}[N_{b \wedge \tau_n}^2 \mathbf{1}_{A_{a,b}}] \leq \mathbb{E}[N_{b \wedge \tau_n}^2] \leq \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0 \quad N_b = 0 \quad a.s. \text{ on } A_{a,b}.$$

We set $E_{a,b} = \{\langle M, M \rangle_b = \langle M, M \rangle_a = M_b \neq M_a\} = A_{[a,b]} \cap \{M_b \neq M_a\}$, which satisfies $\mathbb{P}(E_{a,b}) = 0$. Take

$$E = \bigcup_{a,b \in \mathbb{Q}, 0 \leq a < b} E_{a,b} \Rightarrow \mathbb{P}(E) = 0.$$

On event $\Omega \setminus E$, whenever $\langle M, M \rangle_b = \langle M, M \rangle_a$, one can choose $\mathbb{Q} \ni a_n \downarrow a$ and $\mathbb{Q} \ni b_n \uparrow a$. Then $\langle M, M \rangle_{b_n} = \langle M, M \rangle_{a_n}$ for all $n \in \mathbb{N}$, and $M_{a_n} = M_{b_n}$. By sample-continuity of M , we obtain $M_a = M_b$. \square

Theorem 5.13 (Dambis-Dubins-Schwarz). Let $M = (M_t)_{t \geq 0}$ be a continuous local martingale such that $M_0 = 0$ and $\langle M, M \rangle_\infty = \infty$ a.s.. Then there exists a Brownian motion $\beta = (\beta_s)_{s \geq 0}$ such that

$$a.s. \forall t \geq 0, \quad M_t = \beta_{\langle M, M \rangle_t}.$$

Proof. For every $s \geq 0$, choose the stopping time $\tau_s = \inf\{t \geq 0 : \langle M, M \rangle_t \geq s\}$. Since $\langle M, M \rangle_\infty = \infty$ a.s., we have $\tau_s < \infty$ a.s., and we reset $\tau_s(\omega) = 0$ on the event $E = \{\langle M, M \rangle_\infty < \infty\}$. By completeness of the filtration $(\mathcal{F}_t)_{t \geq 0}$, the variable τ_s remains a stopping time. By construction, for every $\omega \in \Omega$, the function $s \mapsto \tau_s(\omega)$ is increasing. On the event $\Omega \setminus E$, we have

$$\begin{aligned} \lim_{r \uparrow s} \tau_r &= \inf \bigcap_{r < s} \{t \geq 0 : \langle M, M \rangle_t \geq r\} = \inf\{t \geq 0 : \langle M, M \rangle_t \geq s\}, \\ \lim_{r \downarrow s} \tau_r &= \inf \bigcup_{r > s} \{t \geq 0 : \langle M, M \rangle_t \geq r\} = \inf\{t \geq 0 : \langle M, M \rangle_t > s\}. \end{aligned}$$

Hence $s \mapsto \tau_s(\omega)$ is left-continuous, and has right limit $\tau_{s+} = \inf\{t \geq 0 : \langle M, M \rangle_t > s\}$ at $s \geq 0$.

For every $s \geq 0$, we set $\beta_s = M_{\tau_s}$. By Proposition 3.12, $\beta = (\beta_t)_{t \geq 0}$ is an adapted process with respect to the time-changed filtration $\mathcal{G}_s = \mathcal{F}_{\tau_s}$ for every $s \geq 0$, and $\mathcal{G}_\infty = \mathcal{F}_\infty$. The completeness of $(\mathcal{G}_t)_{t \geq 0}$ also follows from $(\mathcal{F}_t)_{t \geq 0}$. Moreover, the sample path $s \mapsto \beta_s$ is càglàd, with the right limit at s given by $\beta_{s+} = M_{\tau_{s+}}$ on the event $\Omega \setminus E$. Note that $\langle M, M \rangle_{\tau_s} = \langle M, M \rangle_{\tau_{s+}} = s$ for every $s \geq 0$. By Lemma 5.12, we have almost surely $M_{\tau_{s+}} = M_{\tau_s}$ for all $s \geq 0$. Therefore, the sample path of $\beta = (\beta_s)_{s \geq 0}$ is almost surely continuous.

Now we verify that $(\beta_s)_{s \geq 0}$ and $(\beta_s^2 - s)_{s \geq 0}$ are martingales with respect to the filtration $(\mathcal{G}_s)_{s \geq 0}$. For every $n \in \mathbb{N}$, by Theorem 3.58, the stopped processes M^{τ_n} and $(M^{\tau_n})^2 - \langle M, M \rangle^{\tau_n}$ are uniformly integrable martingales, since $\langle M, M \rangle_\infty^{\tau_n} = \langle M, M \rangle_{\tau_n} = n < \infty$. By optional stopping theorem, for every $0 \leq s < t \leq n$,

$$\begin{aligned}\mathbb{E}[\beta_t | \mathcal{G}_s] &= \mathbb{E}[M_{\tau_t}^{\tau_n} | \mathcal{F}_{\tau_s}] = M_{\tau_s} = \beta_s, \\ \mathbb{E}[\beta_t^2 - t | \mathcal{G}_s] &= \mathbb{E}[(M_{\tau_t}^{\tau_n})^2 - \langle M^{\tau_n}, M^{\tau_n} \rangle_{\tau_s} | \mathcal{F}_{\tau_t}] = (M_{\tau_s}^{\tau_n})^2 - \langle M, M \rangle_{\tau_s \wedge \tau_n} = \beta_s^2 - s.\end{aligned}$$

The case $d = 1$ of Theorem 5.11 implies that β is a (\mathcal{G}_t) -Brownian motion.

On the other hand, by the very definition of τ_r and τ_{r+} , for all $s \geq 0$, we have $\tau_{\langle M, M \rangle_s} \leq s \leq \tau_{\langle M, M \rangle_{s+}}$, and $\langle M, M \rangle_{\tau_{\langle M, M \rangle_s}} = \langle M, M \rangle_{\tau_{\langle M, M \rangle_{s+}}} = \langle M, M \rangle_s$. According to Lemma 5.12, almost surely, the equality $M_s = M_{\tau_{\langle M, M \rangle_s}} = \beta_{\langle M, M \rangle_s}$ holds for all $s \geq 0$, concluding the proof. \square

Remark. In this theorem, the Brownian motion $\beta = (\beta_t)_{t \geq 0}$ is no longer adapted with respect to the original filtration $(\mathcal{F}_t)_{t \geq 0}$, but with respect to the time-changed filtration $(\mathcal{G}_t)_{t \geq 0}$.

Corollary 5.14. Let M and N be two continuous local martingales such that $M_0 = N_0 = 0$. Assume the following conditions holds almost surely: (i) $\langle M, M \rangle_t = \langle N, N \rangle_t$ for all $t \geq 0$; (ii) $\langle M, N \rangle_t = 0$ for all $t \geq 0$; (iii) $\langle M, M \rangle_\infty = \langle N, N \rangle_\infty = \infty$. Let β and γ be the Brownian motions such that $M_t = \beta_{\langle M, M \rangle_t}$ and $N_t = \gamma_{\langle N, N \rangle_t}$ for all $t \geq 0$ almost surely. Then β and γ are independent.

Proof. We choose the stopping times $\tau_s = \inf\{t \geq 0 : \langle M, M \rangle_t \geq s\} = \inf\{t \geq 0 : \langle N, N \rangle_t \geq s\}$ for all $s \geq 0$, so both $\beta_s = M_{\tau_s}$ and $\gamma_s = N_{\tau_s}$ are (\mathcal{G}_t) -Brownian motions, where $(\mathcal{G}_t)_{t \geq 0}$ is the time-changed filtration $\mathcal{G}_t = \mathcal{F}_{\tau_t}$.

Since the continuous local martingales M and N are orthogonal, the process MN is also a continuous local martingale. By Proposition 3.61 (vi), the stopped process $M^{\tau_n} N^{\tau_n}$ is a uniformly integrable martingale. By optional stopping theorem, for all $0 \leq s < t \leq n$, we have

$$\mathbb{E}[\beta_t \gamma_t | \mathcal{G}_s] = \mathbb{E}[M_{\tau_t}^{\tau_n} N_{\tau_t}^{\tau_n} | \mathcal{F}_{\tau_s}] = M_{\tau_s}^{\tau_n} N_{\tau_s}^{\tau_n} = \beta_s \gamma_s.$$

Hence $(\beta_t \gamma_t)_{t \geq 0}$ is a (\mathcal{G}_t) -martingale, and again by Proposition 3.61 (vi), we have $\langle \beta, \gamma \rangle = 0$. According to Theorem 5.11, (β, γ) is a two-dimensional (\mathcal{G}_t) -Brownian motion, hence β and γ are independent. \square

Now we introduce a useful inequality connecting the local maxima of a continuous local martingale with its quadratic variation.

Theorem 5.15 (Burkholder-Davis-Gundy inequality). Given a continuous martingale M with $M_0 = 0$, we define the local maxima $M_t^* = \sup_{0 \leq s \leq t} |M_s|$ for all $t \geq 0$. Then for every $p > 0$, there exist a constant $C_p > 0$ depending only on p such that for every stopping time τ ,

$$\mathbb{E}[(M_\tau^*)^p] \leq C_p \mathbb{E}[\langle M, M \rangle_\tau^{p/2}]$$

Proof. By replacing a continuous local martingale M with M^τ , it suffices to deal with the case $\tau = \infty$. We can further assume that M is bounded by replacing M with M^{τ_n} , where $\tau_n = \{t \geq 0 : |M_t| = n\}$, and the result of $n \rightarrow \infty$ follows by monotone convergence theorem. When $p \geq 2$. Apply Itô's formula to $|x|^p$:

$$|M_t|^p = \int_0^t p |M_s|^{p-1} \text{sgn}(M_s) dM_s + \frac{1}{2} \int_0^t p(p-1) |M_s|^{p-2} d\langle M, M \rangle_s.$$

Note that M is a bounded. Then we have $M \in \mathbb{H}^2$, and the process $\left(\int_0^t p |M_s|^{p-1} \text{sgn}(M_s) dM_s\right)_{t \geq 0}$ is also a

martingale in \mathbb{H}^2 . Consequently, we have

$$\begin{aligned}\mathbb{E}[|M_t|^p] &\leq \frac{p(p-1)}{2} \mathbb{E} \left[\int_0^t |M_s|^{p-2} d\langle M, M \rangle_s \right] \leq \frac{p(p-1)}{2} \mathbb{E}[|M_t^*|^{p-2} \langle M, M \rangle_t] \\ &\leq \frac{p(p-1)}{2} \mathbb{E}[|M_t^*|^p]^{\frac{p-2}{p}} \mathbb{E}[\langle M, M \rangle_t^{p/2}]^{2/p}\end{aligned}$$

On the other hand, the Doob's L^p -inequality [Proposition 3.40] gives

$$\mathbb{E}[|M_t^*|^p] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|M_t|^p] \quad \Rightarrow \quad \mathbb{E}[|M_t^*|^p] \leq \left(\left(\frac{p}{p-1} \right)^p \frac{p(p-1)}{2} \right)^{p/2} \mathbb{E}[\langle M, M \rangle_t^{p/2}] \quad (5.10)$$

Now we deal with the case $p < 2$. Since $M \in \mathbb{H}^2$, the process $M^2 - \langle M, M \rangle$ is a uniformly integrable martingale. Hence $\mathbb{E}[(M_\tau)^2] = \mathbb{E}[\langle M, M \rangle_\tau]$ for every stopping time τ . Given $x > 0$, consider the stopping time $\tau_x = \inf\{t \geq 0 : M_t^2 \geq x\}$. If τ is a bounded stopping time, we have

$$\mathbb{P}((M_\tau^*)^2 \geq x) = \mathbb{P}(\tau_x \leq \tau) = \mathbb{P}((M_{\tau_x \wedge \tau})^2 \geq x) \leq \frac{\mathbb{E}[(M_{\tau_x \wedge \tau})^2]}{x} = \frac{\mathbb{E}[\langle M, M \rangle_{\tau_x \wedge \tau}]}{x} \leq \frac{\mathbb{E}[\langle M, M \rangle_\tau]}{x}.$$

Consider the stopping time $\sigma_x = \inf\{s \geq 0 : \langle M, M \rangle_s \geq x\}$, so $\{\langle M, M \rangle_t \geq x\} \Leftrightarrow \{\sigma_x \leq t\}$. For every $t \geq 0$, we use the preceding bound with $\tau = \sigma_x \wedge t$:

$$\begin{aligned}\mathbb{P}((M_t^*)^2 \geq x) &\leq \mathbb{P}((M_{\sigma_x \wedge t}^*)^2 \geq x) + \mathbb{P}(\sigma_x \leq t) \\ &\leq \frac{1}{x} \mathbb{E}[\langle M, M \rangle_{\sigma_x \wedge t}] + \mathbb{P}(\langle M, M \rangle_t \geq x) = \frac{1}{x} \mathbb{E}[\langle M, M \rangle_t \mathbf{1}_{\{\langle M, M \rangle_t < x\}}] + 2\mathbb{P}(\langle M, M \rangle_t \geq x).\end{aligned}$$

Now we integrate both side with respect to $\frac{p}{2}x^{p/2-1}dx$. For the left-hand side:

$$\int_0^\infty \mathbb{P}((M_t^*)^2 \geq x) \frac{p}{2}x^{p/2-1}dx = \mathbb{E} \left[\int_0^{(M_t^*)^2} \frac{p}{2}x^{p/2-1}dx \right] = \mathbb{E}[(M_t^*)^p].$$

Similarly, $\int_0^\infty \mathbb{P}(\langle M, M \rangle_t \geq x) \frac{p}{2}x^{p/2-1}dx = \mathbb{E}[\langle M, M \rangle_t^{p/2}]$. Furthermore,

$$\int_0^\infty \mathbb{E}[\langle M, M \rangle_t \mathbf{1}_{\{\langle M, M \rangle_t < x\}}] \frac{p}{2}x^{p/2-2}dx = \mathbb{E} \left[\langle M, M \rangle_t \int_0^{\langle M, M \rangle_t} \frac{p}{2}x^{p/2-2}dx \right] = \frac{p}{2-p} \mathbb{E}[\langle M, M \rangle_t^{p/2}].$$

This gives the bound

$$\mathbb{E}[(M_t^*)^p] \leq \frac{4-p}{2-p} \mathbb{E}[\langle M, M \rangle_t^{p/2}] \quad (5.11)$$

Note both the bounds (5.10) and (5.11) holds for $t = \infty$ by monotone convergence theorem. \square

Remark. In fact, the complete Burkholder–Davis–Gundy inequality gives both upper and lower bounds: For every real $p > 0$, there exist two constants $c_p, C_p > 0$ depending only on p such that, for every continuous local martingale M with $M_0 = 0$ and every stopping time τ , one have

$$c_p \mathbb{E}[\langle M, M \rangle_\tau^{p/2}] \leq \mathbb{E}[(M_\tau^*)^p] \leq C_p \mathbb{E}[\langle M, M \rangle_\tau^{p/2}].$$

Furthermore, the case of $p = 1$ gives a sufficient condition for a continuous local martingale M with $M_0 = 0$ to be a uniformly integrable martingale: $\mathbb{E}[\langle M, M \rangle_\infty^{1/2}] < \infty$.

5.3 The Representation of Martingales as Stochastic Integral

On a probability space (Ω, \mathcal{F}, P) , we choose a filtration $(\mathcal{F}_t)_{t \geq 0}$ on Ω to be the completed canonical filtration of a Brownian motion. An interesting conclusion is that all martingales with respect to this filtration can be represented as stochastic integrals with respect to the Brownian motion.

In this section, we fix a Brownian motion $B = (B_t)_{t \geq 0}$, and assume $(\mathcal{F}_t)_{t \geq 0}$ to be the completed canonical filtration of B . Before formally present this result, we first introduce some technical lemma.

Lemma 5.16. The vector space generated by the random variables of the form

$$\exp \left(i \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}) \right), \quad \text{where } 0 = t_0 < t_1 < \cdots < t_n \text{ and } \lambda_1, \dots, \lambda_n \in \mathbb{R}$$

is dense in the space $L^2_{\mathbb{C}}(\Omega, \mathcal{F}_{\infty}, \mathbb{P})$ of all square-integrable complex-valued \mathcal{F}_{∞} -measurable random variables.

Proof. By elementary Hilbert theory, it suffices to show that $Z = 0$ is the only variable that satisfying

$$\mathbb{E} \left[Z \exp \left(i \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}) \right) \right] = 0$$

for all choices of $0 = t_0 < t_1 < \cdots < t_n$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Define the complex measure μ on \mathbb{R}^n by

$$\mu(F) = \mathbb{E} [Z \mathbb{1}_F(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})], \quad \forall F \in \mathcal{B}(\mathbb{R}^n).$$

Then the Fourier transform of μ satisfies

$$\int_{\mathbb{R}^n} e^{i\lambda^\top x} \mu(dx) = \mathbb{E} \left[Z \exp \left(i \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}) \right) \right] = 0, \quad \forall \lambda \in \mathbb{R}^n.$$

By Lévy's continuity theorem, we have $\mu = 0$. Hence $\mathbb{E}[Z \mathbb{1}_A] = 0$ for all $A \in \sigma(B_{t_1}, \dots, B_{t_n})$. A monotone class argument shows that $\mathbb{E}[Z \mathbb{1}_A] = 0$ holds for all $A \in \sigma(B_t, t \geq 0)$, and further by completion, for all $A \in \mathcal{F}_{\infty}$. Hence $Z = 0$ in $L^2_{\mathbb{C}}(\Omega, \mathcal{F}_{\infty}, \mathbb{P})$. \square

Theorem 5.17. For every $Z \in L^2(\Omega, \mathcal{F}_{\infty}, \mathbb{P})$, there exists a unique progressive process $H \in L^2(B)$ such that

$$Z = \mathbb{E}[Z] + \int_0^\infty H_s dB_s. \quad (5.12)$$

Consequently, for every L^2 -bounded martingale M (resp. for every continuous local martingale M), there exists a unique process $H \in L^2(B)$ (resp. $H \in L^2_{\text{loc}}(B)$) and a constant $C \in \mathbb{R}$ such that

$$M_t = C + \int_0^t H_s dB_s, \quad \forall t \geq 0. \quad (5.13)$$

Proof. Consider the first assertion. If both H and H' satisfy this (5.12), the second moment formula gives

$$0 = \mathbb{E} \left[\left(\int_0^\infty (H_s - H'_s) dB_s \right)^2 \right] = \mathbb{E} \left[\int_0^\infty (H_s - H'_s)^2 ds \right] \Rightarrow H = H' \text{ in } L^2(B).$$

This shows uniqueness. For existence, let \mathcal{H} be the vector space of all $Z \in L^2(\Omega, \mathcal{F}_{\infty}, \mathbb{P})$ for which there exists

an associated $H \in L^2(B)$ satisfying (5.12). By Proposition 5.10, for any step function $h = \sum_{j=1}^n \lambda_j \mathbf{1}_{(t_{j-1}, t_j]}$, where $0 = t_0 < t_1 < \dots < t_n$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, we have

$$\exp\left(i \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}) + \frac{1}{2} \sum_{j=1}^n \lambda_j^2 (t_j - t_{j-1})\right) = \mathcal{E}(ih \cdot B)_t = 1 + i \int_0^t \mathcal{E}(ih \cdot B)_s h(s) dB_s$$

Since $\mathcal{E}(ih \cdot B)_s h(s)$ is a bounded continuous local martingale, both its real and imaginary parts are in $L^2(B)$. This implies that both the real and imaginary parts of $\exp\left(i \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}})\right)$ are in \mathcal{H} . According to Lemma 5.16, \mathcal{H} contains a dense subset of $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$.

To prove the first assertion, it remains to show that \mathcal{H} is closed, which implies $\mathcal{H} = L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. We assume that $Z_n \in \mathcal{H}$ is a sequence of random variables that converges to Z in L^2 . Let $H^{(n)} \in L^2(B)$ be the associated progressive process. Then

$$\begin{aligned} \|H^{(n)} - H^{(m)}\|_{L^2(B)} &= \mathbb{E} \left[\int_0^\infty (H_s^{(n)} - H_s^{(m)})^2 ds \right] = \mathbb{E} \left[\left(\int_0^\infty (H_s^{(n)} - H_s^{(m)}) dB_s \right)^2 \right] \\ &= \mathbb{E} \left[(Z_n - Z_m - \mathbb{E}[Z_n] + \mathbb{E}[Z_m])^2 \right] \leq \|Z_n - Z_m\|_2^2. \end{aligned}$$

Hence $H^{(n)}$ is a Cauchy sequence in $L^2(B)$, which converges to some $H \in L^2(B)$ by completeness of $L^2(B)$. Consequently, $Z_n = \mathbb{E}[Z_n] + \int_0^\infty H_s^{(n)} dB_s$ converges to $Z = \mathbb{E}[Z] + \int_0^\infty H_s dB_s$ in L^2 .

We turn to the second assertion. The uniqueness argument is similar. If M is a L^2 -bounded martingale, then M_t converges *a.s.* and in L^2 to some $M_\infty \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. Since $(\int_0^t H_s dB_s)_{s \geq 0}$ is bounded in L^2 , it is a uniformly integrable martingale. We can find the process $H \in L^2(B)$ that satisfies (5.12) for M_∞ . Then

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty H_s dB_s \quad \Rightarrow \quad M_t = \mathbb{E}[M_\infty | \mathcal{F}_t] = \mathbb{E}[M_\infty] + \int_0^t H_s dB_s, \quad \forall t \geq 0.$$

Finally, if M is a continuous local martingale, we have $M_0 = C$ *a.s.* for some constant $C \in \mathbb{R}$ because \mathcal{F}_0 is \mathbb{P} -trivial. We then choose stopping times $\tau_n = \{t \geq 0 : |M_t| \geq n\}$ so that M^{τ_n} are bounded martingales. According to the preceding result, there exists $H^{(n)} \in L^2(B)$ such that

$$M_{t \wedge \tau_n} = C + \int_0^t H_s^{(n)} dB_s, \quad \forall t \geq 0.$$

By uniqueness of the progressive process in this representation, we have $H^{(m)} = H^{(n)} \mathbf{1}_{[0, \tau_m]}$ in $L^2(B)$ for all $n > m$. Consequently, we can find $H \in L^2(B)$ such that $H^{(n)} = H \mathbf{1}_{[0, \tau_n]}$ in $L^2(B)$ for all $n \in \mathbb{N}$, and the representation formula (5.13) follows by letting $n \rightarrow \infty$. The uniqueness argument is similar. \square

Remark. In this theorem, we do not require the L^2 -bounded martingale M to be sample-continuous. Next we discuss some consequence of this representation theorem.

Proposition 5.18. The filtration $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous.

Proof. Let Z be a bounded \mathcal{F}_{t+} -measurable random variable. By Theorem 5.17, there exists $H \in L^2(B)$ such that $Z = \mathbb{E}[Z] + \int_0^\infty H_s dB_s$. Given $\epsilon > 0$, Z is $\mathcal{F}_{t+\epsilon}$ measurable. By continuity of $H \cdot B$, we have

$$Z = \mathbb{E}[Z | \mathcal{F}_{t+\epsilon}] = \mathbb{E}[Z] + \int_0^{t+\epsilon} H_s dB_s \xrightarrow{a.s.} \mathbb{E}[Z] + \int_0^t H_s dB_s, \quad \epsilon \downarrow 0.$$

As a result, $Z = \mathbb{E}[Z] + \int_0^t H_s dB_s$ *a.s.* By completeness of the filtration $(\mathcal{F}_t)_{t \geq 0}$, Z is also \mathcal{F}_t -measurable. Hence for all $A \in \mathcal{F}_{t+}$, the variable $Z = \mathbf{1}_A$ is \mathcal{F}_t measurable, and $A \in \mathcal{F}_t$. Therefore $\mathcal{F}_{t+} = \mathcal{F}_t$. \square

Remark. We also can define the left limit of $(\mathcal{F}_t)_{t \geq 0}$ at any $t > 0$:

$$\mathcal{F}_{t-} = \sigma \left(\bigcup_{0 \leq s < t} \mathcal{F}_s \right).$$

A similar argument to Proposition 5.18 implies that $\mathcal{F}_{t-} = \mathcal{F}_t$. Namely, the completed filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by a Brownian motion B is also left-continuous.

Proposition 5.19. All martingales with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ have an *a.s.* modification with continuous sample paths.

Proof. Following Theorem 5.17, an L^2 -bounded martingale is continuous according to the representation formula (5.13). Now we consider a uniformly integrable martingale M . This suffice since we can replace a martingale M by the stopped martingale M^a for every $a \geq 0$.

Since M is a uniformly integrable martingale, we have $M_t = \mathbb{E}[M_\infty | \mathcal{F}_t]$ for all $t \geq 0$. By Theorem 3.44 and Proposition 5.1, M has an *a.s.* modification with càdlàg sample paths, which we still denote by M for simplicity. Let $M_\infty^{(n)}$ be a sequence of bounded random variables that converges in L^1 to M_∞ , and introduce the martingales $M_t^{(n)} = \mathbb{E}[M_\infty^{(n)} | \mathcal{F}_t]$. These martingales are then bounded in L^2 , hence sample-continuous. Furthermore, the Doob's maximal inequality Proposition 3.40 (i) gives

$$\mathbb{P} \left(\sup_{t \geq 0} |M_t^{(n)} - M_t| \geq \lambda \right) \leq \frac{3}{\lambda} \mathbb{E}[|M_\infty^{(n)} - M_\infty|], \quad \forall \lambda > 0.$$

We choose a subsequence n_k such that

$$\mathbb{P} \left(\sup_{t \geq 0} |M_t^{(n_k)} - M_t| > \frac{1}{2^k} \right) \leq \frac{1}{2^k}, \quad \forall k \in \mathbb{N} \quad \Rightarrow \quad \mathbb{P} \left(\sup_{t \geq 0} |M_t^{(n_k)} - M_t| > \frac{1}{2^k} \text{ for infinitely many } k \right) = 0.$$

Here we use Borel-Cantelli lemma. Consequently, $\sup_{t \geq 0} |M_t^{(n_k)} - M_t| \rightarrow 0$ *a.s.*, and the sample paths of M , being the uniform limit of a sequence of continuous functions, is continuous. \square

5.4 Stochastic Differential Equations

We start from a deterministic process $(y_t)_{t \geq 0}$, of which the dynamic is specified by the following ordinary differential equation (ODE):

$$dy_t = b(t, y_t) dt \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s) ds$$

To model a noisy system, we simply introduce a random perturbation of the term σdB_t , where $\sigma > 0$ is a constant, and B_t is a Brownian motion. This form implicitly assumes independence of perturbations affecting disjoint time intervals, and we get

$$dy_t = b(t, y_t) dt + \sigma dB_t \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s) ds + \sigma B_t.$$

We generalize the above equation by allowing σ depending on the time t and the state y_t :

$$dy_t = b(t, y_t) dt + \sigma(t, y_t) dB_t \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s) ds + \int_0^t \sigma(s, y_s) dB_s.$$

This gives rise to the following definition of stochastic differential equation.

Definition 5.20 (Stochastic Differential Equation, SDE). Let $p, q \in \mathbb{N}$. Let $\sigma = (\sigma_{ij})_{i \in [p], j \in [q]} : \mathbb{R}_+ \times \mathbb{R}^q \rightarrow \mathbb{R}^{p \times q}$ and $b = (b_i)_{i \in [p]} : \mathbb{R}_+ \times \mathbb{R}^q \rightarrow \mathbb{R}^p$ be locally bounded measurable functions. A *solution of the stochastic equation* $E(\sigma, b)$, which is given by

$$dX_t = \sigma(t, X_t) dB_t + b(t, X_t) dt,$$

consists of:

- A filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a complete filtration $(\mathcal{F}_t)_{t \geq 0}$;
- A q -dimensional (\mathcal{F}_t) -Brownian motion $B = (B^1, \dots, B^q)$ starting from 0;
- An (\mathcal{F}_t) -adapted and sample-continuous process $X = (X^1, \dots, X^p)$ taking values in \mathbb{R}^p , such that

$$X_t = X_0 + \int_0^t \sigma(s, X_s) dB_s + \int_0^t b(s, X_s) ds \quad \stackrel{def}{\Leftrightarrow} \quad X_t^i = X_0^i + \sum_{j=1}^q \int_0^t \sigma_{ij}(s, X_s) dB_s^j + \int_0^t b_i(s, X_s) ds.$$

If $X_0 \sim \delta_x$ for any $x \in \mathbb{R}^p$, we say that X is a solution of $E^x(\sigma, b)$.

There are several notions of existence and uniqueness for stochastic differential equations.

Definition 5.21 For the stochastic differential equation $E(\sigma, b)$, we say that there is

- *weak existence*, if for every $x \in \mathbb{R}^p$, there exists a solution of $E^x(\sigma, b)$;
- *weak existence* and *weak uniqueness*, if in addition, for every $x \in \mathbb{R}^p$, all solutions of $E^x(\sigma, b)$ have the same law;
- *pathwise uniqueness*, if, whenever the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and the (\mathcal{F}_t) -Brownian motion B are fixed, two solutions X and Y such that $X_0 = Y_0$ *a.s.* are indistinguishable.

Furthermore, we say that a solution X of $E(\sigma, b)$ is a *strong solution* if X is adapted with respect to the completed canonical filtration of B .

Remark. We give a simple example of stochastic differential equation where weak existence and weak uniqueness hold, but pathwise uniqueness fails. Let $(\beta_t)_{t \geq 0}$ be a Brownian motion with $\beta_0 = x \in \mathbb{R}$, and consider

$$B_t = \int_0^t \operatorname{sgn}(\beta_s) d\beta_s, \quad \text{where } \operatorname{sgn}(x) = \mathbf{1}_{(0, \infty)}(x) - \mathbf{1}_{(-\infty, 0]}(x).$$

Then $(B_t)_{t \geq 0}$ is a continuous local martingale with quadratic variation $\langle B, B \rangle_t = t$, hence is a Brownian motion by Theorem 5.11. Furthermore, by associativity of stochastic integrals, one have

$$\int_0^t \operatorname{sgn}(\beta_s) dB_s = \int_0^t \operatorname{sgn}(\beta_s)^2 d\beta_s = \beta_t - \beta_0, \quad \forall t \geq 0.$$

Therefore, $(\beta_t)_{t \geq 0}$ solves the following stochastic differential equations:

$$dX_t = \operatorname{sgn}(X_t) dB_t, \quad X_0 = x.$$

For this equation, weak existence holds, and again by Theorem 5.11 we know that any solution of this equation must be a Brownian motion. Hence weak uniqueness also holds. Nevertheless, pathwise uniqueness fails for this equation. For example, if we set $x = 0$, then both β and $-\beta$ solve the preceding SDE with the same Brownian motion B starting from 0.

Now we study the properties of SDE $E(\sigma, b)$ where functions σ and b are continuous on $\mathbb{R}_+ \times \mathbb{R}^q$ and Lipschitz in the variable x . Then there exists a constant L such that for every $t \geq 0$ and $x, y \in \mathbb{R}^d$,

$$\begin{aligned} |\sigma(t, x) - \sigma(t, y)| &\leq L|x - y|, \\ |b(t, x) - b(t, y)| &\leq L|x - y|. \end{aligned}$$

Here we use $|\cdot|$ to denote the Euclidean norm of vectors and the Frobenius norm of matrices.

Lemma 5.22 (Gronwall's lemma). Let $T > 0$ and let $g : [0, T] \rightarrow \mathbb{R}_+$ be a bounded measurable function. If there exist two constants $a \geq 0$ and $b \geq 0$ such that

$$g(t) \leq a + b \int_0^t g(s) ds, \quad \forall t \in [0, T],$$

then we have $g(t) \leq ae^{bt}$ for all $t \in [0, T]$.

Proof. A simple recursion on g gives

$$\begin{aligned} g(t) &\leq a + a(bt) + b^2 \int_0^t \left(\int_0^{s_1} g(s_2) ds_2 \right) ds_1 \leq \dots \\ &\leq a + a(bt) + a \frac{(bt)^2}{2} + \dots + a \frac{(bt)^n}{n!} + b^{n+1} \int_0^t ds_1 \int_0^{s_1} ds_2 \dots \int_0^{s_{n+1}} ds_{n+1} g(s_{n+1}). \end{aligned}$$

Since g is bounded, we let $0 \leq g(t) \leq M$ for all $t \in [0, T]$. Then

$$a \sum_{k=0}^n \frac{(bt)^k}{k!} \leq \dots \leq a \sum_{k=0}^n \frac{(bt)^k}{k!} + \frac{M(bt)^{n+1}}{(n+1)!}.$$

Letting $t \rightarrow \infty$ produces the desired result. □

The following theorem gives the existence of a solution of SDE in the Lipschitz case.

Theorem 5.23. Let functions σ and b be continuous on $\mathbb{R}_+ \times \mathbb{R}^q$ and Lipschitz in the variable x . Then pathwise uniqueness holds for the SDE $E(\sigma, b)$. Furthermore, for every complete filtered $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, every (\mathcal{F}_t) -Brownian motion B and every $x \in \mathbb{R}^p$, there exists a unique strong solution of $E^x(\sigma, b)$.

Proof. We prove the case $p = q = 1$. The multi-dimensional case is similar. To tackle pathwise uniqueness, we fix the complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and the (\mathcal{F}_t) -Brownian motion B with $B_0 = 0$. Let X and Y be two solutions of $E(\sigma, b)$ with $X_0 = Y_0$ a.s.. We fix $M > 0$, and set

$$\tau = \inf \{t \geq 0 : |X_t| \vee |Y_t| \geq M\}$$

Then $t \mapsto \mathbb{E}[(X_{t \wedge \tau} - Y_{t \wedge \tau})^2]$ is a bounded measurable function. Moreover, for every $t \geq 0$,

$$X_{t \wedge \tau} = X_0 + \int_0^{t \wedge \tau} \sigma(s, X_s) dB_s + \int_0^{t \wedge \tau} b(s, X_s) ds,$$

and a similar formula holds for $Y_{t \wedge \tau}$. We fix $T > 0$. For all $t \in [0, T]$, use the bound (5.6):

$$\begin{aligned} \mathbb{E}[(X_{t \wedge \tau} - Y_{t \wedge \tau})^2] &\leq 2\mathbb{E}\left[\left(\int_0^{t \wedge \tau} (\sigma(s, X_s) - \sigma(s, Y_s)) dB_s\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^{t \wedge \tau} (b(s, X_s) - b(s, Y_s)) ds\right)^2\right] \\ &\leq 2\mathbb{E}\left[\int_0^{t \wedge \tau} (\sigma(s, X_s) - \sigma(s, Y_s))^2 ds\right] + 2\mathbb{E}\left[T \int_0^{t \wedge \tau} (b(s, X_s) - b(s, Y_s))^2 ds\right] \\ &\leq 2(1 + T)L^2\mathbb{E}\left[\int_0^{t \wedge \tau} (X_s - Y_s)^2 ds\right] \\ &\leq 2(1 + T)L^2\mathbb{E}\left[\int_0^t (X_{s \wedge \tau} - Y_{s \wedge \tau})^2 ds\right]. \end{aligned}$$

By Lemma 5.22, we have $\mathbb{E}[(X_{t \wedge \tau} - Y_{t \wedge \tau})^2] = 0$, and $X_{t \wedge \tau} = Y_{t \wedge \tau}$ a.s. for all $t \in [0, T]$. Let $M \rightarrow \infty$ and $T \rightarrow \infty$, we then have $X_t = Y_t$ a.s. for all $t \geq 0$. The indistinguishability of X and Y then follows from sample-continuity and a density argument.

Next we construct a solution of $E^x(\sigma, b)$ using Picard's approximation. Define by induction:

$$\begin{aligned} X_t^0 &= x, \quad X_t^1 = x + \int_0^t \sigma(s, x) dB_s + \int_0^t b(s, x) ds, \\ X_t^n &= x + \int_0^t \sigma(s, X_s^{n-1}) dB_s + \int_0^t b(s, X_s^{n-1}) ds, \quad n \in \mathbb{N}. \end{aligned}$$

Clearly, X^n is sample-continuous and adapted to the completed canonical filtration of B . We fix $T > 0$, and find a strong solution on $[0, T]$. Define

$$g_n(t) = \mathbb{E}\left[\sup_{s \in [0, t]} |X_s^n - X_s^{n-1}|^2\right], \quad \forall t \in [0, T].$$

Since $\sigma(\cdot, x)$ is continuous, the process $(\int_0^t \sigma(s, x) dB_s)_{t \geq 0}$ is a continuous local martingale with finite quadratic variation, hence is a martingale by Corollary 3.59. Consequently, we can use Doob's L^2 -inequality [Proposition 3.40 (ii)] and boundedness of functions $\sigma(\cdot, x)$ and $b(\cdot, x)$ on $[0, T]$ to find some constant $C_T > 0$ depending only on T such that $g_1(t) \leq C_T$ for all $t \in [0, T]$.

Now we bound g_n by induction. For any $n \in \mathbb{N}$, one have

$$X_t^{n+1} - X_t^n = \int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1})) dB_s + \int_0^t (b(s, X_s^n) - b(s, X_s^{n-1})) ds.$$

We then use the Burkholder-Davis-Gundy inequality [Theorem 5.15]:

$$\begin{aligned}
& \mathbb{E} \left[\sup_{s \in [0, t]} (X_s^{n+1} - X_s^n)^2 \right] \\
& \leq 2\mathbb{E} \left[\sup_{s \in [0, t]} \left| \int_0^s (\sigma(r, X_r^n) - \sigma(r, X_r^{n-1})) dB_r \right|^2 \right] + 2\mathbb{E} \left[\sup_{s \in [0, t]} \left| \int_0^s (b(r, X_r^n) - b(r, X_r^{n-1})) dr \right|^2 \right] \\
& \leq 2C_2\mathbb{E} \left[\int_0^t (\sigma(r, X_r^n) - \sigma(r, X_r^{n-1}))^2 dr \right] + 2\mathbb{E} \left[T \int_0^t (b(r, X_r^n) - b(r, X_r^{n-1}))^2 dr \right] \\
& \leq 2(C_2 + T)L^2\mathbb{E} \left[\int_0^t (X_r^n - X_r^{n-1})^2 dr \right]
\end{aligned}$$

Consequently, we have

$$g_{n+1}(t) \leq 2(C_2 + T)L^2 \int_0^t g_n(s) ds, \quad \forall t \in [0, T].$$

Since $g_1(t) \leq C_T$ for all $t \in [0, T]$, an induction gives

$$g_n(t) \leq C_T \frac{(2(C_2 + T)L^2 t)^{n-1}}{(n-1)!}, \quad \forall t \in [0, T].$$

Hence we have

$$\sum_{n=1}^{\infty} g_n(T)^{1/2} < \infty \quad \Rightarrow \quad \sum_{n=1}^{\infty} \sup_{t \in [0, T]} |X_t^n - X_t^{n-1}| < \infty \quad a.s..$$

By Weierstrass M-test, the sequence of processes $(X_t^n, t \in [0, T])_{n=1}^{\infty}$ converges *a.s.* uniformly to a limiting process $(X_t, t \in [0, T])$, which also has continuous sample paths on $[0, T]$ and is adapted to the completed canonical filtration of B . Furthermore, using Lipschitz property of $\sigma(t, \cdot)$ and $b(t, \cdot)$ and dominated convergence theorem for stochastic integrals [Theorem 5.7], the following convergences hold in probability:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \left(\int_0^t \sigma(s, X_s) dB_s - \int_0^t \sigma(s, X_s^n) dB_s \right) &= 0, \\
\lim_{n \rightarrow \infty} \left(\int_0^t b(s, X_s) ds - \int_0^t b(s, X_s^n) ds \right) &= 0,
\end{aligned}$$

where we use $\sum_{n=1}^{\infty} \sup_{s \in [0, t]} |X_s^n - X_s^{n-1}|$ to dominate the stochastic parts. By passing these limits to the definition of X_t^n , we conclude that X_t is a strong solution of $E^x(\sigma, b)$ on $[0, T]$. Let $T \rightarrow \infty$, we obtain a process $X = (X_t)_{t \geq 0}$ solving $E^x(\sigma, b)$, and the uniqueness of this strong solution follows from pathwise uniqueness. \square

Theorem 5.24. Equip both spaces $C(\mathbb{R}_+, \mathbb{R}^p)$ and $C(\mathbb{R}_+, \mathbb{R}^q)$ with the Borel σ -algebra of the compact convergence topology, and complete this σ -algebra on $C(\mathbb{R}_+, \mathbb{R}^p)$ by W -negligible sets, where W is the Wiener measure. Under the assumptions of the preceding theorem, there exists a measurable mapping $F_x : C(\mathbb{R}_+, \mathbb{R}^q) \rightarrow C(\mathbb{R}_+, \mathbb{R}^p)$ such that

- (i) for every $t \geq 0$, the mapping $\mathbf{w} \mapsto F_x(\mathbf{w})_t$ coincides W -*a.s.* with a measurable function of $(\mathbf{w}(r))_{0 \leq r \leq t}$;
- (ii) for every $\mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}^q)$, the mapping $x \mapsto F_x(\mathbf{w})$ is continuous;
- (iii) for every $t \geq 0$, and for every choice of the complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and of the (\mathcal{F}_t) -Brownian motion B with $B_0 = 0$, the process $X_t = F_x(B)_t$ is the unique solution of $E^x(\sigma, b)$; furthermore, if U is an \mathcal{F}_0 -measurable \mathbb{R}^p -valued random variable, the process $F_U(B)_t$ is the unique solution of $E(\sigma, b)$ with $X_0 = U$.

Proof. Step I: For brevity, we only consider the case $p = q = 1$. We let \mathcal{G}_t be the σ -algebra generated by projection mappings $\{\pi_s : 0 \leq s \leq t\}$ and all W -negligible sets in $C(\mathbb{R}_+, \mathbb{R})$, so $(\mathcal{G}_t)_{t \geq 0}$ is a complete filtration. By Theorem 5.23, with the filtered probability space $(C(\mathbb{R}_+, \mathbb{R}), \mathcal{G}_\infty, (\mathcal{G}_t)_{t \geq 0}, W)$ fixed, and with the Brownian motion $B_t(\mathbf{w}) = \mathbf{w}(t)$ fixed as the canonical process, there exists a unique (up to indistinguishability) and strong solution $X^x = (X_t^x)_{t \geq 0}$ of $E^x(\sigma, b)$ for every $x \in \mathbb{R}$.

Step II: Let d be a metric on $C(\mathbb{R}_+, \mathbb{R})$ that induces the compact convergence topology. We fix $q \geq 2$, and prove that there exists a constant $C_q^* > 0$ depending only on q , such that for all $x, y \in \mathbb{R}$,

$$\mathbb{E}[d(X^x, X^y)^q] \leq C_q^* |x - y|^q. \quad (5.14)$$

Then by Kolmogorov's lemma [Theorem 4.6] applied to the process $(X^x)_{x \in \mathbb{R}}$ taking values in $C(\mathbb{R}_+, \mathbb{R})$, we find a modification $(\tilde{X}^x)_{x \in \mathbb{R}}$ of $(X^x)_{x \in \mathbb{R}}$ with continuous sample paths. Define $F_x(\mathbf{w}) = \tilde{X}^x(\mathbf{w}) = (\tilde{X}_t^x(\mathbf{w}))_{t \geq 0}$. Then $F_x : (C(\mathbb{R}_+, \mathbb{R}), \mathcal{G}_\infty) \rightarrow (C(\mathbb{R}_+, \mathbb{R}), \mathcal{G}_\infty)$ is a measurable mapping with property (ii).

To this end, we define the stopping time

$$\tau_n = \inf \{t \geq 0 : |X_t^x| \vee |X_t^y| \geq n\}, \quad n = 1, 2, \dots$$

We fix some $T \geq 1$. For every $t \in [0, T]$, we apply Jensen's inequality, Burkholder-Davis-Gundy inequality [Theorem 5.15], Hölder's inequality and Lipschitz property as follows:

$$\begin{aligned} & \mathbb{E} \left[\sup_{s \in [0, t]} |X_{s \wedge \tau_n}^y - X_{s \wedge \tau_n}^x|^q \right] \\ & \leq 3^{q-1} \left(|x - y|^q + \mathbb{E} \left[\sup_{s \in [0, t]} \left| \int_0^{s \wedge \tau_n} (\sigma(r, X_r^x) - \sigma(r, X_r^y)) dB_r \right|^q + \sup_{s \in [0, t]} \left| \int_0^{s \wedge \tau_n} (b(r, X_r^x) - b(r, X_r^y)) dr \right|^q \right] \right) \\ & \leq 3^{q-1} \left(|x - y|^q + C_q \mathbb{E} \left[\left(\int_0^{t \wedge \tau_n} (\sigma(r, X_r^x) - \sigma(r, X_r^y))^2 dr \right)^{q/2} \right] + \mathbb{E} \left[\left(\int_0^{t \wedge \tau_n} |b(r, X_r^x) - b(r, X_r^y)| dr \right)^q \right] \right) \\ & \leq 3^{q-1} \left(|x - y|^q + C_q t^{\frac{q}{2}-1} \mathbb{E} \left[\int_0^t |\sigma(r \wedge \tau_n, X_{r \wedge \tau_n}^x) - \sigma(r \wedge \tau_n, X_{r \wedge \tau_n}^y)|^q dr \right] \right. \\ & \quad \left. + t^{q-1} \mathbb{E} \left[\int_0^t |b(r \wedge \tau_n, X_{r \wedge \tau_n}^x) - b(r \wedge \tau_n, X_{r \wedge \tau_n}^y)|^q dr \right] \right) \\ & \leq 3^{q-1} \left(|x - y|^q + K^q T^{\frac{q}{2}-1} (C_q + T^{q/2}) \mathbb{E} \left[\int_0^t |X_{r \wedge \tau_n}^x - X_{r \wedge \tau_n}^y|^q dr \right] \right) \end{aligned}$$

We let $C'_q = 3^{q-1} K^q (1 + C_q)$. Then we obtain the following estimate by using Gronwall's lemma [Lemma 5.22] on the bounded function $t \mapsto \mathbb{E} \left[\sup_{s \in [0, t]} |X_{s \wedge \tau_n}^y - X_{s \wedge \tau_n}^x|^q \right]$:

$$\mathbb{E} \left[\sup_{s \in [0, t]} |X_{s \wedge \tau_n}^y - X_{s \wedge \tau_n}^x|^q \right] \leq C'_q |x - y|^q \exp(C'_q T^{q-1} t), \quad \forall t \in [0, T].$$

A monotone convergence argument follows by letting $n \rightarrow \infty$:

$$\mathbb{E} \left[\sup_{s \in [0, t]} |X_s^y - X_s^x|^q \right] \leq C'_q |x - y|^q \exp(C'_q t^q), \quad \forall t \geq 0.$$

We define the following metric d on $C(\mathbb{R}_+, \mathbb{R})$, which induces the uniform topology:

$$d(\mathbf{w}, \mathbf{w}') = \sum_{k=1}^{\infty} \alpha_k \left(\sup_{s \in [0, k]} |\mathbf{w}(s) - \mathbf{w}'(s)| \wedge 1 \right),$$

where $\alpha_k > 0$ is a real sequence such that $\sum_{k=1}^{\infty} \alpha_k$ converges. Choose (α_k) such that $\sum_{k=1}^{\infty} \alpha_k \exp(C'_q k^q) < \infty$. Again by Hölder's inequality, one have

$$\mathbb{E} [d(X_s^x, X_s^y)^q] \leq \left(\sum_{k=1}^{\infty} \alpha_k \right)^{q-1} \sum_{k=1}^{\infty} \alpha_k \mathbb{E} \left[\sup_{s \in [0, k]} |X_s^x - X_s^y|^q \right] \leq C_q^* |x - y|^q,$$

where $C_q^* = C'_q (\sum_{k=1}^{\infty} \alpha_k)^{q-1} (\sum_{k=1}^{\infty} \alpha_k \exp(C'_q k^q))$. This complete the proof of (??). For the assertion (i), we point out that for any $t \geq 0$, the mapping $\mathbf{w} \mapsto F_x(\mathbf{w})_t = \tilde{X}_t^x \stackrel{a.s.}{=} X_t^x$ is \mathcal{G}_t -measurable. The result then follows from Doob-Dynkin theorem [Theorem 2.22].

Step III: We prove the first part of assertion (iii). Fix a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and an (\mathcal{F}_t) -Brownian motion B . Clearly, the process $F_x(B) = (F_x(B)_t)_{t \geq 0}$ has continuous paths, and is also adapted since $F_x(B)_t$ coincide *a.s.* with a measurable function of $(B_s)_{s \in [0, t]}$ by (i), and since the filtration $(\mathcal{F}_t)_{t \geq 0}$ is complete. Then it remains to show that $F_x(B)$ solves $E^x(\sigma, b)$.

By construction of F_x and the fact that $X^x = \tilde{X}^x$ W -*a.s.*, for all $t \geq 0$, we have

$$F_x(\mathbf{w})_t = x + \int_0^t \sigma(s, F_x(\mathbf{w})_s) d\mathbf{w}(s) + \int_0^t b(s, F_x(\mathbf{w})_s) ds, \quad \text{for } W\text{-a.s. } \mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}).$$

By Proposition 5.8, we have the following approximation:

$$\int_0^t \sigma(s, F_x(\mathbf{w})_s) d\mathbf{w}(s) = \lim_{n \rightarrow \infty} \sum_{j=1}^{2^n} \sigma \left(\frac{(j-1)t}{2^n}, F_x(\mathbf{w})_{\frac{(j-1)t}{2^n}} \right) \left(\mathbf{w} \left(\frac{jt}{2^n} \right) - \mathbf{w} \left(\frac{(j-1)t}{2^n} \right) \right) \quad (5.15)$$

in probability $W(d\mathbf{w})$. Since W is the law of B , by Proposition 5.8, we have

$$\begin{aligned} F_x(B)_s &= x + \lim_{n \rightarrow \infty} \sum_{j=1}^{2^n} \sigma \left(\frac{(j-1)t}{2^n}, F_x(B)_{\frac{(j-1)t}{2^n}} \right) \left(B_{\frac{jt}{2^n}} - B_{\frac{(j-1)t}{2^n}} \right) + \int_0^t b(s, F_x(B)_s) ds \\ &= x + \int_0^t \sigma(s, F_x(B)_s) dB_s + \int_0^t b(s, F_x(B)_s) ds, \quad a.s., \end{aligned}$$

where the *a.s.* convergence follows by passing the convergence in probability to an appropriate subsequence. Therefore, $F_x(B)$ is the desired solution of $E^x(\sigma, b)$.

Step IV: We prove the first part of assertion (iii). The mapping $x \mapsto F_x(B)_t(\omega)$ is continuous for any fixed $\omega \in \Omega$, and the mapping $\omega \mapsto F_x(B)_t(\omega)$ is \mathcal{F}_t -measurable for any fixed $\omega \in \Omega$. Then $(x, \omega) \mapsto F_x(B)_t(\omega)$ is $\mathcal{B}(\mathbb{R}) \otimes \mathcal{F}_t$ -measurable according to a similar procedure in Proposition 3.9. If U is a \mathcal{F}_0 -measurable random variable, then $F_U(B)_t$ is a composition of $\omega \mapsto (U(\omega), \omega)$ and $(x, \omega) \mapsto F_x(B)_t(\omega)$, hence is \mathcal{F}_t -measurable.

Let $H(x, \mathbf{w})_t = F_x(\mathbf{w})_t - x - \int_0^t b(s, F_x(\mathbf{w})_s) ds$. We use the convergence result (5.15) in probability W . Since $B \sim W$, and U is a \mathcal{F}_0 -measurable variable, which is independent of B , we have that

$$H(U, B)_t = \sum_{j=1}^{2^n} \sigma \left(\frac{(j-1)t}{2^n}, F_U(B)_{\frac{(j-1)t}{2^n}} \right) \left(B_{\frac{jt}{2^n}} - B_{\frac{(j-1)t}{2^n}} \right),$$

where the series converges in probability, and the limit is the stochastic integral $\int_0^t \sigma(s, F_U(B)_s) dB_s$. Hence

$$F_U(B)_t - U - \int_0^t b(s, F_U(B)_s) ds = \int_0^t \sigma(s, F_U(B)_s) dB_s.$$

Consequently, $F_U(B) = (F_U(B)_t)_{t \geq 0}$ solves the SDE $E(\sigma, b)$ with initial value U . □

5.5 Girsanov's Theorem and Cameron–Martin Formula

In Section 5.2, we show that the class of continuous semimartingales is invariant under composition with C^2 -function. In this section, we study the effect on the class of continuous semimartingales of an absolutely continuous transformation of probability measures. We consider two probability measures \mathbb{P} and \mathbb{Q} on the same measurable space (Ω, \mathcal{F}) . To avoid confusion, we write $\mathbb{E}_{\mathbb{P}}$ and $\mathbb{E}_{\mathbb{Q}}$ for the expectation under \mathbb{P} and \mathbb{Q} , respectively. Unless otherwise specified, our notions of semimartingales refer to the underlying probability measure \mathbb{P} . We will point it out explicitly when consider these notions under \mathbb{Q} .

Throughout this section, we assume $(\mathcal{F}_t)_{t \geq 0}$ is a complete and right continuous filtration. Most of the time we may assume that \mathbb{P} and \mathbb{Q} are mutually absolutely continuous, hence the filtration $(\mathcal{F}_t)_{t \geq 0}$, being complete with respect to \mathbb{P} , is also complete with respect to \mathbb{Q} .

Proposition 5.25. Assume that \mathbb{Q} is a probability measure on $(\Omega, \mathcal{F}_{\infty})$ which is absolutely continuous with respect to \mathbb{P} on \mathcal{F}_t for every $t \geq 0$. Let D_t be the Radon-Nikodym derivative of \mathbb{Q} with respect to \mathbb{P} on \mathcal{F}_t :

$$D_t = \left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t}, \quad t \in \mathbb{R}_+.$$

Then $D = (D_t)_{t \geq 0}$ is a \mathbb{P} -martingale, and D has a càdlàg modification as a consequence of Theorem 3.44. Furthermore, the following two assertions are equivalent:

- (i) the martingale D is uniformly integrable;
- (ii) $\mathbb{Q} \ll \mathbb{P}$ on \mathcal{F}_{∞} .

Proof. We fix $t > s \geq 0$. Then for all $A \in \mathcal{F}_s$, we have $\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A D_s]$. For the martingale property,

$$\mathbb{E}_{\mathbb{P}}[D_s \mathbb{1}_A] = \mathbb{E}_{\mathbb{P}}[D_t \mathbb{1}_A] = \mathbb{Q}(A), \quad \forall A \in \mathcal{F}_s \subset \mathcal{F}_t.$$

Hence we have $D_s = \mathbb{E}[D_t | \mathcal{F}_s]$, and D is a martingale. If (i) holds, let Z be the a.s. and L^1 limit of $(D_t)_{t \geq 0}$. Using a monotone class argument, we have $\mathbb{E}_{\mathbb{P}}[Z \mathbb{1}_A] = \mathbb{Q}(A)$ for all $A \in \mathcal{F}_{\infty}$, hence $\mathbb{Q} \ll \mathbb{P}$ on \mathcal{F}_{∞} . If (ii) holds, let D_{∞} be the Radon-Nikodym derivative of \mathbb{Q} with respect to \mathbb{P} on \mathcal{F}_t . Then we have $\mathbb{E}[D_{\infty} | \mathcal{F}_t] = D_t$ for all $t \geq 0$, implying uniform integrability. \square

In the sequel, we assume that the martingale $D = (D_t)_{t \geq 0}$ has càdlàg sample paths.

Proposition 5.26. Under the assumption of the preceding proposition, for every stopping time τ , we have $d\mathbb{Q} = D_{\tau} d\mathbb{P}$ on $\mathcal{F}_{\tau} \cap \{\tau < \infty\}$. Furthermore, if $\mathbb{Q} \ll \mathbb{P}$ on \mathcal{F}_{∞} , we have

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

Proof. For the uniform integrable case where $\mathbb{Q} \ll \mathbb{P}$ on \mathcal{F}_{∞} , by optional stopping theorem [Theorem 3.49],

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[D_{\infty} \mathbb{1}_A] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[D_{\infty} | \mathcal{F}_{\tau}] \mathbb{1}_A] = \mathbb{E}_{\mathbb{P}}[D_{\tau} \mathbb{1}_A], \quad \forall A \in \mathcal{F}_{\tau}.$$

Since D_{τ} is \mathcal{F}_{τ} -measurable, the second assertion follows. For general case, we use the fact that the stopped martingale $(D_{s \wedge t})_{s \geq 0}$ is uniformly integrable for every $t \geq 0$. Then

$$\mathbb{Q}(A \cap \{\tau \leq t\}) = \mathbb{E}_{\mathbb{P}}[D_{\tau \wedge t} \mathbb{1}_{A \cap \{\tau \leq t\}}] = \mathbb{E}_{\mathbb{P}}[D_{\tau} \mathbb{1}_{A \cap \{\tau \leq t\}}], \quad \forall A \in \mathcal{F}_{\tau}.$$

Letting t tends to infinity concludes the proof. \square

Remark. Let $X = (X_t)_{t \geq 0}$ be an adaptive process with càdlàg sample paths. If XD is a \mathbb{P} -local martingale, then X is a \mathbb{Q} -local martingale. We give a clarification below.

Let τ be a stopping time. Then for any $s \geq 0$ and any $A \in \mathcal{F}_s$, we have $A \cap \{\tau > s\} \in \mathcal{F}_s \cap \mathcal{F}_\tau = \mathcal{F}_{\tau \wedge s}$. Let $t > s \geq 0$. If the stopped process $(XD)^\tau$ is a \mathbb{P} -martingale, then

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[X_{t \wedge \tau} \mathbf{1}_A] &= \mathbb{E}_{\mathbb{Q}}[X_{t \wedge \tau} \mathbf{1}_{A \cap \{\tau \leq s\}}] + \mathbb{E}_{\mathbb{Q}}[X_{t \wedge \tau} \mathbf{1}_{A \cap \{\tau > s\}}] \\ &= \mathbb{E}_{\mathbb{Q}}[X_{s \wedge \tau} \mathbf{1}_{A \cap \{\tau \leq s\}}] + \mathbb{E}_{\mathbb{P}}[X_{t \wedge \tau} D_{t \wedge \tau} \mathbf{1}_{A \cap \{\tau > s\}}] \\ &= \mathbb{E}_{\mathbb{Q}}[X_{s \wedge \tau} \mathbf{1}_{A \cap \{\tau \leq s\}}] + \mathbb{E}_{\mathbb{P}}[X_{s \wedge \tau} D_{s \wedge \tau} \mathbf{1}_{A \cap \{\tau > s\}}] = \mathbb{E}_{\mathbb{Q}}[X_{s \wedge \tau} \mathbf{1}_A]. \end{aligned}$$

Hence X^τ is a \mathbb{Q} -martingale. In addition, a sequence of stopping times increasing \mathbb{P} -a.s. to ∞ also increases \mathbb{Q} -a.s. to ∞ . Consequently, if XD is a \mathbb{P} -local martingale, then X is a \mathbb{Q} -local martingale.

Proposition 5.27. Under the preceding assumption, the martingale D is \mathbb{Q} -a.s. strictly positive, i.e.

$$\inf_{t \geq 0} D_t > 0, \quad \mathbb{Q}\text{-a.s.}$$

Proof. For every $n \in \mathbb{N}$, define the stopping time $\tau_n = \inf\{t \geq 0 : D_t < 1/n\}$. Then the event $\{\tau_n < \infty\}$ is \mathcal{F}_{τ_n} -measurable, and $D_{\tau_n} \leq 1/n$ on $\{\tau_n < \infty\}$ by right-continuity of D . Hence

$$\mathbb{Q}(\tau_n < \infty) = \mathbb{E}_{\mathbb{P}}[D_{\tau_n} \mathbf{1}_{\{\tau_n < \infty\}}] \leq \frac{1}{n} \quad \Rightarrow \quad \mathbb{Q}\left(\bigcap_{n=1}^{\infty} \{\tau_n < \infty\}\right) = 0.$$

Then \mathbb{Q} -a.s. there exists $n \in \mathbb{N}$ such that $\tau_n = \infty$. This complete the proof. \square

Remark. If we further assume that \mathbb{P} and \mathbb{Q} are mutually absolutely continuous, then the martingale D is also \mathbb{P} -a.s. strictly positive.

Theorem 5.28 (Girsanov). Suppose the assumption of the preceding proposition holds, and assume that the martingale $D = (D_t)_{t \geq 0}$ is sample-continuous. If $M = (M_t)_{t \geq 0}$ is a continuous \mathbb{P} -local martingale, then

$$\widetilde{M} = M - D^{-1} \cdot \langle M, D \rangle$$

is a continuous \mathbb{Q} -local martingale. Furthermore, if N is another continuous \mathbb{P} -local martingale, then

$$\langle M, N \rangle = \langle \widetilde{M}, N \rangle = \langle \widetilde{M}, \widetilde{N} \rangle.$$

Proof. By Proposition 5.27, the process $D^{-1} \cdot \langle M, D \rangle$ is \mathbb{P} -a.s. of finite variation, and the process \widetilde{M} is a \mathbb{P} -semimartingale. According to the integration by parts formula, we have

$$\begin{aligned} (\widetilde{M}D)_t &= \widetilde{M}_0 D_0 + \int_0^t \widetilde{M}_s dD_s + \int_0^t D_s d\widetilde{M}_s + \langle \widetilde{M}, D \rangle_t \\ &= \widetilde{M}_0 D_0 + \int_0^t \widetilde{M}_s dD_s + \int_0^t D_s dM_s - \langle M, D \rangle_t + \langle \widetilde{M}, D \rangle_t \\ &= \widetilde{M}_0 D_0 + \int_0^t \widetilde{M}_s dD_s + \int_0^t D_s dM_s. \end{aligned}$$

Consequently, the process $\widetilde{M}D$ is a continuous \mathbb{P} -local martingale. By the Remark under Proposition 5.26, the process \widetilde{M} is a continuous \mathbb{Q} -local martingale. The last assertion holds because the bracket of a finite variation process and a semimartingale vanishes. \square

Proposition 5.29. If D is a continuous local martingale taking strictly positive values. There exists a unique continuous local martingale L such that

$$D_t = \mathcal{E}(L)_t = \exp \left(L_t - \frac{1}{2} \langle L, L \rangle_t \right).$$

Moreover, L is given by the formula

$$L_t = \log D_0 + \int_0^t D_s^{-1} dD_s. \quad (5.16)$$

Proof. We first prove uniqueness: If both L and L' has the desired property, then $L - L' = \frac{1}{2} \langle L', L' \rangle - \frac{1}{2} \langle L, L \rangle$ is a continuous martingale of finite variation, hence is constantly zero [Proposition 3.55]. To show the second assertion, use Itô's formula to the process $\log D$:

$$\log D_t = \log D_0 + \int_0^t D_s^{-1} dD_s - \frac{1}{2} \int_0^t D_s^{-2} d\langle D, D \rangle_s = L_t - \frac{1}{2} \langle L, L \rangle_t,$$

where L is given in (5.16). □

Theorem 5.30 (Girsanov). Suppose the assumption of the preceding proposition holds, and assume that the martingale $D = (D_t)_{t \geq 0}$ is sample-continuous. If $M = (M_t)_{t \geq 0}$ is a continuous \mathbb{P} -local martingale, then

$$\widetilde{M} = M - D^{-1} \cdot \langle M, D \rangle = M - \langle M, L \rangle.$$

is a continuous \mathbb{Q} -local martingale, where L is given in (5.16). Moreover, $D^{-1} = \mathcal{E}(-\widetilde{L})$.

Proof. The first identity immediately follows from associativity of stochastic integral. For the second assertion, we use the identity $\widetilde{L} = L - \langle L, L \rangle$:

$$\mathcal{E}(-\widetilde{L})_t = \exp \left(-\widetilde{L}_t - \frac{1}{2} \langle \widetilde{L}, \widetilde{L} \rangle_t \right) = \exp \left(-L_t + \frac{1}{2} \langle L, L \rangle_t \right) = \mathcal{E}(L)_t^{-1}.$$

This proves the second assertion. □

Remark I. According to Theorem 5.30, if P and Q are mutually absolutely continuous on \mathcal{F}_∞ , then the role of P and Q can be exchanged by replacing L with $-\widetilde{L}$.

Remark II. If $M = B$ is an (\mathcal{F}_t) -Brownian motion under P , then $\widetilde{B} = B - \langle B, L \rangle$ is a \mathbb{Q} -continuous local martingale, and $\langle \widetilde{B}, \widetilde{B} \rangle_t = \langle B, B \rangle_t = t$. By Lévy's characterization of Brownian motions [Theorem 5.11], \widetilde{B} is an (\mathcal{F}_t) -Brownian motion under \mathbb{Q} .

Theorem 5.31 Let L be a continuous local martingale such that $L_0 = 0$. Consider the following properties:

- (i) (Novikov's criterion). $\mathbb{E} \left[\exp \left(\frac{1}{2} \langle L, L \rangle_\infty \right) \right] < \infty$.
- (ii) (Kazamaki's criterion). L is a uniformly integrable martingale, and $\mathbb{E} \left[\exp \left(\frac{1}{2} L_\infty \right) \right] < \infty$.
- (iii) $\mathcal{E}(L)$ is a uniformly integrable martingale.

Then (i) \Rightarrow (ii) \Rightarrow (iii).

Proof. (i) \Rightarrow (ii): The process $\mathcal{E}(L)$ is a nonnegative continuous local martingale, hence is a supermartingale by Theorem 3.54 (i). By Fatou's lemma,

$$\mathbb{E}[\mathcal{E}(L)_\infty] \leq \lim_{t \rightarrow \infty} \mathbb{E}[\mathcal{E}(L)_t] \leq \mathbb{E}[\mathcal{E}(L)_0] = 1.$$

By property (i), we have $\mathbb{E}[\langle L, L \rangle_\infty] < \infty$, and L is a continuous martingale that is bounded in L^2 by Theorem 3.48. By Cauchy-Schwarz inequality,

$$\mathbb{E} \left[\exp \left(\frac{1}{2} L_\infty \right) \right] = \mathbb{E} \left[\mathcal{E}(L)_\infty^{1/2} e^{\frac{1}{4} \langle L, L \rangle_\infty} \right] \leq \sqrt{\mathbb{E}[\mathcal{E}(L)_\infty]} \sqrt{\mathbb{E} \left[e^{\frac{1}{2} \langle L, L \rangle_\infty} \right]} \leq \sqrt{\mathbb{E} \left[e^{\frac{1}{2} \langle L, L \rangle_\infty} \right]} < \infty.$$

(ii) \Rightarrow (iii): If L is a uniformly integrable martingale, by optional stopping theorem, for any stopping time τ , one have $L_\tau = \mathbb{E}[L_\infty | \mathcal{F}_\tau]$. By Jensen's inequality,

$$\mathbb{E} \left[\exp \left(\frac{1}{2} L_\tau \right) \right] \leq \mathbb{E} \left[\exp \left(\frac{1}{2} L_\infty \right) | \mathcal{F}_\tau \right].$$

Since $\exp(\frac{1}{2} L_\infty)$ is integrable, the collection of random variables $\exp(\frac{1}{2} L_\tau)$ is uniformly integrable, where τ runs over all stopping times. On the other hand, set $Z_t^{(a)} = \exp\left(\frac{aL_t}{1+a}\right)$. Then for all $0 < a < 1$,

$$\mathcal{E}(aL)_t = (\mathcal{E}(L)_t)^{a^2} (Z_t^{(a)})^{1-a^2}.$$

By Hölder's inequality, for any measurable set $\Gamma \in \mathcal{F}$ and any stopping time τ , one have

$$\mathbb{E}[\mathbf{1}_\Gamma \mathcal{E}(aL)_\tau] \leq \mathbb{E}[\mathcal{E}(L)_\tau]^{a^2} \mathbb{E}[\mathbf{1}_\Gamma Z_\tau^{(a)}]^{1-a^2} \leq \mathbb{E}[\mathbf{1}_\Gamma Z_\tau^{(a)}]^{1-a^2} \leq \mathbb{E} \left[\mathbf{1}_\Gamma \exp \left(\frac{1}{2} L_\tau \right) \right]^{2a(1-a)},$$

where we also use Jensen's inequality and the fact $\frac{1+a}{2a} > 1$ in the last inequality. Consequently, the collection of random variables $\mathcal{E}(aL)_\tau$ is uniformly integrable, where τ runs over all stopping times. Let $\tau_n \wedge \infty$ be a sequence of stopping times reducing $\mathcal{E}(aL)$. Then for all $t > s \geq 0$, by uniform integrability,

$$\mathbb{E}[\mathcal{E}(aL)_t | \mathcal{F}_s] = \lim_{n \rightarrow \infty} \mathbb{E}[\mathcal{E}(aL)_{t \wedge \tau_n} | \mathcal{F}_s] = \lim_{n \rightarrow \infty} \mathcal{E}(aL)_{s \wedge \tau_n} = \mathcal{E}(aL)_s.$$

Hence $\mathcal{E}(aL)$ is a uniformly integrable martingale, and

$$1 = \mathbb{E}[\mathcal{E}(aL)_\infty] \leq \mathbb{E}[\mathcal{E}(L)_\infty]^{a^2} \mathbb{E}[Z_\infty^{(a)}]^{1-a^2} \leq \mathbb{E}[\mathcal{E}(L)_\infty]^{a^2} \mathbb{E} \left[\exp \left(\frac{1}{2} L_\infty \right) \right]^{2a(1-a)} \Rightarrow \mathbb{E}[\mathcal{E}(L)_\infty] = 1.$$

Again, by Fatou's lemma, $\mathbb{E}[\mathcal{E}(L)_\infty | \mathcal{F}_t] \leq \mathcal{E}(L)_t$. On the other hand, we have $\mathbb{E}[\mathcal{E}(L)_\infty] = \mathbb{E}[\mathcal{E}(L)_t] = \mathbb{E}[\mathcal{E}(L)_0] = 1$. Hence $\mathbb{E}[\mathcal{E}(L)_\infty | \mathcal{F}_t] = \mathcal{E}(L)_t$, and $\mathcal{E}(L)$ is a uniformly integrable martingale. \square

Remark. Let L be a continuous \mathbb{P} -local martingale satisfying property (ii). To apply Girsanov's theorem, we let \mathbb{Q} be the probability measure with density $\mathcal{E}(L)_\infty$ with respect to \mathbb{P} . According to Proposition 5.25, the Radon-Nikodym derivative is $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = D_t = \mathcal{E}(L)_t$.

Girsanov transformation and SDE. Let $\beta = (\beta_t)_{t \geq 0}$ be a p -dimensional (\mathcal{F}_t) -Brownian motion under \mathbb{P} . Consider the following continuous local martingale:

$$L_t = \int_0^t b(s, \beta_s) d\beta_s.$$

If there exists $g \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$ such that $|b(t, x)| \leq g(t)$ for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^p$, the Novikov's criterion is satisfied. Then the associated exponential martingale is given by

$$D_t = \mathcal{E}(L)_t = \exp \left(\int_0^t b(s, \beta_s) d\beta_s - \frac{1}{2} \int_0^t |b(s, \beta_s)|^2 ds \right).$$

We set $d\mathbb{Q} = D_\infty d\mathbb{P}$, which is a probability distribution. By Remark II under Theorem 5.30, the following process B is an (\mathcal{F}_t) -Brownian motion under \mathbb{Q} :

$$B_t = \beta_t - \int_0^t b(s, \beta_s) ds$$

Consequently, $X = \beta$ solves the following SDE under probability measure \mathbb{Q} :

$$dX_t = dB_t + b(t, X_t) dt$$

Here we only assume that $b : \mathbb{R}_+ \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ is dominated by a function $g \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$.

Proposition 5.32. Consider the following two SDEs admitting unique strong solution on \mathbb{R}_+ :

$$\begin{cases} dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t, \\ dY_t = (\mu + \nu)(t, Y_t) dt + \sigma(t, Y_t) dB_t, \end{cases}$$

where B_t is a p -dimensional (\mathcal{F}_t) -Brownian motion under \mathbb{P} , and $\sigma : \mathbb{R}_+ \times \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$ is almost everywhere invertible. Furthermore, the *Novikov's condition* is satisfied:

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^\infty |\sigma(s, Y_s)^{-1} \nu(s, Y_s)|^2 ds \right) \right] < \infty.$$

Then, if $X_0 \stackrel{d}{=} Y_0$, the following identity holds for all bounded functional $\Phi : C(\mathbb{R}_+) \rightarrow \mathbb{R}$:

$$\mathbb{E} [\Phi(X)] = \mathbb{E} \left[\Phi(Y) \exp \left(- \int_0^\infty \sigma(s, Y_s)^{-1} \nu(s, Y_s) ds - \frac{1}{2} \int_0^\infty |\sigma(s, Y_s)^{-1} \nu(s, Y_s)|^2 dB_s \right) \right].$$

Proof. Define the following continuous \mathbb{P} -local martingale:

$$L_t = - \int_0^t \sigma(s, Y_s)^{-1} \nu(s, Y_s) dB_s.$$

Since the Novikov's criterion is satisfied, we can use the exponential martingale:

$$\mathcal{E}(L)_t = \exp \left(- \int_0^t \sigma(s, Y_s)^{-1} \nu(s, Y_s) ds - \frac{1}{2} \int_0^t |\sigma(s, Y_s)^{-1} \nu(s, Y_s)|^2 dB_s \right).$$

Define $d\mathbb{Q} = \mathcal{E}(L)_\infty d\mathbb{P}$, and the (\mathcal{F}_t) -Brownian motion under \mathbb{P} :

$$\tilde{B}_t = B_t - \langle B, L \rangle_t = B_t + \int_0^t \sigma(s, Y_s)^{-1} \nu(s, Y_s) ds.$$

Then we have $dY_t = \mu(t, Y_t) dt + \sigma(t, Y_t) d\tilde{B}_t$. Consequently, Y_t solves the first SDE under probability measure \mathbb{Q} and $(\mathcal{F}_t, \mathbb{Q})$ -Brownian motion \tilde{B} . By uniqueness of the strong solution, if $X_0 \stackrel{d}{=} Y_0$, then X and Y has the same law under \mathbb{P} and \mathbb{Q} , respectively. The final result follows by $\mathbb{E}_{\mathbb{P}} [\Phi(X)] = \mathbb{E}_{\mathbb{Q}} [\Phi(Y)]$. \square

We specialize the previous discussion to the case $b(t, x) = g(t)$, where $g \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$. We set

$$h(t) = \int_0^t g(s) ds, \quad \forall t \in \mathbb{R}_+.$$

The set \mathcal{H} of all functions h that can be written in this form is called the *Cameron-Martin space*, and we

write the derivative of $h \in \mathcal{H}$ in sense of distribution by $\dot{h} = g \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), m)$. As a special case of our previous discussion, given a Brownian motion B under \mathbb{P} , we construct the probability measure

$$d\mathbb{Q} = D_\infty d\mathbb{P} = \exp \left(\int_0^\infty g(t) dB_s - \frac{1}{2} \int_0^\infty |g(t)|^2 ds \right) d\mathbb{P}.$$

Then the process $\tilde{B}_t = B_t - h(t)$ is a Brownian motion under \mathbb{Q} . Consequently, for every nonnegative measurable function on $C(\mathbb{R}_+, \mathbb{R})$, one have

$$\mathbb{E}_{\mathbb{P}} [D_\infty \Phi((B_t)_{t \geq 0})] = \mathbb{E}_{\mathbb{Q}} [\Phi((B_t)_{t \geq 0})] = \mathbb{E}_{\mathbb{Q}} \left[\Phi \left((\tilde{B}_t + h(t))_{t \geq 0} \right) \right] = \mathbb{E}_{\mathbb{P}} [\Phi((B_t + h(t))_{t \geq 0})]$$

We rewrite this formula to the following form.

Theorem 5.33 (Cameron-Martin formula). Let W be the Wiener measure on $C(\mathbb{R}_+, \mathbb{R})$, and let h be a function in the Cameron-Martin space \mathcal{H} . Then for every nonnegative measurable function Φ on $C(\mathbb{R}_+, \mathbb{R})$,

$$\int W(d\mathbf{w}) \Phi(\mathbf{w} + h) = \int W(d\mathbf{w}) \exp \left(\int_0^\infty \dot{h}(s) d\mathbf{w}(s) - \frac{1}{2} \int_0^\infty |\dot{h}(s)|^2 ds \right) \Phi(\mathbf{w}).$$

Remark. The integral $\int_0^\infty \dot{h}(s) d\mathbf{w}(s)$ is viewed as the Wiener integral, where $\mathbf{w} \sim W$.

Application: Law of hitting times for Brownian motion with drift. Let B be a real Brownian motion with $B_0 = 0$, and define the hitting time $\tau_a = \inf\{t \geq 0 : B_t = a\}$ for every $a > 0$. Now given a constant $\mu \in \mathbb{R}$, we are interested in the law of the stopping time

$$v_a = \inf\{t \geq 0 : B_t + \mu t = a\}.$$

Clearly, if $\mu = 0$, we have $v_a = \tau_a$, and the desired law is given by Corollary 4.25. For the general case, we fix $t > 0$, and use Cameron-Martin formula to the following function:

$$h(s) = \mu(s \wedge t), \quad \dot{h}(s) = \mu \mathbb{1}_{\{s \leq t\}}, \quad \Phi(\mathbf{w}) = \mathbb{1}_{\{\max_{s \in [0, t]} \mathbf{w}(s) \geq a\}}, \quad \mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}).$$

Then we have

$$\mathbb{P}(v_a \leq t) = \mathbb{E}[\Phi(B + h)] = \mathbb{E} \left[\exp \left(\mu B_t - \frac{\mu^2}{2} t \right) \mathbb{1}_{\{\tau_a \leq t\}} \right].$$

By optional stopping theorem [Theorem 3.48], we have

$$\exp \left(\mu B_{t \wedge \tau_a} - \frac{\mu^2}{2} (t \wedge \tau_a) \right) = \mathbb{E} \left[\exp \left(\mu B_t - \frac{\mu^2}{2} t \right) \middle| \mathcal{F}_{t \wedge \tau_a} \right].$$

Consequently,

$$\begin{aligned} \mathbb{P}(v_a \leq t) &= \mathbb{E} \left[\exp \left(\mu B_{t \wedge \tau_a} - \frac{\mu^2}{2} (t \wedge \tau_a) \right) \mathbb{1}_{\{\tau_a \leq t\}} \right] = \mathbb{E} \left[\exp \left(\mu a - \frac{\mu^2}{2} \tau_a \right) \mathbb{1}_{\{\tau_a \leq t\}} \right] \\ &= \int_0^t \frac{a}{\sqrt{2\pi s^3}} e^{-\frac{1}{2s}(\mu s - a)^2} ds. \end{aligned}$$

Therefore, v_a has a density supported on \mathbb{R}_+ : $\rho_{v_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{1}{2t}(\mu t - a)^2}$, $t > 0$.

6 Markov Processes

6.1 Transition Semigroups and Feller Semigroups

Definition 6.1 (Markovian transition kernels). Let (E, \mathcal{E}) be a measurable space. A *Markovian transition kernel* (or *transition kernel*, for short) on E is a mapping $Q : E \times \mathcal{E} \rightarrow [0, 1]$ satisfying the following properties:

- (i) For every $x \in E$, the mapping $\mathcal{E} \ni A \mapsto Q(x, A)$ is a probability measure on (E, \mathcal{E}) ;
- (ii) For every $A \in \mathcal{E}$, the mapping $E \ni x \mapsto Q(x, A)$ is \mathcal{E} -measurable.

Remark. (i) If E is a finite or countable set equipped with the σ -algebra \mathcal{E} of all its subsets, we can then characterize a transition kernel Q by the matrix $(Q(x, \{y\}))_{x, y \in E}$.

(ii) Let $B(E)$ be the vector space of all bounded measurable real functions on E , and we define the norm $\|f\| = \sup_{x \in E} |f(x)|$ for all $f \in B(E)$. Given a function $f \in B(E)$, we define

$$Qf : E \rightarrow \mathbb{R}, \quad Qf(x) = \int_E Q(x, dy) f(y).$$

For every $A \in \mathcal{E}$, we have $Q\mathbf{1}_A(x) = Q(x, A)$, hence the function $Q\mathbf{1}_A$ is measurable. A simple function approximation argument shows that Qf is measurable for all $f \in B(E)$. Furthermore,

$$\|Qf\| = \sup_{x \in E} \left(\int_E Q(x, dy) f(y) \right) \leq \|f\| \sup_{x \in E} \left(\int_E Q(x, dy) \right) = \|f\|.$$

Clearly, $B(E)$ is complete under the norm $\|\cdot\|$. From this perspective, we can view Q as a bounded linear operator on the Banach space $B(E)$ such that $\|Q\| \leq 1$, which is called a *contraction* on $B(E)$.

Definition 6.2 (Transition semigroups). A collection $(Q_t)_{t \geq 0}$ of transition kernels on E is said to be a *transition semigroup* on E if the following properties hold:

- (i) $Q_0(x, \cdot) = \delta_x$ for every $x \in E$.
- (ii) (Chapman-Kolmogorov identity). For every $s, t \geq 0$ and every $A \in \mathcal{E}$,

$$Q_{t+s}(x, A) = \int_E Q_t(x, dy) Q_s(y, A).$$

- (iii) For every $A \in \mathcal{E}$, the mapping $(t, x) \mapsto Q_t(x, A)$ is measurable with respect to $\mathcal{B}(\mathbb{R}_+) \times \mathcal{E}$.

Remark. If we view $(Q_t)_{t \geq 0}$ as bounded linear operators on $B(E)$, the Chapman-Kolmogorov identity implies that $Q_{t+s} = Q_t Q_s$ for all $s, t \geq 0$. This give rise to the associative property: $(Q_r Q_s) Q_t = Q_r (Q_s Q_t)$ for all $r, s, t \geq 0$. Hence $(Q_t)_{t \geq 0}$ is a semigroup of contractions on $B(E)$.

Definition 6.3 (Resolvent). Let $\lambda > 0$. The λ -*resolvent* of the transition semigroup $(Q_t)_{t \geq 0}$ is the linear operator $R_\lambda : B(E) \rightarrow B(E)$ defined by

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt, \quad \forall f \in B(E), x \in E.$$

Remark. The resolvent has the following properties:

- (i) $R_\lambda : B(E) \rightarrow B(E)$ is a positive and bounded linear operator. Note that $R_\lambda f \geq 0$ for all $f \geq 0$, and

$$\sup_{x \in E} \left(\int_0^\infty e^{-\lambda t} Q_t f(x) dt \right) \leq \int_0^\infty e^{-\lambda t} \|Q_t f\| dt \int_0^\infty e^{-\lambda t} \|f\| dt \Rightarrow \|R_\lambda f\| \leq \frac{1}{\lambda} \|f\|.$$

(ii) (Resolvent equation). For all $\lambda, \mu > 0$, we have $R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu = 0$:

$$\begin{aligned} R_\lambda(R_\mu f)(x) &= \int_0^\infty e^{-\lambda s} \left(\int_E Q_s(x, dy) \int_0^\infty e^{-\mu t} Q_t f(y) dt \right) ds \\ &= \int_0^\infty e^{-\lambda s} \left(\int_0^\infty e^{-\mu t} Q_{s+t} f(x) dt \right) ds \stackrel{r=s+t}{=} \int_0^\infty e^{-(\lambda-\mu)s} \left(\int_s^\infty e^{-\mu r} Q_r f(x) dr \right) ds \\ &= \int_0^\infty Q_r f(x) e^{-\mu r} \left(\int_0^r e^{-(\lambda-\mu)s} ds \right) dr = \int_0^\infty \left(\frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} \right) Q_r f(x) dr. \end{aligned}$$

Consequently, if we fix $f \in B(E)$, it holds

$$\|(R_\lambda - R_\mu)f\| = |\lambda - \mu| \|R_\lambda R_\mu f\| \leq \frac{|\lambda - \mu| \|f\|}{\lambda\mu}, \quad \forall \lambda, \mu \in (0, +\infty).$$

Hence $\lambda \rightarrow R_\lambda f$ is a continuous mapping from $(0, +\infty)$ into $B(E)$.

(iii) For every $\lambda > 0$ and every $n \in \mathbb{N}$, we have

$$R_\lambda^n f(x) = \int_0^\infty \frac{s^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) ds.$$

Clearly, this equality holds for $n = 1$. Hence we can prove the general case by induction:

$$\begin{aligned} R_\lambda^{n+1} f(x) &= R_\lambda(R_\lambda^n f)(x) = \int_0^\infty e^{-\lambda r} \left(\int_0^\infty \frac{t^{n-1}}{(n-1)!} e^{-\lambda t} Q_{r+t} f(x) dt \right) dr \\ &= \int_0^\infty \left(\int_r^\infty \frac{(s-r)^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) ds \right) dr \\ &= \int_0^\infty \int_0^s \frac{(s-r)^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) dr ds = \int_0^\infty \frac{s^n}{n!} e^{-\lambda s} Q_s f(x) ds. \end{aligned}$$

Preliminary: LCCB space. From now on, we deal with a special topological space E , which is *Locally Compact, Hausdorff, and has a Countable Basis* \mathcal{B} (LCCB).

Since E is locally compact, for each $x \in X$ there exists an open neighborhood U_x with compact closure. Consequently, one can find a basis set $B_x \in \mathcal{B}$ such that $x \in B_x \subset U_x$, and $\overline{B_x}$ is compact. We choose $\mathcal{B}_K \subset \mathcal{B}$ to be the collection of all basis sets with compact closure. Then $B_x \in \mathcal{B}_K$ for all $x \in X$, and $E = \bigcup_{B \in \mathcal{B}_K} \overline{B}$ is a countable union of compact sets. Therefore, E is a σ -compact topological space.

By σ -compactness of E , we choose an increasing sequence $(C_n)_{n=1}^\infty$ of compact subsets increasing to E . Then one can construct an increasing sequence $(K_n)_{n=1}^\infty$ of compact subspace of E such that

$$K_1 \subset K_2^\circ \subset K_2 \subset K_3^\circ \subset \cdots \subset K_n \subset K_{n+1}^\circ \subset K_{n+1} \subset \cdots, \quad E = \bigcup_{n=1}^\infty K_n. \quad (*)$$

We start by choosing a neighborhood U_x with compact closure for each $x \in E$ and setting $K_0 = \emptyset$. If K_{n-1} is constructed, then $K_{n-1} \cup C_n$ is compact, and there exists x_1, \dots, x_k such that $K_{n-1} \cup C_n \subset U_{x_1} \cup \cdots \cup U_{x_k}$. We construct $K_n = \overline{U_{x_1}} \cup \cdots \cup \overline{U_{x_k}}$, which is also compact. Then we have $K_{n-1} \subset K_n^\circ$, and $\bigcup_{n=1}^\infty K_n \supset \bigcup_{n=1}^\infty (K_n)^\circ \supset \bigcup_{n=1}^\infty C_n = E$. Clearly, for any compact subset K of E , $\{(K_n)^\circ\}_{n=1}^\infty$ is an open cover of K , hence there exists K_n such that $K \subset K_n$.

Preliminary: Continuous functions vanishing at infinity. Let E be a locally compact Hausdorff space. A continuous function $f : E \rightarrow \mathbb{R}$ is said to be *vanishing at infinity*, if for all $\epsilon > 0$,

$$\exists \text{ compact } K \subset E \text{ such that } |f(x)| < \epsilon, \forall x \in E \setminus K \quad \Leftrightarrow \quad \{x \in E : |f(x)| \geq \epsilon\} \text{ is compact.}$$

In addition, If $(K_n)_{n=1}^\infty$ is a sequence of compact sets specified in (*), then we have

$$\lim_{n \rightarrow \infty} \sup_{x \in E \setminus K_n} |f(x)| = 0.$$

We denote by $C_0(E)$ the vector space of all continuous real-valued functions on E vanishing at infinity. This is a closed subspace of $B(E)$. Consequently, $C_0(E)$ is a Banach space given the supremum norm.

In addition, since E is a locally compact Hausdorff space, it admits the Alexandroff compactification $E^* = E \cup \{\infty\}$, whose topology consists of all open sets in E and all sets of the form $(E \setminus K) \cup \{\infty\}$, where K is a compact subset of E . Consequently, for every $f \in C_0(E)$, we can extend it to a function of $C_b(E^*)$ by setting $f(\infty) = 0$. Conversely, for every $f \in C_b(E^*)$, we have $f|_E - f(\infty) \in C_0(E)$.

Riesz-Markov Theorem. If E is a locally compact Hausdorff space and $L : C_0(E) \rightarrow \mathbb{R}$ is a bounded linear functional, then there exists a unique regular (finite signed) measure μ such that $Lf = \int f d\mu$ for all $f \in C_0(E)$. Furthermore, $\|L\| = |\mu|(E)$.

Using $C_0(E)$ to separate points of a LCCB space E . If X is a second-countable normal space, we choose a basis $\mathcal{B} = \{B_n\}_{n=1}^\infty$ of X . We define $I_c := \{(m, n) \in \mathbb{N}^2 : \overline{B}_m \subsetneq B_n\}$, and for each $(m, n) \in I_c$, by Urysohn's lemma, we can find a continuous function $f_{m,n} : X \rightarrow [0, 1]$ such that $f_{m,n}(\overline{B}_m) = \{1\}$ and $f_{m,n}(X \setminus B_n) = \{0\}$. Since for every pair of distinct points $x_1, x_2 \in X$, there exists disjoint neighborhoods $B_{n_1} \ni x_1$ and $B_{n_2} \ni x_2$. Therefore, $\mathcal{F} = \{f_{m,n} : (m, n) \in I_c\} \subset C(X)$ is a countable collection of functions separating points of X , i.e. for all $x_1, x_2 \in E$ with $x_1 \neq x_2$, there exists $f \in \mathcal{F}$ such that $f(x_1) \neq f(x_2)$.

Now we consider a LCCB space E . If $E^* = E \cup \{\infty\}$ is the Alexandroff compactification of E , then $U_n := E^* \setminus K_n$ is a countable local base of ∞ , where $\{K_n\}_{n=1}^\infty$ is specified in (*), because for any neighborhood V of ∞ in E^* , $\{K_n^c\}_{n=1}^\infty$ is an open cover of the compact set $E^* \setminus K$. Consequently, we can construct a countable collection $\mathcal{F} \subset C(E^*)$ of functions separating points E^* , and we may set these functions to 0 at ∞ . By restricting these function on E , we obtain a countable collection of functions in $C_0(E)$ separating points of E . We again use this conclusion when finding the càdlàg version of a Feller process.

Vague convergence. We equip a metrizable space E with its Borel σ -algebra \mathcal{E} . Let μ_n be a sequence of probability measures on (E, \mathcal{E}) . If there exists a probability measure μ such that

$$\int_E f d\mu_n \rightarrow \int_E f d\mu, \quad \forall f \in C_c(E),$$

then μ_n is said to *vaguely converges to μ* . Clearly, weak convergence implies vague convergence.

In a LCCB space E , we can prove that weak convergence is equivalent vague convergence. Let μ_n be a sequence of probability measures converging vaguely to μ , and fix $\epsilon > 0$. For any $g \in C_b(E)$ with $\|g\| \leq M$, we let $\phi \in C_c(E)$ be a function supported on K such that $\int_E \phi d\mu > 1 - \frac{\epsilon}{3M}$. By vague convergence, there exists N_1 such that $\int_E \phi d\mu_n > 1 - \frac{\epsilon}{3M}$ for all $n \geq N_1$. Since $g\phi \in C_c(E)$, we also choose N_2 such that $|\int_E g\phi d\mu_n - \int_E g\phi d\mu| < \epsilon/3$ for all $n \geq N_2$. Then

$$\begin{aligned} \left| \int_E g d\mu_n - \int_E g d\mu \right| &= \left| \int_E g(1 - \phi) d\mu_n \right| + \left| \int_E g\phi d\mu_n - \int_E g\phi d\mu \right| + \left| \int_E g(1 - \phi) d\mu \right| \\ &\leq 2M \left| \int_E (1 - \phi) d\mu \right| + \left| \int_E g\phi d\mu_n - \int_E g\phi d\mu \right| < \epsilon, \quad \forall n \geq \max\{N_1, N_2\}. \end{aligned}$$

Since ϵ is arbitrarily small, we have $\int_E g d\mu_n \rightarrow \int_E g d\mu$, and the weak convergence is clear.

Definition 6.4 (Feller semigroups). A transition semigroup $(Q_t)_{t \geq 0}$ is said to be a *Feller semigroup* if

- (i) For every $t \geq 0$, we have $Q_t C_0(E) \subset C_0(E)$.
- (ii) For every $f \in C_0(E)$ and every $x \in E$, $\lim_{t \downarrow 0} Q_t f(x) = f(x)$.

Remark. Consider the λ -resolvent. For any sequence $x_n \in E$, by dominated convergence theorem,

$$\lim_{n \rightarrow \infty} R_\lambda f(x_n) = \lim_{n \rightarrow \infty} \int e^{-\lambda s} Q_s f(x_n) ds = \int e^{-\lambda s} \lim_{n \rightarrow \infty} Q_s f(x_n) ds, \quad \forall f \in C_0(E).$$

In the last equality, we use the fact $\sup_{t \geq 0} |e^{-\lambda s} Q_s f(x_n)| \leq \|f\|$ for all $n \in \mathbb{N}$. Consequently, if $(Q_t)_{t \geq 0}$ is a Feller semigroup, we also have $R_\lambda C_0(E) \subset C_0(E)$ for all $\lambda > 0$.

Proposition 6.5. Let $(Q_t)_{t \geq 0}$ be a Feller semigroup, and let R_λ be its λ -resolvent, where $\lambda > 0$. Define $\mathfrak{R} = \{R_\lambda f : f \in C_0(E)\}$. Then \mathfrak{R} does not depend on the choice of λ , and \mathfrak{R} is a dense subspace of $C_0(E)$.

Proof. For any $\mu \neq \lambda$, the resolvent equation gives $R_\lambda f = R_\mu g$, where $g = f - (\lambda - \mu)R_\lambda f \in C_0(E)$. Hence \mathfrak{R} does not depend on the choice of λ . For the second assertion, note that

$$\lim_{\lambda \rightarrow \infty} \lambda R_\lambda f(x) = \lim_{\lambda \rightarrow \infty} \lambda \int_0^\infty e^{-\lambda s} Q_s f(x) ds = \lim_{\lambda \rightarrow \infty} \int_0^\infty e^{-t} Q_{t/\lambda} f(x) dt = \int_0^\infty e^{-t} f(x) dt = f(x),$$

where the last equality holds by dominated convergence, since $\sup_{s \geq 0} |e^{-s} Q_{s/\lambda} f(x)| \leq \|f\|$ for all $\lambda > 0$. Furthermore, for all $\lambda, \mu > 0$ and all $x \in E$, we have

$$\begin{aligned} (\lambda R_\lambda - \mu R_\mu) f(x) &= \int_0^\infty (\lambda e^{-\lambda s} - \mu e^{-\mu s}) Q_s f(x) ds = \int_0^\infty e^{-t} (Q_{t/\lambda} - Q_{t/\mu}) f(x) dt \\ &= \int_0^\infty e^{-t} \left(\frac{t}{\mu} - \frac{t}{\lambda} \right) Q_{t/\mu} Q_{t/\lambda} f(x) dt \leq \left| \frac{1}{\lambda} - \frac{1}{\mu} \right| \|f\|. \end{aligned}$$

Hence $\|(\lambda R_\lambda - \mu R_\mu) f\| \rightarrow 0$ as $\lambda, \mu \rightarrow \infty$, and $\lambda R_\lambda f \in \mathfrak{R}$ converges in $C_0(E)$ as $\lambda \rightarrow \infty$ by completeness, and the pointwise limit f must be the limit $C_0(E)$. Consequently, \mathfrak{R} is dense in $C_0(E)$. \square

Remark. In the proof, we also conclude that $\lim_{\lambda \rightarrow \infty} \|\lambda R_\lambda f - f\| = 0$ for all $f \in C_0(E)$.

Proposition 6.6 (Strong continuity). Let $(Q_t)_{t \geq 0}$ be a Feller semigroup, and fix $f \in C_0(E)$. Then $\lim_{t \downarrow 0} \|Q_t f - f\| = 0$. Consequently, the mapping $t \mapsto Q_t f$ is uniformly continuous from $(0, \infty)$ into $C_0(E)$.

Proof. By Fubini's theorem, since $(s, y) \mapsto e^{-\lambda s} Q_s f(y)$ is dominated by $e^{-\lambda s} \|f\|$, it holds

$$Q_t R_\lambda f(x) = e^{\lambda t} \int_t^\infty e^{-\lambda s} Q_s f(x) ds \leq e^{\lambda t} R_\lambda f - \int_0^t e^{\lambda(t-s)} Q_s f(x) ds$$

Consequently, we have

$$\|Q_t R_\lambda f - R_\lambda f\| = (e^{\lambda t} - 1) \|R_\lambda f\| + t e^{\lambda t} \|f\| \rightarrow 0 \quad \text{as } t \downarrow 0.$$

Therefore we have $\lim_{t \downarrow 0} \|Q_t f - f\| = 0$ for all $f \in \mathfrak{R}$. The continuity of $Q_t - \text{Id}$ and a standard density argument extend this conclusion to all $f \in C_0(E)$. For the second assertion, note that for all $t > s \geq 0$,

$$\|Q_t f - Q_s f\| = \|Q_s(Q_{t-s} f - f)\| = \|Q_{t-s} f - f\| \rightarrow 0 \quad \text{as } t - s \downarrow 0.$$

Since this convergence is uniform for all $s \in (0, \infty)$, the mapping $t \mapsto Q_t f$ is uniformly continuous. \square

Definition 6.7 (Infinitesimal generator). Let $(Q_t)_{t \geq 0}$ be a Feller semigroup. Define the space $\mathfrak{D}(L)$ by

$$\mathfrak{D}(L) = \left\{ f \in C_0(E) : \frac{Q_t f - f}{t} \text{ converges in } C_0(E) \text{ when } t \downarrow 0 \right\}.$$

Then $\mathfrak{D}(L)$ is a subspace of the vector space $C_0(E)$. Define the linear operator $L : \mathfrak{D}(L) \rightarrow C_0(E)$ as follows:

$$L f = \lim_{t \downarrow 0} \frac{Q_t f - f}{t}, \quad \forall f \in \mathfrak{D}(L).$$

The operator $L : \mathfrak{D}(L) \rightarrow C_0(E)$ is called the *infinitesimal generator* (or *generator*, for short) of $(Q_t)_{t \geq 0}$.

Proposition 6.8. Let L be the generator of Feller semigroup $(Q_t)_{t \geq 0}$.

(i) For every $f \in \mathfrak{D}(L)$ and every $s > 0$, $Q_s f \in \mathfrak{D}(L)$, and $L(Q_s f) = Q_s(Lf)$. Furthermore,

$$Q_t f = f + \int_0^t Q_s L f \, ds = f + \int_0^t L Q_s f \, ds \quad \Leftrightarrow \quad \frac{d}{dt} Q_t f = L Q_t f. \quad (6.1)$$

(ii) $\mathfrak{D}(L) = \mathfrak{R}$, and R_λ is the inverse of $\lambda \text{Id} - L$ for all $\lambda > 0$, namely, $(\lambda \text{Id} - L)R_\lambda = R_\lambda(\lambda \text{Id} - L) = \text{Id}$.

(iii) The semigroup $(Q_t)_{t \geq 0}$ is determined by the generator L :

$$Q_t = e^{tL} := \lim_{\lambda \rightarrow \infty} e^{-t\lambda} \sum_{k=0}^{\infty} \frac{t^k \lambda^{2k}}{k!} (\lambda \text{Id} - L)^{-k}$$

Proof. (i) For all $s \geq 0$, Q_s is a bounded linear operator on $C_0(E)$, and $Q_s f \in C_0(E)$ for all $f \in D(E) \subset C_0(E)$. Semigroup property and continuity of Q_s implies

$$\frac{Q_t(Q_s f) - Q_s f}{t} = Q_s \left(\frac{Q_t f - f}{t} \right) \Rightarrow Q_s f \in \mathfrak{D}(L), \quad L(Q_s f) = Q_s(Lf).$$

Similarly, we have $h^{-1}(Q_{t+h}f - Q_t f) = Q_t(Lf)$ when $h \downarrow 0$. Moreover,

$$\begin{aligned} \lim_{h \downarrow 0} \left\| \frac{Q_t f - Q_{t-h} f}{h} - Q_t(Lf) \right\| &= \lim_{h \downarrow 0} \left\| Q_{t-h} \left(\frac{Q_h f - f}{h} \right) - Q_t(Lf) \right\| \\ &\leq \lim_{h \downarrow 0} \left\| Q_{t-h} \left(\frac{Q_h f - f}{h} - Lf \right) \right\| + \lim_{h \downarrow 0} \|(Q_{t-h} - Q_t)(Lf)\| \\ &\leq \lim_{h \downarrow 0} \left\| \frac{Q_h f - f}{h} - Lf \right\| + \lim_{h \downarrow 0} \|Lf - Q_h(Lf)\| = 0. \end{aligned}$$

Consequently, for every $x \in E$, the mapping $t \mapsto Q_t f(x)$ is differentiable. By fundamental theorem of calculus,

$$Q_t f - f = \int_0^t Q_s(Lf) \, ds = \int_0^t L(Q_s f) \, ds, \quad \forall t \in \mathbb{R}_+.$$

(ii) If $f \in \mathfrak{D}(L)$, we use (6.1) and Fubini's theorem:

$$\begin{aligned} R_\lambda(\lambda \text{Id} - L)f &= \int_0^\infty \lambda e^{-\lambda s} Q_s f \, ds - \int_0^\infty e^{-\lambda s} Q_s(Lf) \, ds \\ &= \int_0^\infty \lambda e^{-\lambda s} \left(f + \int_0^s L(Q_t f) \, dt \right) \, ds - \int_0^\infty e^{-\lambda s} L(Q_s f) \, ds \\ &= f + \int_0^\infty \left(\int_t^\infty \lambda e^{-\lambda s} L(Q_t f) \, ds \right) \, dt - \int_0^\infty e^{-\lambda s} L(Q_s f) \, ds = f. \end{aligned}$$

Apparently, we have $\mathfrak{D}(L) \subset \mathfrak{R}$. On the other hand, for $g \in C_0(E)$, by dominated convergence theorem,

$$\begin{aligned} \lim_{h \downarrow 0} \frac{Q_h R_\lambda g - R_\lambda g}{h} &= \lim_{h \downarrow 0} \frac{1}{h} \int_0^\infty e^{-\lambda t} (Q_{t+h} - Q_t) g \, dt \\ &= \lim_{h \downarrow 0} \frac{1}{h} \left((1 - e^{-\lambda h}) \int_0^\infty e^{-\lambda t} Q_{t+h} g \, dt - \int_0^h e^{-\lambda t} Q_t g \, dt \right) = \lambda R_\lambda g - g. \end{aligned}$$

Hence $R_\lambda g \in \mathfrak{D}(L)$, and $\mathfrak{R} = \mathfrak{D}(L)$. Moreover, the first assertion in (i) implies

$$L R_\lambda g = \lambda R_\lambda g - g \quad \Rightarrow \quad (\lambda \text{Id} - L) R_\lambda g = g.$$

(iii) Fix $f \in C_0(E)$. For every $\lambda > 0$, if $(\lambda \text{Id} - L)g = f$ for some $g \in \mathfrak{D}(L)$, then $g = R_\lambda(\lambda \text{Id} - L)g = R_\lambda f$. Hence the resolvent R_λ is uniquely determined by inverting the operator $\lambda \text{Id} - L$, which is defined on $C_0(E)$.

For every $\lambda > 0$, we define $A_\lambda := \lambda R_\lambda L = \lambda^2 R_\lambda - \lambda \text{Id}$. Then $A_\lambda : \mathfrak{D}(L) \rightarrow C_0(E)$ is a bounded linear operator, and $\|A_\lambda\| \leq 2\lambda$. Then we define the following series, which converges in norm:

$$e^{tA_\lambda} = \sum_{k=0}^{\infty} \frac{(tA_\lambda)^k}{k!} \quad \Rightarrow \quad \|e^{tA_\lambda}\| = \left\| e^{t\lambda^2 R_\lambda - \lambda \text{Id}} \right\| \leq \sum_{k=0}^{\infty} e^{-t\lambda} \frac{\|t\lambda^2 R_\lambda\|^k}{k!} \leq \sum_{k=0}^{\infty} e^{-t\lambda} \frac{(t\lambda)^k}{k!} = 1.$$

By commutativity principle (i.e. $e^{T+S} = e^T e^S$ for commutative bounded linear operators $ST = TS$), the collection $(e^{tA_\lambda})_{t \geq 0}$ is a semigroup of contractions. Moreover, since $A_\lambda A_\mu = A_\mu A_\lambda = \frac{\mu - \lambda}{\mu \lambda} L(R_\lambda - R_\mu)L$,

$$e^{tA_\lambda} - e^{tA_\mu} = \int_0^t \frac{d}{ds} e^{sA_\mu + (t-s)A_\lambda} \, ds = \int_0^t e^{(t-s)A_\mu} e^{sA_\lambda} (A_\lambda - A_\mu) \, ds.$$

Hence for all $f \in \mathfrak{D}(L)$, since $A_\lambda f \rightarrow Lf$ as $\lambda \rightarrow \infty$,

$$\|(e^{tA_\lambda} - e^{tA_\mu})f\| \leq \|(A_\lambda - A_\mu)f\| \quad \Rightarrow \quad e^{tL}f := \lim_{\lambda \rightarrow \infty} e^{tA_\lambda}f \text{ exists in } C_0(E).$$

Clearly, the mapping $e^{tL} : \mathfrak{D}(L) \rightarrow C_0(E)$ is a contraction. Since $\mathfrak{D}(L)$ is dense in $C_0(E)$, we can extend the definition of e^{tL} from $\mathfrak{D}(L)$ to $C_0(E)$. The following shows that $(e^{tL})_{t \geq 0}$ is a semigroup: $\forall f \in \mathfrak{D}(L)$,

$$\begin{aligned} \|e^{(t+s)L}f - e^{tL}e^{sL}f\| &= \|(e^{(t+s)L} - e^{(t+s)A_\lambda})f\| + \|e^{tA_\lambda}(e^{sA_\lambda} - e^{sL})f\| + \|(e^{tA_\lambda} - e^{tL})e^{sL}f\| \\ &\leq \|(e^{(t+s)L} - e^{(t+s)A_\lambda})f\| + \|(e^{sA_\lambda} - e^{sL})f\| + \|(e^{tA_\lambda} - e^{tL})e^{sL}f\| \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

We also note strong continuity so that $(e^{tL})_{t \geq 0}$ is a Feller semigroup: $\forall f \in \mathfrak{D}(L)$,

$$\|e^{tL}f - f\| \leq \|(e^{tL} - e^{tA_\lambda})f\| + \|e^{tA_\lambda}f - f\| \leq t\|Lf - A_\lambda f\| + \|e^{tA_\lambda}f - f\| \rightarrow 0 \quad \text{as } t \downarrow 0.$$

Furthermore, the generator of $(e^{tL})_{t \geq 0}$ is L :

$$e^{tA_\lambda}f - f = \int_0^t A_\lambda e^{sA_\lambda}f \, ds \quad \Rightarrow \quad e^{tL}f - f = \int_0^t L e^{sL}f \, ds \quad \Rightarrow \quad \lim_{t \downarrow 0} \frac{e^{tL}f - f}{t} = L.$$

Since the resolvent $R_\lambda f = \int_0^\infty e^{-\lambda t} Q_t f \, dt$ is the Laplacian transform, it has a one-to-one correspondence with the transition semigroup $(Q_t)_{t \geq 0}$. Hence we can recover $Q_t = e^{tL}$ uniquely from the generator L . \square

Remark. Note that the domain of operator $(\lambda \text{Id} - L)R_\lambda$ is $C_0(E)$.

6.2 Markov Processes and Feller Processes

Now we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t \geq 0}$.

Definition 6.9 (Markov processes and Feller processes). Let $(Q_t)_{t \geq 0}$ be a transition semigroup on E . A *Markov process* with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ with transition semigroup $(Q_t)_{t \geq 0}$ is an adapted process $X = (X_t)_{t \geq 0}$ with values in E such that for all $f \in E$ and all $s, t \geq 0$,

$$\mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] = Q_t f(X_s). \quad (6.2)$$

Without specifying a filtration, we implicitly mean that the definition holds for the canonical filtration $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$. Clearly, a Markov process X with respect to any filtration $(\mathcal{F}_t)_{t \geq 0}$ is also a Markov process with respect to the canonical filtration $(\mathcal{F}_t^X)_{t \geq 0}$.

A Markov process with values in E is called a *Feller process* if its transition semigroup is Feller.

Remark. Particularly, for all $A \in \mathcal{E}$, one can take $f = \mathbb{1}_A$ in (6.1) to obtain

$$\mathbb{P}(X_{s+t} \in A | \mathcal{F}_s) = Q_t(X_s, A).$$

It is seen that the conditional distribution of X_{s+t} given the history \mathcal{F}_s only depends on the current state X_s . Furthermore, given the law γ of X_0 , we have for all $0 < t_1 < \dots < t_n$ and all $A_0, A_1, \dots, A_n \in \mathcal{E}$ that

$$\begin{aligned} & \mathbb{P}(X_0 \in A_0, X_{t_1} \in A_1, X_{t_2} \in A_2, \dots, X_{t_n} \in A_n) \\ &= \int_{A_0} \gamma(dx_0) \int_{A_1} Q_{t_1}(x_0, dx_1) \int_{A_1} Q_{t_2-t_1}(x_1, dx_2) \dots \int_{A_n} Q_{t_n-t_{n-1}}(x_{n-1}, dx_n). \end{aligned} \quad (6.3)$$

More generally, if $f_0, f_1, \dots, f_n \in B(E)$, we have

$$\begin{aligned} & \mathbb{E}[f_0(X_0)f_1(X_{t_1}) \dots f_n(X_{t_n})] \\ &= \int_E \gamma(dx_0) f_0(x_0) \int_E Q_{t_1}(x_0, dx_1) f_1(x_1) \int_E Q_{t_2-t_1}(x_1, dx_2) f_2(x_2) \dots \int_E Q_{t_n-t_{n-1}}(x_{n-1}, dx_n) f_n(x_n). \end{aligned}$$

Now we discuss the existence of a Markov process with transition semigroup $(Q_t)_{t \geq 0}$. According to (6.2), with the initial γ given, we obtain a pre-measure $P_{t_1, \dots, t_n}^\gamma = \mathbb{P}(X_0 \in \cdot, X_{t_1} \in \cdot, \dots, X_{t_n} \in \cdot)$ on all measurable rectangles \mathcal{E}^n , which can be uniquely extended to a measure on the product space $(E^n, \mathcal{E}^{\otimes n})$. In addition, the collection of all finite marginals $\{P_{t_1, \dots, t_n}^\gamma : n \in \mathbb{N}, 0 < t_1 < \dots < t_n\}$ satisfies the compatibility condition given in Proposition 4.11, according to the Chapman-Kolmogorov identity.

We further assume that E is a Polish space. Then according to Corollary 4.15, which is a consequence of the Daniell-Kolmogorov extension theorem, the compatible family $\{P_{t_1, \dots, t_n}^\gamma : n \in \mathbb{N}, 0 < t_1 < \dots < t_n\}$ of probability measures has a unique extension P^γ on the canonical space $(E^{\mathbb{R}_+}, \mathcal{E}^{\otimes \mathbb{R}_+})$. Consequently, the canonical process $\{\pi_t\}_{t \geq 0}$ on $(E^{\mathbb{R}_+}, \mathcal{E}^{\otimes \mathbb{R}_+})$ is a Markov process under P^γ with transition semigroup $(Q_t)_{t \geq 0}$ with respect to the canonical filtration, and the law of π_0 is given by γ .

To summarize, if E is a Polish space, we can construct a E -valued Markov process $(X_t)_{t \geq 0}$ with transition semigroup $(Q_t)_{t \geq 0}$ under any given initial distribution γ .

Alternative definition of Feller Semigroup. Let $(Q_t)_{t \geq 0}$ be a transition group, and for every $x \in E$, let $(X_t^x)_{t \geq 0}$ be a Markov process with semigroup $(Q_t)_{t \geq 0}$ starting from $X_0 \sim \delta_x$. Then $(X_t^x)_{t \geq 0, x \in E}$ is a *Markov family*, and the law of every process $(X_t^x)_{t \geq 0}$ is given by $P^x := P^{\delta_x}$.

Clearly, if $(Q_t)_{t \geq 0}$ is a Feller semigroup, then every process $(X_t^x)_{t \geq 0}$ is a Feller process. In fact, we can characterize a Feller semigroup by the following properties of the law of Markov families $(X_t^x)_{t \geq 0, x \in E}$. Following our discussion of *vague convergence*, one can easily show that

$$\begin{aligned} \forall x \in E \text{ and } \forall t \geq 0, \quad X_t^y \xrightarrow{d} X_t^x \text{ as } y \rightarrow x &\Leftrightarrow Q_t C_0(E) \subset C_0(E); \\ \forall x \in E, \quad X_t^x \xrightarrow{\mathbb{P}} x \text{ as } t \downarrow 0 &\Leftrightarrow Q_t f(x) \rightarrow f(x), \quad \forall f \in C_0(E). \end{aligned}$$

Now we study the sample path property of Feller processes. Recall that we assume E to be a LCCB space.

Proposition 6.10. If $(X_t)_{t \geq 0}$ is a Markov process with transition semigroup $(Q_t)_{t \geq 0}$, and $h \in B(E)$ is a nonnegative function, then the process $(e^{-\lambda t} R_\lambda h(X_t))_{t \geq 0}$ is a supermartingale.

Proof. Clearly $e^{-\lambda t} R_\lambda h(X_t)$ is bounded, hence in L^1 . For every $s, t \geq 0$, we have

$$Q_s R_\lambda h = \int_0^\infty e^{-\lambda t} Q_{t+s} h \, dt = e^{\lambda s} \int_0^\infty e^{-\lambda(s+t)} Q_{t+s} h \, dt \leq e^{\lambda s} R_\lambda h.$$

By (6.2), the following inequality holds for $(X_t)_{t \geq 0}$:

$$\mathbb{E}[e^{-\lambda(t+s)} R_\lambda h(X_{t+s}) | \mathcal{F}_t] = e^{-\lambda(t+s)} Q_s R_\lambda h(X_t) \leq e^{\lambda t} R_\lambda h(X_t)$$

Therefore $(e^{-\lambda t} R_\lambda h(X_t))_{t \geq 0}$ is a martingale. □

We first consider the case where E is a compact space.

Lemma 6.11. Let E be a compact Hausdorff space and have countable basis. Let $(X_t)_{t \geq 0}$ be a Feller process with transition semigroup $(Q_t)_{t \geq 0}$, with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Then the process $(X_t)_{t \geq 0}$ has a càdlàg modification.

Proof. Let $\mathcal{H} \in C(E)$ be a countable collection of functions separating points of E . We first show that a sequence $x_n \in E$ converges if $h(x_n)$ converges for all $h \in \mathcal{H}$. By compactness of E , every sequence of points of E has at least one limit point, and a sequence converges if and only if the limit point is unique. If $x, y \in E$ are both limit points of x_n , then $h(x) = \lim_{n \rightarrow \infty} h(x_n) = h(y)$ for all $h \in \mathcal{H}$, and $x = y$ by definition of \mathcal{H} . Consequently, x_n converges to a unique limit point.

We take a sequence $f_n \in C_0^+(E)$ separating the points of E , and take

$$\mathcal{H} = \{R_p f_n : p \in \mathbb{N}, n \in \mathbb{N}\}.$$

This is also a countable subset of $C(E)$ that separates the points of E , since $\|p R_p f - f\| \rightarrow 0$ as $p \rightarrow \infty$.

Let D be a countable dense subset of \mathbb{R}_+ . By Proposition 6.10, if $h \in \mathcal{H}$, there exists $p \in \mathbb{N}$ such that $(e^{-pt} h(X_t))_{t \geq 0}$ is a supermartingale. By Theorem 3.43 (i), the left limit $\lim_{D \ni s \downarrow t} h(X_s)$ [resp. the right limit $\lim_{D \ni s \uparrow t} h(X_s)$] exists for all $t \in \mathbb{R}_{++}$ [resp. $t \in \mathbb{R}_+$] except on an event N_h of probability zero. We take $N = \bigcup_{h \in \mathcal{H}} N_h$, hence $(h(X_t))_{t \in D}$ has side limits on $\Omega \setminus N$. Then we define

$$\begin{cases} \tilde{X}_t(\omega) = \lim_{D \ni s \downarrow t} X_s(\omega), & \omega \in \Omega \setminus N \\ \tilde{X}_t(\omega) = x_0, & \omega \in N \end{cases}$$

where $x_0 \in E$ is a fixed point. Clearly this is a càdlàg process.

It remains to show that $(\tilde{X}_t)_{t \geq 0}$ is a modification of $(X_t)_{t \geq 0}$. For any $t \geq 0$, take $D \ni t_n \downarrow t$. Then for all $f, g \in C(E)$, we have

$$\mathbb{E}[f(X_t)g(\tilde{X}_t)] = \lim_{n \rightarrow \infty} \mathbb{E}[f(X_t)g(X_{t_n})] = \lim_{n \rightarrow \infty} \mathbb{E}[f(X_t)Q_{t_n-t}g(X_t)] = \mathbb{E}[f(X_t)g(X_t)].$$

By functional monotone class theorem, $\mathbb{E}[\varphi(X_t, \tilde{X}_t)] = \mathbb{E}[\varphi(X_t, X_t)]$ for all bounded Borel function on $E \times E$. We take $\varphi(x, y) = \mathbb{1}_{\{x=y\}}$, which gives $\mathbb{P}(X_t = \tilde{X}_t) = 1$. Hence $(\tilde{X}_t)_{t \geq 0}$ is a modification of $(X_t)_{t \geq 0}$. \square

Theorem 6.12 (Regularity of sample paths). Let E be a LCCB space. Let $(X_t)_{t \geq 0}$ be a Feller process with transition semigroup $(Q_t)_{t \geq 0}$, with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Set $\widetilde{\mathcal{F}}_\infty = \mathcal{F}_\infty$, and denote by \mathcal{N} the class of all zero probability sets in $\widetilde{\mathcal{F}}_\infty$. Define the modified filtration: $\widetilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+} \cup \sigma(\mathcal{N}))$, $\forall t \geq 0$.

The process $(X_t)_{t \geq 0}$ has a càdlàg modification $(\tilde{X}_t)_{t \geq 0}$, such that $(\tilde{X}_t)_{t \geq 0}$ is a Feller process with respect to the modified filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$.

Proof. Let $E^* = E \cup \{\infty\}$ be the Alexandroff compactification of E . According to Lemma 6.11, and note that $N_h \in \mathcal{F}_\infty$ for all $h \in \mathcal{H}$, we can find a càdlàg modification $(\tilde{X}_t)_{t \geq 0}$ taking values in E^* and adapted to the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$. We also point out that the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$ is right continuous, so the stopping time we are about to define makes sense.

It is necessary to show $(\tilde{X}_t)_{t \geq 0}$ is also càdlàg as a process taking values in E . We take a strictly positive $g \in C_0(E)$, then the function $h = R_1 g \in C_0(E)$ is also strictly positive, and $(e^{-t}h(\tilde{X}_t))_{t \geq 0}$ is a nonnegative and càdlàg supermartingale with respect to the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$. We set

$$\tau_n = \inf \left\{ t \geq 0 : e^{-t}h(\tilde{X}_t) < \frac{1}{n} \right\}, \quad \text{and} \quad \tau = \lim_{n \rightarrow \infty} \tau_n,$$

which are stopping times with respect to the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$ by Proposition 3.13. By optional sampling theorem [Theorem 3.48], we have

$$\mathbb{E} \left[e^{-t}h(\tilde{X}_t) \right] \leq \mathbb{E} \left[e^{-\tau_n \wedge t}h(\tilde{X}_{\tau_n \wedge t}) \right] \Rightarrow \mathbb{E} \left[\mathbb{1}_{\{\tau_n \leq t\}} e^{-t}h(\tilde{X}_t) \right] \leq \mathbb{E} \left[\mathbb{1}_{\{\tau_n \leq t\}} e^{-\tau_n}h(\tilde{X}_{\tau_n}) \right] \leq \frac{1}{n}.$$

Letting n increase to ∞ , we have

$$\mathbb{E} \left[\mathbb{1}_{\{\tau \leq t\}} e^{-t}h(\tilde{X}_t) \right] \leq 0.$$

Since $\mathbb{P}(\tilde{X}_t = \infty) = 0$, and since h is strictly positive, we have $\tau > t$ a.s.. Note that $t > 0$ is arbitrary. Then $\tau_n \rightarrow \infty$ a.s., and $\inf_{s \in [0, t]} e^{-s}h(\tilde{X}_s) > 0$ a.s. for all $t > 0$. As a result, almost surely, we have $\tilde{X}_{s-} \neq \infty$ and $\tilde{X}_s \neq \infty$ for all $s > 0$. This extends càdlàg property to E .

Finally we verify that $(\tilde{X}_t)_{t \geq 0}$ is a Markov process with semigroup $(Q_t)_{t \geq 0}$ with respect to the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$. It suffices to prove that, for all $s \geq 0, t > 0$ and $A \in \widetilde{\mathcal{F}}_s, f \in C_0(E)$, we have

$$\mathbb{E} \left[\mathbb{1}_A f(\tilde{X}_{s+t}) \right] = \mathbb{E} \left[\mathbb{1}_A Q_t f(\tilde{X}_s) \right]. \quad (6.4)$$

We may assume $A \in \mathcal{F}_{s+}$ since it a.s. equals to some \mathcal{F}_{s+} -set. Taking $D \ni s_n \downarrow s$, we have

$$\mathbb{E} \left[\mathbb{1}_A f(X_{s+t}) \right] = \mathbb{E} \left[\mathbb{E} \left[\mathbb{1}_A f(X_{s+t}) \mid \mathcal{F}_{s_n} \right] \right] = \mathbb{E} \left[\mathbb{1}_A Q_{t+s-s_n} f(X_{s_n}) \right]. \quad (6.5)$$

Since $Q_{t+s-s_n} f$ converges uniformly to $Q_t f$, and $X_{s_n} \xrightarrow{a.s.} \tilde{X}_s \xrightarrow{a.s.} \tilde{X}_s \xrightarrow{a.s.} X_s$, setting $n \rightarrow \infty$ in (6.5) gives $\mathbb{E}[\mathbb{1}_A f(X_{s+t})] = \mathbb{E}[\mathbb{1}_A Q_t f(X_s)]$. As $(\tilde{X}_t)_{t \geq 0}$ is a modification of $(X_t)_{t \geq 0}$, this is equivalent to (6.4). \square

Remark. In fact, we prove the existence of a Feller process with càdlàg sample paths in this theorem. Given a locally compact Polish space E , a starting law γ and a Feller semigroup $(Q_t)_{t \geq 0}$, we can construct a canonical process (X_t^γ) on the space $E^{\mathbb{R}_+}$, which is a càdlàg Feller process with semigroup $(Q_t)_{t \geq 0}$, according to the Remark under Definition 6.9 and Theorem 6.12. Meanwhile, this construction, using Kolmogorov extension, gives a probability measure P^γ on the measurable space $(\mathbb{D}(E), \mathcal{D})$, where $\mathbb{D}(E)$ is the space of càdlàg paths $\mathbb{R}_+ \rightarrow E$, and \mathcal{D} is the σ -algebra on $\mathbb{D}(E)$ generated by all coordinate maps.

Notations. In later discussions, we often make use of this càdlàg property of Feller processes, which, as is indicated by this theorem, is not a harmful assumption. For every $x \in E$, we use P^x to denote the probability measure on $\mathbb{D}(E)$ which is the law of Feller process $(X_t^x)_{t \geq 0}$ starting from $X_0 \sim \delta_x$. Moreover, we use \mathbb{E}_x to denote the expectation taken with respect to P^x .

If $\Phi : \mathbb{D}(E) \rightarrow \mathbb{R}_+$ is a measurable map, the mapping $x \mapsto \mathbb{E}_x[\Phi]$ is also measurable. To see this, it suffices to consider the case $\Phi = \mathbb{1}_A$, where $A \in \mathcal{D}$. If A depends on only finitely many coordinate maps:

$$A = \{f \in \mathbb{D}(E) : f(t_1) \in B_1, \dots, f(t_p) \in B_{t_p}\}, \quad \text{where } 0 \leq t_1 < \dots < t_p, \quad \text{and } B_1, \dots, B_{t_p} \in \mathcal{E}, \quad (6.6)$$

then the mapping $x \mapsto \mathbb{E}_x[\mathbb{1}_A]$ has an explicit form:

$$\mathbb{E}_x[\mathbb{1}_A] = \int_{B_1} Q_{t_1}(x, dy_1) \int_{B_2} Q_{t_2-t_1}(y_1, dy_2) \cdots \int_{B_{t_p}} Q_{t_p-t_{p-1}}(y_{p-1}, dy_p),$$

which is measurable. The remaining case then follows from a π - λ theorem argument.

In addition, for any initial law, the expectation \mathbb{E}_γ taken with respect to P^γ is given by

$$\mathbb{E}_\gamma[\mathbb{1}_A] = \int \gamma(dx) \int_{B_1} Q_{t_1}(x, dy_1) \int_{B_2} Q_{t_2-t_1}(y_1, dy_2) \cdots \int_{B_{t_p}} Q_{t_p-t_{p-1}}(y_{p-1}, dy_p) = \int_E \mathbb{E}_x[\mathbb{1}_A] \gamma(dx).$$

Using an argument of π - λ system and a simple function approximation, we obtain

$$\mathbb{E}_\gamma[\Phi] = \int_E \mathbb{E}_x[\Phi] \gamma(dx), \quad \forall \text{ measurable mapping } \Phi : (\mathbb{D}(E), \mathcal{D}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+)).$$

Theorem 6.13 (Simple Markov property). Let $(X_t)_{t \geq 0}$ be a Markov process with semigroup $(Q_t)_{t \geq 0}$ with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Assume that process $(X_t)_{t \geq 0}$ has càdlàg sample paths. Let $s \geq 0$, and let $\Phi : \mathbb{D}(E) \rightarrow \mathbb{R}_+$ be a measurable map. Then

$$\mathbb{E}[\Phi((X_{s+t})_{t \geq 0}) | \mathcal{F}_s] = \mathbb{E}_{X_s}[\Phi] \quad a.s..$$

Proof. Following our preceding discussion, $\mathbb{E}_{X_s}[\Phi]$ is a composition of X_s and the mapping $x \mapsto \mathbb{E}_x[\Phi]$. It suffices to consider the case $\Phi = \mathbb{1}_A$, where A is given in (6.6). For $\varphi_1, \dots, \varphi_p \in B(E)$, we have

$$\begin{aligned} & \mathbb{E}[\varphi_1(X_{s+t_1}) \cdots \varphi_p(X_{s+t_p}) | \mathcal{F}_s] \\ &= \mathbb{E}[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) \mathbb{E}[\varphi_p(X_{s+t_p}) | \mathcal{F}_{s+t_{p-1}}] | \mathcal{F}_s] \\ &= \mathbb{E}[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) Q_{t_p-t_{p-1}} \varphi_p(X_{s+t_{p-1}}) | \mathcal{F}_s] \\ &= \mathbb{E}\left[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) \int_E Q_{t_p-t_{p-1}}(X_{s+t_{p-1}}, dy_p) \varphi_p(y_p) dy_p | \mathcal{F}_s\right] \\ &= \cdots = \int_E Q_{t_1}(X_s, dy_1) \varphi_1(y_1) \int_E Q_{t_2-t_1}(y_1, dy_2) \varphi_2(y_2) \cdots \int_E Q_{t_p-t_{p-1}}(y_{p-1}, dy_p) \varphi_p(y_p). \end{aligned}$$

Taking $\varphi_j = \mathbb{1}_{B_j}$ immediately gives the desired conclusion according to (6.6). \square

Theorem 6.14 (Strong Markov property). Let $(X_t)_{t \geq 0}$ be a Feller process with semigroup $(Q_t)_{t \geq 0}$ with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Assume that process $(X_t)_{t \geq 0}$ has càdlàg sample paths. Let τ be a stopping time of the filtration $(\mathcal{F}_{t+})_{t \geq 0}$, and let $\Phi : \mathbb{D}(E) \rightarrow \mathbb{R}_+$ be a measurable map. Then

$$\mathbb{E} [\mathbf{1}_{\{\tau < \infty\}} \Phi((X_{\tau+t})_{t \geq 0}) | \mathcal{F}_\tau] = \mathbf{1}_{\{\tau < \infty\}} \mathbb{E}_{X_\tau} [\Phi] \quad a.s..$$

Proof. The right hand side of the last display is a measurable mapping, since the mapping $\{\tau < \infty\} \ni \omega \mapsto X_\tau(\omega)$ is \mathcal{F}_τ -measurable by Proposition 3.12, and the mapping $x \mapsto \mathbb{E}_x[\Phi]$ is measurable. To show the desired conclusion, it suffices to show

$$\mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \Phi((X_{\tau+t})_{t \geq 0})] = \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \mathbb{E}_{X_\tau} [\Phi]] \quad a.s..$$

As is discussed above, we can restrict our discussion to the special case $\Phi(f) = \varphi_1(f(t_1)) \cdots \varphi_p(f(t_p))$, where $0 \leq t_1 < t_2 < \cdots < t_p$, $\varphi_1, \dots, \varphi_p \in B(E)$.

Consider the case $p = 1$. Let $t > 0$ and $\varphi \in B(E)$ fixed. By Riesz-Markov representation, we may assume that $\varphi \in C_0(E)$. Denote by $[\tau]_n$ the estimate $\max\{k2^{-n} : k2^{-n} > \tau, k \in \mathbb{N}\}$. Then

$$\begin{aligned} \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \varphi(X_{\tau+t})] &= \lim_{n \rightarrow \infty} \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \varphi(X_{[\tau]_n+t})] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{k=1}^{\infty} \mathbf{1}_{A \cap \{(k-1)2^{-n} \leq \tau < k2^{-n}\}} \varphi(X_{k2^{-n}+t}) \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{k=1}^{\infty} \mathbf{1}_{A \cap \{(k-1)2^{-n} \leq \tau < k2^{-n}\}} Q_t \varphi(X_{k2^{-n}}) \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} Q_t \varphi(X_{[\tau]_n})] = \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} Q_t \varphi(X_\tau)], \end{aligned}$$

where first and last equalities uses right-continuity of sample paths and the fact that $\varphi, Q_t \varphi \in C_0(E)$, and the third equality uses Proposition 3.10 (ii), which implies $A \cap \{(k-1)2^{-n} \leq \tau < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$.

For the general case $p \in \mathbb{N}$, we have

$$\begin{aligned} &\mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \varphi_1(X_{\tau+t_1}) \cdots \varphi_p(X_{\tau+t_p})] \\ &= \mathbb{E} [\mathbf{1}_{A \cap \{\tau < \infty\}} \varphi_1(X_{\tau+t_1}) \cdots \varphi_{p-1}(X_{\tau+t_{p-1}}) \mathbb{E} [\mathbf{1}_{\{\tau < \infty\}} \varphi_p(X_{\tau+t_p}) | \mathcal{F}_{\tau+t_{p-1}}]] \\ &= \mathbb{E} \left[\mathbf{1}_{A \cap \{\tau < \infty\}} \varphi_1(X_{\tau+t_1}) \cdots \varphi_{p-1}(X_{\tau+t_{p-1}}) \int_E Q_{t_p-t_{p-1}}(X_{\tau+t_{p-1}}, dy_p) \varphi_p(y_p) \right] \\ &= \cdots = \mathbb{E} \left[\mathbf{1}_{A \cap \{\tau < \infty\}} \int_E Q_{t_1}(X_\tau, dy_1) \varphi_1(y_1) \int_E Q_{t_2-t_1}(y_1, dy_2) \varphi_2(y_2) \cdots \int_E Q_{t_p-t_{p-1}}(y_{p-1}, dy_p) \varphi_p(y_p) \right]. \end{aligned}$$

This concludes the proof of strong Markov property. \square

Remark. In this theorem, we assume that τ is a stopping time of the filtration $(\mathcal{F}_{t+})_{t \geq 0}$. This is a more general assumption than a stopping time of the filtration $(\mathcal{F}_t)_{t \geq 0}$.

6.3 The Generator and Feynman-Kac Formula

In this section, we are going to discuss some properties of generators. We first study the Brownian process, whose generator has a simple form.

Example 6.15. A Brownian motion $B = (B_t)_{t \geq 0}$ is a real-valued Markov process with transition group

$$Q_0(x, A) = \mathbb{1}_A(x), \quad Q_t(x, A) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} dy, \quad \forall t > 0, x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R}).$$

We first verify that $(Q_t)_{t \geq 0}$ is a Feller semigroup. We fix $f \in C_0(\mathbb{R})$, and there exists $M > 0$ such that $\|f\| \leq M$. For any $x_0 \in \mathbb{R}$, by Lebesgue dominated convergence theorem,

$$\lim_{x \rightarrow x_0} Q_t f(x) = \lim_{x \rightarrow x_0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} f(y) dy = Q_t f(x_0).$$

Given any $\epsilon > 0$, we choose $K > 0$ such that $|f(y)| < \epsilon/2$ for all $|y| > K$, and choose $\alpha > 0$ such that

$$\int_{-\alpha}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} dz < \frac{\epsilon}{2M} \quad \Leftrightarrow \quad \int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} dz < \frac{\epsilon}{2}.$$

Then for all $x > K + \alpha$, we have $|f(z+x)| < \epsilon/2$ for all $z > -\alpha$, and

$$|Q_t f(x)| = \left| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y) dy \right| \leq M \int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} dz + \int_{-\alpha}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} |f(z+x)| dz < \alpha.$$

Hence we conclude that $Q_t f \in C_0(\mathbb{R})$. To show continuity of $(Q_t)_{t \geq 0}$, we fix $\eta > 0$. Then we choose $\delta > 0$ such that $|f(y) - f(x_0)| < \eta/2$ for all $|y - x_0| \leq \delta$:

$$\left| \int_{|y-x_0| \leq \delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} (f(y) - f(x_0)) dy \right| < \frac{\eta}{2} \left| \int_{|y-x_0| \leq \delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} dy \right| \leq \frac{\eta}{2}, \quad \forall t > 0;$$

and choose $t_0 > 0$ such that

$$\int_{|z| \geq \delta} \frac{1}{\sqrt{2\pi t_0}} e^{-\frac{z^2}{2t_0}} dz < \frac{\eta}{4M} \quad \Rightarrow \quad \left| \int_{|y-x_0| \geq \delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} (f(y) - f(x_0)) dy \right| < \frac{\eta}{2}, \quad \forall t \in (0, t_0).$$

Consequently, we have $|(Q_t f - f)(x_0)| < \eta$ for all $t \in (0, t_0)$, and $Q_t f(x_0) \rightarrow f(x_0)$ as $t \downarrow 0$.

Resolvent and generator. For $\lambda > 0$, the resolvent is

$$\begin{aligned} R_\lambda f(x) &= \int_0^\infty e^{-\lambda t} Q_t f(x) dt = \int_{-\infty}^\infty \underbrace{\left(\int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\lambda t - \frac{(y-x)^2}{2t}} dt \right)}_{r_\lambda(y-x)} f(y) dy \\ &= \int_{-\infty}^\infty \frac{1}{|y-x|} \mathbb{E}[\tau_y e^{-\lambda \tau_y}] f(y) dy, \end{aligned}$$

where the hitting time $\tau_y = \inf\{t \geq 0 : B_t = |y-x|\}$ has density $\frac{|y-x|}{\sqrt{2\pi t^3}} e^{-\frac{(y-x)^2}{2t}} dt$. By Proposition 4.27,

$$r_\lambda(y-x) = \frac{1}{|y-x|} \frac{d}{d\lambda} \mathbb{E}[e^{-\lambda \tau_y}] = \frac{1}{\sqrt{2\lambda}} e^{-|y-x|\sqrt{2\lambda}}.$$

This gives the formula of R_λ :

$$R_\lambda f(x) = \frac{1}{\sqrt{2\lambda}} \int_{-\infty}^{\infty} e^{-|y-x|\sqrt{2\lambda}} f(y) dy.$$

Now we find the generator L of $(Q_t)_{t \geq 0}$. If $h \in \mathfrak{D}(L)$, there exists $f \in C_0(\mathbb{R})$ such that $R_\lambda f = h$. Taking $\lambda = 1/2$, and $\text{sgn}(z) = \mathbb{1}_{\{z \geq 0\}} - \mathbb{1}_{\{z < 0\}}$, we have

$$h(x) = \int_{-\infty}^{\infty} e^{-|y-x|} f(y) dy \quad \Rightarrow \quad h'(x) = \int_{-\infty}^{\infty} \text{sgn}(y-x) e^{-|y-x|} f(y) dy.$$

Furthermore, h' is differentiable: for all $x \in \mathbb{R}$,

$$\begin{aligned} h'(x+\delta) - h'(x) &= \int_{-\infty}^{\infty} \text{sgn}(y-x-\delta) e^{-|y-x-\delta|} f(y) dy - \int_{-\infty}^{\infty} \text{sgn}(y-x) e^{-|y-x|} f(y) dy \\ &= \int_{\mathbb{R} \setminus [x, x+\delta]} \text{sgn}(y-x) \left(e^{-|y-x-\delta|} - e^{-|y-x|} \right) f(y) dy - \int_x^{x+\delta} (e^{y-x-\delta} + e^{x-y}) f(y) dy, \end{aligned}$$

Hence we have

$$\lim_{\delta \downarrow 0} \frac{h'(x+\delta) - h'(x)}{\delta} = h(x) - 2f(x).$$

A similar argument also holds for the left limit, hence $h'' = h - 2f$. By Proposition 6.8, we have

$$\left(\frac{1}{2} \text{Id} - L \right) h = \left(\frac{1}{2} \text{Id} - L \right) R_{1/2} f = f \Rightarrow Lh = \frac{1}{2} h'', \quad \text{where } h \in \mathfrak{D}(L) \subset \{h \in C^2(\mathbb{R}) : g, g'' \in C_0(\mathbb{R})\}.$$

In fact, we can show that $\mathfrak{D}(L) = \{g \in C^2(\mathbb{R}) : g, g'' \in C_0(\mathbb{R})\}$. If g is a twice continuously differentiable function with $g, g'' \in C_0(\mathbb{R})$, we take $f = \frac{1}{2}(g - g'') \in C_0(\mathbb{R})$. Then $h = R_{1/2} f \in \mathfrak{D}(L)$, and the preceding argument gives $h'' = h - 2f$. This yields $(h - g)'' = h - g \Rightarrow (h - g)(x) = C_1 e^x + C_2 e^{-x}$, where $C_1, C_2 \in \mathbb{R}$. Since $h - g \in C_0(\mathbb{R})$, we have $g = h \in \mathfrak{D}(L)$.

Proposition 6.16. Let $d \in \mathbb{N}$, and let $B = (B_t)_{t \geq 0}$ be a d -dimensional Brownian motion. The infinitesimal generator of B is equal to $\frac{1}{2} \Delta$ on the space $C_0^2(\mathbb{R}^d)$.

Proof. For $f \in C_0(\mathbb{R}^d)$, we write

$$Q_t f(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-|z|^2/2} f(x + z\sqrt{t}) dz, \quad (6.7)$$

where $|z|^2 = \sum_{j=1}^d z_j^2$. If $f \in C_0^2(\mathbb{R}^d)$, the chain rule gives

$$\frac{\partial}{\partial u} f(x + zu) = z^\top \nabla f(x + zu), \quad \frac{\partial^2}{\partial u^2} f(x + zu) = z^\top \nabla^2 f(x + zu) z,$$

where $\nabla^2 f = [\frac{\partial^2}{\partial x_i \partial x_j} f]_{i,j \in [n]}$ is Hessian matrix of f . By Taylor's formula, there exists $\theta(t, z) \in [0, 1]$ such that

$$\begin{aligned} Q_t f(x) &= f(x) + (2\pi)^{-d/2} \frac{t}{2} \int_{\mathbb{R}^d} e^{-|z|^2/2} z^\top \nabla^2 f(x + \theta(t, z) z\sqrt{t}) z dz \\ &= f(x) + \frac{t}{2} \Delta f(x) + (2\pi)^{-d/2} \frac{t}{2} J(t, x), \end{aligned}$$

where

$$J(t, x) = \int_{\mathbb{R}^d} e^{-|z|^2/2} \left[\sum_{i,j=1}^d \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x + \theta(t, z)z\sqrt{t}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right) z_i z_j \right] dz.$$

We set the following function F , and use uniform continuity of second partial derivatives of f :

$$F(x, z, t) = \max_{i,j \in [d]} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x + \theta(t, z)z\sqrt{t}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right| \Rightarrow \limsup_{t \downarrow 0} F(x, z, t) = 0.$$

Moreover, for any $R > 0$, one have

$$|J(t, x)| \leq \int_{|z| \leq R} F(x, t, z) e^{-|z|^2/2} \|z\|_1^2 dz + 2 \max_{i,j \in [d]} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\| \int_{|z| > R} e^{-|z|^2/2} \|z\|_1^2 dz.$$

Since the first term converges to 0 as $t \downarrow 0$ by dominated convergence theorem for any $R > 0$, we have

$$\lim_{t \downarrow 0} \left\| \frac{1}{t} (Q_t f - f) - \frac{1}{2} \Delta f \right\| \leq \max_{i,j \in [d]} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\| \int_{|z| > R} e^{-|z|^2/2} \|z\|_1^2 dz, \quad \forall R > 0.$$

Take $R \rightarrow \infty$. Then we have $Lf = \frac{1}{2} \Delta f$ for all $f \in C_0^2(\mathbb{R})$. \square

Remark. For the case $d \geq 2$, the space $C_0^2(\mathbb{R}^d)$ is not equal to $\mathfrak{D}(L)$. One can show that $\mathfrak{D}(L)$ is the subspace of $C_0(\mathbb{R}^d)$ of functions f such that δf taken in weak sense is in $C_0(\mathbb{R}^d)$.

In addition, according to Proposition 6.8, we have for all $\varphi \in C_0^2(\mathbb{R}^d)$ that

$$\frac{d}{dt} Q_t \varphi = \frac{1}{2} Q_t \Delta \varphi = \frac{1}{2} \Delta Q_t \varphi$$

Consequently, the function $u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$, $(t, x) \mapsto Q_t \varphi(x)$ solves the following partial differential equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \Delta_x u(t, x), & t > 0 \\ u(0, x) = \varphi(x) \end{cases}$$

This is called the *heat diffusion equation*. Note that $(Q_t)_{t \geq 0}$ is defined in (6.7).

The following theorem gives a characterization for the domain of the generator of a Feller semigroup $(Q_t)_{t \geq 0}$. For any $x \in E$, we can construct a Markov process $(X_t^x)_{t \geq 0}$ such that $\mathbb{P}(X_0 = x) = 1$ with semigroup $(Q_t)_{t \geq 0}$.

Theorem 6.17. Let $h, g \in C_0(E)$. The following conditions are equivalent:

- (i) $h \in \mathfrak{D}(L)$ and $Lh = g$;
- (ii) For every $x \in E$, the process

$$h(X_t^x) - \int_0^t g(X_s^x) ds$$

is a martingale with respect to the filtration $\mathcal{F}_t^x = \sigma(X_s^x, 0 \leq s \leq t)$.

Proof. (i) \Rightarrow (ii): Let $h \in \mathfrak{D}(L)$ and $Lh = g$. By (6.1), we have

$$Q_t h = h + \int_0^t Q_s g ds \Rightarrow \mathbb{E} [h(X_{t+s}^x) | \mathcal{F}_t^x] = Q_s h(X_t^x) = h(X_t^x) + \int_0^s Q_r g(X_t^x) dr. \quad (6.8)$$

Meanwhile, use the boundedness of g , we have

$$\mathbb{E} \left[\int_t^{t+s} g(X_r^x) dr \middle| \mathcal{F}_t^x \right] = \int_t^{t+s} \mathbb{E} [g(X_r^x) | \mathcal{F}_t] = \int_t^{t+s} Q_{r-t} g(X_t^x) dr = \int_0^s Q_r g(X_t^x) dr. \quad (6.9)$$

Combining (6.8) and (6.9), we have the martingale property:

$$\mathbb{E} \left[h(X_{t+s}^x) - \int_0^{t+s} g(X_r^x) dr \middle| \mathcal{F}_t^x \right] = h(X_t^x) - \int_0^t g(X_r^x) dr, \quad \forall t+s > t \geq 0.$$

(ii) \Rightarrow (i): For every $x \in E$ and every $t \geq 0$, we have

$$Q_t h(x) - \int_0^t Q_r g(x) dr = \mathbb{E} \left[h(X_t^x) - \int_0^t g(X_r^x) dr \right] = h(x),$$

where the first equality follows from Markov property, and the second from martingale property. Hence

$$\lim_{t \downarrow 0} \frac{Q_t h - h}{t} - g = \lim_{t \downarrow 0} \frac{1}{t} \int_0^t (Q_r g - g) dr = 0.$$

Since $g \in C_0(E)$, we have $h \in \mathfrak{D}(L)$, and $Lh = g$. □

Now we introduce one case of Feynman-Kac formula.

Theorem 6.18. (Feynman-Kac formula). Let $v \in C_0(E)$ be a nonnegative function. For every $t \geq 0$, define for every $\varphi \in B(E)$ and every $x \in E$ that

$$K_t \varphi(x) = \mathbb{E} \left[\varphi(X_t^x) \exp \left(- \int_0^t v(X_s^x) ds \right) \right],$$

where $(X_t^x)_{t \geq 0}$ is a càdlàg Feller process with semigroup $(Q_t)_{t \geq 0}$ starting from $X_0 \sim \delta_x$.

(i) $(K_t)_{t \geq 0}$ is a semigroup of contractions on $B(E)$.

(ii) If $\varphi \in \mathfrak{D}(L)$, then

$$\frac{d}{dt} K_t \varphi|_{t=0} = L\varphi - v\varphi.$$

Proof. (i) Fix $x \in E$, and let $(X_t)_{t \geq 0}$ be a càdlàg Feller process with semigroup $(Q_t)_{t \geq 0}$ starting from $X_0 \sim \delta_x$. Clearly, K_t is a contraction: $K_t \varphi(x) \leq \mathbb{E} [\varphi(X_t)] = Q_t \varphi(x) \leq \|\varphi\|$. Next, we let $\pi_t(f) = f(t)$ be the projection map from $\mathbb{D}(E) \rightarrow E$, and fix $s, t \geq 0$. By definition of K_t and simple Markov property,

$$\begin{aligned} K_s(K_t \varphi)(x) &= \mathbb{E} \left[(K_t \varphi)(X_s) \exp \left(- \int_0^s v(X_s) ds \right) \right] \\ &= \mathbb{E} \left[\mathbb{E}_{X_s} \left[(\varphi \circ \pi_t) \exp \left(- \int_0^t v \circ \pi_r dr \right) \right] \exp \left(- \int_0^s v(X_r) dr \right) \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\varphi(X_{s+t}) \exp \left(- \int_0^t v(X_{s+r}) dr \right) \middle| \mathcal{F}_s^X \right] \exp \left(- \int_0^s v(X_r) dr \right) \right] \\ &= \mathbb{E} \left[\varphi(X_{s+t}) \exp \left(- \int_0^{s+t} v(X_r) dr \right) \right] = (K_{s+t} \varphi)(x). \end{aligned}$$

(ii) Since $\frac{d}{ds}e^{-\int_s^t f(r) dr} = f(s)e^{-\int_s^t f(r) dr}$, the fundamental theorem of calculus gives

$$1 - e^{-\int_0^t v(X_r) dr} = \int_0^t v(X_s) e^{-\int_s^t v(X_r) dr} ds.$$

By Fubini's theorem and simple Markov property, for every $\varphi \in B(E)$, we have

$$\begin{aligned} K_t \varphi(x) &= \mathbb{E} \left[\varphi(X_t) \left(1 - \int_0^t v(X_s) \exp \left(- \int_s^t v(X_r) dr \right) ds \right) \right] \\ &= Q_t \varphi(x) - \int_0^t \mathbb{E} \left[\varphi(X_t) v(X_s) \exp \left(- \int_s^t v(X_r) dr \right) \right] ds \\ &= Q_t \varphi(x) - \int_0^t \mathbb{E} \left[v(X_s) \mathbb{E} \left[\varphi(X_t) \exp \left(- \int_s^t v(X_r) dr \right) \middle| \mathcal{F}_s^X \right] \right] ds \\ &= Q_t \varphi(x) - \int_0^t \mathbb{E} [v(X_s) K_{t-s} \varphi(X_s)] ds \\ &= Q_t \varphi(x) - \int_0^t Q_s(v K_{t-s} \varphi)(x) ds. \end{aligned}$$

Hence $\lim_{t \downarrow 0} K_t \varphi = \varphi$. Furthermore,

$$\begin{aligned} \left| \frac{1}{t} \int_0^t Q_s(v K_{t-s} \varphi)(x) ds - v(x) \varphi(x) \right| &= \frac{1}{t} \left| \int_0^t Q_s(v K_{t-s} \varphi - v \varphi)(x) ds \right| + \frac{1}{t} \left| \int_0^t (Q_s - \text{Id})(v \varphi)(x) ds \right| \\ &\leq \frac{1}{t} \int_0^t \|v K_{t-s} \varphi - v \varphi\| ds + \frac{1}{t} \int_0^t \|(Q_s - \text{Id})(v \varphi)\| ds \\ &\leq \frac{1}{t} \int_0^t \|v\| \|K_s \varphi - \varphi\| ds + \frac{1}{t} \int_0^t \|(Q_s - \text{Id})(v \varphi)\| ds \end{aligned}$$

Letting $t \downarrow 0$, the last display converges to zero. Consequently, we have

$$\frac{d}{dt} K_t \varphi|_{t=0} = \frac{d}{dt} Q_t \varphi|_{t=0} - \frac{d}{dt} \int_0^t Q_s(v K_{t-s} \varphi)(x) ds|_{t=0} = L \varphi - v \varphi,$$

which complete the proof. \square

Remark. To find the derivative of $t \mapsto K_t \varphi$, we use the semigroup property:

$$\frac{d}{dt} K_t \varphi = \frac{d}{ds} K_{s+t} \varphi|_{s=0} = \frac{d}{ds} K_s(K_t \varphi)|_{s=0} = L K_t \varphi - v K_t \varphi.$$

Therefore, for any $f \in B(E)$, the function

$$\psi(t, x) = K_t f(x) = \mathbb{E} \left[f(X_t^x) \exp \left(- \int_0^t v(X_s^x) ds \right) \right]$$

solves the following initial value problem:

$$\begin{cases} \frac{\partial \psi}{\partial t}(t, x) = L \psi(t, x) - v(x) \psi(t, x), & t > 0 \\ \psi(0, x) = f(x). \end{cases}$$

Here (X_t^x) is a càdlàg Feller process with generator L starting from $X_0 \sim \delta_x$. Letting $v = 0$ gives the *Kolmogorov backward equation*, which coincides the form of (6.1).

6.4 Diffusion Processes

In this section, we discuss the solution of SDE $E(\sigma, b)$

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt, \quad \sigma = (\sigma_{ij})_{i \in [p], j \in [q]}, \quad b = (b_i)_{i \in [p]}. \quad (6.10)$$

in form of a Markov process. The processes with continuous sample paths that are obtained as solutions of SDE (6.10) are called *diffusion processes*. The function $b : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is called the *drift coefficient*, and $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times q}$ is called the *diffusion coefficient*.

We tackle with the homogeneous case where $\sigma(t, x) = \sigma(x)$ and $b(t, x) = b(x)$, and we maintain the Lipschitz assumption: there exists a constant $K > 0$ such that for all $x, y \in \mathbb{R}^q$,

$$|\sigma(x) - \sigma(y)| \leq K|x - y|, \quad |b(x) - b(y)| \leq K|x - y|.$$

Here we use $|\cdot|$ to denote the Euclidean norm of vectors and the Frobenius norm of matrices.

Theorem 6.19. Assume that $X = (X_t)_{t \geq 0}$ is a solution of SDE $E(\sigma, b)$ on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Then $(X_t)_{t \geq 0}$ is a Markov process with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$, with semigroup $(Q_t)_{t \geq 0}$ defined by

$$Q_t f(x) = \mathbb{E}[f(X_t^x)],$$

where $X^x = (X_t^x)_{t \geq 0}$ is an arbitrary solution of $E^x(\sigma, b)$. Using the notation of Theorem 5.24, we also write

$$Q_t f(x) = \int f(F_x(\mathbf{w}))_t W(d\mathbf{w}). \quad (6.11)$$

Proof. We first prove that, for any $f \in B(\mathbb{R}^p)$ and any $s, t \geq 0$, we have

$$\mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] = Q_t f(X_s),$$

To deal the time shift s , we define filtration $(\mathcal{F}'_t)_{t \geq 0}$ and processes $(X'_t)_{t \geq 0}$, $(B'_t)_{t \geq 0}$ as follows:

$$\mathcal{F}'_t = \mathcal{F}_{s+t}, \quad X'_t = X_{s+t}, \quad B'_t = B_{t+s} - B_s.$$

Then $(\mathcal{F}'_t)_{t \geq 0}$ is still complete, X' is adapted to $(\mathcal{F}'_t)_{t \geq 0}$, and B' is a q -dimensional (\mathcal{F}'_t) -Brownian motion. Furthermore, the approximation formula (5.8) for the integral of sample-continuous adapted processes gives

$$\begin{aligned} X'_t &= X_{s+t} = X_s + \int_s^{s+t} \sigma(X_r) dB_r + \int_s^{s+t} b(X_r) dr \\ &= X_s + \int_0^t \sigma(X'_r) dB'_r + \int_0^t b(X'_r) dr. \end{aligned}$$

Consequently, X' solves $E(\sigma, b)$ on the space $(\Omega, \mathcal{F}, (\mathcal{F}'_t)_{t \geq 0}, \mathbb{P})$ and with Brownian motion B' , with $X'_0 = X_s$. By Theorem 5.24 (iii), we have $X' = F_{X_s}(B')$ a.s., which implies

$$\begin{aligned} \mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] &= \mathbb{E}[f(X'_t) | \mathcal{F}_s] = \mathbb{E}[f(F_{X_s}(B')_t) | \mathcal{F}_s] \\ &= \int f(F_{X_s}(\mathbf{w}))_t W(d\mathbf{w}) = Q_t f(X_s), \end{aligned}$$

where in the third equality we use the independence of B' and \mathcal{F}_s .

Now it remains to verify that $(Q_t)_{t \geq 0}$ is a transition semigroup. Clearly, Q_0 is an identity map, and $(t, x) \mapsto Q_t f(x)$ is a continuous map, hence is measurable. Finally, one have

$$Q_{s+t}f(x) = \mathbb{E}[X_{s+t}^x] = \mathbb{E}[\mathbb{E}[X_{s+t}^x | \mathcal{F}_s]] = \mathbb{E}[Q_t f(X_s^x)] = \int Q_s(x, dy) Q_t f(y)$$

which is the Chapman-Kolmogorov equation. This completes the proof. \square

We then give an estimate for the second moment of a diffusion process.

Lemma 6.20. Fix $x \in \mathbb{R}^p$, and let $(X_t^x)_{t \geq 0}$ be a solution of the SDE $E^x(\sigma, b)$. Then there exists a constant $C_x > 0$ depending only on x , such that for all $t \geq 0$,

$$\mathbb{E}[|X_t^x - x|^2] \leq C_x e^{C_x(t+t^2)}(t+t^2).$$

Proof. By triangle inequality and Lipschitz property, for all $t \geq 0$, we have

$$|\sigma(X_t^x)|^2 \leq (|\sigma(x)| + |\sigma(X_t^x) - \sigma(x)|)^2 \leq 2|\sigma(x)|^2 + 2K^2|X_t^x - x|^2. \quad (6.12)$$

A similar formula also holds for $|b(X_t^x)|^2$.

We define a stopping time $\tau = \inf\{t \geq 0 : |X_t^x - x| > M\}$ for some $M > 0$, and fix $T > 0$. Then the function $t \mapsto \mathbb{E}[|X_{t \wedge \tau}^x - x|^2]$ is bounded on $[0, T]$. For any $t \in [0, T]$, we have

$$\begin{aligned} \mathbb{E}[|X_{t \wedge \tau}^x - x|^2] &\leq 2\mathbb{E}\left[\left(\int_0^{t \wedge \tau} \sigma(X_s^x) dB_s\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^{t \wedge \tau} b(X_s^x) ds\right)^2\right] \\ &\leq 2\mathbb{E}\left[\int_0^{t \wedge \tau} |\sigma(X_s^x)|^2 ds\right] + 2\mathbb{E}\left[T \int_0^{t \wedge \tau} |b(X_s^x)|^2 ds\right] \\ &\leq 4T(|\sigma(x)|^2 + T|b(x)|^2) + 4K^2(1+T) \int_0^{t \wedge \tau} \mathbb{E}[|X_s^x - x|^2] ds \\ &\leq 4T(|\sigma(x)|^2 + T|b(x)|^2) + 4K^2(1+T) \int_0^t \mathbb{E}[|X_{s \wedge \tau}^x - x|^2] ds. \end{aligned}$$

where we use (6.12) in the third inequality. By Gronwall's lemma [Lemma 5.22], we have

$$\mathbb{E}[|X_{t \wedge \tau}^x - x|^2] \leq 4T(|\sigma(x)|^2 + T|b(x)|^2)e^{4K^2(1+T)t}, \quad \forall t \in [0, T].$$

Let $M \rightarrow \infty$, and use the monotone convergence theorem. Then for all $t \geq 0$, we have

$$\mathbb{E}[|X_t^x - x|^2] \leq 4t(|\sigma(x)|^2 + t|b(x)|^2)e^{4K^2(t+t^2)}.$$

Setting $C_x = 4 \max\{|\sigma(x)|^2, |b(x)|^2, K^2\}$ concludes the proof. \square

To move forward, we have the following conclusion.

Theorem 6.21. The semigroup $(Q_t)_{t \geq 0}$ is a Feller semigroup, and its generator L satisfies $\mathfrak{D}(L) \supset C_c^2(\mathbb{R}^p)$. In addition, for every $f \in C_c^2(\mathbb{R}^p)$, we have

$$Lf = \frac{1}{2} \sum_{i,j=1}^p (\sigma \sigma^*)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^p b_i \frac{\partial f}{\partial x_i},$$

where σ^* is the matrix transpose of σ .

Proof. (i) We first fix $f \in C_0(\mathbb{R}^p)$, and verify that $Q_t f \in C_0(\mathbb{R}^p)$. Since $x \mapsto F_x(\mathbf{w})$ is continuous for every $\mathbf{w} \in C(\mathbb{R}^+, \mathbb{R}^q)$, the continuity of $Q_t f$ follows from (6.11) and dominated convergence theorem. Let X_t^x be a solution of $E^x(\sigma, b)$. By Markov's inequality and Lemma 6.20, there exists a constant $C_x > 0$ such that

$$\mathbb{P}(|X_t^x - x| > \lambda) \leq \frac{1}{\lambda^2} \mathbb{E}[|X_t^x - x|^2] \leq \frac{C_x e^{C_x(t+t^2)}(t+t^2)}{\lambda^2}, \quad \forall t \geq 0.$$

Then we have

$$|Q_t f(x)| \leq |\mathbb{E}[f(X_t^x) \mathbf{1}_{\{|X_t^x - x| \leq \lambda\}}]| + \|f\| \mathbb{P}(|X_t^x - x| > \lambda) \leq \sup_{y: |y-x| \leq \lambda} |f(y)| + \frac{C_x e^{C_x(t+t^2)} \|f\| (t+t^2)}{\lambda^2}$$

Since $f \in C_0(\mathbb{R}^p)$, the first term of the last display converges to 0 as $x \rightarrow \infty$ for all $\lambda > 0$. Then we have $\limsup_{x \rightarrow \infty} |Q_t f(x)| \leq C \|f\| (t^2 + t)/\lambda^2$ for all $\lambda > 0$. This implies $Q_t f(x) \rightarrow 0$ as $x \rightarrow \infty$.

Now we fix $x \in \mathbb{R}^p$, and verify that $Q_t f(x) \rightarrow f(x)$ as $t \downarrow 0$. For any $\lambda > 0$,

$$\begin{aligned} |\mathbb{E}[f(X_t^x)] - f(x)| &\leq \sup_{y: |y-x| \leq \lambda} |f(x) - f(y)| + 2\|f\| \mathbb{P}(|X_t^x - x| > \lambda) \\ &\leq \sup_{y: |y-x| \leq \lambda} |f(x) - f(y)| + \frac{2C_x e^{C_x(t+t^2)} \|f\| (t+t^2)}{\lambda^2} \\ &\rightarrow \sup_{y: |y-x| \leq \lambda} |f(x) - f(y)|, \quad \text{as } t \downarrow 0. \end{aligned}$$

Taking $\lambda \downarrow 0$ gives $Q_t f(x) \rightarrow f(x)$.

(ii) To prove the second assertion, we use Itô's formula to $f(X_t^x)$, where $f \in C_c^2(\mathbb{R}^p)$. Recall that

$$\begin{aligned} X_t^{x,i} &= x_i + \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x) dB_s^k + \int_0^t b_i(X_s^x) ds, \\ \langle X^{x,i}, X^{x,j} \rangle_t &= \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x) \sigma_{jk}(X_s^x) ds = \int_0^t (\sigma \sigma^*)_{ij}(X_s^x) ds. \end{aligned}$$

Using Itô's formula [Theorem 5.9] and associativity of stochastic integrals, one have

$$\begin{aligned} f(X_t^x) &= f(x) + \sum_{i=1}^p \int_0^t \frac{\partial f}{\partial x_i}(X_s^x) dX_s^{x,i} + \frac{1}{2} \sum_{i,j=1}^p \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x) d\langle X^{x,i}, X^{x,j} \rangle_s \\ &= f(x) + \sum_{i=1}^p \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x) \frac{\partial f}{\partial x_i}(X_s^x) dB_s^k + \int_0^t \left[\sum_{i=1}^p b_i(X_s^x) \frac{\partial f}{\partial x_i}(X_s^x) + \frac{1}{2} \sum_{i,j=1}^p (\sigma \sigma^*)_{ij}(X_s^x) \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x) \right] ds. \end{aligned}$$

Define $g = \frac{1}{2} \sum_{i,j=1}^p (\sigma \sigma^*)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^p b_i \frac{\partial f}{\partial x_i}$. Then the process

$$M_t = \sum_{i=1}^p \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x) \frac{\partial f}{\partial x_i}(X_s^x) dB_s^k = f(X_t^x) - f(x) - \int_0^t g(X_s^x) ds$$

is a continuous local martingale. Since $f \in C_c^2(\mathbb{R}^p)$, both f and g are compactly supported continuous functions, hence the process $(M_t)_{t \geq 0}$ is uniformly bounded. By Proposition 3.54 (ii), M is a martingale. According to Theorem 6.17, $f \in \mathfrak{D}(L)$, and $Lf = g$, which complete our proof. \square

Remark. Let $\Sigma = \sigma\sigma^*$ be the covariance matrix. We can formally write the generator L of $E(\sigma, b)$ as

$$L = \frac{1}{2}\Sigma \cdot \nabla^2 + b \cdot \nabla,$$

where ∇^2 is the Hessian matrix. Fix $f \in C_c^2(\mathbb{R}^p)$, and let $u(t, x) = Q_t f(x)$. Then $u : \mathbb{R}_+ \times \mathbb{R}^p \rightarrow \mathbb{R}$ solves the following Kolmogorov backward equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \frac{1}{2}\Sigma(x) \cdot \nabla_x^2 u(t, x) + b(x) \cdot \nabla_x u(t, x), & t > 0 \\ u(0, x) = f(x). \end{cases}$$

Example 6.22 (Ornstein-Uhlenbeck process). Let $\lambda > 0$. An one-dimensional *Ornstein-Uhlenbeck* process is the solution of the following SDE:

$$dX_t = dB_t - \lambda X_t dt.$$

By applying Itô's formula to $e^{\lambda t} X_t$, we have

$$e^{\lambda t} X_t = X_0 + \int_0^t e^{\lambda s} dX_t + \lambda \int_0^t e^{\lambda s} X_s dt \Rightarrow X_t = X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)} dB_s.$$

Conditional on X_0 , the mean and covariance of $(X_t)_{t \geq 0}$ are given by

$$\mathbb{E}[X_t | X_0] = X_0 e^{-\lambda t}, \quad \text{Cov}(X_s, X_t | X_0) = \frac{1}{2\lambda} \left(e^{-\lambda|s-t|} - e^{-\lambda(s+t)} \right).$$

The generator of this process is given by

$$Lf(x) = \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(x) - \lambda x \frac{\partial f}{\partial x}(x), \quad f \in C_c^2(\mathbb{R}).$$

Moreover, if we set the starting law as $X_0 \sim N(0, \frac{1}{2\lambda})$, then the covariance function of $X = (X_t)_{t \geq 0}$ is given by $K(s, t) = \frac{1}{2\lambda} e^{-\lambda|s-t|}$. In this case, X is a stationary Gaussian process.

Example 6.23 (Geometric Brownian motion). Let $\sigma > 0$ and $\mu \in \mathbb{R}$. A *geometric Brownian motion* with parameters σ and μ is the solution of the following SDE:

$$dX_t = \sigma X_t dB_t + \mu X_t dt.$$

The quadratic variation of X is

$$X_t = X_0 + \int_0^t \sigma X_s dB_s + \int_0^t \mu X_s ds \Rightarrow \langle X, X \rangle_t = \sigma^2 \int_0^t X_s^2 ds.$$

Assume $X_0 > 0$. By applying Itô's formula to $\log X_t$, we have

$$\log X_t = \log X_0 + \int_0^t \frac{1}{X_s} dX_s - \frac{1}{2} \int_0^t \frac{1}{X_s^2} d\langle X, X \rangle_s = \log X_0 + \sigma B_t + \left(\mu - \frac{\sigma^2}{2} \right) t.$$

Therefore, starting from $X_0 > 0$, we have

$$X_t = X_0 \exp \left(\sigma B_t + \left(\mu - \frac{\sigma^2}{2} \right) t \right).$$

Note that $M_t = \exp\left(\sigma B_t - \frac{\sigma^2}{2}t\right)$ is a martingale. Conditional on X_0 , the mean and covariance of $(X_t)_{t \geq 0}$ are given by

$$\mathbb{E}[X_t|X_0] = X_0 e^{\mu t}, \quad \text{Cov}(X_s, X_t|X_0) = X_0^2 e^{\mu(s+t)} \left(e^{\sigma^2(s \wedge t)} - 1\right).$$

The generator of this process is given by

$$Lf(x) = \frac{\sigma^2}{2} x^2 \frac{\partial^2 f}{\partial x^2}(x) + \mu x \frac{\partial f}{\partial x}(x), \quad f \in C_c^2(\mathbb{R}).$$

Furthermore, for any partition $0 = t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n < \dots$, we have independence of the successive ratios:

$$\frac{X_1 - X_0}{X_0}, \frac{X_2 - X_1}{X_1}, \dots, \frac{X_n - X_{n-1}}{X_{n-1}}, \dots,$$

which is a consequence of independent increment property of Brownian motions.

Hermitian Adjoint of the Generator. Now we discuss the adjoint L^* of generator L in sense that

$$\int_{\mathbb{R}^p} Lf(x)g(x) dx = \int_{\mathbb{R}^p} f(x)L^*g(x) dx.$$

We use integration by parts. Since both f and g are compactly supported on \mathbb{R}^p , we can choose a common compact support $\Gamma \subset \mathbb{R}^p$. Then

$$\begin{aligned} \int_{\Gamma} \nabla f \cdot (bg) dm &= - \int_{\Gamma} f \nabla \cdot (bg) dm, \\ \int_{\Gamma} \nabla^2 f \cdot (\Sigma g) dm &= - \int_{\Gamma} \nabla f \cdot (\nabla \cdot (\Sigma g)) dm = \int_{\Gamma} f \nabla^2 \cdot (\Sigma g) dm. \end{aligned}$$

Consequently, we have

$$L^* = \frac{1}{2} \nabla^2 \cdot \Sigma - \nabla \cdot b \quad \Leftrightarrow \quad L^*g = \frac{1}{2} \sum_{i,j=1}^p \frac{\partial^2 ((\sigma\sigma^*)_{ij}g)}{\partial x_i \partial x_j} - \sum_{i=1}^p \frac{\partial (b_i g)}{\partial x_i}. \quad (6.13)$$

We denote by $\mathfrak{D}(L^*)$ the set of functions g such that (6.13) makes sense.

Theorem 6.24 (Fokker-Planck equation). Let $(X_t)_{t \geq 0}$ be a diffusion process with diffusion $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}$ and drift $b : \mathbb{R}^p \rightarrow \mathbb{R}$, and let ρ_0 be a probability density function on \mathbb{R}^p . Let $\rho(t, \cdot)$ be the probability density function of X_t for every $t \geq 0$. Then $\rho(\cdot, x)$ is differentiable on \mathbb{R}_+ for every $x \in \mathbb{R}^d$, $\rho(t, \cdot) \in \mathfrak{D}(L^*)$ for every $t \geq 0$, and ρ solves the following *Fokker-Planck equation*:

$$\begin{cases} \frac{\partial \rho}{\partial t}(t, x) = \frac{1}{2} \nabla_x^2 \cdot \Sigma(x) \rho(t, x) - \nabla_x \cdot b(x) \rho(t, x), & t > 0 \\ \rho(0, x) = \rho_0(x) \end{cases}$$

Proof. Let $q(\cdot|t, x)$ be the probability density function of $Q_t(\cdot, x)$, where $t > 0$. Let $f \in C_c^2(\mathbb{R}^p)$. Then we have

$$\lim_{t \downarrow 0} \frac{1}{t} \left(\int_{\mathbb{R}^p} f(y) q(y|t, x) dy - f(x) \right) = Lf(x) = \frac{1}{2} \Sigma(x) \cdot \nabla^2 f(x) + b(x) \cdot \nabla f(x), \quad \forall x \in \mathbb{R}^p.$$

For any $t > 0$, we use interchangeability of derivative and integration:

$$\begin{aligned}
\int_{\mathbb{R}^p} f(y) \frac{\partial q}{\partial t}(y|t, x) dy &= \frac{\partial}{\partial t} \int_{\mathbb{R}^p} f(y) q(y|t, x) dy \\
&= \lim_{s \downarrow 0} \frac{1}{s} \int_{\mathbb{R}^p} (q(y|t+s, x) - q(y|t, x)) f(y) dy \\
&= \lim_{s \downarrow 0} \frac{1}{s} \left(\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} q(z|t, x) q(y|s, z) f(y) dz dy - \int_{\mathbb{R}^p} q(z|t, x) f(z) dy \right) \\
&= \lim_{s \downarrow 0} \frac{1}{s} \int_{\mathbb{R}^p} q(z|t, x) \left(\int_{\mathbb{R}^p} q(y|s, z) f(y) dy - f(z) \right) dz \\
&= \int_{\mathbb{R}^p} q(z|t, x) Lf(z) dz = \int_{\mathbb{R}^p} L^* q(z|t, x) f(z) dz.
\end{aligned}$$

Here the third equality uses Chapman-Kolmogorov identity, and the fifth uses dominated convergence. Since the above equation holds for all $f \in C_c^2(\mathbb{R})$, one have the following result:

$$\frac{\partial q}{\partial t}(y|t, x) = L^* q(y|t, x) = \frac{1}{2} \nabla^2 \cdot \Sigma(z) q(y|t, x) - \nabla \cdot b(y) q(y|t, x). \quad (6.14)$$

Now assume that $X_0 \sim \rho_0$. Then we have

$$\rho(t, y) = \int_{\mathbb{R}^p} q(y|t, x) \rho_0(x) dx$$

By applying this integration to both sides of (6.14), one obtain the desired result. \square

References

- [1] Richard M. Dudley. (2002). *Real Analysis and Probability*. Cambridge University Press, Cambridge.
- [2] Achim Klenke. (2013). *Probability Theory: A Comprehensive Course, 2nd Edition*. Springer, Berlin.
- [3] Jean-François Le Gall. (2013). *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, Berlin.
- [4] Samuel Karlin and Howard M. Taylor. (1975). *A First Course in Stochastic Processes, 2nd Edition*. Academic Press, Inc., New York.
- [5] Jacques Neveu. (1964). *Bases Mathématiques du Calcul des Probabilités*. Masson et Cie, Paris.
- [6] Daniel Revuz and Marc Yor. (1999). *Continuous Martingales and Brownian Motion, 3rd Edition*. Springer, Berlin.
- [7] Giuseppe Da Prato. (2014). *Introduction to Stochastic Analysis and Malliavin Calculus*. Edizioni della Normale, Pisa.
- [8] Grigorios A. Pavliotis. (2014). *Stochastic Processes and Applications, Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer, Berlin.