Lecture Notes for Stochastic Analysis

Jyunyi Liao

Contents

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 $\mathscr{A} \subset 2^\Omega$ is called a collection of subsets of Ω.

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

(i) $\Omega \in \mathscr{F}$;

(ii) For all $A \in \mathscr{F}$, the *complement* $A^c := \Omega \backslash A \in \mathscr{F}$;

(iii) For all sequences $(A_n)_{n=1}^\infty$ in $\mathscr{F}$, $\bigcup_{n=1}^\infty A_n \in \mathscr{F}$.

A pair $(\Omega, \mathscr{F})$ consisting of a set Ω and a σ-algebra of subsets of Ω is called a *measurable space.* A set $A \in \mathscr{F}$ is said to be $\mathscr{F}$-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 $\mathscr{F}$ also satisfies the following:

(i) $\emptyset \in \mathscr{F}$;

(ii) For all $A, B \in \mathscr{F}$, $A \backslash B, A \cup B, A \cap B \in \mathscr{F}$;

(iii) For all sequences $(A_n)_{n=1}^\infty$ in $\mathscr{F}$, $\bigcap_{n=1}^\infty A_n \in \mathscr{F}$.

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

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

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

Remark. We can generate the minimal σ-algebra $\mathscr{F}$ from $\mathscr{A}$ as follows: (i) Complete $\mathscr{A}$: $\mathscr{F} \leftarrow \mathscr{A} \cup \{\Omega, \emptyset\}$; (ii) For all $A \in \mathscr{A}$, add A^c to $\mathscr{F}$ if necessary; (iii) For all sequences of sets in $\mathscr{F}$, include their union in $\mathscr{F}$.

Definition 1.3. Let $(X, \mathscr{T})$ 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 $\mathscr{B}(X) = \sigma(\mathscr{T})$.

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

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

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

(i) $\Omega \in \mathscr{L}$;

(ii) For all $A, B \in \mathscr{L}$ such that $A \subset B$, it holds $B \backslash A \in \mathscr{L}$;

(iii) For all increasing sequences $A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots$ in $\mathscr{L}$, it holds $\bigcup_{n=1}^\infty A_n \in \mathscr{L}$.

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

(i) $\Omega \in \mathscr{L}$;

(ii) For all $A \in \mathscr{L}$, it holds $A^c \in \mathscr{L}$;

(iii) For all sequences $(A_n)_{n=1}^\infty$ of disjoint sets in $\mathscr{L}$, it holds $\coprod_{n=1}^\infty A_n \in \mathscr{L}$.

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 $\mathscr{A}$ of subsets of Ω:

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

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 $\mathscr{L} = \{\Omega, \{0,1\}, \{2,3\}, \{0,2\}, \{1,3\}, \{1,2\}, \{0,3\}, \emptyset\}$.

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

Proof. By definition, a σ-algebra is a π-system and λ-system. Conversely, if $\mathscr{F}$ 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 $\mathscr{F}$. Define

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

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

$$B_n = A_n \cup B_{n-1} = \underbrace{(A_n \backslash C_n)}_{\in \mathscr{F}} \cup \underbrace{(B_{n-1} \backslash C_n)}_{\in \mathscr{F}} \cup C_n. \tag{1.1}$$

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

We introduce the Sierpiński-Dynkin π-λ theorem, which is a powerful tool in measure-theoretic analysis.

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

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

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

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

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

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

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

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

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

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

Now we introduce the monotone class theorem.

Definition 1.8 (Monotone class)**.** A collection $\mathscr{M}$ 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 $\mathscr{M}$, it holds $\bigcup_{n=1}^\infty A_n \in \mathscr{M}$; (ii) For all decreasing sequence $(B_n)_{n=1}^\infty$ in $\mathscr{M}$, it holds $\bigcap_{n=1}^\infty B_n \in \mathscr{M}$.

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 $\mathscr{A}$ of subsets of Ω:

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

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

Remark. By (ii) and (iii), we can show that an algebra $\mathscr{A}$ 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 $\mathscr{R}$ of subsets of Ω such that $A\backslash B, A \cap B, A \cup B \in \mathscr{R}$ for all $A, B \in \mathscr{R}$. Following this, an *algebra* is a ring that contain Ω.

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

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

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

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

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

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

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

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

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

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

1.2 Measures, Pre-measures and Carathéodory's Extension

1.2.1 Measure Spaces

Definition 1.12 (Measure)**.** Let $(\Omega, \mathscr{F})$ be a measurable space. A *(nonnegative) measure* μ on $(\Omega, \mathscr{F})$ is a function $\mu : \mathscr{F} \to \overline{\mathbb{R}}_+ := [0, \infty]$ 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 $\mathscr{F}$, then

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

The triple $(\Omega, \mathscr{F}, \mu)$ 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 a countable collection $\{\Omega_n\}_{n=1}^\infty \subset \mathscr{F}$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and $\mu(\Omega_n) < \infty$ for each $n \in \mathbb{N}$.

(iii) μ is called a *semi-finite measure* if every positive measure set E have a finite measure subset.

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

Remark. A measure μ also has the following properties:

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

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^n A_k\right) \le \lim_{n\to\infty} \sum_{k=1}^n \mu(A_k) = \sum_{n=1}^\infty \mu(A_n), \quad \forall \textit{sequences } (A_n)_{n=1}^\infty \textit{ in } \mathscr{F}.$$

We then discuss the limit property of measures.

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

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

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

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

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

(iii) More generally, if $(A_n)_{n=1}^\infty$ is a sequence in $\mathscr{F}$, then

$$\mu\left(\bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k\right) \le \liminf_{n\to\infty} \mu(A_n),$$

and if in addition $\mu(\bigcup_{n=1}^\infty A_n) < \infty$, then

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

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

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

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

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

Since $\mu(A_1) < \infty$, and since $\bigcup_{n=1}^{\infty} C_n = A_1 \backslash \left(\bigcap_{n=1}^{\infty} A_n\right)$, the identity becomes (1.2)

$$\mu(A_1) - \lim_{n\to\infty} \mu(A_n) = \mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

(iii) The set $B_n = \bigcap_{k=n}^{\infty} A_k$ is an increasing sequence in $\mathscr{F}$, and we have $\mu(B_n) \leq \inf_{n\geq k} \mu(A_k)$. By (i),

$$\mu\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} \mu(B_n) \leq \liminf_{n\to\infty} \mu(A_k).$$

Also, the set $B_n = \bigcup_{k=n}^{\infty} A_k$ is a decreasing sequence in $\mathscr{F}$, and we have $\mu(B_n) \geq \sup_{n\geq k} \mu(A_k)$. By (ii),

$$\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} \mu(B_n) \geq \limsup_{n\to\infty} \mu(A_k).$$

Then we conclude the proof. $\square$

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

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu(\emptyset) = 0.$$

The following theorem is often useful in measure-theoretic analysis.

Theorem 1.14 (First Borel-Cantelli Lemma). *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. If $(A_n)_{n=1}^{\infty}$ is a sequence of sets in $\mathscr{F}$ 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 $\mathscr{F}$, and $\mu(B_1) \leq \sum_{k=1}^{\infty} \mu(A_k) < \infty$. By Lemma 1.13,

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

1.2.2 Construction of Measures

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

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

(i) $\mu(\emptyset) = 0$;

(ii) (Countable additivity). If $(A_n)_{n=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathscr{A}$ with $\coprod_{n=1}^{\infty} A_n \in \mathscr{A}$, 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 $\mu^* : 2^{\Omega} \to \overline{\mathbb{R}}_+$ 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 $\mathscr{R}$ be a ring of subsets of Ω, and let $\mu : \mathscr{R} \to \overline{\mathbb{R}}_+$ be a pre-measure on $\mathscr{R}$. Define $\mu^* : 2^{\Omega} \to \overline{\mathbb{R}}_+$ by*

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

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

Proof. It is easy to check that μ^* is an outer measure. For all $E \in \mathscr{R}$, take $A_1 = E$ and $A_2 = A_3 = \cdots = \emptyset$. 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 $B_n = A_n \backslash \left(\bigcup_{k=1}^{n-1} B_k\right)$ for $n \geq 2$. Then $(B_n)_{n=1}^{\infty}$ is a disjoint sequence in $\mathscr{R}$, 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 $\mu^*(E) \geq \mu(E)$. □

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

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \backslash E),\ \forall A \subset \Omega. \tag{1.3}$$

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

Remark. To check (1.3), it suffices to check $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \backslash E)$, 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 $\mathscr{R}^*$ given in Definition 1.18 is a σ-algebra.*

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

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

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

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

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

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

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

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 $G \in \mathscr{R}^*$, as desired. (To show that decreasing sequences in $\mathscr{R}^*$ have their limits in $\mathscr{R}^*$, take the complement.) $\square$

Proposition 1.20. *μ^* is a measure on $(\Omega, \mathscr{R}^*)$.*

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

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

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

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

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

We can certainly assume $\mu^*(A) < \infty$. Then for all $\epsilon > 0$, there exists a sequence $(F_n)_{n=1}^\infty$ in $\mathscr{R}$ 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 $\mathscr{R}$ such that $G_1 = F_1$ and $G_n = F_n \backslash \left(\bigcup_{k=1}^{n-1} F_k\right)$ for all $n \geq 2$. Then $\bigcup_{n=1}^\infty G_n = \bigcup_{n=1}^\infty F_n \supset A$, and

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

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

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

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 $(\Omega, \mathscr{F})$. Let $\mathscr{P} \subset \mathscr{F}$ be a π-system such that $\sigma(\mathscr{P}) = \mathscr{F}$ and that $\mu|_{\mathscr{P}} = \nu|_{\mathscr{P}}$.*

(i) If $\mu(\Omega) = \nu(\Omega)$, then $\mu = \nu$;

(ii) If there exists an increasing sequence $(\Omega_n)_{n=1}^\infty$ of sets in $\mathscr{P}$ 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 $\mathscr{L} = \{A \in \mathscr{F} : \mu(A) = \nu(A)\}$. Then $\mathscr{L}$ is a λ-system that contains π-system $\mathscr{P}$. By Theorem 1.7, $\mathscr{F} \subset \sigma(\mathscr{P}) \subset \mathscr{L} \subset \mathscr{F}$. Hence $\mathscr{L} = \mathscr{F}$, as desired.

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

$$\mu(A) = \lim_{n\to\infty} \mu_n(A \cap \Omega_n) = \lim_{n\to\infty} \nu_n(A \cap \Omega_n) = \nu(A),\ \forall A \in \mathscr{F}.$$ □

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

Proof. Any ring $\mathscr{R}$ of subsets of Ω is a π-system. Apply Lemma 1.22. □

Remark. Define the collection $\mathscr{A}$ 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 $\mathscr{A}$ is an algebra. For each $A \in \mathscr{A}$, define $\ell(A)$ to be the length of A. Then $(\mathbb{R}, \mathscr{A}, \ell)$ 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 $\mathscr{S}$ of subsets of Ω such that for all $A, B \in \mathscr{S}$, there exists finite collection $\{A_k\}_{k=1}^n \subset \mathscr{S}$ of pairwise disjoint sets such that $A \backslash B = \coprod_{k=1}^n A_k$.

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

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

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

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

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

$$\left(\coprod_{k=1}^n A_k\right) \cap \left(\coprod_{l=1}^m B_l\right) = \bigcup_{k=1}^n \bigcup_{l=1}^m \underbrace{(A_k \cap B_l)}_{\in \mathscr{S}} \in \mathscr{R},\ \forall \textit{ disjoint } \{A_k\}_{k=1}^n, \{B_l\}_{l=1}^m \subset \mathscr{S},$$

$$\left(\coprod_{k=1}^n A_k\right) \backslash \left(\coprod_{l=1}^m B_l\right) = \bigcup_{k=1}^n \bigcap_{l=1}^m \underbrace{(A_k \backslash B_l)}_{\in \mathscr{R}} \in \mathscr{R},\ \forall \textit{ disjoint } \{A_k\}_{k=1}^n, \{B_l\}_{l=1}^m \subset \mathscr{S}.$$

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

$$\mu^*\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k),\ \forall \textit{ pairwise disjoint } A_1, \cdots, A_n \in \mathscr{S}.$$

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

1.2.3 Application: Construction of Product Measures

An application of measure extension theorem is the construction of product measures.

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

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

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

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

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

Proof. We use the Carathéodory's extension theorem to prove this. We check that (i) $\mathscr{F}_1 \times \mathscr{F}_2$ 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 $\mathscr{F}_1 \times \mathscr{F}_2$. If (i) and (ii) are satisfied, the existence of an extension on $\mathscr{F}_1 \otimes \mathscr{F}_2$ is ensured.

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

$$A \backslash B = \underbrace{((A_1 \cap B_1^c) \times A_2)}_{\in \mathscr{F}_1 \times \mathscr{F}_2} \amalg \underbrace{((A_1 \cap B_1) \times (A_2 \cap B_2^c))}_{\in \mathscr{F}_1 \times \mathscr{F}_2}$$

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

(ii) Clearly, $(\mu_1 \times \mu_2)(\emptyset) = 0$. Then we need to verify the countable additivity of μ. Let $E \times F \in \mathscr{F}_1 \times \mathscr{F}_2$, 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,

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

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 $\mathscr{F}_1 \times \mathscr{F}_2$.

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 $\{A_n\}_{n=1}^{\infty} \subset \mathscr{F}_1$ and $\{B_n\}_{n=1}^{\infty} \subset \mathscr{F}_2$ 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. In general, the set of measurable rectangles $\mathscr{F}_1 \times \mathscr{F}_2$ is not a σ-algebra, since it is possibly not closed under complements countable intersections. For example, consider $(\mathbb{R}^2, \mathscr{B}(\mathbb{R})^2)$, where

$$\mathscr{B}(\mathbb{R})^2 = \mathscr{B}(\mathbb{R}) \times \mathscr{B}(\mathbb{R})).$$

The union of $(0,1) \times (0,1)$ and $(-1,0) \times (-1,0)$ is not in $\mathscr{B}(\mathbb{R})^2$.

Product topology and product σ-algebra. Let $(X_1, \mathscr{T}_1)$ and $(X_2, \mathscr{T}_2)$ are two second-countable topological spaces. The *product topology* $\mathscr{T}_1 \otimes \mathscr{T}_2$ is the topology generated by all open rectangles $\mathscr{T}_1 \times \mathscr{T}_2$.

Let $\mathscr{B}_1 = \sigma(\mathscr{T}_1)$ and $\mathscr{B}_2 = \sigma(\mathscr{T}_2)$ be the Borel σ-algebras generated by $\mathscr{T}_1$ and $\mathscr{T}_2$, respectively. Then the σ-algebras generated by the product topology $\mathscr{T}_1 \otimes \mathscr{T}_2$ and by Borel rectangles $\mathscr{B}_1 \times \mathscr{B}_2$ coincide. In a nutshell, $\sigma(\mathscr{T}_1) \otimes \sigma(\mathscr{T}_2) = \sigma(\mathscr{T}_1 \otimes \mathscr{T}_2)$.

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

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 $\mathscr{A} = \{U_m \times V_n\}_{m,n\in\mathbb{N}}$ 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 $\mathscr{A}$ contains all open sets in $X_1 \times X_2$, and $\sigma(\mathscr{A}) \supset \sigma(\mathscr{T}_1 \otimes \mathscr{T}_2)$. On the other hand, note that $\mathscr{A} \subset \mathscr{T}_1 \times \mathscr{T}_2$, which is the set of all open rectangles in $X_1 \times X_2$, we have $\sigma(\mathscr{A}) \subset \sigma(\mathscr{T}_1 \times \mathscr{T}_2) \subset \sigma(\mathscr{T}_1 \otimes \mathscr{T}_2)$. Furthermore, since $\mathscr{A} \subset \mathscr{B}_1 \times \mathscr{B}_2$, we have $\sigma(\mathscr{A}) \subset \sigma(\mathscr{B}_1 \times \mathscr{B}_2)$.

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

Since the real line $\mathbb{R}$ given the standard topology is second-countable, we have $\mathscr{B}(\mathbb{R}^2) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$. The same conclusion applies for all Euclidean spaces $\mathbb{R}^n$, where $n \in \mathbb{N}$.

1.3 Measurable Functions and Lebesgue Integration

1.3.1 Measurable Functions

Definition 1.26 (Inverse image)**.** Given a function $T : \Omega_1 \to \Omega_2$ 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 $\mathscr{A}$ is a collection of subsets of Ω_2, define $T^{-1}\mathscr{A} = \{T^{-1}A : A \in \mathscr{A}\}$.

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

(i) $T^{-1}\Omega_2 = \Omega_1$, $T^{-1}\emptyset = \emptyset$;

(ii) For all $A \subset \Omega_2$, $T^{-1}(\Omega_2 \backslash A) = \Omega_1 \backslash T^{-1}A$.

(iii) If $\{A_\alpha\}_{\alpha \in J}$ is a collection of subsets of Ω_2, then $T^{-1}\left(\bigcup_{\alpha\in J} A_\alpha\right) = \bigcup_{\alpha \in J} T^{-1}A_\alpha$.

(iv) If $\mathscr{F}$ is a σ-algebra of subsets of Ω_2, then $T^{-1}\mathscr{F}$ is again a σ-algebra.

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

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

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

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

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

Definition 1.30. Let $(\Omega, \mathscr{F})$ be a measurable space.

(i) A real-valued function $f : \Omega \to \mathbb{R}$ is said to be *measurable* if $f^{-1}(B) \in \mathscr{F}$ for all $B \in \mathscr{B}(\mathbb{R})$. In other words, the function $f : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is measurable.

(ii) An extended real-valued function $f : \Omega \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is said to be *measurable* if the sets $\{\omega : f(\omega) = -\infty\}$ and $\{\omega : f(\omega) = \infty\}$ are measurable, and the real-valued function f_1 is measurable:

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

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

The measurability of a real-valued function can be characterized by its level sets.

Proposition 1.31 (Characterization of real-valued measurable functions)**.** *Let $(\Omega, \mathscr{F})$ be a measurable space, and $f : \Omega \to \mathbb{R}$. 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

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

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

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

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 $f : \Omega \to \overline{\mathbb{R}}$.

Definition 1.32. Given a function $f : \Omega \to \mathbb{R}$, 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 $(\Omega, \mathscr{F})$ 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

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

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, i.e. the operation $\infty - \infty$ are not allowed.)

The limit operation also preserves measurability.

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

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

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 $\alpha \in \mathbb{R}$,

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

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

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

are also measurable. $\square$

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

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

Theorem 1.35. *Let $(\Omega, \mathscr{F})$ be a measurable space. Let X and Y be two second-countable measurable spaces. A vector-valued function $f = (f_X, f_Y) : \Omega \to X \times Y$ 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 $\pi_X : X \times Y \to Y,\ (x, y) \mapsto x$. 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 $\{U_m\}_{n=1}^\infty$ 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:

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

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

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

1.3.2 Simple Function Approximation of Measurable Functions

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

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

Let $(\Omega, \mathscr{F})$ be a measurable space, and let $f : \Omega \to \mathbb{R}$ be a nonnegative measurable function. Then there exists an increasing sequence $(\varphi_n)_{n=1}^\infty$ of measurable functions such that $f(\omega) = \lim_{n\to\infty} \varphi_n(\omega)$ for all $\omega \in \Omega$. Namely, φ_n converges pointwise to f. Furthermore, if there exists $M > 0$ such that $f(\omega) \le M$ 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 $0 \le k < 4^n$, define

$$E_{n,k} = \left\{\omega : 2^{-n}k \le f(\omega) < 2^{-n}(k+1)\right\},\ E_{n,4^n} = \{\omega : f(\omega) \ge 2^n\}.$$

Then choose a nonnegative measurable simple function φ_n as follows:

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

Clearly, φ_n is increasing, and $\varphi_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$. If there exists $M > 0$ such that $f(\omega) \le M$ 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 $\omega \in \Omega$. □

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 (f_n) for a general measurable function f, with $|f_n| \uparrow |f|$.

The following theorem shows that a pointwise convergent function sequence almost converges uniformly. It is also known as the second statement of the Littlewood's three principles for real analysis.

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

Proof. Choose $\Omega_0 \in \mathscr{F}$ such that $\mu(\Omega\backslash\Omega_0) = 0$ and $f_n(x) \to f(x)$ 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 $\bigcup_{k=1}^{\infty} A_k \subset \Omega\backslash\Omega_0$. Since μ is finite, we have

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

Let $E = \Omega\backslash\left(\bigcup_{k=1}^{\infty} B_{k,N_k}\right) \in \mathscr{F}$. Then $\mu(\Omega\backslash E) < \epsilon$. 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. □

We also have a monotone class theorem for measurable functions.

Theorem 1.38 (Monotone class theorem). *Let $\mathscr{A}$ be a π-system that contains Ω, and let $\mathcal{H}$ be a collection of real-valued functions on Ω that satisfies:*

(i) $\{\mathbb{1}_A : A \in \mathscr{A}\} \subset \mathcal{H}$;

(ii) $\mathcal{H}$ is closed under linear operation, i.e. for all $f, g \in \mathcal{H}$ and $c \in \mathbb{R}$, we have $f + g, cf \in \mathcal{H}$.

(iii) If $f_n \in \mathcal{H}$ are nonnegative and increase to a function f, then $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all bounded functions that are measurable with respect to $\sigma(\mathscr{A})$.

Proof. We define

$$\mathscr{H} = \{A \subset \Omega : \mathbb{1}_A \in \mathscr{H}\}.$$

By the assumptions $\Omega \in \mathscr{A}$, (ii) and (iii), $\mathscr{H}$ is a λ-system, which contains $\sigma(\mathscr{A})$ by the π-λ theorem. By (ii), $\mathcal{H}$ contains all simple functions, and (iii) implies that $\mathcal{H}$ contains all bounded functions that are measurable with respect to $\sigma(\mathscr{A})$ by simple function approximation. □

1.3.3 Lebesgue Integration: Nonnegative Measurable Functions

Definition 1.39 (Lebesgue integral for nonnegative measurable functions). A simple function $\varphi : \Omega \to \mathbb{R}$ takes only finitely many values $a_1, \cdots, a_n \in \mathbb{R}$. Hence it has the unique *standard expression*:

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

For a nonnegative simple function $\varphi : \Omega \to \mathbb{R}_+$ defined by (1.6), we define its *Lebesgue integral* as

$$\int \varphi \, \mathrm{d}\mu = \sum_{k=1}^{n} a_n \mu(A_n).$$

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

$$\int f \, \mathrm{d}\mu = \sup\left\{\int \varphi \, \mathrm{d}\mu : 0 \leq \varphi \leq f, \; \varphi \textit{ is a measurable simple function}\right\}.$$

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

$$\int_A f \, \mathrm{d}\mu = \int f \chi_A \, \mathrm{d}\mu.$$

Remark. These integrals are well-defined but may take value ∞, with the convention $\infty \cdot 0 = 0$.

The monotonicity is an important property of Lebesgue integrals.

Proposition 1.40 (Monotonicity)**.** *If f and g are nonnegative measurable functions such that $f \leq g$, then*

$$\int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu.$$

Proof. If φ is a simple function with $0 \leq \varphi \leq f$, we also have $0 \leq \varphi \leq g$, and $S(\varphi) \leq \int g \, d\mu$. By taking the supremum over all such φ, we get the desired inequality. $\square$

Following the monotone property, we introduce one of the most important convergence theorems, which ensure the interchangeability of limit and integration.

Theorem 1.41 (Monotone convergence theorem/Levi's theorem)**.** *Let $(f_n)_{n=1}^{\infty}$ be a monotone increasing sequence of nonnegative measurable functions, and let $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ for all $\omega \in \Omega$. Then*

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

Proof. By Proposition 1.40, $\int f_n \, \mathrm{d}\mu$ is monotone increasing, and

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \sup_{n\geq 1} \int f_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu. \tag{1.7}$$

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 $\mathscr{F}$ such that $\Omega = \bigcup_{n=1}^{\infty} A_n$. Note that f_n is nonnegative, and φ is simple. Then

$$\int f_n \, \mathrm{d}\mu \geq \int_{A_n} f_n \, \mathrm{d}\mu \geq \alpha \int_{A_n} \varphi \, \mathrm{d}\mu$$

Letting $n \to \infty$ and then $\alpha \uparrow 1$, we have

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu \geq \alpha \int \varphi \, \mathrm{d}\mu, \quad \text{and} \quad \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu \geq \int \varphi \, \mathrm{d}\mu.$$

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). $\square$

Many properties of the Lebesgue integral can be proved by applying simple function approximation and monotone convergence theorem.

Proposition 1.42. *For all nonnegative measurable functions f and g and all $\alpha, \beta \in \mathbb{R}_+$,*

$$\int (\alpha f + \beta g) \, \mathrm{d}\mu = \alpha \int f \, \mathrm{d}\mu + \beta \int g \, \mathrm{d}\mu.$$

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

Proposition 1.43. *Let f and g be nonnegative measurable functions. Then*

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

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

$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu.$$

Proof. Let $\varphi = \sum_{k=1}^n a_k \chi_{A_k}$ be a simple function such that $0 \le \varphi \le f$. 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, $\int f \,\mathrm{d}\mu = 0$.

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

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

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, $\int h \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$. Similarly, we have $\int h \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$. □

Now we introduce the second important convergence theorem. The Fatou's lemma is useful when we do not know whether limit and integration are interchangeable.

Theorem 1.44 (Fatou's lemma)**.** *Let $(f_n)_{n=1}^\infty$ be a sequence of nonnegative measurable functions, and let $f(\omega) = \liminf_{n\to\infty} f_n(\omega)$ for all $\omega \in \Omega$. Then*

$$\int f \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu. \tag{1.8}$$

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

$$\int g_n \,\mathrm{d}\mu \le \int f_k \,\mathrm{d}\mu, \quad \forall k \ge n,$$

which implies

$$\int g_n \,\mathrm{d}\mu \le \inf_{k\ge n} \int f_k \,\mathrm{d}\mu.$$

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

$$\int f \,\mathrm{d}\mu = \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu \le \lim_{n\to\infty} \inf_{k\ge n} \int f_k \,\mathrm{d}\mu.$$

This is indeed the inequality (1.8). □

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

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

We will discuss a sufficient condition of interchangeability between limit and integration later in the Lebesgue dominated convergence theorem.

1.3.4 Integrable Functions and Lebesgue Integration

In this section, we extend the definition of Lebesgue integral to signed measurable functions.

Definition 1.45 (Lebesgue integrable functions)**.** A measurable function f is said to be *integrable* if

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

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

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

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

$$\int_A f \,\mathrm{d}\mu = \int f\chi_A \,\mathrm{d}\mu.$$

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

Proposition 1.46 (Linearity of Lebesgue integral)**.** *For all $f, g \in L^1(\Omega, \mathscr{F}, \mu)$ and all $\alpha, \beta \in \mathbb{R}$,*

$$\int (\alpha f + \beta g) \,\mathrm{d}\mu = \alpha \int f \,\mathrm{d}\mu + \beta \int g \,\mathrm{d}\mu.$$

Proof. By Proposition 1.42,

$$\int \alpha f \,\mathrm{d}\mu = \int (\alpha f)^+ \,\mathrm{d}\mu - \int (\alpha f)^- \,\mathrm{d}\mu = \alpha \int f^+ \,\mathrm{d}\mu - \alpha \int f^- \,\mathrm{d}\mu = \alpha \int f \,\mathrm{d}\mu.$$

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

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

□

Proposition 1.47 (Absolute integrability)**.** *Let f be a measurable function. Then $f \in L^1(\Omega, \mathscr{F}, \mu)$ if and only if $|f| \in L^1(\Omega, \mathscr{F}, \mu)$. In that case,*

$$\left| \int f \,\mathrm{d}\mu \right| \leq \int |f| \,\mathrm{d}\mu. \tag{1.9}$$

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. □

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

Theorem 1.48 (Lebesgue's dominated convergence theorem)**.** *Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions such that $f_n \to f$ a.e., where f is also a measurable function. If there exists $g \in L^1(\Omega, \mathscr{F}, \mu)$ such that $|f_n| \le g$ for all $n \in \mathbb{N}$, then all functions f_n and f are integrable, and*

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

Proof. We may assume $f_n \to f$ pointwise by redefining f and f_n on a set of measure zero. For all $n \in \mathbb{N}$, we have $|f_n| \le g$, then $|f| \le g$. By Proposition 1.47, all functions f_n and f are integrable. Since $g + f_n \ge 0$, and $g - f_n \ge 0$, we apply Fatou's lemma [Theorem 1.44] to obtain

$$\int (g+f) \, \mathrm{d}\mu \le \liminf_{n\to\infty} \int (g+f_n) \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu + \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu,$$

and

$$\int (g-f) \, \mathrm{d}\mu \le \liminf_{n\to\infty} \int (g-f_n) \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu - \limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

Hence we have

$$\limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu \le \int f \, \mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu,$$

and the result follows. □

Remark. If $(\Omega, \mathscr{F}, \mu)$ 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 1.49 (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 $f_n \ge g$ for all $n \in \mathbb{N}$, then

$$\int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu$$

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

$$\limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu \le \int \limsup_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Finally we discuss integral transform among different measure spaces.

Theorem 1.50 (Integral transform)**.** *Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be two measure spaces. If function $T : (\Omega_1, \mathscr{F}_1, \mu_1) \to (\Omega_2, \mathscr{F}_2, \mu_2)$ is measure-preserving, i.e. $\mu_2 = T_*\mu_1$ is the pushforward of μ_1, then*

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

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

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

Then (1.10) holds for all nonnegative measurable simple functions f. Similar to the procedure of defining Lebesgue integral (Definition 1.39 and Definition 1.45), it holds for all $f \in L^1(\Omega_2, \mathscr{F}_2, \mu_2)$. □

1.3.5 Integration on Product Spaces and Fubini's Theorem

Now we discuss Lebesgue integral on product spaces and the interchange of integrals. We first present the general conclusion in Theorem 1.51.

Theorem 1.51 (Fubini's theorem). *Let $f \in L^1(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \otimes \mu_2)$.*

(i) For all $\omega_1 \in \Omega_1$, the function $\omega_2 \mapsto f(\omega_1, \omega_2)$ is integrable.

(ii) The function $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2)$ is integrable.

(iii) For all $\omega_2 \in \Omega_1$, the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is integrable.

(iv) The function $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\,\mathrm{d}\mu_1(\omega_1)$ is integrable.

(v) The following integrals are equivalent:

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

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

Proposition 1.52. *Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be σ-finite measure spaces, and denote their product space by $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \otimes \mu_2)$. For all $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$, we define the slices*

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

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

(i) $A_{\omega_1} \in \mathscr{F}_2$ for all $\omega_1 \in \Omega_1$;

(ii) The function $f_A : \omega_1 \mapsto \mu_2(A_{\omega_1})$ is measurable;

(iii) $(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_1} \mu_2(A_{\omega_1})\,\mathrm{d}\mu_1(\omega_1) := \int_{\Omega_1} f_A\,\mathrm{d}\mu_1$.

A similar statement also holds for slices $A^{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$.

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

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

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

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

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

$$\mu_2(A_{\omega_1}) = \lim_{n\to\infty} \mu_2\left((A_n)_{\omega_1}\right) \quad \Rightarrow \quad f_A = \lim_{n\to\infty} f_n \textit{ is measurable.}$$

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

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

Step II: Similar to Step I, we can prove that for any decreasing sequence $(B_n)_{n=1}^\infty$ of sets in $\mathscr{M}$ 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 $B := \bigcap_{n=1}^\infty B_n \in \mathscr{M}$.

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

Step IV: Denote by $\mathscr{A}$ the collection of all finite unions of measurable rectangles in $\Omega_1 \times \Omega_2$. Then $\mathscr{A}$ 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 $\mathscr{A}_n = \{A \in \mathscr{A} : A \subset X_n \times Y_n\}$, and $\mathscr{M}_n = \{M \in \mathscr{M} : M \subset X_n \times Y_n\}$. Then $\mathscr{A}_n \subset \mathscr{M}_n$. Clearly, $\mathscr{A}_n$ is an algebra, and $\mathscr{M}_n$ is a monotone class by Steps I and II. By monotone class theorem, $\sigma(\mathscr{A}_n) \subset \mathscr{M}_n$. Since $\sigma(\mathscr{A}_n)$ contains all measurable subsets of $X_n \times Y_n$, so does $\mathscr{M}$. As a result, $\mathscr{M} \supset \mathscr{M}_n$ contains all measurable subsets of $X_n \times Y_n$ for all $n \in \mathbb{N}$. Since $\mathscr{M}$ is closed under countable unions, $\mathscr{M} \supset \mathscr{F}_1 \otimes \mathscr{F}_2$. $\square$

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

Theorem 1.53 (Tonelli's theorem)**.** *Let $f : (\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \otimes \mu_2) \to \overline{\mathbb{R}}_+$ 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 $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, \mathrm{d}\mu_2(\omega_2)$ is measurable.

(iii) For all $\omega_2 \in \Omega_1$, the function $\omega_1 \mapsto f(\omega_1, \omega_2)$ is measurable.

(iv) The function $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\, \mathrm{d}\mu_1(\omega_1)$ is measurable.

(v) The following integrals are equivalent:

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

Proof. By Theorem 1.52, the theorem holds for all indicator functions χ_A, where $A \in \mathscr{F}_1 \times \mathscr{F}_2$. 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 $f = \lim_{n\to\infty} \varphi_n$. 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 $f \notin L^1(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \otimes \mu_2)$, we cannot change the order of integration. For example, consider the function on $f : [0,1] \times [0,1] \to \overline{\mathbb{R}}$:

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

Then

$$\int_{[0,1]} \left(\int_{[0,1]} \frac{x^2 - y^2}{(x^2+y^2)^2} \, \mathrm{d}y \right) \mathrm{d}x = \int_{[0,1]} \frac{1}{1+x^2} \, \mathrm{d}x = \frac{\pi}{4};$$
$$\int_{[0,1]} \left(\int_{[0,1]} \frac{x^2 - y^2}{(x^2+y^2)^2} \, \mathrm{d}x \right) \mathrm{d}y = \int_{[0,1]} \frac{-1}{1+y^2} \, \mathrm{d}y = -\frac{\pi}{4};$$

and

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

1.4 Signed Measures, Jordan Decomposition and Radon-Nikodym Theorem

Definition 1.54 (Signed measure)**.** Let $(\Omega, \mathscr{F})$ be a measurable space. A *signed measure* μ on $(\Omega, \mathscr{F})$ is a set function $\mu : \mathscr{F} \to \overline{\mathbb{R}}$ that satisfies the following:

(i) $\mu(\emptyset) = 0$;

(ii) (Countable additivity). If $(A_n)_{n=1}^{\infty}$ is a sequence of disjoint sets in $\mathscr{F}$, then

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

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

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 $n \in \mathbb{N}$.

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.

Theorem 1.55 (Hahn decomposition theorem)**.** *Let μ be a signed measure on a measurable space $(\Omega, \mathscr{F})$. Then there exist measurable sets $P, N \in \mathscr{F}$ 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 $A \in \mathscr{F}$ with $A \subset P$, $\mu(A) \geq 0$. (In other words, P is a positive set.)

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

We prove that $N := \Omega \backslash P$ 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, $P \cup E \in \mathscr{P}$, but $\mu(P \cup E) = \mu(P) + \mu(E) > M = \sup_{A \in \mathscr{P}} \mu(A)$, a contradiction!) Hence there exists $B \subset E$ with $\mu(B) < 0$. We choose the smallest positive integer k_1 such that there exists $B_1 \subset E$ 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, $E \backslash B_1$ is not positive. Then we choose the smallest $k_2 \in \mathbb{N}$ such that there exists $B_2 \subset E \backslash B_1$ with $\mu(B_2) < -k_2^{-1}$. Repeat this procedure, we obtain a sequence $k_n \in \mathbb{N}$ and $B_n \subset \mathscr{F}$ such that

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

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

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

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

Remark. We called a set $E \in \mathscr{F}$ 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 $P \triangle P' = N \triangle N' = (P \cap N') \cup (N \cap P')$is a μ-null set.

Corollary 1.56 (Jordan decomposition). *Given a signed measure μ on a measurable space $(\Omega, \mathscr{F})$, take the Hahn decomposition $\Omega = P \amalg N$ on μ. Define*

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

Then μ^+ and μ^- are two (finite) measures on $(\Omega, \mathscr{F})$, 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 1.57. Let μ be a measure on a measurable space $(\Omega, \mathscr{F})$.

(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 $A \in \mathscr{F}$ 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 $A \in \mathscr{F}$ such that $\mu(A) = 0$ and $\nu(\Omega \backslash A) = 0$.

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

Theorem 1.58. *Let $f \in L^1(\Omega, \mathscr{F}, \mu)$. Define $\nu : \mathscr{F} \to \mathbb{R}$ by*

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

Then ν is a (finite) signed measure on $(\Omega, \mathscr{F})$. Furthermore, for any $\epsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathscr{F}$ 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 $\mathscr{F}$. 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,

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

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

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

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

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

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 1.58 also holds true. It is the generalization of the fundamental theorem of calculus on measures, known as Radon-Nikodym theorem.

Theorem 1.59 (Radon-Nikodym theorem). *Let μ and ν be two σ-finite measures defined on a measurable space $(\Omega, \mathscr{F})$. If $\nu \ll \mu$, then there exists a nonnegative measurable function $f : \Omega \to \mathbb{R}_+$ such that*

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

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 $f : \Omega \to \overline{\mathbb{R}}$ such that (i) $f \geq 0$ *a.e.*, and (ii) $\int_A f \, \mathrm{d}\mu \leq \nu(A)$ for all $A \in \mathscr{F}$. Then $f \equiv 0$ is in $\mathcal{F}$, and $\mathcal{F}$ is closed under finite maximum:

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

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

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

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

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

Using Hahn decomposition theorem on signed measure $\nu - \nu'$, there exists a positive set $P \in \mathscr{F}$ 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 $\nu(P \backslash A) - \nu'(P \backslash A) > \delta$.) Then

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

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

$$0 \leq \nu(A) - \int_A g \, \mathrm{d}\mu = \int_{\Omega \backslash A} g \, \mathrm{d}\mu - \nu(\Omega \backslash A) \leq 0, \ \forall A \in \mathscr{F}.$$

Note that $g : \Omega \to \overline{\mathbb{R}}$ is integrable. The set $E = \{\omega : g(\omega) = \infty\}$ is μ-null. Choose $f = g\chi_{\Omega \backslash E}$, 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 $f_n : \Omega_n \to \mathbb{R}_+$ such that

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

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

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

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

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

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

Corollary 1.60. *Let μ (resp. ν) be a σ-finite measure (resp. finite signed measure) on a measurable space $(\Omega, \mathscr{F})$. If $\nu \ll \mu$, then there exists $f \in L^1(\Omega, \mathscr{F}, \mu)$ such that*

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

*The **Radon-Nikodym derivative** $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} := f$ 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 $g, h : \Omega \to \mathbb{R}_+$ such that

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

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

Theorem 1.61 (Lebesgue decomposition theorem)**.** *Let μ and ν be two σ-finite measures on a measurable space $(\Omega, \mathscr{F})$. Then there exist unique measures $\nu_0 \ll \mu$ and $\nu_1 \perp \mu$ such that $\nu = \nu_0 + \nu_1$.*

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

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

Let $E = \{\omega : f(\omega) = 0\}$, and define $\nu_1(A) = \nu(A \cap E)$, $\nu_0 = \nu(A \cap E^c)$ for all $A \in \mathscr{F}$. Clearly, $\nu_1 \perp \mu$, since $\nu_1(X \backslash E) = \nu_1(\emptyset) = 0 = \mu(E)$.

It remains to show $\nu_0 \ll \mu$. If $\mu(A) = 0$, we fix any $n \in \mathbb{N}$ and let $B_n = \{\omega \in A : f(\omega) > n^{-1}\}$. Then

$$0 \le \lambda(B_n) \le n \int_{B_n} f\,d\lambda \le n \int_A f\,d\lambda = n\mu(A) = 0.$$

Then the set $B = A \cap E^c = \bigcup_{n=1}^\infty B_n$ has measure zero. Hence $0 \le \nu_0(A) = \nu(B) \le \lambda(B) = 0$.

Finally we prove uniqueness. If $\nu = \nu_0' + \nu_1'$ with $\nu_0' \ll \mu$ and $\nu_1' \perp \mu$, there exists $E' \in \mathscr{F}$ such that $\nu_1'(X \backslash E') = \mu(E') = 0$. Then for all measurable $A \subset X \backslash (E \cup E')$, 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 $\nu_0' = \nu_0$. □

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

Corollary 1.62 (Lebesgue)**.** *Let μ (resp. ν) be a σ-finite measure (resp. σ-finite signed measure) on $(\Omega, \mathscr{F})$. 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 $\nu_0 \ll \mu$.

1.5 Convergence of Measurable Functions and Measures

1.5.1 Convergence in Measure

Definition 1.63 (Cauchy sequence in measure)**.** Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions on $(\Omega, \mathscr{F}, \mu)$. 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.*

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

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

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.*

The proof of this theorem makes use of a powerful subsequence lemma.

Lemma 1.66. *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

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

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 $\mu(E) = \lim_{N\to\infty} \mu(F_N) = 0$. For each $\omega \in \Omega\backslash E$, 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, $f_n \to f$ *a.e.*, which is measurable. $\square$

Proof of Theorem 1.65. Given $\epsilon > 0$ and $\eta > 0$. If $f_n \xrightarrow{\mu} f$, there exists N such that $\mu\left(|f_n - f| \geq \eta/2\right) < \epsilon/2$ 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 1.66, one of its subsequence f_{n_k} converges a.e. to a measurable function f. Furthermore, if we choose F_k in Lemma 1.66, for all $k \geq N$, we have

$$|f_{n_k}(\omega) - f(\omega)| \leq \sum_{l=k}^{\infty} |f_l(\omega) - f_{l+1}(\omega)| \leq \frac{1}{2^{k-1}}, \ \forall \omega \in \Omega\backslash F_k,$$

which implies

$$\mu\left(|f_{n_k} - f| > \frac{1}{2^{k-1}}\right) \leq \mu(F_k) < \frac{1}{2^{k-1}}.$$

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

$$\mu(|f_n - f| \geq \eta) \leq \mu\left(\left\{|f_n - f_{n_k}| \geq \frac{\eta}{2}\right\} \cup \left\{|f_{n_k} - f| \geq \frac{\eta}{2}\right\}\right) < \epsilon$$

for all $n \geq \max\{n_k, N\}$. Therefore $f_n \xrightarrow{\mu} f$. $\square$

The following theorem shows that a pointwise convergent function sequence also converges in measure.

Theorem 1.67 (Egoroff). *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space. If a sequence of functions (f_n) converges to a function f μ-a.e., then $f_n \xrightarrow{\mu} f$.*

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

For finite measure spaces, the condition of almost sure convergence in Fatou's lemma and Lebesgue dominated convergence theorem can be replaced by convergence in measure.

Theorem 1.68. *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space.*

(i) (Fatou's lemma). If (f_n) is a sequence of nonnegative measurable functions such that $f_n \xrightarrow{\mu} f$, then

$$\int_\Omega f \, d\mu \leq \liminf_{n \to \infty} \int_\Omega f_n \, d\mu. \tag{1.1}$$

(ii) (Dominated convergence theorem). If (f_n) is a sequence of integrable functions such that $f_n \xrightarrow{\mu} f$, and there exists an integrable function g such that $|f_n| \leq g$ for each $n \in \mathbb{N}$, then

$$\int_\Omega f \, d\mu = \lim_{n \to \infty} \int_\Omega f_n \, d\mu. \tag{1.2}$$

Proof. (i) By the very definition of limit infimum, we take a subsequence (f_{n_k}) such that

$$\lim_{k \to \infty} \int_\Omega f_{n_k} \, d\mu = \liminf_{n \to \infty} \int_\Omega f_n \, d\mu,$$

and $f_{n_k} \xrightarrow{\mu} f$ still. By Lemma 1.66, we may further assume $f_{n_k} \to f$ a.e. by passing to a further subsequence. The inequality (1.1) then follows from the classical Fatou's lemma [Theorem 1.44].

(ii) The result follows by applying (i) on sequences $g - f_n$ and $g + f_n$. □

Remark. As we will see in Theorem 1.75, we can further weaken the condition that (f_n) is dominated by an integrable function.

Finally, we study the continuous transformation of μ-convergent sequences.

Theorem 1.69 (Continuous mapping). *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space, and $\phi : \mathbb{R} \to \mathbb{R}$ a continuous function. If $f_n : \Omega \to \mathbb{R}$ is a sequence of measurable functions that converges in measure μ to f, the transformed sequence $(\phi \circ f_n)_{n=1}^\infty$ also converges in measure μ, and the limit equals $\phi \circ f$.*

Proof. Fix $\eta > 0$. For each $k \in \mathbb{N}$, define

$$E_k = \left\{ x \in \mathbb{R} : \text{there exists } y \in \mathbb{R} \text{ such that } |y - x| \leq \frac{1}{k} \text{ and } |\phi(y) - \phi(x)| \geq \eta \right\}.$$

Since ϕ is continuous, the sequence $E_k \downarrow \emptyset$, and $\mu(E_k) \downarrow 0$. Then for each $k \in \mathbb{N}$,

$$\begin{aligned} \mu\left(|\phi \circ f_n - \phi \circ f| \geq \eta\right) &= \mu\left(|f - f_n| \leq \frac{1}{k},\ |\phi \circ f_n - \phi \circ f| \geq \eta\right) + \mu\left(|f - f_n| > \frac{1}{k},\ |\phi \circ f_n - \phi \circ f| \geq \eta\right) \\ &\leq \mu(E_k) + \mu\left(|f - f_n| > \frac{1}{k}\right) \to \mu(E_k), \quad \text{as } n \to \infty. \end{aligned}$$

Since μ is finite, we let $k \to \infty$ and apply Proposition 1.13 to get the desired result. □

1.5.2 L^p Convergence and Uniform Integrability

For completeness of our discussion, we give a brief review of L^p spaces.

Definition 1.70 (L^p-spaces)**.** Let $(\Omega, \mathscr{F}, \mu)$ be a measurable space. For $1 \le p < \infty$, define $\mathcal{L}^p(\Omega, \mathscr{F}, \mu)$ 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

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

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

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

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 $L^p(\Omega, \mathscr{F}, \mu)$. Then the space $(L^p(\Omega, \mathscr{F}, \mu), \|\cdot\|_p)$ is a normed space.

Theorem 1.71 (Chebyshev inequality)**.** *Let $1 \le p < \infty$, and $f \in L^p(\Omega, \mathscr{F}, \mu)$. Then*

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

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

$$\|f\|_p^p = \int |f|^p \, \mathrm{d}\mu \ge \int_E |f|^p \, \mathrm{d}\mu \ge \eta^p \mu(E), \ \forall \eta > 0.$$ □

Remark. As a result of Theorem 1.71, the convergence in L^p-norm implies the convergence in measure.

Theorem 1.72 (Riesz-Fisher)**.** *$L^p(\Omega, \mathscr{F}, \mu)$ is a Banach space. That is, every Cauchy sequence (f_n) in $L^p(\Omega, \mathscr{F}, \mu)$ converges in L^p norm to a function in $L^p(\Omega, \mathscr{F}, \mu)$.*

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 $n, m \ge N$. By Fatou's lemma,

$$\int |f - f_m|^p \, \mathrm{d}\mu = \int \lim_{k \to \infty} |f_{n_k} - f_m|^p \, \mathrm{d}\mu \le \liminf_{k \to \infty} \int |f_{n_k} - f_m|^p \, \mathrm{d}\mu \le \epsilon^p, \ \forall m \ge N.$$

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

Now we introduce uniform integrability of function classes.

Definition 1.73. (Uniform Integrability). Let $(X, \mathscr{F}, \mu)$ be a finite measure space. A collection of integrable functions $\mathcal{F} \subset L^1(\Omega, \mathscr{F}, \mu)$ is said to be *uniformly integrable*, if

$$\lim_{N \to \infty} \sup_{f \in \mathcal{F}} \int_{\{|f| > N\}} |f| \, \mathrm{d}\mu = 0. \tag{1.3}$$

Remark. If g is an integrable function such that $|f| \le g$ for all $f \in \mathcal{F}$, by dominated convergence theorem,

$$\int_\Omega |f| \, \mathbb{1}_{\{|f|>N\}} \, d\mu \le \int_\Omega g \, \mathbb{1}_{\{g>N\}} \, d\mu \to \int_\Omega g \, \mathbb{1}_{\{g=\infty\}} \, d\mu = 0 \quad \text{as} \quad N \to \infty.$$

Therefore, $\mathcal{F}$ is uniformly integrable if it is dominated by an integrable function g.

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

Theorem 1.74. *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space. A collection of functions $\mathcal{F} \subset L^1(\Omega, \mathscr{F}, \mu)$ 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 $A \in \mathscr{F}$ with $\mu(A) < \delta$,

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

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

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

Then we can verify that $\mathcal{F}$ satisfies (i) and (ii):

(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 $\sup_{f\in\mathcal{F}} \int_{\{|f|>N\}} |f| \, \mathrm{d}\mu < \epsilon/2$ and $\delta = \frac{\epsilon}{2N}$.

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

$$\sup_{f\in\mathcal{F}} \mu(|f| \geq N) \leq \frac{1}{N} \sup_{f\in\mathcal{F}} \|f\|_1 \to 0 \quad \text{as} \quad N \to \infty.$$

Given $\epsilon > 0$, we choose the δ specified in (ii), and choose N_0 such that $\mu(|f| \geq N) < \delta$ for all $N \geq N_0$ and all $f \in \mathcal{F}$. By uniform absolute continuity of $\mathcal{F}$, we have $\sup_{f\in\mathcal{F}} \int_{\{|f|>N\}} |f| \, \mathrm{d}\mu < \epsilon$ for all $N \geq N_0$. Since ϵ is arbitrary, $\mathcal{F}$ is uniformly integrable. ☐

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

Theorem 1.75. *Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space, and $f_n \in L^1(\Omega, \mathscr{F}, \mu)$. Let $(f_n)_{n=1}^\infty$ be a sequence of integrable functions that converges to f in measure μ. The following are equivalent:*

(i) $(f_n)_{n=1}^\infty$ is uniformly integrable;

(ii) $\lim_{n\to\infty} \|f_n - f\|_1 = 0$; *(iii)* $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$.

Proof. We prove (i) ⇒ (ii) ⇒ (iii) ⇒ (i). The statement (ii) ⇒ (iii) is trivial.

(i) ⇒ (ii). Given $N > 0$, we define $f^N = \max\{-N, \min\{f, N\}\}$. Then

$$|f_n - f| \leq |f_n - f_n^N| + |f_n^N - f^N| + |f^N - f|.$$

We fix $N > 0$. By continuous mapping theorem [Theorem 1.69], we have $f_n^N \xrightarrow{\mu} f^N$. Since $|f_n^N - f^N| \leq 2N$, by dominated convergence [Theorem 1.68], $\|f_n^N - f^N\|_1 \to 0$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \|f_n - f\|_1 \leq \sup_{n\in\mathbb{N}} \|f_n^N - f_n^N\|_1 + \|f_n^N - f^N\|_1 + \|f^N - f\|_1 \leq \sup_{n\in\mathbb{N}} \|f_n^N - f_n^N\|_1 + \|f^N - f\|_1.$$

Now we control the remaining two terms. By uniform integrability of (f_n),

$$\sup_{n\in\mathbb{N}} \|f_n^N - f_n^N\|_1 \leq \sup_{n\in\mathbb{N}} \int_{\{|f_n|>N\}} |f_n| \, \mathrm{d}\mu \to 0 \quad \text{as } N \to \infty.$$

To control $\|f_n - f\|_1$, we apply Fatou's lemma to get $\|f\|_1 \leq \liminf_{n\to\infty} \|f_n\|_1 \leq \sup_{n\in\mathbb{N}} \|f_n\|_1 < \infty$. By dominated convergence theorem, we also have $\|f^N - f\|_1 \to 0$. Hence $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.

(iii) $\Rightarrow$ (i). We define the continuous function

$$\psi_N(x) = \begin{cases} |x|, & x \in (-M+1, M-1), \\ (M-1)(M-|x|), & x \in [-M, -M+1] \cup [M-1, M], \\ 0, & x > M. \end{cases}$$

Fix $\epsilon > 0$. Note that $\|f\|_1 < \infty$, and $\psi_N(f) \to |f|$ a.e.. By dominated convergence theorem, we choose N so large that $\int_\Omega (|f| - \psi_N(f))\,\mathrm{d}\mu < \epsilon/3$. We also choose n_0 according to the following criteria:

- Since $\|f_n\|_1 \to \|f\|_1$, there exists n_0 such that $\|f_n\|_1 \le \|f\|_1 + \epsilon/3$ for all $n \ge n_0$;
- By continuous mapping theorem and dominated convergence theorem, we have $\|\psi_N(f_n) - \psi_N(f)\|_1 \to 0$. We then choose n_0 such that $\|\psi_N(f_n) - \psi_N(f)\|_1 < \epsilon/3$ for all $n \ge n_0$.

Consequently, for all $n \ge n_0$,

$$\int_{\{|f_n| \ge N\}} |f_n|\,\mathrm{d}\mu \le \int_\Omega (|f_n| - \psi_N(f_n))\,\mathrm{d}\mu \le \int_\Omega |f|\,\mathrm{d}\mu + \frac{\epsilon}{3} - \int_\Omega \psi_N(f)\mathrm{d}\mu + \frac{\epsilon}{3} < \epsilon.$$

By taking N larger, we can make $\int_{\{|f_n| \ge N\}} |f_n| < \epsilon$ for $1 \le n < n_0$, and (f_n) is uniformly integrable. $\square$

Remark. The condition of Lebesgue dominated convergence theorem can be weakened as follows. If $(\Omega, \mathscr{F}, \mu)$ is a finite measure space, $(f_n)_{n=1}^\infty \subset L^1(\Omega, \mathscr{F}, \mu)$ is a uniformly integrable sequence, and $f_n \xrightarrow{\mu} f$, then

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

1.5.3 Weak Convergence of Measures

Definition 1.76 (Weak convergence)**.** Let Ω be a metric space with its Borel σ-algebra $\mathscr{B}$. Let $C_b(\Omega)$ be the set of all bounded continuous functions on Ω. Let μ_n be a sequence of probability measures on $(\Omega, \mathscr{B})$. If there exists a probability measure μ on $(\Omega, \mathscr{B})$ such that

$$\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu, \quad \forall f \in C_b(\Omega),$$

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 $f : \Omega \to \overline{\mathbb{R}}$ 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 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 1.77. *Let Ω be a metric space. For every nonnegative lower semi-continuous function $f : \Omega \to \overline{\mathbb{R}}_+$, there exists a sequence of nonnegative bounded Lipschitz continuous functions f_n such that $f_n \uparrow f$ pointwise.*

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

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

Symmetrically $g_n(y) - g_n(x) \le nd(x,y)$. 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. □

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

Lemma 1.78. *Let Ω be a metric space equipped with its Borel σ-algebra $\mathscr{B}$. Let μ_n be a sequence of probability measures on $(\Omega, \mathscr{B})$. If $\mu_n \xrightarrow{w} \mu$ and $\mu_n \xrightarrow{w} \mu'$, then $\mu = \mu'$.*

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

Let G be a closed set, then χ_G is lower semi-continuous. By Lemma 1.77, we choose a sequence of bounded Lipschitz continuous functions $f_n \nearrow \chi_G$. By monotone convergence theorem,

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

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

The Portmanteau lemma gives multiple equivalent definitions of weak convergence.

Theorem 1.79 (Portmanteau lemma)**.** *Let Ω be a metric space with its Borel σ-algebra $\mathscr{B}$. Let μ_n be a sequence of probability measures on $(\Omega, \mathscr{B})$. The following are equivalent:*

(i) $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$ *for all bounded continuous functions f. In other words, $\mu_n \xrightarrow{w} \mu$;*
(ii) $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$ *for all bounded Lipschitz continuous functions f;*
(iii) $\liminf_{n\to\infty} \int f \, \mathrm{d}\mu_n \geq \int f \, \mathrm{d}\mu$ *for all lower semi-continuous function f bounded from below;*
(iv) $\limsup_{n\to\infty} \int f \, \mathrm{d}\mu_n \leq \int f \, \mathrm{d}\mu$ *for all upper semi-continuous function f bounded from above;*
(v) $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for every open sets G;*
(vi) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for every closed sets F;*
(vii) $\lim_{n\to\infty} \mu_n(B) \to \mu(B)$ *for all Borel sets B with $\mu(\partial B) = 0$, where $\partial B = \overline{B} \setminus \mathring{B}$ 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 1.77, 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

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

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

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

(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,

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

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

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

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

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

$$\mu(\mathring{B}) \overset{\text{(v)}}{\leq} \liminf_{n\to\infty}\mu_n(\mathring{B}) \leq \liminf_{n\to\infty}\mu_n(B) \leq \limsup_{n\to\infty}\mu_n(B) \leq \limsup_{n\to\infty}\mu_n(\overline{B}) \overset{\text{(vi)}}{\leq} \mu(\overline{B}).$$

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 $r \in [0,\infty)\backslash C$.

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 $\omega \in \partial B_F(r) = \overline{\Omega\backslash B_F(r)} \cap B_F(r)$, choose a sequence $\omega_n \in \Omega\backslash B_F(r)$ such that $\omega_n \to \omega$. Again by continuity of $d(\cdot,F)$, $d(\omega,F) = \lim_{n\to\infty} d(\omega_n,F) \geq r$. Hence $d(\omega,F)=r$, which implies $\delta B_F(r) \subset D_F(r)$.

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 $r \in [0,\infty)\backslash C$, 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. □

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

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

Since $B_F(r_k) \downarrow F$,

$$\mu(F) = \lim_{k\to\infty}\mu(B_F(r_k)) \geq \limsup_{n\to\infty}\mu_n(F).$$

Hence (vi) follows. □

2 Random Variables

2.1 Random Variables and Independence

From now on, our discussion builds on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

2.1.1 Random Variables and Distribution Functions

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

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

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

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

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 $X : \Omega \to \overline{\mathbb{R}}$ 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. $F(x) = \lim_{\epsilon \to 0^+} F(x+\epsilon)$ for all $x \in \mathbb{R}$;

(iii) $F(-\infty) := \lim_{x \to -\infty} F(x) = 0$, and $F(\infty) := \lim_{x \to \infty} F(x) = 1$.

In fact, any function $F : \mathbb{R} \to [0,1]$ 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

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

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:

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

Since $\mathbb{P}(X = x) = F(x) - \lim_{\epsilon \to 0^+} F(x-\epsilon)$, 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 $(\Omega, \mathscr{F}, \mathbb{P})$ determines a pushforward measure $\mu_F = \mathbb{P} \circ X^{-1}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$:

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

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

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

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

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

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

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

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

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

(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 $\mu \perp m$. In other words, μ is concentrated on a Lebesgue-null set E, where μ takes zero at each point of E.

We give an example of singular measures on $\mathbb{R}$.

Ternary Cantor sets and devil's staircase. A ternary Cantor set K is obtained by repeatedly removing the middle thirds from the compact unit interval $[0, 1]$:

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

Since ternary Cantor set $K = \bigcap_{n=1}^{\infty} K_n$ is the intersection of a decreasing sequence of non-empty compact sets, it is itself compact. Furthermore, it can be written as the set of numbers in $[0, 1]$ with a ternary expansion omitting the digit 1 (the ternary numbers $0.c_1c_2\cdots$ with $c_n = 1$ digit is 1 are removed from K_n):

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

It is not hard to see that (i) K is an uncountable set, since $\{0, 2\}^{\mathbb{N}}$ is uncountable, and (ii) K is a Lebesgue-null set in $\mathbb{R}$, since $m(K) = \lim_{n\to\infty} m(K_n) = \lim_{n\to\infty} (2/3)^n = 0$.

We define a function sequence (F_n) on $[0, 1]$ by the following recursive formula:

$$F_1(x) = x, \qquad F_{n+1}(x) = \begin{cases} \frac{1}{2}F_n(3x), & 0 \le x < \frac{1}{3}, \\ \frac{1}{2}, & \frac{1}{3} \le x \le \frac{2}{3}, \\ \frac{1}{2} + \frac{1}{2}F_n(3x - 2), & \frac{2}{3} < x \le 1. \end{cases}$$

This is a sequence of continuous monotone increasing functions, with

$$\|F_{n+1} - F_n\|_\infty \le \frac{1}{2}\|F_n - F_{n-1}\|_\infty \le \cdots \le \frac{1}{2^{n-1}}\|F_2 - F_1\|_\infty \le \frac{1}{3 \cdot 2^n}.$$

Therefore, the sequence (F_n) converges uniformly to a function F on $[0,1]$, called the *devil's staircase*, which is also continuous and monotone increasing. Furthermore, for each connected component (a,b) of $[0,1]\backslash K$, we have $F(a)=F(b)$. Hence $F'=0$ almost everywhere on $[0,1]$.

Note that $F(0)=0$ and $F(1)=1$, we define $F=0$ on $(-\infty,0]$ and $F=1$ on $[1,\infty)$. Then F is a c.d.f.. Let μ_F be the distribution measure of F. Since F is continuous, μ_F has no atom. Furthermore,

$$\mu_F(\mathbb{R}\backslash K)=\mu_F([0,1]\backslash K)=\mu_F\left(\bigcup_{n=1}^{\infty}[0,1]\backslash K_n\right)=\sum_{(a,b)\in[0,1]\backslash K}(F(b)-F(a))=0.$$

Therefore $\mu_F\perp m$, and μ_F is a singular measure on $\mathbb{R}$.

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 $X_1,X_2,\cdots,X_n$ are all random variables, the function

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

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

$$F(x_1,\cdots,x_n)=\mathbb{P}(X_1\le x_1,\cdots,X_n\le x_n)$$

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

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

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

Remark. By Theorem 1.35, a random vector is also a measurable function $X:(\Omega,\mathscr{F})\to(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$. Furthermore, the c.d.f. $F:\mathbb{R}^n\to[0,1]$ 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 $x_1,\cdots,x_n\in\mathbb{R}$, $\lim_{x_j\to-\infty}F(x_1,\cdots,x_n)=0$, and $\lim_{x_1,\cdots,x_n\to\infty}F(x_1,\cdots,x_n)=1$.

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

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

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

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

2.1.2 Independence

Definition 2.8 (Independence)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ br a probability space.

(i) (Independent events). A pair of events $A, B \in \mathscr{F}$ are said to be *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Also, n events $A_1, \cdots, A_n \in \mathscr{F}$ are said to be *mutually independent* if for all $I \subset \{1, \cdots, n\}$, it holds

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

(ii) (Independent σ-algebras). Two two sub-σ-algebras $\mathscr{F}_1$ and $\mathscr{F}_2$ of $\mathscr{F}$ are said to be *independent*, if each pair of events $A \in \mathscr{F}_1$ and $B \in \mathscr{F}_2$ are independent.

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

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

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

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

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

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

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

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

This theorem can also be stated as follows:

(i) X and Y are independent if and only if $\sigma(X)$ and $\sigma(Y)$ are independent.

(ii) 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

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

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

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

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

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

Corollary 2.11. *If X and Y are two independent random variables, and $\psi, \varphi : \mathbb{R} \to \mathbb{R}$ 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

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

Then we finish the proof. $\square$

Remark. More generally, we say two collection of random variables $\{X_\alpha\}_{\alpha\in I}$ and $\{Y_\beta\}_{\beta\in J}$ are *independent* if for any two random vectors $(X_{\alpha_1}, \cdots, X_{\alpha_m})$ and $(Y_{\beta_1}, \cdots, Y_{\beta_n})$, it holds

$$\mathbb{P}\left((X_{\alpha_1}, \cdots, X_{\alpha_m}) \in A, (Y_{\beta_1}, \cdots, Y_{\beta_n}) \in B\right) = \mathbb{P}\left((X_{\alpha_1}, \cdots, X_{\alpha_m}) \in A\right) \cdot \mathbb{P}\left((Y_{\beta_1}, \cdots, Y_{\beta_n}) \in B\right)$$

for all Borel sets $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$.

The σ-algebra generated by the random variable collection $\{X_\alpha\}_{\alpha\in I}$ is the smallest sub-σ-algebra of $\mathscr{F}$ such that every X_α is measurable. By definition,

$$\sigma\left(\{X_\alpha\}_{\alpha\in I}\right) \subset \sigma\left(\bigcup_{\alpha\in I} \sigma(X_\alpha)\right).$$

On the other hand, since every $\sigma(X_\alpha)$ is contained in $\sigma\left(\{X_\alpha\}_{\alpha\in I}\right)$, the union $\bigcup_{\alpha\in I} \sigma(X_\alpha)$ is also contained in $\sigma\left(\{X_\alpha\}_{\alpha\in I}\right)$, and so

$$\sigma\left(\{X_\alpha\}_{\alpha\in I}\right) \supset \sigma\left(\bigcup_{\alpha\in I} \sigma(X_\alpha)\right).$$

Hence we have

$$\sigma\left(\{X_\alpha\}_{\alpha\in I}\right) = \sigma\left(\bigcup_{\alpha\in I} \sigma(X_\alpha)\right) = \sigma\left(\bigcup_{m\in\mathbb{N},\, \alpha_1,\cdots,\alpha_m\in I} \sigma(X_{\alpha_1}, \cdots, X_{\alpha_m})\right).$$

Now we claim that $\{X_\alpha\}_{\alpha\in I}$ and $\{Y_\beta\}_{\beta\in J}$ are independent if and only if $\sigma(\{X_\alpha\}_{\alpha\in I})$ and $\sigma(\{Y_\beta\}_{\beta\in J})$ are independent. We fix $\mathscr{G} = \sigma(Y_{\beta_1}, \cdots, Y_{\beta_n})$. For each $A \in \mathscr{G}$, the class

$$\{B \in \mathscr{F} :\ \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)\}$$

is a λ-system containing every $\sigma(X_{\alpha_1}, \cdots, X_{\alpha_m})$. Since the union of σ-algebras $\sigma(X_{\alpha_1}, \cdots, X_{\alpha_m})$ generated by finite subcollections is a π-system, the above λ-system also contains $\sigma(\{X_\alpha\}_{\alpha\in I})$, by Sierpiński-Dynkin π-λ theorem. Since both $A \in \mathscr{G}$ and $\mathscr{G} = \sigma(Y_{\beta_1}, \cdots, Y_{\beta_n})$ is arbitrary, $\sigma(\{X_\alpha\}_{\alpha\in I})$ is independent of all $\sigma(Y_{\beta_1}, \cdots, Y_{\beta_n})$ generated by finite subcollections. Similar to the previous procedure, we can run over all $A \in \sigma(\{X_\alpha\}_{\alpha\in I})$ to conclude that $\sigma(\{X_\alpha\}_{\alpha\in I})$ is independent of $\sigma(\{Y_\beta\}_{\beta\in J})$.

2.2 Expectation

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

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

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

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

Theorem 2.13 (Integral transform)**.** *Using the integral transform formula in Theorem 1.50, we immediately know that the expectation of $\varphi(X)$ equals the Lebesgue-Stieltjes integral*

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

Particularly,

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

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

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

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

$$\mathbb{E}[\varphi(X)] = \int \varphi\rho \, \mathrm{d}m = \int \varphi(x)\rho(x) \, \mathrm{d}x.$$

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

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

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

Proof. By Fubini's theorem,

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

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 $\mathbb{1}_{\{X>x\}}$. $\square$

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 $g : \mathbb{R} \to \mathbb{R}$ 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

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

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

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

(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

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

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 $\mathscr{G}$ be a sub σ-algebra of $\mathscr{F}$, i.e. $\mathscr{G} \subset \mathscr{F}$ and $\mathscr{G}$ 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 $\mathscr{G}$* is defined as any random variable ξ satisfying the following:

(i) ξ is $\mathscr{G}$-measurable, i.e. $\xi^{-1}(\mathscr{B}(\mathbb{R})) \subset \mathscr{G}$;

(ii) $\mathbb{E}[Y\mathbb{1}_A] = \mathbb{E}[\xi\mathbb{1}_A]$ for all $A \in \mathscr{G}$.

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

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

Then ρ is the desired $\mathscr{G}$-measurable function. Furthermore, if ξ is a conditional expectation of Y with respect to $\mathscr{G}$, we define $A_n = \{\omega : \rho(\omega) > \xi(\omega) + n^{-1}\}$. Then A_n is $\mathscr{G}$-measurable, and

$$0 \le \mathbb{P}(\rho > \xi + n^{-1}) \le n\mathbb{E}[(\rho - \xi)\mathbb{1}_{A_n}] = \mathbb{E}[Y\mathbb{1}_{A_n}] - \mathbb{E}[Y\mathbb{1}_{A_n}] = 0$$

for all $n \in \mathbb{N}$, and letting $n \to \infty$ gives $\mathbb{P}(\rho > \xi) = 0$. Similarly, we have $\mathbb{P}(\xi > \rho + n^{-1}) = 0$. Hence $\xi = \rho$ *a.s.*. Therefore, the conditional expectation, written $\mathbb{E}[Y|\mathscr{G}]$, exists and is almost surely unique.

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

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

(i) (Total expectation formula). $\mathbb{E}[\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}X$. In addition, if $\mathscr{H}$ is another sub σ-algebra of $\mathscr{F}$ such that $\mathscr{H} \subset \mathscr{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathscr{G}]|\mathscr{H}] = \mathbb{E}[\mathbb{E}[X|\mathscr{H}]|\mathscr{G}] = \mathbb{E}[X|\mathscr{H}]$ a.s..

(ii) (Monotonicity). If $X \ge 0$ a.s., then $\mathbb{E}[X|\mathscr{G}] \ge 0$ a.s.. Hence $X \le Y$ a.s. implies $\mathbb{E}[X|\mathscr{G}] \le \mathbb{E}[Y|\mathscr{G}]$ a.s.. In particular, $|\mathbb{E}[X|\mathscr{G}]| \le \mathbb{E}[|X|\,|\mathscr{G}]$ a.s..

(iii) ($\mathscr{G}$-linearity). For any $\mathscr{G}$-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) (Independence law). $\sigma(X)$ is independent of $\mathscr{G}$ if and only if $\mathbb{E}[\varphi(X)|\mathscr{G}] = \mathbb{E}[\varphi(X)]$ a.s. $\forall$ measurable φ.

(v) (Conditional Jensen's inequality). If $g : \mathbb{R} \to \mathbb{R}$ is a convex function such that $g(X)$ is integrable, then $\mathbb{E}[g(X)|\mathscr{G}] \ge g(\mathbb{E}[X|\mathscr{G}])$ a.s..

(vi) The following inequalities almost surely hold:

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

$$\mathbb{E}[|XY|^r|\mathscr{G}]^{1/r} \le \mathbb{E}[|X|^p|\mathscr{G}]^{1/p}\,\mathbb{E}[|Y|^q|\mathscr{G}]^{1/q} \ \ a.s..$$

- *(Conditional moment inequality). If $1 \le p < q$, then*

$$\mathbb{E}[|X|^p|\mathscr{G}]^{1/p} \le \mathbb{E}[|X|^q|\mathscr{G}]^{1/q} \ \ a.s..$$

- *(Conditional Minkowski inequality). If $p \ge 1$, then*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(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 $a + b\,\mathbb{E}[X|\mathscr{G}] \leq \mathbb{E}[g(X)|\mathscr{G}]$ *a.s.* for all $a, b \in \mathcal{S}$. Since $\mathcal{S}$ is countable, we have

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

Hence $g(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[g(X)|\mathscr{G}]$ *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 $X \in L^p(\Omega, \mathscr{F}, \mathbb{P})$, where $1 \le p < \infty$. By conditional Jensen's inequality,

$$(\mathbb{E}[X|\mathscr{G}])^p \le \mathbb{E}\left[|X|^p|\mathscr{G}\right] \ a.s. \quad \Rightarrow \quad \|\mathbb{E}[X|\mathscr{G}]\|_p^p \le \mathbb{E}\left[\mathbb{E}\left[|X|^p|\mathscr{G}\right]\right] = \|X\|_p^p \quad \Rightarrow \quad \frac{\|\mathbb{E}[X|\mathscr{G}]\|_p}{\|X\|_p} \le 1.$$

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

The convergence theorems for expectation can be extended to conditional expectation.

Theorem 2.18. *Let $\mathscr{G}$ be a sub σ-algebra of $\mathscr{F}$.*

(i) (Conditional monotone convergence theorem). Let $(X_n)_{n=1}^\infty$ be a increasing sequence of L^1 nonnegative random variables such that $X_n \uparrow X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Then

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] = \mathbb{E}\left[X|\mathscr{G}\right] \ a.s..$$

(ii) (Conditional Fatou's lemma). Let $(X_n)_{n=1}^\infty$ be a sequence of nonnegative L^1 random variables. Then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathscr{G}\right] \le \liminf_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] \ a.s..$$

(iii) (Conditional dominated convergence theorem). If $(X_n)_{n=1}^\infty$ is a sequence of random variables such that $X_n \to X$ a.s., and there exists a integrable random variable $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ such that $|X_n| \le Y$ a.s. for all $n \in \mathbb{N}$, then

$$\mathbb{E}[X|\mathscr{G}] = \lim_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] \ a.s. \text{ and in } L^1.$$

Proof. (i) Define $Y_n = \mathbb{E}[X - X_n|\mathscr{G}]$. By monotonicity of conditional expectation, Y_n is a decreasing sequence. We denote by Y the limit of sequence (Y_n). For each $A \in \mathscr{G}$,

$$\mathbb{E}[Y_n \mathbb{1}_A] = \mathbb{E}[(X - X_n)\mathbb{1}_A].$$

Since $|X - X_n| \le |X| \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, by Lebesgue dominated convergence theorem,

$$\mathbb{E}[Y\mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}[(X - X_n)\mathbb{1}_A] = \mathbb{E}\left[\lim_{n\to\infty}(X - X_n)\mathbb{1}_A\right] = 0.$$

Since $Y \ge 0$ is $\mathscr{G}$-measurable, $Y = 0$ *a.s.*, and the desired limit follows.

(ii) Let $Y_n = \inf_{k\ge n} X_k$, which is a increasing sequence of nonnegative L^1 random variables. By monotonicity of conditional expectation,

$$\mathbb{E}[Y_n|\mathscr{G}] \le \mathbb{E}[X_k|\mathscr{G}], \quad \text{for all } k \ge n.$$

Hence $\mathbb{E}[Y_n|\mathscr{G}] \le \inf_{k\ge n} \mathbb{E}[X_k|\mathscr{G}]$, and by (i),

$$\mathbb{E}\left[\lim_{n\to\infty} Y_n \middle| \mathscr{G}\right] = \lim_{n\to\infty} \mathbb{E}\left[Y_n|\mathscr{G}\right] \le \liminf_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] \ a.s..$$

(iii) The almost sure convergence follows by applying (ii) on sequences $(Y + X_n)$ and $(Y - X_n)$. For the L^1 convergence, note that $0 \le |X_n - X| \le 2Y$. By Lebesgue dominated convergence theorem,

$$\mathbb{E}\left[\,|\mathbb{E}[X_n|\mathscr{G}] - \mathbb{E}[X|\mathscr{G}]|\,\right] \le \mathbb{E}\left[\mathbb{E}[|X_n - X|\,|\mathscr{G}]\right] = \mathbb{E}\,|X_n - X| \to 0.$$

Hence $\mathbb{E}[X_n|\mathscr{G}]$ converges to $\mathbb{E}[X|\mathscr{G}]$ in L^1-norm. $\square$

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

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

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

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

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

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

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

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

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

$$\mathbb{E}_1[\varphi\circ T|T^{-1}\mathscr{G}_1] = \mathbb{E}_2[\varphi|\mathscr{G}_2]\circ T, \tag{2.7}$$

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

Proof. Let $\xi_2 = \mathbb{E}_2[\varphi|\mathscr{G}_2]$. Then $\xi_1 := \xi_2 \circ T$ is a $T^{-1}\mathscr{G}_2$-measurable function on $(\Omega_1, \mathscr{F}_1)$. By Theorem 1.49, we have $\mathbb{E}_1[\psi\circ T] = \mathbb{E}_2[\psi]$ for all $\psi \in L^1(\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$. For any $A_1 \in T^{-1}\mathscr{G}_2$, we have $A_2 := TA_1 \in \mathscr{G}_2$, 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 $A_1 \in T^{-1}\mathscr{G}_2$ is arbitrary, we have $\mathbb{E}_1[\varphi\circ T|T^{-1}\mathscr{G}_1] = \xi_2\circ T$ *a.s.*, which is (2.7). □

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

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

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

Theorem 2.22 (Doob-Dynkin)**.** *Let $(\Omega, \mathscr{F})$ and $(\Gamma, \mathscr{G})$ be measurable spaces. Given a measurable function $T : \Omega \to \Gamma$, let $\sigma(T) := T^{-1}\mathscr{G} \subset \mathscr{F}$ be the σ-algebra generated by T. A real-valued function $g : \Omega \to \mathbb{R}$ is $\sigma(T)$-measurable if and only if there exists a $\mathscr{G}$-measurable function $\varphi : \Gamma \to \mathbb{R}$ such that*

$$g = \varphi\circ T.$$

Proof. The sufficiency is clear, so we only prove the necessity. Let g be 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 $B_n \in \mathscr{G}$ such that $A_n = T^{-1}B_n$, namely, $\mathbb{1}_{A_n} = \mathbb{1}_{B_n} \circ T$. Hence we define

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

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

Remark. By Theorem 2.22, since $\mathbb{E}[Y|\sigma(X)] : (\Sigma, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is a $\sigma(X)$-measurable function, there exists a Borel-measurable function $\varphi : (\mathbb{R}, \mathscr{B}(\mathbb{R})) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mathbb{E}[Y|\sigma(X)] = \varphi(X)$. In other words, $\mathbb{E}[Y|X]$ is the composition of a measurable function φ and the random variable X.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Clearly, for $r \le s$, we have $\mathbb{1}_{\{X\in(-\infty,r]\}} \le \mathbb{1}_{\{X\in(-\infty,s]\}}$. By monotonicity of conditional expectation, we have $F(r|\cdot) \le F(s|\cdot)$ 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 $\{B_r\}_{r\in\mathbb{Q}} \subset \mathscr{F}$ and $C \in \mathscr{F}$ such that

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

as well as

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

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 $\omega \in \Omega\backslash E$, define

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

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

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

$$\omega \mapsto \mu_{X|\mathscr{G}}(B|\mathscr{G})(\omega) = \mathbb{1}_E(\omega)F_0(r) + \mathbb{1}_{\Omega\backslash E}(\omega)F(r|\omega)$$

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

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

This is a λ-system, because

(i) $\mu_{X|\mathscr{G}}(\Omega|\mathscr{G})(\omega) \equiv 1$, which implies $\Omega \in \mathscr{D}$;

(ii) For $E \subset F \in \mathscr{D}$, $\mu_{X|\mathscr{G}}(F\backslash E|\mathscr{G})(\omega) = \mu_{X|\mathscr{G}}(F|\mathscr{G})(\omega) - \mu_{X|\mathscr{G}}(E|\mathscr{G})(\omega)$, hence $F\backslash E \in \mathscr{D}$;

(iii) For increasing sequence $B_n \in \mathscr{D}$, $\mu_{X|\mathscr{G}}(B|\mathscr{G})(\omega) = \lim_{n\to\infty} \mu_{X|\mathscr{G}}(B_n|\mathscr{G})(\omega)$, hence $B := \bigcup_{n=1}^{\infty} B_n \in \mathscr{D}$.

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

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

Thus we complete the full proof. $\square$

Remark. If X is independent of $\mathscr{G}$, the conditional distribution $\mu_{X|\mathscr{G}}(\cdot|\mathscr{G})(\omega)$ 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,

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

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

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

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 $X_n : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 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

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

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

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

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

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

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 $(\Omega, \mathscr{F}, \mathbb{P})$ 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

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

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) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded continuous functions f;*

(ii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded Lipschitz continuous functions f;*

(iii) $\liminf_{n\to\infty} \mathbb{E}[f(X_n)] \geq \mathbb{E}[f(X)]$ *for all lower semi-continuous function f bounded from below;*

(iv) $\limsup_{n\to\infty} \mathbb{E}[f(X_n)] \leq \mathbb{E}[f(X)]$ *for all upper semi-continuous function f bounded from above;*

(v) $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ *for every open sets G;*

(vi) $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *for every closed sets F;*

(vii) $\lim_{n\to\infty} \mathbb{P}(X_n \in B) \to \mathbb{P}(X \in B)$ *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 $\lim_{n\to\infty} \mathbb{P}(X_n \in I) = \mathbb{P}(X \in I)$ 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

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

Since $\lim_{n\to\infty} \mathbb{P}(X_n \in I) = \mathbb{P}(X \in I)$ 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

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

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 $F : \mathbb{R} \to [0, 1]$ 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 $B = (-\infty, x]$. □

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

A continuous mapping preserves several modes of stochastic convergence of random variable sequences.

Theorem 2.30 (Continuous mapping). *Let X be a random variable. If $g : \mathbb{R} \to \mathbb{R}$ 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

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

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

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

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 $x_k \to x$ 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

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

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

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 $\limsup_{n\to\infty} \mathbb{P}(|X_n| > M_0) \leq \mathbb{P}(|X| > M_0)$. 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 $j = 1, \cdots, N-1$. 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 $f_n : \mathbb{R} \to [-M, M]$ 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 $f : \mathbb{R} \to [-M, M]$.*

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 $f_{1n}(r_1) \to f(r_1)$. Then $f_{1n}(r_2)$ is a bounded sequence, and again we choose one of its convergent subsequence $f_{2n}(r_2) \to f(r_2)$.

"diagonal trick": Repeat this procedure, so for each $k \in \mathbb{N}$, we choose a subsequence f_{kn} such that $f_{kn}(r_j) \to f(r_j)$ 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 $\lim_{n\to\infty} f_{nn}(r_k) = f(r_k)$ for all $k \in \mathbb{N}$. Hence we obtain a subsequence f_{nn} that converges to f pointwise on $\mathbb{Q}$. Clearly, $f : \mathbb{Q} \to [-M, M]$ is increasing.

For all irrationals $x \in \mathbb{R}\backslash\mathbb{Q}$, choose a increasing rational sequence $r_{k_j} \to x$, and let $f(x) = \lim_{j\to\infty} f(r_{k_j})$. 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 $f_{nn} \to f$ pointwise on $\mathbb{R}$. Given $\epsilon > 0$. If $x \in \mathbb{R}\backslash\mathbb{Q}$ is a point of continuity of f, we choose rationals $r < x < r'$ with $|f(r_{k_i}) - f(r_{k_j})| < \epsilon$. Then

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

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 $F_n : \mathbb{R} \to [0,1]$ be a sequence of cumulative distribution functions. Then there exists a subsequence F_{n_k} such that $F_{n_k}(x) \to F(x)$ 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 $G : \mathbb{R} \to [0,1]$. Define $F(x) = \lim_{\epsilon \to 0^+} G(x+\epsilon)$ 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. □

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, $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

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 $F_{n_k}(M) \to F(M) > 1-\epsilon$. Since $\epsilon > 0$ is arbitrary, $F(x) \to 1$ as $x \to \infty$. The case $x \to -\infty$ is similar. □

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 $\rho((x_n,y_n),(x,y)) := \sqrt{|x-x_n|^2 + |y-y_n|^2} \le |x-x_n| + |y-y_n|$.

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

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

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

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - c| \ge \epsilon) = \limsup_{n\to\infty} \mathbb{P}(X_n \in \mathbb{R}\backslash(c-\epsilon, c+\epsilon)) \le \mathbb{P}(c \in \mathbb{R}\backslash(c-\epsilon,c+\epsilon)) = 0.$$

(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 $\mathbb{E}[f(X_n,c)] \to \mathbb{E}[f(X,c)]$. Thus $(X_n,c) \xrightarrow{d} (X,c)$, and the result follows. □

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 $c \ne 0$, 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,

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

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, $((1 + X_n)^{-1})_{n=N}^{\infty}$ is uniformly tight, and so is $((1 + X_n)^{-1})_{n=1}^{\infty}$.

(vi) Follows from (iii) and (v). □

Theorem 2.39. *Let* $R : \mathbb{R} \to \mathbb{R}$ *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* $h \to 0$*, then* $R(X_n) = o_{\mathbb{P}}(|X_n|^p)$*;*

(ii) If $R(h) = O(|h|^p)$ *as* $h \to 0$*, 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 $\mathbb{P}(|g(X_n)| > M) \leq \mathbb{P}(|X_n| > \delta) \to 0$, and the sequence $g(X_n)$ is uniformly tight. □

2.5 Characteristic Functions

Definition 2.40 (Characteristic function)**.** Let $X \sim \mu$ be a (real-valued) random variable, where μ is a distribution measure. The *characteristic function of X* is defined as

$$\phi_X : \mathbb{R} \to \mathbb{C},\ \phi_X(\lambda) = \mathbb{E}\left[e^{\mathrm{i}\lambda X}\right] = \int_{\mathbb{R}} e^{\mathrm{i}\lambda x}\,\mathrm{d}\mu(x),\ \mathrm{i}^2 = -1.$$

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) $\phi_X : \mathbb{R} \to \mathbb{C}$ *is bounded and uniformly continuous.*

(iii) If $\mathbb{E}\left[|X|^n\right] < \infty$ *for some* $n \in \mathbb{N}$*, then* ϕ_X *is* n*-differentiable, and its* k*-th derivative is*

$$\phi_X^{(k)}(\lambda) = \mathbb{E}\left[(\mathrm{i}X)^k e^{\mathrm{i}\lambda X}\right],\ k = 1, \cdots, n. \tag{2.13}$$

Furthermore, all these derivatives are uniformly continuous. Particularly, we have $\phi^{(k)}(0) = \mathrm{i}^k\mathbb{E}[X^k]$*.*

(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* $\lim_{\lambda\to\pm\infty} \phi_X(\lambda) = 0$*.*

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:

$$e^{\mathrm{i}\theta} - 1 = 2\mathrm{i}e^{\frac{\mathrm{i}\theta}{2}}\sin\frac{\theta}{2} \ \Rightarrow\ \left|e^{\mathrm{i}\theta} - 1\right| \le 2\left|\sin\frac{\theta}{2}\right| \le |\theta|.$$

Then for all $\lambda_1, \lambda_2 \in \mathbb{R}$, we have

$$|\phi_X(\lambda_1) - \phi_X(\lambda_2)| = \left|\mathbb{E}\left[\left(e^{\mathrm{i}(\lambda_1-\lambda_2)X} - 1\right)e^{\mathrm{i}\lambda_2 X}\right]\right| \le \mathbb{E}\left|e^{\mathrm{i}(\lambda_1-\lambda_2)X} - 1\right| \le 2\int_{\mathbb{R}}\left|\sin\frac{(\lambda_1-\lambda_2)x}{2}\right|\mathrm{d}\mu(x). \tag{2.14}$$

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

$$2\int_{\mathbb{R}}\left|\sin\frac{(\lambda_1-\lambda_2)x}{2}\right|\mathrm{d}\mu(x) \le 2\int_{-R}^{R}\left|\sin\frac{(\lambda_1-\lambda_2)x}{2}\right|\mathrm{d}\mu(x) + \frac{\epsilon}{2} < 2\int_{[-R,R]}\frac{\epsilon}{4}\,\mathrm{d}\mu + \frac{\epsilon}{2} = \epsilon. \tag{2.15}$$

Hence ϕ_X is uniformly continuous.

(iii) Assume $\mathbb{E}|X| < \infty$. By Lebesgue dominated convergence theorem, since $\frac{1}{\epsilon}\left|\exp(\mathrm{i}\epsilon X) - 1\right| \le |X|$,

$$\lim_{\epsilon\to0}\frac{\phi_X(\lambda+\epsilon) - \phi_X(\lambda)}{\epsilon} = \lim_{\epsilon\to0}\mathbb{E}\left[\frac{e^{\mathrm{i}\epsilon X} - 1}{\epsilon}e^{\mathrm{i}\lambda X}\right] = \mathbb{E}\left[\lim_{\epsilon\to0}\frac{e^{\mathrm{i}\epsilon X} - 1}{\epsilon}e^{\mathrm{i}\lambda X}\right] = \mathbb{E}\left[\mathrm{i}Xe^{\mathrm{i}\lambda X}\right].$$

Hence ϕ_X is differentiable. Furthermore, by monotone convergence theorem, we choose $R_1 > 0$ such that $\mathbb{E}[|X|\mathbb{1}_{\{|X|>R_1\}}] < \epsilon/2$. Alike (2.14) and (2.15), whenever $|\lambda_1 - \lambda_2| < \epsilon/(2R_1^2)$, we have

$$|\phi_X'(\lambda_1) - \phi_X'(\lambda_2)| = \mathbb{E}\left[\mathrm{i}X(e^{\mathrm{i}\lambda_1 X} - e^{\mathrm{i}\lambda_2 X})\right] < 2\int_{-R_1}^{R_1}\left|x\sin\frac{(\lambda_1-\lambda_2)x}{2}\right|\mathrm{d}\mu(x) + \frac{\epsilon}{2} < \epsilon.$$

Therefore the derivative ϕ_X' is uniformly continuous. Now assume (2.13) holds for $k-1$. If $\mathbb{E}[|X|^k] < \infty$,

$$\lim_{\epsilon\to0}\frac{\phi_X^{(k-1)}(\lambda+\epsilon) - \phi_X^{(k-1)}(\lambda)}{\epsilon} = \lim_{\epsilon\to0}\mathbb{E}\left[\frac{e^{\mathrm{i}\epsilon X} - 1}{\epsilon}(\mathrm{i}X)^{k-1}e^{\mathrm{i}\lambda X}\right] = \mathbb{E}\left[\lim_{\epsilon\to0}\frac{e^{\mathrm{i}\epsilon X} - 1}{\epsilon}(\mathrm{i}X)^{k-1}e^{\mathrm{i}\lambda X}\right] = \mathbb{E}\left[(\mathrm{i}X)^k e^{\mathrm{i}\lambda X}\right].$$

Therefore ϕ_X is k-differentiable. Again by monotone convergence theorem, we choose $R_k > 0$ such that $\mathbb{E}[|X|^k \mathbb{1}_{\{|X|>R_k\}}] < \epsilon/2$. Whenever $|\lambda_1 - \lambda_2| < \epsilon/(2R_k^{k+1})$, we have

$$\left|\phi_X^{(k)}(\lambda_1) - \phi_X^{(k)}(\lambda_2)\right| = \mathbb{E}\left[(\mathrm{i}X)^k(e^{\mathrm{i}\lambda_1 X} - e^{\mathrm{i}\lambda_2 X})\right] < 2\int_{-R_k}^{R_k} \left|x^k \sin\frac{(\lambda_1-\lambda_2)x}{2}\right| \mathrm{d}\mu(x) + \frac{\epsilon}{2} < \epsilon.$$

Hence $\phi_X^{(k)}$ is uniformly continuous. By induction we finish the proof of (iii).

(iv) We only proves necessity, since (iii) implies sufficiency. By definition, if $\phi_X''(0)$ exists, we have

$$\phi_X''(0) = \lim_{h\to 0}\frac{\phi_X(h)+\phi_X(-h)-2\phi(0)}{h^2} = \lim_{h\to 0}\int_{\mathbb{R}}\frac{2\cos(hx)-2}{h^2}\,\mathrm{d}\mu(x) = -2\lim_{h\to 0}\int_{\mathbb{R}}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x).$$

By Fatou's lemma, we have

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\,\mathrm{d}\mu(x) = 2\int_{\mathbb{R}}\lim_{h\to 0}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x) \le 2\liminf_{h\to 0}\int_{\mathbb{R}}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x) = -\phi_X''(0) < \infty.$$

Generally, if $\phi_X^{(2k-2)}(0)$ exists, then $\mathbb{E}[X^{2k-2}] < \infty$, and by (iii), $\phi_X^{(2k-2)}(\lambda) = \mathbb{E}[(\mathrm{i}X)^{2k-2}e^{\mathrm{i}\lambda X}]$. Then

$$\begin{aligned}\phi_X^{(2k)}(0) = \lim_{h\to 0}\frac{\phi_X^{(2k-2)}(h)+\phi_X^{(2k-2)}(-h)-2\phi(0)}{h^2} &= \lim_{h\to 0}\int_{\mathbb{R}}(\mathrm{i}x)^{2k-2}\frac{2\cos(hx)-2}{h^2}\,\mathrm{d}\mu(x)\\ &= (-1)^{k-1}2\lim_{h\to 0}\int_{\mathbb{R}}x^{2k-2}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x).\end{aligned}$$

By Fatou's lemma, we have

$$\begin{aligned}\mathbb{E}[X^{2k}] = \int_{\mathbb{R}} x^{2k}\,\mathrm{d}\mu(x) &= 2\int_{\mathbb{R}}\lim_{h\to 0}x^{2k-2}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x)\\ &\le 2\liminf_{h\to 0}\int_{\mathbb{R}}x^{2k-2}\frac{1-\cos(hx)}{h^2}\,\mathrm{d}\mu(x) = (-1)^{k-1}\phi_X^{(2k-2)}(0) < \infty.\end{aligned}$$

Then (iv) follows from induction.

(v) Since X is continuous, there exists a density function $\rho \in L^1(\mathbb{R})$ of X, and $\phi_X(\lambda) = \int_{\mathbb{R}}\rho(x)e^{\mathrm{i}\lambda x}\,\mathrm{d}x$. 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 $x \to x - \frac{\pi}{\lambda}$ implies

$$\phi_X(\lambda) = \int_{\mathbb{R}}\rho(x)e^{\mathrm{i}\lambda x}\,\mathrm{d}x = \int_{\mathbb{R}}\rho\left(x-\frac{\pi}{\lambda}\right)e^{\mathrm{i}\lambda x}e^{\mathrm{i}\pi x}\,\mathrm{d}x = -\int_{\mathbb{R}}\rho\left(x-\frac{\pi}{\lambda}\right)e^{\mathrm{i}\lambda x}\,\mathrm{d}x$$

Use the two formulae to compute $\phi_X(\lambda)$, we have

$$|\phi_X(\lambda)| \le \frac{1}{2}\int_{\mathbb{R}}\left|\rho(x)-\rho\left(x-\frac{\pi}{\lambda}\right)\right|\,\mathrm{d}x.$$

Since ρ is continuous, $\left|\rho(x)-\rho\left(x-\frac{\pi}{\lambda}\right)\right| \to 0$ as $|\lambda|\to\infty$ for all $x\in\mathbb{R}$. By Lebesgue dominate convergence theorem, $|\phi_X(\lambda)| \to 0$ as $|\lambda|\to\infty$.

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 $\|f-g\|_1 \le \epsilon$ for any $\epsilon > 0$. Then

$$\limsup_{\lambda\to\pm\infty}|\phi_X(\lambda)| \le \limsup_{\lambda\to\pm\infty}\left|\int(\rho(x)-f(x))e^{\mathrm{i}\lambda x}\,\mathrm{d}x\right| + \limsup_{\lambda\to\pm\infty}\left|\int f(x)e^{\mathrm{i}\lambda x}\,\mathrm{d}x\right| \le \epsilon + 0 = \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $|\phi_X(\lambda)| \to 0$ as $\lambda \to \pm\infty$. $\square$

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,

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t} \phi_F(t)\,\mathrm{d}t.$$

Distinct c.d.f. F have distinct characteristic functions.

(ii) For any $\varphi \in C_c(\mathbb{R})$,

$$\int \varphi\,\mathrm{d}F = \lim_{\delta\to 0^+} \frac{1}{2\pi} \int \varphi(x) \left(\int e^{-\mathrm{i}tx} \phi_F(t) e^{-\frac{\delta}{2}t^2}\,\mathrm{d}t \right) \mathrm{d}x.$$

(iii) If $\|\phi_F\|_1 := \int |\phi_F(t)|\,\mathrm{d}t < \infty$, then F has density function ρ:

$$\rho(x) = \frac{1}{2\pi} \int e^{-\mathrm{i}tx} \phi_F(t)\,\mathrm{d}t,$$

and $\sup_{x\in\mathbb{R}} \rho(x) \le \frac{1}{2\pi}\|\phi_F\|_1 < \infty$.

Proof. (ii) Let $Z \sim N(0,1)$ be independent of $X \sim F$. Then for all $\varphi \in C_c(\mathbb{R})$ and $\delta > 0$,

$$\begin{aligned}
\mathbb{E}\left[\varphi(X + \sqrt{\delta}Z)\right] &= \mathbb{E}\left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) e^{-\frac{\lambda^2}{2}}\,\mathrm{d}\lambda\right] \\
&= \mathbb{E}\left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) \mathbb{E}\left[e^{-\mathrm{i}\lambda Z}\right] \mathrm{d}\lambda\right] = \mathbb{E}\left[\frac{1}{\sqrt{2\pi}} \int \varphi(X + \sqrt{\delta}\lambda) e^{-\mathrm{i}\lambda Z}\,\mathrm{d}\lambda\right] \\
&= \mathbb{E}\left[\frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) e^{-\mathrm{i}\frac{\xi - X}{\sqrt{\delta}}Z}\,\mathrm{d}\xi\right] = \frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) \mathbb{E}\left[e^{-\mathrm{i}\frac{\xi - X}{\sqrt{\delta}}Z}\right] \mathrm{d}\xi \\
&= \frac{1}{\sqrt{2\pi\delta}} \int \varphi(\xi) \mathbb{E}\left[e^{-\mathrm{i}\frac{\xi Z}{\sqrt{\delta}}} \phi_F\left(\frac{Z}{\sqrt{\delta}}\right)\right] \mathrm{d}\xi = \frac{1}{2\pi\sqrt{\delta}} \int \varphi(\xi) \left(\int e^{-\mathrm{i}\frac{\xi z}{\sqrt{\delta}}} \phi_F\left(\frac{z}{\sqrt{\delta}}\right) e^{-\frac{z^2}{2}}\,\mathrm{d}z \right) \mathrm{d}\xi
\end{aligned}$$

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),

$$|F(b) - F(a)| \le \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t} \right| |\phi_F(t)|\,\mathrm{d}t.$$

Since $\left| \frac{\exp(-\mathrm{i}at) - \exp(-\mathrm{i}bt)}{\mathrm{i}t} \right| \le |b - a|$, we have

$$|F(b) - F(a)| \le \frac{|b - a|}{2\pi} \|\phi_F\|_1. \tag{2.16}$$

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

$$\frac{F(b) - F(a)}{b - a} \le \frac{1}{2\pi} \int \frac{e^{\mathrm{i}at} - e^{\mathrm{i}bt}}{\mathrm{i}(b - a)t} \phi_F(t)\,\mathrm{d}t.$$

Let $b \searrow a$, the equation in (iii) holds, and the estimate of upper bound follows from (2.16). □

The proof of (i) requires some technical lemma.

Lemma 2.43 (Dirichlet). *Let $\alpha \in \mathbb{R}$. Then*

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{\sin(\alpha t)}{t}\,\mathrm{d}t = \pi\,\mathrm{sgn}(\alpha),$$

where $\mathrm{sgn}(\alpha) = \mathbb{1}_{(0,\infty)}(\alpha) - \mathbb{1}_{(-\infty,0)}(\alpha)$.

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,

$$\int_0^\infty \frac{\sin t}{t}\,\mathrm{d}t = \int_0^\infty \left(\int_0^\infty e^{-st}\sin t\,\mathrm{d}s\right)\mathrm{d}t = \int_0^\infty \left(\int_0^\infty e^{-st}\sin t\,\mathrm{d}t\right)\mathrm{d}s = \int_0^\infty \frac{1}{1+s^2}\,\mathrm{d}s = \frac{\pi}{2}.$$

The result follows by changing variables. □

Proof of Theorem 2.42 (i). Let $a < b$ be two points of continuity of F, and $X \sim F$. By Fubini's theorem,

$$\begin{aligned}
\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t}\phi_F(t)\,\mathrm{d}t &= \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t}\mathbb{E}\left[e^{\mathrm{i}tX}\right]\mathrm{d}t \\
&= \mathbb{E}\left[\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t}e^{\mathrm{i}tX}\,\mathrm{d}t\right] \\
&= \mathbb{E}\left[\frac{1}{2\pi}\int_{-T}^{T} \frac{\sin(tX - at) - \sin(tX - bt)}{t}e^{\mathrm{i}tX}\,\mathrm{d}t\right] \qquad (2.17)
\end{aligned}$$

Note that $\int_0^T \frac{\sin t}{t}\,\mathrm{d}t \le \int_0^\pi \frac{\sin t}{t}\,\mathrm{d}t \le \pi$ for all $T > 0$, the integrand in (2.17) is bounded by 2. By Lebesgue dominated convergence theorem,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\mathrm{i}at} - e^{-\mathrm{i}bt}}{\mathrm{i}t}\phi_F(t)\,\mathrm{d}t &= \mathbb{E}\left[\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} \frac{\sin(tX - at) - \sin(tX - bt)}{t}e^{\mathrm{i}tX}\,\mathrm{d}t\right] \\
&= \frac{1}{2}\mathbb{E}\left[\mathrm{sgn}(X-a) - \mathrm{sgn}(X-b)\right] \qquad \text{(By Lemma 2.43)} \\
&= \frac{1}{2}\left(1 - 2F(a) - 1 + 2F(b)\right) \\
&= F(b) - F(a),
\end{aligned}$$

where the last row follows from continuity of F at a and b. □

Remark. In high-dimensional case, a similar conclusion follows: Let μ_F be a distribution measure on $\mathscr{B}(\mathbb{R}^p)$, and let $\phi_F(\lambda) = \int_{\mathbb{R}^p} \exp\left(\mathrm{i}\langle x, \lambda\rangle\right)\mathrm{d}\mu_F(x)$ be the characteristic function of μ_F. Let $A \subset \mathbb{R}^d$ be a cell of the form

$$A = \left\{(x_1, \cdots, x_p) : a_j \le x_j \le b_j \textit{ for all } j\right\},$$

where $a_j < b_j$ for all j and $\mu_F(\partial A) = 0$. Then

$$\mu_F(A) = \lim_{T\to\infty}\frac{1}{(2\pi)^p}\int_{[-T,T]^p}\prod_{j=1}^{p}\left\{\frac{e^{\mathrm{i}a_jt_j} - e^{\mathrm{i}b_jt_j}}{\mathrm{i}t_j}\right\}\phi_F(t)\,\mathrm{d}t, \quad \text{where } t = (t_1, \cdots, t_p) \in \mathbb{R}^p. \qquad (2.18)$$

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 $\mu(\delta A) = 0$ form a π-system that generate $\mathscr{B}(\mathbb{R}^p)$. 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 $\mathscr{B}(\mathbb{R}^2)$. 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 $X \overset{d}{=} Y$ if and only $\langle X, \alpha\rangle \overset{d}{=} \langle Y, \alpha\rangle$ for all $\alpha \in \mathbb{R}^p$.*

Proof. The necessity is clear. For sufficiency, note that when $\langle X, \alpha\rangle \overset{d}{=} \langle Y, \alpha\rangle$ 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*

$$\mathbb{P}(|X| \geq \epsilon) \leq K \int_0^1 \left[1 - \mathrm{Re}\left(\phi_X\left(\frac{t\delta}{\epsilon}\right)\right)\right] \mathrm{d}t,$$
$$\mathbb{E}\left[X^2 \mathbb{1}_{\{|X|\leq\epsilon\}}\right] \leq K\epsilon^2 \left[1 - \mathrm{Re}\left(\phi_X\left(\frac{t\delta}{\epsilon}\right)\right)\right].$$

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

$$1 - \frac{\sin x}{x} \geq \frac{1}{K},\ \forall |x| \geq \delta, \quad \textit{and} \quad \frac{1-\cos x}{x^2} \geq \frac{1}{K},\ \forall |x| \leq \delta.$$

Let $X \sim F$, and let $\eta = \delta/\epsilon$. Then

$$\begin{aligned} \mathbb{P}(|X| \geq \epsilon) &= \mathbb{E}\left[\mathbb{1}_{\{|\eta X| \geq \delta\}}\right] \leq K\mathbb{E}\left[1 - \frac{\sin(\eta X)}{\eta X}\right] \\ &= K - \mathbb{E}\left[K \int_0^1 \cos(\eta t X)\,\mathrm{d}t\right] = K \int_0^1 \left[1 - \mathrm{Re}\left(\phi_X(\eta t)\right)\right] \mathrm{d}t, \qquad \text{(By Fubini's theorem)} \end{aligned}$$

and

$$\mathbb{E}\left[X^2 \mathbb{1}_{\{|X|\leq\epsilon\}}\right] = \frac{1}{\eta^2}\mathbb{E}\left[|\eta X|^2 \mathbb{1}_{\{|\eta X|\leq\delta\}}\right] \leq \frac{K}{\eta^2}\mathbb{E}\left[1 - \cos(\eta X)\right] = K\epsilon^2\left[1 - \mathrm{Re}\left(\phi_X(\eta t)\right)\right].$$

Thus we complete the proof. $\square$

2.6 The Continuity Theorem and The Central Limit Theorems

2.6.1 Lévy's Continuity Theorem

Definition 2.47 (Quantile)**.** Given a c.d.f. $F : \mathbb{R} \to [0,1]$, the *quantile* of distribution F is defined as

$$F^{-1}(p) = \inf\{\alpha \in \mathbb{R} : F(\alpha) \geq p\}, \ \forall p \in (0,1).$$

Remark. The quantile $F^{-1} : (0,1) \to \mathbb{R}$ satisfies the following properties:

(i) $F^{-1} : (0,1) \to \mathbb{R}$ 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:

$$\{\alpha \in \mathbb{R} : F(\alpha) \geq p\} = \bigcap_{n=1}^{\infty} \left\{\alpha \in \mathbb{R} : F(\alpha) \geq p - \frac{1}{n}\right\} \quad \Rightarrow \quad F^{-1}(p) = \lim_{n\to\infty} F^{-1}\left(p - \frac{1}{n}\right).$$

Lemma 2.48 (Galois inequality)**.** *Let $\alpha \in \mathbb{R}$ and $p \in (0,1)$. Then $F(\alpha) \geq p$ if and only if $F^{-1}(p) \leq \alpha$. Particularly, 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,

$$F(\alpha) = \lim_{n\to\infty} F\left(\alpha + \frac{1}{n}\right) \geq p.$$

Thus we finish the proof. □

Corollary 2.49 (Quantile transformation)**.** *Let $U \sim \mathrm{Unif}(0,1)$. 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 $F_n^{-1}(p) \to F^{-1}(p)$ 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 $\mathbb{P}(F_n(Z) < p) \to \mathbb{P}(F(Z) < p)$.

Let Φ be the c.d.f. of standard Gaussian variables. By Galois inequality,

$$\Phi(F_n^{-1}(p)) = \mathbb{P}(Z < F_n^{-1}(p)) = \mathbb{P}(F_n(Z) < p) \overset{n\to\infty}{\to} \mathbb{P}(F(Z) < p) = \Phi(F^{-1}(p))$$

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 $F_n^{-1}(p) \to F^{-1}(p)$ for each point p of continuity of F^{-1}.

Conversely, assume that $F_n^{-1}(p) \to F^{-1}(p)$ for each point p of continuity of F^{-1}. Let $U \sim \mathrm{Unif}(0,1)$, 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 $F_n : \mathbb{R} \to [0,1]$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$ 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, $X_n \to X$ on Ω except possibly at countably many points of discontinuity of F^{-1}, which form a null set. ☐

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 $\phi_{F_n} \to \phi_F$ 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 $e^{\mathrm{i}\lambda X_n} \xrightarrow{a.s.} e^{\mathrm{i}\lambda X}$ for all $\lambda \in \mathbb{R}$. By Lebesgue dominated convergence theorem, $\phi_{F_n} \to \phi_F$ pointwise. ☐

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 $\phi : \mathbb{R} \to \mathbb{C}$, the following are equivalent:*

(i) $\{X_n\}_{n=1}^\infty$ is uniformly tight, i.e. $\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}(|X_n| \geq M) = 0$.
(ii) $X_n \xrightarrow{d} X$ for some random variable X.
(iii) ϕ is the characteristic of some random variable X, i.e. $\phi(\lambda) = \mathbb{E}[e^{\mathrm{i}\lambda X}]$;
(iv) ϕ is continuous everywhere on $\mathbb{R}$;
(v) ϕ is continuous at 0.

Proof. (i) ⇒ (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) ⇒ (iii) follows from Corollary 2.52. (iii) ⇒ (iv) and (iv) ⇒ (v) are trivial.

(v) ⇒ (i): Following Proposition 2.46, we set $\delta = 2$ and $K = 5$. The following estimate holds for all $n \in \mathbb{N}$:

$$\mathbb{P}(|X_n| \geq T) \leq 5 \int_0^1 \left[1 - \mathrm{Re}\left(\phi_n\left(\frac{2t}{T}\right)\right)\right] \mathrm{d}t.$$

By Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \geq T) \leq 5 \int_0^1 \left[1 - \mathrm{Re}\left(\phi\left(\frac{2t}{T}\right)\right)\right] \mathrm{d}t.$$

Since $\phi(0) = 1$, and ϕ is continuous at 0, the right-hand side of the above estimate converges to 0 as $T \to \infty$. Given $\epsilon > 0$, we choose T_0 such that $\lim_{n\to\infty} \mathbb{P}(|X_n| \geq T_0) < \epsilon/2$, 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 $\{X_n\}_{n=1}^\infty$. ☐

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 $\phi_{F_n} \to \phi$ pointwise, and ϕ is continuous at 0, then ϕ is the characteristic function of some c.d.f. F, and $F_n \xrightarrow{w} F$.

2.6.2 The Central Limit Theorems

Theorem 2.54 (Khintchine'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*

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j.$$

Then $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$*.*

Proof. Without loss of generality, assume $\mu = 0$. Let $\phi : \mathbb{R} \to \mathbb{C}$ be the characteristic function of X_1. Then the characteristic function of $\overline{X}_n$ is

$$\phi_n(\lambda) = \mathbb{E}\left[\prod_{j=1}^n e^{\mathrm{i}\frac{\lambda}{n}X_j}\right] = \phi\left(\frac{\lambda}{n}\right)^n.$$

Since $\mathbb{E}[|X_1|] < \infty$, ϕ is differentiable, $\phi'(0) = \mathrm{i}\,\mathbb{E}X_1 = 0$, and ϕ' is uniformly continuous. Fix $\lambda \in \mathbb{R}$. Given any $\epsilon > 0$, we can choose N such that $|\phi'(t)| \le \epsilon$ for all $|t| \le |\lambda|/N$. Hence

$$\lim_{n\to\infty} |\phi_n(\lambda) - 1| = \lim_{n\to\infty}\left|\phi\left(\frac{\lambda}{n}\right)^n - 1\right| \le \lim_{n\to\infty}\left|\left(1 + \int_0^{\lambda/n} \phi'(t)\,\mathrm{d}t\right)^n - 1\right| \le \max\left\{e^{|\lambda|\epsilon} - 1, 1 - e^{-|\lambda|\epsilon}\right\}.$$

Since $\epsilon > 0$ is arbitrary, $\phi_n(\lambda) \to 1$ pointwise. By Lévy's continuity theorem, $\overline{X}_n \xrightarrow{d} 0$. By Theorem 2.35 (iv),we have $\overline{X}_n \xrightarrow{\mathbb{P}} 0$. ☐

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* $\mathbb{E}\left[|X_1|^2\right] < \infty$*. Denote* $\mu := \mathbb{E}X_1 < \infty$*, and* $0 < \sigma^2 := \mathrm{Var}(X_1) < \infty$*. Define*

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j,\ Z_n = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$$

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

$$\phi_n(\lambda) = \mathbb{E}\left[\prod_{j=1}^n e^{\mathrm{i}\frac{\lambda}{\sqrt{n}}X_j}\right] = \phi\left(\frac{\lambda}{\sqrt{n}}\right)^n.$$

Since $\mathbb{E}[|X_1|^2] < \infty$, ϕ is twice-differentiable, $\phi'(0) = \mathrm{i}\,\mathbb{E}X_1 = 0$, and $\phi''(0) = -\mathbb{E}[X_1^2] = -1$.

$$\phi\left(\frac{\lambda}{\sqrt{n}}\right) = 1 + \int_0^{\lambda/\sqrt{n}} \phi'(t)\,\mathrm{d}t = 1 + \int_0^{\lambda/\sqrt{n}}\int_0^t \phi''(u)\,\mathrm{d}u\,\mathrm{d}t = 1 - \frac{\lambda^2}{2n} + \int_0^{\lambda/\sqrt{n}}\int_0^t (1 + \phi''(u))\,\mathrm{d}u\,\mathrm{d}t.$$

Note that ϕ'' is uniformly continuous. Given $\epsilon > 0$, choose N such that $|1 + \phi''(u)| < \epsilon$ for all $|u| \le |\lambda|/\sqrt{N}$. Then for all $n \ge N$, we have

$$1 - \frac{\lambda^2(1+\epsilon)}{2n} \le \phi\left(\frac{\lambda}{\sqrt{n}}\right) \le 1 - \frac{\lambda^2(1-\epsilon)}{2n} \ \Rightarrow\ e^{-\frac{\lambda^2(1+\epsilon)}{2}} \le \lim_{n\to\infty}\phi_n(\lambda) \le e^{-\frac{\lambda^2(1-\epsilon)}{2}}.$$

Since $\epsilon > 0$ is arbitrary, $\lim_{n\to\infty}\phi_n(\lambda) = e^{-\lambda^2/2}$, which is the characteristic function of $Z \sim N(0,1)$. ☐

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 $0 < \sigma_n^2 := \mathbb{E}\left[X_n^2\right] < \infty$. 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

$$\lim_{n\to\infty} \frac{\max_{1\le k\le n}\sigma_k^2}{s_n^2} = 0 \tag{2.19}$$

*if and only if the following **Lindeberg's condition** is satisfied:*

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[X_k^2 \mathbb{1}_{\{|X_k|\ge \epsilon s_n\}}\right] = 0, \quad \forall \epsilon > 0. \tag{2.20}$$

Proof of Suffieiency. Suppose the Lindeberg's condition (2.20) holds. Then for $1 \le k \le n$ and all $\epsilon > 0$,

$$\frac{\max_{1\le k\le n}\sigma_k^2}{s_n^2} = \max_{1\le k\le n}\left\{\frac{1}{s_n^2}\mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right] + \frac{1}{s_n^2}\mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right]\right\} \le \epsilon^2 + \frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right].$$

Then (2.19) is true follows by letting $n\to\infty$ and $\epsilon\downarrow 0$. Let ϕ_n be the characteristic function of X_n. To prove $Z_n \to N(0,1)$, we need to show that the characteristic function of Z_n satisfies

$$\phi_{Z_n}(\lambda) = \prod_{k=1}^n \phi_k\left(\frac{\lambda}{s_n}\right) \to e^{-\lambda^2/2} \quad \text{as } n\to\infty. \tag{2.21}$$

The result then follows from Lévy's continuity theorem. We claim that (2.21) holds if and only if

$$\lim_{n\to\infty}\sum_{k=1}^n\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right) + \frac{\lambda^2}{2} = 0, \tag{2.22}$$

We first prove the following (2.23), which together with (2.22) implies (2.21):

$$\lim_{n\to\infty}\left|\exp\left\{\sum_{k=1}^n\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right\} - \prod_{k=1}^n \phi_k\left(\frac{\lambda}{s_n}\right)\right| = 0. \tag{2.23}$$

Claim I. If $\phi : \mathbb{R}\to\mathbb{C}$ 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 $N \sim$ 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

$$\mathbb{E}\left[e^{\mathrm{i}\lambda W}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\mathrm{i}\lambda W}|N\right]\right] = \mathbb{E}\left[\phi(\lambda)^N\right] = \sum_{n=0}^{\infty}\frac{e^{-1}}{n!}\phi(\lambda)^n = e^{\phi(\lambda)-1}.$$

Claim II (Product comparison). Given $\{a_1,\cdots,a_n\},\{b_1,\cdots,b_n\}\subset\{z\in\mathbb{C}:|z|\le 1\}$, it holds

$$\left|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j\right| \le \sum_{j=1}^n |a_j - b_j|.$$

The case $n=1$ is clear. Then prove the case $n=2$:

$$|a_1b_1 - a_2b_2| = |a_1(b_1-b_2) - (a_1-b_1)b_2| \le |a_1-b_1| + |a_2-b_2|.$$

We apply this formula to $\prod_{j=1}^{n-1}a_j, \prod_{j=1}^{n-1}b_j, a_n, b_n$, so the general case follows from induction.

Proof of (2.23). By Claims I and II,

$$\left|\exp\left\{\sum_{k=1}^{n}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right\}-\prod_{k=1}^{n}\phi_k\left(\frac{\lambda}{s_n}\right)\right|\le\sum_{k=1}^{n}\left|\exp\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)-\phi_k\left(\frac{\lambda}{s_n}\right)\right|. \tag{2.24}$$

By Taylor's theorem, we have

$$\left|\phi_k\left(\frac{\lambda}{s_n}\right)-1\right|\le 1+\frac{\lambda^2}{2s_n^2}\sup_{t\in\mathbb{R}}\phi''(t)-1=\frac{\lambda^2\sigma_k^2}{2s_n^2}\le\frac{\lambda^2}{2}\max_{1\le k\le n}\frac{\sigma_k^2}{s_n^2}.$$

Given any $0<\epsilon<1$, by (2.19), we can choose N such that $|\phi_k(\lambda/s_n)-1|\le\epsilon/2$ for all $n\ge N$ and all $1\le k\le n$. Since $|e^z-z-1|\le\epsilon|z|$ for all $|z|\le\epsilon/2$, following (2.24), we have

$$\left|\exp\left\{\sum_{k=1}^{n}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right\}-\prod_{k=1}^{n}\phi_k\left(\frac{\lambda}{s_n}\right)\right|\le\sum_{k=1}^{n}\epsilon\left|\phi_k\left(\frac{\lambda}{s_n}\right)-1\right|\le\frac{\epsilon\lambda^2}{2s_n^2}\sum_{k=1}^{n}\sigma_k^2=\frac{\epsilon\lambda^2}{2},\quad\forall n\ge N.$$

Since $\epsilon>0$ is arbitrary, the limit (2.23) holds.

Proof of (2.22). Given $\epsilon>0$, we use the following expansion:

$$\begin{aligned}\sum_{k=1}^{n}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)+\frac{\lambda^2}{2}&=\sum_{k=1}^{n}\mathbb{E}\Big[\underbrace{e^{i\lambda X_k/s_n}-1-\frac{i\lambda}{s_n}X_k-\frac{(i\lambda)^2}{2s_n^2}X_k^2}_{=:A_{n,k}}\Big]\\&=\sum_{k=1}^{n}\mathbb{E}\left[A_{n,k}\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right]+\sum_{k=1}^{n}\mathbb{E}\left[A_{n,k}\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right]=:S_{n,\epsilon}^{(1)}+S_{n,\epsilon}^{(2)}.\end{aligned}$$

Now we bound the two terms. For the first term,

$$\begin{aligned}\left|S_{n,\epsilon}^{(1)}\right|&\le\sum_{k=1}^{n}\mathbb{E}\left|A_{n,k}\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right|\le\sum_{k=1}^{n}\mathbb{E}\left[\frac{1}{3!}\left|\frac{\lambda}{s_n}X_k\right|^3\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right]\\&=\frac{|\lambda|^3}{6s_n^3}\sum_{k=1}^{n}\mathbb{E}\left[|X_k|^3\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right]\le\frac{|\lambda|^3\epsilon}{6s_n^2}\sum_{k=1}^{n}\mathbb{E}|X_k|^2=\frac{|\lambda|^3\epsilon}{6}.\end{aligned} \tag{2.25}$$

By Lindeberg's condition, we can bound the second term as $n\to\infty$:

$$\left|S_{n,\epsilon}^{(2)}\right|\le\sum_{k=0}^{n}\mathbb{E}\left|A_{n,k}\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right|\le\sum_{k=1}^{n}\mathbb{E}\left[\left|\frac{\lambda}{s_n}X_k\right|^2\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right]=\frac{|\lambda|^2}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right]\to 0. \tag{2.26}$$

Since $\epsilon>0$, we can bound (2.22) by arbitrarily small numbers, and the result follows. ☐

Remark. In estimates (2.25) and (2.26), we used the following estimate in case $n=2$:

$$\left|e^{i\theta}-\sum_{k=1}^{n}\frac{(i\theta)^k}{k!}\right|\le\min\left\{\frac{2\theta^n}{n!},\frac{\theta^{n+1}}{(n+1)!}\right\},\quad\forall\theta\in\mathbb{R}.$$

The following proof of necessity is given by William Feller.

Proof of Necessity (Theorem 2.56). 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 $\lambda\ne0$, we have

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|\ge\epsilon s_n\}}\right]=1-\frac{2}{\lambda^2}\sum_{k=1}^{n}\mathbb{E}\left[\frac{\lambda^2X_k^2}{2s_n^2}\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right].$$

Then we use the estimate $1-\cos x \le \frac{x^2}{2}$ to obtian

$$
\begin{aligned}
\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[X_k^2 \mathbb{1}_{\{|X_k|\ge \epsilon s_n\}}\right] &\le \frac{2}{\lambda^2}\left(\frac{\lambda^2}{2}+\sum_{k=1}^n \mathbb{E}\left[\left(\cos\left(\frac{\lambda X_k}{s_n}\right)-1\right)\mathbb{1}_{\{|X_k|<\epsilon s_n\}}\right]\right) \\
&= \frac{2}{\lambda^2}\left(\frac{\lambda^2}{2}+\sum_{k=1}^n \operatorname{Re}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)+\sum_{k=1}^n \mathbb{E}\left[\left(1-\cos\left(\frac{\lambda X_k}{s_n}\right)\right)\mathbb{1}_{\{|X_k|\ge \epsilon s_n\}}\right]\right) \\
&\le \frac{2}{\lambda^2}\left(\frac{\lambda^2}{2}+\sum_{k=1}^n \operatorname{Re}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right)+\frac{4}{\lambda^2}\sum_{k=1}^n \mathbb{P}(|X_k|\ge \epsilon s_n) \\
&\le \frac{2}{\lambda^2}\left(\frac{\lambda^2}{2}+\sum_{k=1}^n \operatorname{Re}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right)+\frac{4}{\lambda^2}\sum_{k=1}^n \frac{\sigma_k^2}{\epsilon^2 s_n^2} \quad \text{(By Chebyshev's inequality)} \\
&\le \frac{2}{\lambda^2}\left(\frac{\lambda^2}{2}+\sum_{k=1}^n \operatorname{Re}\left(\phi_k\left(\frac{\lambda}{s_n}\right)-1\right)\right)+\frac{4}{\lambda^2\epsilon^2}.
\end{aligned}
$$

By (2.22), the first term converges to 0 as $n\to\infty$. Since $\lambda\neq 0$ is arbitrary, we obtain the Lindeberg's condition (2.20) by letting $\lambda^2\to\infty$. ☐

We also have another form of Lindeberg-Feller theorem which applies to triangular arrays.

Theorem 2.57 (Lindeberg-Feller)**.** *For each $n\in\mathbb{N}$, let $(X_{n,m})_{m=1}^\infty$ be independent random variables with $\mathbb{E}X_{n,m}=0$ and $\mathbb{E}[X_{n,m}^2]<\infty$ for each m. Assume that*

(i) $\sum_{m=1}^n \mathbb{E}[X_{n,m}^2]\to\sigma^2>0$ as $n\to\infty$, and

(ii) (Lindeberg's condition). for all $\epsilon>0$, $\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[X_{n,m}^2\mathbb{1}_{\{|X_{n,m}|>\epsilon\}}]=0$.

Then we have

$$\lim_{n\to\infty}\max_{1\le m\le n}\mathbb{E}[X_{n,m}^2]=0, \tag{2.27}$$

and $S_n=X_{n,1}+X_{n,2}+\cdots+X_{n,m}\xrightarrow{d}N(0,\sigma^2)$ as $n\to\infty$.

Proof. Without loss of generality we assume $\sigma^2=1$. The proof of (2.27) is similar to Theorem 2.56. We let $\phi_{n,m}$ be the characteristic function of $X_{n,m}$. By Lévy's continuity theorem, it suffices prove that

$$\phi_{S_n}(\lambda)=\prod_{m=1}^n \phi_{n,m}(\lambda)\to e^{-\lambda^2/2} \quad \text{as } n\to\infty,$$

which is valid if

$$\lim_{n\to\infty}\sum_{m=1}^n(\phi_{n,m}(\lambda)-1)+\frac{\lambda^2}{2}=0 \quad\text{and}\quad \lim_{n\to\infty}\left|\prod_{m=1}^n e^{\phi_{n,m}(\lambda)-1}-\prod_{m=1}^n\phi_{n,m}(\lambda)\right|=0. \tag{2.28}$$

To prove the first identity in (2.28), we fix any $\epsilon>0$ and use Lindeberg's condition:

$$
\begin{aligned}
&\left|\sum_{m=1}^n(\phi_{n,m}(\lambda)-1)+\frac{\lambda^2}{2}\right| = \left|\sum_{m=1}^n \mathbb{E}\Bigg[\underbrace{e^{i\lambda X_{n,m}}-1-i\lambda X_{n,m}-\frac{(i\lambda X_{n,m})^2}{2}}_{=:A_{n,m}}\Bigg]+\frac{\lambda^2}{2}\left(1-\sum_{m=1}^n\mathbb{E}[X_{n,m}^2]\right)\right| \\
&\le \sum_{m=1}^n \mathbb{E}\left|A_{n,m}\mathbb{1}_{\{|X_{n,m}|\le\epsilon\}}\right|+\sum_{m=1}^n \mathbb{E}\left|A_{n,m}\mathbb{1}_{\{|X_{n,m}|>\epsilon\}}\right|+\frac{\lambda^2}{2}\left|1-\sum_{m=1}^n\mathbb{E}[X_{n,m}^2]\right| \\
&\le \frac{|\lambda|^3\epsilon}{6}\sum_{m=1}^n\mathbb{E}|X_{n,m}|^2+|\lambda|^2\sum_{m=1}^n\mathbb{E}\left[X_{n,m}^2\mathbb{1}_{\{|X_{n,m}|>\epsilon\}}\right]+\frac{\lambda^2}{2}\left|1-\sum_{m=1}^n\mathbb{E}[X_{n,m}^2]\right|\to\frac{|\lambda|^3\epsilon}{6}.
\end{aligned}
$$

Letting $\epsilon \downarrow 0$ yields the desired result. For the second identity in (2.28), note that

$$|\phi_{n,m}(\lambda) - 1| \leq 1 + \frac{\lambda^2}{2} \sup_{t \in \mathbb{R}} |\phi''_{n,m}(t)| - 1 \leq \frac{\lambda^2}{2} \max_{1 \leq m \leq n} \mathbb{E}[X_{n,m}^2].$$

Fix $\epsilon > 0$. For large enough n, we have

$$\begin{aligned}
\left| \prod_{m=1}^{n} e^{\phi_{n,m}(\lambda)-1} - \prod_{m=1}^{n} \phi_{n,m}(\lambda) \right| &\leq \sum_{m=1}^{n} \left| e^{\phi_{n,m}(\lambda)-1} - \phi_{n,m}(\lambda) \right| \\
&\leq \sum_{m=1}^{n} \epsilon \, |\phi_{n,m}(\lambda) - 1| \leq \frac{\lambda^2 \epsilon}{2} \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2] \to \frac{\lambda^2 \epsilon}{2}.
\end{aligned}$$

Again we let $\epsilon \downarrow 0$ to conclude the proof. $\square$

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 2.58 (Lyapunov Condition)**.** *Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent random variables such that $\mathbb{E}X_n = 0$ and $0 < \sigma_n^2 := \mathbb{E}\left[X_n^2\right] < \infty$. Let $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$. If there exists $\delta > 0$ satisfying the **Lyapunov condition***

$$\lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^{2+\delta}\right] = 0, \tag{2.29}$$

then the Lindeberg's condition (2.20) *holds, and so the central limit theorem [Theorem 2.56] applies.*

Proof. If there exists $\delta > 0$ that satisfies the Lyapunov condition (2.29), then

$$\begin{aligned}
\frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}}\right] &= \sum_{k=1}^{n} \mathbb{E}\left[\left|\frac{X_k}{s_n}\right|^2 \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}}\right] \\
&\leq \sum_{k=1}^{n} \mathbb{E}\left[\left|\frac{X_k}{s_n}\right|^2 \left|\frac{X_k}{\epsilon s_n}\right|^{\delta} \mathbb{1}_{\{|X_k| \geq \epsilon s_n\}}\right] \leq \frac{1}{\epsilon^{\delta} s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^{2+\delta}\right] \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Hence the Lindeberg's condition (2.20) is satisfied. $\square$

3 Martingales and Local Martingales

3.1 Processes, Filtrations and Stopping Times

Definition 3.1 (Stochastic processes)**.** Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, 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 $(\Omega, \mathscr{F}, \mathbb{P})$ is a collection $(\mathscr{F}_t)_{t\in\mathcal{T}}$ indexed by elements $\mathcal{T}$ of increasing sub σ-algebras of $\mathscr{F}$, i.e. $\mathscr{F}_s \subset \mathscr{F}_t$ for all $s < t$.

Remark. We can also define the limit of a filtration $(\mathscr{F}_t)_{t\in\mathcal{T}}$ by $\mathscr{F}_\infty = \sigma\left(\bigcup_{t\in\mathcal{T}} \mathscr{F}_t\right)$. If $\mathcal{T} = \mathbb{N}$, then

$$\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_n \subset \mathscr{F}_{n+1} \subset \cdots \subset \mathscr{F}_\infty \subset \mathscr{F}.$$

If $\mathcal{T} = [0, \infty)$, then for all $t > s \geq 0$,

$$\mathscr{F}_0 \subset \mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{F}_\infty \subset \mathscr{F}.$$

Definition 3.3 (Adaptation)**.** Let $\mathcal{T}$ be $\mathbb{N}_0$ or $\mathbb{R}_+$, and let $(\mathscr{F}_t)_{t\in\mathcal{T}}$ be a filtration on $(\Omega, \mathscr{F}, \mathbb{P})$. A stochastic process $(X_t)_{t\in\mathcal{T}}$ is said to be *adapted to* $(\mathscr{F}_t)_{t\in\mathcal{T}}$, if X_t is $\mathscr{F}_t$-measurable for each $t \in \mathcal{T}$.

Remark. A stochastic process $(X_t)_{t\geq 0}$ automatically induces a *canonical filtration*

$$\mathscr{F}_t^X = \sigma(\{X_s\}_{s\leq t}),\ t \in \mathcal{T}.$$

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 $(X_t)_{t\geq 0}$ is adapted to its canonical filtration.

Definition 3.4 (Stopping time)**.** Let $\mathcal{T}$ be $\mathbb{N}_0$ or $\mathbb{R}_+$. A random variable $\tau : \Omega \to \overline{\mathcal{T}} := \mathcal{T} \cup \{\infty\}$ is said to be a *stopping time of the filtration* $\{\mathscr{F}_t\}_{t\in\mathcal{T}}$ if $\{\tau \leq t\} \in \mathscr{F}_t$ 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 $\mathscr{F}_t$-measurable for all $t \in \mathcal{T}$, since

$$\{\tau < t\} = \bigcup_{n=1}^\infty \left\{\tau \leq t - \frac{1}{n}\right\}.$$

Furthermore,

$$\{\tau = \infty\} = \left(\bigcup_{n=1}^\infty \{\tau \leq n\}\right)^c \in \mathscr{F}_\infty.$$

We may modify the definition of stopping time in discrete case. If $\mathcal{T} = \mathbb{N}_0$, then $\tau : \Omega \to \overline{\mathbb{N}}_0$ is a stopping time of the filtration $\{\mathscr{F}_n\}_{n=1}^\infty$ if and only if $\{\tau = n\} \subset \mathscr{F}_n$ 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 $(\mathscr{F}_t)_{t\in\mathcal{T}}$. The *$\sigma$-algebra generated by* τ is defined as

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_\infty : A \cap \{\tau \leq t\} \in \mathscr{F}_t,\ \forall t \in \mathcal{T}\}.$$

Remark. We need to check that $\mathscr{F}_\tau$ is a σ-algebra. Clearly, $\Omega \in \mathscr{F}_\tau$. If $A \in \mathscr{F}_\tau$, then for all $t \in \mathcal{T}$, we have $A^c \cap \{\tau \le t\} = \{\tau \le t\} \cap (A \cap \{\tau \le t\})^c \in \mathscr{F}_t$, which implies $A^c \in \mathscr{F}_\tau$. Finally, given $\{A_n\}_{n=1}^\infty \subset \mathscr{F}_\tau$, we have

$$\left(\bigcup_{n=1}^\infty A_n\right) \cap \{\tau \le t\} = \bigcup_{n=1}^\infty (A_n \cap \{\tau \le t\}) \in \mathscr{F}_t,\ \forall t \in \mathcal{T}.$$

Then $\bigcup_{n=1}^\infty A_n \in \mathscr{F}_\tau$. Therefore $\mathscr{F}_\tau$ is a σ-algebra.

Definition 3.6 (Stopped processes)**.** Let $(X_t)_{t\in\mathcal{T}}$ be an adapted process, and let τ be a stopping time. The *stopped process* $(X_t^\tau)_{t\in\mathcal{T}}$ is defined by

$$X_t^\tau(\omega) = X_{t\wedge\tau(\omega)}(\omega), \quad \forall \omega \in \Omega.$$

On each path, $X_t^\tau(\omega) = X_t(\omega)$ for $t \le \tau(\omega)$, and $X_t^\tau(\omega) = X_{\tau(\omega)}(\omega)$ for $t > \tau(\omega)$. Then this definition can be viewed such that the process $(X_t)_{t\in\mathcal{T}}$ is stopped at the time τ.

Now we work on the case $\mathcal{T} = \mathbb{R}_+$.

Definition 3.7 (Right-continuity)**.** Let $(\mathscr{F}_t)_{t\ge0}$ be a filtration on $(\Omega, \mathscr{F}, \mathbb{P})$. For every $t \in \mathbb{R}_+$, define

$$\mathscr{F}_{t+} = \bigcap_{s>t} \mathscr{F}_s, \quad \textit{and} \quad \mathscr{F}_{\infty+} = \mathscr{F}_\infty.$$

Then $\mathscr{F}_{t+}$ is a σ-algebra, and the collection $(\mathscr{F}_{t+})_{t\ge0}$ is also a filtration on $(\Omega, \mathscr{F}, \mathbb{P})$. If $\mathscr{F}_t = \mathscr{F}_{t+}$ for all $t \in \mathbb{R}_+$, then the filtration $(\mathscr{F}_t)_{t\ge0}$ is said to be *right-continuous*. By construction, the filtration $(\mathscr{F}_{t+})_{t\ge0}$ is automatically right-continuous.

Definition 3.8 (Completeness)**.** Let $(\mathscr{F}_t)_{t\ge0}$ be a filtration on $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{N}$ be the sets of all $(\mathscr{F}_\infty, \mathbb{P})$-negligible sets, i.e. $A \in \mathscr{N}$ if there exists $A' \in \mathscr{F}_\infty$ such that $A' \supset A$ and $\mathbb{P}(A') = 0$. The filtration $\{\mathscr{F}_t\}_{t\ge0}$ is said to be *complete* if $\mathscr{N} \subset \mathscr{F}_0$.

Remark. If $(\mathscr{F}_t)_{t\ge0}$ is not complete, we can complete it by letting $\mathscr{F}_t' = \sigma(\mathscr{F}_t \cup \sigma(\mathscr{N}))$ for every $t \in \mathbb{R}_+$. Apply this completion procedure to the canonical filtration $\mathscr{F}_t = \sigma(\{X_s\}_{s\le t})$ of a stochastic process $\{X_t\}_{t\ge0}$, we obtain the *completed canonical filtration* of $\{X_t\}_{t\ge0}$.

Let $(\mathscr{F}_t)_{t\ge0}$ be a complete filtration on $(\Omega, \mathscr{F}, \mathbb{P})$. By definition, if two random variables $\xi \overset{a.s.}{=} \eta$, and ξ is $\mathscr{F}_t$-measurable, then η is also $\mathscr{F}_t$-measurable.

Definition 3.9 (Measurability and progressiveness)**.** A stochastic process $(X_t)_{t\ge0}$ 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 $\mathscr{F} \otimes \mathscr{B}(\mathbb{R}_+)$ is measurable. In addition, we fix a filtration $(\mathscr{F}_t)_{t\ge0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$. If for each $t \ge 0$, the mapping

$$(\omega, s) \mapsto X_s(\omega)$$

defined on $\Omega \times [0, t]$ equipped with the product σ-field $\mathscr{F}_t \otimes \mathscr{B}([0, t])$ is measurable, then the process $(X_t)_{t\ge0}$ is said to be *progressive*.

Remark. By definition, a progressive process $(X_t)_{t\ge0}$ is both adapted and measurable. In later discussion, we fix the filtration $(\mathscr{F}_t)_{t\ge0}$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ as well as the state space (E, d).

Proposition 3.10. *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

$$X_s^{(n)} = X_{\frac{kt}{n}} \text{ if } s \in \left[\frac{(k-1)t}{n}, \frac{kt}{n}\right), \; k \in \{1, \cdots, n\} \quad \text{and} \quad X_t^{(n)} = X_t.$$

The sample-right-continuity of $\{X_t\}_{t\geq 0}$ implies that for all $\omega \in \Omega$,

$$\lim_{n\to\infty} X_s^{(n)}(\omega) = X_s(\omega), \; \forall s \in [0, t].$$

Furthermore, for every Borel set $B \in \mathscr{B}(E)$,

$$\left\{(\omega, s) \in \Omega \times [0, t] : X_s^{(n)}(\omega) \in B\right\} = \left(\bigcup_{k=1}^{n} \left\{X_{\frac{kt}{n}}(\omega) \in B\right\} \times \left[\frac{(k-1)t}{n}, \frac{kt}{n}\right)\right) \cup (\{X_t \in B\} \times \{t\}).$$

This belongs to the product σ-algebra $\mathscr{F}_t \otimes \mathscr{B}([0, t])$. Hence the mapping $(\omega, s) \mapsto X_s^{(n)}(\omega)$ is measurable on $(\Omega \times [0, t], \mathscr{F}_t \otimes \mathscr{B}([0, t]))$, and so is the pointwise limit $(\omega, s) \mapsto X_s(\omega)$. □

Remark. If the filtration $\{\mathscr{F}_t\}_{t\geq 0}$ 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 3.11. *Write $\mathscr{G}_t = \mathscr{F}_{t+}$ for every $t \in [0, \infty]$.*

(i) A random variable $\tau : \Omega \to [0, \infty]$ is a stopping time of the filtration $(\mathscr{G}_t)_{t\geq 0}$ if and only if $\{\tau < t\} \in \mathscr{F}_t$ for all $t > 0$. This is equivalent to the condition that $\tau \wedge t$ is $\mathscr{F}_t$-measurable for all $t > 0$.

(ii) Let τ be a stopping time of the filtration $(\mathscr{G}_t)_{t\geq 0}$. Then

$$\mathscr{F}_{\tau+} := \{A \in \mathscr{F}_\infty : A \cap \{\tau < t\} \in \mathscr{F}_t, \; \forall t > 0\} = \mathscr{G}_\tau.$$

Proof. (i) Assume $\{\tau < s\} \in \mathscr{F}_s$ for all $s > 0$, and fix $t \geq 0$. Then for all $s > t$,

$$\{\tau \leq t\} = \bigcap_{n\in\mathbb{N}:\, t+n^{-1}<s} \left\{\tau < t + \frac{1}{n}\right\} \in \mathscr{F}_s \;\Rightarrow\; \{\tau \leq t\} \in \bigcup_{s>t} \mathscr{F}_s = \mathscr{G}_t.$$

Conversely, if τ is a stopping time of $\{\mathscr{G}_t\}_{t\geq 0}$, then for all $t > 0$, we have

$$\{\tau < t\} = \bigcup_{n\in\mathbb{N}:\, t-n^{-1}>0}^{\infty} \underbrace{\left\{\tau \leq t - \frac{1}{n}\right\}}_{\in \mathscr{G}_{t-n^{-1}} \subset \mathscr{F}_t} \in \mathscr{F}_t.$$

If $\tau \wedge t$ is $\mathscr{F}_t$-measurable, we have $\{\tau \leq s\} \in \mathscr{F}_t$ for all $s < t$. Then we have $\{\tau < t\} \in \mathscr{F}_t$ 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 $\{\mathscr{G}_n\}_{n\geq 0}$, we have $\{\tau \leq s\} \in \mathscr{G}_s \subset \mathscr{F}_t$ for all $s < t$, and $\tau \wedge t$ is thus $\mathscr{F}_t$-measurable.

(ii) By definition, $\mathscr{G}_\tau := \{A \in \mathscr{F}_\infty : A \cap \{\tau \leq t\} \in \mathscr{G}_t, \; \forall t \geq 0\}$. If $A \cap \{\tau < t\} \in \mathscr{F}_t$ for all $t > 0$, then

$$A \cap \{\tau \leq t\} = \bigcap_{n\in\mathbb{N}:\, t+n^{-1}<s} \left(A \cap \left\{\tau < t + \frac{1}{n}\right\}\right) \in \mathscr{F}_s, \; \forall s > t \geq 0 \;\Rightarrow\; A \cap \{\tau \leq t\} \in \mathscr{G}_t, \; \forall t \geq 0.$$

Conversely, if $A \in \mathscr{G}_\tau$, we have

$$A \cap \{\tau < t\} = \bigcup_{n\in\mathbb{N}:\, t-n^{-1}>0}^{\infty} \underbrace{\left(A \cap \left\{\tau \le t - \frac{1}{n}\right\}\right)}_{\in \mathscr{G}_{t-1/n} \subset \mathscr{F}_t} \in \mathscr{F}_t,\ \forall t > 0.$$

Hence the conclusion follows. $\square$

Proposition 3.12 (Properties of stopping times)**.** *Let τ, σ be two stopping times of the filtration $(\mathscr{F}_t)_{t\ge 0}$.*

(i) τ is a stopping time of $(\mathscr{F}_{t+})_{t\ge 0}$, and $\mathscr{F}_\tau \subset \mathscr{F}_{\tau+}$. If $(\mathscr{F}_t)_{t\ge 0}$ is right-continuous, we have $\mathscr{F}_\tau = \mathscr{F}_{\tau+}$.

(ii) If $\tau = t$ is a constant stopping time, then $\mathscr{F}_\tau = \mathscr{F}_t$, and $\mathscr{F}_{\tau+} = \mathscr{F}_{t+}$.

(iii) τ is $\mathscr{F}_\tau$-measurable.

(iv) Given $A \in \mathscr{F}_\infty$, define

$$\tau^A(\omega) = \begin{cases} \tau(\omega), & \omega \in A, \\ \infty, & \omega \notin A. \end{cases}$$

Then $\tau^A \in \mathscr{F}_\tau$ if and only if τ^A is a stopping time.

(v) If $\sigma \le \tau$, then $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$, and $\mathscr{F}_{\sigma+} \subset \mathscr{F}_{\tau+}$.

(vi) All $\sigma \wedge \tau$, $\sigma \vee \tau$ and $\sigma + \tau$ are stopping times, and $\{\sigma \le \tau\}, \{\sigma = \tau\} \in \mathscr{F}_{\sigma\wedge\tau} = \mathscr{F}_\sigma \cap \mathscr{F}_\tau$.

(vii) A function $\omega \mapsto Y(\omega)$ defined on $\{\tau < \infty\}$ is $\mathscr{F}_\tau$-measurable if and only if for each $t \ge 0$, the restriction of Y to the set $\{\tau \le t\}$ is $\mathscr{F}_t$-measurable.

(viii) If $(\tau_n)_{n=1}^\infty$ is a monotone sequence of increasing stopping times, then $\tau_\infty = \lim_{n\to\infty} \tau_n$ is a stopping time.

(ix) If $(\tau_n)_{n=1}^\infty$ is a monotone sequence of decreasing stopping times, then $\tau_\infty = \lim_{n\to\infty} \tau_n$ is a stopping time of the filtration $(\mathscr{F}_{t+})_{t\ge 0}$, and

$$\mathscr{F}_{\tau_\infty+} = \bigcap_{n=1}^{\infty} \mathscr{F}_{\tau_n+}.$$

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 $n \ge N_\omega$, then τ is a stopping time, and

$$\mathscr{F}_{\tau_\infty} = \bigcap_{n=1}^{\infty} \mathscr{F}_{\tau_n}.$$

Proof. (i) By Remark of Definition 3.4 and Proposition 3.11, τ is also a stopping time of $(\mathscr{F}_{t+})_{t\ge 0}$. The statement $\mathscr{F}_\tau \subset \mathscr{F}_{\tau+}$ follows from $\mathscr{F}_t \subset \mathscr{G}_t$. (ii) immediately follows from definition.

(iii) For all $\alpha \in \mathbb{R}$, we have

$$\{\tau \le \alpha\} \cap \{\tau \le t\} = \{\tau \le \alpha \wedge t\} \in \mathscr{F}_t,\ \forall t \ge 0 \quad \Rightarrow \quad \{\tau \le \alpha\} \in \mathscr{F}_\tau.$$

(iv) The result immediately follows from the definition of $\mathscr{F}_\tau$, since

$$\{\tau^A \le t\} = A \cap \{\tau \le t\},\ \forall t \ge 0.$$

(v) If $A \in \mathscr{F}_\sigma$, then $A \cap \{\sigma \le t\} \in \mathscr{F}_t$ for all $t \ge 0$. Since $\sigma \le \tau$, we have $\{\sigma \le t\} \supset \{\tau \le t\}$, and

$$A \cap \{\tau \le t\} = (A \cap \{\sigma \le t\}) \cap \{\tau \le t\} \in \mathscr{F}_t.$$

(vi) For all $t \geq 0$, we have

$$\{\sigma \wedge \tau \leq t\} = \{\sigma \leq t\} \cup \{\tau \leq t\} \in \mathscr{F}_t, \quad \{\sigma \vee \tau \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathscr{F}_t,$$

$$\textit{and} \quad \{\sigma + \tau > t\} = \{\tau \geq t\} \cup \left(\bigcup_{q \in \mathbb{Q} \cap (0,t]} \{t - q \leq \tau < t\} \cap \{\sigma > q\} \right) \in \mathscr{F}_t.$$

Hence $\sigma \wedge \tau$, $\sigma \vee \tau$ and $\sigma + \tau$ are stopping times. By (v), $\mathscr{F}_{\sigma\wedge\tau} \subset \mathscr{F}_\sigma \cap \mathscr{F}_\tau$. Conversely, if $A \in \mathscr{F}_\sigma \cap \mathscr{F}_\tau$,

$$A \cap \{\sigma \wedge \tau \leq t\} = (\underbrace{A \cap \{\sigma \leq t\}}_{\in \mathscr{F}_t}) \cup (\underbrace{A \cap \{\tau \leq t\}}_{\in \mathscr{F}_t}) \in \mathscr{F}_t, \ \forall t \geq 0 \ \Rightarrow \ A \in \mathscr{F}_{\sigma\wedge\tau}.$$

Hence $\mathscr{F}_{\sigma\wedge\tau} = \mathscr{F}_\sigma \cap \mathscr{F}_\tau$. By Proposition 3.11 (i), for all $t \geq 0$, both $\sigma \wedge t$ and $\tau \wedge t$ are $\mathscr{F}_t$-measurable, and

$$\{\sigma \leq \tau\} \cap \{\sigma \leq t\} = \{\sigma \leq t\} \cap \{\sigma \wedge t \leq \tau \wedge t\} \in \mathscr{F}_t \ \Rightarrow \ \{\sigma \leq \tau\} \in \mathscr{F}_\sigma,$$
$$\{\sigma \leq \tau\} \cap \{\tau \leq t\} = \{\sigma \leq t\} \cap \{\tau \leq t\} \cap \{\sigma \wedge t \leq \tau \wedge t\} \in \mathscr{F}_t \ \Rightarrow \ \{\sigma \leq \tau\} \in \mathscr{F}_\tau.$$

Then $\{\sigma \leq \tau\} \in \mathscr{F}_\sigma \cap \mathscr{F}_\tau = \mathscr{F}_{\sigma\wedge\tau}$, and $\{\sigma = \tau\} = \{\sigma \leq \tau\} \cap \{\sigma \geq \tau\} \in \mathscr{F}_{\sigma\wedge\tau}$.

(vii) We first assume that for each $t \geq 0$, the restriction $Y|_{\{\tau \leq t\}}$ is $\mathscr{F}_t$-measurable. Then for every Borel set $B \in \mathscr{B}(E)$, we have $\{Y \in B\} \cap \{\tau \leq t\} \in \mathscr{F}_t$. Since Y is defined on $\{\tau < \infty\}$, we have

$$\{Y \in B\} = \bigcup_{n=1}^{\infty} (\{Y \in B\} \cap \{\tau \leq n\}) \in \mathscr{F}_\infty.$$

Hence $\{Y \in B\} \in \mathscr{F}_\tau$. Conversely, if Y is $\mathscr{F}_\tau$-measurable, then $\{Y \in B\} \cap \{\tau \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$.

(viii) For every $t \geq 0$, $\{\tau_\infty \leq t\} = \bigcap_{n=1}^\infty \{\tau_n \leq t\} \in \mathscr{F}_t$.

(ix) For every $t \geq 0$, $\{\tau_\infty < t\} = \bigcup_{n=1}^\infty \{\tau_n < t\} \in \mathscr{F}_t$. Hence τ is a stopping time of $\{\mathscr{F}_{t+}\}_{t\geq 0}$ by Proposition 3.11 (i). By (v), we have $\mathscr{F}_{\tau+} \subset \mathscr{F}_{\tau_n+}$ for each $n \in \mathbb{N}$. Conversely, if $A \in \mathscr{F}_{\tau_n+}$ for each $n \in \mathbb{N}$,

$$A \cap \{\tau_\infty < t\} = \bigcup_{n=1}^{\infty} (A \cap \{\tau_n < t\}) \in \mathscr{F}_t, \ \forall t > 0 \ \Rightarrow \ A \in \mathscr{F}_{\tau_\infty+}.$$

Hence $\mathscr{F}_{\tau_\infty+} = \bigcap_{n=1}^\infty \mathscr{F}_{\tau_n+}$. Furthermore, if τ_n is stationary, then $\{\tau_\infty \leq t\} = \bigcup_{n=1}^\infty \{\tau_n \leq t\} \in \mathscr{F}_t$. Thus τ_∞ is a stopping time, and $\mathscr{F}_{\tau_\infty} \subset \mathscr{F}_{\tau_n}$ for each $n \in \mathbb{N}$ by (v). Conversely, if $A \in \mathscr{F}_{\tau_n}$ for each $n \in \mathbb{N}$,

$$A \cap \{\tau_\infty \leq t\} = \bigcup_{n=1}^{\infty} (A \cap \{\tau_n \leq t\}) \in \mathscr{F}_t, \ \forall t > 0 \ \Rightarrow \ A \in \mathscr{F}_{\tau_\infty}.$$

Hence $\bigcap_{n=1}^\infty \mathscr{F}_{\tau_n} = \mathscr{F}_{\tau_\infty}$. □

Proposition 3.13. *Let $X = (X_t)_{t\geq 0}$ be a progressive process of $(\mathscr{F}_t)_{t\geq 0}$. If T is a stopping time of $(\mathscr{F}_t)_{t\geq 0}$, then the function $X_\tau : \omega \mapsto X_{\tau(\omega)}(\omega)$, defined on the set $\{\tau < \infty\}$, is $\mathscr{F}_\tau$-measurable.*

Proof. By Proposition 3.12 (vii), it suffices to show that the restriction of X_τ to $\{\tau \leq t\}$ is $\mathscr{F}_t$-measurable for all $t \geq 0$. The restriction $X_\tau|_{\{\tau \leq t\}}$ is a composition of two measurable mappings:

$$\tau \wedge t \textit{ is } \mathscr{F}_t\textit{-measurable}: \ (\{\tau \leq t\}, \mathscr{F}_t) \to (\{F \leq t\} \times [0,t], \mathscr{F}_t \otimes \mathscr{B}([0,t])), \ \omega \mapsto (\omega, \tau(\omega) \wedge t)\,,$$
$$X \textit{ is progressive}: \ (\{\tau \leq t\} \times [0,t], \mathscr{F}_t \otimes \mathscr{B}([0,t])) \to (E, \mathscr{B}(E)), \ (\omega, s) \mapsto X_s(\omega).$$

Hence $X_\tau|_{\{\tau \leq t\}}$ is $\mathscr{F}_t$-measurable, and the result follows. □

Remark. We may also consider the discrete-time case. Let $(X_n)_{n=0}^{\infty}$ be an adapted sequence. Then for all $B \in \mathscr{B}(E)$, we have $\{X_\tau \in B\} \cap \{\tau \le t\} = \bigcup_{n=0}^{t} \{X_n \in B\} \cap \{\tau = n\} \in \mathscr{F}_t$ for every $t \in \mathbb{N}$. Then we have $\{X_\tau \in B\} \in \mathscr{F}_\tau$, and X_τ is always $\mathscr{F}_\tau$-measurable.

We introduce a common category of stopping times called hitting times.

Proposition 3.14. *Let $(X_t)_{t\ge 0}$ 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*

$$\tau_A = \inf \{t \ge 0 : X_t \in A\}.$$

Given a random time $\sigma : \Omega \to [0, \infty]$, 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 $(X_t)_{t\ge 0}$ is right-continuous and $G \subset E$ is an open set, then τ_G is a stopping time of the filtration $(\mathscr{F}_{t+})_{t\ge 0}$. Furthermore, if σ is a stopping time of the filtration $(\mathscr{F}_{t+})_{t\ge 0}$, so is τ_G^σ.

(ii) If $(X_t)_{t\ge 0}$ is continuous and $F \subset E$ is a closed set, then τ_F is a stopping time of the filtration $(\mathscr{F}_t)_{t\ge 0}$. Furthermore, if σ is a stopping time of the filtration $(\mathscr{F}_{t+})_{t\ge 0}$, 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

$$\{\tau_G < t\} = \bigcup_{q \in \mathbb{Q} \cap [0,t)} \{X_q \in G\} \in \mathscr{F}_t, \ \forall t > 0.$$

Then by Proposition 3.11 (i), τ_G is a stopping time of $\{\mathscr{F}_{t+}\}_{t\ge 0}$. Furthermore, if σ is a stopping time of $\{\mathscr{F}_{t+}\}_{t\ge 0}$, we have

$$\begin{aligned} \{\tau_G^\sigma < t\} &= \bigcup_{q \in (0,t)} \left(\{\sigma < q\} \cap \{\inf\{s \ge q : X_s \in G\} < t\}\right) \\ &= \bigcup_{q \in (0,t)} \left(\{\sigma < q\} \cap \left(\bigcup_{r \in \mathbb{Q} \cap [q,t)} \{X_r \in G\}\right)\right) \in \mathscr{F}_t, \ \forall t > 0. \end{aligned}$$

(ii) Fix $t \ge 0$. If $\tau_G(\omega) \le t$, choose $s_n \downarrow s := \tau_G(\omega)$ such that $X_{s_n} \in F$. Since $t \mapsto X_t(\omega)$ is continuous and F is closed, we have $X_{s_n}(\omega) \to X_s(\omega) \in F$. Hence

$$\{\tau_F \le t\} = \bigcup_{s \in [0,t]} \{X_s \in F\} = \left\{\inf_{q \in \mathbb{Q} \cap [0,t]} d(X_q, F) = 0\right\} \in \mathscr{F}_t, \ \forall t \ge 0.$$

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 $\mathscr{F}_t$-measurable for all $q \in \mathbb{Q} \cap [0, t]$, and countable infimum preserves measurability. Furthermore, if σ is a stopping time of $\{\mathscr{F}_{t+}\}_{t\ge 0}$, we have

$$\begin{aligned} \{\tau_F^\sigma < t\} &= \bigcup_{q \in (0,t)} \left(\{\sigma < q\} \cap \{\inf\{s \ge q : X_s \in F\} < t\}\right) \\ &= \bigcup_{q \in (0,t)} \left(\{\sigma < q\} \cap \left\{\inf_{r \in \mathbb{Q} \cap [q,t]} d(X_r, F) = 0\right\}\right) \in \mathscr{F}_t, \ \forall t > 0. \end{aligned}$$

Therefore τ_F^σ is a stopping time of $(\mathscr{F}_{t+})_{t\ge 0}$. $\square$

Remark. Similarly, given any Borel set $A \in \mathscr{B}(\mathbb{R})$, 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 $\{\tau_A \le n\} = \bigcup_{k=0}^n \{X_k \in A\}$. Furthermore, if σ is a stopping time, then the first hitting time τ_A^σ after σ is also a stopping time:

$$\{\tau_A^\sigma \le n\} = \bigcup_{k=0}^{n-1} (\{\sigma = k\} \cap \{k < \tau_A^\sigma \le n\}) = \bigcup_{k=0}^{n-1} \left(\{\sigma = k\} \cap \left(\bigcup_{j=k+1}^{n} \{X_j \in A\} \right) \right) \in \mathscr{F}_n, \ \forall n \in \mathbb{N}_0.$$

Finally we introduce a technical lemma about stopping times which resembles the form of simple function approximation.

Proposition 3.15. *Let τ be a stopping time.*

(i) If $\sigma : \Omega \to [0, \infty]$ is a $\mathscr{F}_\tau$-measurable random variable such that $\sigma \ge \tau$, then σ is also a stopping time.

(ii) Furthermore,

$$\tau_n = \frac{\lfloor 2^n \tau \rfloor + 1}{2^n} = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{\{k2^{-n} \le \tau < (k+1)2^{-n}\}} + \infty \mathbb{1}_{\{\tau = \infty\}}, \ n \in \mathbb{N}$$

is a sequence of stopping times decreasing to τ.

Proof. (i) Since σ is $\mathscr{F}_\tau$-measurable, we have $\{\sigma \le t\} \in \mathscr{F}_\tau$, and $\{\sigma \le t\} = \{\sigma \le t\} \cap \{\tau \le t\} \in \mathscr{F}_t$ for all $t \ge 0$. Hence σ is also a stopping time.

(ii) Note that $\tau_n(\omega) = \inf\{k2^{-n} : k2^{-n} > \tau(\omega), k \in \mathbb{Z}\}$. Then we have $\tau_n \downarrow \tau$. Since τ_n is a measurable function of τ, it is $\mathscr{F}_\tau$-measurable, hence a stopping time by the first assertion. $\square$

3.2 Discrete-time Martingales

3.2.1 Definition and Properties

Definition 3.16 (Discrete-time martingales)**.** Let $(X_n)_{n=0}^\infty$ be a real-valued and L^1 process that is adapted to the filtration $\{\mathscr{F}_n\}_{n=0}^\infty$. 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 $\mathbb{E}[X_n|\mathscr{F}_m] = X_m$ for all $n > m \geq 0$;

(ii) $(X_n)_{n=0}^\infty$ is said to be a *supermartingale* if $\mathbb{E}[X_n|\mathscr{F}_m] \leq X_m$ for all $n > m \geq 0$;

(iii) $(X_n)_{n=0}^\infty$ is said to be a *submartingale* if $\mathbb{E}[X_n|\mathscr{F}_m] \geq X_m$ for all $n > m \geq 0$;

All these notations depends on the choice of the filtration $\{\mathscr{F}_n\}_{n=0}^\infty$, which is fixed in later discussion.

Remark. The set of all martingales in $\left(L^1(\Omega, \mathscr{F}, \mathbb{P})\right)^{\mathbb{N}_0}$ is a vector space.

Proposition 3.17. *Let $f : \mathbb{R} \to \mathbb{R}$ 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. The proof simply uses conditional Jensen's inequality. Here are some useful corollaries:

(i) If $(X_n)_{n=0}^\infty$ is a martingale and $p \geq 1$, then $(|X_n|^p)_{n=0}^\infty$ is a submartingale;

(ii) If $(X_n)_{n=0}^\infty$ is a submartingale, then $(X_n^+)_{n=0}^\infty$ is a submartingale.

Definition 3.18 (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 $\mathscr{F}_{n-1}$-measurable for all $n \geq 1$. We define the integral (or the *martingale transform*) of $(H_n)_{n=0}^\infty$ with respect to an adapted process $(X_n)_{n=0}^\infty$ by

$$(H \cdot X)_0 = H_0 X_0, \quad (H \cdot X)_n = (H \cdot X)_{n-1} + H_n(X_n - X_{n-1}) = H_0 X_0 + \sum_{k=1}^n H_k(X_k - X_{k-1}), \ \forall n \in \mathbb{N}.$$

Clearly, $(H \cdot X)_n$ is an adapted process. We can easily check the following facts:

- If X_n is a martingale, so is $(H \cdot X)_n$.
- If X_n is a submartingale (or supermartingale) and H_n is nonnegative, so is $(H \cdot X)_n$.

Remark. If τ is a stopping time, the process $H_n = \mathbb{1}_{\{\tau \geq n\}}$ is nonnegative and predictable. Then

$$(H \cdot X)_n = X_0 + \sum_{k=1}^n \mathbb{1}_{\{\tau \geq n\}}(X_k - X_{k-1}) = X_{n \wedge \tau}.$$

Therefore, if $(X_n)_{n=1}^\infty$ is a submartingale, so is the stopped process $(X_n^\tau)_{n=1}^\infty = (X_{n\wedge\tau})_{n=1}^\infty$.

Theorem 3.19 (Doob's decomposition theorem)**.** *Let $(X_n)_{n=1}^\infty$ be a submartingale. Then there exists an increasing predictable L^1 process $(A_n)_{n=0}^\infty$ staring 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

$$M_n = X_0 + \sum_{k=1}^n \left(X_k - \mathbb{E}[X_k|\mathscr{F}_{k-1}]\right), \quad A_n = \sum_{k=1}^n \left(\mathbb{E}[X_k|\mathscr{F}_{k-1}] - X_{k-1}\right), \ \forall n \geq 1.$$

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 $Y_n = \mathbb{E}[Y_n|\mathscr{F}_{n-1}] = Y_{n-1} = \cdots = Y_0 = 0$. Hence $Y_n \equiv 0$. ☐

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 $(A_n)_{n=0}^\infty$ staring 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 3.20 (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 $\mathbb{E}[X_N|\mathscr{F}_\tau] \geq X_\tau$;

(iii) If σ is another bounded stopping time and $\sigma \leq \tau$, then $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] \geq X_\sigma$.

Proof. (i) By definition, we have $\{\tau \geq k\} = \{\tau \leq k-1\}^c \in \mathscr{F}_{k-1}$ for all $k \in \mathbb{N}$. Then

$$\begin{aligned}\mathbb{E}[X_\tau] = \mathbb{E}\left[X_0 + \sum_{k=1}^N (X_k - X_{k-1})\mathbb{1}_{\{\tau\geq k\}}\right] &= \mathbb{E}[X_0] + \sum_{k=1}^N \mathbb{E}\left[(X_k - X_{k-1})\mathbb{1}_{\{\tau\geq k\}}\right] \\ &= \mathbb{E}[X_0] + \sum_{k=1}^N \mathbb{E}\big[\underbrace{\mathbb{E}[(X_k - X_{k-1})|\mathscr{F}_{k-1}]}_{\geq 0}\mathbb{1}_{\{\tau\geq k\}}\big] \geq \mathbb{E}[X_0].\end{aligned}$$

(ii) If $A \in \mathscr{F}_\tau$, we have $A \cap \{\tau = n\} \in \mathscr{F}_n$ for all $N \in \mathbb{N}_0$, and

$$\begin{aligned}\mathbb{E}[X_N\mathbb{1}_A] = \sum_{n=0}^N \mathbb{E}\left[X_N\mathbb{1}_{A\cap\{\tau=n\}}\right] &= \sum_{n=0}^N \mathbb{E}\left[\mathbb{E}\left[X_N|\mathscr{F}_n\right]\mathbb{1}_{A\cap\{\tau=n\}}\right] \\ \geq \sum_{n=0}^N \mathbb{E}\left[X_n\mathbb{1}_{A\cap\{\tau=n\}}\right] &= \sum_{n=0}^N \mathbb{E}\left[X_\tau\mathbb{1}_{A\cap\{\tau=n\}}\right] = \mathbb{E}[X_\tau\mathbb{1}_A].\end{aligned}$$

Since X_τ is $\mathscr{F}_\tau$-measurable, we have $\mathbb{E}[X_N|\mathscr{F}_\tau] \geq X_\tau$.

(iii) Since $\sigma \leq \tau \leq N$, we have $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$. 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),

$$M_\sigma = \mathbb{E}[M_N|\mathscr{F}_\sigma] = \mathbb{E}[\mathbb{E}[M_N|\mathscr{F}_\tau]|\mathscr{F}_\sigma] = \mathbb{E}[M_\tau|\mathscr{F}_\sigma].$$

Clearly, $A_\tau \geq A_\sigma$, and $\mathbb{E}[A_\tau|\mathscr{F}_\sigma] \geq \mathbb{E}[A_\sigma|\mathscr{F}_\sigma] = A_\sigma$. Hence $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] = \mathbb{E}[M_\tau + A_\tau|\mathscr{F}_\sigma] \geq M_\sigma + A_\sigma = X_\sigma$. ☐

We give another characterization of martingales.

Theorem 3.21. *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 $\mathbb{E}[X_\tau] = E[X_0]$ for every bounded stopping time τ.*

Proof. The direction "⇒" follows by Theorem 3.20 (i). To prove the converse, assume that $\mathbb{E}[X_\tau] = E[X_0]$ for every bounded stopping time τ. To prove that $(X_n)_{n=0}^\infty$ is a martingale, we show that for each $n \in \mathbb{N}$,

$$\mathbb{E}[X_n|\mathscr{F}_{n-1}] = X_{n-1} \Leftrightarrow\ \mathbb{E}[X_n\mathbb{1}_A] = \mathbb{E}[X_{n-1}\mathbb{1}_A],\ \forall A \in \mathscr{F}_{n-1}.$$

Let $\tau = (n-1)\mathbb{1}_A + n\mathbb{1}_{A^c}$, so τ is a bounded stopping time, and $\mathbb{E}[X_0] = \mathbb{E}[X_\tau] = \mathbb{E}[X_{n-1}\mathbb{1}_A] + \mathbb{E}[X_n\mathbb{1}_{A^c}]$. Since n is a constant stopping time, we have $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. Then $\mathbb{E}[X_n\mathbb{1}_A] = \mathbb{E}[X_{n-1}\mathbb{1}_A]$. ☐

3.2.2 Martingale Convergence Theorems and Application

Now we discuss the convergence of martingales.

Definition 3.22 (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

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le n$$

of integers such that $x_{s_j} \le a$ and $x_{t_j} \ge b$ for all $j \in \{1, \cdots, k\}$. 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 ∞.

Definition 3.23 (Upcrossing number)**.** Given an adapted random variable 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 $n \in \overline{\mathbb{N}}$. It can be depicted by a sequence of stopping times. Let $\tau_0 = -\infty$, and define

$$\sigma_j = \min\left\{k \in \mathbb{N}_0 : k > \tau_{j-1},\ X_k \le a\right\},\ \tau_j = \min\left\{k \in \mathbb{N}_0 : k > \sigma_j,\ X_k \ge b\right\},\ j \ge 1.$$

Then the upcrossing number of $(x_n)_{n=0}^\infty$ along $[a, b]$ before time $n \in \mathbb{N}$ is $U_{[a,b]}^X(n) = \max\{j : \tau_j \le n\}$, and $U_{[a,b]}^X(\infty) = \lim_{n\to\infty} U_{[a,b]}^X(n) = \max\{j : \tau_j < \infty\} = \min\{j : \tau_j = \infty\}$.

Remark. The upcrossing number $U_{[a,b]}^X(n)$ is an integer-valued random variable.

Proposition 3.24 (Doob's upcrossing inequality)**.** *If $(X_n)_{n=0}^\infty$ is a submartingale, then for all real numbers $a < b$ and all $n \in \mathbb{N}$,*

$$\mathbb{E}\left[U_{[a,b]}^X(n)\right] \le \frac{\mathbb{E}[(X_n - a)^+ - (X_0 - a)^+]}{b - a}. \tag{3.1}$$

Proof. Define stopping times $\{\sigma_j\}_{j=1}^\infty$ and $\{\tau_j\}_{j=1}^\infty$ as in Definition 3.23. Then $\{U_{[a,b]}^X(n) \ge k\} = \{\tau_k \le n\}$, and

$$H_n = \begin{cases} 1, & \text{if } \sigma_k < n \le \tau_k \text{ for some } k \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

defines a nonnegative predictable process, since

$$\{H_n = 1\} = \bigcup_{m=1}^{n-1} \left(\{X_m \le a\} \cap \bigcap_{j=m+1}^{n-1} \{X_j < b\} \right).$$

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, $H \cdot Y$ satisfies

$$(H \cdot Y)_n = 0 + \sum_{k=1}^n H_n(Y_n - Y_{n-1}) \ge (b - a) \cdot U_{[0,b-a]}^Y(n) = (b - a) \cdot U_{[a,b]}^X(n)$$

Note that $(1 - H) \cdot Y$ is also a submartingale. Then

$$(b - a) \cdot \mathbb{E}\left[U_{[a,b]}^X(n)\right] \le \mathbb{E}\left[(H \cdot Y)_n\right] = \mathbb{E}\left[Y_n - ((1 - H) \cdot Y)_n\right] \le \mathbb{E}[Y_n] - \mathbb{E}[(1 - H_0)Y_0] = \mathbb{E}[Y_n] - \mathbb{E}[Y_0].$$

This is indeed the inequality (3.1). ☐

We present Doob's first martingale convergence theorem below.

Theorem 3.25 (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* $X_\infty = \lim_{n\to\infty} X_n$ *a.s. exists, and* $X_\infty \in L^1(\Omega, \mathscr{F}_\infty, \mathbb{P})$.

Proof. Let $M := \sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] < \infty$. Then for all $a \in \mathbb{R}$,

$$(x-a)^+ \le x^+ + a^-,\ \forall x \in \mathbb{R} \ \Rightarrow\ \mathbb{E}[(X_n - a)^+] \le \mathbb{E}[X_n] + a^- \le M + a^-,\ \forall n \in \mathbb{N}.$$

By Fatou's lemma and Proposition 3.24, the total upcrossing number $U_{[a,b]}^X(n) \uparrow U_{[a,b]}^X(\infty)$ satisfies

$$\mathbb{E}\left[U_{[a,b]}^X(\infty)\right] \le \liminf_{n\to\infty} \mathbb{E}\left[U_{[a,b]}^X(n)\right] \le \liminf_{n\to\infty} \frac{\mathbb{E}[(X_n-a)^+]}{b-a} \le \frac{M+|a|}{b-a} < \infty,\ \forall \mathbb{R} \ni b > a.$$

Then for all real numbers $a < b$, we have

$$\mathbb{P}\left(U_{[a,b]}^X(\infty) < \infty\right) = 1.$$

Consequently,

$$\begin{aligned}\mathbb{P}\left(\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right) &= \mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}:a<b} \left\{\liminf_{n\to\infty} X_n \le a < b \le \limsup_{n\to\infty} X_n\right\}\right)\\ &= \mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}:a<b} \left\{U_{[a,b]}^X(\infty) < \infty\right\}\right) = 0.\end{aligned}$$

Therefore $X_\infty = \lim_{n\to\infty} X_n$ a.s. exists. Now we prove that $X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. By Fatou's lemma,

$$\mathbb{E}[X_\infty^+] \le \liminf_{n\to\infty} \mathbb{E}[X_n^+] \le M < \infty,$$

and

$$\mathbb{E}[X_\infty^-] \le \liminf_{n\to\infty} \mathbb{E}[X_n^-] = \liminf_{n\to\infty} \mathbb{E}[X_n^+ - X_n] \le M - \mathbb{E}[X_0] < \infty.$$

Therefore $\mathbb{E}|X_\infty| < \infty$, and $X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Since every X_n is $\mathscr{F}_\infty$-measurable, so is the a.s. limit X_∞. ☐

Remark. The Theorem 3.25 does not imply L^1-convergence. As a counterexample, 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. Let $\mathscr{F}_n$ 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 $\mathbb{E}[Y_n^+] \le 1$, the sequence Y_n converges *a.s.*. Furthermore, $Y_n \to Y_\infty = 1$ *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 3.26 (Supermartingale convergence theorem)**.** *A nonnegative supermartingale* $(X_n)_{n=0}^\infty$ *converges almost surely, and the limit* X_∞ *satisfies* $\mathbb{E}X_\infty \le \mathbb{E}X_0$.

Proof. Since $(-X_n)_{n=0}^\infty$ is a submartingale and $\mathbb{E}[(-X_n)^+] = 0$ for all n, it converges almost surely to some $-X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ by Theorem 3.25. The inequality $\mathbb{E}X_\infty \le \mathbb{E}X_0$ follows from Fatou's lemma. ☐

Next we show some applications of the Martingale convergence theorem.

Proposition 3.27 (Bounded increments)**.** *Let* $(X_n)_{n=0}^\infty$ *be a martingale such that* $|X_n - X_{n-1}| < K$ *a.s. for all* $n \in \mathbb{N}$*, where* $0 < K < \infty$*. Then with probability* 1,

- *either the martingale* $(X_n)_{n=1}^\infty$ *converges to a finite limit,*
- *or* $\liminf_{n\to\infty} X_n = -\infty$ *and* $\limsup_{n\to\infty} X_n = \infty$.

Proof. We may assume $X_0 = 0$ without loss of generality. Let $N \in (0, \infty)$ and $\tau_N = \inf\{n \in \mathbb{N} : X_n \le -N\}$, which is a stopping time. Then the stopped martingale $(X_{n\wedge\tau_N})_{n=1}^\infty$ is bounded below by $-N - K > -\infty$, and $X_{n\wedge\tau_N} + K + N$ converges almost surely by Theorem 3.26. Consequently, $\lim_{n\to\infty} X_n$ exists and is finite everywhere on $\{\tau_N = \infty\}$, except on a $\mathbb{P}$-null set. This statement also holds on

$$\bigcup_{N=1}^{\infty}\{\tau_N = \infty\} = \bigcup_{N=1}^{\infty}\left\{\inf_{n\in\mathbb{N}} X_n > -N\right\} = \left\{\liminf_{n\to\infty} X_n > -\infty\right\}.$$

Therefore, with probability 1, either $(X_n)_{n=1}^\infty$ converges to a finite limit or $\liminf_{n\to\infty} X_n = -\infty$. Applying the same conclusion on $(-X_n)_{n=1}^\infty$, either $(X_n)_{n=1}^\infty$ converges to a finite limit or $\limsup_{n\to\infty} X_n = \infty$. $\square$

We are applying this result to prove the second Borel-Cantelli lemma.

Theorem 3.28 (Second Borel-Cantelli Lemma)**.** *Let $(\mathscr{F}_n)_{n=1}^\infty$ be a filtration with $\mathscr{F}_0 = \{\emptyset, \Omega\}$, and $(E_n)_{n=1}^\infty$ an adapted event sequence, i.e. $E_n \in \mathscr{F}_n$ for each $n \in \mathbb{N}$. Then*

$$\left\{\sum_{n=1}^{\infty}\mathbb{P}(E_n|\mathscr{F}_{n-1}) = \infty\right\} = \{(E_n)_{n=1}^\infty \textit{ occurs infinitely often}\} \quad a.s..$$

Remark. We say two measurable sets $A = B$ a.s., if $A\Delta B = (A\backslash B)\cup(B\backslash A)$ has probability zero.

Proof. Let $X_n = \sum_{k=1}^n \mathbb{1}_{E_k}$, which is a submartingale. By Doob's decomposition theorem, we take $X_n = M_n + A_n$, where M_n is a martingale and A_n is a predictable increasing sequence. To be more specific,

$$M_n = \sum_{k=1}^{n}\mathbb{1}_{E_k} - \mathbb{P}(E_k|\mathscr{F}_{k-1}), \quad A_n = \sum_{k=1}^{n}\mathbb{P}(E_k|\mathscr{F}_{k-1}).$$

Note that both $(X_n)_{n=1}^\infty$ and $(A_n)_{n=1}^\infty$ are monotone increasing. We need to show that

$$\{A_\infty = \infty\} = \{X_\infty = \infty\} \quad a.s..$$

Since $|M_n - M_{n-1}| \le 1$, by Proposition 3.27, with probability 1, exactly one of the following cases holds:

- $M_n = X_n - A_n$ converges to a finite limit. On this event, $X_\infty = \infty$ if and only if $A_\infty = \infty$.
- $\liminf_{n\to\infty} M_n = -\infty$ and $\limsup_{n\to\infty} M_n = \infty$. On this event, we have both $X_\infty = \infty$ and $A_\infty = \infty$.

Then we complete the proof. $\square$

Corollary 3.29 (Second Borel-Cantelli lemma)**.** *If $(E_n)_{n=1}^\infty$ is a sequence of independent events such that $\sum_{n=1}^\infty \mathbb{P}(E_n) = \infty$, then*

$$\mathbb{P}\left((E_n)_{n=1}^\infty \textit{ occurs infinitely often}\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}E_n\right) = 1.$$

Next we discuss the L^p convergence of martingales.

Proposition 3.30 (Maximal inequality)**.** *If $(X_n)_{n=0}^\infty$ is a submartingale, then for every $n \in \mathbb{N}$,*

$$\lambda\mathbb{P}\left(\max_{0\le k\le n} X_k^+ \ge \lambda\right) \le \mathbb{E}\left[X_n^+\mathbb{1}\left\{\max_{0\le k\le n} X_k^+ \ge \lambda\right\}\right] \le \mathbb{E}[X_n^+], \quad \forall\lambda > 0, \tag{3.2}$$

and

$$\lambda\mathbb{P}\left(\max_{0\le k\le n}|X_k| \ge \lambda\right) \le 2\mathbb{E}|X_n| + \mathbb{E}|X_0|, \quad \forall\lambda > 0. \tag{3.3}$$

Proof. Let $\tau = n \wedge \min\{m \in \mathbb{N}_0 : X_m^+ \geq \lambda\}$, and $E = \{\max_{0\leq k\leq n} X_k^+ \geq \lambda\}$. Then τ is a bounded stopping time, and $E = \{X_\tau^+ \geq \lambda\}$. By Markov's inequality,

$$\lambda \mathbb{P}(E) = \lambda \mathbb{P}(X_\tau^+ \geq \lambda) \leq \mathbb{E}\left[X_\tau^+ \mathbb{1}_E\right]. \tag{3.4}$$

Since (X_n^+) is a submartingale, we have $\mathbb{E}[X_\tau^+] \leq \mathbb{E}[X_n^+]$ by Theorem 3.20. Noticing that $X_\tau = X_n$ on E^c, we have $\mathbb{E}[X_\tau^+ \mathbb{1}_E] \leq \mathbb{E}[X_n^+ \mathbb{1}_E]$. Then the first inequality in (3.2) follows, and the second is trivial.

The proof of (3.3) is similar, but we let $\tau = n \wedge \min\{m \in \mathbb{N}_0 : |X_m| \geq \lambda\}$. By Markov's inequality and applying Theorem 3.20 on supermartingales (X_n) and (X_n^+), we have

$$\lambda \mathbb{P}\left(|X_\tau| \geq \lambda\right) \leq \mathbb{E}|X_\tau| = 2\mathbb{E}[X_\tau^+] - \mathbb{E}[X_\tau] \leq 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0] \leq 2\mathbb{E}|X_n| + \mathbb{E}|X_0|.$$

Then we finish the proof. □

Proposition 3.31 (Doob's L^p-inequality)**.** *If $(X_n)_{n=0}^\infty$ is a submartingale and $1 < p < \infty$, then*

$$\mathbb{E}\left[\max_{0\leq k\leq n} (X_n^+)^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[(X_n^+)^p\right], \ \forall n \in \mathbb{N}. \tag{3.5}$$

Proof. We use a corollary of Fubini's theorem: for $p > 0$ and a nonnegative random variable $Y \in L^p(\Omega, \mathscr{F}, \mathbb{P})$,

$$\mathbb{E}[Y^p] = \int\int_0^\infty p\lambda^{p-1}\mathbb{1}_{\{\lambda\leq Y\}}\mathrm{d}\lambda\,\mathrm{d}\mathbb{P} = \int_0^\infty p\lambda^{p-1}\mathbb{P}(Y \geq \lambda)\,\mathrm{d}\lambda.$$

Let $(X_n)_{n=0}^\infty$ be a submartingale in L^p, and $Y := \max_{0\leq k\leq n} X_n^+$. For each $M > 0$,

$$\begin{aligned}\mathbb{E}[(Y\wedge M)^p] &= \int_0^\infty p\lambda^{p-1}\mathbb{P}(Y\wedge M \geq \lambda)\,\mathrm{d}\lambda \leq \int_0^\infty p\lambda^{p-2}\mathbb{E}[X_n^+\mathbb{1}_{\{Y\wedge M\geq\lambda\}}]\,\mathrm{d}\lambda \qquad \text{(By Proposition 3.30)}\\ &= \mathbb{E}\left[X_n^+\int_0^\infty p\lambda^{p-2}\mathbb{1}_{\{Y\wedge M\geq\lambda\}}\,\mathrm{d}\lambda\right] = \left(\frac{p}{p-1}\right)\mathbb{E}\left[X_n^+(Y\wedge M)^{p-1}\right]. \end{aligned}\tag{3.6}$$

Note that $q = p/(p-1)$ is the conjugate exponent of p, by Hölder's inequality,

$$\mathbb{E}\left[X_n^+(Y\wedge M)^{p-1}\right] \leq \left(\mathbb{E}[(X_n^+)^p]\right)^{1/p}\left(\mathbb{E}[(Y\wedge M)^{(p-1)q}]\right)^{1/q} = \left(\mathbb{E}[(X_n^+)^p]\right)^{1/p}\left(\mathbb{E}[(Y\wedge M)^p]\right)^{1-1/p}. \tag{3.7}$$

Combining (3.6) and (3.7), we have $\mathbb{E}[(Y\wedge M)^p] \leq \mathbb{E}[(X_n^+)^p]$ for all $M > 0$. Letting $M \to \infty$, the monotone convergence theorem implies (3.5). □

Remark. As $p \downarrow 1$, the coefficient $\left(\frac{p}{p-1}\right)^p$ blows up, and an estimate of the same form does not exist for $p = 1$. As a counterexample, we consider the Gambler's ruin: A gambler has 1 dollar, and in each play he earns or loses 1 dollar with probability 1/2. He exits the game until he loses all his money. To model the game, let $S_0 = 1$, and $S_n = 1 + \sum_{k=1}^n \xi_k$, where $\xi_1, \xi_2, \cdots$ is a sequence of i.i.d. Rademacher variables, and let $\tau_0 = \inf\{n \geq 0 : S_n = 0\}$. The stopped martingale $X_n = S_{n\wedge\tau_0}$ is the money the gambler holds after n plays.

According to the martingale property, $\mathbb{E}[X_n] = 1$ for each $n \in \mathbb{N}$. If we let $\tau_M = \inf\{n \geq 0 : X_n = M\}$, the martingale $(X_{n\wedge\tau_M})$ will converge to a random variable X_{τ_M} valued in $\{0, M\}$, and $\mathbb{E}[X_{\tau_M}] = 1$ by dominated convergence theorem. Hence $\mathbb{P}(\max_{m\geq 0} X_m \geq M) = M^{-1}$, and

$$\mathbb{E}\left[\max_{1\leq m<\infty} X_m\right] = \sum_{M=1}^\infty \mathbb{P}\left(\max_{m\geq 0} X_m \geq M\right) = \sum_{M=1}^\infty \frac{1}{M} = \infty.$$

By monotone convergence, $\mathbb{E}\left[\max_{1\leq m\leq n} X_m\right] \uparrow \infty$ as $n \to \infty$, which cannot be bounded in terms of $\mathbb{E}[X_n] = 1$.

Theorem 3.32 (Convergence theorem for L^p-bounded 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 $1<p<\infty$, 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 $X_\infty\in L^p(\Omega,\mathscr{F},\mathbb{P})$ such that $X_n\to X_\infty$ *a.s.*. By Doob's L^p-inequality [Proposition 3.22] and monotone convergence theorem,

$$\mathbb{E}\left[\max_{1\le k\le n}|X_k|^p\right]\le\left(\frac{p}{p-1}\right)^p\mathbb{E}|X_n|^p\le\left(\frac{p}{p-1}\right)^p M\quad\Rightarrow\quad\mathbb{E}[Y^p]\le\left(\frac{p}{p-1}\right)^p M<\infty.$$

Since $|X_n-X_\infty|\le 2Y$, by Lebesgue dominated convergence theorem, $\|X_n-X_\infty\|_p\to 0$. ☐

Next we discuss the convergence of uniformly integrable martingales.

Theorem 3.33 (Convergence theorem for uniformly integrable submartingales)**.** *For a submartingale $(X_n)_{n=0}^\infty$, the following are equivalent:*

(i) $(X_n)_{n=0}^\infty$ is uniformly integrable.
(ii) $(X_n)_{n=0}^\infty$ converges a.s. and in L^1.
(iii) $(X_n)_{n=0}^\infty$ converges in L^1.

Proof. We first show that (i) $\Rightarrow$ (ii). Since $(X_n)_{n=0}^\infty$ is uniformly integrable, we have $\sup_{n\in\mathbb{N}_0}\mathbb{E}|X_n|<\infty$. By Theorem 3.25, X_n converges a.s. to some $X_\infty\in L^1(\Omega,\mathscr{F},\mathbb{P})$. By uniform integrability of $(X_n)_{n=1}^\infty$, the convergence also holds in L^1. The implication (ii) $\Rightarrow$ (iii) is trivial, and (iii) $\Rightarrow$ (i) is by Theorem 1.75. ☐

Proposition 3.34. *Given $Z\in L^1(\Omega,\mathscr{F},\mathbb{P})$, the following collection is uniformly integrable:*

$$\{\mathbb{E}[Z|\mathscr{G}]:\mathscr{G}\text{ is a sub-}\sigma\text{-algebra of }\mathscr{F}\}$$

Proof. Since Z is integrable, by Theorem1.58, for every $\epsilon>0$, there exists $\delta>0$ such that $\mathbb{E}[|Z|\mathbb{1}_A]<\epsilon$ for all $A\in\mathscr{F}$ with $\mathbb{P}(A)<\delta$. Given $M>0$, and define $X_\mathscr{G}=\mathbb{E}[Z|\mathscr{G}]$, $Y_\mathscr{G}=\mathbb{E}[|Z||\mathscr{G}]$. Then $|X_\mathscr{G}|\le Y_\mathscr{G}$, and

$$\mathbb{E}[|X_\mathscr{G}|\mathbb{1}_{\{|X_\mathscr{G}|>M\}}]\le\mathbb{E}[Y_\mathscr{G}\mathbb{1}_{\{Y_\mathscr{G}>M\}}]=\mathbb{E}[|Z|\mathbb{1}_{\{Y_\mathscr{G}>M\}}].$$

By Chebyshev inequality, $\mathbb{P}(Y_\mathscr{G}>M)\le\mathbb{E}[Y_\mathscr{G}]/M=\mathbb{E}|Z|/M$. If $M>\mathbb{E}|Z|/\delta$, we have $\mathbb{E}[|Z|\mathbb{1}_{\{Y_\mathscr{G}>M\}}]<\epsilon$. Note the choice of M is independent of $\mathscr{G}$. Since $\epsilon>0$, we have

$$0\le\lim_{M\to\infty}\sup_{\mathscr{G}\subset\mathscr{F}}\mathbb{E}\left[|X_\mathscr{G}|\mathbb{1}_{\{|X_\mathscr{G}|>M\}}\right]\le\lim_{M\to\infty}\sup_{\mathscr{G}\subset\mathscr{F}}\mathbb{E}\left[|Z|\mathbb{1}_{\{|Y_\mathscr{G}|>M\}}\right]=0.$$

Hence the collection of conditional expectations is uniformly integrable. ☐

Theorem 3.35 (Doob's convergence theorem for uniformly integrable martingales)**.** *For a martingale $(X_n)_{n=0}^\infty$, the following are equivalent:*

(i) $(X_n)_{n=0}^\infty$ is uniformly integrable.
(ii) $(X_n)_{n=0}^\infty$ converges a.s. and in L^1.
(iii) $(X_n)_{n=0}^\infty$ converges in L^1.
(iv) $(X_n)_{n=0}^\infty$ is closed, i.e. there exists $Z\in L^1(\Omega,\mathscr{F},\mathbb{P})$ such that $X_n=\mathbb{E}[Z|\mathscr{F}_n]$ for all $n\in\mathbb{N}_0$.

Proof. By Theorem 3.33, we have (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii). To show (iii) $\Rightarrow$ (iv), we let $X_\infty=\lim_{n\to\infty}X_n$ in L^1. Since the conditional expectation is a bounded linear operator on $L^1(\Omega,\mathscr{F},\mathbb{P})$,

$$\mathbb{E}[X_\infty|\mathcal{F}_n]=\lim_{m\to\infty}\mathbb{E}[X_m|\mathcal{F}_n]=X_n.$$

Finally, (iv) $\Rightarrow$ (i) is by Proposition 3.34. ☐

Corollary 3.36 (Lévy's upward theorem). *Let $(\mathscr{F}_n)_{n=0}^{\infty}$ be a filtration. If $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$,*

$$\mathbb{E}[Z|\mathscr{F}_n] \to \mathbb{E}[Z|\mathscr{F}_\infty] \quad \textit{a.s. and in } L^1.$$

Proof. We define a uniformly integrable martingale $X_n = \mathbb{E}[Z|\mathscr{F}_n]$, which converges both a.s. and in L^1. Let $X_\infty \in L^1(\Omega, \mathscr{F}_\infty, \mathbb{P})$ be the a.s. and L^1 limit. It suffices to show that $X_\infty = \mathbb{E}[Z|\mathscr{F}_\infty]$. Let $A \in \bigcup_{n=0}^{\infty}$. Then $A \in \mathscr{F}_n$ for some $n \in \mathbb{N}_0$, and

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X_\infty \mathbb{1}_A|\mathscr{F}_n]] = \mathbb{E}[\mathbb{E}[X_\infty|\mathscr{F}_n]\mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A].$$

Since $\bigcup_{n=0}^{\infty} \mathscr{F}_n$ is a π-system, and the sets E satisfying $\mathbb{E}[X_\infty \mathbb{1}_E] = \mathbb{E}[Z\mathbb{1}_E]$ is a λ-system, by Sierpiński-Dynkin π-λ Theorem, we have $\mathbb{E}[X_\infty \mathbb{1}_E] = \mathbb{E}[Z\mathbb{1}_E]$ for all $E \in \mathscr{F}_\infty$, and the results follows by definition of conditional expectation. ☐

Remark. Given $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, the martingale $X_n = \mathbb{E}[Z|\mathscr{F}_n]$ is also called a *Doob's martingale*. According to Theorem 3.35, every uniformly integrable martingale is a Doob martingale. Furthermore, even if the choice random variable Z in Theorem 3.35 (iv) is not unique, by Corollary 3.36, the conditional expectation $\mathbb{E}[Z|\mathscr{F}_\infty]$ is unique and equals X_∞.

We discuss two 0-1 laws, which can be proved by constructing Doob martingales.

Corollary 3.37 (Levy's 0-1 law). *If $\mathscr{F}_n \uparrow \mathscr{F}_\infty$ and $A \in \mathscr{F}_\infty$, then $\mathbb{P}(A|\mathscr{F}_n) \to \mathbb{1}_A$ a.s. and in L^1.*

Corollary 3.38 (Kolmogorov's 0-1 law). *Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent random variables, and $\mathscr{G}_n = \sigma(X_{n+1}, X_{n+2}, \cdots)$ for all $n \in \mathbb{N}_0$. Define the* ***tail σ-algebra*** *$\mathscr{G}_\infty$ as follows:*

$$\mathscr{G}_\infty = \bigcap_{n \in \mathbb{N}_0} \mathscr{G}_n.$$

Then $\mathscr{G}_\infty$ is $\mathbb{P}$-trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathscr{G}_\infty$.

Proof. Let $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$, which is independent of $\mathscr{G}_n$. Then for all $n \in \mathbb{N}_0$, we have $A \in \mathscr{G}_\infty \subset \mathscr{G}_n$, and $\mathbb{P}(A|\mathscr{F}_n) = \mathbb{P}(A)$. Also $A \in \mathscr{F}_\infty$, and by Corollary 3.37, $\mathbb{P}(A|\mathscr{F}_n) \to \mathbb{1}_A$. Hence $\mathbb{P}(A) \in \{0, 1\}$. ☐

3.2.3 Doob's Optional Stopping Theorem

In this part, we study the expectation of X_τ, where τ is a stopping time.

Theorem 3.39 (Optional stopping theorem for nonnegative supermartingales). *Let $(X_n)_{n=1}^{\infty}$ be a nonngeative supermartingale, and let τ be a stopping time. Then*

$$\mathbb{E}[X_\tau] \le \mathbb{E}[X_0].$$

Proof. The nonnegative supermartingale (X_n) has an a.s. limit X_∞, which is integrable. Also, the stopped process $X_{n\wedge\tau} \to X_\tau$ a.s.. By Fatou's lemma,

$$\mathbb{E}[X_\tau] \le \liminf_{n\to\infty} \mathbb{E}[X_{n\wedge\tau}]$$

Note that $n \wedge \tau$ is a bounded stopping time, for each $n \in \mathbb{N}$,

$$\mathbb{E}[X_{n\wedge\tau}] \le \mathbb{E}[X_0].$$

Hence $\mathbb{E}[X_\tau] \le \mathbb{E}[X_0]$. ☐

Theorem 3.40 (Optional stopping theorem)**.** *Let* $(X_n)_{n=1}^{\infty}$ *be a submartingale, and let* τ *be a stopping time. If* <u>*at least one*</u> *of the following conditions holds:*

(i) τ *is a bounded stopping time;*

(ii) $\mathbb{E}[\tau] < \infty$*, and there exists* $c > 0$ *such that* $\mathbb{E}[|X_{n+1} - X_n| \, |\mathscr{F}_n] \le c$ *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* $\mathbb{E}[X_\tau] \ge \mathbb{E}[X_0]$.

Proof. The case (i) is proved in Theorem 3.20. To prove the case (ii), we write the stopped process as

$$X_n^\tau = X_0 + \sum_{k=1}^{n} (X_k - X_{k-1}) \mathbb{1}_{\{k \le \tau\}}, \quad \forall n \in \mathbb{N}_0,$$

which is dominated by

$$|X_n^\tau| \le |X_0| + \sum_{k=1}^{\infty} |X_k - X_{k-1}| \mathbb{1}_{\{\tau \ge k\}}.$$

Note that

$$\begin{aligned}\mathbb{E}\left[|X_0| + \sum_{k=1}^{\infty} |X_k - X_{k-1}| \mathbb{1}_{\{\tau \ge k\}} \right] &= \mathbb{E}|X_0| + \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left[|X_k - X_{k-1}| |\mathscr{F}_{k-1} \right] \mathbb{1}_{\{\tau \ge k\}} \right] < \infty. \\ &\le \mathbb{E}|X_0| + c\mathbb{E}[\tau].\end{aligned}$$

Hence (X_n^τ) is dominated by an integrable random variable, which implies (iii). Finally, if (iii) holds, we use the following fact:

$$\mathbb{E}[X_n^\tau] = \mathbb{E}[X_{n \wedge \tau}] \ge \mathbb{E}[X_0], \quad \forall n \in \mathbb{N}_0,$$

which holds because $n \wedge \tau$ is a bounded stopping time. By uniform integrability, the convergence $X_n^\tau \to X_\tau$ holds both a.s. and in L^1, and $\mathbb{E}[X_0] \le \mathbb{E}[X_n^\tau] \to \mathbb{E}[X^\tau]$. $\square$

For uniformly integrable martingales, we have a stronger optional stopping theorem.

Theorem 3.41 (Optional stopping theorem for uniformly integrable martingales)**.** *Let* $(X_n)_{n=1}^{\infty}$ *be a uniformly integrable martingale, and* $X_\infty = \lim_{n\to\infty} X_n$ *a.s.. If* τ *is a stopping time, then* $X_\tau \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ *and*

$$X_\tau = \mathbb{E}[X_\infty | \mathscr{F}_\tau]$$

with the convention that $X_\tau = X_\infty$ *on* $\{\tau = \infty\}$*. If* $\sigma \le \tau$ *is another stopping time, then* $\mathbb{E}[X_\tau | \mathscr{F}_\sigma] = X_\sigma$*.*

Proof. By Theorem 3.35, we have $\mathbb{E}[X_\infty | \mathscr{F}_n] = X_n$. Then for all $A \in \mathscr{F}_\tau$, $A \cap \{\tau = n\} \in \mathscr{F}_n$, and

$$\begin{aligned}\mathbb{E}[X_\infty \mathbb{1}_A] &= \sum_{n=0}^{\infty} \mathbb{E}\left[X_\infty \mathbb{1}_{A \cap \{\tau = n\}} \right] + \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{\tau = \infty\}}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left[X_n \mathbb{1}_{A \cap \{\tau = n\}} \right] + \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{\tau = \infty\}}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left[X_\tau \mathbb{1}_{A \cap \{\tau = n\}} \right] + \mathbb{E}[X_\tau \mathbb{1}_{A \cap \{\tau = \infty\}}] = \mathbb{E}[X_\tau \mathbb{1}_A].\end{aligned}$$

Since X_τ is $\mathscr{F}_\tau$-measurable, we have $\mathbb{E}[X_\infty | \mathscr{F}_\tau] = X_\tau$. Furthermore, if $\sigma \le \tau$ is another stopping time, then $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$, and $\mathbb{E}[X_\tau | \mathscr{F}_\sigma] = \mathbb{E}[\mathbb{E}[X_\infty | \mathscr{F}_\tau] | \mathscr{F}_\sigma] = \mathbb{E}[X_\infty | \mathscr{F}_\sigma] = X_\sigma$. $\square$

3.2.4 Backward Martingales and Applications

To end this section, we also introduce backward martingale, which is a powerful tool in some scenarios.

Definition 3.42 (Backward martingales)**.** A *backward filtration* is an increasing sequence of sub σ-algebras $(\mathscr{F}_n)_{n\in-\mathbb{N}_0}$ indexed by nonpositive integers, i.e. $\mathscr{F}_n \supset \mathscr{F}_{n-1}$ 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 $\mathbb{E}[X_n|\mathscr{F}_m] = X_m$ for all $m < n \leq 0$.

(ii) $(Y_n)_{n\in-\mathbb{N}_0}$ is said to be a *backward submartingale*, if $\mathbb{E}[X_n|\mathscr{F}_m] \geq X_m$ for all $m < n \leq 0$.

(iii) $(Y_n)_{n\in-\mathbb{N}_0}$ is said to be a *backward supermartingale*, if $\mathbb{E}[X_n|\mathscr{F}_m] \leq X_m$ for all $m < n \leq 0$.

Remark. Likewise, we define the limit of the backward filtration $(\mathscr{F}_n)_{n\in-\mathbb{N}_0}$ by

$$\mathscr{F}_{-\infty} = \bigcap_{n\in-\mathbb{N}_0} \mathscr{F}_n.$$

Theorem 3.43 (Doob's convergence theorem for backward submartingales)**.** *If $(X_n)_{n\in-\mathbb{N}_0}$ is a backward submartingale such that $\lim_{n\to-\infty} \mathbb{E}[X_n] > -\infty$, then $(X_n)_{n\in-\mathbb{N}_0}$ is uniformly integrable and converges a.s. and in L^1 to a random variable $X_{-\infty} \in L^1(\Omega, \mathscr{F}_{-\infty}, \mathbb{P})$. Moreover, $\mathbb{E}[X_n|\mathscr{F}_{-\infty}] \geq X_{-\infty}$ for all $n \in -\mathbb{N}_0$.*

Proof. Since $X_n, X_{n+1}, \cdots, X_0$ is a submartingale, by Doob's upcrossing inequality [Proposition 3.24],

$$\mathbb{E}\left[U_{[a,b]}^X(n)\right] \leq \frac{\mathbb{E}[(X_0-a)^+ - (X_n-a)^+]}{b-a} \leq \frac{\mathbb{E}|X_0|+|a|}{b-a}, \quad \forall a < b.$$

Since $U_{[a,b]}^X(n) \uparrow U_{[a,b]}^X(-\infty)$, by monotone convergence theorem, we have $\mathbb{P}\left(U_{[a,b]}^X(-\infty) < \infty\right) = 1$, and

$$\mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}:a<b} \left\{U_{[a,b]}^X(-\infty) < \infty\right\}\right) = 1$$

Similar to our proof of Theorem 3.25, X_n converges a.s. when $n \to -\infty$. Since every sequence $(X_k)_{k\leq n}$ is $\mathscr{F}_n$-measurable, the limit $X_{-\infty}$ is $\mathscr{F}_n$-measurable for each $n \in -\mathbb{N}_0$, hence is $\mathscr{F}_{-\infty}$-measurable.

To prove uniform integrability, we fix $\epsilon > 0$. Since $\mathbb{E}[X_n] \downarrow L := \lim_{n\to-\infty} \mathbb{E}[X_n] > -\infty$, 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

$$\begin{aligned}
\mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>M\}}\right] &= \mathbb{E}\left[X_n\mathbb{1}_{\{|X_n|>M\}}\right] - \mathbb{E}[X_n] \leq \mathbb{E}\left[\mathbb{E}[X_N|\mathscr{F}_n]\mathbb{1}_{\{|X_n|>M\}}\right] - \mathbb{E}[X_N] + \frac{\epsilon}{4} \\
&\leq \mathbb{E}\left[X_N\mathbb{1}_{\{|X_n|>M\}}\right] + \frac{\epsilon}{4} \\
&\leq \mathbb{E}\left[|X_N|\mathbb{1}_{\{|X_n|>M,\, |X_N|>M/2\}}\right] + \mathbb{E}\left[|X_N|\mathbb{1}_{\{|X_n|>M,\, |X_N|\leq M/2\}}\right] + \frac{\epsilon}{4} \\
&\leq \mathbb{E}\left[|X_N|\mathbb{1}_{\{|X_N|>M/2\}}\right] + \frac{1}{2}\mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>M\}}\right] + \frac{\epsilon}{4}.
\end{aligned}$$

Hence

$$\mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>M\}}\right] < 2\mathbb{E}\left[|X_N|\mathbb{1}_{\{|X_N|>M/2\}}\right] + \frac{\epsilon}{2}$$

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 $\|X_n - X_{-\infty}\|_1 \to 0$. Moreover, for all $A \in \mathscr{F}_{-\infty} = \bigcap_{n\in-\mathbb{N}_0} \mathscr{F}_n$,

$$\mathbb{E}[X_n\mathbb{1}_A] - \mathbb{E}[X_m\mathbb{1}_A] = \mathbb{E}[X_n\mathbb{1}_A] - \mathbb{E}[\mathbb{E}[X_n|\mathscr{F}_m]\mathbb{1}_A] \geq \mathbb{E}[X_n\mathbb{1}_A] - \mathbb{E}[X_n\mathbb{1}_A] \geq 0, \quad \forall m < n.$$

Let $m \to -\infty$. By Lebesgue dominated convergence theorem, $\mathbb{E}[X_n\mathbb{1}_A] \geq \mathbb{E}[X_{-\infty}\mathbb{1}_A]$ for all $n \in -\mathbb{N}_0$. Hence we have $\mathbb{E}[X_n|\mathscr{F}_{-\infty}] \geq X_{-\infty}$. $\square$

Remark. If $(X_n)_{n\in-\mathbb{N}_0}$ is a backward martingale, the requirement $\lim_{n\to-\infty}\mathbb{E}[X_n]=\mathbb{E}[X_0]>-\infty$ is satisfied. As we can see, *a backward martingale always converges a.s. and in* L^1, with no additional condition required. Moreover, the limit is given by $X_{-\infty}=\mathbb{E}[X_0|\mathscr{F}_{-\infty}]$.

We also have an immediate corollary of this theorem: If $\mathscr{F}_n\downarrow\mathscr{F}_{-\infty}$ is a backward filtration, and Z is a integrable random variable, then $\mathbb{E}[Z|\mathscr{F}_n]\to\mathbb{E}[Z|\mathscr{F}_{-\infty}]$ a.s. and in L^1.

The convergence theorem for backward martingales is useful in probability theory. Typically, we apply it to study the exchangeability of random variables.

Definition 3.44 (Exchangeable σ-algebra)**.** Let $X_1,X_2,\cdots$ be a sequence of random variables, and define $\mathscr{F}_\infty=\sigma(X_1,X_2,\cdots)$. For each $n\in\mathbb{N}$, define the *n-exchangeable σ-algebra* by

$$\mathscr{E}_n=\left\{A\in\mathscr{F}_\infty:\ A\text{ is invariant under permutation of }X_1,X_2,\cdots,X_n\right\}.$$

Also, define the *exchangeable algebra* by

$$\mathscr{E}=\bigcap_{n=1}^\infty\mathscr{E}_n.$$

Remark. If A is included in the algebra $\mathscr{F}_\infty$ generated by the random sequence $(X_n)_{n=1}^\infty$, it is of the form

$$A=\left\{(X_1,X_2,\cdots)\in B\right\},$$

where B is a Borel set in the product space $\mathbb{R}^{\mathbb{N}}$. If $A\in\mathscr{F}_\infty$ is invariant under permutation of $X_1,\cdots,X_n$, then for any bijection $\pi:\{1,\cdots,n\}\to\{1,\cdots,n\}$, we have

$$\pi(B):=\big\{(x_{\pi(1)},x_{\pi(2)},\cdots,x_{\pi(n)},x_{n+1},\cdots):(x_1,x_2,\cdots)\in B\big\}=B.$$

For example, for any $c\in\mathbb{R}$, the set $\{X_1+\cdots+X_n\le c\}$ is n-permutation invariant. Furthermore, for any measurable function $\varphi:\mathbb{R}^n\to\mathbb{R}$, if A is n-permutation invariant,

$$\mathbb{E}\left[\varphi(X_1,\cdots,X_n)\mathbb{1}_A\right]=\mathbb{E}\left[\varphi(X_{\pi(1)},\cdots,X_{\pi(n)})\mathbb{1}_A\right].$$

Therefore, the conditional expectation is also permutation-invariant:

$$\mathbb{E}\left[\varphi(X_{\pi(1)},\cdots,X_{\pi(n)})|\mathscr{E}_n\right]=\mathbb{E}[\varphi(X_1,\cdots,X_n)|\mathscr{E}_n]$$

Theorem 3.45 (Hewitt-Savage 0-1 law)**.** *The exchangeable algebra $\mathscr{E}$ of an i.i.d. sequence $(X_n)_{n=1}^\infty$ is trivial, i.e. $\mathbb{P}(A)\in\{0,1\}$ for all $A\in\mathscr{E}$.*

We first prove the following lemma.

Lemma 3.46. *Let $\varphi:\mathbb{R}^k\to\mathbb{R}$ be a bounded measurable function, and define*

$$A_n\varphi=\frac{1}{(n)_k}\sum_{(i_1,\cdots,i_k)\in I_{n,k}}\varphi(X_{i_1},\cdots,X_{i_k}),$$

where $I_{n,k}$ consists of sequences of distinct integers $1\le i_1<\cdots<i_k\le n$, and $(n)_k=n(n-1)\cdots(n-k+1)$ is the number of such sequences. Then $A_n\varphi\to\mathbb{E}[\varphi(X_1,\cdots,X_k)]$ a.s..

Proof. By definition, $A_n\varphi$ is $\mathscr{E}_n$-measurable, and

$$A_n\varphi=\mathbb{E}[A_n\varphi|\mathscr{E}_n]=\frac{1}{(n)_k}\sum_{(i_1,\cdots,i_k)\in I_{n,k}}\mathbb{E}[\varphi(X_{i_1},\cdots,X_{i_k})|\mathscr{E}_n]=\mathbb{E}[\varphi(X_1,\cdots,X_k)|\mathscr{E}_n].$$

Since $\mathscr{E}_n \downarrow \mathscr{E}$, by backward martingale convergence theorem, $A_n\varphi \to \mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ a.s. and in L^1. To finish the proof, we need to prove that $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}] = \mathbb{E}[\varphi(X_1, \cdots, X_k)]$. Note that

$$A_n\varphi = \frac{1}{(n)_k} \sum_{1 \in i \in I_{n,k}} \varphi(X_{i_1}, \cdots, X_{i_k}) + \frac{1}{(n)_k} \sum_{1 \notin i \in I_{n,k}} \varphi(X_{i_1}, \cdots, X_{i_k}).$$

Since φ is bounded, the first term involving X_1 is bounded by $\frac{k}{n}\|\varphi\|_\infty$, which converges to 0 as $n \to \infty$. Then the a.s. limit $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ is measurable with respect to $\sigma(X_2, X_3, \cdots)$. Repeating the same procedure, $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ is measurable with respect to $\sigma(X_n, X_{n+1}, \cdots)$ for all $n \in \mathbb{N}$. Acutally, the conditional expectation $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ is measurable with respect to the tail σ-algebra $\bigcap_{n=1}^\infty \sigma(X_n, X_{n+1}, \cdots)$, which is trivial by Kolmogorov's 0-1 law [Corollary 3.38]. Consequently,

$$\mathbb{P}\left(\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}] > q\right) \in \{0, 1\}$$

for all $q \in \mathbb{Q}$. Therefore $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ is a constant a.s., which is $\mathbb{E}[\varphi(X_1, \cdots, X_k)]$. $\square$

Proof of Theorem 3.45. By Lemma 3.46, $\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}] = \mathbb{E}[\varphi(X_1, \cdots, X_k)]$ holds for each bounded measurable function φ. Then the exchangeable σ-algebra $\mathscr{E}$ is independent of $\sigma(X_1, \cdots, X_k)$ for all $k \in \mathbb{N}$. Since $\bigcup_{k=1}^\infty \sigma(X_1, \cdots, X_k)$ is a π-system, by Sierpiński-Dynkin π-λ theorem, $\mathscr{E}$ is independent of $\sigma(X_1, X_2, \cdots)$. Then $\mathscr{E}$ is independent of itself, and $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ for all $A \in \mathscr{E}$, which concludes the proof. $\square$

Theorem 3.47 (De Finetti's theorem)**.** *If $X_1, X_2, \cdots$ is a sequence of **exchangeable** random variables, i.e. for all $n \in \mathbb{N}$ and all permutations π of $\{1, \cdots, n\}$,*

$$(X_1, \cdots, X_n) \overset{d}{=} (X_{\sigma(1)}, \cdots, X_{\sigma(n)}),$$

then $X_1, X_2, \cdots$ are i.i.d. conditional on the exchangeable σ-algebra $\mathscr{E}$.

Proof. We let $f : \mathbb{R}^{k-1} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be bounded measurable functions. Then the tensor product $\varphi(x_1, \cdots, x_k) = f(x_1, \cdots, x_{k-1})g(x_k)$ is also bounded, and

$$\begin{aligned}
(n)_{k-1} A_n f \cdot n A_g &= \sum_{i \in I_{n,k-1}} f(X_{i_1}, \cdots, X_{i_{k-1}}) \sum_{j=1}^n g(X_j) \\
&= \sum_{i \in I_{n,k}} f(X_{i_1}, \cdots, X_{i_{k-1}}) g(X_{i_k}) + \sum_{i \in I_{n_k-1}} f(X_{i_1}, \cdots, X_{i_{k-1}}) \sum_{m=1}^{k-1} g(X_{i_m}) \\
&= (n)_k A_n \varphi + (n)_{k-1} \sum_{m=1}^{k-1} A_n \varphi_m,
\end{aligned}$$

where $\varphi_m(x_1, \cdots, x_{k-1}) = f(x_1, \cdots, x_{k-1})g(x_m)$. Rearranging the identity, we have

$$A_n\varphi = \frac{n}{n-k+1} A_n f \cdot A_n g - \frac{1}{n-k+1} \sum_{m=1}^{k-1} A_n \varphi_m.$$

According to the proof of Lemma 3.46, with exchangeability, $A_n\varphi \to \mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}]$ for all bounded functions $\varphi : \mathbb{R}^k \to \mathbb{R}$. We let $n \to \infty$ in the last display to obtain

$$\mathbb{E}[\varphi(X_1, \cdots, X_k)|\mathscr{E}] = \mathbb{E}[f(X_1, \cdots, X_{k-1})|\mathscr{E}] \cdot \mathbb{E}[g(X_k)|\mathscr{E}].$$

By induction, for all $n \in \mathbb{N}$ and bounded measurable functions $f_1, \cdots, f_n$, we have

$$\mathbb{E}[f_1(X_1)f_2(X_2)\cdots f_n(X_n)|\mathscr{E}] = \mathbb{E}[f_1(X_1)|\mathscr{E}] \times \mathbb{E}[f_2(X_2)|\mathscr{E}] \times \cdots \times \mathbb{E}[f_n(X_n)|\mathscr{E}].$$

Hence $X_1, X_2, \cdots$ are i.i.d. conditional on $\mathscr{E}$. □

The backward martingale method gives a beautiful proof of Kolmogorov's strong law of large numbers.

Theorem 3.48 (Kolmogorov's strong law of large numbers)**.** *Let $(\xi_n)_{n=1}^\infty$ be a sequence of i.i.d. random variables such that $\mathbb{E}|\xi_1| < \infty$. Then*

$$\lim_{n\to\infty} \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} = \mathbb{E}\xi_1 \quad a.s..$$

Proof. Let $S_n = \xi_1 + \cdots + \xi_n$, and $X_{-n} = S_n/n$. Let $\mathscr{F}_{-n} = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \cdots) \downarrow \mathscr{F}_{-\infty}$. Then

$$\begin{aligned}\mathbb{E}[X_{-n}|\mathscr{F}_{-n-1}] &= \frac{1}{n}\mathbb{E}\left[S_{n+1} - \xi_{n+1}|\mathscr{F}_{-n-1}\right] = \frac{S_{n+1}}{n} - \frac{\mathbb{E}\left[\xi_{n+1}|\mathscr{F}_{-n-1}\right]}{n} \\ &= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = X_{-n-1},\end{aligned}$$

where in the third inequality we use the exchangeability of $\xi_1, \cdots, \xi_{n+1}$ and the fact $S_{n+1} = \xi_1 + \cdots + \xi_{n+1}$. By backward martingale convergence theorem, $X_{-n} \to X_{-\infty}$ a.s. and in L^1, and $X_{-\infty} = \mathbb{E}[\xi_1|\mathscr{F}_{-\infty}]$. Also, by definition we have $\mathscr{F}_{-n} \subset \mathscr{E}_n$, and $\mathscr{F}_{-\infty} \subset \mathscr{E}$. By Hewitt-Savage 0-1 law, $\mathbb{E}[\xi_1|\mathscr{F}_{-\infty}] = \mathbb{E}\xi_1$ a.s.. □

3.3 Continuous-time Martingales

3.3.1 Definition and Properties

Definition 3.49 (Continuous-time martingales)**.** Let $(X_t)_{t\geq 0}$ be a real-valued and L^1 process that is adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$. 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 $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$ for all $t > s \geq 0$;

(ii) $(X_t)_{t\geq 0}$ is said to be a *supermartingale* if $\mathbb{E}[X_t|\mathscr{F}_t] \leq X_s$ for all $t > s \geq 0$;

(iii) $(X_t)_{t\geq 0}$ is said to be a *submartingale* if $\mathbb{E}[X_t|\mathscr{F}_t] \geq X_s$ for all $t > s \geq 0$;

All these notations depends on the choice of the filtration $\{\mathscr{F}_t\}_{t\geq 0}$, which is fixed in later discussion.

Remark. Similar to Proposition 3.17, we have an immediate corollary of conditional Jensen's inequality. Let $f : \mathbb{R} \to \mathbb{R}$ 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 3.50. *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, $\mathbb{E}[X_s] = \mathbb{E}[\mathbb{E}[X_s|\mathscr{F}_0]] \geq \mathbb{E}[X_0]$. 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

$$\mathbb{E}|X_s| = 2\mathbb{E}[X_s^+] - \mathbb{E}[X_s] \leq 2\mathbb{E}[X_t^+] - \mathbb{E}[X_0] < \infty,\ \forall s \in [0, t].$$

The result immediately follows. □

Proposition 3.51. *Let $(X_t)_{t\geq 0}$ be an L^2 martingale. Then for all reals $0 \leq s < t$ and all finite partitions $s = t_0 < t_1 < \cdots < t_k = t$, we have*

$$\mathbb{E}\left[\sum_{j=1}^k (X_{t_j} - X_{t_{j-1}})^2 \middle| \mathscr{F}_s\right] = \mathbb{E}\left[X_t^2 - X_s^2|\mathscr{F}_s\right] = \mathbb{E}\left[(X_t - X_s)^2|\mathscr{F}_s\right].$$

Proof. For each $j = 1, \cdots, k$,

$$\mathbb{E}\left[(X_{t_j} - X_{t_{j-1}})^2|\mathscr{F}_s\right] = \mathbb{E}\left[\mathbb{E}\left[(X_{t_j} - X_{t_{j-1}})^2|\mathscr{F}_{t_{j-1}}\right]\middle|\mathscr{F}_s\right] = \mathbb{E}\left[\mathbb{E}\left[X_{t_j}^2 - X_{t_{j-1}}^2|\mathscr{F}_{t_{j-1}}\right]\middle|\mathscr{F}_s\right]$$

Then the desired result follows by summing over j. □

Now we extend the inequalities in Proposition 3.30 and Proposition 3.31 to continuous-time martingales.

Proposition 3.52. *Let $(X_t)_{t\geq 0}$ be a right-continuous submartingale.*

(i) (Maximal inequality). For every $t > 0$,

$$\lambda\mathbb{P}\left(\sup_{0\leq s\leq t} |X_s| > \lambda\right) \leq \mathbb{E}|X_0| + 2\mathbb{E}|X_t|,\ \forall\lambda > 0.$$

In addition, if $(X_t)_{t\geq 0}$ is nonnegative, then

$$\lambda\mathbb{P}\left(X_t^* > \lambda\right) \leq \mathbb{E}[X_t],\ \forall\lambda > 0,\ \textit{where } X_t^* = \sup_{0\leq s\leq t} X_s.$$

(ii) (Doob's L^p-inequality). If $(X_t)_{t\geq 0}$ is a right-continuous martingale, then for every $t>0$,

$$\mathbb{E}\left[\sup_{0\leq s\leq t}|X_s|^p\right]\leq\left(\frac{p}{p-1}\right)^p\mathbb{E}|X_t|^p,\ \forall p>1.$$

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 $f(t_{nk})\to f(s_n)$ for every n. Then $f(t_{nn})\to M$, 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 $D_k=\{t_0^k,t_1^k,\cdots,t_k^k\}$, where $0\leq t_0^k<t_1^k<\cdots<t_k^k=t$.

(i) For each $k\in\mathbb{N}$, we can apply the maximal inequality [Proposition 3.30] of discrete form on sequence $Y_n=X_{t^k_{n\wedge k}}$, which is a submartingale of the filtration $\mathscr{G}_n=\mathscr{F}_{t^k_{n\wedge k}}$:

$$\lambda\mathbb{P}\left(\max_{s\in D_k}|X_s|>\lambda\right)\leq\mathbb{E}[|X_0|]+2\mathbb{E}[|X_t|],\ \forall k\in\mathbb{N},\ \lambda>0.$$

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 $k\to\infty$. By monotone convergence theorem,

$$\lambda\mathbb{P}\left(\max_{s\in D_k}|X_s|>\lambda\right)\nearrow\lambda\mathbb{P}\left(\sup_{s\in[0,t]}|X_s|>\lambda\right)\leq\mathbb{E}[|X_0|]+2\mathbb{E}[|X_t|],\ \forall\lambda>0.$$

The case of nonnegative submartingale is similar.

(ii) Similar to the proof of (i), we apply Doob's inequality [Proposition 3.31] of discrete form:

$$\mathbb{E}\left[\max_{s\in D_k}|X_k|^p\right]\leq\left(\frac{p}{p-1}\right)^p\mathbb{E}\left[|X_t|^p\right],\ \forall p>1.$$

Since $\max_{s\in D_k}|X_k|\nearrow\sup_{0\leq s\leq t}|X_s|$, we use monotone convergence theorem to get the desired result. $\square$

Remark. If $(X_t)_{t\geq 0}$ is a submartingale, then for any dense subset $D\subset\mathbb{R}$ and every $t>0$,

$$\mathbb{P}\left(\sup_{s\in D\cap[0,t]}|X_s|>\lambda\right)\leq\frac{1}{\lambda}\left(\mathbb{E}[|X_0|]+2\mathbb{E}[|X_t|]\right),\ \forall\lambda>0.$$

Let $\lambda\to\infty$, we obtain that $\sup_{s\in D\cap[0,t]}|X_s|<\infty$ *a.s.* for all $t>0$.

3.3.2 Martingale Convergence Theorems

Definition 3.53 (Upcrossing number)**.** Given a function $f:E\to\mathbb{R}$ and $a<b$, where $E\subset\mathbb{R}$, the *upcrossing number* of this sequence along $[a,b]$, denoted by $U^f_{[a,b]}(E)$, is the largest $k\in\mathbb{N}$ such that there exists a finite and strictly increasing sequence $s_1<t_1<s_2<t_2<\cdots<s_k<t_k$ of elements of E such that $f(s_j)\leq a$ and $f(t_j)\geq b$ for all $j\in\{1,\cdots,k\}$. If there exists no such sequence, we take $U^f_{[a,b]}(E)=0$. If such sequence exists for all $k\in\mathbb{N}$, we take $U^f_{[a,b]}(E)=\infty$,

A function $f:E\to\mathbb{R}$ is said to be *càdlàg (French: continue à droite, limite à gauche)*, if for all $t\in\mathring{E}$, the left limit $f(t-)<\infty$ exists, and the right limit $f(t+)$ exists and equals $f(t)$.

Lemma 3.54. *Let D be a countable dense subset of $\mathbb{R}_+$, and $f : D \to \mathbb{R}$. Assume that for every $T \in D$,*

(i) the function f is bounded on $D \cap [0,T]$, and

(ii) $U^f_{[a,b]}(D \cap [0,T]) < \infty$ 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 $g : \mathbb{R}^+ \to \mathbb{R}$ 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 $|f(t)| \le M$ 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 $f(s_n) \to a$ and $f(t_n) = b > a$, take two rationals $a < p < q < b$. Then for any $k \in \mathbb{N}$, we can find a $f(t_{n_1}) \ge q$, and $s_{n_1} < t_{n_1}$ such that $f(s_{n_1}) \le p$, and $t_{n_2} < s_{n_1}$ such that $f(t_{n_2}) \ge q$, $\cdots$, and $s_{n_k} < t_{n_k}$ such that $f(s_{n_k}) \le p$. Thus we obtain an upcrossing sequence $s_{n_k} < t_{n_k} < s_{n_{k-1}} < y_{n_{k-1}} < \cdots < s_{n_1} < t_{n_1}$ 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

$$|g(r) - f(t+)| = \lim_{n\to\infty} |f(r_n) - f(t+)| < \epsilon, \quad \textit{and} \quad |g(s) - f(t-)| = \lim_{n\to\infty} |f(s_n) - f(t-)| < \epsilon.$$

Hence $\lim_{r \downarrow t} g(r) = f(t+)$, $\lim_{s \uparrow t} g(s) = f(t-)$, and g is càdlàg. □

Theorem 3.55. *Let $(X_t)_{t\ge 0}$ 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 path $t \mapsto X_t(\omega)$ to D has right and left limits everywhere:

$$X_{t+}(\omega) = \lim_{D \ni s \downarrow t} X_s(\omega), \quad \forall t \in \mathbb{R}_+, \qquad X_{t-}(\omega) = \lim_{D \ni s \uparrow t} X_s(\omega), \quad \forall t \in \mathbb{R}_{++}.$$

(ii) For every $t \in \mathbb{R}_+$, the limit $X_{t+} \in L^1(\Omega, \mathscr{F}_{t+}, \mathbb{P})$, and $\mathbb{E}[X_{t+}|\mathscr{F}_t] \ge X_t$ with equality holds if the mean function $t \mapsto \mathbb{E}[X_t]$ is right-continuous (in particular, if $(X_t)_{t\ge0}$ is a martingale). The process $(X_{t+})_{t\ge0}$ is a submartingale with respect to the filtration $(\mathscr{F}_{t+})_{t\ge0}$.

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 3.52, we take an sequence D_k increasing to $D \cap [0,T]$. Using Doob's upcrossing inequality [Proposition 3.24] and monotone convergence theorem, for all $a < b$, we have

$$\mathbb{E}\left[U^X_{[a,b]}(D \cap [0,T])\right] \le \frac{\mathbb{E}\left[(X_T - a)^+ - (X_0 - a)^+\right]}{b-a} < \infty \quad \Rightarrow \quad U^X_{[a,b]}(D \cap [0,T]) < \infty \ a.s..$$

Set the negligible set N as

$$N = \bigcup_{T \in D} \left(\left\{ \sup_{s \in D \cap [0,T]} |X_s| = \infty \right\} \cap \left(\bigcup_{a,b \in \mathbb{Q}} \left\{ U^X_{[a,b]}(D \cap [0,T]) = \infty \right\} \right) \right) \quad \Rightarrow \quad \mathbb{P}(N) = 0. \tag{3.8}$$

Outside N, the assumptions in Lemma 3.54 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 $\mathscr{F}_{t+}$-measurable. Fix $t \in \mathbb{R}_+$, and choose $t_n \downarrow t$. Then we have $X_{t_n} \to X_{t_+}$ *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 $\mathscr{G}_n = \mathscr{F}_{t_{-n}}$:

$$\mathbb{E}[Y_n|\mathscr{G}_{n-1}] = \mathbb{E}[X_{t_{-n}}|\mathscr{F}_{t_{-(n-1)}}] \geq X_{t_{-(n-1)}} = Y_{n-1},\ \forall n \in -\mathbb{N}_0.$$

By Proposition 3.50, we have $\sup_{s\in D\cap[0,T]} \mathbb{E}[|X_s|] < \infty$, which implies $\lim_{n\to-\infty} \mathbb{E}[Y_n] > -\infty$. Using Doob's convergence theorem for discrete-time backward submartingales [Theorem 3.43], we have $X_{t_n} \to X_{t+}$ in L^1, and $X_{t+} \in L^1(\Omega, \mathscr{F}_{t+}, \mathbb{P})$. Due to convergence in L^1, we have

$$\mathbb{E}[X_{t+}|\mathscr{F}_t] = \lim_{n\to\infty} \mathbb{E}[X_{t_n}|\mathscr{F}_t] \geq X_t, \quad \textit{and} \quad \mathbb{E}[X_{t+}] = \lim_{n\to\infty} \mathbb{E}[X_{t_n}].$$

Note the first equality holds because the conditional expectation operator $\mathbb{E}[\cdot|\mathscr{F}_t]$ is a bounded linear operator on $L^1(\Omega, \mathscr{F}, \mathbb{P})$. 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 $\mathbb{E}[X_{t+}|\mathscr{F}_t] = X_t$.

Let $s < t$, and take $s_n \downarrow s$ such that $s_n \leq t_n$. Then $X_{s_n} \to X_{s+}$ *a.s.* and in L^1. Moreover, if $A \in \mathscr{F}_{s+}$,

$$\mathbb{E}[X_{t+}\mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}[X_{t_n}\mathbb{1}_A] \geq \lim_{n\to\infty} \mathbb{E}[X_{s_n}\mathbb{1}_A] = \mathbb{E}[X_{s+}\mathbb{1}_A].$$

Since X_{s+} is $\mathscr{F}_{s+}$-measurable, we have $\mathbb{E}[X_{t+}|\mathscr{F}_{s+}] \geq X_{s+}$. Therefore, $(X_{t+})_{t\geq 0}$ is a submartingale with respect to the filtration $(\mathscr{F}_{t+})_{t\geq 0}$. ☐

Theorem 3.56 (Càdlàg modification). *Let $(\mathscr{F}_t)_{t\geq 0}$ 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 $(\mathscr{F}_t)_{t\geq 0}$.*

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 3.54, the sample paths of $(Y_t)_{t\geq 0}$ are càdlàg.

Since X_{t+} is $\mathscr{F}_t$-measurable by right-continuity of $(\mathscr{F}_t)_{t\geq 0}$, and since the negligible set N falls in all $\mathscr{F}_t$ by completeness of $(\mathscr{F}_t)_{t\geq 0}$, the function Y_t is $\mathscr{F}_t$-measurable. Furthermore,

$$X_t = \mathbb{E}[X_{t+}|\mathscr{F}_t] = X_{t+} \overset{a.s.}{=} Y_t, \quad \forall t \in \mathbb{R}_+.$$

Hence $(Y_t)_{t\geq 0}$ is an *a.s.* modification of $(X_t)_{t\geq 0}$, which is adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$. Furthermore, we have $\mathbb{E}[Y_t|\mathscr{F}_s] = \mathbb{E}[X_t|\mathscr{F}_s] \geq X_s = Y_s$ for all $t > s \geq 0$. Hence $(Y_t)_{t\geq 0}$ is also a submartingale. ☐

Theorem 3.57 (Doob's first martingale convergence theorem). *If $(X_t)_{t\geq 0}$ is a right-continuous submartingale and $\sup_{t\geq 0} \mathbb{E}[X_t^+] < \infty$, then $X_\infty = \lim_{t\uparrow\infty} X_n$ a.s. exists, and $X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$.*

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 3.55 (i) and use monotone convergence theorem to conclude

$$\mathbb{E}\left[U_{[a,b]}^X(D\cap[0,T])\right] \leq \frac{\mathbb{E}[(X_T - a)^+ - (X_0 - a)^+]}{b-a} \leq \frac{M + |a|}{b-a} \quad \Rightarrow \quad \mathbb{E}\left[U_{[a,b]}^X(D)\right] \leq \frac{M+|a|}{b-a} < \infty.$$

Hence $U_{[a,b]}^X(D) < \infty$ a.s. for all $a, b \in \mathbb{Q}$ with $a < b$, and $X_\infty = \lim_{D\ni t\uparrow\infty} X_t \in [-\infty, \infty]$ a.s. exists. We can further exclude values ∞ and $-\infty$ $\mathbb{P}$-*a.e.*, because the Fatou's lemma gives

$$\mathbb{E}[X_\infty^+] \leq \liminf_{D\ni t\uparrow\infty} \mathbb{E}[X_t^+] \leq M, \qquad \mathbb{E}[X_\infty^-] \leq \liminf_{D\ni t\uparrow\infty} \mathbb{E}[X_t^-] = \liminf_{D\ni t\uparrow\infty} \mathbb{E}[X_t^+ - X_t] \leq M - \mathbb{E}[X_0].$$

Hence $X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Finally, since $(X_t)_{t\geq 0}$ is right-continuous, we can drop the restriction $t \in D$. ☐

Similarly, we need uniform integrability of martingale to obtain convergence in L^1.

Theorem 3.58 (Doob's second martingale convergence theorem)**.** *Let $(X_t)_{t\geq 0}$ be a 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 $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ such that $X_t = \mathbb{E}[Z|\mathscr{F}_t]$ for all $t \in \mathbb{R}_+$.

Proof. (i) $\Rightarrow$ (ii): By Theorem 3.57, the limit $X_\infty = \lim_{t\to\infty} X_t$ *a.s.* exists. Since $\{X_t\}_{t\in\mathbb{R}_+}$ is uniformly integrable, convergence in L^1 also holds by Theorem 1.75.

(ii) $\Rightarrow$ (iii) follows from the continuity of conditional expectation operator on $L^1(\Omega, \mathscr{F}, \mathbb{P})$.

(iii) $\Rightarrow$ (i) follows from Theorem 3.34. □

Remark. If (i)-(iii) are satisfied, the limit $X_\infty = \lim_{t\to\infty} X_t$ satisfies $\mathbb{E}[X_\infty|\mathscr{F}_t] = \lim_{s\to\infty} \mathbb{E}[X_s|\mathscr{F}_t] = X_t$.

Theorem 3.59 (Convergence theorem for L^p-bounded 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 $X_\infty = \lim_{t\to\infty} X_t$ 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 3.58, there exists $X_\infty \in L^p(\Omega, \mathscr{F}, \mathbb{P})$ such that $X_t \to X_\infty$ *a.s.*. By Doob's L^p-inequality [Proposition 3.50 (ii)] and monotone convergence theorem,

$$\mathbb{E}\left[\sup_{0\leq s\leq t} |X_s|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_t|^p] \leq \left(\frac{p}{p-1}\right)^p M \quad \Rightarrow \quad \mathbb{E}[Y^p] \leq \left(\frac{p}{p-1}\right)^p M < \infty.$$

Since $|X_t - X_\infty| \leq 2Y$, by Lebesgue dominated convergence theorem, $\|X_t - X_\infty\|_p \to \infty$. □

3.3.3 Optional Stopping Theorems

Given a 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

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega)\mathbb{1}_{\{\tau<\infty\}}(\omega) + X_\infty(\omega)\mathbb{1}_{\{\tau=\infty\}}(\omega), \text{ where } X_\infty = \lim_{t\to\infty} X_t \text{ a.s.}.$$

By Proposition 3.10 and Proposition 3.13, $(X_t)_{t\geq 0}$ is progressive, and the restriction of X_τ to $\{\tau < \infty\}$ is $\mathscr{F}_\tau$-measurable. Meanwhile, $\{X_\infty \mathbb{1}_{\{\tau=\infty\}} \leq \alpha\} \subset \{\tau \leq t\}$ if $\alpha \geq 0$, and $\{X_\infty \mathbb{1}_{\{\tau=\infty\}} \leq \alpha\} \cap \{\tau \leq t\} = \emptyset$ if $\alpha < 0$. Therefore, $\{X_\infty \mathbb{1}_{\{\tau=\infty\}} \leq \alpha\} \cap \{\tau \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$, and $X_\infty \mathbb{1}_{\{\tau=\infty\}}$ is $\mathscr{F}_\tau$-measurable. As a result, the random variable X_τ is $\mathscr{F}_\tau$-measurable.

Theorem 3.60 (Optional stopping theorem for submartingales)**.** *Let $(X_t)_{t\geq 0}$ be a right-continuous submartingale. Let τ and σ be two stopping times such that $\sigma \leq \tau$. Then $X_\tau, X_\sigma \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, and $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] \geq X_\sigma$, 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 $U \in L^1(\Omega, \mathscr{F}_0, \mathbb{P})$ 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:

$$\sigma_n = \sum_{k=0}^{M2^n-1} \frac{k+1}{2^n} \mathbb{1}_{\{k2^{-n}<\sigma\leq(k+1)2^{-n}\}}, \quad \text{and} \quad \tau_n = \sum_{k=0}^{M2^n-1} \frac{k+1}{2^n} \mathbb{1}_{\{k2^{-n}<\tau\leq(k+1)2^{-n}\}}, \quad n \in \mathbb{N}.$$

We fix $n \geq 2$. Then the sequence $(X_{k2^{-n}})_{k=0}^{\infty}$ is a discrete-time submartingale with respect to the filtration $(\mathscr{F}_{k2^{-n}})_{k=0}^{\infty}$. Furthermore, both $2^n\tau_{n-1}$ and $2^n\tau_n$ are stopping times of $(\mathscr{F}_{k2^{-n}})_{k=0}^{\infty}$, since $2^n\tau_{n-1} = \min\{k \in \{2, 4, \cdots, M2^n\} : \ k2^{-n} \geq \tau\}$ implies

$$\{2^n\tau_{n-1} \leq p\} = \left\{\tau \leq \left\lfloor \frac{p}{2} \right\rfloor 2^{1-n}\right\} \in \mathscr{F}_{p2^{-n}}, \ \forall p \in \mathbb{N},$$

and $2^n\tau_n = \min\{k \in \{1, 2, \cdots, M2^n\} : k2^{-n} \geq \tau\}$ implies

$$\{2^n\tau_n \leq p\} = \{\tau \leq p2^{-n}\} \in \mathscr{F}_{p2^{-n}}, \ \forall p \in \mathbb{N}.$$

By the optional stopping theorem for discrete-time submartingales [Theorem 3.20 (iii)], we know that $\mathbb{E}[X_{\tau_{n-1}}|\mathscr{F}_{\tau_n}] \geq X_{\tau_n}$. Hence $Y_n := X_{\tau_{-n}}$ is a backward submartingale. Furthermore, by [Theorem 3.20 (i, ii)], we have $\mathbb{E}[X_{\tau_n}] \geq \mathbb{E}[X_0]$ for all $n \in \mathbb{N}$. Apply Theorem 3.43, 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 $X_{\tau_n} \to X_\tau$ *a.s.*, and this convergence also holds in L^1. Also, $X_{\sigma_n} \to X_\sigma$ *a.s.* and in L^1. Then for all $A \in \mathscr{F}_\sigma$, we have

$$\text{(By Theorem 3.20)} \quad \mathbb{E}[X_{\tau_n}\mathbb{1}_A] \geq \mathbb{E}[X_{\sigma_n}\mathbb{1}_A], \ \ \forall n \in \mathbb{N} \quad \Rightarrow \quad \mathbb{E}[X_\tau\mathbb{1}_A] \geq \mathbb{E}[X_\sigma\mathbb{1}_A],$$

where the $\Rightarrow$ follows from L^1 convergence. Hence $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] \geq X_\sigma$.

(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, $\mathbb{E}[X_\tau] \geq \limsup_{n\to\infty} \mathbb{E}[X_{\tau\wedge n}] \geq \mathbb{E}[X_0] > -\infty$. Similarly we have $\mathbb{E}[X_\sigma] \geq \mathbb{E}[X_0] > -\infty$. Hence $X_\tau, X_\sigma \in L^1(\Omega, \mathscr{F}, \mathbb{P})$.

Fix $A \in \mathscr{F}_\sigma \subset \mathscr{F}_\tau$, and define $\tau^A(\omega) = \tau(\omega)\mathbb{1}_A + \infty\mathbb{1}_{A^c}$. According to Proposition 3.12 (d), both τ^A and σ^A are also stopping times. By (i), we have

$$\begin{aligned}
0 &\leq \mathbb{E}\left[X_{\tau^A\wedge n}\right] - \mathbb{E}\left[X_{\sigma^A\wedge n}\right] \\
&= \mathbb{E}\left[X_n\mathbb{1}_{A^c} + X_{\tau\wedge n}\mathbb{1}_{A\cap\{\sigma\leq n\}} + X_n\mathbb{1}_{A\cap\{\sigma>n\}}\right] - \mathbb{E}\left[X_n\mathbb{1}_{A^c} + X_\sigma\mathbb{1}_{A\cap\{\sigma\leq n\}} + X_n\mathbb{1}_{A\cap\{\sigma>n\}}\right] \\
&= \mathbb{E}\left[X_{\tau\wedge n}\mathbb{1}_{A\cap\{\sigma\leq n\}}\right] - \mathbb{E}\left[X_\sigma\mathbb{1}_{A\cap\{\sigma\leq n\}}\right]
\end{aligned}$$

Note that $\mathbb{1}_{A\cap\{\sigma\leq n\}} \to \mathbb{1}_{A\cap\{\sigma<\infty\}}$ and $X_{\tau\wedge n} \to X_\tau$ as $n \to \infty$. By Lebesgue dominated convergence theorem, we have $\mathbb{E}[X_\tau\mathbb{1}_{A\cap\{\sigma<\infty\}}] \geq \mathbb{E}[X_\sigma\mathbb{1}_{A\cap\{\sigma<\infty\}}]$. Clearly, $\mathbb{E}[X_\tau\mathbb{1}_{A\cap\{\sigma=\infty\}}] = \mathbb{E}[X_\sigma\mathbb{1}_{A\cap\{\sigma=\infty\}}]$. Hence

$$\mathbb{E}[\mathbb{E}[X_\tau|\mathscr{F}_\sigma]\mathbb{1}_A] = \mathbb{E}[X_\tau\mathbb{1}_A] \geq \mathbb{E}[X_\sigma\mathbb{1}_A], \ \forall A \in \mathscr{F}_\sigma.$$

Since X_σ is $\mathscr{F}_\sigma$-measurable, we have $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] \geq X_\sigma$. □

Theorem 3.61 (Optional stopping theorem for uniformly integrable martingales)**.** *Let $(X_t)_{t\geq 0}$ be a uniformly integrable right-continuous martingale. Let τ be a stopping time. Then we have*

$$\mathbb{E}[X_\infty|\mathscr{F}_\tau] = X_\tau \in L^1(\Omega, \mathscr{F}, \mathbb{P}).$$

Furthermore, if σ is another stopping time such that $\sigma \leq \tau$, then $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] = X_\sigma$.

Proof. Using Proposition 3.15, we define a sequence of stopping times $\tau_n \searrow \tau$ as follows:

$$\tau_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n}\mathbb{1}_{\{k2^{-n}<\tau\leq(k+1)2^{-n}\}} + \infty\mathbb{1}_{\{\tau=\infty\}}, \ n \in \mathbb{N}.$$

Clearly, $2^n\tau_n$ is a stopping time of the filtration $\{\mathscr{F}_{(k+1)2^{-n}}\}_{k=0}^{\infty}$. Apply Theorem 3.41 to discrete-time martingale $\{X_{(k+1)2^{-n}}\}_{k=0}^{\infty}$ with respect to the filtration $\{\mathscr{F}_{(k+1)2^{-n}}\}_{k=0}^{\infty}$, which is uniformly integrable, we have $\mathbb{E}[X_\infty|\mathscr{F}_{\tau_n}] = X_{\tau_n} \in L^1(\Omega,\mathscr{F},\mathbb{P})$. Since $\tau_n(\omega) \searrow \tau(\omega)$, and the sample path $t \mapsto X_t(\omega)$ is right-continuous, we have $X_{\tau_n} \to X_\tau$ *a.s.*. Furthermore, $\{X_{\tau_n}\}_{n=1}^\infty$ is uniformly integrable, so the convergence also holds in L^1. For all $A \in \mathscr{F}_\tau \subset \mathscr{F}_{\tau_n}$, $\mathbb{E}[X_\infty|\mathscr{F}_{\tau_n}] = X_{\tau_n}$ implies $\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[X_{\tau_n}\mathbb{1}_A]$. Then

$$\mathbb{E}[X_\tau \mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}\left[X_{\tau_n}\mathbb{1}_A\right] = \mathbb{E}[X_\infty \mathbb{1}_A],\ \forall A \in \mathscr{F}_\tau \quad \Rightarrow \quad \mathbb{E}[X_\infty|\mathscr{F}_\tau] = X_\tau.$$

Furthermore, if $\sigma \le \tau$ is a stopping time, then $\mathbb{E}[X_\tau|\mathscr{F}_\sigma] = \mathbb{E}\left[\mathbb{E}[X_\infty|\mathscr{F}_\tau]|\mathscr{F}_\sigma\right] = \mathbb{E}[X_\infty|\mathscr{F}_\sigma] = X_\sigma$. □

Given an adapted process $(X_t)_{t\ge 0}$ and a stopping time τ, we denote by $X_t^\tau = X_{t\wedge\tau}$ the stopped process.

Corollary 3.62. *Let $(X_t)_{t\ge0}$ be a right-continuous submartingale. Let τ be a stopping time.*

(i) The stopped process $(X_t^\tau)_{t\ge0}$ is a submartingale.

(ii) In addition, if $(X_t)_{t\ge0}$ is a uniformly integrable martingale, so is the stopped process $(X_t^\tau)_{t\ge0}$. Moreover,

$$X_t^\tau = \mathbb{E}[X_\tau|\mathscr{F}_t].$$

Proof. (i) Fix $t > s \ge 0$. If $A \in \mathscr{F}_s \subset \mathscr{F}_t$, we have $A \cap \{\tau > s\} \in \mathscr{F}_s$, and $A \cap \{\tau > s\} \in \mathscr{F}_\tau$ by the very definition of $\mathscr{F}_\tau$. Hence $A \cap \{\tau > t\} \in \mathscr{F}_s \cap \mathscr{F}_\tau = \mathscr{F}_{\tau\wedge s}$, and

$$\begin{aligned}
\mathbb{E}\left[X_t^\tau \mathbb{1}_A\right] - \mathbb{E}\left[X_s^\tau \mathbb{1}_A\right] &= \mathbb{E}\left[X_{\tau\wedge t}\mathbb{1}_{A\cap\{\tau\le s\}}\right] + \mathbb{E}\left[X_{\tau\wedge t}\mathbb{1}_{A\cap\{\tau>s\}}\right] - \mathbb{E}\left[X_{\tau\wedge s}\mathbb{1}_A\right] \\
&= \mathbb{E}\left[X_{\tau\wedge s}\mathbb{1}_{A\cap\{\tau\le s\}}\right] + \mathbb{E}\left[X_{\tau\wedge t}\mathbb{1}_{A\cap\{\tau>s\}}\right] - \mathbb{E}\left[X_{\tau\wedge s}\mathbb{1}_A\right] \\
&= \mathbb{E}\left[(X_{\tau\wedge t} - X_{\tau\wedge s})\mathbb{1}_{A\cap\{\tau>s\}}\right] = \mathbb{E}\left[\mathbb{E}\left[X_{\tau\wedge t} - X_{\tau\wedge s}|\mathscr{F}_{\tau\wedge s}\right]\mathbb{1}_{A\cap\{\tau>s\}}\right] = 0,
\end{aligned}$$

where the last inequality follows from Theorem 3.60, because $\tau \wedge s \le \tau \wedge t$ are two bounded stopping times.

(ii) Fix $t \ge 0$. If $A \in \mathscr{F}_t$, we have $A \cap \{\tau > t\} \in \mathscr{F}_t$, and $A \cap \{\tau > t\} \in \mathscr{F}_\tau$ by the very definition of $\mathscr{F}_\tau$. Hence $A \cap \{\tau > t\} \in \mathscr{F}_t \cap \mathscr{F}_\tau = \mathscr{F}_{\tau\wedge t}$, and

$$\begin{aligned}
\mathbb{E}\left[X_\tau \mathbb{1}_A\right] - \mathbb{E}\left[X_t^\tau \mathbb{1}_A\right] &= \mathbb{E}\left[X_{\tau\wedge t}\mathbb{1}_{A\cap\{\tau\le t\}}\right] + \mathbb{E}\left[X_\tau\mathbb{1}_{A\cap\{\tau>t\}}\right] - \mathbb{E}\left[X_{\tau\wedge t}\mathbb{1}_A\right] \\
&= \mathbb{E}\left[(X_\tau - X_{\tau\wedge t})\mathbb{1}_{A\cap\{\tau>t\}}\right] = \mathbb{E}\left[\mathbb{E}\left[X_\tau - X_{\tau\wedge t}|\mathscr{F}_{\tau\wedge t}\right]\mathbb{1}_{A\cap\{\tau>t\}}\right] = 0,
\end{aligned}$$

where the last inequality follows from Theorem 3.61, because $\tau \wedge t \le \tau$ is a stopping time. Since $A \in \mathscr{F}_t$ is arbitrary, and $X_t^\tau = X_{\tau\wedge t}$ is $\mathscr{F}_t$-measurable, we have $\mathbb{E}[X_\tau|\mathscr{F}_t] = X_t^\tau$. Since $X_\tau \in L^1(\Omega,\mathscr{F},\mathbb{P})$, the stopped process $(X_t^\tau)_{t\ge0}$ is uniformly integrable. □

3.4 Continuous Semimartingales

3.4.1 Finite Variation Processes

Review: Functions of bounded variation. Given a function $f : [0,T] \to \mathbb{R}$, we define the *total variation* of f on interval $[0,T]$ as

$$V_0^T f = \sup\left\{\sum_{j=1}^n |f(x_j) - f(x_{j-1})| :\ n \in \mathbb{N},\ 0 = x_0 < x_1 < \cdots < x_n = T\right\}$$

If $V_0^T f < \infty$, we say that $f : [0,T] \to \mathbb{R}$ has *bounded variation.*

We can show that every function of bounded variation is the difference of two monotone increasing functions. If $f : [0,T] \to \mathbb{R}$ 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 $f : [0,T] \to \mathbb{R}$ 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 $t = x_0 < x_1 < \cdots < x_n = T$ 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 $0 = x_0 < x_1 < \cdots < x_n = T$, 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 $\mathscr{B}([0,T])$. Let $P \amalg N = [0,T]$ be the Hahn decomposition associated with μ, and define $Y = \mathbb{1}_P - \mathbb{1}_N$. Let $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T$ be an increasing sequence of partitions of interval $[0,t]$ such that the mesh $\max_{1\leq j\leq k_n}(t_j^n - t_{j-1}^n) \to 0$, and let $\mathscr{B}_n$ be the sub σ-algebra generated by intervals $(t_j^n, t_{j-1}^n]$. Then $(\mathscr{B}_n)_{n=1}^\infty$ is a filtration with $\mathscr{B}_\infty = \mathscr{B}([0,T])$, and $X_n = \mathbb{E}[Y|\mathscr{B}_n]$ 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

$$X_n|_{(t_{j-1}^n, t_j^n]} = \frac{\mathbb{E}\left[X_n \mathbb{1}_{(t_{j-1}^n, t_j^n]}\right]}{\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{\mathbb{E}\left[Y \mathbb{1}_{(t_{j-1}^n, t_j^n]}\right]}{\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{\mu((t_{j-1}^n, t_j^n])}{|\mu|([0,T])\mathbb{P}((t_{j-1}^n, t_j^n])} = \frac{f(t_j^n) - f(t_{j-1}^n)}{|\mu|((t_{j-1}^n, t_j^n])},\ \forall 1 \leq j \leq k_n.$$

Now it suffices to prove that $\sum_{j=1}^{k_n} \left|f(t_j^n) - f(t_{j-1}^n)\right| \to \mu([0,T])$. By Doob's convergence theorem for uniformly integrable martingales, we have $X_n \to Y$ *a.s.* and in L^1. As a result,

$$\mathbb{E}|X_n| = \sum_{j=1}^{k_n} \frac{\left|f(t_j^n) - f(t_{j-1}^n)\right|}{\mu([0,T])} \to \mathbb{E}|Y| = 1.$$

Review: Functions of finite variation. A function $f : \mathbb{R}_+ \to \mathbb{R}$ 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 $\mathscr{B}(\mathbb{R}_+)$ 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 $f : [0, T] \to \mathbb{R}$ 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 $\varphi : [0, T] \to \mathbb{R}$ is a measurable function such that $\int_{[0,T]} |\varphi| \, \mathrm{d}|\mu| < \infty$, define the Lebesgue-Stieltjes integral

$$\int_0^T \varphi(s) \, \mathrm{d}f(s) = \int_{[0,T]} \varphi \, \mathrm{d}\mu, \quad \int_0^T \varphi(s) \, |\mathrm{d}f(s)| = \int_{[0,T]} \varphi \, |\mathrm{d}\mu|.$$

It is seen that the function $t \mapsto \int_0^t \varphi(s) \, \mathrm{d}f(s)$ is also of bounded variation on $[0, T]$. To see this, note the associated signed measure is $\nu(A) = \int_A \varphi \, \mathrm{d}\mu$, and the total variation on $[0, T]$ is $|\nu|([0, T]) = \int_{[0,T]} |\varphi| \, |\mathrm{d}\mu| < \infty$.

Also, if $f : \mathbb{R}_+ \to \mathbb{R}$ is a càdlàg function of finite variation with $f(0) = 0$, we can define the Lebesgue-Stieltjes integral

$$\int_0^\infty \varphi(s) \, \mathrm{d}f(s) = \lim_{T \to \infty} \int_0^T \varphi(s) \, \mathrm{d}f(s)$$

for all measurable functions φ such that $\int_0^\infty |\varphi(s)| \, |\mathrm{d}f(s)| < \infty$.

Review: Approximation of Lebesgue Stieltjes integral. We can approximate a Lebesgue-Stieltjes integral of a continuous function by differentiating on a mesh. Let $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T$ be a sequence of partitions whose mesh $\max_{1 \le j \le k_n}(t_j^n - t_{j-1}^n) \to 0$. Then

$$\int_0^T \varphi_n(s) \, \mathrm{d}f(s) = \sum_{j=1}^{k_n} \varphi(t_j^n) \left(f(t_j^n) - f(t_{j-1}^n) \right), \ \textit{where}\ \varphi_n(s) = \varphi(0) \mathbb{1}_{\{0\}}(s) + \sum_{j=1}^{k_n} \varphi(t_j^n) \mathbb{1}_{(t_{j-1}^n, t_j^n]}(s).$$

By continuity of φ, we have $\varphi_n \to \varphi$, and all these functions are dominated by a constant $\max_{s \in [0,T]} |\varphi(s)|$. By Lebesgue dominated convergence theorem, we have

$$\int_0^T \varphi(s) \, \mathrm{d}f(s) = \lim_{n \to \infty} \sum_{j=1}^{k_n} \varphi(t_j^n) \left(f(t_j^n) - f(t_{j-1}^n) \right).$$

Definition 3.63 (Finite variation processes)**.** An adapted process $(X_t)_{t \ge 0}$ 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 $(X_t)_{t \ge 0}$ is said to be an *increasing process.*

Remark. If $(A_t)_{t \ge 0}$ is a finite variation process, then

$$V_t = \int_0^t |\mathrm{d}A_s|, \ \forall t \in \mathbb{R}_+$$

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 $\mathscr{F}_t$-measurable, because

$$V_t = \lim_{n \to \infty} \sum_{j=1}^{k_n} \left| A_{t_j^n} - A_{t_{j-1}^n} \right|, \text{ where } 0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t \text{ is an increasing sequence of partitions of } [0, t].$$

Proposition 3.64. *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 $\int_0^t |H_s(\omega)|\,|\mathrm{d}A_s(\omega)| < \infty$ for all $t \in \mathbb{R}_+$ and all $\omega \in \Omega$. Then the process $H \cdot A$ defined by*

$$(H \cdot A)_t = \int_0^t H_s \,\mathrm{d}A_s, \ \forall t \in \mathbb{R}_+.$$

is also a finite variation process.

Proof. It suffices to show that $H \cdot A$ is an adapted process, namely, if $h : \Omega \times [0,t] \to \mathbb{R}$ is measurable on $\mathscr{F}_t \times \mathscr{B}([0,t])$, and $\int_0^t |h(\omega,s)|\,|\mathrm{d}A_s(\omega)| < \infty$ for all $\omega \in \Omega$, then $\omega \mapsto \int_0^t h(\omega,s)\,\mathrm{d}A_s(\omega)$ is $\mathscr{F}_t$-measurable. Let $h(\omega,s) = \mathbb{1}_F(\omega)\mathbb{1}_{(p,q]}(s)$, where $F \in \mathscr{F}_t$ and $(p,q] \subset [0,t]$. Then

$$\int_0^t h(\omega,s)\,\mathrm{d}A_s(\omega) = \mathbb{1}_F(\omega)\left(A_q(\omega) - A_p(\omega)\right).$$

Clearly, $\mathbb{1}_F(A_q - A_p)$ is $\mathscr{F}_t$-measurable. Now we define

$$\mathscr{L}_t = \left\{G \in \mathscr{F}_t \otimes \mathscr{B}([0,t]) : \omega \mapsto \int_0^t \mathbb{1}_G(\omega,s)\,\mathrm{d}A_s(\omega) \text{ is } \mathscr{F}_t\text{-measurable}\right\}.$$

Note this is a λ-system containing $\{F \times (p,q] : F \in \mathscr{F}_t,\ (p,q] \subset [0,t]\}$, which is a π-system generating $\mathscr{F}_t \otimes \mathscr{B}([0,t])$. By Sierpiński-Dynkin π-λ theorem, we have $\mathscr{L}_t = \mathscr{F}_t \otimes \mathscr{B}([0,t])$. Hence $\omega \mapsto \int_0^t h(\omega,s)\,\mathrm{d}A_s(\omega)$ is $\mathscr{F}_t$-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 $(\mathscr{F}_t)_{t\geq 0}$ is complete, then Proposition 3.64 holds for all progressive process $(H_t)_{t\geq 0}$ such that $\int_0^t |H_s(\omega)|\,|\mathrm{d}A_s(\omega)| < \infty$ for all $t \in \mathbb{R}_+$ and $\mathbb{P}$-*a.e.* $\omega \in \Omega$. To clarify this, we redefine $H \cdot A = 0$ on the $\mathbb{P}$-null set where $\int_0^t |H_s(\omega)|\,|\mathrm{d}A_s(\omega)| < \infty$.

3.4.2 Continuous Local Martingales

The local martingales is a large class of stochastic processes.

Definition 3.65 (Continuous local martingales)**.** An adapted 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 $\tau_n \uparrow \infty$, and the stopped process $X^{\tau_n} = (X_t^{\tau_n})_{t\geq 0}$ is a uniformly integrable martingale. In that case, the sequence $(\tau_n)_{n=1}^\infty$ of stopping times is said to *reduce* process X.

More generally, an adapted 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 3.65, 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 $\tau_n \wedge n \uparrow \infty$.

The following two basic facts about continuous local martingales immediately follow from Corollary 3.62.

Proposition 3.66. *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 $\sigma_n \uparrow \infty$, then $(\sigma_n \wedge \tau_n)_{n=1}^\infty$ also reduces X.

Remark. 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 3.67. *Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale.*

(i) If $(X_t)_{t\geq 0}$ is nonnegative and $X_0 \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, then $(X_t)_{t\geq 0}$ is a supermartingale.

(ii) If there exists random variable $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ 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 $\mathscr{F}_0$-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 $X_0 \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, we have

$$Y_{s\wedge\tau_n} = \mathbb{E}[Y_{t\wedge\tau_n}|\mathscr{F}_s] \quad \Rightarrow \quad X_{s\wedge\tau_n} = \mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] \tag{3.9}$$

Since X is nonnegative, by Fatou's lemma (conditional version), we have

$$X_s = \liminf_{n\to\infty} X_{s\wedge\tau_n} = \liminf_{n\to\infty} \mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] \geq \mathbb{E}\left[\lim_{n\to\infty} X_{t\wedge\tau_n}|\mathscr{F}_s\right] = \mathbb{E}[X_t|\mathscr{F}_s].$$

(ii) Following (3.9), we use Lebesgue dominated convergence theorem, because $|X_{t\wedge\tau_n}| \leq Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ for all $n \in \mathbb{N}$ and $t \geq 0$. Then $X_{t\wedge\tau_n} \to X_t$ in L^1, and $X_s = \mathbb{E}[X_t|\mathscr{F}_s]$ for $0 \leq s < t$.

(iii) By Proposition 3.14 (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 $\mathscr{F}_0$-measurable, we have $\mathbb{E}[WX_{t\wedge\tau_n}|\mathscr{F}_s] = W\mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] = WX_{s\wedge\tau_n}$ for all $t > s \geq 0$. Hence $(\tau_n)_{n=1}^\infty$ also reduces $(WX_t)_{t\geq 0}$, and the conclusion follows. □

Proposition 3.68. *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 $\omega \in \Omega\backslash N$.*

Proof. Since X is a finite variation process, $\int_0^t |\mathrm{d}X_s|$ is an increasing process with continuous sample paths. For every $n \in \mathbb{N}$, define the stopping time

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t |\mathrm{d}X_s| \geq n\right\},$$

and set $Y_t = X_t^{\tau_n}$. Then $Y_t \leq \int_0^{t\wedge\tau_n} |\mathrm{d}X_s| \leq n$. By Proposition 3.67 (ii), Y is a uniformly integrable martingale. Let $0 = t_0 < t_1 < \cdots < t_k = t$ be a partition of $[0, t]$. By Proposition 3.51, we have

$$\mathbb{E}\left[Y_t^2\right] = \sum_{j=1}^p \mathbb{E}\left[\left(Y_{t_j} - Y_{t_{j-1}}\right)^2\right] \leq \mathbb{E}\left[\sup_{1\leq j\leq k} \left|Y_{t_j} - Y_{t_{j-1}}\right| \sum_{j=1}^p \left|Y_{t_j} - Y_{t_{j-1}}\right|\right] \leq n\mathbb{E}\left[\sup_{1\leq j\leq k} \left|Y_{t_j} - Y_{t_{j-1}}\right|\right].$$

Now we take a sequence of increasing partitions $0 = t_0^m < t_1^m < \cdots < t_{k_m}^m = t$ of $[0, t]$ whose mesh converges to 0. By continuity of sample paths of Y and Lebesgue dominated convergence theorem, we have

$$\sup_{1\leq j\leq k_m} \left|Y_{t_j^m}(\omega) - Y_{t_{j-1}^m}(\omega)\right| \to 0 \quad as \quad m \to \infty, \ \forall \omega \in \Omega \quad \Rightarrow \quad \mathbb{E}\left[\sup_{1\leq j\leq k_m} \left|Y_{t_j^m} - Y_{t_{j-1}^m}\right|\right] \to 0.$$

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 $n \to \infty$, we then have $X_t = 0$ *a.s.* for all $\in \mathbb{R}_+$. 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 $\omega \in \Omega\backslash N$. □

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 $\omega \in \Omega\backslash N$, then X and Y are said to be *indistinguishable.* Note this is a stronger condition than *a.s.* modification.

3.4.3 Quadratic Variation and Covariation

From now on we assume that the filtration $(\mathscr{F}_t)_{t\geq 0}$ is complete.

Theorem 3.69 (Quadratic variation)**.** *Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale. There exists an increasing process denoted by $\langle X, X\rangle = (\langle X, X\rangle_t)_{t\geq 0}$, which is unique up to indistinguishability, such that $X^2 - \langle X, X\rangle$ is a continuous local martingale. Furthermore, for every fixed $t > 0$, if $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ is an increasing sequence of partitions of $[0, t]$ with the mesh $\max_{1\leq j\leq k_n} |t_j^n - t_{j-1}^n| \to 0$, then*

$$\langle X, X\rangle_t = \lim_{n\to\infty} \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \tag{3.10}$$

*in probability. The process $\langle X, X\rangle$ is called 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 3.68, $Y_t' - Y_t = 0$ *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 3.67 (ii). We fix $T > 0$ and an increasing sequence of partitions of $[0, T]$ with the mesh $\max_{1\leq j\leq k_n} |t_j^n - t_{j-1}^n| \to 0$. The for every $s > r \geq 0$ and every bounded $\mathscr{F}_r$-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 $0 \leq t' < t$,

$$\mathbb{E}\left[Z\,(X_{s\wedge t} - X_{r\wedge t})\,|\mathscr{F}_{t'}\right] = \begin{cases} \mathbb{E}\left[\mathbb{E}[Z\,(X_{s\wedge t} - X_{r\wedge t})\,|\mathscr{F}_r]|\mathscr{F}_{t'}\right] = 0 & \text{if } \ t' < r, \\ Z\,\mathbb{E}\left[X_{s\wedge t} - X_r|\mathscr{F}_{t'}\right] = Z(X_{s\wedge t'} - X_r) & \text{if } \ t' \geq r. \end{cases}$$

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. $\lim_{n,m\to\infty} \mathbb{E}\left[(M_T^n - M_T^m)^2\right] = 0$.

Proof of the Claim. We fix $m \leq n$ and evaluate the product $\mathbb{E}[M_T^n M_T^m]$:

$$\begin{aligned}
\mathbb{E}[M_T^n M_T^m] &= \sum_{i=1}^{k_m}\sum_{j=1}^{k_n} \mathbb{E}\left[X_{t_{i-1}^m}\left(X_{t_i^m} - X_{t_{i-1}^m}\right) X_{t_{j-1}^n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)\right] \\
&= \sum_{j=1}^{k_n} \sum_{i:(t_{i-1}^m, t_i^m]\supset(t_{j-1}^n, t_j^n]} \mathbb{E}\left[X_{t_{i-1}^m}\left(X_{t_i^m} - X_{t_{i-1}^m}\right) X_{t_{j-1}^n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)\right] \\
&= \sum_{j=1}^{k_n} \sum_{i:(t_{i-1}^m, t_i^m]\supset(t_{j-1}^n, t_j^n]} \sum_{l:(t_{l-1}^n, t_l^n]\subset(t_{i-1}^m, t_i^m]} \mathbb{E}\left[X_{t_{i-1}^m}\left(X_{t_l^n} - X_{t_{l-1}^n}\right) X_{t_{j-1}^n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)\right] \\
&= \sum_{j=1}^{k_n} \sum_{i:(t_{i-1}^m, t_i^m]\supset(t_{j-1}^n, t_j^n]} \mathbb{E}\left[X_{t_{i-1}^m} X_{t_{j-1}^n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right]
\end{aligned} \tag{3.11}$$

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 $\mathscr{F}_{t_{j-1}^n}$ (resp. $\mathscr{F}_{t_{i-1}^m}$) 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 $\mathscr{F}_{t_{j-1}^n}$ (resp. $\mathscr{F}_{t_{l-1}^n}$) to eliminate the corresponding term in the triple sum. As a special case of (3.11), we have

$$\mathbb{E}\left[(M_T^n)^2\right] = \sum_{j=1}^{k_n} \mathbb{E}\left[X_{t_{j-1}^n}^2 \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right]. \tag{3.12}$$

And by Proposition 3.51,

$$\begin{aligned}
\mathbb{E}\left[(M_T^m)^2\right] = \sum_{i=1}^{k_m} \mathbb{E}\left[X_{t_{i-1}^m}^2 \left(X_{t_i^m} - X_{t_{i-1}^m}\right)^2\right] &= \sum_{i=1}^{k_m} \mathbb{E}\left[X_{t_{i-1}^m}^2 \mathbb{E}\left[\left(X_{t_i^m} - X_{t_{i-1}^m}\right)^2 \middle| \mathscr{F}_{t_{i-1}^m}\right]\right] \\
&= \sum_{i=1}^{k_m} \sum_{j:(t_{j-1}^n, t_j^n] \subset (t_{i-1}^m, t_i^m]} \mathbb{E}\left[X_{t_{i-1}^m}^2 \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \middle| \mathscr{F}_{t_{i-1}^m}\right]\right] \\
&= \sum_{j=1}^{k_n} \sum_{i:(t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E}\left[X_{t_{i-1}^m}^2 \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right]. \qquad (3.13)
\end{aligned}$$

Note that for every $j \in \{1, \cdots, k_n\}$, there is a unique $i \in \{1, \cdots, k_m\}$ 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

$$\begin{aligned}
\mathbb{E}\left[(M_T^n - M_T^m)^2\right] &= \sum_{j=1}^{k_n} \sum_{i:(t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \mathbb{E}\left[\left(X_{t_{j-1}^n} - X_{t_{i-1}^m}\right)^2 \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right] \\
&\leq \mathbb{E}\left[\sup_{1 \leq j \leq k_n,\ (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \left(X_{t_{j-1}^n} - X_{t_{i-1}^m}\right)^2 \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right] \\
&\leq \mathbb{E}\left[\sup_{1 \leq j \leq k_n,\ (t_{i-1}^m, t_i^m] \supset (t_{j-1}^n, t_j^n]} \left(X_{t_{j-1}^n} - X_{t_{i-1}^m}\right)^4\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right)^2\right]^{\frac{1}{2}} \qquad (3.14)
\end{aligned}$$

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 $n, m \to \infty$. 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

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right)^2\right] = \sum_{j=1}^{k_n} \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^4\right] + 2 \sum_{1 \leq j < i \leq k_n} \mathbb{E}\left[\left(X_{t_i^n} - X_{t_{i-1}^n}\right)^2 \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right] \\
&\leq 4K^2 \sum_{j=1}^{k_n} \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right] + 2 \sum_{j=1}^{k_n - 1} \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \sum_{i=j+1}^{k_n} \mathbb{E}\left[\left(X_{t_i^n} - X_{t_{i-1}^n}\right)^2 |\mathscr{F}_{t_j^n}\right]\right] \\
&= 4K^2 \mathbb{E}\left[(X_T - X_0)^2\right] + 2 \sum_{j=1}^{k_n - 1} \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \mathbb{E}\left[\left(X_T - X_{t_j^n}\right)^2 |\mathscr{F}_{t_j^n}\right]\right] \qquad \text{(By Proposition 3.51)} \\
&\leq 12K^2 \mathbb{E}\left[(X_T - X_0)^2\right] \leq 48K^4.
\end{aligned}$$

Hence we can bound the second term in (3.14) by $4\sqrt{3}K^2$, which completes the proof. $\square$

Proof of Theorem 3.69 (Cont). *Step III:* By Doob's L^p-inequality [Proposition 3.52 (ii)] and our claim,

$$0 \le \lim_{n,m\to\infty} \mathbb{E}\left[\sup_{0\le t\le T} (M_t^n - M_t^m)^2\right] \le \lim_{n,m\to\infty} 4\mathbb{E}\left[(M_T^n - M_T^m)^2\right] = 0.$$

Hence for all $t \in [0,T]$, $(M_t^n)_{n=1}^\infty$ is a Cauchy sequence in $L^2(\Omega, \mathscr{F}, \mathbb{P})$ and thus converges in L^2. We choose a subsequence $n_k \nearrow \infty$ such that

$$\mathbb{E}\left[\sup_{0\le t\le T} \left(M_t^{n_{k+1}} - M_t^{n_k}\right)^2\right] < 2^{-k},\ \forall k \in \mathbb{N}.$$

As a result,

$$\mathbb{E}\left[\sum_{k=1}^\infty \sup_{0\le t\le T} \left|M_t^{n_{k+1}} - M_t^{n_k}\right|\right] < \sum_{k=1}^\infty 2^{-k} < \infty \quad \Rightarrow \quad \sum_{k=1}^\infty \sup_{0\le t\le T} \left|M_t^{n_{k+1}} - M_t^{n_k}\right| < \infty \ a.s..$$

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 $k \to \infty$. Let $Y_t(\omega) = \lim_{k\to\infty} M_t^{n_k}(\omega)$ for all $t \in [0,T]$ if $\omega \notin N$, and otherwise $Y_t(\omega) = 0$ for all $t \in [0,T]$. Then $(Y_t)_{t\ge 0}$ has continuous sample paths. Also, Y_t is adapted by completeness of our filtration $(\mathscr{F}_t)_{t\ge0}$. Moreover, since $(M_n^t)_{n=1}^\infty$ converges in L^2, it must converges to the *a.s.* limit Y_t in L^2. Also, since the conditional expectation is a bounded linear operator in $L^2(\Omega, \mathscr{F}, \mathbb{P})$, we can pass the martingale property of M_t^n to Y_t to obtain that $\mathbb{E}[Y_t|\mathscr{F}_s] = Y_s$ for all $0 \le s < t \le T$. Hence $(Y_{t\wedge T})_{t\ge0}$ is a 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 $n_k \uparrow \infty$, we have $X_t^2 - 2M_t^{n_k} \rightrightarrows X_t^2 - Y_t$ on $[0,T]$ except possibly on the negligible set N. Set $V_t^T = X_t^2 - 2Y_t$ on $\Omega\backslash N$, and $V_t^T = 0$ on N. Then $V_0^T = 0$, V_t^T is $\mathscr{F}_t$-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 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 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 $M_T^n \to Y_T = \frac{1}{2}(X_T^2 - V_T^T)$ in L^2, we have

$$X_T^2 - 2M_T^n = \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \overset{L^2}{\to} V_T^T = \langle X,X\rangle_T \ a.s..$$

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 $\tau_n = \inf\{t \ge 0 : |X_t| \ge n\}$. By Proposition 3.67 (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 $t \ge 0$, 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 $\eta > 0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^{k_m} \left(X_{t_j} - X_{t_{j-1}}\right)^2 - \langle X,X\rangle_t\right| \ge \eta\right) \le \frac{1}{\eta^2}\left\|\sum_{j=1}^{k_m}\left(X_{t_j}^{\tau_n} - X_{t_{j-1}}^{\tau_n}\right)^2 - \langle X,X\rangle_{t\wedge\tau_n}\right\|_2^2 + \mathbb{P}(\tau_n < t). \quad (3.15)$$

In (3.15), the first term converges to 0 as $m \to \infty$, 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 $n \to \infty$ 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 3.67 (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 $\mathscr{F}_0$-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 3.70. *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 3.67. 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 3.71. *Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale such that $X_0 \in L^2(\Omega, \mathscr{F}, \mathbb{P})$. Then the following are equivalent: (i) X is a martingale, and $\sup_{t\geq 0}\mathbb{E}|X_t|^2 < \infty$; (ii) $\mathbb{E}\left[\langle X, X\rangle_\infty\right] < \infty$. Furthermore, if these properties hold, then $X^2 - \langle X, X\rangle$ 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 3.52] and monotone convergence theorem, we have

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t|^2\right] \leq 4\mathbb{E}|X_T|^2,\quad \forall T\geq 0,\quad \text{and}\quad \mathbb{E}\left[\sup_{t\geq 0}|X_t|^2\right] \leq 4\sup_{t\geq 0}\mathbb{E}|X_t|^2 < \infty$$

Define $\sigma_n = \inf\{t \geq 0 : \langle X, X\rangle_t \geq n\} \uparrow \infty$. 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 $\sup_{t\geq 0}|X_t|^2 + n$. By Proposition 3.67 (ii), this is a uniformly integrable martingale, and

$$\mathbb{E}\left[\langle X, X\rangle_{t\wedge\sigma_n}\right] = \mathbb{E}\left[X_{t\wedge\sigma_n}^2\right] \leq \mathbb{E}\left[\sup_{t\geq 0}|X_t|^2\right] \leq 4\sup_{t\geq 0}\mathbb{E}[|X_t|^2].$$

Let $n\to\infty$ and $t\to\infty$, 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 3.67 (ii), this is a uniformly integrable martingale. By Fatou's lemma, we have

$$\mathbb{E}\left[X_{t\wedge\tau_n}^2\right] = \mathbb{E}\left[\langle X, X\rangle_{t\wedge\tau_n}\right] \leq \mathbb{E}\left[\langle X, X\rangle_\infty\right] < \infty \quad\Rightarrow\quad \mathbb{E}[X_t^2] \leq \liminf_{n\to\infty}\mathbb{E}\left[X_{t\wedge\tau_n}^2\right] \leq \mathbb{E}\left[\langle X, X\rangle_\infty\right] < \infty.$$

Meanwhile, the sequence $|X_{t\wedge\tau_n}| \uparrow |X_t|$ as $n\to\infty$, and $(X_{t\wedge\tau_n})_{n=1}^\infty$ is uniformly integrable:

$$\begin{aligned}\lim_{M\to\infty}\sup_{n\in\mathbb{N}}\mathbb{E}\left[|X_{t\wedge\tau_n}|\mathbb{1}_{\{|X_{t\wedge\tau_n}|\geq M\}}\right] &\leq \lim_{M\to\infty}\sup_{n\in\mathbb{N}}\mathbb{E}\left[X_{t\wedge\tau_n}^2\right]\mathbb{E}\left[\mathbb{1}_{\{|X_{t\wedge\tau_n}|\geq M\}}\right]\\ &\leq \mathbb{E}\left[\langle X, X\rangle_\infty\right]\lim_{M\to\infty}\mathbb{E}\left[\mathbb{1}_{\{|X_t|\geq M\}}\right] = 0.\end{aligned}$$

As a result, $X_{t\wedge\tau_n} \to X_t$ *a.s.* and in L^1. By Proposition 3.67 (iii), $(X_{t\wedge\tau_n})_{t\geq 0}$ is a martingale, and we have $\mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] = X_{s\wedge\tau_n}$ for all $t > s \geq 0$. Convergence in L^1 implies $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$, 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 3.67 (ii), this is a uniformly integrable martingale. $\square$

The following corollary is derived by applying Theorem 3.71 on $(X_{t\wedge T})_{t\geq 0}$ for each $T \geq 0$.

Corollary 3.72. *Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale such that $X_0 \in L^2(\Omega, \mathscr{F}, \mathbb{P})$. Then the following are equivalent: (i) X is a martingale, and $X_t \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ for all $t \geq 0$; (ii) $\mathbb{E}[\langle X, X\rangle_t] < \infty$ for all $t \geq 0$. Furthermore, if these properties hold, then $X_t^2 - \langle X, X\rangle_t$ is a martingale.*

Corollary 3.73. *Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale such that $\langle X, X\rangle_\infty < \infty$ a.s.. Then X converges a.s..*

Proof. If $\langle X, X\rangle_\infty < \infty$, the stopping time $T_n = \inf\{t \geq 0 : \langle X, X\rangle_t \geq n\}$ a.s. increases to ∞ as $n \to \infty$. By Theorem 3.71, the local martingale X^{T_n} is a L^2-bounded martingale, which converges a.s.. On the event $\{\langle X, X\rangle_\infty < \infty\}$ we have $T_n = \infty$ a.s. from some n on, which completes the proof. □

Definition 3.74 (Bracket)**.** Let $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ be two continuous local martingales. The *bracket* (or *quadratic covariation*) $\langle X, Y\rangle$ is defined as the following finite variation process:

$$\langle X, Y\rangle_t = \frac{1}{2}\left(\langle X+Y, X+Y\rangle_t - \langle X, X\rangle_t - \langle Y, Y\rangle_t\right), \quad t \geq 0.$$

We have the following properties and approximation formula for the bracket.

Proposition 3.75. *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 $X_tY_t - \langle X, Y\rangle_t$ 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 $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ of $[0, t]$ with mesh tending to 0,

$$\langle X, Y\rangle_t = \lim_{n\to\infty} \sum_{j=1}^{k_n} (X_{t_j^n} - X_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n}) \quad \textit{in probability.}$$

(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 continuous martingales, $X_tY_t - \langle X, Y\rangle_t$ 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 $t \to \infty$, and $\mathbb{E}[X_\infty Y_\infty] = \mathbb{E}[X_0Y_0] + \mathbb{E}[\langle X, Y\rangle_\infty]$.

(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) Since $XY = \frac{1}{2}(X+Y)^2 - X^2 - Y^2$, the process $X_tY_t - \langle X, Y\rangle_t$ is a continuous local martingale. The uniqueness argument is similar to Theorem 3.69.

(ii) is a consequence of the uniqueness argument. (iii) follows from (3.10).

(iv) According to (iii), we have

$$\begin{aligned} \langle X^\tau, Y^\tau\rangle_t = \langle X^\tau, Y\rangle_t = \langle X, Y\rangle_t \quad &\textit{on } \{\tau \geq t\}, \\ \langle X^\tau, Y^\tau\rangle_t - \langle X^\tau, Y^\tau\rangle_\tau = \langle X^\tau, Y\rangle_t - \langle X^\tau, Y\rangle_\tau = 0 \quad &\textit{on } \{\tau < t\}. \end{aligned}$$

(v) is a consequence of (iv), since $X_t^\tau(Y_t - Y_t^\tau) = X_t^\tau Y_t - \langle X^\tau, Y\rangle_t - (X_t^\tau Y_t^\tau - \langle X^\tau, Y^\tau\rangle_t)$.

(vi) is a consequence of Theorem 3.71.

(vii) If XY is a 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 local martingale. □

Theorem 3.76 (Kunita-Watanabe)**.** *Let X and Y be two continuous local martingales, and let H and K be two measurable processes. Then*

$$\int_0^\infty |H_s|\,|K_s|\,|\mathrm{d}\langle X,Y\rangle_s| \le \left(\int_0^\infty H_s^2\,\mathrm{d}\langle X,X\rangle_s\right)^{1/2}\left(\int_0^\infty K_s^2\,\mathrm{d}\langle Y,Y\rangle_s\right)^{1/2}.$$

Proof. Given $t > s \ge 0$, we abuse the notation $\langle X,Y\rangle_s^t = \langle X,Y\rangle_t - \langle X,Y\rangle_s$. Let $s = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ be a increasing sequence of partitions of $[s,t]$ with the mesh tending to 0. Let $S_{XX}^n = \sum_{j=1}^{k_n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2$, $S_{YY}^n = \sum_{j=1}^{k_n}\left(Y_{t_j^n} - Y_{t_{j-1}^n}\right)^2$, and $S_{XY}^2 = \sum_{j=1}^{k_n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)\left(Y_{t_j^n} - Y_{t_{j-1}^n}\right)$. By Cauchy-Schwarz inequality, we have $S_{XY}^n \le \sqrt{S_{XX}^n S_{YY}^n}$. 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 $\eta > 0$,

$$\mathbb{P}\left(\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} > \eta\right) \le \mathbb{P}\left(\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} - S_{XY}^n + \sqrt{S_{XX}^n S_{YY}^n} > \eta\right) \to 0.$$

By taking the union of all rationals $\eta > 0$, we have $\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} \le 0$ *a.s.*. Since $t > s \ge 0$ are arbitrary, $\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} \le 0$ holds for all rationals $t > s \ge 0$ for *a.s.* $\omega \in \Omega$. By continuity of X and Y, we have for *a.s.* $\omega \in \Omega$ that $\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} \le 0$ for all reals $t > s \ge 0$.

Now we fix $\omega \in \Omega$ with $\langle X,Y\rangle_s^t - \sqrt{\langle X,X\rangle_s^t\langle Y,Y\rangle_s^t} \le 0$ for all reals $t > s \ge 0$. Then all remaining results are deterministic. For any subdivisions $s = t_0 < t_1 < \cdots < t_k = t$, we have

$$\sum_{j=1}^k \left|\langle X,Y\rangle_{t_{j-1}}^{t_j}\right| \le \sum_{j=1}^k \sqrt{\langle X,X\rangle_{t_{j-1}}^{t_j}}\sqrt{\langle Y,Y\rangle_{t_{j-1}}^{t_j}} \le \sqrt{\sum_{j=1}^k \langle X,X\rangle_{t_{j-1}}^{t_j}}\sqrt{\sum_{j=1}^k \langle Y,Y\rangle_{t_{j-1}}^{t_j}} = \sqrt{\langle X,X\rangle_s^t}\sqrt{\langle Y,Y\rangle_s^t}.$$

Let the mesh of our partition tends to 0, we obtain

$$\int_s^t |\mathrm{d}\langle X,Y\rangle_u| \le \sqrt{\langle X,X\rangle_s^t}\sqrt{\langle Y,Y\rangle_s^t}$$

Fix $T > 0$, and let $\mathscr{M}_T$ be the collection of all $A \in \mathscr{B}([0,T])$ such that

$$\int_A |\mathrm{d}\langle X,Y\rangle_u| \le \sqrt{\int_A \mathrm{d}\langle X,X\rangle_u}\sqrt{\int_A \mathrm{d}\langle Y,Y\rangle_u}. \tag{3.16}$$

By monotone convergence theorem, $\mathscr{M}_T$ 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 $\mathscr{M}_T = \mathscr{B}([0,T])$. As a result, (3.16) holds for all bounded Borel sets $A \in \mathscr{B}(\mathbb{R}_+)$. Also, for all nonnegative simple functions h,k on $[0,T]$, choose finite partition $A_1,\cdots,A_m$ of $[0,T]$ such that $h = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$ and $k = \sum_{i=1}^m \beta_j \mathbb{1}_{A_i}$. Then we have

$$\begin{aligned}\int h(s)k(s)\,|\mathrm{d}\langle X,Y\rangle_s| = \sum_{i=1}^m \alpha_i\beta_i \int_{A_i} |\mathrm{d}\langle X,Y\rangle_s| &\le \sqrt{\sum_{i=1}^m \alpha_i^2 \int_{A_i} \mathrm{d}\langle X,X\rangle_u}\sqrt{\sum_{i=1}^m \beta_i^2 \int_{A_i} \mathrm{d}\langle Y,Y\rangle_u} \\ &= \sqrt{\int h(s)^2\,\mathrm{d}\langle X,X\rangle_s}\sqrt{\int k(s)^2\,\mathrm{d}\langle Y,Y\rangle_s}\end{aligned}$$

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 $T \to \infty$. Hence an application of monotone convergence theorem finishes the proof. $\square$

3.4.4 Continuous Semimartingales

Definition 3.77 (Continuous semimartingales)**.** A process $X = (X_t)_{t\geq 0}$ is said to be a *continuous semimartingale* if it can be written as

$$X_t = M_t + A_t, \ \forall t \in \mathbb{R}_+, \tag{3.17}$$

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 3.68, the decomposition (3.17) is unique up to indistinguishability. We call this the *canonical decomposition* of a continuous semimartingale X.

Definition 3.78 (Bracket)**.** 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 3.79. *Let $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ be two continuous semimartingales. Let $t > 0$. Let $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ be any increasing sequence of partitions of $[0, t]$ whose mesh tends to 0. Then*

$$\langle X, Y\rangle_t = \lim_{n\to\infty} \sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)\left(Y_{t_j^n} - Y_{t_{j-1}^n}\right) \quad \textit{in probability.}$$

Proof. Let $X = M + A$ and $Y = M' + A'$ be the canonical decompositions. Then

$$\begin{aligned}\sum_{j=1}^{k_n} \left(X_{t_j^n} - X_{t_{j-1}^n}\right)\left(Y_{t_j^n} - Y_{t_{j-1}^n}\right) &= \sum_{j=1}^{k_n} \left(M_{t_j^n} - M_{t_{j-1}^n}\right)\left(M'_{t_j^n} - M'_{t_{j-1}^n}\right) + \sum_{j=1}^{k_n} \left(M_{t_j^n} - M_{t_{j-1}^n}\right)\left(A'_{t_j^n} - A'_{t_{j-1}^n}\right) \\ &\quad + \sum_{j=1}^{k_n} \left(A_{t_j^n} - A_{t_{j-1}^n}\right)\left(M'_{t_j^n} - M'_{t_{j-1}^n}\right) + \sum_{j=1}^{k_n} \left(A_{t_j^n} - A_{t_{j-1}^n}\right)\left(A'_{t_j^n} - A'_{t_{j-1}^n}\right)\end{aligned}$$

According to Proposition 3.75 (iii),

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \left(M_{t_j^n} - M_{t_{j-1}^n}\right)\left(M'_{t_j^n} - M'_{t_{j-1}^n}\right) = \langle M, M'\rangle_t = \langle X, Y\rangle_t \quad \text{in probability.}$$

Also, note that

$$\begin{aligned}\left|\sum_{j=1}^{k_n} \left(M_{t_j^n} - M_{t_{j-1}^n}\right)\left(A'_{t_j^n} - A'_{t_{j-1}^n}\right)\right| &\leq \left(\int_0^t |\mathrm{d}A'_s|\right) \sup_{1\leq j\leq k_n} \left|M_{t_j^n} - M_{t_{j-1}^n}\right| \to 0, \ \textit{a.s.}, \\ \left|\sum_{j=1}^{k_n} \left(A_{t_j^n} - A_{t_{j-1}^n}\right)\left(M'_{t_j^n} - M'_{t_{j-1}^n}\right)\right| &\leq \left(\int_0^t |\mathrm{d}A_s|\right) \sup_{1\leq j\leq k_n} \left|M'_{t_j^n} - M'_{t_{j-1}^n}\right| \to 0, \ \textit{a.s.}, \\ \left|\sum_{j=1}^{k_n} \left(A_{t_j^n} - A_{t_{j-1}^n}\right)\left(A'_{t_j^n} - A'_{t_{j-1}^n}\right)\right| &\leq \left(\int_0^t |\mathrm{d}A_s|\right) \sup_{1\leq j\leq k_n} \left|A'_{t_j^n} - A'_{t_{j-1}^n}\right| \to 0, \ \textit{a.s.},\end{aligned}$$

where the *a.s.* convergence holds by continuity of sample paths of M, M' and A'. □

4 Brownian Motions: Part I

4.1 Pre-Brownian Motions and Brownian Motions

Definition 4.1 (Gaussian spaces)**.** A (centered) *Gaussian space* H is a closed subspace of $L^2(\Omega, \mathscr{F}, \mathbb{P})$ that contains only centered Gaussian variables.

Remark. To justify the closedness of a Gaussian space $H \subset L^2(\Omega, \mathscr{F}, \mathbb{P})$, we let $H \ni X_n \sim N(0, \sigma_n^2) \xrightarrow{L^2} X$. Convergence in L^2 implies $\mathbb{E}[X_n^2] = \sigma_n^2 \to \sigma^2$. Then for all $\lambda \in \mathbb{R}$, by dominated convergence theorem,

$$\mathbb{E}\left[e^{\mathrm{i}\lambda X}\right] = \lim_{n\to\infty} \mathbb{E}\left[e^{\mathrm{i}\lambda X_n}\right] = \lim_{n\to\infty} \exp\left(-\frac{\sigma_n^2}{2}\lambda^2\right) = \exp\left(-\frac{\sigma^2}{2}\lambda^2\right).$$

Hence $X \sim N(\mu, \sigma^2) \in H$. Furthermore, since $L^2(\Omega, \mathscr{F}, \mathbb{P})$ 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 $s, t \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\mathrm{i}(sX+tY)}\right] = \exp\left(-\frac{s^2}{2}\mathbb{E}[X^2] - st\,\mathbb{E}[XY] - \frac{t^2}{2}\mathbb{E}[Y^2]\right) = \exp\left(-\frac{s^2}{2}\mathbb{E}[X^2] - \frac{t^2}{2}\mathbb{E}[Y^2]\right) = \mathbb{E}\left[e^{\mathrm{i}sX}\right]\mathbb{E}\left[e^{\mathrm{i}tY}\right].$$

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, $\mathbb{E}[X|\sigma(G)] = \mathbb{E}[\xi + (X-\xi)|\sigma(G)] = \mathbb{E}[\xi|\sigma(G)] = \xi$.

4.1.1 Gaussian White Noises and Pre-Brownian Motions

Definition 4.2 (Gaussian white noise)**.** Let $(E, \mathscr{E})$ be a measurable space, and let μ be a σ-finite measure on $(E, \mathscr{E})$. A *Gaussian white noise with intensity* μ is an isometry W from $L^2(E, \mathscr{E}, \mu)$ into a Gaussian space.

Remark. (i) According to the polarization identity, an isometry W also preserves inner product. Therefore, if $f, g \in L^2(E, \mathscr{E}, \mu)$, then we have

$$\mathbb{E}[W(f)W(g)] = \langle f, g\rangle = \int fg\,\mathrm{d}\mu, \quad \text{and in particular,} \quad \mathbb{E}[W(f)^2] = \|f\|_2^2 = \int |f|^2\,\mathrm{d}\mu.$$

If $f = \mathbb{1}_A$ with $\mu(A) < \infty$, we write $W(A) = W(\mathbb{1}_A)$, and $W(A) \sim N(0, \mu(A))$.

(ii) Given any σ-finite measure μ on $(E, \mathscr{E})$, we can always find a Gaussian white noise with intensity μ on an appropriate probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{e_\lambda, \lambda \in \Lambda\}$ be an orthonormal basis of $L^2(E, \mathscr{E}, \mu)$. According to Corollary 4.19, we define $(\Omega, \mathscr{F}, \mathbb{P}) = (\mathbb{R}^\Lambda, \mathscr{B}(\mathbb{R})^{\otimes\Lambda}, \mathbb{G})$, where $\mathbb{G}$ extends Gaussian measures

$$\mathbb{G}_{(t_1,\cdots,t_n)}(A) = \frac{1}{(2\pi)^{n/2}}\int_A e^{-\frac{1}{2}(z_1^2+\cdots+z_n^2)}\,\mathrm{d}z, \quad \forall t_1, \cdots, t_n \in \mathcal{T}, \ \ \forall A \in \mathscr{B}(\mathbb{R}^n).$$

Then the coordinate maps $(\pi_\lambda)_{\lambda\in\Lambda}$ is a collection of independent standard Gaussian variables. For every $f \in L^2(E, \mathscr{E}, \mu)$, 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 $L^2(E, \mathscr{E}, \mu)$. Hence W takes values in the Gaussian space $H = \overline{\operatorname{span}}\,(\pi_\lambda)_{\lambda \in \Lambda}$. Since W maps an orthonormal basis in $L^2(E, \mathscr{E}, \mu)$ to one in H, it is an isometry. Thus we find a Gaussian white noise W with intensity μ.

(iii) Given a measurable set A in $(E, \mathscr{E}, \mu)$ 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

$$\lim_{n\to\infty} \left(\max_{j \in \{1, \cdots, k_n\}} \mu(A_j^n) \right) = 0.$$

Then $W(A_j^n),\ j = 1, \cdots, k_n$ are independent Gaussian variables, and $\mathbb{E}[W(A_j^n)^2] = \mu(A_j)$. Furthermore,

$$\mathbb{E}\left[\left(\sum_{j=1}^{k_n} W(A_j^n)^2 - \mu(A)\right)^2\right] = \sum_{j=1}^{k_n} \mathbb{E}\left[\left(W(A_j^n)^2 - \mu(A_j^n)\right)^2\right] = 2\sum_{j=1}^{k_n} \mu(A_j^n)^2 \le 2\mu(A) \max_{1 \le j \le k_n} \mu(A_j^n) \to 0.$$

This implies

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} W(A_j^n)^2 = \mu(A) \quad \textit{in } L^2.$$

Definition 4.3 (Pre-Brownian motion)**.** Give $\mathbb{R}_+$ the Borel σ-algebra $\mathscr{B}(\mathbb{R}_+)$ and the Lebesgue measure m, and let W be a Gaussian white noise on $\mathbb{R}_+$ with intensity m. The process $(B_t)_{t\ge 0}$ defined by

$$B_t = W(\mathbb{1}_{[0,t]}),\ \forall t \in \mathbb{R}_+$$

is said to be a *pre-Brownian motion.*

Remark. By definition, a pre-Brownian motion $B = (B_t)_{t\ge0}$ is a Gaussian process, i.e. the linear combination of any finitely many observations $B_{t_1}, \cdots, B_{t_n}$ is Gaussian. The covariance function of this process is given by

$$K(s,t) = \mathbb{E}\left[B_s B_t\right] = \int \mathbb{1}_{[0,s]\cap[0,t]}\,\mathrm{d}m = s \wedge t, \quad \forall s, t \in \mathbb{R}_+.$$

Proposition 4.4 (Characterization of pre-Brownian motions)**.** *Let $B = (B_t)_{t\ge0}$ be a (real-valued) stochastic process. The following are equivalent:*

(i) $(B_t)_{t\ge0}$ is a pre-Brownian motion.

(ii) $(B_t)_{t\ge0}$ is a centered Gaussian process with covariance $K(s,t) = s \wedge t$.

(iii) $B_0 = 0$ a.s., and for every $t > s \ge 0$, 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 $0 = t_0 < t_1 < \cdots < t_n$, the variables $\{B_{t_j} - B_{t_{j-1}}, j = 1, \cdots, n\}$ are independent, and for every $j = 1, \cdots, p$, 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.

(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

$$\mathbb{E}\left[(B_t - B_s)^2\right] = t - 2(s \wedge t) + s = t - s,$$
$$\mathbb{E}\left[(B_t - B_s)B_r\right] = t \wedge r - s \wedge r = r - r = 0,\ \forall r \in [0,s].$$

Hence $X_t - X_s \sim N(0, t-s)$, and X_t is independent of all X_r with $r \in [0,s]$.

(iv) $\Rightarrow$ (i). It suffices to show that there exists an isometry W between $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$ and a Gaussian space H. For all step functions $f = \sum_{j=1}^n \lambda_j \mathbb{1}_{(t_{j-1},t_j]}$ in $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$, define

$$W(f) = \sum_{j=1}^n \lambda_j (B_{t_j} - B_{t_{j-1}}).$$

If $f, g \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$ are two step functions, we can find a partition $0 = t_0 < t_1 < \cdots < t_n$ such that $f = \sum_{j=1}^n \lambda_j \mathbb{1}_{(t_{j-1},t_j]}$ and $g = \sum_{j=1}^n \nu_j \mathbb{1}_{(t_{j-1},t_j]}$. According to (iii), we have

$$\mathbb{E}[W(f)W(g)] = \mathbb{E}\left[\sum_{j=1}^n \sum_{k=1}^n \lambda_j \nu_k (B_{t_j} - B_{t_{j-1}})(B_{t_k} - B_{t_{k-1}})\right] = \sum_{j=1}^n \lambda_j \nu_j (t_j - t_{j-1}) = \int fg \,\mathrm{d}m.$$

Therefore W is an isometry from the vector space of all step functions in $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$ into the Gaussian space spanned by $\{B_t, t \in \mathbb{R}_+\}$. Since the step functions are dense in $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$, we immediately extend W to an isometry between $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$ and $\operatorname{span}\{B_t, t \geq 0\}$. $\square$

Remark. According to our proof of (iv) $\Rightarrow$ (i), we can determine a Gaussian noise W with intensity m given a pre-Brownian motion $B = (B_t)_{t\geq 0}$. For all $f \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$, we write the notation

$$W(f) = \int_0^\infty f(s)\,\mathrm{d}B_s, \quad W(f\mathbb{1}_{[0,t]}) = \int_0^t f(s)\,\mathrm{d}B_s,$$

and

$$W(f\mathbb{1}_{(s,t]}) = \int_s^t f(r)\,\mathrm{d}B_r, \ \forall t > s \geq 0.$$

The mapping $W : f \mapsto \int_0^\infty f(s)\,\mathrm{d}B_s$ is called the *Wiener integral* with respect to $B = (B_t)_{t\geq 0}$. Clearly, we have $W(f) \sim N\left(0, \int_0^\infty |f|^2\,\mathrm{d}m\right)$.

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

$$B_t^{(s)} = B_{s+t} - B_s$$

is a pre-Brownian motion that is independent of $\sigma(B_r, r \in [0, s])$.

(iv) (Time inversion). The process $\widehat{B}$ defined by $\widehat{B}_0 = 0$ and

$$\widehat{B}_t = tB_{\frac{1}{t}}$$

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). $\square$

4.1.2 Sample Path Continuity and Brownian Motions

Before introducing Brownian motions, we first discuss continuity of sample paths of a stochastic process. For the convenience of a chaining argument, we consider the set of *dyadic rationals*

$$\mathbb{Q}_2 = \left\{ \frac{m}{2^n} : m, n \in \mathbb{N}_0 \right\},$$

which is a dense subset of $\mathbb{R}_+$.

Theorem 4.6 (Kolmogorov continuity lemma)**.** *Let $X = (X_t)_{t \in \mathbb{Q}_2 \cap I}$ be a process taking values in a metric space (E, d), where $I \subset \mathbb{R}_+$ is a compact interval. Assume there exist $\epsilon, q, C \in (0, \infty)$ such that*

$$\mathbb{E}\left[d(X_s, X_t)^q\right] \leq C|t-s|^{1+\epsilon}, \quad \forall s, t \in \mathbb{Q}_2 \cap I. \tag{4.1}$$

Then for each $\alpha \in (0, \frac{\epsilon}{q})$, there exists a random variable M_α with $\mathbb{P}(M_\alpha < \infty) = 1$ such that

$$d(X_s, X_t) \leq M_\alpha |t-s|^\alpha \quad \text{for all} \quad t, s \in \mathbb{Q}_2 \cap I.$$

Proof. Without loss of generality, we take $I = [0,1]$ and fix $\alpha \in (0, \frac{\epsilon}{q})$. Let

$$G_n = \left\{ d\left(X_{i2^{-n}}, X_{(i-1)2^{-n}}\right) \leq 2^{-\alpha n} \text{ for all } 1 \leq i \leq 2^n \right\}.$$

We apply Markov's inequality and a union bound to obtain

$$\begin{aligned} \mathbb{P}(G_n^c) &\leq \sum_{i=1}^{2^n} \mathbb{P}\left(d\left(X_{i2^{-n}}, X_{(i-1)2^{-n}}\right) \leq 2^{-\alpha n}\right) \\ &\leq 2^{-\alpha q n} \sum_{i=1}^{2^n} \mathbb{E}\left[d\left(X_{i2^{-n}}, X_{(i-1)2^{-n}}\right)^q\right] \leq C 2^{-n(\epsilon - \alpha q)}, \end{aligned} \tag{4.2}$$

where the last inequality follows from (4.1). Now we introduce a useful chaining argument.

Lemma 4.7 (Chaining)**.** *On the event $H_N = \bigcap_{n=N}^\infty G_N$, for all $s, t \in \mathbb{Q}_2 \cap [0,1]$ with $|s-t| < 2^{-N}$,*

$$d(X_s, X_t) \leq \frac{3}{1 - 2^{-\alpha}} |t-s|^\alpha.$$

Proof of the lemma. We fix $t > s > 0$ with $s, t \in \mathbb{Q}_2 \cap [0,1]$. We take $m \geq N$ and $1 \leq j \leq 2^m$ such that $s \leq (j-1)2^{-m} < j2^{-m} \leq t$. Then we can write s and t as binary expansions

$$s = (j-1)2^{-m} - \sum_{i=1}^k \delta_i 2^{-m-i}, \quad t = j2^{-m} + \sum_{i=1}^l \delta_i' 2^{-m-i}, \quad \text{where } \delta_1, \cdots, \delta_k, \delta_1', \cdots, \delta_l' \in \{0,1\}.$$

We take the finite sequences $s_i \downarrow s_k = s$ and $t_i \uparrow t_l = t$ defined by partial sums. On the event H_N,

$$\begin{aligned} d(X_s, X_t) &\leq d(X_s, X_{(j-1)2^{-m}}) + d(X_{(j-1)2^{-m}}, X_{j2^{-m}}) + d(X_{j2^{-m}}, X_t) \\ &\leq \sum_{i=1}^k d\left(X_{s_{i-1}}, X_{s_i}\right) + d(X_{(j-1)2^{-m}}, X_{j2^{-m}}) + \sum_{i=1}^l d\left(X_{t_{i-1}, X_{t_i}}\right) \\ &\leq \sum_{i=1}^\infty 2^{-\alpha(m+i)} + 2^{-m\alpha} + \sum_{i=1}^\infty 2^{-\alpha(m+i)} \leq \frac{3 \cdot 2^{-m\alpha}}{1 - 2^{-\alpha}}. \end{aligned}$$

Since $|t-s| \geq 2^{-m}$, the result follows. □

Proof of Theorem 4.6 (Cont). Applying a union bound on (4.2), we have

$$\mathbb{P}(H_N^c) \le \sum_{n=N}^{\infty} \mathbb{P}(G_N^c) \le C \sum_{n=N}^{\infty} 2^{-n(\epsilon-\alpha q)} \lesssim 2^{-N(\epsilon-\alpha q)}.$$

Since $\epsilon - \alpha q > 0$, we have

$$\sum_{N=1}^{\infty} \mathbb{P}(H_N^c) < \infty.$$

Using the Borel-Cantelli lemma, we know that there exists a random N_α with $\mathbb{P}(N_\alpha < \infty) = 1$ such that H_{N_α} occurs. When $N_\alpha < \infty$, we have

$$d(X_s, X_t) \le C|t-s|^\alpha \quad \text{for all } s,t \in \mathbb{Q}_2 \cap [0,1] \text{ with } |t-s| \le 2^{-N_\alpha}.$$

We can extend this to all $s,t \in \mathbb{Q}_2 \cap [0,1]$ by a similar chaining argument similar to Lemma 4.7:

$$d(X_s, X_t) \le \frac{2^{N_\alpha}+2}{1-2^{-\alpha}}|t-s|^\alpha, \quad \text{for all } s,t \in \mathbb{Q}_2 \cap [0,1].$$

Taking $M_\alpha = \frac{2^{N_\alpha}+2}{1-2^{-\alpha}}$ concludes our proof. ☐

Corollary 4.8 (Kolmogorov's continuity lemma)**.** *Let $X = (X_t)_{t\ge 0}$ be a process taking values in a* ***complete*** *metric space (E,d). Assume there exist $\epsilon, q, C \in (0,\infty)$ such that*

$$\mathbb{E}\left[d(X_s, X_t)^q\right] \le C|t-s|^{1+\epsilon}, \quad \forall s,t \ge 0. \tag{4.3}$$

Then there exists an a.s. modification $\widetilde{X}$ of X that is locally α-Hölder continuous for each $\alpha \in (0, \frac{\epsilon}{q})$.

Proof. We first consider the process $(X_t)_{t\in I}$, where $I \subset \mathbb{R}$ is a compact interval. According to Theorem 4.6, the process $(X_t)_{t\in I}$ is a.s. Hölder continuous of exponent α on $\mathbb{Q}_2 \cap I$. By completeness of (E,d), we define

$$\widetilde{X}_t(\omega) = \begin{cases} \lim_{\mathbb{Q}_2 \cap I \ni s \to t} X_s(\omega) & \text{if } M_\alpha(\omega) < \infty \\ x_0 & \text{otherwise,} \end{cases}$$

where $x_0 \in E$ is an arbitrary fixed point. Then $(\widetilde{X}_t)_{t\in I}$ has Hölder continuous sample paths of exponent α.

Next, we need to show that the process $\widetilde{X}$ is an a.s. modification of X. We fix $t \in I$, and take a dyadic sequence $(t_n) \subset \mathbb{Q}_2 \cap I$ converging to t. The assumption (4.3) and Markov's inequality imply $X_{t_n} \xrightarrow{\mathbb{P}} X_t$, and we also have $X_{t_n} \to \widetilde{X}_t$ a.s. by definition of $\widetilde{X}$. Hence $X_t = \widetilde{X}_t$ a.s..

Finally, we apply our conclusion repeatedly on $I_n = [0,n]$ for $n \in \mathbb{N}$. Then $(X_t)_{t\ge 0}$ has an a.s. modification $(\widetilde{X}_t)_{t\ge 0}$ whose sample paths are locally Hölder continuous of exponent α for all $\alpha \in (0, \frac{\epsilon}{q})$. ☐

Corollary 4.9. *Let $B = (B_t)_{t\ge 0}$ 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 $s,t \ge 0$, we have

$$\mathbb{E}|B_t - B_s|^{2+\delta} = |t-s|^{1+\frac{\delta}{2}}\mathbb{E}|Z|^{2+\delta}, \quad \textit{where } Z \in N(0,1).$$

By the last corollary, process B has an a.s. modification $\widetilde{B}$ 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 $\frac{1}{2}$. ☐

This Corollary justifies the existence of a Brownian motion, which is specified by the following definition.

Definition 4.10 (Brownian motion/Wiener process)**.** If $B = (B_t)_{t\geq 0}$ is a pre-Brownian motion and B is continuous, then process B is said to be a *(standard) Brownian motion/Wiener process.* More generally, if $B - B_0$ is a continuous pre-Brownian motion and B_0 is independent of the process $B - B_0$, then B is also called a *Brownian process.*

Remark. If B is a standard Brownian motion starting at $B_0 = 0$ and Z is a random variable independent of B, we can obtain a Brownian motion $B + Z$ starting from Z.

Proposition 4.11. *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion starting from $B_0 = 0$. The following statements are true:*

(i) (Symmetry). $-B$ is also a Brownian motion.

(ii) (Scale invariance). For all $\lambda > 0$, the process $B_t^\lambda = \frac{1}{\lambda}B_{\lambda^2 t}$ is a Brownian motion.

(iii) (Simple Markov property). For each $s \geq 0$, the process $(B_{s+t})_{t\geq 0}$ is a Brownian motion. Furthermore, $(B_{s+t} - B_s)_{t\geq 0}$ is a Brownian motion starting from 0 and independent of $\mathscr{F}_s = \sigma(B_r, r \leq s)$.

(iv) (Time inversion). The process $\widehat{B}$ defined by $\widehat{B}_0 = 0$ and $\widehat{B}_t = tB_{\frac{1}{t}}$ is a Brownian motion.

Proof. The Proposition mostly follows from Proposition 4.5 and the continuity of transforms applied. The only unclear thing is the continuity of $\widehat{B}$ at point 0 in (iv). We need to show that $\lim_{t\uparrow\infty} \frac{1}{t}B_t = 0$.

If $t \in \mathbb{N}$, the conclusion is clear by the Strong Law of Large Numbers [Theorem 3.48]. For the general case, we need the following lemma.

Lemma 4.12 (Kolmogorov's maximal inequality)**.** *Let $(X_n)_{n=1}^\infty$ be an independent sequence of random variables with $\mathbb{E}X_n = 0$ and $\mathbb{E}[X_n^2] < \infty$ for all $n \in \mathbb{N}$. The partial sum sequence $S_n = \sum_{j=1}^n X_j$ satisfies*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{\mathbb{E}[S_n^2]}{\lambda^2}, \quad \lambda > 0.$$

Proof of the lemma. By definition, the sequence $(S_n)_{n=1}^\infty$ is a martingale sequence. We define the stopping time $\tau = \min\{m \in \mathbb{N} : |S_m| \geq \lambda\}$. Then

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) &= \mathbb{P}\left(|S_{n\wedge\tau}| \geq \lambda\right) \leq \frac{1}{\lambda^2}\mathbb{E}|S_{n\wedge\tau}|^2 \\ &= \frac{1}{\lambda^2}\sum_{m=1}^{n\wedge\tau} \mathbb{E}|X_m|^2 \leq \frac{1}{\lambda^2}\sum_{m=1}^{n} \mathbb{E}|X_m|^2 = \frac{\mathbb{E}[S_n^2]}{\lambda^2}.\end{aligned}$$ ☐

Proof continued. For any $m, n \in \mathbb{N}$, we apply Kolmogorov's maximal inequality for $\lambda = n^{-2/3}$ to obtain

$$\mathbb{P}\left(\sup_{0\leq k\leq 2^m} |B_{n+k2^{-m}} - B_n| \geq n^{2/3}\right) \leq n^{-4/3}\mathbb{E}|B_{n+1} - B_n|^2 = n^{-4/3}.$$

We then take $m \uparrow \infty$ and apply Borel-Cantelli Lemma to conclude

$$\mathbb{P}\left(\sup_{t\in[n,n+1]} |B_t - B_n| \geq n^{2/3} \text{ for infinitely many } n \in \mathbb{N}\right) = 0.$$

For any $\epsilon > 0$, almost surely, we can find $N > \frac{1}{\epsilon^3}$ such that $\frac{|B_n|}{n} < \epsilon$ and $\sup_{s\in[n,n+1]} |B_s - B_n| < n^{2/3}$ for all $n \geq N$. Consequently, for all $t \geq N$,

$$\frac{|B_t|}{t} \leq \frac{|B_{\lfloor t\rfloor}|}{t} + \frac{|B_{\lfloor t\rfloor} - B_t|}{t} \leq \frac{|B_{\lfloor t\rfloor}|}{\lfloor t\rfloor} + \frac{\lfloor t\rfloor^{2/3}}{t} < \epsilon + t^{-1/3} \leq 2\epsilon.$$

Hence $\frac{B_t}{t} \to 0$ as $t \uparrow \infty$, and $\widehat{B}$ is continuous at 0. ☐

4.2 Canonical Construction and Wiener's Construction

4.2.1 Kolmogorov Extension Theorem

Definition 4.13. Let $\{(\Omega_\alpha, \mathscr{F}_\alpha), \alpha \in J\}$ be a collection of measurable spaces. We define the collection of all *measurable rectangles* by

$$\prod_{\alpha\in J} \mathscr{F}_\alpha = \left\{ \prod_{\alpha\in J} A_\alpha : A_\alpha \in \mathscr{F}_\alpha \text{ for all } \alpha \in J, \text{and } A_\alpha = \Omega_\alpha \text{ except for finitely many } \alpha \in J \right\}.$$

Akin to the proof (i) of Theorem 1.25, we can prove that $\prod_{\alpha\in J}\mathscr{F}_\alpha$ is a semi-ring. Similar to the definition of product of two measurable spaces, we define the *product σ-algebra*:

$$\bigotimes_{\alpha\in J} \mathscr{F}_\alpha = \sigma\left(\prod_{\alpha\in J}\mathscr{F}_\alpha\right)$$

The measurable space $\left(\bigcup_{\alpha\in J}\Omega_\alpha, \bigotimes_{\alpha\in J}\mathscr{F}_\alpha\right)$ is said to be the *product space* of $\{(\Omega_\alpha, \mathscr{F}_\alpha), \alpha \in J\}$.

Remark. Every coordinate mapping $\pi_\beta : (\omega_\alpha)_{\alpha\in J} \mapsto \omega_\beta$ is measurable when defined on $\{(\Omega_\alpha, \mathscr{F}_\alpha), \alpha \in J\}$. 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 4.14. *Let $\mathcal{I}_F$ be the collection of all finite subsets of J. For $I \in \mathcal{I}_F$, define $\bigotimes_{\alpha\in I}^{J}\mathscr{F}_\alpha$ 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*

$$\bigotimes_{\alpha\in J}\mathscr{F}_\alpha = \sigma\left(\bigcup_{I\in\mathcal{I}_F}\bigotimes_{\alpha\in I}^{J}\mathscr{F}_\alpha\right)$$

Proof. Clearly $\bigotimes_{\alpha\in I}^{J}\mathscr{F}_\alpha \subset \bigotimes_{\alpha\in J}\mathscr{F}_\alpha$ for all $I\in\mathcal{I}_F$. On the other hand, every element of $\prod_{\alpha\in J}\mathscr{F}_\alpha$ is contained by some $\bigotimes_{\alpha\in I}^{J}\mathscr{F}_\alpha$. □

Proposition 4.15. *Let $\{\mathbb{P}_I, I\in\mathcal{I}_F\}$ be a collection of probability measures defined on finite product spaces $\left\{\left(\prod_{\alpha\in I}\Omega_\alpha, \bigotimes_{\alpha\in I}\mathscr{F}_\alpha\right), I\in\mathcal{I}_F\right\}$. The following compatibility condition is necessary and sufficient for the existence of a* ***finitely additive*** *probability measure $\mathbb{P}$ on $\bigotimes_{\alpha\in J}\mathscr{F}_\alpha$ 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 $(\pi_{I_2\to I_1})_*\mathbb{P}_{I_2} = \mathbb{P}_{I_1}$.

Proof. We only prove sufficiency, since necessity is clear. For every cylinder $A = B\times\prod_{\alpha\in I^c}\Omega_\alpha$ such that $B\in\bigotimes_{\alpha\in I}\mathscr{F}_\alpha$, define $\mathbb{P}(A) = \mathbb{P}_I(B)$. By compatibility condition, we obtain a finitely additive function $\mathbb{P}$ on $\mathscr{A} = \bigcup_{I\in\mathcal{I}_F}\bigotimes_{\alpha\in I}^{J}\mathscr{F}_\alpha$, which is an algebra. We extend $\mathbb{P}$ to $\bigotimes_{\alpha\in J}\mathscr{F}_\alpha = \sigma(\mathscr{A})$. For $A\in\bigotimes_{\alpha\in J}\mathscr{F}_\alpha$, define

$$\mathbb{P}(A) = \sup\{\mathbb{P}(F) : F\subset A,\ F\in\mathscr{A}\}\in[0,1].$$

Then for any collection of disjoint sets $A_1,\cdots,A_n\in\bigotimes_{\alpha\in J}\mathscr{F}_\alpha$, we have

$$\begin{aligned}\sum_{j=1}^{n}\mathbb{P}(A) &= \sum_{j=1}^{n}\sup\{\mathbb{P}(F_j) : F_j\subset A_j,\ F\in\mathscr{A}\}\\ &= \sup\left\{\mathbb{P}\left(\bigcup_{j=1}^{n}F_j\right) : F_j\subset A_j, F_j\in\mathscr{A},\ \forall j\in\{1,\cdots,n\}\right\} = \mathbb{P}\left(\bigcup_{j=1}^{n}A_j\right).\end{aligned}$$

Thus we complete the proof of finite additivity of $\mathbb{P}$ on $\bigotimes_{\alpha\in J}\mathscr{F}_\alpha$. □

Proposition 4.16 (Compact class). *A class $\mathscr{C}$ of subsets of Ω is said to be* ***compact****, if every countable subclasses of $\mathscr{C}$ with* ***finite intersection property*** *has nonempty intersection. That is, for all sequences $(C_n)_{n=1}^\infty \subset \mathscr{C}$ 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 $\mathscr{C}$ is a compact class, so are the following: (i) The class $\mathscr{C}_\delta$ containing all countable intersections of elements of $\mathscr{C}$; (ii) The class $\mathscr{C}_s$ containing all finite unions of elements of $\mathscr{C}$.

Proof. (i) Since every countable intersection of elements of $\mathscr{C}_\delta$ is also a countable intersection of elements of the compact class $\mathscr{C}$, the result follows immediately.

(ii) We take a sequence $D_n = \bigcup_{j=1}^{m_n} C_j^n \in \mathscr{C}_s$ 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 $I_n = \prod_{k=1}^n \{1, \cdots, m_k\}$. 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

$$J_n = \left\{ \alpha \in \prod_{k=1}^\infty \{1, \cdots, m_k\} : \bigcap_{k=1}^n C_{\alpha_k}^k \neq \emptyset \right\}, \quad \forall n \in \mathbb{N}.$$

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 $\alpha_k^* \in \{1, \cdots, m_k\}$ 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 $\mathscr{C}$ we have $\bigcap_{k=1}^\infty D_k \supset \bigcap_{k=1}^\infty C_{\alpha_k^*}^k \neq \emptyset$, completing the proof. ☐

Remark. This definition is also in accordance with compactness in topology. If X is a topological space, and $\mathscr{K}$ is the collection of all compact subspaces of X. If $(K_n)_{n=1}^\infty \subset \mathscr{K}$ 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 $(K_1 \backslash L_n)_{n=1}^\infty$ 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 4.17 (Daniell-Kolmogorov extension). *Let $\{\mathbb{P}_I, I \in \mathcal{I}_F\}$ be a collection of probability measures defined on finite product spaces $\left\{\left(\prod_{\alpha\in I} \Omega_\alpha, \bigotimes_{\alpha \in I} \mathscr{F}_\alpha\right), I \in \mathcal{I}_F\right\}$ that satisfies the compatibility condition in Proposition 4.15. If for each $\alpha \in J$, there exists a compact class $\mathscr{C}_\alpha \subset \mathscr{F}_\alpha$ such that*

$$\mathbb{P}_\alpha(A) = \sup\left\{\mathbb{P}_\alpha(C) : C \in \mathscr{C}_\alpha, C \subset A\right\}, \quad \forall A \in \mathscr{F}_\alpha.$$

Then there exists a unique probability measure $\mathbb{P}$ on $\left(\prod_{\alpha\in J} \Omega_\alpha, \bigotimes_{\alpha\in J} \mathscr{F}_\alpha\right)$ that extends each $\mathbb{P}_I$.

Proof. Step I: Let $\mathbb{P}$ be the finitely additive set function found in Proposition 4.15. We first prove that there exists a compact subclass $\mathscr{C}$ of the semiring $\mathscr{S} = \prod_{\alpha\in J} \mathscr{F}_\alpha$ of all measurable rectangles such that

$$\mathbb{P}(A) = \sup\left\{\mathbb{P}(C) : C \in \mathscr{C}, C \subset A\right\}, \quad \forall A \in \mathscr{S}. \tag{4.1}$$

Let $\mathscr{D} = \{C \times \prod_{\alpha\neq\beta} \Omega_\alpha : \beta \in J,\ C \in \mathscr{C}_\beta\}$. Then every countable intersection $D = \bigcap_{n\in\mathbb{N}} (C_n \times \prod_{\alpha\neq\beta_n} \Omega_\alpha)$ of elements of $\mathscr{D}$ 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 $\mathscr{C}_\beta$, 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 $\mathscr{D}$ is a compact subclass of $\mathscr{S}$. Again, the class $\mathscr{C}$ of all finite intersections of elements of $\mathscr{D}$ 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 $\mathscr{C}_{t_j} \ni C_j \subset A_{\alpha_j}$ such that $\mathbb{P}_{\alpha_j}(C_j) \geq \mathbb{P}_{\alpha_j}(A_{t_j}) - \frac{\epsilon}{n}$. Then $\mathscr{C} \ni C = \bigcap_{j=1}^n (C_j \times \prod_{\alpha \neq \alpha_j} \Omega_\alpha) \subset A$, and

$$\mathbb{P}(A \backslash C) = \mathbb{P}\left(\bigcup_{j=1}^n \left((A_{\alpha_j} \backslash C_j) \times \prod_{\alpha \neq \alpha_j} \Omega_\alpha\right)\right) \leq \sum_{j=1}^n \mathbb{P}_{\alpha_j}(A_{\alpha_j} \backslash C_j) = \epsilon \downarrow 0.$$

Step II: We prove the σ-additivity of $\mathbb{P}$ on $\mathscr{S}$. We take the class $\mathscr{C}_s$ consisting of all finite intersection of elements of $\mathscr{C}$, which is again a compact class by Proposition 4.16 (ii) and is contained in the ring $\mathscr{R}$ generated by $\mathscr{S}$ according to the Remark under Definition 1.24. Similar to the proof of (4.1) in Step I, we can prove that $\mathbb{P}(A) = \sup\{\mathbb{P}(C) : C \in \mathscr{C}_s, C \subset A\}$ for all $A = \coprod_{k=1}^n A_k \in \mathscr{R}$ by taking $C = \coprod_{k=1}^n C_k \in \mathscr{C}_s$ with $\mathbb{P}(A_k \backslash C_k) \leq \epsilon/n$ for arbitrarily small ϵ. We prove that $\mathbb{P}$ is σ-additive on $\mathscr{R}$, hence on $\mathscr{S}$.

Given $\epsilon > 0$, we take a sequence $\mathscr{R} \ni A_n \downarrow \emptyset$, and take $\mathscr{C}_s \ni C_n \subset A_n$ 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 $\mathscr{C}_s$. As a result,

$$\mathbb{P}(A_N) = \mathbb{P}\left(A_N \backslash \left(\bigcap_{n=1}^N C_n\right)\right) = \mathbb{P}\left(\bigcup_{n=1}^N (A_N \backslash C_n)\right) \leq \sum_{n=1}^N \mathbb{P}(A_n \backslash C_n) \leq \epsilon \downarrow 0.$$

Therefore $\mathbb{P}$ is continuous at $\emptyset$. If $(B_n)_{n=1}^\infty \subset \mathscr{R}$ 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

$$\mathbb{P}\left(\bigcup_{n=1}^\infty B_n\right) - \sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(A_N) \to 0 \quad \textit{as } N \to \infty.$$

Step III: According to Step II, $\mathbb{P}$ is a finite pre-measure on the semiring $\mathscr{S}$ which generates $\bigotimes_{\alpha \in J} \mathscr{F}_\alpha$. By Carathéodory's extension theorem, $\mathbb{P}$ can be uniquely extended to a probability measure $\bigotimes_{\alpha \in J} \mathscr{F}_\alpha$. On the other hand, the finite additive function $\mathbb{P}$ is uniquely defined on $\mathscr{S}$, 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. ☐

Proposition 4.18. *Let Ω be a Hausdorff topological space, and equip Ω with the Borel σ-algebra $\mathscr{B}$. Let $\mathscr{C}$ be the collection of all closed sets in Ω, and $\mathscr{K}$ the collection of all compact sets in Ω. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathscr{B})$. We define the collections $\mathscr{R}_c$ of closed regular sets and $\mathscr{R}_k$ of regular sets as*

$$\mathscr{R}_c = \left\{B \in \mathscr{B} : \mathbb{P}(B) = \sup_{C \in \mathscr{C} : C \subset B} \mathbb{P}(C)\right\}, \quad \mathscr{R}_k = \left\{B \in \mathscr{B} : \mathbb{P}(B) = \sup_{K \in \mathscr{K} : K \subset B} \mathbb{P}(K)\right\}.$$

We say $\mathbb{P}$ is tight if $\Omega \in \mathscr{R}_k$. We say $\mathbb{P}$ is closed inner regular (resp. inner regular) if $\mathscr{R}_c = \mathscr{B}$ (resp. $\mathscr{R}_k = \mathscr{B}$).

- *(i) The collection $\mathscr{R}_c^* = \{B \in \mathscr{R}_c : \Omega \backslash B \in \mathscr{R}_c\}$ is a σ-algebra. In addition, if $\mathbb{P}$ is tight, then the collection $\mathscr{R}_k^* = \{B \in \mathscr{R}_k : \Omega \backslash B \in \mathscr{R}_k\}$ is also a σ-algebra.*
- *(ii) If Ω is metrizable, then $\mathbb{P}$ is closed inner regular. In addition, if $\mathbb{P}$ is tight, then it is inner regular.*
- *(iii) (Ulam). If Ω is a Polish space (a separable completely metrizable space), then $\mathbb{P}$ is inner regular.*

Proof. (i) Clearly, $\Omega \in \mathscr{R}_c^*$, and $B \in \mathscr{R}_c^*$ implies $\Omega \backslash B \in \mathscr{R}_c^*$. Given any sequence $(A_n)_{n=1}^\infty \subset \mathscr{R}_c^*$, we prove $A = \bigcup_{n=1}^\infty A_n \in \mathscr{R}_c^*$. 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 $D_n \subset \Omega \backslash A_n$ such that $\mathbb{P}(D_n) \geq \mathbb{P}(\Omega \backslash A_n) - \epsilon 2^{-n}$. Then there exists $N \in \mathbb{N}$ such that $\mathbb{P}(A) > \mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) - \epsilon/2$. Note that $\bigcup_{n=1}^N C_n \subset A$ is closed, and

$$\mathbb{P}(A) - \mathbb{P}\left(\bigcup_{n=1}^N C_n\right) < \frac{\epsilon}{2} + \mathbb{P}\left(\bigcup_{n=1}^N A_n\right) - \mathbb{P}\left(\bigcup_{n=1}^N C_n\right) \leq \frac{\epsilon}{2} + \mathbb{P}\left(\bigcup_{n=1}^N (A_n \backslash C_n)\right) < \epsilon.$$

Meanwhile, $\bigcap_{n=1}^{\infty} D_n \subset \Omega\backslash A$ is also closed, and

$$\mathbb{P}(\Omega\backslash A) - \mathbb{P}\left(\bigcap_{n=1}^{\infty} D_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} ((\Omega\backslash A)\backslash D_n)\right) \leq \sum_{n=1}^{\infty} \mathbb{P}((\Omega\backslash A_n)\backslash D_n) < \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $A \in \mathscr{R}_c^*$. Hence $\mathscr{R}_c^*$ is a σ-algebra.

If $\mathbb{P}$ is tight, we have $\Omega, \emptyset \in \mathscr{R}_k^*$, and $B \in \mathscr{R}_k^*$ implies $\Omega\backslash B \in \mathscr{R}_k^*$. Similar to the above proof, since finite unions and countable intersections of compact sets are still compact, we conclude $\mathscr{R}_k^*$ is a σ-algebra.

(ii) Let d be the metric of Ω. If $U \subset \Omega$ is an open set, take its complement $F = \Omega\backslash U$, and define $F_n = \{x \in \Omega : d(x, F) \geq 1/n\}$. Then $F_n \uparrow U$, and $U \in \mathscr{R}_c^*$ defined in (i). Since $\mathscr{R}_c^*$ is a σ-algebra containing all open sets in Ω, we have $\mathscr{R}_c^* = \mathscr{B}$, and $\mathbb{P}$ is closed inner regular.

In addition, if $\mathbb{P}$ is tight, we can take a compact set K such that $\mathbb{P}(\Omega\backslash K) < \epsilon/2$ for every $\epsilon > 0$. For any Borel set $B \in \mathscr{B}$, take closed set $F \subset B$ with $\mu(B\backslash F) < \epsilon/2$. Then $F \cap K \subset B$ is a compact set, and $\mathbb{P}(B\backslash(K \cap F)) \leq \mathbb{P}(B\backslash F) + \mathbb{P}(\Omega\backslash K) < \epsilon$. Since $\epsilon > 0$ is arbitrary, $B \in \mathscr{R}_k$. 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

$$\mathbb{P}\left(\Omega\backslash \bigcup_{n=1}^{N_m} B\left(\omega_n, \frac{1}{m}\right)\right) < \frac{\epsilon}{2^m}, \quad \forall m \in \mathbb{N}.$$

Let $K = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{N_k} B\left(\omega_n, \frac{1}{k}\right)$. Then K is both closed and totally bounded. Since Ω is a complete metric space, K is compact. Furthermore, we have

$$\mathbb{P}(\Omega\backslash K) \leq \sum_{m=1}^{\infty} \frac{\epsilon}{2^m} = \epsilon \downarrow 0.$$

Hence $\mathbb{P}$ is tight, and the result follows from (ii). $\square$

Theorem 4.17 and Proposition 4.18 together imply the following conclusion.

Corollary 4.19. *Let $\{(\Omega_\alpha, \mathscr{B}_\alpha), \alpha \in J\}$ 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 $\left\{\left(\prod_{\alpha\in I} \Omega_\alpha, \bigotimes_{\alpha\in I} \mathscr{B}_\alpha\right), I \in \mathcal{I}_F\right\}$, there exists a unique probability measure $\mathbb{P}$ on $\left(\prod_{\alpha\in J} \Omega_\alpha, \bigotimes_{\alpha\in J} \mathscr{B}_\alpha\right)$ that extends each $\mathbb{P}_I$.*

Remark. Let E be a metric space equipped with its Borel σ-algebra $\mathscr{B}$. The product space $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$ is called the *canonical space*. Since every evaluation map $\pi_t : (E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}}) \to (E, \mathscr{B}),\ x \mapsto x(t)$ is measurable, we can define a process $(\pi_t)_{t\in\mathscr{T}}$ on $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$, which is called the *canonical process*. Given a probability measure μ on $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$, 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 $(\Omega, \mathscr{F}, \mathbb{P})$ whose state space is the metric space E, the mapping $\Phi : (\Omega, \mathscr{F}) \to (E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$ 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 $\Phi^{-1}(A) = \bigcap_{j=1}^n X_{t_j}^{-1}(A_{t_j}) \in \mathscr{F}$. Since Φ^{-1} preserves complement and countable union operations, and $\mathscr{B}^{\otimes\mathcal{T}}$ 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 $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$. Furthermore, the canonical process $(\pi_t)_{t\in\mathcal{T}}$ defined on $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}}, \mu)$ 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 $(E^{\mathcal{T}}, \mathscr{B}^{\otimes\mathcal{T}})$.

4.2.2 Construction of Brownian Motions

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 $d(f_n, f) \to 0$. Then for each $t \in \mathbb{R}_+$, the coordinate mapping $\pi_t : C(\mathbb{R}_+) \to \mathbb{R},\ f \mapsto f(t)$ is continuous. Consequently, if we give $C(\mathbb{R}_+)$ the Borel σ-algebra $\mathscr{C}$ generated by c.c.topology, then all coordinate mappings $\{\pi_t,\ t \in \mathbb{R}_+\}$ are measurable.

Proposition 4.20. *$\mathscr{C}$ coincides the σ-algebra generated by coordinate maps $\pi_t : f \mapsto f(t)$.*

Proof. Let $\mathscr{B}_p$ be the smallest σ-algebra on $C(\mathbb{R}_+)$ for which the coordinate mappings $\pi_t : f \mapsto f(t)$ are measurable for all $t \in \mathbb{R}_+$. It is clear that $\mathscr{B}_p \subset \mathscr{C}$.

We know 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 base sets of the form

$$B_{[0,n]}(f_0, \epsilon) = \left\{ f \in C(\mathbb{R}_+) : \sup_{t\in[0,n]} |f(t) - f_0(t)| < \epsilon \right\},$$

where $f_0 \in C(\mathbb{R}_+)$ and $\epsilon > 0$. By continuity, every base set

$$B_{[0,n]}(f_0, \epsilon) = \bigcap_{t\in\mathbb{Q}\cap[0,n]} \pi_t^{-1} B(f_0(t), \epsilon) \in \mathscr{B}_p.$$

Hence every open set in $C(\mathbb{R}_+)$ in the c.c.topology is contained in $\mathscr{B}_p$. Therefore $\mathscr{B}_p$ coincides $\mathscr{C}$. $\square$

Wiener Measure. To construct a measure on the space of continuous functions on $\mathbb{R}_+$, we first consider the space $\mathbb{R}^{\mathbb{Q}_2}$ of functions on dyadic rationals $\mathbb{Q}_2$. For each measurable rectangle $A = \{f(t_i) \in A_i, i = 0, 1, \cdots, n\}$, where $0 = t_0 < t_1 < \cdots < t_n$ in $\mathbb{Q}_2$ and $A_0, A_1, \cdots, A_n \in \mathscr{B}(\mathbb{R})$, define

$$\nu(A) = \frac{\mathbb{1}_{A_0}(0)}{\sqrt{(2\pi)^n \prod_{j=1}^n (t_j - t_{j-1})}} \int_{A_1\times\cdots\times A_n} \exp\left\{ -\sum_{j=1}^n \frac{(z_j - z_{j-1})^2}{2(t_j - t_{j-1})} \right\} \mathrm{d}z_1 \cdots \mathrm{d}z_n, \quad \text{where } z_0 = 0.$$

This is a pre-measure on the semi-ring $\mathscr{S}$ of measurable rectangles, which extends uniquely to a measure ν on the product σ-algebra $\mathscr{B}(\mathbb{R})^{\mathbb{Q}_2}$, which is generated by $\mathscr{S}$.

We denote by $C(\mathbb{Q}_2)$ the set of all functions $f : \mathbb{Q}_2 \to \mathbb{R}$ that is uniformly continuous on $[0, T]$ for each $T > 0$, and consider the process $B_t(\omega) = \omega_t$, where $\omega \in \mathbb{R}^{\mathbb{Q}_2}$. Then for all $s, t \in \mathbb{Q}_2$, we have

$$\mathbb{E}_\nu |B_t - B_s|^4 = \frac{1}{\sqrt{2\pi|t-s|}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2|t-s|}}\, dx = 3|t-s|^2.$$

By Kolmogorov's continuity lemma, with probability 1, the process $(B_t)_{t\in\mathbb{Q}_2}$ is uniformly continuous on each compact interval $[0, T]$. Therefore ν is concentrated on the subset $C(\mathbb{Q}_2)$, i.e. $\nu(C(\mathbb{Q}_2)) = 1$.

On the restricted measure space $(C(\mathbb{Q}_2), \mathscr{C}(\mathbb{Q}_2), \nu)$, we define a mapping $\psi : C(\mathbb{Q}_2) \to C(\mathbb{R}_+)$ that sends each $f \in C(\mathbb{Q}_2)$ to its continuous extension on $\mathbb{R}_+$. Then ψ is a measurable mapping, and we define W to be the pushforward of ν, i.e. $W = \nu \circ \psi^{-1}$. This is called the *Wiener measure* on $(C(\mathbb{R}_+), \mathscr{C}(\mathbb{R}_+))$.

For any $0 = t_0 < t_1 < \cdots < t_n$, by dominated convergence theorem, we have

$$W(A) = \frac{\mathbb{1}_{A_0}(0)}{\sqrt{(2\pi)^n \prod_{j=1}^n (t_j - t_{j-1})}} \int_{A_1 \times \cdots \times A_n} \exp\left\{ -\sum_{j=1}^n \frac{(z_j - z_{j-1})^2}{2(t_j - t_{j-1})} \right\} \mathrm{d}z_1 \cdots \mathrm{d}z_n, \quad \text{where } z_0 = 0.$$

If we choose $(\Omega, \mathscr{F}, \mathbb{P}) = (C(\mathbb{R}_+), \mathscr{C}(\mathbb{R}_+), W)$, 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 the Wiener measure 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 4.19, it is possible to construct a collection of independent standard Gaussian variables $(Z_n)_{n=1}^\infty$ on an appropriate probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We define a process $(B_t)_{t\in[0,1]}$ as follows:

$$B_t = \sum_{n=1}^\infty \langle \mathbb{1}_{[0,t]}, e_n \rangle Z_n, \quad \forall t \in [0,1]. \tag{4.2}$$

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 $0 = t_0 < t_1 < \cdots < t_p = 1$, we have

$$\begin{aligned}
\mathbb{E}\left[(B_{t_j} - B_{t_{j-1}})(B_{t_k} - B_{t_{k-1}})\right] &= \sum_{m=1}^\infty \sum_{n=1}^\infty \langle \mathbb{1}_{(t_{j-1},t_j]}, e_m\rangle \langle \mathbb{1}_{(t_{k-1},t_k]}, e_n\rangle \underbrace{\langle e_m, e_n\rangle}_{=\mathbb{E}[Z_m Z_n]} \\
&= \left\langle \sum_{m=1}^\infty \langle \mathbb{1}_{(t_{j-1},t_j]}, e_m\rangle e_m, \sum_{n=1}^\infty \langle \mathbb{1}_{(t_{k-1},t_k]}, e_n\rangle e_n \right\rangle \\
&= \langle \mathbb{1}_{(t_{j-1},t_j]}, \mathbb{1}_{(t_{k-1},t_k]}\rangle = \delta_{jk}(t_j - t_{j-1}).
\end{aligned}$$

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 $S_m(t) = \sum_{k=m}^{2m-1} \frac{\sqrt{2}\sin(k\pi t)}{k} Z_k$, 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

$$\begin{aligned}
T_m^2 \leq \sqrt{2} \sup_{t\in[0,1]} \left| \sum_{k=m}^{2m-1} \frac{e^{\mathrm{i}k\pi t}}{k} Z_k \right|^2 &= \sqrt{2} \sup_{t\in[0,1]} \sum_{j=m}^{2m-1} \sum_{k=m}^{2m-1} \frac{e^{\mathrm{i}(k-j)\pi t}}{jk} Z_j Z_k \\
&\leq \sqrt{2} \sum_{k=m}^{2m-1} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sup_{t\in[0,1]} \left| \sum_{k=m}^{2m-1} \sum_{l=1}^{2m-k-1} e^{\mathrm{i}l\pi t} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \\
&= \sqrt{2} \sum_{k=m}^{2m-1} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sup_{t\in[0,1]} \left| \sum_{l=1}^{m-1} e^{\mathrm{i}l\pi t} \sum_{k=m}^{2m-l-1} \frac{Z_k Z_{k+l}}{k(k+l)} \right| \\
&\leq \sqrt{2} \sum_{k=m}^{2m-1} \frac{Z_k^2}{k^2} + 2\sqrt{2} \sum_{l=1}^{m-1} \left| \sum_{k=m}^{2m-l-1} \frac{Z_k Z_{k+l}}{k(k+l)} \right|
\end{aligned}$$

Let us bound the expectation of the second term:

$$\mathbb{E}\left[\left|\sum_{k=m}^{2m-l-1} \frac{Z_k Z_{k+l}}{k(k+l)}\right|\right] \leq \sqrt{\mathbb{E}\left[\left(\sum_{k=m}^{2m-l-1} \frac{Z_k Z_{k+l}}{k(k+l)}\right)^2\right]} = \left(\sum_{k=m}^{2m-l-1} \frac{1}{k^2(k+l)^2}\right)^{1/2} \leq \frac{1}{m^{3/2}},$$

$$\Rightarrow \quad \mathbb{E}\left[\sum_{l=1}^{m-1}\left|\sum_{k=m}^{2m-l-1} \frac{Z_k Z_{k+l}}{k(k+l)}\right|\right] \leq \frac{1}{\sqrt{m}}.$$

Since $\sum_{k=m}^{2m-1} k^{-2} \leq m^{-1}$, 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 $(\mathscr{F}_t)_{t\geq 0}$ of $(B_t)_{t\geq 0}$:

$$\mathscr{F}_t = \sigma\left(B_s, 0 \leq s \leq t\right),\ \forall t \geq 0.$$

Then $(B_t)_{t\geq 0}$ is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$.

4.3.1 Blumenthal's 0-1 Law, Recurrence and Time Inversion

Theorem 4.21 (Blumenthal's 0-1 law)**.** *The **germ σ-algebra** $\mathscr{F}_{0+} = \bigcap_{t>0} \mathscr{F}_t$ is $\mathbb{P}$-trivial.*

Proof. Let $0 < t_1 < \cdots < t_n$, and let g be a bounded continuous function on $\mathbb{R}^n$. According to continuity and dominated convergence theorem, for all $A \in \mathscr{F}_{0+}$, we have

$$\mathbb{E}\left[\mathbb{1}_A g(B_{t_1}, \cdots, B_{t_n})\right] = \lim_{\epsilon\downarrow 0} \mathbb{E}\left[\mathbb{1}_A g(B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\epsilon)\right].$$

By the simple Markov property of Brownian motions [Proposition 4.11 (iii)], $B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\epsilon$ is independent of $\mathscr{F}_\epsilon \supset \mathscr{F}_{0+}$ whenever $0 < \epsilon < t_1$. Hence

$$\mathbb{E}\left[\mathbb{1}_A g(B_{t_1}, \cdots, B_{t_n})\right] = \mathbb{P}(A) \lim_{\epsilon\downarrow 0} \mathbb{E}\left[g(B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\epsilon)\right] = \mathbb{P}(A)\, \mathbb{E}[g(B_{t_1}, \cdots, B_{t_n})].$$

For any open set $U \in \mathscr{B}(\mathbb{R}^n)$, 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

$$\mathbb{E}\left[\mathbb{1}_A \mathbb{1}_U(B_{t_1}, \cdots, B_{t_n})\right] = \mathbb{P}(A)\, \mathbb{E}[\mathbb{1}_U(B_{t_1}, \cdots, B_{t_n})].$$

Since $\mathscr{B}(\mathbb{R}^n)$ is generated by all open sets in $\mathbb{R}^n$, an argument of π-λ theorem implies that $\mathscr{F}_{0+}$ is independent of $\sigma(B_{t_1}, \cdots, B_{t_n})$. This holds for all finite marginals $0 < t_1 < \cdots < t_n$, hence $\mathscr{F}_{0+}$ is independent of $\sigma(B_t, t > 0)$. By right-continuity of $t \mapsto B_t(\omega)$, $B_0 = \lim_{t\to 0} B_t$ is measurable with respect to $\sigma(B_t, t > 0)$, and we have $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0) \supset \mathscr{F}_{0+}$. Therefore $\mathscr{F}_{0+}$ is independent of itself, and the result follows. $\square$

Proposition 4.22. *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$.*

(i) Almost surely, for each $\epsilon > 0$,

$$\sup_{0\leq s\leq\epsilon} B_s > 0 \quad \textit{and} \quad \inf_{0\leq s\leq\epsilon} B_s < 0.$$

(ii) (Recurrence). For every $\alpha \in \mathbb{R}$, define stopping time $\tau_\alpha = \inf\{t > 0 : B_t = \alpha\}$ (with respect to the canonical filtration, with the convention $\inf\emptyset = \infty$*). Then we have $\tau_\alpha < \infty$ a.s.. Consequently, it holds*

$$\limsup_{t\to\infty} B_t = \infty \quad \textit{and} \quad \liminf_{t\to\infty} B_t = -\infty \quad \textit{a.s.}.$$

Proof. (i) We choose a sequence $0 < \epsilon_n \downarrow 0$, and define the decreasing intersection

$$A = \bigcap_{n=1}^{\infty} \left\{\sup_{0\leq s\leq\epsilon_n} B_s > 0\right\}.$$

Then for each $t > 0$, there exists $\epsilon_n < t$, which implies

$$A = \bigcap_{k=n}^{\infty} \left\{\sup_{0\leq s\leq\epsilon_k} B_s > 0\right\} \in \mathscr{F}_t.$$

Hence $A \in \mathscr{F}_{0+}$. By Blumenthal's 0-1 law, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. We claim $\mathbb{P}(A) = 1$, so the result follows. To show $\mathbb{P}(A) = 1$, note that

$$\mathbb{P}\left(\sup_{0\leq s\leq \epsilon_n} B_s > 0\right) \geq \mathbb{P}\left(B_{\epsilon_n} > 0\right) = \frac{1}{2}, \quad \forall n \in \mathbb{N} \quad \Rightarrow \quad \mathbb{P}(A) \geq \frac{1}{2}.$$

(ii) By the simple Markov property of Brownian motions [Proposition 4.11 (ii)], for each $\lambda > 0$, the process $B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$ is also a Brownian motion. Since all Brownian motions starting at 0 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

$$\begin{aligned} 1 = \mathbb{P}\left(\sup_{0\leq s\leq 1} B_s > 0\right) &= \mathbb{P}\left(\sup_{s\geq 0} B_s > 1\right) = \mathbb{P}\left(\sup_{s\geq 0} B_s^\alpha > 1\right) \\ &= \mathbb{P}\left(\sup_{s\geq 0} B_s > \alpha\right) \leq \mathbb{P}(\tau_\alpha < \infty), \quad \forall \alpha > 0. \end{aligned}$$

Therefore $\tau_\alpha < \infty$ a.s. for each $\alpha > 0$, which holds only if $\limsup_{t\uparrow\infty} B_t = \infty$ a.s.. Symmetric arguments of (i) and (ii) follow by replacing B_s by $-B_s$. $\square$

Remark. We fix a number $M > 0$. By the second statement, with probability 1, we can find $s_1 > 0$ such that $B_{s_1} > M$, then $s_2 > s_1$ such that $B_{s_2} < -M$, and then $s_3 > s_2$ such that $B_{s_3} > M$, etc. Following this procedure, we find a sequence $0 < s_1 < s_2 < \cdots$ where B crosses the interval $[-M, M]$ during each $[s_{n-1}, s_n]$. By continuity of $t \mapsto B_t$, there exists a sequence $0 < t_n \uparrow \infty$ such that $B_{t_n} = 0$ at each t_n. Therefore B returns to 0 infinitely often. In words, one-dimensional Brownian motions are recurrent.

We have a stronger recurrence statement ragarding the return time of Brownian motions.

Proposition 4.23. *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$, and let $\tau_0 = \inf\{t > 0 : B_t = 0\}$ be the (first) return time. Then $\tau_0 = 0$ a.s..*

Proof. We define $\tau_+ = \inf\{t > 0 : B_t > 0\}$, and $\tau_- = \inf\{t > 0 : B_t < 0\}$. By Proposition 4.22 (i), we have $\tau_+, \tau_- \in [0, \epsilon)$ for each $\epsilon > 0$. Hence $\tau_+ = \tau_- = 0$. Since B hits both $(0, \infty)$ and $(-\infty, 0)$ a.s. immediately, by continuity of the path $t \mapsto B_t$, we have $\tau_0 = 0$ a.s.. $\square$

Up to now we only discuss the behavior of Brownian motions near $t = 0$. By using a time inversion trick, we extend our result to get information about the behavior as $t \to \infty$.

Proposition 4.24. *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$. The tail σ-algebra*

$$\mathscr{T} = \bigcap_{0<t<\infty} \sigma(B_s, s \geq t)$$

is trivial, i.e. if $A \in \mathscr{T}$, then either $\mathbb{P}_x(A) = 0$ or $\mathbb{P}_x(A) = 1$.

Proof. $\mathscr{T}$ is exactly the same as the germ σ-algebra for the process $\widehat{B}_t = tB_{\frac{1}{t}}$, and the result follows from Proposition 4.11 (iv) and Blumenthal's 0-1 law. $\square$

Proposition 4.25 (Growth Rate). *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$. Then with probability 1,*

$$\limsup_{t\uparrow\infty} \frac{B_t}{\sqrt{t}} = \infty, \quad \textit{and} \quad \liminf_{t\uparrow\infty} \frac{B_t}{\sqrt{t}} = -\infty.$$

Proof. For each $M \in (0, \infty)$,

$$\mathbb{P}\left(\frac{B_n}{\sqrt{n}} \geq M \text{ i.o.}\right) = \mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{\frac{B_n}{\sqrt{n}} \geq M\right\}\right)$$
$$\geq \limsup_{n\to\infty} \mathbb{P}\left(\frac{B_n}{\sqrt{n}} \geq M\right) = \limsup_{n\to\infty} \mathbb{P}\left(B_1 \geq M\right) > 0.$$

By Proposition 4.24, the probability of this event is 1. Since $M > 0$ is arbitrary, the first result holds. The second result follows from symmetry. $\square$

4.3.2 Monotonicity, Continuity, Non-Differentiability and Law of the Iterated Logarithm

The following proposition follows immediately from Remark III under Definition 4.2, which is a property of Gaussian white noise.

Proposition 4.26. *Let $(B_t)_{t\geq 0}$ be a Brownian motion. Let $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ be a sequence of partitions of $[0, t]$ with the mesh $\max_{1\leq j\leq k_n}(t_j^n - t_{j-1}^n) \to 0$. Then*

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} (B_{t_j^n} - B_{t_{j-1}^n})^2 = t \quad \textit{in } L^2.$$

This statement implies a lot of properties of the sample path of Brownian motions.

Corollary 4.27. *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 nondegenerate interval $[a, b]$.

(ii) Almost surely, the sample path $t \mapsto B_t$ has infinite total variation on any nondegenerate 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 4.22, we have

$$\sup_{q\leq t\leq q+\epsilon} B_t > B_q, \quad \inf_{q\leq t\leq q+\epsilon} B_t < B_q, \ \forall \epsilon > 0, \ \forall q \in \mathbb{Q}, \ a.s..$$

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 4.26, we have

$$\left(\sup_{1\leq j\leq k_n} \left|B_{t_j^n} - B_{t_{j-1}^n}\right|\right) \sum_{j=1}^{k_n} \left|B_{t_j^n} - B_{t_{j-1}^n}\right| \geq \sum_{j=1}^{k_n} \left(B_{t_j^n} - B_{t_{j-1}^n}\right)^2 \xrightarrow{L^2} t.$$

As $n \to \infty$, we have $\sup_{1\leq j\leq k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \to 0$ by continuity, which implies $\sum_{j=1}^{k_n} |B_{t_j^n} - B_{t_{j-1}^n}| \to \infty$ *a.s.*. The result follows by taking intersection of all $0 < t \in \mathbb{Q}$. (iii) is a consequence of Theorem 3.69. $\square$

Proposition 4.28 (Non-differentiability)**.** *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$. Then*

$$\limsup_{t\downarrow 0} \frac{B_t}{\sqrt{t}} = \infty \quad \textit{and} \quad \liminf_{t\downarrow 0} \frac{B_t}{\sqrt{t}} = -\infty \quad \textit{a.s.}.$$

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,

$$\sup_{0\le s\le \epsilon} \frac{B_s}{\sqrt{s}} > \alpha, \quad \forall \epsilon > 0. \tag{4.3}$$

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

$$A = \bigcap_{n=1}^{\infty} \left\{ \sup_{0\le s\le \epsilon_n} \frac{B_s}{\sqrt{s}} > \alpha \right\} \in \mathscr{F}_{0+}.$$

According to Blumenthal's 0-1 law, we have either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Note that

$$\mathbb{P}\left(\sup_{0\le s\le \epsilon_n} \frac{B_s}{\sqrt{s}} > \alpha \right) \ge \mathbb{P}\left(\frac{B_{\epsilon_n}}{\sqrt{\epsilon_n}} > \alpha \right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \quad \Rightarrow \quad \mathbb{P}(A) \ge \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x > 0.$$

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 4.29 (Global non-differentiability)**.** *Let $(B_t)_{t\ge 0}$ be a Brownian motion. Almost surely, the sample path $t \mapsto B_t$ is not Hölder continuous with exponent γ at any point for each $\gamma > \frac{1}{2}$. In particular, the sample path $t \mapsto B_t$ is not Lipschitz continuous at any point, hence is nowhere differentiable.*

Proof. We fix $\gamma > \frac{1}{2}$, and take $k \in \mathbb{N}$ such that $\gamma > \frac{1}{2} + \frac{1}{k}$. Given $C > 0$, let

$$A_n = \left\{ \omega \in \Omega : \text{there exists } s \in [0,1] \text{ such that } |B_t - B_s| \le C|t-s|^{\gamma} \text{ for all } |t-s| < \frac{k}{n} \right\}.$$

For $1 \le m \le n-k+1$, define

$$Y_{m,n} = \max_{0\le j\le k-1} \left| B_{\frac{m+j}{n}} - B_{\frac{m+j-1}{n}} \right|,$$

and

$$E_n = \left\{ \text{there exists } 1 \le m \le n-k+1 \text{ such that } Y_{m,n} \le \frac{(2k-1)C}{n^{\gamma}} \right\}.$$

On the event A_n, we apply the triangle inequality to get

$$\left| B_{\frac{m+j}{n}} - B_{\frac{m+j-1}{n}} \right| \le \left| B_{\frac{k+j}{n}} - B_s \right| + \left| B_s - B_{\frac{k+j-1}{n}} \right| \le \frac{(2k-1)C}{n^{\gamma}}, \quad j = 0, 1, \cdots, k-1$$

for some $1 \le m \le n-k+1$. Therefore $A_n \subset E_n$, and

$$\begin{aligned}
\mathbb{P}(A_n) \le \mathbb{P}(E_n) &\le (n-k+1) \cdot \mathbb{P}\left(\left| B_{1/n} \right| \le \frac{(2k-1)C}{n^{\gamma}} \right)^k \\
&= (n-k+1) \cdot \mathbb{P}\left(|B_1| \le \frac{(2k-1)C}{n^{\gamma-\frac{1}{2}}} \right)^k \le (2k-1)^k C^k n^{\frac{k}{2}-k\gamma+1}.
\end{aligned}$$

Since $\gamma > \frac{1}{2} + \frac{1}{k}$, we have $\mathbb{P}(A_n) \to 0$ as $n \to \infty$. Noticing that (A_n) is a increasing sequence, we have $\mathbb{P}(A_n) = 0$ for all $n \in \mathbb{N}$. Also, since C is arbitrary chosen, with probability 1, the path of (B_t) is not Hölder continuous with exponent γ at any point in $[0,1]$. $\square$

Theorem 4.30 (Law of the iterated logarithm). *Let $(B_t)_{t\geq 0}$ be a Brownian motion with $B_0 = 0$. Then*

$$\limsup_{t\uparrow\infty} \frac{B_t}{\sqrt{2t\log\log t}} = 1 \quad a.s..$$

Proof. We first give a tail bound of $B_1 \sim N(0,1)$, which is a standard Gaussian variable:

$$\frac{\alpha e^{-\alpha^2/2}}{(1+\alpha^2)\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty \frac{e^{-t^2/2}}{(1+t^2)^2}\,\mathrm{d}t \leq \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-t^2/2}\,\mathrm{d}t \leq \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty \frac{te^{-t^2/2}}{\alpha}\,\mathrm{d}t = \frac{e^{-\alpha^2/2}}{\alpha\sqrt{2\pi}}, \quad \forall \alpha > 0.$$

Let $h(t) = \sqrt{2t\log\log t}$. Using the law of $M_t := \sup_{0\leq s\leq t} B_s \overset{d}{=} |B_t|$ given in Corollary 4.36, for all $r > 1$ and all $\delta > 0$, whenever $n \geq n_0 := 1 + \frac{\mathrm{e}}{\log r}$, 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

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{B_s}{h(s)} \leq \sqrt{r+\delta}\right) = \mathbb{P}\left(\sup_{s\in[r^{n-1},r^n]} \frac{B_s}{h(s)} > \sqrt{r+\delta} \quad \text{for finitely many } n\right) = 1.$$

Let $\delta \downarrow 0$ and $r \downarrow 1$, we have $\mathbb{P}\left(\limsup_{n\to\infty} \frac{B_s}{h(s)} \leq 1\right) = 1$.

Now we prove $\limsup_{n\to\infty} \frac{B_s}{h(s)} \geq 1$ *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 second Borel-Cantelli lemma [Corollary 3.29] implies that a.s. $B_{r^n} - B_{r^{n-1}} \geq \sqrt{\frac{r-1}{r}}h(r^n)$ for infinitely many n. A symmetric argument of the first part of our proof implies $\liminf_{n\to\infty} \frac{B_s}{h(s)} \geq -1$ *a.s.*. Hence for *a.s.* $\omega \in \Omega$, we can find some $N(\omega)$ such that for all $n \geq N(\omega)$,

$$B_{r^{n-1}} \geq -2h(r^{n-1}) \geq -\frac{2}{\sqrt{r}}h(r^n).$$

Hence $B_{r^n} \geq \left(1 - \frac{3}{\sqrt{r}}\right)h(r^n)$ occurs infinitely often. Consequently, for all $r > 1$, we have

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{B_t}{h(t)} \geq 1 - \frac{3}{\sqrt{r}}\right) = 1.$$

Letting $r \to \infty$ suffices. $\square$

4.4 Strong Markov Property and Applications

In this section we discuss the strong Markov property of a Brownian motion $(B_t)_{t\geq 0}$ starting from $B_0 = 0$. Let τ be a stopping time (with respect to the canonical filtration $(\mathscr{F}_t)_{t\geq 0}$ of $(B_t)_{t\geq 0}$). We define

$$\mathbb{1}_{\{\tau<\infty\}}B_\tau(\omega) = \mathbb{1}_{\{\tau(\omega)<\infty\}}B_{\tau(\omega)}(\omega).$$

This is a $\mathscr{F}_\tau$-measurable variable. To see this, note that $(B_t)_{t\geq 0}$ is an adaptive and continuous process, hence is progressive [Proposition 3.10]. Then the desired result follows from Proposition 3.13.

4.4.1 Strong Markov Property

Theorem 4.31 (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 $\mathscr{F}_\tau$.

Proof. We first deal with the case $\tau < \infty$ *a.s.*. Fix $A \in \mathscr{F}_\tau$ and $0 = t_0 < t_1 < \cdots < t_n$. We claim that

$$\mathbb{E}\big[\mathbb{1}_A g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \cdots, B_{t_n}^{(\tau)})\big] = \mathbb{P}(A)\,\mathbb{E}\,[g(B_{t_0}, B_{t_1}, \cdots, B_{t_n})]$$

for all bounded continuous functions $g : \mathbb{R}^n \to \mathbb{R}$. If we take $A = \Omega$, a similar argument to the proof of Theorem 4.21 implies that $(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \cdots, B_{t_n}^{(\tau)}) \overset{d}{=} (B_{t_0}, B_{t_1}, \cdots, B_{t_n})$ for all choices $0 = t_0 < t_1 < \cdots < t_n$. 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, $(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \cdots, B_{t_n}^{(\tau)})$ is independent of $\mathscr{F}_\tau$, and $(B_t^{(\tau)})_{t\geq 0}$ is independent of $\mathscr{F}_\tau$.

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 $B_t^{(\tau_p)} \to B_t^{(\tau)}$ *a.s.*, and dominated convergence implies

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_A g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \cdots, B_{t_n}^{(\tau)})\right] = \lim_{p\to\infty} \mathbb{E}\left[\mathbb{1}_A g(B_{t_0}^{(\tau_p)}, B_{t_1}^{(\tau_p)}, \cdots, B_{t_n}^{(\tau_p)})\right] \\
&= \lim_{p\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{A\,\cap\,\{(k-1)2^{-p}<\tau\leq k2^{-p}\}} g\left(B_{k2^{-p}+t_0} - B_{k2^{-p}}, B_{k2^{-p}+t_1} - B_{k2^{-p}}, \cdots, B_{k2^{-p}+t_n} - B_{k2^{-p}}\right)\right] \\
&= \lim_{p\to\infty} \sum_{k=0}^{\infty} \mathbb{P}(A \cap \{(k-1)2^{-p} < \tau \leq k2^{-p}\})\mathbb{E}\left[g\left(B_{t_0}, B_{t_1}, \cdots, B_{t_n}\right)\right] = \mathbb{P}(A)\mathbb{E}\left[g\left(B_{t_0}, B_{t_1}, \cdots, B_{t_n}\right)\right],
\end{aligned}$$

where the last row follows from the fact that

$$A \cap \{(k-1)2^{-p} < \tau \leq k2^{-p}\} = \big(A \cap \{\tau \leq k2^{-p}\}\big) \cap \{\tau \leq (k-1)2^{-p}\}^c \in \mathscr{F}_{k2^{-p}}$$

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

$$\mathbb{E}\big[\mathbb{1}_{A\cap\{\tau<\infty\}} g(B_{t_0}^{(\tau)}, B_{t_1}^{(\tau)}, \cdots, B_{t_n}^{(\tau)})\big] = \mathbb{P}(A \cap \{\tau < \infty\})\,\mathbb{E}\,[g(B_{t_0}, B_{t_1}, \cdots, B_{t_n})]\,.$$

Then the desired result follows in a straightforward way. ☐

Remark. In Section 6.2.2, we will introduce a stronger statement. We show that, for any stopping time τ with respect to the filtration $(\mathscr{F}_{t+})_{\geq 0}$ and measurable function $\Phi_\cdot : \mathbb{R}_+ \times C(\mathbb{R}_+) \to \mathbb{R}_+$,

$$\mathbb{E}\,[\Phi_\tau((X_{\tau+t})_{t\geq 0})|\mathscr{F}_{\tau+}] = \mathbb{E}_{X_\tau}\Phi_\tau.$$

4.4.2 Zero Set

Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from $B_0 = 0$. We study the property of the zero set

$$\mathcal{Z} = \{t \geq 0 : B_t = 0\}.$$

Since $t \mapsto B_t$ is continuous, this random set is almost surely closed. By Proposition 4.22, we know that 0 is a limit point of $\mathcal{Z}$, and $\mathcal{Z}$ is unbounded. To be specific, we define $\tau_0 = \inf\{s > 0 : B_s = 0\}$ and $\upsilon_t = \inf\{s > t : B_s = 0\}$. Then $\tau_0 = 0$ a.s., and $\upsilon_t < \infty$ a.s. for each $t \in \mathbb{R}_+$.

Proposition 4.32. *With probability* 1*, the zero set* $\mathcal{Z}$ *has no isolated points.*

Proof. The recurrence of $(B_t)_{t\geq 0}$ implies $\mathbb{P}(\upsilon_t < \infty) = 1$, and

$$\mathbb{P}\left(\tau_0\left((B_{\upsilon_t+s})_{s\geq 0}\right) > 0 \,|\, \mathscr{F}_{\upsilon_t}\right) = \mathbb{P}_0\left(\tau_0 > 0\right) = 0.$$

We take expectation and then take a union bound to obtain

$$\mathbb{P}\left(\tau_0\left((B_{\upsilon_t+s})_{s\geq 0}\right) > 0 \text{ for some } t \in \mathbb{Q}\right) = 0.$$

If a point $u \in \mathcal{Z}(\omega)$ is isolated from the left, i.e. $u = \upsilon_t(\omega)$ for some rational t, the above result implies that u is a limit point from the right. Therefore $\mathcal{Z}$ has no isolated points almost surely. $\square$

Remark. In fact, the zero $\mathcal{Z}$ is an *uncountable* set. To see this, we note that $\mathcal{Z}(\omega)$ is a closed subset of $\mathbb{R}$, hence is a complete metric space. By Baire's category theorem, $\mathcal{Z}(\omega)$ is not a countable union of nowhere dense sets. If $\mathcal{Z}(\omega)$ is a countable set, then at least one singleton $\{u\} \subset \mathcal{Z}$ contains an open ball in $\mathcal{Z}(\omega)$. This implies that u is an isolated point, which contradicts Proposition 4.32.

Proposition 4.33. *With probability* 1*, the Lebesgue measure of* $\mathcal{Z}$ *is zero.*

Proof. For all $t \neq 0$, we have $\mathbb{P}(t \in \mathcal{Z}) = \mathbb{P}(B_t = 0) = 0$. Then

$$\mathbb{E}\left[m(\mathcal{Z})\right] = \mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{t\in\mathcal{Z}\}}\, dt\right] = \mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{B_t=0\}}\, dt\right] = \int_0^\infty \mathbb{P}(B_t = 0)\, dt = 0.$$

Hence $m(\mathcal{Z}) = 0$ almost surely. $\square$

4.4.3 Hitting Times and Reflection Principle

In this part, we study the hitting time

$$\tau_a = \inf\{s \geq 0 : B_s = a\}, \quad a \geq 0$$

of a Brownian motion $(B_t)_{t\geq 0}$ starting from $B_0 = 0$. Then we can view $(\tau_a)_{a\geq 0}$ as an increasing random process, which has jumps and is not continuous.

Theorem 4.34 (First passage process)**.** *The process* $(\tau_a)_{a\geq 0}$ *has stationary and independent increments.*

Proof. We fix $0 \leq a < b$. Then $\tau_b((B_{\tau_a+t})_{t\geq 0}) = \tau_b - \tau_a$. By Proposition 4.22, $\tau_a < \infty$ a.s. for all $a > 0$. For any bounded measurable function f, by the strong Markov property,

$$\mathbb{E}_0\left[f(\tau_b - \tau_a)|\mathscr{F}_{\tau_a}\right] = \mathbb{E}_0\left[f(\tau_b((B_{\tau_a+t})_{t\geq 0}))|\mathscr{F}_{\tau_a}\right] = \mathbb{E}_a[f(\tau_b)] = \mathbb{E}_0[f(\tau_{b-a})].$$

Hence $\tau_b - \tau_a \overset{d}{=} \tau_{b-a}$, which proves stationarity.

Now we let $0 = a_0 < a_1 < \cdots < a_n$ and let $f_1, \cdots, f_n$ be bounded measurable functions. Then

$$
\begin{aligned}
\mathbb{E}_0\left[\prod_{i=1}^{n} f_i(\tau_{a_i}-\tau_{a_{i-1}})\right] &= \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(\tau_{a_i}-\tau_{a_{i-1}})\mathbb{E}_0\left[f(\tau_{a_n}-\tau_{a_{n-1}})\middle|\mathscr{F}_{\tau_{a_{n-1}}}\right]\right] \\
&= \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(\tau_{a_i}-\tau_{a_{i-1}})\mathbb{E}_{a_{n-1}}\left[f(\tau_{a_n})\right]\right] = \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(\tau_{a_i}-\tau_{a_{i-1}})\right]\mathbb{E}_0\left[f(\tau_{a_n-a_{n-1}})\right] \\
&= \cdots = \prod_{i=1}^{n}\mathbb{E}_0\left[f(\tau_{a_i-a_{i-1}})\right] = \prod_{i=1}^{n}\mathbb{E}_0\left[f(\tau_{a_i}-\tau_{a_{i-1}})\right],
\end{aligned}
$$

where the last equality follows from the fact $\tau_{a_i} - \tau_{a_{i-1}} \overset{d}{=} \tau_{a_i - a_{i-1}}$. This proves independence. □

Theorem 4.35 (Reflection principle)**.** *Given $t > 0$, let $M_t = \sup_{0\le s\le t} B_s$. For any $a > 0$ and $b \in (-\infty, a]$,*

$$\mathbb{P}(M_t \ge a, B_t \le b) = \mathbb{P}(B_t \ge 2a - b). \tag{4.5}$$

In particular, we have $M_t \overset{d}{=} |B_t|$.

Proof. Consider the hitting time $\tau_a = \inf\{s > 0 : B_s = a\}$. By Proposition 4.22 (ii), we have $\tau_a < \infty$ a.s.. And by Theorem 4.31, the process $B_t^{(\tau_a)} = B_{\tau_a+t} - B_{\tau_a} = B_{\tau_a+t} - a$ is a Brownian process independent of $\mathscr{F}_{\tau_a}$. Consequently, $(\tau_a, B^{(\tau_a)}) \overset{d}{=} (\tau_a, -B^{(\tau_a)})$, whose distribution equals the product of the law of τ_a and the Wiener measure W on $C(\mathbb{R}_+)$. Let $H = \{(s, f) \in \mathbb{R} \times C(\mathbb{R}_+) : s \le t, f(t-s) \le b - a\}$. Hence

$$
\begin{aligned}
\mathbb{P}(M_t \ge a, B_t \le b) &= \mathbb{P}(\tau_a \le t, B_t \le b) = \mathbb{P}\left(\tau_a \le t, B_{t-\tau_a}^{(\tau_a)} \le b - a\right) = \mathbb{P}\left((\tau_a, B^{(\tau_a)}) \in H\right) \\
&= \mathbb{P}\left((\tau_a, -B^{(\tau_a)}) \in H\right) = \mathbb{P}(\tau_a \le t, -B_{t-\tau_a}^{(\tau_a)} \le b - a) \\
&= \mathbb{P}(\tau_a \le t, B_t \ge 2a - b) = \mathbb{P}(B_t \ge 2a - b).
\end{aligned}
$$

For the last assertion, note that

$$\mathbb{P}(M_t \ge a) = \mathbb{P}(M_t \ge a, B_t \ge a) + \mathbb{P}(M_t \ge a, B_t \le a) = \mathbb{P}(B_t \ge a) + \mathbb{P}(B_t \ge 2a - a) = 2\mathbb{P}(B_t \ge a).$$

Thus we complete the proof. □

Remark. According to (4.5), we also have

$$\mathbb{P}(M_t \le y, B_t \le x) = \begin{cases} \mathbb{P}(B_t \le x) - \mathbb{P}(B_t \ge 2y - x) & \textit{if } y > 0,\ x \le y \\ \mathbb{P}(M_t \le y) & \textit{if } y > 0,\ x > y. \end{cases}$$

This gives the density of (M_t, B_t):

$$\rho_{M_t,B_t}(y, x) = \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{2\pi t}}\left(e^{-\frac{x^2}{2t}} + e^{-\frac{(2y-x)^2}{2t}}\right)\right) = \frac{2(2y-x)}{t^{3/2}\sqrt{2\pi}}e^{-\frac{(2y-x)^2}{2t}}, \quad y > 0,\ x \le y.$$

Corollary 4.36 (Law of hitting times)**.** *Given $a > 0$, the hitting time τ_a is identically distributed to $a^2 B_1^{-2}$.*

Proof. By the last assertion of Theorem 4.35, we have

$$\mathbb{P}(\tau_a \le t) = \mathbb{P}(M_t \ge a) = \mathbb{P}(|B_t| \ge a) = \mathbb{P}(B_t^2 \ge a^2) = \mathbb{P}\left(\frac{a^2}{B_1^2} \le t\right),$$

which holds for all $t \ge 0$. □

Remark. Since $B_1 \sim N(0,1)$, we can derive the density of τ_a:

$$\rho_{\tau_a}(t) = \frac{2\mathbb{1}_{\{t>0\}}}{\sqrt{2\pi}} e^{-\frac{a^2}{2t}} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{a}{\sqrt{t}}\right) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}} \mathbb{1}_{\{t>0\}}.$$

This also implies $\mathbb{E}[\tau_a] = \infty$ for all $a > 0$.

To derive more property of the hitting times, we make use of the martingale property of Browninan motions. The following proposition is obtained by direct calculation.

Proposition 4.37. *Let $(\mathscr{F}_t)_{t\geq 0}$ be the canonical filtration of a Brownian motion $(B_t)_{t\geq 0}$. All these processes are martingales with respect to $(\mathscr{F}_t)_{t\geq 0}$: (i) B_t; (ii) $B_t^2 - t$; (iii) $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$, $\theta \in \mathbb{R}$.*

Proposition 4.38 (Laplacian transform of hitting times)**.** *Given $a > 0$, the hitting time τ_a satisfies*

$$\mathbb{E}\left[e^{-\lambda \tau_a}\right] = e^{-a\sqrt{2\lambda}}, \quad \forall \lambda > 0.$$

Proof. We consider the martingale $N_t^\theta = \exp(\theta B_t - \frac{1}{2}\theta^2 t)$, where $\theta > 0$. By Corollary 3.62, $N_{t\wedge\tau_a}^\theta$ is a martingale bounded by $e^{\theta a}$ from above, hence is uniformly integrable. As a result,

$$\mathbb{E}\left[N_{\tau_a}^\theta\right] = \lim_{t\to\infty} \mathbb{E}\left[N_{t\wedge\tau_a}^\theta\right] = \mathbb{E}[N_0^\theta] = 1.$$

Since $\tau_a < \infty$ *a.s.*, we have

$$1 = \mathbb{E}\left[N_{\tau_a}^\theta\right] = \mathbb{E}\left[\exp\left(\theta a - \frac{1}{2}\theta^2 \tau_a\right)\right].$$

Setting $\theta = \sqrt{2\lambda}$, the above reads $\mathbb{E}\left[e^{-\lambda\tau_a}\right] = e^{-a\sqrt{2\lambda}}$. ☐

To move forward, we alsop study the exit time of the Brownian motion from an interval.

Proposition 4.39 (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

$$\mathbb{P}(\tau_a < \tau_b) = \frac{b}{b-a}, \quad \textit{and} \quad \mathbb{P}(\tau_b < \tau_a) = \frac{-a}{b-a}.$$

(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

$$\mathbb{E}\left[e^{-\lambda\tau}\right] = \frac{\cosh\left(\frac{b+a}{2}\sqrt{2\lambda}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right)}.$$

Proof. (i) We define a stopping time $\tau = \tau_a \wedge \tau_b$. By Corollary 3.62, 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, $\mathbb{E}\left[B_\tau\right] = \mathbb{E}\left[B_0\right] = 0$. 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 $\mathbb{E}[t\wedge\tau] \to \mathbb{E}[\tau]$ as $t \to \infty$. On the other hand, since $B_{t\wedge\tau}^2 \leq a^2 \vee b^2$, we have $\mathbb{E}[B_{t\wedge\tau}^2] \to \mathbb{E}[B_\tau^2]$ as $t \to \infty$ by dominated convergence theorem. Note that

$$\mathbb{E}\left[B_\tau^2\right] = \mathbb{E}\left[a^2\mathbb{1}_{\{B_\tau = a\}} + b^2\mathbb{1}_{\{B_\tau = b\}}\right] = a^2\mathbb{P}(\tau_a < \tau_b) + b^2\mathbb{P}(\tau_b < \tau_a) = -ab.$$

(iii) Similar to Proposition 4.37 (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,

$$\mathbb{E}\left[N_\tau\right] = \lim_{n\to\infty}\left[N_{t\wedge\tau}\right] = \mathbb{E}[N_0] = \cosh\left(\frac{a+b}{2}\sqrt{2\lambda}\right).$$

On the other hand, since $B_\tau \in \{a, b\}$ *a.s.*, we have

$$\mathbb{E}[N_\tau] = \mathbb{E}\left[e^{-\lambda\tau}\left(\mathbb{1}_{\{B_\tau = a\}}\cosh\left(\frac{a-b}{2}\sqrt{2\lambda}\right) + \mathbb{1}_{\{B_\tau=b\}}\cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right)\right)\right] = \mathbb{E}\left[e^{-\lambda\tau}\right]\cosh\left(\frac{b-a}{2}\sqrt{2\lambda}\right).$$

Then the desired result follows. □

Finally, we discuss the time reversal of Brownian motions.

Proposition 4.40 (Time reversal). *Set $\widetilde{B}_t = B_1 - B_{1-t}$ for every $t \in [0,1]$. Then $(\widetilde{B}_t)_{t\in[0,1]}$ is a Brownian motion on $[0,1]$, which has the same law as $(B_t)_{t\in[0,1]}$.*

Proof. Clearly, $\widetilde{B}_0 = 0$, and $(\widetilde{B}_t)_{t\in[0,1]}$ has continuous sample paths. We show that $(B_t)_{t\in[0,1]}$ and $(\widetilde{B}_t)_{t\in[0,1]}$ has the same finite-dimensional marginal distributions, which are extended to the Wiener measure on $C([0,1])$. Take a partition $0 = t_0 < t_1 < \cdots < t_n = 1$. Then the increments

$$(\widetilde{B}_{t_1}, \widetilde{B}_{t_2} - \widetilde{B}_{t_1}, \cdots, \widetilde{B}_{t_n} - \widetilde{B}_{t_{n-1}}) = (B_1 - B_{1-t_1}, B_{1-t_1} - B_{1-t_2}, \cdots, B_{1-t_{n-1}})$$

are jointly Gaussian and independent, and the desired result follows. □

Corollary 4.41. *Let $M_t = \sup_{0\leq s\leq t} B_t$. Then $M_t - B_t \overset{d}{=} M_t \overset{d}{=} |B_t|$ for every $t > 0$.*

Proof. Fix $t > 0$. Akin to Proposition 4.40, we define $\widetilde{B}_s = B_t - B_{t-s}$ for every $s \in [0,t]$. By symmetry and time reversal property of Brownian motions, all $(B_s)_{s\in[0,t]}$, $(\widetilde{B}_s)_{s\in[0,t]}$ and $(-\widetilde{B}_s)_{s\in[0,t]}$ have the same law. Consequently, $\sup_{0\leq s\leq t} B_t \overset{d}{=} \sup_{0\leq s\leq t}(-\widetilde{B}_t)$, which is in fact $M_t \overset{d}{=} M_t - B_t$. □

4.4.4 The Local Maxima

In this part, we study the times where a Brownian motion attains its local maxima:

$$\mathcal{M} = \left\{t \geq 0 : B_t = \max_{s\in[t-\epsilon,t+\epsilon]} B_s \ \text{ for some } \epsilon > 0\right\}.$$

Lemma 4.42. *Let $[a,b]$ and $[c,d]$ be two disjoint intervals in $[0,\infty)$. Then the maxima of $(B_t)_{t\geq 0}$ on them are almost surely different.*

Proof. Without loss of generality, assume $b < c$. By the simple Markov property, the maximal increment $\sup_{t\in[c,d]} B_t - B_c \overset{d}{=} \sup_{t\in[0,d-c]} B_t$ and is independent of $(B_s)_{s\in[0,c]}$. Then

$$\mathbb{P}\left(\sup_{t\in[a,b]} B_t = \sup_{t\in[c,d]} B_t\right) = \mathbb{P}\left(\sup_{t\in[a,b]} B_t - B_c = \sup_{t\in[c,d]} B_t - B_c\right) = 0.$$ □

Lemma 4.43. *For a standard Brownian motion, almost surely, every local maximum is a strict local maximum.*

Proof. By Lemma 4.42, with probability 1, $(B_t)_{t\geq 0}$ has different maxima on any pair of disjoint rational intervals. If some $t^* \in \mathcal{M}$ is not strict, we derive a contradiction by selecting another local maximum $s^* \neq t^*$ in $(t^*-\epsilon, t^*+\epsilon)$ with $B_{s^*} = B_{t^*}$, and finding two disjoint rational intervals containing s^* and t^*, respectively. $\square$

Theorem 4.44. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion.*

(i) Almost surely, $\mathcal{M}$ is a countable dense subset of $[0,\infty)$.

(ii) For each $t > 0$, almost surely, $t \notin \mathcal{M}$, i.e. the Brownian motion does not attain a local maximum at t.

(iii) For each $\lambda \in \mathbb{R}$, almost surely, λ is not a local maximum of $(B_t)_{t\geq 0}$.

Proof. (i) By Lemma 4.43, almost surely, $\mathcal{M}$ is contained in the range of the mapping

$$[a,b] \mapsto \inf\left\{t \in [a,b] : B_t = \max_{s\in[a,b]} B_s\right\}$$

from all rational intervals into $[0,\infty)$, which is countable. By Corollary 4.27 (i), $(B_t)_{t\geq 0}$ is not monotone in any nondegenerate interval, hence it almost surely has a local maximum in every nondegenerate interval.

(ii) Fix $t > 0$. By Proposition 4.22 and simple Markov property of Brownian motions, with probability 1, we have $\sup_{t\leq s\leq t+\epsilon} B_s - B_t > 0$ for each $\epsilon > 0$, which implies that $t \notin \mathcal{M}$.

(iii) For every rational interval $[a,b] \subset [0,\infty)$, we have

$$\mathbb{P}\left(\max_{s\in[a,b]} B_s = \lambda\right) = \mathbb{E}\left[\mathbb{P}\left(\max_{s\in[a,b]} B_s = \lambda \,\middle|\, \mathscr{F}_a\right)\right] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi a}} e^{-\frac{x^2}{2a}} \mathbb{P}\left(\max_{s\in[0,b-a]} B_s = \lambda - x\right) = 0.$$

Then with probability 1, $(B_t)_{t\geq 0}$ does not attains the maximum λ on any rational interval $[a,b]$, and the result follows from Lemma 4.43. $\square$

4.4.5 The Arcsine Laws

The arcsine laws are a collection of results for one-dimensional Brownian motion.

Proposition 4.45 (Arcsine laws)**.** *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from $B_0 = 0$.*

(i) (Sign change). Let $L = \sup\{t \in [0,1] : B_t = 0\}$.

(ii) (Leftest maximum). Let $M_1 = \sup_{0\leq s\leq 1} B_s$, and define $U = \inf\{t \geq 0 : B_t = M_1\}$. Then $U < 1$ a.s.. The laws of L and U are both given by

$$\mathbb{P}(L \leq t) = \mathbb{P}(U \leq t) = \frac{2}{\pi}\arcsin(\sqrt{t}), \quad 0 \leq t \leq 1.$$

Proof. (i) The case for L is rather straightforward:

$$\begin{aligned}
\mathbb{P}(L \leq t) &= 2\int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \mathbb{P}\left(\sup_{s\in[0,1-t]} (B_t - B_{t+s}) \geq y\right) \mathrm{d}y = 2\int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \mathbb{P}\left(\tau_y \geq 1-t\right) \mathrm{d}y \\
&= 2\int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \left(\int_{1-t}^\infty \frac{y}{\sqrt{2\pi z^3}} e^{-\frac{y^2}{2z}} \mathrm{d}z\right) \mathrm{d}y = \frac{1}{\pi}\int_{1-t}^\infty \left(\int_0^\infty \frac{y}{\sqrt{tz^3}} e^{-\frac{y^2}{2t}-\frac{y^2}{2z}} \mathrm{d}y\right) \mathrm{d}z \\
&= \frac{1}{\pi}\int_{1-t}^\infty \frac{1}{\sqrt{tz^3}} \frac{tz}{t+z} \mathrm{d}z \overset{s=\frac{t}{t+z}}{=} \frac{1}{\pi}\int_0^t \frac{s^{3/2}}{t\sqrt{1-s}} \frac{\partial z}{\partial s} \mathrm{d}s \\
&= \frac{1}{\pi}\int_0^t \frac{1}{\sqrt{s(1-s)}} \mathrm{d}s = \frac{2}{\pi}\arcsin(\sqrt{t}).
\end{aligned}$$

(ii) Define $\widetilde{B}_t = B_1 - B_{1-t}$ for all $t \in [0,1]$, which is a Brownian motion on $[0,1]$. Then for all $\epsilon > 0$,

$$B_1 - M_1 \le B_1 - \sup_{1-\epsilon \le s \le 1} B_s = \inf_{0 \le s \le \epsilon} \widetilde{B}_s < 0 \quad a.s..$$

Clearly, we have $0 \le U \le 1$, and $\{U = 1\} \subset \{B_1 = M_1\}$. Hence $\mathbb{P}(U < 1) = 1$.

Fix $0 < t < 1$, and let $M_t = \sup_{0 \le s \le t} B_s$. By simple Markov property, $(B_{t+s} - B_t)_{s \in [0,1-t]}$ is independent of $(B_s)_{s \in [0,t]}$. Define $N_t = \sup_{0 \le s \le 1-t}(B_{t+s} - B_t)$. Then N_t is independent of $(M_t, B_t, M_t - B_t)$, and

$$\mathbb{P}(U \le t) = \mathbb{P}(M_t \ge B_t + N_t) = \mathbb{P}(N_t \le M_t - B_t) = \mathbb{P}(|B_1 - B_t| \le |B_t|).$$

Let Z_1, Z_2 be $N(0,1)$ *i.i.d.*, and θ is uniformly distributed on $[0, 2\pi)$. By calculus,

$$\mathbb{P}(U \le t) = \mathbb{P}\left(\sqrt{1-t}|Z_1| \le \sqrt{t}|Z_2|\right) = \mathbb{P}\left(\frac{|Z_1|}{\sqrt{Z_1^2 + Z_2^2}} \le t\right) = \mathbb{P}\left(|\sin\theta| \le \sqrt{t}\right) = \frac{2}{\pi}\arcsin(\sqrt{t}).$$

Hence we complete the proof. $\square$

Remark. We also point out some details about random times L and U not discussed in the arcsin laws:

(i) Both L and U are not stopping times;

(ii) By Theorem 4.44, almost surely, 0 is neither a local maximum nor a local minimum of $(B_t)_{t \ge 0}$. As a result, $(B_t)_{t \ge 0}$ always changes its sign at L;

(iii) By Lemma 4.42, almost surely, the maximum of $(B_t)_{t \ge 0}$ on $[0,1]$ is unique and strict. Consequently, U is the unique moment at which $(B_t)_{t \ge 0}$ achieves its maximum on $[0,1]$.

We will introduce one more arcsine law for Brownian motions in Theorem 9.12.

5 Stochastic Integration

In this chapter, our discussion is based on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a complete filtration $(\mathscr{F}_t)_{t\geq 0}$. All processes we study are indexed by $\mathbb{R}_+$ and take real values.

5.1 Construction of Stochastic Integrals

5.1.1 From Elementary Processes to L^2-Bounded Martingales

Preliminary: Space $\mathbb{H}^2$. Given a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, we denote by $\mathcal{H}^2$ the vector space of all 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 3.71, 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 $\mathbb{E}[M_\infty|\mathscr{F}_t] = M_t$ 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 3.70. 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

$$\lim_{n,m\to\infty} \mathbb{E}\left[\left(M_\infty^{(n)} - M_\infty^{(m)}\right)^2\right] = \lim_{n,m\to\infty} \left\|M^{(n)} - M^{(m)}\right\|_{\mathbb{H}^2}^2 = 0.$$

Then $M_\infty^{(n)}$ is a Cauchy sequence in L^2, and we denote $Z = \lim_{n\to\infty} M_\infty^{(n)}$ in L^2. On the other hand, the Doob's L^2-inequality [Proposition 3.52 (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 $M_t = \lim_{k\to\infty} M_t^{(n_k)}$ 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 $(\mathscr{F}_t)_{t\geq 0}$, the process $M = (M_t)_{t\geq 0}$ is adapted. Also, the continuity of conditional expectation passes $M_t^{(n_k)} = \mathbb{E}[M_\infty^{(n_k)}|\mathscr{F}_t]$ to $M_t = \mathbb{E}[Z|\mathscr{F}_t]$ as $k \to \infty$, 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

$$\lim_{n\to\infty} \left\|M^{(n)} - M\right\|_{\mathbb{H}^2}^2 = \lim_{n\to\infty} \mathbb{E}\left[\left(M_\infty^{(n)} - M_\infty\right)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(M_\infty^{(n)} - Z_\infty\right)^2\right] = 0.$$

Thus $M \in \mathbb{H}^2$ is indeed the limit of sequence $M^{(n)}$, completing the proof. $\square$

Preliminary: Progressive σ-fields. Given $(\Omega, \mathscr{F}, \mathbb{P})$, we define the progressive σ-field on $\Omega \times \mathbb{R}_+$ as

$$\mathscr{P} = \{A \subset \Omega \times \mathbb{R}_+ : A \cap (\Omega \times [0,t]) \in \mathscr{F}_t \otimes \mathscr{B}([0,t]),\ \forall t \in \mathbb{R}_+\} \subset \mathscr{F} \otimes \mathscr{B}(\mathbb{R}_+),$$

where the inclusion holds since $A = \bigcup_{n=1}^{\infty}(A \cap (\Omega \times [0,n]))$. Clearly, if $A \in \mathscr{P}$, the process $X_t(\omega) = \mathbb{1}_A(\omega, t)$ is a progressive process. Furthermore, one can verify that $\mathscr{P}$ 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 $\mathscr{P}$-measurable.

Preliminary: Space $L^2(M)$. Given a process $M \in \mathbb{H}^2$, the Theorem 3.69 determines to an increasing process $(\langle M, M\rangle_s)_{s\geq 0}$, which is called the quadratic variation of M. Then for every $A \in \mathscr{P}$, one can define

$$\mu_M(A) = \mathbb{E}\left[\int_0^\infty \mathbb{1}_A(\cdot, s)\,\mathrm{d}\langle M, M\rangle_s\right].$$

This is a measure on $(\Omega \times \mathbb{R}_+, \mathscr{P})$. We denote by $\mathcal{L}^2(M)$ the space of all progressive progresses H such that

$$\|H\|^2_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_s^2\,\mathrm{d}\langle M, M\rangle_s\right] < \infty,$$

and choose the quotient space $L^2(M)$ that makes $\|\cdot\|_{L^2(M)}$ a proper norm. Then $L^2(M) = L^2(\Omega\times\mathbb{R}_+, \mathscr{P}, \mu_M)$ can be viewed as an ordinary L^2-space, and we can define the inner product

$$\langle H, K\rangle_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_s K_s\,\mathrm{d}\langle M, M\rangle_s\right]$$

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 $\mathbb{1}_{[0,\tau]}H$ defined by $(\mathbb{1}_{[0,\tau]}H)_s(\omega) = \mathbb{1}_{[0,\tau(\omega)]}(s)H_s(\omega)$ also belongs to $L^2(M)$. Note that $\mathbb{1}_{[0,\tau]}H$ is progressive since it has left-continuous sample paths.

Definition 5.1 (Elementary processes)**.** An *elementary process* is a progressive process of the form

$$H_s(\omega) = \sum_{j=1}^n H_{(j)}(\omega)\mathbb{1}_{(t_{j-1},t_j]}(s),$$

where $0 = t_0 < t_1 < \cdots < t_n$, and $H_{(j)}$ is a bounded $\mathscr{F}_{t_{j-1}}$-measurable random variable for all $j \in \{1, \cdots, n\}$. Clearly, the set $\mathscr{E}$ of all (equivalence classes of) elementary processes is a subspace of $L^2(M)$.

Proposition 5.2. *For every $M \in \mathbb{H}^2$, $\mathscr{E}$ is dense in $L^2(M)$.*

Proof. Fix $M \in \mathbb{H}^2$. By elementary Hilbert space theory, it suffices to show that $L^2(M) \ni K \perp \mathscr{E}$ implies $K = 0$. Assume that $K \in L^2(M)$ is orthogonal to $\mathscr{E}$, and set

$$X_t = \int_0^t K_s\,\mathrm{d}\langle M, M\rangle_s, \quad \forall t \geq 0.$$

According to Proposition 3.64, since

$$\mathbb{E}\left[\int_0^t |K_s|\,|\mathrm{d}\langle M, M\rangle_s|\right] \leq \left(\mathbb{E}\left[\int_0^t K_s^2\,\mathrm{d}\langle M, M\rangle_s\right]\right)^{1/2}\left(\mathbb{E}\left[\int_0^t d\langle M, M\rangle_s\right]\right)^{1/2} \leq \|K\|_{L^2(M)}\,\|M\|_{\mathbb{H}^2},$$

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 $\mathscr{F}_s$-measurable random variable. Then

$$0 = \langle H, K\rangle_{L^2(M)} = \mathbb{E}\left[Y\int_s^t K_u \,\mathrm{d}\langle M, M\rangle_u\right] = \mathbb{E}[Y(X_t - X_s)]$$

Since $\mathbb{E}[Y(X_t - X_s)] = 0$ for all bounded $\mathscr{F}_s$-measurable random variable Y, we have $\mathbb{E}[X_t - X_s|\mathscr{F}_s] = 0$. 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 3.68, we have $X = 0$ *a.s.*, i.e.

$$\int_0^t K_s \,\mathrm{d}\langle M, M\rangle_s = 0 \quad \forall t \geq 0, \quad a.s. \quad \Rightarrow \quad K_s = 0 \quad d\langle M, M\rangle_s\text{-}a.e., \quad a.s..$$

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 $H \in \mathscr{E}$, we define the following formula:*

$$H_s = \sum_{j=1}^n H_{(j)}\mathbb{1}_{(t_{j-1},t_j]}(s) \quad \Rightarrow \quad (H\cdot M)_t = \sum_{j=1}^n H_{(j)}\left(M_{t_j\wedge t} - M_{t_{j-1}\wedge t}\right)$$

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

$$\langle H\cdot M, N\rangle = H \cdot \langle M, N\rangle, \quad \forall N \in \mathbb{H}^2, \tag{5.1}$$

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

$$(\mathbb{1}_{[0,\tau]}H)\cdot M = (H\cdot M)^\tau = H\cdot M^\tau. \tag{5.2}$$

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 depends 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 $\mathscr{E}$ into $L^2(M)$.

We fix the process $H = (H_s)_{s\geq 0}$ of the form given in the theorem. For every $j \in \{1, \cdots, n\}$, 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

$$\begin{aligned}
\left\langle M^{(j)}, M^{(j)}\right\rangle_t &= \sum_{j=1}^n H_{(j)}^2\left(\langle M, M\rangle_{t_j\wedge t} - \langle M, M\rangle_{t_{j-1}\wedge t}\right), && \text{(By the approximation formula)}\\
\left\langle M^{(j)}, M^{(k)}\right\rangle_{\mathbb{H}^2} &= \mathbb{E}\left[M_\infty^{(j)}M_\infty^{(k)}\right] = \mathbb{E}\left[H_{(j)}H_{(k)}(M_{t_j} - M_{t_{j-1}})(M_{t_k} - M_{t_{k-1}})\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[H_{(j)}H_{(k)}(M_{t_j} - M_{t_{j-1}})(M_{t_k} - M_{t_{k-1}})|\mathscr{F}_{t_{j-1}}\right]\right] = 0, \quad \forall 1 \leq j < k \leq n.
\end{aligned}$$

By orthogonality of $M^{(j)}$'s and bilinearity of quadratic variation, it holds

$$\langle H\cdot M, H\cdot M\rangle_t = \sum_{j=1}^n H_{(j)}^2\left(\langle M, M\rangle_{t_j\wedge t} - \langle M, M\rangle_{t_{j-1}\wedge t}\right) = \int_0^t H_s^2 \,\mathrm{d}\langle M, M\rangle_s.$$

Consequently, we have

$$\|H\cdot M\|_{\mathbb{H}^2}^2=\mathbb{E}\left[\langle H\cdot M,H\cdot M\rangle_\infty\right]=\mathbb{E}\left[\int_0^\infty H_s^2\,\mathrm{d}\langle M,M\rangle_s\right]=\|H\|_{L^2(M)}^2,\quad \forall t\geq 0.$$

By linearity, if $H=H'$ in $L^2(M)$, then $H\cdot M=H'\cdot M$ in $\mathbb{H}^2$. Therefore the mapping $\mathscr{E}\to\mathbb{H}^2$ 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 3.76] implies

$$\mathbb{E}\left[\int_0^\infty |H_s|\,|\mathrm{d}\langle M,N\rangle_s|\right]\leq \|H\|_{L^2(M)}\|N\|_{\mathbb{H}^2}<\infty,\quad \forall H\in L^2(M).$$

Then the variable $(H\cdot\langle M,N\rangle)_\infty=\int_0^\infty H_s\,\mathrm{d}\langle M,N\rangle_s$ is well-defined and in L^1. For the case where H is an elementary process, we have

$$\begin{cases}\langle H\cdot M,N\rangle=\sum_{j=1}^n\langle M^{(j)},N\rangle\\ \langle M^{(j)},N\rangle_t=H_{(j)}^2\left(\langle M,N\rangle_{t_j\wedge t}-\langle M,N\rangle_{t_{j-1}\wedge t}\right),\quad \forall t\geq 0.\end{cases}$$

This gives (5.1) in the case $H\in\mathscr{E}$:

$$\langle H\cdot M,N\rangle_t=\sum_{j=1}^n H_{(j)}^2\left(\langle M,N\rangle_{t_j\wedge t}-\langle M,N\rangle_{t_{j-1}\wedge t}\right)=\int_0^t H_s\,\mathrm{d}\langle M,N\rangle_s=(H\cdot\langle M,N\rangle)_t,\quad \forall t\geq 0.$$

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 $L^1(\Omega,\mathscr{F},\mathbb{P})$:

$$\mathbb{E}\left[|\langle X,N\rangle_\infty|\right]=\mathbb{E}\left[\langle X,X\rangle_\infty\right]^{1/2}\mathbb{E}\left[\langle N,N\rangle_\infty\right]^{1/2}=\|X\|_{\mathbb{H}^2}\|N\|_{\mathbb{H}^2}.$$

Let $H^{(n)}\in\mathscr{E}$ be a sequence that converges to H in $L^2(M)$. Then $H^{(n)}\cdot M\to H\cdot M$ in $\mathbb{H}^2$, and

$$\langle H\cdot M,N\rangle_\infty=\lim_{n\to\infty}\left\langle H^{(n)}\cdot M,N\right\rangle_\infty=\lim_{n\to\infty}\left(H^{(n)}\cdot\langle M,N\rangle\right)_\infty=(H\cdot\langle M,N\rangle)_\infty,$$

where the last equality holds in L^1 by Kunita-Watanabe:

$$\mathbb{E}\left[\left|\int_0^\infty (H_s^{(n)}-H_s)\,\mathrm{d}\langle M,N\rangle_s\right|\right]\leq\left\|H_s^{(n)}-H_s\right\|_{L^2(M)}\|N\|_{\mathbb{H}^2}.$$

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 3.70 (ii), we have $H\cdot M-X=0$ *a.s.*.

Finally it remains to show (5.2). By Proposition 3.75 (iv), for all $N\in\mathbb{H}^2$, we have

$$\langle (H\cdot M)^\tau,N\rangle_t=\langle H\cdot M,N\rangle_{t\wedge\tau}=(H\cdot\langle M,N\rangle)_{t\wedge\tau}=\left(\mathbb{1}_{[0,\tau]}H\cdot\langle M,N\rangle\right)_t,$$

which implies the first inequality. The proof for the second one is similar:

$$\langle H\cdot M^\tau,N\rangle=H\cdot\langle M^\tau,N\rangle=H\cdot\langle M,N\rangle^\tau=\mathbb{1}_{[0,\tau]}H\cdot\langle M,N\rangle.$$

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:

$$\int_0^t H_s \,\mathrm{d}M_s = (H \cdot M)_t, \quad \forall t \geq 0, \quad \textit{and} \quad \int_0^\infty H_s \,\mathrm{d}M_s = (H \cdot M)_\infty.$$

The property (5.1) gives commutativity of stochastic integral and bracket:

$$\left\langle \int_0^\cdot H_s \,\mathrm{d}M_s, N \right\rangle_t = \int_0^t H_s \,\mathrm{d}\langle M, N\rangle_s$$

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 $\int_0^\infty |H_s|\,|\mathrm{d}A_s| < \infty$ a.s.. If $\int_0^\infty |K_s H_s|\,|\mathrm{d}A_s| < \infty$ 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:

$$\mathbb{E}\left[\left(K^2 H^2 \cdot \langle M, M\rangle\right)_\infty\right] = \mathbb{E}\left[\left(K^2 \cdot \langle H \cdot M, H \cdot M\rangle\right)_\infty\right]$$

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:

$$\langle H \cdot M, K \cdot N\rangle_t = \left\langle \int_0^\cdot H_s \,\mathrm{d}M_s, \int_0^\cdot K_s \,\mathrm{d}N_s \right\rangle_t = \int_0^t H_s K_s \,\mathrm{d}\langle M, N\rangle_s.$$

According to Proposition 3.75 (vi), we have

$$\mathbb{E}\left[\left(\int_0^t H_s \,\mathrm{d}M_s\right)\left(\int_0^t K_s \,\mathrm{d}N_s\right)\right] = \mathbb{E}\left[\int_0^t H_s K_s \,\mathrm{d}\langle M, N\rangle_s\right].$$

Note that $H \cdot M = \int_0^\cdot H_s \,\mathrm{d}M_s$ is a martingale of $\mathbb{H}^2$. For all $0 \leq s < t \leq \infty$, we have

$$\mathbb{E}\left[\int_0^t H_s \,\mathrm{d}M_s\right] = 0, \quad \textit{and} \quad \mathbb{E}\left[\int_0^t H_u \,\mathrm{d}M_u \middle| \mathscr{F}_s\right] = \int_0^s H_u \,\mathrm{d}M_u.$$

According to Theorem 3.71, the second moment of the stochastic integral is given by

$$\mathbb{E}\left[\left(\int_0^t H_s \,\mathrm{d}M_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 \,\mathrm{d}\langle M, M\rangle_s\right].$$

Next we will discuss stochastic integrals for local martingales and semimartingales.

5.1.2 Stochastic Integrals for Local Martingales and Semimartingales

Preliminaries: Space $L^2_{\text{loc}}(M)$. Let M be a continuous local martingale. Similar to the case $M \in \mathbb{H}^2$, we can define a Hilbert space $L^2(M)$ associated with M containing all progressive processes H such that $\mathbb{E}\left[\int_0^\infty H_s^2 \,\mathrm{d}\langle M, M\rangle_s\right] < \infty$. Furthermore, we denote by $L^2_{\text{loc}}(M)$ the set of all progressive processes such that

$$\int_0^t H_s^2 \,\mathrm{d}\langle M, M\rangle_s < \infty, \quad \forall t \geq 0, \quad a.s..$$

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,*

$$\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle. \tag{5.3}$$

If τ is a stopping time, we then have

$$(\mathbb{1}_{[0,\tau]} H) \cdot M = (H \cdot M)^\tau = H \cdot M^\tau. \tag{5.4}$$

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 $\int_0^s H_s^2 \,\mathrm{d}\langle M, M\rangle_s < \infty$ 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:

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t (1 + H_s^2) \,\mathrm{d}\langle M, M\rangle_s \geq n\right\}.$$

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,

$$\int_0^\infty H_s^2 \,\mathrm{d}\langle M^{\tau_n}, M^{\tau_n}\rangle_s = \int_0^{\tau_n} H_s^2 \,\mathrm{d}\langle M, M\rangle_s \leq n$$

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 $(H \cdot M)_t = \lim_{n\to\infty}(H \cdot M^{\tau_n})_t$ 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 $(H \cdot M)^{\tau_n} = \lim_{m\to\infty}(H \cdot M^{\tau_m})^{\tau_n} = H \cdot M^{\tau_n} \in \mathbb{H}^2$. 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 $\sigma_n \to \infty$ as $n \to \infty$, 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. ☐

Remark. (i) (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 $\langle H \cdot M, H \cdot M\rangle = H^2 \cdot \langle M, M\rangle$.

(ii) (Connection to Wiener's integral). A Brownian motion $B = (B_t)_{t\geq 0}$ is a continuous martingale with respect to the natural filtration $\mathscr{F}_t = \sigma(B_s, 0 \leq s \leq t)$. According to the Remark of Proposition 4.4, for each $h \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$, we can define the Wiener integral $\int_0^t h(s)\,\mathrm{d}B_s = W(h\mathbb{1}_{[0,t]})$, 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

$$\int_0^t h(s)\,\mathrm{d}B_s = W(h\mathbb{1}_{[0,t]}) = \sum_{j=1}^n \lambda_j \left(B_{t_j\wedge t} - B_{t_{j-1}\wedge t}\right).$$

Then for all continuous local martingales N, we have

$$\left\langle \int_0^{\cdot} h(s)\,\mathrm{d}B_s, N \right\rangle_t = \sum_{j=1}^n \lambda_j(\langle B_{t_j\wedge\cdot}, N\rangle_t - \langle B_{t_{j-1}\wedge\cdot}, N\rangle_t) = \sum_{j=1}^n \lambda_j(\langle B, N\rangle_{t_j\wedge t} - \langle B, N\rangle_{t_{j-1}\wedge t}) = (h \cdot \langle B, N\rangle)_t.$$

Therefore we have $\int_0^t h(s)\,\mathrm{d}B_s = (h \cdot B)_t$ for all step functions h. For the general case $h \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$, the quadratic variation $\langle B, B\rangle_t = t$ implies $h \in L^2(B)$, and the result follows from a density argument.

(iii) (Moment formulae). In the setting of Theorem 5.5, we again write $\int_0^t H_s\,\mathrm{d}M_s = (H \cdot M)_t$. If M, N are continuous local martingales, $H \in L^2_{\mathrm{loc}}(M)$ and $K \in L^2_{\mathrm{loc}}(N)$, the first formula in the Remark under Proposition 5.4 still holds true for $\langle H \cdot M, K \cdot N\rangle$:

$$\langle H \cdot M, K \cdot N\rangle_t = \int_0^t H_s K_s\,\mathrm{d}\langle M, N\rangle_s, \quad \forall t \geq 0;$$

whereas the formulae for moments of $\int_0^t H_s\,\mathrm{d}M_s$ may fail.

We can make an extension. For a continuous local martingale M and a progressive process $H \in L^2_{\mathrm{loc}}(M)$, and for some fixed $t > 0$, assume the following condition holds:

$$\mathbb{E}\left[\int_0^t H_s^2\,\mathrm{d}\langle M, M\rangle_s\right] < \infty. \tag{5.5}$$

According to Theorem 3.71, the stopped process $(H \cdot M)^t$ is a martingale of $\mathbb{H}^2$. As a result, we have

$$\mathbb{E}\left[\int_0^t H_s\,\mathrm{d}M_s\right] = 0, \quad \mathbb{E}\left[\left(\int_0^t H_s\,\mathrm{d}M_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2\,\mathrm{d}\langle M, M\rangle_s\right].$$

Consequently, regardless of whether the condition (5.5) holds, we have the following bound:

$$\mathbb{E}\left[\left(\int_0^t H_s\,\mathrm{d}M_s\right)^2\right] \leq \mathbb{E}\left[\int_0^t H_s^2\,\mathrm{d}\langle M, M\rangle_s\right]. \tag{5.6}$$

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

$$\int_0^t |H_s|\,|\mathrm{d}V_s| \leq \sup_{0\leq s\leq t} |H_s| \left(\int_0^t |\mathrm{d}V_s|\right) < \infty \ \ a.s., \quad \forall t > 0.$$

In particular, an adapted and 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 $0 = t_0 < t_1 < \cdots < t_n$, and $H_{(j)}$ is a $\mathscr{F}_{t_{j-1}}$-measurable random variable for every $j \in \{1, \cdots, n\}$, then

$$(H \cdot X)_t = \sum_{j=1}^n H_{(j)} \left(X_{t_j \wedge t} - X_{t_{j-1} \wedge t} \right).$$

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 $\inf \emptyset = \infty$:

$$\tau_k = \inf \{t \geq 0 : |H_s| \geq k\} = \inf \left\{ t_{j-1} : |H_{(j)}| \geq k \right\}.$$

Then $\tau_k \uparrow \infty$ as $k \to \infty$. Furthermore, for every k,

$$H_s \mathbb{1}_{[0,\tau_k]}(s) = \sum_{j=1}^n H_{(j)}^k \mathbb{1}_{(t_{j-1},t_j]}(s), \quad \textit{where } H_{(j)}^k = H_{(j)} \mathbb{1}_{\{\tau_k \geq t_{j-1}\}} \leq k.$$

Consequently, $H\mathbb{1}_{[0,\tau_k]}$ is an elementary process, and its stochastic integral with respect to $M \in \mathbb{H}^2$ is

$$(H \cdot M)_{t \wedge \tau_k} = (H \mathbb{1}_{[0,\tau_n]} \cdot M)_t = \sum_{j=1}^n H_{(j)}^k \left(X_{t_j \wedge t} - X_{t_{j-1} \wedge t} \right).$$

Then the desired result follows by letting $k \to \infty$. ☐

We introduce two important convergence results 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 $\lim_{n \to \infty} H_s^n = H_s$ a.s. for every $s \in [0, T]$, where H is a locally bounded progressive process. Let $K = (K_s)_{s \geq 0}$ be a nonnegative progressive process such that*

$$\int_0^T K_s^2 \, \mathrm{d}\langle M, M \rangle_s < \infty \ a.s., \quad \textit{and} \quad \int_0^T K_s \, |\mathrm{d}A_s| < \infty \ a.s.. \tag{5.7}$$

*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 $H^n \cdot X \to H \cdot X$ **uniformly on** $[0, T]$ **in probability**, i.e.*

$$\lim_{n \to \infty} \mathbb{P} \left(\sup_{t \in [0,T]} \left| \int_0^t H_s^n \, \mathrm{d}X_s - \int_0^t H_s \, \mathrm{d}X_s \right| > \epsilon \right) = 0, \quad \textit{for each } \epsilon > 0.$$

Note the property (5.7) *holds if K is a locally bounded progressive progress.*

Proof. Since $|H_s^n - H_s| \le 2K_s$ a.s. and $\int_0^T K_s |\mathrm{d}A_s| < \infty$, by Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \int_0^T |H_s^n - H_s|\,|\mathrm{d}A_s| = 0 \quad \text{a.s..}$$

So we only need to deal with the case when $X = M$ is a continuous local martingale. For every $N \in \mathbb{N}$, choose the following stopping time:

$$\tau_N = \inf\left\{t \in [0,T] : \int_0^t K_s^2\,\mathrm{d}\langle M,M\rangle_s \ge N\right\} \wedge T.$$

By assumption (5.7), we have $\mathbb{P}(\tau_N = T) \to 1$ as $N \to \infty$. Furthermore, on $[0,T]$,

$$|(H^n - H)\cdot M|^2 \le \langle (H^n - H)\cdot M, (H^n - H)\cdot M\rangle = (H^n - H)^2 \cdot \langle M,M\rangle.$$

Hence for every $t \in [0,T]$,

$$\left|\int_0^{t\wedge\tau_N} (H_s^n - H_s)\,\mathrm{d}M_s\right|^2 \le \int_0^{t\wedge\tau_N} (H_s^n - H_s)^2\,\mathrm{d}\langle M,M\rangle_s = \int_0^{\tau_N} (H_s^n - H_s)^2\,\mathrm{d}\langle M,M\rangle_s.$$

Since $\int_0^{\tau_N} K_s^2\,\mathrm{d}\langle M,M\rangle_s \le N$, and $|H_s^n - H_s| \le 2K_s$ and $H_s^n \to H_s$ a.s. for each $s \in [0,T]$, by Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^{t\wedge\tau_N} (H_s^n - H_s)\,\mathrm{d}M_s\right|^2\right] \le \lim_{n\to\infty} \mathbb{E}\left[\int_0^{\tau_N} (H_s^n - H_s)^2\,\mathrm{d}\langle M,M\rangle_s\right] = 0.$$

Then for every $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,T]} \left|\int_0^t (H_s^n - H_s)\,\mathrm{d}M_s\right| \ge \epsilon\right) &\le \mathbb{P}(\tau_N < t) + \mathbb{P}\left(\sup_{t\in[0,T]} \left|\int_0^{t\wedge\tau_N} (H_s^n - H_s)\,\mathrm{d}M_s\right| \ge \epsilon\right) \\ &\le \mathbb{P}(\tau_N < t) + \frac{1}{\epsilon^2}\mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^{t\wedge\tau_N} (H_s^n - H_s)\,\mathrm{d}M_s\right|^2\right].\end{aligned}$$

Letting $n \to \infty$ and $N \to \infty$, we obtain the desired result. □

Corollary 5.8 (Dominated convergence theorem for stochastic integrals)**.** *Let $X = (X_t)_{t\ge 0}$ be a continuous semimartingale. Let $(H^n)_{n=1}^\infty$ be a sequence of locally bounded progressive processes such that $\lim_{n\to\infty} H_s^n = H_s$ a.s. for every $s \ge 0$, where H is a locally bounded progressive process. Let $K = (K_s)_{s\ge 0}$ be a nonnegative progressive process satisfying* (5.7) *for every $T > 0$. If $|H_s^n| \le K_s$ a.s. for every $n \in \mathbb{N}$ and every $s \ge 0$, then for every stopping time τ with $\tau < \infty$ a.s.,*

$$\lim_{n\to\infty} \int_0^\tau H_s^n\,\mathrm{d}X_s = \int_0^\tau H_s\,\mathrm{d}X_s \quad \textit{in probability.}$$

Proof. We fix $N > 0$. Then $H^n \cdot X \to H \cdot X$ uniformly on $[0,T]$ in probability, and for every $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\int_0^\tau H_s^n\,\mathrm{d}X_s - \int_0^\tau H_s\,\mathrm{d}X_s\right| > \epsilon\right) &\le \mathbb{P}\left(\left|\int_0^\tau H_s^n\,\mathrm{d}X_s - \int_0^\tau H_s\,\mathrm{d}X_s\right| > \epsilon,\ \tau \le N\right) + \mathbb{P}(\tau > N) \\ &\le \mathbb{P}\left(\sup_{t\in[0,N]} \left|\int_0^t H_s^n\,\mathrm{d}X_s - \int_0^t H_s\,\mathrm{d}X_s\right| > \epsilon\right) + \mathbb{P}(\tau > N).\end{aligned}$$

We then let $n \to \infty$ and $N \to \infty$ to obtain the desired result. □

For continuous integrands, we have the following useful approximation result.

Proposition 5.9. *Let X be a continuous martingale, and let H be an continuous adapted process. For every $t > 0$ and every sequence of partitions $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ whose mesh tends to 0, we have*

$$\int_0^t H_s \,\mathrm{d}X_s = \lim_{n\to\infty} \sum_{j=1}^{k_n} H_{t_{j-1}^n}(X_{t_j^n} - X_{t_{j-1}^n}) \quad \textit{in probability.} \tag{5.8}$$

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

$$\int_0^t H_s^n \,\mathrm{d}X_s = \sum_{j=1}^{k_n} H_{t_{j-1}^n}(X_{t_j^n} - X_{t_{j-1}^n}), \quad \forall n \in \mathbb{N}.$$

We take $K_s = \max_{0\le r\le s} |H_r|$. This is a locally bounded progressive process dominating H^n. Since H is sample continuous, we have $H_s^n \to H_s$ *a.s.* for all $s \in [0, t]$. Using Theorem 5.7 concludes the proof. ☐

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 $\lim_{n\to\infty} \sum_{j=1}^{k_n} Y_{t_j^n}(X_{t_j^n} - X_{t_{j-1}^n}) = \int_0^t Y_s \,\mathrm{d}X_s + \langle X, Y\rangle_t$, and

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s \,\mathrm{d}Y_s + \int_0^t Y_s \,\mathrm{d}X_s + \langle X, Y\rangle_t. \tag{5.9}$$

This is known as the *formula of integration by parts.*

5.2 Itô's Formula and its Consequences

Theorem 5.10 (Itô's Lemma)**.** *Let $X^1, \cdots, X^p$ be p continuous semimartingales, and let $F \in C^2(\mathbb{R}^p)$. Then for every $t \geq 0$, we have*

$$F(X_t^1, \cdots, X_t^p) = F(X_0^1, \cdots, X_0^p) + \sum_{j=1}^{p} \int_0^t \frac{\partial F(X_s^1, \cdots, X_s^p)}{\partial x^j} \, \mathrm{d}X_s^j + \frac{1}{2} \sum_{i,j=1}^{p} \int_0^t \frac{\partial^2 F(X_s^1, \cdots, X_s^p)}{\partial x^i \partial x^j} \, \mathrm{d}\langle X^i, X^j \rangle_s.$$

Proof. We write $(X^1, \cdots, X^p) = X$ for brevity. Fix $t > 0$, and consider an increasing sequence of partitions $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ whose mesh tends to 0. According to Taylor's theorem,

$$\begin{aligned} F(X_t) &= F(X_0) + \sum_{l=1}^{k_n} \Big(F(X_{t_j^n}) - F(X_{t_{j-1}^n}) \Big) \\ &= F(X_0) + \underbrace{\sum_{j=1}^{p} \sum_{l=1}^{k_n} \frac{\partial F}{\partial x^j}(X_{t_{l-1}^n})(X_{t_l^n}^j - X_{t_{l-1}^n}^j)}_{\text{(a)}} + \underbrace{\frac{1}{2} \sum_{i,j=1}^{p} \sum_{l=1}^{k_n} f_{n,l}^{i,j}(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j)}_{\text{(b)}}, \end{aligned}$$

where the quantity $f_{n,l}^{i,j}$ can be written as

$$f_{n,l}^{i,j} = \frac{\partial^2 F}{\partial x^i \partial x^j}((1-\xi)X_{t_{l-1}^n} + \xi X_{t_l^n})$$

for some random variable $\xi : \Omega \to [0,1]$. By Proposition 5.9, the term (a) converges to $\sum_{j=1}^p \int_0^t \frac{\partial F}{\partial x^j}(X_s) \, \mathrm{d}X_s^j$ in probability as $n \to \infty$. 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 $i, j \in \{1, \cdots, p\}$ that

$$\sup_{1 \leq l \leq k_n} \left| f_{n,l}^{i,j} - D_{ij}F(X_{t_{l-1}^n}) \right| \leq \sup_{1 \leq l \leq k_n} \left(\sup_{x \in \left[X_{t_{l-1}^n} \wedge X_{t_l^n}, X_{t_{l-1}^n} \vee X_{t_l^n} \right]} \left| D_{ij}F(x) - D_{ij}F(X_{t_{l-1}^n}) \right| \right) \to 0 \quad a.s..$$

By Proposition 3.79, $\sum_{l=1}^{k_n}(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \xrightarrow{\mathbb{P}} \langle X^i, X^j \rangle_t < \infty$. This gives an estimate of (b):

$$\left| \sum_{l=1}^{k_n} D_{ij}F(X_{t_{l-1}^n})(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) - \sum_{l=1}^{k_n} f_{n,l}^{i,j}(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \right| \xrightarrow{\mathbb{P}} 0 \quad as \;\; n \to \infty.$$

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

$$\begin{aligned} &\lim_{n \to \infty} \sum_{l=1}^{k_n} D_{ij}F(X_{t_{l-1}^n})(X_{t_l^n}^i - X_{t_{l-1}^n}^i)(X_{t_l^n}^j - X_{t_{l-1}^n}^j) \\ &\quad = \int_0^t D_{ij}F(X_s) \, \mathrm{d}(X^i X^j)_s - \int_0^t D_{ij}F(X_s) X_s^i \, \mathrm{d}X_s^j - \int_0^t D_{ij}F(X_s) X_s^j \, \mathrm{d}X_s^i && \textit{(in probability)} \\ &\quad = \int_0^t D_{ij}F(X_s) \, \mathrm{d}(X^i X^j)_s - \int_0^t D_{ij}F(X_s) \, \mathrm{d}(X^i \cdot X^j)_s - \int_0^t D_{ij}F(X_s) \, \mathrm{d}(X^j \cdot X^i)_s && \textit{(by associativity)} \\ &\quad = \int_0^t D_{ij}F(X_s) \, \mathrm{d}\langle X^i, X^j \rangle_s. && \textit{(by linearity and (5.9))} \end{aligned}$$

Thus we finish the proof of Itô's formula. □

Remark. The formula (5.9) of integration by parts is a special case of Itô's lemma.

Proposition 5.11. *Take a twice continuously differentiable function $F(r,x)$ in $\mathbb{R}^2$. Itô's formula implies*

$$F(\langle X,X\rangle_t, X_t) = F(0,X_0) + \int_0^t \frac{\partial F}{\partial x}(\langle X,X\rangle_s, X_s)\,\mathrm{d}X_s + \int_0^t \left(\frac{\partial F}{\partial r} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)(\langle X,X\rangle_s, X_s)\,\mathrm{d}\langle M,M\rangle_s$$

For a continuous local martingale M, $F(\langle M,M\rangle, M)$ 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 $F : \mathbb{R}^2 \to \mathbb{C}$ satisfies the above condition, and $\frac{\partial F}{\partial x} = \lambda F$. We define

$$\mathscr{E}(\lambda M)_t = \exp\bigg(\lambda M_t - \frac{\lambda^2}{2}\langle M,M\rangle_t\bigg), \quad \forall t \geq 0.$$

Consequently, $\mathscr{E}(\lambda M)$ is a complex continuous local martingale (i.e. both its real and imaginary parts), and

$$\mathscr{E}(\lambda M)_t = \mathrm{e}^{\lambda M_0} + \lambda \int_0^t \mathscr{E}(\lambda M)_s\,\mathrm{d}M_s.$$

5.2.1 Multidimensional Brownian motions

A *d-dimensional Brownian motion* is a stochastic process $\{B_t = (B_t^1, \cdots, B_t^d), t \geq 0\}$ with values in $\mathbb{R}^d$ whose component processes $B^1, \cdots, B^d$ are independent Brownian motions. A Brownian motion $(B_t)_{t\geq 0}$ is called a *$(\mathscr{F}_t)$-Brownian motion* if it is adapted and has independent increments with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$.

Theorem 5.12 (Lévy's characterization of multi-dimensional Brownian motions)**.** *An adapted and continuous process $B = (B^1, \cdots, B^d)$ is a $(\mathscr{F}_t)$-Brownian motion if and only if its component processes $B_1, \cdots, B_d$ are continuous local martingales such that $\langle B^i, B^j\rangle_t = \delta_{ij}t$ for all $i,j \in \{1, \cdots, d\}$ and all $t \geq 0$.*

Proof. We only prove the sufficiency part, since the other direction is clear. Take $\alpha = (\alpha^1, \cdots, \alpha_d) \in \mathbb{R}^d$ 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 5.11, the process $\big(\exp(\mathrm{i}\alpha^\top B_t - \frac{1}{2}|\alpha|^2 t)\big)_{t\geq 0}$ is a continuous local martingale bounded on each compact interval $[0,t]$, $t > 0$, hence is a martingale. As a result, for all $t \geq s > 0$,

$$\mathbb{E}\left[\exp\left(i\alpha^\top B_t - \frac{1}{2}|\alpha|^2 t\right)\middle|\mathscr{F}_s\right] = \exp\left(i\alpha^\top B_s - \frac{1}{2}|\alpha|^2 s\right) \quad \Rightarrow \quad \mathbb{E}\left[\exp\big(i\alpha^\top (B_t - B_s)\big)|\mathscr{F}_s\right] = e^{-\frac{1}{2}|\alpha|^2(t-s)}.$$

Given any $A \in \mathscr{F}_s$, we take the measure $\mathbb{P}_A(E) = \mathbb{P}(A \cap E)/\mathbb{P}(A)$, $\forall E \in \mathscr{F}$. Comparing the characteristic functions, the law of $B_t - B_s$ does not change from $\mathbb{P}$ to $\mathbb{P}_A$. Then $\mathbb{E}\left[f(B_t - B_s)\mathbb{1}_A\right] = \mathbb{P}(A)\mathbb{E}\left[f(B_t - B_s)\right]$ for all nonnegative measurable functions. Choosing f to be indicator functions do we obtain that $B_t - B_s$ is independent of $\mathscr{F}_s$, and $B_t - B_s \sim N(0, (t-s)I_d)$. Since B is adapted and continuous and has independent Gaussian increments, it is a $(\mathscr{F}_t)$-Brownian motion. □

Remark. (i) (Rotational invariance of Brownian motions). If B is a d-dimensional Brownian motion and $Q \in \mathbb{R}^{d\times d}$ is an orthogonal matrix, then QB is also a d-dimensional Brownian motion. To see this, we note

$$\langle X^i, X^j\rangle_t = \left\langle \sum_{k=1}^d Q_{ik}B^k, \sum_{l=1}^d Q_{jl}B^l \right\rangle_t = \sum_{k,l=1}^d Q_{ik}Q_{jl}\langle B^k, B^l\rangle_t = \sum_{k=1}^d Q_{ik}Q_{jk}t = \delta_{ij}t.$$

(ii) Let $B = (B^1, \cdots, B^d)$ be a d-dimensional $(\mathscr{F}_t)$-Brownian motion. By Itô's formula, for a twice continuously differentiable real-valued function $F(x_1, \cdots, x_d)$ on $\mathbb{R}^d$,

$$F(B_t^1, \cdots, B_t^d) = F(B_0) + \sum_{j=1}^d \int_0^t \frac{\partial F}{\partial x_j}(B_s^1, \cdots, B_s^d)\,\mathrm{d}B_s^j + \frac{1}{2}\sum_{j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_j^2}(B_s^1, \cdots, B_s^d)\,\mathrm{d}s.$$

We denote by Δ the Laplacian operator. Then

$$F(B_t) = F(B_0) + \int_0^t \nabla F(B_s) \cdot \mathrm{d}B_s + \frac{1}{2}\int_0^t \Delta F(B_s)\,\mathrm{d}s.$$

Likewise, for a twice continuously differentiable real-valued function $G(t, x_1, \cdots, x_d)$ on $\mathbb{R}_+ \times \mathbb{R}^d$, we have

$$G(t, B_t) = G(0, B_0) + \int_0^t \nabla_x G(s, B_s) \cdot \mathrm{d}B_s + \int_0^t \left(\frac{\partial G}{\partial t} + \frac{1}{2}\Delta_x G\right)(s, B_s)\,\mathrm{d}s.$$

5.2.2 The Dambis-Dubins-Schwarz Theorem

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 5.13. *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

$$\mathbb{E}\left[N_{t\wedge\tau_n}^2\right] = \mathbb{E}\left[\langle N, N\rangle_{t\wedge\tau_n}\right] \leq \frac{1}{n}, \quad \forall t \in [a, b].$$

On the event $A_{a,b} := \{\langle M, M\rangle_b = \langle M, M\rangle_a\} \subset \{\tau_n \geq b\}$, we have

$$\mathbb{E}\left[N_b^2 \mathbb{1}_{A_{a,b}}\right] = \mathbb{E}\left[N_{b\wedge\tau_n}^2 \mathbb{1}_{A_{a,b}}\right] \leq \mathbb{E}\left[N_{b\wedge\tau_n}^2\right] \leq \frac{1}{n} \quad \overset{n\to\infty}{\Rightarrow} \quad N_b = 0 \ \ a.s. \text{ on } A_{a,b}.$$

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

$$E = \bigcup_{a,b\in\mathbb{Q},\, 0\leq a<b} E_{a,b} \quad \Rightarrow \quad \mathbb{P}(E) = 0.$$

On event $\Omega\backslash E$, 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 $M_a = M_b$. ☐

Theorem 5.14 (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 almost surely,*

$$(M_t)_{t\geq 0} = (\beta_{\langle M,M\rangle_t})_{t\geq 0}.$$

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 $(\mathscr{F}_t)_{t\geq 0}$, 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 $\Omega\backslash E$, 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 3.13, $\beta = (\beta_t)_{t\geq 0}$ is an adapted process with respect to the time-changed filtration $\mathscr{G}_s = \mathscr{F}_{\tau_s}$ for every $s \geq 0$, and $\mathscr{G}_\infty = \mathscr{F}_\infty$. The completeness of $(\mathscr{G}_t)_{t\geq 0}$ also follows from $(\mathscr{F}_t)_{t\geq 0}$. 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 $\Omega\backslash E$. Note that $\langle M, M\rangle_{\tau_s} = \langle M, M\rangle_{\tau_{s+}} = s$ for every $s \geq 0$. By Lemma 5.13, 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 $(\beta_s)_{s>0}$ and $(\beta_s^2 - s)_{s\geq 0}$ are martingales with respect to the filtration $(\mathscr{G}_s)_{s\geq 0}$. For every $n \in \mathbb{N}$, by Theorem 3.71, 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 $0 \leq s < t \leq n$,

$$\begin{aligned}
&\mathbb{E}\left[\beta_t|\mathscr{G}_s\right] = \mathbb{E}\left[M_{\tau_t}^{\tau_n}|\mathscr{F}_{\tau_s}\right] = M_{\tau_s} = \beta_s,\\
&\mathbb{E}\left[\beta_t^2 - t|\mathscr{G}_s\right] = \mathbb{E}\left[(M_{\tau_t}^{\tau_n})^2 - \langle M^{\tau_n}, M^{\tau_n}\rangle_{\tau_s}|\mathscr{F}_{\tau_t}\right] = (M_{\tau_s}^{\tau_n})^2 - \langle M, M\rangle_{\tau_s\wedge\tau_n} = \beta_s^2 - s.
\end{aligned}$$

The case $d = 1$ of Theorem 5.12 implies that β is a $(\mathscr{G}_t)$-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 5.13, 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 $(\mathscr{F}_t)_{t\geq 0}$, but with respect to the time-changed filtration $(\mathscr{G}_t)_{t\geq 0}$.

Corollary 5.15. *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 $(\mathscr{G}_t)$-Brownian motions, where $(\mathscr{G})_{t\geq 0}$ is the time-changed filtration $\mathscr{G}_t = \mathscr{F}_{\tau_t}$.

Since the continuous local martingales M and N are orthogonal, the process MN is also a continuous local martingale. By Proposition 3.75 (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

$$\mathbb{E}\left[\beta_t\gamma_t|\mathscr{G}_s\right] = \mathbb{E}\left[M_{\tau_t}^{\tau_n}N_{\tau_t}^{\tau_n}|\mathscr{F}_{\tau_s}\right] = M_{\tau_s}^{\tau_n}N_{\tau_s}^{\tau_n} = \beta_s\gamma_s.$$

Hence $(\beta_t\gamma_t)_{t\geq 0}$ is a $(\mathscr{G}_t)$-martingale, and again by Proposition 3.75 (vi), we have $\langle \beta, \gamma\rangle = 0$. According to Theorem 5.12, (β, γ) is a two-dimensional $(\mathscr{G}_t)$-Brownian motion, hence β and γ are independent. $\square$

5.2.3 The Burkholder-Davis-Gundy inequality

Now we introduce a useful inequality connecting the local maxima of a continuous local martingale with its quadratic variation.

Theorem 5.16 (Burkholder-Davis-Gundy inequality)**.** *Given a continuous martingale M with $M_0 = 0$, we define the local maxima $M_t^* = \sup_{0\le s\le t}|M_s|$ for all $t \ge 0$. Then for every $p > 0$, there exist constants $C_p, c_p > 0$ depending only on p such that for every stopping time τ,*

$$c_p\mathbb{E}\big[\langle M,M\rangle_\tau^{p/2}\big] \le \mathbb{E}\,|M_\tau^*|^p \le C_p\mathbb{E}\big[\langle M,M\rangle_\tau^{p/2}\big].$$

Remark. By the case $p = 1$ of the Burkholder-Davis-Gundy inequality, if M is a continuous local martingale with $M_0 = 0$ such that $\mathbb{E}[\langle M,M\rangle_\infty^{1/2}] < \infty$, then M is a uniformly integrable martingale.

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 $\tau_n = \{t \ge 0 : |M_t| = n\}$, and the result of $n \to \infty$ follows by monotone convergence theorem.

Step I. We first prove the inequality $\mathbb{E}\,|M_\infty^*|^p \le C_p\mathbb{E}\big[\langle M,M\rangle_\infty^{p/2}\big]$.

Case I: $p \ge 2$**.** Apply Itô's formula to $|x|^p$:

$$|M_t|^p = \int_0^t p|M_s|^{p-1}\mathrm{sgn}(M_s)\,\mathrm{d}M_s + \frac{1}{2}\int_0^t p(p-1)|M_s|^{p-2}\,\mathrm{d}\langle M,M\rangle_s.$$

Note that M is a bounded. Then $M \in \mathbb{H}^2$, and the process $\left(\int_0^t p|M_s|^{p-1}\mathrm{sgn}(M_s)\,\mathrm{d}M_s\right)_{t\ge0}$ is also a martingale in $\mathbb{H}^2$. Consequently, we have

$$\begin{aligned}\mathbb{E}\left[|M_t|^p\right] &\le \frac{p(p-1)}{2}\mathbb{E}\left[\int_0^t |M_s|^{p-2}\,\mathrm{d}\langle M,M\rangle_s\right] \le \frac{p(p-1)}{2}\mathbb{E}\left[|M_t^*|^{p-2}\langle M,M\rangle_t\right]\\ &\le \frac{p(p-1)}{2}\mathbb{E}\left[|M_t^*|^p\right]^{\frac{p-2}{p}}\,\mathbb{E}\left[\langle M,M\rangle_t^{p/2}\right]^{2/p}\end{aligned}$$

On the other hand, the Doob's L^p-inequality Proposition 3.52 gives

$$\mathbb{E}|M_t^*|^p \le \left(\frac{p}{p-1}\right)^p\mathbb{E}|M_t|^p$$

Combining the last two displays, we have

$$\mathbb{E}|M_t^*|^p \le \left(\left(\frac{p}{p-1}\right)^p\frac{p(p-1)}{2}\right)^{p/2}\mathbb{E}\left[\langle M,M\rangle_t^{p/2}\right]. \tag{5.10}$$

Case II: $p < 2$**.** Since $M \in \mathbb{H}^2$, the process $M^2 - \langle M,M\rangle$ is a uniformly integrable martingale, and $\mathbb{E}[(M_\tau)^2] = \mathbb{E}[\langle M,M\rangle_\tau]$ for every stopping time τ. Given $x > 0$, consider the stopping time $\tau_x = \inf\{t \ge 0 : M_t^2 \ge x\}$. If τ is a bounded stopping time, we have

$$\mathbb{P}\left((M_\tau^*)^2 \ge x\right) = \mathbb{P}(\tau_x \le \tau) = \mathbb{P}\left((M_{\tau_x\wedge\tau})^2 \ge x\right) \le \frac{\mathbb{E}\left[(M_{\tau_x\wedge\tau})^2\right]}{x} = \frac{\mathbb{E}\left[\langle M,M\rangle_{\tau_x\wedge\tau}\right]}{x} \le \frac{\mathbb{E}\left[\langle M,M\rangle_\tau\right]}{x}.$$

Consider the stopping time $\sigma_x = \inf\{s \ge 0 : \langle M,M\rangle_s \ge x\}$, so $\{\langle M,M\rangle_t \ge x\} = \{\sigma_x \le t\}$. For every $t \ge 0$, we use the preceding bound with $\tau = \sigma_x \wedge t$:

$$\begin{aligned}\mathbb{P}\left((M_t^*)^2 \ge x\right) &\le \mathbb{P}((M_{\sigma_x\wedge t}^*)^2 \ge x) + \mathbb{P}(\sigma_x \le t) \le x^{-1}\mathbb{E}\left[\langle M,M\rangle_{\sigma_x\wedge t}\right] + \mathbb{P}\left(\langle M,M\rangle_t \ge x\right)\\ &= x^{-1}\mathbb{E}\left[\langle M,M\rangle_t\mathbb{1}_{\{\langle M,M\rangle_t<x\}}\right] + 2\mathbb{P}\left(\langle M,M\rangle_t \ge x\right).\end{aligned}$$

Now we integrate both side with respect to $\frac{p}{2}x^{p/2-1}\,\mathrm{d}x$. For the left-hand side:

$$\int_0^\infty \mathbb{P}\left((M_t^*)^2 \geq x\right)\frac{p}{2}x^{p/2-1}\,\mathrm{d}x = \mathbb{E}\left[\int_0^{(M_t^*)^2}\frac{p}{2}x^{p/2-1}\,\mathrm{d}x\right] = \mathbb{E}\left[(M_t^*)^p\right].$$

Similarly, $\int_0^\infty \mathbb{P}\left(\langle M,M\rangle_t \geq x\right)\frac{p}{2}x^{p/2-1}\,\mathrm{d}x = \mathbb{E}\left[\langle M,M\rangle_t^{p/2}\right]$. Furthermore,

$$\int_0^\infty \mathbb{E}\left[\langle M,M\rangle_t \mathbb{1}_{\{\langle M,M\rangle_t<x\}}\right]\frac{p}{2}x^{p/2-2}\,\mathrm{d}x = \mathbb{E}\left[\langle M,M\rangle_t\int_0^{\langle M,M\rangle_t}\frac{p}{2}x^{p/2-2}\,\mathrm{d}x\right] = \frac{p}{2-p}\mathbb{E}\left[\langle M,M\rangle_t^{p/2}\right].$$

This gives the bound

$$\mathbb{E}\,|M_t^*|^p \leq \frac{4-p}{2-p}\mathbb{E}\left[\langle M,M\rangle_t^{p/2}\right] \tag{5.11}$$

Note both (5.10) and (5.11) hold for $t=\infty$ by monotone convergence theorem.

Step II. Now we prove the inequality $\mathbb{E}\,|M_\infty^*|^p \geq c_p\mathbb{E}\big[\langle M,M\rangle_\infty^{p/2}\big]$. By Itô's lemma,

$$M_t^2 = 2\int_0^t M_s\,dM_s + \langle M,M\rangle_t.$$

For $x,y\geq 0$, we note the inequality

$$|x+y|^p \leq \begin{cases} 2^{p-1}\left(|x|^p+|y|^p\right), & p\geq 1,\\ |x|^p+|y|^p, & 0<p<1.\end{cases}$$

We let $x=M_\infty^2$ and $y=-2\int_0^t M_t\,dM_t$ to get

$$\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right] \leq \max\left\{1,2^{\frac{p}{2}-1}\right\}\left(\mathbb{E}|M_\infty|^p + 2^{\frac{p}{2}}\mathbb{E}\left|\int_0^\infty M_t\,dM_t\right|^{p/2}\right) \leq 2^p\left(\mathbb{E}|M_\infty^*|^p + \mathbb{E}\left|\int_0^\infty M_t\,dM_t\right|^{p/2}\right).$$

We then apply the Burkholder-Davis-Gundy inequality to the continuous local martingale $\int_0^t M_s\,dM_s$ and the Cauchy-Schwartz inequality to get

$$\begin{aligned}\mathbb{E}\left|\int_0^\infty M_t\,dM_t\right|^{p/2} &\leq C_p\mathbb{E}\left[\left(\int_0^\infty M_t^2\,d\langle M,M\rangle_t\right)^{p/4}\right]\\ &\leq C_p\mathbb{E}\left[|M_\infty^*|^{p/2}\langle M,M\rangle_\infty^{p/4}\right] \leq C_p\sqrt{\mathbb{E}|M_\infty^*|^p\,\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right]}.\end{aligned}$$

Therefore

$$\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right] \leq 2^pC_p\left(\mathbb{E}|M_\infty^*|^p + \sqrt{\mathbb{E}|M_\infty^*|^p\,\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right]}\right).$$

We rearrage the above inequality to obtain

$$\left(\sqrt{\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right]} - 2^{p-1}C_p\sqrt{\mathbb{E}|M_\infty^*|^p}\right)^2 \leq \left(2^{2p-2}C_p^2+2^pC_p\right)\mathbb{E}|M_\infty^*|^p,$$

which implies

$$\mathbb{E}\left[\langle M,M\rangle_\infty^{p/2}\right] \leq \left(2^{p-1}C_p + \sqrt{2^{2p-2}C_p^2+2^pC_p}\right)^2\mathbb{E}|M_\infty^*|^p.$$

Then we finish the proof. $\square$

5.3 The Representation of Martingales as Stochastic Integral

On a probability space $(\Omega, \mathscr{F}, P)$, we choose a filtration $(\mathscr{F}_t)_{t\geq 0}$ 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 $(\mathscr{F}_t)_{t\geq 0}$ to be the completed canonical filtration of B. Before formally presenting this result, we first introduce some technical lemma.

Lemma 5.17. *The vector space generated by the random variables of the form*

$$\exp\left(i\sum_{j=1}^n \lambda_j(B_{t_j} - B_{t_{j-1}})\right), \quad \textit{where } \ 0 = t_0 < t_1 < \cdots < t_n \quad \textit{and} \quad \lambda_1, \cdots, \lambda_n \in \mathbb{R}$$

is dense in the space $L^2_{\mathbb{C}}(\Omega, \mathscr{F}_\infty, \mathbb{P})$ of all square-integrable complex-valued $\mathscr{F}_\infty$-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 $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$. Define the complex measure μ on $\mathbb{R}^n$ by

$$\mu(F) = \mathbb{E}\left[Z\mathbb{1}_F(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}})\right], \quad \forall F \in \mathscr{B}(\mathbb{R}^n).$$

Then the Fourier transform of μ satisfies

$$\int_{\mathbb{R}^n} e^{i\lambda^\top x}\,\mu(\mathrm{d}x) = \mathbb{E}\left[Z\exp\left(i\sum_{j=1}^n \lambda_j(B_{t_j} - B_{t_{j-1}})\right)\right] = 0, \quad \forall \lambda \in \mathbb{R}^n.$$

By Lévy's continuity theorem, we have $\mu = 0$. Hence $\mathbb{E}[Z\mathbb{1}_A] = 0$ for all $A \in \sigma(B_{t_1}, \cdots, B_{t_n})$. A monotone class argument shows that $\mathbb{E}[Z\mathbb{1}_A]$ holds for all $A \in \sigma(B_t, t \geq 0)$, and further by completion, for all $A \in \mathscr{F}_\infty$. Hence $Z = 0$ in $L^2_{\mathbb{C}}(\Omega, \mathscr{F}_\infty, \mathbb{P})$. ☐

Theorem 5.18. *For every $Z \in L^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$, there exists a unique progressive process $H \in L^2(B)$ such that*

$$Z = \mathbb{E}[Z] + \int_0^\infty H_s\,\mathrm{d}B_s. \tag{5.12}$$

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. $H \in L^2_{\mathrm{loc}}(B)$) and a constant $C \in \mathbb{R}$ such that

$$M_t = C + \int_0^t H_s\,\mathrm{d}B_s, \quad \forall t \geq 0. \tag{5.13}$$

Proof. Consider the first assertion. If both H and H' satisfy this (5.12), the second moment formula gives

$$0 = \mathbb{E}\left[\left(\int_0^\infty (H_s - H'_s)\,\mathrm{d}B_s\right)^2\right] = \mathbb{E}\left[\int_0^\infty (H_s - H'_s)^2\,\mathrm{d}s\right] \quad \Rightarrow \quad H = H' \ \textit{ in } \ L^2(B).$$

This shows uniqueness. For existence, let $\mathscr{H}$ be the vector space of all $Z \in L^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$ for which there exists an associated $H \in L^2(B)$ satisfying (5.12). By Proposition 5.11, for any step function $h = \sum_{j=1}^n \lambda_j \mathbb{1}_{(t_{j-1}, t_j]}$, where $0 = t_0 < t_1 < \cdots < t_n$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$, we have

$$\exp\bigg(i\sum_{j=1}^n \lambda_j(B_{t_j} - B_{t_{j-1}}) + \frac{1}{2}\sum_{j=1}^n \lambda_j^2(t_j - t_{j-1})\bigg) = \mathscr{E}(ih\cdot B)_t = 1 + \mathrm{i}\int_0^t \mathscr{E}(ih\cdot B)_s h(s)\,\mathrm{d}B_s$$

Since $\mathscr{E}(ih\cdot B)_s h(s)$ 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 $\mathscr{H}$. According to Lemma 5.17, $\mathscr{H}$ contains a dense subset of $L^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$.

To prove the first assertion, it remains to show that $\mathscr{H}$ is closed, which implies $\mathscr{H} = L^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$. We assume that $Z_n \in \mathscr{H}$ 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

$$\begin{aligned}\|H^{(n)} - H^{(m)}\|_{L^2(B)} &= \mathbb{E}\left[\int_0^\infty (H_s^{(n)} - H_s^{(m)})^2\,\mathrm{d}s\right] = \mathbb{E}\left[\left(\int_0^\infty (H_s^{(n)} - H_s^{(m)})\,\mathrm{d}B_s\right)^2\right]\\ &= \mathbb{E}\left[(Z_n - Z_m - \mathbb{E}[Z_n] + \mathbb{E}[Z_m])^2\right] \le \|Z_n - Z_m\|_2^2.\end{aligned}$$

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, $Z_n = \mathbb{E}[Z_n] + \int_0^\infty H_s^{(n)}\,\mathrm{d}B_s$ converges to $Z = \mathbb{E}[Z] + \int_0^\infty H_s\,\mathrm{d}B_s$ 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 $M_\infty \in L^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$. Since $(\int_0^t H_s\,\mathrm{d}B_s)_{s\ge 0}$ 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

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty H_s\,\mathrm{d}B_s \quad \Rightarrow \quad M_t = \mathbb{E}[M_\infty|\mathscr{F}_t] = \mathbb{E}[M_\infty] + \int_0^t H_s\,\mathrm{d}B_s, \quad \forall t \ge 0.$$

Finally, if M is a continuous local martingale, we have $M_0 = C$ *a.s.* for some constant $C \in \mathbb{R}$ because $\mathscr{F}_0$ is $\mathbb{P}$-trivial. We then choose stopping times $\tau_n = \{t \ge 0 : |M_t| \ge n\}$ so that M^{τ_n} are bounded martingales. According to the preceding result, there exists $H^{(n)} \in L^2(B)$ such that

$$M_{t\wedge\tau_n} = C + \int_0^t H_s^{(n)}\,\mathrm{d}B_s, \quad \forall t \ge 0.$$

By uniqueness of the progressive process in this representation, we have $H^{(m)} = H^{(n)}\mathbb{1}_{[0,\tau_m]}$ in $L^2(B)$ for all $n > m$. Consequently, we can find $H \in L^2(B)$ such that $H^{(n)} = H\mathbb{1}_{[0,\tau_n]}$ in $L^2(B)$ for all $n \in \mathbb{N}$, and the representation formula (5.13) follows by letting $n \to \infty$. The uniqueness argument is similar. $\square$

Remark. In this theorem, we do not require the L^2-bounded martingale M to be continuous. Next we discuss some consequence of this representation theorem.

Proposition 5.19. *The filtration $(\mathscr{F}_t)_{t\ge 0}$ is right-continuous.*

Proof. Let Z be a bounded $\mathscr{F}_{t+}$-measurable random variable. By Theorem 5.18, there exists $H \in L^2(B)$ such that $Z = \mathbb{E}[Z] + \int_0^\infty H_s\,\mathrm{d}B_s$. Given $\epsilon > 0$, Z is $\mathscr{F}_{t+\epsilon}$ measurable. By continuity of $H\cdot B$, we have

$$Z = \mathbb{E}[Z|\mathscr{F}_{t+\epsilon}] = \mathbb{E}[Z] + \int_0^{t+\epsilon} H_s\,\mathrm{d}B_s \overset{a.s.}{\to} \mathbb{E}[Z] + \int_0^t H_s\,\mathrm{d}B_s, \quad \epsilon \downarrow 0.$$

As a result, $Z = \mathbb{E}[Z] + \int_0^t H_s\,\mathrm{d}B_s$ *a.s.*. By completeness of the filtration $(\mathscr{F}_t)_{t\ge 0}$, Z is also $\mathscr{F}_t$-measurable. Hence for all $A \in \mathscr{F}_{t+}$, the variable $Z = \mathbb{1}_A$ is $\mathscr{F}_t$ measurable, and $A \in \mathscr{F}_t$. Therefore $\mathscr{F}_{t+} = \mathscr{F}_t$. $\square$

Remark. We also can define the left limit of $(\mathscr{F}_t)_{t\geq 0}$ at any $t>0$:

$$\mathscr{F}_{t-}=\sigma\left(\bigcup_{0\leq s<t}\mathscr{F}_s\right).$$

A similar argument to Proposition (5.19) implies that $\mathscr{F}_{t-}=\mathscr{F}_t$. Namely, the completed filtration $(\mathscr{F}_t)_{t\geq 0}$ generated by a Brownian motion B is also left-continuous.

Proposition 5.20. *All martingales with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$ have an a.s. modification with continuous sample paths.*

Proof. Following Theorem 5.18, 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 $M_t=\mathbb{E}[M_\infty|\mathscr{F}_t]$ for all $t\geq 0$. By Theorem 3.56 and Proposition 5.19, 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 $M_t^{(n)}=\mathbb{E}[M_\infty^{(n)}|\mathscr{F}_t]$. These martingales are then bounded in L^2, hence continuous. Furthermore, the Doob's maximal inequality Proposition 3.52 (i) gives

$$\mathbb{P}\left(\sup_{t\geq 0}|M_t^{(n)}-M_t|\geq\lambda\right)\leq\frac{3}{\lambda}\mathbb{E}\left|M_\infty^{(n)}-M_\infty\right|,\quad\forall\lambda>0.$$

We choose a subsequence n_k such that

$$\mathbb{P}\left(\sup_{t\geq 0}|M_t^{(n_k)}-M_t|>\frac{1}{2^k}\right)\leq\frac{1}{2^k},\ \forall k\in\mathbb{N},$$

which implies

$$\mathbb{P}\left(\sup_{t\geq 0}|M_t^{(n_k)}-M_t|>\frac{1}{2^k}\ \text{ for infinitely many } k\right)=0.$$

Here we use Borel-Cantelli lemma. Consequently, $\sup_{t\geq 0}|M_t^{(n_k)}-M_t|\to 0$ *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):

$$\mathrm{d}y_t = b(t, y_t)\,\mathrm{d}t \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s)\,\mathrm{d}s$$

To model a noisy system, we simply introduce a random perturbation of the term $\sigma\,\mathrm{d}B_t$, 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

$$\mathrm{d}y_t = b(t, y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s)\,ds + \sigma B_t.$$

We generalize the above equation by allowing σ depending on the time t and the state y_t:

$$\mathrm{d}y_t = b(t, y_t)\,\mathrm{d}t + \sigma(t, y_t)\,\mathrm{d}B_t \quad \Leftrightarrow \quad y_t = y_0 + \int_0^t b(s, y_s)\,ds + \int_0^t \sigma(s, y_s)\,\mathrm{d}B_s.$$

This gives rise to the following definition of stochastic differential equation.

Definition 5.21 (Stochastic Differential Equation, SDE)**.** Let $\sigma = (\sigma_{ij})_{i\in[p],j\in[q]} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^{p\times q}$ and $b = (b_i)_{i\in[p]} : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$ be locally bounded measurable functions. A *solution of the stochastic equation* $E(\sigma, b)$, which is given by

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t,$$

consists of:

- A filtered probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a complete filtration $(\mathscr{F})_{t\geq 0}$;
- A q-dimensional $(\mathscr{F}_t)$-Brownian motion $B = (B^1, \cdots, B^q)$ starting from 0;
- An $(\mathscr{F}_t)$-adapted and continuous process $X = (X^1, \cdots, X^p)$ taking values in $\mathbb{R}^p$, such that

$$X_t = X_0 + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s + \int_0^t b(s, X_s)\,\mathrm{d}s \quad \overset{def}{\Leftrightarrow} \quad X_t^i = X_0^i + \sum_{j=1}^q \int_0^t \sigma_{ij}(s, X_s)\,\mathrm{d}B_s^j + \int_0^t b_i(s, X_s)\,\mathrm{d}s.$$

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 5.22. 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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and the $(\mathscr{F}_t)$-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 an 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

$$B_t = \int_0^t \operatorname{sgn}(\beta_s)\,\mathrm{d}\beta_s, \quad \textit{where } \operatorname{sgn}(x) = \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0]}(x).$$

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 5.12. Furthermore, by associativity of stochastic integrals, one have

$$\int_0^t \operatorname{sgn}(\beta_s)\,\mathrm{d}B_s = \int_0^t \operatorname{sgn}(\beta_s)^2\,\mathrm{d}\beta_s = \beta_t - \beta_0, \quad \forall t \geq 0.$$

Therefore, $(\beta_t)_{t\geq 0}$ solves the following stochastic differential equations:

$$\mathrm{d}X_t = \operatorname{sgn}(X_t)\,\mathrm{d}B_s, \quad X_0 = x.$$

For this equation, weak existence holds, and again by Theorem 5.12 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.

5.4.1 Existence Theory for SDEs with Lipschitz Coefficients

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 5.23 (Gronwall's lemma)**.** *Let $T > 0$ and let $g : [0,T] \to \mathbb{R}_+$ be a bounded measurable function. If there exist two constants $a \geq 0$ and $b \geq 0$ such that*

$$g(t) \leq a + b\int_0^t g(s)\,\mathrm{d}s, \quad \forall t \in [0,T],$$

then we have $g(t) \leq ae^{bt}$ for all $t \in [0,T]$.

Proof. A simple recursion on g gives

$$\begin{aligned} g(t) &\leq a + a(bt) + b^2\int_0^t \left(\int_0^{s_1} g(s_2)\,\mathrm{d}s_2\right)\mathrm{d}s_1 \leq \cdots \\ &\leq a + a(bt) + a\frac{(bt)^2}{2} + \cdots + a\frac{(bt)^n}{n!} + b^{n+1}\int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \cdots \int_0^{s_{n+1}} \mathrm{d}s_{n+1} g(s_{n+1}). \end{aligned}$$

Since g is bounded, we let $0 \leq g(t) \leq M$ for all $t \in [0,T]$. Then

$$g(t) \leq a\sum_{k=0}^n \frac{(bt)^n}{n!} + \frac{M(bt)^{n+1}}{(n+1)!}.$$

Letting $t \to \infty$ produces the desired result. $\square$

The following theorem gives the existence of a solution of SDE in the Lipschitz case.

Theorem 5.24. *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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, every $(\mathscr{F}_t)$-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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and the $(\mathscr{F}_t)$-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 $t \geq 0$,

$$X_{t\wedge\tau} = X_0 + \int_0^{t\wedge\tau} \sigma(s, X_s)\,\mathrm{d}B_s + \int_0^{t\wedge\tau} b(s, X_s)\,\mathrm{d}s,$$

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):

$$\begin{aligned}
\mathbb{E}\left[(X_{t\wedge\tau} - Y_{t\wedge\tau})^2\right] &\leq 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau} (\sigma(s, X_s) - \sigma(s, Y_s))\,\mathrm{d}B_s\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau} (b(s, X_s) - b(s, Y_s))\,\mathrm{d}s\right)^2\right] \\
&\leq 2\mathbb{E}\left[\int_0^{t\wedge\tau} (\sigma(s, X_s) - \sigma(s, Y_s))^2\,\mathrm{d}s\right] + 2\mathbb{E}\left[T\int_0^{t\wedge\tau} (b(s, X_s) - b(s, Y_s))^2\,\mathrm{d}s\right] \\
&\leq 2(1+T)L^2\mathbb{E}\left[\int_0^{t\wedge\tau} (X_s - Y_s)^2\,\mathrm{d}s\right] \\
&\leq 2(1+T)L^2\mathbb{E}\left[\int_0^{t} (X_{s\wedge\tau} - Y_{s\wedge\tau})^2\,\mathrm{d}s\right].
\end{aligned}$$

By Lemma 5.23, we have $\mathbb{E}\left[(X_{t\wedge\tau} - Y_{t\wedge\tau})^2\right] = 0$, and $X_{t\wedge\tau} = Y_{t\wedge\tau}$ *a.s.* for all $t \in [0, T]$. Let $M \to \infty$ and $T \to \infty$, 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:

$$\begin{aligned}
X_t^0 = x, \quad X_t^1 &= x + \int_0^t \sigma(s, x)\,\mathrm{d}B_s + \int_0^t b(s, x)\,\mathrm{d}s, \\
X_t^n &= x + \int_0^t \sigma(s, X_s^{n-1})\,\mathrm{d}B_s + \int_0^t b(s, X_s^{n-1})\,\mathrm{d}s, \quad n \in \mathbb{N}.
\end{aligned}$$

Clearly, X^n is 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 $\left(\int_0^t \sigma(s, x)\,\mathrm{d}B_s\right)_{t\geq 0}$ is a continuous local martingale with finite quadratic variation, hence is a martingale by Corollary 3.72. Consequently, we can use Doob's L^2-inequality [Proposition 3.52 (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

$$X_t^{n+1} - X_t^n = \int_0^t \left(\sigma(s, X_s^n) - \sigma(s, X_s^{n-1})\right)\mathrm{d}B_s + \int_0^t \left(b(s, X_s^n) - b(s, X_s^{n-1})\right)\mathrm{d}s.$$

We then use the Burkholder-Davis-Gundy inequality [Theorem 5.16]:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\in[0,t]}(X_s^{n+1}-X_s^n)^2\right]\\
&\leq 2\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s\left(\sigma(r,X_r^n)-\sigma(r,X_r^{n-1})\right)\mathrm{d}B_r\right|^2\right]+2\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s\left(b(r,X_r^n)-b(r,X_r^{n-1})\right)\mathrm{d}r\right|^2\right]\\
&\leq 2C_2\mathbb{E}\left[\int_0^t\left(\sigma(r,X_r^n)-\sigma(r,X_r^{n-1})\right)^2\mathrm{d}r\right]+2\mathbb{E}\left[T\int_0^t\left(b(r,X_r^n)-b(r,X_r^{n-1})\right)^2\mathrm{d}r\right]\\
&\leq 2(C_2+T)L^2\mathbb{E}\left[\int_0^t(X_r^n-X_r^{n-1})^2\,\mathrm{d}r\right]
\end{aligned}$$

Consequently, we have

$$g_{n+1}(t)\leq 2(C_2+T)L^2\int_0^t g_n(s)\,\mathrm{d}s,\quad \forall t\in[0,T].$$

Since $g_1(t)\leq C_T$ for all $t\in[0,T]$, an induction gives

$$g_n(t)\leq C_T\frac{(2(C_2+T)L^2t)^{n-1}}{(n-1)!},\quad \forall t\in[0,T].$$

Hence we have

$$\sum_{n=1}^{\infty}g_n(T)^{1/2}<\infty\quad\Rightarrow\quad\sum_{n=1}^{\infty}\sup_{t\in[0,T]}\left|X_t^n-X_t^{n-1}\right|<\infty\ \ a.s..$$

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:

$$\begin{aligned}
\lim_{n\to\infty}\left(\int_0^t\sigma(s,X_s)\,\mathrm{d}B_s-\int_0^t\sigma(s,X_s^n)\,\mathrm{d}B_s\right)&=0,\\
\lim_{n\to\infty}\left(\int_0^t b(s,X_s)\,\mathrm{d}s-\int_0^t b(s,X_s^n)\,\mathrm{d}s\right)&=0,
\end{aligned}$$

where we use $\sum_{n=1}^{\infty}\sup_{s\in[0,t]}\left|X_s^n-X_s^{n-1}\right|$ 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 $T\to\infty$, 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 5.25. *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 $F_x:C(\mathbb{R}_+,\mathbb{R}^q)\to C(\mathbb{R}_+,\mathbb{R}^p)$ 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 $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$ and of the $(\mathscr{F}_t)$-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 $\mathscr{F}_0$-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 $\mathscr{G}_t$ be the σ-algebra generated by projection mappings $\{\pi_s : 0 \le s \le t\}$ and all W-negligible sets in $C(\mathbb{R}_+, \mathbb{R})$, so $(\mathscr{G}_t)_{t\ge 0}$ is a complete filtration. By Theorem 5.24, with the filtered probability space $(C(\mathbb{R}_+, \mathbb{R}), \mathscr{G}_\infty, (\mathscr{G}_t)_{t\ge0}, W)$ 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 $X^x = (X^x_t)_{t\ge0}$ 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 $q \ge 2$, and prove that there exists a constant $C^*_q > 0$ depending only on q, such that for all $x, y \in \mathbb{R}$,

$$\mathbb{E}\left[d(X^x, X^y)^q\right] \le C^*_p |x-y|^q. \tag{5.14}$$

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 $(\widetilde{X}^x)_{x\in\mathbb{R}}$ of $(X^x)_{x\in\mathbb{R}}$ with continuous sample paths. Define $F_x(\mathbf{w}) = \widetilde{X}^x(\mathbf{w}) = (\widetilde{X}^x_t(\mathbf{w}))_{t\ge0}$. Then $F_x : (C(\mathbb{R}_+, \mathbb{R}), \mathscr{G}_\infty) \to (C(\mathbb{R}_+, \mathbb{R}), \mathscr{G}_\infty)$ is a measurable mapping with property (ii).

To this end, we define the stopping time

$$\tau_n = \inf\left\{t \ge 0 : |X^x_t| \vee |X^y_t| \ge n\right\}, \quad n = 1, 2, \cdots.$$

We fix some $T \ge 1$. For every $t \in [0, T]$, we apply Jensen's inequality, Burkholder-Davis-Gundy inequality [Theorem 5.16], Hölder's inequality and Lipschitz property as follows:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\in[0,t]} |X^y_{s\wedge\tau_n} - X^y_{s\wedge\tau_n}|^q\right] \\
&\le 3^{q-1}\left(|x-y|^q + \mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau_n} (\sigma(r, X^x_r) - \sigma(r, X^y_r))\,\mathrm{d}B_r\right|^q + \sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau_n} (b(r, X^x_r) - b(r, X^y_r))\,\mathrm{d}r\right|^q\right]\right) \\
&\le 3^{q-1}\left(|x-y|^q + C_q\mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} (\sigma(r, X^x_r) - \sigma(r, X^y_r))^2\,\mathrm{d}r\right)^{q/2}\right] + \mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} |b(r, X^x_r) - b(r, X^y_r)|\,\mathrm{d}r\right)^q\right]\right) \\
&\le 3^{q-1}\Bigg(|x-y|^q + C_q t^{\frac{q}{2}-1}\mathbb{E}\left[\int_0^t \left|\sigma(r\wedge\tau_n, X^x_{r\wedge\tau_n}) - \sigma(r\wedge\tau_n, X^y_{r\wedge\tau_n})\right|^q\,\mathrm{d}r\right] \\
&\qquad + t^{q-1}\mathbb{E}\left[\int_0^t \left|b(r\wedge\tau_n, X^x_{r\wedge\tau_n}) - b(r\wedge\tau_n, X^y_{r\wedge\tau_n})\right|^q\,\mathrm{d}r\right]\Bigg) \\
&\le 3^{q-1}\left(|x-y|^q + K^q T^{\frac{q}{2}-1}\left(C_q + T^{q/2}\right)\mathbb{E}\left[\int_0^t \left|X^x_{r\wedge\tau_n} - X^y_{r\wedge\tau_n}\right|^q\,\mathrm{d}r\right]\right)
\end{aligned}$$

We let $C'_q = 3^{q-1}K^q(1 + C_q)$. Then we obtain the following estimate by using Gronwall's lemma [Lemma 5.23] on the bounded function $t \mapsto \mathbb{E}\left[\sup_{s\in[0,t]} |X^y_{s\wedge\tau_n} - X^y_{s\wedge\tau_n}|^q\right]$:

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X^y_{s\wedge\tau_n} - X^y_{s\wedge\tau_n}|^q\right] \le C'_q|x-y|^q \exp\left(C'_q T^{q-1} t\right), \quad \forall t \in [0, T].$$

A monotone convergence argument follows by letting $n \to \infty$:

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X^y_s - X^y_s|^q\right] \le C'_q|x-y|^q \exp\left(C'_q t^q\right), \quad \forall t \ge 0.$$

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

$$\mathbb{E}\left[d(X_s^x, X_s^y)^q\right] \leq \left(\sum_{k=1}^{\infty} \alpha_k\right)^{q-1} \sum_{k=1}^{\infty} \alpha_k \mathbb{E}\left[\sup_{s\in[0,k]} |X_s^x - X_s^y|^q\right] \leq C_q^* |x-y|^q,$$

where $C_q^* = C_q' \left(\sum_{k=1}^{\infty} \alpha_k\right)^{q-1} \left(\sum_{k=1}^{\infty} \alpha_k \exp(C_q' k^q)\right)$. This complete the proof of (5.14). For the assertion (i), we point out that for any $t \geq 0$, the mapping $\mathbf{w} \mapsto F_x(\mathbf{w})_t = \widetilde{X}_t^x \overset{a.s.}{=} X_t^x$ is $\mathscr{G}_t$-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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and an $(\mathscr{F}_t)$-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 $(\mathscr{F}_t)_{t\geq 0}$ 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 $X^x = \widetilde{X}^x$ W-*a.s.*, for all $t \geq 0$, we have

$$F_x(\mathbf{w})_t = x + \int_0^t \sigma(s, F_x(\mathbf{w})_s)\,\mathrm{d}\mathbf{w}(s) + \int_0^t b(s, F_x(\mathbf{w})_s)\,\mathrm{d}s, \quad \textit{for } W\textit{-a.s. } \mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}).$$

By Proposition 5.9, we have the following approximation:

$$\int_0^t \sigma(s, F_x(\mathbf{w})_s)\,\mathrm{d}\mathbf{w}(s) = \lim_{n\to\infty} \sum_{j=1}^{2^n} \sigma\left(\frac{(j-1)t}{2^n}, F_x(\mathbf{w})_{\frac{(j-1)t}{2^n}}\right)\left(\mathbf{w}\left(\frac{jt}{2^n}\right) - \mathbf{w}\left(\frac{(j-1)t}{2^n}\right)\right) \tag{5.15}$$

in probability $W(\mathrm{d}\mathbf{w})$. Since W is the law of B, by Proposition 5.9, we have

$$\begin{aligned} F_x(B)_s &= x + \lim_{n\to\infty} \sum_{j=1}^{2^n} \sigma\left(\frac{(j-1)t}{2^n}, F_x(B)_{\frac{(j-1)t}{2^n}}\right)\left(B_{\frac{jt}{2^n}} - B_{\frac{(j-1)t}{2^n}}\right) + \int_0^t b(s, F_x(B)_s)\,\mathrm{d}s \\ &= x + \int_0^t \sigma(s, F_x(B)_s)\,\mathrm{d}B_s + \int_0^t b(s, F_x(B)_s)\,\mathrm{d}s, \quad a.s., \end{aligned}$$

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 $\mathscr{F}_t$-measurable for any fixed $\omega \in \Omega$. Then $(x, \omega) \mapsto F_x(B)_t(\omega)$ is $\mathscr{B}(\mathbb{R}) \otimes \mathscr{F}_t$-measurable according to a similar procedure in Proposition 3.10. If U is a $\mathscr{F}_0$-measurable random variable, then $F_U(B)_t$ is a composition of $\omega \mapsto (U(\omega), \omega)$ and $(x, \omega) \to F_x(B)_t(\omega)$, hence is $\mathscr{F}_t$-measurable.

Let $H(x, \mathbf{w})_t = F_x(\mathbf{w})_t - x - \int_0^t b(s, F_x(\mathbf{w})_s)\,\mathrm{d}s$. We use the convergence result (5.15) in probability W. Since $B \sim W$, and U is a $\mathscr{F}_0$-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 $\int_0^t \sigma(s, F_U(B)_s)\,\mathrm{d}B_s$. Hence

$$F_U(B)_t - U - \int_0^t b(s, F_U(B))\,\mathrm{d}s = \int_0^t \sigma(s, F_U(B)_s)\,\mathrm{d}B_s.$$

Consequently, $F_U(B) = (F_U(B))_{t\geq 0}$ solves the SDE $E(\sigma, b)$ with initial value U. $\square$

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 P and Q on the same measurable space $(\Omega, \mathscr{F})$. To avoid confusion, we write $\mathbb{E}_P$ and $\mathbb{E}_Q$ for the expectation under P and Q, respectively. Unless otherwise specified, our notions of semimartingales refer to the underlying probability measure P. We will point it out explicitly when consider these notions under Q.

5.5.1 Girsanov's Theorem

Throughout this subsection, we assume $(\mathscr{F}_t)_{t\geq 0}$ is a complete and right continuous filtration. Most of the time we may assume that P and Q are mutually absolutely continuous, hence the filtration $(\mathscr{F}_t)_{t\geq 0}$, being complete with respect to P, is also complete with respect to Q.

Proposition 5.26. *Assume that Q is a probability measure on $(\Omega, \mathscr{F}_\infty)$ which is absolutely continuous with respect to P on $\mathscr{F}_t$ for every $t \geq 0$. Let D_t be the Radon-Nikodym derivative of Q with respect to P on $\mathscr{F}_t$:*

$$D_t = \frac{\mathrm{d}Q}{\mathrm{d}P}\bigg|_{\mathscr{F}_t}, \quad t \in \mathbb{R}_+.$$

Then $D = (D_t)_{t\geq 0}$ is a P-martingale, and D has a càdlàg modification thanks to Theorem 3.56. Furthermore, the following two assertions are equivalent:

(i) the martingale D is uniformly integrable;

(ii) $Q \ll P$ on $\mathscr{F}_\infty$.

Proof. We fix $t > s \geq 0$. Then for all $A \in \mathscr{F}_s$, we have $Q(A) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A D_s]$. For the martingale property,

$$\mathbb{E}_P\left[D_s \mathbb{1}_A\right] = \mathbb{E}_{\mathbb{P}}\left[D_t \mathbb{1}_A\right] = Q(A), \quad \forall A \in \mathscr{F}_s \subset \mathscr{F}_t.$$

Hence we have $D_s = \mathbb{E}[D_t|\mathscr{F}_s]$, 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 $\mathbb{E}_P[Z\mathbb{1}_A] = Q(A)$ for all $A \in \mathscr{F}_\infty$, hence $Q \ll P$ on $\mathscr{F}_\infty$. If (ii) holds, let D_∞ be the Radon-Nikodym derivative of Q with respect to P on $\mathscr{F}_t$. Then $\mathbb{E}_P[D_\infty|\mathscr{F}_t] = D_t$ for all $t \geq 0$, which implies uniform integrability. $\square$

In the sequel, we assume that the martingale $D = (D_t)_{t\geq 0}$ has càdlàg sample paths.

Proposition 5.27. *Under the assumption of the preceding proposition, for every stopping time τ, we have $\mathrm{d}Q = D_\tau \mathrm{d}P$ on $\mathscr{F}_\tau \cap \{\tau < \infty\}$. Furthermore, if $Q \ll P$ on $\mathscr{F}_\infty$, we have*

$$D_\tau = \frac{\mathrm{d}Q}{\mathrm{d}P}\bigg|_{\mathscr{F}_\tau}$$

Proof. For the uniform integrable case where $Q \ll P$ on $\mathscr{F}_\infty$, by optional stopping theorem [Theorem 3.61],

$$Q(A) = \mathbb{E}_P[D_\infty \mathbb{1}_A] = \mathbb{E}_P\left[\mathbb{E}_P[D_\infty|\mathscr{F}_\tau]\mathbb{1}_A\right] = \mathbb{E}_P\left[D_\tau \mathbb{1}_A\right], \quad \forall A \in \mathscr{F}_\tau.$$

Since D_τ is $\mathscr{F}_\tau$-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

$$Q(A \cap \{\tau \leq t\}) = \mathbb{E}_P\left[D_{\tau\wedge t}\mathbb{1}_{A\cap\{\tau\leq t\}}\right] = \mathbb{E}_P[D_\tau \mathbb{1}_{A\cap\{\tau\leq t\}}], \quad \forall A \in \mathscr{F}_\tau.$$

Letting t tends to infinity concludes the proof. $\square$

Remark. Let $X = (X_t)_{\geq 0}$ be an adaptive process with càdlàg sample paths. If XD is a P-local martingale, then X is a Q-local martingale. To see this, we take a stopping time τ. Then for any $s \geq 0$ and any $A \in \mathscr{F}_s$, $A \cap \{\tau > s\} \in \mathscr{F}_s \cap \mathscr{F}_\tau = \mathscr{F}_{\tau \wedge s}$. Let $t > s \geq 0$. If the stopped process $(XD)^\tau$ is a P-martingale, then

$$\begin{aligned}
\mathbb{E}_Q[X_{t\wedge\tau}\mathbb{1}_A] &= \mathbb{E}_Q[X_{t\wedge\tau}\mathbb{1}_{A\cap\{\tau\leq s\}}] + \mathbb{E}_Q[X_{t\wedge\tau}\mathbb{1}_{A\cap\{\tau>s\}}] \\
&= \mathbb{E}_Q[X_{s\wedge\tau}\mathbb{1}_{A\cap\{\tau\leq s\}}] + \mathbb{E}_P[X_{t\wedge\tau}D_{t\wedge\tau}\mathbb{1}_{A\cap\{\tau>s\}}] \\
&= \mathbb{E}_Q[X_{s\wedge\tau}\mathbb{1}_{A\cap\{\tau\leq s\}}] + \mathbb{E}_P[X_{s\wedge\tau}D_{s\wedge\tau}\mathbb{1}_{A\cap\{\tau>s\}}] = \mathbb{E}_Q[X_{s\wedge\tau}\mathbb{1}_A].
\end{aligned}$$

Hence X^τ is a Q-martingale. In addition, a sequence of stopping times increasing P-a.s. to ∞ also increases Q-a.s. to ∞. Consequently, if XD is a P-local martingale, then X is a Q-local martingale.

Proposition 5.28. *Under the preceding assumption, the martingale D is Q-a.s. strictly positive, i.e.*

$$\inf_{t\geq 0} D_t > 0, \qquad Q\text{-a.s.}.$$

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 $\mathscr{F}_{\tau_n}$-measurable, and $D_{\tau_n} \leq 1/n$ on $\{\tau_n < \infty\}$ by right-continuity of D. Hence

$$Q(\tau_n < \infty) = \mathbb{E}_P\left[D_{\tau_n}\mathbb{1}_{\{\tau_n<\infty\}}\right] \leq \frac{1}{n},$$

which implies

$$Q\left(\bigcap_{n=1}^{\infty}\{\tau_n < \infty\}\right) = 0.$$

Then Q-*a.s.* there exists $n \in \mathbb{N}$ such that $\tau_n = \infty$. This complete the proof. ☐

Remark. If we further assume that P and Q are mutually absolutely continuous, then the martingale D is also P-*a.s.* strictly positive.

Theorem 5.29 (Girsanov)**.** *Suppose the assumption of the preceding proposition holds, and assume that the martingale $D = (D_t)_{t\geq 0}$ is continuous. If $M = (M_t)_{t\geq 0}$ is a continuous P-local martingale, then*

$$\widetilde{M} = M - D^{-1} \cdot \langle M, D\rangle$$

is a continuous 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 5.28, the process $D^{-1} \cdot \langle M, D\rangle$ is 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

$$\begin{aligned}
(\widetilde{M}D)_t &= \widetilde{M}_0D_0 + \int_0^t \widetilde{M}_s\,\mathrm{d}D_s + \int_0^t D_s\,\mathrm{d}\widetilde{M}_s + \langle\widetilde{M}, D\rangle_t \\
&= \widetilde{M}_0D_0 + \int_0^t \widetilde{M}_s\,\mathrm{d}D_s + \int_0^t D_s\,\mathrm{d}M_s - \langle M, D\rangle_t + \langle\widetilde{M}, D\rangle_t \\
&= \widetilde{M}_0D_0 + \int_0^t \widetilde{M}_s\,\mathrm{d}D_s + \int_0^t D_s\,\mathrm{d}M_s.
\end{aligned}$$

Consequently, the process $\widetilde{M}D$ is a continuous P-local martingale. By the Remark under Proposition 5.27, the process $\widetilde{M}$ is a continuous Q-local martingale. The last assertion holds because the bracket of a finite variation process and a semimartingale vanishes. ☐

Proposition 5.30. *If D is a continuous local martingale taking strictly positive values. There exists a unique continuous local martingale L such that*

$$D_t = \mathscr{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L, L\rangle_t\right).$$

Moreover, L is given by the formula

$$L_t = \log D_0 + \int_0^t D_s^{-1}\,\mathrm{d}D_s. \tag{5.16}$$

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 3.68]. To show the second assertion, use Itô's formula to the process $\log D$:

$$\log D_t = \log D_0 + \int_0^t D_s^{-1}\,\mathrm{d}D_s - \frac{1}{2}\int_0^t D_s^{-2}\,\mathrm{d}\langle D, D\rangle_s = L_t - \frac{1}{2}\langle L, L\rangle_t,$$

where L is given in (5.16). □

We then have another form of Girsanov's theorem.

Theorem 5.31 (Girsanov)**.** *Suppose the assumption of the preceding proposition holds, and assume that the martingale $D = (D_t)_{t\geq 0}$ is continuous. If $M = (M_t)_{t\geq 0}$ is a continuous P-local martingale, then*

$$\widetilde{M} = M - D^{-1}\cdot\langle M, D\rangle = M - \langle M, L\rangle.$$

is a continuous Q-local martingale, where L is given in (5.16)*. Moreover, $D^{-1} = \mathscr{E}(-\widetilde{L})$.*

Proof. The first identity immediately follows from associativity of stochastic integral. For the second assertion, we use the identity $\widetilde{L} = L - \langle L, L\rangle$:

$$\mathscr{E}(-\widetilde{L})_t = \exp\left(-\widetilde{L}_t - \frac{1}{2}\langle\widetilde{L}, \widetilde{L}\rangle_t\right) = \exp\left(-L_t + \frac{1}{2}\langle L, L\rangle_t\right) = \mathscr{E}(L)_t^{-1}.$$

This proves the second assertion. □

Remark. (i) According to Theorem 5.30, if P and Q are mutually absolutely continuous on $\mathscr{F}_\infty$, then the role of P and Q can be exchanged by replacing L with $-\widetilde{L}$.

(ii) In particular, if $M = B$ is an $(\mathscr{F}_t)$-Brownian motion under P, then $\widetilde{B} = B - \langle B, L\rangle$ is a Q-continuous local martingale, and $\langle\widetilde{B}, \widetilde{B}\rangle_t = \langle B, B\rangle_t = t$. By Lévy's characterization of multi-dimensional Brownian motions [Theorem 5.11], $\widetilde{B}$ is an $(\mathscr{F}_t)$-Brownian motion under Q.

Theorem 5.32. *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) $\mathscr{E}(L)$ is a uniformly integrable martingale.

Then (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii).

Proof. (i) $\Rightarrow$ (ii): The process $\mathscr{E}(L)$ is a nonnegative continuous local martingale, hence is a supermartingale by Proposition 3.67 (i). By Fatou's lemma,

$$\mathbb{E}[\mathscr{E}(L)_\infty] \leq \lim_{t\to\infty}\mathbb{E}[\mathscr{E}(L)_t] \leq \mathbb{E}[\mathscr{E}(L)_0] = 1.$$

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 3.60. By Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\exp\left(\frac{1}{2}L_\infty\right)\right] = \mathbb{E}\left[\mathscr{E}(L)_\infty^{1/2} e^{\frac{1}{4}\langle L,L\rangle_\infty}\right] \le \sqrt{\mathbb{E}\left[\mathscr{E}(L)_\infty\right]}\sqrt{\mathbb{E}\left[e^{\frac{1}{2}\langle L,L\rangle_\infty}\right]} \le \sqrt{\mathbb{E}\left[e^{\frac{1}{2}\langle L,L\rangle_\infty}\right]} < \infty.$$

(ii) $\Rightarrow$ (iii): If L is a uniformly integrable martingale, by optional stopping theorem, for any stopping time τ, one have $L_\tau = \mathbb{E}\left[L_\infty|\mathscr{F}_\tau\right]$. By Jensen's inequality,

$$\mathbb{E}\left[\exp\left(\frac{1}{2}L_\tau\right)\right] \le \mathbb{E}\left[\exp\left(\frac{1}{2}L_\infty\right)|\mathscr{F}_\tau\right].$$

Since $\exp\left(\frac{1}{2}L_\infty\right)$ is integrable, the collection of random variables $\exp\left(\frac{1}{2}L_\tau\right)$ 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 $0 < a < 1$,

$$\mathscr{E}(aL)_t = \left(\mathscr{E}(L)_t\right)^{a^2} (Z_t^{(a)})^{1-a^2}.$$

By Hölder's inequality, for any measurable set $\Gamma \in \mathscr{F}$ and any stopping time τ, one have

$$\mathbb{E}\left[\mathbb{1}_\Gamma \mathscr{E}(aL)_\tau\right] \le \mathbb{E}\left[\mathscr{E}(L)_\tau\right]^{a^2} \mathbb{E}\left[\mathbb{1}_\Gamma Z_\tau^{(a)}\right]^{1-a^2} \le \mathbb{E}\left[\mathbb{1}_\Gamma Z_\tau^{(a)}\right]^{1-a^2} \le \mathbb{E}\left[\mathbb{1}_\Gamma \exp\left(\frac{1}{2}L_\tau\right)\right]^{2a(1-a)},$$

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 $\mathscr{E}(aL)_\tau$ is uniformly integrable, where τ runs over all stopping times. Let $\tau_n \wedge \infty$ be a sequence of stopping times reducing $\mathscr{E}(aL)$. Then for all $t > s \ge 0$, by uniform integrability,

$$\mathbb{E}[\mathscr{E}(aL)_t|\mathscr{F}_s] = \lim_{n\to\infty} \mathbb{E}\left[\mathscr{E}(aL)_{t\wedge\tau_n}|\mathscr{F}_s\right] = \lim_{n\to\infty} \mathscr{E}(aL)_{s\wedge\tau_n} = \mathscr{E}(aL)_s.$$

Hence $\mathscr{E}(aL)$ is a uniformly integrable martingale, and

$$1 = \mathbb{E}[\mathscr{E}(aL)_\infty] \le \mathbb{E}\left[\mathscr{E}(L)_\infty\right]^{a^2} \mathbb{E}\left[Z_\infty^{(a)}\right]^{1-a^2} \le \mathbb{E}\left[\mathscr{E}(L)_\infty\right]^{a^2} \mathbb{E}\left[\exp\left(\frac{1}{2}L_\infty\right)\right]^{2a(1-a)},$$

which implies $\mathbb{E}\left[\mathscr{E}(L)_\infty\right] = 1$. Again, by Fatou's lemma, $\mathbb{E}[\mathscr{E}(L)_\infty|\mathscr{F}_t] \le \mathscr{E}(L)_t$. On the other hand, we have

$$\mathbb{E}[\mathscr{E}(L)_\infty] = \mathbb{E}[\mathscr{E}(L)_t] = \mathbb{E}[\mathscr{E}(L)_0] = 1.$$

Hence $\mathbb{E}[\mathscr{E}(L)_\infty|\mathscr{F}_t] = \mathscr{E}(L)_t$, and $\mathscr{E}(L)$ is a uniformly integrable martingale. □

Remark. Let L be a continuous P-local martingale satisfying property (ii). To apply Girsanov's theorem, we let Q be the probability measure with density $\mathscr{E}(L)_\infty$ with respect to P. According to Proposition 5.25, the Radon-Nikodym derivative is $\frac{\mathrm{d}Q}{\mathrm{d}P}|_{\mathscr{F}_t} = D_t = \mathscr{E}(L)_t$.

5.5.2 The Cameron-Martin Formula

Motivation. Girsanov transformation and SDE. Let $\beta = (\beta_t)_{t\ge0}$ be a p-dimensional $(\mathscr{F}_t)$-Brownian motion under P. Consider the following continuous local martingale:

$$L_t = \int_0^t b(s,\beta_s)\,\mathrm{d}\beta_s.$$

If there exists $g \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$ such that $|b(t,x)| \le g(t)$ 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

$$D_t = \mathscr{E}(L)_t = \exp\left(\int_0^t b(s,\beta_s)\,\mathrm{d}\beta_s - \frac{1}{2}\int_0^t |b(s,\beta_s)|^2\,\mathrm{d}s \right).$$

We set $\mathrm{d}Q = D_\infty \mathrm{d}P$, which is a probability distribution. By Remark II under Theorem 5.30, the following process B is an $(\mathscr{F}_t)$-Brownian motion under Q:

$$B_t = \beta_t - \int_0^t b(s,\beta_s)\,\mathrm{d}s$$

Consequently, $X = \beta$ solves the following SDE under probability measure Q:

$$\mathrm{d}X_t = \mathrm{d}B_t + b(t,X_t)\,\mathrm{d}t$$

Here we only assume that $b : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^p$ is dominated by a function $g \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$.

Proposition 5.33. *Consider the following two SDEs admitting unique strong solution on $\mathbb{R}_+$:*

$$\begin{cases} \mathrm{d}X_t = \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}B_t, \\ \mathrm{d}Y_t = (\mu+\nu)(t,Y_t)\,\mathrm{d}t + \sigma(t,Y_t)\,\mathrm{d}B_t, \end{cases}$$

where B_t is a p-dimensional $(\mathscr{F}_t)$-Brownian motion under P, and $\sigma : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^{p\times p}$ is almost everywhere invertible. Furthermore, the Novikov's condition is satisfied:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^\infty |\sigma(s,Y_s)^{-1}\nu(s,Y_s)|^2\,\mathrm{d}s \right) \right] < \infty.$$

Then, if $X_0 \overset{d}{=} Y_0$, the following identity holds for all bounded functional $\Phi : C(\mathbb{R}_+) \to \mathbb{R}$:

$$\mathbb{E}\left[\Phi(X)\right] = \mathbb{E}\left[\Phi(Y) \exp\left(-\int_0^\infty \sigma(s,Y_s)^{-1}\nu(s,Y_s)\,\mathrm{d}s - \frac{1}{2}\int_0^\infty |\sigma(s,Y_s)^{-1}\nu(s,Y_s)|^2\,\mathrm{d}B_s \right) \right].$$

Proof. Define the following continuous P-local martingale:

$$L_t = -\int_0^t \sigma(s,Y_s)^{-1}\nu(s,Y_s)\,\mathrm{d}B_s.$$

Since the Novikov's criterion is satisfied, we can use the exponential martingale:

$$\mathscr{E}(L)_t = \exp\left(-\int_0^t \sigma(s,Y_s)^{-1}\nu(s,Y_s)\,\mathrm{d}s - \frac{1}{2}\int_0^t |\sigma(s,Y_s)^{-1}\nu(s,Y_s)|^2\,\mathrm{d}B_s \right).$$

Define $\mathrm{d}Q = \mathscr{E}(L)_\infty\,\mathrm{d}P$, and the $(\mathscr{F}_t)$-Brownian motion under P:

$$\widetilde{B}_t = B_t - \langle B, L\rangle_t = B_t + \int_0^t \sigma(s,Y_s)^{-1}\nu(s,Y_s)\,\mathrm{d}s.$$

Then we have $\mathrm{d}Y_t = \mu(t,Y_t)\,\mathrm{d}t + \sigma(t,Y_t)\,\mathrm{d}\widetilde{B}_t$. Consequently, Y_t solves the first SDE under probability measure Q and $(\mathscr{F}_t, Q)$-Brownian motion $\widetilde{B}$. By uniqueness of the strong solution, if $X_0 \overset{d}{=} Y_0$, then X and Y has the same law under P and Q, respectively. The final result follows by $\mathbb{E}_P\left[\Phi(X)\right] = \mathbb{E}_Q\left[\Phi(Y)\right]$. □

We specialize the previous discussion to the case $b(t,x) = g(t)$, where $g \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$. We set

$$h(t) = \int_0^t g(s)\,\mathrm{d}s, \quad \forall t \in \mathbb{R}_+.$$

The set $\mathscr{H}$ of all functions h that can be written in this form is called the *Cameron-Martin space*, and we write the derivative of $h \in \mathscr{H}$ in sense of distribution by $\dot{h} = g \in L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), m)$. As a special case of our previous discussion, given a Brownian motion B under P, we construct the probability measure

$$\mathrm{d}Q = D_\infty \,\mathrm{d}P = \exp\left(\int_0^\infty g(t)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t |g(t)|^2\,\mathrm{d}s\right)\mathrm{d}P.$$

Then the process $\widetilde{B}_t = B_t - h(t)$ is a Brownian motion under Q. Consequently, for every nonnegative measurable function on $C(\mathbb{R}_+, \mathbb{R})$, one have

$$\mathbb{E}_P\left[D_\infty \Phi((B_t)_{t\geq 0})\right] = \mathbb{E}_Q\left[\Phi((B_t)_{t\geq 0})\right] = \mathbb{E}_Q\left[\Phi\left((\widetilde{B}_t + h(t))_{t\geq 0}\right)\right] = \mathbb{E}_P\left[\Phi((B_t + h(t))_{t\geq 0})\right]$$

We rewrite this formula to the following form.

Theorem 5.34 (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 $\mathscr{H}$. Then for every nonnegative measurable function Φ on $C(\mathbb{R}_+, \mathbb{R})$,*

$$\int W(\mathrm{d}\mathbf{w})\Phi(\mathbf{w} + h) = \int W(\mathrm{d}\mathbf{w}) \exp\left(\int_0^\infty \dot{h}(s)\mathrm{d}\mathbf{w}(s) - \frac{1}{2}\int_0^\infty |\dot{h}(s)|^2\,\mathrm{d}s\right)\Phi(\mathbf{w}).$$

Remark. The integral $\int_0^\infty \dot{h}(s)\mathrm{d}\mathbf{w}(s)$ is viewed as the Wiener integral, where $\mathbf{w} \sim W$.

Application: Law of hitting times for Brownian motion with drift. Let $(B_t)_{t\geq 0}$ 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 $\mu \in \mathbb{R}$, consider the stopping time

$$\upsilon_a = \inf\{t \geq 0 : B_t + \mu t = a\}.$$

Clearly, if $\mu = 0$, we have $\upsilon_a = \tau_a$, and the desired law is given by Corollary 4.36. For the general case, we fix $t > 0$, and use Cameron-Martin formula to the following function:

$$h(s) = \mu(s \wedge t), \quad \dot{h}(s) = \mu \mathbb{1}_{\{s \leq t\}}, \qquad \Phi(\mathbf{w}) = \mathbb{1}_{\{\max_{s\in[0,t]} \mathbf{w}(s) \geq a\}}, \ \mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}).$$

Then we have

$$\mathbb{P}\left(\upsilon_a \leq t\right) = \mathbb{E}\left[\Phi(B + h)\right] = \mathbb{E}\left[\exp\left(\mu B_t - \frac{\mu^2}{2}t\right)\mathbb{1}_{\{\tau_a \leq t\}}\right].$$

By optional stopping theorem [Theorem 3.60], we have

$$\exp\left(\mu B_{t\wedge\tau_a} - \frac{\mu^2}{2}(t \wedge \tau_a)\right) = \mathbb{E}\left[\exp\left(\mu B_t - \frac{\mu^2}{2}t\right)|\mathscr{F}_{t\wedge\tau_a}\right].$$

Consequently,

$$\begin{aligned}\mathbb{P}\left(\upsilon_a \leq t\right) &= \mathbb{E}\left[\exp\left(\mu B_{t\wedge\tau_a} - \frac{\mu^2}{2}(t \wedge \tau_a)\right)\mathbb{1}_{\{\tau_a \leq t\}}\right] = \mathbb{E}\left[\exp\left(\mu a - \frac{\mu^2}{2}\tau_a\right)\mathbb{1}_{\{\tau_a \leq t\}}\right] \\ &= \int_0^t \frac{a}{\sqrt{2\pi s^3}} e^{-\frac{1}{2s}(\mu s - a)^2}\mathrm{d}s.\end{aligned}$$

Therefore, υ_a has a density supported on $\mathbb{R}_+$: $\rho_{\upsilon_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{1}{2t}(\mu t - a)^2}, \ t > 0.$

6 Markov Processes

6.1 Transition Semigroups and Feller Semigroups

Definition 6.1 (Markovian transition kernels)**.** Let $(E,\mathscr{E})$ be a measurable space. A *Markovian transition kernel* (or *transition kernel*, for short) on E is a mapping $Q: E\times\mathscr{E}\to[0,1]$ satisfying the following properties:

(i) For every $x\in E$, the mapping $\mathscr{E}\ni A\mapsto Q(x,A)$ is a probability measure on $(E,\mathscr{E})$;

(ii) For every $A\in\mathscr{E}$, the mapping $E\ni x\mapsto Q(x,A)$ is $\mathscr{E}$-measurable.

Remark. (i) If E is a finite or countable set equipped with the σ-algebra $\mathscr{E}$ 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

$$Qf: E\to\mathbb{R},\quad Qf(x)=\int_E Q(x,\mathrm{d}y)f(y).$$

For every $A\in\mathscr{E}$, we have $Q\mathbb{1}_A(x)=Q(x,A)$, hence the function $Q\mathbb{1}_A$ is measurable. A simple function approximation argument shows that Qf is measurable for all $f\in B(E)$. Furthermore,

$$\|Qf\|=\sup_{x\in E}\left(\int_E Q(x,\mathrm{d}y)f(y)\right)\leq\|f\|\sup_{x\in E}\left(\int_E Q(x,\mathrm{d}y)\right)=\|f\|.$$

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 $A\in\mathscr{E}$,

$$Q_{t+s}(x,A)=\int_E Q_t(x,\mathrm{d}y)Q_s(y,A).$$

(iii) For every $A\in\mathscr{E}$, the mapping $(t,x)\mapsto Q_t(x,A)$ is measurable with respect to $\mathscr{B}(\mathbb{R}_+)\times\mathscr{E}$.

Remark. If we view $(Q_t)_{t\geq 0}$ as bounded linear operators on $B(E)$, the Chapman-Kolmogorov identity implies that $Q_{t+s}=Q_tQ_s$ for all $s,t\geq 0$. This give rise to the associative property: $(Q_rQ_s)Q_t=Q_r(Q_sQ_t)$ 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 $R_\lambda: B(E)\to B(E)$ defined by

$$R_\lambda f(x)=\int_0^\infty e^{-\lambda t}Q_tf(x)\,\mathrm{d}t,\quad \forall f\in B(E),\ x\in E.$$

Remark. The resolvent has the following properties:

(i) $R_\lambda: B(E)\to B(E)$ is a positive and bounded linear operator. Note that $R_\lambda f\geq 0$ for all $f\geq 0$, and

$$\sup_{x\in E}\left(\int_0^\infty e^{-\lambda t}Q_tf(x)\,\mathrm{d}t\right)\leq\int_0^\infty e^{-\lambda t}\|Q_tf\|\,\mathrm{d}t\leq\int_0^\infty e^{-\lambda t}\|f\|\,\mathrm{d}t\quad\Rightarrow\quad\|R_\lambda f\|\leq\frac{1}{\lambda}\|f\|.$$

(ii) (Resolvent equation). For all $\lambda, \mu > 0$, we have $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$:

$$\begin{aligned}
R_\lambda(R_\mu f)(x) &= \int_0^\infty e^{-\lambda s} \left(\int_E Q_s(x, \mathrm{d}y) \int_0^\infty e^{-\mu t} Q_t f(y) \,\mathrm{d}t \right) \mathrm{d}s \\
&= \int_0^\infty e^{-\lambda s} \left(\int_0^\infty e^{-\mu t} Q_{s+t} f(x) \,\mathrm{d}t \right) \mathrm{d}s \overset{r=s+t}{=} \int_0^\infty e^{-(\lambda-\mu)s} \left(\int_s^\infty e^{-\mu r} Q_r f(x) \,\mathrm{d}r \right) \mathrm{d}s \\
&= \int_0^\infty Q_r f(x) e^{-\mu r} \left(\int_0^r e^{-(\lambda-\mu)s} \,\mathrm{d}s \right) \mathrm{d}r = \int_0^\infty \left(\frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} \right) Q_r f(x) \,\mathrm{d}r.
\end{aligned}$$

Consequently, if we fix $f \in B(E)$, it holds

$$\|(R_\lambda - R_\mu) f\| = |\lambda - \mu| \, \|R_\lambda R_\mu f\| \le \frac{|\lambda - \mu| \, \|f\|}{\lambda \mu}, \quad \forall \lambda, \mu \in (0, +\infty).$$

Hence $\lambda \to R_\lambda f$ is a continuous mapping from $(0, +\infty)$ into $B(E)$.

(iii) For every $\lambda > 0$ and every $n \in \mathbb{N}$, we have

$$R_\lambda^n f(x) = \int_0^\infty \frac{s^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) \,\mathrm{d}s.$$

Clearly, this equality holds for $n = 1$. Hence we can prove the general case by induction:

$$\begin{aligned}
R_\lambda^{n+1} f(x) = R_\lambda(R_\lambda^n f)(x) &= \int_0^\infty e^{-\lambda r} \left(\int_0^\infty \frac{t^{n-1}}{(n-1)!} e^{-\lambda t} Q_{r+t} f(x) \,\mathrm{d}t \right) \mathrm{d}r \\
&= \int_0^\infty \left(\int_r^\infty \frac{(s-r)^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) \,\mathrm{d}t \right) \mathrm{d}r \\
&= \int_0^\infty \int_0^s \frac{(s-r)^{n-1}}{(n-1)!} e^{-\lambda s} Q_s f(x) \,\mathrm{d}r \,\mathrm{d}s = \int_0^\infty \frac{s^n}{n!} e^{-\lambda s} Q_s f(x) \,\mathrm{d}s.
\end{aligned}$$

Preliminary: LCCB space. From now on, we deal with a special topological space E, which is *Locally Compact, Hausdorff, and has a Countable Basis $\mathscr{B}$ (LCCB).*

Since E is locally compact, for each $x \in E$ there exists an open neighborhood U_x with compact closure. Consequently, one can find a basis set $B_x \in \mathscr{B}$ such that $x \in B_x \subset U_x$, and $\overline{B}_x$ is compact. We choose $\mathscr{B}_K \subset \mathscr{B}$ to be the collection of all basis sets with compact closure. Then $B_x \in \mathscr{B}_K$ for all $x \in E$, and $E = \bigcup_{B \in \mathscr{B}_K} \overline{B}$ 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

$$K_1 \subset K_2^\circ \subset K_2 \subset K_3^\circ \subset \cdots \subset K_n \subset K_{n+1}^\circ \subset K_{n+1} \subset \cdots, \quad E = \bigcup_{n=1}^\infty K_n. \tag{*}$$

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 $x_1, \cdots, x_k$ 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 $f : E \to \mathbb{R}$ is said to be *vanishing at infinity*, if for all $\epsilon > 0$, there exists a compact $K \subset E$ such that $|f(x)| < \epsilon$ for all $x \in E \backslash K$, or equivalently, $\{x \in E : |f(x)| \ge \epsilon\}$ is compact. In addition, If

$(K_n)_{n=1}^\infty$ is a sequence of compact sets specified in (*), then we have

$$\lim_{n\to\infty} \sup_{x\in E\setminus K_n} |f(x)| = 0.$$

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 $L : C_0(E) \to \mathbb{R}$ is a bounded linear functional, then there exists a unique regular (finite signed) measure μ such that $Lf = \int f \,\mathrm{d}\mu$ 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 $\mathscr{B} = \{B_n\}_{n=1}^\infty$ 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 $f_{m,n} : X \to [0,1]$ 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^*, $\{K_n^\circ\}_{n=1}^\infty$ is an open cover of the compact set $E^*\setminus K$. Consequently, we can construct a countable collection $\mathcal{F} \in C(E^*)$ 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 $\mathscr{E}$. Let μ_n be a sequence of probability measures on $(E, \mathscr{E})$. If there exists a probability measure μ such that

$$\int_E f \,\mathrm{d}\mu_n \to \int_E f \,\mathrm{d}\mu, \quad \forall f \in C_c(E),$$

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 $\int_E \phi \,\mathrm{d}\mu > 1 - \frac{\epsilon}{3M}$. By vague convergence, there exists N_1 such that $\int_E \phi \,\mathrm{d}\mu_n > 1 - \frac{\epsilon}{3M}$ for all $n \geq N$. Since $g\phi \in C_c(E)$, we also choose N_2 such that $\left|\int_E g\phi \,\mathrm{d}\mu_n - \int_E g\phi \,\mathrm{d}\mu\right| < \epsilon/3$ for all $n \geq N_2$. Then

$$\begin{aligned}\left|\int_E g \,\mathrm{d}\mu_n - \int_E g \,\mathrm{d}\mu\right| &= \left|\int_E g(1-\phi) \,\mathrm{d}\mu_n\right| + \left|\int_E g\phi \,\mathrm{d}\mu_n - \int_E g\phi \,\mathrm{d}\mu\right| + \left|\int_E g(1-\phi) \,\mathrm{d}\mu\right| \\ &\leq 2M \left|\int_E (1-\phi) \,\mathrm{d}\mu\right| + \left|\int_E g\phi \,\mathrm{d}\mu_n - \int_E g\phi \,\mathrm{d}\mu\right| < \epsilon, \quad \forall n \geq \max\{N_1, N_2\}.\end{aligned}$$

Since ϵ is arbitrarily small, we have $\int_E g \,\mathrm{d}\mu_n \to \int_E g \,\mathrm{d}\mu$, 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,

$$\lim_{n\to\infty} R_\lambda f(x_n) = \lim_{n\to\infty} \int e^{-\lambda s} Q_s f(x_n)\,\mathrm{d}s = \int e^{-\lambda s} \lim_{n\to\infty} Q_s f(x_n)\,\mathrm{d}s, \quad \forall f \in C_0(E).$$

In the last equality, we use the fact $\sup_{t\geq 0} \left|e^{-\lambda s} Q_s f(x_n)\right| \leq \|f\|$ 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

$$\lim_{\lambda\to\infty} \lambda R_\lambda f(x) = \lim_{\lambda\to\infty} \lambda \int_0^\infty e^{-\lambda s} Q_s f(x)\,\mathrm{d}s = \lim_{\lambda\to\infty} \int_0^\infty e^{-t} Q_{t/\lambda} f(x)\,\mathrm{d}t = \int_0^\infty e^{-t} f(x)\,\mathrm{d}t = f(x),$$

where the last equality holds by dominated convergence, since $\sup_{s\geq 0} \left|e^{-s} Q_{s/\lambda} f(x)\right| \leq \|f\|$ for all $\lambda > 0$. Furthermore, for all $\lambda, \mu > 0$ and all $x \in E$, we have

$$\begin{aligned}(\lambda R_\lambda - \mu R_\mu) f(x) &= \int_0^\infty \left(\lambda e^{-\lambda s} - \mu e^{-\mu s}\right) Q_s f(x)\,\mathrm{d}s = \int_0^\infty e^{-t}\left(Q_{t/\lambda} - Q_{t/\mu}\right) f(x)\,\mathrm{d}t \\ &= \int_0^\infty e^{-t} \left(\frac{t}{\mu} - \frac{t}{\lambda}\right) Q_{t/\mu} Q_{t/\lambda} f(x)\,\mathrm{d}t \leq \left|\frac{1}{\lambda} - \frac{1}{\mu}\right| \|f\|.\end{aligned}$$

Hence $\|(\lambda R_\lambda - \mu R_\mu) f\| \to 0$ as $\lambda, \mu \to \infty$, and $\lambda R_\lambda f \in \mathfrak{R}$ converges in $C_0(E)$ as $\lambda \to \infty$ by completeness, and the pointwise limit f must be the limit $C_0(E)$. Consequently, $\mathfrak{R}$ is dense in $C_0(E)$. ☐

Remark. In the proof, we also conclude that $\lim_{\lambda\to\infty} \|\lambda R_\lambda f - f\| = 0$ 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

$$Q_t R_\lambda f(x) = e^{\lambda t} \int_t^\infty e^{-\lambda s} Q_s f(x)\,\mathrm{d}s \leq e^{\lambda t} R_\lambda f - \int_0^t e^{\lambda(t-s)} Q_s f(x)\,\mathrm{d}s$$

Consequently, we have

$$\|Q_t R_\lambda f - R_\lambda f\| = \left(e^{\lambda t} - 1\right) \|R_\lambda f\| + t e^{\lambda t} \|f\| \to 0 \quad \textit{as} \;\; t \downarrow 0.$$

Therefore we have $\lim_{t\downarrow 0} \|Q_t f - f\| = 0$ for all $f \in \mathfrak{R}$. The continuity of $Q_t - \mathrm{Id}$ and a standard density argument extend this conclusion to all $f \in C_0(E)$. For the second assertion, note that for all $t > s \geq 0$,

$$\|Q_t f - Q_s f\| = \|Q_s(Q_{t-s} f - f)\| = \|Q_{t-s} f - f\| \to 0 \quad \textit{as} \;\; t - s \downarrow 0.$$

Since this convergence is uniform for all $s \in (0, \infty)$, the mapping $t \mapsto Q_t f$ is uniformly continuous. ☐

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 $L : \mathfrak{D}(L) \to C_0(E)$ as follows:

$$Lf = \lim_{t\downarrow 0} \frac{Q_t f - f}{t}, \quad \forall f \in C_0(E).$$

The operator $L : \mathfrak{D}(L) \to C_0(E)$ 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,

$$Q_t f = f + \int_0^t Q_s L f \,\mathrm{d}s = f + \int_0^t L Q_s f \,\mathrm{d}s \qquad \Leftrightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t} Q_t f = L Q_t f. \tag{6.1}$$

(ii) $\mathfrak{D}(L) = \mathfrak{R}$, and R_λ is the inverse of $\lambda \operatorname{Id} - L$ for all $\lambda > 0$, namely, $(\lambda \operatorname{Id} - L) R_\lambda = R_\lambda(\lambda \operatorname{Id} - L) = \operatorname{Id}$.

(iii) The semigroup $(Q_t)_{t\geq 0}$ is determined by the generator L:

$$Q_t = e^{tL} := \lim_{\lambda\to\infty} e^{-t\lambda} \sum_{k=0}^{\infty} \frac{t^k \lambda^{2k}}{k!} (\lambda \operatorname{Id} - L)^{-k}$$

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

$$\frac{Q_t(Q_s f) - Q_s f}{t} = Q_s \left(\frac{Q_t f - f}{t} \right) \quad \Rightarrow \quad Q_s f \in \mathfrak{D}(L), \;\; L(Q_s f) = Q_s(Lf).$$

Similarly, we have $h^{-1}(Q_{t+h} f - Q_f) = Q_t(Lf)$ 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,

$$Q_t f - f = \int_0^t Q_s(Lf) \,\mathrm{d}s = \int_0^t L(Q_s f) \,\mathrm{d}s, \quad \forall t \in \mathbb{R}_+.$$

(ii) If $f \in \mathfrak{D}(L)$, we use (6.1) and Fubini's theorem:

$$\begin{aligned}
R_\lambda(\lambda \operatorname{Id} - L) f &= \int_0^\infty \lambda e^{-\lambda s} Q_s f \,\mathrm{d}s - \int_0^\infty e^{-\lambda s} Q_s(Lf) \,\mathrm{d}s \\
&= \int_0^\infty \lambda e^{-\lambda s} \left(f + \int_0^s L(Q_t f) \,\mathrm{d}t \right) \mathrm{d}s - \int_0^\infty e^{-\lambda s} L(Q_s f) \,\mathrm{d}s \\
&= f + \int_0^\infty \left(\int_t^\infty \lambda e^{-\lambda s} L(Q_t f) \,\mathrm{d}s \right) \mathrm{d}t - \int_0^\infty e^{-\lambda s} L(Q_s f) \,\mathrm{d}s = f.
\end{aligned}$$

Apparently, we have $\mathfrak{D}(L) \subset \mathfrak{R}$. On the other hand, for $g \in C_0(E)$, by dominated convergence theorem,

$$\begin{aligned}\lim_{h\downarrow 0} \frac{Q_h R_\lambda g - R_\lambda g}{h} &= \lim_{h\downarrow 0} \frac{1}{h} \int_0^\infty e^{-\lambda t} \left(Q_{t+h} - Q_t\right) g \,\mathrm{d}t \\ &= \lim_{h\downarrow 0} \frac{1}{h} \left(\left(1 - e^{-\lambda h}\right) \int_0^\infty e^{-\lambda t} Q_{t+h} g \,\mathrm{d}t - \int_0^h e^{-\lambda t} Q_t g \,\mathrm{d}t \right) = \lambda R_\lambda g - g.\end{aligned}$$

Hence $R_\lambda g \in \mathfrak{D}(L)$, and $\mathfrak{R} = \mathfrak{D}(L)$. Moreover, the first assertion in (i) implies

$$LR_\lambda g = \lambda R_\lambda g - g \quad \Rightarrow \quad (\lambda \operatorname{Id} - L) R_\lambda g = g.$$

(iii) Fix $f \in C_0(E)$. For every $\lambda > 0$, if $(\lambda \operatorname{Id} - L) g = f$ for some $g \in \mathfrak{D}(L)$, then $g = R_\lambda (\lambda \operatorname{Id} - L) g = R_\lambda f$. Hence the resolvent R_λ is uniquely determined by inverting the operator $\lambda \operatorname{Id} - L$, which is defined on $C_0(E)$.

For every $\lambda > 0$, we define $A_\lambda := \lambda R_\lambda L = \lambda^2 R_\lambda - \lambda \operatorname{Id}$. Then $A_\lambda : \mathfrak{D}(L) \to C_0(E)$ is a bounded linear operator, and $\|A_\lambda\| \le 2\lambda$. Then we define the following series, which converges in norm:

$$e^{tA_\lambda} = \sum_{k=0}^\infty \frac{(tA_\lambda)^k}{k!} \quad \Rightarrow \quad \left\|e^{tA_\lambda}\right\| = \left\|e^{t\lambda^2 R_\lambda - \lambda \operatorname{Id}}\right\| \le \sum_{k=0}^\infty e^{-t\lambda} \frac{\|t\lambda^2 R_\lambda\|^k}{k!} \le \sum_{k=0}^\infty e^{-t\lambda} \frac{(t\lambda)^k}{k!} = 1.$$

By commutativity principle (i.e. $e^{T+S} = e^T e^S$ for commutative bounded linear operators $ST = TS$), the collection $(e^{tA_\lambda})_{t\ge 0}$ is a semigroup of contractions. Moreover, since $A_\lambda A_\mu = A_\mu A_\lambda = \frac{\mu-\lambda}{\mu\lambda} L(R_\lambda - R_\mu) L$,

$$e^{tA_\lambda} - e^{tA_\mu} = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} e^{tA_\mu + s(A_\lambda - A_\mu)} \,\mathrm{d}s = \int_0^t e^{(t-s)A_\mu} e^{sA_\lambda} (A_\lambda - A_\mu) \,\mathrm{d}s.$$

Hence for all $f \in \mathfrak{D}(L)$, since $A_\lambda f \to Lf$ as $\lambda \to \infty$,

$$\left\|\left(e^{tA_\lambda} - e^{tA_\mu}\right) f\right\| \le \|(A_\lambda - A_\mu) f\| \quad \Rightarrow \quad e^{tL} f := \lim_{\lambda\to\infty} e^{tA_\lambda} f \ \textit{exists in}\ C_0(E).$$

Clearly, the mapping $e^{tL} : \mathfrak{D}(L) \to C_0(E)$ 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 $(e^{tL})_{t\ge 0}$ is a semigroup: $\forall f \in \mathfrak{D}(L)$,

$$\begin{aligned}\left\|e^{(t+s)L} f - e^{tL} e^{sL} f\right\| &= \left\|(e^{(t+s)L} - e^{(t+s)A_\lambda}) f\right\| + \left\|e^{tA_\lambda} (e^{sA_\lambda} - e^{sL}) f\right\| + \left\|(e^{tA_\lambda} - e^{tL}) e^{sL} f\right\| \\ &\le \left\|(e^{(t+s)L} - e^{(t+s)A_\lambda}) f\right\| + \left\|(e^{sA_\lambda} - e^{sL}) f\right\| + \left\|(e^{tA_\lambda} - e^{tL}) e^{sL} f\right\| \to 0 \quad \textit{as}\ \lambda \to \infty.\end{aligned}$$

We also note strong continuity so that $(e^{tL})_{t\ge 0}$ is a Feller semigroup: $\forall f \in \mathfrak{D}(L)$,

$$\|e^{tL} f - f\| \le \|(e^{tL} - e^{tA_\lambda}) f\| + \|e^{tA_\lambda} f - f\| \le t\|Lf - A_\lambda f\| + \|e^{tA_\lambda} f - f\| \to 0 \quad \textit{as}\ t \downarrow 0.$$

Furthermore, the generator of $(e^{tL})_{t\ge 0}$ is L:

$$e^{tA_\lambda} f - f = \int_0^t A_\lambda e^{sA_\lambda} f \,\mathrm{d}s \quad \Rightarrow \quad e^{tL} f - f = \int_0^t L e^{sL} f \,\mathrm{d}s \quad \Rightarrow \quad \lim_{t\downarrow 0} \frac{e^{tL} f - f}{t} = L.$$

Since the resolvent $R_\lambda f = \int_0^\infty e^{-\lambda t} Q_t f \,\mathrm{d}t$ is the Laplacian transform, it has a one-to-one correspondence with the transition semigroup $(Q_t)_{t\ge 0}$. Hence we can recover $Q_t = e^{tL}$ uniquely from the generator L. ☐

Remark. Note that the domain of operator $(\lambda \operatorname{Id} - L) R_\lambda$ is $C_0(E)$.

6.2 Markov Processes and Feller Processes

Now we fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $(\mathscr{F}_t)_{t\geq 0}$.

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 $(\mathscr{F}_t)_{t\geq 0}$ 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 $s, t \geq 0$,

$$\mathbb{E}\left[f(X_{s+t})|\mathscr{F}_s\right] = Q_t f(X_s). \tag{6.2}$$

Without specifying a filtration, we implicitly mean that the definition holds for the canonical filtration $\mathscr{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$. Clearly, a Markov process X with respect to any filtration $(\mathscr{F}_t)_{t\geq 0}$ is also a Markov process with respect to the canonical filtration $(\mathscr{F}_t^X)_{t\geq 0}$.

A Markov process with values in E is called a *Feller process* if its transition semigroup is Feller.

Remark. Particularly, for all $A \in \mathscr{E}$, one can take $f = \mathbb{1}_A$ in (6.1) to obtain

$$\mathbb{P}(X_{s+t} \in A|\mathscr{F}_s) = Q_t(X_s, A).$$

It is seen that the conditional distribution of X_{s+t} given the history $\mathscr{F}_s$ only depends on the current state X_s. Furthermore, given the law γ of X_0, we have for all $0 < t_1 < \cdots < t_n$ and all $A_0, A_1, \cdots, A_n \in \mathscr{E}$ that

$$\begin{aligned} &\mathbb{P}(X_0 \in A_0, X_{t_1} \in A_1, X_{t_2} \in A_2, \cdots, X_{t_n} \in A_n) \\ &\quad= \int_{A_0} \gamma(\mathrm{d}x_0) \int_{A_1} Q_{t_1}(x_0, \mathrm{d}x_1) \int_{A_1} Q_{t_2-t_1}(x_1, \mathrm{d}x_2) \cdots \int_{A_n} Q_{t_n-t_{n-1}}(x_{n-1}, \mathrm{d}x_n). \end{aligned} \tag{6.3}$$

More generally, if $f_0, f_1, \cdots, f_n \in B(E)$, we have

$$\begin{aligned} &\mathbb{E}\left[f_0(X_0) f_1(X_{t_1}) \cdots f_n(X_{t_n})\right] \\ &\quad= \int_E \gamma(\mathrm{d}x_0) f_0(x_0) \int_E Q_{t_1}(x_0, \mathrm{d}x_1) f_1(x_1) \int_E Q_{t_2-t_1}(x_1, \mathrm{d}x_2) f_2(x_2) \cdots \int_E Q_{t_n-t_{n-1}}(x_{n-1}, \mathrm{d}x_n) f_n(x_n). \end{aligned}$$

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 $P^\gamma_{t_1,\cdots,t_n} = \mathbb{P}(X_0 \in \cdot, X_{t_1} \in \cdot, \cdots, X_{t_n} \in \cdot)$ on all measurable rectangles $\mathscr{E}^n$, which can be uniquely extended to a measure on the product space $(E^n, \mathscr{E}^{\otimes n})$. In addition, the collection of all finite marginals $\{P^\gamma_{t_1,\cdots,t_n} : n \in \mathbb{N},\ 0 < t_1 < \cdots < t_n\}$ satisfies the compatibility condition given in roposition 4.15, according to the Chapman-Kolmogorov identity.

We further assume that E is a Polish space. Then according to Corollary 4.19, which is a consequence of the Daniell-Kolmogorov extension theorem, the compatible family $\{P^\gamma_{t_1,\cdots,t_n} : n \in \mathbb{N},\ 0 < t_1 < \cdots < t_n\}$ of probability measures has a unique extension P^γ on the canonical space $(E^{\mathbb{R}_+}, \mathscr{E}^{\otimes \mathbb{R}_+})$. Consequently, the canonical process $\{\pi_t\}_{t\geq 0}$ on $(E^{\mathbb{R}_+}, \mathscr{E}^{\otimes \mathbb{R}_+})$ 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

$$\begin{aligned} \forall x \in E \ \ and \ \ \forall t \geq 0, \ \ X_t^y \xrightarrow{d} X_t^x \ \ as \ \ y \to x \qquad &\Leftrightarrow \qquad Q_t C_0(E) \subset C_0(E); \\ \forall x \in E, \ \ X_t^x \xrightarrow{\mathbb{P}} x \ \ as \ \ t \downarrow 0 \qquad &\Leftrightarrow \qquad Q_t f(x) \to f(x), \ \ \forall f \in C_0(E). \end{aligned}$$

6.2.1 Sample Path Regularity

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

$$Q_s R_\lambda h = \int_0^\infty e^{-\lambda t} Q_{t+s} h \,\mathrm{d}t = e^{\lambda s} \int_0^\infty e^{-\lambda(s+t)} Q_{t+s} h \,\mathrm{d}t \leq e^{\lambda s} R_\lambda h.$$

By (6.2), the following inequality holds for $(X_t)_{t\geq 0}$:

$$\mathbb{E}[e^{-\lambda(t+s)} R_\lambda h(X_{t+s}) | \mathscr{F}_t] = e^{-\lambda(t+s)} Q_s R_\lambda h(X_t) \leq e^{\lambda t} R_\lambda h(X_t)$$

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 with a countable basis. Let $(X_t)_{t\geq 0}$ be a Feller process with semigroup $(Q_t)_{t\geq 0}$, with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$. Then the process $(X_t)_{t\geq 0}$ has a càdlàg modification.*

Proof. Let $\mathscr{H} \in C(E)$ 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 $h \in \mathscr{H}$. 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 $h(x) = \lim_{n\to\infty} h(x_n) = h(y)$ for all $h \in \mathscr{H}$, and $x = y$ by definition of $\mathscr{H}$. Consequently, x_n converges a unique limit point.

We take a sequence $f_n \in C_0^+(E)$ separating the points of E, and take $\mathscr{H} = \{R_p f_n : p \in \mathbb{N}, n \in \mathbb{N}\}$. This is also a countable subset of $C(E)$ that separates the points of E, since $\|pR_p f - f\| \to 0$ as $p \to \infty$.

Let D be a countable dense subset of $\mathbb{R}_+$. By Proposition 6.10, if $h \in \mathscr{H}$, there exists $p \in \mathbb{N}$ such that $(e^{-pt} h(X_t))_{t\geq 0}$ is a supermartingale. By Theorem 3.55 (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 $N = \bigcup_{h\in\mathscr{H}} N_h$, hence $(h(X_t))_{t\in D}$ has side limits on $\Omega \backslash N$. Then we define

$$\begin{cases} \widetilde{X}_t(\omega) = \lim_{D\ni s\downarrow t} X_s(\omega), & \omega \in \Omega \backslash N \\ \widetilde{X}_t(\omega) = x_0, & \omega \in N \end{cases}$$

where $x_0 \in E$ is a fixed point. Clearly this is a càdlàg process.

Finally, it remains to show that $(\widetilde{X}_t)_{t\geq 0}$ 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

$$\mathbb{E}[f(X_t) g(\widetilde{X}_t)] = \lim_{n\to\infty} \mathbb{E}[f(X_t) g(X_{t_n})] = \lim_{n\to\infty} \mathbb{E}[f(X_t) Q_{t_n - t} g(X_t)] = \mathbb{E}[f(X_t) g(X_t)].$$

By functional monotone class theorem, $\mathbb{E}[\varphi(X_t, \widetilde{X}_t)] = \mathbb{E}[\varphi(X_t, X_t)]$ for all bounded Borel function on $E \times E$. We take $\varphi(x, y) = \mathbb{1}_{\{x=y\}}$, which gives $\mathbb{P}(X_t = \widetilde{X}_t) = 1$. Hence $(\widetilde{X}_t)_{t\geq 0}$ is a modification of $(X_t)_{t\geq 0}$. □

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 $(\mathscr{F}_t)_{t\geq 0}$. Set $\widetilde{\mathscr{F}}_\infty = \mathscr{F}_\infty$, and denote by $\mathscr{N}$ the class of all zero probability sets in $\widetilde{\mathscr{F}}_\infty$. Define the modified filtration: $\widetilde{\mathscr{F}}_t = \sigma\left(\mathscr{F}_{t+} \cup \sigma(\mathscr{N})\right),\ \forall t \geq 0$.*

The process $(X_t)_{t\geq 0}$ has a càdlàg modification $(\widetilde{X}_t)_{t\geq 0}$, such that $(\widetilde{X}_t)_{t\geq 0}$ is a Feller process with respect to the modified filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$.

Proof. Let $E^* = E \cup \{\infty\}$ be the Alexandroff compactification of E. According to Lemma 6.11, and note that $N_h \in \mathscr{F}_\infty$ for all $h \in \mathscr{H}$, we can find a càdlàg modification $(\widetilde{X}_t)_{t\geq 0}$ taking values in E^* and adapted to the filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$. We also point out that the filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$ is right continuous, so the stopping time we are about to define makes sense.

It is necessary to show $(\widetilde{X}_t)_{t\geq 0}$ 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 $\left(e^{-t}h(\widetilde{X}_t)\right)_{t\geq 0}$ is a nonnegative and càdlàg supermartingale with respect to the filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$. We set

$$\tau_n = \inf\left\{t \geq 0 : e^{-t}h(\widetilde{X}_t) < \frac{1}{n}\right\}, \quad \textit{and} \quad \tau = \lim_{t\to\infty} \tau_n,$$

which are stopping times with respect to the filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$ by Proposition 3.14. By optional sampling theorem [Theorem 3.60], we have

$$\mathbb{E}\left[e^{-t}h(\widetilde{X}_t)\right] \leq \mathbb{E}\left[e^{-\tau_n\wedge t}h(\widetilde{X}_{\tau_n\wedge t})\right] \quad \Rightarrow \quad \mathbb{E}\left[\mathbb{1}_{\{\tau_n\leq t\}}e^{-t}h(\widetilde{X}_t)\right] \leq \mathbb{E}\left[\mathbb{1}_{\{\tau_n\leq t\}}e^{-\tau_n}h(\widetilde{X}_{\tau_n})\right] \leq \frac{1}{n}.$$

Letting n increase to ∞, we have

$$\mathbb{E}\left[\mathbb{1}_{\{\tau\leq t\}}e^{-t}h(\widetilde{X}_t)\right] \leq 0.$$

Since $\mathbb{P}(\widetilde{X}_t = \infty) = 0$, and since h is strictly positive, we have $\tau > t$ *a.s.*. Note that $t > 0$ is arbitrary. Then $\tau_n \to \infty$ *a.s.*, and $\inf_{s\in[0,t]} e^{-s}h(\widetilde{X}_s) > 0$ *a.s.* for all $t > 0$. As a result, almost surely, we have $\widetilde{X}_{s-} \neq \infty$ and $\widetilde{X}_s \neq \infty$ for all $s > 0$. This extends càdlàg property to E.

Finally we verify that $(\widetilde{X}_t)_{t\geq 0}$ is a Markov process with semigroup $(Q_t)_{t\geq 0}$ with respect to the filtration $(\widetilde{\mathscr{F}}_t)_{t\geq 0}$. It suffices to prove that, for all $s \geq 0, t > 0$ and $A \in \widetilde{\mathscr{F}}_s$, $f \in C_0(E)$, we have

$$\mathbb{E}\left[\mathbb{1}_A f(\widetilde{X}_{s+t})\right] = \mathbb{E}\left[\mathbb{1}_A Q_t f(\widetilde{X}_s)\right]. \tag{6.4}$$

We may assume $A \in \mathscr{F}_{s+}$ since it *a.s.* equals to some $\mathscr{F}_{s+}$-set. Taking $D \ni s_n \downarrow s$, we have

$$\mathbb{E}\left[\mathbb{1}_A f(X_{s+t})\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_A f(X_{s+t})|\mathscr{F}_{s_n}\right]\right] = \mathbb{E}\left[\mathbb{1}_A Q_{t+s-s_n} f(X_{s_n})\right]. \tag{6.5}$$

Since $Q_{t+s-s_n}f$ converges uniformly to $Q_t f$, and $X_{s_n} \overset{a.s.}{=} \widetilde{X}_{s_n} \overset{a.s.}{\to} \widetilde{X}_s \overset{a.s.}{=} X_s$, setting $n \to \infty$ in (6.5) gives $\mathbb{E}[\mathbb{1}_A f(X_{s+t})] = \mathbb{E}[\mathbb{1}_A Q_t(X_s)]$. As $(\widetilde{X}_t)_{t\geq 0}$ is a modification of $(X_t)_{t\geq 0}$, this is equivalent to (6.4). ☐

Remark. In fact, we prove the existence of a Feller process with càdlàg sample paths in this theorem. Assume we are given a process $(X_t)_{t\geq 0}$ together with a family $(P_x)_{x\in E}$ of probability measures such that, under P_x, $(X_t)_{t\geq 0}$ is a Markov process with semigroup $(Q_t)_{t\geq 0}$ with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$, and $X_0 = x$ a.s.. Then we define a new filtration $(\mathscr{F}_t)_{t\in[0,\infty]}$ by $\widetilde{\mathscr{F}}_t = \sigma\left(\mathscr{F}_{t+} \cup \sigma(\mathscr{N}')\right)$, where $\mathscr{N}'$ is the class of all $\mathscr{F}_\infty$-sets that have zero P_x probability for each $x \in E$. By the same arguments as in the preceding proof, we can then constructed an $(\widetilde{\mathscr{F}}_t)$-adapted càdlàg process $(\widetilde{X}_t)_{t\geq 0}$ such that for all $x \in E$,

- $P_x(\widetilde{X}_t = X_t) = 1$ for all $t \geq 0$, and
- under P_x, $(\widetilde{X}_t)_{t\geq 0}$ is a Feller process with semigroup $(Q_t)_{t\geq 0}$ with respect to the filtration $(\widetilde{\mathscr{F}}_t)_{t\in[0,\infty]}$.

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 $\Phi : \mathbb{D}(E) \to \mathbb{R}_+$ 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 $A \in \mathscr{D}$. If A depends on only finitely many coordinate maps:

$$A = \left\{ f \in \mathbb{D}(E) : f(t_1) \in B_1, \cdots, f(t_p) \in B_{t_p} \right\}, \quad \text{where } \ 0 \leq t_1 < \cdots < t_p, \ \text{ and } \ B_1, \cdots, B_p \in \mathscr{E}, \tag{6.6}$$

then the mapping $x \mapsto \mathbb{E}_x[\mathbb{1}_A]$ has an explicit form:

$$\mathbb{E}_x[\mathbb{1}_A] = \int_{B_1} Q_{t_1}(x, \mathrm{d}y_1) \int_{B_2} Q_{t_2 - t_1}(y_1, \mathrm{d}y_2) \cdots \int_{B_p} Q_{t_p - t_{p-1}}(y_{p-1}, \mathrm{d}y_p),$$

which is measurable. The remaining case then follows from a π-λ theorem argument.

In addition, for any initial law, the expectation $\mathbb{E}_\gamma$ taken with respect to P^γ is given by

$$\mathbb{E}_\gamma[\mathbb{1}_A] = \int \gamma(\mathrm{d}x) \int_{B_1} Q_{t_1}(x, \mathrm{d}y_1) \int_{B_2} Q_{t_2 - t_1}(y_1, \mathrm{d}y_2) \cdots \int_{B_p} Q_{t_p - t_{p-1}}(y_{p-1}, \mathrm{d}y_p) = \int_E \mathbb{E}_x[\mathbb{1}_A]\gamma(\mathrm{d}x).$$

Using an argument of π-λ system and a simple function approximation, we obtain

$$\mathbb{E}_\gamma[\Phi] = \int_E \mathbb{E}_x[\Phi]\gamma(\mathrm{d}x), \quad \forall \text{ measurable mapping } \Phi : (\mathbb{D}(E), \mathscr{D}) \to (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+)).$$

6.2.2 Markov Properties

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 $(\mathscr{F}_t)_{t\geq 0}$. Assume that process $(X_t)_{t\geq 0}$ has càdlàg sample paths. Let $s \geq 0$, and let $\Phi : \mathbb{D}(E) \to \mathbb{R}_+$ be a measurable map. Then*

$$\mathbb{E}\left[\Phi((X_{s+t})_{t\geq 0}) | \mathscr{F}_s\right] = \mathbb{E}_{X_s}\Phi \quad a.s..$$

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 $\varphi_1, \cdots, \varphi_p \in B(E)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\varphi_1(X_{s+t_1}) \cdots \varphi_p(X_{s+t_p}) | \mathscr{F}_s\right] \\
&= \mathbb{E}\left[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) \mathbb{E}\left[\varphi_p(X_{s+t_p}) | \mathscr{F}_{s+t_{p-1}}\right] | \mathscr{F}_s\right] \\
&= \mathbb{E}\left[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) Q_{t_p - t_{p-1}} \varphi_p(X_{s+t_{p-1}}) | \mathscr{F}_s\right] \\
&= \mathbb{E}\left[\varphi_1(X_{s+t_1}) \cdots \varphi_{p-1}(X_{s+t_{p-1}}) \int_E Q_{t_p - t_{p-1}}(X_{s+t_{p-1}}, \mathrm{d}y_p) \varphi_p(y_p) \, \mathrm{d}y_p | \mathscr{F}_s\right] \\
&= \cdots = \int_E Q_{t_1}(X_s, \mathrm{d}y_1)\varphi_1(y_1) \int_E Q_{t_2 - t_1}(y_1, \mathrm{d}y_2)\varphi_2(y_2) \cdots \int_E Q_{t_p - t_{p-1}}(y_{p-1}, \mathrm{d}y_p)\varphi_p(y_p).
\end{aligned}$$

Taking $\varphi_j = \mathbb{1}_{B_j}$ immediately gives the desired conclusion according to (6.6). □

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 $(\mathscr{F}_t)_{t\geq 0}$. Assume that process $(X_t)_{t\geq 0}$ has càdlàg sample paths. Let τ be a stopping time of the filtration $(\mathscr{F}_{t+})_{t\geq 0}$, and let $\Phi. : \mathbb{R}_+ \times \mathbb{D}(E) \to \mathbb{R}_+$ be a measurable map. Then*

$$\mathbb{E}\left[\mathbb{1}_{\{\tau < \infty\}} \Phi_\tau((X_{\tau+t})_{t\geq 0}) | \mathscr{F}_{\tau+}\right] = \mathbb{1}_{\{\tau < \infty\}} \mathbb{E}_{X_\tau}\Phi_\tau \quad a.s..$$

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 $\mathscr{F}_\tau$-measurable by Proposition 3.13, and the mapping $x \mapsto \mathbb{E}_x[\Phi]$ is measurable. To show the desired conclusion, it suffices to show that for any $A \in \mathscr{F}_{\tau+}$,

$$\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\Phi_\tau((X_{\tau+t})_{t\geq 0})\right] = \mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\mathbb{E}_{X_\tau}\Phi_\tau\right].$$

We approximate τ by the smallest multiple of 2^{-n} greater than τ, i.e. $\tau_n = \frac{\lfloor 2^n\tau\rfloor+1}{2^n}$. We consider the mapping $\Phi_s(f) = \varphi_0(s)\prod_{i=1}^m \varphi_i(f(t_i))) \geq 0$, where φ_0 is nonnegative, bounded and measurable on $\mathbb{R}_+$, and $\varphi_1, \cdots, \varphi_m \in B(E)$. By the previous conclusion,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau_n<\infty\}}\Phi_{\tau_n}((X_{\tau_n+t})_{t\geq 0})\right] &= \sum_{k=1}^{\infty}\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau_n=k2^{-n}\}}\Phi_{k2^{-n}}((X_{k2^{-n}+t})_{t\geq 0})\right] \\
&= \sum_{n=1}^{\infty}\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau=k2^{-n}\}}\mathbb{E}\left[\Phi_{k2^{-n}}((X_{k2^{-n}+t})_{t\geq 0})|\mathscr{F}_{k2^{-n}}\right]\right] \\
&= \sum_{n=1}^{\infty}\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau=k2^{-n}\}}\mathbb{E}_{X_{k2^{-n}}}\Phi_{k2^{-n}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\mathbb{E}_{X_{\tau_n}}\Phi_{\tau_n}\right] = \mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\psi(\tau_n, X_{\tau_n})\right].
\end{aligned}$$

where $\psi(s,x) = \mathbb{E}_x\Phi_s = \varphi_0(s)\prod_{i=1}^m Q_{t_i}\varphi_i(x)$. According to Riesz-Markov Theorem, since a finite signed measure on E is determined by its values against all C_0 functions on E, we may assume $\varphi_1, \cdots, \varphi_n \in C_0(E)$. We also assume $\varphi_0 \in C_0(\mathbb{R}_+)$. Then ψ is a bounded and continuous map on $\mathbb{R}_+\times E$. By dominated convergence theorem and right-continuity of $t \mapsto X_t$, we let $n \to \infty$ to obtain

$$\mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\Phi_\tau((X_{\tau+t})_{t\geq 0})\right] = \mathbb{E}\left[\mathbb{1}_{A\cap\{\tau<\infty\}}\mathbb{E}_{X_\tau}\Phi_\tau\right].$$

By monotone class theorem [Theorem 1.38], we concludes the proof of strong Markov property. ☐

Remark. In this theorem, we assume that τ is a stopping time of the filtration $(\mathscr{F}_{t+})_{t\geq 0}$. This is a more general assumption than a stopping time of the filtration $(\mathscr{F}_t)_{t\geq 0}$. Where we use this assumption is that by Proposition 3.11,

$$\{\tau_n = k2^{-n}\} = \{(k-1)2^{-n} \leq \tau < k2^{-n}\} \in \mathscr{F}_{k2^{-n}},$$

and since $A \in \mathscr{F}_{\tau+}$,

$$A \cap \{(k-1)2^{-n} \leq \tau < k2^{-n}\} \in \mathscr{F}_{k2^{-n}}.$$

We can use this statement to generalize the Blumenthal's 0-1 law.

Theorem 6.15 (Blumenthal's 0-1 law)**.** *Let $(X_t)_{t\geq 0}$ be a Feller process with càdlàg sample paths, and let $(\mathscr{F}_t)_{t\geq 0}$ be the canonical filtration. Then for each $x \in E$, the germ σ-algebra $\mathscr{F}_{0+}$ is trivial under $\mathbb{P}_x$, i.e. $\mathbb{P}_x(A) \in \{0, 1\}$ for each $A \in \mathscr{F}_{0+}$.*

Proof. For each $A \in \mathscr{F}_{0+}$, by the strong Markov property,

$$\mathbb{1}_A = \mathbb{E}_x\left[\mathbb{1}_A|\mathscr{F}_{0+}\right] = \mathbb{E}_x\mathbb{1}_A = \mathbb{P}_x(A).$$

Hence $\mathbb{P}_x(A)$ is either 0 or 1. ☐

6.3 The Generators and the Feynman-Kac Formula

In this section, we are going to discuss some properties of generators. We first study the Brownian motion, whose generator has a simple form.

Example 6.16. *A Brownian motion $B = (B_t)_{t\geq 0}$ is a real-valued Markov process with transition semigroup*

$$Q_0(x,A) = \mathbb{1}_A(x), \quad Q_t(x,A) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}\,\mathrm{d}y, \quad \forall t>0,\ x\in\mathbb{R},\ A\in\mathscr{B}(\mathbb{R}).$$

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,

$$\lim_{x\to x_0} Q_t f(x) = \lim_{x\to x_0}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} f(y)\,\mathrm{d}y = Q_t f(x_0).$$

Given any $\epsilon>0$, we choose $K>0$ such that $|f(y)|<\epsilon/2$ for all $|y|>K$, and choose $\alpha>0$ such that

$$\int_\alpha^\infty \frac{1}{\sqrt{2\pi t}} e^{\frac{-z^2}{2t}}\,\mathrm{d}z < \frac{\epsilon}{2M} \quad\Leftrightarrow\quad \int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi t}} e^{\frac{-z^2}{2t}}\,\mathrm{d}z < \frac{\epsilon}{2}.$$

Then for all $x>K+\alpha$, we have $|f(z+x)|<\epsilon/2$ for all $z>-\alpha$, and

$$|Q_t f(x)| = \left|\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y)\,\mathrm{d}y\right| \leq M\int_{-\infty}^{-\alpha} \frac{1}{\sqrt{2\pi t}} e^{\frac{-z^2}{2t}}\,\mathrm{d}z + \int_{-\alpha}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{\frac{-z^2}{2t}} |f(z+x)|\,\mathrm{d}z < \alpha.$$

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 $|y-x_0|\leq\delta$:

$$\left|\int_{|y-x_0|\leq\delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} (f(y)-f(x_0))\,\mathrm{d}y\right| < \frac{\eta}{2}\left|\int_{|y-x_0|\leq\delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}}\,\mathrm{d}y\right| \leq \frac{\eta}{2}, \quad \forall t>0;$$

and choose $t_0>0$ such that

$$\int_{|z|\geq\delta} \frac{1}{\sqrt{2\pi t_0}} e^{-\frac{z^2}{2t_0}}\,\mathrm{d}z < \frac{\eta}{4M} \quad\Rightarrow\quad \left|\int_{|y-x_0|\geq\delta} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}} (f(y)-f(x_0))\,\mathrm{d}y\right| < \frac{\eta}{2}, \quad \forall t\in(0,t_0).$$

Consequently, we have $|(Q_t f - f)(x_0)|<\eta$ for all $t\in(0,t_0)$, and $Q_t f(x_0)\to f(x_0)$ as $t\downarrow 0$.

Resolvent and generator. For $\lambda>0$, the resolvent is

$$\begin{aligned} R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x)\,\mathrm{d}t &= \int_{-\infty}^{\infty} \underbrace{\left(\int_0^\infty \frac{1}{\sqrt{2\pi t}} e^{-\lambda t - \frac{(y-x)^2}{2t}}\,\mathrm{d}t\right)}_{r_\lambda(y-x)} f(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{1}{|y-x|} \mathbb{E}\left[\tau_y e^{-\lambda\tau_y}\right] f(y)\,\mathrm{d}y, \end{aligned}$$

where the hitting time $\tau_y = \inf\{t\geq 0: B_t = |y-x|\}$ has density $\frac{|y-x|}{\sqrt{2\pi t^3}} e^{-\frac{(y-x)^2}{2t}}\,\mathrm{d}t$. By Proposition 4.38,

$$r_\lambda(y-x) = \frac{1}{|y-x|}\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[e^{-\lambda\tau_y}] = \frac{1}{\sqrt{2\lambda}} e^{-|y-x|\sqrt{2\lambda}}.$$

This gives the formula of R_λ:

$$R_\lambda f(x) = \frac{1}{\sqrt{2\lambda}} \int_{-\infty}^{\infty} e^{-|y-x|\sqrt{2\lambda}} f(y)\,\mathrm{d}y.$$

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 $\mathrm{sgn}(z) = \mathbb{1}_{\{z\geq 0\}} - \mathbb{1}_{\{z<0\}}$, we have

$$h(x) = \int_{-\infty}^{\infty} e^{-|y-x|} f(y)\,\mathrm{d}y \quad \Rightarrow \quad h'(x) = \int_{-\infty}^{\infty} \mathrm{sgn}(y-x) e^{-|y-x|} f(y)\,\mathrm{d}y.$$

Furthermore, h' is differentiable: for all $x \in \mathbb{R}$,

$$\begin{aligned} h'(x+\delta) - h'(x) &= \int_{-\infty}^{\infty} \mathrm{sgn}(y-x-\delta) e^{-|y-x-\delta|} f(y)\,\mathrm{d}y - \int_{-\infty}^{\infty} \mathrm{sgn}(y-x) e^{-|y-x|} f(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}\setminus[x,x+\delta]} \mathrm{sgn}(y-x) \left(e^{-|y-x-\delta|} - e^{-|y-x|}\right) f(y)\,\mathrm{d}y - \int_x^{x+\delta} \left(e^{y-x-\delta} + e^{x-y}\right) f(y)\,\mathrm{d}y, \end{aligned}$$

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

$$\left(\frac{1}{2}\,\mathrm{Id} - L\right) h = \left(\frac{1}{2}\,\mathrm{Id} - L\right) R_{1/2} f = f \;\Rightarrow\; Lh = \frac{1}{2} h'', \quad \textit{where}\;\; h \in \mathfrak{D}(L) \subset \left\{h \in C^2(\mathbb{R}) : g, g'' \in C_0(\mathbb{R})\right\}.$$

In fact, we can show that $\mathfrak{D}(L) = \left\{g \in C^2(\mathbb{R}) : g, g'' \in C_0(\mathbb{R})\right\}$. 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 6.17. *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

$$Q_t f(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-|z|^2/2} f(x + z\sqrt{t})\,\mathrm{d}z, \tag{6.7}$$

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

$$\begin{aligned} Q_t f(x) &= f(x) + (2\pi)^{-d/2} \frac{t}{2} \int_{\mathbb{R}^d} e^{-|z|^2/2} z^\top \nabla^2 f(x + \theta(t,z) z\sqrt{t}) z\,\mathrm{d}z \\ &= f(x) + \frac{t}{2}\Delta f(x) + (2\pi)^{-d/2} \frac{t}{2} J(t,x), \end{aligned}$$

where

$$J(t,x)=\int_{\mathbb{R}^d} e^{-|z|^2/2}\left[\sum_{i,j=1}^{d}\left(\frac{\partial^2 f}{\partial x_i\partial x_j}(x+\theta(t,z)z\sqrt{t})-\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\right)z_iz_j\right]\mathrm{d}z.$$

We set the following function F, and use uniform continuity of second partial derivatives of f:

$$F(x,z,t)=\max_{i,j\in[d]}\left|\frac{\partial^2 f}{\partial x_i\partial x_j}(x+\theta(t,z)z\sqrt{t})-\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\right|\quad\Rightarrow\quad\limsup_{t\downarrow 0}F(x,z,t)=0.$$

Moreover, for any $R>0$, one have

$$|J(t,x)|\le\int_{|z|\le R}F(x,t,z)\,e^{-|z|^2/2}\|z\|_1^2\,\mathrm{d}z+2\max_{i,j\in[d]}\left\|\frac{\partial^2 f}{\partial x_i\partial x_j}\right\|\int_{|z|>R}e^{-|z|^2/2}\|z\|_1^2\,\mathrm{d}z.$$

Since the first term converges to 0 as $t\downarrow 0$ by dominated convergence theorem for any $R>0$, we have

$$\lim_{t\downarrow 0}\left\|\frac{1}{t}(Q_tf-f)-\frac{1}{2}\Delta f\right\|\le\max_{i,j\in[d]}\left\|\frac{\partial^2 f}{\partial x_i\partial x_j}\right\|\int_{|z|>R}e^{-|z|^2/2}\|z\|_1^2\,\mathrm{d}z,\quad\forall R>0.$$

Take $R\to\infty$. Then we have $Lf=\frac{1}{2}\Delta f$ for all $f\in C_0^2(\mathbb{R})$. □

Remark. For the case $d\ge 2$, 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 $C_0(\mathbb{R}^d)$.

Generally, we have the following relationship between Brownian motions and heat equations.

Theorem 6.18 (Brownian motions and the heat equation)**.** *Let $\varphi\in C(\mathbb{R}^d)$. Then the function $u:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}$, $(t,x)\mapsto Q_t\varphi(x)$ solves the following heat equation:*

$$\begin{cases}\frac{\partial u}{\partial t}(t,x)=\frac{1}{2}\Delta_x u(t,x), & t>0\\ u(0,x)=\varphi(x),\end{cases}$$

where $(Q_t)_{t\ge 0}$ is defined in (6.7)*.*

Proof. We write $u_t=Q_t\varphi$ for $t>0$, so $u_t\in C^\infty(\mathbb{R}^d)$. For each $t>0$, we fix $0<\epsilon<t$. By Proposition 6.8,

$$u_t=Q_\epsilon\varphi+\int_0^{t-\epsilon}\frac{1}{2}\Delta(Q_{s+\epsilon}\varphi)\,\mathrm{d}s=u_\epsilon+\int_\epsilon^t\frac{1}{2}\Delta u_s\,\mathrm{d}s$$

Since $\frac{1}{2}\Delta u_s=Q_{s-\epsilon}(\frac{1}{2}\Delta u_\epsilon)$ depends (uniformly) continuously on $s\in[\epsilon,\infty)$ we have

$$\frac{\partial u_t}{\partial t}=\lim_{s\to t}\frac{Q_t\varphi-Q_s\varphi}{t-s}=\frac{1}{2}\Delta u_t.$$

Furthermore, since $(Q_t)_{t\ge 0}$ is a Feller semigroup,

$$\lim_{t\downarrow 0}u_t=\lim_{t\downarrow 0}Q_t\varphi=\varphi.$$

Thus we conclude the proof. □

The following theorem gives a characterization for the generator domain of a Feller semigroup $(Q_t)_{t\ge 0}$. For any $x\in E$, we can construct a Markov process $(X_t^x)_{t\ge 0}$ such that $\mathbb{P}(X_0=x)=1$ with semigroup $(Q_t)_{t\ge 0}$.

Theorem 6.19. *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

$$h(X_t^x) - \int_0^t g(X_s^x)\,\mathrm{d}s$$

is a martingale with respect to the filtration $\mathscr{F}_t^x = \sigma(X_s^x, 0 \le s \le t)$.

Proof. (i) $\Rightarrow$ (ii): Let $h \in \mathfrak{D}(L)$ and $Lh = g$. By (6.1), we have

$$Q_t h = h + \int_0^t Q_s g\,\mathrm{d}s,$$

which implies

$$\mathbb{E}\left[h(X_{t+s}^x)|\mathscr{F}_t^x\right] = Q_s h(X_t^x) = h(X_t^x) + \int_0^s Q_r g(X_t^x)\,\mathrm{d}r. \tag{6.8}$$

Meanwhile, use the boundedness of g, we have

$$\mathbb{E}\left[\int_t^{t+s} g(X_r^x)\,\mathrm{d}r \middle| \mathscr{F}_t^x\right] = \int_t^{t+s} \mathbb{E}\left[g(X_r^x)|\mathscr{F}_t\right] = \int_t^{t+s} Q_{r-t} g(X_t^x)\,\mathrm{d}r = \int_0^s Q_r g(X_t^s)\,\mathrm{d}r. \tag{6.9}$$

Combining (6.8) and (6.9), we have the martingale property:

$$\mathbb{E}\left[h(X_{t+s}^x) - \int_0^{t+s} g(X_r^x)\,\mathrm{d}r \middle| \mathscr{F}_t^x\right] = h(X_t^x) - \int_0^t g(X_r^x), \quad \forall t+s > t \ge 0.$$

(ii) $\Rightarrow$ (i): For every $x \in E$ and every $t \ge 0$, we have

$$Q_t h(x) - \int_0^t Q_r g(x)\,\mathrm{d}r = \mathbb{E}\left[h(X_t^x) - \int_0^t g(X_r^x)\,\mathrm{d}r\right] = h(x),$$

where the first equality follows from Markov property, and the second from martingale property. Hence

$$\lim_{t\downarrow 0} \frac{Q_t h - h}{t} - g = \lim_{t\downarrow 0} \frac{1}{t} \int_0^t (Q_r g - g)\,\mathrm{d}r = 0.$$

Since $g \in C_0(E)$, we have $h \in \mathfrak{D}(L)$, and $Lh = g$. ☐

Now we introduce one special case of Feynman-Kac formula, which reveals the relation between parabolic partial differential equations and stochastic differential equations.

Theorem 6.20 (Feynman-Kac formula)**.** *Let $\upsilon \in C_0(E)$ be a nonnegative function. For every $t \ge 0$, define for every $\varphi \in B(E)$ and every $x \in E$ that*

$$K_t\varphi(x) = \mathbb{E}\left[\varphi(X_t^x)\exp\left(-\int_0^t \upsilon(X_s^x)\,\mathrm{d}s\right)\right],$$

where $(X_t^x)_{t\ge 0}$ is a càdlàg Feller process with semigroup $(Q_t)_{t\ge 0}$ starting from $X_0 \sim \delta_x$.

(i) $(K_t)_{t\ge 0}$ is a semigroup of contractions on $B(E)$.

(ii) If $\varphi \in \mathfrak{D}(L)$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} K_t\varphi|_{t=0} = L\varphi - \upsilon\varphi.$$

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: $K_t\varphi(x) \leq \mathbb{E}\left[\varphi(X_t)\right] = Q_t\varphi(x) \leq \|\varphi\|$. Next, we let $\pi_t(f) = f(t)$ be the projection map from $\mathbb{D}(E) \to E$, and fix $s, t \geq 0$. By definition of K_t and simple Markov property,

$$\begin{aligned}
K_s(K_t\varphi)(x) &= \mathbb{E}\left[(K_t\varphi)(X_s)\exp\left(-\int_0^s \upsilon(X_s)\,\mathrm{d}s\right)\right]\\
&= \mathbb{E}\left[\mathbb{E}_{X_s}\left[(\varphi\circ\pi_t)\exp\left(-\int_0^t \upsilon\circ\pi_r\,\mathrm{d}r\right)\right]\exp\left(-\int_0^s \upsilon(X_r)\,\mathrm{d}r\right)\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\varphi(X_{s+t})\exp\left(-\int_0^t \upsilon(X_{s+r})\,\mathrm{d}r\right)\middle|\,\mathscr{F}_s^X\right]\exp\left(-\int_0^s \upsilon(X_r)\,\mathrm{d}r\right)\right]\\
&= \mathbb{E}\left[\varphi(X_{s+t})\exp\left(-\int_0^{s+t} \upsilon(X_r)\,\mathrm{d}r\right)\right] = (K_{s+t}\varphi)(x).
\end{aligned}$$

(ii) Since $\frac{\mathrm{d}}{\mathrm{d}s}e^{-\int_s^t f(r)\,\mathrm{d}r} = f(s)e^{-\int_s^t f(r)\,\mathrm{d}r}$, the fundamental theorem of calculus gives

$$1 - e^{-\int_0^t \upsilon(X_r)\,\mathrm{d}r} = \int_0^t \upsilon(X_s)e^{-\int_s^t \upsilon(X_r)\,\mathrm{d}r}\,\mathrm{d}s.$$

By Fubini's theorem and simple Markov property, for every $\varphi \in B(E)$, we have

$$\begin{aligned}
K_t\varphi(x) &= \mathbb{E}\left[\varphi(X_t)\left(1 - \int_0^t \upsilon(X_s)\exp\left(-\int_s^t \upsilon(X_r)\,\mathrm{d}r\right)\,\mathrm{d}s\right)\right]\\
&= Q_t\varphi(x) - \int_0^t \mathbb{E}\left[\varphi(X_t)\upsilon(X_s)\exp\left(-\int_s^t \upsilon(X_r)\,\mathrm{d}r\right)\right]\,\mathrm{d}s\\
&= Q_t\varphi(x) - \int_0^t \mathbb{E}\left[\upsilon(X_s)\,\mathbb{E}\left[\varphi(X_t)\exp\left(-\int_s^t \upsilon(X_r)\,\mathrm{d}r\right)\middle|\,\mathscr{F}_s^X\right]\right]\,\mathrm{d}s\\
&= Q_t\varphi(x) - \int_0^t \mathbb{E}\left[\upsilon(X_s)K_{t-s}\varphi(X_s)\right]\,\mathrm{d}s\\
&= Q_t\varphi(x) - \int_0^t Q_s(\upsilon\, K_{t-s}\varphi)(x)\,\mathrm{d}s.
\end{aligned}$$

Hence $\lim_{t\downarrow 0} K_t\varphi = \varphi$. Furthermore,

$$\begin{aligned}
\left|\frac{1}{t}\int_0^t Q_s(\upsilon\, K_{t-s}\varphi)(x)\,\mathrm{d}s - \upsilon(x)\varphi(x)\right| &= \frac{1}{t}\left|\int_0^t Q_s(\upsilon\, K_{t-s}\varphi - \upsilon\varphi)(x)\,\mathrm{d}s\right| + \frac{1}{t}\left|\int_0^t (Q_s - \mathrm{Id})(\upsilon\varphi)(x)\,\mathrm{d}s\right|\\
&\leq \frac{1}{t}\int_0^t \|\upsilon K_{t-s}\varphi - \upsilon\varphi\|_\infty\,\mathrm{d}s + \frac{1}{t}\int_0^t \|(Q_s - \mathrm{Id})(\upsilon\varphi)\|_\infty\,\mathrm{d}s\\
&\leq \frac{1}{t}\int_0^t \|\upsilon\|_\infty\,\|K_s\varphi - \varphi\|_\infty\,\mathrm{d}s + \frac{1}{t}\int_0^t \|(Q_s - \mathrm{Id})(\upsilon\varphi)\|_\infty\,\mathrm{d}s
\end{aligned}$$

Letting $t \downarrow 0$, the last display converges to zero. Consequently, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}K_t\varphi|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}Q_t\varphi|_{t=0} - \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t Q_s(\upsilon\, K_{t-s}\varphi)(x)\,\mathrm{d}s|_{t=0} = L\varphi - \upsilon\varphi,$$

which complete the proof. □

Remark. To find the derivative of the mapping $t \mapsto K_t\varphi$, we use the semigroup property:

$$\frac{\mathrm{d}}{\mathrm{d}t}K_t\varphi = \frac{\mathrm{d}}{\mathrm{d}s}K_{s+t}\varphi|_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}K_s(K_t\varphi)|_{s=0} = LK_t\varphi - \upsilon K_t\varphi.$$

Therefore, for any $f \in B(E)$, the function

$$u(t,x) = K_t f(x) = \mathbb{E}\left[f(X_t^x) \exp\left(- \int_0^t \upsilon(X_s^x)\,\mathrm{d}s \right) \right]$$

solves the following initial value problem (which is often a parabolic PDE):

$$\begin{cases} \dfrac{\partial u}{\partial t} = Lu - \upsilon u, \quad t > 0 \\ u(0,x) = f(x). \end{cases}$$

Here (X_t^x) is a càdlàg Feller process with generator L starting from $X_0 \sim \delta_x$. Letting $\upsilon = 0$ 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 $E(\sigma, b)$

$$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t + b(X_t)\,\mathrm{d}t, \quad \sigma = (\sigma_{ij})_{i\in[p],j\in[q]},\ b = (b_i)_{i\in[p]}. \tag{6.10}$$

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 $b : \mathbb{R}^p \to \mathbb{R}^p$ is called the *drift coefficient*, and $\sigma : \mathbb{R}^p \to \mathbb{R}^{p\times q}$ 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.

6.4.1 Markovianity of Time-Independent SDEs

Theorem 6.21. *Assume that $X = (X_t)_{t\geq0}$ is a solution of SDE (6.10) on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq0}, \mathbb{P})$. Then $(X_t)_{t\geq0}$ is a Markov process with respect to the filtration $(\mathscr{F}_t)_{t\geq0}$, with semigroup*

$$Q_tf(x) = \mathbb{E}\left[f(X_t^x)\right], \quad t \geq 0,$$

where $X^x = (X_t^x)_{t\geq0}$ is an arbitrary solution of $E^x(\sigma, b)$. Using the notation of Theorem 5.24, we also write

$$Q_tf(x) = \int f(F_x(\mathbf{w})_t)\,W(\mathrm{d}\mathbf{w}). \tag{6.11}$$

Proof. We first prove that, for any $f \in B(\mathbb{R}^p)$ and any $s, t \geq 0$, we have

$$\mathbb{E}\left[f(X_{s+t})|\mathscr{F}_s\right] = Q_tf(X_s),$$

To deal the time shift s, we define filtration $(\mathscr{F}'_t)_{t\geq0}$ and processes $(X'_t)_{t\geq0}$, $(B'_t)_{t\geq0}$ as follows:

$$\mathscr{F}'_t = \mathscr{F}_{s+t}, \quad X'_t = X_{s+t}, \quad B'_t = B_{t+s} - B_s.$$

Then $(\mathscr{F}'_t)_{t\geq0}$ is still complete, X' is adapted to $(\mathscr{F}'_t)_{t\geq0}$, and B' is a q-dimensional $(\mathscr{F}'_t)$-Brownian motion. Furthermore, the approximation formula (5.8) for the integral of continuous adapted processes gives

$$\begin{aligned} X'_t = X_{s+t} &= X_s + \int_s^{s+t} \sigma(X_r)\,\mathrm{d}B_r + \int_s^{s+t} b(X_r)\,\mathrm{d}r \\ &= X_s + \int_0^t \sigma(X'_r)\,\mathrm{d}B'_r + \int_0^t b(X'_r)\,\mathrm{d}r. \end{aligned}$$

Consequently, X' solves $E(\sigma, b)$ on the space $(\Omega, \mathscr{F}, (\mathscr{F}'_t)_{t\geq0}, \mathbb{P})$ 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

$$\begin{aligned} \mathbb{E}\left[f(X_{s+t})|\mathscr{F}_s\right] = \mathbb{E}\left[f(X'_t)|\mathscr{F}_s\right] &= \mathbb{E}\left[f(F_{X_s}(B')_t)|\mathscr{F}_s\right] \\ &= \int f(F_{X_s}(\mathbf{w})_t)\,W(\mathrm{d}\mathbf{w}) = Q_tf(X_s), \end{aligned}$$

where in the third equality we use the independence of B' and $\mathscr{F}_s$.

Now it remains to verify that $(Q_t)_{t\geq 0}$ is a transition semigroup. Clearly, Q_0 is an identity map, and $(t,x)\mapsto Q_tf(x)$ is a continuous map, hence is measurable. Finally, one have

$$Q_{s+t}f(x)=\mathbb{E}\left[X_{s+t}^x\right]=\mathbb{E}\left[\mathbb{E}\left[X_{s+t}^x|\mathscr{F}_s\right]\right]=\mathbb{E}\left[Q_tf(X_s^x)\right]=\int Q_s(x,\mathrm{d}y)Q_tf(y)$$

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 6.22. *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 $t\geq 0$,*

$$\mathbb{E}\left[|X_t^x-x|^2\right]\leq C_xe^{C_x(t+t^2)}(t+t^2).$$

Proof. By triangle inequality and Lipschitz property, for all $t\geq 0$, we have

$$|\sigma(X_t^x)|^2\leq(|\sigma(x)|+|\sigma(X_t^x)-\sigma(x)|)^2\leq 2|\sigma(x)|^2+2K^2|X_t^x-x|^2. \tag{6.12}$$

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 $t\mapsto\mathbb{E}\left[|X_{t\wedge\tau}^x-x|^2\right]$ is bounded on $[0,T]$. For any $t\in[0,T]$, we have

$$\begin{aligned}\mathbb{E}\left[|X_{t\wedge\tau}^x-x|^2\right]&\leq 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau}\sigma(X_s^x)\,\mathrm{d}B_s\right)^2\right]+2\mathbb{E}\left[\left(\int_0^{t\wedge\tau}b(X_s^x)\,\mathrm{d}s\right)^2\right]\\&\leq 2\mathbb{E}\left[\int_0^{t\wedge\tau}|\sigma(X_s^x)|^2\,\mathrm{d}s\right]+2\mathbb{E}\left[T\int_0^{t\wedge\tau}|b(X_s^x)|^2\,\mathrm{d}s\right]\\&\leq 4T(|\sigma(x)|^2+T|b(x)|^2)+4K^2(1+T)\int_0^{t\wedge\tau}\mathbb{E}\left[|X_s^x-x|^2\right]\mathrm{d}s\\&\leq 4T(|\sigma(x)|^2+T|b(x)|^2)+4K^2(1+T)\int_0^{t}\mathbb{E}\left[|X_{s\wedge\tau}^x-x|^2\right]\mathrm{d}s.\end{aligned}$$

where we use (6.12) in the third inequality. By Gronwall's lemma [Lemma 5.22], we have

$$\mathbb{E}\left[|X_{t\wedge\tau}^x-x|^2\right]\leq 4T(|\sigma(x)|^2+T|b(x)|^2)e^{4K^2(1+T)t},\quad\forall t\in[0,T].$$

Let $M\to\infty$, and use the monotone convergence theorem. Then for all $t\geq 0$, we have

$$\mathbb{E}\left[|X_t^x-x|^2\right]\leq 4t\left(|\sigma(x)|^2+t|b(x)|^2\right)e^{4K^2(t+t^2)}.$$

Setting $C_x=4\max\left\{|\sigma(x)|^2,|b(x)|^2,K^2\right\}$ concludes the proof. $\square$

To move forward, we have the following conclusion.

Theorem 6.23. *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

$$\mathbb{P}(|X_t^x - x| > \lambda) \leq \frac{1}{\lambda^2}\mathbb{E}\left[|X_t^x - x|^2\right] \leq \frac{C_x e^{C_x(t+t^2)}(t+t^2)}{\lambda^2}, \quad \forall t \geq 0.$$

Then we have

$$|Q_t f(x)| \leq \left|\mathbb{E}\left[f(X_t^x)\mathbb{1}_{\{|X_t^x - x| \leq \lambda\}}\right]\right| + \|f\|\,\mathbb{P}(|X_t^x - x| > \lambda) \leq \sup_{y:\,|y-x|\leq\lambda} |f(y)| + \frac{C_x e^{C_x(t+t^2)}\,\|f\|\,(t+t^2)}{\lambda^2}$$

Since $f \in C_0(\mathbb{R}^p)$, the first term of the last display converges to 0 as $x \to \infty$ for all $\lambda > 0$. Then we have $\limsup_{x\to\infty} |Q_t f(x)| \leq C\,\|f\|\,(t^2+t)/\lambda^2$ for all $\lambda > 0$. This implies $Q_t f(x) \to 0$ as $x \to \infty$.

Now we fix $x \in \mathbb{R}^p$, and verify that $Q_t f(x) \to f(x)$ as $t \downarrow 0$. For any $\lambda > 0$,

$$\begin{aligned}
|\mathbb{E}\left[f(X_t^x)\right] - f(x)| &\leq \sup_{y:|y-x|\leq\lambda} |f(x) - f(y)| + 2\,\|f\|\,\mathbb{P}\left(|X_t^x - x| > \lambda\right) \\
&\leq \sup_{y:|y-x|\leq\lambda} |f(x) - f(y)| + \frac{2C_x e^{C_x(t+t^2)}\|f\|(t+t^2)}{\lambda^2} \\
&\to \sup_{y:|y-x|\leq\lambda} |f(x) - f(y)|, \quad \textit{as} \ \ t \downarrow 0.
\end{aligned}$$

Taking $\lambda \downarrow 0$ gives $Q_t f(x) \to f(x)$.

(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

$$\begin{aligned}
X_t^{x,i} &= x_i + \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x)\,\mathrm{d}B_s^k + \int_0^t b_i(X_s^x)\,\mathrm{d}s, \\
\langle X^{x,i}, X^{x,j}\rangle_t &= \sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x)\sigma_{jk}(X_s^x)\,\mathrm{d}s = \int_0^t (\sigma\sigma^*)_{ij}(X_s^x)\,\mathrm{d}s.
\end{aligned}$$

Using Itô's formula [Theorem 5.9] and associativity of stochastic integrals, one have

$$\begin{aligned}
&f(X_t^x) = f(x) + \sum_{i=1}^p \int_0^t \frac{\partial f}{\partial x_i}(X_s^x)\,\mathrm{d}X_s^{x,i} + \frac{1}{2}\sum_{i,j=1}^p \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x)\,\mathrm{d}\langle X^{x,i}, X^{x,j}\rangle_s \\
&= f(x) + \sum_{i=1}^p\sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x)\frac{\partial f}{\partial x_i}(X_s^x)\,\mathrm{d}B_s^k + \int_0^t \left[\sum_{i=1}^p b_i(X_s^x)\frac{\partial f}{\partial x_i}(X_s^x) + \frac{1}{2}\sum_{i,j=1}^p (\sigma\sigma^*)_{ij}(X_s^x)\frac{\partial^2 f}{\partial x_i \partial x_j}(X_s^x)\right]\mathrm{d}s.
\end{aligned}$$

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

$$M_t = \sum_{i=1}^p\sum_{k=1}^q \int_0^t \sigma_{ik}(X_s^x)\frac{\partial f}{\partial x_i}(X_s^x)\,\mathrm{d}B_s^k = f(X_t^x) - f(x) - \int_0^t g(X_s^x)\,\mathrm{d}s$$

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 3.67 (ii), M is a martingale. According to Theorem 6.19, $f \in \mathfrak{D}(L)$, 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 $u : \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}$ solves the following Kolmogorov backward equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\Sigma(x) \cdot \nabla_x^2 u(t,x) + b(x) \cdot \nabla_x u(t,x), \ t > 0 \\ u(0,x) = f(x). \end{cases}$$

Now we discuss two typical SDEs and their generators.

Example 6.24 (Ornstein-Uhlenbeck process)**.** *Let $\lambda > 0$. An one-dimensional Ornstein-Uhlenbeck process is the solution of the following SDE:*

$$\mathrm{d}X_t = \mathrm{d}B_t - \lambda X_t \,\mathrm{d}t.$$

By applying Itô's formula to $e^{\lambda t} X_t$, we have

$$e^{\lambda t} X_t = X_0 + \int_0^t e^{\lambda s} \mathrm{d}X_t + \lambda \int_0^t e^{\lambda s} X_s \,\mathrm{d}t,$$

which implies

$$X_t = X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)} \,\mathrm{d}B_t.$$

Conditional on X_0, the mean and covariance of $(X_t)_{t\ge 0}$ are given by

$$\mathbb{E}[X_t | X_0] = X_0 e^{-\lambda t}, \quad \mathrm{Cov}(X_s, X_t | X_0) = \frac{1}{2\lambda}\left(e^{-\lambda|s-t|} - e^{-\lambda(s+t)}\right).$$

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 $X = (X_t)_{t\ge 0}$ is given by $K(s,t) = \frac{1}{2\lambda} e^{-\lambda|s-t|}$. In this case, X is a stationary Gaussian process.

Example 6.25 (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:*

$$\mathrm{d}X_t = \sigma X_t \,\mathrm{d}B_t + \mu X_t \,\mathrm{d}t.$$

Since

$$X_t = X_0 + \int_0^\infty \sigma X_t \,\mathrm{d}B_t + \int_0^t \mu X_t \,\mathrm{d}t,$$

the quadratic variation of X is

$$\langle X, X \rangle_t = \sigma^2 \int_0^t X_t^2 \,\mathrm{d}t.$$

Assume $X_0 > 0$. By applying Itô's formula to $\log X_t$, we have

$$\log X_t = \log X_0 + \int_0^t \frac{1}{X_t} \,\mathrm{d}X_t - \frac{1}{2}\int_0^t \frac{1}{X_t^2}\,\mathrm{d}\langle X, X\rangle_t = \log X_0 + \sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right)t.$$

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

$$\mathbb{E}[X_t|X_0] = X_0 e^{\mu t}, \quad \mathrm{Cov}(X_s, X_t|X_0) = X_0^2 e^{\mu(s+t)}\left(e^{\sigma^2(s\wedge t)} - 1\right).$$

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 $0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n < \cdots$, we have independence of the successive ratios:

$$\frac{X_1 - X_0}{X_0}, \frac{X_2 - X_1}{X_1}, \cdots, \frac{X_n - X_{n-1}}{X_{n-1}}, \cdots,$$

which is a consequence of independent increment property of Brownian motions.

6.4.2 The Fokker-Planck Equation

Hermitian Adjoint of the Generator. Now we discuss the adjoint L^* of generator L in sense that

$$\int_{\mathbb{R}^p} Lf(x)g(x)\,\mathrm{d}x = \int_{\mathbb{R}^p} f(x)L^*g(x)\,\mathrm{d}x.$$

Since both f and g are compactly supported on $\mathbb{R}^p$, we can choose a common compact support $\Gamma \subset \mathbb{R}^p$. Using integration by parts,

$$\int_\Gamma \nabla f \cdot (bg)\,\mathrm{d}m = -\int_\Gamma f\nabla \cdot (bg)\,\mathrm{d}m,$$

and

$$\int_\Gamma \nabla^2 f \cdot (\Sigma g)\,\mathrm{d}m = -\int_\Gamma \nabla f \cdot (\nabla \cdot (\Sigma g))\,\mathrm{d}m = \int_\Gamma f\nabla^2 \cdot (\Sigma g)\,\mathrm{d}m.$$

Consequently, we have

$$L^* = \frac{1}{2}\nabla^2 \cdot \Sigma - \nabla \cdot b. \tag{6.13}$$

This operator is useful in our derivation of the Fokker-Planck equation, which reveals the dynamics of the probability density flow of a diffusion process.

Theorem 6.26 (Fokker-Planck equation)**.** *Let $(X_t)_{t\geq 0}$ be a diffusion process with diffusion $\sigma : \mathbb{R}^p \to \mathbb{R}$ and drift $b : \mathbb{R}^p \to \mathbb{R}$, 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 $t \geq 0$. If $\rho \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^p)$, then ρ solves the following Fokker-Planck equation:*

$$\begin{cases} \dfrac{\partial \rho}{\partial t} = L^* \rho, \quad t > 0 \\ \rho(0, x) = \rho_0(x). \end{cases}$$

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

$$\lim_{t \downarrow 0} \frac{1}{t} \left(\int_{\mathbb{R}^p} f(y) q(y|t, x) \, \mathrm{d}s - f(x) \right) = Lf(x) = \frac{1}{2} \Sigma(x) \cdot \nabla^2 f(x) + b(x) \cdot f(x), \quad \forall x \in \mathbb{R}.$$

For any $t > 0$, we use interchangeability of derivative and integration:

$$\begin{aligned} \int_{\mathbb{R}^p} f(y) \frac{\partial q}{\partial t}(y|t, x) \, \mathrm{d}y &= \frac{\partial}{\partial t} \int_{\mathbb{R}^p} f(y) q(y|t, x) \, \mathrm{d}y \\ &= \lim_{s \downarrow 0} \frac{1}{s} \int_{\mathbb{R}^p} \left(q(y|t + s, x) - q(y|t, x) \right) f(y) \, \mathrm{d}y \\ &= \lim_{s \downarrow 0} \frac{1}{s} \left(\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} q(z|t, x) q(y|s, z) f(y) \, \mathrm{d}z \, \mathrm{d}y - \int_{\mathbb{R}^p} q(z|t, x) f(z) \, \mathrm{d}y \right) \\ &= \lim_{s \downarrow 0} \frac{1}{s} \int_{\mathbb{R}^p} q(z|t, x) \left(\int_{\mathbb{R}^p} q(y|s, z) f(y) \, \mathrm{d}y - f(z) \right) \mathrm{d}z \\ &= \int_{\mathbb{R}^p} q(z|t, x) Lf(z) \, \mathrm{d}z = \int_{\mathbb{R}^p} L^* q(z|t, x) f(z) \, \mathrm{d}z. \end{aligned}$$

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:

$$\frac{\partial q}{\partial t}(y|t, x) = L^* q(y|t, x) = \frac{1}{2} \nabla^2 \cdot \Sigma(z) q(y|t, x) - \nabla \cdot b(y) q(y|t, x). \tag{6.14}$$

Now assume that $X_0 \sim \rho_0$. Then we have

$$\rho(t, y) = \int_{\mathbb{R}^p} q(y|t, x) \rho_0(x) \, \mathrm{d}x$$

By applying this integration to both sides of (6.14), one obtain the desired result. $\square$

6.5 Jump Processes

In this subsection, we study the Markov processes where the state space E is finite or countable and equipped with the discrete topology. Note that a càdlàg function $f \in \mathbb{D}(E)$ satisfies that for every $0 \le t < \infty$, there exists $\epsilon > 0$ such that $f(s) = f(t)$ for all $s \in [t, t+\epsilon)$.

Consider a Feller semigroup $(Q_t)_{t\ge 0}$ on E. According to our discussion in Section 6.2.1, we can construct a probability space Ω, a right-continuous filtration $(\mathscr{F}_t)_{t\in[0,\infty]}$, a family of probability measures $(P_x)_{x\in E}$ and a càdlàg process $(X_t)_{t\ge 0}$ such that under each P_x, $(X_t)_{t\ge0}$ is a Markov process with semigroup $(Q_t)_{t\ge0}$ with respect to the filtration $(\mathscr{F}_t)_{t\in[0,\infty]}$, and $P_x(X_0 = x) = 1$.

By the càdlàg property of sample paths, for P_x-a.e. $x \in E$, there exists a sequence of times

$$0 = T_0(\omega) < T_1(\omega) \le T_2(\omega) \le \cdots \le \infty$$

such that

(i) $X_t(\omega) = X_0(\omega)$ for each $t \in [0, T_1(\omega))$,

(ii) for each $j \ge 1$, the condition $T_j(\omega) < \infty$ implies that $X_{T_{j+1}}(\omega) \ne X_{T_j}(\omega)$, $T_{j+1}(\omega) > T_j(\omega)$ and $X_t(\omega) = X_{T_j}(\omega)$ for each $t \in [T_j(\omega), T_{j+1}(\omega))$, and

(iii) $T_j(\omega) \uparrow \infty$ as $j \to \infty$.

In other words, $T_j(\omega)$ are the moments at which $(X_t)_{t\ge0}$ jumps to another point. One can easily verify that $(T_n)_{n\in\mathbb{N}}$ are stopping times:

$$\{T_1 < t\} = \bigcup_{q\in[0,t)\cap\mathbb{Q}} \{X_q \ne X_0\} \in \mathscr{F}_t, \quad \{T_n < t\} = \left(\bigcap_{j=0}^{n-1} \{T_n < t\}\right) \cap \left(\bigcup_{q\in[T_{n-1},t)\cap\mathbb{Q}} \{X_q \ne X_0\}\right) \in \mathscr{F}_t.$$

We first study the law of the first jumping tine T_1.

Lemma 6.27. *For each $x \in E$, there exists $\lambda(x) \ge 0$ such that T_1 is exponentially distributed with parameter $\lambda(x)$ under P_x. (We make the convention that an exponential variable with parameter 0 is equal to ∞ a.s..) Furthermore, if $\lambda(x) > 0$, then X_{T_1} and T_1 are independent under P_x.*

Proof. We let $s, t \ge 0$, and define $\Phi(f) = \mathbb{1}_{\{f(r)=f(0),\ \forall r\in[0,t]\}}$ for $f \in \mathbb{D}(E)$. Using the simple Markov property [Theorem 6.13] and the fact that $X_s = x$ on the event $\{T_1 > s\}$, we have

$$\begin{aligned} P_x(T_1 > s+t) &= \mathbb{E}_x\left[\mathbb{1}_{\{T_1>s\}}\Phi((X_{s+r})_{r\ge0})\right] = \mathbb{E}_x\left[\mathbb{1}_{\{T_1>s\}}\mathbb{E}_{X_s}\Phi((X_r)_{r\ge0})\right] \\ &= \mathbb{E}_x\left[\mathbb{1}_{\{T_1>s\}}P_x(T_1>t)\right] = P_x(T_1>s)P_x(T_1>t). \end{aligned}$$

Since $T_1 > 0$ a.s., this memoryless property implies that T_1 is an exponential variable under P_x with parameter $\lambda(x) = \log\frac{1}{P_x(T_1>1)}$. If $\lambda(x) > 0$, we have $P_x(T_1 < \infty) = 1$. We fix $y \in E$ and consider the mapping

$$\Psi(f) = \begin{cases} 0, & f \text{ is constant} \\ \mathbb{1}_{\{\gamma_1(f)=y\}}, & f \text{ is non-constant, and } \gamma_1(f) \text{ is the value of } f \text{ after its first jump.} \end{cases}$$

Then

$$P_x(T_1 > t, X_{T_1} = y) = \mathbb{E}_x\left[\mathbb{1}_{\{T_1>t\}}\mathbb{E}_{X_t}\left[\Psi((X_r)_{r\ge0})\right]\right] = P_x(T_1>t)P_x(X_{T_1}=y),$$

which gives the desired independece. □

Remark. If $\lambda(x) = 0$, the point $x \in E$ is said to be an *absorbing state* of the Markov process, in the sense that

$$P_x(X_t = 0, \forall t \ge 0) = 1.$$

For each point $x \in E$ with $\lambda(x) > 0$, we set

$$\Pi(x, y) = P_x(X_{T_1} = y), \quad y \in E \backslash \{x\}.$$

Then $\Pi(x, \cdot)$ is a probability measure on E with $\Pi(x, x) = 0$.

Proposition 6.28. *Let L be the generator of $(Q_t)_{t\geq 0}$. Then $\mathfrak{D}(L) = C_0(E)$. For every $x \in E$ and $\varphi \in C_0(E)$,*

- *if $\lambda(x) = 0$, then $L\varphi(x) = 0$;*
- *if $\lambda(x) > 0$, then*

$$L\varphi(x) = \lambda(x) \sum_{y \in E\backslash\{x\}} \Pi(x, y)(\varphi(y) - \varphi(x)) = \sum_{y \in E} L(x, y)\varphi(y),$$

where

$$L(x, y) = \begin{cases} \lambda(x)\Pi(x, y), & y \in E\backslash\{x\}, \\ -\lambda(x), & y = x. \end{cases} \tag{6.15}$$

Proof. Let $\varphi \in C_0(E)$. If $\lambda(x) = 0$, it is trivial that $Q_t\varphi(x) = \mathbb{E}_x [\varphi(X_t)] = \varphi(x)$ and so

$$\lim_{t\downarrow 0} \frac{Q_t\varphi(x) - \varphi(x)}{t} = 0.$$

If $\lambda(x) > 0$, we use the strong Markov property of $(X_t)_{t\geq 0}$ to obtain

$$\begin{aligned} P_x(T_2 \leq t) &\leq P_x(T_1 \leq t, T_2 \leq T_1 + t) = \mathbb{E}_x \left[\mathbb{1}_{\{T_1 \leq t\}} P_{X_{T_1}}(T_1 \leq t)\right] \\ &= \mathbb{E}_x \left[\mathbb{1}_{\{T_1 \leq t\}} \left(1 - e^{-t\lambda(X_{T_1})}\right)\right] = P_x(T_1 \leq t)\, \mathbb{E}_x \left[1 - e^{-t\lambda(X_{T_1})}\right] \\ &= \left(1 - e^{-t\lambda(x)}\right) \sum_{y \in E\backslash\{x\}} \Pi(x, y) \left(1 - e^{-t\lambda(y)}\right). \end{aligned}$$

We fix any $\epsilon > 0$. Then there exists a finite subset $F \subset E\backslash\{x\}$ such that $\Pi(x, F) > 1 - \epsilon$, and

$$\lim_{t\downarrow 0} \frac{P_x(T_2 \leq t)}{t} \leq \lim_{t\downarrow 0} \frac{1 - e^{-t\lambda(x)}}{t} \left(\sum_{y \in F} \Pi(x, y) \left(1 - e^{-t\lambda(y)}\right) + \epsilon \right) = \lambda(x)\epsilon.$$

Next, we decompost $Q_t\varphi(x)$ as follows and use the independence of T_1 and X_{T_1} under P_x to obtain

$$\begin{aligned} Q_t\varphi(x) = \mathbb{E}_x[\varphi(X_t)] &= \mathbb{E}_x \left[\varphi(X_0)\mathbb{1}_{\{t<T_1\}}\right] + \mathbb{E}_x \left[\varphi(X_1)\mathbb{1}_{\{T_1\leq t\}}\right] - \mathbb{E}_x \left[\varphi(X_1)\mathbb{1}_{\{T_2\leq t\}}\right] + \mathbb{E}_x \left[\varphi(X_t)\mathbb{1}_{\{T_2\leq t\}}\right] \\ &= \varphi(x)e^{-t\lambda(x)} + \left(1 - e^{-t\lambda(x)}\right) \sum_{y \in E\backslash\{x\}} \Pi(x, y)\varphi(y) + \mathbb{E}_x \left[(\varphi(X_t) - \varphi(X_1))\, \mathbb{1}_{\{T_2\leq t\}}\right]. \end{aligned}$$

Combining the last two displays, we have

$$\lim_{t\downarrow 0} \left| \frac{Q_t\varphi(x) - \varphi(x)}{t} - \lambda(x) \sum_{y \in E\backslash\{x\}} \Pi(x, y) \left(\varphi(y) - \varphi(x)\right) \right| \leq \lim_{t\downarrow 0} 2\|\varphi\|_\infty \frac{P_x(T_2 \leq t)}{t} \leq 2\|\varphi\|_\infty \lambda(x)\epsilon.$$

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

Remark. By taking $\varphi = \mathbb{1}_{\{y\}}$, we can interprete $L(x, y)$ as the instantaneous transition from x to y:

$$L(x, y) = L\mathbb{1}_{\{y\}}(x) = \frac{\mathrm{d}}{\mathrm{d}t} P_x(X_t = y)\big|_{t=0}, \quad y \in E.$$

Construction of a jump process. In practice, one usually starts from the transition rates of the process. Assume we are given a collection $(\lambda(x))_{x\in E}$ of nonnegative numbers, and for every $x \in E$ with $\lambda(x) > 0$, a probability measure $\Pi(x,\cdot)$ on E with $\Pi(x,x) = 0$. Then we define the generator $L : E \times E \to \mathbb{R}$ via (6.15). If a corresponding semigroup $(Q_t)_{t\geq 0}$ exists, by Proposition 6.8, it must satisfy the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t(x,y) = Q_t L(x,y).$$

Hence $Q_t = e^{tL}$ in the sense of the exponential of matrices.

- Since $L\mathbb{1}_E = 0$, we have $Q_t\mathbb{1}_E = \mathbb{1}_E$. Hence $Q_t(x,\cdot)$ is a transition kernel.
- The property $e^{(t+s)L} = e^{tL}e^{sL}$ implies the Chapman-Kolmogorov equation for $(Q_t)_{t\geq 0}$. Hence $(Q_t)_{t\geq 0}$ is a transition semigroup on E.
- Clearly, $Q_t = e^{tL} \to \mathrm{Id}$ as $t \downarrow 0$. Hence $Q_t(x,\cdot)$ is a Feller semigroup.

After getting $(Q_t)_{t\geq 0}$, we can immediately construct a Feller process $(X_t)_{t\geq 0}$ with càdlàg sample paths.

The next proposition provides a complete description of the sample paths of $(X_t)_{t\geq 0}$ under P_x. For the sake of simplicity, we assume that there are no absorbing states.

Proposition 6.29. *Assume that $\lambda(y) > 0$ for every $y \in E$. Let $x \in E$. Then*

(i) P_x-a.s., the jump times $T_1 < T_2 < \cdots$ are all finite;

(ii) under P_x, the sequence $(X_0, X_{T_1}, X_{T_2}, \cdots)$ is a discrete-time Markov chain with transition kernel Π started from x;

(iii) conditional on $(X_0, X_{T_1}, X_{T_2}, \cdots)$, the interval times $(T_n - T_{n-1})_{n=1}^{\infty}$ are independent, and for every $n \in \mathbb{N}$, the conditional distribution of $T_n - T_{n-1}$ is exponential with parameter $\lambda(X_{T_{n-1}})$.

Proof. (i) Since $\lambda(x) > 0$, it is clear that $P_x(T_1 < \infty) = 1$. Let $\Phi = \mathbb{1}_{\{T_1<\infty\}}$. By the strong Markov property,

$$\begin{aligned} P_x(T_{n+1} < \infty) &= \mathbb{E}_x\left[\mathbb{1}_{\{T_n<\infty\}}\mathbb{1}_{\{T_{n+1}-T_n<\infty\}}\right] = \mathbb{E}_x\left[\mathbb{1}_{\{T_n<\infty\}}\Phi((X_{T_n+t})_{t\geq 0})\right] \\ &= \mathbb{E}_x\left[\mathbb{1}_{\{T_n<\infty\}}\mathbb{E}_{X_{T_n}}\Phi\right] = \mathbb{E}_x\left[\mathbb{1}_{\{T_n<\infty\}}P_{X_{T_n}}(T_1<\infty)\right] = P_x(T_n<\infty). \end{aligned}$$

where the last equality holds because $\lambda(y) > 0$ for all $y \in E$. By induction, every T_n is P_x-a.s. finite.

(ii) & (iii). Let $y, z \in E$ and $f_1, f_2 \in B(\mathbb{R}_+)$. By the strong Markov property,

$$\begin{aligned} &\mathbb{E}_x\left[\mathbb{1}_{\{X_{T_1}=y\}}f_1(T_1)\mathbb{1}_{\{X_{T_2}=z\}}f_2(T_2-T_1)\right] = \mathbb{E}_x\left[\mathbb{1}_{\{X_{T_1}=y\}}f_1(T_1)\mathbb{E}_{X_1}[\mathbb{1}_{\{X_{T_1}=z\}}f_2(T_1)]\right] \\ &= \Pi(x,y)\Pi(y,z)\int_0^\infty \mathrm{d}s_1 e^{-s_1\lambda(x)}f_1(s_1)\int_0^\infty \mathrm{d}s_2 e^{-s_2\lambda(y)}f_2(s_2). \end{aligned}$$

By induction, with the convention $y_0 = x$, for every $n \in \mathbb{N}$, $y_1, \cdots, y_n \in E$ and $f_1, \cdots, f_n \in B(\mathbb{R}_+)$,

$$\begin{aligned} &\mathbb{E}_x\left[\mathbb{1}_{\{X_{T_1}=y_1\}}\mathbb{1}_{\{X_{T_2}=y_2\}}\cdots\mathbb{1}_{\{X_{T_n}=y_n\}}f_1(T_1)f_2(T_2-T_1)\cdots f_n(T_n-T_{n-1})\right] \\ &= \Pi(y_0,y_1)\Pi(y_1,y_2)\cdots\Pi(y_{n-1},y_n)\prod_{k=1}^n\int_0^\infty \mathrm{d}s_k e^{-s_k\lambda(y_{k-1})}f_k(s_k). \end{aligned} \tag{6.16}$$

Hence the desired result follows. ☐

Remark. For a jump process with absorbing states, we have an easy extension of the above theorem. Let $A = \{y \in E : \lambda(y) = 0\}$, and let $\Pi(y,y) = 1$ for each $y \in A$. We define $X_\infty = \lim_{t\uparrow\infty} X_t$ if the limit exists, i.e. $(X_t)_{t\geq 0}$ hits an absorbing state. Then under P_x, the sequence $(X_0, X_{T_1}, X_{T_2}, \cdots)$ is a discrete-time Markov chain with transition kernel Π started from x. Furthermore, with the convention $\infty - \infty = \infty$, conditional on $(X_0, X_{T_1}, X_{T_2}, \cdots)$, the interval times $(T_n - T_{n-1})_{n=1}^\infty$ are independent, and for every $n \in \mathbb{N}$, the conditional distribution of $T_n - T_{n-1}$ is exponential with parameter $\lambda(X_{T_{n-1}})$.

6.5.1 Poisson Point Process.

Definition 6.30 (Lévy processes)**.** A *Lévy process* is a real-valued stochastic process $(X_t)_{t\geq 0}$ satisfying:

(i) $X_0 = 0$ a.s..

(ii) (Independent and stationary increments). For every $0 \leq s < t$, the increment $X_t - X_s$ is independent of $(X_r, 0 \leq r \leq s)$ and has the same law as X_{t-s}.

(iii) X_t converges in probability to 0 when $t \downarrow 0$.

For each $t \geq 0$, we denote by $Q_t(0, \mathrm{d}y)$ the law of X_t. For every $x \in \mathbb{R}$, we let $Q_t(x, \mathrm{d}y)$ be the image of $Q_t(0, \mathrm{d}y)$ under the translation $y \mapsto x + y$.

Proposition 6.31 (Markovianity of Lévy processes)**.** *The collection $(Q_t)_{t\geq 0}$ is a Feller semigroup, and $(X_t)_{t\geq 0}$ is a Markov process with semigroup $(Q_t)_{t\geq 0}$.*

Proof. Step I. We first verify that $(Q_t)_{t\geq 0}$ is a transition semigroup. Let $\varphi \in B(\mathbb{R})$ and $x \in \mathbb{R}$. For any $s, t \geq 0$, by property (ii), the law of $(X_t, X_{t+s} - X_t)$ is given by the product measure $Q_t(0, \cdot) \otimes Q_s(0, \cdot)$, and

$$\int_{\mathbb{R}} Q_t(x, \mathrm{d}y) \int_{\mathbb{R}} Q_s(y, \mathrm{d}z)\varphi(z) = \int_{\mathbb{R}} Q_t(0, \mathrm{d}y) \int_{\mathbb{R}} Q_s(0, \mathrm{d}z)\varphi(x + y + z)$$
$$= \mathbb{E}\left[\varphi(x + X_t + (X_{t+s} - X_t))\right] = \mathbb{E}\left[\varphi(x + X_{t+s})\right] = \int_{\mathbb{R}} Q_{t+s}(x, dz)\varphi(z).$$

Thus we establish the Chapman-Kolmogorov equation. The measurability of the mapping $(t, x) \mapsto Q_t(x, A)$ follows from the strong continuity we are going to establish in order to verify the Feller property.

Step II. We next verify the Feller property. If $\varphi \in C_0(\mathbb{R})$, by dominated convergence theorem, the mapping $x \mapsto Q_t\varphi(x) = \mathbb{E}\left[\varphi(x + X_t)\right]$ is continuous, and $\mathbb{E}[\varphi(x + X_t)] \to 0$ as $|x| \to \infty$. Hence $Q_t\varphi \in C_0(\mathbb{R})$.

For each $\epsilon > 0$, by uniform continuity of φ, there exists $\delta > 0$ such that $|\varphi(x) - \varphi(y)| < \epsilon$ for all $x, y \in \mathbb{R}$ with $|x - y| < \delta$. By property (iii),

$$\lim_{t\downarrow 0} |Q_t\varphi(x) - \varphi(x)| \leq \lim_{t\downarrow 0} \mathbb{E}\,|\varphi(x + X_t) - \varphi(x)|$$
$$\leq \lim_{t\downarrow 0} \mathbb{E}\left[|\varphi(x + X_t) - \varphi(x)|\, \mathbb{1}_{\{|X_t|\leq\delta\}}\right] + 2\|\varphi\|_\infty \lim_{t\downarrow 0} \mathbb{P}(|X_t| > \delta) \leq \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $Q_t\varphi \to \varphi$ as $t \downarrow 0$, and the convergence is uniform.

Step III. Finally we verify the second assertion. For every $s, t \geq 0$ and $\varphi \in B(\mathbb{R})$,

$$\mathbb{E}\left[\varphi(X_{t+s})|X_r, 0 \leq r \leq s\right] = \mathbb{E}\left[\varphi(X_s + (X_{t+s} - X_s))|X_r, 0 \leq r \leq s\right]$$
$$= \int_{\mathbb{R}} Q_t(0, \mathrm{d}y)\varphi(X_s + y) = \int_{\mathbb{R}} Q_t(X_s, \mathrm{d}y)\varphi(y) = Q_t\varphi(X_s).$$

Therefore $(X_t)_{t\geq 0}$ is a Markov process with semigroup $(Q_t)_{t\geq 0}$ □

The Poisson point process lies in the intersection of Lévy processes and jump processes.

Definition 6.32 (Poisson point processes)**.** Let $\lambda > 0$. A *Poisson point process* with intensity λ is an integer-valued stochastic process $(X_t)_{t\geq 0}$ satisfying:

(i) $X_0 = 0$ a.s..

(ii) (Independent and stationary increments). For every $0 \leq s < t$, the increment $X_t - X_s$ is independent of $(X_r, 0 \leq r \leq s)$ and has the same law as X_{t-s}.

(iii) For every $t > 0$, the law of X_t is Poisson:

$$\mathbb{P}(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \cdots.$$

Remark. According to Proposition 6.31, a Poisson point process $(X_t)_{t\geq 0}$ with intensity λ is a Markov process with Feller semigroup

$$Q_t(n, n+m) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \quad n, m = 0, 1, 2, \cdots.$$

For every $n \in \mathbb{N}_0$, the first jump time T_1 satisfies

$$P_n(T_1 > t) = P_n(X_t = n) = P_0(X_t = 0) = e^{-\lambda t}.$$

Hence the jump rate $\lambda(n) = \lambda$ for all $n \in \mathbb{N}_0$. Furthermore, the generator of $(Q_t)_{t\geq 0}$ is

$$L(n, n+m) = \frac{\mathrm{d}}{\mathrm{d}t} P_n(X_t = n+m)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} P_0(X_t = m)\Big|_{t=0} = \begin{cases} -\lambda, & m = 0, \\ \lambda, & m = 1, \\ 0, & m \geq 2. \end{cases}$$

According to (6.31), the transition probability Π is

$$\Pi(n, n+1) = 1, \quad \text{and} \quad \Pi(n, m) = 0 \quad \text{for all } m \neq n+1.$$

Therefore $X_{T_1} = 1$ a.s.. By Proposition 6.29, all jump times $T_1 < T_2 < \cdots$ are finite. By strong Markov property, we have $X_{T_n} = n$ a.s. for all $n \in \mathbb{N}$.

Proposition 6.33. *Let $(X_t)_{t\geq 0}$ be a Poisson point process with intensity $\lambda > 0$. Define the n^{th} arrival*

$$T_n = \inf\{t \geq 0 : X_t = n\}.$$

Then $(T_n - T_{n-1})_{n=1}^{\infty}$ are i.i.d. exponential variables with parameter λ.

Proof. According to our previous discussion, almost surely, T_n is the n^{th} jump time of the process $(X_t)_{t\geq 0}$. Furthermore, we have

$$(X_0, X_{T_1}, X_{T_2}, \cdots, X_{T_n}, \cdots) = (0, 1, 2, \cdots, n, \cdots) \quad a.s..$$

We then let $y_k = k$ for each $k \in \mathbb{N}_0$ in (6.16) to get

$$\mathbb{E}\left[f_1(T_1) f_2(T_2 - T_1) \cdots f_n(T_n - T_{n-1})\right] = \prod_{k=1}^{n} \int_0^{\infty} e^{-\lambda s} f_k(s)\,\mathrm{d}s, \quad \forall f_1, \cdots, f_n \in B(\mathbb{R}_+).$$

Thus we complete the proof. $\square$

Remark. Consequently, the law of the n^{th} arrival is the distribution Gamma(n, λ).

Proposition 6.34 (Conditioning)**.** *Let $U_1, \cdots, U_n$ be i.i.d. and uniform on $[0, t]$, and let $U^{(1)} \leq \cdots \leq U^{(n)}$ be the corresponding ordered statistics. Let $(X_t)_{t\geq 0}$ be a Poisson point process with intensity $\lambda > 0$. Conditioning on the event $\{X_t = n\}$, the arrival times $(T_1, \cdots, T_n)$ and $(U^{(1)}, \cdots, U^{(n)})$ are identically distributed.*

Proof. The event $\{X_t = n\}$ has probability $\frac{(\lambda t)^n}{n!} e^{-\lambda t}$. On this event, the joint density of $(T_1, \cdots, T_n)$ is

$$p\left(T_1 = t_1, \cdots, T_n = t_n, X_t = n\right) = \left(\prod_{k=1}^{n} \lambda e^{-\lambda(t_k - t_{k-1})}\right) e^{-\lambda(t - t_n)} = \lambda^n e^{-\lambda t}.$$

Dividing the above result by $\mathbb{P}(X_t = n)$, we know that the conditional density of $(T_1, \cdots, T_n)$ is $n!/t^n$ on the region $\{0 \leq t_1 \leq \cdots \leq t_n \leq t\}$. This is the distribution of $(U^{(1)}, \cdots, U^{(n)})$. $\square$

7 Local Times

In this section, we study the theory of local times of continuous semimartingales. Throughout this section, we fix a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ with a complete filtration $(\mathscr{F}_t)_{t\geq 0}$. Before we proceed, we review some properties of convex functions.

Proposition 7.1 (Convex functions)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then*

(i) f is locally Lipschitz continuous;

(ii) the right derivative

$$D^+ f(x) = \lim_{h\downarrow 0} \frac{f(x+h) - f(x)}{h}$$

exists for each $x \in \mathbb{R}$, and $D^+ f$ is right-continuous on $\mathbb{R}$; and

(iii) the left derivative

$$D^- f(x) = \lim_{h\downarrow 0} \frac{f(x) - f(x-h)}{h}$$

exists for each $x \in \mathbb{R}$, and $D^- f$ is left-continuous on $\mathbb{R}$.

Proof. (i) We first prove that f is locally bounded. Fix $0 < N < \infty$. Then for every $x \in [-N, N]$, we have

$$f(x) \leq \frac{x+N}{2N} f(N) + \frac{N-x}{2N} f(-N) \leq \max\{f(N), f(-N)\} < \infty,$$

and

$$f(0) \leq \frac{1}{2} f(x) + \frac{1}{2} f(-x) \leq \frac{1}{2} f(x) + \frac{1}{2} \max\{f(N), f(-N)\}.$$

Hence

$$2f(0) - \max\{f(N), f(-N)\} \leq \sup_{x\in[-N,N]} |f(x)| \leq \max\{f(N), f(-N)\}.$$

To prove local Lipschitz continuity, we fix $x, y \in [-N, N]$ with $x < y$. By convexity of f, we have

$$f(y) \leq f(x) + \frac{y-x}{2N-x}(f(2N) - f(y)) \leq f(x) + \frac{2(y-x)}{2N-x} \sup_{z\in[-2N,2N]} f(z).$$

and

$$f(x) \leq f(y) + \frac{y-x}{2N+y}(f(-2N) - f(y)) \leq f(y) + \frac{2(y-x)}{2N+y} \sup_{z\in[-2N,2N]} f(z).$$

Hence

$$\frac{|f(y) - f(x)|}{y-x} \leq \frac{2}{3N} \sup_{z\in[-2N,2N]} f(z),$$

which proves Lipschitz continuity.

(ii) For each $x \in \mathbb{R}$, we fix $x < y < z$. By convexity of f,

$$f(y) \leq \frac{y-x}{z-x} f(z) + \frac{z-y}{z-x} f(x).$$

Hence we have

$$\frac{f(y) - f(x)}{y-x} \leq \frac{f(z) - f(x)}{z-x}.$$

By local Lipschitz continuity of f, the net $\left(\frac{f(x+h)-f(x)}{h}\right)_{h>0}$ is bounded. Hence the limit

$$D^+ f(x) = \lim_{h\downarrow 0} \frac{f(x+h) - f(x)}{h}$$

exists for every $x \in \mathbb{R}$. Furthermore, $D^+ f$ is monotone increasing on $\mathbb{R}$, since

$$\frac{f(w)-f(x)}{w-x} \le \frac{f(z)-f(x)}{z-x} \le \frac{f(z)-f(y)}{z-y} \quad \text{for all } x < w < y < z.$$

Letting $w \downarrow x$ and $z \downarrow y$, we have $D^+ f(x) \le D^+ f(y)$ for all $x < y$. To prove right continuity, we assume $\alpha = \lim_{y \downarrow x} D^+ f(y)$. Then $D^+ f(x) \le \alpha$, and $D^+ f(y) \ge \alpha$ for all $y > x$. Hence it suffices to show that $D^+ f(x) \ge \alpha$. By convexity of f, we take $0 < h < z - y$ to obtain

$$\frac{f(z)-f(y)}{z-y} \ge \lim_{h \downarrow 0} \frac{f(y+h)-f(y)}{h} = D^+ f(y) \ge \alpha.$$

We then let $y \downarrow x$ to obtain

$$\frac{f(z)-f(x)}{z-x} \ge \alpha, \quad \text{for all } z > x.$$

Finally, we let $z \downarrow x$ to conclude that $D^+ f(x) \ge \alpha$. Thus we complete the proof. $\square$

7.1 Tanaka's Formula and Local Times

Motivation. Let $X = (X_t)_{t \ge 0}$ be a continuous semimartingale. If $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function, the Itô's formula asserts that $(f(X_t))_{t \ge 0}$ is still a semimartingale, and

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X, X\rangle_s. \tag{7.1}$$

In fact, if $f : \mathbb{R} \to \mathbb{R}$ is not twice continuously differentiable but convex, then $(f(X_t))_{t \ge 0}$ is still a semimartingale, and we can obtain a representation of $f(X_t)$ similar to (7.1).

Theorem 7.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, and let $(X_t)_{t \ge 0}$ be a semimartingale. Then $(f(X_t))_{t \ge 0}$ is also a semimartingale. Furthermore, there exists a continuous increasing process $(A_t^f)_{t \ge 0}$ such that*

$$f(X_t) = f(X_0) + \int_0^t D^- f(X_s)\,\mathrm{d}X_s + A_t^f. \tag{7.2}$$

Proof. **Step I.** We let ϕ be a nonnegative C^∞ function supported on $[0, 1]$ such that $\int_0^1 \phi(x)\,dx = 1$, and let

$$f_n(x) = n(\phi(n\cdot) * f)(x) = n\int_{-\infty}^{\infty} \phi(ny) f(x-y)\,\mathrm{d}y.$$

Then f_n is C^∞ on $\mathbb{R}$, and by standard mollification results, $f_n \to f$ pointwise, and $f_n' = n\phi(n\cdot) * D^- f$. By left continuity of $D^- f$, for every $x \in \mathbb{R}$,

$$f_n'(x) = n\int_0^{1/n} \phi(ny) D^- f(x-y)\,\mathrm{d}y \to D^- f(x).$$

Finally, since f_n is also convex, we have $f_n'' \ge 0$ on $\mathbb{R}$ for all $n \in \mathbb{N}$.

Step II. Let $X = M + V$ be the canonical decomposition of semimartingale X. For each $N > 0$, define

$$\tau_N = \left\{ t \ge 0 : |X_t| + \int_0^t |\mathrm{d}V_s| + \langle M, M\rangle_t \ge N \right\}.$$

By Itô's formula,

$$f_n(X_{t\wedge\tau_N}) = f_n(X_0) + \int_0^{t\wedge\tau_N} f_n'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^{t\wedge\tau_N} f_n''(X_s)\,\mathrm{d}\langle X, X\rangle_s. \tag{7.3}$$

By definition of τ_N, we know $\langle M, M\rangle_{\tau_N} \le N$. Since f_n' is uniformly bounded on every compact interval $[0,t]$, by the Dominated Convergence Theorem 5.7 for stochastic integrals, as $n \to \infty$,

$$\int_0^{t\wedge\tau_N} f_n'(X_s)\,\mathrm{d}X_s \to \int_0^{t\wedge\tau_N} D^- f(X_s)\,\mathrm{d}X_s \quad \text{in probability.} \tag{7.4}$$

For every $t \ge 0$, define

$$A_t^{f,N} = f(X_{t\wedge\tau_N}) - f(X_0) - \int_0^{t\wedge\tau_N} D^- f(X_s)\,\mathrm{d}X_s.$$

By definition, $t \mapsto A_t^{f,N}$ is continuous. Since $f_n \to f$ and $f_n' \to D^- f$ pointwise, by (7.3) and (7.4),

$$A_t^{f,N} = \lim_{n\to\infty} \frac{1}{2}\int_0^{t\wedge\tau_N} f_n''(X_s)\, d\langle X, X\rangle_s \quad \text{in probability.}$$

Since $f_n'' \ge 0$, the process $A^{f,N}$ is increasing. Furthermore, for $N' > N$, we have $A_{t\wedge\tau_N}^{f,N'} = A_t^{f,N}$. Therefore there exists a continuous increasing process $A^f = (A_t^f)_{t\ge 0}$ such that $A_{t\wedge\tau_N}^f = A_t^{f,N}$ for every $N > 0$. Furthermore, we can obtain (7.2) by letting $N \uparrow \infty$ in the definition of $A_t^{f,N}$. □

Remark. We can adapt our proof to show that

$$f(X_t) = f(X_0) + \int_0^t D^+ f(X_s)\,\mathrm{d}X_s + \widetilde{A}_t^f$$

for some continuous increasing process $(\widetilde{A}_t^f)_{t\ge 0}$. If f is twice continuously differentiable, we have $D^+ f = D^- f$ and $\widetilde{A}_t^f = A_t^f = \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X, X\rangle_s$. However, in general we may have $\widetilde{A}^f \ne A^f$.

Theorem 7.3 (Tanaka)**.** *Let $(X_t)_{t\ge 0}$ be a semimartingale and $a \in \mathbb{R}$. Then there exists an increasing process $(L_t^a(X))t \ge 0$ that satisfies the following* ***Tanaka's formula****:*

$$|X_t - a| = |X_0 - a| + \int_0^t \operatorname{sgn}(X_s - a)\,\mathrm{d}X_s + L_t^a(X), \tag{7.5}$$

$$(X_t - a)^+ = (X_0 - a)^+ + \int_0^t \mathbb{1}_{\{X_s > a\}}\,\mathrm{d}X_s + \frac{1}{2}L_t^a(X), \tag{7.6}$$

$$(X_t - a)^- = (X_0 - a)^- - \int_0^t \mathbb{1}_{\{X_s \le a\}}\,\mathrm{d}X_s + \frac{1}{2}L_t^a(X). \tag{7.7}$$

The increasing process $(L_t^a(X))_{t\ge 0}$ is called the ***local time of X at level*** *a. Furthermre, for every stopping time τ, we have $(L_{t\wedge\tau}^a(X))_{t\ge 0} = (L_t^a(X^\tau))_{t\ge 0}$.*

Proof. We apply Theorem 7.2 on the convex function $f(x) = |x - a|$. Then the process $(L_t^a(X))_{t\ge 0}$ defined by

$$L_t^a(X) = |X_t - a| - |X_0 - a| - \int_0^t \operatorname{sgn}(X_s - a)\,\mathrm{d}X_s \tag{7.8}$$

is an increasing process, and (7.5) follows from definition. Also $L_{t\wedge\tau}^a(X) = L_t^a(X^\tau)$ for all stopping times τ, since $\int_0^t \operatorname{sgn}(X_s^\tau - a)\,\mathrm{d}X_s^\tau = \int_0^{t\wedge\tau} \operatorname{sgn}(X_s - a)\,\mathrm{d}X_s$. To show (7.6) and (7.7), we apply Theorem 7.2 to convex functions $(x-a)^+$ and $(x-a)^-$ to obtain

$$(X_t - a)^+ = (X_0 - a)^+ + \int_0^t \mathbb{1}_{\{X_s > a\}}\,\mathrm{d}X_s + A_t^{a+}, \quad (X_t - a)^- = (X_0 - a)^- - \int_0^t \mathbb{1}_{\{X_s \le a\}}\,\mathrm{d}X_s + A_t^{a-},$$

where $(A_t^{a+})_{t\ge 0}$ and $(A_t^{a-})_{t\ge 0}$ are two increasing processes. Taking the difference of the above two identities, we see that $A_t^{a+} = A_t^{a-}$. By comparing the sum of the above two identities with (7.8), $A_t^{a+} + A_t^{a-} = L_t^a(X)$. Hence $A_t^{a+} = A_t^{a-} = \frac{1}{2}L_t^a(X)$, and we prove (7.6) and (7.7). □

Next, we study the variation property of local times. We write by $\mathrm{d}L_s^a(X)$ the random measure associated with the function $s \mapsto L_s^a(X)$.

Proposition 7.4. *Let X be a continuous semimartingale and let $a \in \mathbb{R}$. Then almost surely, the random measure $\mathrm{d}L_s^a(X)$ is supported on $\{s \geq 0 : X_s = a\}$.*

Proof. We set $Y_t = |X_t - a|$. By (7.5), since $\operatorname{sgn}(x) = 1$ for all $x \in \mathbb{R}$, we have $\langle Y, Y\rangle_t = \langle X, X\rangle_t$, and

$$\mathrm{d}Y_s = \operatorname{sgn}(X_s - a)\mathrm{d}X_s + \mathrm{d}L_s^a(X).$$

By applying Itô's formula to Y_t^2, it holds

$$\begin{aligned}(X_t - a)^2 &= Y_0^2 + 2\int_0^t Y_s\,\mathrm{d}Y_s + \langle Y, Y\rangle_t \\ &= (X_0 - a)^2 + 2\int_0^t (X_s - a)\,\mathrm{d}X_s + 2\int_0^t |X_s - a|\,\mathrm{d}L_s^a(X) + \langle X, X\rangle_t.\end{aligned}$$

In the other hand, if we directly apply Itô's formula to $(X_t - a)^2$, we have

$$(X_t - a)^2 = (X_0 - a)^2 + 2\int_0^t (X_s - a)\,\mathrm{d}X_s + \langle X, X\rangle_t.$$

Comparing the two results, we have

$$\int_0^t |X_s - a|\,\mathrm{d}L_s^a(X) = 0.$$

Then we finish the proof. □

Remark. This proposition shows that $(L_t^a(X))_{t\geq 0}$ may only increase when $X_t = a$. To some degree, $(L_t^a(X))_{t\geq 0}$ measures how long the process stay at level a before time t, which justifies the name "local time".

7.2 Continuity of Local Times and Generalized Itô's Formula

In this subsection we discuss the continuity of local times $L^a(X)$ with respect to the space variable a. It is often helpful to view $L^a(X) = (L_t^a(X))_{t\geq 0}$ as a random function in $C(\mathbb{R}_+, \mathbb{R}_+)$, which is equipped with the compact convergence topology. Throughout this section, we let $X = M + V$ be a continuous semimartingale with its canonical decomposition.

7.2.1 Continuity of Local Times

Theorem 7.5 (Càdlàg)**.** *The process $(L^a(X))_{a\in\mathbb{R}}$ with values in $C(\mathbb{R}_+, \mathbb{R}_+)$ has a càdlàg modification, which we consider from now on and for which we keep the same notation $(L^a(X))_{a\in\mathbb{R}}$. For each $a \in \mathbb{R}$, we denote by $L^{a-}(X) = \lim_{b\uparrow a} L^b(X)$ the left limit of $b \mapsto L^b(X)$ at level a. Then*

$$L_t^a(X) - L_t^{a-}(X) = 2\int_0^t \mathbb{1}_{\{X_s = a\}}\,\mathrm{d}V_s = 2\int_0^t \mathbb{1}_{\{X_s = a\}}\,\mathrm{d}X_s, \quad t \geq 0. \tag{7.9}$$

In particular, if X is a continuous local martingale, the process $(L_t^a(X))_{a\in\mathbb{R}, t\geq 0}$ is jointly continuous.

The proof of this theorem uses Tanaka's formula and the following technical lemma.

Lemma 7.6. *Let $p \geq 1$. There exists a constant $C_p > 0$ such that for every $-\infty < a < b < \infty$,*

$$\mathbb{E}\left[\left(\int_0^t \mathbb{1}_{\{a < X_s \leq b\}}\,\mathrm{d}\langle M, M\rangle_s\right)^p\right] \leq C_p (b-a)^p \left(\mathbb{E}\left[\langle M, M\rangle_t^{p/2}\right] + \mathbb{E}\left[\left(\int_0^t |\mathrm{d}V_s|\right)^p\right]\right). \tag{7.10}$$

Furthermore, for each $a \in \mathbb{R}$, write $Y^a = (Y^a_t)_{t\geq 0}$ for the random function in $C(\mathbb{R}_+, \mathbb{R})$ defined by

$$Y^a_t = \int_0^t \mathbb{1}_{\{X_s > a\}} \,\mathrm{d}M_s.$$

Then the process $(Y^a)_{a\in\mathbb{R}}$ has a continuous modification.

Proof. (i) We first prove the inequality (7.10). We may assume $b = -a = r$ for some $r > 0$, otherwise we may assume $r = \frac{b-a}{2}$ and replace X with $X - \frac{b+a}{2}$. Let f be the unique function in $C^2(\mathbb{R})$ such that

$$f(0) = f'(0) = 0, \quad \text{and} \quad f''(x) = \left(2 - \frac{|x|}{r}\right)^+.$$

Then $|f'| \leq 2r$ on $\mathbb{R}$. Since $f'' \geq 0$ on $\mathbb{R}$ and $f''(x) \geq 1$ for $x \in [-r, r]$, and by Itô's formula, we have

$$\frac{1}{2}\int_0^t \mathbb{1}_{\{-r<X_s\leq r\}} \,\mathrm{d}\langle M, M\rangle_s \leq \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle M, M\rangle_s = f(X_t) - f(X_0) - \int_0^t f'(X_s)\,\mathrm{d}X_s. \tag{7.11}$$

Recalling that $|f'| \leq 2r$. By the Burkholder-Davis-Gundy inequality [Theorem 5.16],

$$\begin{aligned}\mathbb{E}\,|f(X_t) - f(X_0)|^p &\leq (2r)^p\mathbb{E}\,|X_t - X_0|^p \leq (2r)^p\mathbb{E}\left[\left(|M_t - M_0| + \int_0^t |\mathrm{d}V_s|\right)^p\right] \\ &\leq C_p(2r)^p\left(\mathbb{E}\left[\langle M, M\rangle_t^{p/2}\right] + \mathbb{E}\left[\left(\int_0^t |\mathrm{d}V_s|\right)^p\right]\right),\end{aligned} \tag{7.12}$$

and we henceforce use C_p to denote any constant depending only on p, which may vary from line to line. Next we control $\int_0^t f'(X_s)\,\mathrm{d}X_s$, which can decomposed as

$$\int_0^t f'(X_s)\,\mathrm{d}X_s = \int_0^t f'(X_s)\,\mathrm{d}M_s + \int_0^t f'(X_s)\,\mathrm{d}V_s.$$

Note that

$$\mathbb{E}\left[\left|\int_0^t f'(X_s)\,\mathrm{d}V_s\right|^p\right] \leq (2r)^p\int_0^t \mathbb{E}\left[\left(\int_0^t |\mathrm{d}V_s|\right)^p\right],$$

and again by the Burkholder-Davis-Gundy inequality,

$$\mathbb{E}\left[\left|\int_0^t f'(X_s)\,\mathrm{d}M_s\right|^p\right] \leq C_p\mathbb{E}\left[\left(\int_0^t f'(X_s)\,\mathrm{d}\langle M, M\rangle_t\right)^{p/2}\right] \leq C_p(2r)^p\mathbb{E}\left[\langle M, M\rangle_t^{p/2}\right]$$

Combining the last three displays and the estimates (7.11), (7.12), we obtain the inequality (7.10).

(ii) Fix $p > 2$. By the Burkholder-Davis-Gundy inequality, for every $-\infty < a < b < \infty$ and $t \geq 0$,

$$\mathbb{E}\left[\sup_{s\in[0,t]} |Y^b_s - Y^a_s|^p\right] \leq C_p\mathbb{E}\left[\left(\int_0^t \mathbb{1}_{\{a<X_s\leq b\}}\,\mathrm{d}\langle M, M\rangle_s\right)^{p/2}\right]. \tag{7.13}$$

We define stopping time

$$\tau_n = \inf\left\{t \geq 0 : \langle M, M\rangle_t + \int_0^t |\mathrm{d}V_s| \geq n\right\}.$$

By (7.10),

$$\mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} \mathbb{1}_{\{a<X_s\leq b\}}\,\mathrm{d}\langle M, M\rangle_s\right)^{p/2}\right] \leq C_p(b-a)^{p/2}\left(n^{p/4} + n^{p/2}\right).$$

We then use (7.13) with t replaced by $t \wedge \tau_n$, and letting $t \to \infty$, to obtain

$$\mathbb{E}\left[\sup_{s\geq 0} |Y^b_{s\wedge\tau_n} - Y^a_{s\wedge\tau_n}|^p\right] \leq C_p(b-a)^{p/2}\left(n^{p/4} + n^{p/2}\right). \tag{7.14}$$

Since $p > 2$, by Kolmogorov's continuity lemma [Corollary 4.8], there exists an a.s. modification of the process $a \mapsto (Y^a_{s\wedge\tau_n})_{s\geq 0}$ valued in $C(\mathbb{R}_+, \mathbb{R})$. We write $(Y^{(n),a})_{a\in\mathbb{R}}$ for this continuous modification.

If $1 \leq n \leq m$, for every fixed $a \in \mathbb{R}$, we have $(Y^{(n),a}_s)_{s\geq 0} = (Y^{(m),a}_{s\wedge\tau_n})_{s\geq 0}$ a.s.. By a continuity argument (choose a dense subset of $\mathbb{R}$ and use continuity on the space variable a), the equality holds for all $s \geq 0$ and all $a \in \mathbb{R}$ a.s.. Therefore, we can find a continuous process $(\widetilde{Y}^a)_{a\in\mathbb{R}}$ valued in $C(\mathbb{R}_+, \mathbb{R})$ such that for each $n \geq 1$, $(Y^{(n),a}_s)_{s\geq 0} = (\widetilde{Y}^a_{s\wedge\tau_n})_{s\geq 0}$ for all $a \in \mathbb{R}$ a.s., which is the desired continuous modification. □

Remark. We apply the inequality (7.10) on the semimartingale X^{τ_n}, where τ_n has the same definition as in the above proof. We then let $a \uparrow b$ and apply dominated convergence theorem to obtain

$$\mathbb{E}\left|\int_0^{t\wedge\tau_n} \mathbb{1}_{\{X_s=b\}}\, \mathrm{d}\langle M, M\rangle_s\right| = 0.$$

Finally we let $\tau_n \to \infty$ to obtain that $\int_0^t \mathbb{1}_{\{X_s=b\}}\, \mathrm{d}\langle M, M\rangle_s = 0$ a.s. for every $t \geq 0$. Hence

$$\int_0^t \mathbb{1}_{\{X_s=b\}}\, \mathrm{d}M_s = 0 \;\; a.s., \quad b \in \mathbb{R},\; t \geq 0. \tag{7.15}$$

Proof of Theorem 7.5. We slightly abuse the notation and write $(Y^a)_{a\in\mathbb{R}}$ for the continuous modification obtained in the second assertion of Lemma 7.6. We also define

$$Z^a_t = \int_0^t \mathbb{1}_{\{X_s>a\}}\, \mathrm{d}V_s, \quad t \geq 0.$$

For each $a_0 \in \mathbb{R}$, by the dominated convergence theorem, for all $T > 0$,

$$\lim_{a\downarrow a_0} \int_0^T |\mathbb{1}_{\{X_s>a\}} - \mathbb{1}_{\{X_s>a_0\}}|\, \mathrm{d}V_s = 0, \qquad \lim_{b\uparrow a_0} \int_0^T |\mathbb{1}_{\{X_s>b\}} - \mathbb{1}_{\{X_s\geq a_0\}}|\, \mathrm{d}V_s = 0,$$

It is seen that $Z^a_t \to Z^{a_0}_t$ as $a \downarrow a_0$ and $Z^b_t \to \int_0^t \mathbb{1}_{\{X_s\geq a_0\}}\, \mathrm{d}V_s$ as $b \uparrow a_0$, and both convergences are uniformly on each compact interval $[0, T]$. Hence the process $a \mapsto Z^a$ has càdlàg sample paths. Since $a \mapsto Y_a$ is continuous, By Tanaka's formula [Theorem 7.3], for each $a \in \mathbb{R}$, we have

$$(L^a_t)_{t\geq 0} = 2\left((X_t - a)^+ - (X_0 - a)^+ - Y^a_t - Z^a_t\right)_{t\geq 0} \quad a.s.. \tag{7.16}$$

which provides the desired càdlàg modification, because

$$a \mapsto (X_t - a)^+ - (X_0 - a)^+ - Y^a_t$$

has continuous sample paths. Furthermore, one can evaluate the jump by

$$L^a_t - L^{a-}_t = Z^{a-}_t - Z^a_t = \int_0^t \mathbb{1}_{\{X_s=a\}}\, \mathrm{d}V_s.$$

By (7.15), we finish the proof of the second identity of (7.9). □

Remark. Our càdlàg modification $(\widetilde{L}^a(X))_{a\in\mathbb{R}}$ is done for the spatial process $(L^a(X))_{a\in\mathbb{R}}$ taking values in $C(\mathbb{R}_+, \mathbb{R})$. Hence for each fixed $a \in \mathbb{R}$, the processes $(L^a_t(X))_{t\geq 0}$ and $(\widetilde{L}^a_t(X))_{t\geq 0}$ are indistinguishable.

7.2.2 Itô-Tanaka Formula

Now we study an extension of Itô's formula using the càdlàg property of local times. If f is a convex function on $\mathbb{R}$, the left derivative D^-f is a left-continuous increasing function on $\mathbb{R}$, and there exists a unique Radon measure $D^2f(\mathrm{d}x)$ on $\mathbb{R}_+$ such that $D^2f([a,b)) = D^-f(b) - D^-f(a)$ for every $a < b$, which can be interpret as the second derivative of f in the sense of distributions. For all $-\infty < a < b < \infty$,

$$D^-f(b) = D^-f(a) + \int_{\mathbb{R}} \mathbb{1}_{\{a\leq x<b\}} D^2f(\mathrm{d}x).$$

By the fundamental theorem of calculus for absolute continuous functions and Fubini's theorem,

$$\begin{aligned} f(b) &= f(a) + \int_a^b D^-f(y)\,\mathrm{d}y = f(a) + \int_a^b \left(D^-f(a) + \int_{\mathbb{R}} \mathbb{1}_{\{a\leq x<y\}} D^2f(\mathrm{d}x)\right)\mathrm{d}y \\ &= f(a) + (b-a)D^-f(a) + \int_{\mathbb{R}}\int_a^b \mathbb{1}_{\{a\leq x<y\}}\,\mathrm{d}y\, D^2f(\mathrm{d}x). \end{aligned}$$

Hence

$$f(b) = f(a) + (b-a)D^-f(a) + \int_{[a,\infty)} (b-x)^+\, D^2f(\mathrm{d}x), \quad -\infty < a < b < \infty. \tag{7.17}$$

We can then identify the increasing process $(A_t^f)_{t\geq 0}$ in Theorem 7.2 using the local times $(L_t^a(X))_{a\in\mathbb{R}}$ and the distributional derivative D^2f. This is a generalization of the Itô's formula.

Theorem 7.7 (Itô-Tanaka)**.** *Let $f:\mathbb{R}\to\mathbb{R}$ be a convex function. Then for every $t\geq 0$,*

$$f(X_t) = f(X_0) + \int_0^t D^-f(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a(X)D^2f(\mathrm{d}a). \tag{7.18}$$

Proof. We first assume that $(X_t)_{t\geq 0}$ is bounded, so there exists $K > 0$ such that $|X_t| \leq K$ for all $t \geq 0$, and $L^a(X) \equiv 0$ for $|a| > K$ by Proposition 7.4. By Tanaka's formula, for every $a \in \mathbb{R}$,

$$(X_t - a)^+ = (X_0 - a)^+ + Y_t^a + Z_t^a + \frac{1}{2}L_t^a(X), \quad t \geq 0. \tag{7.19}$$

where

$$Y_t^a = \int_0^t \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s, \quad \text{and} \quad Z_t^a = \int_0^t \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}V_s, \quad t\geq 0.$$

By Fubini's Theorem,

$$\begin{aligned} \int_{[-K,K]} Z_t^a D^2f(\mathrm{d}a) &= \int_0^t \int_{[-K,K]} \mathbb{1}_{\{X_s>a\}} D^2f(\mathrm{d}a)\,\mathrm{d}V_s \\ &= \int_0^t \int_{\mathbb{R}} \mathbb{1}_{\{-K\leq a<X_s\}} D^2f(\mathrm{d}a)\,\mathrm{d}V_s = \int_0^t \left(D^-f(X_s) - D^-f(-K)\right)\mathrm{d}V_s. \end{aligned} \tag{7.20}$$

Next, we introduce the stopping times $\tau_n = \{t\geq 0 : \langle M,M\rangle_t \geq n\}$, and consider the continuous modification of $a \mapsto \int_0^{t\wedge\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s$ provided by Lemma 7.6. Define

$$M_t^{f,n} = \int_{[-K,K]} \left(\int_0^{t\wedge\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s\right) D^2f(\mathrm{d}a), \quad t\geq 0.$$

Then $\int_0^t \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s$ is a local martingale reduced by τ_n, and by Fubini's theorem,

$$\mathbb{E}[M_{t+\epsilon}^{f,n}|\mathscr{F}_t] = M_t^{f,n} + \int_{[-K,K]} \mathbb{E}\left[\left(\int_{t\wedge\tau_n}^{(t+\epsilon)\wedge\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s\right)\,\middle|\,\mathscr{F}_t\right] D^2f(\mathrm{d}a) = M_t^{f,n}, \quad t+\epsilon > t \geq 0.$$

Hence $(M_t^{f,n})_{t\geq 0}$ defines a continuous martingale in $\mathbb{H}^2$. For any $N \in \mathbb{H}^2$,

$$
\begin{aligned}
\mathbb{E}\left[\langle M^{f,n}, N\rangle_\infty\right] &= \mathbb{E}\left[M_\infty^{f,n} N_\infty\right] = \mathbb{E}\left[\int_{[-K,K]} \left(\int_0^{t\wedge\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s\right) N_\infty D^2 f(\mathrm{d}a)\right] \\
&= \mathbb{E}\left[\int_{[-K,K]} \left(\int_0^{\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}\langle M, N\rangle_s\right) D^2 f(\mathrm{d}a)\right] = \mathbb{E}\left[\int_0^{\tau_n} \left(\int_{[-K,K]} \mathbb{1}_{\{X_s>a\}}\, D^2 f(\mathrm{d}a)\right) \mathrm{d}\langle M, N\rangle_s\right] \\
&= \mathbb{E}\left[\left(\int_0^{\tau_n} \left(\int_{[-K,K]} \mathbb{1}_{\{X_s>a\}}\, D^2 f(\mathrm{d}a)\right) \mathrm{d}M_s\right) N_\infty\right].
\end{aligned}
$$

Since $(M,N) \mapsto \mathbb{E}[\langle M,N\rangle_\infty] = \mathbb{E}[M_\infty N_\infty]$ is an inner product on $\mathbb{H}^2$, we have the following Fubini's theorem for stochastic integrals:

$$
M_t^{f,n} = \int_{[-K,K]} \left(\int_0^{t\wedge\tau_n} \mathbb{1}_{\{X_s>a\}}\,\mathrm{d}M_s\right) D^2 f(\mathrm{d}a) = \int_0^{t\wedge\tau_n} \left(\int_{[-K,K]} \mathbb{1}_{\{X_s>a\}}\, D^2 f(\mathrm{d}a)\right) \mathrm{d}M_s.
$$

Letting $n \to \infty$, and apply the monotone convergence theorem, we have

$$
\int_{[-K,K]} Y_t^a\, D^2 f(\mathrm{d}a) = \int_0^t \left(D^- f(X_s) - D^- f(-K)\right) \mathrm{d}M_s, \quad t \geq 0. \tag{7.21}
$$

We then integrate (7.19) with respect to $D^2 f(\mathrm{d}a)$ on $[-K,K]$, and apply (7.17), (7.20) and (7.21) to obtain

$$
f(X_t) + (X_t - X_0) D^- f(-K) = f(X_0) + \int_0^t \left(D^- f(X_s) - D^- f(-K)\right) \mathrm{d}X_s + \frac{1}{2}\int_{[-K,K]} L_t^a(X) D^2 f(\mathrm{d}a).
$$

Note that $\int_0^t \mathrm{d}X_s = X_t - X_0$, and $L_t^a(X) = 0$ for $|a| > K$, we obtain (7.18) for bounded semimartingales. For the case when X is unbounded, we stop X when it first leaves $[-K,K]$ at the stopping time T_K. Then

$$
f(X_{t\wedge T_K}) = f(X_0) + \int_0^{t\wedge T_K} D^- f(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_{\mathbb{R}} L_{t\wedge T_K}^a(X) D^2 f(\mathrm{d}a).
$$

By continuity of f and the monotone convergence theorem, we let $K \uparrow \infty$ to conclude (7.18). $\square$

Remark. The Itô-Tanaka formula (7.18) also holds for each f that is a difference of two convex functions.

Corollary 7.8 (Occupation times formula)**.** *Almost surely, for all $t \geq 0$ and all nonnegative Borel functions Φ on $\mathbb{R}$,*

$$
\int_0^t \Phi(X_s)\,\mathrm{d}\langle X, X\rangle_s = \int_{\mathbb{R}} \Phi(a) L_t^a(X)\,\mathrm{d}a. \tag{7.22}
$$

More generally, we have a.s. for all $t \geq 0$ and all nonnegative Borel functions Φ on $\mathbb{R}_+ \times \mathbb{R}$ that

$$
\int_0^t F(s, X_s)\,\mathrm{d}\langle X, X\rangle_s = \int_{\mathbb{R}} \int_0^\infty F(s,a)\,\mathrm{d}L_s^a(X)\,\mathrm{d}a. \tag{7.23}
$$

Proof. We first fix $\Phi \in C_c(\mathbb{R})$ with $\Phi \geq 0$ and $f \in C^2(\mathbb{R})$ with $f'' = \Phi$. Then (7.22) holds for each $t \geq 0$ (by a continuity argument) outside a zero probability set $\mathcal{N}_\Phi$ by comparing Itô's formula and the Itô-Tanaka formula. Next, we take a countable dense subset $\{\Phi_n\}_{n=1}^\infty$ of $C_c(\mathbb{R})$ and take $\mathcal{N} = \bigcup_{n=1}^\infty N_{\Phi_n}$. Then the formula (7.22) holds for all $\Phi \in C_c(\mathbb{R})$ outside the zero-probability set $\mathcal{N}$. An application of the Monotone Class Theorem 1.38 gives the general result for nonnegative measurable functions Φ. Consequently, (7.23) holds for all functions F of the type $F(s,a) = \mathbb{1}_{[\alpha,\beta]}(s)\mathbb{1}_A(a)$, where $0 \leq \alpha \leq \beta$ and $A \in \mathscr{B}(\mathbb{R})$. Again an application of Theorem 1.38 gives the general result. $\square$

7.2.3 Approximation of Local Times

We have the following proposition which gives another reason for the terminology "local time".

Proposition 7.9. *Let X be a continuous semimartingale. Then almost surely for all $t \geq 0$ and $a \in \mathbb{R}$,*

$$L_t^a(X) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^t \mathbb{1}_{[a,a+\epsilon)}(X_s)\, \mathrm{d}\langle X, X\rangle_s. \tag{7.24}$$

Proof. By the occupation times formula (7.22), almost surely for all $t \geq 0$ and $a \in \mathbb{R}$,

$$\int_0^t \mathbb{1}_{[a,a+\epsilon)}(X_s)\, \mathrm{d}\langle X, X\rangle_s = \int_a^{a+\epsilon} L_t^x(X)\, \mathrm{d}x.$$

Since $a \mapsto L^a(X)$ is right-continuous on $\mathbb{R}$, the result follows. $\square$

Remark. An analogue of the above argument also gives

$$L_t^{a-}(X) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^t \mathbb{1}_{(a-\epsilon,a]}(X_s)\, \mathrm{d}\langle X, X\rangle_s.$$

In particular, if X is a continuous local martingale, we have

$$L_t^a(X) = \lim_{\epsilon \downarrow 0} \frac{1}{2\epsilon} \int_0^t \mathbb{1}_{(a-\epsilon,a+\epsilon)}(X_s)\, \mathrm{d}\langle X, X\rangle_s.$$

We also have the following estimate for the moments of local times.

Corollary 7.10. *Let $p > 1$. There exists a constant C_p such that, for any continuous semimartingale X with canonical decomposition $X = M + V$, we have for every $a \in \mathbb{R}$ and $t \geq 0$ that*

$$\mathbb{E}|L_t^a(X)|^p \leq C_p \left(\mathbb{E}\left[\langle M, M\rangle_t^{p/2} \right] + \mathbb{E}\left[\left(\int_0^t |\mathrm{d}V_s| \right)^p \right] \right).$$

Proof. This estimate follow from (7.10) in Lemma 7.6, the approximation (7.24) and Fatou's lemma. $\square$

Next, we introduce the downcrossing approximation of local time. We let X be a continuous semimartingale, and introduce two sequences of stopping times

$$\sigma_0^\epsilon = 0, \quad \tau_n^\epsilon = \inf\{t \geq \sigma_n : X_t = \epsilon\}, \quad \text{and} \quad \sigma_n^\epsilon = \inf\{t \geq \tau_{n-1}^\epsilon : X_t = 0\}.$$

Define the *downcrossing number of X* from level 0 to level ϵ before time t by

$$N_{[0,\epsilon]}^X(t) = \inf\{n \in \mathbb{N}_0 : \sigma_n \leq t\}.$$

We have introduced this notation with slight difference in the proof of martingale convergence.

Theorem 7.11 (Downcrossing representation of the local time at zero)**.** *For each $t \geq 0$, we have*

$$\lim_{\epsilon \downarrow 0} \epsilon N_{[0,\epsilon]}^X(t) = \frac{1}{2} L_t^0(X) \quad \textit{in probability.} \tag{7.25}$$

Furthermore, if there exists $p \geq 1$ such that

$$\mathbb{E}\left[\langle M, M\rangle_\infty^{p/2} + \left(\int_0^\infty |\mathrm{d}V_s| \right)^p \right] < \infty,$$

then

$$\lim_{\epsilon\downarrow 0}\mathbb{E}\left[\sup_{t\geq 0}\left|\epsilon N^X_{[0,\epsilon]}(t)-\frac{1}{2}L^0_t(X)\right|^p\right]=0. \tag{7.26}$$

Proof. For notation simplicity, we write L_s for $L^0_s(X)$ in our proof. By Tanaka's formula,

$$(X_{t\wedge\tau^\epsilon_n})^+-(X_{t\wedge\sigma^\epsilon_n})^+=\int_{t\wedge\sigma^\epsilon_n}^{t\wedge\tau^\epsilon_n}\mathbb{1}_{\{X_s>0\}}\,\mathrm{d}X_s+\frac{1}{2}\left(L_{t\wedge\tau^\epsilon_n}-L_{t\wedge\sigma^\epsilon_n}\right) \tag{7.27}$$

Note that $L^0(X)$ does not increase on the intervals of the type $[\tau^\epsilon_{n-1},\sigma^\epsilon_n)$, $n=1,2,\cdots$. By (7.27) we have

$$\begin{aligned}\frac{1}{2}L_t&=\frac{1}{2}\sum_{n=1}^\infty\left(L_{t\wedge\sigma^\epsilon_{n+1}}-L_{t\wedge\sigma^\epsilon_n}\right)=\frac{1}{2}\sum_{n=1}^\infty\left(L_{t\wedge\tau^\epsilon_n}-L_{t\wedge\sigma^\epsilon_n}\right)\\&=\sum_{n=1}^\infty\left((X_{t\wedge\tau^\epsilon_n})^+-(X_{t\wedge\sigma^\epsilon_n})^+-\int_{t\wedge\sigma^\epsilon_n}^{t\wedge\tau^\epsilon_n}\mathbb{1}_{\{X_s>0\}}\,\mathrm{d}X_s\right)\\&=\sum_{n=1}^\infty\left((X_{t\wedge\tau^\epsilon_n})^+-(X_{t\wedge\sigma^\epsilon_n})^+\right)-\int_0^t\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}X_s.\end{aligned} \tag{7.28}$$

Noting that $(X_{t\wedge\tau^\epsilon_n})^+-(X_{t\wedge\sigma^\epsilon_n})^+=\epsilon\mathbb{1}_{\{\tau^\epsilon_n\leq t\}}$, we have

$$\sum_{n=1}^\infty\left((X_{t\wedge\tau^\epsilon_n})^+-(X_{t\wedge\sigma^\epsilon_n})^+\right)=\epsilon N^n_\epsilon(t)+u(\epsilon),$$

where $0\leq u(\epsilon)=(X_t)^+-(X_{\sigma^\epsilon_{n(t)}})^+\leq\epsilon$, with $n(t)=N^X_{[0,\epsilon]}(t)$. Recalling (7.28), we have

$$\frac{1}{2}L_t-\epsilon N^X_{[0,\epsilon]}(t)=u(\epsilon)-\int_0^t\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}X_s. \tag{7.29}$$

(i) Since $0\leq\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\leq\mathbb{1}_{(0,\epsilon]}(X_s)$ for all $\epsilon>0$, by Theorem 5.7, we have

$$\lim_{\epsilon\downarrow 0}\int_0^t\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}X_s=0\quad\text{in probability.}$$

Recalling (7.29), we obtain (7.25).

(ii) By the Burkholder-Davis-Gundy inequality [Theorem 5.16], for every $\epsilon>0$ and $T>0$,

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}X_s\right|^p\right]\\&\leq C_p\mathbb{E}\left[\left(\int_0^T\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}\langle M,M\rangle_s\right)^{p/2}+\left(\int_0^T\left|\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\right|\,|\mathrm{d}V_s|\right)^p\right]\\&\leq C_p\mathbb{E}\left[\langle M,M\rangle_T^{p/2}+\left(\int_0^T|\mathrm{d}V_s|\right)^p\right].\end{aligned}$$

By (7.11) and the dominated convergence theorem, we have

$$\lim_{\epsilon\downarrow 0}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t\sum_{n=1}^\infty\mathbb{1}_{(\sigma^\epsilon_n,\tau^\epsilon_n]}(s)\mathbb{1}_{(0,\epsilon]}(X_s)\,\mathrm{d}X_s\right|^p\right]=0.$$

Letting $T\uparrow\infty$ and by (7.29), we conclude (7.26). □

7.3 Brownian Local Times

In this subsection, we study the local times of the standard Brownian motion $B = (B_t)_{t\geq 0}$ on $\mathbb{R}$. We fix $(\mathscr{F}_t)_{t\geq 0}$ to be the completed canonical filtration of B.

Theorem 7.12 (Trotter)**.** *There exists a unique process $(L_t^a(B))_{t\geq 0, a\in\mathbb{R}}$, called the* ***local time*** *of the Brownian motion $B = (B_t)_{t\geq 0}$ such that*

(i) the map $(t, a) \mapsto L_t^a(B)$ is continuous, and the map $a \mapsto L_t^a(B)$ is Hölder continuous with exponent γ for each $\gamma < \frac{1}{2}$ and uniformly in t on every compact interval;

(ii) for every fixed $a \in \mathbb{R}$, the map $t \mapsto L_t^a(B)$ is increasing;

(iii) a.s. for every $t \geq 0$ and every nonnegative measurable function on $\mathbb{R}$,

$$\int_0^t \Phi(B_s)\,\mathrm{d}s = \int_{\mathbb{R}} \Phi(a) L_t^a(B)\,\mathrm{d}a;$$

(iv) a.s. for every $a \in \mathbb{R}$, $\mathrm{supp}(\mathrm{d}L_s^a(B)) \subset \{s \geq 0 : B_s = a\}$, and for fixed $a \in \mathbb{R}$,

$$\mathrm{supp}(\mathrm{d}L_s^a(B)) = \{s \geq 0 : B_s = a\} \quad a.s.. \tag{7.30}$$

Proof. (i) and (ii) are properties of local times [Theorem 7.5]. To prove Hölder continuity, the estimate (7.14) implies that for every $p = 2 + \delta > 2$,

$$\mathbb{E}\left[\sup_{0\leq s\leq t} |Y_s^b - Y_s^a|^{2+\delta}\right] \leq C_\delta (b-a)^{1+\frac{\delta}{2}} \left(t^{\frac{2+\delta}{4}} + t^{\frac{2+\delta}{2}}\right), \quad a, b \in \mathbb{R},$$

where $Y_t^a = \int_0^t \mathbb{1}_{\{B_s > a\}}\,\mathrm{d}B_s$. By Kolmogorov continuity lemma [Corollary 4.8], there exists an a.s. modification of $(Y^a)_{a\in\mathbb{R}}$ that is locally γ-Hölder continuous for each $\gamma \in (0, \frac{\delta}{4+2\delta})$, and so is

$$L_t^a(B) = (B_t - a)^+ - (B_0 - a)^+ - Y_t^a.$$

Letting $\delta \to \infty$, we conclude that $a \mapsto L_t^a(B)$ is Hölder continuous with exponent γ for each $\gamma < \frac{1}{2}$ and uniformly in t on every compact interval. Note that $\langle B, B\rangle_t = t$, (iii) follows from Corollary 7.8.

(iv) The inclusion $\mathrm{supp}(\mathrm{d}L_s^a(B)) \subset \{s \geq 0 : B_s = a\}$ holds a.s. if $a \in \mathbb{R}$ is fixed, hence simultaneously for all rational a a.s.. The continuity argument (i) allows us to get that the inclusion holds for all $a \in \mathbb{R}$ outside a zero probability set. In fact, if there exists $a \in \mathbb{R}$ such that $L_t^a(B) > L_s^a(B)$ for some $0 \leq s < t$ and $B_r \neq a$ for all $r \in [s, t]$, we can find a rational b sufficiently close to a such that the same properties hold when a is replaced by b, which gives a contradiction.

Finally we verify the a.s. equality (7.30) for fixed $a \in \mathbb{R}$. For each $q \in \mathbb{Q}$, let $\tau_q := \inf\{t \geq q : B_t = a\}$. Then our claim will follow if we can verify that a.s. for every $\epsilon > 0$, $L_{\tau_q+\epsilon}^a(B) > L_{\tau_q}^a(B)$. By the strong Markov property at time τ_q, it suffices to prove that if $(\beta_t)_{t=1}^\infty$ is a Brownian motion started from $\beta_0 = a$, then a.s. for every $\epsilon > 0$, $L_\epsilon^a(\beta) > 0$. Without loss of generality we can take $a = 0$. By Tanaka's formula and an scaling argument,

$$L_\epsilon^0(\beta) = |\beta_\epsilon| - \int_0^\epsilon \mathrm{sgn}(\beta_s)\,\mathrm{d}\beta_s \overset{d}{=} \sqrt{\epsilon}\,|\beta_1| - \sqrt{\epsilon}\int_0^1 \mathrm{sgn}(\beta_s)\,\mathrm{d}\beta_s = \sqrt{\epsilon} L_1^0(\beta).$$

Since $\mathbb{E}[L_1^0(\beta)] = \mathbb{E}|\beta_1| > 0$, we have $\mathbb{P}(L_\epsilon^0(\beta) > 0) = \mathbb{P}(L_1^0(\beta) > 0) > 0$. By Blumenthal's 0-1 law, the event

$$A := \bigcap_{n=1}^\infty \left\{L_{2^{-n}}^0(\beta) > 0\right\}$$

has probability 1, which concludes the proof. $\square$

7.3.1 Laws of Brownian Local Times

In this part we study the law of local times of Brownian motions. The following lemma gives an integral representation of Brownian local times.

Lemma 7.13. *Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion, and*

$$\beta_t = -\int_0^t \operatorname{sgn}(B_s)\,\mathrm{d}B_s, \quad t \geq 0.$$

Then $\beta = (\beta_t)_{t\geq 0}$ is also a standard Brownian motion, and $L_t^0(B) = \sup_{0\leq s\leq t} \beta_s$.

Proof. Since $\langle \beta, \beta\rangle_t = t$, by Lévy's characterization of Brownian motions [Theorem 5.12], $(\beta_t)_{t\geq 0}$ is a Brownian motion. By Tanaka's formula, $|B_t| = L_t^0(B) - \beta_t$, which immediately shows that $L_t^0(B) \geq \sup_{0\leq s\leq t} \beta_t$, since

$$L_t^0(B) \geq L_s^0(B) = \beta_s + |B_s| \geq \beta_s, \quad \text{for all } s \geq 0.$$

To show the opposite inequality, let U_t be the rightest zero of B in $[0, t]$. By the support peroperty of local times, we have $L_t^0(B) = L_{U_t}^0(B) = \beta_{U_t} \leq \sup_{0\leq s\leq t} \beta_s$, which concludes the proof. □

Recalling that we write

$$M_t = \sup_{0\leq s\leq t} B_s.$$

Corollary 4.41 asserts that $M_t \overset{d}{=} M_t - B_t \overset{d}{=} |B_t|$ for every $t > 0$. We have a stronger conclusion.

Theorem 7.14 (Lévy)**.** *The two processes $(M_t, M_t - B_t)_{t\geq 0}$ and $(L_t^0(B), |B_t|)_{t\geq 0}$ have the same law.*

Proof. By Lemma 7.13 and Tanaka's formula,

$$(L_t^0(B), |B_t|)_{t\geq 0} = \left(\sup_{s\in[0,t]} \beta_s, \sup_{s\in[0,t]} \beta_s - \beta_t\right)_{t\geq 0} \quad \text{a.s..}$$

Since $(\beta_s)_{s\geq 0}$ is a standard Brownian motion, the result follows. □

Remark. By the remark under Theorem 4.35, we can obtain an explicit formula for the density of $(L_t^0(B), |B_t|)$.

Corollary 7.15. $\mathbb{P}(L_\infty^a(B) = \infty) = 1$ *for every $a \in \mathbb{R}$.*

Proof. By the point recurrence of 1-dimensional Brownian motions, $\mathbb{P}(M_\infty = \infty) = 1$, and by Theorem 7.14, we have $\mathbb{P}(L_\infty^0(B) = \infty) = 1$. If $a \neq 0$, by the strong Markov property, $B_{T_a+t} - a$ is a standard Brownian motion, where $T_a := \inf\{t \geq 0 : B_t = a\}$. Hence $\mathbb{P}(L_\infty^a(B) = \infty) = 1$. □

Next we study the law of local times indexed by stopping times.

Proposition 7.16. *Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion.*

(i) Let $a \neq 0$ and $T_a = \inf\{t \geq 0 : B_t = a\}$. Then $L_{T_a}^0(B)$ has an exponential distribution with mean $2|a|$.

(ii) Let $a > 0$ and $U_a = \inf\{t \geq 0 : B_t = |a|\}$. Then $L_{U_a}^0(B)$ has an exponential distribution with mean a.

Proof. By simple scaling and symmetry arguments, we may take $a = 1$. Since $L_\infty^0(B) = \infty$ by Corollary 7.15, we fix $s > 0$ and take $\tau_s = \inf\{t \geq 0 : L_t^0(B) \geq s\}$, which is an a.s. finite stopping time. Furthermore, $B_{\tau_s} = 0$ by the support property of local time. By the strong Markov property, $(B'_t)_{t\geq 0} = (B_{\tau_s+t})_{t\geq 0}$ is a standard Brownian motion started from 0 and independent of $\mathscr{F}_{\tau_s}$. By Proposition 7.24,

$$L_t^0(B') = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} \int_{\tau_s}^{\tau_s+t} \mathbb{1}_{[0,\epsilon)}(B_s)\,\mathrm{d}s = L_{\tau_s+t}^0(B) - L_{\tau_s}^0(B) = L_{\tau_s+t}^0(B) - s.$$

(i) On the event $\{L^0_{T_1}(B) \geq s\} = \{\tau_s \leq T_1\}$,

$$L^0_{T_1}(B) - s = L^0_{T_1 - \tau_s}(B') = L^0_{T_1'}(B'),$$

where $T_1' = \{t \geq 0 : B_t' = 1\}$. Since the event $\{\tau_s \leq T_1\} \in \mathscr{F}_{\tau_s}$ and B' is independent of $\mathscr{F}_{\tau_s}$, the conditional distribution of $L^0_{T_1}(B) - s$ given that $L^0_{T_1}(B) \geq s$ is the same as the unconditional distribution of $L^0_{T_1}(B)$. Hence the distribution of $L^0_{T_1}(B)$ is exponential. Furthermore, by the monotone convergence theorem, Tanaka's formula (7.6), and dominated convergence theorem,

$$\mathbb{E}[L^0_{T_1}(B)] = \lim_{t\uparrow\infty} \mathbb{E}\left[L^0_{t\wedge T_1}(B)\right] = 2\lim_{t\uparrow\infty} \mathbb{E}[(B_{t\wedge T_1})^+] = 2\mathbb{E}[(B_{T_1})^+] = 2.$$

(ii) The proof is similar to (i), but we apply Tanaka's formula (7.5) to show that $\mathbb{E}[L^0_{U_1}(B)] = 1$. $\square$

Now we turn to the result on the support of the random measure $\mathrm{d}L^0_s(B)$. We consider the time change associated with $(L^0_s(B))_{s\geq 0}$, i.e.

$$\tau_t = \inf\{s \geq 0 : L^0_s(B) > t\}, \quad t \geq 0.$$

By construction, $(\tau_t)_{t\geq 0}$ has càdlàg increasing sample paths, since

$$\tau_t = \inf \bigcup_{h>0} \{s \geq 0 : L^0_s(B) > t + h\} = \inf_{h>0} \tau_{t+h} = \lim_{h\downarrow 0} \tau_{t+h}.$$

Furthermore, by Theorem 7.14 and Theorem 4.34, $(\tau_t)_{t\geq 0} \overset{d}{=} (T_t)_{t\geq 0}$ and has independent increments.

Proposition 7.17. *Let D be the countable set of jump times of $(\tau_t)_{t\geq 0}$. With probability* 1,

$$\operatorname{supp}(\mathrm{d}L^0_s(B)) = \{s \geq 0 : B_s = 0\} = \{\tau_s : s \geq 0\} \cup \{\tau_{s-} : s \in D\},$$

Remark. We may write

$$\mathcal{C} = \bigcup_{s\geq 0} (\tau_{s-}, \tau_s),$$

where (τ_{s-}, τ_s) is nonempty if and only if the local time $L^0(B)$ has a constant stretch at level s, and in that case the stretch is exactly $[\tau_{s-}, \tau_s]$. Then $\mathcal{C}$ is a countable union of open intervals, and

$$\{\tau_s : s \geq 0\} \cup \{\tau_{s-} : s \in D\}$$

is the complement of $\mathcal{C}$.

Proof. The first equality is (7.30). Next, for each $s \geq 0$ and $\epsilon > 0$, we have $L^0_{\tau_s}(B) = s$ and $L^0_{\tau_s+\epsilon}(B) > s$, which implies $\tau_s \in \operatorname{supp}(\mathrm{d}L^0_s(B))$. Since $\operatorname{supp}(\mathrm{d}L^0_s(B))$ is closed we also have $\tau_{s-} \in \operatorname{supp}(\mathrm{d}L^0_s(B))$. Hence

$$\{\tau_s : s \geq 0\} \cup \{\tau_{s-} : s \in D\} \subset \operatorname{supp}(\mathrm{d}L^0_s(B)). \tag{7.31}$$

Finally, for every $t \in \operatorname{supp}(\mathrm{d}L^0_s(B))$, we have either

$$L^0_{t+\epsilon}(B) > L^0_t(B) \quad \text{for every } \epsilon > 0,$$

or, if $t > 0$,

$$L^0_{t-\epsilon}(B) < L^0_t(B) \quad \text{for every } \epsilon > 0,$$

or both simultaneously, which implies $t = \tau_{L^0_t(B)}$ or $t = \tau_{L^0_t(B)-}$, respectively. Hence the opposite inclusion of (7.31) holds, and the second equality is valid. $\square$

8 Brownian Motions: Part II

8.1 Brownian Motions and Harmonic Functions

Brownian motions are closedly related to harmonic functions. We are going to address this particular topic.

Definition 8.1 (Harmonic functions)**.** In this section we always assume U is an open subset of $\mathbb{R}^d$. A function $u \in C^2(U)$ is said to be *harmonic*, if

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_d^2} = 0 \quad \text{in } U.$$

Remark. Let V be a bounded open subset of U with $\overline{V} \subset U$. We define the stopping time $\tau = \{t > 0 : B_t \notin V\}$. By Itô's formula, for every $x \in V$, the process $u(B_{t\wedge\tau})$ is a local martingale under $\mathbb{P}_x$:

$$u(B_{t\wedge\tau}) = u(B_0) + \int_0^{t\wedge\tau} \nabla u(B_s) \cdot \mathrm{d}B_s.$$

The quadratic variation of this process is given by

$$\langle u(B^\tau), u(B^\tau)\rangle_t = \int_0^{t\wedge\tau} |\nabla u(B_s)|^2 \,\mathrm{d}s, \quad t \geq 0.$$

Since $\overline{V}$ is a compact set, by Corollary 3.72, the stopped process $(u(B_{t\wedge\tau}))_{t\geq 0}$ is a true martingale.

Theorem 8.2. *Let $u \in C^2(U)$ be a harmonic function, and let V be a bounded open subset of U with $\overline{V} \subset U$. For every $x \in V$,*

$$u(x) = \mathbb{E}_x[u(B_\tau)].$$

Proof. Since V is bounded, $(B_t)_{t\geq 0}$ leaves V with probability 1. For each $t > 0$ and $x \in V$, by optional stoppping theorem,

$$\mathbb{E}_x\left[u(B_{t\wedge\tau})\right] = \mathbb{E}_x[u(B_0)] = u(x).$$

Since $u \in C^2(U)$ and $V \Subset U$, we let $t \uparrow \infty$ and use dominated convergence theorem to conclude the proof. □

8.1.1 Mean Value Property

In this section, we study the mean value property of harmonic functions. For any open ball $B(x,r) \subset U$, we can relate the mean value of harmonic function $u \in C^2(U)$ both on the ball $B(x,r)$ and on the sphere $\partial B(x,r)$ to its value at x. We denote by $\Sigma_{x,r}$ the uniform probability measure on the sphere $\partial B(x,r)$, i.e.

$$\mathrm{d}\Sigma_{x,r} = \frac{1}{2}\pi^{-\frac{d}{2}} r^{1-d}\Gamma\left(\frac{d}{2}\right)\mathrm{d}S,$$

Theorem 8.3 (Mean value property)**.** *Let $U \subset \mathbb{R}^d$ be an open set, and $u \in C^2(U)$. Then the following assertions are equivalent:*

(i) u is harmonic in U.

(ii) For all open balls $B(x,r) \subset U$,

$$u(x) = \int u \,\mathrm{d}\Sigma_{x,r}. \tag{8.1}$$

(iii) For all open balls $B(x,r) \subset U$,

$$u(x) = \frac{1}{m(B(x,r))}\int_{B(x,r)} u(y)\,\mathrm{d}y. \tag{8.2}$$

Proof. (i) ⇒ (ii). We let $V = B(x,r)$ in Theorem 8.2. By rotational variance of Brownian motions, the law of B_τ under $\mathbb{P}_x$ is the uniform distribution on the sphere $\partial B(x,r)$, and the desired result follows.

(ii) ⇒ (i). Fix $\epsilon > 0$ and $U_\epsilon = \{x \in U : d(x, U^c) > 2\epsilon\}$. It suffices to prove that $u \in C^2(U_\epsilon)$ and $\Delta u = 0$ on U_ϵ. We take a standard mollifier $\phi \in C_c^\infty(B(0, 2\epsilon))$, for example, $\phi(x) = \exp\left(-\frac{1}{\epsilon^2 - |x|^2}\right)\mathbb{1}_{B(0,\epsilon)}(x)$. Then

$$\begin{aligned} u(x) &= C_1 \int_0^\epsilon r^{d-1} e^{-\frac{1}{\epsilon^2 - r^2}} u(x)\,\mathrm{d}r = C_1 \int_0^\epsilon r^{d-1} \int \phi(y-x)u(y)\,\mathrm{d}\Sigma_{x,r}(y)\,\mathrm{d}r \\ &= C_2 \int_{B(x,\epsilon)} \phi(y-x)u(y)\,\mathrm{d}y = C_2(\phi * u)(x), \end{aligned}$$

where C_1, C_2 are constants depending only on ϵ and d, and we switch from Cartesian coordinate to spherical coordinate. Since ϕ is C^∞, the convolution $\phi * u$ of the last display is in fact a C^∞ function on U_ϵ.

Next, we apply Itô's formula to $u(B_t)$ under $\mathbb{P}_x$ for $x \in U_\epsilon$ and $0 < r < \epsilon$ to obtain

$$\mathbb{E}_x\left[u(B_{t\wedge\tau_{x,r}})\right] = u(x) + \frac{1}{2}\mathbb{E}_x\left[\int_0^{t\wedge\tau_{x,r}} \Delta u(B_s)\,\mathrm{d}s\right],$$

where $\tau_{x,r} = \inf\{t > 0 : B_t \notin B(x,r)\}$. Also, $\mathbb{E}_x\tau_{x,r} \le \mathbb{E}_x\left[\inf\{s > 0 : |B_s^1 - x| = r\}\right] < \infty$. By dominated convergence theorem, we let $t \uparrow \infty$ on both sides of the last display to obtain

$$\mathbb{E}_x\left[u(B_{\tau_{x,r}})\right] = u(x) + \frac{1}{2}\mathbb{E}_x\left[\int_0^{\tau_{x,r}} \Delta u(B_s)\,\mathrm{d}s\right].$$

By the mean value property of u,

$$\mathbb{E}_x\left[\int_0^{\tau_{x,r}} \Delta u(B_s)\,\mathrm{d}s\right] = 0 \tag{8.3}$$

for all $r \in (0,\epsilon)$. If $\Delta u(x) > 0$, by continuity of Δu, we can take $r_0 \in (0,\epsilon)$ such that $\Delta u > \delta$ in $B(x, r_0)$ for some $\delta > 0$. Then we have

$$\mathbb{E}_x\left[\int_0^{\tau_{x,r_0}} \Delta u(B_s)\,\mathrm{d}s\right] \ge \delta\,\mathbb{E}_x\tau_{x,r_0} > 0,$$

which contradicts (8.3). Hence $\Delta u(x) \le 0$. Similarly we have $\Delta u(x) \ge 0$. Therefore $\Delta u(x) = 0$.

(ii) ⇒ (iii). Fix $x \in U$ and $r > 0$ with $B(x,r) \subset U$. Then

$$\int_{B(x,r)} u(y)\,\mathrm{d}y = \int_0^r \int_{\partial B(x,\lambda)} u\,\mathrm{d}S\,d\lambda = \int_0^r \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\lambda^{d-1} \int u\,d\Sigma_{x,\lambda}\,d\lambda = m(B(x,r))u(x),$$

where we use the fact

$$m(B(x,r)) = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} r^d.$$

(iii) ⇒ (ii). Assume u has the mean value property (8.2). Define $\psi : (0,\infty) \to \mathbb{R}$ by

$$\psi(r) = \frac{1}{r^{d-1}} \int_{\partial B(x,r)} u\,\mathrm{d}S = C_d \int u\,\mathrm{d}\Sigma_{x,r}, \quad \text{where } C_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}.$$

Then for all $r > 0$ with $B(x,r) \subset U$, we switch from Cartesian coordinates to polar coordinates to obtain

$$r^d m(B(x,1))u(x) = m(B(x,r))u(x) = \int_{B(x,r)} u(y)\,\mathrm{d}y = \int_0^r s^{d-1}\psi(s)\,\mathrm{d}s.$$

Differentiating with respect to r, we know that $\psi(r)$ is constant on $0 < r < d(x, U^c)$. Using the well-known fact that $\mathrm{d}m(B(x,r))/\mathrm{d}r = C_d r^{d-1}$, we have $\psi(r) = C_d u(x)$, which complete the proof. □

An immediate corollary of this property is the maximum principle for harmonic functions.

Theorem 8.4 (Strong maximum principle)**.** *Let U be a bounded, open subset of $\mathbb{R}^d$, and let $u \in C^2(U) \cap C(\overline{U})$ be a harmonic function on U. Then*

$$\max_{\overline{U}} u = \max_{\partial U} u. \tag{8.4}$$

Moreover, if U is connected and there exists $x^* \in U$ such that $u(x^*) = \max_{\overline{U}} u$, then u is constant within U.

Remark. According to our proof, a harmonic function $u \in C^2(U)$ must be smooth.

Proof. Suppose such a point $x^* \in U$ exists. For all $0 < r < d(x^*, \partial U)$, the mean value property implies

$$M = u(x^*) = \frac{1}{m(B(x^*, r))} \int_{B(x^*,r)} u(y)\,\mathrm{d}y \le M,$$

which holds only if $u \equiv M$ within $B(x^*, r)$. Hence the set $\{x \in U : u(x) = M\}$ is both open and relatively closed in U, which equals U if U is connected. The identity (8.4) follows from this. $\square$

8.1.2 Recurrence and Transience of Multi-dimensional Brownian Motions

Radial harmonic functions. In this subsection, we apply harmonic functions to study the recurrence and transience of Brownian motions in $\mathbb{R}^d$, where $d \ge 2$. A harmonic radial function $x \mapsto \phi(|x|)$ on $\mathbb{R}^d \backslash \{0\}$ satisfies

$$0 = \Delta_x \phi(|x|) = \phi''(|x|) + \frac{d-1}{|x|}\phi'(|x|) \quad \Leftrightarrow \quad \phi''(r) + \frac{d-1}{r}\phi'(r) = 0.$$

By solving the differential equation, we know that ϕ must be of the form

$$\phi(r) = \begin{cases} a + b\log r, & d = 2, \\ a + br^{2-d}, & d \ge 3, \end{cases} \quad r > 0. \tag{8.5}$$

In our following discussion, we use the function

$$\phi(r) = \begin{cases} \log r, & d = 2, \\ r^{2-d}, & d \ge 3, \end{cases} \quad r > 0. \tag{8.6}$$

Then $x \mapsto \phi(|x|)$ is a harmonic function in $\mathbb{R}^d \backslash \{0\}$.

Theorem 8.5. *For each $a \ge 0$, we define the hitting time $\tau_a = \inf\{t \ge 0 : |B_t| = a\}$, with the convention $\inf \emptyset = \infty$. Let $x \in \mathbb{R}^d \backslash \{0\}$, and let ϵ and R be such that $0 < \epsilon < |x| < R$. Then*

$$\mathbb{P}_x(\tau_\epsilon < \tau_R) = \begin{cases} \dfrac{\log R - \log|x|}{\log R - \log \epsilon}, & d = 2, \\[2ex] \dfrac{R^{2-d} - |x|^{2-d}}{R^{2-d} - \epsilon^{2-d}}, & d \ge 3. \end{cases} \tag{8.7}$$

Consequently, we have $\mathbb{P}_x(\tau_0 < \infty) = 0$, and for every $\epsilon \in (0, |x|)$,

$$\mathbb{P}_x(\tau_\epsilon < \infty) = \begin{cases} 1, & d = 2, \\ \left(\dfrac{\epsilon}{|x|}\right)^{d-2}, & d \ge 3. \end{cases} \tag{8.8}$$

Proof. Let ϕ be the function defined in (8.6), which is harmonic in the annulus $A_{\epsilon,R} = \{y \in \mathbb{R}^d : \epsilon < |y| < R\}$. Let $\tau = \inf\{t \geq 0 : B_t \notin A_{\epsilon,R}\}$. Then $\mathbb{P}_x(\tau < \infty) = 1$. Furthermore, by Theorem 8.2,

$$\phi(|x|) = \mathbb{E}_x \phi(|B_\tau|) = \mathbb{P}_x(\tau_\epsilon < \tau_R)\, \phi(\epsilon) + (1 - \mathbb{P}_x(\tau_\epsilon < \tau_R))\, \phi(R).$$

This implies

$$\mathbb{P}_x(\tau_\epsilon < \tau_R) = \frac{\phi(R) - \phi(|x|)}{\phi(R) - \phi(\epsilon)}.$$

We fix $R > |x|$. As $\epsilon \downarrow 0$, the limit $\tau_\epsilon \uparrow \tau_0$ holds $\mathbb{P}_x$ a.s.. Hence we pass $\epsilon \downarrow 0$ to assert $\mathbb{P}_x(\tau_0 < \tau_R) = 0$. Since $R > |x|$ is arbitrary, and $\tau_R \uparrow \infty$, $\mathbb{P}_x$ a.s., we let $R \uparrow \infty$ to conclude that $\mathbb{P}_x(\tau_0 < \infty) = 0$. Finally, we fix $0 < \epsilon < |x|$ and $\mathbb{P}_x(\tau_\epsilon < \infty) = \lim_{R\to\infty}(\tau_\epsilon < \tau_R)$ to conclude (8.8). □

Remark. By translation invariance of Brownian motions, for any pair of distinct points $x, y \in \mathbb{R}^d$, we have

$$\mathbb{P}_x(\exists t \geq 0 \text{ such that } B_t = y) = \mathbb{P}_{x-y}(\tau_0 < \infty) = 0.$$

By this theorem, multi-dimensional Brownian motions are *point-transient.*

Theorem 8.6. *Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion.*

(i) If the dimension $d = 2$, then $(B_t)_{t\geq 0}$ is neighborhood recurrent, meaning that for every nonempty set $U \subset \mathbb{R}^d$, the set $\{t \geq 0 : B_t \in U\}$ is unbounded.

(ii) If the dimension $d \geq 3$, then $(B_t)_{t\geq 0}$ is transient, meaning that

$$\lim_{t\to\infty} |B_t| = \infty \quad a.s..$$

In other words, $(B_t)_{t\geq 0}$ leaves any bounded set with probability 1.

Proof. (i) We first consider an open ball $B(0, \epsilon)$. By Theorem 8.5, starting from $x \neq 0$, the Brownian motion $(B_t)_{t\geq 0}$ never hits 0 but hits any open ball centered at 0. Thus, almost surely, for every $M > 0$, $(B_t)_{t\geq 0}$ leave $B(0, 2\epsilon)$ at some time later than M, and then visit $B(0, \epsilon)$ by strong Markov property. By translation invariance, given any open ball $B(x, \epsilon)$ in $\mathbb{R}^d$, the Brownian motion B hits it at arbitrarily large times, a.s..

Note that every contains an open ball of rational radius centered at a point with rational coordinates. The conclusion follows from a countable union argument.

(ii) Assume without loss of generality that the starting point of $(B_t)_{t\geq 0}$ is $x \neq 0$. Since $x \mapsto |x|^{2-d}$ is harmonic in $\mathbb{R}^d \backslash \{0\}$ and $\mathbb{P}_x(\tau_0 < \infty) = 0$, the process $(|B_t|^{2-d})_{t\geq 0}$ is a local martingale under $\mathbb{P}_x$. By Proposition 3.67, $(|B_t|^{2-d})_{t\geq 0}$ is a nonnegative supermartingale, which a.s. converges as $t \to \infty$. The a.s. limit must be 0, otherwise the path $t \mapsto B_t$ would be bounded. Hence $|B_t| \to \infty$ as $t \to \infty$. □

Remark. (I) According to the neighborhood recurrence property of planar Brownian motions ($d = 2$), the sample path $\{B_t(\omega)\}_{t\geq 0}$ is almost surely dense in $\mathbb{R}^2$.

(II) In regard of the growth rate of $|B_t|_{t\geq 0}$ when $d \geq 3$, we fix a sequence $t_n \uparrow \infty$. For each $\epsilon > 0$,

$$\mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{\frac{|B_{t_n}|}{\sqrt{t_n}} \leq \epsilon\right\}\right) \geq \limsup_{n\to\infty} \mathbb{P}\left(\frac{|B_{t_n}|}{\sqrt{t_n}} \leq \epsilon\right) = \mathbb{P}(|B_1| \leq \epsilon) > 0.$$

By Blumenthal's 0-1 law, the probability on the left-hand side must therefore be one, and

$$\liminf_{t\to\infty} \frac{|B_t|}{\sqrt{t}} = 0 \quad \text{a.s..}$$

8.1.3 The Dirichlet Problem

In this part, we assume U is an open and bounded subset of $\mathbb{R}^d$, and study the boundary-value problem

$$\begin{cases} \Delta u = 0 & \text{in } U, \\ u = g & \text{on } \partial U, \end{cases} \tag{8.9}$$

where g is a continuous function on ∂U, and $u \in C^2(U) \cap C(\overline{U})$ is the unknown.

Proposition 8.7. *Define exit time $\tau = \{t \geq 0 : B_t \notin U\}$, and let g be a bounded measurable function on ∂U.*

(i) The function $u(x) = \mathbb{E}_x[g(B_\tau)]$ is harmonic in U;

(ii) In addition, if g is continuous on ∂U, and u solves the Dirichlet problem (8.9)*, then $u(x) = \mathbb{E}_x[g(B_\tau)]$ for each $x \in U$.*

Proof. (i) It suffices to verify that $u(x) = \mathbb{E}_x[g(B_\tau)]$ satisfies the mean value property. For each $x \in U$, we fix $B(x,r) \subset U$ and define $\tau_{x,r} = \inf\{t \geq 0 : |B_t - x| = r\}$. For each $\mathbf{w} \in C(\mathbb{R}_+, \mathbb{R}^d)$ with $\mathbf{w}(0) \in U$, we define $\Phi(\mathbf{w})$ to be the value of g at the first exit time of $\mathbf{w}$ from D, i.e. $\Phi(\mathbf{w}) = g(\inf\{t \geq 0 : \mathbf{w}(t) \notin D\})$. Then

$$g(B_\tau) = \Phi((B_t)_{t\geq 0}) = \Phi((B_{\tau_{x,r}+t})_{t\geq 0}) \quad \mathbb{P}_x \ a.s..$$

By the strong Markov property of Brownian motions,

$$u(x) = \mathbb{E}_x[g(B_\tau)] = \mathbb{E}_x\left[\Phi((B_{\tau_{x,r}+t})_{t\geq 0})\right] = \mathbb{E}_x\left[\mathbb{E}_{B_{\tau_{x,r}}}\left[\Phi((B_t)_{t\geq 0})\right]\right] = \mathbb{E}_x\left[u(B_{\tau_{x,r}})\right].$$

Since the law of $B_{\tau_{x,r}}$ under $\mathbb{P}_x$ is the uniform probability measure $\Sigma_{x,r}$ on $\partial B(x,r)$, we conclude the proof.

(ii) For each $x \in U$, we fix $B(x,r) \subset U$. For each $0 < \epsilon < r$, we set $U_\epsilon = \{x \in U : d(x, U^c) > \epsilon\}$, and define $\tau_\epsilon = \inf\{t \geq 0 : B_t \notin U_\epsilon\}$. Since u is harmonic in U, by Theorem 8.2,

$$u(x) = \mathbb{E}_x\left[u(B_{\tau_\epsilon})\right].$$

It is clear that τ_ϵ is monotone increasing as $\epsilon \downarrow 0$, and the limit $\tau_0 \leq \tau$. On the other hand, we have $B_{\tau_0} \in \partial U$ by te continuity of sample paths, which implies $\tau_0 \geq \tau$. Therefore $\tau_\epsilon \downarrow \tau$ as $\epsilon \downarrow 0$. By the dominated convergence theorem, we have $u(x) = \mathbb{E}_x[u(B_\tau)] = \mathbb{E}_x[g(B_\tau)]$. □

Remark. The second assertion implies that if a solution to the Dirichlet problem (8.9) exists, it must be unique and has the form $u(x) = \mathbb{E}_x[g(B_\tau)]$. We next study the existence of solutions.

Definition 8.8 (Exterior cone condition)**.** Let $U \subset \mathbb{R}^d$ be open. If $y \in \partial U$, we say U satisfies the *exterior cone condition at* y if there exists an open cone C with apex y and $\epsilon > 0$ such that $C \cap B(y, \epsilon) \subset U^c$.

Lemma 8.9 (Brownian motions avoiding a cone)**.** *Define exit time $\tau = \{t \geq 0 : B_t \notin U\}$. Under the exterior cone condition, we have for every $y \in \partial U$ and every $\eta > 0$ that*

$$\lim_{U \ni x \to y} P_x(\tau > \eta) = 0.$$

Proof. For every $\xi \in \mathbb{R}^d$ with $|v| = 1$ and every $\gamma \in (0,1)$, consider the circular cone

$$C(\xi, \gamma) = \left\{x \in \mathbb{R}^d : x^\top \xi > (1-\gamma)|x|\right\}.$$

By the exterior cone condition, there exists $\xi \in \mathbb{R}^d$, $\gamma \in (0,1)$ and $\epsilon > 0$ such that $y + C(\xi,\gamma) \cap B(0,r) \subset U^c$. For notation simplicity we define the truncated cone $C = C(\xi,\gamma) \cap B(0,r)$, and fix a smaller truncated cone $D = C(\xi, \frac{\gamma}{2}) \cap B(0, \frac{r}{2})$. For an open $V \subset \mathbb{R}^d$, let $\tau_V = \inf\{t \geq 0 : B_t \in V\}$.

By Proposition 3.14, the event $\{\tau_{C(\xi,\gamma/2)} = 0\}$ is contained in the germ σ-algebra $\bigcap_{t>0} \mathscr{F}_t$, where $(\mathscr{F}_t)_{t\geq 0}$ is the canonical filtration $(B_t)_{t\geq 0}$. For any $s > 0$,

$$\mathbb{P}_0\left(\tau_{C(\xi,\gamma/2)} \leq s\right) \geq \mathbb{P}_0\left(B_s^\top \xi - \left(1 - \frac{\gamma}{2}\right)|B_s| > 0\right) = \mathbb{P}_0\left(Y^\top \xi - \left(1 - \frac{\gamma}{2}\right) > 0\right) =: \delta > 0,$$

where the law of Y is the uniform probability measure on the unit sphere $\partial B(0,1)$. We let $s \downarrow 0$ and apply Blumenthal's 0-1 law [Theorem 6.15] to obtain that $\tau_{C(\xi,\gamma/2)} = 0$, $\mathbb{P}_0$-a.s.. By continuity of Brownian motions, we also have $\tau_D = 0$, $\mathbb{P}_0$-a.s.. On the other hand, for each $r \in (0, \frac{\epsilon}{2})$, we set $D_r = \{x \in D : |x| > r\}$. Then $D_r \uparrow D$ as $r \downarrow 0$, and $\tau_{D_r} \downarrow \tau_D = 0$. Thus, for any $\beta > 0$, we fix $r > 0$ so small that $\mathbb{P}_0(\tau_{D_r} \leq \eta) \geq 1 - \beta$. Since $y + C \subset U^c$, we have the estimate

$$\mathbb{P}_x(\tau \leq \eta) \geq \mathbb{P}_x(\tau_{y+C} \leq \eta) = \mathbb{P}_0(\tau_{y-x+C} \leq \eta).$$

We also note that $\overline{D}_r \subset C$. When $|y - x| < \frac{1}{2} d(\overline{D}_r, \partial C)$, the shifted cone $y - x + C \supset D_r$. Therefore

$$\mathbb{P}_x(\tau \leq \eta) \geq \mathbb{P}_0(\tau_{D_r} \leq \eta) \geq 1 - \beta.$$

Since $\beta > 0$ is arbitrary, we conclude the proof. □

The exterior cone condition is sufficient for the existence of solution to the Dirichlet problem.

Theorem 8.10 (Solution of the Dirichlet problem)**.** *Let U be an open bounded subset of $\mathbb{R}^d$ such that each $y \in \partial U$ satisfies the exterior cone condition. Then for each continuous function g on ∂U, the solution of the Dirichlet problem* (8.9) *is uniquely given by the function*

$$u(x) = \mathbb{E}_x[g(B_\tau)], \quad \textit{where } \tau = \inf\{t \geq 0 : B_t \notin U\}.$$

Proof. Following Proposition 8.7 (i), it suffices to show that $\lim_{U \ni x \to y} u(x) = g(y)$ for each $y \in \partial U$. We fix $\epsilon > 0$. By continuity of g, there exists $\delta > 0$ such that $|g(z) - g(y)| < \epsilon/3$ for all $z \in \partial U \cap B(y, \delta)$. Also, we fix $M > 0$ such that $|g(z)| \leq M$ for all $z \in \partial U$. Then for all $\eta > 0$,

$$\begin{aligned}|u(x) - g(y)| &\leq \mathbb{E}_x\left[|g(B_\tau) - g(y)|\mathbb{1}_{\{\tau \leq \eta\}}\right] + \mathbb{E}_x\left[|g(B_\tau) - g(y)|\mathbb{1}_{\{\tau > \eta\}}\right] \\ &\leq \mathbb{E}_x\left[|g(B_\tau) - g(y)|\mathbb{1}_{\{\tau \leq \eta\} \cap \{\sup_{t\in[0,\eta]} |B_t - x| \leq \delta/2\}}\right] + 2M\mathbb{P}_x\left(\sup_{t\in[0,\eta]} |B_t - x| > \frac{\delta}{2}\right) + 2M\mathbb{P}_x(\tau > \eta).\end{aligned}$$

(i) Under the event $\{\tau \leq \eta\} \cap \{\sup_{t\in[0,\eta]} |B_t - x| \leq \delta/2\}$, we have $|B_\tau - y| \leq |B_t - x| + |y - x| < \delta$ for all $|y - x| < \delta/2$. Then

$$\mathbb{E}_x\left[|g(B_\tau) - g(y)|\mathbb{1}_{\{\tau \leq \eta\} \cap \{\sup_{t\in[0,\eta]} |B_t - x| \leq \delta/2\}}\right] \leq \frac{\epsilon}{3}.$$

(ii) By translation invariance and continuity of Brownian motions, we apply dominated convergence theorem to conclude $\mathbb{P}_0\left(\sup_{t\in[0,\eta]} |B_t| > \frac{\delta}{2}\right) \downarrow 0$ as $\eta \downarrow 0$. Hence we fix $\eta > 0$ so small that

$$\mathbb{P}_x\left(\sup_{t\in[0,\eta]} |B_t - x| > \frac{\delta}{2}\right) = \mathbb{P}_0\left(\sup_{t\in[0,\eta]} |B_t| > \frac{\delta}{2}\right) \leq \frac{\epsilon}{6M}.$$

(iii) By Lemma 8.9, we can fix $r \in (0, \delta/2)$ such that $\mathbb{P}_x(\tau > \eta) < \epsilon/(6M)$ for all $x \in U \cap B(y, r)$. Combining the last three estimates, we conclude that $|u(x) - g(y)| < \epsilon$ for all $x \in U \cap B(y, r)$. Since $\epsilon > 0$ is arbitrary, we have $\lim_{U \ni x \to y} u(x) = g(y)$, which completes the proof. □

Remark. In fact, our proof can be extended to certain unbounded open sets. For example, if $U = \{x \in \mathbb{R}^d : x_d > 0\}$ is the upper half-space, the above theorem also applies if g is bounded and continuous on ∂U.

8.1.4 The Poisson Kernel and Exit Distributions

In this part, we study two special cases Dirichlet problem (8.9) and derive the corresponding exit distributions.

Half-space. We define the *Poisson kernel of the upper half space* $U = \{(x,y) \in \mathbb{R}^d : x \in \mathbb{R}^{d-1}, y > 0\}$ to be

$$p(x,y) = \frac{c_d y}{\left(|x|^2 + y^2\right)^{d/2}}, \quad \text{where } c_d = \frac{\Gamma(\frac{d}{2})}{\pi^{d/2}}.$$

The choice of c_d implies $\int_{\mathbb{R}^{d-1}} p(x,1)\,\mathrm{d}x = 1$. We next compute the partial derivatives:

$$\frac{\partial p}{\partial x_i} = -\frac{dx_i}{|x|^2 + y^2} p(x,y), \quad \frac{\partial^2 p}{\partial x_i^2} = \frac{d}{|x|^2 + y^2}\left(\frac{(d+2)x_i^2}{|x|^2 + y^2} - 1\right) p(x,y), \quad i = 1, \cdots, d-1,$$

and

$$\frac{\partial p}{\partial y} = \left(\frac{1}{y} - \frac{dy}{|x|^2 + y^2}\right) p(x,y), \frac{\partial^2 p}{\partial y^2} = \frac{d}{|x|^2 + y^2}\left(\frac{(d+2)y^2}{|x|^2 + y^2} - 3\right) p(x,y).$$

As a result, $\Delta p(x,y) = 0$. Therefore p is harmonic on U.

Theorem 8.11. *Let $g : \mathbb{R}^{d-1} \to \mathbb{R}$ be bounded and continuous, and*

$$u(x,y) = \int_{\mathbb{R}^{d-1}} p(x-\xi, y) g(\xi)\,\mathrm{d}\xi, \quad x \in \mathbb{R}^{d-1},\ y > 0.$$

Thne u solves the Dirichlet problem (8.9) *on $U = \{(x,y) : x \in \mathbb{R}^{d-1}, y > 0\}$ with boundary value g.*

Proof. We write $p_\xi(x,y) = p(x-\xi,y)$ for $\xi \in \mathbb{R}^{n-1}$. Then p_ξ is a harmonic function for every $\xi \in \mathbb{R}^{n-1}$. Interchanging the integral and derivative (which is justified by the dominated convergence theorem), we have

$$\Delta u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Now it remains to verify the boundary condition. Since g is bounded, we may assume $|g| \le 1$ by scaling. For each $x \in \mathbb{R}^{n-1}$ and $\epsilon > 0$, by continuity of g, we take $\delta > 0$ such that $|g(\xi) - g(x)| < \epsilon$ for all $|\xi - x| \le \delta$. The choice of c_d ensures that $\int_{\mathbb{R}^{n-1}} p_\xi(x,y)\,\mathrm{d}x = 1$ for all $y > 0$. By definition of the Poisson kernel p, there exists $y_\delta > 0$ such that

$$\int_{B(0,\delta)} p(\xi,y)\,d\xi = \int_{B(0,\frac{\delta}{y})} p(\xi,1)\,d\xi > 1 - \epsilon$$

for all $y \in (0, y_\epsilon)$. Then we have

$$\begin{aligned} |u(x,y) - g(x)| &\le \int_{|\xi - x| \le \delta} p(x-\xi,y)|g(\xi) - g(x)|\,\mathrm{d}\xi + \int_{|\xi - x| > \delta} p(x-\xi,y)|g(\xi) - g(x)|\,\mathrm{d}\xi \\ &\le \int_{|\xi - x| \le \delta} p(x-\xi,y)\epsilon\,\mathrm{d}\xi + 2\int_{|\xi| > \delta} p(\xi,y)\,\mathrm{d}\xi \le 3\epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, $u(x,y) \to g(x)$ as $y \downarrow 0$. This complete the proof. □

Next, we study the exit distribution on the half-space U, i.e. the law of the position where a Brownian motion B exits from the half-space U. According to Theorem 8.10, for all bounded continuous functions g,

$$\mathbb{E}_{(x,y)}[g(B_\tau)] = \int_{\mathbb{R}^{d-1}} p(x-\xi,y) g(\xi)\,\mathrm{d}\xi,$$

where $\tau = \inf\{t \ge 0 : B_t \in \partial U\}$. This expectation determines the law of B_τ.

Corollary 8.12. *The law of B_τ under $\mathbb{P}_{(x,y)}$ is given by the probability density function*

$$\rho_{(x,y)}(\xi,0)=p_\xi(x,y)=\frac{c_d y}{\left(|x-\xi|^2+y^2\right)^{d/2}},\quad \xi\in\mathbb{R}^{d-1}.$$

In fact, by translation and rotation invariance of Brownian motions, we can adapt the above conclusion to any hitting time for a hyperplane in $\mathbb{R}^d$ (the starting point should not be on the hyperplane). This fact inspires a probabilistic proof of Liouville's theorem for harmonic functions.

Theorem 8.13 (Liouville's Theorem)**.** *If $u\in C^2(\mathbb{R}^d)$ is a bounded harmonic function, then u is constant.*

Proof. Since u is a bounded harmonic function on $\mathbb{R}^d$, the process $(u(B_t))_{t\geq 0}$ is a continuous local martingale. By Proposition 3.67 (ii), $(u(B_t))_{t\geq 0}$ is a uniformly integrable martingale.

Let x and y be two distinct points in $\mathbb{R}^d$, and take H the hyperplane in $\mathbb{R}^d$ such that the reflection in H maps x to y. Define $\tau_H=\inf\{t\geq 0: B_t\in H\}$. Then B_{τ_H} has the same exit distribution under $\mathbb{P}_x$ and $\mathbb{P}_y$. By the optional stopping theorem [Theorem 3.61] for uniformly integrable martingales,

$$u(x)=\mathbb{E}_x[u(B_{\tau_H})]=\mathbb{E}_y[u(B_{\tau_H})]=u(y).$$

Since $x,y\in\mathbb{R}^d$ are arbitrary, we conclude the proof. $\square$

Unit ball. We define the *Poisson kernel of the unit ball* $B(0,1)=\{x\in\mathbb{R}^d:|x|<1\}$ to be

$$K(x,y)=K_y(x)=\frac{1-|x|^2}{|x-y|^d},\quad \text{where } y\in S^{d-1}=\partial B(0,1).$$

Then the partial derivatives are

$$\frac{\partial K_y}{\partial x_i}=-\frac{d(x_i-y_i)(1-|x|^2)}{|x-y|^{d+2}}-\frac{2x_i}{|x-y|^d},$$

and

$$\frac{\partial^2 K_y}{\partial x_i^2}=\frac{d(d+2)(x_i-y_i)^2(1-|x|^2)}{|x-y|^{d+4}}-\frac{d(1-|x|^2)}{|x-y|^{d+2}}+\frac{4dx_i(x_i-y_i)}{|x-y|^{d+2}}-\frac{2}{|x-y|^d},$$

Note that $|y|^2=1$. Then

$$\Delta K_y=\frac{2d(1-|x|^2)}{|x-y|^{d+2}}+\frac{2d}{|x-y|^{d+2}}\left(|x|^2+|x-y|^2-|y|^2\right)-\frac{2d}{|x-y|^d}=0.$$

Therefore $y\mapsto K(x,y)$ is a harmonic function on the unit ball $B(0,1)$ for each $y\in S^{d-1}$. Next, we show that for each $x\in B(0,1)$, the mapping $y\mapsto K(x,y)$ is a density function on the unit sphere S^{d-1}, i.e.

$$\int_{\partial B(0,1)}K(x,y)\,\mathrm{d}\Sigma(y)=1,\tag{8.10}$$

where $\Sigma=\Sigma_{0,1}$ is the uniform probability measure on S^{d-1}. Define

$$F(x)=\int_{S^{d-1}}K(x,y)\,\mathrm{d}\Sigma(y),\quad x\in B.$$

By Fubini's theorem and the mean value property of $z\mapsto K(z,y)$, we can verify mean-value property of F:

$$\int F\,\mathrm{d}\Sigma_{x,r}=\iint K(z,y)\,\mathrm{d}\Sigma(y)\,\mathrm{d}\Sigma_{x,r}(z)=\iint K(z,y)\,\mathrm{d}\Sigma_{x,r}(z)\,\mathrm{d}\Sigma(y)=\int K(x,y)\,\mathrm{d}\Sigma(y)=F(x).\tag{8.11}$$

By rotation invariance of $(x, y) \mapsto K(x, y)$ and Σ, we know that F is a radial harmonic function, which is of form (8.5). Moreover, noticing that F is bounded in any punctured neighborhood of 0, we know that F must be a constant. Hence $F \equiv F(0) = 1$.

Theorem 8.14. *Let g be a continuous function on $S^{d-1} = \partial B(0,1)$, and*

$$u(x) = \int_{S^{d-1}} K(x, y) g(y) \, \mathrm{d}\Sigma(y), \quad x \in B(0,1).$$

Then u solves the Dirichlet problem (8.9) *on the unit ball $B(0,1)$ with boundary value g.*

Proof. Similar to (8.11), we can prove that u has the mean-value property, hence is harmonic in $B(0,1)$. To verify the boundary condition, we fix $y_0 \in S^{d-1}$ and $\epsilon > 0$. By continuity of g, we may assume $|g| \leq 1$ on S^{d-1} by an scaling argument, and take $\delta > 0$ such that $|g(y) - g(y_0)| < \epsilon$ for all $y \in S^{d-1} \cap B(y_0, \delta)$. Meanwhile, for $x \in B(0,1)$ with $|x - y_0| \leq \frac{\delta}{2}$ and $y \in S^{d-1}$ with $|y - y_0| \geq \delta$, one have

$$K(x, y) = \frac{1 - |x|^2}{|x - y|^d} \leq \left(\frac{2}{\delta}\right)^d (1 - |x|^2) \downarrow 0, \quad \text{as } x \to y_0..$$

Then there exists $\delta_1 > 0$ such that for all $x \in B(0,1)$ with $|x - y| \leq \delta_1$,

$$\int_{S^{d-1} \setminus B(y_0, \delta)} K(x, y) \, \mathrm{d}\Sigma(y) \leq \epsilon.$$

Therefore, for all $x \in B(0,1)$ with $|x - y| \leq \delta_1$,

$$\begin{aligned} |u(x) - g(y_0)| &\leq \int_{S^{d-1} \cap B(y_0, \delta)} K(x, y) |g(x) - g(y_0)| \, \mathrm{d}\Sigma(y) + \int_{S^{d-1} \setminus B(y_0, \delta)} K(x, y) |g(x) - g(y_0)| \, \mathrm{d}\Sigma(y) \\ &\leq \int_{S^{d-1} \cap B(y_0, \delta)} \epsilon K(x, y) \, \mathrm{d}\Sigma(y) + 2 \int_{S^{d-1} \setminus B(y_0, \delta)} K(x, y) \, \Sigma(\mathrm{d}y) \leq 3\epsilon. \end{aligned}$$

Since $\epsilon > 0$, we prove the boundary condition $u(x) \to g(y_0)$ as $B \ni x \to y_0$. □

Again, we compare this result with Theorem 8.10 to get the exit distribution of the Brownian motion $(B_t)_{t \geq 0}$ from the unit ball $B(0,1)$.

Corollary 8.15. *Let $\tau = \inf\{t : B_t \notin B(0,1)\}$ be the exit time of Brownian motion $(B_t)_{t \geq 0}$ from unit ball $B(0,1)$. For every $x \in B(0,1)$, the law of B_τ under $\mathbb{P}_x$ has density $y \mapsto K(x, y)$ with respect to the uniform probability measure $\mathrm{d}\Sigma(y)$ on S^{d-1}.*

8.2 Occupation Times and Green's Functions

8.2.1 Green's Functions

We start from the following fundamental result about the occupation time of the Brownian motion.

Proposition 8.16. *Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion and $t>0$, and let $U\subset\mathbb{R}^d$ be a nonempty bounded set. Then for any $x\in\mathbb{R}^d$,*

(i) if $d\leq 2$, we have
$$\mathbb{P}_x\left(\int_0^\infty \mathbb{1}_U(B_t)\,\mathrm{d}t=\infty\right)=1;$$
(ii) if $d\geq 3$, we have
$$\mathbb{E}_x\left[\int_0^\infty \mathbb{1}_U(B_t)\,\mathrm{d}t\right]<\infty.$$

Proof. Since a bounded set is contained in an open ball and contains an open ball, we may assume that U is an open ball. By shifting, we can assume $U=B(0,r)$.

(i) Let $d\leq 2$ and $D=B(0,2r)$. Define $T_0=\inf\{t>0:B_t\notin D\}$. For each $k\in\mathbb{N}$, define
$$S_k=\inf\{t>T_{k-1}:B_t\in U\},\quad\text{and}\quad T_k=\inf\{t>S_k:B_t\notin D\}.$$
Almost surely, these stopping times are finite. By the strong Markov property, for each $k\geq 1$,
$$\begin{aligned}\mathbb{P}_x\left(\int_{S_k}^{T_k}\mathbb{1}_U(B_t)\,\mathrm{d}t\geq\epsilon\,\middle|\,\mathscr{F}_{S_k+}\right)&=\mathbb{P}_{B_{S_k}}\left(\int_0^{T_0}\mathbb{1}_U(B_t)\,\mathrm{d}t\geq\epsilon\right)\\&=\mathbb{E}_x\left[\mathbb{P}_{B_{S_k}}\left(\int_0^{T_0}\mathbb{1}_U(B_t)\,\mathrm{d}t\geq\epsilon\right)\right]=\mathbb{P}_x\left(\int_{S_k}^{T_k}\mathbb{1}_U(B_t)\,\mathrm{d}t\geq\epsilon\right),\end{aligned}$$
where we get the second inequality by rotation invariance. Since the second expression does not depend on k, the random variables
$$\int_{S_k}^{T_k}\mathbb{1}_U(B_t)\,\mathrm{d}t,\quad k=1,2,\cdots$$
are i.i.d.. Since these random variables are not identically zero and nonnegative, they have positive expectation. By the strong law of large numbers,
$$\int_0^\infty\mathbb{1}_U(B_t)\,\mathrm{d}t\geq\lim_{n\to\infty}\sum_{k=1}^n\int_{S_k}^{T_k}\mathbb{1}_U(B_t)\,\mathrm{d}t=\infty,\quad\text{a.s..}$$

(ii) Let $d\geq 3$ and $p_t(x,y)=(2\pi t)^{-d/2}\exp\left(-\frac{|x-y|^2}{2t}\right)$ the transition kernel of the Brownian motion. Then
$$\begin{aligned}\int_0^\infty p_t(x,y)\,\mathrm{d}t&=\int_0^\infty\frac{1}{(2\pi t)^{d/2}}e^{-\frac{|x-y|^2}{2t}}\,\mathrm{d}t=\int_\infty^0\left(\frac{s}{\pi|x-y|^2}\right)^{d/2}e^{-s}\left(-\frac{|x-y|^2}{2s^2}\right)\mathrm{d}s\\&=\frac{|x-y|^{2-d}}{2\pi^{d/2}}\int_0^\infty s^{\frac{d}{2}-2}e^{-s}\,\mathrm{d}s=\frac{\Gamma\left(\frac{d}{2}-1\right)}{2\pi^{\frac{d}{2}}|x-y|^{d-2}}\end{aligned}$$
Apply Fubini's theorem and switch from Cartesian to polar coordinates, we obtain
$$\begin{aligned}\mathbb{E}_0\left[\int_0^\infty\mathbb{1}_{B(0,r)}(B_t)\,\mathrm{d}t\right]&=\int_0^\infty\mathbb{P}_0\left(B_t\in B(0,r)\right)\mathrm{d}t=\int_0^\infty\int_{B(0,r)}p_t(0,y)\,\mathrm{d}y\,\mathrm{d}t\\&=\frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\int_0^r\rho^{d-1}\int_0^\infty p_t(0,y)\,\mathrm{d}t\,\mathrm{d}\rho=\frac{r^2}{d-2}<\infty.\end{aligned}$$

To extend the conclusion to $x \neq 0$, we apply the strong Markov property at the exit time τ from $B(0, |x|)$ of a Brownian motion starting at 0 and the rotational invariance to obtain

$$\mathbb{E}_x \left[\int_0^\infty \mathbb{1}_{B(0,r)}(B_t)\,\mathrm{d}t \right] = \mathbb{E}_0 \left[\int_\tau^\infty \mathbb{1}_{B(0,r)}(B_t)\,\mathrm{d}t \right] \leq \mathbb{E}_0 \left[\int_0^\infty \mathbb{1}_{B(0,r)}(B_t)\,\mathrm{d}t \right] < \infty,$$

which completes the proof. □

Remark. We write

$$\Phi(x,y) = \int_0^\infty p_t(x,y)\,\mathrm{d}t$$

the *Green function* or the *potential kernel*, because $\Phi(x,\cdot)$ is the electrostatic potential of a unit charge at x. For the case $d \geq 3$, we have

$$\Phi(x,y) = \frac{\Gamma(\frac{d}{2}-1)}{2\pi^{\frac{d}{2}}|x-y|^{d-2}}, \quad d \geq 3.$$

In the case $d \leq 2$, we have $\int_0^\infty p_t(x,y)\,\mathrm{d}t \equiv \infty$. Hence we have to take another approach to define a useful Φ.

Definition 8.17. Let $(B_t)_{t\geq 0}$ be a d-dimensional Brownian motion. A *transient Brownian motion* is the process $(B_t)_{t\in[0,\tau]}$ in either of the following two cases:

(i) $d \geq 3$ and $\tau = \infty$;

(ii) $d \geq 2$ and τ is the first exit time from a bounded open domain $U \subset \mathbb{R}^d$.

We use the convention that $U = \mathbb{R}^d$ in case (i).

Proposition 8.18 (Transition subdensity)**.** *For a transient Brownian motion $(B_t)_{t\in[0,\tau]}$, there exist a family of transition (sub)densities $\mathfrak{p}_t^*(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+,\ 0 \leq t < \infty$ such that*

$$\mathbb{P}_x\left(B_t \in A \textit{ and } t \leq \tau\right) = \int_A \mathfrak{p}_t^*(x,y)\,\mathrm{d}y \quad \textit{for every Borel set } A \subset \mathbb{R}^d.$$

Moreover,

(i) for each $t \geq 0$, we have $\mathfrak{p}_t^*(x,y) = \mathfrak{p}_t^*(y,x)$ for almost every $x, y \in \mathbb{R}^d$;

(ii) if τ is the first exit time from a bounded open domain $U \subset \mathbb{R}^d$, then for each $t \geq 0$ and each $x \in U$, we have $\mathfrak{p}_t^*(x,y) = 0$ for almost every $y \notin U$.

Proof. We fix $t \geq 0$ throughout the proof. For the existence of the density, by the Radon-Nikodym theorem, it suffices to check that $\mathbb{P}_x(B_t \in A,\ t \leq \tau) = 0$ for every Borel set $A \subset \mathbb{R}^d$ of Lebesgue measure 0.

(i) If $d \geq 3$ and $\tau = \infty$, we can drop the requirement $t \leq \tau$ and choose the heat kernel $\mathfrak{p}_t^* = p_t$.

(ii) If $d \geq 2$ and τ is the first exit time from a bounded open domain $U \subset \mathbb{R}^d$, for each compact subset $K \subset U$, $x \in K$ and $n \in \mathbb{N}$, define

$$\mathfrak{p}_{t,K,n}^*(x,y) = \int_K \cdots \int_K \prod_{k=1}^{2^n} p_{t2^{-n}}(z_{k-1}, z_k)\,\mathrm{d}z_1 \cdots \mathrm{d}z_{2^n-1},$$

where $z_0 = x$ and $z_{2^n} = y$, and p is the transition density of d-dimensional Brownian motion. Then

$$\mathbb{P}_x\left(B_t \in A, \text{ and } B_{kt2^{-n}} \in K \text{ for all } k = 0, 1, \cdots, 2^n - 1\right) = \int_A \mathfrak{p}_{t,K,n}^*(x,y)\,\mathrm{d}y$$

for every Borel set $A \subset \mathbb{R}^d$. Since $p_{t,K,n}^*$ is decreasing in n, by the monotone convergence theorem,

$$\mathbb{P}_x\left(B_t \in A \text{ and } t \leq \tau_K\right) = \lim_{n\to\infty} \int_A \mathfrak{p}_{t,K,n}^*(x,y)\,\mathrm{d}y = \int_A \mathfrak{p}_{t,K}^*(x,y)\,\mathrm{d}y, \tag{8.12}$$

where $p^*_{t,K}(x,y) = \lim_{n\to\infty} \mathfrak{p}^*_{t,K,n}(x,y)$ and $\tau_K = \inf\{s \geq 0 : B_s \notin K\}$. The symmetry of p_t implies that $\mathfrak{p}^*_{t,K,n}$ and $\mathfrak{p}^*_{t,K}$ are both symmetric. We construct an increasing sequence $(K_m)_{m=1}^\infty$ of compact sets exhausting U, e.g. $K_m = \{x \in U : d(x, U^c) \geq \frac{1}{m}\}$. Then

$$\bigcup_{m=1}^\infty \{t \leq \tau_{K_m}\} = \bigcup_{m=1}^\infty \bigcap_{s<t} \{B_s \in K_m\} = \bigcap_{s<t} \bigcup_{m=1}^\infty \{B_s \in K_m\} = \bigcap_{s<t} \{B_s \in U\} = \{t \leq \tau\}.$$

Taking a monotone limit in (8.12) produces a symmetric version $\mathfrak{p}^*_t(x,y)$ of the transition density. For the second statement, by (8.12), we note that $\mathfrak{p}^*_{t,K}(x,\cdot) = 0$ a.e. on K^c for each compact $K \subset U$. Hence the monotone limit $\mathfrak{p}^*_t(x,\cdot) = 0$ a.e. on U^c. $\square$

Remark. Let $f \in C_c^\infty(\mathbb{R}^d)$ be a nonnegative function. Then for each $t > 0$ and $x \in U$,

$$\begin{aligned}
\int_{\mathbb{R}^d} \mathfrak{p}^*_t(x,y) f(y)\, \mathrm{d}y &= \mathbb{E}_x\left[f(B_t)\mathbb{1}_{\{t\leq\tau\}}\right] = \mathbb{E}_x\left[f(B_t)\left(1 - \mathbb{1}_{\{t>\tau\}}\right)\right] = \mathbb{E}_x\left[f(B_t)\right] - \mathbb{E}_x\left[\mathbb{1}_{\{t>\tau\}}\mathbb{E}_{B_\tau}[f(B_{t-\tau})]\right] \\
&= \int_{\mathbb{R}^d} p_t(x,y) f(y)\, \mathrm{d}y - \mathbb{E}_x\left[\mathbb{1}_{\{t>\tau\}} \int_{\mathbb{R}^d} p_{t-\tau}(B_\tau, y) f(y)\, \mathrm{d}y\right] \\
&= \int_{\mathbb{R}^d} \left[p_t(x,y) - \mathbb{E}_x\left[p_{t-\tau}(B_\tau, y)\mathbb{1}_{\{\tau<t\}}\right]\right] f(y)\, \mathrm{d}y.
\end{aligned}$$

Therefore, we can choose a version of density $\mathfrak{p}^*_t(x,\cdot) = p_t(x,\cdot) - \mathbb{E}_x\left[p_{t-\tau}(B_\tau,\cdot)\mathbb{1}_{\{\tau<t\}}\right]$. To summarize, we use the following typical version of transition subdensities in Proposition (8.18):

(i) If $d \geq 3$ and $\tau = \infty$, we have $\mathfrak{p}^*_t(x,y) = p_t(x,y)$.

(ii) If $d \geq 2$ and τ is the first exit time from a bounded open domain $U \subset \mathbb{R}^d$,

$$\mathfrak{p}^*_t(x,y) = p_t(x,y) - \mathbb{E}_x\left[p_{t-\tau}(B_\tau, y)\mathbb{1}_{\{\tau<t\}}\right]. \tag{8.13}$$

Definition 8.19 (Green's function)**.** For a transient Brownian motion $(B_t)_{t\in[0,\tau]}$ with transition (sub)densities $(\mathfrak{p}^*_t)$ as above, we define the *Green's function* $G : \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty]$ by

$$G(x,y) = \int_0^\infty \mathfrak{p}^*_t(x,y)\, \mathrm{d}t, \quad x, y \in \mathbb{R}^d.$$

Remark. By Proposition 8.18, if τ is the first exit time of a bounded open domain U, we can choose for each $x \in U$ a version of Green's function such that $G(x,\cdot) = 0$ on U^c.

In probabilistic terms, with $x \in U$ is fixed, the Green function $G(x,\cdot)$ is the density of the expected occupation measure for the transient Brownian motion $(B_t)_{0\leq t\leq\tau}$ started in x.

Proposition 8.20. *Let G be the Green function for a transient Brownian motion $(B_t)_{t\in[0,\tau]}$. For every meausrable functtion $f : \mathbb{R}^d \to [0,\infty]$ and $x \in \mathbb{R}^d$,*

$$\mathbb{E}_x\left[\int_0^\tau f(B_t)\, \mathrm{d}t\right] = \int_{\mathbb{R}^d} G(x,y) f(y)\, \mathrm{d}y.$$

Proof. If $f : \mathbb{R}^d \to [0,\infty]$ is measurable, Tonelli's theorem implies

$$\begin{aligned}
\mathbb{E}_x\left[\int_0^\tau f(B_t)\, \mathrm{d}t\right] &= \int_0^\infty \mathbb{E}_x\left[f(B_t)\mathbb{1}_{\{\tau\geq t\}}\right] \mathrm{d}t = \int_0^\infty \int_{\mathbb{R}^d} \mathfrak{p}^*_t(x,y) f(y)\, \mathrm{d}y\, \mathrm{d}t \\
&= \int_{\mathbb{R}^d} \left[\int_0^\infty \mathfrak{p}^*_t(x,y)\, \mathrm{d}t\right] f(y)\, \mathrm{d}y = \int_{\mathbb{R}^d} G(x,y) f(y)\, \mathrm{d}y.
\end{aligned}$$

Then we finish the proof. $\square$

Proposition 8.21. *Let $(B_t)_{0\le t\le\tau}$ be a d-dimensional transient Brownian motion.*

(i) If $d \ge 3$ and $\tau = \infty$, the Green's function is

$$G(x,y) = \Phi(x,y) = \frac{\Gamma(\frac{d}{2}-1)}{2\pi^{d/2}}|x-y|^{2-d}, \quad x,y \in \mathbb{R}^d.$$

(ii) If $d \ge 2$ and τ is the first exit time from a bounded open domain U, the Green's function is

$$G(x,y) = \Phi(x,y) - \mathbb{E}_x\left[\Phi(B_\tau, y)\right], \quad x,y \in \mathbb{R}^d, \tag{8.14}$$

where

$$\Phi(x,y) = \begin{cases} -\dfrac{1}{\pi}\log|x-y|, & d=2, \\ \dfrac{\Gamma(\frac{d}{2}-1)}{2\pi^{d/2}}|x-y|^{2-d}, & d \ge 3. \end{cases}$$

Remark. In fact, if $d \ge 3$, we have $|B_\infty| = \infty$ a.s., and the results in (i) and (ii) cocincide.

Proof. The result in (i) is already proved, so we focus on (ii). For every $x, y \in \mathbb{R}^d$, we take $(a_t)_{t\ge 0}$ such that

$$\int_0^\infty |p_t(x,y) - a_t|\,\mathrm{d}t < \infty.$$

Assume $\Phi(x,y) = \int_0^\infty (p_t(x,y) - a_t)\,\mathrm{d}t$. By (8.13),

$$\begin{aligned} G(x,y) &= \int_0^\infty \mathfrak{p}_t^*(x,y)\,\mathrm{d}t = \int_0^\infty (p_t(x,y)-a_t)\,\mathrm{d}t - \mathbb{E}_x\left[\int_0^\infty (p_{t-\tau}(B_\tau,y) - a_{t-\tau})\,\mathbb{1}_{\{\tau<t\}}\mathrm{d}t\right] \\ &= \int_0^\infty (p_t(x,y)-a_t)\,\mathrm{d}t - \mathbb{E}_x\left[\int_\tau^\infty (p_{t-\tau}(B_\tau,y)-a_{t-\tau})\,\mathrm{d}t\right] = \Phi(x,y) - \mathbb{E}_x\left[\Phi(B_\tau,y)\right]. \end{aligned}$$

If $d \ge 3$, we simply take $a_t \equiv 0$, and the result follows from (i). Otherwise, if $d = 2$, we let $a_t = \frac{1}{2\pi t}e^{-\frac{1}{2t}}$, so

$$\begin{aligned} \Phi(x,y) &= \int_0^\infty \frac{1}{2\pi t}\left(e^{-\frac{|x-y|^2}{2t}} - e^{-\frac{1}{2t}}\right)\mathrm{d}t = \frac{1}{2\pi}\int_0^\infty \frac{1}{t}\left(\int_{|x-y|^2/(2t)}^{1/(2t)} e^{-s}\,\mathrm{d}s\right)\mathrm{d}t \\ &= \frac{1}{2\pi}\int_0^\infty e^{-s}\left(\int_{|x-y|^2/(2s)}^{1/(2s)} \frac{\mathrm{d}t}{t}\right)\mathrm{d}s = -\frac{1}{\pi}\log|x-y|. \end{aligned}$$

Combining the last two displays completes the proof. ☐

Finally, we study some analytic properties of the Green's function.

Proposition 8.22. *Let $G : \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty]$ be the Green's function for a transient Brownian motion $(B_t)_{t\in[0,\tau]}$ in $U \subset \mathbb{R}^d$. Then*

(i) $G(x,y) < \infty$ for all $x \ne y$;

(ii) for each $y \in U$, the function $x \mapsto G(x,y)$ is harmonic on $U\backslash\{y\}$;

(iii) $G(x,y) = G(y,x)$ for all $x,y \in \mathbb{R}^d$.

Proof. These results are clear by the expression of G when $d \ge 3$ and $\tau = \infty$. Hence we focus on the case that $U \subset \mathbb{R}^d$ is a bounded open domain and $d \ge 2$. Since G vanishes outside $U \times U$, we may assume $x, y \in U$.

(i) For each $y \in U$, the function $\Phi(\cdot, y)$ is bounded on ∂U. Since $B_\tau \in \partial U$ a.s., $\mathbb{E}_x\left[\Phi(B_\tau,y)\right] < \infty$.

(ii) By Proposition 8.7, the function $h^y(x) = \mathbb{E}_x\left[\Phi(B_\tau,y)\right]$ is harmonic.

(iii) The symmetry of G follows from the almost-everywhere symmetry of $\mathfrak{p}_t^*$ together with the continuity. ☐

8.2.2 Poisson's Equation

In this part, we assume U is an open and bounded subset of $\mathbb{R}^d$, fix $f \in C_b(U)$, and study the boundary-value problem

$$\begin{cases} \Delta u = f & \text{in } U, \\ u = 0 & \text{on } \partial U. \end{cases} \tag{8.15}$$

Similar to our discussion in the remark after Definition 8.1, by Itô's formula, if $u \in C^2(U) \cap C(\overline{U})$ satisfies (8.15), the process

$$M_t = u(B_{t\wedge\tau}) - \frac{1}{2}\int_0^{t\wedge\tau} f(B_s)\,\mathrm{d}s, \quad t \geq 0$$

is a martingale under $\mathbb{P}_x$ for each $x \in U$, where $\tau = \inf\{t \geq 0 : B_t \notin U\}$.

Proposition 8.23. *Let $f : U \to \mathbb{R}$ be a bounded continuous function.*

(i) If there exists a bounded solution of (8.15)*, it must be*

$$u(x) = -\frac{1}{2}\mathbb{E}_x\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right], \quad x \in U. \tag{8.16}$$

(ii) If the above $u \in C^2(U)$, it satisfies $\Delta u = f$. In addition, if every $y \in \partial U$ satisfies the exterior cone condition [Definition 8.8], then u is a solution of (8.15)*.*

Proof. (i) Since U is bounded, we have $\mathbb{E}_x\tau < \infty$ for each $x \in U$. If both u and f are bounded continuous functions on U, then $|M_t| \leq \|u\|_\infty + \frac{1}{2}\tau\|f\|_\infty$. Since

$$\lim_{t\uparrow\infty} M_t = u(B_\tau) - \frac{1}{2}\int_0^\tau f(B_t)\,\mathrm{d}t = -\frac{1}{2}\int_0^\tau f(B_t)\,\mathrm{d}t,$$

by the dominated convergence theorem and martingale property,

$$u(x) = \mathbb{E}_x[M_0] = \lim_{t\uparrow\infty}\mathbb{E}_x[M_t] = -\frac{1}{2}\mathbb{E}_x\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right].$$

(ii) We let u be defined as in (8.16), fix $x \in U$ and $B(x,\epsilon) \subset U$. Using the strong Markov property at the stopping time $\tau_\epsilon = \inf\{t \geq 0 : B_t \notin B(x,\epsilon)\}$, we have

$$\mathbb{E}\left[\int_0^\tau f(B_s)\,\mathrm{d}s \,\middle|\, \mathscr{F}_{\tau_\epsilon}\right] = \int_0^{\tau_\epsilon} f(B_s)\,\mathrm{d}s + \mathbb{E}_{B_{\tau_\epsilon}}\left[\int_0^\tau f(B_s)\,\mathrm{d}s\right] = \int_0^{\tau_\epsilon} f(B_s)\,\mathrm{d}s - \frac{1}{2}u(B_{\tau_\epsilon}).$$

Therefore

$$\mathbb{E}_x\left[u(B_{\tau_\epsilon})\right] - u(x) = \frac{1}{2}\mathbb{E}_x\left[\int_0^{\tau_\epsilon} f(B_s)\,\mathrm{d}s\right] = \left(\frac{f(x)}{2} + o(1)\right)\mathbb{E}_x\tau_\epsilon.$$

On the other hand, by Taylor's theorem and the martingale property of $((B_t^j - x)^2 - t)_{t\geq 0}$ under $\mathbb{P}_x$,

$$\begin{aligned}\mathbb{E}_x\left[u(B_{\tau_\epsilon})\right] - u(x) &= \mathbb{E}_x\left[\nabla u(x)^\top(B_{\tau_\epsilon} - x) + \frac{1}{2}(B_{\tau_\epsilon} - x)^\top D^2u(x)(B_{\tau_\epsilon} - x)\right] + o(\epsilon^2) \\ &= \sum_{j=1}^2 \frac{\partial^2 u}{\partial x_j^2}(x)\,\mathbb{E}_x[(B_{\tau_\epsilon}^j - x)^2] + o(\epsilon^2) = \frac{\Delta u(x)}{2}\mathbb{E}_x\tau_\epsilon + o(\epsilon^2).\end{aligned}$$

Since $\mathbb{E}_x\tau_\epsilon = \frac{\epsilon^2}{d}$, the last two displays imply

$$|f(x) - \Delta u(x)| \leq \frac{o(\epsilon^2)}{\mathbb{E}_x\tau_\epsilon}.$$

Letting $\epsilon \downarrow 0$ gives $\Delta u(x) = f(x)$. Finally, to verify the boundary condition under the extorior cone condition, we fix $y \in \partial U$ and $U \ni x_n \to y$. Then

- by Lemma 8.9, $\mathbb{P}_{x_n}(\tau > \eta) = 0$ as $n \to \infty$ for every $\eta > 0$;
- since U is bounded, $\sup_{x \in U} \mathbb{E}_x \tau \le (\operatorname{diam} U)^2 < \infty$ and $\|u\|_\infty \le \frac{(\operatorname{diam} U)^2}{2} \|f\|_\infty < \infty$.

Then for any $\eta > 0$, by the simple Markov property at η,

$$\begin{aligned} |u(x_n)| &\le \frac{1}{2}\mathbb{E}_{x_n}\left[\int_0^{\tau\wedge\eta} |f(B_s)|\,\mathrm{d}s\right] + \frac{1}{2}\left|\mathbb{E}_{x_n}\left[\mathbb{1}_{\{\tau>\eta\}}\int_\eta^\tau f(B_s)\,\mathrm{d}s\right]\right| \\ &\le \frac{\eta}{2}\|f\|_\infty + \left|\mathbb{E}_{x_n}\left[\mathbb{1}_{\{\tau>\eta\}}u(B_\eta)\right]\right| \le \frac{\eta}{2}\|f\|_\infty + \|u\|_\infty \mathbb{P}_{x_n}(\tau>\eta) \to \frac{\eta}{2}\|f\|_\infty. \end{aligned}$$

Since $\eta > 0$ is arbitrary, we let $\eta \downarrow 0$ to conclude that $u(x) \to 0$ as $U \ni x \to y \in \partial U$. □

Poisson's equation and Green's function. To determine if (8.16) solves the boundary-value problem (8.15), it remains to study the differentiability of (8.16). By Proposition 8.20,

$$\mathbb{E}\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right] = \int_U G(x,y)f(y)\,\mathrm{d}y = \int_U \Phi(x,y)f(y)\,\mathrm{d}y - \mathbb{E}_x\left[\int_U \Phi(B_\tau,y)f(y)\,\mathrm{d}y\right].$$

For simplicity, we assume $f \in C^\infty(\overline{U})$. We write

$$w(x) = \int_U \Phi(x,y)f(y)\,\mathrm{d}y,$$

so $u(x) = w(x) - \mathbb{E}_x[w(B_\tau)]$. Some fundamental results about convolution imply that $w \in C^\infty(U)$, and by Proposition 8.7, the function $x \mapsto \mathbb{E}_x[w(B_\tau)]$ is also smooth. Therefore u is the solution of (8.15). Further studies show that one only requires f to be Hölder continuous to ensure that $w \in C^2(U)$.

Half-space. We let the dimension $d \ge 2$, and consider the Brownian motion $(B_t)_{0\le t\le\tau}$ in the upper half space $U = \{(x_1, \cdots, x_d) \in \mathbb{R}^d : x_d > 0\}$, where $\tau = \{t \ge 0 : B_t \notin U\}$. Let $\tilde{y} = (y_1, \cdots, y_{d-1}, -y_d)$ be the reflection of $y = (y_1, \cdots, y_d)$ through the hyperplane $\{x \in \mathbb{R}^d : x_d = 0\}$.

Theorem 8.24. *If $x \in U$ and $f \in C_c(U)$ is nonnegative, then*

$$\mathbb{E}_x\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right] = \int_U \Phi(x,y)f(y)\,\mathrm{d}y - \int_U \Phi(x,\tilde{y})f(y)\,\mathrm{d}y.$$

In other words, the Green's function for the transient Brownian motion $(B_t)_{0\le t\le\tau}$ is

$$G(x,y) = \Phi(x,y) - \Phi(x,\tilde{y}), \quad x, y \in U.$$

Proof. By the reflection principle, $B_t \overset{d}{=} \widetilde{B}_t$ on the event $\{\tau \ge t\}$. Since $\operatorname{supp} f \subset U$,

$$\begin{aligned} \mathbb{E}_x\left[f(B_t) - f(\widetilde{B}_t)\right] &= \mathbb{E}_x\left[\left(f(B_t) - f(\widetilde{B}_t)\right)\mathbb{1}_{\{\tau>t\}}\right] + \mathbb{E}_x\left[\left(f(B_t) - f(\widetilde{B}_t)\right)\mathbb{1}_{\{\tau\le t\}}\right] \\ &= \mathbb{E}_x\left[\left(f(B_t) - f(\widetilde{B}_t)\right)\mathbb{1}_{\{\tau>t\}}\right] = \mathbb{E}_x\left[f(B_t)\mathbb{1}_{\{\tau>t\}}\right]. \end{aligned}$$

Since $f \ge 0$, we apply Tonelli's theorem to obtain

$$\begin{aligned} \mathbb{E}_x\left[\int_0^\tau f(B_t)\mathrm{d}t\right] &= \int_0^\infty \mathbb{E}_x\left[f(B_t)\mathbb{1}_{\{\tau>t\}}\right]\mathrm{d}t = \int_0^\infty \int_U (p_t(x,y) - p_t(x,\tilde{y}))f(y)\,\mathrm{d}y\,\mathrm{d}t \\ &= \int_U \int_0^\infty (p_t(x,y) - p_t(x,\tilde{y}))f(y)\,\mathrm{d}t\,\mathrm{d}y = \int_U \Phi(x,y)f(y)\,\mathrm{d}y - \int_U \Phi(x,\tilde{y})f(y)\,\mathrm{d}y. \end{aligned}$$

Thus we finish the proof. □

Unit ball. We let the dimension $d \geq 2$, and consider the Brownian motion $(B_t)_{0\leq t\leq\tau}$ in the unit ball $B(0,1) = \{x \in \mathbb{R}^d : |x| < 1\}$, where $\tau = \{t \geq 0 : B_t \notin B(0,1)\}$.

Theorem 8.25. *If f is bounded and measurable then*

$$\mathbb{E}_x\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right] = \int_{B(0,1)} G(x,y)f(y)\,\mathrm{d}y,$$

where

$$G(x,y) = \Phi(x,y) - \int_{\partial B(0,1)} \frac{1-|x|^2}{|x-z|^d}\,\Phi(z,y)\,\mathrm{d}\Sigma(z) = \Phi(x,y) - \Phi\left(x|y|, \frac{y}{|y|}\right). \tag{8.17}$$

We use the continuous extension at $y = 0$.

Proof. By Theorem 8.14 and Fubini's theorem, for each $x, y \in B(0,1)$,

$$\begin{aligned}
\mathbb{E}_x\left[\int_0^\tau f(B_t)\,\mathrm{d}t\right] &= \int_{B(0,1)} \Phi(x,y)f(y)\,\mathrm{d}y - \mathbb{E}_x\left[\int_{B(0,1)} \Phi(B_\tau,y)f(y)\,\mathrm{d}y\right] \\
&= \int_{B(0,1)} \Phi(x,y)f(y)\,\mathrm{d}y - \int_{B(0,1)}\left[\int \frac{1-|x|^2}{|x-z|^d}\,\Phi(z,y)\,\mathrm{d}\Sigma(z)\right]f(y)\,\mathrm{d}y.
\end{aligned}$$

Then the first equality in (8.17) is valid. To show the second equality, ny Theorem 8.10, it suffices to show that the second term is harmonic in x on $B(0,1)$ and equals $\Phi(x,y)$ on the boundary $S^{d-1} = \partial B(0,1)$. Indeed, when $|x| = 1$,

$$\left|x|y| - \frac{y}{|y|}\right|^2 = |x|^2|y|^2 - 2x^\top y + 1 = |y|^2 - 2x^\top y + |x|^2 = |x-y|^2.$$

We fix $0 < |y| < 1$. Then the mapping

$$x \mapsto \Phi\left(x|y| - \frac{y}{|y|}\right) = \begin{cases} \Phi\left(x, \frac{y}{|y|^2}\right) - \frac{1}{\pi}\log|y|, & d = 2 \\ |y|^{2-d}\Phi\left(x, \frac{y}{|y|^2}\right), & d = 3 \end{cases}$$

is harmonic on $B(0,1)$. For the case $y = 0$, note that

$$\Phi(B_\tau, 0) = \mathbb{E}_x\left[\Phi(B_\tau, 0)\right] = \lim_{y\to 0}\Phi\left(x|y|, \frac{y}{|y|}\right)$$

is constant. Thus we finish the proof. $\square$

8.3 Planar Brownian Motions and Holomorphic Functions

In this subsection, we focus on the planar case $d = 2$ and study the relation between holomorphic functions and planar Brownian motions. We let $B = (B_t)_{t\geq 0}$ be a 2-dimensional Brownian motion, where it is helpful to identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$. We write

$$B_t = X_t + \mathrm{i}Y_t, \quad t \geq 0,$$

and say that B is a complex Brownian motion.

Let $U \subset \mathbb{R}^2$ be open. A function $\Phi : U \to \mathbb{C}$ is said to be *holomorphic* if it is complex differentiable in U. A holomorphic function $\Phi(x, y) = u(x, y) + \mathrm{i}v(x, y)$ satisfies the *Cauchy-Riemann equation*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial u}.$$

By the Cauchy-Riemann equation, both the real and imaginary parts of Φ are harmonic. Consequently, both $(u(B_t))_{t\geq 0}$ and $(v(B_t))_{t\geq 0}$ are continuous local martingales.

8.3.1 Conformal Martingales

We first introduce a special class of complex-valued local martingales.

Proposition 8.26. *Let $Z = X + \mathrm{i}Y$ be a continuous complex local martingale. Then there exists a unique continuous complex finite variation process $\langle Z, Z\rangle$ with $\langle Z, Z\rangle_0 = 0$, such that $Z^2 - \langle Z, Z\rangle$ is a complex local martingale. Furthermore, the following are equivalent:*

(i) Z^2 is a complex local martingale;

(ii) $\langle Z, Z\rangle = 0$;

(iii) $\langle X, X\rangle = \langle Y, Y\rangle$ and $\langle X, Y\rangle = 0$.

Remark. A complex local martingale $Z = (Z_t)_{t\geq 0}$ satisfying the equivalent properties of the above statement is called a *conformal local martingale.*

Proof. It suffices to define $\langle Z, Z\rangle$ by $\mathbb{C} \times \mathbb{C}$-linearity:

$$\langle X + \mathrm{i}Y, X + \mathrm{i}Y\rangle = \langle X, X\rangle - \langle Y, Y\rangle + 2\mathrm{i}\langle X, Y\rangle.$$

The uniqueness easily follows from Proposition 3.68. □

The conformal local martingales have some nice properties.

Proposition 8.27. *Let Z be a conformal local martingale, and $\Phi : \mathbb{C} \to \mathbb{C}$ be twice continuously differentiable function (as a function of two real variables). Then*

$$\Phi(Z_t) = \Phi(Z_0) + \int_0^t \frac{\partial \Phi}{\partial z}(Z_s)\,\mathrm{d}Z_s + \int_0^t \frac{\partial \Phi}{\partial \overline{z}}(Z_s)\,\mathrm{d}\overline{Z}_s + \frac{1}{4}\int_0^t \Delta\Phi(Z_s)\,\mathrm{d}\langle Z, \overline{Z}\rangle_s. \tag{8.18}$$

In particular, if Φ is harmonic, then $(\Phi(Z_t))_{t\geq 0}$ is a local martingale. In addition, if Φ is holomorphic, then

$$\Phi(Z_t) = \Phi(Z_0) + \int_0^t \Phi'(Z_s)\,\mathrm{d}Z_s. \tag{8.19}$$

Furthermore, if $\langle X, X\rangle_\infty = \infty$ a.s., there exists (possibly on an enlargement of the probability space) a complex Brownian motion $(\beta_t)_{t\geq 0}$ such that

$$Z_t = \beta_{\langle X,X\rangle_t}, \quad t \geq 0. \tag{8.20}$$

Proof. We write $\Phi = u + \mathrm{i}v$. By Itô's formula,

$$u(Z_t) = u(Z_0) + \int_0^t \frac{\partial u}{\partial x}(Z_s)\,\mathrm{d}X_s + \int_0^t \frac{\partial u}{\partial y}(Z_s)\,\mathrm{d}Y_s + \frac{1}{2}\left(\int_0^t \frac{\partial^2 u}{\partial x^2}(Z_s)\,\mathrm{d}\langle X, X\rangle_s + \int_0^t \frac{\partial^2 u}{\partial y^2}(Z_s)\,\mathrm{d}\langle Y, Y\rangle_s\right)$$

Since $\frac{\partial u}{\partial z} = \frac{1}{2}\frac{\partial u}{\partial x} - \frac{\mathrm{i}}{2}\frac{\partial u}{\partial y}$ and $\frac{\partial u}{\partial \bar{z}} = \frac{1}{2}\frac{\partial u}{\partial x} + \frac{\mathrm{i}}{2}\frac{\partial u}{\partial y}$, and $\langle X, X\rangle = \langle Y, Y\rangle = \frac{1}{2}\langle Z, \overline{Z}\rangle$, we have

$$\begin{aligned} u(Z_t) &= u(Z_0) + \int_0^t \frac{\partial u}{\partial x}(Z_s)\,\mathrm{d}\frac{Z_s + \overline{Z}_s}{2} + \int_0^t \frac{\partial u}{\partial y}(Z_s)\,\mathrm{d}\frac{Z_s - \overline{Z}_s}{2\mathrm{i}} + \frac{1}{4}\int_0^t \Delta u(Z_s)\,\mathrm{d}\langle Z, \overline{Z}\rangle_s \\ &= u(Z_0) + \int_0^t \frac{\partial u}{\partial z}(Z_s)\,\mathrm{d}Z_s + \int_0^t \frac{\partial u}{\partial \bar{z}}(Z_s)\,\mathrm{d}\overline{Z}_s + \frac{1}{4}\int_0^t \Delta u(Z_s)\,\mathrm{d}\langle Z, \overline{Z}\rangle_s. \end{aligned}$$

The same formula holds if we replace u with v. Hence (8.18) holds.

Furthremore, since $\langle X, Y\rangle = 0$ and $\langle X, X\rangle = \langle Y, Y\rangle$, we apply Theorem 5.14 to both X and Y and use Corollary 5.15 to conclude the existence of a complex Brownian motion β such that (8.20) holds. ☐

Finally, we study the conformal invariance property of complex Brownian motion, which asserts that the image of complex Brownian motion under a holomorphic function is a time-changed complex Brownian motion.

Theorem 8.28 (Conformal invariance)**.** *Let $\Phi : \mathbb{C} \to \mathbb{C}$ be a non-constant holomorphic function, and let $B = (B_t)_{t\geq 0}$ be a complex Brownian motion starting from $z \in \mathbb{C}$. There exists a complex Brownian motion $\beta = (\beta_t)_{t\geq 0}$ such that*

$$\Phi(B_t) = \beta_{\langle X,X\rangle_t} \quad \textit{for every } t \geq 0,\ \ P_z\textit{-a.s.},$$

where $X = \operatorname{Re}\Phi(B)$ and

$$\langle X, X\rangle_t = \int_0^t |\Phi'(B_s)|^2\,\mathrm{d}s, \quad t \geq 0.$$

Furthermore, the mapping $t \mapsto \langle X, X\rangle_t$ is strictly increasing.

Proof. If Φ is an entire function, so is Φ^2, and $\Phi(B)^2$ is a continuous local martingale. By Proposition 8.26, $\Phi(B)$ is a conformal local martingale. By (8.19),

$$\Phi(B_t) = \Phi(z) + \int_0^t \Phi'(B_s)\,\mathrm{d}B_s.$$

For $X = \operatorname{Re}\Phi(B)$, we have

$$\langle X, X\rangle_t = \int_0^t |\Phi'(B_s)|^2\,\mathrm{d}s, \quad t \geq 0.$$

Since Φ' is holomorphic and not identically zero, it has at most countably zeroes in $\mathbb{C}$, and

$$\mathbb{P}\left(\text{there exists } t > 0 \text{ such that } \Phi'(B_t) = 0\right) = 0.$$

Therefore $t \mapsto \langle X, X\rangle_t$ is strictly increasing.

Finally, following Proposition 8.27, it remains to show that $\langle X, X\rangle_\infty = \infty$ a.s.. If $\langle X, X\rangle_\infty < \infty$, by Corollary 3.73, the process X_t would a.s. converge as t tends to infinity. On the other hand, since Φ is a non-constant entire function, one can find two disjoint open sets $U_1, U_2 \subset \mathbb{C}$ with $\Phi(\overline{U}_1) \cap \Phi(\overline{U}_2) = \emptyset$. By recurrence of planar Brownian motion, both $\{t \geq 0 : B_t \in U_1\}$ and $\{t \geq 0 : B_t \in U_2\}$ are a.s. unbounded, and $\Phi(B_t)$ cannot has a limit as $t \to \infty$. Thus we conclude the proof. ☐

Remark. This conclusion remains true if Φ is a non-constant holomorphic function on $\mathbb{C}\backslash N$, where N is a set satisfying that B hits N with zero probability. For example, we consider the holomorphic function $\Phi(z) = \frac{1}{z}$ on $\mathbb{C}\backslash\{0\}$ and a complex Brownian motion B started from some $z_0 \neq 0$.

8.3.2 The Skew-product Representation

We next study the decomposition of the planar Brownian motion in polar coordinates.

Theorem 8.29. *Let $B = (B_t)_{t\geq 0}$ be a complex Brownian motion started from $z = re^{\mathrm{i}\theta} \in \mathbb{C}\backslash\{0\}$, where $r > 0$ and $\theta \in (-\pi, \pi]$. Then there is a planar Brownian motion (β, γ) started from $(\log r, \theta)$ under P_z, such that*

$$B_t = \exp\left(\beta_{H_t} + \mathrm{i}\gamma_{H_t}\right) \quad \textit{for every } t \geq 0,\ P_z\textit{-a.s.},$$

where

$$H_t = \int_0^t \frac{\mathrm{d}s}{|B_s|^2}, \quad t \geq 0.$$

Proof. By the scaling and the rotational invariance of the Brownian motion, we may assume $z = 1$, where $r = 1$ and $\theta = 0$. We let $W = (W^1, W^2)$ be a complex Brownian motion started from 0. By Theorem (8.28), there exists a complex Brownian motion Z such that

$$e^{W_t} = Z_{C_t}, \quad \text{where } C_t = \int_0^t e^{2W_s^1}\,\mathrm{d}s, \quad t \geq 0.$$

Consider the inverse function C_t of $\mathbb{R}_+ \to \mathbb{R}_+ : t \mapsto D_t$, which, by the formula for inverse functions, is

$$H_t = \int_0^t \exp\left(-2W^1_{H_s}\right)\mathrm{d}s = \int_0^t \frac{\mathrm{d}s}{|Z_s|^2}.$$

Therefore

$$Z_t = \exp\left(W_{H_t}\right) = \exp\left(W^1_{H_t} + \mathrm{i}W^2_{H_t}\right),$$

which is the desired result except we did not get it for B but for the complex Brownian motion Z introduced in the course of the argument. To complete the proof, we let

$$\beta_t = \log\left|B_{\inf\{s\geq 0: \int_0^s |B_r|^{-2}\,\mathrm{d}r > t\}}\right| \quad \text{and} \quad \gamma_t = \arg B_{\inf\{s\geq 0: \int_0^s |B_r|^{-2}\,\mathrm{d}r > t\}}, \quad t \geq 0.$$

Since β and γ are deterministic functions of B, there laws should be same if we replace B by another complex Brownian motion Z started from 1. Hence $(\beta, \gamma) \overset{d}{=} (W_1, W_2)$, and (β, γ) are the desired Brownian motions. ☐

Remark. We can write H_t as the inverse of its inverse, which is

$$H_t = \inf\left\{s \geq 0 : \int_0^s e^{2\beta_r}\,\mathrm{d}r > t\right\}.$$

Consequently,

$$\log|B_t| = \beta_{\inf\left\{s\geq 0: \int_0^s e^{2\beta_r}\,\mathrm{d}r > t\right\}}.$$

Therefore $|B_t|$ is completely determined by the Brownian motion β and independent of γ. Furthermore, the the smaller the modulus of B, the more rapidly the argument of B varies.

8.3.3 Asymptotic Laws of Planar Brownian Motions

In this section, we apply the skew-product decomposition to certain asymptotic results for planar Brownian motion. Let $B = (B_t)_{t\geq 0}$ be a planar Brownian motion that starts from $z \in \mathbb{C}\backslash\{0\}$. We write

$$\theta_t = \arg B_t = \gamma_{H_t},$$

which is the continuous determination of the argument of B_t.

Winding and Unwinding. The argument θ_t of B_t is a time-changed one-dimensional Brownian motion γ_{H_t}, where $H_t = \int_0^t |B_s|^{-2}\, \mathrm{d}s$. If $B_0 = z \neq 0$, we fix the open disc $D = B(z, \frac{|z|}{2})$. By Proposition 8.16, $(B_t)_{t\geq 0}$ a.s. stays in the disc D for arbitrarily long time, and $|B_s|^{-2}$ is bounded from below. Therefore $H_\infty = \infty$ a.s., and we have

$$\limsup_{t\to\infty} \theta_t = \infty, \qquad \liminf_{t\to\infty} \theta_t = -\infty, \quad \text{a.s..}$$

In other words, the planar Brownian motion winds itself arbitrarily large numbers of times around 0, then unwinds itself and does this infinitely often.

Next, we study the asymptotics of the random time change $(H_t)_{t\geq 0}$. We fix β to be the Brownian motion obtained in the skew-product representation of $(B_t)_{t\geq 0}$. For each $a > 0$, write $\beta_t^a = a^{-1}\beta_{a^2 t}$ for the time-scaling. For each such Brownian motion we look at the first hitting time of level b:

$$T_b^a = \inf\{t \geq 0 : \beta_t^a = b\}, \quad b \in \mathbb{R}.$$

Lemma 8.30. *For every $\epsilon > 0$ and $|z| = 1$, we have*

$$\lim_{t\uparrow\infty} \mathbb{P}_z\left(\left|\frac{4H_t}{(\log t)^2} - T_1^{\frac{1}{2}\log t}\right| > \epsilon\right) = 0.$$

Proof. Step I (Laplace's method). We show that, for every continuous function $f : [0,t] \to \mathbb{R}$,

$$\lim_{a\uparrow\infty} \frac{1}{a} \log \int_0^t e^{af(s)}\, \mathrm{d}s = \max_{0\leq s\leq t} f(s).$$

By replacing f by its maximum in the limit, we obtain the upper bound. For the lower bound, we assume $f(r) = \max_{0\leq s\leq t} f(s) = M$ and by continuity fix $\delta > 0$ such that $f(s) > M - \eta$ for all $|s - r| < \delta$. Then

$$\frac{1}{a}\log \int_0^t e^{af(s)}\, \mathrm{d}s \geq \frac{1}{a}\log \int_{0\vee(r-\delta)}^{t\wedge(r+\delta)} e^{a(M-\eta)}\, \mathrm{d}s = \frac{\log\delta + a(M-\eta)}{a} \to M - \eta \quad \text{as} \quad a \uparrow \infty.$$

Since $\eta > 0$ is arbitrary, we let $\eta \downarrow 0$ to obtain the opposite inequality.

Step II. Since $|z| = 1$, we have $\beta_0 = 0$ under $\mathbb{P}_z$. By scaling we have $T_{1+\epsilon}^{\frac{1}{2}\log t} - T_{1-\epsilon}^{\frac{1}{2}\log t} \overset{d}{=} T_{1+\epsilon}^1 - T_{1-\epsilon}^1$. Then, by the strong Markov property,

$$\lim_{\epsilon\downarrow 0} \mathbb{P}_z\left(T_{1+\epsilon}^{\frac{1}{2}\log t} - T_{1-\epsilon}^{\frac{1}{2}\log t} > \eta\right) = \lim_{\epsilon\downarrow 0} \mathbb{P}_z\left(T_{1+\epsilon}^1 - T_{1-\epsilon}^1 > \eta\right) = \lim_{\epsilon\downarrow 0} \mathbb{P}_z\left(T_{2\epsilon}^1 > \eta\right) = 0.$$

Since $T_{1+\epsilon}^{\frac{1}{2}\log t} < T_1^{\frac{1}{2}\log t} < T_{1-\epsilon}^{\frac{1}{2}\log t}$, it suffices to show that

$$\lim_{t\uparrow\infty} \mathbb{P}_z\left(\frac{4H_t}{(\log t)^2} > T_{1+\epsilon}^{\frac{1}{2}\log t}\right) = 0 \quad \text{and} \quad \lim_{t\uparrow\infty} \mathbb{P}_z\left(\frac{4H_t}{(\log t)^2} < T_{1-\epsilon}^{\frac{1}{2}\log t}\right) = 0.$$

We are going to prove the first result.

Step III. We write $a = \frac{1}{2}\log t$. Then

$$\left\{\frac{4H_t}{(\log t)^2} > T_{1+\epsilon}^{\frac{1}{2}\log t}\right\} = \left\{\int_0^{a^2 T_{1+\epsilon}^a} e^{2\beta_s}\, \mathrm{d}s < t\right\} = \left\{\frac{1}{2a}\log \int_0^{a^2 T_{1+\epsilon}^a} e^{2\beta_s}\, \mathrm{d}s < 1\right\}. \tag{8.21}$$

Note that

$$\frac{1}{2a}\log \int_0^{a^2 T_{1+\epsilon}^a} e^{2\beta_s}\, \mathrm{d}s = \frac{\log a}{a} + \frac{1}{2a}\log \int_0^{T_{1+\epsilon}^a} \exp\left(2a\beta_s^a\right)\, \mathrm{d}s \overset{d}{=} \frac{\log a}{a} + \frac{1}{2a}\log \int_0^{T_{1+\epsilon}^1} \exp\left(2a\beta_s\right)\, \mathrm{d}s.$$

By the Laplace's method (Step I), almost surely,

$$\lim_{a\uparrow\infty} \frac{1}{2a} \log \int_0^{T^1_{1+\epsilon}} \exp\left(2a\beta_s\right) \, \mathrm{d}s = \sup_{0\le s\le T^1_{1+\epsilon}} \beta_s = 1+\epsilon.$$

By (8.21), for each $\epsilon > 0$, we have

$$\begin{aligned}\lim_{t\uparrow\infty} \mathbb{P}_z \left(\frac{4H_t}{(\log t)^2} > T_{1+\epsilon}^{\frac{1}{2}\log t} \right) &= \lim_{a\uparrow\infty} \mathbb{P}_z \left(\frac{1}{2a} \log \int_0^{a^2 T^a_{1+\epsilon}} e^{2\beta_s} \, \mathrm{d}s < 1 \right) \\ &= \lim_{a\to\infty} \mathbb{P}_z \left(\left| \frac{\log a}{a} + \frac{1}{2a} \log \int_0^{T^1_{1+\epsilon}} \exp\left(2a\beta_s\right) \, \mathrm{d}s - (1+\epsilon) \right| > \epsilon \right) = 0.\end{aligned}$$

Step IV. In the same way one can show that

$$\lim_{t\uparrow\infty} \mathbb{P}_z \left(\frac{4H_t}{(\log t)^2} < T_{1-\epsilon}^{\frac{1}{2}\log t} \right) = 0,$$

which finishes the proof. □

Remark. Since the law of T^a_1 does not depend on the choice of a, this lemma implies that

$$\frac{4H_t}{(\log t)^2} \xrightarrow{d} T_1 := \{t > 0 : \beta_t = 1\}$$

as $t \uparrow \infty$. The density of the limiting distribution is given by Corollary 4.36.

Next, we study the typical size of the argument θ_t of a planar Brownian motion $(B_t)_{t\ge0}$ when t is large.

Theorem 8.31 (Spitzer's law)**.** *Let $(\theta_t)_{t\ge0}$ be the continuous determination of the argument of the complex Brownian motion B started from $z \in \mathbb{C}\backslash\{0\}$. Then for every $x \in \mathbb{R}$,*

$$\lim_{t\uparrow\infty} \mathbb{P}_z \left(\frac{2\theta_t}{\log t} \le x \right) \le \int_{-\infty}^x \frac{1}{\pi(1+y^2)} \, \mathrm{d}y.$$

In other words, as $t\uparrow\infty$, the law of $\frac{2\theta_t}{\log t}$ converges to a standard symmetric Cauchy distribution.

Proof. By scaling we may assume $|z| = 1$. Given $a > 0$, we define $\gamma^a_t = a^{-1}\gamma_{a^2t}$, $t \ge 0$. Then

$$a^{-1}\theta_t = a^{-1}\gamma_{H_t} = \gamma^a_{a^{-2}H_t}.$$

By Lemma 8.30, for $a = \frac{1}{2}\log t$, we have $a^{-2}H_t - T^a_1 \to 0$ in probability as $t\uparrow\infty$, and

$$\lim_{t\uparrow\infty} \mathbb{P}_z \left(\left| \frac{2\theta_t}{\log t} - \gamma^a_{T^a_1} \right| > \epsilon \right) = 0, \quad \text{for all } \epsilon > 0.$$

Since β and γ are independent, the law of $\gamma^a_{T^a_1}$ does not depend on the choice of $a > 0$. Hence $\frac{2\theta_t}{\log t}$ converges to γ_{T_1} in distribution, where $T_1 = \{t \ge 0 : \beta_t = 1\}$. By Proposition 4.38, the characteristic function of γ_{T_1} is

$$\mathbb{E}\left[e^{\mathrm{i}\lambda\gamma_{T_1}}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\mathrm{i}\lambda\gamma_{T_1}} \,|\, T_1\right]\right] = \mathbb{E}\left[e^{-\frac{1}{2}\lambda^2 T_1}\right] = e^{-|\lambda|},$$

which is the characteristic function of the standard Cauchy distribution. □

Next we study the law of minimum modulus. We know that a planar Brownian motion stated from $z \ne 0$ hits the origin with probability zero, but $\inf_{0\le s\le t} |B_s| \to 0$ as $t\uparrow\infty$ by the neighborhood recurrence property.

Proposition 8.32. *Let $B = (B_t)_{t\geq 0}$ be a planar Brownian motion started from $z \neq 0$. Then for every $b > 0$,*

$$\lim_{t\uparrow\infty} \mathbb{P}_z\left(\min_{0\leq s\leq t} |B_s| \leq t^{-b}\right) \leq \frac{1}{1+2b}.$$

Proof. We may also take $|z| = 1$ by scaling. Then

$$\log\left(\min_{0\leq s\leq t} |B_s|\right) = \min_{0\leq s\leq t} \beta_{H_s} = \min_{0\leq s\leq H_t} \beta_s.$$

Let $a = \frac{1}{2}\log t$. Then

$$\frac{2}{\log t}\log\left(\min_{0\leq s\leq t} |B_s|\right) = \frac{1}{a}\min_{0\leq s\leq H_t} \beta_s = \min_{0\leq s\leq a^{-2}H_t} \beta_s^a.$$

By Lemma 8.30,

$$\min_{0\leq s\leq a^{-2}H_t} \beta_s^a - \min_{0\leq s\leq T_1^a} \beta_s^a \to 0 \quad \text{in probability as } t\uparrow\infty.$$

Since the law of $\min_{0\leq s\leq T_1^a} \beta_s^a$ does not depend on a,

$$\frac{2}{\log t}\log\left(\min_{0\leq s\leq t} |B_s|\right) \to \min_{0\leq s\leq T_1} \beta_s \quad \text{in distribution.}$$

To get the desired result, we fix $b > 0$, define $T_{-2b} = \inf\{t \geq 0 : \beta_t = -2b\}$ and note that

$$\mathbb{P}_z\left(\min_{0\leq s\leq T_1} \beta_s \leq -2b\right) = \mathbb{P}_z\left(T_{-2b} < T_1\right) = \frac{1}{1+2b}.$$

Then combining the last two displays completes the proof. ☐

Finally, we introduce the Kallianpur-Robbins law for the time spent by Brownian motion in a disc.

Theorem 8.33 (Kallianpur-Robbins)**.** *Let $B = (B_t)_{t\geq 0}$ be a complex Brownian motion started from $z \in \mathbb{C}$. Then for any $R > 0$,*

$$\frac{2}{\log t}\int_0^t \mathbb{1}_{\{|B_s|<R\}}\,\mathrm{d}s$$

converges in distribution as $t\uparrow\infty$ to an exponential distribution with mean R^2.

Proof. We fix $t > 0$ and let $a = \frac{1}{2}\log t$. Then

$$\begin{aligned}\frac{2}{\log t}\int_0^t \mathbb{1}_{\{|B_s|<R\}}\mathrm{d}s &= \frac{1}{a}\int_0^t \mathbb{1}_{\{\beta_{H_s}<\log R\}}\,\mathrm{d}s = \frac{1}{a}\int_0^{H_t} \mathbb{1}_{\{\beta_s<\log R\}}e^{2\beta_s}\,\mathrm{d}s \\ &= a\int_0^{a^{-2}H_t} \mathbb{1}_{\{\beta_s^a<a^{-1}\log R\}}e^{2a\beta_s^a}\,\mathrm{d}s = a\int_{-\infty}^{a^{-1}\log R} e^{2ax}L^x_{a^{-2}H_t}(\beta^a)\,\mathrm{d}x = \int_0^R rL^{a^{-1}\log r}_{a^{-2}H_t}(\beta^a)\,\mathrm{d}r, \end{aligned} \tag{8.22}$$

where the last second equality follows from the occupation time formula [Proposition 7.8], and we apply the change of variable $r = e^{ax}$ in the last one. As $t\uparrow\infty$, we have $a^{-1}\log r \downarrow 0$ for every $r > 0$, and $a^{-2}H_t - T_1^a \to 0$ in probability by Lemma 8.30. From the joint continuity of Brownian local times [Theorem 7.12], for every $\epsilon \in (0, R)$, as $t\uparrow\infty$,

$$\sup_{\epsilon\leq r\leq R}\left|L^{a^{-1}\log r}_{a^{-2}H_t}(\beta^a) - L^0_{T_1^a}(\beta^a)\right| \to 0 \quad \text{in probability.}$$

By (8.22), we have

$$\left|\frac{2}{\log t}\int_0^t \mathbb{1}_{\{|B_s|<R\}}\mathrm{d}s - \frac{R^2}{2}L^0_{T_1^a}(\beta^a)\right| \to 0 \quad \text{in probability.}$$

Since $L^0_{T_1^a}(\beta^a) \overset{d}{=} L^0_{T^1}(\beta)$ for each $a > 0$, the desired result follows from Proposition 7.16. ☐

9 General Random Walks

9.1 Donsker's Invariance Principle

In this section, we discuss the approximation of general random walks by Brownian motions. What we are interested in is the behavior of the partial sum sequence

$$S_n = X_1 + X_2 + \cdots + X_n,$$

where $X_1, X_2, \cdots \overset{\text{i.i.d.}}{\sim} F$ are square-integrable i.i.d. random variables. One important conclusion we will use is the Skorokhod's embedding theorem, which asserts that any L^2 random variable can be viewed as the Brownian motion evaluated at an appropriate stopping time.

Theorem 9.1 (Skorokhod's embedding theorem). *Let X be a random variable with $\mathbb{E}X = 0$ and $\mathbb{E}[X^2] < \infty$, and $(B_t)_{t\geq 0}$ a Brownian motion starting at $B_0 = 0$. Then there exists stopping time τ for the Brownian motion such that $B_\tau \overset{d}{=} X$ and $\mathbb{E}[\tau] = \mathbb{E}[X^2]$.*

Proof. We first assume X is supported on a two point set $\{a, b\}$, where $a < 0 < b$. Then

$$\mathbb{E}[X^2] = a^2\mathbb{P}(X = a) + b^2\mathbb{P}(X = b) = -ab.$$

We choose $\tau = \tau_{a,b} = \inf\{t > 0 : B_t \in \{a, b\}\}$, which is the exit time from $[a, b]$. By Proposition 4.39, we have $B_\tau \overset{d}{=} X$ and $\mathbb{E}[\tau] = -ab = \mathbb{E}[X^2]$. To handle the general case, we approximate the distribution of X by a mixture of two-point distributions. Here we use the binary splitting martingale.

Definition 9.2 (Binary splitting martingale). A discrete-time martingale $(X_n)_{n=0}^\infty$ is said to be *binary splitting* if, whenever the event

$$A(x_0, x_1, \cdots, x_n) = \{X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n\}$$

for some $x_0, x_1, \cdots, x_n \in \mathbb{R}$ has positive probability, the random variable X_{n+1} conditioned on the event $A(x_0, x_1, \cdots, x_n)$ is supported on at most two values.

One can approximate a square-integrable random variable with a binary splitting martingale.

Lemma 9.3 (Dubins' embedding theorem). *Let X be a random variable with $\mathbb{E}[X^2] < \infty$. Then there exists a binary splitting martingale $(X_n)_{n=1}^\infty$ such that $X_n \to X$ a.s. and in L^2.*

Proof. Let $X_0 = \mathbb{E}X$, and define, iteratively, for all $n \in \mathbb{N}$ that

$$\xi_n = \mathbb{1}_{\{X\geq X_n\}} - \mathbb{1}_{\{X<X_n\}}, \qquad \mathscr{G}_n = \sigma(\xi_0, \xi_1, \cdots, \xi_{n-1}), \quad \text{and} \quad X_n = \mathbb{E}[X|\mathscr{G}_n].$$

Then $(X_n)_{n=0}^\infty$ is a binary splitting martingale. Since $\sup_{n\in\mathbb{N}} \mathbb{E}[X_n^2] \leq \mathbb{E}[X^2] < \infty$, by martingale convergence theorems [Theorem 3.32 and 3.36], X_n converges a.s. and in L^2 to $X_\infty := \mathbb{E}[X|\mathscr{G}_\infty]$, where $\mathscr{G}_\infty = \sigma\left(\bigcup_{n=0}^\infty \mathscr{G}_n\right)$.

Now it remains to show that $X_\infty = X$. We claim that

$$|X - X_\infty| = \lim_{n\to\infty} \xi_n(X - X_{n+1}), \quad a.s.. \tag{9.1}$$

In fact, if $X(\omega) = X_\infty(\omega)$ then both sides are 0. If $X(\omega) < X_\infty(\omega)$, there exists $N(\omega) > 0$ so that $\xi_n(\omega) = -1$ for all $n \geq N(\omega)$. The case $X(\omega) > X_\infty(\omega)$ is symmetric. Then

$$\mathbb{E}\left[\xi_n(X - X_{n+1})\right] = \mathbb{E}\left[\xi_n\mathbb{E}\left[X - X_{n+1}|\mathscr{G}_n\right]\right] = 0.$$

Since the collection $(|X - X_{n+1}|)_{n\in\mathbb{N}}$ is uniformly integrable, (9.1) implies $\mathbb{E}|X - X_\infty| = 0$. $\square$

Proof of Theorem 9.1 (Continued). We take the binary splitting martingale constructed in Lemma 9.3. Then X_1 is supported on the two points $a^1 = \mathbb{E}[X\mathbb{1}_{\{X<0\}}]$ and $b^1 = \mathbb{E}[X\mathbb{1}_{\{X\geq 0\}}]$. We take the exit time $\tau_1 = \inf\{t \geq 0 : B_t \notin (a^1, b^1)\}$, then $B_{\tau_1} \overset{d}{=} X_1$, and $\mathbb{E}[\tau_1] = \mathbb{E}[X_1^2]$.

Conditional on $B_{\tau_1} = a^1$, we take $a_1^2 = \mathbb{E}[X\mathbb{1}_{\{X<a^1\}}]$ and $b_1^2 = \mathbb{E}[X\mathbb{1}_{\{a^1\leq X<0\}}]$, so X_2 is supported on the two-point set $\{a_1^2, b_1^2\}$. On the event $\{B_{\tau_1} = a^1\}$, we may take $\tau_2 = \inf\left\{t \geq \tau_1 : B_t \notin (a_1^2, b_1^2)\right\}$. Then $(B_{\tau_2}|B_{\tau_1} = a^1) \overset{d}{=} (X_2|X_1 = a^1)$, and $\mathbb{E}[\tau_2\mathbb{1}_{\{B_{\tau_1}=a^1\}}] = \mathbb{E}[X_2^2\mathbb{1}_{\{X_1=a^1\}}]$. Similarly, conditional on $B_{\tau_1} = b^1$, we take $\tau_2 = \inf\left\{t \geq \tau_1 : B_t \notin (a_2^2, b_2^2)\right\}$, where $a_2^2 = \mathbb{E}[X\mathbb{1}_{\{0\leq X<b^1\}}]$ and $b_2^2 = \mathbb{E}[X\mathbb{1}_{\{X\geq b^1\}}]$. Then similar properties hold on the event $\{B_{\tau_1} = a^1\}$, and we conclude that $B_{\tau_2} \overset{d}{=} X_2$, and $\mathbb{E}[\tau_2] = \mathbb{E}[X_2^2]$.

Repeating this approach, we can find an increasing sequence of stopping times $\tau_1 \leq \tau_2 \leq \cdots$ such that $B_{\tau_n} \overset{d}{=} X_n$ and $\mathbb{E}[\tau_n] = \mathbb{E}[X_n^2]$ for all $n \in \mathbb{N}$. In fact, B_{τ_n} determines which of the 2^n regions of the real line the limit $\lim_{m\to\infty} B_{\tau_m}$ should lie in. By Proposition 3.12, $\tau_n \uparrow \tau$ a.s. for some stopping time τ. Furthermore, by monotone convergence theorem and Lemma 9.1,

$$\mathbb{E}[\tau] = \lim_{n\to\infty} \mathbb{E}[\tau_n] = \lim_{n\to\infty} \mathbb{E}[X_n^2] = \mathbb{E}[X^2].$$

Since B_{τ_n} converges in distribution to X and a.s. to B_τ by continuity, we have $B_\tau \overset{d}{=} X$. $\square$

Skorokhod's second embedding theorem concerns about extracting a random walk from a Brownian motion.

Theorem 9.4 (Skorokhod's embedding theorem)**.** *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting at $B_0 = 0$. Let $X_1, X_2, \cdots$ be an i.i.d. sequence with a distribution F, which has mean 0 and finite variance. Define*

$$S_n = X_1 + \cdots + X_n, \quad n = 0, 1, 2, \cdots.$$

Then there exists an increasing sequence of stopping times $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$ for $(B_t)_{t\geq 0}$ satisfying

(i) $B_{\tau_n} \overset{d}{=} S_n$ for all $n = 0, 1, 2, \cdots$, and

(ii) The increments $(\tau_n - \tau_{n-1})_{n=1}^\infty$ are i.i.d..

Proof. Following Theorem 9.1, let τ_1 be a stopping time with $B_{\tau_1} \overset{d}{=} X_1$ and $\mathbb{E}[\tau_1] = \mathbb{E}[X_1^2]$. By the strong Markov property, $B_{\tau_1+\cdot} = (B_{\tau_1+t} - B_{\tau_1})_{t\geq 0}$ is a Brownian motion independent of $\mathscr{F}_{\tau_1}$.

With τ_{n-1} fixed, we follow the same approach on the Brownina motion $B_{\tau_{n-1}+\cdot}$ and take τ_n be a stopping time such that $B_{\tau_n} - B_{\tau_{n-1}} \overset{d}{=} X_n$ and $\mathbb{E}[\tau_n - \tau_{n-1}] = \mathbb{E}[X_n^2]$, and $\tau_n - \tau_{n-1} \overset{d}{=} \tau_1$. The increment $\tau_n - \tau_{n-1}$ independent of $\mathscr{F}_{\tau_{n-1}}$. Thus we find the desired sequence (τ_n). $\square$

Corollary 9.5 (Central limit theorem)**.** *If $(X_n)_{n=1}^\infty$ is an i.i.d. sequence with mean 0 and variance 1, then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j \overset{d}{\to} N(0,1).$$

Proof. By Brownian scaling, for each $n \in \mathbb{N}$, we define a Brownian motion by $W_n(t) = \frac{B_{nt}}{\sqrt{n}} \overset{d}{=} B_t$. Then

$$\frac{S_n}{\sqrt{n}} \overset{d}{=} \frac{B_{\tau_n}}{\sqrt{n}} = W_n\left(\frac{\tau_n}{n}\right).$$

By the weak law of large numbers, $\frac{\tau_n}{n} \to \mathbb{E}[X_1^2] = 1$ in probability. For any $\epsilon > 0$, we pick $\delta > 0$ such that

$$\mathbb{P}\left(\exists t \in (1-\delta, 1+\delta) \text{ such that } |B_t - B_1| > \epsilon\right) < \frac{\epsilon}{2}.$$

Next, we take $N > 0$ great enough so that $\mathbb{P}\left(\left|\frac{\tau_n}{n} - 1\right| \geq \delta\right) < \frac{\epsilon}{2}$. Then $\mathbb{P}\left(\left|W_n\left(\frac{\tau_n}{n}\right) - W_n(1)\right| > \epsilon\right) < \epsilon$ for all $n \geq N$. Since $\epsilon > 0$ is arbitrary, $W_n\left(\frac{\tau_n}{n}\right) - W_n(1) \overset{\mathbb{P}}{\to} 0$, and $W_n\left(\frac{\tau_n}{n}\right) \to N(0,1)$ by Slutsky's lemma. $\square$

Next, we will derive a functional central limit theorem for general random walks. For an i.i.d. sequence $X_1, X_2, \cdots$ with mean 0 and variance 1, we consider the random walk $S_n = X_1 + X_2 + \cdots + X_n$ as a continuous time process by defining

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) S_{\lfloor t \rfloor + 1}, \quad t \geq 0.$$

In other words, we fix $S(n) = S_n$ at all nonnegative integer points $n \in \mathbb{N}$ and take $S(\cdot)$ to be the linear interpolation on each interval $[n, n+1]$.

Theorem 9.6 (Donsker's invariance principle)**.** *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean 0 and variance 1, and $S_n = X_1 + X_2 \cdots + X_n$. Define*

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) S_{\lfloor t \rfloor + 1}, \quad t \geq 0.$$

Then on the space $C([0,1])$ equipped with the uniform topology,

$$\left(\frac{S(nt)}{\sqrt{n}} \right)_{t \in [0,1]} \to (B_t)_{t \in [0,1]} \quad \textit{weakly}$$

as $n \to \infty$. Essentially, for each bounded contionuous function ψ on $C([0,1])$,

$$\mathbb{E}\psi\left(\frac{S(n\cdot)}{\sqrt{n}} \right) \to \mathbb{E}\psi(B). \tag{9.2}$$

Proof. For each $n \in \mathbb{N}$ and $1 \leq m \leq n$, we take $X_{n,m} \stackrel{d}{=} X_m/\sqrt{n}$ and $S_{n,m} = X_{n,1} + \cdots + X_{n,m} \stackrel{d}{=} S_m/\sqrt{n}$. By Skorokhod embedding theorem, we take stopping times $\tau_1^n, \cdots, \tau_n^n$ such that $S_{n,m} = B_{\tau_m^n}$. In fact, we at first fix $(B_t)_{t \geq 0}$ to be a Brownian motion independent of n, and then define the triangular random variable array $(X_{n,m})_{n \in \mathbb{N}, 1 \leq m \leq n}$ in the same probability space.

Let $\tau_1, \tau_2, \cdots$ be the stopping times constructed in Theorem 9.4. By the scaling invariance of Brownian motions, $B_{nt}/\sqrt{n}$ is also a Brownian motion, and $\tau_m^n \stackrel{d}{=} \tau_m/n$. Hence for each $s \in [0,1]$, by the weak law of large numbers, we have $\tau_{\lfloor ns \rfloor}^n \to s$ in distribution, and also in probability.

Lemma 9.7. *If $\tau_{\lfloor ns \rfloor}^n \to s$ in probability for each $s \in [0,1]$, then*

$$\sup_{t \in [0,1]} |S_n^*(nt) - B_t| \to 0 \quad \textit{in probability}$$

as $n \to \infty$, where $S_n^*(t) = S_{n,\lfloor t \rfloor} + (t - \lfloor t \rfloor) S_{n, \lfloor t \rfloor + 1} \stackrel{d}{=} S(t)/\sqrt{n}$.

Proof. Since $t \mapsto B_t$ is uniformly continuous on the compact interval $[0,1]$ (in fact, Hölder continuous with exponent $0 < \gamma < \frac{1}{2}$), for any $\epsilon > 0$, there exists $\delta > 0$ such that $1/\delta \in \mathbb{N}$ and

$$\mathbb{P}\left(|B_t - B_s| < \epsilon \text{ for all } t, s \in [0,1] \text{ and } |t - s| < 2\delta \right) > 1 - \epsilon. \tag{9.3}$$

Note that $\tau_{\lfloor ns \rfloor}^n \to s$ in probability. We take $N_\delta \geq 1/\delta$ such that for all $n \geq N_\delta$,

$$\mathbb{P}\left(\left| \tau_{\lfloor nk\delta \rfloor}^n - k\delta \right| < \delta \text{ for all } k = 1, 2, \cdots, 1/\delta \right) \geq 1 - \epsilon. \tag{9.4}$$

Since $m \mapsto \tau_m^n$ is increasing, for $s \in (k\delta - \delta, k\delta)$, we have

$$\tau_{\lfloor n(k-1)\delta \rfloor}^n - k\delta \leq \tau_{\lfloor ns \rfloor}^n - s \leq \tau_{\lfloor nk\delta \rfloor}^n - (k-1)\delta.$$

Under the event in (9.4), the above bound is controlled by $\pm 2\delta$. Hence for all $n \geq N_\delta$,

$$\mathbb{P}\left(\left|\tau^n_{\lfloor ns \rfloor} - s\right| < 2\delta \text{ for all } s \in [0,1]\right) \geq 1 - \epsilon. \tag{9.5}$$

On the intersection of the two events (9.3) and (9.5), we have

$$\left|S_{n,m} - B_{\frac{m}{n}}\right| = \left|B_{\tau^m_n} - B_{\frac{m}{n}}\right| < \epsilon, \quad \text{for all } 1 \leq m \leq n.$$

To deal with the interpolation, we let $t = \frac{m+\theta}{n}$, where $m \in \mathbb{N}$ and $\theta \in (0,1)$. By the triangle inequality,

$$|S^*_n(nt) - B_t| \leq (1-\theta)|S_{n,m} - B_{\frac{m}{n}}| + \theta|S_{n,m+1} - B_{\frac{m+1}{n}}| + (1-\theta)|B_t - B_{\frac{m}{n}}| + \theta|B_t - B_{\frac{m+1}{n}}| < 2\epsilon$$

for all $n \geq N_\delta$. Since ϵ is arbitrary, we conclude the proof. □

Proof of Theorem 9.6. We take ψ to be a bounded continuous function on $C([0,1])$. It remains to show that

$$\mathbb{E}\psi(S^*_n(n\cdot)) - \mathbb{E}\psi(B) \to 0.$$

We fix $\epsilon > 0$, and let $G_{\epsilon,\delta} = \{\omega \in C([0,1]) : \text{ if } \|\omega - \widetilde{\omega}\|_\infty < \delta \text{ then } |\psi(\omega) - \psi(\widetilde{\omega})| < \epsilon\}$ for $\delta > 0$. Since ψ is continuous, $G_{\epsilon,\delta} \uparrow C([0,1])$ as $\delta \downarrow 0$. Let $R_n = \|S^*_n(n\cdot) - B\|_\infty$. Then

$$\begin{aligned} |\mathbb{E}\psi(S^*_n(n\cdot)) - \mathbb{E}\psi(B)| &\leq \mathbb{E}\left[|\psi(S^*_n(n\cdot)) - \psi(B)|\mathbb{1}_{G_{\epsilon,\delta}\cap\{R_n<\delta\}}\right] + \mathbb{E}\left[|\psi(S^*_n(n\cdot)) - \psi(B)|\mathbb{1}_{G^c_{\epsilon,\delta}\cup\{R_n\geq\delta\}}\right] \\ &\leq \epsilon + 2\|\psi\|_\infty\left(\mathbb{P}(G^c_{\epsilon,\delta}) + \mathbb{P}\left(R_n \geq \delta\right)\right). \end{aligned}$$

We can bound $|\mathbb{E}\psi(S^*_n(n\cdot)) - \mathbb{E}\psi(B)|$ by 2ϵ by choosing small enough δ and then large enough n. Since $\epsilon > 0$ is arbitrary, we complete the proof of Theorem 9.6. □

Remark. (I) For any $M \in \mathbb{N}$, by considering a similar triangular array $(X_{n,m})_{n\in\mathbb{N},1\leq m\leq nM}$, we can conclude

$$\left(\frac{S(nt)}{\sqrt{n}}\right)_{t\in[0,M]} \to (B_t)_{t\in[0,M]} \quad \text{weakly}$$

on the space $C([0,M])$ equipped with the uniform topology.

(II) According to our proof, the identity (9.2) remains valid if ψ is bounded and W-a.s. continuous on $C([0,1])$, where W is the Wiener measure on $C([0,1])$.

With a subtle remark on the topology of the continous function space, we can extend this result to $[0,\infty)$.

Theorem 9.8 (Donsker's invariance principle)**.** *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean* 0 *and variance* 1*, and $S_n = X_1 + X_2 \cdots + X_n$. Define*

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)S_{\lfloor t \rfloor + 1}, \quad t \geq 0.$$

Then on the space $C([0,\infty))$ equipped with the ***compact convergence topology****, as $n \to \infty$,*

$$\left(\frac{S(nt)}{\sqrt{n}}\right)_{t\geq 0} \to (B_t)_{t\geq 0} \quad \textit{weakly.}$$

Proof. The compact convergence topology on $C([0,\infty))$ is induced by the metric

$$d(\omega, \widetilde{\omega}) = \sum_{M=1}^{\infty} 2^{-M} \frac{\sup_{t\in[0,M]} |\omega(t) - \widetilde{\omega}(t)|}{1 + \sup_{t\in[0,M]} |\omega(t) - \widetilde{\omega}(t)|}, \quad \omega, \widetilde{\omega} \in C((0,\infty]).$$

We claim that $d\big(\frac{S_n^*(n\cdot)}{\sqrt{n}}, B\big) \to 0$ in probability. In fact, by Lemma 9.7, for all $M > 0$,

$$\sup_{t\in[0,M]} \left| \frac{S_n^*(nt)}{\sqrt{n}} - B_t \right| \to 0 \quad \text{in probability.}$$

Given any $\epsilon > 0$, we choose $M_\epsilon > 0$ such that $\sum_{M=M_\epsilon+1}^{\infty} 2^{-M} < \epsilon$, and $N_\epsilon > 0$ such that if $n \geq N_\epsilon$,

$$\mathbb{P}\left(\sup_{t\in[0,M]} \left| \frac{S_n^*(nt)}{\sqrt{n}} - B_t \right| > \epsilon \right) < \frac{\epsilon}{M_\epsilon}, \quad \text{for } M = 1, 2, \cdots, M_\epsilon.$$

We take a union bound over all $1 \leq M \leq M_\epsilon$ to conclude that, with probability at least $1 - \epsilon$,

$$d\left(\frac{S_n^*(n\cdot)}{\sqrt{n}}, B \right) \leq \sum_{M=1}^{M_\epsilon} 2^{-M} \frac{\sup_{t\in[0,M]} \left| \frac{S_n^*(nt)}{\sqrt{n}} - B_t \right|}{1 + \sup_{t\in[0,M]} \left| \frac{S_n^*(nt)}{\sqrt{n}} - B_t \right|} + \sum_{M=M_\epsilon+1}^{\infty} 2^{-M} < 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, our claim is true.

Let ψ be a bounded continuous function on the space $C((0,\infty])$ with the compact convergence topology. Then the final part of our proof of Theorem 9.6 can be adapted by replacing

$$G_{\epsilon,\delta} = \{\omega \in C((0,\infty]) : \text{ if } d(\omega, \widetilde{\omega}) < \delta \text{ then } |\psi(\omega) - \psi(\widetilde{\omega})| < \epsilon\}$$

and $R_n = d\big(\frac{S_n^*(n\cdot)}{\sqrt{n}}, B\big)$. Thus we complete our proof. □

9.1.1 The Law of Iterated Logarithm

We can generalize the *Law of Iterated Logarithm* [Theorem 4.30] to random walks.

Theorem 9.9 (Hartman-Wintner). *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean 0 and variance 1, and $S_n = X_1 + X_2 \cdots + X_n$. Then*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \quad a.s..$$

Proof. As in the proof of Donsker's invariance principle, we choose stopping times $\tau_1, \tau_2, \cdots$ with $S_n = B_{\tau_n}$ and $\tau_n/n \to 1$ a.s.. Following Theorem 4.30, it suffices to show that

$$\limsup_{t\uparrow\infty} \frac{S_{\lfloor t\rfloor} - B_t}{\sqrt{t \log\log t}} = 0 \quad \text{a.s..}$$

Step I. Fix $\epsilon > 0$. With probability 1, there exists $t_0(\omega)$ such that $\frac{t}{1+\epsilon} \leq \tau_{\lfloor t\rfloor} \leq (1+\epsilon)t$ for all $t > t_0(\omega)$. We let $M_t = \sup\{|B_s - B_t|, \frac{t}{1+\epsilon} \leq s \leq t(1+\epsilon)\}$, and $t_k = (1+\epsilon)^k$. If $t_k \leq t \leq t_{k+1}$,

$$M_t \leq \sup_{s,r\in[t_{k-1},t_{k+2}]} |B_s - B_r| \leq 2 \sup_{s\in[t_{k-1},t_{k+2}]} |B_s - B_{t_{k-1}}|. \tag{9.6}$$

Note that $t_{k+2} - t_{k-1} = \delta t_{k-1}$, where $\delta = (1+\epsilon)^3 - 1$. By scaling, for $h > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{s\in[t_{k-1},t_{k+2}]} |B_s - B_{t_{k-1}}| > \sqrt{\delta h} \right) &= \mathbb{P}\left(\sup_{s\in[0,1]} |B_s| > \sqrt{h} \right) \leq 2\mathbb{P}\left(\sup_{s\in[0,1]} B_s > \sqrt{h} \right) \\ &= \frac{4}{\sqrt{2\pi}} \int_{\sqrt{h}}^{\infty} e^{-x^2/2}\,\mathrm{d}x \leq \sqrt{\frac{8}{\pi h}} e^{-h/2},\end{aligned}$$

where the last estimate is from the tail bound in the proof of Theorem 4.30. We set $h = 3t_{k-1}\log\log t_{k-1}$ for large enough k in the last display to obtain

$$\mathbb{P}\left(\sup_{s\in[t_{k-1},t_{k+2}]}|B_s - B_{t_{k-1}}| > \sqrt{3\delta t_{k-1}\log\log t_{k-1}}\right) \le \sqrt{\frac{8((k-1)\log(1+\epsilon))^{-3t_{k-1}}}{3(1+\epsilon)^{k-1}(\log(k-1)+\log\log(1+\epsilon))}}.$$

By Borel-Cantelli lemma, the above events occur for only finitely many k. By (9.6), for large enough t,

$$|S_{\lfloor t\rfloor} - B_t| = |B_{\tau_{\lfloor t\rfloor}} - B_t| \le M_t \le \sqrt{3\delta t_{k-1}\log\log t_{k-1}} \le \sqrt{3\delta t\log\log t},$$

where we choose k to be such that $(1+\epsilon)^k = t_k \le t \le t_{k+1} = (1+\epsilon)^{k+1}$. Hence with probability 1,

$$\limsup_{t\uparrow\infty}\frac{S_{\lfloor t\rfloor} - B_t}{\sqrt{t\log\log t}} \le \sqrt{3\delta}.$$

Since $\delta = (1+\epsilon)^3 - 1$ and $\epsilon > 0$ is arbitrary, we set $\epsilon \downarrow 0$ to conclude the proof. $\square$

9.1.2 The Arcsine Laws

In this part we extend the arcsine laws we discuss in Proposition 4.45 to random walks, and also introduce another arcsin law for the positive set of one-dimensional Brownian motions.

Proposition 9.10 (Last sign change). *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean 0 and variance 1, and the associated random walk $S_n = X_1 + X_2 \cdots + X_n$. Define L_n to be the last time the random walk changes its sign before time n, i.e.*

$$L_n = \max\{1 \le k \le n : S_k S_{k-1} \le 0\}.$$

Then $\frac{L_n}{n}$ converges in law to the arcsine distribution.

Proof. We define a function $\psi : C([0,1]) \to [0,1]$ by

$$\psi(\omega) = \sup\{t \in [0,1] : \omega(t) = 0\}, \quad \omega \in C([0,1]).$$

Then ψ is continuous at every function $\omega \in C([0,1])$ such that ω takes positive and negative values in every neighbourhood of every zero and $\omega(1) \neq 0$. To see this, we fix such a function $\omega \in C([0,1])$ and $\epsilon > 0$. Let $\delta_0 = \inf_{t\in[\psi(\omega)+\epsilon,1]}|\omega(t)|$, and take $\delta_1 > 0$ with $(-\delta_1, \delta_1) \subset \omega((\psi(\omega)-\epsilon, \psi(\omega)+\epsilon))$. For any $0 < \delta < \delta_0 \wedge \delta_1$ and $\widetilde{\omega} \in C([0,1])$ with $\|\widetilde{\omega} - \omega\|_\infty < \delta$, the function $\widetilde{\omega}$ has no zero in $(\psi(\omega)+\epsilon, 1]$ but has at least one zero in $(\psi(\omega)-\epsilon, \psi(\omega)+\epsilon)$, since there exists $s, t \in (\psi(\omega)-\epsilon, \psi(\omega)+\epsilon)$ with $\widetilde{\omega}(s) < 0$ and $\widetilde{\omega}(t) > 0$. Thus $|\psi(\omega) - \psi(\widetilde{\omega})| < \epsilon$, which shows that ψ is almost everywhere on $C([0,1])$ under the Wiener measure.

By Donsker's invariance principle and the Portmanteau lemma,

$$\psi\left(\frac{S(n\cdot)}{\sqrt{n}}\right) \xrightarrow{d} \psi(B) \quad \text{as } n \to \infty.$$

Also note that

$$\left|\psi\left(\frac{S(n\cdot)}{\sqrt{n}}\right) - \frac{L_n}{n}\right| \le \frac{1}{n},$$

which converges to 0. By Slutsky's lemma, $\frac{L_n}{n} \to \psi(B)$, which is arcsine distributed. $\square$

Now we are going to introduce another arcsine law for Brownian motions.

Lemma 9.11 (Richard)**.** *Let $(S_n)_{n=0}^{\infty}$ be the simple, symmetric random walk on integers. Then for all $n \in \mathbb{N}$,*

$$|\{k \in \{1, 2, \cdots, n\} : S_k > 0\}| \overset{d}{=} \min\left\{k \in \{0, 1, \cdots, n\} : S_k = \max_{1 \le j \le n} S_j\right\} \tag{9.7}$$

Proof. We let $X_k = S_k - S_{k-1}$ for each $k = 1, \cdots, n$. We rearrange the tuple $(X_1, \cdots, X_n)$ as follows:

- For the terms X_k with $S_k > 0$, place them in *decreasing* order of index k.
- For the terms X_k with $S_k \le 0$, append them in *increasing* order of index k.

We denote by $(Y_1, \cdots, Y_n)$ the rearranged tuple, and write $S'_k = Y_1 + \cdots + Y_k$ for the corresponding partial sums. We first claim that $(X_1, \cdots, X_n) \overset{d}{=} (Y_1, \cdots, Y_n)$.

(i) If all partial sums are nonpositive, then trivially the conditional distributions are the same.

(ii) Otherwise, we condition on the event $\max\{j \in \{1, \cdots, n\} : S_j > 0\} = k$, so $Y_1 = X_k$. Then the tuples $(X_1, \cdots, X_k)$ and $(X_{k+1}, \cdots, X_n) \overset{d}{=} (Y_{k+1}, \cdots, Y_n)$ are conditionally independent. Furthermore, the i.i.d. increments $X_1, \cdots, X_k$ are conditionally exchageable (conditioned on $S_k = 1$ if $k < n$, or $S_n \ge 1$ if $k = n$). Hence the conditional law of $(X_k, X_1, \cdots, X_{k-1})$ remains the same. Repeating this argument now for $(X_1, \cdots, X_{k-1})$, we see after finitely many steps that the two tuples have the same law.

As a consequence, we have $(S_1, \cdots, S_n) \overset{d}{=} (S'_1, \cdots, S'_n)$. We check that

$$|\{k \in \{1, 2, \cdots, n\} : S_k > 0\}| \overset{d}{=} \min\left\{k \in \{0, 1, \cdots, n\} : S'_k = \max_{1 \le j \le n} S'_j\right\}.$$

Indeed, this holds trivially for $n = 1$. When X_{n+1} is appended there are two possibilities:

- if $S_{n+1} \le 0$, then $Y_{n+1} = X_{n+1}$, and the position of the leftest maximum in $(S'_k)_{k=1}^{n+1}$ does not change.
- if $S_{n+1} > 0$, then $Y_1 = X_{n+1}$, and the position of the leftest maximum in $(S'_k)_{k=1}^{n+1}$ is shifted by one position to the right.

Thus we complete thr proof by induction. □

Theorem 9.12 (Lévy's Arcsine law)**.** *Let $(B_t)_{t \ge 0}$ be a Brownian motion starting at $B_0 = 0$. Then the Lebesgue measure of the positive set of B in $[0, 1]$ satisfies the arcsine law, i.e.*

$$\mathbb{P}\left(m\left\{s \in [0, 1] : B_s > 0\right\} \le t\right) = \frac{2}{\pi} \arcsin(\sqrt{t}),$$

where m is the Lebesgue measure on $[0, 1]$.

Proof. We define a function $\phi : C([0, 1]) \to [0, 1]$ by

$$\phi(\omega) = \inf\left\{t \in [0, 1] : \omega(t) = \max_{s \in [0,1]} \omega(s)\right\}, \quad \omega \in C([0, 1]).$$

Then ϕ is continuous at every function $\omega \in C([0, 1])$ with a unique maximum, hence almost everywhere continuous under the Wiener measure. To see this, we fix such a function with $\omega \in C([0, 1])$ and $\epsilon > 0$. Let M_1 be the supremum of ω on $[0, \phi(\omega) - \epsilon] \cup [\phi(\omega) + \epsilon]$ and $M_0 = \max_{t \in [0,1]} \omega(t)$. Then for all $\widetilde{\omega} \in C([0, 1])$ with $\|\widetilde{\omega} - \omega\|_\infty < \frac{1}{2}(M_0 - M_1)$, we have $|\phi(\widetilde{\omega}) - \phi(\omega)| < \epsilon$, which shows the almost everywhere continuity of ϕ. By Donsker's invariance principle, the right-hand side of (9.7) divide by n converges in distribution to $\phi(B)$, which has the arcsine distribution by Theorem 4.45. Next, we define a function $\psi : C([0, 1]) \to [0, 1]$ by

$$\psi(\omega) = m\left\{t \in [0, 1] : \omega(t) > 0\right\}, \quad \omega \in C([0, 1]),$$

which is continuous at every $\omega \in C([0, 1])$ with the property

$$\lim_{\epsilon \downarrow 0} m\left\{t \in [0, 1] : \omega(t) \in (-\epsilon, \epsilon)\right\} = 0, \tag{9.8}$$

because for every $\widetilde{\omega} \in C([0,1])$ with $\|\widetilde{\omega} - \omega\|_\infty < \epsilon$, one have

$$m\{t \in [0,1] : \omega(t) > \epsilon\} \leq \psi(\widetilde{\omega}) \leq m\{t \in [0,1] : \omega(t) > -\epsilon\}.$$

The property (9.8) is equivalent to the property $m\{t \in [0,1] : \omega(t) = 0\}$ which the Brownian motion has almost surely, hence ψ is almost everywhere continuous on $C([0,1])$ under the Wiener measure. Note that

$$\left| \frac{|k \in \{1, \cdots, n\} : S_k > 0|}{n} - m\left\{t \in [0,1] : \frac{S(nt)}{\sqrt{n}} > 0\right\} \right| \leq \frac{|k \in \{1, \cdots, n\} : S_k = 0|}{n},$$

which converges to 0 in probability, since

$$\frac{1}{2n} \sum_{k=0}^{n} \mathbb{P}(S_{2k} = 0) = \frac{1}{2n} \sum_{k=0}^{n} 2^{-2k} \binom{2k}{k} \sim \frac{1}{2n} \sum_{k=0}^{n} \sqrt{\frac{1}{\pi k}} \to 0.$$

Again by Donsker's invariance principle, the left-hand side of (9.7) divided by n converges in distribution to $\psi(B) = m\{t \in [0,1] : B_t > 0\}$. This complete the proof. $\square$

Corollary 9.13 (Arcsine laws for random walks)**.** *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean 0 and variance 1, and the associated random walk $S_n = X_1 + X_2 \cdots + X_n$.*

(i) (First maximum). Let

$$U_n = \inf\left\{1 \leq k \leq n : S_k = \max_{1 \leq j \leq n} S_j\right\}.$$

(ii) (Occupation times of half-lines). Let

$$M_n = \sum_{j=1}^{n} \mathbb{1}_{\{S_j > 0\}}.$$

Then both U_n/n and M_n/n converge in law to the arcsine distribution as $n \to \infty$.

Proof. We use the almost everywhere continuity of the mappings ϕ and ψ we constructed in the proof of Theorem 9.12 and apply Donsker's invariance principle. $\square$

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] Rick Durrett. (2019). *Probability: Theory and Examples, Version 5*. Cambridge University Press, Cambridge.

[5] Samuel Karlin and Howard M. Taylor. (1975). *A First Course in Stochastic Processes, 2nd Edition*. Academic Press, Inc., New York.

[6] Jacques Neveu. (1964). *Bases Mathématiques du Calcul des Probabilités*. Masson et Cie, Paris.

[7] Daniel Revuz and Marc Yor. (1999). *Continuous Martingales and Brownian Motion, 3rd Edition*. Springer, Berlin.

[8] Giuseppe Da Prato. (2014). *Introduction to Stochastic Analysis and Malliavin Calculus*. Edizioni della Normale, Pisa.

[9] Grigorios A. Pavliotis. (2014). *Stochastic Processes and Applications, Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer, Berlin.