Stochastic Analysis

Zhiyuan Zhan
<thaleszhan@gmail.com>

October 24, 2025

Contents

Chapter 1

Basic Measure Theory

1.1 Conditional Expectation

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1.1 (Conditional Expectation)**.** Let $X : \Omega \to \mathbb{R}$ be a L^1 random variable and $\mathcal{G} \subset \mathcal{F}$ be a σ-sub-field. A random variable Y is called the conditional expectation of X given $\mathcal{G}$ if

(i) Y is $\mathcal{G}$-measurable,

(ii) for any $A \in \mathcal{G}$,
$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P},$$

Theorem 1.1.2. *For given X and $\mathcal{G}$, such Y exists and is unique, denoted by $Y = \mathbb{E}[X \mid \mathcal{G}]$.*

Proof. For the uniqueness, let Y' be another conditional expectation. Let
$$A_\varepsilon = \{Y - Y' \geq \varepsilon\} \in \mathcal{G}.$$
for any $\varepsilon > 0$. So
$$\varepsilon \mathbb{P}(A_\varepsilon) \leq \int_{A_\varepsilon} Y - Y' d\mathbb{P} = \int_{A_\varepsilon} X d\mathbb{P} - \int_{A_\varepsilon} X d\mathbb{P} = 0.$$
As $\varepsilon \to 0$, $Y \leq Y'$ a.e.. Similarly, we have $Y' \leq Y$. So $Y = Y'$.
For existence, WTLG, assume $X \geq 0$. Let
$$\nu(A) = \int_A X d\mathbb{P}, \quad A \in \mathcal{G}.$$
Then ν is a measure on $\mathcal{G}$, which is absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{G}$. So by the Radon-Nikodym theorem, there exists a $\mathcal{G}$-measurable Y such that
$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}, \quad A \in \mathcal{G}.$$
□

Example 1.1.3. Suppose $X \in L^2$. Then
$$\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2\right] = \inf\left\{\mathbb{E}[(X - Y)^2] \mid Y \text{ is } \mathcal{G} - \text{measurable.}\right\}$$

1.2 Change of Measures

Fix $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual condition. Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Denote $\mathbb{E} = \mathbb{E}_{\mathbb{P}}$.

Proposition 1.2.1. *Let $\mathbb{P}, \mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$. Suppose $\mathbb{Q} \ll \mathbb{P}$ and*

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

Then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$, $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{G}$ and

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{G}} = \mathbb{E}\left[Z \mid \mathcal{G}\right].$$

Proof. Absolutely continuity is obvious. For any $A \in \mathcal{G}$, by the property of conditional expectation,

$$\mathbb{Q}(A) = \int_A Z d\mathbb{P} = \int_A \mathbb{E}[Z \mid \mathcal{G}] d\mathbb{P}.$$

So by the uniqueness in Radon-Nikodym Theorem,

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{G}} = \mathbb{E}\left[Z \mid \mathcal{G}\right].$$

□

Proposition 1.2.2. *Let $\mathbb{P}, \mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$. Suppose $\mathbb{P} \sim \mathbb{Q}$ with $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$. Let $\mathcal{G} \subset \mathcal{F}$ σ-subalgebra. Then for any $\mathcal{F}$-measurable $Y \geq 0$.*

$$\mathbb{E}_{\mathbb{Q}}[Y \mid \mathcal{G}] = \frac{\mathbb{E}[YZ \mid \mathcal{G}]}{\mathbb{E}[Z \mid \mathcal{G}]}.$$

Remark 1.2.3. For general Y, we need $Y \in L^1$ then $Y = Y^+ - Y^-$.

Proof. For any $A \in \mathcal{G}$,

$$\begin{aligned}
\int_A \mathbb{E}_{\mathbb{Q}}[Y \mid \mathcal{G}] d\mathbb{Q} &= \int_A Y d\mathbb{Q} = \int_A YZ d\mathbb{P} \\
&= \int_A \mathbb{E}[YZ \mid \mathcal{G}] d\mathbb{P} = \int_A \mathbb{E}[YZ \mid \mathcal{G}] \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} \\
&= \int_A \mathbb{E}[YZ \mid \mathcal{G}] \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^{-1} d\mathbb{Q} = \int_A \mathbb{E}[YZ \mid \mathcal{G}] \left(\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{G}}\right)^{-1} d\mathbb{Q} \\
&= \int_A \mathbb{E}[YZ \mid \mathcal{G}] \mathbb{E}\left[Z \mid \mathcal{G}\right]^{-1} d\mathbb{Q}.
\end{aligned}$$

So

$$\mathbb{E}_{\mathbb{Q}}[Y \mid \mathcal{G}] = \frac{\mathbb{E}[YZ \mid \mathcal{G}]}{\mathbb{E}[Z \mid \mathcal{G}]}.$$

□

Chapter 2

Discrete Martingale Theory

2.1 Stochastic Process

Definition 2.1.1 (Stochastic Process)**.** A family of $\{X_t : t \in I\}$ of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a stochastic process, where

(1) $I = \mathbb{N} \cup \{0\}$ or

(2) $I = [0, \infty)$.

Definition 2.1.2 (Finite Dimensional Distribution)**.** A finite distribution of a stochastic process $\{X_t : t \in I\}$ for a give sequence of time $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ is the probability law of $(X_{t_1}, \cdots, X_{t_n})$. If two stochastic processes has same finite distributions, then they are called having the same law.

2.2 Discrete Martingale

Theorem 2.2.1. *Let $X_1, \cdots, X_n$ be a sequence of random variables. Then Y is $\sigma(X_1, \cdots, X_n)$-measurable if and only if $Y = g(X_1, \cdots, X_n)$ for some measurable function g.*

Example 2.2.2. (1) Let $\{\mathcal{F}_n\}$ be a filtration and Y be an integrable random variable. Let $Z_n = \mathbb{E}[Y \mid \mathcal{F}_n]$. Clearly, Z_n is $\mathcal{F}_n$-measurable and by Jensen's Inequality,

$$\mathbb{E}\left[|Z_n|\right] \leq \mathbb{E}|Y| < \infty.$$

Furthermore,

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \mathbb{E}\left[\mathbb{E}[Y \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n\right] = \mathbb{E}[Y \mid \mathcal{F}_n] = Z_n,$$

i.e. Z_n is a $\{\mathcal{F}_n\}$-martingale.

(2) Assume $X_1, X_n, \cdots$ are independent, integrable random variables with $a_n = \mathbb{E}[X_n] \neq 0$. Define

$$Z_n = \frac{X_1 X_2 \cdots X_n}{a_1 a_2 \cdots a_n}$$

and $Z_0 = 1$. Let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. So Z_n is $\mathcal{F}_n$-measurable and is integrable by the independence. Moreover,

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \frac{X_1 \cdots X_n}{a_1 \cdots a_n a_{n+1}} \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \frac{X_1 \cdots X_n}{a_1 \cdots a_n} = Z_n$$

i.e. Z_n is a $\{\mathcal{F}_n\}$-martingale.

(3) Assume $X_1, X_n, \cdots$ are independent, integrable random variables valued $1, -1$ with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$. Define

$$Z_n = S_n^2 - n, \quad S_n = \sum_{i=1}^{n} X_n.$$

Let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. Then clearly Z_n is $\mathcal{F}_n$-measurable and is integrable.

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[(S_n + X_{n+1})^2 \mid \mathcal{F}_n\right] - n - 1 \\
&= S_n^2 + 2S_n\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] - n - 1 \\
&= S_n^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] - n - 1 \\
&= S_n^2 - n = Z_n.
\end{aligned}$$

So Z_n is a $\{\mathcal{F}_n\}$-martingale.

Proposition 2.2.3. *(1) If $(X_n)_{n\geq 0}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$ and φ is a convex function such that $\varphi(X_n) \in L^1$, then $\{\varphi(X_n)\}_{n\geq 0}$ is a submartingale.*

(2) If $(X_n)_{n\geq 0}$ is a sub(sup)-martingale w.r.t. $\{\mathcal{F}_n\}$ and φ is a increasing(decreasing) convex function such that $\varphi(X_n) \in L^1$, then $\{\varphi(X_n)\}_{n\geq 0}$ is a submartingale. In particular, $\{(X_n - a)_+\}$ is a submartingale.

2.3 Stopping Time

Definition 2.3.1 (Stopping Time)**.** Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration. A random variable $T(\omega) \in \mathbb{N} \cup \{0\} \cup \{\infty\}$ is called a stopping time w.s.t $\{\mathcal{F}_n\}_{n\geq 0}$ if

$$\{T \leq n\} \in \mathcal{F}_n, \quad n \in \mathbb{N} \cup \{0\} \cup \{\infty\}.$$

Remark 2.3.2. By the definition, it is clear that $\{T > n\} \in \mathcal{F}_n$ and so $\{T = n\} = \{T \leq n\} \cap \{T > n-1\} \in \mathcal{F}_n$ and also $\{T < n\} = \cup_{k=1}^{n-1} \{T = k\} \in \mathcal{F}_n$.

Definition 2.3.3 (Stopping Process)**.** Let T be a stopping time and $\{Z_n\}_{n\geq 0}$ be a stochastic process. Define

$$Z_{T\wedge n}(\omega) = \begin{cases} Z_n(\omega), n \leq T(\omega) \\ Z_T(\omega), n > T(\omega) \end{cases}$$

Then the process $\{Z_{T\wedge n}\}_{n\geq 0}$ is called the stopping process of Z at T.

Theorem 2.3.4. *If $\{Z_n\}_{n\geq 0}$ is a (sub or sup-)martingale w.s.t. $\mathbb{F} = \{\mathcal{F}_n\}_{n\geq 0}$ and T is a stopping time w.s.t. $\mathbb{F}$, then the stopping process $\{Z_{T\wedge n}\}_{n\geq 0}$ is also a (sub or sup-)martingale w.s.t. $\mathbb{F}$.*

Proof. Let $Y_n = Z_{T\wedge n}$. Then

$$\begin{aligned}
Y_n &= Z_{T\wedge n}\mathbb{I}_{\{T\geq n\}} + Z_{T\wedge n}\mathbb{I}_{\{T<n\}} \\
&= Z_n\mathbb{I}_{\{T\geq n\}} + Z_T\mathbb{I}_{\{T<n\}} \\
&= Z_n\mathbb{I}_{\{T\geq n\}} + \sum_{k=0}^{n-1} Z_k\mathbb{I}_{\{T=k\}}
\end{aligned}$$

Therefore, Y_n is $\mathcal{F}_n$-measurable and L^1. For the martingale property, first note that

$$\begin{aligned} Y_{n+1} &= Z_{T\wedge n+1} \\ &= Z_{T\wedge n} + \mathbb{I}_{\{T\geq n+1\}}(Z_{n+1} - Z_n) \\ &= Y_n + \mathbb{I}_{\{T\geq n+1\}}(Z_{n+1} - Z_n). \end{aligned}$$

Therefore, by $[\mathbb{I}_{\{T\geq n+1\}} = \mathbb{I}^c_{\{T<n\}} \in \mathcal{F}_n$,

$$\begin{aligned} \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[Y_n \mid \mathcal{F}_n] + \mathbb{E}[\mathbb{I}_{\{T\geq n+1\}}(Z_{n+1} - Z_n) \mid \mathcal{F}_n] \\ &= Y_n + \mathbb{I}_{\{T\geq n+1\}}\mathbb{E}[(Z_{n+1} - Z_n) \mid \mathcal{F}_n] \\ &= Y_n. \end{aligned}$$

Similarly, the reasoning is true for sub or sup-martingale. □

Lemma 2.3.5. *Suppose (X_n) is a supermartingale. Let T, S be two bounded stopping times with $S \leq T \leq N$. Then*

$$\int_{S\leq N} X_T \, \mathrm{d}\mathbb{P} \leq \int_{S\leq N} X_S \, \mathrm{d}\mathbb{P}.$$

Proof. Let $Y_n = X_{T\wedge n} - X_{S\wedge n}$ and note that

$$Y_n - Y_{n-1} = \mathbb{I}_{\{T\geq n>S\}}(X_n - X_{n-1}).$$

It follows that

$$\mathbb{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}\left[\mathbb{I}_{\{T\geq n>S\}}(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}\right] = \mathbb{I}_{\{T\geq n>S\}}\mathbb{E}\left[(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}\right] \leq 0$$

because $\{T \geq n > S\} \in \mathcal{F}_{n-1}$. Therefore, Y_n is a supermartingale, i.e.,

$$\mathbb{E}[Y_N \mid \mathcal{F}_n] \leq Y_n, \quad \forall\, n \leq N.$$

It implies that

$$\mathbb{I}_{S=n}Y_n \geq \mathbb{E}[\mathbb{I}_{S=n}Y_N \mid \mathcal{F}_n]$$

and by taking expectation we have

$$\mathbb{E}\left[\mathbb{I}_{S=n}Y_N\right] \leq \mathbb{E}[\mathbb{I}_{S=n}Y_n] = 0,$$

i.e.,

$$\mathbb{E}\left[\mathbb{I}_{S=n}Y_N\right] = \mathbb{E}\left[\mathbb{I}_{S=n}(X_T - X_S)\right] \leq 0$$

Taking summation of n from 1 to N, we get

$$\int_{S\leq N} X_T \, \mathrm{d}\mathbb{P} \leq \int_{S\leq N} X_S \, \mathrm{d}\mathbb{P}.$$ □

2.4 Martingale Convergence Theorem

For a real-valued $\mathbb{F} = (\mathcal{F}_n)$-adapted process (X_n) and $a < b$, define a sequence of stopping times (τ_n) as follows,

$$\begin{aligned} &\tau_1 := \min\{n \colon X_n \leq a\}, \quad \tau_2 := \{n \geq \tau_1 \colon X_n \geq b\}, \cdots \\ &\tau_{2k+1} := \min\{n \geq \tau_{2k} \colon X_n \leq a\}, \quad \tau_{2k+2} := \min\{n \geq \tau_{2k+1} \colon X_n \geq b\}, \cdots \end{aligned}$$

Set a random variable

$$U_N^X(a,b) := \max\{k \colon \tau_{2k} \leq N\}$$

Then $U_N^X(a,b)$ is the number of up-crossing of $(X_n)_{n=0}^N$ for the interval $[a,b]$.

Theorem 2.4.1. *Let (X_n) is a supermartingale. We have*

$$\mathbb{P}\left(U_N^X(a,b) > j\right) \leq \frac{1}{b-a}\int_{U_N^X(a,b)=j} (X_N - a)^- \, \mathrm{d}\mathbb{P}$$

and

$$\mathbb{E}\left[U_N^X(a,b)\right] \leq \frac{1}{b-a}\mathbb{E}\left[(X_N-a)^-\right].$$

Proof. WTLG, assume $a = 0$ and $(X_n)_{n=0}^N$. Set

$$S = \tau_{2j+1} \wedge (N+1), \quad T = \tau_{2(j+1)} \wedge (N+1).$$

Then

$$\{\tau_{2j+1} \leq N\} = \{S \leq N\},$$

on $\{S \leq N\}$, $X_S = X_{\tau_{2j+1}} \leq a = 0$, and

$$\left\{\tau_{2(j+1)} \leq N\right\} = \left\{U_N^X(a,b) \geq j+1\right\} = \left\{U_N^X(a,b) > j\right\},$$

which follows that

$$\left\{U_N^X(0,b) > j\right\} = \left\{\tau_{2(j+1)} \leq N\right\} = \{S < N,\ X_T \geq b\}.$$

On the other hand,

$$\{S < N,\ X_T < b\} = \{S < N,\ T = N+1\} \subset \left\{U_N^X(a,b) = j\right\}.$$

Then

$$\begin{aligned}
b\mathbb{P}\left(U_N^X(0,b) > j\right) &= \int_{\left\{U_N^X(0,b)>j\right\}} b \, \mathrm{d}\mathbb{P} = \int_{\{S<N,\ X_T \geq b\}} b \, \mathrm{d}\mathbb{P} \\
&\leq \int_{\{S<N,\ X_T\geq b\}} X_T \, \mathrm{d}\mathbb{P} \\
&= \int_{\{S<N\}} X_T \, \mathrm{d}\mathbb{P} - \int_{\{S<N,\ X_T<b\}} X_T \, \mathrm{d}\mathbb{P} \\
&\leq \int_{\{S<N\}} X_S \, \mathrm{d}\mathbb{P} - \int_{\{S<N,\ T=N+1\}} X_T \, \mathrm{d}\mathbb{P} \\
&\leq 0 - \int_{\{S<N,\ T=N+1\}} X_N \, \mathrm{d}\mathbb{P} \leq \int_{\{S<N,\ T=N+1\}} X_N^- \, \mathrm{d}\mathbb{P} \\
&\leq \int_{\left\{U_N^X(a,b)=j\right\}} X_N^- \, \mathrm{d}\mathbb{P}.
\end{aligned}$$

and the second result is by taking the sum of j from 0 to ∞. ☐

Theorem 2.4.2 (Martingale Convergence Theorem)**.** *Let (X_n) be a supermartingale with*

$$\sup_n \mathbb{E}[X_n^-] < \infty.$$

Then

$$X(\omega) := \lim_{n\to\infty} X_n(\omega)$$

exists almost everywhere. In particular, if $\sup_n \mathbb{E}[|X_n|] < \infty$, $X \in L^1$.

Proof. For any $a < b$, let

$$U^X(a,b) = \lim_{N\to\infty} U_N^X(a,b),$$

which always exists by taking value in $[0,\infty]$ since $U_N^X(a,b)$ is monotone increasing. By MCT,

$$\mathbb{E}\left[U^X(a,b)\right] = \lim_{N\to\infty} \mathbb{E}\left[U_N^X(a,b)\right] \le \frac{1}{b-a}\sup_N \mathbb{E}\left[(X_N-a)^-\right] < \infty.$$

Set $W_{a,b} = \left\{U^X(a,b) = \infty\right\}$, so $\mathbb{P}(W_{a,b}) = 0$. Define

$$V_{a,b} = \left\{\liminf_n X_n < a,\ \limsup_n X_n > b\right\},$$

and so $V_{a,b} \subset W_{a,b}$ and $\mathbb{P}(V_{a,b}) = 0$. Next,

$$\left\{\liminf_n X_n < \limsup_n X_n\right\} = \bigcup_{a<b\in\mathbb{Q}} V_{a,b}$$

So

$$\mathbb{P}\left\{\liminf_n X_n < \limsup_n X_n\right\} = 0.$$

When $\sup_n \mathbb{E}[|X_n|] < \infty$, by Fatou's lemma, it is obvious $X \in L^1$. $\square$

Example 2.4.3. (1) Let (X_n) be a martingale with $|X_{n+1} - X_n| \le M$ for any n. Let

$$C = \left\{\lim_n X_n \text{ exists and finite.}\right\}$$

and

$$D = \left\{\liminf_n X_n = -\infty,\ \limsup_n X_n = \infty\right\}.$$

Then we have

$$\mathbb{P}\left(C \cup D\right) = 1.$$

Proof. WTLG, assume $X_0 = 0$. For any $k \in \mathbb{N}$, let

$$N_k := \inf\left\{n\colon X_n \le -k\right\},$$

which is a stopping time. So $(X_{n\wedge N_k})$ is also a martingale. Note that

$$X_{N_k} = X_{N_k} - X_{N_k-1} + X_{N_k-1} \ge -M - k,$$

so

$$X_{n\wedge N_k} \ge -k - M \ \Rightarrow\ X_{n\wedge N_k} + a + M \ge 0$$

Then by the martingale convergence theorem, $\lim_n X_{n\wedge N_k}$ exists. So $\lim_n X_n$ exists on $\{N_k = \infty\}$.

$$\left\{\liminf_n X_n > -\infty\right\} = \bigcup_k \{N_k = \infty\}$$

It implies that $\lim_n X_n$ exists on $\{\liminf_n X_n > -\infty\}$. Similarly, by considering $-X_n$, $\lim_n X_n$ exists on $\{\limsup_n X_n < \infty\}$. $\square$

(2) Let $(\mathcal{F}_n)$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Let $B_n \in \mathcal{F}_n$ be a sequence of events.

$$\bigcap_n \bigcup_{k \geq n} B_k = \left\{ \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{I}_{B_n} \mid \mathcal{F}_{n-1}\right] = \infty \right\}$$

Proof. Set $X_0 = 0$ and $X_n = \sum_{m=1}^{n} \mathbb{I}_{B_m}$. Note that

$$\bigcap_n \bigcup_{k \geq n} B_k = \left\{ \sum_{m=1}^{\infty} \mathbb{I}_{B_m} = \infty \right\}$$

Define $M_0 = 0$ and

$$M_n = X_n - \sum_{m=1}^{n} \mathbb{E}\left[\mathbb{I}_{B_m} \mid \mathcal{F}_{m-1}\right] = \sum_{m=1}^{n} \left(\mathbb{I}_{B_m} - \mathbb{E}\left[\mathbb{I}_{B_m} \mid \mathcal{F}_{m-1}\right]\right)$$

and so (M_n) is a martingale w.s.t. $(\mathcal{F}_n)$. Moreover,

$$|M_{n+1} - M_n| = \left|\mathbb{I}_{B_{n+1}} - \mathbb{E}\left[\mathbb{I}_{B_{n+1}} \mid \mathcal{F}_n\right]\right| \leq 2$$

By above, it suffices to prove that on C and D. For C, because $\lim_n M_n$ exists,

$$\sum_{m=1}^{\infty} \mathbb{I}_{B_m} = \infty \;\Leftrightarrow\; \sum_{m=1}^{\infty} \mathbb{E}\left[\mathbb{I}_{B_m} \mid \mathcal{F}_{m-1}\right] = \infty.$$

On D, it is also true. □

2.5 Doob's Decomposition

Definition 2.5.1. Let $(\mathcal{F}_n)$ be a filtration.

(1) A stochastic process (H_n) is called adapted w.s.t. $(\mathcal{F}_n)$ if H_n is $\mathcal{F}_n$-measurable.

(2) A stochastic process (H_n) is called predictable w.s.t. $(\mathcal{F}_n)$ if H_n is $\mathcal{F}_{n-1}$-measurable.

Theorem 2.5.2 (Doob's Decomposition Theorem)**.** *Any submartingale (X_n) can be uniquely written as*

$$X_n = M_n + A_n$$

where M_n is a martingale and A_n is a predictable increasing process with $A_0 = 0$.

Proof. If $X_n = M_n + A_n$, then

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[M_n + A_n \mid \mathcal{F}_{n-1}] = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n.$$

So

$$A_n - A_{n-1} = \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}],$$

which implies that by setting A_0

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

that is predictable and increasing because (X_n) is a submartingale. Let $M_n := X_n - A_n$.

$$\begin{aligned}
\mathbb{E}\left[M_n \mid \mathcal{F}_{n-1}\right] &= \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - A_n \\
&= \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] - A_n + X_{n-1} \\
&= A_n - A_{n-1} - A_n + X_{n-1} = M_{n-1}.
\end{aligned}$$

So M_n is a martingale. □

Remark 2.5.3. Note that if (X_n) is a supermartingale then it can be uniquely written as

$$X_n = M_n - A_n,$$

for a martingale M_n and a predictable increasing process A_t with $A_0 = 0$.

2.6 L^p Convergence

Lemma 2.6.1 (Bounded Optional Stopping Time Theorem)**.** *If (X_n) is a submartingale and N is a finite stopping time with $N \leq K$, then*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_K]$$

Proof. We have known $(X_{n\wedge N})$ is a submartingale, i.e.,

$$\mathbb{E}[X_0] = \mathbb{E}[X_{0\wedge N}] \leq \mathbb{E}[X_{N\wedge K}] = \mathbb{E}[X_N].$$

For the second part, because $N \leq K$, $\Omega = \bigcup_{n=0}^{K} \{N = n\}$. It follows that

$$\mathbb{E}[X_N] = \sum_{n=0}^{K} \mathbb{E}[X_N \mathbb{I}_{\{N=n\}}] = \sum_{n=0}^{K} \mathbb{E}[X_n \mathbb{I}_{\{N=n\}}].$$

Moreover, for any $n \leq K$, because (X_n) is a submartingale

$$X_n \leq \mathbb{E}[X_K \mid \mathcal{F}_n].$$

Because N is a stopping time,

$$X_n \mathbb{I}_{\{N=n\}} \leq \mathbb{E}[X_K \mathbb{I}_{\{N=n\}} \mid \mathcal{F}_n] \;\Rightarrow\; \mathbb{E}[X_n \mathbb{I}_{\{N=n\}}] \leq \mathbb{E}[X_K \mathbb{I}_{\{N=n\}}].$$

So

$$\mathbb{E}[X_N] \leq \sum_{n=0}^{K} \mathbb{E}[X_K \mathbb{I}_{\{N=n\}}] = \mathbb{E}[X_K].$$ □

Theorem 2.6.2 (Doob's Martingale Inequality)**.** *Let (X_n) be a submartingale. Define*

$$A = \left\{ \max_{0\leq m\leq n} X_m \geq \lambda \right\}, \;\; \lambda > 0.$$

Then we have

$$\mathbb{P}(A) \leq \frac{1}{\lambda}\mathbb{E}[X_n \mathbb{I}_A] \leq \frac{1}{\lambda}\mathbb{E}[X_n^+ \mathbb{I}_A] \leq \frac{1}{\lambda}\mathbb{E}[X_n^+]$$

Proof. Define

$$N = \min\{m \colon X_m \geq \lambda\} \wedge n$$

Clearly, $N \leq n$ is a stopping time. On A, $X_N \geq \lambda$. It follows that

$$\lambda \mathbb{P}(A) = \int_A \lambda \; \mathrm{d}\mathbb{P} \leq \int_A X_N \; \mathrm{d}\mathbb{P} = \mathbb{E}[X_N \mathbb{I}_A]$$

By above lemma, $\mathbb{E}[X_N] \leq \mathbb{E}[X_n]$. Note that

$$\begin{aligned}\mathbb{E}[X_N] &= \mathbb{E}[X_N \mathbb{I}_A] + \mathbb{E}[X_N \mathbb{I}_{A^c}] \\ &= \mathbb{E}[X_N \mathbb{I}_A] + \mathbb{E}[X_n \mathbb{I}_{A^c}]\end{aligned}$$

and

$$\mathbb{E}[X_n] = \mathbb{E}[X_n \mathbb{I}_A] + \mathbb{E}[X_n \mathbb{I}_{A^c}].$$

So $\mathbb{E}[X_N \mathbb{I}_A] \leq \mathbb{E}[X_n \mathbb{I}_A]$. □

Theorem 2.6.3. *Let (X_n) be a submartingale. Set*

$$\bar{X}_n = \max_{0\le m\le n} X_n.$$

Then for any $p > 1$,

$$\mathbb{E}\left[\bar{X}_n^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[(X_n^+)^p\right].$$

In particular, if (Y_n) is a martingale and set $Y_n^* = \max_{0\le m\le n}|Y_m|$, then for $p > 1$

$$\mathbb{E}\left[(Y_n^*)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|Y_n|^p\right].$$

Proof. For $M > 0$, note that

$$\left\{\bar{X}_n \wedge M \ge \lambda\right\} = \left\{\bar{X}_n \ge \lambda\right\} \text{ or } \emptyset.$$

First,

$$\begin{aligned}
\mathbb{E}\left[(\bar{X}_n \wedge M)^p\right] &= \mathbb{E}\left[\int_0^{\bar{X}_n\wedge M} p\lambda^{p-1}\,\mathrm{d}\lambda\right] \\
&= \mathbb{E}\left[\int_0^\infty \mathbb{I}_{\left\{\bar{X}_n\wedge M\ge\lambda\right\}} p\lambda^{p-1} d\lambda\right] \\
&= \int_0^\infty p\lambda^{p-1}\mathbb{P}(\left\{\bar{X}_n \wedge M \ge \lambda\right\})\,\mathrm{d}\lambda \\
&= \int_0^\infty p\lambda^{p-1}\mathbb{P}(\left\{\bar{X}_n \ge \lambda\right\})\mathbb{I}_{\{M\le\lambda\}}\,\mathrm{d}\lambda \\
&\le \int_0^\infty p\lambda^{p-1}\frac{1}{\lambda}\int X_n^+\mathbb{I}_{\left\{\bar{X}_n\wedge M\ge\lambda\right\}}\,\mathrm{d}\mathbb{P}\,\mathrm{d}\lambda \\
&= \int X_n^+\int_0^{\bar{X}_n\wedge M} p\lambda^{p-2}\,\mathrm{d}\lambda\,\mathrm{d}\mathbb{P} \\
&= \frac{p}{p-1}\int X_n^+(\bar{X}_n\wedge M)^{p-1}\,\mathrm{d}\mathbb{P} \\
&\le \frac{p}{p-1}\mathbb{E}\left[\left|X_n^+\right|^p\right]^{\frac{1}{p}}\mathbb{E}\left[(\bar{X}_n\wedge M)^p\right]^{\frac{p-1}{p}},
\end{aligned}$$

where the final inequality is by the Hölder's Inequality. So

$$\mathbb{E}\left[(\bar{X}_n \wedge M)^p\right] \le \mathbb{E}\left[\left|X_n^+\right|^p\right].$$

As $M \to \infty$, by Fatou's lemma,

$$\mathbb{E}\left[\bar{X}_n^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[(X_n^+)^p\right].$$

In particular, when (Y_n) is a martingale, $|Y_n|$ is a submartingale by Jensen's Inequality. □

Theorem 2.6.4. *Let (X_n) be a martingale with $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for $p > 1$. Then $X_n \to X$ a.e. and it is in L^p ($p > 1$), i.e.,*

$$\mathbb{E}[|X_n - X|^p] \to 0.$$

Proof. Define $Y = \sup_n |X_n|$. Then by MCT and by above theorem,

$$\mathbb{E}[Y^p] = \lim_n \mathbb{E}\left[\sup_{0\le m\le n}|X_m|^p\right] \le \limsup_n \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|X_n|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}[|X_n|^p]$$

Because $|X_n - X|^p \to 0$ a.e. and $|X_n - X|^p \le c(|X_n|^p + |X|^p) \le c(|Y|^p + |X|^p)$, by DCT

$$\mathbb{E}[|X_n - X|^p] \to 0.$$

□

2.7 UI and L^1 Convergence

Definition 2.7.1 (Uniform Integrability)**.** A family of random variables $(X_i, i \in I)$ is said uniformly integrable (UI) if

$$\lim_{M\to\infty} \sup_{i\in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i|\geq M\}}\right] = 0.$$

Example 2.7.2. If $|X_i| \leq Y$ and $Y \in L^1$, then

$$\sup_i \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i|\geq M\}}\right] \leq \mathbb{E}\left[|Y| \mathbb{I}_{\{|X_i|\geq M\}}\right] \leq \mathbb{E}\left[|Y| \mathbb{I}_{\{|Y|\geq M\}}\right] \to 0.$$

Lemma 2.7.3. *If $X \in L^1$, then*

$$\lim_{\mathbb{P}(A)\to 0} \int_A |X| \, \mathrm{d}\mathbb{P} = 0$$

Proof. Since $X \in L^1$,

$$\int |X| \mathbb{I}_{\{|X|\geq M\}} \, \mathrm{d}\mathbb{P} \to 0, \text{ as } M \to \infty.$$

For any $M > 0$,

$$\begin{aligned} \int_A |X| \, \mathrm{d}\mathbb{P} &= \int_{A\cap\{|X|\leq M\}} |X| \, \mathrm{d}\mathbb{P} + \int_{A\cap\{|X|>M\}} |X| \, \mathrm{d}\mathbb{P} \\ &\leq M\mathbb{P}(A) + \int_{\{|X|>M\}} |X| \, \mathrm{d}\mathbb{P} \end{aligned}$$

For any $\varepsilon > 0$, it can choose M such that $\int_{\{|X|>M\}} |X| \, \mathrm{d}\mathbb{P} \leq \varepsilon/2$. For such M, by choosing $\delta \leq \varepsilon/(2M)$, then for any A with $\mathbb{P}(A) \leq \delta$, $\int_A |X| \, \mathrm{d}\mathbb{P} \leq \varepsilon$. □

Example 2.7.4. Let $X \in L^1$. Then

$$\{\mathbb{E}[X \mid \mathcal{G}] \colon \mathcal{G} \subset \mathcal{F}\}$$

is UI.

Proof. For any $\varepsilon > 0$, by above lemma, it can choose $\delta > 0$ so that if $\mathbb{P}(A) < \delta$, then

$$\int_A |X| \, \mathrm{d}\mathbb{P} < \delta.$$

Note that

$$\begin{aligned} \int_{\{|\mathbb{E}[X|\mathcal{G}]|\geq M\}} |\mathbb{E}[X \mid \mathcal{G}]| \, \mathrm{d}\mathbb{P} &\leq \mathbb{E}\left[\mathbb{E}[|X| \mathbb{I}_{\{\mathbb{E}[|X||\mathcal{G}]\geq M\}} \mid \mathcal{G}]\right] \\ &= \mathbb{E}\left[\mathbb{I}_{\{\mathbb{E}[|X||\mathcal{G}]\geq M\}} \mathbb{E}[|X| \mid \mathcal{G}]\right] = \mathbb{E}\left[|X| \mathbb{I}_{\{\mathbb{E}[|X||\mathcal{G}]\geq M\}}\right] \\ &= \int_{\{\mathbb{E}[|X||\mathcal{G}]\geq M\}} |X| \, \mathrm{d}\mathbb{P} \end{aligned}$$

On the other hand, by Chebyshev's Inequality

$$\mathbb{P}\left(\{\mathbb{E}[|X| \mid \mathcal{G}] \geq M\}\right) \leq \frac{1}{M}\mathbb{E}\left[\mathbb{E}[|X| \mid \mathcal{G}]\right] = \frac{\mathbb{E}[|X|]}{M}$$

Therefore, it can choose $M \geq M_0$ such that

$$\mathbb{P}\left(\{\mathbb{E}[|X| \mid \mathcal{G}] \geq M\}\right) \leq \delta$$

So

$$\sup_{\mathcal{G}} \int_{\{|\mathbb{E}[X|\mathcal{G}]|\geq M\}} |\mathbb{E}[X \mid \mathcal{G}]| \, \mathrm{d}\mathbb{P} \leq \varepsilon.$$ □

Theorem 2.7.5. *Let $\varphi(x) \geq 0$ and $\frac{\varphi(x)}{x} \to \infty$ as $x \to \infty$. If*

$$\sup_{i \in I} \mathbb{E}\left[\varphi(X_i)\right] < \infty,$$

then $(X_i, i \in I)$ is UI.

Proof. Let

$$\varepsilon_M := \sup\left\{\frac{x}{\varphi(x)} : x \geq M\right\},$$

so $\varepsilon_M \to 0$ as $M \to \infty$.

$$\begin{aligned}\mathbb{E}\left[|X_i|, |X_i| \geq M\right] &= \mathbb{E}\left[\frac{|X_i|}{\varphi(X_i)}\varphi(X_i), |X_i| \geq M\right] \\ &\leq \varepsilon_M \mathbb{E}[\varphi(X_i)] \leq C\varepsilon_M \to 0.\end{aligned}$$

So it is UI. $\square$

Remark 2.7.6. In particular, for $p > 1$, $\varphi(x) = |x|^p$ is valid. So if

$$\sup_i \|X_i\|_p < \infty$$

for some $p > 1$, $(X_i, i \in I)$ is UI.

Theorem 2.7.7. *$(X_i, i \in I)$ is UI if and only if it satisfies the following two conditions:*

(1) $\sup_{i \in I} \mathbb{E}\left[|X_i|\right] < \infty$,

(2) for any $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that for any $E \in \mathcal{F}$,

$$\mathbb{P}(E) \leq \delta \quad \Rightarrow \quad \int_E |X_i| \, d\mathbb{P} \leq \varepsilon, \ \forall \ i \in I.$$

Proof. First, assume $(X_i, i \in I)$ is UI. Then for (i), there exists $M > 0$ such that $\sup_{i \in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i| \geq M\}}\right] < 1$

$$\sup_{i \in I} \mathbb{E}\left[|X_i|\right] = \sup_{i \in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i| < M\}}\right] + \sup_{i \in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i| \geq M\}}\right] \leq M + 1 < \infty.$$

For (ii), choose $M > 0$ such that $\sup_{i \in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i| \geq M\}}\right] \leq \varepsilon/2$. Then for $\delta = \varepsilon/(2M)$ and $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq \delta$,

$$\begin{aligned}\mathbb{E}\left[|X_i| \mathbb{I}_A\right] &= \mathbb{E}\left[|X_i| \mathbb{I}_{A \cap \{|X_i| < M\}}\right] + \mathbb{E}\left[|X_i| \mathbb{I}_{A \cap \{|X_i| \geq M\}}\right] \\ &\leq M\mathbb{P}(A) + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Conversely, for $M = C/\delta$, by Markov's inequality

$$\mathbb{P}(|X_i| \geq M) \leq \frac{1}{M}\mathbb{E}\left[|X_i|\right] \leq \delta.$$

So

$$\sup_{i \in I} \mathbb{E}\left[|X_i| \mathbb{I}_{\{|X_i| \geq M\}}\right] \leq \varepsilon.$$ $\square$

Theorem 2.7.8. *Let (X_n) be a sequence of L^1 random variables. If $X_n \to X$ in probability, then TFAE*

(1) (X_n) *is UI,*

(2) $X_n \to X$ *in* L^1*,*

(3) $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$.

Proof. (3) $\Rightarrow$ (2): Let $Y_n = |X_n|$, $Y = |X|$, and

$$Z_n = Y_n + Y - |X_n - X| \geq 0.$$

By (3), $Z_n \to 2Y$ in measure. Then by Fatou's lemma,

$$2\mathbb{E}[Y] \leq \liminf_n \mathbb{E}[Z_n] \leq 2Y - \limsup_n \mathbb{E}[|X_n - X|]$$

Therefore,

$$\limsup_n \mathbb{E}[|X_n - X|] \to 0$$

(2) $\Rightarrow$ (1):

$$\begin{aligned}\mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] &\leq \mathbb{E}\left[|X_n - X|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] + \mathbb{E}\left[|X|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] \\ &\leq \mathbb{E}\left[|X_n - X|\right] + \mathbb{E}\left[|X|\,\mathbb{I}_{\{|X_n|\geq M\}}\right]\end{aligned}$$

$X_n \to X$ in L^1, so $\mathbb{E}[|X_n - X|] \leq \varepsilon/2$ for sufficiently large $n > N_0$. For the other term, by Chebyshev's Inequality

$$\mathbb{P}(\{|X_n| \geq M\}) \leq \frac{\mathbb{E}[|X_n|]}{M} \leq \frac{C}{M}$$

because X_n is convergent in L^1. So for $M > \bar{M}$, $\mathbb{P}(\{|X_n| \geq M\}) < \delta$ and $\mathbb{E}\left[|X|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] \leq \varepsilon/2$. For $n = 0, 1, \cdots, N_0$, choose M_n such that $\mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] \leq \varepsilon$. Then let $M^* = \max\left\{M_0, \cdots, M_{N_0}, \bar{M}\right\}$. Then for all n,

$$\mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] \leq \varepsilon, \quad \forall\, M > M^*.$$

(1) $\Rightarrow$ (3): By Fatou's Lemma, $\mathbb{E}[|X|] \leq \liminf_n \mathbb{E}[|X_n|]$. So it suffices to prove that

$$\limsup_n \mathbb{E}[|X_n|] \leq \mathbb{E}[|X|].$$

For any $\varepsilon > 0$, by UI, there exists M such that

$$\mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] < \varepsilon.$$

Then

$$\begin{aligned}\mathbb{E}[|X_n|] &= \mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|<M\}}\right] + \mathbb{E}\left[|X_n|\,\mathbb{I}_{\{|X_n|\geq M\}}\right] \\ &\leq \mathbb{E}\left[(|X_n - X|)\mathbb{I}_{\{|X_n|<M\}}\right] + \mathbb{E}\left[|X|\right] + \varepsilon,\end{aligned}$$

which follows that

$$\limsup_n \mathbb{E}[|X_n|] \leq \varepsilon + \mathbb{E}\left[|X|\right] + \limsup_n \mathbb{E}\left[(|X_n - X|)\mathbb{I}_{\{|X_n|<M\}}\right]$$

Because $X_n \to X$ in probability, by DCT,

$$\limsup_n \mathbb{E}\left[(|X_n - X|)\mathbb{I}_{\{|X_n|<M\}}\right] = 0.$$

Therefore,

$$\limsup_n \mathbb{E}[|X_n|] \leq \mathbb{E}\left[|X|\right].$$ □

Example 2.7.9 (Random Walk). Let $S_0 = 1$ and $S_n = S_0 + \xi_1 + \cdots + \xi_n$, where ξ_i are i.i.d with $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi = -1) = 1/2$. Define

$$N = \inf \{n \colon S_n = 0\} .$$

that is a stopping time. Because (S_n) is a martingale, $(S_{n\wedge N})$ is also a martingale. Moreover, $S_{n\wedge N} \geq 0$. Then by the Martingale Convergence Theorem, $X_n = S_{n\wedge N} \to 0$ a.e.. However,

$$\mathbb{E}[X_n] = \mathbb{E}[S_{n\wedge N}] = \mathbb{E}[S_0] = 1,$$

so (X_n) is not UI.

Theorem 2.7.10. *For a submartingale (X_n), TFAE*

(1) (X_n) is UI.

(2) $X_n \to X$ in L^1 and a.e..

(3) $X_n \to X$ in L^1.

Furthermore, if (X_n) is a martingale, then $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$.

Proof. It suffices to prove (1) ⇒ (2). If (X_n) is UI, then

$$\sup_n \mathbb{E}[|X_n|] < \infty.$$

So by the Martingale Convergence Theorem, $X_n \to X$ a.e.. Then by above theorem, $X_n \to X$ in L^1.
For martingale, for any $n \leq k$ and $A \in \mathcal{F}_n$

$$\mathbb{E}\left[X_k \mathbb{I}_A\right] = \mathbb{E}[\mathbb{I}_A \mathbb{E}[X_k \mid \mathcal{F}_n]] = \mathbb{E}[X_n \mathbb{I}_A]$$

Therefore, for any $A \in \mathcal{F}_n$, because $X_n \to X$ in L^1, by DCT,

$$\mathbb{E}[X \mathbb{I}_A] = \mathbb{E}[X_n \mathbb{I}_A].$$

Then by the uniqueness of conditional expectation, $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. □

Theorem 2.7.11 (Lévy's upward theorem). *Suppose a sequence of σ-fields $\mathcal{F}_n \uparrow \mathcal{F}_\infty = \sigma(\cup \mathcal{F}_n)$. Assume $\mathbb{E}[|X|] < \infty$. Then*

$$\mathbb{E}[X \mid \mathcal{F}_n] \to \mathbb{E}[X \mid \mathcal{F}_\infty]$$

in L^1 and a.e..

Proof. Let $Y_n = \mathbb{E}[X \mid \mathcal{F}_n]$. Then Y_n is a martingale w.s.t. $(\mathcal{F}_n)$ and (Y_n) is UI. So

$$Y_n \to Y_\infty$$

in L^1 and a.e.. It suffices to prove $Y_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$. First, it is obvious that Y_∞ is $\mathcal{F}_\infty$-measurable. For any n and any $A \in \mathcal{F}_n \subset \mathcal{F}_m$ $(m > n)$,

$$\mathbb{E}[Y_m \mathbb{I}_A] = \mathbb{E}[X \mathbb{I}_A].$$

As $m \to \infty$, because $Y_n \to Y_\infty$ in L^1, $\mathbb{E}[Y_\infty \mathbb{I}_A] = \mathbb{E}[X \mathbb{I}_A]$. By the uniqueness of conditional expectation, $Y_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$. □

Theorem 2.7.12. *Suppose a sequence of σ-fields $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Assume that $Y_n \to Y$ a.e.. If $|Y_n| \le Z$ for some $Z \in L^1$, then*

$$\mathbb{E}[Y_n \mid \mathcal{F}_n] \to \mathbb{E}[Y \mid \mathcal{F}_\infty]$$

a.e..

Proof. Let $W_N = \sup\{|Y_n - Y_m| : n, m \ge N\}$. So (W_N) is decreasing to 0 and $|W_N| \le 2Z$. Then

$$\mathbb{E}[W_N \mid \mathcal{F}_\infty] \to 0.$$

For any N, by Fatou's Lemma and above theorem

$$\begin{aligned}\limsup_n \mathbb{E}\left[|Y_n - Y| \mid \mathcal{F}_n\right] &\le \limsup_n \liminf_m \mathbb{E}\left[|Y_n - Y_m| \mid \mathcal{F}_n\right] \\ &\le \limsup_n \mathbb{E}[W_N \mid \mathcal{F}_n] = \mathbb{E}[W_N \mid \mathcal{F}_\infty] \to 0, \text{ as } N \to \infty\end{aligned}$$

Therefore,

$$\lim_n \mathbb{E}\left[Y_n \mid \mathcal{F}_n\right] = \lim_n \mathbb{E}\left[Y \mid \mathcal{F}_n\right] = \mathbb{E}\left[Y \mid \mathcal{F}_\infty\right].$$

□

Example 2.7.13. Let (Y_n) and (Z_m) be independent random variables with the same distribution

$$\mathbb{P}(Y_n = 1) = \frac{1}{n}, \quad \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n},$$
$$\mathbb{P}(Z_n = n) = \frac{1}{n}, \quad \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}.$$

Let $X_n = Z_n Y_n$. Then $\mathbb{P}(X_n \ge 0) = 1/n^2$. By the Borel–Cantelli lemma, $X_n \to 0$ a.e.. Moreover,

$$\mathbb{E}[X_n \mathbb{I}_{\{X_n \ge 1\}}] = \frac{1}{n} \to 0,$$

which means (X_n) is UI. Let $\mathcal{F}_n = \sigma(Y_0, \cdots, Y_n,)$.

$$\mathbb{E}[X_n \mid \mathcal{F}_n] = Y_n \mathbb{E}[Z_n] = Y_n$$

But by the Borel–Cantelli lemma, Y_n does not converges to 0.

2.8 Backward Martingale

Fix $n \le 0$, let $(\mathcal{F}_n)_{n \le 0}$ be a family of decreasing σ-field as $n \to -\infty$.

Definition 2.8.1 (Backward Martingale)**.** We say a stochastic process $(X_n)_{n \le 0}$ is a backward martingale w.s.t. $(\mathcal{F}_n)_{n \le 0}$ if

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n, \quad n \le -1.$$

Moreover, "$\ge$" is a backward submartingale and "$\le$" is a backward supermartingale.

Theorem 2.8.2. *If $(X_n)_{n \le 0}$ is a backward martingale, then*

$$X_{-\infty} = \lim_{n \to -\infty} X_n$$

a.e. and in L^1.

Proof. Let $U_n(a,b)$ be the number of up-crossing of $X_n, \cdots, X_{-1}, X_0$ on $[a,b]$. As before,

$$\mathbb{E}[U_n(a,b)] \le \frac{1}{b-a}\mathbb{E}[(X_0-a)^-]$$

Therefore, similarly, we always have

$$X_{-\infty} = \lim_{n\to-\infty} X_n$$

a.e. by the Martingale Convergence Theorem. By the backwark martingale property,

$$X_n = \mathbb{E}[X_0 \mid \mathcal{F}_n].$$

So $(X_n)_{n\le 0}$ is UI and it implies that $X_{-\infty} = \lim_{n\to-\infty} X_n$ in L^1. □

Theorem 2.8.3. *Let $(X_n)_{n\le 0}$ be a backward martingale. Let $\mathcal{F}_{-\infty} = \cap \mathcal{F}_n$. Then*

$$X_{-\infty} = \lim_{n\to-\infty} X_n = \mathbb{E}[X_0 \mid \mathcal{F}_{-\infty}].$$

Proof. First, because X_n is $\mathcal{F}_{-\infty}$-measurable, $X_{-\infty}$ is $\mathcal{F}_{-\infty}$-measurable. For any $A \in \mathcal{F}_{-\infty}$,

$$\mathbb{E}[X_{-\infty}\mathbb{I}_A] = \lim_{n\to-\infty} \mathbb{E}[X_n\mathbb{I}_A] = \lim_{n\to-\infty} \mathbb{E}[\mathbb{I}_A\mathbb{E}[X_0 \mid \mathcal{F}_n]] = \mathbb{E}[X_0\mathbb{I}_A].$$ □

Remark 2.8.4. If $(X_n)_{n\le 0}$ is a backward submartingale with $\sup_n \mathbb{E}[|X_n|] < \infty$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = X_{-\infty},$$

a.e. and in L^1 and

$$X_{-\infty} \le \mathbb{E}[X_m \mid \mathcal{F}_{-\infty}], \quad \forall\, m \in -\mathbb{N}_0.$$

Theorem 2.8.5 (Lévy's downward theorem)**.** *If $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ as $n \to -\infty$, then for any $Y \in L^1$,*

$$\mathbb{E}[Y \mid \mathcal{F}_n] \to \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]$$

a.e. and in L^1.

Proof. Let $X_n = \mathbb{E}[Y \mid \mathcal{F}_n]$ that is a backward martingale. Then by above

$$X_n \to \mathbb{E}[X_0 \mid \mathcal{F}_{-\infty}] = \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]$$

a.e. and in L^1. □

Example 2.8.6 (Strong Law of Large Number)**.** Let $\xi_1, \xi_2, \cdots$ be a sequence of i.i.d. L^1 random variables. Define

$$X_{-n} = \frac{S_n}{n}, \quad S_n = \xi_1 + \cdots + x_n.$$

Let $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \cdots)$. By symmetry, for any $j, k \le n+1$,

$$\mathbb{E}[\xi_k \mid \mathcal{F}_{-n-1}] = \mathbb{E}[\xi_j \mid \mathcal{F}_{-n-1}]$$

It follows that

$$\mathbb{E}[\xi_{n+1} \mid \mathcal{F}_{-n-1}] = \frac{1}{n+1}\sum_{k=1}^{n+1} \mathbb{E}[\xi_k \mid \mathcal{F}_{-n-1}] = \frac{1}{n+1}\mathbb{E}[S_{n+1} \mid \mathcal{F}_{-n-1}] = \frac{S_{n+1}}{n+1}$$

So X_{-n} is a backward martingale. So

$$\frac{S_n}{n} \to \mathbb{E}[X_0 \mid \mathcal{F}_{-\infty}] = \mathbb{E}[X_0].$$

2.9 Doob's Optional Stopping Theorem

Theorem 2.9.1. *If $X = (X_n)$ is a UI submartingale, then for any stopping time N, $(X_{n\wedge N})$ is also UI.*

Proof. Note that (X_n^+) is also a submartingale. So $(X_{n\wedge N}^+)$ is a submartingale. Moreover, because $n \wedge N$ is a bounded stopping time

$$\mathbb{E}\left[X_{n\wedge N}^+\right] \leq \mathbb{E}\left[X_n^+\right],$$

and because (X_n) is UI,

$$\sup_n \mathbb{E}\left[X_{n\wedge N}^+\right] \leq \sup_n \mathbb{E}\left[X_n^+\right] = \sup_n \mathbb{E}[|X_n|] < \infty$$

Then by the Martingale Convergence Theorem,

$$X_{n\wedge N} \to X_N, \ a.e.$$

On the other hand,

$$\mathbb{E}\left[X_{N\wedge n}^-\right] = \mathbb{E}\left[X_{N\wedge n}^+\right] - \mathbb{E}\left[X_{N\wedge n}\right] \leqslant \mathbb{E}\left[X_{N\wedge n}^+\right] - \mathbb{E}\left[X_0\right].$$

So

$$\sup_n \mathbb{E}\left[X_{N\wedge n}^-\right] \leqslant \sup_n \mathbb{E}\left[X_{N\wedge n}^+\right] - \mathbb{E}\left[X_0\right] < +\infty.$$

It follows that $\sup_n \mathbb{E}[|X_{n\wedge N}|] < \infty$. So by Fatou's Lemma,

$$\mathbb{E}[|X_N|] \leq \liminf_n \mathbb{E}[|X_{n\wedge N}|] < \infty$$

To verify the uniform integrability,

$$\begin{aligned}\mathbb{E}\left[|X_{n\wedge N}|, |X_{n\wedge N}| \geq K\right] &= \mathbb{E}\left[|X_N|, |X_N| \geq K, N \leq n\right] + \mathbb{E}\left[|X_n|, |X_n| \geq K, N > n\right] \\ &\leq \mathbb{E}\left[|X_N|, |X_N| \geq K\right] + \mathbb{E}\left[|X_n|, |X_n| \geq K\right]\end{aligned}$$

Because $X_N \in L^1$ and (X_n) is UI, for any $\varepsilon > 0$, it can find K_1, K_2 such that

$$\mathbb{E}\left[|X_N|, |X_N| \geq K_1\right], \ \mathbb{E}\left[|X_n|, |X_n| \geq K_2\right] \leq \frac{1}{2}\varepsilon.$$

Therefore, $(X_{n\wedge N})$ is also UI. □

Remark 2.9.2. For the positive and negative part, because $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$, we know $|f| = f^+ + f^-$ and

$$(-f)^- = f^+, \quad (-f)^+ = f^-$$

Moreover, if $f \leq g$, then

$$f^+ \leq g^+, \quad f^- \geq g^-.$$

Theorem 2.9.3 (Doob's Optional Theorem)**.** *Suppose (X_n) is a submartingale and N is a stopping time. If $X_N \in L^1$ and $(X_n \mathbb{I}_{N>n})$ is UI, then $(X_{n\wedge N})$ is UI and*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N].$$

Proof. The uniform integrability is directly obtained by the proof of above theorem. Because of it,

$$X_{n\wedge N} \to X_N$$

a.e. and in L^1. Moreover, because

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{n\wedge N}],$$

as $n \to \infty$, $\mathbb{E}[X_0] \leq \mathbb{E}[X_N]$. □

Theorem 2.9.4. *If (X_n) is a UI submartingale, then for any stopping time N, we have*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_\infty],$$

where $X_\infty = \lim_n X_n$.

Proof. First, we have $X_\infty = \lim_n X_n$ a.e. and in L^1. Fix any n,

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{n\wedge N}] \leq \mathbb{E}[X_n]$$

By above theorem, as $n \to \infty$, we have

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_\infty].$$ □

Corollary 2.9.5. *If (X_n) is a UI submartingale and $M \leq N$ are two stopping times, then*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_M] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_\infty]$$

Proof. Consider the submartingale $Y_n = X_{N\wedge n}$ that is a UI submartingale. By applying above theorem to (Y_n),

$$\mathbb{E}[Y_0] = \mathbb{E}[X_0] \leq \mathbb{E}[Y_M] = \mathbb{E}[X_M] \leq \mathbb{E}[Y_\infty] = \mathbb{E}[X_N].$$ □

Theorem 2.9.6. *Suppose (X_n) is a submartingale with $\sup_n \mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \leq B$ for a constant B. If N is a stopping time with $\mathbb{E}[N] < \infty$, then $(X_{N\wedge n})$ is UI and so we have $\mathbb{E}[X_0] \leq \mathbb{E}[X_N]$.*

Proof. Observe that

$$\begin{aligned}
|X_{N\wedge n}| &= \left| X_0 + \sum_{m=0}^{N\wedge n-1} (X_{m+1} - X_m) \right| \\
&\leq |X_0| + \sum_{m=0}^{N-1} |X_{m+1} - X_m| = |X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| \mathbb{I}_{\{m\leq N-1\}} =: Y
\end{aligned}$$

It suffices to prove $Y \in L^1$. Note that $\{N \geq m+1\} = \{N < m\}^c \in \mathcal{F}_m$

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[|X_0|] + \sum_{m=0}^{\infty} \mathbb{E}\left[|X_{m+1} - X_m| \mathbb{I}_{\{N\geq m+1\}}\right] \\
&= \mathbb{E}[|X_0|] + \sum_{m=0}^{\infty} \mathbb{E}\left[\mathbb{I}_{\{N\leq m+1\}} \mathbb{E}\left[|X_{m+1} - X_m| \mid \mathcal{F}_m\right]\right] \\
&\leq \mathbb{E}[|X_0|] + B \sum_{m=0}^{\infty} \mathbb{P}\left(\{N \leq m+1\}\right) = \mathbb{E}[|X_0|] + B\mathbb{E}[N] < \infty
\end{aligned}$$

So $Y \in L^1$ and $(X_{N\wedge n})$ is UI. □

Example 2.9.7 (Gambler's Ruin Problem)**.** Consider A, B play a series of games against each other in which a fair coin in tossed respectively. In each game, gambler A wins or losses 1 dollar with probability $1/2$ and $1/2$. The initial capital of gambler A is a dollars, and that of gambler B is b dollars. They continue play until one of them is ruined. Determine the probability of that A will be ruined and the expected number of games.
Solution: Let $\hat{S}_n$ be the fortune if A after n-th games, so

$$\hat{S}_n = a + X_1 + \cdots + X_n = a + S_n,$$

where X_i are i.i.d. $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. The game will stop of

$$T = \min\{n\colon S_n = -a \text{ or } S_n = b\}$$

that is a stopping time. Then

$$\{\text{ Gambler } A \text{ is ruined }\} = \{S_T = -a\}.$$

So it needs to find $\mathbb{P}(\{S_T = -a\})$ and $\mathbb{E}[T]$.
We already know (S_n) is a martingale with $S_0 = 0$, so is the stopping process $(S_{T\wedge n})$. Moreover, because

$$|S_{T\wedge n}| \leq a + b,$$

$(S_{T\wedge n})$ is UI. So

$$\mathbb{E}[S_T] = \mathbb{E}[S_0] = 0$$

Note that

$$\mathbb{E}[S_T] = -a\mathbb{P}(S_T = -a) + b\mathbb{P}(S_T = b) = 0$$

Moreover,

$$\mathbb{P}(S_T = -a) + \mathbb{P}(S_T = b) = 1.$$

So

$$\mathbb{P}(S_T = -a) = \frac{b}{a+b}, \quad \mathbb{P}(S_T = b) = \frac{a}{a+b}.$$

First, we need to check $\mathbb{E}[T] < \infty$. By induction it can have

$$\mathbb{P}(T > m(a+b)) \leqslant \left(1 - \left(\frac{1}{2}\right)^{a+b}\right)^m, m \geqslant 1$$

So

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}\left[T\mathbb{I}_{\{T\leqslant a+b\}}\right] + \mathbb{E}\left[T\mathbb{I}_{\{T>a+b\}}\right] \\
&= \mathbb{E}\left[T\mathbb{I}_{\{T\leqslant a+b\}}\right] + \sum_{m=1}^{\infty}\mathbb{E}\left[T\mathbb{I}_{\{m(a+b)<T\leqslant(m+1)(a+b)\}}\right] \\
&\leqslant a+b+\sum_{m=1}^{\infty}(m+1)(a+b)\mathbb{P}(m(a+b)<T\leqslant(m+1)(a+b)) \\
&\leqslant a+b+\sum_{m=1}^{\infty}(m+1)(a+b)\mathbb{P}(T>m(a+b)) \\
&\leqslant a+b+\sum_{m=1}^{\infty}(m+1)(a+b)\left(1-\left(\frac{1}{2}\right)^{a+b}\right)^m < +\infty
\end{aligned}$$

Note that $(Y_n = S_n^2 - n)$ is also a martingale. So by the bounded optional stopping time theorem

$$\mathbb{E}\left[S_{T\wedge n}^2 - T\wedge n\right] = \mathbb{E}[S_0^2] = 0 \;\Rightarrow\; \mathbb{E}\left[S_{T\wedge n}^2\right] = \mathbb{E}\left[T\wedge n\right]$$

Because $|S_{T\wedge n}^2| \leq (a+b)^2$, by DCT, as $n \to \infty$,

$$\mathbb{E}[T] = \mathbb{E}\left[S_T^2\right] = ab.$$

Example 2.9.8 (Random Walk). $\xi_1, \cdots, \xi_n \cdots$ is i.i.d with $\mathbb{P}(\xi = 1) = p$ and $\mathbb{P}(\xi = -1) = q$. $S_0 = k < N$ and $S_n = S_0 + \xi_1 + \cdots + \xi_n$. Find the probability that the random walk hints 0 before N.
Solution: Let

$$T = \inf\{n \colon S_n = 0 \text{ or } S_n = N\}.$$

Define $Z_n = \left(\frac{q}{p}\right)^{S_n}$ that can be proved a martingale w.s.t. $\mathcal{F}_n = \sigma(S_0, \cdots, S_n)$. Therefore, $(Z_{T\wedge n})$ is also a martingale and it is UI because $|S_T| < \infty$. Then

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = \left(\frac{q}{p}\right)^k$$

Note that

$$\mathbb{E}[Z_T] = \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_T}\right] = \mathbb{P}(S_T = 0) + \left(\frac{q}{p}\right)^N \mathbb{P}(S_T = N)$$

Combining it with $\mathbb{P}(S_T = N) + \mathbb{P}(S_T = 0) = 1$, we have

$$\mathbb{P}(S_T = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}.$$

Example 2.9.9. Let (X_n) be i.i.d. with $\mathbb{E}[X_n] = \mu$ and $N \in L^1$ be a stopping time w.s.t. $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. How to calculate $\mathbb{E}\left[\sum_{i=1}^N X_i\right]$.
Solution: Let $Y_0 = 0$ and

$$Y_n = \sum_{i=1}^n X_i - n\mu$$

that is clear a martingale. Then $(Y_{N\wedge n})$ is also a martingale. Because

$$\mathbb{E}\left[|Y_{n+1} - Y_n| \mid \mathcal{F}_n\right] \le \mu + \mathbb{E}\left[|X_{n+1}| \mid \mathcal{F}_n\right] = \mu + \mathbb{E}\left[|X_{n+1}|\right] = 2\mu < \infty,$$

$(Y_{N\wedge n})$ is UI and so

$$\mathbb{E}\left[Y_N\right] = \mathbb{E}\left[\sum_{i=1}^N X_i - N\mu\right] = \mathbb{E}[Y_0] = 0.$$

It follows that

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}\left[N\right]\mu.$$

Chapter 3

Discrete Time Markov Theory

3.1 Markov Chain

Definition 3.1.1 (Markov Chain)**.** If the state space is at most countable, a stochastic process $(X_n)_{n\geq 0}$ is said to have Markov property if for any n and any $i_0, \cdots, i_{n-1}, i, j \in S$

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

and $(X_n)_{n\geq 0}$ is called a Markov chain. Furthermore,

$$p_{ij}(n) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

is called transition probability. In particular, if $p_{ij}(n) \equiv p_{ij}$, such Markov chain is called time-homogeneous, otherwise, it is called time-inhomogeneous.

In following, we mainly consider the time-homogeneous Markov chain.

Example 3.1.2 (Ehrenfest Chain)**.** Let A, B be two bottles such that A contains k balls and B contains $r - k$ balls. Each operation is to randomly choose a ball from the r balls and then transfer it from its original bottle into another one. Let X_n be the number of balls in A after n-th operation. Note that the state space $S = \{0, 1, \cdots, r\}$. Then

$$p_{kj} = \mathbb{P}\left(X_{n+1} = j \mid X_n = k\right) = \begin{cases} 0 & , |k - j| \neq 1 \\ \frac{k}{r} & , j = k + 1 \\ \frac{r-k}{r} & , j = k - 1 \end{cases}.$$

3.2 Constructing Markov Chain

Definition 3.2.1 (Transition Probability)**.** Let $(S, \mathcal{S})$ be the state space (measurable space). A function

$$p \colon S \times \mathcal{S} \to \mathbb{R}$$

is called a transition probability if

(i) For any $x \in S$, $A \to p(x, A)$ is a probability measure defined on $\mathcal{S}$,

(ii) For any $A \in \mathcal{S}$, $x \to p(x, A)$ is a measurable function.

(X_n) is a Markov chain with the transition probability p if

$$\mathbb{P}(X_{n+1} \in B \mid \mathcal{F}_n) = p(X_n, B)$$

When given a transition probability p, let

$$\mathbb{P}_n(B_0 \times \cdots \times B_n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \int_{B_2} p(x_1, dx_2) \cdots \int_{B_n} p(x_{n-1}, dx_n)$$

for $B_0, \cdots, B_n \in \mathcal{S}$. Then $\mathbb{P}_n$ on $(S^{n+1}, \mathcal{S}^{n+1})$ are consistent, i.e.,

$$\mathbb{P}_{n+1}(B_0 \times \cdots B_n \times S) = \mathbb{P}_n(B_0 \times \cdots \times B_n).$$

So by Kolmogorov Extension Theorem, there exists a measure $\mathbb{P}_\mu$ on $(S^\infty, \mathcal{S}^\infty)$ such that it is the finite dimensional distribution of the coordinate process $(X_n)_{n\geq 0}$ on $(S^\infty, \mathcal{S}^\infty)$, i.e. $X_n(\omega) = \omega_n$ for $\omega \in S^\infty$,

$$\mathbb{P}_\mu(X_0 \in B_0, \cdots, X_n \in B_n) = \mathbb{P}_n(B_0 \times \cdots \times B_n)$$

In fact, such $(X_n)_{n\geq 0}$ is a Markov chain with transition probability p, where $\mathcal{F}_n = \mathcal{S}^n$, i.e.

$$\mathbb{P}_\mu(X_{n+1} \in B \mid \mathcal{F}_n) = \mathbb{E}\left[\mathbb{I}_{\{X_{n+1}\in B\}} \mid \mathcal{F}_n\right] = p(X_n, B)$$

Proof. First, clearly $p(X_n, B)$ is $\mathcal{F}_n$-measurable. So it suffices to check for any $A \in \mathcal{F}_n = \mathcal{S}^n$,

$$\mathbb{E}\left[\mathbb{I}_{\{X_{n+1}\in B\}}\mathbb{I}_A\right] = \int_A p(X_n, B) d\mathbb{P}_\mu.$$

Because $\mathcal{F}_n = \sigma(X_0, \cdots, X_n)$, we can assume

$$A = \{X_0 \in B_0, \cdots, X_n \in B_n\}$$

for $B_i \in \mathcal{S}$. So

$$\begin{aligned}
\mathbb{E}\left[\mathbb{I}_{\{X_{n+1}\in B\}}\mathbb{I}_A\right] &= \mathbb{P}_\mu\left(X_0 \in B_0, \cdots, X_n \in B_n, \{X_{n+1} \in B\}\right) \\
&= \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n) \int_B p(x_n, dx_{n+1}) \\
&= \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_n, B) p(x_{n-1}, dx_n)
\end{aligned}$$

In fact, we can prove that for any measurable function f

$$\int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} f(x_n) p(x_{n-1}, dx_n) = \int_A f(X_n) d\mathbb{P}_\mu.$$

By approximation, it can assume $f = \mathbb{I}_C$ for some $C \in \mathcal{S}$. Then

$$\begin{aligned}
\text{LHS} &= \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_{n-1}} p(x_{n-1}, B_n \cap C) p(x_{n-2}, dx_{n-1}) \\
&= \mathbb{P}_\mu(X_0 \in B_0, \cdots, X_{n-1} \in B_{n-1}, X_n \in B_n \cap C).
\end{aligned}$$

and

$$\begin{aligned}
\int_A f(X_n) d\mathbb{P}_\mu &= \int_A \mathbb{I}_{\{X_n\in C\}} d\mathbb{P}_\mu \\
&= \int_S \mathbb{I}_{\{X_n\in C\}\cap A} d\mathbb{P}_\mu \\
&= \mathbb{P}_\mu(X_0 \in B_0, \cdots, X_{n-1} \in B_{n-1}, X_n \in B_n \cap C).
\end{aligned}$$

So LHS = RHS. □

Let $\Omega_0 = S^\infty$ with $\mathcal{F}_\infty = \mathcal{S}^\infty$.

Definition 3.2.2 (Shift Operator). For any $n \in \mathbb{N}$, define

$$\theta_n \colon \Omega_0 \to \Omega_0$$

by

$$\theta_n(\omega) = (\omega_n, \omega_{n+1}, \cdots)$$

for $\omega = (\omega_0, \omega_1, \cdots) \in \Omega_0$.

Proposition 3.2.3. *If $(X_n)_{n\geq 0}$ is a Markov chain with transition probability p, then for any bounded measurable function f on $(S, \mathcal{S})$, we have*

$$\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \int_S f(y)p(X_n, dy).$$

Proof. It is clear true for $f = \mathbb{I}_C$ and so for any simple function f. Therefore, by the following Monotone Class Theorem, it is true for any bounded measurable functions. $\square$

Theorem 3.2.4 (Monotone Class Theorem). *Let $\mathcal{A} \subset \mathcal{P}(S)$ be a π-system (i.e. closed under intersection) that contains S. Let $\mathcal{H}$ be a collection of real-valued functions satisfying*

(1) if $A \in \mathcal{A}$, then $\mathbb{I}_A \in \mathcal{H}$,

(2) if $f, g \in \mathcal{H}$, then $f + g, cf \in \mathcal{H}$ for any real c,

(3) if $f_n \in \mathcal{H}$ are nonnegative and $f_n \uparrow f$ for a bounded measurable f, then $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{A})$-measurable functions.

Theorem 3.2.5 (Markov Property). *Given the μ on S, the corresponding $\mathbb{P}_\mu$ on Ω_0, and the Markov chain $(X_n)_{n\geq 0}$. Let $Y \colon \Omega_0 \to \mathbb{R}$ be a bounded and $\sigma(X_0, \cdots, X_n, \cdots)$-measurable random variable. Then*

$$\mathbb{E}_\mu[Y \circ \theta_m \mid \mathcal{F}_m] = \mathbb{E}_{X_m}[Y],$$

where $\mathbb{E}_\mu$ is the expectation w.s.t. $\mathbb{P}_\mu$ and $\mathbb{E}_{X_m} = \mathbb{E}_x|_{x=X_m}$. $\mathbb{E}_x$ is the expectation w.s.t. $\mathbb{P}_x$, where $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

Proof. By the Monotone Class Theorem, we can assume

$$Y = \prod_{k=0}^{n} g_k(X_k),$$

where g_i is bounded measurable function on $(S, \mathcal{S})$. Because $\mathbb{E}_{X_m}[Y]$ is a function of X_m, it is clear $\mathcal{F}_m$-measurable. For $A \in \mathcal{F}_m$, it suffices to check

$$\mathbb{E}_\mu \left[Y \circ \theta_m \cdot \mathbb{I}_A\right] = \mathbb{E}_\mu \left[\mathbb{E}_{X_m}[Y] \cdot \mathbb{I}_A\right].$$

It can assume

$$A = \left\{X_0 \in A_1, \cdots, X_m \in A_m\right\},$$

Note that $X_k(\theta(\omega)) = X_{k+m}(\omega)$. So

$$\begin{aligned}
\mathbb{E}_\mu \left[Y \circ \theta_m \cdot \mathbb{I}_A\right] &= \mathbb{E}_\mu \left[g_0\left(X_m\right) \cdots g_n\left(X_{m+n}\right) \mathbb{I}_{A_0}\left(X_0\right) \cdots \mathbb{I}_{A_m}\left(X_m\right)\right] \\
&= \int_{A_0} \mu(dx_0) \int_{A_1} p(x_0, dx_1) \cdots \int_{A_m} g_0(x_m) p(x_{m-1}, dx_m) \\
&\quad \int_S g_1(x_{m+1}) p(x_m, dx_{m+1}) \cdots \int_S g_n(x_{m+n}) p(x_{m+n-1}, dx_{m+n})
\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}_{X_m}[Y] &= \int_S g_0(x_0)\delta_{X_m}(dx_0)\int_S g(x_1)p(x_0,dx_1)\cdots\int_S g(x_n)p(x_{n-1},dx_n)\\ &= g_0(X_m)\int_S g(x_{m+1})p(X_m,dx_{m+1})\cdots\int_S g(x_{m+n})p(x_{m+n-1},dx_{m+n})\end{aligned}$$

by replacing x_i by x_{m+i}. So

$$\begin{aligned}\mathbb{E}_\mu\left[Y\circ\theta_m\cdot\mathbb{I}_A\right] &= \int_{A_0}\mu(dx_0)\int_{A_1}p(x_0,dx_1)\cdots\int_{A_m}\mathbb{E}_{X_m}[Y]p(x_{m-1},dx_m)\\ &= \mathbb{E}_\mu\left[\mathbb{E}_{X_m}[Y]\cdot\mathbb{I}_A\right].\end{aligned}$$ □

Remark 3.2.6. By the definition of $\mathbb{P}_\mu$, for $f_i(x)=\mathbb{I}_{B_i}(x)$,

$$\mathbb{E}_\mu\left[f_0(X_0)f_1(X_1)\cdots f_n(X_n)\right] = \int_S f_0(x_0)\mu(dx_0)\int_S f(x_1)p(x_0,dx_1)\cdots\int_S f_n(x_n)p(x_{n-1},dx_n).$$

Corollary 3.2.7. *We have*

$$\mathbb{E}_\mu[Y\circ\theta_m \mid \mathcal{F}_m] = \mathbb{E}_\mu[Y\circ\theta_m \mid \sigma(X_m)]$$

Proof. Because $\sigma(X_m)\subset\mathcal{F}_m$,

$$\begin{aligned}\mathbb{E}_\mu[Y\circ\theta_m \mid \sigma(X_m)] &= \mathbb{E}_\mu\left[\mathbb{E}_\mu[Y\circ\theta_m \mid \mathcal{F}_m] \mid \sigma(X_m)\right]\\ &= \mathbb{E}_\mu\left[\mathbb{E}_{X_m}[Y] \mid \sigma(X_m)\right]\\ &= \mathbb{E}_{X_m}[Y] = \mathbb{E}_\mu[Y\circ\theta_m \mid \mathcal{F}_m].\end{aligned}$$ □

Remark 3.2.8. For any Markov chain $X=(X_n)_{n\ge0}$ on a space $(\Omega,\mathcal{F})$ and taking values on $(S,\mathcal{S})$, we can still obtain a $\mathbb{P}_\mu$ on $(S^\infty,\mathcal{S}^\infty)$, which is $\mathbb{P}_\mu = X_\#\mathbb{P}$. Or on the other hand, $\mu=(X_0)_\#\mathbb{P}$ and $\mathbb{P}_\mu$ is deduced from such μ. And $(X_n)_{n\ge0}$ on $(\Omega,\mathbb{P})$ is equivalent to the coordinate process $(\pi_n)_{n\ge0}$ on $(\mathcal{S}^\infty,\mathbb{P}_\mu)$. Then the Markov property is described as the above theorem.

Definition 3.2.9. If N is a stopping time w.s.t. $(\mathcal{F}_n)_{n\ge0}$, then define

$$\mathcal{F}_N := \{A\colon A\cap\{N\le n\}\in\mathcal{F}_n,\ \forall\, n\},$$

which is a σ-field.

Remark 3.2.10. (1) If $A\in\mathcal{F}_N$, then for any n,

$$A\cap\{N=n\} = (A\cap\{N\le n\})\backslash(A\cap\{N\le n-1\})\in\mathcal{F}_n.$$

(2) Note that X_N is $\mathcal{F}_N$-measurable, because for any n,

$$\{X_N\in B\}\cap\{N=n\} = \{X_n\in B\}\cap\{N=n\}\in\mathcal{F}_n.$$

For a stopping time N, define $\theta_N\colon\Omega_0\to\Omega_0$ by

$$\theta_N(\omega) = \begin{cases}\theta_n(\omega), & \omega\in\{N=n\},\ n<\infty\\ *, & \omega\in\{N=\infty\},\end{cases}$$

where $*$ is an extra point adding to Ω_0.

Theorem 3.2.11 (Strong Markov Property)**.** *On* $\{N < \infty\}$,

$$\mathbb{E}_\mu \left[Y \circ \theta_N \mid \mathcal{F}_N\right] = \mathbb{E}_{X_N}[Y]$$

Proof. First, because $\mathbb{E}_{X_N}[Y]$ is a function of X_N, it is $\mathcal{F}_N$-measurable. It suffices to check for any $A \in \mathcal{F}_N$,

$$\mathbb{E}_\mu \left[Y \circ \theta_N, A \cap \{N < \infty\}\right] = \mathbb{E}_\mu \left[\mathbb{E}_{X_N}[Y], A \cap \{N < \infty\}\right].$$

Note that

$$\begin{aligned}
\text{LHS} &= \sum_{n=0}^{\infty} \mathbb{E}_\mu \left[Y \circ \theta_n, A \cap \{N = n\}\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}_\mu \left[\mathbb{E}_{X_n}[Y, A \cap \{N = n\}]\right] \\
&= \mathbb{E}_\mu \left[\mathbb{E}_{X_N}[Y, A \cap \{N < \infty\}]\right].
\end{aligned}$$

□

Define

$$p^k(x, y) = \mathbb{P}(X_k = y \mid X_0 = x) = \mathbb{P}_x(X_k = y)$$

where the second equality is by the Markov property.

Theorem 3.2.12 (Chapman-Kolmogorov Equation)**.** *For any* $x, y, z \in S$,

$$p^{m+n}(x, z) = \mathbb{P}_x(X_{m+n} = z) = \sum_{y \in S} \mathbb{P}_x(X_m = y)\mathbb{P}_y(X_n = z) = \sum_{y \in S} p^m(x, y)p^n(y, z).$$

Proof. By the Markov property

$$\begin{aligned}
\mathbb{P}_x(X_{m+n} = z) &= \mathbb{E}_x[\mathbb{I}_{\{X_{m+n}=z\}}] \\
&= \mathbb{E}_x \left[\mathbb{E}_x[\mathbb{I}_{\{X_{m+n}=z\}} \mid \mathcal{F}_m]\right] \\
&= \mathbb{E}_x \left[\mathbb{E}_x[\mathbb{I}_{\{z\}}(X_n \circ \theta_m) \mid \mathcal{F}_m]\right] \\
&= \mathbb{E}_x \left[\mathbb{E}_{X_m}[X_n = z]\right] \\
&= \sum_{y \in S} \mathbb{E}_y[X_n = z]\mathbb{P}_x(X_m = y) \\
&= \sum_{y \in S} \mathbb{P}_y(X_n = z)\mathbb{P}_x(X_m = y)
\end{aligned}$$

□

Remark 3.2.13. For any $x \in S$, by the definition of $\mathbb{P}_x$, we have

$$\mathbb{P}_x(X_1 = y_1, X_2 = y_2, \cdots, X_n = y_n) = p(x, y_1)p(y_1, y_2) \cdots p(y_{n-1}, y_n).$$

Note that it can be also obtained by the property of conditional probability.

3.3 Classification of States

Let $(X_n)_{n\geq 0}$ be a Markov chain with discrete state space S. Let $y \in S$. Define $T_y^0 = 0$ and for any $k \in \mathbb{N}$

$$T_y^k = \inf\left\{n > T_y^{k-1} \colon X_n = y\right\},$$

i.e., the time of the k-th returning to y. Note that T_y^k is a stopping time. For simplicity, let $T_y = T_y^1$. Define

$$\rho_{xy} = \mathbb{P}_x(T_y < \infty),$$

i.e. the probability of the chain that can reach y with starting from x.

Theorem 3.3.1. *The probability*

$$\mathbb{P}_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}.$$

Proof. By induction, for $k = 1$, it is obvious. Assume $k \geq 2$ and it is true for $k-1$. Let $Y(\omega) = 1$ if $X_n(\omega) = y$ for some n, otherwise $Y(\omega) = 0$. So

$$\{Y = 1\} = \{T_y < \infty\}.$$

Let $N = T_y^{k-1}$. Then

$$\{Y \circ \theta_N = 1\} = \left\{T_y^k < \infty\right\}.$$

By the strong Markov property, on $\{N < \infty\}$,

$$\mathbb{E}_x\left[Y \circ \theta_N \mid \mathcal{F}_N\right] = \mathbb{E}_{X_N}\left[Y\right] = \mathbb{E}_y\left[Y\right] = \rho_{yy}$$

because $X_N = y$ and $\mathbb{E}_y\left[Y\right] = \mathbb{P}_y(Y = 1)$. Therefore,

$$\begin{aligned}
\mathbb{P}_x(T_y^k < \infty) &= \mathbb{P}_x\left(N < \infty, Y \circ \theta_N = 1\right) \\
&= \mathbb{E}_x[Y \circ \theta_N = 1, N < \infty] \\
&= \mathbb{E}_x\left[\mathbb{E}_{X_N}\left[Y\right], N < \infty\right] \\
&= \rho_{yy}\mathbb{E}_x[\mathbb{I}_{\{N<\infty\}}] = \rho_{yy}\mathbb{P}_x(T_y^{k-1} < \infty).
\end{aligned}$$

The by assumption of induction, it is true for k. ☐

Definition 3.3.2 (Classification of States)**.** Given Markov chain (X_n) valued on discrete $(S, \mathcal{S})$, let $y \in S$.

(1) y is called recurrent if $\rho_{yy} = 1$.

(2) y is called transient if $\rho_{yy} < 1$. In this case, there is a positive probability $1 - \rho_{yy}$ that the Markov chain starting from y never return y.

For $y \in S$, let

$$N(y) = \sum_{n=1}^{\infty} \mathbb{I}_{\{X_n = y\}}$$

that is the number of visits to y.

Lemma 3.3.3. *If y is transient,*

$$\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1 - \rho_{yy}}.$$

Proof. By definition,

$$\begin{aligned}
\mathbb{E}_x[N(y)] &= \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) \\
&= \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \frac{\rho_{xy}}{1 - \rho_{yy}}.
\end{aligned}$$

☐

Theorem 3.3.4. *$y \in S$ is recurrent if and only if $\mathbb{E}_x[N(y)] = \infty$.*

Proof. If y is recurrent, because

$$\mathbb{E}_x[N(y)] = \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \infty.$$

Conversely, assume $y \in S$ is not recurrent, then

$$\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1-\rho_{yy}} = \infty$$

implies that $\rho_{yy} = 1$, inducing a contradiction. □

Theorem 3.3.5. *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$.*

Proof. Assume $\rho_{yx} < 1$. Let

$$K = \inf\left\{k \colon p^k(x,y) > 0\right\}.$$

$\rho_{xy} > 0$ implies that $K < \infty$ and there is a $y_1, \cdots, y_{K-y}$ such that

$$p(x, y_0)p(y_1, y_2)\cdots p(y_{K-1}, y) > 0,$$

because

$$p^K(x,y) = \sum_{y_1,\cdots,y_{K-1}\in S} \mathbb{P}_x(X_1 = y_1, \cdots, X_{K-1} = y_{K-1}, X_K = y) > 0.$$

Note that $y_i \neq x$ for $i = 1, \cdots, K-1$. Because $\rho_{yx} < 1$,

$$\begin{aligned}\mathbb{P}_x(T_x = \infty) &\geq p(x,y_1)\cdots p(y_{K-1},y)(1-\rho_{yx})\\ &= \mathbb{P}_x(X_1 = y_1, \cdots, X_K = y, T_x \circ \theta_K = \infty) > 0,\end{aligned}$$

contradicting to the recurrence of x. Therefore, $\rho_{yx} = 1$.
To check that y is recurrent, it suffices to prove $\mathbb{E}_y[N(y)] = \infty$. Since $\rho_{yx} = 1 > 0$, there exist an $\ell \in \mathbb{N}$ such that

$$p^\ell(y,x) > 0.$$

Note that for $n \geq 1$, by the Chapman-Kolmogorov Equation,

$$p^{\ell+n+K}(y,y) \geq p^\ell(y,x)p^n(x,x)p^K(x,y).$$

So

$$\sum_{n=1}^{\infty} p^{\ell+n+K}(y,y) \geq p^\ell(y,x)p^K(x,y)\sum_{n=1}^{\infty} p^n(x,x).$$

Moreover,

$$\mathbb{E}_x[N(x)] = \sum_{n=1}^{\infty} \mathbb{E}_x[\mathbb{I}_{X_n=x}] = \sum_{n=1}^{\infty} p^n(x,x) = \infty,$$

because of the recurrence of x. It follows that

$$\mathbb{E}_y[N(y)] = \sum_{n=1}^{\infty} p^n(y,y) \geq \sum_{n=1}^{\infty} p^{\ell+n+K}(y,y) = \infty.$$ □

Definition 3.3.6 (Closedness)**.** Let $C \subset S$. C is called closed if for any $x \in C$, $\rho_{xy} > 0$ implies $y \in C$.

Remark 3.3.7. If C is closed and $x \in C$, then $\mathbb{P}_x(X_n \in C) = 1$ for all n. Otherwise, there is a $y \notin C$ such that

$$\mathbb{P}_x(X_n = y) > 0$$

which implies that $\rho_{xy} \geq \mathbb{P}_x(X_n = y) > 0$ and so $y \in C$, contradicting to the assumption.

Definition 3.3.8 (Irreducibility)**.** $D \subset S$ is called irreducible if for any $x, y \in D$, $\rho_{xy} > 0$.

Theorem 3.3.9. *Assume $C \subset S$ is finite and closed. Then C contains a recurrent state. In particular, if C is also irreducible, then every state in C is recurrent.*

Proof. Assume C contains no recurrent state. Then for all $y \in C$, $\rho_{yy} < 1$ and so

$$\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty.$$

It implies that

$$\begin{aligned}
\infty > \sum_{y \in C} \mathbb{E}_x[N(y)] &= \sum_{y \in C} \sum_{n=1}^{\infty} p^n(x, y) \\
&= \sum_{n=1}^{\infty} \sum_{y \in C} p^n(x, y) \\
&= \sum_{n=1}^{\infty} 1 = \infty
\end{aligned}$$

because C is finite and closed, which induces a contradiction. $\square$

Example 3.3.10. Consider a Markov chain with $|S| = 7$ and the transition matrix $P = (p_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i))$

$$P = \begin{pmatrix}
0.3 & 0 & 0 & 0 & 0.7 & 0 & 0 \\
0.1 & 0.2 & 0.3 & 0.4 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\
0 & 0 & 0 & 0.1 & 0 & 0.1 & 0.8 \\
0 & 0 & 0 & 1 & 0 & 0 & 0
\end{pmatrix}$$

Find all recurrent and transient states.
Solution: First, because $\rho_{21} > 0$ but $\rho_{12} = 0$, 2 is transient. Similarly, $\rho_{34} > 0$ with $\rho_{43} = 0$ implies that 3 is transient. Note that $\{1, 5\}$ is closed and irreducible, so $\{1, 5\}$ are recurrent. $\{4, 6, 7\}$ is also closed and irreducible, so they are transient.

Theorem 3.3.11 (Decomposition Theorem)**.** *Let $R = \{x \in S \colon \rho_{xx} = 1\}$ be the set of all recurrent states. Then*

$$R = \bigcup_i R_i,$$

where R_i is closed and irreducible.

Proof. For any $x \in R$, let

$$C_x = \{y \colon \rho_{xy} > 0\}.$$

By above theorem, $C_x \subset R$.
Claim: Either $C_x \cap C_y = \emptyset$ or $C_x = C_y$.
Suppose $C_x \cap C_y \neq \emptyset$. If $z \in C_x \cap C_y$, then

$$\rho_{xy} \geq \rho_{xz} \rho_{zy} = \rho_{xz} > 0,$$

because $\rho_{xz}, \rho_{yz} > 0$ and $y \in R$. For any $w \in C_y$, we have

$$\rho_{xw} \geq \rho_{xy}\rho_{yw} > 0,$$

which implies that $w \in C_x$. So $C_y \subset C_x$. By symmetry, $C_y = C_x$.
Moreover, C_x is closed and irreducible. So

$$R = \bigcup_x C_x. \qquad \square$$

Example 3.3.12 (Birth and Death Chain)**.** Let $S = \{0, 1, 2, \cdots\}$ and X_n be the size of certain population at time n with

$$\mathbb{P}(X_1 = i+1 \mid X_0 = i) = p_i,\ \mathbb{P}(X_1 = i-1 \mid X_0 = i) = p_i, \mathbb{P}(X_1 = i \mid X_0 = i) = r_i = 1 - p_i - q_i.$$

Note that $q_0 = 0$. Determine under which condition that the state 0 is recurrent.
Solution: Step 1. Construction a function $\varphi\colon S \to \mathbb{R}$ such that $(\varphi(X_n))_{n\geq 0}$ is a martingale.
Let $\varphi(0) = 0$ and $\varphi(1) = 1$. In order that $\varphi(X_n)$ is a martingale, we have

$$\begin{aligned}
\varphi(X_n) &= \mathbb{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \\
&= \mathbb{E}[\varphi(X_1 \circ \theta_n) \mid \mathcal{F}_n] \\
&= \mathbb{E}_{X_n}[\varphi(X_1)]
\end{aligned}$$

If $X_n = k$, then

$$\varphi(k) = \mathbb{E}_k[\varphi(X_1)] = p_k\varphi(k+1) + q_k\varphi(k-1) + r_k\varphi(k)$$

which implies that

$$p_k(\varphi(k+1) - \varphi(k)) = q_k(\varphi(k) - \varphi(k-1)) \ \Rightarrow\ \varphi(k+1) - \varphi(k) = \prod_{j=1}^{k} \frac{q_j}{p_j}$$

and so

$$\varphi(n) = \sum_{m=0}^{n-1} \prod_{j=1}^{m} \frac{q_j}{p_j}$$

with $\prod_{j=1}^{0} \frac{q_j}{p_j} = 1$, that is a increasing function.
Step 2. Let $T_c = \inf\{n \geq 1\colon X_n = c\}$. Then we will prove that if $a < x < b$

$$\mathbb{P}_x(T_a < T_b) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)}$$

and so

$$\mathbb{P}_x(T_b < T_a) = \frac{\varphi(x) - \varphi(a)}{\varphi(b) - \varphi(a)}.$$

Let $T = T_a \wedge T_b$ that is a stopping time. Note that $(\varphi(X_{T\wedge}))_{n\geq 0}$ is a martingale. Moreover,

$$|\varphi(X_{T\wedge n})| \leq \varphi(a) + \varphi(b).$$

So it is UI. Then

$$\mathbb{E}_x[\varphi(X_0)] = \mathbb{E}_x[\varphi(X_T)].$$

It follows that

$$\begin{aligned}
\varphi(x) &= \varphi(a)\mathbb{P}_x(X_T = a) + \varphi(b)\mathbb{P}_x(X_T = b) \\
&= \varphi(a)\mathbb{P}_x(T_a < T_b) + \varphi(b)(1 - \varphi(a)\mathbb{P}_x(T_a < T_b)
\end{aligned}$$

Step 3. Assume $a = 0$ and $b = M$. Then

$$\mathbb{P}_x(T_M < T_0) = \frac{\varphi(x) - \varphi(0)}{\varphi(M) - \varphi(0)}$$

Note that $T_M \geq M \to \infty$ as $M \to \infty$. So

$$\mathbb{P}_x(T_0 = \infty) = \frac{\varphi(x)}{\varphi(\infty)}$$

Claim: 0 is recurrent if and only if $\varphi(\infty) = \infty$.
If 0 is recurrent, then because $\rho_{0x} > 0$, $\rho x0 = 1$ that implies that $\mathbb{P}_x(T_0 = \infty) = 0$. Conversely, if $\mathbb{P}_x(T_0 = \infty) = 0$, then $\mathbb{P}_x(T_0 < \infty) = 1$. It is also true for $x = 1$.

$$\{T_0 < \infty\} = \{X_1 = 0,\ T_0 < \infty\} \cup \{X_1 = 1,\ T_0 < \infty\},$$

So

$$\begin{aligned}
\mathbb{P}_0(T_0 < \infty) &= \mathbb{P}_0(X_1 = 0,\ T_0 < \infty) + \mathbb{P}_0(X_1 = 1,\ T_0 < \infty) \\
&\leq \mathbb{P}_0(X_1 = 0) + \mathbb{P}_0(X_1 = 1)\mathbb{P}_1(T_0 < \infty) < \infty.
\end{aligned}$$

Example 3.3.13 (Symmetric Random Walk)**.** $(X_n)_{n\geq 0}$ is called a random walk if $X_n = x_0 + \sum_{i=1}^n \xi_i$, where ξ_i are i.i.d.. In general, X_n represents the position of a particle at n. A symmetric (simple) random walk on $\mathbb{Z}^d$ is that each transition probability is equal. Note that a symmetric random walk is Markov and irreducible. So one state is recurrent and all states are recurrent. For a symmetric random walk with $x_0 = 0$, let $\tau_0 = 0$ and

$$\tau_n = \inf\{k > \tau_{n-1} : X_k = 0\},$$

i.e., the n-th returning time of 0. By the strong Markov property,

$$\mathbb{P}_0(\tau_n < \infty) = \mathbb{P}_0(\tau_1 < \infty)^n$$

Proof. It is true for $n = 1$. Assume it is true for n. Note that

$$\tau_{n+1} = \tau_n + \tau_1 \circ \theta_{\tau_n}$$

Because $\{\tau_{n+1} < \infty\} \subset \{\tau_n < \infty\}$,

$$\begin{aligned}
\mathbb{P}_0(\tau_{n+1} < \infty) &= \mathbb{P}_0(\tau_n < \infty,\ \tau_{n+1} < \infty) \\
&= \mathbb{P}_0(\tau_n < \infty,\ \tau_n + \tau_1 \circ \theta_{\tau_n} < \infty) \\
&= \mathbb{P}_0(\tau_n < \infty,\ \tau_1 \circ \theta_{\tau_n} < \infty) \\
&= \mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{I}_{\{\tau_1<\infty\}} \circ \theta_{\tau_n}\right] \\
&= \mathbb{E}_0\left[\mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{I}_{\{\tau_1<\infty\}} \circ \theta_{\tau_n} \mid \mathcal{F}_{\tau_n}\right]\right] \\
&= \mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{E}_0\left[\mathbb{I}_{\{\tau_1<\infty\}} \circ \theta_{\tau_n} \mid \mathcal{F}_{\tau_n}\right]\right] \\
&= \mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{E}_{X_{\tau_n}}[\mathbb{I}_{\{\tau_1<\infty\}}]\right] \\
&= \mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{E}_0[\mathbb{I}_{\{\tau_1<\infty\}}]\right] \\
&= \mathbb{E}_0\left[\mathbb{I}_{\{\tau_n<\infty\}}\mathbb{P}_0\{\tau_1 < \infty\}\right] \\
&= \mathbb{P}_0\{\tau_1 < \infty\}\mathbb{P}_0(\tau_n < \infty).
\end{aligned}$$

Then by induction, it is true. □

Theorem 3.3.14. *For any random walk, TFAE.*

(1) $\mathbb{P}_0(\tau_1 < \infty) = 1$.

(2) $\mathbb{P}_0(X_n = 0,\ i.o.) = 1$.

(3) $\sum_{n=0}^{\infty} \mathbb{P}_0(X_n = 0) = \infty$.

Proof. (1) $\Rightarrow$ (2): By above

$$\mathbb{P}_0(\tau_n < \infty) = \mathbb{P}_0(\tau_1 < \infty)^n = 1,$$

which implies (2).
(2) $\Rightarrow$ (3): Let $N(0) = \sum_{m=0}^{\infty} \mathbb{I}_{\{X_m=0\}}$. Then (2) means

$$\mathbb{P}_0(N(0) = \infty) = 1.$$

Then

$$\mathbb{E}_0[N(0)] = \sum_{m=0}^{\infty} \mathbb{P}_0(X_m = 0) = \infty.$$

(3) $\Rightarrow$ (1): Note that

$$N(0) = \sum_{n=0}^{\infty} \mathbb{I}_{\tau_n < \infty},$$

which implies that

$$\mathbb{E}_0[N(0)] = \sum_{n=0}^{\infty} \mathbb{P}_0(\tau_n < \infty) = \sum_{n=0}^{\infty} \mathbb{P}_0(\tau_1 < \infty)^n = \infty$$

So $\mathbb{P}_0(\tau_1 < \infty) = 1$. □

Theorem 3.3.15. *Let $(X_n)_{n\geq 0}$ be a simple random walk in $\mathbb{Z}^d$. $(X_n)_{n\geq 0}$ is recurrent if $d \leq 2$. $(X_n)_{n\geq 0}$ is transient if $d \geq 3$.*

Proof. By above theorem, $(X_n)_{n\geq 0}$ is recurrent if and only if

$$\sum_{n=0}^{\infty} \mathbb{P}_0(X_n = 0) = \infty.$$

(i) $d = 1$: $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$. First, it's obvious

$$\mathbb{P}_0(X_{2m+1} = 0) = 0.$$

For $n = 2m$,

$$\mathbb{P}_0(X_m = 0) = \binom{2m}{m} \left(\frac{1}{2}\right)^{2m}$$

By Stirling's formula,

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \text{ as } n \to \infty,$$

we have

$$\binom{2m}{m} \left(\frac{1}{2}\right)^{2m} \sim m^{-\frac{1}{2}}.$$

So

$$\sum_{n=0}^{\infty} \mathbb{P}_0(X_n = 0) = \infty.$$

(ii) $d = 2$: First, it's obvious

$$\mathbb{P}_0(X_{2n+1} = 0) = 0.$$

To make $X_{2n} = 0$, there exists $0 \le m \le n$ such that m steps up with m steps down, and $n - m$ steps left with $n - m$ steps right. So

$$\mathbb{P}_0(X_{2n} = 0) = \sum_{m=0}^{n} \frac{(2n)!}{m!m!(n-m)!(n-m)!} \left(\frac{1}{4}\right)^{2n} = \left(\frac{1}{4}\right)^{2n} \frac{(2n)!}{n!n!} \sum_{m=0}^{n} \frac{n!n!}{m!m!(n-m)!(n-m)!}$$

Let

$$C_n = \frac{(2n)!}{n!n!} = \binom{2n}{n}.$$

Then

$$\sum_{m=0}^{n} \frac{n!n!}{m!m!(n-m)!(n-m)!} = \sum_{m=0}^{n} \binom{n}{m}\binom{n}{n-m} = \binom{2n}{n} = C_n.$$

So

$$\mathbb{P}_0(X_{2n} = 0) = \left(\frac{1}{4}\right)^{2n} C_n^2 \sim \frac{1}{n} 4^{2n}$$

(iii) $d = 3$: First,

$$\mathbb{P}_0(X_{2n+1} = 0) = 0.$$

Similarly, we have

$$\begin{aligned}
\mathbb{P}_0(X_{2n} = 0) &= \sum_{j,k=0} \frac{(2n)!}{j!j!k!k!(n-j-k)!(n-j-k)!} \left(\frac{1}{6}\right)^{2n} \\
&= 2^{-2n}\binom{2n}{n} \sum_{j,k} \left(3^{-n} \frac{n!}{j!k!(n-j-k)!}\right)^2 \\
&\le 2^{-2n}\binom{2n}{n} \max_{j,k} \left(3^{-n} \frac{n!}{j!k!(n-j-k)!}\right) \sum_{j,k} 3^{-n} \frac{n!}{j!k!(n-j-k)!} \\
&= 2^{-2n}\binom{2n}{n} \max_{j,k} \left(3^{-n} \frac{n!}{j!k!(n-j-k)!}\right)
\end{aligned}$$

because

$$(a+b+c)^n = \sum_{j,k} a^j b^k c^{n-j-k} \frac{n!}{j!k!(n-j-k)!}.$$

Moreover, because the maximum should be taken at $i = j \approx n/3$, by Stirling's formula,

$$\max_{j,k} \frac{n!}{j!k!(n-j-k)!} \le C3^n$$

Then

$$\mathbb{P}_0(X_{2n} = 0) \le C' n^{-\frac{3}{2}}.$$

So it is transient. $\square$

3.4 Stationary Measure

Definition 3.4.1 (Stationary/Invariant Measure)**.** A measure μ on $(S, \mathcal{S})$ is said to be a stationary or invariant measure if

$$\sum_x \mu(x)p(x,y) = \mu(y), \quad \forall\ y \in S.$$

In matrix form, $\mu P = \mu$ for $P = (p(x,y))$ and $\mu = (\mu(x))$. Furthermore, if μ is a probability measure, it is called a stationary distribution.

Remark 3.4.2. Note that

$$\sum_x \mu(x)p(x,y) = \mathbb{P}_\mu(X_1 = y) = \mathbb{E}_\mu\left[\mathbb{P}_{X_0}(X_1 = y)\right] = \mu(y),$$

i.e., starting from μ, $X_1 \sim \mu$. Then by Markov property, $X_n \sim \mu$.

Example 3.4.3 (Random Walk)**.** $X_n = x_0 + \xi_1 + \cdots + \xi_n$ on $\mathbb{Z}^d$ with ξ_i i.i.d $\mathbb{P}(\xi = z) = f(z)$. In such case,

$$p(x,y) = \mathbb{P}_x(X_1 = y) = \mathbb{P}(\xi_1 = y - x) = f(y - x).$$

Let $\mu(x) \equiv 1$ for any $x \in S$. Then μ is a stationary measure because

$$\sum_x \mu(x)p(x,y) = \sum_x f(y-x) = \sum_x f(x) = 1.$$

Example 3.4.4 (1-dim Random Walk)**.** $X_n = \xi_1 + \cdots + \xi_n$ on $\mathbb{Z}$ with ξ_i i.i.d $\mathbb{P}(\xi = 1) = p$ and $\mathbb{P}(\xi = -1) = q$. Assume $p \neq q$. Let

$$\mu(x) = \left(\frac{p}{q}\right)^x, \quad \forall\ x \in \mathbb{Z}.$$

Then μ is a stationary measure.
First, the transition probability

$$p(x,y) = \begin{cases} p, & y = x+1, \\ q, & y = x-1, \\ 0, & \text{otherwise}. \end{cases}$$

So

$$\begin{aligned} \sum_x \mu(x)p(x,y) &= \mu(y-1)p(y-1,y) + \mu(y+1)p(y+1,y) \\ &= \left(\frac{p}{q}\right)^y = \mu(y). \end{aligned}$$

Example 3.4.5 (Birth and Death Process)**.** Let $S = \{0, 1, 2, \cdots\}$ and X_n be the size of certain population at time n with

$$p(x, x+1) = p_x,\ p(x, x-1) = q_x,\ p(x,x) = r_x = 1 - p_x - q_x.$$

with $q_0 = 0$. Let

$$\mu(x) = \prod_{k=1}^{x} \frac{p_{k-1}}{q_k}.$$

Then μ is an invariant measure.

Assume $y > 0$.

$$\begin{aligned}\sum_x \mu(x)p(x,y) &= \mu(y-1)p(y-1,y) + \mu(y+1)p(y+1,y) + \mu(y)p(y,y) \\ &= p_{y-1}\prod_{k=1}^{y-1}\frac{p_{k-1}}{q_k} + q_{y+1}\prod_{k=1}^{y+1}\frac{p_{k-1}}{q_k} + r_y\prod_{k=1}^{y}\frac{p_{k-1}}{q_k} \\ &= \mu(y).\end{aligned}$$

It is also true for $y = 0$.

Definition 3.4.6 (Reversible Markov Chain)**.** A measure μ on $(S, \mathcal{S})$ is said to be a reversible or symmetric measure if

$$\mu(x)p(x,y) = \mu(y)p(y,x), \quad \forall\ x, y \in S.$$

Remark 3.4.7. Note that if μ is reversible, then it is obvious invariant.

Theorem 3.4.8. *Assume μ is invariant and the Markov chain $(X_n)_{n\geq 0}$ with $X_0 \sim \mu$ and transition probability p. Then for any fixed n, let*

$$Y_m = X_{n-m}, \quad m = 0, 1, \cdot, n$$

Then (Y_m) is also a Markov chain with $Y_0 \sim \mu$. Moreover, its transition probability is

$$q(x,y) = \frac{\mu(y)p(y,x)}{\mu(x)}.$$

In particular, if μ is reversible, $p = q$.

Proof. The Markov property can be easily obtained by using the Bayesian rule. For the transition probability, because $X_n \sim \mu$ for all n,

$$\begin{aligned}q(x,y) &= \mathbb{P}(Y_{m+1} = y \mid Y_m = x) \\ &= \mathbb{P}(X_{n-m-1} = y \mid X_{n-m} = x) \\ &= \frac{\mathbb{P}(X_{n-m} = x \mid X_{n-m-1} = y)\mathbb{P}(X_{n-m-1} = y)}{\mathbb{P}(X_{n-m}) = x} \\ &= \frac{p(y,x)\mu(y)}{\mu(x)}.\end{aligned}$$

The followings are obvious. □

Theorem 3.4.9 (Existence)**.** *Assume x is a recurrent state. Let $T = \inf\{m \geq 1 \colon X_m = x\}$. Then*

$$\mu_x(y) := \mathbb{E}_x\left[\sum_{n=0}^{T-1}\mathbb{I}_{\{X_n = y\}}\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty} x\mathbb{I}_{\{X_n = y, n < T\}}\right] = \sum_{n=0}^{\infty}\mathbb{P}_x\left(X_n = y, T > n\right)$$

is an invariant measure.

Proof. It can see $\mu_x(x) = 1$. Let

$$\bar{p}_n(x,y) = \mathbb{P}_x\left(X_n = y, T > n\right).$$

Then $\mu_x(y) = \sum_{n=0}^{\infty} \bar{p}_n(x,y)$. It should to check

$$\sum_y \mu_x(y)p(y,z) = \mu_x(z).$$

By Markov property,

$$\begin{aligned}
\mathbb{P}_x\left(X_n = y, X_{n+1} = z, T > n\right) &= \mathbb{E}_x\left[\mathbb{I}_{\{X_n=y,T>n\}}\mathbb{I}_{\{X_{n+1}=z\}}\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_x\left[\mathbb{I}_{\{X_n=y,T>n\}}\mathbb{I}_{\{X_{n+1}=z\}} \mid \mathcal{F}_n\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{X_n=y,T>n\}}\mathbb{E}_x\left[\mathbb{I}_{\{X_{n+1}=z\}} \mid \mathcal{F}_n\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{X_n=y,T>n\}}\mathbb{E}_{X_n}\left[\mathbb{I}_{\{X_1=z\}}\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{X_n=y,T>n\}}p\left(X_n, z\right)\right] \\
&= \mathbb{P}_x\left(X_n = y, T > n\right)p(y,z)
\end{aligned}$$

Consider the following two cases:
Case 1: $z \neq x$. So $\mathbb{P}_x(X_n = y, X_{n+1} = z, T > n+1) = \mathbb{P}_x(X_n = y, X_{n+1} = z, T > n)$, we have

$$\begin{aligned}
\sum_y \mu_x(y)p(y,z) &= \sum_y \sum_{n=0}^{\infty} \bar{p}_n(x,y)p(y,z) \\
&= \sum_{n=0}^{\infty}\left(\sum_y \mathbb{P}_x\left(X_n = y, T > n\right)p(y,z)\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n+1} = z, T > n+1) \\
&= \mu_x(z).
\end{aligned}$$

Case 2: $z = x$. The right-hand side is 1.

$$\begin{aligned}
\sum_y \mu_x(y)p(y,z) &= \sum_{n=0}^{\infty}\sum_y \mathbb{P}_x\left(X_n = y, X_{n+1} = x, T > n\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x\left(X_{n+1} = x, T > n\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(T = n+1) \\
&= \mathbb{P}_x(T < \infty) = 1,
\end{aligned}$$

because x is recurrent. □

Remark 3.4.10. If S is finite, by above theorem, it always has a recurrent state. So for finite case, Markov chain always has an invariant measure.

Theorem 3.4.11 (Uniqueness). *If the Markov chain is irreducible and recurrent, then the invariant measure is unique up to constant multiples.*

Proof. Fix $a \in S$ that is obvious recurrent. So we have an invariant measure μ_a. Given any invariant measure ν. First, we have for any z,

$$\nu(z) = \sum_y \nu(y)p(y,z) = \nu(a)p(a,z) + \sum_{y \neq a} \nu(y)p(y,z).$$

Using that for multiple times, we have

$$
\begin{aligned}
\nu(z) &= \nu(a)p(a,z) + \sum_{y\neq a}\left(\nu(a)p(a,y) + \sum_{x\neq a}\nu(x)p(x,y)\right)p(y,z) \\
&= \nu(a)p(a,z) + \sum_{y\neq a}\nu(a)p(a,y)p(y,z) + \sum_{y\neq a}\sum_{x\neq a}\nu(x)p(x,y)p(y,z) \\
&= \nu(a)\mathbb{P}_z(X_1=z) + \nu(a)\sum_{y\neq a}\mathbb{P}_a(X_1=y,\ X_2=z) \\
&\quad + \sum_{y\neq a}\sum_{x\neq a}\left(\nu(a)p(a,x) + \sum_{w\neq x}\nu(w)p(w,x)\right)p(x,y)p(y,z) \\
&= \nu(a)\mathbb{P}_a\left(X_1=z\right) + \nu(a)\mathbb{P}_a\left(X_1\neq a, X_2=z\right) + \nu(a)\mathbb{P}_a\left(X_1\neq a, X_2\neq a, X_3=z\right) \\
&\quad + \sum_{y\neq a}\sum_{x\neq a}\sum_{w\neq a}\nu(w)p(w,x)p(x,y)p(y,z) \\
&= \cdots \\
&\geq \nu(a)\sum_{n=1}^{\infty}\mathbb{P}_a\left(X_k\neq a, 1\leqslant k<n, X_n=z\right) \\
&= \nu(a)\sum_{n=1}^{\infty}\mathbb{P}_a\left(T_a>n, X_n=z\right) \\
&= \nu(a)\mu_a(z).
\end{aligned}
$$

Conversely, because ν is invariant

$$
\begin{aligned}
\nu(a) &= \sum_x \nu(x)p^n(x,a) \\
&\geq \sum_x \nu(a)\mu_a(x)p^n(x,a) \\
&= \nu(a)\mu_a(a) = \nu(a).
\end{aligned}
$$

Therefore,

$$\sum_x\left(\nu(x)-\nu(a)\mu_a(x)\right)p^n(x,a)=0.$$

For any $y\in S$,

$$\left(\nu(y)-\nu(a)\mu_a(y)\right)p^n(y,a) + \sum_{x\neq y}\left(\nu(x)-\nu(a)\mu_a(x)\right)p^n(x,a)=0.$$

Because of the irreducibility, there exists a n such that $p^n(y,a)>0$. So

$$\nu(y)=\nu(a)\mu_a(y).$$

□

Theorem 3.4.12. *If π is a stationary distribution, then all states y that $\pi(y)>0$ is recurrent.*

Proof. For any $n\in\mathbb{N}$, because π is stationary,

$$\sum_{n=1}^{\infty}\sum_{x\in S}\pi(x)p^n(x,y) = \sum_{n=1}^{\infty}\pi(y)=\infty.$$

On the other hand,

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{x\in S}\pi(x)p^n(x,y) &= \sum_{x\in S}\pi(x)\sum_{n=1}^{\infty}p^n(x,y)\\ &= \sum_{x\in S}\pi(x)\frac{\rho_{xy}}{1-\rho_{yy}}\\ &\leq \frac{1}{1-\rho_{yy}}.\end{aligned}$$

So $\rho_{yy}=1$. □

Theorem 3.4.13. *If the Markov chain is irreducible and it has a stationary distribution π, then $\pi(x)>0$ for all x and*

$$\pi(x)=\frac{1}{\mathbb{E}_x[T_x]},$$

where $T_x=\inf\{n>0\colon X_n=x\}$.

Proof. **Claim:** For all x, $\pi(x)>0$.
Because π is invariant,

$$\pi(x)=\sum_{y\in S}\pi(y)p^n(y,x).$$

Because there is a z such that $\pi(z)>0$ and the irreducibility implies that $p^n(z,x)>0$ for some n,

$$\pi(x)>\pi(z)p^n(z,x)>0.$$

So all x are recurrent, which means the Markov chain is irreducible and recurrent. By the uniqueness of the stationary distribution, for any x,

$$\mu_x(y)=c\pi(y).$$

So

$$\sum_{y\in S}\mu_x(y)=\sum_{n=0}^{\infty}\sum_{y\in S}\mathbb{P}_x\left(X_n=y,T_x>n\right)=\sum_{n=0}^{\infty}\mathbb{P}_x\left(T_x>n\right)=\mathbb{E}_x\left[T_x\right]$$

On the other hand,

$$c=\mathbb{E}_x\left[T_x\right]=\frac{\mu_x(y)}{\pi(y)}.$$

In particular, let $y=x$.

$$\pi(x)=\frac{1}{\mathbb{E}_x[T_x]}.$$ □

Definition 3.4.14 (Positive Recurrence)**.** Let x be a recurrent state. If $\mathbb{E}_x[T_x]<\infty$, then x is called positively recurrent, otherwise, it is called null recurrent.

Theorem 3.4.15. *If the Markov chain is irreducible, then TFAE.*

(1) There exists a positively recurrent state.

(2) There exists a stationary distribution.

(3) All states are positively recurrent state.

Proof. (1) $\Rightarrow$ (2): If x is positively recurrent, then define

$$\pi(y) := \frac{\mu_x(y)}{\mathbb{E}_x[T_x]},$$

which is a stationary distribution.
(2) $\Rightarrow$ (3): By above theorem,

$$\pi(y) = \frac{1}{\mathbb{E}_y[T_y]} > 0,$$

so $\mathbb{E}_y[T_y] < \infty$. $\square$

Chapter 4

Poisson Process

4.1 Construction

Definition 4.1.1 (Poisson Process)**.** A stochastic process $(N_t)_{t\geq 0}$ with $N_0 = 0$ is called a Poisson process of rate λ if

(i) (Independent increasing) for any $t_1 < s_1 < t_2 < s_2 < \cdots < t_n < s_n$,

$$N_{s_1} - N_{t_1},\ N_{s_2} - N_{t_2}, \cdots N_{s_n} - N_{t_n}$$

are independent.

(ii) for any $s < t$, $N_t - N_s \sim \text{Pois}(\lambda(t-s))$.

Remark 4.1.2. In general, $N_t =$ the number of times an event occurs in $[0, t]$.

The next problem is how to construct a Poisson process: Given a $\lambda > 0$, let $\xi_1, \cdots, \xi_n, \cdots$ be i.i.d. with exponential distribution $\exp(\lambda)$, i.e.

$$\mathbb{P}(\xi_i > t) = e^{-\lambda t},$$

(In fact, ξ_i is the time between incidents). Let $T_0 = 0$ and $T_n = \xi_1 + \cdots + \xi_n$ that is the time at which the n-th incident occurs. Define

$$N_t = \sup\left\{n > 0 \colon T_n \leq t\right\}.$$

Then $(N_t)_{t\geq 0}$ is a Poisson process.

Proof. Step 1: $N_t \sim \text{Pois}(\lambda)$.
Note that $T_n \sim \Gamma(n, \lambda)$, i.e., its density function is

$$f_{T_n}(s) = \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s}.$$

As we know,

$$\{N_t = 0\} = \{T_1 > t\} = \{\xi_1 > t\}$$

which implies

$$\mathbb{P}(N_t = 0) = e^{-\lambda t}.$$

For $N_t = n$ with $n \geq 1$,

$$\begin{aligned}
\mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t < T_{n+1}) \\
&= \mathbb{P}(T_n \leq t < T_n + \xi_{n+1}) \\
&= \iint_{s \leq t < s+u} f_{T_n}(s) f_{\xi_{n+1}}(u) ds du \\
&= \frac{(\lambda t)^n}{n!} e^{-\lambda t}.
\end{aligned}$$

Step 2: Fix t, let

$$T_1' = T_{N_t+1} - t,\ T_2' = T_{N_t+2} - T_{N_t+1},\ \cdots T_k' = T_{N_t+k} - T_{N_t+k-1}, \cdots$$

Claim: $T_1',\ T_2',\ \cdots$ are i.i.d. $\exp(\lambda)$ and they are independent with N_t.
First,

$$\begin{aligned}
\mathbb{P}(T_{n+1} - t \geq s \mid N_t = n) &= \frac{\mathbb{P}(T_{n+1} - t \geq s, N_t = n)}{\mathbb{P}(N_t = n)} \\
&= \frac{\mathbb{P}(T_{n+1} - t \geq s, T_n \leq t)}{\mathbb{P}(N_t = n)} \\
&= \frac{\mathbb{P}(T_n + \xi_{n+1} - t \geq s, T_n \leq t)}{\mathbb{P}(N_t = n)} \\
&= \frac{e^{-\lambda(t+s)} \frac{(\lambda t)^2}{n!}}{\mathbb{P}(N_t = n)} \\
&= e^{-\lambda s}.
\end{aligned}$$

Then consider

$$\begin{aligned}
&\mathbb{P}(T_n \leq t,\ T_{n+1} - t \geq s,\ T_{n+k} - T_{n+k-1} \geq v_k, k = 2, 3, \cdots, m) \\
&= \mathbb{P}(T_n \leq t,\ T_{n+1} - t \geq s,\ \xi_{n+k} \geq v_k, k = 2, 3, \cdots, m) \\
&= \mathbb{P}(T_n \leq t,\ T_{n+1} - t \geq s) \prod_{k=2}^{m} \mathbb{P}(\xi_{n+k} \geq v_k),
\end{aligned}$$

which implies that

$$\begin{aligned}
&\mathbb{P}(T_{n+1} - t \geq s,\ T_{n+k} - T_{n+k-1} \geq v_k, k = 2, 3, \cdots, m \mid N_t = n) \\
&= \frac{\mathbb{P}(T_n \leq t,\ T_{n+1} - t \geq s,\ \xi_{n+k} \geq v_k, k = 2, 3, \cdots, m)}{\mathbb{P}(N_t = n)} \\
&= \frac{\mathbb{P}(T_n \leq t,\ T_{n+1} - t \geq s)}{\mathbb{P}(N_t = n)} \prod_{k=2}^{m} \mathbb{P}(\xi_{n+k} \geq v_k) \\
&= e^{-\lambda s} \prod_{k=2}^{m} e^{-\lambda v_k}
\end{aligned}$$

For the independence,

$$\begin{aligned}
&\mathbb{P}(T_1' \geq s,\ T_k' \geq v_k, k = 2, 3, \cdots m,\ N_t \leq \ell) \\
&= \sum_{n=0}^{\ell} \mathbb{P}(T_1' \geq s,\ T_k' \geq v_k, k = 2, 3, \cdots m,\ N_t = \ell) \\
&= \sum_{n=0}^{\ell} \mathbb{P}(T_{N_t+1} - t \geq s,\ T_{N_t+k} - T_{N_t+k-1} \geq v_k, k = 2, 3, \cdots m \mid N_t = n) \mathbb{P}(N_t = n)
\end{aligned}$$

$$= \sum_{n=0}^{\ell} \mathbb{P}(T_{n+1} - t \geq s,\ T_{n+k} - T_{n+k-1} \geq v_k, k = 2, 3, \cdots m \mid N_t = n)\mathbb{P}(N_t = n)$$
$$= e^{-\lambda s} \prod_{k=2}^{m} e^{-\lambda v_k} \mathbb{P}(N_t \leq \ell).$$

Step 3: For any $t_0 < t_1 < t_2 < \cdots < t_n$, it suffices to check

$$\mathbb{P}(N_{t_i} - N_{t_i-1} \geq k_i, i = 1, \cdots, n) = \prod_{i=1}^{n} e^{-\lambda(t_i - t_{i-1})} \frac{(\lambda(t_i - t_{i-1}))^{k_i}}{k_i!}.$$

It only needs to prove for $N_{t_2} - N_{t_1}$ and N_{t_1}. Let $T_1' = T_{N_{t_1}+1} - t_1$ and $T_k' = T_{N_{t_1}+k} - T_{N_{t_1}+k-1}$. Then by Step 2, $T_1', \cdots, T_k'$ are independent with N_{t_1}. Note that

$$\begin{aligned}
\{N_{t_2} - N_{t_1} = m\} &= \left\{T_{N_{t_1}+m} \leq t_2,\ T_{N_{t_1}+m+1} > t_2\right\} \\
&= \left\{T_{N_{t_1}+m} - t_1 \leq t_2 - t_1,\ T_{N_{t_1}+m+1} - t_1 > t_2 - t_1\right\} \\
&= \left\{T_1' + \sum_{k=2}^{m} T_k' \leq t_2 - t_1,\ T_1' + \sum_{k=2}^{m+1} T_k' > t_2 - t_1\right\},
\end{aligned}$$

which follows that $N_{t_2} - N_{t_1}$ is independent with N_{t_1} and moreover

$$\mathbb{P}\left(N_{t_2} - N_{t_1} = m\right) = e^{-\lambda(t_2 - t_1)} \frac{(\lambda(t_2 - t_1))^m}{m!}$$

by Step 2. □

4.2 Compound Poisson Process

Definition 4.2.1 (Compound Poisson Process)**.** Let $(N_t)_{t\geq 0}$ be a Poisson process with λ and $Y_1, \cdots, Y_n, \cdots$ be i.i.d. and independent with N_t. Then

$$S(t) = \sum_{k=1}^{N_t} Y_k$$

is called a compound Poisson process.

Theorem 4.2.2. *Let $Y_1, \cdots, Y_n, \cdots$ be i.i.d. and $N \geq 0$ be an integer-valued and independent random variable. Let*

$$S = Y_1 + Y_2 + \cdots + Y_N,$$

and $S = 0$ if $N = 0$. Then

(i) $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[Y_i]$.

(ii) $\mathrm{Var}(S) = \mathbb{E}[N]\,\mathrm{Var}(Y_i) + \mathrm{Var}(N)\,(\mathbb{E}[Y_i])^2$*. In particular, if* $N \sim \mathrm{Pois}(\lambda)$*, then* $\mathrm{Var}(S) = \lambda\mathbb{E}[Y_i^2]$.

Proof. First, by independence,

$$\begin{aligned}
\mathbb{E}[S] &= \sum_{n=0}^{\infty} \mathbb{E}[S\mathbb{I}_{\{N=n\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[Y_i]\mathbb{E}[\mathbb{I}_{\{N=n\}}] \\
&= \mathbb{E}[Y_i]\mathbb{E}[N].
\end{aligned}$$

For $\mathrm{Var}(S) = \mathbb{E}[S^2] - (\mathbb{E}[S])^2$,

$$\begin{aligned}
\mathbb{E}[S^2] &= \sum_{n=1}^{\infty} \mathbb{E}[S^2 \mathbb{I}_{\{N=n\}}] \\
&= \sum_{n=1}^{\infty} \mathbb{E}[(Y_1 + \cdot + Y_n)^2 \mathbb{I}_{\{N=n\}}] \\
&= \sum_{n=1}^{\infty} \mathbb{E}[S_n^2]\mathbb{E}[\mathbb{I}_{\{N=n\}}] \\
&= \sum_{n=1}^{\infty} \left(n\,\mathrm{Var}(Y_i) + (n\mathbb{E}[Y_i])^2\right) \mathbb{P}(N=n) \\
&= \mathrm{Var}(Y_i)\mathbb{E}[N] + \mathbb{E}[Y_i]^2\mathbb{E}[N^2]
\end{aligned}$$

for $S_n = Y_1 + \cdots + Y_n$. Furthermore, $(\mathbb{E}[S])^2 = \mathbb{E}[Y_i]^2\mathbb{E}[N]^2$. So it is obtained. □

Theorem 4.2.3. *Suppose $(N_t)_{t\geq 0}$ is a Poisson process with rate λ, which describes the number of points come by time t. We keep a point that lands at s with probability p_s. Let $\bar{N}_t$ be the number of points landing at s by time t. Then $(\bar{N}_t)$ is also a Poisson process with rate λp_s.*

Proof. **Independent Increasing:** Because $\bar{N}_{t_1} - \bar{N}_{t_2}$ is determined by $N_{t_1} - N_{t_2}$.
Poisson Distribution: First,

$$\begin{aligned}
\mathbb{P}(\bar{N}_t = m) &= \mathbb{P}(\bar{N}_t = m,\ N_t \geq m) \\
&= \sum_{k=m}^{\infty} \mathbb{P}(\bar{N}_t = m,\ N_t = k) \\
&= \sum_{k=m}^{\infty} \mathbb{P}(\bar{N}_t = m \mid N_t = k)\mathbb{P}(N_t = k) \\
&= \sum_{k=m}^{\infty} \binom{k}{m} (p_s)^m (1-p_s)^{k-m} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \\
&= \frac{e^{-\lambda t}(\lambda t)^m}{m!} (p_s)^m \sum_{k=m}^{\infty} \frac{\left(\lambda t(1-p_s)^{k-m}\right)}{(k-m)!} \\
&= \frac{e^{-\lambda t}(\lambda t)^m}{m!} (p_s)^m e^{\lambda t(1-p_s)} = e^{-\lambda p_s t}\frac{(\lambda p_s t)^m}{m!}.
\end{aligned}$$

It is similar for others. □

Chapter 5

Brownian Motion

5.1 Definition and Properties

Definition 5.1.1. Let $(B_t)_{t\geq 0}$ be a stochastic process. It is called a (standard when $B_0 = 0$) Brownian motion if

(1) $t \to B_t(\omega)$ is continuous a.e.

(2) it is independent increments.

(3) for any $s < t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$.

Remark 5.1.2. Note that for $s \leq t$

$$\mathbb{E}[B_s B_t] = \mathbb{E}[B_s]\mathbb{E}[B_t - B_s] + \mathbb{E}[B_s^2] = s.$$

So for any $s, t \geq 0$, $\mathbb{E}[B_s B_t] = s \wedge t$.

Theorem 5.1.3 (Finite-dimensional Distribution)**.** *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. For any $0 = t_0 < t_1 < \cdots < t_n$,*

$$(B_{t_1}, \cdots, B_{t_n}) \sim \mathcal{N}(\mu, \Sigma)$$

with

$$p(x_1, x_2, \cdots, x_n) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{t_1(t_2-t_1)\cdots(t_n-t_{n-1})}} \exp\left(-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right),$$

where $x_0 = 0$.

Proof. For any measurable f, let $X_i = B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$,

$$\begin{aligned}
\mathbb{E}[f(B_{t_1}, \cdots, f(B_{t_n}))] &= \mathbb{E}\left[f\left(B_{t_1}, B_{t_2} - B_{t_1} + B_{t_1}, \cdots, \sum_{i=1}^n B_{t_i} - B_{t_{i-1}}\right)\right] \\
&= \mathbb{E}\left[f\left(X_1, X_1 + X_2, \cdots, \sum_i X_i\right)\right] \\
&= \int f(y_1, y_1 + y_2, \cdots) \prod_i f_{X_i}(y_i) dy_1 \cdots dy_n.
\end{aligned}$$

Let $x_1 = y_1$, $x_2 = y_1 + y_2$, $\cdots$, and $x_n = \sum_i y_i$. Then

$$f_{X_i}(y_i) = f_{X_i}(x_i - x_{i-1}) = \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left(-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right),$$

which implies the desired property. $\square$

Theorem 5.1.4. *Suppose $(B_t)_{t\geq 0}$ is a Brownian motion.*

(1) $(-B_t)_{t\geq 0}$ is also a Brownian motion.

(2) For any $\lambda > 0$, the process

$$B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$$

is also a Brownian motion.

(3) For any $s > 0$,

$$B_t^s = B_{t+s} - B_s$$

is also a Brownian motion and independent of $\mathcal{F}_s = \sigma(B_t : t \leq s)$.

Proof. (1) is obvious. For (2), because

$$B_{\lambda^2 t} - B_{\lambda^2 s} \sim \mathcal{N}(0, \lambda^2(t-s)),$$

$B_t^\lambda - B_s^\lambda \sim \mathcal{N}(0, t-s)$. For (3), it is obvious a Brownian motion. The independence is directly obtained by

$$(B_{s+t_1} - B_s, B_{s+t_2} - B_s) = (B_{s+t_1} - B_s, B_{s+t_2} - B_{s+t_1} + B_{s+t_1} - B_s)$$

independent of B_s. □

Remark 5.1.5. A direct corollary for (3) is, for any $t_0 < t_1 < \cdots < t_n$, the joint $(B_{t_1}, \cdots, B_{t_n})$ is independent of B_{t_0}.

Given a Brownian motion $B = (B_t)_{t\geq 0}$, a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is called a Brownian filtration if it is a filtration and B_t is $\mathcal{F}_t$-adapted and $B_t - B_s$ is $\mathcal{F}_s$-independent. It is not hard to see B is a $\mathbb{F}$-martingale.

5.2 Properties of Path

Theorem 5.2.1 (0-1 Law)**.** *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion with the nature filtration $\mathcal{F}_t = \sigma(B_s : s \leq t)$. Define*

$$\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t.$$

Then for any $A \in \mathcal{F}_{0+}$, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proof. First, for any $0 < t_1 < \cdots < t_n$ and any bounded continuous function $f : \mathbb{R}^n \to \mathbb{R}$, we will show that $\mathbb{I}_A$ is independent with $f(B_{t_1}, \cdots, B_{t_n})$. By the continuity of path and the continuity of f,

$$\mathbb{E}[\mathbb{I}_A f(B_{t_1}, \cdots, B_{t_n})] = \lim_{\epsilon\to 0} \mathbb{E}[\mathbb{I}_A f(B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\varepsilon)].$$

For $0 < \varepsilon < t_1$, $B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\varepsilon$ are independent with B_s for all $s \leq \varepsilon$, which means they are independent with $\mathcal{F}_\varepsilon$. Because $A \in \mathcal{F}_\varepsilon$, $f(B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\varepsilon)$ is independent with $\mathbb{I}_A$. So

$$\begin{aligned}\mathbb{E}[\mathbb{I}_A f(B_{t_1}, \cdots, B_{t_n})] &= \lim_{\epsilon\to 0} \mathbb{E}[\mathbb{I}_A]\mathbb{E}[f(B_{t_1} - B_\epsilon, \cdots, B_{t_n} - B_\varepsilon)] \\ &= \mathbb{E}[\mathbb{I}_A]\mathbb{E}[f(B_{t_1}, \cdots, B_{t_n})]\end{aligned}$$

So $\mathbb{I}_A$ is independent of $\sigma(B_s : s > 0) = \sigma(B_s : s \geq 0)$ because $B_0 = \lim_{t\to 0} B_t$, which implies that A is independent of itself. □

Remark 5.2.2. In general, if

$$Y = \limsup_n X_n = \inf_{n\geq 1} \sup_{k\geq n} X_k$$

then Y is $\sigma(X_n : n \in \mathbb{N})$-measurable, because

$$\{Y > a\} = \left\{\inf_n \sup_{k\geq n} X_k > a\right\} = \bigcap_{n=1}^{\infty} \left\{\sup_{k\geq n} X_k > a\right\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{X_k > a\}.$$

Theorem 5.2.3. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion.*

(1) We have almost surely for every $\varepsilon > 0$

$$\sup_{0\leq s\leq \varepsilon} B_s > 0, \quad \inf_{0\leq s\leq \varepsilon} B_s < 0.$$

(2) For every $a \in \mathbb{R}$, let

$$T_a = \inf\{t \geq 0 \colon B_t = a\}.$$

Then $\mathbb{P}(T_a < \infty) = 1$.

Proof. (1) Let ε_p be a sequence of positive numbers decreasing to 0. Let

$$A = \bigcap_{p>0} \left\{\sup_{0\leq \varepsilon\leq \varepsilon_p} B_\varepsilon > 0\right\}.$$

Note for any $p_0 > 0$,

$$A = \bigcap_{p\geq p_0} \left\{\sup_{0\leq \varepsilon\leq \varepsilon_p} B_\varepsilon > 0\right\}.$$

So $A \in \mathcal{F}_{\varepsilon_{p_0}}$ because ε_p is decreasing in p, which follows that

$$A \in \bigcap_{p_0} \mathcal{F}_{\varepsilon_{p_0}} = \mathcal{F}_{0^+}.$$

On the other hand,

$$\mathbb{P}(A) = \lim_{p\to\infty} \mathbb{P}\left(\sup_{0\leq \varepsilon\leq \varepsilon_p} B_\varepsilon > 0\right).$$

Moreover, for any p,

$$\mathbb{P}\left(\sup_{0\leq \varepsilon\leq \varepsilon_p} B_\varepsilon > 0\right) \geq \mathbb{P}(B_{\varepsilon_p} > 0) = \frac{1}{2}$$

So $\mathbb{P}(A) \geq \frac{1}{2}$. By 0-1 Law,

$$\mathbb{P}(A) = 1 \;\Rightarrow\; \mathbb{P}\left(\sup_{0\leq \varepsilon\leq \varepsilon_p} B_\varepsilon > 0\right) = 1.$$

For the other one, it is because of $(-B_t)$ also a Brownian motion.

(2) First,

$$\{T_a < \infty\} = \bigcup_{t=0}^{\infty} \{B_t = a\}.$$

It follows that we only need

$$\mathbb{P}\left(\bigcup_{t=0}^{\infty}\{B_t = a\}\right) = 1.$$

Claim: For any $M > 0$,

$$\mathbb{P}(\sup_s B_s > M) = 1, \quad \mathbb{P}(\inf_s B_s < -M) = 1.$$

By (1), we have

$$1 = \mathbb{P}(\sup_{0\le s\le 1} B_s > 0) = \lim_{\delta\to 0}\mathbb{P}(\sup_{0\le s\le 1} B_s > \delta).$$

For the right hand side

$$\mathbb{P}(\sup_{0\le s\le 1} B_s > \delta) = \mathbb{P}\left(\sup_{0\le s\le 1}\frac{1}{\delta}B_s > 1\right).$$

Because $\frac{1}{\delta}B_s \overset{d}{=} B_{\frac{1}{\delta^2}s}$,

$$\mathbb{P}(\sup_{0\le s\le 1} B_s > \delta) = \mathbb{P}\left(\sup_{0\le s\le 1} B_{\frac{1}{\delta^2}s} > 1\right) = \mathbb{P}\left(\sup_{0\le u\le 1/\delta^2} B_u > 1\right).$$

Therefore,

$$\lim_{\delta\to 0}\mathbb{P}\left(\sup_{0\le u\le 1/\delta^2} B_u > 1\right) = \mathbb{P}\left(\sup_s B_s > 1\right) = 1.$$

Then for any $M > 0$,

$$\mathbb{P}\left(\sup_s B_s > M\right) = \mathbb{P}\left(\sup_s \frac{1}{M}B_s > 1\right) = \mathbb{P}\left(\sup_s B_{\frac{1}{M^2}s} > 1\right) = \mathbb{P}\left(\sup_s B_s > 1\right) = 1.$$

For the infimum, it is because $(-B_t)$ is also a Brownian motion.

Then if $a > 0$, there is an M such that $a < M$. By the continuity of path and $B_0 = 0$,

$$\left\{\sup_s B_s > M\right\} \subset \bigcup_{t=0}^{\infty}\{B_t = a\}$$

So $\mathbb{P}\left(\cup_t \{B_t = a\}\right) = 1$. Similarly, for $a \le 0$, it can get by $\inf B_s$. $\square$

Corollary 5.2.4. *For a standard Brownian motion $(B_t)_{t\ge 0}$,*

$$\limsup_{t\to\infty} B_t = \infty, \quad \liminf_{t\to\infty} B_t = -\infty.$$

Proposition 5.2.5. *Let $0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t$ be a sequence of partition of $[0,t]$ such that $\max_i(t_i^n - t_{i-1}^n) \to 0$ as $n \to \infty$. Then*

$$\lim_{n\to\infty}\sum_{i=1}^{p_n}\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 = t$$

in $L^2(\Omega)$.

Proof. To show L^2 convergence, we need

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{p_n}\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - t\right)^2\right] = 0.$$

Let $t = \sum_i t_i^n - t_{i-1}^n$. Then

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=1}^{p_n}\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - t\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{p_n}\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i,j}\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)\left(\left(B_{t_j^n} - B_{t_{j-1}^n}\right)^2 - (t_j^n - t_{j-1}^n)\right)\right] \\
&= \sum_{i\neq j}\mathbb{E}\left[\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right]\mathbb{E}\left[\left(B_{t_j^n} - B_{t_{j-1}^n}\right)^2 - (t_j^n - t_{j-1}^n)\right] \\
&\quad + \sum_i \mathbb{E}\left[\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)^2\right] \\
&= \sum_i \mathbb{E}\left[\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)^2\right] \\
&\leq 2\sum_i\left(\mathbb{E}\left[\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^4\right] + (t_i^n - t_{i-1}^n)^2\right) \\
&= 2(c+1)\sum_i (t_i^n - t_{i-1}^n)^2 \leq 2(c+1)t\max_i(t_i^n - t_{i-1}^n) \to 0
\end{aligned}$$

Note the $X \sim \mathcal{N}(0,\sigma^2)$, $\mathbb{E}[X^4] = c\sigma^4$. $\square$

Corollary 5.2.6. *For a.e. $t \mapsto B_t$ has infinite variation on any finite interval.*

Proof. Let $0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t$ be a sequence of partition of $[0,t]$.

$$\sum_i\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 \leq \max_i\left|B_{t_i^n} - B_{t_{i-1}^n}\right|\sum_i\left|B_{t_i^n} - B_{t_{i-1}^n}\right|$$

By the continuity, $\max_i\left|B_{t_i^n} - B_{t_{i-1}^n}\right| \to 0$. If $\sum_i\left|B_{t_i^n} - B_{t_{i-1}^n}\right| < \infty$,

$$\sum_i\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 \to 0,$$

which induces a contradiction. $\square$

Theorem 5.2.7. *Given a Brownian motion $B = (B_t)_{t\geq 0}$, for a.e. $\omega \in \Omega$,*

$$\limsup_{t\downarrow 0}\frac{W_t(\omega)}{\sqrt{2t\log\log(1/t)}} = 1, \quad \liminf_{t\downarrow 0}\frac{W_t(\omega)}{\sqrt{2t\log\log(1/t)}} = -1$$

and

$$\limsup_{t\to\infty}\frac{W_t(\omega)}{\sqrt{2t\log\log(1/t)}} = 1, \quad \liminf_{t\to\infty}\frac{W_t(\omega)}{\sqrt{2t\log\log(1/t)}} = -1.$$

5.3 Strong Markov Property

Given a standard Brownian motion $(B_t)_{t\geq 0}$, let $\mathcal{F}_t = \sigma(B_s \colon s \leq t)$ and $\mathcal{F}_\infty = \sigma(B_t \colon t \geq 0)$, i.e. $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of B_t.
First, for Markov property, we already know $B_{t+s} - B_s$ is independent with B_s, which directly implies the Markov property by the following lemma.

Lemma 5.3.1. *Let X and Y be two random variables on a probability space Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$. If X is $\mathcal{G}$-measurable and Y is independent with $\mathcal{G}$, then for any Borel measurable function $g \colon \mathbb{R}^2 \to \mathbb{R}$,*

$$\mathbb{E}[g(X,Y) \mid \mathcal{G}] = \mathbb{E}[g(X,Y) \mid \sigma(X)]$$

Remark 5.3.2. First, assume $g(x,y) = \mathbb{I}_A(x)\mathbb{I}_B(y)$ and it can clearly true so that it is also true for all simple function g. Then by applying the Monotone Class Theorem, it can prove that.

So

$$\begin{aligned}\mathbb{E}\left[f\left(B_{t+s}\right) \mid \mathcal{F}_t\right] &= \mathbb{E}\left[f\left(B_t + \left(B_{t+s} - B_t\right)\right) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}\left[f\left(B_t + \left(B_{t+s} - B_t\right)\right) \mid \sigma\left(B_t\right)\right] \\ &= \mathbb{E}\left[f\left(B_{t+s}\right) \mid \sigma\left(B_t\right)\right]\end{aligned}$$

For the strong Markov property, first, we need the stopping time.

Definition 5.3.3 (Stopping Time)**.** A random time $T \colon \Omega \to [0,\infty]$ is a stopping time with respect to $(\mathcal{F}_t)_{t\geq 0}$ if for any t,

$$\{T \leq t\} \in \mathcal{F}_t.$$

Remark 5.3.4. Note that

$$\{T < t\} = \bigcup_{q\in\mathbb{Q}, q<t} \{T \leq q\} \in \mathcal{F}_t.$$

so that $\{T \geq t\} \in \mathcal{F}_t$.

Example 5.3.5. (1) For any $a \in \mathbb{R}$,

$$T_a = \inf\left\{s \geq 0 \colon B_s = a\right\}$$

is a stopping time because

$$\{T \leq t\} = \left\{\inf_{0\leq s\leq t} |B_s - a| = 0\right\} \in \mathcal{F}_t.$$

(2) Let

$$T = \sup\left\{s \leq 1 \colon B_s = 0\right\}.$$

Then it is not a stopping time because it needs information in $[0,1]$.

Definition 5.3.6. Given a stopping time T,

$$\mathcal{F}_T = \left\{A \in \mathcal{F}_\infty \colon A \cap \{T \leq t\} \in \mathcal{F}_t,\ \forall\ t \geq 0\right\},$$

which is a σ-field.

Remark 5.3.7. (1) T is $\mathcal{F}_T$-measurable, where the reasoning is as same as that of the discrete case.

(2) For any $s \geq 0$, $B_s \mathbb{I}_{s \leq T}$ is $\mathcal{F}_T$-measurable.

For any $A \in \mathcal{R}$ and WTLG assuming $0 \notin A$ (otherwise considering A^c),

$$\{B_s \mathbb{I}_{s \leq T} \in A\} \cap \{T \leq t\} = \begin{cases} \emptyset, & t < s \\ \{B_s \in A\} \cap \{s \leq T \leq t\}, & s \leq T \leq t. \end{cases}$$

Because $s \leq t$, $\{B_s \in A\} \in \mathcal{F}_t$. Furthermore, $\{T \geq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$, so $\{B_s \mathbb{I}_{s \leq T} \in A\} \cap \{T \leq t\} \in \mathcal{F}_t$.

For a stopping time T, consider $\mathbb{I}_{T<\infty} B_T$, which is $\mathcal{F}_T$-measurable. Let $n \in \mathbb{N}$. If

$$\frac{k}{2^n} \leq T \leq \frac{k+1}{2^n},$$

then define $B_T^n = B_{k/2^n}$. By the continuity of path, $\lim_n B_T^n = B^T$. So

$$\begin{aligned}\mathbb{I}_{\{T<\infty\}} B_T &= \lim_{n\to\infty} \sum_{i=0}^{\infty} \mathbb{I}_{\left\{\frac{i}{2^n} \leq T \leq \frac{i+1}{2^n}\right\}} B_{\frac{i}{2^n}} \\ &= \lim_{n\to\infty} \sum_{i=0}^{\infty} \mathbb{I}_{\left\{T \geq \frac{i}{2^n}\right\}} \mathbb{I}_{\left\{T \leq \frac{i+1}{2^n}\right\}} B_{\frac{i}{2^n}}.\end{aligned}$$

Both $\mathbb{I}_{\left\{T \leq \frac{i+1}{2^n}\right\}} B_{\frac{i}{2^n}}$ and $\mathbb{I}_{\left\{T \geq \frac{i}{2^n}\right\}}$ are $\mathcal{F}_T$-measurable, which implies that $\mathbb{I}_{T<\infty} B_T$ is $\mathcal{F}_T$-measurable.

Theorem 5.3.8 (Strong Markov Property)**.** *Give a stopping time T. Assume $\mathbb{P}(T < \infty) > 0$. Set*

$$B_t^{(T)} = \mathbb{I}_{\{T<\infty\}} (B_{T+t} - B_T), \quad t \geq 0.$$

Then under the probability $\mathbb{P}(\cdot \mid T < \infty)$, $(B_t^{(T)})_{t\geq 0}$ is a Brownian motion and independent of $\mathcal{F}_T$.

Proof. WTLG assume $\mathbb{P}(T < \infty) = 1$. For any $A \in \mathcal{F}_T$ and $0 \leq t_1 < t_2 < \cdots < t_p$ and any bounded continuous function $F \colon \mathbb{R}^p \to \mathbb{R}$, it suffices to show that

$$\mathbb{E}\left[\mathbb{I}_A F(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)})\right] = \mathbb{P}(A) \mathbb{E}\left[F(B_{t_1}, \cdots, B_{t_p})\right].$$

Define $[t]_n = k/2^n$ if $(k-1)/2^n < t \leq k/2^n$. Observe that

$$F(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)}) = \lim_{n\to\infty} F\left(B_{t_1}^{([T]_n)}, \cdots, B_{t_p}^{([T]_n)}\right)$$

by the continuity of F and B_t.

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_A F(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)})\right] &= \lim_{n\to\infty} \mathbb{E}\left[\mathbb{I}_A F\left(B_{t_1}^{([T]_n)}, \cdots, B_{t_p}^{([T]_n)}\right)\right] \\ = \lim_{n\to\infty} \sum_{k=0}^{\infty} &\mathbb{E}\left[\mathbb{I}_A \mathbb{I}_{\left\{\frac{k-1}{2^n} \leq T \leq \frac{k}{2^n}\right\}} F\left(B_{t_1}^{(k/2^n)}, \cdots, B_{t_p}^{(k/2^n)}\right)\right] \\ = \lim_{n\to\infty} \sum_{k=0}^{\infty} &\mathbb{E}\left[\mathbb{I}_A \mathbb{I}_{\left\{\frac{k-1}{2^n} \leq T \leq \frac{k}{2^n}\right\}} F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \cdots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}\right)\right]\end{aligned}$$

Because T is a stopping time and $A \in \mathcal{F}_T$,

$$A \cap \left\{\frac{k-1}{2^n} \leq T \leq \frac{k}{2^n}\right\} \in \mathcal{F}_{\frac{k}{2^n}}.$$

Because $B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \cdots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}$ are independent of $\mathcal{F}_{\frac{k}{2^n}}$,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{I}_A F(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)})\right] &= \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{I}_A \mathbb{I}_{\left\{\frac{k-1}{2^n} \le T \le \frac{k}{2^n}\right\}}\right] \mathbb{E}\left[F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \cdots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}\right)\right] \\
&= \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{I}_A \mathbb{I}_{\left\{\frac{k-1}{2^n} \le T \le \frac{k}{2^n}\right\}}\right] \mathbb{E}\left[F\left(\tilde{B}_{t_1}, \cdots, \tilde{B}_{t_p}\right)\right] \\
&= \mathbb{P}(A)\mathbb{E}\left[F\left(B_{t_1}, \cdots, B_{t_p}\right)\right],
\end{aligned}$$

where the final equality is because $\tilde{B}_t = B_{\frac{k}{2^n}+t} - B_{\frac{k}{2^n}}$ is also a Brownian motion. □

Remark 5.3.9. A direct corollary of this is

$$\mathbb{E}[f(B_{T+s}) \mid \mathcal{F}_T] = \mathbb{E}[f(B_{T+s}) \mid X_T],$$

which is the strong Markov property.

Theorem 5.3.10 (Reflexive Principle)**.** *For any $t > 0$, let*

$$S_t = \sup_{0\le s\le t} B_s \ge 0.$$

If $a \ge 0$ and $b \le a$, then

$$\mathbb{P}(S_t \ge a, B_t \le b) = \mathbb{P}(B_t \ge 2a - b).$$

In particular, S_t has the same distribution as $|B_t|$.

Proof. Let $T_a = \inf\{t \ge 0 \colon B_t = a\}$. By the continuity of B_t, $\{S_t \ge a\} = \{T_a \le t\}$. So

$$\begin{aligned}
\mathbb{P}(S_t \ge a,\ B_t \le b) &= \mathbb{P}(T_a \le t,\ B_{t-T_a+T_a} \le b) \\
&= \mathbb{P}\left(T_a \le t,\ B_{t-T_a+T_a} - B_{T_a} \le b - a\right) \\
&= \mathbb{P}\left(T_a \le t,\ B_{t-T_a}^{(T_a)} \le b - a\right).
\end{aligned}$$

Let $B'_t = B_{t-T_a}^{(T_a)}$ that is a Brownian motion independent of T_a because T_a is $\mathcal{F}_{T_a}$-measurable. So

$$\begin{aligned}
\mathbb{P}\left(T_a \le t,\ B_t^{(T_a)} \le b - a\right) &= \mathbb{P}\left(T_a \le t\right)\mathbb{P}\left(-B'_t \ge a - b\right) \\
&= \mathbb{P}\left(T_a \le t\right)\mathbb{P}\left(B'_t \ge a - b\right) \\
&= \mathbb{P}(T_a \le t,\ B'_t \ge a - b) \\
&= \mathbb{P}(T_a \le t,\ B_t - B_{T_a} \ge a - b) \\
&= \mathbb{P}(T_a \le t,\ B_t \ge 2a - b).
\end{aligned}$$

But $\{T_a \le t\} \subset \{B_t \ge 2a - b\}$ because of $B_t \ge 2a - b \ge a$ and the continuity of B_t. So

$$\mathbb{P}(S_t \ge a,\ B_t \le b) = \mathbb{P}(B_t \ge 2a - b).$$

For the other one,

$$\begin{aligned}
\mathbb{P}(S_t \ge a) &= \mathbb{P}(S_t \ge a,\ B_t \ge a) + \mathbb{P}(S_t \ge a,\ B_t \le a) \\
&= \mathbb{P}(B_t \ge a) + \mathbb{P}(B_t \ge 2a - a) \\
&= 2\mathbb{P}(B_t \ge a) = \mathbb{P}(|B_t| \ge a).
\end{aligned}$$

□

Corollary 5.3.11. *T_a has the same distribution as $\frac{a^2}{B_1^2}$ with the density function*

$$f(t) = \frac{a}{\sqrt{2\pi t}} e^{-\frac{a^2}{2t}} \mathbb{I}_{\{t>0\}}$$

Proof. Because $\{S_t \ge a\} = \{T_a \le t\}$,

$$\mathbb{P}\left(T_a \le t\right) = \mathbb{P}\left(S_t \ge a\right) = \mathbb{P}(|B_t| \ge a).$$

□

5.4 High-dimensional Brownian Motion

Definition 5.4.1. A d-dimensional stochastic process $(\boldsymbol{B}_t = (B_t^1, \cdots, B_t^d))_{t\geq 0}$ is called a d-dimensional Brownian motion if for each i, $(B_t^i)_{t\geq 0}$ is a Brownian motion and $(B_t^i)_{t\geq 0}$ $(i = 1, \cdots, d)$ are independent of each other.

Remark 5.4.2. A d-dimensional Brownian motion is a martingale with

$$\left\langle B^i, B^j \right\rangle_t = \delta_{ij} t.$$

Theorem 5.4.3 (Lévy Theorem)**.** *Let $\boldsymbol{M} = (M^1, \cdots, M^d)$ be d-dimensional continuous local martingale with respect to $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and $M_0 = 0$. If*

$$\left\langle M^i, M^j \right\rangle = \delta_{ij} t,$$

then $\boldsymbol{M}$ is a d-dimensional Brownian motion.

Theorem 5.4.4. *Let M be a continuous local martingale w.s.t. $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ with $M_0 = 0$ and*

$$\lim_{t\to\infty} \langle M \rangle_t = \infty.$$

For each $t \geq 0$, define the stopping time

$$\tau_t := \inf\left\{s \colon \langle M \rangle_s > t\right\}.$$

Then $(M_{\tau_t})_{t\geq 0}$ is a Brownian motion.

Remark 5.4.5. Let $\boldsymbol{B}$ be a d-dimensional Brownian motion.

(1) If $d = 1$, we have seen $B_t = 0$ for infinitely many t.

(2) If $d = 2$, $\boldsymbol{B}_t \neq 0$ for $t \neq 0$ but it hits every ball centered at 0.

(3) If $d \geq 3$, $\|\boldsymbol{B}_t(\omega)\| \to \infty$ as $t \to \infty$.

Chapter 6

Continuous Time Martingale

6.1 Filtration

Definition 6.1.1 (Filtration)**.** A filtration is a family of increasing σ-fields $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and let $\mathcal{F}_\infty = \bigcup_t \mathcal{F}_t$.

Example 6.1.2. For a process $X = (X_t)_{t\geq 0}$, let $\mathcal{F}_t^X = \sigma(X_s \colon s \leq t)$. Then $(\mathcal{F}_t^X)_{t\geq 0}$ is a filtration, called the natural filtration of X.

Definition 6.1.3 (Right Continuity)**.** Given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. For any $t \geq 0$, define

$$\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$$

and $\mathcal{F}_{\infty+} = \mathcal{F}_\infty$. In general, $\mathcal{F}_t \subset \mathcal{F}_{t+}$. If for any $t \geq 0$,

$$\mathcal{F}_t = \mathcal{F}_{t+},$$

then the filtration is called right-continuous.

Definition 6.1.4 (Completeness)**.** Given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. Let

$$N = \{A \colon \exists\ A' \supset A,\ A' \in \mathcal{F}_\infty,\ \mathbb{P}(A') = 0\}.$$

If $N \subset \mathcal{F}_0$, then the filtration is called complete.

Remark 6.1.5. If a filtration is not complete, then define

$$\mathcal{F}_t' = \mathcal{F}_t \vee \sigma(N),$$

the smallest σ-field containing $\mathcal{F}_t$ and $\sigma(N)$. Then the filtration $(\mathcal{F}_t')_{t\geq 0}$ is complete. So in the following, we always consider complete filtration.

Remark 6.1.6. A filtration $\mathbb{F}$ is said to satisfy the usual condition if $\mathbb{F}$ is right-continuous and complete.

Given a stochastic process $X = (X_t)_{t\geq 0}$, note that

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

viewed by $X(\omega, t) = X_t(\omega)$. On $\Omega \times [0,\infty)$, we can consider the product σ-field $\mathcal{F} \times \mathcal{B}([0,\infty))$.

Definition 6.1.7 (Measurability)**.** Given a stochastic process $X = (X_t)_{t\geq 0}$.

(1) X is said to be measurable if $X(\omega, t)$ is $\mathcal{R} \times \mathcal{F}$-measurable.

(2) X is said to be $(\mathcal{F}_t)_{t\geq 0}$-progressively measurable if for every t, $X\colon \Omega \times [0,t] \to \mathbb{R}$ is $\mathcal{F}_t \times \mathcal{B}([0,t])$-measurable.

(3) X is called $(\mathcal{F}_t)_{t\geq 0}$-adapted if X_t is $\mathcal{F}_t$-measurable for all t.

Remark 6.1.8. Note that if $f\colon E_1 \times E_2 \to \mathbb{R}$ is measurable, then for any $x \in E_1$, $f(x,\cdot)\colon E_2 \to \mathbb{R}$ is also measurable. So if X is progressively measurable, then it is adapted. Moreover, it is also measurable. Conversely, if X is measurable and adapted, then it has a progressively measurable modification.

Proposition 6.1.9. *Suppose a stochastic process $X = (X_t)_{t\geq 0}$ is $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$-adapted and $t \mapsto X_t(\omega)$ is right-continuous (or left-continuous) a.e.. Then X is $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$-progressively measurable.*

Proof. Fix $t > 0$. Consider the process X on $\Omega \times [0,t]$. Define for $s < t$

$$X_s^n := X_{\frac{kt}{n}}, \quad \text{if } s \in \left[\frac{(k-1)t}{n}, \frac{kt}{n}\right)$$

and $X_t^n = X_t$. Then by the right-continuity, $X_s^n \to X_s$ a.e. (Similar for the left-continuity by taking the left-end point). Now for $A \in \mathcal{R}$,

$$\{(\omega,s)\colon X_s^n(\omega) \in A\} = (\{X_t \in A\} \times \{t\}) \cup \bigcup_{k=1}^n \left(\left\{ X_{\frac{kt}{n}} \in A \right\} \times \left[\frac{(k-1)t}{n}, \frac{kt}{n}\right) \right) \in \mathcal{F}_t \times \mathcal{B}([0,t]),$$

because X is adapted. So $X^n = (X_t^n)_{t\geq 0}$ is progressively measurable. By taking limits, X is progressively measurable. ☐

Remark 6.1.10. A σ-field on $\Omega \times [0,\infty)$ is generated by all $(\mathcal{F}_t)_{t\geq 0}$-progressively measurable process, called $(\mathcal{F}_t)_{t\geq 0}$-progressively measurable σ-field.

6.2 Stopping Time

Fix a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$.

Definition 6.2.1 (Stopping Time)**.** A random variable $T\colon \Omega \to [0,\infty]$ is called a stopping time w.s.t. $(\mathcal{F}_t)_{t\geq 0}$ if $\{T \leq t\} \in \mathcal{F}_t$ for all t.

Remark 6.2.2. As mentioned before, $\{T < t\} \in \mathcal{F}_t$ and so $\{T \geq t\} \in \mathcal{F}_t$. Note that the converse is not true.

Definition 6.2.3. Given a stopping time T, let

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty \colon \forall\, t \geq 0,\ A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

Let $\mathcal{G}_t = \mathcal{F}_{t+}$ that is also a filtration and $\mathcal{G}_{t+} = \mathcal{G}_t$.

Proposition 6.2.4. *(1) T is a stopping time w.s.t. $(G_t)_{t\geq 0}$ if and only if for any $t > 0$,*

$$\{T < t\} \subset \mathcal{F}_t,$$

which is also equivalent to $T \wedge t$ is $\mathcal{F}_t$-measurable.

(2) Let T be a $(G_t)_{t\geq 0}$-stopping time. Then

$$\mathcal{G}_T = \{A \in \mathcal{F}_\infty : \forall\ t > 0,\ A \cap \{T < t\} \in \mathcal{F}_t\}$$

In some way, $\mathcal{F}_{T+} := \mathcal{G}_T$.

Proof. (1) Assume T is a stopping time w.s.t. $(G_t)_{t\geq 0}$.

$$\{T < t\} = \bigcup_{q\in\mathbb{Q}, q<t} \{T \leq q\}\,.$$

Because

$$\{T \leq q\} \in \mathcal{G}_q = \mathcal{F}_{q+} \subset \mathcal{F}_t,$$

$\{T < t\} \in \mathcal{F}_t$.

Conversely, for any $t \geq 0$ and any $s > t$,

$$\{T \leq t\} = \bigcap_{q\in\mathbb{Q}, t<q<s} \{T < q\} \in \mathcal{F}_s,$$

because

$$\{T < q\} \in \mathcal{F}_q \subset \mathcal{F}_s.$$

So

$$\{T \leq t\} \in \bigcap_{s>t} \mathcal{F}_s = \mathcal{G}_t.$$

(2) Let $A \in \mathcal{G}_T$. Then for all $t > 0$,

$$A \cap \{T \leq t\} \in \mathcal{G}_t.$$

Hence

$$A \cap \{T < t\} = \bigcup_{q\in\mathbb{Q}, q<t} A \cap \{T \leq q\} \in \mathcal{F}_t.$$

Conversely, for any $A \in \mathcal{F}_\infty$ with $A \cap \{T < t\} \in \mathcal{F}_t$ for all $t > 0$,

$$A \cap \{T \leq t\} = \bigcap_{q\in\mathbb{Q}, t<q<s} A \cap \{T < q\} \in \mathcal{F}_s.$$

So

$$A \cap \{T \leq t\} \in \bigcap_{s>t} \mathcal{F}_s = \mathcal{G}_t.$$

□

Proposition 6.2.5. *(1) For any stopping time T,*

$$\mathcal{F}_T \subset \mathcal{G}_T = \mathcal{F}_{T+}$$

(2) If $T = t$, $\mathcal{F}_T = \mathcal{F}_t$.

(3) T is $\mathcal{F}_T$-measurable.

Proof. (1) It is because

$$A \cap \{T < t\} = \bigcup_{q\in\mathbb{Q}, q<t} A \cap \{T \leq q\}\,.$$

(2) It is because
$$A \cap \{T \le s\} = A \text{ or } \emptyset.$$

(3) For any $a > 0$, let $A = \{T > a\}$.
$$\begin{aligned} A \cap \{T \le t\} &= \{T > a\} \cap \{T \le t\} \\ &= \begin{cases} \emptyset, & a \ge t \\ \{a < T \le t\}, & a < t. \end{cases} \end{aligned}$$

Because
$$\{a < T \le t\} = \{T \le a\}^c \cap \{T \le t\} \in \mathcal{F}_a \cap \mathcal{F}_t \subset \mathcal{F}_t,$$
$A \cap \{T \le t\} \in \mathcal{F}_t$. □

Proposition 6.2.6. *(1) Let T be a stopping time and $A \in \mathcal{F}_\infty$. Define*
$$T^A(\omega) := \begin{cases} T(\omega), & \omega \in A, \\ \infty, & \omega \notin A. \end{cases}$$

Then T^A is a stopping time if and only if $A \in \mathcal{F}_T$.

(2) If T is a stopping time, then $T + s$ is a stopping time for any constant $s \ge 0$.

(3) Let S, T be stopping times with $S \le T$. Then
$$\mathcal{F}_S \subset \mathcal{F}_T, \quad \mathcal{F}_{S+} \subset \mathcal{F}_{T+}.$$

(4) Let S, T be stopping times. Then $S \vee T$ and $S \wedge T$ are stopping times. Moreover, $\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$ and
$$\{S \le T\} \in \mathcal{F}_{S \wedge T}.$$
So $\{T \le S\} \in \mathcal{F}_{S \wedge T}$ and $\{S = T\} \in \mathcal{F}_{S \wedge T}$.

Proof. (1) Note that for any t,
$$\left\{T^A \le t\right\} = A \cap \{T \le t\} \in \mathcal{F}_t.$$

(2) It is because
$$\{T + s \le t\} = \{T \le t - s\} \in \mathcal{F}_{t-s} \subset \mathcal{F}_t.$$

(3) Note that $\{T \le t\} \subset \{S \le t\}$. For any $A \in \mathcal{F}_S$,
$$A \cap \{T \le t\} = (A \cap \{S \le t\}) \cap \{T \le t\} \in \mathcal{F}_t,$$
and similarly for $\mathcal{F}_{S+} \subset \mathcal{F}_{T+}$.

(4) Note that
$$\{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t.$$
Similarly,
$$\{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathcal{F}_t.$$
By (3),
$$\mathcal{F}_{S \wedge T} \subset \mathcal{F}_S \cap \mathcal{F}_T.$$

Conversely, let $A \in \mathcal{F}_S \cap \mathcal{F}_T$. Then

$$A \cap \{S \wedge T \leq t\} = (A \cap \{S \leq t\}) \cup (A \cap \{T \leq t\}) \in \mathcal{F}_t.$$

So $\mathcal{F}_S \cap \mathcal{F}_T \subset \mathcal{F}_{S \wedge T}$.

For any $t > 0$,

$$\{S \leq T\} \cap \{S \wedge T \leq t\} = (\{S \leq T\} \cap \{S \leq t\}) \cup (\{S \leq T\} \cap \{T \leq t\}).$$

Because

$$\{S \leq T\} \cap \{T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \cap \{S \wedge t \leq T \wedge t\}$$

and note that $S \wedge t, T \wedge t$ are $\mathcal{F}_t$-measurable,

$$\{S \leq T\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

For the other term, similarly

$$\{S \leq T\} \cap \{S \leq t\} = \{S \wedge t \leq T \wedge t\} \cap \{S \leq t\} \in \mathcal{F}_t.$$

So

$$\{S \leq T\} \cap \{S \wedge T \leq t\} \in \mathcal{F}_t. \qquad \square$$

Remark 6.2.7. For stopping times $S \leq T$, it can define the stochastic interval

$$(S, T] := \{(t, \omega) \in [0, \infty] \times \Omega \colon S(\omega) < t \leq T(\omega)\},$$

so is similarly $[S, T], (S, T)$.

Proposition 6.2.8. *(1) If $\{S_n\}_{n \geq 0}$ is an increasing stopping time and $S_n \to S$, then S is a stopping time.*

(2) If $\{S_n\}_{n \geq 0}$ is a decreasing stopping time and $S_n \to S$, then S is a $(\mathcal{F}_{t+})$-stopping time and

$$\mathcal{F}_{S+} = \bigcap_n \mathcal{F}_{S_n+}.$$

Proof. (1) Because $S_n \uparrow S$,

$$\{S \leq t\} = \bigcap_n \{S_n \leq t\} \in \mathcal{F}_t.$$

So S is a stopping time.

(2) Because $S_n \downarrow S$,

$$\{S < t\} = \bigcup_n \{S_n < t\} \in \mathcal{F}_t, \text{ and } A \cap \{S < t\} = \bigcup_n (A \cap \{S_n < t\}).$$

So S is a $(\mathcal{F}_{t+})$-stopping time and $\mathcal{F}_{S+} \supset \bigcap_n \mathcal{F}_{S_n+}$. For the other side,

$$A \cap \{S_n < t\} = A \cap \{S < t\} \cap \{S_n < t\} \in \mathcal{F}_t. \qquad \square$$

Proposition 6.2.9. *Let T be a stopping time. A random variable Y defined on $\{T < \infty\}$ is $\mathcal{F}_T$-measurable if and only if for any $t \geq 0$, $Y|_{\{T \leq t\}}$ is $\mathcal{F}_t$-measurable.*

Proof. For any $A \in \mathcal{R}$,

$$\{Y \in A\} \cap \{T \leq t\} = \left\{Y|_{\{T \leq t\}} \in A\right\}. \qquad \square$$

Remark 6.2.10. If $X = (X_t)_{t\geq 0}$ is progressively measurable and T is a stopping time, then $X_T \mathbb{I}_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable.

Proposition 6.2.11. *Let T be a stopping time.*

(1) Let S be a $\mathcal{F}_T$-measurable random variable with values $[0,\infty]$ such that $S \geq T$. Then S is also a stopping time.

(2) Define

$$T_n(\omega) = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{I}_{\left\{\frac{k}{2^n} < T(\omega) \leq \frac{k+1}{2^n}\right\}} + \infty \mathbb{I}_{\{T(\omega)=\infty\}}.$$

Then T_n is a sequence of stopping times that decreases to T.

Proof. (1) For any $t \geq 0$, because $T \leq S$,

$$\{S \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

(2) Note that $T_n \geq T$ is $\mathcal{F}_T$-measurable so by (1) it is a stopping time.

□

Example 6.2.12. Let $(X_t)_{t\geq 0}$ be an adapted stochastic process.

(1) Assume $t \mapsto X_t(\omega)$ is right-continuous. Let O be a open set.

$$T_O := \inf\{t \geq 0 \colon X_t \in O\}$$

is a stopping time with respect to $(\mathcal{F}_{t+})_{t\geq 0}$.

(2) Assume $t \mapsto X_t(\omega)$ is continuous. Let F be a closed set.

$$T_F = \inf\{t \geq 0 \colon X_t \in F\}$$

is a stopping time (w.s.t. $(\mathcal{F}_t)$).

Proof. (1) For any $t \geq 0$,

$$\{T_O < t\} = \bigcup_{q\in[0,t)\cap\mathbb{Q}} \{X_q \in O\}.$$

To prove that, first, for any ω such that $T_O(\omega) = \alpha = \inf\{t \colon X_t(\omega) \in O\} < t$, we can choose a rational sequence $t > t_n \downarrow \alpha$. By the right continuity of X_t, $X_{t_n}(\omega) \to X_\alpha(\omega) \in O$. Because O is open, there is a large n such that $X_{t_n}(\omega) \in O$. So $\{T_O < t\} \subset \bigcup_{q\in[0,t)\cap\mathbb{Q}} \{X_q \in O\}$. The converse is obvious.
So it directly has

$$\{T_O < t\} = \bigcup_{q\in[0,t)\cap\mathbb{Q}} \{X_q \in O\} \in \mathcal{F}_t.$$

(2) For any $t \geq 0$,

$$\{T_F \leq t\} = \left\{\inf_{0\leq s\leq t} d(X_s, F) = 0\right\}.$$

First, "$\subset$" is obvious. Conversely, there is a $\{t_n\} \subset [0,t]$ such that $t_n \to t_0 \leq t$ and

$$\lim_n d(X_{t_n}(\omega), F) = d(\lim_n X_{t_n}(\omega), F) = d(X_{t_0}(\omega), F) = 0$$

because of the continuity. Since F is closed, $X_{t_0}(\omega) \in F$ so $T_F(\omega) \leq t$.

Then

$$\{T_F \leq t\} = \left\{\inf_{0\leq s\leq t} d(X_s, F) = 0\right\} = \left\{\inf_{s\in[0,t]\cap\mathbb{Q}} d(X_s, F) = 0\right\} \in \mathcal{F}_t.$$ □

6.3 Martingale

Fix a filtration $(\mathcal{F}_t)_{t\geq 0}$.

Definition 6.3.1. A stochastic process $(X_t)_{t\geq 0}$ is called a submartingale if

(1) $X_t \in L^1$ and $(X_t)_{t\geq 0}$ is adapted,

(2) for any $s < t$,
$$\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s.$$

If $\leq$ in (2), it is called a supermartingale. If $(X_t)_{t\geq 0}$ is a sub and super martingale, it is called a martingale.

Example 6.3.2. Let $(B_t)_{t\geq 0}$ be a Brownian motion and $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration.

(1) $(B_t)_{t\geq 0}$ is a martingale.

(2) Let
$$Y_t = B_t^2 - t.$$
Then $(Y_t)_{t\geq 0}$ is a martingale.

(3) Let.
$$Z_t = \exp\left(\theta B_t - \frac{1}{2}\theta^2 t\right)$$
Then $(Z_t)_{t\geq 0}$ is also a martingale.

Proof. (1) For any $s < t$,
$$\begin{aligned}\mathbb{E}[B_t \mid \mathcal{F}_s] &= \mathbb{E}[B_t - B_s \mid \mathcal{F}_s] + \mathbb{E}[B_s \mid \mathcal{F}_s] \\ &= \mathbb{E}[B_t - B_s] + B_s \\ &= B_s.\end{aligned}$$

(2) For $s < t$, by (1),
$$\begin{aligned}\mathbb{E}[Y_t \mid \mathcal{F}_s] &= \mathbb{E}\left[B_t^2 \mid \mathcal{F}_s\right] - t \\ &= \mathbb{E}\left[(B_t - B_s)^2 \mid \mathcal{F}_s\right] + 2\mathbb{E}\left[B_t B_s \mid \mathcal{F}_s\right] - \mathbb{E}\left[B_s^2 \mid \mathcal{F}_s\right] - t \\ &= \mathbb{E}\left[(B_t - B_s)^2\right] + 2B_s\mathbb{E}[B_t \mid \mathcal{F}_s] - B_s^2 - t \\ &= t - s + 2B_s^2 - B_s^2 - t \\ &= B_s^2 - s = Y_s.\end{aligned}$$

(3) For any $s < t$,
$$\begin{aligned}\mathbb{E}[Z_t \mid \mathcal{F}_s] &= \exp\left(-\frac{1}{2}\theta^2 t\right)\mathbb{E}\left[\exp(\theta B_t - \theta B_s)\exp(\theta B_s) \mid \mathcal{F}_s\right] \\ &= \exp\left(\theta B_s - \frac{1}{2}\theta^2 t\right)\mathbb{E}\left[\exp(\theta B_t - \theta B_s) \mid \mathcal{F}_s\right] \\ &= \exp\left(\theta B_s - \frac{1}{2}\theta^2 t\right)\mathbb{E}\left[\exp(\theta(B_t - B_s))\right].\end{aligned}$$

Because $B_t - B_s \sim \mathcal{N}(0, t-s)$,

$$\mathbb{E}\left[\exp(\theta(B_t - B_s))\right] = \frac{1}{\sqrt{2\pi(t-s)}} \int \exp(\theta x) \exp(-\frac{1}{2(t-s)} x^2) dx = \exp\left(\frac{1}{2}\theta^2(t-s)\right).$$

So

$$\mathbb{E}[Z_t \mid \mathcal{F}_s] = \exp\left(\theta B_s - \frac{1}{2}\theta^2 t\right) \exp\left(\frac{1}{2}\theta^2(t-s)\right) = Z_s. \qquad \square$$

Example 6.3.3. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ and $(\mathcal{F}_t)_{t\geq 0}$ be the nature filtration.

(1) $(N_t - \lambda t)$ is a martingale.

(2) Let

$$Z_t = (N_t - \lambda t)^2 - \lambda t.$$

Then $(Z_t)_{t\geq 0}$ is also a martingale.

(3) Given $\alpha > 0$, set β such that

$$L_t = \exp(\alpha N_t - \beta t)$$

is a martingale.

Proof. (1) For $s < t$,

$$\begin{aligned}
\mathbb{E}\left[N_t - \lambda t \mid \mathcal{F}_s\right] &= \mathbb{E}\left[N_t - N_s \mid \mathcal{F}_s\right] + \mathbb{E}\left[N_s \mid \mathcal{F}_s\right] - \lambda t \\
&= \mathbb{E}[N_t - N_s] + N_s - \lambda t \\
&= \lambda(t-s) + N_s - \lambda t = N_s - \lambda s.
\end{aligned}$$

(2) For $s < t$,

$$\begin{aligned}
\mathbb{E}\left[Z_t \mid \mathcal{F}_s\right] &= \mathbb{E}\left[(N_t - \lambda t)^2 \mid \mathcal{F}_s\right] - \lambda t \\
&= \mathbb{E}\left[N_t^2 \mid \mathcal{F}_s\right] - 2\lambda t \mathbb{E}\left[N_t \mid \mathcal{F}_s\right] + (\lambda t)^2 - \lambda t \\
&= \mathbb{E}\left[(N_t - N_s)^2 \mid \mathcal{F}_s\right] + 2N_s \mathbb{E}[N_t \mid \mathcal{F}_s] - \mathbb{E}[N_s^2 \mid \mathcal{F}_s] \\
&\quad - 2\lambda t \mathbb{E}\left[N_t - \lambda t \mid \mathcal{F}_s\right] - (\lambda t)^2 - \lambda t \\
&= \mathbb{E}\left[(N_t - N_s)^2\right] + 2N_s \mathbb{E}[N_t - N_s \mid \mathcal{F}_s] + N_s^2 \\
&\quad - 2\lambda t(N_s - \lambda s) - (\lambda t)^2 - \lambda t \\
&= \lambda(t-s) + \lambda^2(t-s)^2 + 2\lambda(t-s)N_s + N_s^2 \\
&\quad - 2\lambda t(N_s - \lambda s) - (\lambda t)^2 - \lambda t \\
&= (N_s - \lambda s)^2 - \lambda s.
\end{aligned}$$

(3) For $s < t$,

$$\begin{aligned}
\mathbb{E}\left[L_t \mid \mathcal{F}_s\right] &= \exp(-\beta t)\mathbb{E}\left[\exp(\alpha N_t) \mid \mathcal{F}_s\right] \\
&= \exp(-\beta t)\mathbb{E}\left[\exp(\alpha(N_t - N_s)) \exp(\alpha N_s) \mid \mathcal{F}_s\right] \\
&= \exp(\alpha N_s - \beta t)\mathbb{E}\left[\exp(\alpha(N_t - N_s))\right] \\
&= \exp(\alpha N_s - \beta t) \exp\left(\lambda(t-s)(e^\alpha - 1)\right).
\end{aligned}$$

because $N_t - N_s \sim \text{Pois}(\lambda(t-s))$. So when $\beta = \lambda(e^\alpha - 1)$,

$$\mathbb{E}\left[L_t \mid \mathcal{F}_s\right] = \exp(\alpha N_s - \beta s) = L_s. \qquad \square$$

Proposition 6.3.4. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be convex function.*

(1) If $X = (X_t)_{t\geq 0}$ is a martingale and $f(X_t) \in L^1$, then $\{f(X)_t\}_{t\geq 0}$ is a submartingale.

(2) If $X = (X_t)_{t\geq 0}$ is a submartingale and f is increasing and $f(X_t) \in L^1$, then $\{f(X)_t\}_{t\geq 0}$ is a submartingale.

Remark 6.3.5. In general, we take $f(x) = |x|^p$ with $p \geq 1$ and $f(x) = x^+$.

Theorem 6.3.6. *Let $X = (X_t)_{t\geq 0}$ be a sub(super)martingale. Then*

$$\sup_{s\in[0,t]} \mathbb{E}[|X_s|] < \infty.$$

Remark 6.3.7. Note that if X is a martingale, by above $|X_t|$ is a submartingale so

$$\mathbb{E}\left[|X_s|\right] \leq \mathbb{E}\left[\mathbb{E}\left[|X_t| \mid \mathcal{F}_s\right]\right] \leq \mathbb{E}[|X_t|] < \infty$$

Proof. Assume X is a submartingale. So $(X_t^+)_{t\geq 0}$ is also a submartingale. So

$$\mathbb{E}[X_s^+] \leq \mathbb{E}[X_t^+] < \infty.$$

On the other hand,

$$\mathbb{E}[X_s^-] = \mathbb{E}[X_s^+] - \mathbb{E}[X_s] \leq \mathbb{E}[X_t^+] - \mathbb{E}[X_0] < \infty.$$

So $\sup_{s\in[0,t]} \mathbb{E}[|X_s|] = \sup_{s\in[0,t]} \mathbb{E}[X_s^+] + \mathbb{E}[X_s^-] \leq 2\mathbb{E}[X_t^+] - \mathbb{E}[X_0] < \infty$. ☐

Theorem 6.3.8. *Let $X = (X_t)_{t\geq 0}$ be a positive submartingale (or martingale) with right-continuous paths.*

(1) Maximum inequality: For any $\lambda > 0$,

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s| > \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}[|X_t|].$$

(2) Doob's inequality: For any $p > 1$ and $t > 0$,

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X_s|^p\right] \leq \left(\frac{p}{1-p}\right)^p \mathbb{E}\left[|X_t|^p\right].$$

Proof. Fix $t > 0$. Consider a countable dense subset $D \subset [0,t]$ containing 0 and t. Let $D_m = \{0 = t_0^m < t_1^m < \cdots < t_m^m = t\}$ such that $D_m \uparrow D$. By the continuity of path

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s| > \lambda\right) = \mathbb{P}\left(\sup_{s\in D} |X_s| > \lambda\right), \quad \mathbb{E}\left[\sup_{s\in[0,t]} |X_s|^p\right] = \mathbb{E}\left[\sup_{s\in D} |X_s|^p\right].$$

And by the convergence,

$$\mathbb{P}\left(\sup_{s\in D} |X_s| > \lambda\right) = \lim_{m\to\infty} \mathbb{P}\left(\sup_{s\in D_m} |X_s| > \lambda\right), \quad \mathbb{E}\left[\sup_{s\in D} |X_s|^p\right] = \lim_{m\to\infty} \mathbb{E}\left[\sup_{s\in D_m} |X_s|^p\right].$$

When $X = (X_t)_{t\geq 0}$ be a positive submartingale (or martingale), $(|X_{t_n^m}|)$ is also a submartingale. Then by the discrete case, we have the inequalities. ☐

Remark 6.3.9. If X is a nonnegative supermartingale with right-continuous paths, by the proof of discrete case, we clearly have

$$\mathbb{P}\left(\sup_{s\in[0,t]} X_s > \lambda\right) \le \frac{1}{\lambda}\mathbb{E}[X_0].$$

Moreover, if X is a supermartingale with right-continuous paths,

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s| > \lambda\right) \le \frac{1}{\lambda}(\mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|]).$$

Proof. By above, we only need to prove the discrete case for a supermartingale $(X_n)_{n\ge0}$. First, let $N = \min\{n\colon X_n > \lambda\} \wedge k$. Then by the optional stopping time theorem,

$$\begin{aligned}
\mathbb{E}[X_0] \ge \mathbb{E}[X_N] &= \mathbb{E}\left[X_N \mathbb{I}_{\{\max_{0\le n\le k} X_n > \lambda\}}\right] + \mathbb{E}\left[X_N \mathbb{I}_{\{\max_{0\le n\le k} X_n \le \lambda\}}\right] \\
&\ge \lambda \mathbb{P}\left(\max_{0\le n\le k} X_n > \lambda\right) + \mathbb{E}\left[X_k \mathbb{I}_{\{\max_{0\le n\le k} X_n \le \lambda\}}\right] \\
&\ge \lambda \mathbb{P}\left(\max_{0\le n\le k} X_n > \lambda\right) - \mathbb{E}\left[|X_k|\right].
\end{aligned}$$

So we have

$$\lambda \mathbb{P}\left(\max_{0\le n\le k} X_n > \lambda\right) \le \mathbb{E}[X_0] + \mathbb{E}\left[|X_k|\right].$$

On the other hand, let $T = \min\{n\colon X_n \le -\lambda\} \wedge k$. Then

$$\begin{aligned}
\mathbb{E}[X_k] \le \mathbb{E}[X_T] &= \mathbb{E}\left[X_T \mathbb{I}_{\{\max_{0\le n\le k} X_n < -\lambda\}}\right] + \mathbb{E}\left[X_T \mathbb{I}_{\{\max_{0\le n\le k} X_n \ge -\lambda\}}\right] \\
&\le -\lambda \mathbb{P}\left(\max_{0\le n\le k} X_n < -\lambda\right) + \mathbb{E}\left[X_k \mathbb{I}_{\{\max_{0\le n\le k} X_n \ge -\lambda\}}\right]
\end{aligned}$$

So

$$\begin{aligned}
\lambda \mathbb{P}\left(\max_{0\le n\le k} X_n < -\lambda\right) &\le -\left(\mathbb{E}[X_k] - \mathbb{E}\left[X_k \mathbb{I}_{\{\max_{0\le n\le k} X_n \ge -\lambda\}}\right]\right) \\
&= \mathbb{E}\left[-X_k \mathbb{I}_{\{\max_{0\le n\le k} X_n < -\lambda\}}\right] \\
&\le \mathbb{E}\left[|X_k|\right].
\end{aligned}$$

Therefore, we have

$$\lambda \mathbb{P}\left(\max_{0\le n\le k} |X_n| > \lambda\right) \le 2\mathbb{E}\left[|X_k|\right] + \mathbb{E}\left[|X_0|\right]. \qquad \square$$

6.4 Path Regularity

Definition 6.4.1. Let $f\colon I \to \mathbb{R}$ with $I \subset \mathbb{R}_+$. For $a < b$, define

$$M_{a,b}^f(I) \overset{\text{def}}{=} \sup\{k \in \mathbb{N}_+ : \exists\{s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k\} \subset I \text{ s.t. } f(s_i) \leqslant a, f(t_i) \geqslant b, \forall i \in [k]\}$$

be the number of up-crossing of f on (a,b).

Lemma 6.4.2. *Let $D \subset \mathbb{R}_+$ be a countable dense subset of $\mathbb{R}_+$ and $f\colon D \to \mathbb{R}$. Assume that for any $T \in D$,* *(i) f is bounded on $D \cap [0,T]$,*

(ii) $M^f_{a,b}(D) < \infty$ for any $a, b \in \mathbb{Q}$ with $a < b$.

Then

(1) for any $t \geq 0$,

$$\lim_{s\downarrow t, s\in D} f(s) =: f(t+)$$

exits.

(2) for any $t > 0$,

$$\lim_{s\uparrow t, s\in D} f(s) =: f(t-)$$

exits.

Furthermore, define $g(t) = f(t+)$ then g is càdlàg (or RLCC), i.e. right-continuous with left-limit.

Proof. Assume for $t > 0$,

$$\lim_{s\downarrow t, s\in D} f(s)$$

does not exists. Then by the boundedness of f,

$$\liminf_{s\downarrow t, s\in D} f(s) < \limsup_{s\downarrow t, s\in D} f(s),$$

which implies that there exist $a, b \in \mathbb{Q}$ such that

$$\liminf_{s\downarrow t, s\in D} f(s) < a < b < \limsup_{s\downarrow t, s\in D} f(s).$$

It follows that

$$M^f_{a,b}(D) = \infty.$$

□

Theorem 6.4.3. *Let $X = (X_t)_{t\geq 0}$ be a supermartingale and D be a countable dense subset of $\mathbb{R}_+$. Then*

(1) for a.e. $\omega \in \Omega$,

$$X_{t+}(\omega) := \lim_{s\downarrow t, s\in D} X_s(\omega), \quad X_{t-}(\omega) := \lim_{s\uparrow t, s\in D} X_s(\omega)$$

exist.

(2) for every $t \in \mathbb{R}_+$, $X_{t+} \in L^1$ and

$$X_t \geq \mathbb{E}\left[X_{t+} \mid \mathcal{F}_t\right],$$

where "=" if and only if $t \mapsto \mathbb{E}[X_t]$ is right-continuous.

(3) $(X_{t+})_{t\geq 0}$ is a supermartingale w.s.t. $(\mathcal{F}_{t+})_{t\geq 0}$. Moreover, it is a martingale if X is a martingale.

Proof. (1) First, give any $T > 0$ and $\lambda > 0$, we have

$$\mathbb{P}\left(\sup_{t\in[0,T]} |X_t| > \lambda\right) \leq \frac{1}{\lambda}\left(\mathbb{E}[|X_0|] + 2\mathbb{E}\left[|X_T|\right]\right).$$

As $\lambda \to \infty$,

$$\mathbb{P}\left(\sup_{t\in[0,T]} |X_t| < \infty\right) = 1.$$

Therefore, $\sup_{t\in[0,T]} |X_t| < \infty$ a.e.. Second, choose a sequence (D_m) of finite subsets of D such that $D_m \uparrow D$ and $T \in D_m$. By the upcrossing inequality of the discrete case of $(X_{s_k}, k \in D_m \cap [0,T])$,

$$\mathbb{E}\left[M^X_{a,b}(D_m \cap [0,T])\right] \leq \frac{1}{b-a}\mathbb{E}\left[(X_T - a)^-\right].$$

As $m \to \infty$, by DCT,

$$\mathbb{E}\left[M^X_{a,b}(D \cap [0,T])\right] \leq \frac{1}{b-a}\mathbb{E}\left[(X_T - a)^-\right] < \infty,$$

which implies that $M^X_{a,b}(D\cap[0,T]) < \infty$ a.e.. By choosing the union of such zero measure set, the above lemma implies X_{t+}, X_{t-} exist a.e..

(2) First, choose a sequence $(t_n)_{n\in\mathbb{N}} \subset D$ such that $t_n \downarrow t$ and $t_n \leq t+1$, so $X_{t+} = \lim_{n\to\infty} X_{t_n}$. Note that (X_{t_n}) is a backward supermartingale (by let $Y_{-n} = X_{t_n}$) and

$$\sup_n \mathbb{E}[|X_{t_n}|] \leq \sup_{s\in[0,t+1]} \mathbb{E}[|X_t|] < \infty.$$

By Martingale Convergence Theorem for discrete case (backward case),

$$\lim_{n\to\infty} X_{t_n} = X_{t+}$$

in L^1 and so $X_{t+} \in L^1$. Note that

$$X_t \geq \mathbb{E}\left[X_{t_n} \mid \mathcal{F}_t\right].$$

Because $X_{t_n} \to X_{t+}$ in L^1,

$$X_t \geq \mathbb{E}\left[X_{t+} \mid \mathcal{F}_t\right].$$

Note that if $X_1 \geq X_2$ and $\mathbb{E}[X_1] = \mathbb{E}[X_2]$, then $X_1 = X_2$. Assume $t \mapsto \mathbb{E}[X_t]$ is right-continuous. Then

$$\mathbb{E}[X_t] = \lim_{t_n\to t} \mathbb{E}[X_{t_n}] = \mathbb{E}[X_{t+}] = \mathbb{E}\left[\mathbb{E}\left[X_{t+} \mid \mathcal{F}_t\right]\right].$$

It follows that $X_t = \mathbb{E}\left[X_{t+} \mid \mathcal{F}_t\right]$. Conversely, it is obvious.

(3) Let $s < t$. Choose $(s_n)_n \subset D$ and $(t_n)_n \subset D$ such that $s_n \downarrow s$ and $t_n \downarrow t$ and $s_n \leq t_n$. Then

$$X_{s_n} \geq \mathbb{E}\left[X_{t_n} \mid \mathcal{F}_{s_n}\right], \quad \forall\, n.$$

Now for any $A \in \mathcal{F}_{s+} = \bigcap_n \mathcal{F}_{s_n}$, we have

$$\mathbb{E}[X_{s_n}\mathbb{I}_A] \geq \mathbb{E}[X_{t_n}\mathbb{I}_A].$$

As $n \to \infty$, by (2),

$$\mathbb{E}[X_{s+}\mathbb{I}_A] \geq \mathbb{E}[X_{t+}\mathbb{I}_A] = \mathbb{E}\left[\mathbb{I}_A\mathbb{E}[X_{t+} \mid \mathcal{F}_{s+}]\right].$$

Because X_{s+} and $\mathbb{E}[X_{t+} \mid \mathcal{F}_{s+}]$ are $\mathcal{F}_{s+}$-measurable,

$$X_{s+} \geq \mathbb{E}[X_{t+} \mid \mathcal{F}_{s+}]. \qquad \square$$

Theorem 6.4.4 (Regularizing Path). *Assume $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous and complete. Let $X = (X_t)_{t\geq 0}$ be a supermartingale such that $t \mapsto \mathbb{E}[X_t]$ is right-continuous. Then there is a $(Y_t)_{t\geq 0}$ such that it has càdlàg path and it is a supermartingale and $Y_t = X_t$ a.e. (called a modification of X).*

Proof. Let $Y_t = X_{t+}$. Then it is a supermartingale w.s.t. $(\mathcal{F}_{t+})_{t\geq 0} = (\mathcal{F}_t)_{t\geq 0}$, which has càdlàg path. Moreover, by the right-continuity of $t \mapsto \mathbb{E}[X_t]$,

$$X_t = \mathbb{E}[X_{t+} \mid \mathcal{F}_t] = \mathbb{E}[Y_t \mid \mathcal{F}_t] = Y_t.$$

□

6.5 Convergence Theorem

Theorem 6.5.1 (Martingale Convergence Theorem). *Let $X = (X_t)_{t\geq 0}$ be a supermartingale with right-continuous paths such that*

$$\sup_t \mathbb{E}[|X_t|] < \infty$$

Then there exists a $X_\infty \in L^1$ such that

$$\lim_{t\to\infty} X_t = X_\infty$$

a.e..

Proof. Let D be a countable and dense subset of $\mathbb{R}_+$. For any $T \in D$ and $a < b$,

$$\mathbb{E}\left[M^X_{a,b}(D \cap [0,T])\right] \leq \frac{1}{b-a}\mathbb{E}[(X_T - a)^-].$$

So

$$\mathbb{E}\left[M^X_{a,b}(D \cap [0,T])\right] \leq \frac{1}{b-a}\sup_T \mathbb{E}[(X_T - a)^-] = M < \infty.$$

As $T \to \infty$, by MCT,

$$\mathbb{E}\left[M^X_{a,b}(D)\right] \leq M < \infty \quad \Rightarrow \quad M^X_{a,b}(D) < \infty,\ \forall\ a, b$$

Therefore, by above lemma,

$$X_\infty = \lim_{D\ni t\to\infty} X_t.$$

By Fatou's lemma,

$$\mathbb{E}\left[|X_\infty|\right] \leq \liminf_{D\ni t\to\infty} \mathbb{E}\left[|X_t|\right] \leq \sup_t \mathbb{E}[|X_t|] < \infty.$$

Therefore, for any $\varphi > 0$, there exists N such that for all $D \ni t \geq N$,

$$|X_t - X_\infty| < \varepsilon.$$

Then for any $s \geq N$, let $s_n \downarrow s$ in D, so

$$|X_{s_n} - X_\infty| < \varepsilon.$$

As $n \to \infty$, because (X_t) is right-continuous,

$$|X_s - X_\infty| < \varepsilon,$$

which implies that

$$X_\infty = \lim_{t\to\infty} X_t.$$

□

Definition 6.5.2 (Closedness)**.** A martingale $X = (X_t)_{t\geq 0}$ is called closed if there exists a $Z \in L^1$ such that

$$X_t = \mathbb{E}[Z \mid \mathcal{F}_t], \quad \forall\, t \geq 0.$$

Theorem 6.5.3. *Let $X = (X_t)_{t\geq 0}$ be a martingale with RLCC path. TFAE.*

(1) X is closed.

(2) X is UI.

(3) X_t converges a.e. and in L^1.

Moreover, in such cases, $X_t = \mathbb{E}[X_\infty \mid \mathcal{F}_t]$ for $X_\infty = \lim_t X_t$.

Proof. (1) $\Rightarrow$ (2): It has been proved in above chapter.
(2) $\Rightarrow$ (3): Because X is UI,

$$\sup_t \mathbb{E}[|X_t|] < \infty.$$

Then by above theorem,

$$\lim_{t\to\infty} X_t = X_\infty$$

a.e.. Then because of UI, it is in L^1.
(3) $\Rightarrow$ (1): For $t < T$,

$$X_t = \mathbb{E}[X_T \mid \mathcal{F}_t].$$

As $T \to \infty$, because $X_\infty = \lim_t X_t$ in L^1,

$$X_t = \mathbb{E}\left[X_\infty \mid \mathcal{F}_t\right]. \qquad \square$$

Remark 6.5.4. In such case, for any stopping time T, we can define

$$X_T(\omega) := \mathbb{I}_{T(\omega)<\infty} X_{T(\omega)}(\omega) + \mathbb{I}_{T(\omega)=\infty} X_\infty(\omega).$$

6.6 Optional Stopping Time

Theorem 6.6.1. *Let $(Y_n)_{n\geq 0}$ be a discrete UI martingale. Then for any stopping times $M \leq N$,*

$$Y_M = \mathbb{E}[Y_N \mid \mathcal{F}_M].$$

In particular, when M, N are bounded, no UI is required.

Proof. First, $Y_M \in L^1$ is by $(Y_{n\wedge M})$ is a UI martingale and $Y_{n\wedge M} \to Y_M$ (If M, N bounded, $Y_N = \sum_k \mathbb{I}_{\{N=k\}} Y_k \in L^1$). We already know Y_M is $\mathcal{F}_M$ measurable. For any $A \in \mathcal{F}_M \subset \mathcal{F}_N$, consider

$$M^A = \begin{cases} M, & \omega \in A \\ \infty, & \omega \in A^c \end{cases}, \quad N^A = \begin{cases} N, & \omega \in A \\ \infty, & \omega \in A^c, \end{cases}$$

and they are stopping times because $A \in \mathcal{F}_M, \mathcal{F}_N$. $M^A \leq N^A$. Moreover, by optional stopping time theorem

$$\mathbb{E}[Y_0] = \mathbb{E}\left[Y_{M^A}\right] = \mathbb{E}\left[Y_{N^A}\right].$$

Note that

$$\mathbb{E}\left[Y_{M^A}\right] = \mathbb{E}[Y_M \mathbb{I}_A] + \mathbb{E}[Y_\infty \mathbb{I}_{A^c}], \quad \mathbb{E}\left[Y_{N^A}\right] = \mathbb{E}[Y_N \mathbb{I}_A] + \mathbb{E}[Y_\infty \mathbb{I}_{A^c}].$$

So

$$\mathbb{E}[Y_M \mathbb{I}_A] = \mathbb{E}[Y_N \mathbb{I}_A].$$

It follows that

$$Y_M = \mathbb{E}[Y_N \mid \mathcal{F}_M]. \qquad \square$$

Remark 6.6.2. For super(sub)martingale, it has similar result. Let $(Z_n)_{n\geq 0}$ be a discrete UI supermartingale. Then for any stopping times $S \leq T$,

$$\mathbb{E}[Z_T \mid \mathcal{F}_S] \leq Z_S$$

In particular, when $S, T \leq m$ are bounded, no UI is required. First, $Z_S, Z_T \in L^1$ as same as above. Let $A \in \mathcal{F}_S$. Note that $\{Z_{T\wedge n}\}_{n\geq 0}$ is still a supermartingale. First, $A \cap \{S = k\} \in \mathcal{F}_k$. So

$$\begin{aligned}
\mathbb{E}[Z_T \mathbb{I}_A] &= \mathbb{E}[Z_{T\wedge m}\mathbb{I}_A] \\
&= \sum_{k=0}^{m} \mathbb{E}\left[\mathbb{I}_{A\cap\{S=k\}} Z_{T\wedge m}\right] \\
&= \sum_{k=0}^{m} \mathbb{E}\left[\mathbb{I}_{A\cap\{S=k\}} \mathbb{E}\left[Z_{T\wedge m} \mid \mathcal{F}_k\right]\right] \\
&\leq \sum_{k=0}^{m} \mathbb{E}\left[\mathbb{I}_{A\cap\{S=k\}} Z_{T\wedge k}\right] \\
&= \sum_{k=0}^{m} \mathbb{E}\left[\mathbb{I}_{A\cap\{S=k\}} Z_{T\wedge S}\right] \\
&= \sum_{k=0}^{m} \mathbb{E}\left[\mathbb{I}_{A\cap\{S=k\}} Z_{S}\right] \\
&= \mathbb{E}\left[\mathbb{I}_A Z_S\right].
\end{aligned}$$

Theorem 6.6.3 (Doob's Optional Stopping Time)**.** *Let $X = (X_t)_{t\geq 0}$ be a UI martingale with right-continuous paths. Let S and T be two stopping times with $S \leq T$. Then $X_S, X_T \in L^1$ and*

$$X_S = \mathbb{E}[X_T \mid \mathcal{F}_S].$$

Proof. First, it is obvious X_S is L^1 and $\mathcal{F}_S$ measurable. Set for any integer $n > 0$,

$$T_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{I}_{\frac{k}{2^n} < T \leq \frac{k+1}{2^n}} + \infty \mathbb{I}_{T=\infty},$$

and

$$S_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{I}_{\frac{k}{2^n} < S \leq \frac{k+1}{2^n}} + \infty \mathbb{I}_{S=\infty}.$$

Note that T_n and S_n are stopping times with $T_n \downarrow T$ and $S_n \downarrow S$ and $S_n \leq T_n$. Since $(X_{\frac{k}{2^n}})_{k\geq 0}$ is a UI martingale, by above theorem,

$$X_{S_n} = \mathbb{E}[X_{T_n} \mid \mathcal{F}_{S_n}].$$

Let $A \in \mathcal{F}_S$. Because $\mathcal{F}_S \subset \mathcal{F}_{S_n}$,

$$\mathbb{E}[\mathbb{I}_A X_{S_n}] = \mathbb{E}[\mathbb{I}_A X_{T_n}].$$

Let $n \to \infty$. By the right-continuity of path and UI,

$$\mathbb{E}[\mathbb{I}_A X_S] = \mathbb{E}[\mathbb{I}_A X_T].$$

□

Corollary 6.6.4 (Bounded Optional Stopping Time)**.** *Let $X = (X_t)_{t\geq 0}$ be a martingale with right-continuous paths. Let $S \leq T$ be two bounded stopping times. Then $X_S, X_T \in L^1$ and*

$$X_S = \mathbb{E}[X_T \mid \mathcal{F}_S].$$

Proof. Assume $S \leq T \leq a$ for some constant a. Consider the martingale $Y_t = X_{t\wedge a}$. Then because $Y_t = \mathbb{E}[X_a \mid \mathcal{F}_t]$, (Y_t) is UI. So

$$Y_S = X_{S\wedge a} = X_S, \quad Y_T = X_{T\wedge a} = X_T \in L^1,$$

and $X_S = \mathbb{E}[X_T \mid \mathcal{F}_S]$. □

Corollary 6.6.5. *Let $X = (X_t)_{t\geq 0}$ be a martingale with right-continuous paths. Let T be a stopping time. Then*

(1) the process $(X_{t\wedge T})_{t\geq 0}$ is also a martingale.

(2) if X is UI, then $(X_{t\wedge T})_{t\geq 0}$ is UI and

$$X_{t\wedge T} = \mathbb{E}[X_T \mid \mathcal{F}_t].$$

Proof. For (2), note $t \wedge T$ is also a stopping time with $t \wedge T \leq T$ and $X_{t\wedge T}$ is $\mathcal{F}_{t\wedge T}$-measurable and sp $\mathcal{F}_t$-measurable. For any $A \in \mathcal{F}_t$, let

$$A = (A \cap \{T \leq t\}) \cup (A \cap \{T > t\}).$$

Then

$$\begin{aligned}
\mathbb{E}[X_{t\wedge T}\mathbb{I}_A] &= \mathbb{E}[X_{t\wedge T}\mathbb{I}_{A\cap\{T\leq t\}}] + \mathbb{E}[X_{t\wedge T}\mathbb{I}_{A\cap\{T>t\}}], \\
&= \mathbb{E}[X_T\mathbb{I}_{A\cap\{T\leq t\}}] + \mathbb{E}[X_{t\wedge T}\mathbb{I}_{A\cap\{T>t\}}], \\
\mathbb{E}[X_T\mathbb{I}_A] &= \mathbb{E}[X_T\mathbb{I}_{A\cap\{T\leq t\}}] + \mathbb{E}[X_T\mathbb{I}_{A\cap\{T>t\}}],
\end{aligned}$$

so to prove $\mathbb{E}[X_{t\wedge T}\mathbb{I}_A] = \mathbb{E}[X_T\mathbb{I}_A]$, it suffices to show

$$\mathbb{E}[X_{t\wedge T}\mathbb{I}_{A\cap\{T>t\}}] = \mathbb{E}[X_T\mathbb{I}_{A\cap\{T>t\}}].$$

Because X is UI, by above

$$X_{t\wedge T} = \mathbb{E}[X_T \mid \mathcal{F}_{t\wedge T}].$$

So we only need to show $A \cap \{T > t\} \in \mathcal{F}_{t\wedge T}$.

$$(A \cap \{T > t\}) \cap \{t \wedge T \leq s\} = A \cap \{T > t\} \cap \{t \leq s\} \in \mathcal{F}_t \subset \mathcal{F}_s$$

It follows

$$X_{t\wedge T} = \mathbb{E}[X_T \mid \mathcal{F}_t].$$

And $X_T \in L^1$ by above theorem, so $(X_{t\wedge T})_{t\geq 0}$ is UI.
For (1), let $a > 0$ and consider $(X_{t\wedge a})_{t\geq 0}$ that is obvious a UI martingale. So by (2),

$$X_{t\wedge a\wedge T} = X_{t\wedge T}$$

is a martingale for any $t \geq a$. □

Remark 6.6.6. In fact, if $X = (X_t)_{t\geq 0}$ be a submartingale with right-continuous paths, then $(X_{t\wedge T})_{t\geq 0}$ is also a submartingale for any stopping time T by discretization. Moreover, by the same reasoning as the discrete case, if $X = (X_t)_{t\geq 0}$ is a UI submartingale, so is $(X_{t\wedge T})_{t\geq 0}$.

Example 6.6.7. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Let $a \in \mathbb{R}$ and define

$$T_a = \inf\{t \geq 0\colon B_t = a\}.$$

Assume $a < 0 < b$. Define $T = T_a \wedge T_b$.

(1) Determine the following probability: $\mathbb{P}(T_a < T_b)$ and $\mathbb{P}(T_b \leq T_a)$.

(2) Find $\mathbb{E}[T]$.

(3) For $\lambda > 0$, find $\mathbb{E}\left[e^{-\lambda T_a}\right]$.

Solution:

(1) Because $N_t = B_{t\wedge T}$ is a martingale and bounded by $|a| + b$, it is UI. So by optional stopping time theorem,

$$\mathbb{E}\left[N_T\right] = \mathbb{E}\left[B_T\right] = \mathbb{E}[N_0] = 0$$

On the other hand,

$$\mathbb{E}[B_T] = a\mathbb{P}(T_a < T_b) + b\mathbb{P}(T_b \leq T_a) = 0$$

So

$$\mathbb{P}(T_a < T_b) = \frac{b}{b-a}, \quad \mathbb{P}(T_b \leq T_a) = \frac{-a}{b-a}.$$

(2) We already know $M_t = B_t^2 - t$ is a martingale. So $M_{t\wedge T} = B_{t\wedge T}^2 - t\wedge T$ is also a martingale.

$$\mathbb{E}[M_{t\wedge T}] = 0 \;\Rightarrow\; \mathbb{E}\left[B_{t\wedge T}^2\right] = \mathbb{E}\left[t\wedge T\right].$$

As $t \to \infty$, by the DCT on the LHS and the MCT on the RHS,

$$\mathbb{E}[T] = \mathbb{E}\left[B_T^2\right] = -ab.$$

(3) Fix a $b \in \mathbb{R}$. Consider the martingale

$$N_t^b = \exp\left(bB_t - \frac{1}{2}b^2 t\right).$$

Assume $b > 0$. Note that $(N_{t\wedge T_a}^b)_{t\geq 0}$ is also a martingale. Moreover, because

$$\left|N_{t\wedge T_a}^b\right| = \exp\left(bB_{t\wedge T_a} - \frac{1}{2}b^2 t\wedge T_a\right) \leq \exp\left(bB_{t\wedge T_a}\right) \leq \exp(b\,|a|),$$

$Y_t = N_{t\wedge T_a}^b$ is UI. So by the optional stopping theorem,

$$\mathbb{E}[Y_\infty] = \mathbb{E}\left[N_{T_a}^b\right] = \mathbb{E}[Y_0] = 1.$$

On the other hand,

$$\mathbb{E}\left[N_{T_a}^b\right] = \mathbb{E}\left[\exp\left(bB_{T_a} - \frac{1}{2}b^2 T_a\right)\right] = \mathbb{E}\left[\exp\left(ba - \frac{1}{2}b^2 T_a\right)\right] = \exp(ab)\mathbb{E}\left[e^{-\frac{1}{2}b^2 T_a}\right].$$

So let $b = \sqrt{2\lambda}$.

$$\mathbb{E}\left[e^{-\lambda T_a}\right] = e^{-\sqrt{2\lambda}a}.$$

Theorem 6.6.8. *Assume $Z = (Z_t)_{t\geq 0}$ is a nonnegative supermartingale with right-continuous paths. Let U, V be stopping times with $U \leq V$. Then $Z_U, Z_V \in L^1$ and*

$$Z_U \geq \mathbb{E}\left[Z_V \mid \mathcal{F}_U\right].$$

Proof. (i) First, assume $U \leq V \leq P$ for some integer P. Let

$$U_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{I}_{\left\{\frac{k}{2^n} < U \leq \frac{k+1}{2^n}\right\}}, \quad V_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{I}_{\left\{\frac{k}{2^n} < V \leq \frac{k+1}{2^n}\right\}}.$$

So they are stopping times with $U_n \downarrow U$, $V_n \downarrow V$, and $U_n \leq V_n$. Because of the right-continuity of paths,

$$Z_{U_n} \to Z_U, \quad Z_{V_n} \to Z_V.$$

Moreover, we can consider $(Y_k = Z_{U_{-k}})_{k\leq 0}$. Because $U_{n+1} \leq U_n$, by the optional stopping time theorem (bounded case),

$$\mathbb{E}\left[Z_{U_n} \mid \mathcal{F}_{U_{n+1}}\right] \leq Z_{U_{n+1}},$$

so (Y_k) is a backward supermartingale with $\mathbb{E}[Y_k] \leq \mathbb{E}[Z_0] < \infty$. It follows that (Y_k) is UI and so $Z_{U_n} \to Z_U$ in L^1, so is $Z_{V_n} \to Z_V$ in L^1.

Because $U_n \leq V_n$, by the optional stopping time theorem (bounded case),

$$\mathbb{E}\left[Z_{V_n} \mid \mathcal{F}_{U_n}\right] \leq Z_{U_n},$$

which implies that

$$\mathbb{E}[Z_{V_n}] \leq Z_{U_n}.$$

As $n \to \infty$, by L^1-convergence

$$\mathbb{E}[Z_V] \leq \mathbb{E}\left[Z_U\right].$$

(ii) Next, consider general $U \leq V$. It is obvious Z_U is $\mathcal{F}_U$-measurable. To prove

$$\mathbb{E}[Z_V \mid \mathcal{F}_U] \leq Z_U,$$

it suffices to prove that for any $A \in \mathcal{F}_U$,

$$\mathbb{E}[Z_V \mathbb{I}_A] \leq \mathbb{E}[Z_U \mathbb{I}_A].$$

Define

$$U^A = \begin{cases} U, & \omega \in A \\ \infty, & \omega \in A^c \end{cases}, \quad V^A = \begin{cases} V, & \omega \in A \\ \infty, & \omega \in A^c \end{cases}.$$

They are stopping times with $U^A \leq V^A$ because $A \in \mathcal{F}_U \subset \mathcal{F}_V$. For any $p \geq 1$, by above step, we have

$$\mathbb{E}\left[Z_{V^A\wedge p}\right] \leq \mathbb{E}\left[Z_{U^A\wedge p}\right]$$

For the RHS,

$$\begin{aligned}\mathbb{E}\left[Z_{U^A\wedge p}\right] &= \mathbb{E}\left[Z_{U^A\wedge p}\mathbb{I}_A\right] + \mathbb{E}\left[Z_{U^A\wedge p}\mathbb{I}_{A^c}\right] \\ &= \mathbb{E}\left[Z_{U\wedge p}\mathbb{I}_A\right] + \mathbb{E}\left[Z_p\mathbb{I}_{A^c}\right],\end{aligned}$$

and we have the similar formula for the LHS. So

$$\mathbb{E}\left[Z_{V\wedge p}\mathbb{I}_A\right] \leq \mathbb{E}\left[Z_{U\wedge p}\mathbb{I}_A\right]$$

Note that

$$\mathbb{I}_A = \mathbb{I}_{A\cap\{U\le p\}} + \mathbb{I}_{A\cap\{U>p\}}.$$

So

$$\mathbb{E}\left[Z_{U\wedge p}\mathbb{I}_A\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{U\le p\}}Z_U\right] + \mathbb{E}\left[\mathbb{I}_{A\cap\{U>p\}}Z_p\right]$$

and

$$\mathbb{E}\left[Z_{V\wedge p}\mathbb{I}_A\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{U\le p\}}Z_{V\wedge p}\right] + \mathbb{E}\left[\mathbb{I}_{A\cap\{U>p\}}Z_p\right].$$

It follows that

$$\mathbb{E}\left[\mathbb{I}_{A\cap\{U\le p\}}Z_U\right] \ge \mathbb{E}\left[\mathbb{I}_{A\cap\{U\le p\}}Z_{V\wedge p}\right] \ge \mathbb{E}\left[\mathbb{I}_{A\cap\{V\le p\}}Z_{V\wedge p}\right]$$

because Z is nonnegative. By MCT, as $p\to\infty$,

$$\mathbb{E}[Z_V\mathbb{I}_A] \le \mathbb{E}[Z_U\mathbb{I}_A]. \qquad \square$$

Remark 6.6.9. Note that if $Z=(Z_t)_{t\ge 0}$ is a UI supermartingale, then above is also clearly true by as $p\to\infty$ in (ii).

Proposition 6.6.10. *Let $X=(X_t)_{t\ge 0}$ be a adapted and right-continuous and integrable stochastic process satisfying $X_T\in L^1$ for all bounded stopping time T. Then X is a martingale if and only if*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0],$$

Proof. For any $0\le s<t$ and any $A\in\mathcal{F}_s$, let

$$T = s\mathbb{I}_{A^c} + t\mathbb{I}_A.$$

Then T is a bounded stopping time and

$$\mathbb{E}[X_0] = \mathbb{E}[X_T] = \mathbb{E}\left[X_s\mathbb{I}_{A^c}\right] + \mathbb{E}\left[X_t\mathbb{I}_A\right] = \mathbb{E}\left[X_s\right] + \mathbb{E}\left[(X_t - X_s)\mathbb{I}_A\right].$$

Because $\mathbb{E}[X_0]=\mathbb{E}[X_s]$,

$$\mathbb{E}\left[X_t\mathbb{I}_A\right] = \mathbb{E}\left[X_s\mathbb{I}_A\right].$$

So

$$X_s = \mathbb{E}[X_t \mid \mathcal{F}_s]. \qquad \square$$

Remark 6.6.11. Furthermore, if above conditions are satisfied for all stopping times T, then X is UI.

Chapter 7

Continuous Time Markov Theory

7.1 Transition Semigroup

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space. Let $(E, \mathcal{E})$ be a measurable space.

Definition 7.1.1 (Markov Process)**.** A E-valued stochastic process $(X_t)_{t\geq 0}$ is called $(\mathcal{F}_t)_{t\geq 0}$-Markov process if

(i) X_t is $\mathcal{F}_t$-adapted,

(ii) for any $t > s$ and any $f \in \mathcal{B}_b(E)$ (bounded measurable function),

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = \mathbb{E}\left[f(X_t) \mid \sigma(X_s)\right].$$

Remark 7.1.2. If $(X_t)_{t\geq 0}$ is a Markov process, it is obvious a Markov process w.s.t. its natural filtration $\mathcal{F}_t^X = \sigma(X_s \colon s \leq t)$.

Definition 7.1.3 (Transition Kernel)**.** A Markov transition kernel from E to E is a map

$$Q\colon E \times \mathcal{E} \to [0,1]$$

such that

(i) for any $x \in E$, $Q(x, \cdot)$ is a probability measure on $(E, \mathcal{E})$.

(ii) for any $A \in \mathcal{E}$, $Q(\cdot, A)$ is $\mathcal{E}$-measurable.

Remark 7.1.4. Given a Markov transition kernel Q, it can define

$$Q\colon \mathcal{B}_b(E) \to \mathcal{B}_b(E)$$

as

$$Qf(x) := \int_E f(y) Q(x, dy),$$

which is a linear operator.

Definition 7.1.5 (Transition Semigroup)**.** A collection $(Q_t)_{t\geq 0}$ of transition kernels on E is called a transition semigroup if

(i) for $x \in E$, $Q_0(x, dy) = \delta_x(dy)$,

(ii) for $s, t \geq 0$ and $A \in \mathcal{E}$,

$$Q_{t+s}(x, A) = \int_E Q_t(x, dy) Q_s(y, A).$$

(Chapman-Kolmogorov equation)

(iii) for any $A \in \mathcal{E}$, $(t, x) \mapsto Q_t(x, A)$ is measurable.

Remark 7.1.6. Note that a transition semigroup induces a semigroup of operators $(Q_t)_{t \geq 0}$. Let $\mathcal{B}_b(E)$ be equipped with $\|\cdot\| = \|\cdot\|_\infty$. Then

(i) $Q_0 f(x) = \int_E f(y) \delta_x(dy) = f(x)$, i.e., $Q_0 = \mathrm{Id}$.

(ii) $Q_t \mathbb{1}(x) = \int_E \mathbb{1}(y) Q_t(x, dy) = 1$, i.e., $Q_t \mathbb{1} = \mathbb{1}$.

(iii) for any $f \geq 0$, $Q_t f \geq 0$.

(iv) for any $s, t \geq 0$,

$$\begin{aligned}
Q_{t+s} f(x) &= \int_E f(y) Q_{t+s}(x, dy) \\
&= \int_E f(y) \int_E Q_t(x, dz) Q_s(z, dy) \\
&= \int_E \left(\int_E f(y) Q_s(z, dy) \right) Q_t(x, dz) \\
&= \int_E Q_s f(z) Q_t(x, dz) \\
&= Q_t \left(Q_s f \right)(x),
\end{aligned}$$

i.e., $Q_{t+s} = Q_t \circ Q_s = Q_t Q_s$.

Definition 7.1.7. A Markov process $X = (X_t)_{t \geq 0}$ with transition semigroup $(Q_t)_{t \geq 0}$ is a $(\mathcal{F}_t)_{t \geq 0}$-adapted process with values in E such that for any $s, t \geq 0$ and any $f \in \mathcal{B}_b(E)$,

$$\mathbb{E}\left[f(X_{t+s}) \mid \mathcal{F}_s\right] = Q_t f(X_s).$$

Remark 7.1.8. Note that it is true Markov because

$$\begin{aligned}
\mathbb{E}\left[f(X_{t+s}) \mid \sigma(X_s)\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X_{t+s}) \mid \mathcal{F}_s\right] \mid \sigma(X_s)\right] \\
&= \mathbb{E}\left[Q_t f(X_s) \mid \sigma(X_s)\right] \\
&= Q_t f(X_s) = \mathbb{E}\left[f(X_{t+s}) \mid \mathcal{F}_s\right].
\end{aligned}$$

Moreover, if $f = \mathbb{I}_A$, then

$$Q_t(X_s, A) = \mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_s).$$

Theorem 7.1.9 (Finite-dimensional Distribution)**.** *Given a Markov process $X = (X_t)_{t \geq 0}$ with transition semigroup $(Q_t)_{t \geq 0}$ and $X_0 \sim \gamma(dx)$. For any $0 < t_1 < \cdots < t_p$,*

$$\begin{aligned}
&\mathbb{P}(X_0 \in A_0, X_{t_1} \in A_1, \cdots X_{t_p} \in A_p) \\
&= \int_{A_0} \gamma(dx) \int_{A_1} Q_{t_1}(x, dx_1) \int_{A_2} Q_{t_2 - t_1}(x_1, dx_2) \cdots \int_{A_p} Q_{t_p - t_{p-1}}(x_{p-1}, dx_p).
\end{aligned}$$

More generally, for any $f_i \in \mathcal{B}_b(E)$ $(i = 0, 1, \cdots, p)$,

$$\begin{aligned}
&\mathbb{E}[f_0(X_0) f_1(X_{t_1}) \cdots f_p(X_{t_p})] \\
&= \int_E f_0(X_0) \gamma(dx) \int_E f_1(x_1) Q_{t_1}(x, dx_1) \int_E f_2(x_2) Q_{t_2 - t_1}(x_1, dx_2) \cdots \int_E f_p(x_p) Q_{t_p - t_{p-1}}(x_{p-1}, dx_p).
\end{aligned}$$

Proof. For $p = 1$,

$$\begin{aligned}
\mathbb{E}[f_0(X_0)f_1(X_{t_1})] &= \mathbb{E}\left[f_0(X_0)\left[f_1(X_{t_1}) \mid \mathcal{F}_0\right]\right] \\
&= \mathbb{E}\left[f_0(X_0)Q_{t_1}f_1(X_0)\right] \\
&= \int_E f_0(x)Q_{t_1}f_1(x)\gamma(dx) \\
&= \int_E f_0(x)\gamma(dx)\int_E f_1(x_1)Q_{t_1}(x, dx_1).
\end{aligned}$$

Assume it is true for $p-1$. Then

$$\begin{aligned}
&\mathbb{E}[f_0(X_0)f_1(X_{t_1})\cdots f_p(X_{t_p})] \\
&= \mathbb{E}\left[\mathbb{E}\left[f_0(X_0)f_1(X_{t_1})\cdots f_p(X_{t_p}) \mid \mathcal{F}_{t_{p-1}}\right]\right] = \mathbb{E}\left[f_0(X_0)\cdots f_{p-1}(X_{t_{p-1}})\mathbb{E}\left[f_p(X_{t_p}) \mid \mathcal{F}_{t_{p-1}}\right]\right] \\
&= \mathbb{E}\left[f_0(X_0)\cdots f_{p-1}(X_{t_{p-1}})Q_{t_p-t_{p-1}}f_p(X_{t_{p-1}})\right] \\
&= \int_E f_0(X_0)\gamma(dx)\cdots\int_E f_{p-1}(x_{p-1})Q_{t_p-t_{p-1}}f_p(x_{p-1})Q_{t_{p-1}-t_{p-2}}(x_{p-2}, dx_{p-1}) \\
&= \int_E f_0(X_0)\gamma(dx)\cdots\int_E f_{p-1}(x_{p-1})Q_{t_{p-1}-t_{p-2}}(x_{p-2}, dx_{p-1})\int_E f_p(x_p)Q_{t_p-t_{p-1}}(x_{p-1}, dx_p).
\end{aligned}$$

□

Construction of Markov Process: Given a transition semigroup $(Q_t)_{t\geq 0}$ and an initial distribution γ. First, let

$$\Omega^* = E^{[0,\infty)} := \{\omega\colon \omega(\cdot)\colon [0,\infty) \to E\}$$

with the coordinate process $X = (X_t)_{t\geq 0}$ defined as

$$X_t\colon \Omega^* \to E, \quad X_t(\omega) = \omega(t).$$

Then σ-field $\mathcal{F}^* := \sigma(X_t\colon t \geq 0)$. For any finite subset $U = \{0 \leq t_1 < t_2 < \cdots < t_p\}$ of $[0,\infty)$, define a probability measure on $E^U \cong E^p$,

$$\mu^U(A_1 \times \cdots A_p) := \int_{A_0}\gamma(dx)\int_{A_1}Q_{t_1}(x, dx_1)\int_{A_2}Q_{t_2-t_1}(x_1, dx_2)\cdots\int_{A_p}Q_{t_p-t_{p-1}}(x_{p-1}, dx_p).$$

Note that for $\left\{\mu^U : U \text{ finite.}\right\}$, if $U \subset V$ and let $\pi_U^V\colon E^V \to E^U$ be the natural projection, then

$$\mu^U = (\pi_U^V)_\#\mu^V, \text{ i.e. } \mu^U(A_1 \times \cdots \times A_{p_U}) = \mu^V(A_1 \times \cdots \times A_{p_U} \times E \times \cdots \times E).$$

Then by the Kolmogorov Extension Theorem, there exists a unique $\mathbb{P}^*$ on $(\Omega^*, \mathcal{F}^*)$ such that

$$\mathbb{P}^*(X_0 \in A_0, X_{t_1} \in A_1, \cdots, X_{t_p} \in A_p) = \mu^U(A_1 \times \cdots A_p).$$

Therefore, the coordinate process $(X_t)_{t\geq 0}$ is a Markov process on $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ with semigroup $(Q_t)_{t\geq 0}$. Because $\mathbb{P}^*$ is determined by γ, $\mathbb{P}^* = \mathbb{P}_\gamma$. In particular, $\gamma(dy) = \delta_x(dy)$, $\mathbb{P}_\gamma = \mathbb{P}_x$.

Remark 7.1.10. For $A \in \mathcal{E}^U$, let

$$\{\omega \in \Omega^*\colon (\omega(t_1), \cdots, \omega(t_p)) \in A\}$$

be called a finite-dimensional cylinder. Let $\mathcal{C}$ be the set of finite-dimensional cylinders. Then in fact $\mathcal{F}^* = \sigma(\mathcal{C})$.

Remark 7.1.11. For any Markov process $(X_t)_{t\geq 0}$ on $(\Omega, \mathbb{P})$ with semigroup $(Q_t)_{t\geq 0}$ and $X_0 \sim \gamma$, we can construct $\mathbb{P}_\gamma$ on $(\Omega^*, \mathcal{F}^*)$ by $(Q_t)_{t\geq 0}$. Then we have $X_\#\mathbb{P} = \mathbb{P}_\gamma$ and $(X_t)_{t\geq 0}$ has the same finite-dimensional distribution as the coordinate process $(\pi_t)_{t\geq 0}$ on $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$.

Example 7.1.12. If

$$Q_t(x, dy) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-y)^2}{2t}}dy$$

then the Markov process with $X_0 = 0$, then the corresponding Markov process is a standard Brownian motion.

7.2 Resolvent

Definition 7.2.1 (Resolvent)**.** Let $\lambda > 0$. The λ-resolvent of the transition semigroup $(Q_t)_{t\geq 0}$ is a linear operator $R_\lambda \colon \mathcal{B}_b(E) \to \mathcal{B}_b(E)$ defined as

$$R_\lambda f(x) := \int_0^\infty e^{-\lambda t} Q_t f(x) dt,$$

or formally, $R_\lambda = \int_0^\infty e^{-\lambda t} Q_t dt$.

Proposition 7.2.2. *Given a transition semigroup $(Q_t)_{t\geq 0}$ and the corresponding R_λ.*

(1) $\|R_\lambda f\| \leq \frac{1}{\lambda} \|f\|$.

(2) If $0 \leq f \leq 1$, $0 \leq \lambda R_\lambda f \leq 1$.

(3) If $\lambda, \mu > 0$, *then*
$$R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0.$$

Proof. (1) For $t \geq 0$,

$$\begin{aligned} \|Q_t f\| &= \sup_x \left| \int_E f(y) Q_t(x, dy) \right| \\ &\leq \sup_x \int_E |f(y)| \, Q_t(x, dy) \\ &\leq \|f\| \,. \end{aligned}$$

Therefore,

$$\begin{aligned} \|R_\lambda f\| &= \left\| \int_0^\infty e^{-\lambda t} Q_t f dt \right\| \\ &\leq \int_0^\infty e^{-\lambda t} \|Q_t f\| \, dt \\ &\leq \|f\| \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda} \|f\| \,. \end{aligned}$$

(2) It is obvious by (1).

(3) By definition,

$$\begin{aligned} R_\lambda R_\mu f(x) &= \int_0^\infty e^{-\lambda t} Q_t \left(\int_0^\infty e^{-\mu s} Q_s f(x) ds \right) dt \\ &= \int_0^\infty \int_0^\infty e^{-\lambda t} e^{-\mu s} Q_{t+s} f(x) ds dt \\ &= \int_0^\infty \frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} Q_r f dr \\ &= \frac{R_\mu - R_\lambda}{\lambda - \mu} f(x). \end{aligned}$$

□

Lemma 7.2.3. *Let $(X_t)_{t\geq 0}$ be a Markov process with semigroup $(Q_t)_{t\geq 0}$. Let $h \in \mathcal{B}_b(E)$ and $h \geq 0$. For $\lambda > 0$,*
$$Y_t = e^{-\lambda t} R_\lambda h(X_t)$$
is a supermartingale.

Proof. For $s > 0$,

$$\begin{aligned} Q_s(R_\lambda h) &= Q_s\left(\int_0^\infty e^{-\lambda t} Q_t h dt\right) \\ &= \int_0^\infty e^{-\lambda t} Q_{t+s} h dt \\ &= e^{\lambda s}\int_s^\infty e^{-\lambda u} Q_u h du. \end{aligned}$$

So

$$e^{-\lambda s} Q_s(R_\lambda h) = \int_s^\infty e^{-\lambda u} Q_u h du \le \int_0^\infty e^{-\lambda u} Q_u h du = R_\lambda h.$$

Then

$$\begin{aligned} \mathbb{E}\left[Y_{t+s} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[e^{-\lambda(t+s)} R_\lambda h(X_{t+s}) \mid \mathcal{F}_s\right] \\ &= e^{-\lambda(t+s)}\mathbb{E}\left[R_\lambda h(X_{t+s}) \mid \mathcal{F}_s\right] \\ &= e^{-\lambda(t+s)} Q_t R_\lambda h(X_s) = e^{-\lambda s} e^{-\lambda t} Q_t R_\lambda h(X_s) \\ &\le e^{-\lambda s} R_\lambda h(X_s) = Y_s. \end{aligned}$$

So $(Y_t)_{t\ge 0}$ is a supermartingale. □

7.3 Feller Semigroup and Generator

Let E be a metric space that is locally compact. Moreover, assume E is a union of countably many compact sets, which implies that there exists compact $K_n \uparrow E$ and any compact subset of E is contained in some K_n. That is E is a σ-compact metric space. In such case, a function $f\colon E \to \mathbb{R}$ is called trending to 0 at infinity if for any $\varepsilon > 0$, there exists a compact K such that $|f(x)| \le \varepsilon$ for all $x \notin K$, which is equivalent to

$$\lim_{n\to\infty} \sup_{x\in E\setminus K_n} |f(x)| < \varepsilon.$$

Let

$$C_0(E) := \{f\colon f \in C(E),\ f \text{ trends to 0 at infinity.}\}$$

Then $C_0(E)$ is a Banach space with the norm defined as

$$\|f\| = \sup_{x\in E} |f(x)|.$$

Definition 7.3.1. A transition semigroup $(Q_t)_{t\ge 0}$ if

(i) for any $f \in C_0(E)$, $Q_t \in C_0(E)$,

(ii) for any $f \in C_0(E)$, $\|Q_t f - f\| \to 0$ as $t \to 0$.

Remark 7.3.2. (i) It follows that for $f \in C_0(E)$,

$$R_\lambda f = \int_0^\infty e^{-\lambda t} Q_t f dt \in C_0(E).$$

(ii) Note that given $f \in C_0(E)$, $t \mapsto Q_t f$ is uniformly continuous because

$$\|Q_{t+s} f - Q_t f\| \le \|Q_s f - f\| \to 0$$

which is independent of t as $s \to 0$.

Example 7.3.3. Consider a standard Brownian motion $(B_t)_{t\geq 0}$,

$$Q_t f(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y) dy.$$

For any $f \in C_0(\mathbb{R})$, it is obvious $Q_t f \in C(\mathbb{R})$. Moreover, choose a K such that

$$Q_t f(x) = \int_K + \int_{\mathbb{R}\setminus K} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y) dy \leq \int_K + \varepsilon$$

Then as $|x| \to \infty$, by DCT, $|Q_t f(x)| \to 0$. So $Q_t f \in C_0(\mathbb{R})$.

Proposition 7.3.4. *Let $(Q_t)_{t\geq 0}$ be a Feller semigroup. For any $\lambda > 0$, let*

$$\mathcal{D} = \{R_\lambda f \colon f \in C_0(E)\}.$$

Then $\mathcal{D}$ is independent of λ and $\mathcal{D} \subset C_0(E)$ is dense.

Proof. For any $\lambda, \mu > 0$, because

$$R_\lambda f = R_\mu f + (\mu - \lambda) R_\mu R_\lambda f = R_\mu (f + f + (\mu - \lambda) R_\lambda f),$$

$\operatorname{Im} R_\lambda \subset \operatorname{Im} R_\mu$. So $\operatorname{Im} R_\lambda = \operatorname{Im} R_\mu$. For density, for any $f \in C_0(E)$,

$$\begin{aligned} R_\lambda(\lambda f) = \lambda R_\lambda f &= \lambda \int_0^\infty e^{-\lambda t} Q_t f dt \\ &= \int_0^\infty e^{-u} Q_{\frac{u}{\lambda}} f du \to f \end{aligned}$$

as $\lambda \to \infty$ by MCT. □

Definition 7.3.5 (Generator)**.** Let $(Q_t)_{t\geq 0}$ be a Feller semigroup. Set

$$\mathcal{D}(L) := \left\{ f \in C_0(E) \colon \lim_{t\to 0} \frac{Q_t f - f}{t} \text{ converges in } C_0(E) \right\}.$$

that is a linear subspace. Then for any $f \in \mathcal{D}(L)$,

$$Lf := \lim_{t\to 0} \frac{Q_t f - f}{t}.$$

L is called the generator of $(Q_t)_{t\geq 0}$, a linear operator.

Proposition 7.3.6. *Let $f \in \mathcal{D}(L)$. Then for any $s \geq 0$, $Q_s f \in D(L)$ and*

$$L(Q_s f) = Q_s(Lf).$$

Proof. Because Q_s is bounded,

$$\lim_{t\to 0} \frac{Q_t Q_s f - Q_s f}{t} = \lim_{t\to 0} Q_s \frac{Q_t f - f}{t} = Q_s L f.$$

So $Q_s f \in \mathcal{D}(L)$ and the LHS

$$L(Q_s f) = Q_s L f.$$

□

Corollary 7.3.7. *If $f \in \mathcal{D}(L)$, for any $t \geq 0$,*

$$Q_t f - f = \int_0^t Q_s(Lf) ds = \int_0^t L(Q_s f) ds.$$

Proof. Consider $t \mapsto Q_t f$,

$$\frac{d}{dt}Q_t f = \lim_{s\to 0}\frac{Q_{t+s}f - Q_t f}{s} = Q_t L f. \qquad \square$$

Proposition 7.3.8. *Let $\lambda > 0$.*

(1) For any $g \in C_0(E)$, $R_\lambda g \in \mathcal{D}(L)$ and

$$(\lambda - L)R_\lambda g = g.$$

(2) If $f \in D(L)$,

$$R_\lambda(\lambda - L)f = f.$$

It follows that $\operatorname{Im} R_\lambda = \mathcal{D}(L)$ *and* $R_\lambda = (\lambda - L)^{-1}$.

Proof. (1) Note that

$$\begin{aligned} Q_\varepsilon R_\lambda g &= Q_\varepsilon \int_0^\infty e^{-\lambda t} Q_t g dt \\ &= \int_0^\infty e^{-\lambda t} Q_{\varepsilon+t} g dt \\ &= \int_\varepsilon^\infty e^{-\lambda(u-\varepsilon)} Q_u g du. \end{aligned}$$

So

$$\frac{1}{\varepsilon}(Q_\varepsilon R_\lambda g - R_\lambda g) = \frac{\mathrm{e}^{\lambda\varepsilon} - 1}{\varepsilon} R_\lambda g - \mathrm{e}^{\lambda\varepsilon}\frac{1}{\varepsilon}\int_0^\varepsilon \mathrm{e}^{-\lambda t} Q_t g dt.$$

As $\varepsilon \to 0$,

$$\lim_{\varepsilon\to 0}(Q_\varepsilon R_\lambda g - R_\lambda g) = \lambda R_\lambda g - g = L R_\lambda g.$$

and so $R_\lambda g \in \mathcal{D}(L)$.

(2) First, for $f \in \mathcal{D}(L)$,

$$\frac{d}{dt}Q_t f = Q_t L f = L Q_t f.$$

and so

$$Q_t f - f = \int_0^t Q_s(Lf) ds.$$

Therefore,

$$\begin{aligned} R_\lambda f &= \int_0^\infty e^{-\lambda t} Q_t f dt \\ &= \int_0^\infty e^{-\lambda t}\left(f + \int_0^t Q_s(Lf) ds\right) dt \\ &= \frac{1}{\lambda} f + \int_0^\infty \frac{e^{-\lambda s}}{\lambda} Q_s L f ds \\ &= \frac{1}{\lambda} f + \frac{1}{\lambda} R_\lambda L f. \end{aligned} \qquad \square$$

Remark 7.3.9. If we have $(L, \mathcal{D}(L))$, then define for $\lambda \geq 0$

$$R_\lambda = (\lambda - L)^{-1}.$$

Such R_λ determines $(Q_t)_{t\geq 0}$ because R_λ is the Laplace transform of $(Q_t)_{t\geq 0}$.

Example 7.3.10. Consider a standard Brownian motion $(B_t)_{t\geq 0}$. Then

$$Q_t f(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y)dy.$$

So the resolvent

$$\begin{aligned}
R_\lambda f &= \int_0^\infty e^{-\lambda t} \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y)dy \right) dt \\
&= \int_{\mathbb{R}} f(y) \left(\int_0^\infty e^{-\lambda t} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} dt \right) dy \\
&= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{2\lambda}|y-x|} dy.
\end{aligned}$$

Assume $f \in C_0(E)$ and f'' exists.

$$\begin{aligned}
&\frac{d}{dt} Q_t f(x) \bigg|_{t=0} \\
&= \lim_{t\to 0} \frac{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} (f(y) - f(x))dy}{t} \\
&= \lim_{t\to 0} \frac{1}{t\sqrt{2\pi t}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{2t}} \left(f'(x)(y-x) + f''(x)\frac{(y-x)^2}{2} + f'''(\theta_{x,y}) \frac{(y-x)^3}{6} \right) dy \\
&= \lim_{t\to 0} \frac{1}{t} \left(f'(x)\mathbb{E}\left[B_t^0\right] + \frac{1}{2} f''(x)\mathbb{E}\left[\left(B_t^0\right)^2\right] + \int_{\mathbb{R}} \left[f'''(\theta_{x,y}) \frac{(y-x)^3}{6} \right] \frac{1}{\sqrt{2\pi t}} \exp\left\{ -\frac{(y-x)^2}{2t} \right\} \mathrm{d}y \right) \\
&= \frac{1}{2} f''(x) + \lim_{t\to 0} \frac{1}{t} \int_{\mathbb{R}} \left[f'''(\theta_{x,y}) \frac{(y-x)^3}{6} \right] \frac{1}{\sqrt{2\pi t}} \exp\left\{ -\frac{(y-x)^2}{2t} \right\} \mathrm{d}y \\
&= \frac{1}{2} f''(x),
\end{aligned}$$

(It is because $f \in C_0(E)$ implies that $f^{(n)} \in C_0(E)$). So

$$Lf(x) = f''(x).$$

Theorem 7.3.11. *Given a Markov process $(X_t^x)_{t\geq 0}$ with $X_0^x = x$ and it has RLCC paths. Let $h, g \in C_0(E)$. TFAE.*

(1) $h \in \mathcal{D}(L)$ and $Lh = g$.

(2) For any $x \in E$,

$$M_t = h(X_t^x) - \int_0^t g(X_s^x)ds$$

is a martingale.

Proof. (1) $\Rightarrow$ (2) : Note that

$$\begin{aligned}
\mathbb{E}\left[M_{t+s} \mid \mathcal{F}_t\right] &= \mathbb{E}\left[h(X_{t+s}^x) - \int_0^{t+s} g(X_u^x)du \mid \mathcal{F}_t \right] \\
&= \mathbb{E}\left[h(X_{t+s}^x) \mid \mathcal{F}_t\right] - \mathbb{E}\left[\int_0^t g(X_u^x)du \mid \mathcal{F}_t \right] - \mathbb{E}\left[\int_t^{t+s} g(X_u^x)du \mid \mathcal{F}_t \right] \\
&= Q_s h(X_t^x) - \int_0^t g(X_u^x)du - \int_t^{t+s} \mathbb{E}\left[g(X_u^x) \mid \mathcal{F}_t\right] du
\end{aligned}$$

$$
\begin{aligned}
&= Q_s h(X_t^x) - \int_0^t g(X_u^x)du - \int_t^{t+s} Q_{u-t} g(X_t^x)du \\
&= Q_s h(X_t^x) - \int_0^t g(X_u^x)du - \int_0^s Q_u g(X_t^x)du.
\end{aligned}
$$

By above,

$$
Q_t h = h + \int_0^t Q_s g ds.
$$

So

$$
\mathbb{E}\left[M_{t+s} \mid \mathcal{F}_t\right] = h(X_t^x) - \int_0^t g(X_u^x)du = M_t. \qquad \square
$$

(2) $\Rightarrow$ (1) : First,

$$
\mathbb{E}\left[M_t\right] = \mathbb{E}[M_0] = h(x).
$$

On the other hand,

$$
\begin{aligned}
\mathbb{E}\left[M_t\right] &= \mathbb{E}\left[h(X_t^x)\right] - \mathbb{E}\left[\int_0^t g(X_s^x)ds\right] \\
&= Q_t h(x) - \int_0^t Q_s g(x)ds
\end{aligned}
$$

Therefore,

$$
\int_0^t Q_s g(x)ds = \int_0^t Q_s Lh(x)ds \;\Rightarrow\; Q_t(g - Lh) = 0
$$

because $t \mapsto Q_t f$ is uniform continuous. Because Q_t is invertible, $g = Lh$.

7.4 Markov Property

Definition 7.4.1. For two processes $(X_t)_{t\geq 0}$ and $(X_t')_{t\geq 0}$,

(1) If for any $t \geq 0$,

$$
\mathbb{P}(X_t = X_t') = 1,
$$

then $(X_t)_{t\geq 0}$ is called a modification of $(X_t')_{t\geq 0}$.

(2) If

$$
\mathbb{P}(X_t = X_t',\ t \geq 0) = 1,
$$

then $(X_t)_{t\geq 0}$ and $(X_t')_{t\geq 0}$ are called indistinguishable.

Theorem 7.4.2. *Assume $(X_t)_{t\geq 0}$ is a Markov process with Feller semigroup $(Q_t)_{t\geq 0}$. Then $(X_t)_{t\geq 0}$ has a Markov modification $(X_t')_{t\geq 0}$ that is càdlàg.*

Sketch of Proof. Consider $\{R_\lambda f\}$ such that $Y_t = e^{-\lambda} R_\lambda f(X_t)$ is a supermartingale that has a càdlàg modification. Because such family is rich enough, $(X_t)_{t\geq 0}$ has a càdlàg modification $\square$

Assume E is a metric space. Given a semigroup $(Q_t)_{t\geq 0}$. For $x \in E$, $(X_t^x)_{t\geq 0}$ is a Markov process with $X_0^x = x$ associated with $(Q_t)_{t\geq 0}$. Assume $(X_t^x)_{t\geq 0}$ is càdlàg. Let

$$
D(E) := \{f\colon [0,\infty) \to E\colon f \text{ is càdlàg.}\}\ (= E^{[0,\infty)}).
$$

equipped with the σ-field $\mathcal{D}$ generated by the coordinate process $W_t(\omega) = \omega(t)$ for $\omega \in D(E)$.

Remark 7.4.3. If $X = (X_t)_{t\geq 0}$ is a càdlàg process on $(\Omega, \mathcal{F}, \mathbb{P})$, then $X\colon \Omega \to D(E)$ i.e. X can be viewed as $D(E)$-value random variable. Furthermore, let $\mathbb{P}_X = X_{\#}\mathbb{P}$ be the law of X on $D(E)$.

Definition 7.4.4 (Shift Operator)**.** Fix $t \geq 0$,

$$\theta_t\colon D(E) \to D(E)$$

is defined as for any $\omega \in D(E)$,

$$\theta_t(\omega)(s) := \omega(t+s).$$

Theorem 7.4.5 (Markov Property)**.** *Let $X = (X_t)_{t\geq 0}$ be a càdlàg Markov process associated with semigroup $(Q_t)_{t\geq 0}$. Let $s \geq 0$ and $\Phi\colon D(E) \to \mathbb{R}$ be a measurable and bounded function. Then*

$$\mathbb{E}\left[\Phi(\theta_s \circ X) \mid \mathcal{F}_s\right] = \mathbb{E}_{X_s}[\Phi].$$

Remark 7.4.6. Note that because $\mathbb{E}_{X_s}[\Phi]$ is $\sigma(X_s)$-measurable,

$$\mathbb{E}\left[\Phi(\theta_s \circ X) \mid \mathcal{F}_s\right] = \mathbb{E}\left[\Phi(\theta_s \circ X) \mid \sigma(X_s)\right].$$

Proof. By Monotone Class Theorem, assume

$$\Phi(f) = \varphi_1(f(t_1)) \cdots \varphi_p(f(t_p)).$$

So the RHS is

$$\begin{aligned}
\mathbb{E}_{X_s}[\Phi] &= \mathbb{E}_{X_s}\left[\varphi_1(W_{t_1}) \cdots \varphi_p(W_{t_p})\right] \\
&= \int_E \delta_{X_s}(dx_0) \int_E \varphi_1(x_1) Q_{t_1}(x_0, dx_1) \int_E \varphi_2(x_2) Q_{t_2-t_1}(x_1, dx_2) \cdots \int_E \varphi_p(x_p) Q_{t_p - t_{p-1}}(x_{p-1}, dx_p) \\
&= \int_E \varphi_1(x_1) Q_{t_1}(X_s, dx_1) \int_E \varphi_2(x_2) Q_{t_2-t_1}(x_1, dx_2) \cdots \int_E \varphi_p(x_p) Q_{t_p - t_{p-1}}(x_{p-1}, dx_p)
\end{aligned}$$

the LHS is

$$\mathbb{E}\left[\Phi(\theta_s \circ X) \mid \mathcal{F}_s\right] = \mathbb{E}\left[\varphi_1(X_{t_1+s}) \cdots \varphi_p(X_{t_p+s}) \mid \mathcal{F}_s\right].$$

For $p = 1$,

$$\mathbb{E}\left[\varphi_1(X_{t_1+s}) \mid \mathcal{F}_s\right] = Q_{t_1}\varphi_1(X_s) = \int_E \varphi_1(x_1) Q_{t_1}(X_s, dx_1).$$

So it is true. Assume it is true for $p - 1$.

$$\begin{aligned}
&\mathbb{E}\left[\varphi_1(X_{t_1+s}) \cdots \varphi_p(X_{t_p+s}) \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\varphi_1(X_{t_1+s}) \cdots \varphi_p(X_{t_p+s}) \mid \mathcal{F}_{t_{p-1}+s}\right] \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\varphi_1(X_{t_1+s}) \cdots \varphi_{p-1}(X_{t_{p-1}+s}) \mathbb{E}\left[\varphi_p(X_{t_p+s}) \mid \mathcal{F}_{t_{p-1}+s}\right] \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\varphi_1(X_{t_1+s}) \cdots \varphi_{p-1}(X_{t_{p-1}+s}) Q_{t_p - t_{p-1}}\varphi_p(X_{t_{p-1}+s}) \mid \mathcal{F}_s\right] \\
&= \int_E \varphi_1(x_1) Q_{t_1}(X_s, dx_1) \cdots \int_E \varphi_{p-1}(x_{p-1}) Q_{t_p - t_{p-1}}\varphi_p(x_{p-1}) Q_{t_{p-1}-t_{p-2}}(x_{p-2}, dx_{p-1}) \\
&= \int_E \varphi_1(x_1) Q_{t_1}(X_s, dx_1) \cdots \int_E \varphi_{p-1}(x_{p-1}) Q_{t_{p-1}-t_{p-2}}(x_{p-2}, dx_{p-1}) \int_E \varphi_p(x_p) Q_{t_p - t_{p-1}}(x_{p-1}, dx_p). \square
\end{aligned}$$

Theorem 7.4.7 (Strong Markov Property)**.** *Let $(Q_t)_{t\geq 0}$ be a Feller semigroup and $(X_t)_{t\geq 0}$ be the corresponding Markov process with RLCC paths. Let T be a stopping time and $\Phi\colon D(E) \to \mathbb{R}$ be a measurable and bounded function.*

$$\mathbb{E}\left[\mathbb{I}_{\{T<\infty\}}\Phi(\theta_T \circ X) \mid \mathcal{F}_T\right] = \mathbb{I}_{\{T<\infty\}}\mathbb{E}_{Y_T}[\Phi].$$

Proof. By Monotone Class Theorem, assume

$$\Phi(f) = \varphi_1(f(t_1)) \cdots \varphi_p(f(t_p)).$$

The integrability and measurability are obvious. It suffices to show that for any $A \in \mathcal{F}_T$,

$$\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\Phi(\theta_T \circ X)\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\mathbb{E}_{X_T}[\Phi]\right].$$

and it is sufficient to consider $p = 1$, i.e.

$$\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\varphi_1(X_{t_1+T})\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\mathbb{E}_{X_T}[\Phi]\right].$$

Note that

$$\mathbb{E}_{X_T}[\Phi] = Q_{t_1}\varphi_1(X_T).$$

So the RHS is

$$\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\mathbb{E}_{X_T}[\Phi]\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}Q_{t_1}\varphi_1(X_T)\right],$$

and our goal is to show

$$\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\varphi_1(X_{t_1+T})\right] = \mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}Q_{t_1}\varphi_1(X_T)\right].$$

Let

$$T_n = \sum_{i=0}^{\infty} \frac{i+1}{2^n}\mathbb{I}_{\left\{\frac{i}{2^n}<T\le\frac{i+1}{2^n}\right\}} + \infty\mathbb{I}_{\{T=\infty\}}.$$

Then $T_n \downarrow T$ stopping time. By Monotone Class Theorem, we further assume φ_1 is continuous. So by the continuity of X_t and Feller property of Q_t,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\varphi_1(X_{t_1+T})\right] &= \lim_{n\to\infty}\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}\varphi_1(X_{t_1+T_n})\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{\infty}\mathbb{E}\left[\mathbb{I}_{A\cap\left\{\frac{i}{2^n}<T\le\frac{i+1}{2^n}\right\}}\varphi_1(X_{t_1+\frac{i}{2^n}})\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{\infty}\mathbb{E}\left[\mathbb{I}_{A\cap\left\{\frac{i}{2^n}<T\le\frac{i+1}{2^n}\right\}}\mathbb{E}\left[\varphi_1(X_{t_1+\frac{i+1}{2^n}}) \mid \mathcal{F}_{\frac{i+1}{2^n}}\right]\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{\infty}\mathbb{E}\left[\mathbb{I}_{A\cap\left\{\frac{i}{2^n}<T\le\frac{i+1}{2^n}\right\}}Q_{t_1}\varphi_1(X_{\frac{i+1}{2^n}})\right] \\
&= \lim_{n\to\infty}\mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}Q_{t_1}(X_{T_n})\right] \\
&= \mathbb{E}\left[\mathbb{I}_{A\cap\{T<\infty\}}Q_{t_1}\varphi_1(X_T)\right].
\end{aligned}$$

For $p > 1$, it can be done by the Markov property and induction. □

7.5 Jump Process and Lévy Process

Jump Markov Process. Assume the state space E is finite equipping with the discrete metric $d(x, y) = \delta_x(y)$ and σ-field $\mathcal{P}(E)$. Let $f \in D(E)$, i.e. $f\colon [0,\infty) \to E$ is càdlàg. Note that for $\{y_n\} \in E$,

$$y_n \to y \quad \Leftrightarrow \quad \exists\, m,\ y_n = y,\ \forall\, n \ge m.$$

Therefore, there exists $t \in (0,\infty]$ such that $f(s) = f(0)$ for all $s \in (0,t)$. Let

$$t_1 = \max\{t > 0\colon f(s) = f(0), \quad \forall\, s \in (0,t)\}.$$

If $t_1 < \infty$, there exists $t_2 > t_1$ such that

$$t_2 = \max\{t > t_1 \colon f(s) = f(t_1),\ \forall\, s \in (t_1, t)\}.$$

Therefore, there exist $0 < t_1 < t_2 < \cdots$ such that

$$f(t) = f(t_n), \quad \forall\, t \in [t_n, t_{n+1}).$$

Let $(Q_t)_{t\geq 0}$ be a semigroup on E. Because $C(E) = B(E)$, $(Q_t)_{t\geq 0}$ is a Feller semigroup. So we can construct a measure space $(\Omega, \mathcal{F})$ on which there is a family $(\mathbb{P}_x \colon x \in E)$ and a process $(X_t)_{t\geq 0}$ with càdlàg paths such that $(X_t)_{t\geq 0}$ is a Markov process associated with $(Q_t)_{t\geq 0}$ when $X_0 = x$. For every $\omega \in \Omega$, there exists a sequence

$$0 = T_0(\omega) < T_1(\omega) < \cdots < T_n(\omega) < \cdots,$$

such that

$$X_t(\omega) = X_{T_n}(\omega), \quad \forall\, t \in [T_n, T_{n+1}).$$

Moreover, T_n is a stopping time, like

$$\{T_1 < t\} = \bigcup_{q\in[0,t)\cap\mathbb{Q}} \{X_q \neq X_0\} \in \mathcal{F}_t^X.$$

Note that for a t, on the set of $\{\omega : t < T_1(\omega)\}$, then $T_1(\omega) = t + T_1 \circ \theta_t$. So we have

$$T_2 = T_1 + T_1 \circ \theta_{T_1}.$$

Lemma 7.5.1. *Let $x \in E$. There exists a $q(x) \geq 0$ such that T_1 is exponentially distributed with parameter $q(x)$ under $\mathbb{P}_x$. Furthermore, if $q(x) > 0$, then T_1 and X_{T_1} are independent.*

Proof. First,

$$\begin{aligned}
\mathbb{P}_x(T_1 > s+t) &= \mathbb{P}_x(T_1 > s+t,\ T_1 > s) \\
&= \mathbb{P}_x(s + T_1\circ\theta_s > s+t,\ T_1 > s) \\
&= \mathbb{P}_x(T_1\circ\theta_s > t,\ T_1 > s) \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1\circ\theta_s > t\}}\mathbb{I}_{\{T_1>s\}}\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_x\left[\mathbb{I}_{\{T_1\circ\theta_s > t\}}\mathbb{I}_{\{T_1>s\}} \mid \mathcal{F}_s\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1>s\}}\mathbb{E}_x\left[\mathbb{I}_{\{T_1\circ\theta_s > t\}} \mid \mathcal{F}_s\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1>s\}}\mathbb{E}_{X_s}\left[\mathbb{I}_{T_1>t}\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1>s\}}\mathbb{E}_x\left[\mathbb{I}_{T_1>t}\right]\right] \\
&= \mathbb{P}_x(T_1 > s)\mathbb{P}_x(T_1 > t),
\end{aligned}$$

which implies that there exists a $q(x) \geq 0$ such that

$$\mathbb{P}_x(T_1 > t) = e^{-q(x)t}.$$

When $q(x) > 0$, $T_1 < \infty$. Let $y \in E$. Consider

$$\begin{aligned}
\mathbb{P}_x(T_1 > t,\ X_{T_1} = y) &= \mathbb{P}_x(T_1 > t,\ X_{t+T_1\circ\theta_t} = y) \\
&= \mathbb{P}_x(T_1 > t,\ X_{T_1}\circ\theta_t = y) \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1>t\}}\mathbb{I}_{\left\{X_{T_1}\circ\theta_t = y\right\}}\right] \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1>t\}}\mathbb{E}_{X_t}\left[\mathbb{T}_{X_{T_1}=Y}\right]\right] \\
&= \mathbb{P}_x(T_1 > t)\mathbb{P}_x\left(X_{T_1} = y\right).
\end{aligned}$$

□

Note that if $q(x)=0$, $X_t \equiv x$. If $q(x)>0$, for $x, y \in E$, define

$$\pi(x,y) = \mathbb{P}_x(X_{T_1} = y).$$

So $(\pi(x,y))_{x,y\in E}$ is a transition matrix.

Proposition 7.5.2. *Let L be the generator of $(Q_t)_{t\geq 0}$. Then $\mathcal{D}(L) = C(E) = B(E)$. And for any $\varphi \in C(E)$, $x \in E$, if $q(x)=0$, then $L\varphi(x) = 0$, and if $q(x) > 0$,*

$$L\varphi(x) = q(x) \sum_{y\in E, y\neq x} \pi(x,y)\left(\varphi(y) - \varphi(x)\right).$$

Proof. Note that

$$L\varphi(x) = \lim_{t\to 0} \frac{Q_t\varphi(x) - \varphi(x)}{t}.$$

If $q(x)=0$, $X_t \equiv x$ and $T_1 = \infty$. So

$$Q_t\varphi(x) = \mathbb{E}_x[\varphi(X_t)] = \mathbb{E}_x[\varphi(x)] = \varphi(x).$$

So $L\varphi(x) = 0$.
Assume $q(x)>0$. Then $T_1 < \infty$.
Claim: We claim

$$\mathbb{P}_x(T_2 \leq t) = O(t^2), \quad t \to 0$$

In fact,

$$\begin{aligned}
\mathbb{P}_x(T_2 \leq t) &= \mathbb{P}_x(T_1 \leq t,\ T_1 + T_1\circ\theta_{T_1} \leq t) \\
&\leq \mathbb{P}_x(T_1 \leq t,\ T_1 + T_1\circ\theta_{T_1} \leq t + T_1) \\
&= \mathbb{P}_x(T_1 \leq t,\ T_1\circ\theta_{T_1} \leq t) \\
&= \mathbb{E}_x\left[\mathbb{I}_{\{T_1\leq t\}}\mathbb{E}_{X_{T_1}}\left[\mathbb{I}_{\{T_1\leq t\}}\right]\right] \\
&\leq \mathbb{E}_x\left[\mathbb{I}_{\{T_1\leq t\}} \sup_{y\in E}\mathbb{P}_y(T_1\leq t)\right] \\
&= \sup_{y\in E}\mathbb{P}_y(T_1\leq t)\mathbb{P}_x(T_1\leq t) \\
&= \sup_{y\in E}\left(1-e^{-q(y)t}\right)\left(1-e^{-q(x)t}\right) \leq Ct^2,
\end{aligned}$$

when $t\to 0$.
Then we have

$$\begin{aligned}
Q_t\varphi(x) &= \mathbb{E}_x\left[\varphi(X_t)\right] \\
&= \mathbb{E}_x\left[\varphi(X_t)\mathbb{I}_{\{t<T_1\}}\right] + \mathbb{E}_x\left[\varphi(X_t)\mathbb{I}_{\{t\geq T_1\}}\right] \\
&= \varphi(x)\mathbb{P}_x(T_1>t) + \mathbb{E}_x\left[\varphi(X_t)\mathbb{I}_{\{t\geq T_1\}}\mathbb{I}_{\{t<T_2\}}\right] + \mathbb{E}_x\left[\varphi(X_t)\mathbb{I}_{\{t\geq T_1\}}\mathbb{I}_{\{t\geq T_2\}}\right] \\
&= \varphi(x)\mathbb{P}_x(T_1>t) + E_x\left[\varphi(X_{T_1})\mathbb{I}_{\{T_1\leq t<T_2\}}\right] + O(t^2) \\
&= \varphi(x)\mathbb{P}_x(T_1>t) + E_x\left[\varphi(X_{T_1})\mathbb{I}_{\{T_1\leq t\}}\right] - E_x\left[\varphi(X_{T_1})\mathbb{I}_{\{T_2\leq t\}}\right] + O(t^2) \\
&= \varphi(x)\mathbb{P}_x(T_1>t) + E_x\left[\varphi(X_{T_1})\mathbb{I}_{\{T_1\leq t\}}\right] + O(t^2) \\
&= \varphi(x)e^{-q(x)t} + \mathbb{E}_x[\varphi(X_{T_1})]\mathbb{P}_x(T_1\leq t) + O(t^2) \\
&= \varphi(x)e^{-q(x)t} + (1-e^{-q(x)t})\sum_{y\in E, y\neq x}\pi(x,y)\varphi(y) + O(t^2).
\end{aligned}$$

Therefore,

$$L\varphi(x) = q(x)\sum_{y\in E, y\neq x}\pi(x,y)\left(\varphi(y)-\varphi(x)\right).$$

□

Theorem 7.5.3. *If $q(y) > 0$ for all y, then $T_1 < T_2 < \cdots < \infty$ a.e.. Moreover, $(X_{T_n})_{n\geq 0}$ is a Markov chain with the transition matrix $\pi(x, y)$.*

Proof. First,

$$\begin{aligned}
\mathbb{P}_x\left(X_{T_1} = z_1, X_{T_2} = z_2\right) &= \mathbb{P}_x\left(X_{T_1} = z_1, X_{T_1 + T_1 \circ \theta_{T_1}} = z_2\right) \\
&= \mathbb{P}_x\left(X_{T_1} = z_1, X_{T_1} \circ \theta_{T_1} = z_2\right) \\
&= \mathbb{P}_x\left(X_{T_1} = z_1, E_{X_{T_1}}\left[X_{T_1} = z_2\right]\right) \\
&= \mathbb{P}_x\left(X_{T_1} = z_1, E_{z_1}\left[X_{T_1} = z_2\right]\right) \\
&= \mathbb{P}_x\left(X_{T_1} = z_1\right) P_{z_1}\left(X_{T_1} = z_2\right) \\
&= \pi\left(x, z_1\right) \pi\left(z_1, z_2\right)
\end{aligned}$$

Then by induction, we have

$$P_x\left(X_{T_1} = z_1, X_{T_2} = z_2, \cdots, X_{T_n} = z_n\right) = \pi\left(x, z_1\right) \pi\left(z_1, z_2\right) \cdots \pi\left(z_{n-1}, z_n\right). \qquad \square$$

Lévy Process. Let $Y = (Y_t)_{t\geq 0}$ be a stochastic process such that

(i) $Y_0 = 0$ a.e.

(ii) for any $s \leq t$, $Y_t - Y_s$ is independent of $\sigma(Y_r : r \leq s)$,

(iii) $Y_t \to 0$ in probability as $t \to 0$.

Then Y is called a Léy process.

Theorem 7.5.4. *For $t \geq 0$, let $Q_t(x, dy)$ be the law of $Y_t + x$, i.e.,*

$$Q_t f(x) = \mathbb{E}\left[f(Y_t + x)\right].$$

$(Q_t)_{t\geq 0}$ is a Feller semigroup and Y is a Markov process associated with $(Q_t)_{t\geq 0}$.

Chapter 8

Stochastic Integral

8.1 Local Martingale

Definition 8.1.1 (Local Martingale)**.** An $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ adapted stochastic process $X = (X_t)_{t\geq 0}$ is called a local martingale if there exists a sequence stopping times $\{T_n\}_{n\in\mathbb{N}}$ with $T_n \uparrow \infty$ such that $(X_{t\wedge T_n}\mathbb{I}_{\{T_n>0\}})_{t\geq 0}$ is a UI martingale w.s.t. $\mathbb{F}$.

Remark 8.1.2. If $X_0 = 0$, we only need to require $(X_{t\wedge T_n})_{t\geq 0}$ is a UI martingale w.s.t. $\mathbb{F}$. Let $T_n = n$, clearly martingales are local martingales but the converse is not true.

Notation: Let $\mathcal{M}^{loc}$ be the set of all local martingales and $\mathcal{M}_0^{loc} \subset \mathcal{M}^{loc}$ be the set of all local martingales with $X_0 = 0$.

Lemma 8.1.3. *A local martingale with*

$$\sup_{s\leq t} |X_s| \in L^1$$

for any t is a martingale.

Proof. Assume $X_0 = 0$. Let $T_n \uparrow \infty$ be stopping times such that $(X_{t\wedge T_n})_{t\geq 0}$ be UI martingales. Then for $s \leq t$,

$$\mathbb{E}\left[X_{t\wedge T_n} \mid \mathcal{F}_s\right] = X_{s\wedge T_n}$$

By Martingale Convergence Theorem, as $n \to \infty$,

$$X_{t\wedge T_n} \to X_t, \quad X_{s\wedge T_n} \to X_s$$

a.e.. Because

$$\sup_n |X_{t\wedge T_n}| \leq \sup_{0\leq r\leq t} |X_r| \in L^1,$$

by DCT,

$$\mathbb{E}\left[X_t \mid \mathcal{F}_s\right] = X_s.$$

□

Remark 8.1.4. In fact, from the proof, it is not difficult to see a local martingale $X = (X_t)_{t\geq 0}$ is a martingale if

$$\{X_T : T \in \mathcal{S}_t\},$$

is UI for all t, where $\mathcal{S}_t := \{T \text{ stopping time} : T \leq t\}$. Such condition is called class (DL).

Proposition 8.1.5. *Every nonnegative local martingale is a supermartingale.*

Proof. Assume $X_0 = 0$. Let $T_n \uparrow \infty$ be stopping times such that $(X_{t\wedge T_n})_{t\geq 0}$ be UI martingales. Then for $s \leq t$,

$$\begin{aligned} X_s &= \lim_{n\to\infty} X_{s\wedge T_n} = \lim_{n\to\infty} \mathbb{E}\left[X_{t\wedge T_n} \mid \mathcal{F}_s\right] \\ &\geq \mathbb{E}\left[\lim_{n\to\infty} X_{t\wedge T_n} \mid \mathcal{F}_s\right] = \mathbb{E}\left[X_t \mid \mathcal{F}_s\right] \end{aligned}$$

by Fatou's lemma. □

Definition 8.1.6. Let $\bar{\Omega} = [0,\infty) \times \Omega$.

(1) A σ-field $\mathcal{P}$ over $\bar{\Omega}$ is called predictable if it is generated by all left-continuous and adapted process $X\colon [0,\infty)\times\Omega \to \mathbb{R}$.

(2) A stochastic process $X = (X_t)_{t\geq 0}$ is called predictable if X is $\mathcal{P}$-measurable on $\bar{\Omega}$.

Remark 8.1.7. (i) Every predictable process is adapted.

(ii) Every continuous and increasing process is predictable.

(iii) If filtration $\mathbb{F}$ satisfies the usual condition, every predictable process is adapted to $\mathbb{F}_- = (\mathcal{F}_{t-})_{t\geq 0}$.

Notation: Let $\mathcal{M}^2$ be the set of all càdlàg martingales $X = (X_t)_{t\geq 0}$ such that

$$\sup_{t\geq 0} \mathbb{E}\left[X_t^2\right] < \infty.$$

Let $\mathcal{M}_0^2 \subset \mathcal{M}^2$ be all $X \in \mathcal{M}^2$ with $X_0 = 0$. Let $\mathcal{M}_0^{2,c} \subset \mathcal{M}_0^2$ be all $X \in \mathcal{M}^2$ that is continuous.

Remark 8.1.8. Note that if $M \in \mathcal{M}^2$, then M is UI. So by convergence theorem, $M_\infty = \lim_t M_t$ in L^1 and $M_t = \mathbb{E}\left[M_\infty \mid \mathcal{F}_t\right]$.

Theorem 8.1.9 (Doob-Meyer Decomposition)**.** *Let $X = (X)_{t\geq 0}$ be a right-continuous supermartingale. Assume*

$$\{X_T \colon T \in \mathcal{S}\},$$

is UI, where $\mathcal{S} := \{T \text{ stopping time} : T < \infty\}$ (X is called of class (D)). Then X admits a unique decomposition

$$X_t = X_0 + M_t - A_t,$$

where M is a right-continuous UI martingale with $M_0 = 0$ and A is a increasing and right-continuous and predictable process with $A_0 = 0$.

Remark 8.1.10. Note that any càdlàg martingale is of class (DL), but a càdlàg martingale is of class (D) if and only if it is UI. These results are also true for a càdlàg positive submartingale.

Corollary 8.1.11. *Let $M \in \mathcal{M}^2$. Then there exists a unique right-continuous predictable process $\langle M\rangle = (\langle M\rangle_t)_{t\geq 0}$ with $\langle M\rangle = 0$ such that $M^2 - \langle M\rangle$ is a martingale.*

Proof. M is a martingale so $-M^2$ is a supermartingale. To show of class (D), it suffices to show the UI of $(M_t^2)_{t\geq 0}$ because it is a positive submartingale. Because $M \in M^2$, M is UI so $M_\infty = \lim_t M_t$ in L^1. So $M_\infty^2 = \lim_t M_t^2$ in L^1, which implies that M^2 is UI. Then by Doob-Meyer decomposition.

$$-M^2 = \text{martingale } - \langle M\rangle.$$

□

Remark 8.1.12. Because $\langle M\rangle$ is uniquely determined by M, it is called the quadratic variation of M. And $\sup_t \mathbb{E}[\langle M\rangle_t] = \mathbb{E}[M_0^2] < \infty$.

Lemma 8.1.13. *Let $M \in M^{2,c}$. For partition Π of $[0,t]$, we have*

$$\lim_{\|\Pi\|\to 0} \sum_{t_k\in\Pi} \left|M_{t_k} - M_{t_{k-1}}\right|^2 = \langle M\rangle_t$$

in probability.

Remark 8.1.14. In fact, it is the general definition of quadratic variation.

Definition 8.1.15. For $M, N \in \mathcal{M}^2$, the process

$$\langle M, N\rangle = \frac{1}{4}\left(\langle M+N\rangle - \langle M-N\rangle\right).$$

is called the cross variation (or quadratic covariation) of M, N.

Remark 8.1.16. (i) Note that $\langle M, M\rangle_t = \langle M\rangle_t$.

(ii) By definition, it is not hard to see $MN - \langle M, N\rangle$ is a martingale.

(iii) If $M, N \in M^{2,loc}$ right-continuous, then there exists a unique increasing right-continuous predictable process $\langle M\rangle$ and there exists a unique increasing right-continuous predictable process $\langle M, N\rangle$ of bounded variation such that

$$M^2 - \langle M\rangle, \quad MN - \langle M, N\rangle,$$

are local martingales.

(iv) Moreover,

$$\lim_{\|\Pi\|\to 0} \sum_{k=1}^{m} \left(M_{t_k} - M_{t_{k-1}}\right)\left(N_{t_k} - N_{t_{k-1}}\right) = \langle M, N\rangle_t.$$

Example 8.1.17 (Variation of Brownian Motion)**.** Given a Brownian motion $B = (B_t)_{t\geq 0}$, there are two ways to calculate it quadratic variation.

(1) By directly calculating the quadratic total variation,

$$\langle B\rangle_t = \lim_{n\to\infty} \sum_{i=1}^{p_n} \left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 = t.$$

(2) We already know $(B_t^2 - t)_{t\geq 0}$ is a martingale. Then, by Doob-Meyer decomposition, we directly has

$$\langle B\rangle_t = t.$$

So we usually denote

$$dB_t dB_t = t, \ dB_t dt = dt dt = 0.$$

Example 8.1.18. If $\boldsymbol{B} = (B^1, \cdots, B^d)$ is a d-dim Brownian motion, then

$$\left\langle B^i, B^j\right\rangle_t = \delta_{ij} t.$$

Given a partition Π of $[0,t]$ and let

$$S_\Pi = \sum_{k=1}^{n} \left(B_{t_k}^i - B_{t_{k-1}}^i\right)\left(B_{t_k}^j - B_{t_{k-1}}^j\right), \quad i \neq j.$$

Then $S_\Pi \to \langle B^i, B^j\rangle_t$ by definition. But by independence

$$\mathbb{E}\left[S_\Pi^2\right] = \sum_{k=1}^{n} \mathbb{E}\left[\left(\Delta_k B^i\right)^2\right]\mathbb{E}\left[\left(\Delta_k B^j\right)^2\right] = \sum_{k=1}^{n} \left(\Delta t_k\right)^2 \leq t\,\|\Pi\| \to 0.$$

Definition 8.1.19 (Semimartingale)**.** A process $X = (X_t)_{t\geq 0}$ is called a semimartingale if

$$X_t = X_0 + M_t + A_t,$$

where $M \in \mathcal{M}_0^{loc}$ and A is adapted and càdlàg of bounded variation, i.e. there exists increasing adapted process A^+, A^- such that

$$A = A^+ - A^-.$$

Remark 8.1.20. This decomposition may be not unique. But if X is continuous, it is unique.

Lemma 8.1.21. *A continuous local martingale with bounded variation is constant a.e..*

Remark 8.1.22. A continuous non-constant local martingale is of unbounded variation. Therefore, we cannot use the usual Riemannian-Stieltjes to define the stochastic integral w.s.t. martingale.

8.2 Integral w.s.t. Martingale

For $M \in \mathcal{M}^2$, the goal is to define $\int_0^T H_t dM_t$. By employing the idea from Riemannian-Stieltjes

$$\int_0^T H_t dM_t := \lim_n \sum_{i=0}^n H_{\alpha_i}(M_{t_{i+1}} - M_{t_i}).$$

But because M is not of bounded variation, it is not well-defined. So there are three choices:

(i) $\alpha_i = t_i$: Itô integral.

(ii) $\alpha_i = (t_{i+1} - t_i)/2$: Stratonovich integral.

(iii) $\alpha_i = t_{i+1}$: backward Itô integral.

In the following, we mainly consider the Itô integral.
Notation: Let $\mathcal{E}^b$ be the set of all bounded predictable simple process, i.e., if $H_t \in \mathcal{E}^b$, then

$$H_t(\omega) = \sum_{i=0}^{n-1} h^i(\omega)\mathbb{I}_{(t_i,t_{i+1}]}(t),$$

where h^i is bounded and $\mathcal{F}_{t_i}$-measurable.

Integral of simple process.

Definition 8.2.1. For $H \in \mathcal{E}^b$ with $H_t(\omega) = \sum_{i=0}^{n-1} h^i(\omega)\mathbb{I}_{(t_i,t_{i+1}]}(t)$, the stochastic integral with respect to M is

$$(H \cdot M)_t = \int_0^t H_s dM_s = \sum_{i=0}^{n-1} h^i(M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).$$

Lemma 8.2.2. *Let $H^1, H^2 \in \mathcal{E}^b$ and $c_1, c_2 \in \mathbb{R}$. Then $c_1 H_1 + c_2 H_2 \in \mathcal{E}^b$ and*

$$(c_1 H^1 + c_2 H^2) \cdot M = c_1(H^1 \cdot M) + c_2(H^2 \cdot M).$$

Proposition 8.2.3. *For $H \in \mathcal{E}^b$ and $M \in \mathcal{M}^2$, $H \cdot M \in \mathcal{M}_0^{2,c}$. Moreover,*

$$\mathbb{E}\left[(H\cdot M)_\infty^2\right] = \mathbb{E}\left[\left(\int_0^\infty H_u dM_u\right)^2\right] = \mathbb{E}\left[\int_0^\infty H_u^2 d\langle M\rangle_u\right].$$

Proof. Let $s \le t$. If $s = t_k$ and $t = t_\ell$ with $k < \ell$, then

$$\begin{aligned}
\mathbb{E}[(H\cdot M)_t - (H\cdot M)_s \mid \mathcal{F}_s] &= \sum_{i=k}^{\ell-1} \mathbb{E}\left[h^i(M_{t_{i+1}} - M_{t_i}) \mid \mathcal{F}_{t_k}\right] \\
&= \sum_{i=k}^{\ell-1} \mathbb{E}\left[\mathbb{E}\left[h^i(M_{t_{i+1}} - M_{t_i}) \mid \mathcal{F}_{t_i}\right] \mid \mathcal{F}_{t_k}\right] \\
&= \sum_{i=k}^{\ell-1} \mathbb{E}\left[h^i\mathbb{E}\left[(M_{t_{i+1}} - M_{t_i}) \mid \mathcal{F}_{t_i}\right] \mid \mathcal{F}_{t_k}\right] \\
&= 0
\end{aligned}$$

It is similar for any $s \le t$. So $H \cdot M \in \mathcal{M}_0^2$. Next,

$$\begin{aligned}
\mathbb{E}\left[(H\cdot M)_\infty^2\right] &= \mathbb{E}\left[\left(\sum_{i=0}^{n-1} h^i(M_{t_{i+1}} - M_{t_i})\right)^2\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[(h^i)^2(M_{t_{i+1}} - M_{t_i})^2\right] + 2\sum_{i<j} \mathbb{E}\left[h_i h_j (M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\left[h_i h_j (M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})\right] &= \mathbb{E}\left[\mathbb{E}\left[h_i h_j (M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j}) \mid \mathcal{F}_{t_j}\right]\right] \\
&= \mathbb{E}\left[h_i h_j (M_{t_{i+1}} - M_{t_i})\mathbb{E}\left[(M_{t_{j+1}} - M_{t_j}) \mid \mathcal{F}_{t_j}\right]\right] \\
&= 0.
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}\left[(H\cdot M)_\infty^2\right] &= \sum_{i=0}^{n-1} \mathbb{E}\left[(h^i)^2(M_{t_{i+1}} - M_{t_i})^2\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[(h^i)^2\mathbb{E}\left[(M_{t_{i+1}} - M_{t_i})^2 \mid \mathcal{F}_{t_i}\right]\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\left[(M_{t_{i+1}} - M_{t_i})^2 \mid \mathcal{F}_{t_i}\right] &= \mathbb{E}\left[M_{t_{i+1}}^2 \mid \mathcal{F}_{t_i}\right] + M_{t_i}^2 - 2M_{t_i}\mathbb{E}\left[M_{t_i} \mid \mathcal{F}_{t_i}\right] \\
&= \mathbb{E}\left[M_{t_{i+1}}^2 - M_{t_i}^2 \mid \mathcal{F}_{t_i}\right].
\end{aligned}$$

Because $M \in \mathcal{M}^2$, $M^2 - \langle M\rangle$ is a martingale, which implies that

$$\mathbb{E}\left[M_{t_{i+1}}^2 - M_{t_i}^2 \mid \mathcal{F}_{t_i}\right] = \mathbb{E}\left[\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i} \mid \mathcal{F}_{t_i}\right].$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[(H\cdot M)_\infty^2\right] &= \sum_{i=0}^{n-1} \mathbb{E}\left[(h^i)^2\mathbb{E}\left[\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i} \mid \mathcal{F}_{t_i}\right]\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[(h^i)^2\left(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}\right)\right]
\end{aligned}$$

$$= \mathbb{E}\left[\int_0^\infty H_u^2 d\left\langle M\right\rangle_u\right].$$

Also because H is bounded and $\mathbb{E}[\langle M\rangle] < \infty$,

$$\mathbb{E}\left[(H\cdot M)_t^2\right] \leq \mathbb{E}\left[(H\cdot M)_\infty^2\right] < \infty. \qquad \square$$

We omit the proof of continuity.

Corollary 8.2.4. *For $H \in \mathcal{E}^b$ and $B = (B_t)_{t\geq 0}$ a Brownian motion,*

$$\mathbb{E}\left[\int_a^b H_u dB_u\right] := \mathbb{E}\left[(H\cdot M)_b\right] - \mathbb{E}\left[(H\cdot M)_a\right] = 0,$$

and

$$\mathbb{E}\left[\left(\int_a^b H_u dB_u\right)^2\right] = \mathbb{E}\left[\int_a^b H_u^2 du\right].$$

Integral of L^2 integrable process.

Theorem 8.2.5. *If $M \in \mathcal{M}_0^{2,c}$ and H is a progressively measurable process such that*

$$\mathbb{E}\left[\int_0^T H_s^2 d\left\langle M\right\rangle_s\right] < \infty, \tag{8.1}$$

for all $T \geq 0$, then there exists a sequence of predictable simple processes $H^{(n)}$ such that

$$\sup_{T>0}\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left|H_s^{(n)} - H_s\right|^2 d\left\langle M\right\rangle_s\right] = 0.$$

Remark 8.2.6. Note that because $\langle M\rangle$ is increasing, it can use Riemannian-Stieltjes to define

$$\int_0^T H_s^2 d\left\langle M\right\rangle_s.$$

Remark 8.2.7. For $M \in \mathcal{M}_0^{2,c}$ and any H satisfies above condition and a corresponding sequence of predictable simple processes $H^{(n)}$, because

$$\mathbb{E}\left[\left(\int_0^T H_s^{(n)} dM_s - \int_0^T H_s^{(m)} dM_s\right)^2\right] = \mathbb{E}\left[\int_0^T \left|H_s^{(n)} - H_s^{(m)}\right|^2 d\langle M\rangle_s\right] \longrightarrow 0,$$

due to $\int_0^T H_s^{(n)} d\left\langle M\right\rangle_s$ converges in L^2. So $\int_0^T H_s^{(n)} dM_s$ is Cauchy in L^2 and it is convergent in L^2.

Notation: For any $0 \leq T < \infty$, let $\mathcal{L}_T^*(M)$ be the set of all bounded progressively measurable process satisfying condition (8.1) and $\mathcal{L}^*(M) = \bigcap_{T\geq 0} \mathcal{L}_T^*$

Definition 8.2.8 (Stochastic Integral)**.** For $H \in \mathcal{L}_T^*$, the stochastic integral w.s.t. $M \in \mathcal{M}^{2,c}$ is defined by

$$\int_0^T H_s dM_s = \lim_{n\to\infty} \int_0^T H_s^{(n)} dM_s,$$

where and $H^{(n)}$ is a sequence satisfying (8.1) and the convergence is in L^2.

Remark 8.2.9. (i) Note that the convergence is also in L^1.

(ii) This definition is well-defined, i.e., independent of the choice of $H^{(n)}$. If there is another $K^{(n)}$, then we can construct $Z^{(n)}$ such that $Z^{(2n)} = H^{(n)}$ and $Z^{(2n+1)} = K^{(n)}$. So

$$\sup_{T>0} \lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left|Z_s^{(n)} - H_s\right|^2 d\langle M\rangle_s\right] = 0,$$

which implies that $\int_0^T Z_s^{(n)} dM_s$ is also Cauchy in L^2 and thus

$$\lim_{n\to\infty} \int_0^T Z_s^{(n)} dM_s = \lim_{n\to\infty} \int_0^T H_s^{(n)} dM_s = \lim_{n\to\infty} \int_0^T K_s^{(n)} dM_s.$$

(iii) If $t \mapsto \langle M\rangle_t$ is absolutely continuous a.e., then $\int_0^T H_s dM_s$ is well-defined if H is bounded, measurable and $\mathbb{F}$-adapted.

Proposition 8.2.10. *Let $M \in \mathcal{M}^{2,c}$ and $H, K \in \mathcal{L}_T^*(M)$ and $\alpha, \beta \in \mathbb{R}$.*

(1) Since H is $\mathbb{F}$-adapted, $\left(\int_0^t H_u dM_u\right)_{0\le t\le T} \in \mathcal{M}_0^{2,c}$

(2) Linearity:

$$\int_0^T \alpha H_u + \beta K_u dM_u = \alpha \int_0^T H_u dM_u + \beta \int_0^T K_u dM_u.$$

(3) Isometry:

$$\mathbb{E}\left[\left|\int_0^T H_u dM_u\right|^2\right] = \mathbb{E}\left[\int_0^T H_u^2 d\langle M\rangle_u\right].$$

(4) Moreover,

$$\mathbb{E}\left[\left|\int_s^t H_u dM_u\right|^2 \mid \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t H_u^2 d\langle M\rangle_u \mid \mathcal{F}_s\right].$$

(5)

$$\left\langle \int_0^\cdot H_u dM_u \right\rangle_t = \int_0^t H_u^2 d\langle M\rangle_u.$$

Proof. (1) Choose a simple process $H^{(n)}$ to approximate H. Because

$$\int_0^T H_s^{(n)} dM_s = \sum_{i=0}^{n-1} h^i (M_{T\wedge t_{i+1}} - M_{T\wedge t_i}) \in \mathcal{F}_T,$$

and

$$\int_0^T H_s dM_s = \lim_{n\to\infty} \int_0^T H_s^{(n)} dM_s,$$

it is $\mathcal{F}_T$-measurable. Moreover, because above convergence is in L^2 (L^2 implies L^1), $\int_0^T H_s dM_s \in L^1$. For any $0 \le s < t$ and any $A \in \mathcal{F}_s$, it suffice to prove

$$\mathbb{E}\left[\int_0^t H_u dM_u \mathbb{I}_A\right] = \mathbb{E}\left[\int_0^s H_u dM_u \mathbb{I}_A\right] \Leftrightarrow \mathbb{E}\left[\int_s^t H_u dM_u \mathbb{I}_A\right] = 0$$

Note that by L^1-convergence,

$$\mathbb{E}\left[\int_s^t H_u dM_u \mathbb{I}_A\right] = \lim_{n\to\infty} \mathbb{E}\left[\int_s^t H_u^{(n)} dM_u \mathbb{I}_A\right].$$

But because $(\int_0^t H_u^{(n)} dM_u)$ is a martingale, $\mathbb{E}\left[\int_s^t H_u^{(n)} dM_u \mathbb{I}_A\right] = 0$. So

$$\mathbb{E}\left[\int_s^t H_u dM_u \mathbb{I}_A\right] = 0.$$

(2) It is directly obtained by the linearity of $\int_0^t H_u^{(n)} dM_u$ and L^1-convergence.

(3) It is directly obtained by the same property of $\int_0^t H_u^{(n)} dM_u$ and also the L^1-convergence.

(4) For $A \in \mathcal{F}_s$, because $\mathbb{I}_A^2 = \mathbb{I}_A$, by (3),

$$\mathbb{E}\left[\left|\int_s^t H_u dM_u\right|^2 \mathbb{I}_A\right] = \mathbb{E}\left[\left|\int_s^t H_u \mathbb{I}_A dM_u\right|^2\right] = \mathbb{E}\left[\int_s^t H_u^2 d\left\langle M\right\rangle_u \mathbb{I}_A\right].$$

(5) For $0 \leq s < t$, by (1), $(\int_0^t H_u dM_u)_{t \geq 0}$ is a martingale. So by (4),

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t H_u dM_u\right)^2 - \left(\int_0^s H_u dM_u\right)^2 \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\left(\int_0^t H_u dM_u - \int_0^s H_u dM_u\right)^2 \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\int_s^t H_u d\left\langle M\right\rangle_u \mid \mathcal{F}_s\right].\end{aligned}$$

It follows that

$$\mathbb{E}\left[\left(\int_0^t H_u dM_u\right)^2 - \int_0^t H_u^2 d\langle M\rangle_u \mid \mathcal{F}_s\right] = \left(\int_0^s H_u dM_u\right)^2 - \int_0^s H_u^2 d\langle M\rangle_u.$$

Because $\int_0^t H_u^2 d\langle M\rangle_u$ is an increasing process and $\left(\int_0^t H_u dM_u\right)^2 - \int_0^t H_u^2 d\langle M\rangle_u$ is a martingale, by the uniqueness of Doob-Meyer decomposition,

$$\left\langle \int_0^\cdot H_u dM_u \right\rangle_t = \int_0^t H_u^2 d\left\langle M\right\rangle_u. \qquad \square$$

Remark 8.2.11. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if $\mathcal{G} \subset \mathcal{F}$, then by Jensen's inequality,

$$\left|\mathbb{E}\left[X \mid \mathcal{G}\right]\right|^p \leq \mathbb{E}\left[|X|^p \mid \mathcal{G}\right] \;\Rightarrow\; \left\|\mathbb{E}[X \mid \mathcal{G}]\right\|_p \leq \left\|X\right\|_p$$

for $1 \leq p < \infty$, which is also true for $p = \infty$. So for $\mathcal{G}$-measurable $X_n \to X$ in $L^p(\mathcal{F})$, by

$$\|\mathbb{E}[X \mid \mathcal{G}] - X\|_p \leq \left\|\mathbb{E}\left[X - X_n \mid \mathcal{G}\right]\right\|_p + \left\|X_n - X\right\|_p \leq 2\left\|X_n - X\right\|_p \to 0,$$

$X = \mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.

Corollary 8.2.12. *Consider a Brownian motion B and $H \in \mathcal{L}^*(B)$,*

$$\mathbb{E}\left[\int_s^t H_u dM_u \mid \mathcal{F}_s\right] = 0,$$

$$\mathbb{E}\left[\left(\int_s^t H_u dM_u\right)^2 \mid \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t H_u^2 du \mid \mathcal{F}_s\right] = \int_s^t \mathbb{E}\left[H_u^2\right] du$$

Theorem 8.2.13. *Let $M, N \in \mathcal{M}^{2,c}$ and $H \in \mathcal{L}^*(M)$ and $K \in \mathcal{L}^*(N)$.*

(1) *For stopping times $S \leq T$,*

$$\mathbb{E}\left[\int_0^{t\wedge T} H_u dM_u \mid \mathcal{F}_S\right] = \int_0^{t\wedge S} H_u dM_u.$$

(2) *For a stopping time T,*

$$\int_0^{t\wedge T} H_u dM_u = \int_0^t H_u \mathbb{I}_{[0,T]} dM_u = \int_0^t H_u dM_{u\wedge T}.$$

(3) *For stopping times $S \leq T$,*

$$\mathbb{E}\left[\left(\int_{t\wedge S}^{t\wedge T} H_u dM_u\right)\left(\int_{t\wedge S}^{t\wedge T} K_u dN_u\right) \mid \mathcal{F}_S\right] = \mathbb{E}\left[\left(\int_{t\wedge S}^{t\wedge T} H_u K_u d\left\langle M, N\right\rangle_u\right) \mid \mathcal{F}_S\right].$$

In particular, if S, T are constant

$$\mathbb{E}\left[\left(\int_s^t H_u dM_u\right)\left(\int_s^t K_u dN_u\right) \mid \mathcal{F}_s\right] = \mathbb{E}\left[\left(\int_s^t H_u K_u d\left\langle M, N\right\rangle_u\right) \mid \mathcal{F}_s\right].$$

Moreover, it follows that

$$\left\langle \int_0^{\cdot} H_u dM_u, \int_0^{\cdot} K_u dN_u \right\rangle_t = \mathbb{E}\left[\int_0^t H_u K_u d\left\langle M, N\right\rangle_u\right].$$

In particular,

$$\left\langle \int_0^{\cdot} H_u dM_u, N \right\rangle_t = \mathbb{E}\left[\int_0^t H_u d\left\langle M, N\right\rangle_u\right].$$

(4) *If $G \in \mathcal{L}^*\left(\int_0^{\cdot} H_u dM_u\right)$, then $GH \in \mathcal{L}^*(M)$ and*

$$\int_0^t G_s d\left(\int_0^s H_u dM_u\right) = \int_0^t G_u H_u dM_u.$$

Proposition 8.2.14 (Kunita-Watanabe)**.** *Let $M, N \in \mathcal{M}_0^{2,c}$ and $H \in \mathcal{L}^*(M)$ and $K \in \mathcal{L}^*(N)$. Then*

$$\int_0^t \left|H_u K_u\right| d\left\langle M, N\right\rangle_u \leq \left(\int_0^t H_u^2 d\left\langle M\right\rangle_u\right)^{\frac{1}{2}} \left(\int_0^t K_u^2 d\left\langle N\right\rangle_u\right)^{\frac{1}{2}}.$$

Remark 8.2.15. Condition (8.1) can be weaker as

$$\mathbb{P}\left(\int_0^T H_u^2 d\left\langle M\right\rangle_u < \infty\right) = 1,$$

but the convergence

$$\int_0^T H_u dM_u = \lim_{n\to\infty} \int_0^T H_u^{(n)} dM_u$$

is weaker to in probability. In such case, $\left(\int_0^t H_u dM_u\right)_{t\geq 0}$ is not a martingale, but a local martingale.

8.3 Integral w.s.t. Local (Semi) Martingale

Local martingale.

Definition 8.3.1. For $M \in \mathcal{M}^{c,loc}$ and $X \in \mathcal{L}^*(M)$, i.e.

$$\mathbb{E}\left[\int_0^T X_u^2 d\left\langle M\right\rangle_u\right] < \infty, \quad \forall\, T,$$

the stochastic integral of X w.s.t. M is defined by

$$\int_0^t X_s dM_s := \int_0^t X_s \mathbb{I}_{\{T_n \geq s\}} dM_{s\wedge T_n}$$

on $\{0 \leq t \leq T_n\}$, where $T_n \uparrow \infty$ is the sequence such that $(M_{t\wedge T_n})_{t\geq 0}$ is a UI martingale.

Remark 8.3.2. Here we do not need the condition of L^2-integrability for martingale because such T_n can be chosen such that $(M_{t\wedge T_n})_{t\geq 0}$ is L^2-integrable.

Theorem 8.3.3. *Let $M \in \mathcal{M}^{c,loc}$ and $X, Y \in \mathcal{L}^*(M)$.*

(1) $\left(\int_0^t X_u dM_u\right)$ *is a continuous local martingale, i.e., in* $\mathcal{M}_0^{c,loc}$.

(2) Linearity:

$$\int_0^t \left(\alpha X_s + \beta Y_s\right) dM_s = \alpha \int_0^t X_s dM_s + \beta \int_0^t Y_s dM_s.$$

(3) Quadratic variation:

$$\left\langle \int_0 X_s dM_s \right\rangle_t = \int_0^t X_s^2 d\langle M\rangle_s.$$

(4) For stopping time T,

$$\int_0^{t\wedge T} X_s dM_s = \int_0^t X_s I_{\{s\leq T\}} dM_s.$$

Remark 8.3.4. Note that for the properties related to expectation cannot be extended to local martingale, like

$$\mathbb{E}\left[\left(\int_0^t X_u dM_u\right)^2\right] \neq \mathbb{E}\left[\int_0^t X_u^2 d\langle M\rangle_u\right], \quad \mathbb{E}\left[\left(\int_s^t X_u dM_u\right)^2 \mid \mathcal{F}_s\right] \neq \mathbb{E}\left[\int_s^t X_u^2 d\langle M\rangle_u \mid \mathcal{F}_s\right],$$

in general.

Semimartingale. Recall $X = (X_t)_{t\geq 0}$ is a semimartingale if

$$X_t = X_0 + M_t^X + A_t^X,$$

where $M^X = (M_t^X)_{t\geq 0}$ is a local martingale with $M_0^X = 0$ and $A^X = (A_t^X)_{t\geq 0}$ is a càdlàg, adapted process of bounded variation. Note that this decomposition if not unique unless X is continuous. In the following, we consider continuous X.

Definition 8.3.5. Let X be a continuous semimartingale. For $H \in \mathcal{L}^*(M^X)$, define

$$\int_0^t H_s dX_s = \int_0^t H_s dM_s^X + \int_0^t H_s dA_s^X,$$

where the second integral is the Riemannian-Stieltjes integral.

Next, we need to define the quadratic variation for general case.

Definition 8.3.6. Let X, Y be semimartingales.

(1) The quadratic variation of X is defined as

$$[X, X]_t := \lim_{n\to\infty} \sum_{t_i\in\Pi_n, t_i\le t} \left(X_{t_{i+1}\wedge t} - X_{t_i\wedge t}\right)^2.$$

(2) The cross variation of X, Y is

$$[X, Y]_t = \frac{1}{4}\left([X+Y, X+Y]_t - [X-Y, X-Y]_t\right).$$

Remark 8.3.7. If X, Y are two continuous local martingale, then $[X, X]_t = \langle X\rangle_t$ and $[X, Y]_t = \langle X, Y\rangle_t$.

Theorem 8.3.8. *If X, Y are semimartingales and let $M^{X,c}, M^{Y,c}$ be their continuous local martingale parts, then*

$$[X, Y]_t = \left\langle M^{X,c}, M^{Y,c}\right\rangle_t + \sum_{s\le t} \Delta X_s \Delta Y_s,$$

where $\Delta X_s = X_s - X_{s-}$.

Remark 8.3.9. In particular, if X, Y are continuous semimartingales, then

$$[X, Y]_t = \left\langle M^X, M^Y\right\rangle_t.$$

Corollary 8.3.10. *If X, Y are continuous semimartingale and $Y \in \mathcal{L}^*(M^X)$, then*

$$\left[\int_0^{\cdot} H_s dX_s, Y\right]_t = \int_0^t H_s d[X, Y]_s.$$

Theorem 8.3.11 (DCT)**.** *Let X be a continuous semimartingale with the decomposition $X_t = X_0 + M_t + A_t$. Let $H^{(n)}$ and H be locally bounded progressive processes, and let K be a nonnegative progressive process. If*

(i) $H_s^{(n)} \to H_s$ *a.e. for any* $s \in [0, t]$,

(ii) $\left|H_s^{(n)}\right| \le K_s$ *a.e. for any n and* $s \in [0, t]$,

(iii) $K_s \in \mathcal{L}^*$ *and* $\int_0^t |K_s|\,|dA_s| < \infty$,

then

$$\int_0^t H_s^{(n)} dX_s \to \int_0^t H_s dX_s.$$

8.4 Itô Formula

Theorem 8.4.1 (1-dim, Continuous Form)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a C^2-function and $X = (X_t)_{t\ge 0}$ be a continuous semimartingale with the decomposition $X_t = X_0 + M_t + A_t$. Then*

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_0) dX_u + \frac{1}{2}\int_0^t f''(X_u) d[X, X]_u \\ &= f(X_0) + \int_0^t f'(X_0) dM_u + \int_0^t f'(X_0) dA_u + \frac{1}{2}\int_0^t f''(X_u) d\langle M\rangle_u. \end{aligned}$$

Sketch of Proof. By Taylor formula,

$$f(X_{t_{i+1}\wedge t}) - f(X_{t_i}) = f'(X_{t_i})\Delta_i X + \frac{1}{2}f''(X_{t_i})(\Delta_i X)^2 + R_i.$$

By taking the summation,

$$f(X_t) - f(X_0) = \sum_i f'(X_{t_i})\Delta_i X + \frac{1}{2}\sum_i f''(X_{t_i})(\Delta_i X)^2 + \sum_i R_i.$$

By the definition of integral,

$$\sum_i f'(X_{t_i})\Delta_i X \to \int_0^t f'(X_u)dX_u.$$

By the definition of quadratic variation,

$$\sum_i f''(X_{t_i})(\Delta_i X)^2 \to \int_0^t f''(X_u)d[X,X]_u.$$

For R_i, because $f \in C^2$

$$\left|\sum_i R_i\right| \le \frac{1}{2}\sum_i |f''(\xi_i) - f''(X_{t_i})|\,|\Delta_i X| \le \frac{1}{2}\sum_i |\Delta_i X|^2 \to 0. \qquad \square$$

Remark 8.4.2. (i) Note that $\int_0^t f'(X_0)dM_u$ is a continuous local martingale, and $\int_0^t f'(X_0)dA_u + \int_0^t f''(X_u)d\langle M\rangle_u$ is of bounded variation (because integral of bounded variation is still of bounded variation). So $f(X_t)$ is also a continuous semi-martingale.

(ii) It has the differential form

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)d[X,X]_t \\ &= f'(X_t)dM_t + f'(X_t)dA_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t, \end{aligned}$$

which is just a formula.

(iii) If X is a continuous semimartingale, we denote $(dX_t)^2 = d[X,X]_t$. Therefore, the different form becomes

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2.$$

In particular, for $X = B$ a Brownian motion, $(dB_t)^2 = dt$. For example, if $dX_t = fdB_t + gdt$, then

$$(dX_t)^2 = f^2(dB_t)^2 + 2fgdB_tdt + g^2(dt)^2 = f^2(dB_t)^2 = f^2dt.$$

Example 8.4.3. (1) Let $f(x) = x^2$ and $X = B$ a Brownian motion. Then

$$B_t^2 = B_0^2 + 2\int_0^t B_s dB_s + \int_0^t ds \;\Rightarrow\; \int_0^t B_s dB_s = \frac{1}{2}\left(B_t^2 - t\right).$$

(2) Let W be a Brownian motion and $X \in \mathcal{L}^*(W)$. Consider the process

$$Z_t = \exp\left(\int_0^t X_u dW_u - \frac{1}{2}\int_0^t X_u^2 du\right).$$

Let

$$Y_t = \int_0^t X_u dW_u - \frac{1}{2}\int_0^t X_u^2 du,$$

which is a semimartingale, or informally, $dY_t = X_t dW_t - \frac{1}{2}X_t^2 dt$. Let $f(x) = e^x$. Then

$$\begin{aligned} dZ_t &= f'(Y_t)dY_t + \frac{1}{2}f''(Y_t)(dY_t)^2 \\ &= Z_t\left(X_t dW_t - \frac{1}{2}X_t^2 dt\right) + \frac{1}{2}Z_t X_t^2 dt \\ &= Z_t X_t dW_t. \end{aligned}$$

Therefore,

$$Z_t = Z_0 + \int_0^t Z_u X_u dW_u = 1 + \int_0^t Z_u X_u dW_u$$

Moreover, $Z_t = \exp\left(\int_0^t X_u dW_u - \frac{1}{2}\int_0^t X_u^2 du\right)$ is a solution of SDE

$$dZ_t = Z_t X_t dW_t.$$

In particular, if $X_t \equiv \sigma$, then

$$Z_t = \exp\left(\sigma W_t - \frac{1}{2}\sigma^2 t\right)$$

is a solution of $dZ_t = \sigma Z_t dW_t$.

Remark 8.4.4. Note that Z is a continuous local martingale. In fact, for X with $\mathbb{P}(\int_0^T X_u^2 du < \infty) = 1$, it is also a local martingale.

Theorem 8.4.5 (Multi-dim, Local Martingale, Continuous Form)**.** *Let $\boldsymbol{X} = (X^1, \cdots, X^n)$ be a vector of continuous local martingales. Let $f\colon [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$ be $C^{1,2}$.*

$$\begin{aligned} f(\boldsymbol{X}_t) = f(0, \boldsymbol{X}_0) &+ \int_0^t \frac{\partial}{\partial t} f(s, \boldsymbol{X}_s)dt + \sum_{i=1}^n \int_0^t \frac{\partial}{\partial x_i} f(s, \boldsymbol{X}_s) dX_s^i \\ &+ \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s, \boldsymbol{X}_s) d\left\langle X^i, X^j \right\rangle_s. \end{aligned}$$

or in differential form

$$df(X_t) = \frac{\partial}{\partial t} f(t, \boldsymbol{X}_t)dt + \sum_{i=1}^n \frac{\partial}{\partial x_i} f(t, \boldsymbol{X}_t) dX_t^i + \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} f(t, \boldsymbol{X}_t) d\left\langle X^i, X^j \right\rangle_t.$$

Example 8.4.6. Let $\boldsymbol{W} = (W^1, \cdots, W^n)$ be a n-dimensional standard Brownian motion with $n \geq 2$. Let

$$R_t = \|\boldsymbol{W}\| = \sqrt{\sum_{i=1}^n (W^1)^2 + \cdots + (W^n)^2},$$

called the Bessel process. Let $f(\boldsymbol{x}) = \|x\|$. Then $\frac{\partial}{\partial x_i} f(\boldsymbol{x}) = x_i / f(\boldsymbol{x})$ and

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{x}) = \begin{cases} -\frac{x_i x_j}{f(\boldsymbol{x})^3}, & i \neq j \\ \frac{f(\boldsymbol{x})^2 - x_i^2}{f(\boldsymbol{x})^3}, & i = j. \end{cases}$$

Then we have

$$\begin{aligned} dR_t &= \sum_{i=1}^n \frac{\partial}{\partial x_i} f(\boldsymbol{W}_t) dW_t^i = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{W}_t) d\left\langle B^i, B^j \right\rangle_t \\ &= \sum_{i=1}^n \frac{W_t^i}{R_t} dW_t^i + \frac{1}{2} \sum_{i=1}^n \frac{R_t^2 - (W_i^t)^2}{R_t^3} dt \\ &= \sum_{i=1}^n \frac{W_t^i}{R_t} dW_t^i + \frac{1}{2} \frac{n-1}{R_t} dt. \end{aligned}$$

Therefore,

$$R_t dR_t = \sum_{i=1}^n W_t^i dW_t^i + \frac{n-1}{2} dt.$$

Theorem 8.4.7 (Itô Formula)**.** *Let $\boldsymbol{X} = (X^1, \cdots, X^n)$ be a n-dim semimartingale with decomposition*

$$X_t^i = X_0^i + M_t^i + A_t^i, \quad i = 1, \cdots, n$$

(Note that because the decomposition is not unique, it should be given.) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be C^2. Then $f(\boldsymbol{X})$ is a semimartingale and

$$\begin{aligned} f(\boldsymbol{X}_t) = f(\boldsymbol{X}_0) &+ \sum_{i=1}^n \int_0^t \frac{\partial}{\partial x_i} f(\boldsymbol{X}_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{X}_s) d\left\langle M^{i,c}, M^{j,c} \right\rangle_s \\ &+ \sum_{s \leq t} \left(f(\boldsymbol{X}_s) - f(\boldsymbol{X}_{s-}) - \sum_{i=1}^n \frac{\partial}{\partial x_i} f(\boldsymbol{X}_{s-}) \Delta X_{s-}^i \right). \end{aligned}$$

Application: Integration by parts.

Example 8.4.8. Given a standard Brownian motion W. Consider $\int_0^t s dW_s$. Let $f(t,x) = tx$. Then

$$\frac{\partial}{\partial t} f(t,x) = x, \; \frac{\partial}{\partial x} f(t,x) = t, \; \frac{\partial^2}{\partial t \partial x} f(t,x) = 0.$$

So

$$tW_t = \int_0^t W_s ds + \int_0^t s dW_s \;\Rightarrow\; \int_0^t s dW_s = tW_t - \int_0^t W_s ds.$$

Theorem 8.4.9. *Let $f(s,\omega)$ is continuous of bounded variation w.s.t. for a.e. ω. Then*

$$\int_0^t f(s) dW_s = f(t) W_t - \int_0^t W_s df(s).$$

Theorem 8.4.10 (Integration by parts)**.** *Suppose X, Y are continuous semimartingale, then*

$$\int_0^t X_s dY_s = X_t Y_t - X_0 Y_0 + \int_0^t Y_s dX_s - [X,Y]_t.$$

on informally,

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d[X,Y]_t.$$

Proof. It can be obtained by Itô formula on $f(x,y)=xy$. $\square$

Remark 8.4.11. In general, it can be written as

$$d(X_tY_t)=X_tdY_t+Y_tdX_t+(dX_t)(dY_t),$$

where X_t, Y_t can be continuous semimartingales ($M_t=0$, i.e., as above theorem, or $M_t\neq 0$) or deterministic and for calculating $(dX_t)(dY_t)$, we can use

$$(dB_t)^2=0,\ dtdB_t=(dt)^2=0.$$

Remark 8.4.12. If X,Y are semimartingale,

$$X_tY_t=X_0Y_0+\int_0^t X_{u-}dY_u+\int_0^t Y_{u-}dX_u+[X,Y]_t.$$

In particular,

$$X_t^2=X_0^2+2\int_0^t X_{u-}dX_u+[X,X]_t.$$

8.5 Martingale Representation Theorem

The problem is whether a martingale can be represented as

$$M_t=M_0+\int_0^t H_sdB_s.$$

Example 8.5.1. Let W^1,W^2 be two independent Brownian motions and $\mathcal{F}_t=\sigma(W_s^1,W_s^2\colon s\leq t)$. Then W^1,W^2 are two martingales w.s.t. $\mathbb{F}=(\mathcal{F}_t)$. If

$$W_t^2=\int_0^t H_sdW_s^1$$

for some $H\in\mathcal{L}^*(W^1)$, then

$$t=\left\langle W^2,W^2\right\rangle_t=\left\langle\int_0^\cdot H_sdW^1,W^2\right\rangle_t=\int_0^t H_sd\left\langle W^1,W^2\right\rangle_s=0,$$

which induces a contradiction.

Theorem 8.5.2 (Martingale Representation Theorem)**.** *Let $\boldsymbol{B}$ be a n-dimensional Brownian motion w.s.t. its natural filtration $\mathbb{F}^B$. Let M be a martingale w.s.t. $\mathbb{F}$ and $\mathbb{F}^B$ that is in $\mathcal{M}^2$ and càdlàg. Then there exist $H^i\in\mathcal{L}^*$ for all i such that*

$$M_t=M_0+\sum_{i=1}^n\int_0^t H_s^idB_s^i.$$

Remark 8.5.3. Note that $\mathcal{F}_t^B=\sigma(B_s^1,\cdots,B_s^n\colon s\leq t)$.

8.6 Girsanov Theorem

Fix $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual condition. Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Denote $\mathbb{E} = \mathbb{E}_{\mathbb{P}}$.

Example 8.6.1. Let $Z_1, \cdots, Z_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$. Let

$$d\tilde{\mathbb{P}}(\omega) = \exp\left(\sum_{i=1}^n \mu_i Z_i - \frac{1}{2}\sum_{i=1}^n \mu_i^2\right) d\mathbb{P}(\omega).$$

Denote $\tilde{\mathbb{E}} = \mathbb{E}_{\tilde{\mathbb{P}}}$. Consider the characteristic equation,

$$\begin{aligned}
\tilde{\mathbb{E}}\left[\exp(it_1Z_1 + \cdots it_nZ_n)\right] &= \int_\Omega \exp(it_1Z_1 + \cdots it_nZ_n) d\tilde{\mathbb{P}} \\
&= \int_\Omega \exp(it_1Z_1 + \cdots it_nZ_n)\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} d\mathbb{P} \\
&= \int_\Omega \exp\left(\sum_{j=1}^n ((\,)\,(it_j + \mu_j)Z_j - \frac{1}{2}\mu_j^2)\right) d\mathbb{P} \\
&= \mathbb{E}\left[\prod_{j=1}^n \exp\left((it_j + \mu_j)Z_j - \frac{1}{2}\mu_j^2\right)\right] \\
&= \prod_{j=1}^n \mathbb{E}\left[\exp\left((it_j + \mu_j)Z_j - \frac{1}{2}\mu_j^2\right)\right] \\
&= \prod_{j=1}^n \exp\left(-\frac{t_j^2}{2} + it_j\mu_j\right).
\end{aligned}$$

It follows that $Z_1, \cdots, Z_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_i, 1)$ on $\tilde{\mathbb{P}}$.

Let X be a measurable, $\mathbb{F}$-adapted stochastic process such that

$$\mathbb{P}\left(\int_0^T X_u^2 du < \infty\right) = 1, \quad \forall\, T \geq 0.$$

Define

$$Z_t := \exp\left(\int_0^t X_u dW_u - \frac{1}{2}\int_0^t X_u^2 du\right).$$

Then by above

$$Z_t = 1 + \int_0^t Z_u X_u dW_u,$$

$Z = (Z_t)_{t\geq 0}$ is a continuous local martingale.

Proposition 8.6.2 (Novikov Condition)**.** *If X be a measurable, $\mathbb{F}$-adapted stochastic process such that*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T X_u^2 du\right)\right] < \infty, \quad \forall\, 0 \leq T < \infty,$$

then Z is a martingale.

Remark 8.6.3. Note that by Jensen's inequality, if X satisfies Novikov condition, it is in $\mathcal{L}^*(W)$.

Define $\tilde{\mathbb{P}}_t$ on $\mathcal{F}_t$ by,

$$\tilde{\mathbb{P}}_t(A) := \int_A Z_t d\mathbb{P}, \quad \forall\ A \in \mathcal{F}_t,\ \Rightarrow\ Z_t = \frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}}.$$

Remark 8.6.4. If $Z = (Z_t)_{t\geq 0}$ is a martingale, then

(i) $\tilde{\mathbb{P}}_t$ is a probability measure because $\mathbb{E}[Z_t] = \mathbb{E}[Z_0] = 1$.

(ii) for any $s \leq t$ and any $A \in \mathcal{F}_s$,

$$\tilde{\mathbb{P}}_s(A) = \tilde{\mathbb{P}}_t(A)$$

because of the martingale property of Z.

Theorem 8.6.5 (Girsanov Theorem). *Assume that $Z = (Z_t)_{t\geq 0}$ defined as above is a martingale. Define a process $\tilde{W}$ as*

$$\tilde{W}_t = W_t - \int_0^t X_u du.$$

Then for each fixed $T \in [0, \infty)$, $(\tilde{W}_t)_{t\in[0,T]}$ is a Brownian motion on $(\Omega, \mathcal{F}_T, \tilde{P}_T)$.

Corollary 8.6.6. *Under the same assumption of above theorem, suppose f is a measurable function such that $f(W_t) \in L^1$. Then*

$$\mathbb{E}_{\mathbb{Q}}[f(\tilde{W}_t)] = \mathbb{E}[f(W_t)].$$

Example 8.6.7. (1) Suppose $X_t = h(t)$, a deterministic function, that is in $L^2([0,T])$. Since

$$\mathbb{E}\left[\frac{1}{2}\exp\left(\int_0^T h^2(u)du\right)\right] = \frac{1}{2}\exp\left(\int_0^T h^2(u)du\right) < \infty,$$

by above theorem

$$Z_t = \exp\left(\int_0^t h(u)dW_u - \frac{1}{2}\int_0^t h^2(u)du\right)$$

is a martingale. So

$$\tilde{W}_t = W_t - \int_0^t h(u)du$$

is a Brownian motion w.s.t. $\tilde{\mathbb{P}}$ defined as,

$$d\tilde{\mathbb{P}} = \exp\left(\int_0^T h(u)dW_u - \frac{1}{2}\int_0^T h^2(u)du\right) d\mathbb{P}.$$

(2) Consider $X_t = \mathrm{sign}(W_t)$ (i.e. $X_t = 1$ for $W_t \geq 0$ and $X_t = -1$ for $W_t < 0$).

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\right)\int_0^T X_u^2 du\right] = \exp\left(\frac{T}{2}\right) < \infty.$$

So

$$\tilde{W}_t = W_t - \int_0^t \mathrm{sign}(W_u)du$$

is a Brownian motion w.s.t. $\mathbb{Q}$ given by

$$d\mathbb{Q} = \exp\left(\int_0^T \mathrm{sign}(W_u)du - \frac{T}{2}\right) d\mathbb{P}.$$

Remark 8.6.8. Suppose $Y = (Y_t)_{t\geq 0}$ is a martingale w.s.t. to $d\mathbb{Q} = Z_T d\mathbb{P}$, i.e.,

$$Y_s = \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_s\right] = \frac{\mathbb{E}[Y_t Z_T \mid \mathcal{F}_s]}{\mathbb{E}[Z_T \mid \mathcal{F}_s]} = \frac{\mathbb{E}[Y_t Z_T \mid \mathcal{F}_s]}{Z_s}.$$

In particular,

$$\mathbb{E}\left[Y_T Z_T \mid \mathcal{F}_s\right] = Y_s Z_s.$$

8.7 Local Times

Let $W = (W_t)$ be a standard Brownian motion. Define the level set

$$\{0 \le t < \infty \colon W_t(\omega) = x\}.$$

Obviously, its Lebesgue measure is 0.

Definition 8.7.1. For any Borel set $B \in \mathcal{B}$, define the occupation time of B as

$$\Gamma_t(B) = \int_0^t \mathbb{I}_{W_s \in B} ds = m\left(\{s \in [0,t] \colon W_s \in B\}\right).$$

Note that it is a random variable.

Remark 8.7.2. The stochastic process $\Gamma(B) = (\Gamma_t(B))_{t \ge 0}$ is adapted and continuous.

Definition 8.7.3 (Local Time)**.** For a given Brownian motion W, the local time is defined as

$$L_t(x) = L_t(x,\omega) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \Gamma_t([x-\varepsilon, x+\varepsilon]).$$

Note it is a random variable.

Remark 8.7.4. Local time serves as a density function w.s.t. the Lebesgue measure for the occupation time, i.e.,

$$\Gamma_t(B,\omega) = \int_B L_t(x,\omega) dx.$$

Theorem 8.7.5. *The local times of a Brownian motion exist.*

Consider $g(x) = |x|$. For $\varepsilon > 0$, let

$$g_\varepsilon(x) = \begin{cases} |x|, & |x| > \varepsilon \\ \frac{1}{2}\left(\varepsilon + \frac{x^2}{\varepsilon}\right), & |x| \le \varepsilon. \end{cases}$$

Then $g_\varepsilon \in C^1$ and

$$g'_\varepsilon(x) = \begin{cases} 1, & x \ge \varepsilon \\ \frac{x}{\varepsilon}, & |x| < \varepsilon \\ -1, & x < -\varepsilon. \end{cases}$$

Moreover, for $|x| \ne \varepsilon$,

$$g''_\varepsilon(x) = \begin{cases} 0, & |x| > \varepsilon \\ \frac{1}{\varepsilon}, & |x| < \varepsilon. \end{cases}$$

Therefore,

$$\begin{aligned} g_\varepsilon(W_t) &= g_\varepsilon(W_0) + \int_0^t g'_\varepsilon(W_s) dW_s + \frac{1}{2}\int_0^t g''(W_s) ds \\ &= g_\varepsilon(0) + \int_0^t g'_\varepsilon(W_s) dW_s + \frac{1}{2}\int_0^t \frac{1}{\varepsilon} \mathbb{I}_{\{|W_s| < \varepsilon\}} ds \\ &= \frac{\varepsilon}{2} + \int_0^t g'_\varepsilon(W_s) dW_s + \frac{1}{2\varepsilon}\Gamma_t([-\varepsilon], -\varepsilon). \end{aligned}$$

For the second term,

$$\begin{aligned}\int_0^t g_\varepsilon'(W_s)dW_s &= \int_0^t g_\varepsilon'(W_s)\mathbb{I}_{\{|W_s|\le\varepsilon\}}dW_s + \int_0^t g_\varepsilon'(W_s)\mathbb{I}_{\{|W_s|>\varepsilon\}}dW_s \\ &= \int_0^t g_\varepsilon'(W_s)\mathbb{I}_{\{|W_s|\le\varepsilon\}}dW_s + \int_0^t \operatorname{sign}(W_s)\mathbb{I}_{\{|W_s|>\varepsilon\}}dW_s.\end{aligned}$$

Note that as $\varepsilon \to 0$, by the DCT for stochastic integral,

$$\int_0^t \operatorname{sign}(W_s)\mathbb{I}_{\{|W_s|>\varepsilon\}}dW_s \to \int_0^t \operatorname{sign}(W_s)dW_s.$$

For the other one,

$$I = \int_0^t g_\varepsilon'(W_s)\mathbb{I}_{\{|W_s|\le\varepsilon\}}dW_s = \int_0^t \frac{W_s}{\varepsilon}\mathbb{I}_{\{|W_s|\le\varepsilon\}}dW_s$$

Note that because I is a martingale, $\mathbb{E}[I]=0$ and

$$\begin{aligned}\mathbb{E}[I^2] &= \mathbb{E}\left[\int_0^t \frac{W_s^2}{\varepsilon^2}\mathbb{I}_{\{|W_s|\le\varepsilon\}}ds\right] \\ &= \int_0^t \frac{1}{\varepsilon^2}\mathbb{E}\left[W_s^2\mathbb{I}_{\{|W_s|\le\varepsilon\}}\right]ds \\ &\le \int_0^t \mathbb{E}\left[\mathbb{I}_{\{|W_s|\le\varepsilon\}}\right]ds \\ &= \int_0^t \mathbb{P}(|W_s|\le\varepsilon)ds \\ &= \int_0^t \mathbb{P}(|W_1|\le\varepsilon/\sqrt{s})ds \\ &= \int_0^t \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon/\sqrt{s}}^{\varepsilon/\sqrt{s}} e^{-y^2}dyds \to 0.\end{aligned}$$

Therefore, as $\varepsilon \to 0$, we get the Tanaka formula,

$$|W_t| = \int_0^t \operatorname{sign}(W_s)dW_s + L_t.$$

Corollary 8.7.6. *Fix $a \in \mathbb{R}$,*

$$\begin{aligned}|W_t - a| &= |a| + \int_0^t \operatorname{sign}(W_s - a)\,dW_s + L_t(a) \\ (W_t - a)^+ &= (-a)^+ + \int_0^t \mathbb{I}_{(a,\infty)}(W_s)\,dW_s + \frac{1}{2}L_t(a) \\ (W_t - a)^- &= (-a)^- - \int_0^t \mathbb{I}_{(-\infty,a]}(W_s)\,dW_s + \frac{1}{2}L_t(a).\end{aligned}$$

Remark 8.7.7. For every Borel measurable function $f\colon \mathbb{R} \to [0,\infty)$,

$$\int_0^t f(W_s)ds = \int_{-\infty}^{\infty} f(x)L_t(x)dx.$$

Remark 8.7.8. For any semimartingale $X = X_0 + M + A$, we have the similar definition of local time Λ_t, which satisfies

$$\int_0^t f(X_s)\, d\langle M\rangle_s = \int_{-\infty}^{\infty} f(x)\Lambda_t(x)dx, \quad 0 \leq t < \infty,$$

and the Tanaka-Meyer formula

$$|X_t - a| = |X_0 - a| + \int_0^t \operatorname{sign}(X_s - a)\, dX_s + \Lambda_t(a).$$

Chapter 9

Stochastic Differential Equation

Consider a stochastic process satisfying:

$$dX_t = b(t, X_t, W_t)dt + \sigma(t, X_t, W_t)dW_t. \tag{$*$}$$

for two problems:

(1) existence and uniqueness and properties of solutions,

(2) how to solve for particular cases.

9.1 Examples

Example 9.1.1 (Geometric Brownian Motion)**.** Solving

$$dX_t = \alpha X_t dt + \sigma X_t dW_t.$$

where the initial value X_0 is given, and α, σ are constant.
Solution: Diving X_t on the both side and integrating,

$$\int_0^t \frac{dX_u}{X_u} = \alpha t + \sigma W_t.$$

Let $f(x) = \log x$. Then by Itô formula,

$$\begin{aligned} d\log X_t &= \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}(dX_t)^2 \\ &= \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 X_t^2 dt \\ &= \frac{1}{X_t}dX_t - \frac{1}{2}\sigma^2 dt. \end{aligned}$$

So

$$\log X_t - \log X_0 = \int_0^t \frac{dX_u}{X_u} - \frac{1}{2}\sigma^2 t.$$

It follows that

$$X_t = X_0 \exp\left(\sigma W_t + \left(\alpha - \frac{1}{2}\sigma^2\right) t\right).$$

Remark 9.1.2. (1) If $(W_t)_{t\geq 0}$ is independent of X_0, then

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[X_0]\mathbb{E}\left[\exp\left(\sigma W_t + \left(\alpha - \frac{1}{2}\sigma^2\right)t\right)\right] \\ &= e^{\alpha t}\mathbb{E}[X_0]\mathbb{E}\left[\exp\left(\sigma W_t - \frac{1}{2}\sigma^2 t\right)\right] \\ &= e^{\alpha t}\mathbb{E}[X_0],\end{aligned}$$

because $\exp\left(\sigma W_t - \frac{1}{2}\sigma^2 t\right)$ is a martingale.

(2) If $\alpha > \frac{1}{2}\sigma^2$, then $X_t \to \infty$ as $t \to \infty$. If $\alpha < \frac{1}{2}\sigma^2$, $X_t \to 0$ as $t \to \infty$. If $\alpha = \frac{1}{2}\sigma^2$, X_t will fluctuate between arbitrary large and arbitrary small values.

Definition 9.1.3. A stochastic process $(X_t)_{t\geq 0}$ of the form

$$X_t = X_0 \exp(\sigma W_t + \mu t)$$

is called the geometric Brownian motion.

Example 9.1.4 (Hull-White Interest Rate Model)**.** Consider SDE

$$dR_t = (a_t - b_t R_t) + \sigma_t dW_t, \quad R_0 = r,$$

where a_t, b_t, σ_t are deterministic.
Solution:

$$dR_t + b_t dR_t = a_t dt + \sigma_t dW_t.$$

Multiplying $e^{\int_0^t b_u du}$,

$$e^{\int_0^t b_u du} dR_t + e^{\int_0^t b_u du} b_t dR_t = e^{\int_0^t b_u du} a_t dt + \sigma_t e^{\int_0^t b_u du} dW_t.$$

Because $e^{\int_0^t b_u du}$ is of bounded variation on any interval, by Itô formula

$$d\left(e^{\int_0^t b_u du}\right) dR_t = e^{\int_0^t b_u du} a_t dt + \sigma_t e^{\int_0^t b_u du} dW_t.$$

So

$$R_t = re^{-\int_0^t b_u du} + r\int_0^t e^{\int_t^s b_u du} a_s ds + \int_0^t \sigma_s e^{\int_t^s b_u du} dW_s.$$

Example 9.1.5. Consider SDE

$$dX_t = rX_t(K - X_t)dt + \beta X_t dW_t, \quad X_0 = x > 0.$$

Solution:

$$\frac{1}{X_t} dX_t + rX_t dt = rK dt + \beta dW_t.$$

Therefore,

$$\int_0^t \frac{1}{X_t} dX_t + \int_0^t rX_t dt = rKt + \beta W_t.$$

For the left hand side, first

$$\int_0^t \frac{1}{X_t} dX_t = \log X_t - \log X_0 + \frac{1}{2}\int_0^t \frac{1}{X_s^2}(dX_t)^2$$

$$
\begin{aligned}
&= \log X_t - \log X_0 + \frac{1}{2}\int_0^t \frac{1}{X_s^2}\beta^2 X_s^2 ds \\
&= \log \frac{X_t}{x} + \frac{1}{2}\beta^2 t.
\end{aligned}
$$

Therefore,

$$
X_t \exp\left(r\int_0^t X_s ds\right) = x\exp\left(\beta W_t + \left(rK - \frac{1}{2}\beta^2\right)t\right).
$$

Integrating w.s.t. t,

$$
\begin{aligned}
\int_0^t \exp\left(r\int_0^t X_s ds\right) d\left(\int_0^t X_s ds\right) &= \frac{1}{r}\left(\exp\left(r\int_0^t X_s ds\right) - 1\right) \\
&= x\int_0^t \exp\left(\beta W_s + \left(rK - \frac{1}{2}\beta^2\right)s\right) ds.
\end{aligned}
$$

It follows that

$$
\exp\left(r\int_0^t X_s ds\right) = 1 + rx\int_0^t \exp\left(\beta W_s + \left(rK - \frac{1}{2}\beta^2\right)s\right) ds,
$$

and so

$$
\int_0^t X_s ds = \frac{1}{r}\log\left(1 + rx\int_0^t \exp\left(\beta W_s + \left(rK - \frac{1}{2}\beta^2\right)s\right) ds\right).
$$

Differentiating w.s.t. t,

$$
X_t = \frac{x\exp\left(\beta W_t + \left(rK - \frac{1}{2}\beta^2\right)t\right)}{1 + rx\int_0^t \exp\left(\beta W_s + \left(rK - \frac{1}{2}\beta^2\right)s\right) ds}.
$$

Example 9.1.6. Consider SDE

$$
dX_t = \alpha_t dt + b_t X_t dW_t.
$$

Solution: Trying to find an integrator ρ_t for

$$
\rho_t dX_t - b_t\rho_t X_t dW_t = \alpha_t\rho_t dt.
$$

By Itô formula,

$$
d(\rho_t X_t) = \rho_t dX_t + X_t d\rho_t + (dX_t)(d\rho_t)
$$

First, try to find ρ_t such that

$$
X_t d\rho_t = -b_t\rho_t X_t dW_t \;\Rightarrow\; \frac{d\rho_t}{\rho_t} = -b_t dW_t.
$$

Then by above

$$
\log\frac{\rho_t}{\rho_0} + \frac{1}{2}\int_0^t \frac{1}{\rho_t^2}(d\rho_t)^2 = \log\frac{\rho_t}{\rho_0} + \frac{1}{2}\int_0^t b_u^2 du = -\int_0^t b_u dW_u,
$$

which implies that

$$
\rho_t = \exp\left(-\int_0^t b_u dW_u - \frac{1}{2}\int_0^t b_u^2 du\right)
$$

by setting $\rho_0 = 1$. Then

$$
d(\rho_t X_t) = \rho_t dX_t + X_t d\rho_t + (dX_t)(d\rho_t)
$$

$$
\begin{aligned}
&= \rho_t dX_t - -b_t\rho_t X_t dW_t - b_t^2\rho_t X_t dt \\
&= \alpha_t\rho_t dt - b_t^2\rho_t X_t dt.
\end{aligned}
$$

So

$$d(\rho_t X_t) + b_t^2\rho_t X_t dt = \alpha_t\rho_t dt.$$

Multiplying $e^{\int_0^t b_u^2 du}$, then we have

$$d\left(e^{\int_0^t b_u^2 du}\rho_t X_t\right) = e^{\int_0^t b_u^2 du}\alpha_t\rho_t dt.$$

It follows that

$$\rho_t X_t - X_0 = \int_0^t e^{\int_t^s b_u^2 du}\alpha_s\rho_s ds,$$

i.e.,

$$X_t = X_0 \exp\left(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds\right) + \int_0^t \alpha_s \exp\left(\int_s^t b_u dW_u - \frac{1}{2}\int_s^t b_u^2 du\right) ds.$$

Example 9.1.7. Consider SDE

$$LX_t'' + RX_t' + \frac{1}{2}X_t = G_t + \alpha\tilde{W}_t,$$

where $\tilde{W}_t$ is white noise, i.e., $\tilde{W}_t dt = dW_t$ (in distribution meaning).
Solution: Introduce

$$\boldsymbol{X}_t = \begin{pmatrix} X_t \\ X_t' \end{pmatrix}$$

and it follows that

$$d\boldsymbol{X}_t = A\boldsymbol{X}_t dt + \boldsymbol{H}_t dt + \boldsymbol{K} dW_t,$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{CL} & -\frac{R}{L} \end{pmatrix}, \quad \boldsymbol{H}_t = \begin{pmatrix} 0 \\ \frac{G_t}{L} \end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix} 0 \\ \frac{\alpha}{L} \end{pmatrix}.$$

So

$$\boldsymbol{X}_t = \exp(At)\boldsymbol{X}_0 + \exp(At)\int_0^t \exp(-As)\left(\boldsymbol{H}_s ds + \boldsymbol{K} dW_s\right).$$

9.2 Weak and Strong Solution

Theorem 9.2.1 (Existence and Uniqueness)**.** *Fix $T > 0$. Let $b\colon [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ and $\sigma\colon [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable such that*

$$\|b(t,x)\| + \|\sigma(t,x)\|_{\mathrm{F}} \le C(1+\|x\|), \quad \forall\, x\in\mathbb{R}^n,\ t\in[0,t] \tag{9.1}$$

for some constant C, and

$$\|b(t,x)-b(t,y)\| + \|\sigma(t,x)-\sigma(t,y)\|_{\mathrm{F}} \le D\,\|x-y\|\,, \quad \forall\, x,y\in\mathbb{R}^n,\ t\in[0,t] \tag{9.2}$$

for some constant D. Let $\boldsymbol{Z}$ be a random variable independent of $\mathcal{F}_\infty^W$ and $\boldsymbol{Z}\in L^2$. Then

$$d\boldsymbol{X}_t = b(t,\boldsymbol{X}_t)dt + \sigma(t,\boldsymbol{X}_t)d\boldsymbol{W}_t, \quad 0\le t\le T,\ \boldsymbol{X}_0 = \boldsymbol{Z},$$

has a unique (strong) solution $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,T]}$ that is continuous w.s.t. t with properties *(i)* $\boldsymbol{X}_t$ *is measurable w.s.t.* $\mathcal{F}_t^W \vee \sigma(\boldsymbol{Z})$ *for all* t,

(ii) $\mathbb{E}\left[\int_0^T \|\boldsymbol{X}_u\|^2\, du\right] < \infty$.

Remark 9.2.2. (1) Condition (9.1) guarantees that the existence of global solution w.s.t. t.

(2) Condition (9.2), as similar as ODE, is to make the uniqueness of the solution.

Definition 9.2.3. Given SDE,

$$d\boldsymbol{X}_t = b(t, \boldsymbol{X}_t)dt + \sigma(t, \boldsymbol{X}_t)d\boldsymbol{W}_t. \tag{9.3}$$

(1) A strong solution of (9.3) on a give probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and w.s.t. the fixed Brownian motion $\boldsymbol{W}$ and initial value $\boldsymbol{Z}$, is a stochastic process $\boldsymbol{X}$ with continuous paths and with the following properties:

   (i) $\boldsymbol{X}$ is adapted to $\mathbb{F}^W = (\mathcal{F}_t^W)_{t\geq 0}$,

   (ii) $\mathbb{P}(\boldsymbol{X}_0 = \boldsymbol{Z}) = 1$,

   (iii) $\mathbb{P}\left(\int_0^T |b_i(s, \boldsymbol{X}_s)| + |\sigma_{ij}(s, \boldsymbol{X}_s)|\, ds < \infty\right) = 1$, for all i, j,

   (iv) $\boldsymbol{X}$ satisfies the integral version

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t b(s, \boldsymbol{X}_s)ds + \int_0^t \sigma(s, \boldsymbol{X}_s)d\boldsymbol{W}_s.$$

(2) A weak solution of (9.3) is a triple $((\boldsymbol{X}, \boldsymbol{W}), (\Omega, \mathcal{F}, \mathbb{P}), \mathbb{F} = (\mathcal{F}_t))$ such that

   (i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space,

   (ii) $\mathbb{F}$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ with the usual condition,

   (iii) $\boldsymbol{X}$ is a continuous $\mathbb{F}$-adapted process,

   (iv) $\boldsymbol{W} = (\boldsymbol{W}_t, \mathcal{F}_t)$ is a Brownian motion,

   (v) $\mathbb{P}\left(\int_0^T |b_i(s, \boldsymbol{X}_s)| + |\sigma_{ij}(s, \boldsymbol{X}_s)|\, ds < \infty\right) = 1$, for all i, j,

   (vi) $\boldsymbol{X}$ satisfies the integral version

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t b(s, \boldsymbol{X}_s)ds + \int_0^t \sigma(s, \boldsymbol{X}_s)d\boldsymbol{W}_s.$$

Remark 9.2.4. It is obviously that the existence of strong solution implies that the solution is also a weak solution. But the existence of weak solution does not implies the existence of strong solution.

Example 9.2.5 (Tanaka Equation)**.** Consider the SDE

$$dX_t = \operatorname{sign}(X_t)dW_t.$$

Note that $\sigma(t, x) = \operatorname{sign}(x)$ does not satisfy the Lipschitz condition.
Claim: (9.2.5) has no strong solution.
Suppose (9.2.5) has a strong solution X. Then

$$X_t = \int_0^t \operatorname{sign}(X_s)dW_s \;\Rightarrow\; \langle X\rangle_t = t.$$

By Lévy theorem, X is a Brownian motion. On the other hand,

$$dW_t = \mathrm{sign}(X_t)dX_t \;\Rightarrow\; W_t = \int_0^t \mathrm{sign}\, X_s ds,$$

which means W is a Brownian motion w.s.t. $\mathbb{F}^X$, i.e., $\mathbb{F}^W \subset \mathbb{F}^X$. By the Tanaka equation,

$$W_t = |X_t| - L_t^X,$$

which implies that $\mathcal{F}_t^W \neq \mathcal{F}_t^X$, contradicting to X adapted to $\mathbb{F}^W$.
(9.2.5) has a weak solution. Choose $B = (B_t)_{t\geq 0}$ be a Brownian motion. Define

$$\tilde{W}_t = \int_0^t \mathrm{sign}(B_u)dB_u,$$

which is a Brownian motion. Then let $X = B$,

$$d\tilde{W}_t = \mathrm{sign}(X_t)dX_t \;\Rightarrow\; dX_t = \mathrm{sign}(X_t)d\tilde{W}_t.$$

9.3 Feynman-Kac Formula

Theorem 9.3.1 (Feynman-Kac Formula)**.** *Consider SDE*

$$dX_t = \beta(t, X_t)dt + \sigma(t, X_t)dW_t.$$

Let f be a Borel-measurable function. Fix $T > 0$ and $t \in [0, T]$. Define

$$g(t, x) = \mathbb{E}\left[f(X_T) \mid X_t = x\right] = \mathbb{E}^{t,x}[f(X_T)].$$

Assume $g(t, x) < \infty$. Then $g(t, x)$ satisfies PDE

$$\frac{\partial}{\partial t}g + \beta\frac{\partial}{\partial x}g + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}g = 0$$

with terminal $g(T, x) = f(x)$.

Remark 9.3.2. Note that $(g(t, X_t))_{0\leq t\leq T}$ is a martingale, because by Itô formula

$$\begin{aligned}
dg(t, X_t) &= g_t(t, X_t)dt + g_x(t, X_t)dX_t + \frac{1}{2}g_{xx}(t, X_t)(dX_t)^2 \\
&= g_t dt + g_x(\beta dt + \sigma dW_t) + \frac{1}{2}g_{xx}\sigma^2 dt \\
&= g_x\sigma dW_t + \left(g_t + \beta g_x + \frac{1}{2}\sigma^2 g_{xx}\right)dt \\
&= g_x\sigma dW_t.
\end{aligned}$$

Theorem 9.3.3 (Discounted Feynman-Kac Formula)**.** *Consider SDE*

$$dX_t = \beta(t, X_t)dt + \sigma(t, X_t)dW_t.$$

Let f be a Borel-measurable function and r be a constant. Fix $T > 0$ and $t \in [0, T]$. Define

$$h(t, x) = \mathbb{E}^{t,x}\left[e^{-r(T-t)}f(X_T)\right].$$

Then h satisfies

$$h_t(t, x) + \beta(t, x)h_x(t, x) + \frac{1}{2}\sigma(t, x)h_{xx}(t, x) = rh(t, x)$$

with terminal $h(T, x) = f(x)$.

Chapter 10

Diffusion Process

Consider SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{10.1}$$

where $X_t \in \mathbb{R}^n$, $b\colon [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma\colon [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ and $B = (B_t)_{t\geq 0}$ be a m-dimensional Brownian motion. Any process $X = (X_t)_t$ satisfies (10.1) is called a (Itô) diffusion.

10.1 Markov Property

We mainly consider the time-homogeneous case, i.e.,

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq s,\ X_s = x$$

where b, σ is time-independent. To guarantee the existence and uniqueness of solution, we only require the Lipschitz condition,

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|, \quad \forall\ x, y \in \mathbb{R}^n.$$

Then denote the unique solution $X_t = X_t^{s,x}$ for $t \geq 0$, and for $s = 0$, $X_t = X_t^x$. The time-homogeneity means that $\left\{X_{s+h}^{s,x}\right\}_{h\geq 0}$ and $\left\{X_h^{0,x}\right\}_{h\geq 0}$ have the same diffusion (by the uniqueness of weak solution). So let $\mathbb{Q}^x$ be the law of $X^x = (X_t^x)_{t\geq 0}$ on $(\mathbb{R}^n)^{[0,\infty)}$ and the $\mathbb{E}_{\mathbb{Q}^x} = \mathbb{E}_x$. Moreover, $\mathbb{F}^B = (\mathcal{F}_t^B)_{t\geq 0}$ be the natural filtration of B and $\mathbb{F}^X$ be the natural filtration of X. Note that X is $\mathbb{F}^B$-adapted and so $\mathcal{F}_t^X \subset \mathcal{F}_t^B$.

Theorem 10.1.1 (Markov Property)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a bounded Borel function. Then for any $t, s \geq 0$, we have*

$$\mathbb{E}_x\left[f(X_{t+s}) \mid \mathcal{F}_t^B\right] = \mathbb{E}_{X_t}[f(X_s)].$$

Theorem 10.1.2 (Strong Markov Property)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a bounded Borel function and $\tau < \infty$ be a stopping time w.s.t. $\mathbb{F}^B$. Then*

$$\mathbb{E}_x\left[f(X_{\tau+s}) \mid \mathcal{F}_\tau^B\right] = \mathbb{E}_{X_\tau}[f(X_s)].$$

10.2 Generator

Let X_t be the Itô diffusion

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

Since X is a Markov process, it has the corresponding transition semigroup $(Q_t)_{t\geq 0}$, i.e.,

$$Q_t g(x) = \mathbb{E}_x[g(X_t)] = \mathbb{E}[g(X_t) \mid X_0 = x].$$

And the generator

$$Lg(x) = \lim_{t \downarrow 0} \frac{Q_t g(x) - g(x)}{t}.$$

Lemma 10.2.1. *Let g be a lower bounded, measurable function on $\mathbb{R}^n$.*

(1) If g is lower semi-continuous, then $Q_t g$ is lower semi-continuous for all $t \geq 0$.

(2) If g is bounded and continuous, then $Q_t g$ is continuous. In other words, any Itô diffusion X is Feller-continuous.

Note that by Itô formula, for any $f \in C_c^2(\mathbb{R}^n)$

$$\begin{aligned}
f(X_t) &= f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i} dX_s^i + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j} d[X^i, X^j]_s \\
&= f(X_0) + \sum_{i=1}^n \int_0^t b^i \frac{\partial f}{\partial x_i} ds + \sum_{i,j=1}^n \int_0^t \sigma_{ij} \frac{\partial f}{\partial x_i} dB_s^j \\
&\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t (\sigma\sigma^\top)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} dt \\
&= f(X_0) + \int_0^t \left(\sum_{i=1}^n b^i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n (\sigma\sigma^\top)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right) ds + \sum_{i,j=1}^n \int_0^t \sigma_{ij} \frac{\partial f}{\partial x_i} dB_s^j,
\end{aligned}$$

because

$$d[X^i, X^j]_s = \left(\sum_k \sigma_{ik} dB^k \right) \left(\sum_\ell \sigma_{j\ell} dB^\ell \right) = (\sigma\sigma^\top)_{ij} dt.$$

This implies the following theorem.

Theorem 10.2.2. *If $f \in C_c^2(\mathbb{R}^n)$, then $f \in \mathcal{D}(L)$ and*

$$Lf(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} \left(\sigma\sigma^\top\right)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j},$$

where L is the generator of Markov process X.

For $f \in C_c^2(\mathbb{R}^n)$, we have

$$f(X_t) = f(X_0) + \int_0^t Lf(X_s) ds + \int_0^t \nabla f(X_s)^\top \sigma(X_s) dB_s.$$

So

$$M_t = f(X_t) - \int_0^t Lf(X_s) ds$$

is a martingale, which is a particular case of Theorem 7.3.11. Note that here X_0 is not fixed to a point. Moreover, if $f \in C^2$, then we just know $(M_t)_{t \geq 0}$ is a local martingale.

Remark 10.2.3. A Feller semigroup $(Q_t)_{t \geq 0}$ is called a Feller-Dynkin diffusion semigroup if the domain $\mathcal{D}(L)$ of its generator L contains $C_c^2(\mathbb{R}^n)$. A A continuous Markov process $X = (X_t)_{t \geq 0}$ is said to be a Feller-Dynkin diffusion process if its associated semigroup is a Feller-Dynkin diffusion semigroup. So by above theorem, we know an Itô diffusion is a Feller-Dynkin diffusion process.

Theorem 10.2.4 (Dynkin's formula)**.** *If $f \in C_c^2(\mathbb{R}^n)$ and τ is a stopping time with $\mathbb{E}_x[\tau] < \infty$, then*

$$\mathbb{E}_x\left[f(X_\tau)\right] = f(x) + \mathbb{E}_x\left[\int_0^\tau Lf(X_s)ds\right].$$

Example 10.2.5 (Bessel Process)**.** Let B be a m-dimensional standard Brownian motion. Consider

$$R_t = \|B_t\| = \sqrt{(B_t^1)^2 + \cdots + (B_t^m)^2},$$

the Bessel process. Then we know

$$dR_t = \frac{n-1}{2R_t} + \sum_{i=1}^m \frac{B^i}{R_t} dB^i.$$

Let

$$\tilde{B}_t = \sum_{i=1}^m \int_0^t \frac{B_s^i}{\|B_s\|} dB_s^i.$$

Then by Lévy's theorem, $\tilde{B} = (\tilde{B}_t)_{t\geq 0}$ is a 1-dimensional Brownian motion. Therefore,

$$dR_t = \frac{n-1}{2R_t} dt + d\tilde{B}_t.$$

So by the uniqueness of weak solution, $R = (R_t)_{t\geq 0}$ is also an Itô diffusion with generator

$$Lf(x) = \frac{1}{2} f''(x) + \frac{n-1}{2x} f'(x).$$

Example 10.2.6. Let $U \in C^1(\mathbb{R}^n)$ and

$$L = \Delta \cdot + \langle \nabla U, \nabla \cdot \rangle$$

on $C_c^\infty(\mathbb{R}^n)$. Then by the divergence theorem,

$$\mu(dx) = e^{U(x)} dx$$

is symmetric for L. Moreover, L is essentially self-adjoint on $L^2(\mathbb{R}^n, \mu)$.

Chapter 11

Symmetric Markov Operator

In this chapter, let E be a Polish space that is a separable complete metric space and let E be equipped with the Borel σ-field $\mathcal{F}$. Then the measure decomposition theorem implies that for any probability measure μ on the product σ-field $\mathcal{F} \otimes \mathcal{F}$ on $E \times E$ with $\mu_1 = \pi_1^{\#}\mu$, the first projection, then

$$\mu(dx, dy) = k(x, dy)\mu_1(dx)$$

for some probability transition kernel $k\colon E \times \mathcal{F} \to [0, 1]$. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space.

Remark 11.0.1. Moreover, because of the existence of kernels, by Ionescu–Tulcea theorem, for any probability measure μ on E^n, there are kernels k_i from E^{i-1} to E such that

$$\mu(dx_1, dx_2, \cdots, dx_n) = \mu_1(dx_1)k_2(x_1, dx_2)k_3(x_1, x_2, dx_3) \cdots k_n(x_1, \cdots, x_{n-1}, dx_n).$$

For now on any measure μ is assumed to be σ-finite.

11.1 Markov Operator

Definition 11.1.1. A Markov operator P on $(E, \mathcal{F})$ is a linear operator $P\colon \mathcal{B}_b(E) \to \mathcal{B}_b(E)$ such that

(1) (*Mass conversation*) $P\mathbb{1} = \mathbb{1}$ for constant function $\mathbb{1}(x) \equiv 1$,

(2) (Positivity preserving) for $f \geq 0$, $Pf \geq 0$.

Remark 11.1.2. For $0 \leq f \leq 1$,

$$P(\mathbb{1} - f) \geq 0 \;\Rightarrow\; 0 \leq Pf \leq P\mathbb{1} \leq \mathbb{1}.$$

Therefore, $\|Pf\|_\infty \leq \|f\|_\infty$ for all $f \in \mathcal{B}_b(E)$.

Proposition 11.1.3 (Jensen's inequality)**.** *For any convex $\phi\colon \mathbb{R} \to \mathbb{R}$ and any $f \in \mathcal{B}_b(E)$, if P is a Markov operator, then*

$$P(\phi(f)) \geq \phi(Pf)$$

Proof. Because ϕ is convex, for any $b \in \mathbb{R}$, there is $a = a(b)$ such that

$$\phi(c) \geq \phi(b) + a(b)(c - b), \quad \forall\, c \in \mathbb{R}$$

For any $x \in E$, let $c = f(x)$. We have

$$\phi(f(x)) \geq \phi(b) + a(b)(f(x) - b) \quad \Rightarrow \quad \phi(f) \geq \phi(b) + a(b)(f - b)$$

By the positivity and mass properties of P, we have

$$P(\phi(f)) \geq \phi(b) + a(b)(Pf - b)$$

So for any $x \in E$,

$$P(\phi(f))(x) \geq \phi(b) + a(b)(Pf(x) - b)$$

Then let $b = Pf(x)$, we get

$$P(\phi(f))(x) \geq \phi(Pf(x))$$

which is true for any $x \in E$. $\square$

Definition 11.1.4 (Invariant Measure). A measure μ on $(E, \mathcal{F})$ is called invariant for a Markov operator P if

$$\int_E Pf d\mu = \int_E f d\mu,$$

for all $f \in \mathcal{B}_b(E)$.

Remark 11.1.5. When $f \in \mathcal{B}_b(E)$ is 0 μ-a.e., $Pf = 0$ μ-a.e.. Therefore, P can be extended on $L^\infty(\mu)$. Moreover, μ is invariant for P if

$$\int_E Pf d\mu = \int_E f d\mu, \quad \forall\ f \in L^1(\mu) \cap L^\infty(\mu).$$

Note that for $1 \leq p < \infty$, $L^1(\mu) \cap L^\infty(\mu) \subset L^\infty(\mu)$ that is because

$$\|f\|_p^p = \int |f|^p d\mu \leq \|f\|_\infty^{p-1} \int |f| d\mu = \|f\|_\infty^{p-1} \|f\|_1.$$

So by Jensen's inequality for $\phi(x) = |x|^p$ $(1 \leq p < \infty)$,

$$\int |Pf|^p d\mu \leq \int P\left(|f|^p\right) d\mu = \int |f|^p d\mu, \quad \forall\ f \in L^1(\mu) \cap L^\infty(\mu).$$

Lemma 11.1.6. *For any* $1 \leq p < \infty$,

$$L^1(\mu) \cap L^\infty(\mu) \subset L^p(\mu)$$

is dense.

Proof. Let $f \in L^p$ and $\varepsilon > 0$.

- Step 1: For any $n \in \mathbb{N}$, let

  $$g_n(x) := \max\left(-n, \min(f(x), n)\right) \in [-n, n].$$

  Therefore, $g_n \in L^\infty$ and

  $$\|f - g_n\|_p^p = \int_{|f|>n} ||f| - n|^p d\mu \leq \int_{|f|>n} |f|^p d\mu \underset{n\to\infty}{\longrightarrow} 0.$$

  So let n be sufficiently large such that $\|f - g_n\|_p \leq \varepsilon/2$.

- By σ-finiteness, choose $E_k \uparrow E$ with $\mu(E_k) < \infty$ and put $h_k = g_n \mathbb{I}_{E_k}$. So $h_k \in L^1 \cap L^\infty$. Because $\mathbb{I}_{E_k^x} \to 0$ as $k \to \infty$, by DCT,

  $$\|g_n - h_k\|_p^p = \int_{E_k^c} |g_n|^p\, d\mu \leq \int_{E_k^c} |f|^p d\mu \underset{k\to\infty}{\longrightarrow} 0.$$

  Therefore,

  $$\|f - h_k\|_p \leq \|f - g_n\|_p + \|g_n - h_k\|_p < \varepsilon.$$ $\square$

Then because

$$\|Pf\|_p \le \|f\|_p, \quad \forall\, f \in L^1(\mu) \cap L^\infty(\mu),$$

and the density,

$$P \colon L^p(\mu) \to L^p(\mu)$$

for all $1 \le p \le \infty$. Note that this definition should fix an invariant μ.

Definition 11.1.7 (Reversible Measure)**.** A measure μ is called reversible for a Markov operator P if

$$\int fPgd\mu = \int gPfd\mu, \quad \forall\, f, g \in L^2(\mu).$$

Remark 11.1.8. It is obviously that if μ is reversible, then it is invariant, because it can choose $g_n \in L^2(\mu)$ such that $g_n \uparrow \mathbb{1}$ by the σ-finiteness of μ.

Definition 11.1.9. A symmetric Markov semigroup on $(E, \mathcal{F}, \mu)$ is a family of $(P_t)_{t\ge 0}$ of Markov operators such that

(i) (*Initial Condition*) $P_0 f = f$ for all $f \in L^\infty$;

(ii) (*Semigroup*) for every $t, s \ge 0$, $P_t P_s = P_{t+s}$;

(iii) (*Symmetry*) for every $t \ge 0$, μ is reversible for P_t;

(iv) (*Strong Continuity*) for all $f \in L^2(\mu)$, $P_t f \to f$ in $L^2(\mu)$ as $t \to 0$.

Remark 11.1.10. Note that strong continuity implies that $P_t \to P_{t_0}$ in the strong operator topology on $L^2(\mu)$ as $t \to t_0$ with the help of the initial condition and the semigroup property.

Theorem 11.1.11 (Kernel Representation)**.** *Let P be a Markov operator on $(E, \mathcal{F})$ that is continuous on $L^1(\nu)$. Then there exists a probability kernel p on $(E, \mathcal{F})$ such that for every $f \in L^\infty(\nu)$ and ν-a.e. $x \in E$,*

$$Pf(x) = \int_E f(y)p(x, dy).$$

11.2 Generator

For a given symmetric Markov semigroup $(P_t)_{t\ge 0}$ on $(E, \mathcal{F}, \mu)$, we can similarly define the generator but the domain is different,

$$\mathcal{D}(L) := \left\{ f \in L^2(\mu) \colon \lim_{t\to 0} \frac{Q_t f - f}{t} \text{ exists in } L^2(\mu) \right\}.$$

And we also define $\mathcal{D}_p(L)$ for considering the convergence in $L^p(\mu)$. Except for the domain, some properties are as same as the generator of a Feller semigroup, like, $\mathcal{D}(L) \subset L^2(\mu)$ dense, and $L(P_t f) = P_t(Lf)$. So

$$P_t f - f = \int_0^t P_s(Lf)ds = \int_0^t L(P_s f)ds.$$

Moreover, by the symmetry of $(P_t)_{t\ge 0}$, for any $f, g \in \mathcal{D}(L)$

$$\int_E fLgd\mu = \int_E gLfd\mu,$$

and for every $f \in \mathcal{D}_1(L)$,

$$\int Lfd\mu = 0.$$

Assume there exists an algebra $\mathcal{A} \subset \mathcal{D}(L)$, for example, $\mathcal{A} = C_c^\infty(\mathbb{R}^n)$.

Definition 11.2.1 (Carré du Champ). The carré du champ associated to L is the bilinear form Γ on $\mathcal{A} \times \mathcal{A}$ defined by

$$\Gamma(f,g) = \frac{1}{2}(L(fg) - fLg - gLf).$$

$\Gamma(f) = \Gamma(f,f)$.

Note that

$$\frac{d}{dt}(P_t f)^2 = 2f\frac{d}{dt}P_t f = 2fLP_t f,$$

and by Jensen's inequality,

$$L\left(f^2\right) = \lim_{t\to 0}\frac{P_t\left(f^2\right) - f^2}{t} \geq \lim_{t\to 0}\frac{(P_t f)^2 - f^2}{t} = \left.\frac{d\left(P_t f\right)^2}{dt}\right|_{t=0} \leq 2fLf,$$

which implies that $\Gamma(f) \geq 0$. Then by the Cauchy-Schwartz inequality,

$$\Gamma(f,g)^2 \leq \sqrt{\Gamma(f)\Gamma(g)}.$$

Proposition 11.2.2. *Let $(P_t)_{t\geq 0}$ on $(E, \mathcal{F}, \mu)$ be a given symmetric Markov semigroup and L be its generator. Then L is a self-adjoint operator on L^μ and so it is closed.*

Moreover, because

$$0 \leq \int_E \Gamma(f)d\mu = -\int_E fLfd\mu,$$

L is non-positive definite.

Construct semigroup from generator L. Let's assume

$$L = \sum_{i,j=1}^n \sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x)\frac{\partial}{\partial x_i},$$

where b_i and σ_{ij} are continuous functions and $\sigma = (\sigma_{ij}(x)) \in \mathbb{R}^{n\times n}$ is symmetric and nonnegative. $\mathcal{D}(L) = C_c^\infty(\mathbb{R}^n)$. Moreover, of σ is invertible, L is called an elliptic diffusion operator. A Borel measure μ is called symmetric for L if for any $f, g \in C_c^\infty(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} gLfd\mu = \int_{\mathbb{R}^n} fLgd\mu.$$

In the following, let's fix a measure μ symmetric for L.
Note that because $\mathcal{D}(L) = C_c^\infty(\mathbb{R}^n) \subset L^2(\mathbb{R}^n, \mu)$ is dense, L is a non-positive symmetric operator that is densely defined on $L^2(\mathbb{R}^n, \mu)$. But it is not self-adjoint. However, it can be extended to a self-adjoint operator.

Theorem 11.2.3 (Friedrichs Extension). *On the Hilbert space $L^2(\mathbb{R}^n, \mu)$, for L defined above, there exists a densely defined non-positive self-adjoint extension of L.*

In fact, if L is essentially self-adjoint, then the Friedrichs extension is the closed operator $\overline{L}$. In such case,

$$\ker(-L^* + \lambda I) = \{0\}, \ \lambda > 0.$$

It means

$$-Lf + \lambda f = 0 \ \Rightarrow \ f = 0,$$

where Lf, the differential in the sense of distribution.

Therefore, in the following, we assume L is essentially self-adjoint and replace $\overline{L}$ by L. Then L is self-adjoint on $L^2(\mathbb{R}^n, \mu)$. So we can define

$$P_t = e^{tL} = \int_{\mathbb{R}} e^{t\lambda} dE_L(\lambda) = \int_0^\infty e^{-t\lambda} dE_L(\lambda), \forall\, t \geq 0,$$

where E_L is the spectral measure associated with L. The $P_t\colon L^2(\mathbb{R}^n, \mu) \to L^2(\mathbb{R}^n, \mu)$ is a bounded operator. Note that

(i) $P_t P_s = P_{t+s}$ for all $t, s \geq 0$.

(ii) for all $f \in L^2$,
$$\|P_t f\|_2 \leq \|f\|_2 .$$

(iii) for all $f \in L^2$, $t \mapsto P_t f$ is continuous in $L^2(\mu, \mathbb{R}^n)$.

(iv) for all $f, g \in L^2$,
$$\int_{\mathbb{R}^n} f P_t g d\mu = \int_{\mathbb{R}^n} g P_t f d\mu,$$
i.e., μ is reversible for P_t.

(v) for all $f \in L^2$,
$$\lim_{t \to 0} \|P_t f - f\|_2 = 0$$

(vi) for all any $f \in \mathcal{D}(L)$,
$$\lim_{t \to 0} \left\| \frac{P_t f - f}{t} - Lf \right\|_2 = 0.$$

(vii) if $\mathbb{1} \in \mathcal{D}(L)$ and $L\mathbb{1} = 0$, then $P_t \mathbb{1} = \mathbb{1}$.

11.3 Compact Markov Operators

Definition 11.3.1 (Diffusion Carré du Champ)**.** Let $\mathcal{A} \subset \mathbb{R}^E$ be an algebra such that for any $k \in \mathbb{N}$, any $f_1, \cdots, f_k \in \mathcal{A}$, and any $\Psi \in C^\infty(\mathbb{R}^k)$, $\Psi(f_1, \cdots, f_k) \in \mathcal{A}$. We say a bilinear form $\Gamma\colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is called a diffusion carré du champ if for any Ψ and f_i as above,

$$\Gamma\left(\Psi\left(f_1, \ldots, f_k\right), g\right) = \sum_{i=1}^k \partial_i \Psi\left(f_1, \ldots, f_k\right) \Gamma\left(f_i, g\right).$$

Consider a symmetric Markov semigroup with generator L and the corresponding carré du champ

$$\Gamma(f, g) = \frac{1}{2}(L(fg) - fLg - gLf).$$

If Γ is a diffusion carré du champ, then

$$L\Psi\left(f_1, \ldots, f_k\right) = \sum_{i=1}^k \partial_i \Psi\left(f_1, \ldots, f_k\right) Lf_i + \sum_{i,j=1}^k \partial_i \partial_j \Psi\left(f_1, \ldots, f_k\right) \Gamma\left(f_i, f_j\right). \tag{11.1}$$

In particular, for $k = 1$,

$$\begin{aligned} \Gamma(\psi(f), g) &= \psi'(f)\Gamma(f, g) \\ L\psi(f) &= \psi'(f)Lf + \psi''(f)\Gamma(f). \end{aligned}$$

Definition 11.3.2 (Diffusion Semigroup)**.** An operator L satisfying (11.1) is called a diffusion generator. A symmetric Markov semigroup whose generator is a diffusion generator is called a diffusion semigroup.

Definition 11.3.3 (Dirichlet Form)**.** A bilinear form $\mathcal{E}\colon \mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E}) \to \mathbb{R}$ is called a Dirichlet form if

(i) $\mathcal{D}(\mathcal{E}) \subset L^2(\mu)$ dense for some μ,

(ii) $\mathcal{E}(f,g) = \mathcal{E}(g,f)$ for $f,g \in \mathcal{D}(\mathcal{E})$,

(iii) $\mathcal{E}(f) = \mathcal{E}(f,f) \geq 0$ for $f \in \mathcal{D}(\mathcal{E})$,

(iv) $\mathcal{D}(\mathcal{E})$ is complete w.s.t.
$$\langle f,g\rangle_{\mathcal{E}} := \int_E fg d\mu + \mathcal{E}(f,g),$$

(v) for any $f \in \mathcal{D}(\mathcal{E})$, $0 \vee f \wedge 1 \in \mathcal{D}(\mathcal{E})$ and
$$\mathcal{E}(0 \vee f \wedge 1) \leq \mathcal{E}(f).$$

Remark 11.3.4. Note that for any symmetric, non-negative bilinear form $\mathcal{E}$ defined on some dense $D \subset L^2(\mu)$, if $\mathcal{E}$ satisfies that for any $f_n \to 0$ in $L^2(\mu)$ and f_n Cauchy w.s.t. $\langle\cdot,\cdot\rangle_{\mathcal{E}}$, $\mathcal{E}(f_n) \to 0$, then D can be extended to the closure of D w.s.t. $\|\cdot\|_2 + \mathcal{E}(\cdot)$, and such $\mathcal{E}$ is called closable.

If Γ is a diffusion carré du champ on an algebra $\mathcal{A} \subset L^2(\mu)$ dense and $\Gamma(f,f) \geq 0$ for all $f \in \mathcal{A}$, then let
$$\mathcal{E}(f,g) := \int_E \Gamma(f,g) d\mu$$
and taking $\mathcal{D}(\mathcal{E})$ be the closure of $\mathcal{A}$ w.s.t. $\langle\cdot,\cdot\rangle_{\mathcal{E}}$. It can prove that such $\mathcal{E}$ is a Dirichlet form. Moreover, by the symmetric and positivity of $\mathcal{E}$, Riesz representation theorem implies that we can define a a non-positive, symmetric operator L by
$$\int gLf d\mu = -\mathcal{E}(f,g)$$
on the domain
$$\mathcal{D}(L) = \{f \in \mathcal{D}(\mathcal{E}) : \exists C \text{ such that } \mathcal{E}(f,g) \leq C\|g\|_2 \text{ for all } g \in \mathcal{D}(\mathcal{E})\}.$$
Moreover, it can be extended to a self-adjoint operator L by using Friedrichs extension.

Definition 11.3.5 (Compact Markov Diffusion Triple)**.** Let $(E,\mathcal{F},\mu)$ be a polished measure space and μ be a probability measure. For $\mathcal{A} \subset L^2(\mu)$, let
$$\Gamma\colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$$
be a symmetric bilinear form. We say (E,μ,Γ) a compact Markov diffusion triple the followings are satisfied.

(a) $\mathcal{A}$ is dense in $L^2(\mu)$,

(b) $\mathcal{A}$ is an algebra closed under composition with smooth functions,

(c) $\Gamma(f) = \Gamma(f,f) \geq 0$ for all $f \in \mathcal{A}$,

(d) Γ is a diffusion carré du champ,

(e) $\Gamma(f) = 0$ implies that f is a constant,

and let $\mathcal{E}(f, g) = \int_E \Gamma(f, g) d\mu$ for all $f, g \in \mathcal{A}$, which satisfies

(f) for every $f \in \mathcal{A}$, there exist a $C > 0$ such that $\mathcal{E}(f, g) \leq C \|g\|_2$ for all $g \in \mathcal{A}$.

It follows that $\mathcal{E}$ can be extended to a Dirichlet form. Let L be the self-adjoint operator defined on $\mathcal{D}(L)$ such that

$$\int gLf d\mu = -\mathcal{E}(f, g).$$

Note that $\mathcal{A} \subset \mathcal{D}(L)$. Let $P_t = e^{tL}$ called the semigroup be assumed that

(g) $L\mathcal{A} \subset \mathcal{A}$,

(h) $P_t\mathcal{A} \subset \mathcal{A}$.

Proposition 11.3.6. *Let $(E, \mathcal{F}, \mu)$ be a compact Markov diffusion triple and P_t be its semigroup.*

(1) P_t is a symmetric Markov semigroup for μ.

(2) For any $f \in L^2(\mu)$,

$$\lim_{t\to\infty} P_t f = \int_E f d\mu$$

in L^2, which is called the ergodic property.

Curvature.

Definition 11.3.7. Given a compact Markov diffusion triple $(E, \mathcal{F}, \mu)$. For any $f, g \in \mathcal{A}$,

$$\Gamma_2(f, g) = \frac{1}{2}\left(L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf)\right),$$

and $\Gamma_2(f) = \Gamma_2(f, f)$.

Definition 11.3.8 (Curvature Dimension)**.** For $\rho \in \mathbb{R}$ and $n \in [1, \infty]$, a compact Markov diffusion triple $(E, \mathcal{F}, \mu)$ is to said to satisfy the curvature-dimension condition $\mathrm{CD}(\rho, n)$ if

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{1}{n}(Lf)^2.$$

11.4 Poincaré Inequality

Proposition 11.4.1. *Let P_t be the semigroup of a compact Markov triple. TFAE.*

(1) $\mathrm{CD}(\rho, \infty)$ holds for some $\rho \in \mathbb{R}$.

(2) For every $f \in \mathcal{A}$ and $t \geq 0$,

$$\Gamma(P_t f) \leq e^{-2\rho t} P_t \Gamma(f).$$

(3) For every $f \in \mathcal{A}$ and $t \geq 0$,

$$P_t(f^2) - (P_t f)^2 \leq \frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma(f).$$

(4) For every $f \in \mathcal{A}$ and $t \geq 0$,

$$P_t(f^2) - (P_t f)^2 \geq \frac{e^{2\rho t} - 1}{\rho} \Gamma(P_t f).$$

For the last two conditions, if $\rho = 0$, then the coefficients in RHS can be taken as $2t$.

Proof. (1) $\Rightarrow$ (2): for $f \in \mathcal{A}$, let

$$\Lambda(s) = e^{-2\rho} P_s \Gamma(P_{t-s} f).$$

Then by chain rule,

$$\Lambda'(s) = 2e^{-2\rho s} P_s \left(\Gamma_2\left(P_{t-s} f\right) - \rho \Gamma\left(P_{t-s} f\right)\right) \geq 0,$$

because of $\mathrm{CD}(\rho, \infty)$. Therefore, $\Lambda(t) \geq \Lambda(0)$.
(2) $\Rightarrow$ (3): Let

$$\Lambda(s) = P_s \left(P_{t-s} f\right)^2.$$

So $\Lambda'(s) = 2P_s \Gamma\left(P_{t-s} f\right)$ and

$$\begin{aligned}
\Lambda(t) - \Lambda(0) &= 2 \int_0^t P_s \Gamma\left(P_{t-s} f\right) ds \\
&\leq 2 \int_0^t e^{-2\rho(t-s)} P_t \Gamma(f) ds \\
&= \frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma(f).
\end{aligned}$$

(2) $\Rightarrow$ (4): Similarly, as above, by using $P_s \Gamma\left(P_{t-s} f\right) \geq e^{2\rho s} \Gamma\left(P_t f\right)$,

$$\Lambda(t) - \Lambda(0) \geq 2 \int_0^t e^{2\rho s} \Gamma\left(P_t f\right) ds = \frac{e^{2\rho t} - 1}{\rho} \Gamma\left(P_t f\right).$$

(3) $\Rightarrow$ (1): Note that for any $h \in \mathcal{A}$,

$$P_t h = h + tLh + \frac{t^2}{2} L^2 h + o\left(t^2\right), \quad t \to 0.$$

Therefore choosing $h = f$ and $h = f^2$, we have

$$\begin{aligned}
P_t\left(f^2\right) - \left(P_t f\right)^2 &= tL\left(f^2\right) + \frac{t^2}{2} L^2\left(f^2\right) - 2tfLf - t^2(Lf)^2 - t^2 f L^2 f + o\left(t^2\right) \\
&= 2t\Gamma(f) + \frac{t^2}{2} L^2\left(f^2\right) - t^2(Lf)^2 - t^2 f L^2 f + o\left(t^2\right).
\end{aligned}$$

On the other hand,

$$\frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma(f) = 2t\Gamma(f) - 2\rho t^2 \Gamma(f) + 2t^2 L\Gamma(f) + o\left(t^2\right).$$

Therefore, by (3),

$$\frac{1}{2} L^2\left(f^2\right) - (Lf)^2 - f L^2 f + o(1) \leq -2\rho \Gamma(f) + 2L\Gamma(f) + o(1).$$

As $t \to 0$, we have

$$\frac{1}{2} L^2\left(f^2\right) - (Lf)^2 - f L^2 f \leq -2\rho \Gamma(f) + 2L\Gamma(f).$$

Then by arranging,

$$L\Gamma(f) - 2\Gamma(f, Lf) \geq 2\rho \Gamma(f).$$

(4) $\Rightarrow$ (1): It is similarly as above. $\square$

For (3) in above proposition, which is called local Poincaré inequality, if $\rho > 0$, by the ergodic property, as $t \to \infty$,

$$\int f^2 d\mu - \left(\int f d\mu\right)^2 \leq \frac{1}{\rho} \int \Gamma(f) d\mu,$$

which is called a Poincaré inequality.

Definition 11.4.2 (Poincaré inequality)**.** Let μ be a probability measure and $\mathcal{E}$ be a Dirichlet form on $L^2(\mu)$. We say that μ and $\mathcal{E}$ satisfy a Poincaré inequality with constant C (PI(C)) if

$$\int_E f^2 d\mu - \left(\int_E f d\mu\right)^2 \leq C\mathcal{E}(f),$$

for any $f \in \mathcal{D}(\mathcal{E})$. The Poincaré constant of μ and $\mathcal{E}$ is the smallest C such that above inequality holds for all $f \in \mathcal{D}(\mathcal{E})$.

Remark 11.4.3. When considering a compact Markov triple, because $\mathcal{A} \subset \mathcal{D}(\mathcal{E})$ is dense, it suffices to check PI on $\mathcal{A}$. Moreover, if a compact Markov triple satisfies CD(ρ, ∞), it satisfies PI$(1/\rho)$.

Corollary 11.4.4. *The compact Markov triple (E, μ, Γ) satisfies* CD(ρ, ∞) *if and only if for any $t \geq 0$ and $x \in E$ μ-a.e., the measure $p_t(x, \cdot)$ satisfies* PI *with constant $(1 - e^{-2\rho t})/\rho$.*

Proposition 11.4.5 (Spectral Gap)**.** *If the compact Markov triple (E, μ, Γ) satisfies* PI(C) *for some constant C, then the spectrum of L*

$$\sigma(L) \subset (-\infty, -\frac{1}{C}] \cup \{0\}.$$

Proof. Let $\lambda \in \sigma(L)$ such that $\lambda \neq 0$. Because L is self-adjoint, i.e., $\sigma(L) = \sigma_{ap}(L)$, there exists $f_n \in \mathcal{D}(L)$ such that $\|f\|_2 = 1$ and

$$\|Lf_n - \lambda f_n\|_2 \to 0, \quad n \to \infty.$$

Note that because μ is a probability measure, $\|Lf_n - \lambda f_n\|_1 \to 0$. It follows that $\int f_n d\mu \to 0$ by $\int Lf_n d\mu = 0$ for all n. Then PI implies that

$$\int_E f_n^2 d\mu - \left(\int_E f_n d\mu\right)^2 \leq C \int_E \Gamma(f_n) d\mu = -C \int_E f_n L f_n d\mu.$$

As $n \to \infty$,

$$\lambda = \int_E f_n L f_n d\mu \leq -\frac{1}{C}.$$

□

PI **under** CD(ρ, n)**.**

Lemma 11.4.6. *Suppose the compact Markov triple (E, μ, Γ) satisfies* CD(ρ, ∞) *with some $\rho > 0$. It satisfies* PI(C) *for some $C > 0$ if and only if*

$$\int_E \Gamma(f) d\mu \leq C \int_E (Lf)^2 d\mu, \quad \forall\ f \in \mathcal{D}(L).$$

Proof. $\Rightarrow$: Let

$$\Lambda(t) = \int_E (P_t f)^2 d\mu.$$

Then

$$\Lambda'(t) = -2\int_E \Gamma(P_t f)\, d\mu, \quad \Lambda''(t) = 4\int_E (LP_t f)^2\, d\mu.$$

Because it satisfies $\mathrm{CD}(\rho,\infty)$ with some $\rho > 0$, by

$$\Gamma(P_t f) \le e^{-2\rho t} P_t\Gamma(f) \le P_t\Gamma(f) \le \Gamma(f)$$

Then by DCT, $\lim_{t\to\infty} \Lambda'(t)$ exists. And by ergodicity,

$$\lim_{t\to\infty} \Lambda(t) = \int_E f d\mu,$$

$\lim_{t\to\infty} \Lambda'(t) = 0$. By assumption,

$$\Lambda''(t) \ge -\frac{2}{C}\Lambda'(t).$$

Therefore,

$$\begin{aligned}
\int f^2 d\mu - \left(\int f d\mu\right)^2 &= -\int_0^\infty \Lambda'(t)dt \\
&\le \frac{C}{2}\int_0^\infty \Lambda''(t)dt \\
&= -\frac{C}{2}\Lambda'(0) \\
&= C\int_E \Gamma(f)d\mu.
\end{aligned}$$

$\Leftarrow$: Choosing $f \in \mathcal{D}(L)$ with mean 0 (otherwise, let $f - \int f d\mu$ and note that $\Gamma(c, g) = 0$ for any constant c and function g). By Cauchy-Schwartz inequality,

$$\begin{aligned}
\int_E \Gamma(f)d\mu &= \int_E f(-Lf)d\mu \\
&\le \sqrt{\int_E f^2 d\mu \int_E (Lf)^2 d\mu} \\
&\le \sqrt{C\int_E \Gamma(f)d\mu \int_E (Lf)^2 d\mu}.
\end{aligned}$$

□

Theorem 11.4.7. *Let (E, μ, Γ) be a compact Markov triple. If it satisfies $\mathrm{CD}(\rho, n)$ for some $\rho > 0$ and $n > 1$, then μ satisfies $\mathrm{PI}(C)$ with $C = \frac{n-1}{\rho n}$.*

Proof. Because of $\mathrm{CD}(\rho, n)$,

$$\int_E \Gamma_2(f)d\mu \ge \rho\int_E \Gamma(f)d\mu + \frac{1}{n}\int_E (Lf)^2 d\mu.$$

Because $\int Lh d\mu = 0$,

$$\begin{aligned}
\int_E \Gamma_2(f) &= \frac{1}{2}\left(\int_E L\Gamma(f)d\mu - \int_E \Gamma(f, Lf)d\mu\right) \\
&= \frac{1}{2}\int_E L\Gamma(f)d\mu - \frac{1}{2}\int_E L(fLf)d\mu + \frac{1}{2}\int_E (Lf)^2 + fL^2 f d\mu \\
&= \int_E (Lf)^2 d\mu.
\end{aligned}$$

Therefore,

$$\frac{n-1}{\rho n}\int_E (Lf)^2 d\mu \ge \int_E \Gamma(f)d\mu.$$

Then by above lemma, it has the result. □

11.5 Applications with PI

Decay of Variance. For a probability measure μ and $f \in L^2(\mu)$, let

$$\operatorname{Var}_\mu(f) = \int_E f^2 d\mu - \left(\int_E f d\mu\right)^2.$$

Proposition 11.5.1. *The compact Markov triple* (E, μ, Γ) *satisfies* PI(C) *if and only if*

$$\operatorname{Var}_\mu(P_t f) \leq e^{-\frac{2t}{C}} \operatorname{Var}_\mu(f), \quad f \in L^2(\mu).$$

Proof. $\Rightarrow$: For $f \in \mathcal{A}$,

$$\frac{d}{dt} \int_E (P_t f)^2 \, d\mu = 2 \int_E P_t f L P_t f d\mu = -2\mathcal{E}\left(P_t f\right).$$

Define

$$\Lambda(t) = e^{2t/C} \operatorname{Var}_\mu\left(P_t f\right),$$

so

$$\Lambda'(t) = \frac{2}{C} \operatorname{Var}_\mu\left(P_t f\right) - 2\mathcal{E}\left(P_t f\right) \leq 0$$

by PI(C). It follows that $\Lambda(t) \leq \Lambda(0)$. For general $f \in L^2(\mu)$, it can get by density.
$\Leftarrow$: It suffices to prove that for $f \in \mathcal{A}$ with $\int f d\mu = 0$. Note that

$$P_t f = f + tLf + o(t),$$

and so

$$\operatorname{Var}\left(P_t f\right) = \int_E f^2 d\mu + 2t \int_E f L f d\mu + o(t).$$

On the other hand,

$$e^{-2t/C} \operatorname{Var}_\mu(f) = \left(1 - \frac{2t}{C} + o(t)\right) \operatorname{Var}_\mu(f).$$

Therefore,

$$2t \int_E f L f d\mu + o(t) \leq \left(-\frac{2t}{C} + o(t)\right) \operatorname{Var}_\mu(f).$$

Then dividing t on the both sides and taking $t \to 0$,

$$2 \int_E f L f d\mu \leq -\frac{2}{C} \operatorname{Var}_\mu(f).$$

□

Log-concave measures.

Definition 11.5.2. The probability measure μ on $\mathbb{R}^n$ defined by

$$d\mu(x) = e^{-W(x)} dx$$

is called log-concave if $W \colon \mathbb{R}^n \to \mathbb{R}$ is convex. For $\rho > 0$, μ is called ρ-strongly log-concave if $W(x) - \rho\,|x|^2$ is convex.

Assume $W \in C^\infty(\mathbb{R}^n)$. And on $\mathbb{R}^n$,

$$\Gamma(f,g) = \langle \nabla f, \nabla g \rangle$$

is a carré du champ. Then by divergence theorem,

$$-\int_{\mathbb{R}^n} \Gamma(f,g) d\mu = -\int_{\mathbb{R}^n} \left\langle e^{-W} \nabla f, \nabla g \right\rangle dx = \int_{\mathbb{R}^n} f(\Delta g - \langle \nabla W, \nabla g \rangle) d\mu.$$

Therefore,

$$Lg = \Delta g - \langle \nabla W, \nabla g \rangle.$$

If all derivatives of $W(x)$ grow at most polynomially fast as $|x| \to \infty$, then $(\mathbb{R}^n, \mu, \Gamma)$ is a compact Markov triple with $\mathcal{A}$ being the class of smooth, bounded functions whose derivatives all vanish super-polynomially fast.
Moreover,

$$\Gamma_2(f,g) = \left\langle \nabla^2 f, \nabla^2 g \right\rangle + (\nabla f)^\top \left(\nabla^2 W\right) \nabla g,$$

By the strongly convexity of W,

$$\Gamma_2(f,f) \geq \rho \, \|f\|_2 = \rho \Gamma(f),$$

i.e., $(\mathbb{R}^n, \mu, \Gamma)$ is $\mathrm{CD}(\rho, \infty)$.

Corollary 11.5.3. *Every ρ-strongly log-concave probability measure satisfies* $\mathrm{PI}(1/\rho)$*, i.e.,*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\rho} \mathbb{E}_\mu[\|\nabla f\|^2].$$

11.6 Log-Sobolev Inequality

Theorem 11.6.1 (Strong Gradient Bound)**.** *Let (E, μ, Γ) be a compact Markov triple that satisfies* $\mathrm{CD}(\rho, \infty)$*. Then for every $f \in \mathcal{A}$ and $t \geq 0$,*

$$\sqrt{\Gamma(P_t f)} \leq e^{-\rho t} P_t \sqrt{\Gamma(f)}.$$

Proof. First, assume $f > 0$. Fix t and define

$$\Lambda(s) = P_s \sqrt{\Gamma(P_{t-s} f)}.$$

by the chain rule for L,

$$\begin{aligned}
\Lambda'(s) &= P_s L \sqrt{\Gamma\left(P_{t-s} f\right)} + P_s \frac{\frac{d}{ds} \Gamma\left(P_{t-s} f\right)}{2\sqrt{\Gamma\left(P_{t-s} f\right)}} \\
&= P_s \left[\frac{L\Gamma\left(P_{t-s} f\right)}{2\sqrt{\Gamma\left(P_{t-s} f\right)}} - \frac{\Gamma\left(\Gamma\left(P_{t-s} f\right)\right)}{4\Gamma\left(P_{t-s} f\right)^{3/2}} - \frac{\Gamma\left(P_{t-s} f, L P_{t-s} f\right)}{\sqrt{\Gamma\left(P_{t-s} f\right)}} \right] \\
&= P_s \left[\frac{\Gamma_2\left(P_{t-s} f\right)}{\sqrt{\Gamma\left(P_{t-s} f\right)}} - \frac{\Gamma\left(\Gamma\left(P_{t-s} f\right)\right)}{4\Gamma\left(P_{t-s} f\right)^{3/2}} \right].
\end{aligned}$$

and so

$$\frac{d}{ds}\left(e^{-\rho s} \Lambda(s)\right) = e^{-\rho s}\left(\Lambda'(s) - \rho \Lambda(s)\right)$$

$$= e^{-\rho s} P_s \left[\frac{\Gamma_2 (P_{t-s} f) \Gamma (P_{t-s} f) - \rho \Gamma (P_{t-s} f)^2 - \frac{1}{4} \Gamma (\Gamma (P_{t-s} f))}{4 \Gamma (P_{t-s} f)^{3/2}} \right].$$

Let $g = P_{t-s} f$.

$$\frac{d}{ds} \left(e^{-\rho s} \Lambda(s) \right) \geq 0 \Leftrightarrow \Gamma(g) (\Gamma_2(g) - \rho \Gamma(g)) \geq \frac{1}{4} \Gamma(\Gamma(g)).$$

Therefore, it suffices to prove above inequality, which is followed by the diffusion property of Γ and $\mathrm{CD}(\rho, \infty)$.
For general f, let $\psi(x) = \sqrt{x + \varepsilon}$ and replace $\Lambda(s)$ by

$$\Lambda(s) = P_s \psi \left(e^{-2\rho s} \Gamma (P_{t-s} f) \right). \qquad \square$$

Let μ be a probability measure on E and $f \colon E \to [0, \infty)$ measurable. Define the entropy by

$$\mathrm{Ent}_\mu(f) := \int_E f \log f d\mu - \int_E f d\mu \log \left(\int_E f d\mu \right),$$

where we adopt the convention that $0 \log 0 = 0$. By Jensen's inequality for $\psi(x) = x \log x$, $\mathrm{Ent}_\mu f \geq 0$. Because ψ is strictly convex, $\mathrm{Ent}_\mu f = 0$ if and only if f is constant μ-a.e. Also,

$$\mathrm{Ent}_\mu(cf) = c \, \mathrm{Ent}_\mu(f), \quad \forall\, c > 0.$$

Remark 11.6.2. Note that if $\nu \ll \mu$ is another probability measure and let $f = \frac{d\nu}{d\mu}$,

$$\mathrm{Ent}_\mu(f) = \int_E f \log f d\mu = \mathrm{KL}(\nu \parallel \mu).$$

Definition 11.6.3 (Log-Sobolev Inequality)**.** If μ is a probability measure and $\mathcal{E}$ is a Dirichlet form, we say they satisfy a log-Sobolev inequality with constant C ($\mathrm{LSI}(C)$) if for all $f \in \mathcal{D}(\mathcal{E})$,

$$\mathrm{Ent}_\mu(f^2) \leq 2C\mathcal{E}(f).$$

The smallest C for which μ and $\mathcal{E}$ satisfy a $\mathrm{LSI}(C)$ is called the log-Sobolev constant of $\mu, \mathcal{E}$.

Assume $f > \varepsilon > 0$, i.e., f is bounded below. Then

$$\mathcal{E}(\sqrt{f}) = \int_E \Gamma(\sqrt{f}) d\mu = \frac{1}{4} \int_E \frac{\Gamma(f)}{f} d\mu.$$

Then $\mathrm{LSI}(C)$ is equivalent to

$$\mathrm{Ent}_\mu(f) \leq 2C\mathcal{E}(\sqrt{f}) = \frac{C}{2} \int_E \frac{\Gamma(f)}{f} d\mu.$$

Definition 11.6.4. Let $\nu \ll \mu$ be another probability measure and $f = \frac{d\nu}{d\mu}$. The Fisher information of ν w.s.t. μ is defined as

$$\mathrm{I}(\nu \mid \mu) = \mathrm{I}_\mu(f) = \int_E \frac{\Gamma(f)}{f} d\mu$$

and the entropy of ν w.s.t. μ (i.e., KL divergence) is defined as

$$H(\nu \mid \mu) = \mathrm{Ent}_\mu(f).$$

By $\operatorname{Ent}_\mu(cf) = c\operatorname{Ent}_\mu(f)$, μ, Γ satisfy $\mathrm{LSI}(C)$ if and only if

$$H(\nu \mid \mu) \le \frac{C}{2} I(\nu \mid \mu)$$

for every probability measure $\nu \ll \mu$, where we allow infinity on the both sides. Moreover, if $\frac{d\nu}{d\mu} \notin \mathcal{D}(\mathcal{E})$, the RHS is ∞.

Proposition 11.6.5. *If $\mu, \mathcal{E}$ satisfy* $\mathrm{LSI}(C)$*, then they satisfy* $\mathrm{PI}(C)$*.*

Proof. Given $f \in \mathcal{A}$ with mean 0 and $\varepsilon > 0$. Because $\log(1+x) = x - \frac{1}{2}x^2 + o(x^2)$,

$$\begin{aligned}\operatorname{Ent}_\mu((1+\epsilon f)^2) &= \int_E (1+\epsilon f)^2 \left(\epsilon f - \frac{\epsilon^2}{2} f^2\right) d\mu - \int_E (1+\epsilon f)^2 d\mu \log \int_E (1+\epsilon f)^2 d\mu + o\left(\epsilon^2\right) \\ &= 2\epsilon^2 \int_E f^2 d\mu + o\left(\epsilon^2\right)\end{aligned}$$

Moreover,

$$\mathcal{E}(1+\epsilon f) = \epsilon^2 \mathcal{E}(f)$$

Apply LSI to $1 + \epsilon f$.

$$\mathcal{E}(1+\epsilon f) \le 2C\epsilon^2 \mathcal{E}(f).$$

Dividing ϵ and taking $\epsilon \to 0$,

$$\int_E f^2 d\mu \le C\mathcal{E}(f).$$

□

LSI **under** $\mathrm{CD}(\rho, \infty)$**.**

Proposition 11.6.6. *For a compact Markov triple (E, μ, Γ), TFAE.*

(1) It satisfies $\mathrm{CD}(\rho, \infty)$ *for some $\rho \in \mathbb{R}$.*

(2) For all $f \in \mathcal{A}$,

$$\Gamma(f)\left(\Gamma_2(f) - \rho\Gamma(f)\right) \ge \frac{1}{4}\Gamma(\Gamma(f)).$$

(3) For every $f \in \mathcal{A}$ and $t \ge 0$,

$$\sqrt{\Gamma(P_t f)} \le e^{-\rho t} P_t \sqrt{\Gamma(f)}.$$

(4) For every positive $f \in \mathcal{A}$ and $t \ge 0$,

$$P_t(f \log f) - P_t f \log P_t f \le \frac{1 - e^{-2\rho t}}{2\rho} P_t \frac{\Gamma(f)}{f}.$$

(5) For every positive $f \in \mathcal{A}$ and $t \ge 0$,

$$P_t(f \log f) - P_t f \log P_t f \ge \frac{e^{2\rho t} - 1}{2\rho} \frac{\Gamma\left(P_t f\right)}{P_t f}.$$

Note that for (4), it is called a local LSI. As $t \to \infty$, by ergodicity, it obtains $\mathrm{LSI}(1/\rho)$.

Corollary 11.6.7. *If (E, μ, Γ) is a compact Markov triple satisfies* $\mathrm{CD}(\rho, \infty)$ *for some $\rho > 0$, then μ, Γ satisfy a* $\mathrm{LSI}(1/\rho)$*.*

Proof of Proposition 11.6.6. (1) $\Rightarrow$ (2) $\Rightarrow$ (3) is by the strong gradient bound. (3) together with Jensen's inequality implies that

$$\Gamma\left(P_t f\right) \leq e^{-2\rho t} P_t \Gamma(f),$$

which follows that (1).
(3) $\Rightarrow$ (4): Fix $t > 0$ and define

$$\Lambda(s) = P_s\left[P_{t-s} f \log P_{t-s} f\right] = P_s \psi\left(P_{t-s} f\right),$$

for $\psi(x) = x \log x$. So by (3),

$$\begin{aligned}
\Lambda'(s) &= P_s\left[\psi''\left(P_{t-s} f\right) \Gamma\left(P_{t-s} f\right)\right] \\
&= P_s \frac{\Gamma\left(P_{t-s} f\right)}{P_{t-s} f} \\
&\leq e^{-2\rho(t-s)} P_s \frac{\left(P_{t-s} \sqrt{\Gamma(f)}\right)^2}{P_{t-s} f},
\end{aligned}$$

Note that by Cauchy-Schwartz inequality,

$$\mathbb{E} X = \mathbb{E}\left[\sqrt{Y} \frac{X}{\sqrt{Y}}\right] \leq \sqrt{\mathbb{E} Y \mathbb{E} \frac{X^2}{Y}} \Rightarrow \frac{(\mathbb{E} X)^2}{\mathbb{E} Y} \leq \mathbb{E} \frac{X^2}{Y}.$$

By setting $X = \sqrt{\Gamma(f)}, Y = f$ and taking expectation w.s.t. $p_{t-s}(x, \cdot)$,

$$\Lambda'(s) \leq e^{-2\rho(t-s)} P_s P_{t-s} \frac{\Gamma(f)}{f} = e^{-2\rho(t-s)} P_t \frac{\Gamma(f)}{f}.$$

So

$$\Lambda(t) - \Lambda(0) \leq P_t \frac{\Gamma(f)}{f} \int_0^t e^{-2\rho(t-s)} ds = \frac{1 - e^{-2\rho t}}{2\rho} P_t \frac{\Gamma(f)}{f}.$$

(3) $\Rightarrow$ (5) : It is similar as (3) $\Rightarrow$ (4), except for taking $X = \sqrt{\Gamma\left(P_{t-s} f\right)}, Y = P_{t-s} f$ and the expectation w.s.t. $p_s(x, \cdot)$, which impllies that

$$\Lambda'(s) \geq \frac{\left(P_s \sqrt{\Gamma\left(P_{t-s} f\right)}\right)^2}{P_t f} \geq e^{2\rho s} \frac{\Gamma\left(P_t f\right)}{P_t f}.$$

(4) $\Rightarrow$ (1) : (4) is local LSI(ρ), which implies local PI as similar as above proposition. Then it implies CD(ρ, ∞). (5) $\Rightarrow$ (1) is as similar as above. □

LSI **under** CD(ρ, n)**.**

Theorem 11.6.8. *If a compact Markov triple (E, μ, Γ) satisfies* CD(ρ, n) *for some $\rho > 0$, then μ, Γ satisfy a* LSI(C).

Lemma 11.6.9. *Suppose that*

$$\int_E f \Gamma(\log f) d\mu \leq C \int_E f \Gamma_2(\log f) d\mu$$

for some $C > 0$ and all positive $f \in \mathcal{A}$. Then μ, Γ satisfy a LSI(C).

Proof. For fix some positive $f \in \mathcal{A}$. Let

$$\Lambda(t) = \int_E P_t f \log P_t f d\mu.$$

Then

$$\Lambda'(t) = -\int_E \frac{\Gamma(P_t f)}{P_t f} d\mu = -\int_E P_t f \Gamma(\log P_t f) d\mu,$$

and

$$\Lambda''(t) = \int_E \frac{\Gamma(P_t f) L P_t f}{(P_t f)^2} - \frac{2\Gamma(P_t f, L P_t f)}{P_t f} d\mu.$$

Taking $g = P_t f$, by the diffusion property of L,

$$L\frac{\Gamma(g)}{g} = -\frac{\Gamma(g) Lg}{g^2} + \frac{L\Gamma(g)}{g} - \frac{2\Gamma(g, \Gamma(g))}{g^2} + \frac{2\Gamma(g)^2}{g^3}.$$

Since $\int Lh d\mu = 0$,

$$\int_E \frac{\Gamma(g) Lg}{g^2} = \int_E \frac{L\Gamma(g)}{g} - 2\frac{\Gamma(g, \Gamma(g))}{g^2} + 2\frac{\Gamma(g)^2}{g^3} d\mu.$$

Note that $L\Gamma(g) - 2\Gamma(g, Lg) = 2\Gamma_2(g)$. So

$$\Lambda''(t) = 2\int_E \frac{\Gamma_2(g)}{g} - \frac{\Gamma(g, \Gamma(g))}{g^2} + \frac{\Gamma(g)^2}{g^3} d\mu = \int_E g\Gamma_2(\log g).$$

Therefore,

$$\begin{aligned} \Lambda'(t) &= -\int_E P_t f \Gamma(\log P_t f) d\mu \\ \Lambda''(t) &= 2\int_E P_t f \Gamma_2(\log P_t f) d\mu. \end{aligned}$$

By assumption of $-\Lambda'(t) \leq \frac{C}{2}\Lambda''(t)$, $\Lambda'(t) \geq \Lambda'(0)\exp(-\frac{2t}{C})$, i.e.,

$$\int_E \frac{\Gamma(P_t f)}{P_t f} d\mu \leq e^{-2t/C} \int_E \frac{\Gamma(f)}{f} d\mu.$$

Therefore,

$$\Lambda(0) - \Lambda(t) = -\int_0^t \Lambda'(s) ds \leq -\Lambda'(0) \int_0^\infty e^{-2s/C} ds = \frac{C\left(1 - e^{-2t/C}\right)}{2} \mathrm{I}_\mu(f),$$

then it can prove that by taking $t \to \infty$. □

Proof of Theorem 11.6.8. It suffices to check the condition of above lemma. By the diffusion property of Γ,

$$\begin{aligned} \Gamma_2(e^{ag}) &= a^2 e^{2ag}\left[\Gamma_2(g) + a\Gamma(g, \Gamma(g)) + a^2\Gamma(g)^2\right] \\ \Gamma(e^{ag}) &= a^2 e^{2ag}\Gamma(g) \\ Le^{ag} &= ae^{ag}[Lg + a\Gamma(g)]. \end{aligned}$$

Therefore, $\mathrm{CD}(\rho, n)$ implies that

$$\Gamma_2(e^{ag}) = a^2 e^{2ag}\left[\Gamma_2(g) + a\Gamma(g, \Gamma(g)) + a^2\Gamma(g)^2\right]$$

$$\begin{aligned}\Gamma\left(e^{ag}\right) &= a^2 e^{2ag}\Gamma(g)\\ Le^{ag} &= ae^{ag}[Lg + a\Gamma(g)].\end{aligned}$$

So

$$\int_E e^g\left[\Gamma_2(g) + a\Gamma(g,\Gamma(g)) + a^2\Gamma(g)^2 - \rho\Gamma(g) - \frac{1}{n}[Lg + a\Gamma(g)]^2\right]d\mu \geq 0.$$

Note that

$$4\left(Le^{g/2}\right)^2 = e^g\left[Lg + \frac{1}{2}\Gamma(g)\right]^2, \quad \text{and} \quad \int_E (Lf)^2 d\mu = \int_E \Gamma_2(f)d\mu.$$

It implies that

$$\begin{aligned}&\int_E [Lg + a\Gamma(g)]^2 d\mu\\ &= \int_E 4\left(Le^{g/2}\right)^2 + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right]d\mu\\ &= \int_E 4\Gamma_2\left(e^{g/2}\right) + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right]d\mu\\ &= \int_E 4\Gamma_2\left(e^{g/2}\right) + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right]d\mu\\ &= \int_E e^g\left[\Gamma_2(g) + \frac{1}{2}\Gamma(g,\Gamma(g)) + (2a-1)\Gamma(g)Lg + a^2\Gamma(g)^2\right]d\mu,\end{aligned}$$

and so

$$\int_E e^g\Gamma(g)Lgd\mu = -\int_E \Gamma\left(g, e^g\Gamma(g)\right)d\mu = -\int_E e^g\Gamma(g,\Gamma(g)) + e^g\Gamma(g)^2 d\mu.$$

Then the inequality implies that

$$\int_E e^g\left[\frac{n-1}{n}\Gamma_2(g) + b_n\Gamma(g,\Gamma(g)) + c_n\Gamma(g)^2 - \rho\Gamma(g)\right]d\mu \geq 0$$

for

$$b_n = \frac{2an + 4a - 3}{2n}, \quad c_n = \frac{na^2 - (a-1)^2}{n}.$$

By choosing $a = \frac{3}{2n+4}$, it has

$$\int_E e^g\Gamma_2(g)d\mu \geq \frac{n\rho}{n-1}\int_E e^g\Gamma(g)d\mu. \qquad \square$$

11.7 Applications with LSI

Decay of Entropy.

Proposition 11.7.1. *The compact Markov triple (E,μ,Γ) satisfies a* LSI(C) *if and only if*

$$\operatorname{Ent}_\mu\left(P_t f\right) \leq e^{-2t/C}\operatorname{Ent}_\mu(f)$$

for every $t \geq 0$ and every $f \in L^\mu$ with finite entropy.

Proof. It suffices to consider $f \in \mathcal{A}$ with finite entropy. Define

$$\Lambda(t) = \operatorname{Ent}_\mu\left(P_t f\right) = \int_E P_t f \log P_t f d\mu - \int_E f d\mu \log \int_E f d\mu.$$

$\Rightarrow$: Note that

$$\Lambda'(t) = -\int_E \frac{\Gamma(P_t f)}{P_t f} d\mu = -I_\mu(P_t f),$$

so by LSI, $\Lambda'(t) \le -\frac{2}{C}\Lambda(t)$ that implies that $\Lambda(t) \le e^{-2t/C}\Lambda(0)$.

$\Leftarrow$: By Taylor expansion,

$$\Lambda(t) = \Lambda(0) + t\Lambda'(0) + o(t) = \Lambda(0) - tI_\mu(f) + o(t).$$

Because

$$\Lambda(t) \le e^{-2t/C}\Lambda(0) = \left(1 - \frac{2t}{C} + o(t)\right)\Lambda(0),$$

as $t \to 0$, we have

$$I_\mu(f) \ge \frac{2}{C}\Lambda(0).$$

□

If $f = \frac{d\nu_0}{d\mu}$, then for $\nu_t = P_t^*\mu_0$,

$$d\nu_t = P_t f d\mu.$$

Suppose μ, Γ satisfies LSI(C). Then we have

$$H(\nu_t \mid \mu) \le e^{-\frac{2t}{C}} H(\nu_0 \mid \mu).$$

Moreover, by the following Pinsker-Csizsár-Kullback inequality,

$$d_{\mathrm{TV}}(\mu, \nu_t)^2 \le \frac{1}{2} e^{-\frac{2t}{C}} H(\nu_0 \mid \mu),$$

where

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| = \frac{1}{2}\int_E \left|1 - \frac{d\nu}{d\mu}\right| d\mu.$$

Moreover, PI can also be applied to consider the convergence, becasue

$$d_{\mathrm{TV}}(\mu, \nu)^2 \le \frac{1}{4}\operatorname{Var}_\mu\left(\frac{d\nu}{d\mu}\right),$$

but it needs $\frac{d\nu}{d\mu} \in L^2(\mu)$.

Proposition 11.7.2 (Pinsker-Csizsár-Kullback)**.** *For any probability measure μ, ν on the same space,*

$$d_{\mathrm{TV}}(\mu, \nu)^2 \le \frac{1}{2} H(\nu \mid \mu),$$

where $H(\nu \mid \mu) = \infty$ if ν is not absolutely continuous to μ.

Proof. WTLG, assume $f = \frac{d\nu}{d\mu} \in L^1$. Therefore, it suffices to show

$$\left(\int_E |1 - f| d\mu\right)^2 \le 2\operatorname{Ent}_\mu(f).$$

Define $f_s = 1 + s(f-1)$ for $s \in [0,1]$ and

$$\Lambda(s) = 2\operatorname{Ent}_\mu(f_s) - \left(\int_E |1 - f_s| d\mu\right)^2 = 2\operatorname{Ent}_\mu(f_s) - s^2\left(\int_E |1 - f| d\mu\right)^2.$$

Since $\int f_s d\mu = 1$ for all s, it follows that

$$\frac{d}{ds}\operatorname{Ent}_\mu(f_s) = \int_E (f-1)(1 + \log f_s) d\mu, \quad \frac{d^2}{ds^2}\operatorname{Ent}_\mu(f_s) = \int_E \frac{(f-1)^2}{f_s} d\mu.$$

In particular, $\Lambda(0) = \Lambda'(0) = 0$ and

$$\Lambda''(s) = 2\int_E \frac{(f-1)^2}{f_s} d\mu - 2\left(\int_E |1 - f| d\mu\right)^2 \ge 0,$$

by Cauchy-Schwartz inequality. Hence, $\Lambda(s) \ge 0$ for all $s \in [0,1]$. □

Hypercontractivity. We already shown that if μ is an invariant measure, then $P_t\colon L^p(\mu) \to L^p(\mu)$ is contractive.

Theorem 11.7.3. *For a compact Markov triple with semigroup P_t, TFAE.*

(1) *It satisfies* LSI(C).

(2) *(Hypercontractivity) For some (or every) $1 < p < \infty$, every $t \geq 0$ and $f \in L^p(\mu)$,*

$$\|P_t f\|_{q(t)} \leq \|f\|_p \,,$$

where $q(t)$ satisfies $\frac{q(t)-1}{p-1} = e^{2t/C}$.

(3) *(Reverse hypercontractivity) For some (or every) $-\infty < p < 1$, every $t \geq 0$ and every positive, bounded $f \in \mathcal{A}$,*

$$\|P_t f\|_{q(t)} \geq \|f\|_p \,,$$

where $q(t)$ is as above.

Remark 11.7.4. All eigenvectors of $L\colon L^2 \to L^2$ belong to L^q for all $2 < q < \infty$. If $Lf = -\lambda f$ for $\lambda > 0$ then $P_t f = e^{-\lambda t} f$. By above theorem with $p = 2$ and $t = \frac{C}{2}\log(q-1)$,

$$e^{-\lambda t}\|f\|_q = \|P_t f\|_q \leq \|f\|_2 \;\Rightarrow\; \|f\|_q \leq (q-1)^{C\lambda/2}\|f\|_2.$$

11.8 Riemannian Markov Operator

Let (M, g) be a compact Riemannian manifold and $W \in C^\infty(M)$. WTLG, assume $\int_M e^{-W} dV = 1$, where dV is the canonical volume form on M. Let M be equipped with Borel σ-algebra and for any Borel set A, define

$$\mu(A) := \int_M \mathbb{I}_A e^{-W} dV,$$

which is a probability measure. Let $\mathcal{A} = C^\infty(M)$ and define $\Gamma\colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ by

$$\Gamma(f, g) = \langle \nabla f, \nabla g \rangle \,.$$

Locally, $\Gamma(f, g) = g^{ij}\partial_i f \partial_j g$, which obviously a diffusion carré du champ. We can obtain

$$Lg = \Delta g - \langle \nabla W, \nabla g \rangle \,, \quad \forall\; g \in C^\infty(M),$$

because by the convergence theorem

$$\int_M f L g d\mu = -\int_M \Gamma(f, g) d\mu.$$

Moreover,

$$|\mathcal{E}(f, g)| = \left|\int_M \Gamma(f, g) d\mu\right| \leq \|Lf\|_{L^2(\mu)} \|g\|_{L^2(\mu)} \,,$$

which implies that $\mathcal{E}$ is closable. So it is a Dirichlet form.

Theorem 11.8.1 (Parabolic regularity theorem)**.** *Suppose that $u\colon M \times \mathbb{R} \to \mathbb{R}$ is a bounded solution to*

$$\begin{cases} \dfrac{\partial u(x,t)}{\partial t} = \Delta u - \langle \nabla W, \nabla u \rangle \\ u(x, 0) = f(x), \end{cases}$$

where $f, W \in C^\infty(M)$. Then $u(t, \cdot) \in C^\infty(M)$.

Because $u(x,t) = (P_t f)(x)$ satisfies above PDE, $P_t f \in C^\infty(M)$ when $f \in C^\infty(M)$. Therefore, we have $P_t \mathcal{A} \subset \mathcal{A}$. It follows that (M, μ, Γ) is a compact Markov triple.

Lemma 11.8.2. *Let (M, g) be a Riemannian manifold with the Levi-Civita connection ∇ and $f \in C^\infty(M)$.*

(1) Let E_i be an orthonormal frame in TM. Then

$$\left\|\nabla^2 f\right\|^2 = \sum_i \|\nabla_{E_i}(\nabla f)\|^2,$$

where note that $\|\nabla^2 f\|^2 = g^{kh} g^{lj} f_{;hl} f_{;kj}$.

(2) For any $p \in M$, choosing a normal coordinate centered p,

$$(\Delta f)(p) = \sum_{i=1}^n (\partial_i^2 f)(p).$$

Proof. (1) Locally, let

$$E_i = E_i^\ell \frac{\partial}{\partial x^\ell}.$$

Let

$$\nabla_{\frac{\partial}{\partial x^\ell}} dx^k = -\Gamma^k_{\ell h} dx^h,$$

where $\Gamma^k_{\ell h}$ is the Christoffel symbol. Then

$$\begin{aligned}
\nabla_{E_i}(\nabla f) &= \nabla_{E_i}\left(\frac{\partial f}{\partial x^k} dx^k\right) \\
&= E_i^\ell \frac{\partial^2 f}{\partial x^k \partial x^\ell} dx^k - E_i^\ell \frac{\partial f}{\partial x^k} \Gamma^k_{\ell h} dx^h \\
&= E_i^\ell \left(\frac{\partial^2 f}{\partial x^h \partial x^\ell} - \frac{\partial f}{\partial x^k} \Gamma^k_{\ell h}\right) dx^h \\
&= E_i^\ell f_{;h\ell} dx^h.
\end{aligned}$$

So we have

$$\|\nabla_{E_i}(\nabla f)\|^2 = g^{kh} \left(E_i^\ell f_{;h\ell}\right) \left(E_i^j f_{;kj}\right),$$

and

$$\sum_i \|\nabla_{E_i}(\nabla f)\|^2 = g^{kh} \left(\sum_i E_i^\ell E_i^j\right) f_{;h\ell} f_{;kj}.$$

Because E_i is orthonormal,

$$\begin{aligned}
\langle E_i, E_m \rangle &= \left\langle E_i^\ell \frac{\partial}{\partial x^\ell}, E_m^j \frac{\partial}{\partial x^j} \right\rangle \\
&= E_i^\ell g_{\ell j} E_m^j = \delta_{im},
\end{aligned}$$

which means that $E^\top g E = I$ for matrix $E = (E_j^m)_{m \times j}$. It follows that

$$g^{-1} = E E^\top \ \Rightarrow\ g^{\ell j} = \sum_i E_i^\ell E_i^j.$$

Therefore,

$$\sum_i \|\nabla_{E_i}(\nabla f)\|^2 = g^{kh} g^{\ell j} f_{;h\ell} f_{;kj} = \left\|\nabla^2 f\right\|^2.$$

(2) Choose a normal coordinate centered at p, i.e. $g^{ij}(p) = \delta_{ij}$ and $\Gamma^h_{ij}(p) = 0$. Then

$$\begin{aligned}(\Delta f)(p) &= g^{ij}(p) f_{;ij}(p) \\ &= g^{ij}(p)\left(\frac{\partial^2 f}{\partial x^i \partial x^j}(p) - \frac{\partial f}{\partial x^h}(p)\Gamma^h_{ij}(p)\right) \\ &= \frac{\partial^2 f}{(\partial x^i)^2}(p).\end{aligned}$$

□

Remark 11.8.3. In general, for $\omega_1, \omega_2 \in \Gamma(T^*M)$,

$$\langle \omega_1, \omega_2 \rangle := \langle \sharp\omega_1, \sharp\omega_2 \rangle = g(\sharp\omega_1, \sharp\omega_2).$$

Note that for $\omega_1 = f^1_i dx^i$ and $\omega_2 = f^2_j dx^j$,

$$\begin{aligned}\langle \omega_1, \omega_2 \rangle &= \left\langle g^{\ell i} f^1_i \frac{\partial}{\partial x^\ell}, g^{kj} f^2_j \frac{\partial}{\partial x^k} \right\rangle \\ &= f^1_i f^2_j g^{\ell i} g^{kj} g_{\ell k} \\ &= f^1_i f^2_j g^{ji},\end{aligned}$$

which means that $\langle \cdot, \cdot \rangle$ on T^*M with the matrix expression g^{-1}. In this notation,

$$\langle \omega, X \rangle := \langle \sharp\omega, X \rangle = \langle \omega, \flat X \rangle = \omega(X).$$

Note that $\langle X, \operatorname{grad} f \rangle = X(f)$ and by above notation

$$\langle \omega, \nabla f \rangle = \langle \omega, \operatorname{grad} f \rangle = \omega(\operatorname{grad} f).$$

Moreover, for any $\omega \in T^*M$ and $X \in \Gamma(TM)$, because ∇ is Levi-Civita,

$$\nabla_X(\sharp\omega) = \sharp(\nabla_X \omega).$$

It is because for any $Y \in \Gamma(TM)$,

$$\begin{aligned}Xg(\sharp\omega, Y) &= g(\nabla_X \sharp\omega, Y) + g(\sharp\omega, \nabla_X Y) \\ &= g(\nabla_X \sharp\omega, Y) + \omega(\nabla_X Y).\end{aligned}$$

On the other hand,

$$\begin{aligned}Xg(\sharp\omega, Y) = X(\omega(Y)) &= (\nabla_X \omega)(Y) + \omega(\nabla_X Y) \\ &= g(\sharp(\nabla_X \omega), Y) + \omega(\nabla_X Y).\end{aligned}$$

Therefore,

$$g(\sharp(\nabla_X \omega), Y) = g(\nabla_X \sharp\omega, Y) \ \Rightarrow\ \nabla_X(\sharp\omega) = \sharp(\nabla_X \omega).$$

It follows that

$$\langle \nabla_X(\sharp\omega_1), \sharp\omega_2 \rangle = \langle \nabla_X \omega_1, \omega_2 \rangle,$$

which implies that

$$\nabla_X \langle \omega_1, \omega_2 \rangle = \langle \nabla_X \omega_1, \omega_2 \rangle + \langle \omega_1, \nabla_X \omega_2 \rangle.$$

Lemma 11.8.4. *Choose a orthonormal basis $\{E_i\}$ and let*

$$H_{ij} = \nabla^2 f(E_i, E_j).$$

Then

$$\left\|\nabla^2 f\right\|^2 = \sum_{i,j} H_{ij}^2, \quad \Delta f = \sum_j H_{ii},$$

i.e, $\left\|\nabla^2 f\right\|^2 = \|H\|_{\mathrm{F}}^2$ *and* $\Delta f = \operatorname{tr} H$*, which induces*

$$(\Delta f)^2 \leq n \left\|\nabla^2 f\right\|^2$$

if $\dim M = n$*.*

Proof. First, for any $\omega \in \Gamma(T^*M)$,

$$\begin{aligned}
\sum_j \omega(E_j)^2 &= \sum_j \left(\sum_i \omega_i E_j^i\right)\left(\sum_\ell \omega_\ell E_j^\ell\right) \\
&= \sum_{i,\ell} \omega_i \omega_\ell \left(\sum_j E_j^i E_j^\ell\right) \\
&= \sum_{i,\ell} \omega_i \omega_\ell g^{i\ell} = \|\omega\|^2 .
\end{aligned}$$

By above lemma,

$$\left\|\nabla^2 f\right\| = \sum_{i,j} \left(\nabla^2 f(E_i, E_j)\right)^2 .$$

And $\Delta f = \operatorname{tr} H$ is just by definition. $\square$

Definition 11.8.5 (Weighted Ricci Curvature)**.** Given $W \in C^\infty(M)$, the weighted Ricci curvature is a $(2,0)$-tensor Ric_W defined by

$$\mathrm{Ric}_W(X,Y) := \mathrm{Ric}(X,Y) + (\nabla^2 W)(X,Y).$$

Proposition 11.8.6. *Let* (M,g) *be a compact Riemannian manifold and* $W \in C^\infty(M)$*. Define*

$$\begin{aligned}
Lf &= \Delta f - \langle \nabla W, \nabla f\rangle \\
\Gamma(f,g) &= \langle \nabla f, \nabla g\rangle \\
\Gamma_2(f) &= \frac{1}{2} L\Gamma(f) - \Gamma(f, Lf).
\end{aligned}$$

Then

$$\Gamma_2(f) = \mathrm{Ric}_W(\operatorname{grad} f, \operatorname{grad} f) + \left\|\nabla^2 f\right\|^2 .$$

Proof. By Bochner's formula, we have

$$\frac{1}{2}\Delta \|\nabla f\|^2 = \mathrm{Ric}(\operatorname{grad} f, \operatorname{grad} f) + \langle \nabla f, \nabla(\Delta f)\rangle + \left\|\nabla^2 f\right\|^2 .$$

By definition,

$$\begin{aligned}
\Gamma_2(f) &= \frac{1}{2} L(\|\nabla f\|^2) - \langle \nabla f, \nabla L f\rangle \\
&= \frac{1}{2}\Delta(\|\nabla f\|^2) - \frac{1}{2}\left\langle \nabla W, \nabla(\|\nabla f\|^2)\right\rangle - \langle \nabla f, \nabla(\Delta f)\rangle + \langle \nabla f, \nabla \langle \nabla W, \nabla f\rangle\rangle \\
&= \mathrm{Ric}(\operatorname{grad} f, \operatorname{grad} f) + \left\|\nabla^2 f\right\|^2 - \frac{1}{2}\left\langle \nabla W, \nabla(\|\nabla f\|^2)\right\rangle + \langle \nabla f, \nabla \langle \nabla W, \nabla f\rangle\rangle .
\end{aligned}$$

Moreover,

$$\begin{aligned}\frac{1}{2}\left\langle \nabla W, \nabla(\|\nabla f\|^2)\right\rangle &= \frac{1}{2}\nabla_{\operatorname{grad} W}\left\langle \nabla f, \nabla f\right\rangle \\ &= \left\langle \nabla_{\operatorname{grad} W}\nabla f, \nabla f\right\rangle.\end{aligned}$$

On the other hand,

$$\begin{aligned}\left\langle \nabla f, \nabla\left\langle \nabla W, \nabla f\right\rangle\right\rangle &= \nabla_{\operatorname{grad} f}\left\langle \nabla W, \nabla f\right\rangle \\ &= \left\langle \nabla_{\operatorname{grad} f}\nabla W, \nabla f\right\rangle + \left\langle \nabla W, \nabla_{\operatorname{grad} f}\nabla f\right\rangle \\ &= (\nabla^2 W)(\operatorname{grad} f, \operatorname{grad} f) + \left\langle \nabla W, \nabla_{\operatorname{grad} f}\nabla f\right\rangle,\end{aligned}$$

because

$$(\nabla^2 W)(\operatorname{grad} f, \operatorname{grad} f) = (\nabla_{\operatorname{grad} f}\nabla W)(\operatorname{grad} f) = \left\langle \nabla_{\operatorname{grad} f}\nabla W, \nabla f\right\rangle$$

Moreover,

$$\left\langle \nabla_{\operatorname{grad} W}\nabla f, \nabla f\right\rangle = \nabla^2 f(\operatorname{grad} f, \operatorname{grad} W)$$

and

$$\left\langle \nabla W, \nabla_{\operatorname{grad} f}\nabla f\right\rangle = \nabla^2 f(\operatorname{grad} W, \operatorname{grad} f),$$

they are same by the symmetry of $\nabla^2 f$. Therefore,

$$\begin{aligned}\Gamma_2(f) &= \operatorname{Ric}(\operatorname{grad} f, \operatorname{grad} f) + \left\|\nabla^2 f\right\|^2 + (\nabla^2 W)(\operatorname{grad} f, \operatorname{grad} f) \\ &= \operatorname{Ric}_W(\operatorname{grad} f, \operatorname{grad} f) + \left\|\nabla^2 f\right\|^2.\end{aligned}$$

□

Corollary 11.8.7. *Consider a compact Markov triple (M, μ, Γ) defined as above, where M has dimension n.*

(1) If $W \equiv c$ and $\operatorname{Ric} \geq \rho g$, then (M, μ, Γ) satisfies $\operatorname{CD}(\rho, n)$.

(2) If $\operatorname{Ric}_W \geq \rho g$, then (M, μ, Γ) satisfies $\operatorname{CD}(\rho, \infty)$.

(3) If $\operatorname{Ric}_W \geq \rho g$ and $m > n$ such that

$$\operatorname{Ric}_W \geq \rho g + \frac{1}{m-n}\nabla W \otimes \nabla W,$$

then (M, μ, Γ) satisfies $\operatorname{CD}(\rho, m)$.

Proof. (1) It is directly obtained by

$$(\Delta f)^2 \leq n\left\|\nabla f\right\|^2.$$

(2) It is obvious.

(3) Similarly,

$$\begin{aligned}\Gamma_2(f) &\geq \rho\Gamma(f) + \frac{1}{m-n}\langle \nabla W, \nabla f\rangle^2 + \frac{1}{n}(\Delta f)^2 \\ &\geq \rho\Gamma(f) + \frac{1}{m}\left(\Delta f - \langle \nabla W, \nabla f\rangle\right),\end{aligned}$$

where the last inequality is because $\frac{a^2}{m} + \frac{b^2}{m-n} \geq \frac{1}{m}(a+b)^2$. □