Stochastic Analysis

Zhan Zhiyuan

1 Review: Probability and Stochastic

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{R}^d$ be the Borel sets of $\mathbb{R}^d$.

- Independence: First, a collection of σ-algebras $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ is called independent if for any $k \in \mathbb{N}$ and any $i_1, i_2, \cdots, i_k \in \Lambda$,

$$\mathbb{P}(C_1 \cap C_2 \cap \cdots \cap C_k) = \prod_{j=1}^{k} \mathbb{P}(C_j),\ \forall\ C_j \in \mathcal{F}_{i_j}$$

  And for a collection of random variables $(X_\lambda)_{\lambda\in\Lambda}$, each

$$\sigma(X_\lambda) := \left\{X_\lambda^{-1}(A) \colon A \in \mathcal{R}\right\}$$

  is a σ-algebra. So $(X_\lambda)_{\lambda\in\Lambda}$ is said independent if the corresponding $(\sigma(X_\lambda))_{\lambda\in\Lambda}$ is independent.

  Theorem 1.1. *For independent $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ and $N \in \mathbb{N} \cup \{\infty\}$, let $\Lambda_n \subset \Lambda$ for $n = 1, 2, \cdots, N$ be disjoint, then*

$$\left(\sigma\left(\bigcup_{\lambda\in\Lambda_n} \mathcal{F}_\lambda\right)\right)_{n=1}^{N}$$

  is independent.

- Convergence of random variables: Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables.

  (1) $X_n \xrightarrow{a.e} X$ almost everywhere if $\mathbb{P}\left(\{\omega \in \Omega \colon \lim_n X_n(\omega) = X(\omega)\}\right) = 1$.

  (2) $X_n \xrightarrow{\mathbb{P}} X$ in probability if

$$\forall\ \varepsilon > 0,\ \lim_n \mathbb{P}\left(\{\omega \in \Omega \colon |X_n(\omega) - X(\omega)| > \varepsilon\}\right) = 0$$

  In particular, if $X_n \xrightarrow{\mathbb{P}} 0$, then $X_n = o_p(1)$.
  Remark. X_n is said tight denoted by $X_n = \mathcal{O}_p(1)$ if

$$\forall\ \varepsilon > 0,\ \exists\ M > 0,\ s.t.\ \sup_n \mathbb{P}\left(\{\omega \in \Omega \colon |X_n(\omega)| > M\}\right) < \varepsilon$$

  (3) For $\alpha > 0$, let $X_n \in L^\alpha(\Omega, \mathbb{P})$. $X_n \to X$ in α-mean if $\lim_n \mathbb{E}[|X_n - X|^\alpha] = 0$.

  Theorem 1.2. *For the convergence of X_n,*

  (1) $X_n \xrightarrow{a.e} X$ implies $X_n \xrightarrow{\mathbb{P}} X$;

  (2) $X_n \to X$ in α-mean implies $X_n \xrightarrow{\mathbb{P}} X$.

  Theorem 1.3 (Strong Law of Large Number)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be* i.i.d *random variables and $X_1 \in L^1$. Then*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{a.e.} \mathbb{E}[X_1]$$

- Convergence in distribution: Let $\{X_n\}$ be a sequence of $\mathbb{R}^k$-valued random variables with distribution functions $F_n(x)$ and X be a $\mathbb{R}^k$-valued random variable with distribution functions $F(x)$. $X_n \xrightarrow{d} X$ in distribution if for any continuous point x of $F(x)$

$$\lim_{n\to\infty} F_n(x) = F(x)$$

Remark. Note that it does not require X_n and X are in the same probability space. In fact, convergence in distribution is equivalent to the weak convergence of $\mathbb{P}_n \to \mathbb{P}$ when considering the metric space.

Theorem 1.4. *The following statements are equivalent.*

(1) $X_d \xrightarrow{d} X$ $\mathbb{R}^k$*-valued random variables.*

(2) For all $f \in \mathcal{C}_b(\mathbb{R}^k)$, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

(3) For all $f \in \mathcal{C}_b(\mathbb{R}^k)$ *uniformly continuous,* $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

(4) For any closed $A \in \mathbb{R}^k$,

$$\limsup_n \mathbb{P}_n(A) \leqslant \mathbb{P}(A)$$

(5) For any open $A \in \mathbb{R}^k$,

$$\liminf_n \mathbb{P}_n(A) \geqslant \mathbb{P}(A)$$

Theorem 1.5. *Let* $\{X_n\}_{n\in\mathbb{N}}$ *and* X *be* $\mathbb{R}^k$*-valued random variables.*

(1) If $X_d \xrightarrow{d} X$*, then for any* $g \in \mathcal{C}(\mathbb{R}^k, \mathbb{R}^m)$, $g(X_k) \xrightarrow{d} g(X)$.

(2) $X_d \xrightarrow{\mathbb{P}} X$ *implies* $X_d \xrightarrow{d} X$.

By the characteristic function $\phi_X(\lambda) = \mathbb{E}[e^{i\lambda X}]$, it can prove the following central limit theorem.

Theorem 1.6 (Central Limit Theorem)**.** *Let* $\{X_n\}_{n=1}^{\infty}$ *be* i.i.d *with mean* μ *and variance* σ^2*. Let* $S_n = X_1 + X_2 + \cdots + X_n$*. Then*

$$\frac{\sqrt{n}}{\sigma}(\frac{S_n}{n} - \mu) \xrightarrow{d} \mathcal{N}(0,1)$$

- Conditional Expectation: Let X be a random variable and $X \in L^1$. For $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, let

$$\mathbb{E}[X \mid B] := \frac{\mathbb{E}[X, B]}{\mathbb{P}(B)}$$

where $\mathbb{E}[X, B] = \mathbb{E}[X\mathbb{1}_B]$.

(1) For $\{B_i\}_{i=1}^n \subset \mathcal{F}$ disjoint and $\Omega = \bigcup_i B_i$, define a random variable

$$\mathbb{E}[X \mid \{B_i\}](\omega) = \sum_{i=1}^n \mathbb{E}[X \mid B_i]\mathbb{1}_{B_i}(\omega)$$

(2) Let $\mathcal{B} = \sigma(B_1, B_2, \cdots, B_n)$ be the σ-algebra generated by $\{B_i\}_{i=1}^n$. We denote

$$\mathbb{E}[X \mid \{B_i\}] = \mathbb{E}[X \mid \mathcal{B}]$$

Then we have $\mathbb{E}[X \mid \mathcal{B}]$ is $\mathcal{B}$-measurable and $\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{B}], B\right] = \mathbb{E}[X, B]$ for any $B \in \mathcal{B}$.

Definition 1.7 (Conditional Expectations)**.** Let $X \in L^1$ and $\mathcal{G} \subset \mathcal{F}$ be a σ-subalgebra. The conditional expectation of X over $\mathcal{G}$ is a random variable Y such that

(1) $Y \in L^1$;
(2) Y is $\mathcal{G}$-measurable;
(3) for any $A \in \mathcal{G}$, $\mathbb{E}[Y, A] = \mathbb{E}[X, A]$.

Denote $Y = \mathbb{E}[X \mid \mathcal{G}]$. In particular, if $X = \mathbb{1}_A$, then $\mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}[\mathbb{1}_A \mid \mathcal{G}]$.

For any $X \in L^1$ and σ-subalgebra $\mathcal{G}$, by the Radon-Nikodym theorem, $\mathbb{E}[X \mid \mathcal{G}]$ uniquely exists.

Theorem 1.8. *Let $X, Y \in L^1$ and $\mathcal{G}$ be a σ-subalgebra.*

(1) For any $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$.
(2) $X \geqslant Y$ implies $\mathbb{E}[X \mid \mathcal{G}] \geqslant \mathbb{E}[Y \mid \mathcal{G}]$.
(3) If $X \in L^p$ $(p \geqslant 1)$, then $\mathbb{E}[X \mid \mathcal{G}] \in L^p$ and

$$\|\mathbb{E}[X \mid \mathcal{G}]\|_{L^p} \leqslant \|X\|_{L^p}$$

In particular, $|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}[|X| \mid \mathcal{G}]$.

(4) For $0 \leqslant X_1 \leqslant X_2 \leqslant \cdots$ random variables

$$X_n \xrightarrow{a.e} X \quad \Rightarrow \quad \mathbb{E}[X_n \mid \mathcal{G}] \xrightarrow{a.e.} \mathbb{E}[X \mid \mathcal{G}]$$

And also $\mathbb{E}[\cdot \mid \mathcal{G}]$ satisfies the DCT and Fatou's lemma.

(5) If Y is $\mathcal{G}$-measurable, then $\mathbb{E}[YX \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$. In particular, $\mathbb{E}[Y \mid \mathcal{G}] = Y$.
(6) If $\mathcal{G}'$ is also a σ-subalgebra $\mathcal{G}$, then

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}'] = \mathbb{E}[X \mid \mathcal{G}']$$

(7) If $\mathcal{H}$ is independent with $\sigma(X, \mathcal{G})$, then

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

In particular, if $\sigma(X)$ is independent with $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ a.e..

(8) If $\phi \colon \mathbb{R} \to \mathbb{R}$ is convex and $\phi(X) \in L^1$, then $\phi(\mathbb{E}[X \mid \mathcal{G}]) \leqslant \mathbb{E}[\phi(X) \mid \mathcal{G}]$.
(9) If $X \in L^2$, then $Y^* = \mathbb{E}[X \mid \mathcal{G}]$ minimizes $\mathbb{E}[(X - Y)^2]$ over all $\mathcal{G}$-measurable Y.

Lemma 1.9. *Let $X, Y : \Omega \to \mathbb{R}$ two random variables. Then it can prove that X is $\sigma(Y)$-measurable if and only if there is a measurable function $g \colon \mathbb{R} \to \mathbb{R}$ such that $X = g(Y)$.*

Proof. It's clear that $X = g(Y)$ implies X is $\sigma(Y)$-measurable. In converse, for fixed n, let

$$A_{m,n} := \left\{\omega \colon m2^{-n} \leqslant X(\omega) \leqslant (m+1)2^{-n}\right\}, \ \forall\, m \in \mathbb{Z}$$

Since X is $\sigma(Y)$-measurable, $A_{m,n} \in \sigma(Y)$. Therefore, there is a $B_{m,n} \in \mathcal{R}$ such that $A_{m,n} = X^{-1}(B_{m,n})$. Let

$$g_n(x) = \sum_m m2^{-n}\mathbb{1}_{B_{m,n}}(x)$$

First, for any x, $g_n(x)$ is monotonically increasing so $g(x) = \lim_n g_n(x)$. Second,

$$f_n(X) \leqslant Y \leqslant f_n(X) + \frac{1}{2^n}$$

Therefore, $Y = f(X)$ by taking limits. $\square$

Remark. This result is also true for multidimensional case, that is, if Z is $\sigma(X,Y)$-measurable, then there is a $g\colon \mathbb{R}^2 \to \mathbb{R}$ such that $Z = g(X,Y)$.

Therefore, if $\mathcal{G} = \sigma(Y)$, then $\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)] = g(Y)$ for some g.

Lemma 1.10. *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 $g\colon \mathbb{R}^2 \to \mathbb{R}$,*

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

- Stochastic Process: Let (S, β_S) be a measurable space and $T = \mathbb{R}_+$ or $\mathbb{Z}_+$. If for any $t \in T$, $X_t\colon \Omega \to S$ is measurable, then $X = (X_t)_{t\in T}$ is called a S-values stochastic process. And for any $\omega \in \Omega$, $X_\cdot(\omega)\colon T \to S$, *i.e.* $(X_t(\omega))_{t\in T}$ is called a sample path.

  Considering a map $p\colon S \times \beta_S \to [0,1]$ such that

  (1) fix any $A \in \beta_S$, $x \mapsto p(x,A)$ is measurable;
  (2) fix any $x \in S$, $p(x,\cdot)$ is a probability measure defined on S,

  then p is called a transition probability.

- Markov chain: Let $T = \mathbb{Z}_+$. For discrete states case, let $S = \mathbb{Z}$. A stochastic process $(X_t)_{t\in T}$ is called a Markov chain if for any $t \in \mathbb{Z}_+$, for any $x_0, x_1, \cdots, x_{t-1} \in S$ and any $A \in \beta_S$,

  (1) $\mathbb{P}(X_0 = x_0, X_1 = x_1, \cdots, X_{t-1} = x_{t-1}) > 0$;
  (2) $\mathbb{P}(X_t \in A \mid X_0 = x_0, X_1 = x_1, \cdots, X_{t-1} = x_{t-1}) = p(x_{t-1}, A)$

  where p is a transition probability. Then $(X_t)_{t\in T}$ is called a Markov chain.
  Remark. If $S = \mathbb{R}$, let $\sigma(X_0, X_1, \cdots, X_{t-1}) = \sigma(\sigma(X_0), \sigma(X_1), \cdots, \sigma(X_{t-1}))$,

  $$\mathbb{P}(X_t \in A \mid \sigma(X_0, X_1, \cdots, X_{t-1})) = \mathbb{P}(X_t \in A \mid \sigma(X_{t-1}))$$

  then $(X_t)_{t\in T}$ is called a Markov chain.
  Note that $\mathbb{P}(X_t \in A \mid \sigma(X_0, X_1, \cdots, X_{t-1})) =: \mathbb{P}(X_t \in A \mid X_0, X_1, \cdots, X_{t-1})$

  Proposition 1.11. *Let $(X_t)_{t\in T}$ be a Markov chain.*

  (1) For Y nonnegative $\sigma(X_t)$-measurable, $\mathbb{E}[Y \mid X_0, \cdots, X_{t-1}] = \mathbb{E}[Y \mid X_{t-1}]$.
  (2) If $n \in \mathbb{N}$ and $m \in \mathbb{N}_+$, $\mathbb{P}(X_{n+m} \in A \mid X_0, X_1, \cdots, X_n) = \mathbb{P}(X_{n+m} \in A \mid X_n)$.

  Let $S = \mathbb{Z}$ and $T = \mathbb{Z}_+$. Note if $\mathbb{P}(X_0 = x) = 1$, we denote $(X_t)_{t\in T}$ by $(X_t^x)_{t\in T}$. And we denote $p(x,\{y\}) = p(x,y)$. So $\mathbb{P}(X_1^x = y) = p(x,y)$. Moreover,

  $$\begin{aligned}\mathbb{P}(X_n^x = y) &= \sum_{x_1,\cdots,x_{n-1}} \mathbb{P}(X_1^x = x_1, \cdots, X_{n-1}^x = x_{n-1}, X_n^x = y)\\ &= \sum_{x_1,\cdots,x_{n-1}} p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y)\\ &=: p^{(n)}(x,y)\end{aligned}$$

  From this denotation, $p^{(n+m)}(x,y) = \sum_{x_m} p^{(n)}(x,x_m)p^{(m)}(x_m,y)$. Moreover, for $x,y \in S$, $x \to y$ means there is a $n \in \mathbb{N}$ such that $p^{(n)}(x,y) > 0$. And if for any $x,y \in S$, $x \leftrightarrow y$, then the Markov chain is irreducible.

For $x \in S$, if

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} X_n^x = x\right) = 1$$

then $x \in S$ is called recurrent. Moreover, let

$$\tau_y^x := \inf\{n \geqslant 1 \colon X_n^x = y\}$$
$$N_y^x := \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n^x = y\}}$$

Remark. Let $\tau_x^{(k)} = \inf\left\{N \geqslant 1 \colon \sum_{j=1}^{N} \mathbb{1}_{\{X_j^x = x\}} \geqslant k\right\}$ and $t_k = \tau_x^{(k)} - \tau_x^{(k-1)}$. And the Markov property guarantees $\{t_k\}_{k=1}^{\infty}$ is independent.

So x is recurrent $\Leftrightarrow \mathbb{P}(N_x^x = \infty) = 1 \Leftrightarrow \sum_{n=1}^{\infty} p^{(n)}(x,x) = \infty$, where the first equivalence is because

$$\mathbb{P}(N_x^x \geqslant n) = \mathbb{P}(\tau_x^{(n)} < \infty) = \mathbb{P}(t_1 < \infty)^n = \mathbb{P}(\tau_x^x < \infty)^n$$

and the second equivalence is because

$$\sum_{n=1}^{\infty} p^{(n)}(x,x) = \mathbb{E}[N_x^x] = \mathbb{E}[\sum_{k=1}^{\infty} \mathbb{1}_{\{N_x^x \geqslant k\}}] = \sum_{k=1}^{\infty} \mathbb{P}(\tau_x^x < \infty)^k$$

Moreover, if $x \leftrightarrow y$, then the recurrence of x implies the recurrence of y.

The average recurrent time of a state x is defined as

$$m(x) = \sum_{n=1}^{\infty} n\mathbb{P}(\tau_x^x = n), \text{ if } \mathbb{P}(\tau_x^x < \infty) = 1$$

otherwise, $m(x) = \infty$. Then x is called positively recurrence if $m(x) < \infty$ and x is called zero recurrence if x is recurrent but $m(x) = \infty$. And for $x, y \in S$,

$$\lim_n \frac{1}{n} \sum_{l=1}^{n} p^{(l)}(x,y) = \frac{\mathbb{P}(\tau_y^x < \infty)}{m(y)}$$

Let π be a distribution on S. If

$$\pi(y) = \sum_{x \in S} \pi(x) p(x,y), \ \forall\ y \in S$$

then π is called a stationary distribution. Moreover, for a distribution π_0 defined on S, let

$$\pi_n(y) = \sum_{x \in S} \pi_0(x) p^{(n)}(x,y)$$

Note that $\pi_n(y) = \sum_{z \in S} \pi_{n-1}(z) p(z,y)$. If there is a π on S such that for any π_0

$$\lim_n \pi_n(x) = \pi(x)$$

such π is called a limit distribution, which is also stationary.

Theorem 1.12. *For an irreducible Markov chain, TFAE*

(1) There is a positively recurrent x.

(2) All x is positively recurrent.

(3) There is a unique stationary distribution π and

$$\pi(x) = \lim_n \frac{1}{n} \sum_{l=1}^{n} p^{(l)}(x,x) = \frac{1}{m(x)}$$

Theorem 1.13. *The Markov chain is irreducible.*

(1) If there is a stationary distribution π with $\pi(y) > 0$, then all states are recurrent.

(2) If y is recurrent, then $\mathbb{P}(\tau_y^x < \infty) = 1$ for any $x \in S$.

- Poisson Random Measure: Let μ be a σ-finite measure defined on $(\mathbb{R}^d, \mathcal{R}^d)$. And for $E \in \mathbb{R}^d$ with $\mu(E) > 0$, let $\nu_E(A) := \mu(A \cap E)/\mu(E)$ for any $A \in \mathcal{R}^d$ is a probability measure.

  The Poisson random measure is a family of random variables $X = (X(A))_{A \in \mathcal{R}^d}$ (or $X \colon \mathcal{R}^d \times \Omega \to \mathbb{R}$) such that

  (1) For *a.e.* $\omega \in \Omega$, $X(\cdot, \omega)$ is a σ-finite measure on $\mathbb{R}^d$;

  (2) For any $E \in \mathbb{R}^d$, $X(E)$ is a random variable obeying the Poisson distribution with mean $\mu(E)$;

  (3) For disjoint $E_1, E_2, \cdots, E_n \in \mathbb{R}^d$, $X(E_1), X(E_2), \cdots, X(E_n)$ are independent.

  Existence of $X = (X(A))_{A \in \mathcal{R}^d}$. First, let $\{S_n\}_{n=1}^{\infty} \subset \mathcal{R}^d$ be disjoint partition of $\mathbb{R}^d$ such that $\mu(S_n) < \infty$.

  I. For any n, let $X(S_n) \sim P_o(\mu(S_n))$. And $\{X(S_n)\}_n$ is independent.

  II. Fix S_n, let $\{E_j\}_{j=1}^{\infty}$ be a disjoint partition of S_n. For any $k = \sum_{j=1}^{m} k_j$,

  $$\mathbb{P}(X(E_1) = k_1, \cdots, X(E_m) = k_m \mid X(S_n) = k) := \frac{k!}{k_1! k_2! \cdots k_m!} \prod_{j=1}^{m} \left(\frac{\mu(E_j)}{\mu(S_n)} \right)^{k_j}$$

  Based on this, it is not hard to see

  $$\mathbb{P}(X(E_1) = k_1, \cdots, X(E_m) = k_m) = \prod_{j=1}^{m} e^{-\mu(E_j)} \frac{\mu(E_j)^{k_j}}{k_j!}$$

  So $X(E_j) \sim P_o(\mu(E_j))$ and they are independent. Therefore, for any $A \in \mathcal{R}^d$ and $A \in S_n$, then $\{A, A^c\}$ is a partition so that

  $$\mathbb{P}(X(A) = k_1 \mid X(S_n) = k) = \binom{k}{k_1} p^{k_1} (1-p)^{1-k_1} \quad \Rightarrow \quad \mathbb{P}(X(A) = k_1) = e^{-p} \frac{p^{k_1}}{k_1!}$$

  where $p = \mu(A)/\mu(S_n)$.
  Or equivalently let $N \sim P_o(\mu(S_n))$ and $\{X_j\}_{j=1}^{\infty}$ independent with $\mathbb{P}(X_j \in A) = \mu(A)/\mu(S_n)$ for any $A \subset S_n$. Then define

  $$X(A) := \sum_{j=1}^{N} \mathbb{1}_A(X_j), \ \forall \ A \subset S_n$$

III. Finally, for any $A \in \mathcal{R}^d$,

$$X(A) := \sum_{n=1}^{\infty} X(A \cap S_n)$$

- Poisson Process: Let $Y = (Y(A))_{A \in \mathcal{R}}$ be a Poisson random measure defined on $(\mathbb{R}, \mathcal{R}, \mu)$. Define

$$X_t(\omega) := Y([0,t])(\omega), \ \forall \ t \geqslant 0$$

Then $X = (X_t)_{t \geqslant 0}$ is called a Poisson process. In particular, if $\mu([a,b]) = \lambda(b-a)$, then X is called a Poisson process with intensity λ.

Theorem 1.14. *Let X be a Poisson process with intensity λ.*

(1) $X_0 = 0$ a.e.

(2) For any $0 \leqslant t_0 < t_1 < t_2 < \cdots < t_n$,

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$$

are independent (such property is called independent increments).

(3) For $t > s$, $X_t - X_s \sim P_o(\lambda(t-s))$.

(4) For almost every $\omega \in \Omega$, $t \mapsto X_t(\omega)$ is right continuous, monotonously increasing and has the left limit.

Remark. If a stochastic process $X = (X_t)_{t \geqslant 0}$ just satisfies (4) except for monotonously increasing, it is called càdlàg or RCLL. If a stochastic process $X = (X_t)_{t \geqslant 0}$ satisfies (1)(2) and (4) except for monotonously increasing and two more properties: $X_t - X_s$ has the same distribution as X_{t-s} for any $t > s$; $X_s \xrightarrow{\mathbb{P}} X_t$ for any $t \geqslant 0$, then X is called a Lévy process.

Theorem 1.15. *Let X be a Poisson process with intensity λ and*

$$\tau_i := \inf\{t \geqslant 0 \colon X_t \geqslant i\}$$

Then $\tau_1, \tau_2 - \tau_1, \cdots$ are i.i.d *with the exponential distribution* $\exp(\lambda)$.

- Brownian Motion: A stochastic process $X = (X_t)_{t \geqslant 0}$ $(T = \mathbb{R}_+ \cup \{0\}, S = \mathbb{R})$ is called a Gaussian process if for any $0 < t_1 < t_2 < \cdots < t_n$,

$$(X_{t_1}, X_{t_2}, \cdots, X_{t_n}) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

The (standard) Brownian motion or Wiener process is a stochastic process $B = (B_t)_{t \geqslant 0}$ $(T = \mathbb{R}_+ \cup \{0\}, S = \mathbb{R})$ such that

(1) B is a Gaussian process;

(2) $\forall \ t, s \in [0, \infty)$, $\mathbb{E}[B_t] = 0$ and $\mathbb{E}[B_t B_s] = t \wedge s$;

(3) For *a.e.* $\omega \in \Omega$, the map $t \mapsto B_t(\omega)$ is continuous.

Corollary 1.16. *Let $B = (B_t)_{t \geqslant 0}$ be a Brownian motion.*

I. $B_0 = 0$ a.e..

II. B is independent increments.

III. $\forall \ t \geqslant 0$ and $\forall \ s > 0$, $B_{t+s} - B_t \sim \mathcal{N}(0, s)$.

Note that it can also define the Brownian motion by $I - III$ and (3).

Construction of B: Let $X_1, X_2, \cdots$ be *i.i.d.* and $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. Define

$$B_t^{(n)} := \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{\lfloor nt \rfloor} X_k + (t - \frac{\lfloor nt \rfloor}{n}) X_{\lfloor nt \rfloor + 1} \right)$$

Then $B_t^{(n)} \xrightarrow{d} B_t$, which is a Brownian motion.

Construction by RKHS: Let $K \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a positive semi-definite kernel and $\mathcal{H}$ be the corresponding reproducing kernel Hilbert space with induced inner product $\langle \cdot, \cdot \rangle_K$. Then let $\{e_j\}_{j=1}^{\infty}$ be an orthonormal basis.

$$K(t,s) = \sum_{j=1}^{\infty} \langle K(t,\cdot), e_j \rangle_K \, e_j(s) = \sum_{j=1}^{\infty} e_j(t) e_j(s)$$

Let $\{Z_j\}_{j=1}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and define

$$X_t = \sum_{j=1}^{\infty} \langle K(t,\cdot), e_j \rangle_K \, Z_j$$

Then $(X_t)_{t \geqslant 0}$ is a Gaussian process with mean 0 and $\operatorname{cov}(X_t, X_s) = \mathbb{E}[X_t X_s] = K(t,s)$. Therefore, let

$$K(t,s) := t \wedge s$$

So it is positive semi-definite. And then such $(X_t)_{t \geqslant 0}$ is a Brownian motion.

Corollary 1.17. *Let $B = (B_t)_{t \geqslant 0}$ be a Brownian motion. The following are all Brownian motions.*

(1) $X_t = -B_t, \ t \geqslant 0$;

(2) $X_t = B_{t+c} - B_c$ *for* $c > 0, \ t \geqslant 0$;

(3) $X_t = \sqrt{c} B_{t/c}, \ t \geqslant 0$;

(4) $X_0 = 0$ *and* $X_t = t B_{\frac{1}{t}}$ *for* $t > 0$;

(5) $X_t = B_1 - B_{1-t}, \ t \in [0,1]$

If $X = (X_t)_{t \geqslant 0}$ is a Gaussian process with mean 0 and

$$\operatorname{cov}(X_t, X_s) = \frac{1}{2} \left(s^{2H} + t^{2H} - |t-s|^{2H} \right)$$

then X is called a fractional Brownian motion. When $H = \frac{1}{2}$, it is a Brownian motion.

Theorem 1.18 (Lévy-Itô Decomposition)**.** *Let $X = (X_t)_{t \geqslant 0}$ be a Lévy process. There is a $c \in \mathbb{R}$ and $\sigma > 0$ and a Poison random measure p and a Brownian motion $(B_t)_{t \geqslant 0}$ such that*

$$X_t = ct + \sigma B_t + \lim_{K \to \infty} \left\{ \iint_{D_K} u p(dsdu) + \iint_{D_K} \frac{u}{1+u^2} \mu(dsdu) \right\}$$

where $D_K = \left\{ (s,u) \colon 0 \leqslant s \leqslant t, |u| \geqslant \frac{1}{K} \right\}$

- Stochastic Integral: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For $(\mathcal{F}_t)_{t\geqslant 0}$, if $\mathcal{F}_t \subset \mathcal{F}$ for all t and $\mathcal{F}_s \subset \mathcal{F}_t$ for any $0 \leqslant s < t$, then $\mathcal{F}_t$ is called a filtration. For a filtration $(\mathcal{F}_t)$ and a stochastic process $(X_t)_{t\geqslant 0}$, $(X_t)_{t\geqslant 0}$ is called $(\mathcal{F}_t)_t$-adapted if X_t is $\mathcal{F}_t$-measurable for any t.

  Consider a stochastic process $f = (f_t)_{t\in[0,T]}$,

  (1) $\forall\, t \in [0,T]$, the map

  $$\begin{array}{ccc} [0,T]\times\Omega & \longrightarrow & \mathbb{R} \\ (s,\ \omega) & \mapsto & f_s(\omega) \end{array}$$

  is $\mathcal{B}([0,T]) \otimes \mathcal{F}$-measurable.

  Remark. This condition can be relaxed to $(\mathcal{F}_t)_{t\geqslant 0}$-predictable process $f = (f_t)$. First, the $(\mathcal{F}_t)_{t\geqslant 0}$-predictable σ-algebra $\mathcal{P}$ is the smallest σ-algebra on $[0,T]\times\Omega$ such that all simple processes and their L^2 limits are measurable. Then a stochastic process is called predictable if it is measurable with respect to $\mathcal{P}$.

  (2) and

  $$\|f\|_{\mathcal{L}^2} := \sqrt{\mathbb{E}\left[\int_0^T f_t^2 dt\right]} < \infty$$

  Then we say $f = (f_t)_{t\in[0,T]} \in \mathcal{L}^2([0,T])$ (Similarly, it can define $\mathcal{L}^p([0,T])$).

  Then define the stochastic integral. Let $B = (B_t)_{t\geqslant 0}$ be a Brownian motion. If it is $\mathcal{F}_t$-adopted and $B_t - B_s$ is independent with $\mathcal{F}_s$ for any $0 \leqslant s < t$, $B = (B_t)_{t\geqslant 0}$ is called a $(\mathcal{F}_t)_t$-adapted Brownian motion. (In particular, $\mathcal{F}_t = \sigma(\{B_s \colon s \in [0,t]\})$). Consider so called simple process, for $0 = t_0 < t_1 < \cdots < t_n = T$,

  $$f_t(\omega) = \sum_{j=1}^n e_j(\omega)\mathbb{1}_{[t_{j-1},t_j]}(t)$$

  where e_j is $\mathcal{F}_{t_{j-1}}$-measurable and $e_j \in L^2$. Then define

  $$\int_0^T f_t(\omega)dB_t := \sum_{j=1}^n e_j(\omega)\left(B_j(\omega) - B_{j-1}(\omega)\right)$$

  and clearly it is $\mathcal{F}_T$-measurable.

  Theorem 1.19 (Itô Isometry). *Let $f = (f_t)_{t\in[0,T]}$ be a simple process.*

  $$\mathbb{E}\left[\left(\int_0^T f_t(\omega)dB_t\right)^2\right] = \|f\|_{\mathcal{L}^2}^2 = \mathbb{E}\left[\int_0^T f_t^2 dt\right]$$

  i.e. *the map $f \mapsto \int_0^T f_t dB_t$ from $\mathcal{L}^2([0,T])$ to $L^2(\Omega)$ is an isometry map.*

  Theorem 1.20 (Dense). *For any $f \in \mathcal{L}^2([0,T])$, there is a sequence of simple process $(f^n)_{n=1}^\infty$ such that*

  $$\lim_n \left\|f - f^{(n)}\right\|_{\mathcal{L}^2}^2 = 0$$

  So in general, for any $f \in \mathcal{L}^2([0,T])$, choose any simple process sequence $\left\{f^{(n)}\right\}_{n=1}^\infty$ convergent to f in $\|\cdot\|_{\mathcal{L}^2}$ so it is Cauchy in this norm. Because of the Itô Isometry, the sequence

  $$\left\{\int_0^T f_t^{(n)}(\omega)dB_t\right\}_{n=1}^\infty$$

is Cauchy in $L^2(\Omega)$. Therefore, there is a limit point and it is defined as

$$\int_0^T f_t dB_t := \lim_n \int_0^T f_t^{(n)} dB_t$$

and also by the isometry property it is independent with the choice of the sequence. Besides, because $\int_0^T f_t^{(n)} dB_t$ converges to $\int_0^T f_t dB_t$ in L^2, it is also convergent pointwisely almost everywhere. So $\int_0^T f_t dB_t$ is $\mathcal{F}_T$-measurable. And the isometry property can be also extended

$$\mathbb{E}\left[\left(\int_0^T f_t(\omega) dB_t\right)^2\right] = \|f\|_{\mathcal{L}^2}^2, \ \forall\, f \in \mathcal{L}^2([0,T])$$

Theorem 1.21 (Properties of Integral)**.** *For $f, g \in \mathcal{L}^2([0,T])$ and $\alpha, \beta \in \mathbb{R}$,*

(1) $\mathbb{E}\left[\int_0^T f_t dB_t\right] = 0$*;*

(2) $\int_0^T (\alpha f_t + \beta g_t) dB_t = \alpha \int_0^T f_t dB_t + \beta \int_0^T f_t dB_t$*.*

Remark. When $f \in \mathcal{L}^2([0,T])$,

$$\int_0^T f(t) dB_t \overset{d}{=} \mathcal{N}\left(0, \int_0^T f(t)^2 dt\right)$$

Considering two stochastic process $X = (X_t)$ and $Y = (Y_t)$, Y is called a modification of X if for any t,

$$\mathbb{P}(X_t = Y_t) = 1$$

If $t \mapsto Y_t(\omega)$ is continuous *a.e.*, Y is called a continuous modification.

Remark. Let $f \in \mathcal{L}^2([0,T])$. For $t \in [0,T]$,

$$\mathbb{P}\left(X_t = \int_0^t f_s dB_s\right) = 1$$

X_t may not be continuous but $\int_0^t f_s dB_s$ has the continuous path.

- Discrete Martingale: Let $X = (X_n)_{n=0}^\infty$ be a stochastic process and $\mathbb{F} = (\mathcal{F}_n)_{n=0}^\infty$ be a filtration. X is called a $\mathbb{F}$-martingale if

  (1) $X_n \in L^1$ for all n;

  (2) X is $\mathbb{F}$-adopted;

  (3) $\forall\, n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m$ *a.e.*.

  Replacing (3) in the martingale by

  (1) $\forall\, n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] \geqslant X_m$ *a.e.*, then it is called submartingale.

  (2) $\forall\, n > m$, $\mathbb{E}[X_n \mid \mathcal{F}_m] \leqslant X_m$ *a.e.*, then it is called supermartingale.

  Let $\mathbb{F} = (\mathcal{F}_n)_{n=0}^\infty$ be a filtration. If $\tau\colon \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ satisfies

  $$\forall\, n \in \mathbb{Z}_+,\ \{\tau = n\} \in \mathcal{F}_n$$

  then τ is called a stopping time.

Theorem 1.22 (Bounded Optional Stopping Theorem). *Let $X = (X_n)_{n=0}^{\infty}$ be a submartingale and τ be a stopping time for the filtration $\mathbb{F} = (\mathcal{F}_n)_{n=0}^{\infty}$. If there is a $n \in \mathbb{N}$ such that $\tau(\omega) \leqslant n$ a.e., then*

$$\mathbb{E}[X_0] \leqslant \mathbb{E}[X_\tau] \leqslant \mathbb{E}[X_n]$$

where $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$. In particular, if X is a martingale, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

Remark. For general stopping time T, let $\tau = T \wedge n$. So τ is a bounded stopping time and thus the above theorem is called the bounded optional stopping theorem. The general version of stopping theorem is in the following.

Proof. For X_τ,

$$\mathbb{E}[X_\tau] = \mathbb{E}\left[X_\tau \sum_{i=1}^{n} \mathbb{1}_{\{\tau=i\}}\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_\tau \mathbb{1}_{\{\tau=i\}}\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{\tau=i\}}\right]$$

Because X is a submartingale, *i.e.* $X_i \leqslant \mathbb{E}[X_n \mid \mathcal{F}_i]$ for $i \leqslant n$,

$$\begin{aligned}
\mathbb{E}[X_\tau] &= \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{\tau=i\}}\right] \\
&\leqslant \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}[X_n \mid \mathcal{F}_i], \{\tau = i\}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_n, \{\tau = i\}] \\
&= \mathbb{E}[X_n]
\end{aligned}$$

And by

$$\{\tau \geqslant k\} = \{\tau \leqslant k-1\}^c = \left(\bigcup_{i=0}^{k-1} \{\tau = i\}\right)^c \in \mathcal{F}_{k-1}$$

we have

$$\begin{aligned}
\mathbb{E}[X_0] &= \mathbb{E}[X_0, \{\tau = 0\}] + \mathbb{E}[X_0, \{\tau \geqslant 1\}] \\
&\leqslant \mathbb{E}[X_0, \{\tau = 0\}] + \mathbb{E}\left[\mathbb{E}[X_1 \mid \mathcal{F}_0], \{\tau \geqslant 1\}\right] \\
&= \mathbb{E}[X_0, \{\tau = 0\}] + \mathbb{E}[X_1, \{\tau \geqslant 1\}] \\
&= \mathbb{E}[X_0, \{\tau = 0\}] + \mathbb{E}[X_1, \{\tau = 1\}] + \mathbb{E}[X_1, \{\tau \geqslant 2\}] \\
&\leqslant \mathbb{E}[X_0, \{\tau = 0\}] + \mathbb{E}[X_1, \{\tau = 1\}] + \mathbb{E}\left[\mathbb{E}[X_2 \mid \mathcal{F}_1], \{\tau \geqslant 2\}\right] \\
&= \cdots \\
&\leqslant \sum_{i=0}^{n} \mathbb{E}[X_i, \{\tau = i\}] \\
&= \mathbb{E}[X_\tau]
\end{aligned}$$

If $(X_n)_{n=0}^{\infty}$ is a martingale, then $(-X_n)_{n=0}^{\infty}$ is also a martingale, *i.e.* submartingale, so

$$\mathbb{E}[X_0] = \mathbb{E}[X_\tau] = \mathbb{E}[X_n]$$

$\square$

Theorem 1.23 (Doob's Inequality). *$(X_n)_{n=0}^\infty$ is a submartingale and $a > 0$.*

$$\mathbb{P}\left(\max_{0\leqslant i\leqslant n} X_i \geqslant a\right) \leqslant \frac{\mathbb{E}[X_n^+]}{a}$$

where $X_n^+ = \max\{X_n, 0\}$ (and $X_n^- = \max\{-X_n, 0\}$).

Proof. Let

$$\tau = \begin{cases} \min\{0 \leqslant i \leqslant n \colon X_i \geqslant a\}, & \exists\, i,\ X_i \geqslant a \\ n, & \text{otherwise} \end{cases}$$

Then $\tau \leqslant n$ is a stopping time. Let $A = \{\max_{0\leqslant i\leqslant n} X_i \geqslant a\} = \{X_\tau \geqslant a\}$. By the Optional Stopping Theorem,

$$\mathbb{E}[X_n] \geqslant \mathbb{E}[X_\tau] = \mathbb{E}[X_\tau \mathbb{1}_A] + \mathbb{E}[X_\tau \mathbb{1}_{A^c}] = \mathbb{E}[X_\tau \mathbb{1}_A] + \mathbb{E}[X_n \mathbb{1}_{A^c}] \geqslant a\mathbb{P}(A) + \mathbb{E}[X_n \mathbb{1}_{A^c}]$$

So

$$\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_n] - \mathbb{E}[X_n \mathbb{1}_{A^c}] \geqslant \mathbb{E}[X_\tau \mathbb{1}_A] \geqslant a\mathbb{P}(A)$$

And because $X_n^+ \geqslant X_n \mathbb{1}_A$,

$$\mathbb{P}(A) \leqslant \frac{1}{a}\mathbb{E}[X_\tau \mathbb{1}_A] \leqslant \frac{\mathbb{E}[X_n^+]}{a}$$

$\square$

And by the Jensen's inequality of the conditional expectations, if $(X_n)_{n=0}^\infty$ is a martingale and $\varphi\colon \mathbb{R} \to \mathbb{R}$ is convex and $\varphi(X_n) \in L^1$, the $(\varphi(X_n))_{n=0}^\infty$ is a submartingale. It implies

$$\mathbb{P}\left(\max_{0\leqslant i\leqslant n} |X_i| \geqslant a\right) = \mathbb{P}\left(\max_{0\leqslant i\leqslant n} |X_i|^p \geqslant a^p\right) \leqslant \frac{\mathbb{E}[|X_n|^p]}{a^p}$$

Example 1.24 (Kolmogorov's Inequality). Let $\{\xi_n\}_{n\in\mathbb{N}}$ be *i.i.d.* with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = \sigma^2$. Then let $X_0 = 0$ and $X_n = \sum_{k=1}^n \xi_k$. So $(X_n)_{n\in\mathbb{N}_0}$ is a martingale and

$$\mathbb{P}\left(\sup_{m\leqslant n} |X_m| \geqslant \lambda\sqrt{n}\right) \leqslant \frac{\sigma^2}{\lambda^2}$$

For $n \in \mathbb{N}$, $(\mathcal{F}_{m,n})_{m=0}^n$ is a filtration. And if $X_{m,n} \in L^1$ and $\mathcal{F}_{m,n}$-measurable for $1 \leqslant m \leqslant n$ and $\mathbb{E}[X_{m,n} \mid \mathcal{F}_{m-1,n}] = 0$, then $(X_{m,n})_{m,n}$ is called a martingale difference array. Then

$$\left(\sum_{1\leqslant l\leqslant m} X_{l,n}\right)_{m=0}^n$$

is $(\mathcal{F}_{m,n})_{m=0}^n$-martingale.

Theorem 1.25 (Martingale CLT). *Let $(X_{m,n})_{m,n}$ be a martingale difference array. Assume that*

(1) $\sum_{m=0}^n \mathbb{E}\left[X_{m,n}^2 \mid \mathcal{F}_{m-1,n}\right] \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$.

(2) $\varepsilon > 0$, $\sum_{m=0}^n \mathbb{E}\left[X_{m,n}^2 \mathbb{1}_{\{|X_{m,n}|>\varepsilon\}} \mid \mathcal{F}_{m-1,n}\right] \xrightarrow{\mathbb{P}} 1$

Then

$$\sum_{m=0}^n X_{n,m} \xrightarrow{d} \mathcal{N}(0,1), \ \text{as } n \to \infty$$

- Continuous Martingale: The definition is to replace $(X_n)_{n=0}^{\infty}$ and $(\mathcal{F}_n)_{n=0}^{\infty}$ by $(X_t)_{t\geqslant 0}$ and $(\mathcal{F}_t)_{t\geqslant 0}$, and $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ for $t > s$. Similarly, it is for submartingale and supermartingale. And also it is true for the Optional Stopping Theorem ($\{\tau \leqslant t\} \in \mathcal{F}_t$) and Doob's Inequality.

  Theorem 1.26. *Let $(B_t)_{t\geqslant 0}$ be a Brownian motion.*

  (1) $B = (B_t)_{t\geqslant 0}$ is a $\mathcal{F}_t = \sigma(\{B_s : s \in [0,t]\})$-martingale.
  (2) For $f \in \mathcal{L}^2([0,T])$, let $X_t = \int_0^t f_s dB_s$. Then $(X_t)_{t\in[0,T]}$ is a martingale with respect to $(\mathcal{F}_t)_{t\in[0,T]}$.

  Proof. For (1), if $t > s$, then $B_t - B_s$ is independent with $\mathcal{F}_s$, so

  $$\mathbb{E}[B_t \mid \mathcal{F}_s] = \mathbb{E}[B_s + B_t - B_s \mid \mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s$$

  For (2), it is equivalent to prove $\mathbb{E}[B_t - B_s \mid \mathcal{F}_s] = 0$ for $t > s$. By the definition of conditional expectation (uniqueness), it is necessary to show $\mathbb{E}[B_t - B_s, A] = 0$ for any $A \in \mathcal{F}_s$. Let $X_t^{(n)}$ and $X_s^{(n)}$ be sequences defined by integral of simple processes and they are convergent to X_t and X_s in L^2 respectively.

  $$\mathbb{E}[X_t - X_s, A] = \mathbb{E}[X_t - X_t^{(n)}, A] - \mathbb{E}[X_s - X_s^{(n)}, A] + \mathbb{E}[X_t^{(n)} - X_s^{(n)}, A]$$

  The first two terms converge to 0 because of the L^2 convergence, and the third term converges to 0 is because of (1). □

  Remark. The converse is the Martingale Representation Theorem. If $X_t \in L^2$ is a $\mathcal{F}_t$-martingale, then there is a $f \in \mathcal{L}^2([0,T])$ such that

  $$X_t = X_0 + \int_0^t f_s dB_s$$

- Itô Formula: Let $(\mathcal{F}_t)_{t\geqslant 0}$ be a filtration and $B = (B_t)_{t\geqslant 0}$ be a $(\mathcal{F}_t)_{t\geqslant 0}$-Brownian motion. Let $v = (v_t)_{t\in[0,T]} \in \mathcal{L}^2([0,T])$ and $u = (u_t)_{t\in[0,T]} \in \mathcal{L}([0,T])$.

  $$X_t = X_0 + \int_0^t u_s ds + \int_0^t v_s dB_s$$

  Then $(X_t)_{t\in[0,T]}$ is called an Itô process. Formally,

  $$dX_t = u_t dt + v_t dB_t$$

  Theorem 1.27 (Itô's Formula). *Let $(X_t)_{t\in[0,T]}$ be an Itô process and $\varphi(t,x) \in \mathcal{C}^{1,2}([0,\infty) \times \mathbb{R})$. If*

  $$\left(\frac{\partial\varphi}{\partial x}(t, X_t) v_t\right)_{t\in[0,T]} \in \mathcal{L}^2([0,T])$$

  then the Itô's formula is

  $$\begin{aligned}\varphi(T, X_T) = \varphi(0, X_0) &+ \int_0^T \frac{\partial\varphi}{\partial t}(t, X_t)dt + \int_0^T \frac{\partial\varphi}{\partial x}(t, X_t)u_t dt \\ &+ \int_0^T \frac{\partial\varphi}{\partial x}(t, X_t) v_t dB_t + \frac{1}{2}\int_0^T \frac{\partial^2\varphi}{\partial x^2}(t, X_t) v_t^2 dt\end{aligned}$$

  Or formally,

  $$\begin{aligned}d\varphi(t, X_t) &= \frac{\partial\varphi}{\partial t}dt + \frac{\partial\varphi}{\partial x}dX_t + \frac{v_t^2}{2}\frac{\partial^2\varphi}{\partial x^2}(t, X_t)dt \\ &= \left(\frac{\partial\varphi}{\partial t} + u_t\frac{\partial\varphi}{\partial x} + \frac{v_t^2}{2}\frac{\partial^2\varphi}{\partial x^2}(t, X_t)\right)dt + v_t\frac{\partial\varphi}{\partial x}(t, X_t)dB_t\end{aligned}$$

Remark. This is intuitively because the Taylor expansion of φ and omit the higher order terms except for dX_t and dt. Here, $\frac{v_t^2}{2}\frac{\partial^2\varphi}{\partial x^2}(t, X_t)dt$ is because $\mathbb{E}[dB_t^2] = dt$.

Example 1.28 (Black–Scholes Model)**.** Let $X_t = \mu t + \sigma B_t$ or equivalently $dX_t = \mu dt + \sigma dB_t$. And let $\varphi(t, x) = ce^x$, then

$$\frac{\partial\varphi}{\partial x} = \frac{\partial^2\varphi}{\partial x^2} = \varphi,\ \frac{\partial\varphi}{\partial t} = 0$$

So by the Itô's formula,

$$d\varphi(t, X_t) = \left(\mu\varphi(t, X_t) + \frac{\sigma^2}{2}\varphi(t, X_t)\right) dt + \sigma\varphi(t, X_t)dB_t$$

Therefore if let $Y_t = c\exp(\mu t + \sigma B_t)$, then we have

$$dY_t = \left(\mu Y_t + \frac{\sigma^2}{2}Y_t\right) dt + \sigma Y_t dB_t$$

or equivalently

$$Y_t = c + \left(\mu + \frac{\sigma^2}{2}\right)\int_0^t Y_s ds + \sigma\int_0^t Y_s dB_s$$

So we can see that Y_t is a martingale $\Leftrightarrow \mu + \frac{\sigma^2}{2} = 0$.

Theorem 1.29 (High Dimensional Itô's Formula)**.** *Let $m, d \in \mathbb{N}$. Let $\boldsymbol{B} = (\boldsymbol{B}_t)_{t\geqslant 0}$, where $\boldsymbol{B}_t = \left(B_t^{(1)}, B_t^{(2)}, \cdots, B_t^{(m)}\right)$, be a m-dimensional Brownian motion. Let*

$$\boldsymbol{u} = (u_t^{(1)}, \cdots, u^{(d)}(t)) \in \mathcal{L}([0, T]^d),\ V = \left[V_t^{(ij)}\right]_{d\times m} \in \mathcal{L}^2([0, T]^{d\times m})$$

Let $\mathbb{R}^d$-valued stochastic process $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\geqslant 0}$ be a d-dimensional Itô process, i.e.

$$d\boldsymbol{X}_t = \boldsymbol{u}_t dt + V_t d\boldsymbol{B}_t$$

Let $\varphi(\boldsymbol{x}) \in \mathcal{C}^2(\mathbb{R}^d)$ and $J\varphi(\boldsymbol{x}_t)V_t \in \mathcal{L}^2([0, T]^m)$. Then

$$d\varphi(\boldsymbol{X}_t) = J\varphi(\boldsymbol{X}_t)d\boldsymbol{X}_t + \sum_{i,l=1}^{d}\sum_{j=1}^{m}\frac{\partial^2\varphi}{\partial x_i\partial x_l}(\boldsymbol{X}_t)V_t^{(i,j)}V_t^{(l,j)}dt$$

Remark. $\boldsymbol{B} = (\boldsymbol{B}_t)_{t\geqslant 0}$ a high dimensional Brownian motion means for any fixed $i = 1, 2, \cdots, m$, $(B_t^{(i)})_{t\geqslant 0}$ is a Brownian motion and for any $i \neq j$ and any s, t, $B_t^{(i)}$ and $B_s^{(j)}$ are independent. Or in other words, $\boldsymbol{B}_t - \boldsymbol{B}_s \sim \mathcal{N}(\boldsymbol{0}, (t-1)I_m)$.

Corollary 1.30 (Integral by Parts)**.** *Let X_t and Y_t be Itô processes.*

$$dX_t = \mu_t^x dt + \sigma_t^x dB_t,\ dY_t = \mu_t^y dt + \sigma_t^y dB_t$$

Let $\varphi(x, y) = xy$. Then by the Itô's formula,

$$X_tY_t = X_0Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t \sigma_s^x\sigma_s^y ds$$

Or equivalently,

$$\int_0^t X_s dY_s = X_tY_t - X_0Y_0 - \int_0^t Y_s dX_s - \int_0^t \sigma_s^x\sigma_s^y ds$$

- Stochastic Differential Equation: Let $\mu_t(x)$ and $\sigma_t(x)$ be measurable functions such that $b_t(X_t) \in \mathcal{L}([0,T])$ and $\sigma_t(X_t) \in \mathcal{L}^2([0,T])$. The equation

$$dX_t = \mu_t(X_t)dt + \sigma_t(X_t)dB_t$$

is called a stochastic differential equation. If there is a solution $(X_t)_{t\in[0,T]}$, it is called a diffusion process.

Example 1.31 (Ornstein-Uhlenbeck Process)**.** Let $\mu \in \mathbb{R}$ and $\sigma > 0$. Consider the SDE

$$dX_t = \mu X_t dt + \sigma dB_t$$

its solution is called a Ornstein-Uhlenbeck process. Let $\varphi(t,x) = xe^{-\mu t}$. By the Itô's formula, if X_t is a solution, then

$$\begin{aligned} e^{-\mu t}X_t &= X_0 + \int_0^t -\mu e^{-\mu s}X_s ds + \int_0^t e^{-\mu s}(\mu X_s ds + \sigma dB_s) \\ &= X_0 + \int_0^t \sigma e^{-\mu s} dB_s \end{aligned}$$

So

$$X_t = e^{\mu t}X_0 + \int_0^t \sigma e^{\mu(t-s)} dB_s$$

Theorem 1.32. *If there is a $C > 0$, $\forall\ t \in [0,T]$ and $\forall\ x \in \mathbb{R}$,*

(1) μ_t and σ_t are Lipschitz continuous with C;

(2) $|\mu_t(x)| \leqslant C(1+|x|)$ and $|\sigma_t(x)| \leqslant C(1+|x|)$, called the linear growth.

then the SDE has a unique solution.

- Relations to PDE: Considering time-independent diffusion process,

$$X_t^x = x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s,$$

where μ and σ are time-independent and satisfy (1) and (2).
Define

$$Af(x) := \lim_{\varepsilon\to 0} \frac{\mathbb{E}[f(X_\varepsilon^x)] - f(x)}{\varepsilon}$$

then A is called the generating operator for X_t^x.

Lemma 1.33. *For $f \in \mathcal{C}^2(\mathbb{R})$ and a stopping time τ such that $\mathbb{E}[\tau] < \infty$,*

$$\mathbb{E}[f(X_\tau^x)] = f(x) + \mathbb{E}\left[\int_0^\tau \mu(X_t^x)\frac{\partial f}{\partial x}(X_t^x) + \frac{1}{2}\sigma(X_t^x)^2\frac{\partial^2 f}{\partial x^2}(X_t^x)dt\right]$$

It is by the Itô's formula and the Optimal Stopping Theorem.

Theorem 1.34. *For $f \in \mathcal{C}^2(\mathbb{R})$,*

$$Af(x) = \mu(x)\frac{\partial f}{\partial x}(x) + \frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2}(x)$$

2 More Probability

- Let $(S, \mathcal{S})$ be a measurable space.

  Theorem 2.1 (Lebesgue Decomposition). *Let μ and ν be two σ-finite measures on S. The there exists a unique decomposition*

  $$\nu = \nu_a + \nu_s,$$

  *where $\nu_a \ll \mu$ and $\nu_s \perp \mu$ (*i.e. *$\exists\ D \in \mathcal{S}$ s.t. $\nu_s(D) = \mu(D^c) = 0$). Furthermore, there is a* a.e. *unique $h \in L^0_+$ such that*

  $$\nu_a = \int_A h d\mu$$

  When $\nu \ll \mu$, it is the Radon-Nikodym theorem, and denote $h = \frac{d\nu}{d\mu}$. In the following, let μ, ν, ρ be σ-finite measures on S.

  (1) If $\nu \ll \mu$ and $\rho \ll \mu$, then $\nu + \rho \ll \mu$ and

  $$\frac{d\nu}{d\mu} + \frac{d\rho}{d\mu} = \frac{d(\nu+\rho)}{d\mu}$$

  (2) If $\nu \ll \mu$ and $h \in L^0_+$, then

  $$\int h d\nu = \int g d\mu,\ g = h\frac{d\nu}{d\mu}$$

  (3) If $\nu \ll \mu$ and $\mu \ll \rho$, then $\nu \ll \rho$ and

  $$\frac{d\nu}{d\mu}\frac{d\mu}{d\rho} = \frac{d\nu}{d\rho}$$

  (4) If $\nu \sim \mu$ *i.e.* $\nu \ll \mu$ and $\mu \ll \nu$, then

  $$\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$$

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Here are some properties.

  (1) $L^q \subset L^p$ for $1 \leqslant p < q$

  Proof. For $f \in L^q$, then by Hölder's Inequality

  $$\int |f|^p \, d\mathbb{P} \leqslant \left(\int (|f|^p)^{p'} d\mathbb{P}\right)^{\frac{1}{p'}} \left(\int 1^{q'} d\mathbb{P}\right)^{\frac{1}{q'}}$$

  Let $p' = \frac{q}{p} > 1$ and q' such that $\frac{1}{p'} + \frac{1}{q'} = 1$. Then

  $$\|f\|^p_{L^p} \leqslant \|f\|^p_{L^q} < \infty$$

  So $f \in L^p$. □

  Note that it is true for any measurable space with finite measure.

  (2) About the convergence, here are their relations, where $1 \leqslant p < q$.

$$
\begin{tikzcd}
& \mathcal{L}^\infty \arrow[dl] \arrow[dd] \arrow[dr] \arrow[drr] & & \\
\text{a.s.} \arrow[dr] & & \mathcal{L}^p \arrow[dl] & \mathcal{L}^q \arrow[l] \arrow[dll] \\
& \mathbb{P} & &
\end{tikzcd}
$$

- Weak Convergence: If S is a metric space with metric d and $\mathcal{S} = \mathcal{B}(d)$, the σ-algebra generated by d, then for a sequence of probability measures $\{\mu_n\}_n$ and μ, $\mu_n \xrightarrow{w} \mu$ weakly convergent means

$$\int f d\mu_n \longrightarrow \int f d\mu, \ \forall \ f \in \mathcal{C}_b(S)$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For a random variable $X \colon \Omega \to \mathbb{R}$, the induced measure μ_X defined on $(\mathbb{R}, \mathcal{R})$ is

$$\mu_X(A) = \mathbb{P}(X \in A), \ \forall \ A \in \mathcal{R}$$

So for a sequence of random variables X_n, the distribution convergence $X_n \xrightarrow{d} X$ means $\mu_{X_n} \xrightarrow{w} \mu_X$.

Total-variation Convergence: a sequence of probability measures $\{\mu\}_n$ over $(\mathbb{R}, \mathcal{R})$ converges μ in total-variation if

$$\sup_{A \in \mathcal{R}} |\mu_n(A) - \mu(A)| \to 0, \ \text{as } n \to \infty$$

- Characteristic Function over $\mathbb{R}$: For any probability measure μ on $(\mathbb{R}, \mathcal{R})$, the characteristic function is

$$\phi_\mu(t) = \int e^{itx} \mu(dx)$$

For a random variable X, ϕ_X.

(1) If $\mu_n \xrightarrow{w} \mu$, then $\phi_{\mu_n}(t) \to \phi_\mu(t)$ for all $t \in \mathbb{R}$.

(2) Continuity Theorem: Let $\{\mu_n\}_n$ be probability measures and $\{\phi_n\}_n$ be their characteristic functions. Suppose there is a ϕ such that ϕ is continuous at $t = 0$ and $\phi_n \to \phi$ for all t. Then ϕ is a characteristic function with probability measure μ and $\mu_n \xrightarrow{w} \mu$.

(3) Inversion problem: μ is a probability measure

$$\mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) = \frac{1}{2\pi} \lim_{T \to \infty} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi_\mu(t) dt$$

(4) For two probability measure μ_1 and μ_2, $\phi_{\mu_1} = \phi_{\mu_2}$ implies $\mu_1 = \mu_2$.

(5) If $\int |\phi_\mu(t)| \, dt < \infty$, then $\mu \ll \lambda$ and

$$\frac{d\mu}{d\lambda} = f(x), \ \text{where } f(x) = \frac{1}{2\pi} \int e^{-itx} \phi_\mu(x) dt$$

where λ is the Lebesgue measure.

(6) Riemannian-Lebesgue theorem: If $\mu \ll \lambda$, then $\lim_{t\to\pm\infty} \phi_\mu(t) = 0$

- Conditional Distribution: Let $(\mathcal{B}, \|\cdot\|)$ be a Banach space and $(S, \mathcal{S})$ be measurable space. A map $\mu\colon S \to B$ is called a vectored measure if $\mu(\varnothing) = 0$ and

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n), \text{ where } A_n \in \mathcal{S} \text{ disjoint}$$

So $A \mapsto \mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}[\mathbb{1}_A \mid \mathcal{G}]$ is L^1-vectored measure.

If $(E, \mathcal{E})$ and $(S, \mathcal{S})$ be two measurable spaces. A map $v\colon E \times \mathcal{S} \to \mathbb{R}$ is called a (measurable) kernel if

(1) $x \mapsto v(x, B)$ is $\mathcal{E}$-measurable for each $B \in \mathcal{S}$;

(2) $B \mapsto v(x, B)$ is a measure on S for each $x \in E$.

For $e\colon \Omega \to S$ measurable, $\mu_{e|\mathcal{G}}\colon \Omega \times \mathcal{S} \to [0,1]$ is

$$\mu_{e|\mathcal{G}}(\omega, B) := \mathbb{P}(e \in B \mid \mathcal{G})(\omega)$$

is called the regular conditional distribution of e given $\mathcal{G}$. If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$ (or Borel space, *i.e.* isomorphic in measure meaning to $(\mathbb{R}, \mathcal{R})$), then any e introduces a regular conditional distribution. Let $\boldsymbol{X}$ be a $\mathbb{R}^n$-valued random vector and $g\colon \mathbb{R}^n \to \mathbb{R}$ such that $g(\boldsymbol{X}) \in L^1$. Then

$$\mathbb{E}\left[g(\boldsymbol{X}) \mid \mathcal{G}\right] = \int_{\mathbb{R}^n} g(\boldsymbol{x}) \mu_{\boldsymbol{X}|\mathcal{G}}(\cdot, d\boldsymbol{x})$$

More specifically,

$$\mu_{\boldsymbol{X}|\mathcal{G}}\colon \Omega \times \mathcal{R}^n \longrightarrow [0,1]$$

(1) Conditional cdf: $F_{\boldsymbol{X}|\mathcal{G}}\colon \Omega \times \mathbb{R}^n \to [0,1]$ is defined by

$$F(\omega, \boldsymbol{x}) = \mu_{\boldsymbol{X}|\mathcal{G}}(\omega, \{\boldsymbol{y}\colon \boldsymbol{y} \leqslant_n \boldsymbol{x}\})$$

(2) Conditional pdf: $f_{\boldsymbol{X}|\mathcal{G}}\colon \Omega \times \mathbb{R}^n \to [0, \infty)$ satisfies

(i) $f_{\boldsymbol{X}|\mathcal{G}}(\omega, \cdot)$ is Borel measurable for all $\omega \in \Omega$;

(ii) $f_{\boldsymbol{X}|\mathcal{G}}(\cdot, \boldsymbol{x})$ is $\mathcal{G}$-measurable for each $\boldsymbol{x} \in \mathbb{R}^n$;

(iii) $\int_A f_{\boldsymbol{X}|\mathcal{G}}(\omega, \boldsymbol{x}) d\boldsymbol{x} = \mu_{\boldsymbol{X}|\mathcal{G}}(\omega, A)$ for $A \in \mathcal{R}^n$.

(3) Conditional characteristic function: $\phi_{\boldsymbol{X}|\mathcal{G}}\colon \Omega \times \mathbb{R}^n \to \mathbb{C}$ is given by

$$\phi_{\boldsymbol{X}|\mathcal{G}}(\omega, \boldsymbol{t}) = \int_{\mathbb{R}^n} e^{i\boldsymbol{t}\cdot\boldsymbol{x}} \mu_{\boldsymbol{X}|\mathcal{G}}(\omega, d\boldsymbol{x}),\ \forall\ \boldsymbol{t} \in \mathbb{R}^n, \omega \in \Omega$$

Moreover, $\phi_{\boldsymbol{X}|\mathcal{G}}(\omega, \boldsymbol{t}) = \phi(\boldsymbol{t})$ for some $\phi\colon \mathbb{R}^n \to \mathbb{C}$ if and only if $\sigma(\boldsymbol{X})$ is independent with $\mathcal{G}$.

Theorem 2.2. *Let $\boldsymbol{X} = (\boldsymbol{X}^0, \boldsymbol{X}^u)$ be $\mathbb{R}^n$-vectored valued random variable, where $\boldsymbol{X}^0$ represents the first d components and $\boldsymbol{X}^u$ represents the last components. Let $f_{\boldsymbol{X}} : \mathbb{R}^n \to \mathbb{R}$ be the pdf of $\boldsymbol{X}$ and $\mathcal{G} = \sigma(\boldsymbol{X}^0)$. Then there is a conditional pdf*

$$f_{\boldsymbol{X}^u|\mathcal{G}}\colon \Omega \times \mathbb{R}^{n-d} \to [0, \infty)$$

of $\boldsymbol{X}^u$ given $\mathcal{G}$,

$$f_{\boldsymbol{X}^u|\mathcal{G}}\left(\omega, \boldsymbol{x}^u\right) = \begin{cases} \frac{f_{\boldsymbol{X}}(\boldsymbol{X}^o(\omega), \boldsymbol{x}^u)}{\int_{\mathbb{R}^{n-d}} f_{\boldsymbol{X}}(\boldsymbol{X}^o(\omega), \boldsymbol{y}) d\boldsymbol{y}}, & \int_{\mathbb{R}^{n-d}} f\left(\boldsymbol{X}^o, \boldsymbol{y}\right) d\boldsymbol{y} > 0, \\ f_0\left(\boldsymbol{x}^u\right), & \textit{otherwise,} \end{cases}$$

for $\boldsymbol{x} \in \mathbb{R}^{n-d}$ and $\omega \in \Omega$, where $f_0 : \mathbb{R}^{n-d} \to \mathbb{R}$ is an arbitrary density function.

Theorem 2.3. *Suppose Ω, S are complete separable metric spaces with $\mathcal{F} = \mathcal{B}(\Omega)$. $Y\colon \Omega \to Y$ is measurable and μ is the distribution of Y on S. Then there is a measurable kernel*

$$P\colon S \times \mathcal{F} \longrightarrow [0,1]$$

such that

$$P_x(A \backslash \{Y = x\}) = 0,\ a.e.\ x \in S$$

and for any $X \in L^1(\Omega, S)$

$$\mathbb{E}[X \mid Y](\omega) = \mathbb{E}_{Y(\omega)}[X] = \int_\Omega X(\omega) P_{x_0}(d\omega)$$

where $x_0 = Y(\omega)$ and $\mathbb{E}_x$ is the expectation w.s.t. P_x.

Remark. It is a simplified version of above results.

3 Discrete Martingale

- Uniform Integrability: A family of random variables $\mathcal{X} \subset L^0$ is said UI if

$$\lim_{K\to\infty} \sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\,\mathbb{1}_{\{|X|>K\}}\right] = 0$$

It is equivalent to L^1-bounded and uniformly absolutely continuous, *i.e.*

(1) $\exists\, C > 0$ such that $\mathbb{E}[|X|] \leqslant C$ for all $X \in \mathcal{X}$

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

$$\mathbb{P}(A) < \delta \ \Rightarrow\ \sup_{X\in\mathcal{X}} \mathbb{E}\left[|X|\,\mathbb{1}_A\right] \leqslant \varepsilon$$

Or if there is a function $\varphi\colon [0,\infty) \to [0,\infty)$ with $\lim_{x\to\infty} \varphi(x)/x = \infty$ such that

$$\sup_{X\in\mathcal{X}} \mathbb{E}\left[\varphi(|X|)\right] < \infty$$

then $\mathcal{X}$ is UI. So $\sup_{X\in\mathcal{X}} \|X\|_{L^p} < \infty$ $(p > 1)$ implies UI.

If $\{\mathcal{X}_n\}_{n\in\mathbb{N}_0} \subset L^p$ $(p \geqslant 1)$ and $X \in L^0$ with $X_n \xrightarrow{\mathbb{P}} X$, then

$$\{\mathcal{X}_n\}_{n\in\mathbb{N}_0} \text{ UI} \ \Leftrightarrow\ X_n \xrightarrow{L^p} X \ \Leftrightarrow\ \|X_n\|_{L^p} \to \|X\| < \infty$$

- Discrete Martingale: Fix a discrete filtered probability $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$. Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a discrete stochastic process.

(1) X is adapted if X_n is $\mathcal{F}_n$-measurable for $n \in \mathbb{N}_0$.

(2) X is predicted if X_n is $\mathcal{F}_{n-1}$-measurable for $n \in \mathbb{N}$.

Let X and H be two discrete stochastic processes. $((H\cdot X)_n)_{n\in\mathbb{N}_0}$ is defined by

$$(H\cdot X)_0 = 0,\ (H\cdot X)_n := \sum_{k=1}^{n} H_k(X_k - X_{k-1}),\ \forall\, n \in \mathbb{N}$$

called the martingale transform of X by H. Moreover, if H is predictable and $H_n(X_n - X_{n-1}) \in L^1$, then X (sub)-martingale implies $H\cdot X$ (sub)-martingale. Note that it is the discrete version of stochastic integral.

There are some properties of stopping times.

(1) A random time T is stopping time if and only if the process $X_n = \mathbb{1}_{\{n \geqslant T\}}$ is adapted.

(2) If S and T are stopping times, then $S + T$ and $\max S, T$ and $\min S, T$ are stopping times.

(3) If $\{T_n\}_{n\in\mathbb{N}_0}$ is a sequence of stopping times with $T_1 \leqslant T_2 \leqslant \cdots$ *a.s.*, then $T = \sup_n T_n$ is also a stopping time. (Similarly for decreasing and inf).

Example 3.1 (Stopping Times)**.**

For a fixed $m \in \mathbb{N}_0 \cup \{\infty\}$, let $T = m$ be the constant random time so it is clear a stopping time.

Let $(X_n)_{n\in\mathbb{N}_0}$ be a $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$-adopted process, then for a fixed Borel set B

$$T_B := \min \{n \in \mathbb{N}_0 \colon X_n \in B\}$$

is a stopping time because

$$\{T_B \leqslant n\} = \{X_0 \in B\} \cup \{X_1 \in B\} \cup \cdots \cup \{X_n \in B\}$$

Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a stochastic process and T be a stopping time. The process $X^T = (X_n^T)_{n\in\mathbb{N}_0}$ is called X stopped at T, defined as

$$X_n^T(\omega) := X_{T(\omega)\wedge n} = X_n(\omega)\mathbb{1}_{\{n\leqslant T(\omega)\}} + X_{T(\omega)}\mathbb{1}_{\{n>T(\omega)\}}$$

Note that X (sub)martingale implies X^T (sub)martingale.

Theorem 3.2 (Martingale Convergence)**.** *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a martingale with*

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

Then there is a random variable $X^* \in L^1$ such that $X_n \to X$ a.e..

Proposition 3.3 (Doob-Meyer Decomposition)**.** *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a submartingale. There exists a martingale M and a predictable A with $A_0 = 0$ such that $A_n \in L^1$ and $A_n \leqslant A_{n+1}$ and*

$$X_n = M_n + A_n$$

Here

$$A_n := \sum_{k=1}^{n} \mathbb{E}\left[X_k - X_{k-1} \mid \mathcal{F}_{k-1}\right],\ n \in \mathbb{N}$$

and $M_n = X_n - A_n$. Predictability of A_n is clear and $A_{n+1} \geqslant A_n$ is by the submartingale property of X_n. And $\mathbb{E}[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] = 0$ by simply calculating.

Then for a submartingale with $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$, then there is a $X^* \in L^1$ such that $X_n \to X^*$ *a.e.*. Moreover, if (sub)martingale $\{X_n\}_{n\in\mathbb{N}_0}$ is UI, then $X_n \xrightarrow{L^1} X^*$ because $X_n \in L^1$ and $X_n \to X^*$ *a.e* implies $X_n \xrightarrow{\mathbb{P}} X^*$.

A martingale X is called a Lévy martingale if there is a random variable $\tilde{X} \in L^1$ such that $X_n = \mathbb{E}[\tilde{X} \mid \mathcal{F}_n]$ for every $n \in \mathbb{N}_0$. A martingale X is Lévy if and only if it is UI. And thus it has a L^1-limit, that is $\mathbb{E}[\tilde{X} \mid \mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

- Backward Martingale: Let $-\mathbb{N}_0 = \{\cdots, -2, -1, 0\}$. Filtration $(\mathcal{F}_n)_{n\in-\mathbb{N}_0}$ is σ-algebras with $\mathcal{F}_{n-1} \subset \mathcal{F}_n$. A stochastic process $(X_n)_{n\in-\mathbb{N}_0}$ is called backward submartingale with respect to $(\mathcal{F}_n)_{n\in-\mathbb{N}_0}$ if it is adapted and in L^1 and

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] \geqslant X_{n-1},\ \forall\ n \in -\mathbb{N}_0$$

$(X_n)_{n\in-\mathbb{N}_0}$ is called backward supermartingale if $(-X_n)_{n\in-\mathbb{N}_0}$ is a backward submartingale. When taking the equality, it is a backward martingale.

Theorem 3.4. *Suppose $(X_n)_{n\in-\mathbb{N}_0}$ is a backward submartingale such that*

$$\lim_{n\to-\infty} \mathbb{E}[X_n] > -\infty$$

Then $\{X_n\}_{n\in-\mathbb{N}_0}$ is UI and there is a $X_{-\infty} \in L^1(\cap_n \mathcal{F}_n)$ such that

$$X_n \to X_{-\infty}$$

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

$$X_{-\infty} \leqslant \mathbb{E}\left[X_m \mid \cap_n \mathcal{F}_n\right],\ \forall\ m \in \mathbb{N}_0$$

If $(X_n)_{n\in-\mathbb{N}_0}$ is a backward martingale, then

$$X_n \to X_{-\infty} = \mathbb{E}[X_0 \mid \cap_n \mathcal{F}_n]$$

$a.e$ and in L^1.

- Maximal Process: A random variable Y is called weak L^1 (wL^1) if there is $C \geqslant 0$ such that

$$\lambda\mathbb{P}(|Y| > \lambda) \leqslant C,\ \forall\ \lambda > 0$$

and the smallest C is called $\|Y\|_{wL^1}$. Note that it is not a norm. Besides, by the Markov's inequality, $L^1 \subset wL^1$.

For a stochastic process $(X_n)_{n\in\mathbb{N}_0}$, the maximal process is $(X_n^*)_{n\in\mathbb{N}_0}$ defined as

$$X_n^* = \sup_{0\leqslant m\leqslant n} |X_m|$$

and $X_\infty^* = \sup_n |X_n|$. By the Doob's Inequality, if $(X_n)_{n\in\mathbb{N}_0}$ is a martingale or non-negative submartingale, then

$$\mathbb{P}(X_n^* \geqslant \lambda) \leqslant \frac{\mathbb{E}[|X_n|\, \mathbb{1}_{\{X_n^*\geqslant\lambda\}}]}{\lambda} \leqslant \frac{\mathbb{E}[|X_n|]}{\lambda}$$

for any $n \in \mathbb{N}_0$ and $\lambda > 0$. And it implies

$$\|X_\infty^*\|_{wL^1} \leqslant \sup_{n\in\mathbb{N}_0} \|X_n\|_{L^1}$$

Theorem 3.5 (Maximal Inequality)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a martingale or non-negative submartingale, and $p \in (1, \infty)$. Then, for $n \in \mathbb{N}_0$,*

$$\|X_n^*\|_{L^p} \leqslant \frac{p}{p-1} \|X_n\|_{L^p}$$

A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ is said bounded in L^p for $p \in [1,\infty]$ if

$$\sup_n \|X_n\|_{L^p} < \infty$$

If $(X_n)_{n\in\mathbb{N}_0}$ is a martingale or non-negative submartingale, and L^p-bounded for $(1,\infty)$, then

(1) the family $\{|X_n|\}_{n\in\mathbb{N}_0}$ is UIs,
(2) there exists a $X_\infty \in L^p$ such that $X_n \to X_\infty$ in L^p.

- Stopping Time Theorems: Let T be a stopping time and $(X_n)_{n\in\mathbb{N}_0}$ be a stochastic process. When $\mathbb{P}(T=\infty)=0$,

$$X_T(\omega) := X_{T(\omega)}(\omega)$$

But if $\mathbb{P}(T=\infty)\neq 0$

$$X_T := \lim_{n\to\infty} X_{T\wedge n}$$

when the limit exists *a.e.*.

Proposition 3.6 (UI conditions)**.** *For a martingale $(X_n)_{n\in\mathbb{N}_0}$, let T be a stopping time such that $\left(X_n^T\right)_{n\in\mathbb{N}_0}$ is UI, then*

$$X_T = \lim_{n\to\infty} X_{T\wedge n} \in L^1 \ a.e.\ , \ \mathbb{E}[X_T] = \mathbb{E}[X_0] \tag{1}$$

Conversely, if $(X_n)_{n\in\mathbb{N}_0}$ is a nonnegative martingale and T is a stopping time such that (1) *holds, then $\left(X_n^T\right)_{n\in\mathbb{N}_0}$ is UI.*

In fact, for a martingale $(X_n)_{n\in\mathbb{N}_0}$ and a stopping time T, each of the following situations can make $\left(X_n^T\right)_{n\in\mathbb{N}_0}$ be UI.

I. T is bounded.
II. X is UI.
III. $\mathbb{E}[T] < \infty$ and there is a $C > 0$ such that $\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] < C$ for all n.

Proposition 3.7. *Let $(X_n)_{n\in\mathbb{N}_0}$ be a nonnegative supermartingale and T be a stopping time. Then $X_T = \lim_n X_{T\wedge n}$ is well-defined, $X_T \in L^1$ and $\mathbb{E}[X_T] \leqslant \mathbb{E}[X_0]$. If $(X_n)_{n\in\mathbb{N}_0}$ is a submartingale and $\left\{X_{T\wedge n}^+\right\}_n$ is UI, then $X_T \in L^1$ is well-defined and $\mathbb{E}[X_T] \geqslant \mathbb{E}[X_0]$.*

For a stopping time T, let

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty \colon A \cap \{T \leqslant n\} \in \mathcal{F}_n,\ \forall\ n \in \mathbb{N}_0\}$$

where $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Then $\mathcal{F}_T$ is the σ-algebra generated by

$$\mathcal{S} = \{X_T \colon X \in \mathcal{X}\}$$

where $\mathcal{X}$ is the set of all $\mathcal{F}_n$-adapted process for which X_T is well-defined.

Proposition 3.8. *Let $(X_n)_{n\in\mathbb{N}_0}$ be a UI martingale and S, T be two stopping times with $T \leqslant S$* a.e.. *Then $X_T, X_S \in L^1$ and*

$$\mathbb{E}[X_S \mid \mathcal{F}_T] = X_T,\ a.e$$

In particular,

$$\mathbb{E}[X_S] = \mathbb{E}[X_T]$$

- Square-integrability: A martingale $(X_n)_{n\in\mathbb{N}_0}$ is called a L^2-martingale if $X_n \in L^2$ for all $n \in \mathbb{N}_0$. For any $m_2 > n_2 \geqslant m_1 > n_1$,

$$\mathbb{E}[(X_{m_2} - X_{n_2})(X_{m_1} - X_{n_1}) \mid \mathcal{F}_{n_1}] = 0$$

In particular, $X_{m_2} - X_{n_2}$ is orthogonal to $X_{m_1} - X_{n_1}$ in L^2. Moreover,

$$\mathbb{E}[(X_{m_1} - X_{n_1})^2 \mid \mathcal{F}_{n_1}] = \mathbb{E}[X_{m_1}^2 - X_{n_1}^2 \mid \mathcal{F}_{n_1}]$$

L^2-martingale $(X_n)_{n\in\mathbb{N}_0}$ implies that $(X_n^2)_{n\in\mathbb{N}_0}$ is submartingale. So by the Doob-Meyer decomposition, there are a martingale M_n and a predictable process A_n that is denoted by $\langle X\rangle_n$. By the construction of A_n and the property of L^2-martingale,

$$\langle X\rangle_n - \langle X\rangle_{n-1} = \mathbb{E}\left[(X_n - X_{n-1})^2 \mid \mathcal{F}_{n-1}\right]$$

called the quadratic variation. And define $\langle X\rangle_\infty = \lim_{n\to\infty}\langle X\rangle_n \in [0,\infty]$, and it is well-defined because $\langle X\rangle_n$ is increasing. Then $(X_n)_{n\in\mathbb{N}_0}$ is L^2-bounded if and only if $\mathbb{E}[\langle X\rangle_\infty] < \infty$. In such case,

$$\mathbb{E}[(X_\infty^*)^2] \leqslant \mathbb{E}[X_0^2] + 4\mathbb{E}[\langle X\rangle_\infty]$$

Proposition 3.9. *Let $(X_n)_{n\in\mathbb{N}_0}$ be L^2 martingale. Then on $\{\langle X\rangle_\infty < \infty\}$, $\lim_n X_n$ exists.*

Proof. For $N \in \mathbb{N}$, define a random time

$$T_N = \min\left\{n \in \mathbb{N}_0 \colon \langle X\rangle_{n+1} \geqslant N\right\}$$

Because $\langle X\rangle$ is predictable, T_N is a stopping time. And since $X^2 - \langle X\rangle$ is a martingale, by the bounded optional stopping theorem

$$\mathbb{E}[X_{T_N\wedge n}^2] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X\rangle_{T_N\wedge n}]$$

And because $\langle X\rangle_n$ increasing, $\langle X\rangle_{T_N\wedge n} \leqslant N$ for any N and n. Therefore,

$$\mathbb{E}[X_{T_N\wedge n}^2] \leqslant \mathbb{E}[X_0^2] + N$$

So the martingale $(X_n^{T_N})_{n\in\mathbb{N}_0}$ is L^2-bounded. In particular, $\lim_n X_n^{T_N}$ exists.

For almost all $\omega \in \{\langle X\rangle_\infty < \infty\}$, there is a sufficiently large N such that $T_N(\omega) = \infty$. So choose such N, then $X_n^{T_N} = X_n$. □

4 Brownian Motion

- Gaussian Measure: A measure γ on $\mathbb{R}^n$ is called Gaussian if its characteristic function is

$$\varphi_\gamma(\boldsymbol{t}) = \exp(i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$$

where $\boldsymbol{t} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ is a positive semi-definite matrix. A $\mathbb{R}^n$-valued random vector $\boldsymbol{X}$ is called Gaussian if $\mu_{\boldsymbol{X}} = \gamma$.

Theorem 4.1 (Structure Theorem of Gaussian)**.** *TFAE for a random vector $\boldsymbol{X} \colon \Omega \to \mathbb{R}^n$.*

(1) $\boldsymbol{X}$ is Gaussian.

(2) For each $\boldsymbol{a} \in \mathbb{R}^n$, the linear combination $\boldsymbol{a}^T\boldsymbol{X}$ is Gaussian.

(3) *Either $\boldsymbol{X}$ is constant or there exist*

   (i) *an integer $d \in \mathbb{N}$,*

   (ii) *a vector $\boldsymbol{\mu} \in \mathbb{R}^n$,*

   (iii) *a rank d matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$,*

   (iv) *a $\mathbb{R}^d$-valued $\boldsymbol{Y}$ defined on the same probability space as $\boldsymbol{X}$ and with the standard d-dimensional Gaussian*

   such that

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Y}$$

Remark. Then support of $\boldsymbol{X}$ is the affine space

$$\boldsymbol{\mu} + \operatorname{Im}\boldsymbol{\Sigma} = \{\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{x} \colon \boldsymbol{x} \in \mathbb{R}^n\}$$

In fact, $d = \operatorname{rank}\boldsymbol{\Sigma}$, called the rank of the Gaussian. If $d = n$, it is called nondegenerate, otherwise, it is degenerate. $\boldsymbol{X}$ is nondegenerate if and only if it admits a pdf

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

An n-dimensional random vector $\boldsymbol{X}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is called conditionally normally distributed with respect to a σ-subalgebra $\mathcal{G} \subset \mathcal{F}$ if there exist

(I) a random vector $\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}}$,

(II) a random $n \times n$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{G}}$, that is positive semi-definite,

such that $\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{G}}$ are $\mathcal{G}$-measurable and

$$\mathbb{E}[e^{i\boldsymbol{t}^T\boldsymbol{x}} \mid \mathcal{G}] = \exp\left(i\boldsymbol{t}^T\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{G}}\boldsymbol{t}\right)$$

Remark. We can let

$$\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}} = \mathbb{E}[\boldsymbol{X} \mid \mathcal{G}],\ \boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{G}} = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}})^T(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}}) \mid \mathcal{G}\right]$$

Note that $\boldsymbol{X}$ is independent with $\mathcal{G}$ if and only if $\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{G}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{G}}$ are constant.

- Brownian Motion: Let $(B_t)_{t\geqslant 0}$ be a Brownian motion, *i.e.*

  (1) $B_0 = 0$ *a.e.* and $B_t - B_s \sim \mathcal{N}(0, t-s)$ for all $0 \leqslant s < t$;

  (2) $(B_t)_{t\geqslant 0}$ is independent increments.

  (3) $\exists\ \Omega^*$ with $\mathbb{P}(\Omega^*) = 1$ such that $t \mapsto B_t(\omega)$ is continuous for all $\omega \in \Omega^*$.

  The existence of Brownian motion is based on the following theories.

  (1) The Kolmogorov Extension Theorem: Let $\mathcal{T}$ be a nonempty index set, and

  $$\{\mu_{t_1,t_2,\cdots,t_n} \colon n \in \mathbb{N},\ t_1, t_2, \cdots, t_n \in \mathcal{T}\}$$

  be a family of probability measures on $\mathbb{R}^n$ with the property

  (i) $\forall\ n \in \mathbb{N}$, $(t_1, t_2, \cdots, t_n) \in \mathcal{T}^n$ and a permutation $\sigma \in S_n$ and $\{A_i\}_{i=1}^n \in \mathcal{R}$

  $$\mu_{(t_1,\cdots,t_n)}(A_1 \times \cdots \times A_n) = \mu_{(t_{\sigma(1)},\cdots,t_{\sigma(n)})}(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)})$$

(ii) $\forall\, n \in \mathbb{N}$ and $A \in \mathcal{R}^n$ and $(t_1, t_2, \cdots, t_{n+1}) \in \mathcal{T}^{n+1}$,

$$\mu_{(t_1,\cdots,t_{n+1})}(A \times \mathbb{R}) = \mu_{(t_1,\cdots,t_n)}(A)$$

Then there is a probability measure $\mathbb{P}$ defined on

$$\left(\prod_{t\in\mathcal{T}} \mathbb{R}, \prod_{t\in\mathcal{T}} \mathcal{R}\right)$$

such that the finite-dimensional distribution of the coordinate process $\{x_t\}_{t\in\mathcal{T}}$ is given by μ.

Definition 4.2. For $n \in \mathbb{N}$, the family of probability measures $\mu_{t_1,\dots,t_n}$ on $\mathcal{B}(\mathbb{R}^n)$ indexed by all n-tuples $(t_1, \dots, t_n) \in [0,\infty)^n$ is said to be the (family of) n-dimensional distributions of the stochastic process $\{\boldsymbol{X}_t\}_{t\in[0,\infty)}$ if

$$\mu_{t_1,\dots,t_n}(x_1, \cdots, x_n) = \mathbb{P}\left(\{\omega \in \Omega \colon \boldsymbol{X}_{t_i}(\omega) \le x_i,\ \forall\, i\}\right)$$

for all $t_1, \dots, t_n \in [0,\infty)$. The collection of all n-dimensional distribution, with n ranging through $\mathbb{N}$ is called the (family of) finite-dimensional distributions of the process $\{\boldsymbol{X}_t\}_{t\in[0,\infty)}$.

Remark. Note that for $A = A_1 \times \cdots \times A_n$

$$\mathbb{P}\left((\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}, \cdots, \boldsymbol{X}_{t_n})^{-1}(A)\right) = \mathbb{P}\left(\boldsymbol{X}_{t_1} \in A_1, \boldsymbol{X}_{t_2} \in A_2, \cdots, \boldsymbol{X}_{t_n} \in A_n\right)$$

because

$$(\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}, \cdots, \boldsymbol{X}_{t_n})^{-1}(A) = \bigcap_{i=}^{n} \boldsymbol{X}_{t_i}^{-1}(A_i)$$

So

$$\mu_{t_1,\dots,t_n}(x_1, \cdots, x_n) = \mathbb{P}\left(\boldsymbol{X}_{t_1} \le x_1, \boldsymbol{X}_{t_2} \le x_2, \cdots, \boldsymbol{X}_{t_n} \le x_n\right)$$

which means it is the joint distribution of $(\boldsymbol{X}_{t_1}, \cdots, \boldsymbol{X}_{t_n})$.

Therefore, there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a process $(X_t)_{t\geqslant 0}$ which satisfies (1)(2) for the definition of Brownian motion by letting $\mu_{(t_1,\cdots,t_n)}$ be a centered Gaussian measure on $\mathbb{R}^n$ with covariance $\Sigma_{ij} = \min\{t_i, t_j\}$.

(2) For (3), a real function $f \colon [0,\infty) \to \mathbb{R}$ is said local Hölder continuous with $\gamma \in (0,1]$ if $\exists\, \delta, K > 0$ such that

$$|f(t) - f(s)| \leqslant K\,|t-s|^\gamma\,,\ \forall\, t, s \text{ with } |t-s| < \delta$$

Kolmogorov-Cĕntsov Theorem: If a stochastic process $(X_t)_{t\in[0,1]}$ satisfies

$$\mathbb{E}[|X_t - X_s|^\alpha] \leqslant C\,|t-s|^{1+\beta}$$

for some $\alpha, \beta, C > 0$. Then there is modification $(\tilde{X}_t)_{t\in[0,1]}$ on Ω^* such that for $\omega \in \Omega^*$, $t \mapsto \tilde{X}_t(\omega)$ is γ-local Hölder continuous for $0 < \gamma < \frac{\beta}{\alpha}$.
Because $X_t - X_s \sim \mathcal{N}(0, t-s)$ and $X_1 \sim \mathcal{N}(0,1)$,

$$\mathbb{E}[|X_t - X_s|^{2n}] = C_n\,|t-s|^n$$

where $C_n = \mathbb{E}[|X_1|^{2n}] = \frac{(2n)!}{2^n n!} < \infty$, there is a modification satisfying the continuous path condition. So $(B_t)_{t\in[0,1]}$ is a Brownian motion. This Brownian motion can be extended to $[0,\infty)$.

- Canonical Space: Equipping $\mathcal{C}([0,\infty))$ with the uniform continuous topology that is $\alpha_n \to \alpha$ if $\alpha_n|_{[0,k]} \to \alpha|_{[0,k]}$ uniformly for all $k \in \mathbb{N}$. Moreover, this topology is metrizable wit a metric d. $(\mathcal{C}([0,\infty)), d)$ is separable and complete. And let $\mathcal{B}(\mathcal{C}([0,\infty)))$ be the Borel algebra generated by this topology.

  For any $t \in [0,\infty)$, let $\pi_t\colon \mathcal{C}([0,\infty)) \to \mathbb{R}$ by $\pi_t(\alpha) = \alpha(t)$. Fix $n \in \mathbb{N}$ and $A \in \mathcal{R}^n$ and $t_1, \cdots, t_n \geqslant 0$,
$$\{\alpha \in \mathcal{C}([0,\infty))\colon (\alpha(t_1), \alpha(t_2), \cdots, \alpha(t_n)) \in A\}$$
  is called a finite-dimensional cylinder. Let $\mathcal{C}$ be the set of all finite-dimensional cylinders. In fact, $\mathcal{B}(\mathcal{C}([0,\infty))) = \sigma(\mathcal{C})$. Moreover, it is also the smallest σ-algebra such that all π_t are measurable.

  Proposition 4.3. *Let stochastic process $X = (X_t)_{t\geqslant 0}$ be path-continuous. Then there exists a probability measure $\mathbb{P}_X$ define on $\mathcal{C}([0,\infty)$ such that the coordinate process $(x_t)_{t\geqslant 0}$ has the same finite-dimensional distributions as $(X_t)_{t\geqslant 0}$.*

  Remark. Since $X = (X_t)_{t\geqslant 0}$ is path-continuous, it can be viewed as a measurable map
$$X\colon (\Omega, \mathcal{F}) \longrightarrow (\mathcal{C}([0,\infty)), \sigma(\mathcal{C}))$$
  Therefore, $\mathbb{P}_X$ defined on $\mathcal{C}([0,\infty))$ is the push-forward probability measure of $\mathbb{P}$ with respect to X.

  Such $\mathbb{P}_X$ is called the law or the distribution of X. In particular, if X is a Brownian motion, then $\mathbb{P}_X$ is called a Wiener measure (because Brownian motion is also called Wiener process).

- More for path space $\mathcal{C}([0,T], \mathbb{R}^d)$: On $\mathcal{C}([0,T], \mathbb{R}^d)$, for any $t \in [0,T]$, let
$$\begin{aligned} \pi_t\colon \quad \mathcal{C}([0,T], \mathbb{R}^d) &\longrightarrow \mathbb{R}^d \\ (\alpha(t))_{t\in[0,T]} &\mapsto \pi_t\left((\alpha(t))_{t\in[0,T]}\right) := \alpha(t) \end{aligned}$$
  and as above mention define the set of Borel sets $\mathcal{B} = \mathcal{B}(\mathcal{C}([0,T], \mathbb{R}^d))$ be the smallest set such that each π_t is measurable and also it is generated by the finite cylinders. Next, let's see its connections to random process

  (1) For a process $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,T]}$ over $(\Omega, \mathbb{P}, \mathcal{F})$, $\boldsymbol{X}$ can be viewed as
$$\begin{aligned} \boldsymbol{X}\colon \quad (\Omega, \mathcal{F}) &\longrightarrow \left(\mathcal{C}([0,T], \mathbb{R}^d), \mathcal{B}\right) \\ \omega &\mapsto (\boldsymbol{X}_t(\omega))_{t\in[0,T]} \end{aligned}$$
  and by the definition of $\mathcal{B}$ (finite cylinders), $\boldsymbol{X}$ is measurable because for any $\prod_{i=1}^n A_i$ $(A_i \in \mathcal{R}^d)$,
$$\begin{aligned} &\boldsymbol{X}^{-1}\left(\left\{\alpha \in \mathcal{C}([0,\infty), \mathbb{R}^d)\colon (\alpha(t_1), \alpha(t_2), \cdots, \alpha(t_n)) \in \textstyle\prod_{i=1}^n A_i\right\}\right) \\ &= \boldsymbol{X}^{-1}\left(\left\{\alpha \in \mathcal{C}([0,\infty), \mathbb{R}^d)\colon \alpha(t_i) \in A_i,\ i = 1, 2, \cdots, n\right\}\right) \\ &= \{\omega \in \Omega\colon \boldsymbol{X}_{t_i}(\omega) \in A_i,\ i = 1, 2, \cdots, n\} \\ &== \textstyle\bigcap_{i=}^n \boldsymbol{X}_{t_i}^{-1}(A_i) \in \mathcal{F} \end{aligned}$$
  For $\boldsymbol{X}$, define $\mathbb{P}_X$ over $\mathcal{C}([0,\infty), \mathbb{R}^d)$ by push-forward $\mathbb{P}_X = \boldsymbol{X}_{\#}\mathbb{P}$. Therefore, for
$$C = \left\{\alpha \in \mathcal{C}([0,\infty), \mathbb{R}^d)\colon (\alpha(t_1), \alpha(t_2), \cdots, \alpha(t_n)) \in \prod_{i=1}^n A_i\right\} \in \mathcal{B}$$
  we have
$$\mathbb{P}_X(C) = \mathbb{P}(\boldsymbol{X} \in C) = \mathbb{P}(\boldsymbol{X}^{-1}(C))$$

(2) Next, let research $\mathbb{P}_X$ from another view. Let $\pi = (\pi_t)_{t\in[0,T]}$ be the process defined as above on $\mathcal{C}([0,T],\mathbb{R}^d)$, called the coordinate process. Then for any measure μ on $\mathcal{C}([0,T],\mathbb{R}^d)$, let's consider the finite-dimensional distribution of π given by μ. For $0 \leq t_1 < t_2 < \cdots < t_n \leq T$,

$$\begin{array}{rccc} (\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_n}) \colon & \mathcal{C}([0,T],\mathbb{R}^d) & \longrightarrow & \prod_{i=1}^n \left(\mathbb{R}^d\right) \\ & (\alpha(t))_{t\in[0,T]} & \mapsto & (\alpha(t_1), \alpha(t_2), \cdots, \alpha(t_n)) \end{array}$$

Then define $\mu_{t_1,t_2,\cdots,t_n} = (\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_n})_{\#}\mu$, that is, for $A = \prod_{i=1}^n A_i \in \left(\mathcal{R}^d\right)^n$,

$$\mu_{t_1,t_2,\cdots,t_n}(A) := \mu\left((\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_n})^{-1}(A)\right) = \mu\left(\pi_{t_1} \in A_1, \pi_{t_2} \in A_2, \cdots, \pi_{t_n} \in A_n\right)$$

For example, for a process $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,T]}$, it induces $\mathbb{P}_X$ on $\mathcal{C}([0,T],\mathbb{R}^d)$, which induces

$$\begin{aligned} (\mathbb{P}_X)_{t_1,t_2,\cdots,t_n}(A) &= \mathbb{P}_X\left((\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_n})^{-1}(A)\right) \\ &= \mathbb{P}\left(((\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_n}) \circ \boldsymbol{X})^{-1}(A)\right) \\ &= \mathbb{P}\left((\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}, \cdots, \boldsymbol{X}_{t_n})^{-1}(A)\right) \end{aligned}$$

The Kolmogorov Extension Theorem is basically consider the converse part, that is, for a process $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in[0,T]}$, the finite-dimensional distribution uniquely determine $\mathbb{P}_X$ on $\mathcal{C}([0,\infty),\mathbb{R}^d)$.

Remark. First, the path-space is $\mathcal{W} = \mathcal{C}([0,T],\mathbb{R}^d)$ with the canonical the set of Borel sets $\mathcal{F} = \mathcal{B}(\mathcal{W})$, which can be filtered by $(\mathcal{F}_t)_{t\in[0,T]}$ and $\mathcal{F}_t$ is generated by

$$\left\{\alpha \in \mathcal{C}([0,\infty),\mathbb{R}^d) \colon \alpha(t_1) \in A_1, \alpha(t_2) \in A_2, \cdots, \alpha(t_n) \in A_n\right\}$$

for any $n \in \mathbb{N}$, $A_i \in \mathcal{R}^d$ and $0 \leq t_1 < t_2 < \cdots < t_n \leq t$. Then $(\mathcal{W}, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]})$ is called the Wiener space or a sample space of Brownian motion (In many cases, $\mathcal{C}_0([0,T],\mathbb{R}^d) \subset \mathcal{W}$ that restricts $\alpha(0) = 0$ refers to the Wiener space). For a $\mathbb{P}$-Brownian motion $B = (B_t)_{t\in[0,T]}$, consider $\mathbb{P}_0 = B_{\#}\mathbb{P}$ on $\mathcal{W}$, called the Wiener measure. Then for the filtered space $(\mathcal{W}, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}_0)$, the coordinate process $(W_t)_{t\in[0,T]}$ is a Brownian motion and is adapted to $(\mathcal{F}_t)_{t\in[0,T]}$ by the above induction. Such $(W_t)_{t\in[0,T]}$ is called a $\mathbb{P}_0$-Brownian motion. This is an example to construct a Brownian motion, because $\mathbb{P}_0 = B_{\#}\mathbb{P}$ can be defined by a series of Gaussian integrals.

- Convergence: A sequence of probability measures $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ defined on $\mathcal{C}([0,\infty))$ is called tight if for any $\varepsilon > 0$, there exists a compact $K \subset \mathcal{C}([0,\infty))$ such that $\sup_n \mathbb{P}_n(K^c) < \varepsilon$. If the sequence is tight, there is a subsequence that converges weakly.

  For $m \in \mathbb{N}$ and $t_1, \cdots, t_m \geqslant 0$, $\pi_{(t_1,\cdots,t_m)} \colon \mathcal{C}([0,\infty)) \to \mathbb{R}^m$ defined by

$$\pi_{(t_1,\cdots,t_m)}(\alpha) = (\alpha(t_1), \cdots, \alpha(t_m)) \in \mathbb{R}^m$$

  which are measurable. Then for any probability $\mathbb{P}$ on $\mathcal{C}([0,\infty))$, the push-forward of $\mathbb{P}$ to $\mathbb{R}^m$ with respect to $\pi_{(t_1,\cdots,t_m)}$ is denoted by

$$\mathbb{P}_{(t_1,\cdots,t_m)}(A) = \mathbb{P}(\pi^{-1}_{(t_1,\cdots,t_m)}(A))$$

  called the finite-dimensional distribution. A sequence of probability measures $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ defined on $\mathcal{C}([0,\infty))$ converges of finite-dimensional distributions

$$\mathbb{P}^n \xrightarrow{f.d.d.} \mathbb{P} \;\Leftrightarrow\; \mathbb{P}^n_{(t_1,\cdots,t_m)} \xrightarrow{wk} \mathbb{P}_{(t_1,\cdots,t_m)}$$

  for any $m \in \mathbb{N}$ and $t_1, \cdots, t_m \geqslant 0$.

Remark. Let $\mathbb{P}_n$ be the laws of $(X_t^n)_{t\geqslant 0}$ and $\mathbb{P}$ be the law of of $(X_t)_{t\geqslant 0}$. Then $\mathbb{P}_n \xrightarrow{f.d.d.} \mathbb{P}$ if and only if for any $m \in \mathbb{N}$ and $t_1, \cdots, t_m \geqslant 0$, $(X_{t_1}^n, \cdots, X_{t_m}^n) \xrightarrow{d} (X_{t_1}, \cdots, X_{t_m})$

It's clear that $\mathbb{P}^n \xrightarrow{wk} \mathbb{P}$ implies $\mathbb{P}^n \xrightarrow{f.d.d.} \mathbb{P}$. But the converse is not true. However, if $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ is tight, the converse can be true.

Convergence of Brownian Motion: Let $(B_t)_{t\geqslant 0}$ be a Brownian motion.

(1) $\lim_{t\to\infty} B_t/t = 0$.

(2) $\limsup_{t\to\infty} B_t/\sqrt{t} = \infty$, $\liminf_{t\to\infty} B_t/\sqrt{t} = -\infty$.

(3) It is recurrent, *i.e.* for each $a \in \mathbb{R}$,

$$\{t \in [0,\infty) \colon B_t(\omega) = a\}$$

is unbounded *a.e.*.

Total variation: Let $\mathcal{P}_{[0,t]}$ be the set of all partitions of $[0,t]$, *i.e.* $\Delta \in \mathcal{P}_{[0,t]}$ if $\Delta = \{t_0, t_1, \cdots, t_k\}$ with $0 = t_0 < \cdots < t_k = t$. Let $|\Delta| = \sup_i (t_i - t_{i-1})$. For a sequence Δ_n in $\mathcal{P}_{[0,t]}$, $\Delta_n \to \mathrm{Id}$ if $|\Delta_n| \to 0$, moreover if $\sum_n |\Delta_n| < \infty$, $\Delta_n \xrightarrow{sum} \mathrm{Id}$. For a function $f \colon [0,t] \to \infty$, let $\Delta = \{t_0, t_1, \cdots, t_k\} \in \mathcal{P}_{[0,t]}$, the p-variation of Δ is

$$\mathrm{Var}_p(f;\Delta) := |f(0)|^p + \sum_{i=1}^{k} |f(t_i) - f(t_{i-1})|$$

And the total p-variation is

$$\mathrm{Var}_p(f;[0,t]) := \sup_{\Delta\in\mathcal{P}_{[0,t]}} \mathrm{Var}_p(f;\Delta)$$

Let Δ_n in $\mathcal{P}_{[0,t]}$ with $\Delta_n \to \mathrm{Id}$, then

$$\lim_{n\to\infty} \mathrm{Var}(B_\cdot(\omega);\Delta_n) = t$$

convergent in $L^2(\omega)$. Moreover, if $\Delta_n \xrightarrow{sum} \mathrm{Id}$, the convergence is also *a.e.*. But note that for any $t > 0$,

$$\mathrm{Var}_2(B_\cdot(\omega);[0,t]) = \infty,\ \ \mathrm{Var}_1(B_\cdot(\omega);[0,t]) = \infty,\ \ a.e.$$

Note that the Brownian motion is not differentiable with respect to t, but the irregularity can be estimated.

$$\limsup_{h\downarrow 0} \frac{|B_{t+h} - B_t|}{\sqrt{2h\log(\log(1/h))}} = 1,\ \ a.e.$$

5 Continuous Process

- Continuous Process: A stochastic process $(X_t)_{t\geqslant 0}$ is said measurable if the map $X \colon [0,\infty) \to \mathbb{R}$ is measurable from $\mathcal{B}([0,\infty)) \otimes \mathcal{F}$ to $\mathcal{R}$. We assume all stochastic process are measurable from now. Note that if there is filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geqslant 0}$ and $(X_t)_{t\geqslant 0}$ is $\mathbb{F}$-adapted, then X is measurable. But the converse is clearly not true.

  For the path, a stochastic process $(X_t)_{t\geqslant 0}$ is said

  (1) continuous if all path $t \mapsto X_t(\omega)$ is continuous;

  (2) right(left)-continuous if all path $t \mapsto X_t(\omega)$ is right(left)-continuous;

(3) RCLL(LCRL) if all path $t \mapsto X_t(\omega)$ is right(left)-continuous and has left(right) limits;

(4) of finite variation if all path $t \mapsto X_t(\omega)$ has finite variation on any segment;

(5) bounded if there is a K such that all path $t \mapsto X_t(\omega)$ is bounded by K on $[0, \infty)$.

Remark. RCLL is also called "càdlàg" and LCRL is also called "càglàg".

- Random time: Given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$, a random time $\tau \colon \Omega \to [0, \infty)$ is called

  (1) a $\mathbb{F}$-stopping time if
$$\{\tau \leqslant t\} \in \mathcal{F}_t, \ \forall\, t \geqslant 0$$

  (2) a $\mathbb{F}$-optional time if
$$\{\tau < t\} \in \mathcal{F}_t, \ \forall\, t > 0$$

  Note that stopping time is a optional because $\{\tau < t\} = \bigcup_n \{\tau \leqslant t_n\}$ when $t_n \uparrow t$ and $\{\tau \leqslant t_n\} \in \mathcal{F}_{t_n} \subset \mathcal{F}_t$. But the converse is not true. For example, let $(B_t)_{t \geqslant 0}$ be a Brownian motion and $\mathbb{F}$ be the induced filtration and
$$\tau = \inf\{t \geqslant 0 \colon B_t > 1\}$$
  Then τ is a optional time but not a stopping time.

  Given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$, let
$$\mathcal{F}_{t+} := \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}, \ \mathbb{F}_+ = (\mathcal{F}_{t+})_{t \geqslant 0}$$
  If $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, *i.e.* $\mathbb{F} = \mathbb{F}_+$. Then $\mathbb{F}$ is called right continuous.

  We can see τ is an $\mathbb{F}$-optional time if and only if it is an $\mathbb{F}_+$-stopping time. So if $\mathbb{F}$ is right continuous, stopping time and optional time are coincide.

  If τ and σ are stopping(optional) times, so are
$$\sigma + \tau, \ \max(\sigma, \tau), \ \min(\sigma, \tau)$$
  For any subset $A \subset \mathbb{R}$, the hitting time is defined as
$$\tau_A := \inf\{t \geqslant 0 \colon X_t \in A\}$$
  Then if A is closed, then τ_A is a stopping time and if A is closed or open, τ_A is a optional time.

- Progressive Measurability: A stochastic process $(X_t)_{t \geqslant 0}$ is said progressive if when it viewed as the map $X \colon [0, \infty) \times \Omega \to \mathbb{R}$, it is measurable with respect to the σ-algebra of Prog, where
$$\mathrm{Prog} := \{A \in \mathcal{B}([0, \infty)) \times \mathcal{F} \colon A \cap ([0, T] \times \Omega) \in \mathcal{B}([0, \infty)) \times \mathcal{F}_T, \ \forall\, T \geqslant 0\}$$
  Note that it is stronger than measurable and adapted. However, if a stochastic process $(X_t)_{t \geqslant 0}$ is $\mathbb{F}$-adapted and is right(left) continuous, then $(X_t)_{t \geqslant 0}$ is progressively measurable.

  If $(X_t)_{t \geqslant 0}$ is progressive and τ is progressive, then the stopped process $(X_t^\tau)_{t \geqslant 0}$ that is defined as $X_t^\tau(\omega) := X_{\tau(\omega) \wedge t}(\omega)$ is also progressively measurable.

  If τ is a stopping time, define
$$\mathcal{F}_\tau := \sigma\{X_t^\tau \colon t \geqslant 0, \ X \text{ is a progressively-measurable process}\}$$

Proposition 5.1 (Properties of $\mathcal{F}_\tau$)**.** *Let σ, τ and $\{\tau_n\}_{n\in\mathbb{N}}$ be stopping times of $\mathbb{F} = (\mathcal{F}_t)_{t\geqslant 0}$.*

(1) Let $\bigvee_{t\geqslant 0} \mathcal{F}_t = \sigma(\bigcup_t \mathcal{F}_t)$.

$$\mathcal{F}_\tau = \left\{ A \in \bigvee_{t\geqslant 0} \mathcal{F}_t \colon A \cap \{\tau \leqslant t\} \in \mathcal{F}_t, \ \forall \ t \right\}$$

(2) If $\sigma \leqslant \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

(3) If $\tau_1 \geqslant \tau_2 \geqslant \cdots \geqslant \tau$ and $\tau = \lim_n \tau_n$ and $\mathbb{F}$ is right-continuous, then $\mathcal{F}_\tau = \cap_n \mathcal{F}_{\tau_n}$.

(4) If there is a countable set $\mathcal{Q} = \{q_k \colon k \in \mathbb{N}\} \subset [0, \infty)$ such that $\tau(\omega) \in \mathcal{Q}$, then

$$\mathcal{F}_\tau = \left\{ A \in \bigvee_{t\geqslant 0} \mathcal{F}_t \colon A \cap \{\tau = q_k\} \in \mathcal{F}_{q_k}, \ \forall \ k \in \mathbb{N} \right\}$$

(5) If X is a (progressively) measurable process, then $X_\tau \mathbb{1}_{\tau<\infty}$ is a ($\mathcal{F}_\tau$-measurable) random variable.

Moreover, if X is a bounded progressive process, then there is a sequence of adapted and continuous process $\{X^n\}_{n\in\mathbb{N}}$ such that

$$X_t^n(\omega) \to X_t(\omega)$$

a.e. in t for each $\omega \in \Omega$.

6 Continuous Martingale

- Martingale: A stochastic process $(X_t)_{t\geqslant 0}$ is called a $\mathbb{F} = (\mathcal{F}_t)_{t\geqslant 0}$-supermartingale if it is adapted and all X_t are in L^1 and $\mathbb{E}[X_t \mid \mathcal{F}_s] \leqslant X_s$ for any $t \geqslant s$. And it is called a submartingale if $-X_t$ is a supermartingale and a martingale if it is both sub and super martingale.

- Let $\mathcal{S}_t$ be the set of all stopping times τ with $\tau \leqslant t$. So $\mathcal{S}_b = \bigcup_t \mathcal{S}_t$ is the set of all bounded stopping times. Let $\mathcal{S}$ be the set of all stopping times. Then a measurable process $(X_t)_{t\geqslant 0}$ is called

  (1) of class (DL) if $\{X_\tau \colon \tau \in \mathcal{S}_t, \ t \geqslant 0\}$ is UI.

  (2) of class (D) if $\left\{X_\tau \mathbb{1}_{\{\tau<\infty\}} \colon \tau \in \mathcal{S}\right\}$ is UI.

Theorem 6.1 (Bounded Optional Stopping Time Theorem)**.** *Let $M = (M_t)_{t\geqslant 0}$ be a right-continuous martingale. Then M is of class (DL) and*

$$\mathbb{E}[M_t \mid \mathcal{F}_\tau] = M_\tau, \ \mathbb{E}[M_\tau] = \mathbb{E}[M_0], \ \forall \ \tau \in \mathcal{S}_t, \ t \geqslant 0$$

Proposition 6.2. *Let $M = (M_t)_{t\geqslant 0}$ be adapted and right-continuous with the property $M_\tau \in L^1$ for all $\tau \in \mathcal{S}_b$. Then M is a martingale if and only if*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0], \ \forall \ \tau \in \mathcal{S}_b$$

This implies that if M is a right-continuous martingale, then so is the stopped process M^τ for each stopping time τ.

Convergence Property of Martingale:

Theorem 6.3. *Let $M = (M_t)_{t\geqslant 0}$ be a right-continuous martingale. TFAE:*

(1) $\{M_t\}_{t\geqslant 0}$ *is UI,*

(2) $\lim_t M_t$ *exists in* L^1*,*

(3) There is a last element M^* *such that*

$$M_t = \mathbb{E}[M^* \mid \mathcal{F}_t],\ \forall\ t \geqslant 0$$

Theorem 6.4. *Let* $(X_t)_{t\geqslant 0}$ *be a right-continuous submartingale with a last element* X^**, i.e.* $X^* \in L^1$ *and* $X_t \leqslant \mathbb{E}[X^* \mid \mathcal{F}_t]$ a.e.. *Then*

(1) $X_\infty = \lim_{t\to\infty} X_t$ a.e. *and* $X_\infty \in L^1$ *is the* a.e.*-minimal last element,*

(2) $X_\tau \leqslant \mathbb{E}[X_\infty \mid \mathcal{F}_\tau]$ *for* $\tau \in \mathcal{S}$*.*

Similarly as the discrete case, if $(M_t)_{t\geqslant 0}$ is a right-continuous (sub)martingale, then so is the stopped process $(M_t^\tau)_{t\geqslant 0}$ for all stopping time τ.

- RCLL Modification: A filtration is said to satisfy the usual conditions if it is right continuous and complete, *i.e.* $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ implies $A \in \mathcal{F}_0$.

Theorem 6.5. *Under the usual conditions, a submartingale* $(X_t)_{t\geqslant 0}$ *has a RCLL modification if and only if the mapping* $t \mapsto \mathbb{E}[X_t]$ *is right-continuous. In particular, each martingale has a RCLL modification.*

- Doob's and Maximal Inequalities: For a stochastic process $X = (X_t)_{t\geqslant 0}$, the maximal process

$$X_t^* = \sup\left\{|X|_q : q = t \text{ or } q \text{ is a rational in } [0,t)\right\}$$

Note that if X is RCLL or LCRL, $X_t^* = \sup_{s\in[0,t]} |X_s|$ *a.e.*.

Theorem 6.6. *Let* $X = (X_t)_{t\geqslant 0}$ *be a RCLL process that is either martingale or positive submartingale. Then*

$$\mathbb{P}(X^* \geqslant M) \leqslant \frac{1}{M^p} \sup_{t\geqslant 0} \mathbb{E}[|X_t|^p],\ \forall\ M > 0,\ p \geqslant 0$$

and

$$\|X^*\|_{L^p} \leqslant \frac{p}{1-p} \sup_{t\geqslant 0} \|X_t\|_{L^p},\ p \geqslant 1$$

- Predictable and Optional: A stochastic process $X = (X_t)_{t\geqslant 0}$ is said to be

  (1) optional, if it is measurable with respect to the σ-algebra $\mathcal{O}$, which is the smallest σ-algebra on $[0,\infty) \times \Omega$ such that all RCLL and adapted processes are measurable.

  (2) predicable, if is measurable with respect to the σ-algebra $\mathcal{P}$, which is the smallest σ-algebra on $[0,\infty) \times \Omega$ such that all LCRL and adapted processes are measurable.

Note that $\mathcal{P} \subset \mathcal{O} \subset \sigma(\text{Prog}) \subset \mathcal{B}([0,\infty)) \otimes \mathcal{F}$.

7 Markov Property

- Markov Property: Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a filtered probability space. A stochastic process $X = (X_t)_{t\geqslant 0}$ is a $(\mathcal{F}_t)_{t\geqslant 0}$-Markov process if for all $B \in \mathcal{R}$ and $t, h \geqslant 0$

$$\mathbb{P}(X_{t+h} \in B \mid \mathcal{F}_t) = \mathbb{P}(X_{t+h} \in B \mid \sigma(X_t))$$

Remark. The strong Markov proper is to replace the t in above by a stopping time τ. That is, for a stopping time τ,

$$\mathbb{P}(X_{\tau+h} \in B \mid \mathcal{F}_\tau) = \mathbb{P}(X_{\tau+h} \in B \mid \sigma(X_\tau))$$

Example 7.1. Let $(B_t)_{t\geqslant 0}$ be a Brownian motion and $\mathbb{F}^B = (\mathcal{F}_t^B)_{t\geqslant 0}$ be the natural filtration. Fix any $t \geqslant 0$ and $s > 0$. $B_{t+s} - B_t$ is independent with $\mathcal{F}_t^B$. Based on 1.10, for any Borel function $f : \mathbb{R} \to \mathbb{R}$,

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

Taking $f = \mathbb{1}_B$, it induces the Markov property.

- $(\mathcal{F}_t)_{t\geqslant 0}$-adapted Brownian motion is a $(\mathcal{F}_t)_{t\geqslant 0}$-adapted process $(B_t)_{t\geqslant 0}$ such that

  (1) $B_0 = 0$ and $B_t - B_s \sim \mathcal{N}(0, t-s)$ for $0 \leqslant s < t$,

  (2) $B_t - B_s$ is independent with $\mathcal{F}_s$ for $0 \leqslant s < t$,

  (3) for all $\omega \in \Omega$, $t \mapsto B_t(\omega)$ is continuous.

An $(\mathcal{F}_t)_{t\geqslant 0}$-adapted progress $(X_t)_{t\geqslant 0}$ with continuous is an $(\mathcal{F}_t)_{t\geqslant 0}$-adapted Brownian motion if and only if the complex-values process $(Y_t^r)_{t\geqslant 0}$, given by

$$Y_t^r = \exp\left(irX_t + \frac{1}{2}rt^2\right), \ \forall\, t \geqslant 0$$

is an $(\mathcal{F}_t)_{t\geqslant 0}$-martingale for all $r \in \mathbb{R}$.

Let $(B_t)_{t\geqslant 0}$ be a Brownian motion and $\mathbb{F}^B = (\mathcal{F}_t^B)_{t\geqslant 0}$ be the natural filtration. Note that $\mathbb{F}^B$ may be not right continuous. So let's consider $\mathbb{F}_+^B$ the right-continuous augmentation. Fortunately, the Brownian motion $(B_t)_{t\geqslant 0}$ is a $\mathbb{F}_+^B$-Brownian motion, and a $\mathbb{F}_+^B$-Markov process. In fact, for all t, $\mathcal{F}_t^B$ and $\mathcal{F}_{t+}^B$ are *a.s.*-equal, *i.e.*

$$\forall\, A \in \mathcal{F}_{t+}^B,\ \exists\, A' \in \mathcal{F}_t^B,\ s.t.\ \mathbb{P}(A \Delta A') = 0$$

Besides, for any $\varepsilon > 0$, there is a $t \in [0, \varepsilon)$ such that $\mathbb{P}(B_t < 0) = 1$. Similarly, for any $\varepsilon > 0$, there is a $t' \in [0, \varepsilon)$ such that $\mathbb{P}(B_{t'} > 0) = 1$.

Strong Markov Property: For a Brownian motion $(B_t)_{t\geqslant 0}$, let $\tau < \infty$ be a $(\mathcal{F}_{t+}^B)$-stopping time. Then the process defined by

$$W_t = B_{\tau+t} - B_\tau$$

is a Brownian motion, independent of $\mathcal{F}_{\tau+}^B$.

- Let $(B_t)_{t\geqslant 0}$ be a Brownian motion and τ be a stopping time with $\tau(\omega) < \infty$ for all $\omega \in \Omega$. Then process $(\tilde{B}_t)_{t\geqslant 0}$ given by

$$\tilde{B}_t = \begin{cases} B_t, & t \leqslant \tau \\ B_\tau - (B_t - B_\tau) & t > \tau \end{cases}$$

is a Brownian motion.

8 Stochastic Integration

- Integral of Finite Variation Process: For $T > 0$, $F\colon [0,T] \to \mathbb{R}$ be càdlàg function of finite variation and let $|F| : [0,T] \to [0,\infty)$ be the total variation. For $a < b \in [0,T]$, let $\mu((a,b]) = F(b) - F(a)$. Then μ can be extended to a unique signed measure on $\mathcal{B}([0,T])$, called the Stieltjes measure, denoted by dF, with the Hahn-Jordan decomposition

  $$dF = dF_+ - dF_-,$$

  where $F_+, F_-\colon [0,T] \to \mathbb{R}$ càdlàg with finite positive measures dF_+, dF_- and $d|F| = dF_+ + dF_-$. A function $h \in L^1(dF)$ if

  $$\int_0^T |h(x)|\, d|F|(x) < \infty$$

  and then

  $$\int_0^T h(x)dF(x) = \int_0^T h(x)dF_+(x) - \int_0^T h(x)dF_-(x)$$

  For an adapted càdlàg process $(X_t)_{t\geqslant 0}$ with finite variation *w.s.t.* t, let $(H_t)_{t\geqslant 0}$ be a progressively measurable process *s.t.*

  $$\int_0^T |H_t(\omega)|\, d|X|_t(\omega) < \infty,\ a.e.\ \omega$$

  Then for all $t \geqslant 0$,

  $$Y_t(\omega) := \int_0^t H_s(\omega)dX_s(\omega)$$

  It is called the stochastic integral, denoted by $Y = (H \cdot X)$.

- Simple Process: Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$.

  A stochastic process $(H_t)_{t\geqslant 0}$ is called simple predictable if there is a partition $\Delta = \{t_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}_{[0,\infty)}$ and a sequence of random variables $\{K_n\}_{n\in\mathbb{N}_0}$ *s.t.* $K_n \in \mathcal{F}_{t_n}$ and

  $$H_t = \sum_{n=0}^{\infty} K_n \mathbb{1}_{(t_n, t_{n+1}]}(t)$$

  Note that it is predictable. $\mathcal{H}_{simp}$ is the set of all simple process and $H \in \mathcal{H}_{simp}^{\infty} \subset \mathcal{H}_{simp}$ is

  $$\|H\|_{\mathcal{H}_{simp}^{\infty}} := \sup_{n\in\mathbb{N}_0} \|K_n\|_{L^\infty} < \infty$$

  Let $(M_t)_{t\geqslant 0}$ be arbitrary process and $(H_t)_{t\geqslant 0} \in \mathcal{H}_{simp}$. Then

  $$(H \cdot M)_t = \int_0^t H_u dM_u := \sum_{n=0}^{\infty} K_n (M_{t\wedge t_{n+1}} - M_{t\wedge t_n})$$

- Square-Integrable Martingale: Let $\mathcal{M}_0^{2,c}$ be the set of all continuous martingale $M = (M_t)_{t\geqslant 0}$ with

  $$\|M\|_{\mathcal{M}_0^{2,c}} := \sqrt{\sup_{t\geqslant 0} \mathbb{E}\left[M_t^2\right]} < \infty$$

  Note that $\left(\mathcal{M}_0^{2,c}, \|\cdot\|_{\mathcal{M}_0^{2,c}}\right)$ is a Banach space.

Proposition 8.1. *For $M \in \mathcal{M}_0^{2,c}$ and $H \in \mathcal{H}_{simp}^\infty$, we have $H \cdot M \in \mathcal{H}_{simp}^\infty$ and*

$$\mathbb{E}\left[(H \cdot M)_t^2\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[K_n^2 (M_{t\wedge t_{n+1}} - M_{t\wedge t_n})^2\right]$$

Let $\mathcal{A}_0^c$ be the set of all continuous, adapted, and nondecreasing process $(A_t)_{t\geqslant 0}$ with $A_0 = 0$.

Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of RCLL or LCRL processes and X be a RCLL or LCRL process. We say $X^n \xrightarrow{ucp} X$ uniformly on compacts in probability if

$$(X^n - X)_t^* \to 0, \text{ in probability for each } t \geqslant 0$$

Theorem 8.2 (Quadratic Variation)**.** *For $M \in \mathcal{M}_0^{2,c}$, there exists a unique $\langle M\rangle \in \mathcal{A}_0^c$ such that $M_t^2 - \langle M\rangle$ is a UI martingale. Furthermore, for each sequence $\{\Delta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_{[0,\infty)}$ with $\Delta_n \to Id$, we have*

$$\langle M\rangle^{\Delta_n} \xrightarrow{ucp} \langle M\rangle$$

where for $\Delta = \{t_k\}_{k\in\mathbb{N}} \in \mathcal{P}_{[0,\infty)}$, $\langle M\rangle_t^\Delta := \sum_{k=0}^\infty \left(M_{t_{k+1}\wedge t} - M_{t_k\wedge t}\right)^2$

Because both M and $M^2 - \langle M\rangle$ are martingale,

$$\begin{aligned}
\mathbb{E}[(M_{t_{k+1}} - M_{t_k})^2 \mid \mathcal{F}_{t_k}] &= \mathbb{E}[M_{t_{k+1}}^2 - M_{t_k}^2 \mid \mathcal{F}_{t_k}] \\
&= \mathbb{E}[\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k} \mid \mathcal{F}_{t_k}]
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}\left[(H \cdot M)_\infty^2\right] &= \sum_{k=0}^{\infty} \mathbb{E}\left[K_n^2 (M_{t_{n+1}} - M_{t_n})^2\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[K_k^2 \mathbb{E}\left[(M_{t_{k+1}} - M_{t_k})^2 \mid \mathcal{F}_{t_k}\right]\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[K_k^2 \mathbb{E}[\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k} \mid \mathcal{F}_{t_k}]\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{\infty} K_k^2 (\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k})\right] \\
&= \mathbb{E}\left[\int_0^\infty H_t^2 d\langle M\rangle_t\right]
\end{aligned}$$

For $H \in \mathcal{H}_{simp}$,

$$\|H\|_{\mathcal{L}^2(M)} := \sqrt{\mathbb{E}\left[\int_0^\infty H_t^2 d\langle M\rangle_t\right]}$$

Note that after taking the limits let the integral defined on all progressively measurable processes and let $\mathcal{L}^2(M)$ be the set of finite $\|\cdot\|_{\mathcal{L}^2(M)}$.

And for any $M \in \mathcal{M}_0^{2,c}$, element in $\mathcal{H}_{simp}^\infty$ is with finite $\|\cdot\|_{\mathcal{L}^2(M)}$. So for $M \in \mathcal{M}_0^{2,c}$

$$\|H \cdot M\|_{\mathcal{M}_0^{2,c}} = \|H\|_{\mathcal{L}^2(M)}, \ \forall\ H \in \mathcal{H}_{simp}^\infty$$

It is called the Itô's isometry.

Theorem 8.3. *For all $M \in \mathcal{M}_0^{2,c}$, $\mathcal{H}_{simp}^\infty$ is dense in $\mathcal{L}^2(M)$.*

Theorem 8.4. *Given $M \in \mathcal{M}_0^{2,c}$, there exists a unique linear isometry*

$$\mathcal{L}^2(M) \to \mathcal{M}_0^{2,c} \ \ni \ H \mapsto H \cdot M = \int H_t dM_t$$

Example 8.5. Let $M = B = (B_t)_{t\in[0,T]}$ be a Brownian motion. Then the decomposition of $(B_t^2)_{t\in[0,T]}$ is

$$B_t^2 = M_t + \langle B\rangle_t$$

where $\langle B\rangle_t = t$. It is by the result of $B_t^2 - t$ is a UI martingale and the uniqueness of the decomposition.

Proof. First,

$$B_t^2 = (B_s + (B_t - B_s))^2 = B_s^2 + 2(B_t - B_s)B_s + (B_t - B_s)^2$$

Because $B_t - B_s$ is independent with $\mathcal{F}_s$,

$$\begin{aligned}\mathbb{E}[B_t^2 \mid \mathcal{F}_s] &= \mathbb{E}[B_s^2 \mid \mathcal{F}_s] + 2\mathbb{E}[(B_t - B_s)B_s \mid \mathcal{F}_s] + \mathbb{E}[(B_t - B_s)^2 \mid \mathcal{F}_s] \\ &= B_s^2 + 2B_s\mathbb{E}[B_t - B_s] + \mathbb{E}[(B_t - B_s)^2] \\ &= B_s^2 + t - s\end{aligned}$$

So

$$\mathbb{E}[B_t^2 - t \mid \mathcal{F}_s] = B_s^2 - s$$

□

Then $\mathcal{L}^2(B)$ is the above $\mathcal{L}^2([0,T])$ with

$$\|H\|_{\mathcal{L}^2(B)} = \sqrt{\mathbb{E}\left[\int_0^T H_t^2 dt\right]}$$

So the Itô's isometry is as same as above.

9 Semi-Martingales

- Continuous local martingales: A continuous adapted process $(M_t)_{t\geqslant 0}$ is called a continuous local martingale if there is a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ such that

  (1) $\tau_1 \leqslant \tau_2 \leqslant \cdots \leqslant \tau_n$ *a.e.*,

  (2) $\{M^{\tau_n}\}_{t\geqslant 0}$ is a UI martingale for all $n \in \mathbb{N}$.

  The set of all such M with $M_0 = 0$ is denoted by $\mathcal{M}_0^{loc,c}$.

  Proposition 9.1. *There are some properties for local martingales.*

  (1) A local martingale is a martingale if and only if it is of class (DL).

  (2) A bounded local martingale is a martingale of class (D).

  (3) A local martingale bounded from below is a supermartingale.

  (4) For $M \in \mathcal{M}_0^{loc,c}$ and stopping time τ, $M^\tau \in \mathcal{M}_0^{loc,c}$.

(5) If M is a continuous local martingale with finite variation, then $M_t = 0$ a.e. for all t.

Theorem 9.2 (Quadratic Variation)**.** *Let $M = \{M_t\}_{t\geqslant 0}$ be a continuous local martingale. Then there exists a unique process $\langle M\rangle \in \mathcal{A}_0^c$ such that $M^2 - \langle M\rangle$ is a local martingale. For each sequence $\{\Delta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_{[0,\infty)}$ with $\Delta_n \to Id$, we have*

$$\langle M\rangle^{\Delta_n} \xrightarrow{ucp} \langle M\rangle$$

Theorem 9.3. *For $M, N \in \mathcal{M}_0^{loc,c}$, there is a unique process $(\langle M, N\rangle_t)_{t\geqslant 0}$ such that*

(1) $\langle M, N\rangle_0 = 0$,

(2) the process $(M_tN_t - \langle M, N\rangle_t)$ is a local martingale.

$\langle M, N\rangle$ is called the quadratic covariation of M, N.

For $M, N \in \mathcal{M}_0^{loc,c}$,

$$\langle M, N\rangle_t = \frac{1}{2}\left(\langle M+N\rangle_t - \langle M\rangle_t - \langle N\rangle_t\right)$$

Lemma 9.4. *If $X = (X_t)$ is a $\mathcal{F}^X$-adapted martingale and $Y = (Y_t)$ is a $\mathcal{F}^Y$-adapted martingale and $\mathcal{F}^X, \mathcal{F}^Y$ independent, then $(X_t + Y_t)$ a martingale with $\mathcal{F}^Z$ with $\mathcal{F}_t^Z = \sigma(X_0, Y_0, \cdots, X_t, Y_t)$. Moreover, $\langle X+Y\rangle = \langle X\rangle + \langle Y\rangle$.*

There is a Cauchy inequality of it, called Kunita-Watanabe inequality. For $M, N \in \mathcal{M}_0^{loc,c}$, and two measurable process H and K,

$$\int_0^\infty |H_t|\,|K_t|\,d\,|\langle M, N\rangle|_t \leqslant \left(\int_0^\infty H_t^2 d\langle M\rangle_t\right)^{\frac{1}{2}} \left(\int_0^\infty K_t^2 d\langle N\rangle_t\right)^{\frac{1}{2}}$$

Moreover, for $p, q \in [1,\infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}\left[\int_0^\infty |H_t|\,|K_t|\,d\,|\langle M, N\rangle|_t\right] \leqslant \left\|\left(\int_0^\infty H_t^2 d\langle M\rangle_t\right)^{\frac{1}{2}}\right\|_{L^p} \left\|\left(\int_0^\infty K_t^2 d\langle N\rangle_t\right)^{\frac{1}{2}}\right\|_{L^q}$$

- Integration of Local Martingale: For a continuous local martingale M, let $H \in L(M)$ be

$$\int_0^t H_u^2 d\langle M\rangle_u < \infty,\ \forall\ t \geqslant 0\ a.e.$$

For $H \in L(M)$, let $\{\tau_n\}_{n\in\mathbb{N}}$ be a sequence of stopping time deined as

$$\tau_n := \inf\{t \geqslant 0\colon\ |M_t| \geqslant n\} \wedge \inf\left\{t \geqslant 0\colon\ \int_0^t H_u^2 d\langle M\rangle_u\right\}$$

Note that $M^{\tau_n} \in \mathcal{M}_0^{loc,c}$ and $H\mathbb{1}_{[0,\tau_n]} \in \mathcal{L}^2(M^{\tau_n})$. So $H\mathbb{1}_{[0,\tau_n]} \cdot M^{\tau_n}$ is well-defined with the property

$$H\mathbb{1}_{[0,\tau_n]} \cdot M^{\tau_n}|_{[0,\tau_n]} = H\mathbb{1}_{[0,\tau_{n+1}]} \cdot M^{\tau_{n+1}}|_{[0,\tau_n]}$$

So let denote $H \cdot M$ with stopping times $\{\tau_n\}_{n\in\mathbb{N}}$,

$$(H\cdot M)^{\tau_n} := H\mathbb{1}_{[0,\tau_n]} \cdot M^{\tau_n}$$

So $H \cdot M \in \mathcal{M}_0^{loc,c}$.

Theorem 9.5. *For $M \in \mathcal{M}_0^{loc,c}$ and $H \in L(M)$, there exists a stochastic process $H \cdot M \in \mathcal{M}_0^{loc,c}$ with the following properties:*

(1) If $M \in \mathcal{M}_0^{2,c}$ and $H \in \mathcal{L}^2(M)$, $H \cdot M$ is coincide with the $\mathcal{L}^2(M)$-integral.

(2) For each stopping time τ, $(H \cdot M)^\tau = H\mathbb{1}_{[0,\tau]} \cdot M^\tau$.

Proposition 9.6. *For $M, L \in \mathcal{M}_0^{loc,c}$ and $H \in L(M)$,*

$$L = H \cdot M \;\Leftrightarrow\; \langle L, N \rangle = \int_0^\cdot H_u d\langle M, N \rangle_u, \;\forall\; N \in \mathcal{M}_0^{loc,c}$$

- Let $\mathcal{V}_0^c$ be the set of all continuous and adapted processes with paths of finite variation which vanish at $t = 0$.

  A stochastic process X is called a continuous semi-martingale if there exist processes $A \in \mathcal{V}_0^c$ and $M \in \mathcal{M}_0^{loc,c}$ such that

$$X_t = X_0 + M_t + A_t, \;\forall\; t \geqslant 0 \; a.e.$$

  It is also called the semi-martingale decomposition.

  For $X = X_0 + M + A$, we can see

$$\langle X \rangle^{\Delta_n} \xrightarrow{ucp} \langle M \rangle$$

  for any $\{\Delta_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_{[0,\infty)}$ with $\Delta_n \to Id$. Then define

$$\langle X \rangle_t := \langle M \rangle_t$$

  And let $H \in L(X)$ *s.t.*

$$\int_0^t |H_u|\, d\,|A|_u + \int_0^t |H_u|^2\, d\langle M \rangle_u < \infty$$

  Then for $H \in L(X)$,

$$(H \cdot X)_t := (H \cdot A)_t + (H \cdot M)_t$$

  So $H \cdot X$ is also a semi-martingale.

Proposition 9.7. *Let X be a semi-martingale with $X = A + M$.*

(1) Both $H \mapsto H \cdot X$ and $X \mapsto H \cdot X$ are linear.

(2) For $H \in L(X)$ and $K \in L(H \cdot X)$, then $KH \in L(X)$ and

$$KH \cdot X = K \cdot (H \cdot X)$$

(3) For $H \in L(X)$ and a stopping time T, we have $H^T \in L(X)$, X^T is also a semi-martingale and

$$(H \cdot X)^T = H\mathbb{1}_{[0,T]} \cdot X = H \cdot X^T$$

(4) For $H \in \mathcal{H}_{simp}$ with $H = \sum_{n=0}^{\infty} K_n \mathbb{1}_{(t_n,t_{n+1}]}(t)$, we have $H \in L(X)$ and

$$(H \cdot X)_t = \sum_{n=0}^{\infty} K_n (X_{t\wedge t_{n+1}} - X_{t\wedge t_n})$$

- Itô's formula: First the one-dimensional case is as the following.

Theorem 9.8. *Let X be a continuous semi-martingale taking values in $[a,b]$ and let $f\colon [a,b] \to \mathbb{R}$ in $\mathcal{C}^2$. Then $f(X)$ is also a continuous semi-martingale and*

$$f(X_t) - f(X_0) = \int_0^t f'(X_u)dM_u + \int_0^t f'(X_u)dA_u + \frac{1}{2}\int_0^t f''(X_u)d\langle M\rangle_u$$

Example 9.9. For example, let

$$X_t = X_0 + \int_0^t u_s ds + \int_0^t v_s dB_s$$

Then it is a continuous semi-martingale with $A_t = \int_0^t u_s ds$ and $M_t = \int_0^t v_s dB_s$. So $\langle M\rangle_t = \int_0^t v_s^2 ds$. Then we can see

$$dA_t = u_t dt,\ dM_t = v_t dB_t,\ d\langle M\rangle_t = v_t^2 dt$$

Therefore, the Itô's formula provides

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)v_s dB_s + \int_0^t f'(X_s)u_s ds + \frac{1}{2}\int_0^t f''(X_s)v_s^2 ds$$

which is the Itô's formula appeared in above.

Proposition 9.10. *(1) If $M \in \mathcal{M}_0^{2,c}$, then*

$$M \cdot M = \frac{1}{2}M^2 - \frac{1}{2}\langle M\rangle$$

(2) Let X and Y be two continuous semi-martingales. Then XY is also a semi-martingale and

$$X_tY_t = X_0Y_0 + \int_0^t Y_u dX_u + \int_0^t X_u dY_u + \langle X, Y\rangle_t$$

Note that the above two results can be proved without using the Itô's formula.

Remark. If

$$X_t = X_0 + \int_0^t u_s^x ds + \int_0^t v_s^x dB_s$$
$$Y_t = Y_0 + \int_0^t u_s^y ds + \int_0^t v_s^y dB_s$$

by the definition of $\langle\cdot\rangle$ for the semi-martingale

$$\langle X\rangle_t = \left\langle \int_0^t v_s^x dB_s \right\rangle_t = \int_0^t (v_s^x)^2 ds$$

and so

$$\langle Y\rangle_t = \left\langle \int_0^t v_s^x dB_s \right\rangle_t = \int_0^t (v_s^y)^2 ds$$

Moreover, since

$$X_t + Y_t = \int_0^t (u_s^x + u_s^y)ds + \int_0^t (v_s^x + v_s^y)dB_s$$

$\langle X+Y\rangle_t = \int_0^t (v_s^x + v_s^y)^2 ds$. So

$$\langle X, Y\rangle_t = \int_0^t v_s^x v_s^y ds$$

Remark. Let's consider a more general case. For

$$X_t = X_0 + \int_0^t u_s^x ds + \int_0^t v_s^x dB_s^{(1)}$$
$$Y_t = Y_0 + \int_0^t u_s^y ds + \int_0^t v_s^y dB_s^{(2)}$$

the quadratic covariation of X_t, Y_t is

$$d\langle X, Y\rangle_t = v^x v^y d\left\langle B^{(1)}, B^{(2)}\right\rangle_t$$

And for $d\left\langle B^{(1)}, B^{(2)}\right\rangle_t$, there are following three cases:

(i) $B_t^{(1)}$ and $B_t^{(2)}$ are independent,

$$d\left\langle B^{(1)}, B^{(2)}\right\rangle_t = 0$$

(ii) $B_t^{(1)} = B_t^{(2)} = B_t$,

$$d\langle B, B\rangle_t = dt$$

(iii) $B_t^{(1)}$ and $B_t^{(2)}$ have the correlation ρ_t,

$$d\left\langle B^{(1)}, B^{(2)}\right\rangle_t = \rho_t dt$$

Theorem 9.11. *Let $X^1, X^2, \cdots, X^d$ be d continuous semi-martingales and let $D \subset \mathbb{R}^d$ be an open set such that $\boldsymbol{X}_t = (X_t^1, X_t^2, \cdots, X_t^d) \in D$ for all $t \geqslant 0$ a.e.. Moreover, let $f\colon D \to \mathbb{R}$ be $\mathcal{C}^2$. Then the process $\{f(\boldsymbol{X}_t)\}_{t\geqslant 0}$ is a continuous semi-martingale and*

$$f(\boldsymbol{X}_t) - f(\boldsymbol{X}_0) = \sum_{k=1}^d \int_0^t \frac{\partial f}{\partial x_k}(\boldsymbol{X}_u) dX_u^k + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f(\boldsymbol{X}_u)}{\partial x_i \partial x_j} d\left\langle X^i, X^j\right\rangle_u$$

Or in the differential formula,

$$\begin{aligned} df(\boldsymbol{X}_t) &= \sum_{k=1}^d \frac{\partial f}{\partial x_k}(\boldsymbol{X}_t) dX_t^k + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 f(\boldsymbol{X}_t)}{\partial x_i \partial x_j} d\left\langle X^i, X^j\right\rangle_t \\ &= \nabla f(\boldsymbol{X}_t)\cdot d\boldsymbol{X}_t + \frac{1}{2}\operatorname{Tr}\left(\operatorname{Hess} f(\boldsymbol{X}_t) d\langle \boldsymbol{X}\rangle_t\right) \end{aligned}$$

where $d\langle \boldsymbol{X}\rangle_t = (d\langle X^i, X^j\rangle_t)_{ij}$.

Example 9.12. Let

$$X_t^i = X_0^i + \int_0^t u_s^{(i)} ds + \sum_{j=1}^m \int_0^t V_s^{(ij)} dB_s^{(j)}$$

where $\boldsymbol{B}_t = \left(B_t^{(1)}, B_t^{(2)}, \cdots, B_t^{(m)}\right)$ is a standard m-dimensional Brownian motion (independent).

So by Lemma 9.4, for each $i = 1, 2, \cdots, d$, $(X_t^i)_{t\geqslant 0}$ is a continuous semi-martingale. And

$$\begin{aligned} 2\left\langle X^i, X^j\right\rangle_t &= \left\langle X^i + X^j\right\rangle_t - \left\langle X^i\right\rangle_t - \left\langle X^j\right\rangle_t \\ &= \sum_{k=1}^m \int_0^t \left(V^{(ik)} + V^{(jk)}\right)^2 dt - \sum_{k=1}^m \int_0^t (V^{(ik)})^2 dt - \sum_{k=1}^m \int_0^t (V^{(jk)})^2 dt \\ &= 2\sum_{k=1}^m \int_0^t \left(V^{(ik)} V^{(jk)}\right) dt \end{aligned}$$

Therefore, by above Itô's formula, it can get the same formula as above.

Example 9.13. Consider when V is square and diagonal, that is

$$\begin{aligned} dX_t^{(1)} &= u_t^{(1)}dt + v_t^{(1)}dB_t^{(1)} \\ dX_t^{(2)} &= u_t^{(2)}dt + v_t^{(2)}dB_t^{(2)} \\ &\vdots \\ dX_t^{(d)} &= u_t^{(d)}dt + v_t^{(d)}dB_t^{(d)} \end{aligned}$$

Here we don't need $B_t^{(1)}, B_t^{(2)}, \cdots, B_t^{(d)}$ are not necessary independent. Let $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)}, \cdots, X_t^{(d)})$. Then we have the

$$\begin{aligned} df(\boldsymbol{X}_t) &= \nabla f(\boldsymbol{X}_t) \cdot d\boldsymbol{X}_t + \frac{1}{2}\operatorname{Tr}\left(\operatorname{Hess} f(\boldsymbol{X}_t) d\left\langle \boldsymbol{X}_t \right\rangle\right) \\ &= \nabla f(\boldsymbol{X}_t) \cdot d\boldsymbol{X}_t + \frac{1}{2}\sum_{i,j=1}^{d} v_t^{(i)} v_t^{(j)} \frac{\partial^2 f}{\partial x_i x_j}(\boldsymbol{X}_t) d\left\langle B^{(i)}, B^{(j)} \right\rangle_t \end{aligned}$$

Theorem 9.14 (Characterization of Brownian Motion)**.** *Let $M \in \mathcal{M}_0^{loc,c}$. If $\langle M \rangle_t = t$ for all t* a.e.. *Then M is an $(\mathcal{F}_t)_{t\geqslant 0}$ Brownian motion.*

- Itô Process: For two process u_t, v_t with $\int_0^t (|u_s| + v_s^2)ds < \infty$,

$$X_t = X_0 + \int_0^t u_s ds + \int_0^t v_s dB_s$$

is called a Itô process. If $H \in L(X)$ is progressively measurable, then $H \cdot X$ is an Itô process. If Y is also an Itô process and $F\colon [0,\infty) \times \mathbb{R}^2 \to \mathbb{R}$ is $\mathcal{C}^{1,2,2}$, then $F(t, X_t, Y_t)$ is also Itô.

Let X be an Itô process. If

$$u_t = \mu(t, X_t), \;\; v_t = \sigma(t, X_t)$$

for some measurable functions μ, σ, then X is called an inhomogeneous diffusion process. If μ and σ do not depend on t, it is called a homogeneous diffusion process.

10 Stratonovich Integral

- Definition: Let $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ be two semi-martingales. Then the Stratonovich integral

$$\int_0^t Y_s \circ dX_s := \int_0^t Y_s dX_s + \frac{1}{2}\left\langle X, Y \right\rangle_t$$

Or denoting $\circ dX_t$ by ∂X_t,

$$\int_0^t Y_s \partial X_s := \int_0^t Y_s dX_s + \frac{1}{2}\left\langle X, Y \right\rangle_t$$

and the differential form

$$Y_t \partial X_t = Y_t dX_t + \frac{1}{2} d\left\langle X, Y \right\rangle_t$$

In fact,

$$\int_0^t Y_s \circ dX_t = \lim_{|\Delta_n| \to 0} \sum_{i=1}^{n} \frac{Y_{t_i} + Y_{t_{i-1}}}{2}(X_{t_i} - X_{t_{i-1}})$$

Remark. By the definition, it induces

$$\int_0^t X_s \circ dX_s = \frac{1}{2} X_t^2$$

- Properties: By the definition, it has the simplified

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s \partial Y_s + \int_0^t Y_s \partial X_s$$

that is

$$\partial X_t Y_t = X_t \partial Y_t + Y_t \partial X_t$$

Besides, because of the Itô's formula,

$$f(\boldsymbol{X}_t) - f(\boldsymbol{X}_0) = \sum_{k=1}^{d} \int_0^t \frac{\partial f}{\partial x_k}(\boldsymbol{X}_s) \partial X_s$$

or equivalently

$$\partial f(\boldsymbol{X}_t) = \sum_{k=1}^{d} \frac{\partial f}{\partial x_k}(\boldsymbol{X}_t) \partial X_t \tag{1}$$

Remark. Therefore, if X_t satisfies the SDE in Stratonovich sense

$$d\boldsymbol{X}_t = \boldsymbol{\mu}(\boldsymbol{X}_t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t)\partial \boldsymbol{B}_t$$

then it is equivalent to the SDE

$$d\boldsymbol{X}_t = \left[\boldsymbol{\mu}(\boldsymbol{X}_t) + \frac{1}{2}\boldsymbol{a}(\boldsymbol{X}_t) \nabla \cdot \boldsymbol{\sigma}(\boldsymbol{X}_t)\right] dt + \boldsymbol{\sigma}(\boldsymbol{X}_t) d\boldsymbol{B}_t$$

where

- $\boldsymbol{a}(\boldsymbol{X}_t) = \boldsymbol{\sigma}(\boldsymbol{X}_t)\boldsymbol{\sigma}(\boldsymbol{X}_t)^\top$ is the diffusion matrix,
- $\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{X}_t)$ is the divergence of the matrix $\boldsymbol{\sigma}(\boldsymbol{X}_t)$ with respect to the spatial variables.

Or equivalently,

$$dX_t^i = \left[\mu^i(\boldsymbol{X}_t) + \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{n} \sigma^{ik}(\boldsymbol{X}_t) \frac{\partial \sigma^{jk}(\boldsymbol{X}_t)}{\partial X^k}\right] dt + \sum_{j=1}^{m} \sigma^{ij}(\boldsymbol{X}_t) dB_t^j$$

It is because of $d\langle X_t, Y_t\rangle = dX_t dY_t$.

11 Representations of Martingales

- Right-continuous inverses: Let A_0 be the set of all càdlàg and nondecreasing functions $f\colon [0,\infty) \to [0,\infty]$ with $f(\infty) := \lim_{t\to\infty} f(t)$. The right-continuous inverse $g\colon [0,\infty) \to [0,\infty]$ is given by

$$g(s) := \inf\{t \geqslant 0\colon f(t) > s\}$$

and denoted by $g = f^{-1}$. Then we have the following statements.

(1) $g = f^{-1} \in A_0$.

(2) $g^{-1} = f$.

(3) $g(f(t)) = \sup \mathrm{Flag}_f(t)$, where $\mathrm{Flag}_f(t) := \{t' \geqslant 0\colon f(t) = f(t')\}$.

(4) $f(g(s)) = \sup \mathrm{Vert}_f(s)$, where for each $s \in f([0,\infty))$, $\mathrm{Vert}_f(s)$ is the set of all values in the interval of the form $[f(t-), f(t)]$ that contains s. In particular, if f is continuous, then $f(g(s)) = s$.

A stochastic process $\{\tau_s\}_{s\geqslant 0}$ with trajectories in A_0 is called a time change if τ_s is a stopping time for each s.

(1) For a filtration $(\mathcal{F}_t)_{t\geqslant 0}$, $(\mathcal{G}_s)_{s\geqslant 0}$ defined by $\mathcal{G}_s = \mathcal{F}_{\tau_s}$ is called the time change filtration.

(2) For an $(\mathcal{F}_t)_{t\geqslant 0}$-adapted and a predictable measurable process $X = (X_t)_{t\geqslant 0}$, $(X_{\tau_s})_{s\geqslant 0}$ is called the time change of X.

Note that $(\mathcal{F}_t)_{t\geqslant 0}$ right-continuous implies $(\mathcal{G}_s)_{s\geqslant 0}$ right-continuous and $(X_t)_{t\geqslant 0}$ càdlàg implies $(X_{\tau_s})_{s\geqslant 0}$ càdlàg.

Proposition 11.1. *If σ is a $(\mathcal{G}_s)_{s\geqslant 0}$-stopping time, then the random variable τ_σ is a $(\mathcal{F}_t)_{t\geqslant 0}$-stopping time.*

Let $\{\tau_s\}_{s\geqslant 0}$ be a time-change. A process $(X_t)_{t\geqslant 0}$ is said to be τ-continuous if it is continuous and is constant on $[\tau_{s-}, \tau_s]$ for all $s \geqslant 0$ *a.e.*. If $\{\tau_s\}_{s\geqslant 0}$ is a time-change and $(X_t)_{t\geqslant 0}$ is τ-continuous, then it is clear $(X_{\tau_s})_{s\geqslant 0}$ is continuous.

Proposition 11.2. *Let $\{\tau_s\}_{s\geqslant 0}$ be a time-change and $M \in \mathcal{M}_0^{loc,c}$ be τ-continuous. Then the time-changed process $(M_{\tau_s})_{s\geqslant 0}$ is a continuous local martingale for the time-changed filtration $(\mathcal{G}_s)_{s\geqslant 0}$.*

Theorem 11.3. *Let $(X_t)_{t\geqslant 0}$ be a càdlàg semi-martingale and $\{\tau_s\}_{s\geqslant 0}$ be a time-change. Then $(X_{\tau_s})_{s\geqslant 0}$ is a càdlàg semi-martingale.*

Theorem 11.4. *Let $(X_s)_{s\geqslant 0}$ be a càdlàg semi-martingale. Then there is a filtered probability space $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\geqslant 0}, \mathbb{P}')$ which supports an $(\mathcal{F}'_t)_{t\geqslant 0}$-Brownian motion $(B_t)_{t\geqslant 0}$ such that $(X_s)_{s\geqslant 0}$ and $(B_{\tau_s})_{s\geqslant 0}$ has same finite-dimensional distribution.*

- Let $(M_t)_{t\geqslant 0}$ be a continuous local martingale with quadratic variation $\langle M\rangle$ and

$$\tau_s := \inf\{t \geqslant 0\colon \langle M\rangle_t > s\}$$

If $(M_t)_{t\geqslant 0}$ is divergent, *i.e.* $\langle M\rangle_\infty = \infty$, then τ will not take value ∞. Note that τ is τ-continuous.

Theorem 11.5. *For a divergent $M \in \mathcal{M}_0^{loc,c}$, we define $\tau_s := \inf\{t \geqslant 0\colon \langle M\rangle_t > s\}$ and $\mathcal{G}_s = \mathcal{F}_{\tau_s}$. Then the time-changed process $(B_s)_{s\geqslant 0}$ given by*

$$B_s = M_{\tau_s},\ s \geqslant 0$$

is a $(\mathcal{G}_s)_{s\geqslant 0}$-Brownian motion and the local martingale is a time-change of B, i.e.

$$M_t = B_{\langle M\rangle_t},\ \forall\, t \geqslant 0$$

- Let $M = (M_t)_{t\geqslant 0}$ be a continuous local martingale. Then a continuous local martingale $(Z_t)_{t\geqslant 0}$ is called the stochastic exponential of M if

$$dZ_t = Z_t dM_t$$

By the Itô formula,

$$Z_t = \mathcal{E}(M)_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right)$$

defines a stochastic exponential of M.

Remark. When M is a continuous local martingale, for any $\lambda \in \mathbb{R}$,

$$\mathcal{E}(\lambda M)_t = \exp\left(\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\right)$$

then $\mathcal{E}(\lambda M)$ is also a local martingale and by the following,

$$\mathcal{E}(\lambda M)_t = e^{\lambda M_0} + \lambda \int_0^t \mathcal{E}(\lambda M)_s dM_s$$

Example 11.6. Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then

$$\mathcal{E}(B)_t = \exp\left(B_t - \frac{1}{2}t\right)$$

Example 11.7. Let $B = (B_t)_{t\geq 0}$ be a Brownian motion and $(\theta_t)_{t\geq 0}$ be a predictable process in $\mathcal{L}^2(B)$. Let

$$M_t = \int_0^t \theta_u dB_u, \quad Z = \mathcal{E}(M) = \mathcal{E}\left(\int_0^\cdot \theta_u dB_u\right)$$

Then we have

$$\mathcal{E}\left(\int_0^\cdot \theta_u dB_u\right)_t = \exp\left(\int_0^t \theta_u dB_u - \frac{1}{2}\int_0^t \theta_u^2 du\right)$$

Moreover, if

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

then

$$(\mathcal{E}(M)_t)_{t\in[0,T]}$$

is a martingale with

$$\mathbb{E}\left[\mathcal{E}(M)_t\right] = 1, \quad \forall\, t < T$$

Theorem 11.8. *Let $(B_t)_{t\geqslant 0}$ be a Brownian motion with natural filtration $\mathcal{F}^B = \mathcal{F}$. Let $\mathcal{I}$ be the set of simple functions from $[0,\infty)$ to $\mathbb{R}$. Then*

$$E = \left\{\mathcal{E}\left(\int_0^\cdot B_u du\right)_\infty : f \in \mathcal{I}\right\}$$

is total in $\mathcal{L}(\mathcal{F}_\infty)$, i.e. *the linear expansion of E is dense.*

Theorem 11.9. *Let $(B_t)_{t\geqslant 0}$ be a Brownian motion with natural filtration $\mathcal{F}^B = \mathcal{F}$.*

$$\mathcal{F}_\infty = \sigma(\mathcal{F}_t \colon t \geqslant 0)$$

*For any random variable $X \in \mathcal{L}^2(\mathcal{F}_\infty)$, there is a unique (*a.e.*) predictable process $(H_t)_{t\geqslant 0}$ such that $\mathbb{E}\left[\int_0^\infty H_t^2 dt\right] < \infty$ and*

$$X = \mathbb{E}[X] + \int_0^\infty H_t dB_t, \ \textit{a.e.}$$

Theorem 11.10. *Let $(B_t)_{t\geqslant 0}$ be a Brownian motion with natural filtration $\mathcal{F}^B$ and $\mathcal{F}$ be its augmentation.*

(1) Let M be a $\mathcal{L}^2$-bounded, càdlàg $\mathcal{F}$-martingale. Then there is a unique predictable process $H \in \mathcal{L}^2(B)$ i.e. $\mathbb{E}\left[\int_0^\infty H_t^2 dt\right] < \infty$ such that

$$M_t = M_0 + \int_0^t H_u dB_u, \ \forall\ t \geqslant 0$$

(2) If M is a càdlàg $\mathcal{F}$-local martingale, then there is a predictable process $H \in L(B)$ such that

$$M_t = M_0 + \int_0^t H_u dB_u, \ \forall\ t \geqslant 0$$

12 Girsanov's Theorem

Consider filtered probability space $(\Omega, \mathcal{F} = (\mathcal{F}_t), \mathbb{P})$.

Definition 12.1. Let $\mathbb{Q}$ be a probability measure on $\mathcal{F}$ and be equivalent to $\mathbb{P}$. Then let

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}} \in L^1, \quad \mathbb{E}[Z] = 1$$

Then the uniformly-integrable martingale

$$Z_t = \mathbb{E}[Z \mid \mathcal{F}_t], \ t \geq 0$$

is called the density of $\mathbb{Q}$ with respect to $\mathbb{P}$.

Proposition 12.2. *Let $X = (X_t)_{t\geq 0}$ be a càdlàg and adapted process. Then X is a $\mathbb{Q}$-martingale if and only if $(Z_t X_t)_{t\geq 0}$ is a càdlàg $\mathbb{P}$-local martingale.*

Lemma 12.3. *Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, and $\mathcal{G} \subset \mathcal{H}$ be a sub-σ-algebra. Given a probability measure $\mathbb{Q}$ that is equivalent $\mathbb{P}$, let $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$. Then for any $X \in L^1(\mathbb{Q})$, $XZ \in L^1(\mathbb{P})$ and*

$$\mathbb{E}^{\mathbb{Q}}\left[X \mid \mathcal{G}\right] = \frac{1}{\mathbb{E}[Z \mid \mathcal{G}]}\mathbb{E}[XZ \mid \mathcal{G}]$$

Example 12.4. Setting $\mathcal{G} = \langle B \rangle$ and $X = \mathbb{1}_A$.

$$\mathbb{E}^{\mathbb{Q}}\left[X \mid \mathcal{G}\right] = \mathbb{Q}(A \mid B)\mathbb{1}_B + \mathbb{Q}(A \mid B^c)\mathbb{1}_{B^c}$$

and

$$\mathbb{E}[Z \mid \mathcal{G}] = \frac{\mathbb{Q}(B)}{\mathbb{P}(B)}\mathbb{1}_B + \frac{\mathbb{Q}(B^c)}{\mathbb{P}(B^c)}\mathbb{1}_{B^c}$$

and

$$\mathbb{E}[XZ \mid \mathcal{G}] = \frac{\mathbb{Q}(AB)}{\mathbb{P}(B)}\mathbb{1}_B + \frac{\mathbb{Q}(AB^c)}{\mathbb{P}(B^c)}\mathbb{1}_{B^c}$$

Therefore, the lemma basically said

$$\mathbb{Q}(A \mid B) = \frac{1}{\frac{\mathbb{Q}(B)}{\mathbb{P}(B)}}\frac{\mathbb{Q}(AB)}{\mathbb{P}(B)}$$

Proposition 12.5. *Suppose that the density process $(Z_t)_{t\geq 0}$ is continuous. Let X be a continuous $\mathbb{P}$-semi-martingale with the decomposition*

$$X = X_0 + M + A$$

Then X is also a $\mathbb{Q}$-semi-martingale with the decomposition $X = X_0 + N + B$, where

$$N = M - F, \quad B = A + F$$

and

$$F_t = \int_0^t \frac{1}{Z_s} d\langle M, Z\rangle_s$$

Theorem 12.6 (Girsanov)**.** *Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the usual augmentation of the natural filtration generated by a Brownian motion $(B_t)_{t\geq 0}$.*

I. Let $\mathbb{Q} \sim \mathbb{P}$ be a probability measure on $\mathcal{F}$ and $(Z_t)_{t\geq 0}$ be the density process $Z_t = \mathbb{E}[Z \mid \mathcal{F}_t]$. Then there is a predictable process $(\theta_t)_{t\geq 0}$ in $\mathcal{L}^2(B)$ such that

$$Z = \mathcal{E}\left(\int_0^\cdot \theta_u dB_u\right)$$

and

$$B_t - \int_0^t \theta_u du$$

is a $\mathbb{Q}$-Brownian motion.

II. Conversely, let $(\theta_t)_{t\geq 0}$ in $\mathcal{L}^2(B)$ with the property $Z = \mathcal{E}\left(\int_0^\cdot \theta_u dB_u\right)$ a uniformly-integrable martingale of $Z_\infty > 0$ a.s. Then for any $\mathbb{Q} \sim \mathbb{P}$ such that $\mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \mathcal{F}_\infty] = Z_\infty$,

$$B_t - \int_0^t \theta_u du$$

is a $\mathbb{Q}$-Brownian motion.

Remark. Usually, we consider the Girsanov's theorem on finite interval $[0, T]$. Here is an example, let $\theta_u = \mu$.

$$Z_t = \exp\left(\mu B_t - \frac{1}{2}\mu^2 t\right)$$

Because this process is UI on $[0, T]$, there is a probability measure $\mathbb{P}^{\mu,T}$ on $\mathcal{F}_T$ with the property

$$B_t - \mu t$$

a $\mathbb{P}^{\mu,T}$-Brownian motion and

$$\frac{d\mathbb{P}^{\mu,T}}{d\mathbb{P}} = \exp\left(\mu B_T - \frac{1}{2}\mu^2 T\right)$$

Theorem 12.7 (Novikov's criterion)**.** *If $M \in \mathcal{M}_0^{loc,c}$ is such that*

$$\mathbb{E}[e^{\frac{1}{2}\langle M\rangle_\infty}] < \infty$$

then $\mathcal{E}(M)$ is a UI martingale.

Let's see an easier expression of the Girsanov's theorem that is more understandable for the following applications.

Theorem 12.8 (Liptser and Shiryaev 2013)**.** *For $t \in [0,T]$, let $M_t = \int_0^t \theta_u dB_u$ where B is a $\mathbb{P}$-Brownian motion. Assume that $\theta_t \in \mathcal{L}^2(B)$. Then M is a $\mathbb{P}$-martingale in $\mathcal{L}^2(\mathbb{P})$. Moreover, if*

$$\mathbb{E}_{\mathbb{P}}[\mathcal{E}(M)_T] = 1, \quad \mathcal{E}(M)_t = \exp\left(\int_0^t \theta_u dB_u - \frac{1}{2}\int_0^t \|\theta_u\|^2 du\right)$$

(such that Novikov's condition can be satisfied). Then the process

$$t \mapsto B_t - \int_0^t \theta_u du$$

is a $\mathbb{Q}$-Brownian motion for

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(M)_T$$

Remark. In fact, under above condition, $\mathcal{E}(M)$ is also a $\mathbb{P}$-martingale (Le Gall 2016).

Here is an applications of the Girsanov's theorem to measuring the bound of the difference of generated distributions of two SDEs with same Brownian part but different drift terms.

Consider a filtered probability space $(\Omega, \mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, where $\{\mathcal{F}_t\}_{t\geq 0}$ be the usual augmentation of the natural filtration generated by a Brownian motion $(B_t)_{t\geq 0}$. For the following two SDEs,

$$\begin{aligned} dX_t^{(1)} &= b_t^{(1)}(X_t^{(1)})dt + \sqrt{2}dB_t, \quad X_0^{(1)} \sim \rho_0 \\ dX_t^{(2)} &= b_t^{(2)}(X_t^{(2)})dt + \sqrt{2}dB_t, \quad X_0^{(2)} \sim \rho_0 \end{aligned}$$

let $\mu^{(i)} = (X_T^{(i)})_\#\mathbb{P}$ be the distribution of $X_T^{(i)}$ for $i = 1, 2$. There are two proofs to estimate the bound of $\mathrm{KL}(\mu^{(1)}\|\mu^{(2)})$.

(1) Let

$$\theta_t = \frac{1}{\sqrt{2}}\left(b_t^{(1)}(X_t^{(2)}) - b_t^{(2)}(X_t^{(2)})\right)$$

and $M_t = \int_0^t \theta_u du$. Then by the Girsanov's theorem (when the conditions are satisfied), there is a new probability measure $\mathbb{Q}$ such that

$$W_t = B_t - \frac{1}{\sqrt{2}}\int_0^t \left(b_u^{(1)}(X_u^{(2)}) - b_u^{(2)}(X_u^{(2)})\right) du$$

a $\mathbb{Q}$-Brownian motion. Then we have

$$\sqrt{2}dB_t = \sqrt{2}dW_t + \left(b_t^{(1)}(X_t^{(2)}) - b_t^{(2)}(X_t^{(2)})\right) dt$$

and thus

$$\begin{aligned} dX_t^{(2)} &= b_t^{(2)}(X_t^{(2)})dt + \sqrt{2}dB_t \\ &= b_t^{(1)}(X_t^{(2)})dt + \sqrt{2}dW_t \end{aligned}$$

So this equation under $\mathbb{Q}$ is as same as the first equation under $\mathbb{P}$ (by the uniqueness of solution of SDE (Liptser and Shiryaev 2013)). Therefore,

$$\begin{aligned} \mu^{(1)} &= (X_T^{(1)})_\#\mathbb{P} \\ \mu^{(2)} &= (X_T^{(2)})_\#\mathbb{P} = (X_T^{(1)})_\#\mathbb{Q} \end{aligned}$$

which implies

$$\mathrm{KL}(\mu^{(1)} \| \mu^{(2)}) \leq \mathrm{KL}(\mathbb{P} \| \mathbb{Q}) = \mathbb{E}^{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}}\right]$$

To calculate this, by the Girsanov's theorem, we have

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\int_0^T \theta_t dB_t - \frac{1}{2}\int_0^T \|\theta_t\|^2 dt\right) \Rightarrow \frac{d\mathbb{P}}{d\mathbb{Q}} = \exp\left(-\int_0^T \theta_t dB_t + \frac{1}{2}\int_0^T \|\theta_t\|^2 dt\right)$$

Therefore,

$$\begin{aligned}
\mathrm{KL}(\mu^{(1)} \| \mu^{(2)}) &\leq \mathbb{E}^{\mathbb{P}}\left[-\int_0^T \theta_t dB_t + \frac{1}{2}\int_0^T \|\theta_t\|^2 dt\right] \\
&= \frac{1}{4}\int_0^T \mathbb{E}^{\mathbb{P}}\left[\left\|b_t^{(1)}(X_t^{(2)}) - b_t^{(2)}(X_t^{(2)})\right\|^2\right] dt
\end{aligned}$$

Remark. Let $(\Omega, \mathcal{F})$ be a measurable set and $\mathbb{P} \ll \mathbb{Q}$ be two probability measures. Let $X : \Omega \to \mathbb{R}^d$ be a random variable with $\mathbb{P}_X = X_\#\mathbb{P}, \mathbb{Q}_X = X_\#\mathbb{Q}$. Then by the property that the relative entropy decreases under push-forward map (Lemma 9.4.5 in Ambrosio et al. 2008),

$$\mathrm{KL}(\mathbb{P}_X \| \mathbb{Q}_X) \leq \mathrm{KL}(\mathbb{P} \| \mathbb{Q})$$

Proof. First, $\mathbb{P} \ll \mathbb{Q}$ implies $\mathbb{P}_X \ll \mathbb{Q}_X$ by definition. Next, we prove that

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right] = \frac{d\mathbb{P}_X}{d\mathbb{Q}_X} \circ X$$

First, because $\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right]$ is $\sigma(X)$-measurable, there is a measurable function $h \colon \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right] = h(X)$. Then for any $B \in \mathcal{R}^d$,

$$\begin{aligned}
\int_{\mathbb{R}^d} \mathbb{1}_B \frac{d\mathbb{P}_X}{d\mathbb{Q}_X} d\mathbb{Q}_X &= \int_{\mathbb{R}^d} \mathbb{1}_B d\mathbb{P}_X \\
&= \int_\Omega \mathbb{1}_B \circ X d\mathbb{P} \\
&= \int_\Omega (\mathbb{1}_B \circ X) \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q}
\end{aligned}$$

Clearly, $\mathbb{1}_B \circ X$ is $\sigma(X)$-measurable so

$$\mathbb{E}^{\mathbb{Q}}\left[(\mathbb{1}_B \circ X)\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right] = (\mathbb{1}_B \circ X)\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right]$$

and by $\Omega \in \sigma(X)$,

$$\begin{aligned}
\int_{\mathbb{R}^d} \mathbb{1}_B \frac{d\mathbb{P}_X}{d\mathbb{Q}_X} d\mathbb{Q}_X &= \int_\Omega (\mathbb{1}_B \circ X) \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} \\
&= \int_\Omega \mathbb{E}^{\mathbb{Q}}\left[(\mathbb{1}_B \circ X)\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right] d\mathbb{Q} \\
&= \int_\Omega (\mathbb{1}_B \circ X)\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \;\middle|\; \sigma(X)\right] d\mathbb{Q} \\
&= \int_\Omega (\mathbb{1}_B \circ X)(h \circ X) d\mathbb{Q} \\
&= \int_{\mathbb{R}^d} \mathbb{1}_B h d\mathbb{Q}_X
\end{aligned}$$

Therefore, $\mathbb{Q}_X$-surely we have $h = \frac{d\mathbb{P}_X}{d\mathbb{Q}_X}$ and thus we have the desired result. By this and the Jensen's inequality of the conditional expectation for the convex function

$$\eta(x) = (x \log x)\mathbb{1}_{(0,\infty)}(x),$$

defined on $\mathbb{R}$, we have

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_X \| \mathbb{Q}_X) &= \int_{\mathbb{R}^d} \eta\left(\frac{d\mathbb{P}_X}{d\mathbb{Q}_X}\right) d\mathbb{Q}_X \\
&= \int_\Omega \eta\left(\frac{d\mathbb{P}_X}{d\mathbb{Q}_X} \circ X\right) d\mathbb{Q} \\
&= \int_\Omega \eta\left(\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\middle|\, \sigma(X)\right]\right) d\mathbb{Q} \\
&\leq \int_\Omega \mathbb{E}^{\mathbb{Q}}\left[\eta\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) \,\middle|\, \sigma(X)\right] d\mathbb{Q} \\
&= \int_\Omega \eta\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} = \mathrm{KL}(\mathbb{P}\|\mathbb{Q})
\end{aligned}$$

□

Remark. The next problem is, in order to apply the Girsanov's theorem, we need the Novikov's condition, or equivalently,

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T \|\theta_t\|^2 \, dt\right] < \infty, \quad \mathbb{E}^{\mathbb{P}}[\mathcal{E}(M)_T] = 1$$

For applications, our assumptions usually guarantee

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T \|\theta_t\|^2 \, dt\right] = \frac{1}{2}\mathbb{E}^{\mathbb{P}}\left[\int_0^T \left\|b_t^{(1)}(X_t) - b_t^{(2)}(X_t)\right\|^2 dt\right] \leq M < \infty$$

But it cannot guarantee $\mathbb{E}^{\mathbb{P}}[\mathcal{E}(M)_T] = 1$, equivalently $\mathcal{E}(M)$ is a $\mathbb{P}$-martingale. So we use the approximation technique introduced in Chen et al. 2023.

First, by the (3.4)Proposition of Chapter IV in Revuz and Yor 2013, we have known $\mathcal{E}(M)$ is a local martingale, which means there is a nondecreasing sequence of stopping times T_n with the property $T_n \uparrow T$ such that $(\mathcal{E}(M)_{t\wedge T_n})_{t\in[0,T_n]}$ is a $\mathbb{P}$-martingale(see the (1.5)Definition of Chapter IV in Revuz and Yor 2013). Besides, let $M^n = M^{T_n}$, that is

$$(M^n)_t := M_{t\wedge T_n} = \int_0^{t\wedge T_n} \theta_u du = \begin{cases} \int_0^t \theta_u dB_u, & t \leq T_n \\ \int_0^{T_n} \theta_u dB_u, & t > T_n \end{cases}$$

Therefore. by the definition of exponential of a martingale,

$$\mathcal{E}(M^n)_t = \begin{cases} \exp\left(\int_0^t \theta_u dB_u - \frac{1}{2}\int_0^t \|\theta_u\|^2 du\right), & t \leq T_n \\ \exp\left(\int_0^{T_n} \theta_u dB_u - \frac{1}{2}\int_0^{T_n} \|\theta_u\|^2 du\right), & t > T_n \end{cases}$$

and so $\mathcal{E}(M^n)_t = \mathcal{E}(M)_{t\wedge T_n}$. Note that

$$M_t^n = \int_0^{t\wedge T_n} \theta_u dB_u = \int_0^t \theta_u \mathbb{1}_{t\in[0,T_n]} dB_u$$

So martingale M^n satisfies the condition before applying the Girsanov's Theorem. There is a probability measure $\mathbb{Q}^n$ on Ω such that

$$W_t^n = B_t - \int_0^t \theta_u \mathbb{1}_{t\in[0,T_n]} du = B_t - \frac{1}{\sqrt{2}} \int_0^t \left(b_u^{(1)} - b_u^{(2)}\right) \mathbb{1}_{t\in[0,T_n]}(u) du$$

is a Brownian motion and we have

$$\begin{aligned}\frac{d\mathbb{Q}^n}{d\mathbb{P}} &= \exp\left(\int_0^T \theta_t \mathbb{1}_{t\in[0,T_n]} dB_t - \frac{1}{2}\int_0^T \|\theta_t\|^2 \mathbb{1}_{t\in[0,T_n]} dt\right) \\ &= \exp\left(\int_0^{T_n} \theta_t dB_t - \frac{1}{2}\int_0^{T_n} \|\theta_t\|^2 dt\right)\end{aligned}$$

which implies

$$\begin{aligned}\mathrm{KL}\left(\mathbb{P}\|\mathbb{Q}^n\right) = \mathbb{E}^{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}^n}\right] &= \int_0^{T_n} \mathbb{E}^{\mathbb{P}}\left[\|\theta_t\|^2\right] dt \\ &= \frac{1}{2}\int_0^{T_n} \mathbb{E}^{\mathbb{P}}\left[\left\|b_t^{(1)} - b_t^{(2)}\right\|^2\right] dt \\ &\leq \frac{1}{2}\int_0^{T} \mathbb{E}^{\mathbb{P}}\left[\left\|b_t^{(1)} - b_t^{(2)}\right\|^2\right] dt \leq M\end{aligned}$$

Next, reconsidering the second SDE

$$\begin{aligned}dX_t^{(2)} &= b_t^{(2)}(X_t^{(2)})dt + \sqrt{2}dB_t \\ &= b_t^{(1)}(X_t^{(2)})\mathbb{1}_{t\in[0,T_n]}(t)dt + b_t^{(2)}(X_t^{(2)})\mathbb{1}_{t\in[T_n,T]}(t)dt + \sqrt{2}dW_t^n, \quad X_0^{(2)} \sim \rho_0\end{aligned}$$

and the equation

$$dX_t^n = b_t^{(1)}(X_t^n)\mathbb{1}_{t\in[0,T_n]}(t)dt + b_t^{(2)}(X_t^n)\mathbb{1}_{t\in[T_n,T]}(t)dt + \sqrt{2}dB_t, \quad X_0^n \sim \rho_0$$

Let $\mu_n^{(1)}$ be the distribution of X_T^n under $\mathbb{P}$. But we can see it has the same formula as $X_t^{(2)}$ under $\mathbb{Q}^n$. So

$$\begin{aligned}\mu^{(2)} &= (X_T^{(2)})_\#\mathbb{P} \\ \mu_n^{(1)} &= (X_T^n)_\#\mathbb{P} = (X_T^{(2)})_\#\mathbb{Q}^n\end{aligned}$$

And by the decreasing property of the relative entropy under push-forward map,

$$\mathrm{KL}\left(\mu^{(2)}\|\mu_n^{(1)}\right) \leq \mathrm{KL}\left(\mathbb{P}\|\mathbb{Q}^n\right) \leq M$$

Note that for all $t \leq T_n$, $X_t^n = X_t^{(1)}$ by the uniqueness of solution of SDE. By the Lemma 13 in Chen et al. 2023, for any $\varepsilon > 0$,

$$\left(X_{t\wedge(T-\varepsilon)}^n\right)_{t\in[0,T]} \to \left(X_{t\wedge(T-\varepsilon)}^{(1)}\right)_{t\in[0,T]} \quad a.s., \quad \text{as } n \to \infty$$

Therefore, $X_{T-\varepsilon}^n \to X_{T-\varepsilon}^{(1)}$ $a.s.$ as $n \to \infty$. Let $\mu_{n,\varepsilon}^{(1)} = (X_{T-\varepsilon}^n)_\#\mathbb{P}$ and $\mu_\varepsilon^{(1)} = (X_{T-\varepsilon}^{(1)})_\#\mathbb{P}$. Then for any continuous and bounded f define on $\mathbb{R}^d$,

$$\int_{\mathbb{R}^d} f d\mu_{n,\varepsilon}^{(1)} = \int_\Omega f \circ X_{T-\varepsilon}^n d\mathbb{P} \ \to \ \int_\Omega f \circ X_{T-\varepsilon}^{(1)} d\mathbb{P} = \int_{\mathbb{R}^d} f d\mu_\varepsilon^{(1)}$$

as $n \to \infty$, which means $\mu_{n,\varepsilon}^{(1)} \to \mu_\varepsilon^{(1)}$ weakly as $n \to \infty$. Besides, let $\mu_\varepsilon^{(2)} = (X_{T-\varepsilon}^{(2)})_\#\mathbb{P}$. Then by the lower semicontinuity of KL divergence (Lemma 9.4.3 in Ambrosio et al. 2008),

$$\mathrm{KL}\left(\mu_\varepsilon^{(2)}\|\mu_\varepsilon^{(1)}\right) \leq \liminf_{n\to\infty} \mathrm{KL}\left(\mu_\varepsilon^{(2)}\|\mu_{n,\varepsilon}^{(1)}\right)$$

Similarly as above, by comparing the equation in W_t^n and B_t,

$$\mu_{n,\varepsilon}^{(1)} = (X_{T-\varepsilon}^n)_{\#}\mathbb{P} = (X_{T-\varepsilon}^{(2)})_{\#}\mathbb{Q}^n$$

So we have

$$\begin{aligned}\mathrm{KL}\left(\mu_\varepsilon^{(2)}\|\mu_\varepsilon^{(1)}\right) &\le \liminf_{n\to\infty} \mathrm{KL}\left((X_{T-\varepsilon}^{(2)})_{\#}\mathbb{P}\|(X_{T-\varepsilon}^{(2)})_{\#}\mathbb{Q}^n\right)\\ &\le \liminf_{n\to\infty} \mathrm{KL}\left(\mathbb{P}\|\mathbb{Q}^n\right)\\ &\le M\end{aligned}$$

And because $X_{T-\varepsilon}^{(i)} \to X_T^{(i)}$ *a.s.* as $\varepsilon \to 0^+$, $\mu_\varepsilon^{(i)} \to \mu^{(i)}$ weakly for $i = 1, 2$. Using the same property, we have our final result

$$\mathrm{KL}\left(\mu^{(2)}\|\mu^{(1)}\right) \le \liminf_{\varepsilon\to 0^+} \mathrm{KL}\left(\mu_\varepsilon^{(2)}\|\mu_\varepsilon^{(1)}\right) \le M$$

Therefore, even if the Novikov's condition cannot be fully satisfied, we have the same upper bound.

(2) Here is another method to estimate the same bound by considering the process in the path space.

Theorem 12.9 (Theorem 7.7 in Liptser and Shiryaev 2013). *Let $\xi = (\xi_t, \mathcal{F}_t), 0 \le t \le T$, be a process of diffusion type with the differential*

$$d\xi_t = \alpha_t(\xi)dt + dW_t, \quad \xi_0 = 0, \quad 0 \le t \le T.$$

Then

$$\left.\begin{matrix} P\left(\int_0^T \alpha_t^2(\xi)dt < \infty\right) = 1 \\ P\left(\int_0^T \alpha_s^2(W)dt < \infty\right) = 1 \end{matrix}\right\} \Leftrightarrow \mu_\xi \sim \mu_W \tag{2}$$

Here (P-a.s.)

$$\begin{aligned}\frac{d\mu_\xi}{d\mu_W}(t, W) &= \exp\left(\int_0^t \alpha_s(W)dW_s - \frac{1}{2}\int_0^t \alpha_s^2(W)ds\right)\\ \frac{d\mu_W}{d\mu_\xi}(t, \xi) &= \exp\left(-\int_0^t \alpha_s(\xi)d\xi_s + \frac{1}{2}\int_0^t \alpha_s^2(\xi)ds\right)\end{aligned}$$

Remark. In fact, condition (2) can be equivalent to $P\left(\int_0^T \alpha_t^2(\xi)dt < \infty\right) = 1$ and

$$P\left(\int_0^t \alpha_s^2(W)ds < \infty\right) = \mathbb{E}\left[\exp\left(-\int_0^t \alpha_s(\xi)dW_s - \frac{1}{2}\int_0^t \alpha_s^2(\xi)ds\right)\right] = 1$$

Therefore, if we want to apply this to our problem, the conditions are as same as the conditions in the Theorem 12.8. Note that here $W = B$ denoted by the Brownian motion.

But this theorem needs the process starting from 0.

Problem: How to prove that in the path-space as many papers do.

13 Markov Process

From this section, random variable X denotes random vector $X\colon \Omega \to \mathbb{R}^n$, and we don't use black bold symbol to distinguish random variable and random vector.

- A stochastic process $X = (X_t)_{t\geqslant 0}$ is memoryless if for any Borel measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathbb{E}\left[f(X_{t+s}) \mid \mathcal{F}_s^X\right] = \mathbb{E}\left[f(X_{t+s}) \mid \sigma(X_s)\right],\ \forall\ s,t \geqslant 0$$

A transition function $P = \{P_t\}_{t\geqslant 0}$ is a family of kernels, that means

$$P_t\colon \mathbb{R}^n \times \mathcal{R}^n \to [0,1]$$

such that

(1) For $t \geqslant 0$ and $x \in \mathbb{R}^n$, $P_t(x,\cdot)$ is a probability measure on $\mathbb{R}^n$;

(2) For $t \geqslant 0$ and $A \in \mathcal{R}^n$, $x \mapsto P_t(x,A)$ is measurable on $\mathbb{R}$;

(3) For $s,t \geqslant 0$, $x \in \mathbb{R}^n$ and $A \in \mathcal{R}^n$,

$$P_{t+s}(x,A) = \int_{\mathbb{R}^n} P_t(y,A)P_s(dy,A)$$

called the Chapman-Kolmogorov relation.

Give a transition function, define a family of operators $(\boldsymbol{P}_t)_{t\geqslant 0}$ on the bounded Borel measurable function space $L^\infty(\mathbb{R}^n)$ (equipped with the supremum norm), defined by

$$(\boldsymbol{P}_t f)(x) := \int_{\mathbb{R}^n} f(y)P_t(x,dy)$$

and it has the following properties.

(1) $\boldsymbol{P}_t 1 = 1$;

(2) For $t \geqslant 0$, $\boldsymbol{P}_t$ maps nonnegative functions to nonnegative functions;

(3) $(\boldsymbol{P}_t)_{t\geqslant 0}$ is an one-parameter semi-group, *i.e.*

$$\boldsymbol{P}_t \circ \boldsymbol{P}_s = \boldsymbol{P}_{t+s},\ \forall, s,t \geqslant 0$$

Then a stochastic process $(X_t)_{t\geqslant 0}$ is called a Markov process if there is a transition function $P = \{P_t\}_{t\geqslant 0}$ on $\mathbb{R}^n$ such that for every $f \in L^\infty(\mathbb{R}^n)$,

$$\mathbb{E}\left[f(X_{t+s}) \mid \mathcal{F}_s^X\right] = (\boldsymbol{P}_t f)(X_s)$$

Proposition 13.1. *Let $(B_t)_{t\geqslant 0}$ be a n-dimensional stochastic process defined on $(\Omega,\mathcal{F},\mathbb{P})$ such that $B_0 = 0$ a.e.. Then $(B_t)_{t\geqslant 0}$ is an Brownian motion if and only if it is a Markov process with*

$$\boldsymbol{P}_0 = Id,\ (\boldsymbol{P}_t f)(x) = \frac{1}{(2\pi t)^{\frac{n}{2}}}\int_{\mathbb{R}^n} f(y)e^{-\frac{\|x-y\|^2}{2t}}dy, \forall\ t > 0$$

Theorem 13.2. *Let $(B_t)_{t\geqslant 0}$ be a n-dimensional Brownian motion and $f\colon \mathbb{R}_{\geqslant 0}\times\mathbb{R}^n \to \mathbb{C}$ be such that* *(i)* $f \in \mathcal{C}^{1,2}(\mathbb{R}_{\geqslant 0} \times \mathbb{R}^n, \mathbb{C})$*;*

(ii) For $t \geqslant 0$*, there is a constant* $K, \alpha > 0$ *such that for every* $x \in \mathbb{R}$

$$\sup_{0 \leqslant s \leqslant t} |f(s,t)| \leqslant K e^{\alpha \|x\|}$$

The process $\{f(t, B_t)\}_{t \geqslant 0}$ *is a martingale if and only if*

$$\frac{\partial f}{\partial t} + \frac{1}{2}\Delta f = 0$$

Theorem 13.3. *Let* $\{P_t\}_{t \geqslant 0}$ *be a transition function on* $\mathbb{R}^n$ *and* ν *be a probability measure on* $\mathbb{R}^n$*. Then there is a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a stochastic process* $(X_t)_{t \geqslant 0}$ *such that*

(1) $X_0 \sim \nu$*;*

(2) if $f \in L^\infty(\mathbb{R}^n)$*, then*

$$\mathbb{E}[f(X_{t+s}) \mid \mathcal{F}_s^X] = (\boldsymbol{P}_t f)(X_s)$$

- Strong Markov Property: Let $(X_t)_{t \geqslant 0}$ be a Markov process with a transition function $\{P_t\}_{t \geqslant 0}$. We say $(X_t)_{t \geqslant 0}$ has a strong Markov property if for any $f \in L^\infty(\mathbb{R}^n)$ and any stopping time S of $\mathcal{F}^X$ such that

$$\mathbb{E}[f(X_{S+t}) \mid \mathcal{F}_S^X] = (\boldsymbol{P}_t f)(X_S)$$

  Let $(B_t)_{t \geqslant 0}$ be a Brownian motion and T be a finite stopping time. The process

$$(B_{T+t} - B_T)_{t \geqslant 0}$$

  is also a Brownian motion independent from $\mathcal{F}_T^B$. So $(B_t)_{t \geqslant 0}$ is a strong Markov process.

- A contraction semi-group of operators $(\boldsymbol{P}_t)_{t \geqslant 0}$ on $L^\infty(\mathbb{R}^n)$ such that $\boldsymbol{P}_t 1 = 1$ and $0 \leqslant \boldsymbol{P}_t f \leqslant 1$ for $0 \leqslant f \leqslant 1$ is said Feller-Dynkin if

  (1) $\boldsymbol{P}_t \colon \mathcal{C}_0(\mathbb{R}^n) \to \mathcal{C}_0(\mathbb{R}^n)$

  (2) $\forall\ f \in \mathcal{C}_0(\mathbb{R}^n)$,

$$\lim_{t \to 0} \|\boldsymbol{P}_t f - f\|_\infty = 0$$

  If the transition function of a Markov process is Feller-Dynkin, then the Markov process is also called Feller-Dynkin.

Theorem 13.4. *Let* $\{P_t\}_{t \geqslant 0}$ *be a Feller-Dynkin transition function. Then for each probability measure* ν *on* $\mathbb{R}^n$*, there is a filtered probability space and an adapted process* $(X_t)_{t \geqslant 0}$ *defined on that such that*

(1) $X_0 \sim \nu$*;*

(2) the path of $(X_t)_{t \geqslant 0}$ *is càdlàg;*

(3) $(X_t)_{t \geqslant 0}$ *is a strong Markov process.*

Proposition 13.5. *Let $(\boldsymbol{P}_t)_{t\geqslant 0}$ be a Feller-Dynkin semi-group. Then there is a densely defined*

$$L\colon \mathcal{D}(L) \subset \mathcal{C}_0(\mathbb{R}^n) \to \mathcal{C}_0(\mathbb{R}^n)$$

where

$$\mathcal{D}(L) = \left\{ f \in \mathcal{C}_0(\mathbb{R}^n)\colon \lim_{t\to 0} \frac{\boldsymbol{P}_t f - f}{t} \text{ exists} \right\}$$

such that $\forall\ f \in \mathcal{D}(L)$,

$$\lim_{t\to 0} \left\| \frac{\boldsymbol{P}_t f - f}{t} - Lf \right\| = 0$$

In such case, we say L generates $(\boldsymbol{P}_t)_{t\geqslant 0}$.

Example 13.6. Let $(B_t)_{t\geqslant 0}$ be n-dimensional Brownian motion. Then it is a Feller-Dynkin and $\mathcal{D}(L) = \mathcal{C}_c(\mathbb{R}^n)$ and $\forall\ f \in \mathcal{C}_c(\mathbb{R}^n)$,

$$Lf = \frac{1}{2}\Delta f$$

where $\mathcal{C}_c$ is the smooth and compact supported.

Let $(\boldsymbol{P}_t)_{t\geqslant 0}$ be a Feller-Dynkin semi-group. It is called a (Feller-Dynkin) diffusion semi-group if $\mathcal{C}_c(\mathbb{R}^n) \subset \mathcal{D}(L)$ and for any probability measure ν on $\mathbb{R}^n$, there is a probability measure $\mathbb{P}^\nu$ on filtered canonical space $(\mathcal{C}([0,\infty)), \sigma\mathcal{C}, (\mathcal{F}_t)_{t\geqslant 0})$, where $\mathcal{F}_t = \sigma(\pi_s\colon 0 \leqslant s \leqslant t)$, such that

(1) the distribution of π_0 is ν,
(2) $(\pi_t)_{t\geqslant 0}$ is a Markov process with semi-group $(\boldsymbol{P}_t)_{t\geqslant 0}$.

Then a Markov process $(X_t)_{t\geqslant 0}$ is called a (Feller-Dynkin) diffusion process if its semi-group is a (Feller-Dynkin) diffusion semi-group.

Example 13.7. Let $(B_t)_{t\geqslant 0}$ be a Brownian motion. For $\mu \in \mathbb{R}$, then the diffusion process $(B_t + \mu t)_{t\geqslant 0}$ has the generator L defined on $\mathcal{C}_c(\mathbb{R})$

$$Lf = \frac{df}{dx} + \frac{1}{2}\frac{d^2 f}{dx^2}$$

Example 13.8. Consider a particular OU process,

$$X_t = e^{\theta t} B_{\frac{1-e^{-2\theta t}}{2\theta}}$$

It's generator defined on $\mathcal{C}_c(\mathbb{R})$ and

$$Lf = \theta x \frac{df}{dx} + \frac{1}{2}\frac{d^2 f}{dx^2}$$

Proposition 13.9. *Let $(X_t)_{t\geqslant 0}$ be a diffusion process with transition function $(P_t)_{t\geqslant 0}$ and generator L. For $f \in \mathcal{C}_c(\mathbb{R}^n)$, the process*

$$\left(f(X_t) - \int_0^t (Lf)(X_s)ds \right)_{t\geqslant 0}$$

is a martingale.

Theorem 13.10 (Dynkin theorem)**.** *Let $(X_t)_{t\geqslant 0}$ be a diffusion process with generator L. Then there is $b_i, \sigma_{ij} \in \mathcal{C}(\mathbb{R}^n)$ such that for all $f \in \mathcal{C}_c(\mathbb{R}^n)$,*

$$Lf = \sum_{i=1}^{n} b_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} \sigma_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}$$

- An adapted stochastic process $(X_t)_{t\geqslant 0}$ is called a Lévy process if it is càdlàg and
  (1) $X_0 = 0$ *a.e.*
  (2) For every $T \geqslant 0$, the process $(X_{T+t} - X_T)_{t\geqslant 0}$ is independent from $\mathcal{F}_T$
  (3) For every $t, T \geqslant 0$, $X_{t+T} - X_T$ has the same distribution as X_t

14 Diffusion Operator

- Considering the diffusion operator,

$$L = \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^{n} \sigma_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}$$

defined on Banach space $L^\infty(\mathbb{R}^n)$ or Hilbert space $L^2(\mathbb{R}^n)$, which is a densely defined unbounded operator.x

Assume there is a Borel measure μ on $\mathbb{R}^n$ that is equivalent to the Lebesgue measure such that L is symmetric, *i.e.* for all $f, g \in \mathcal{C}_c(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} (Lf)g d\mu = \int_{\mathbb{R}^n} f(Lg)d\mu$$

Example 14.1. On $\mathcal{C}_c(\mathbb{R}^n)$, the operator defined as

$$Lf := \operatorname{div}(\sigma\nabla f)$$

is a diffusion operator that is symmetric with respect to the Lebesgue measure.

For $f, g \in \mathcal{C}_c(\mathbb{R}^n)$, define

$$\Gamma(f,g) = \frac{1}{2}(L(fg) - (Lf)g - f(Lg)) = \sum_{i,j=1}^{n} \sigma_{ij}(x)\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}$$

called carré du champ. Note that $\Gamma(f,f) \geqslant 0$. And the Dirichlet form of L is

$$\mathcal{E}(f,g) := \int_{\mathbb{R}^n} L(f,g)d\mu$$

Note that $\mathcal{E}(f,g) = \mathcal{E}(g,f)$ and $\mathcal{E}(f,f) \geqslant 0$ and

$$\mathcal{E}(f,g) = -\int_{\mathbb{R}^n} fLgd\mu = -\int_{\mathbb{R}^n} gLfd\mu$$

L is densely defined on $\mathcal{D}(L) = \mathcal{C}_c(\mathbb{R}^n)$ and a non-positive symmetric operator on the Hilbert space $L^2(\mathbb{R}^n)$ but not self-adjoint. However, on the Hilbert space $L^2(\mathbb{R}^n)$, there is a densely defined non-positive self-adjoint extension $\overline{L}$ of L. If $\overline{L}$ is the unique non-positive self-adjoint extension of L, then L is called essentially self-adjoint. If for some $\lambda > 0$,

$$\ker(-L^* + \lambda Id) = 0$$

then L is essentially self-adjoint, which equivalent to for any $f \in L^2(\mathbb{R}^n)$, $Lf = \lambda f$ implies $f = 0$.

15 Stochastic Differential Equation

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and a m-dimensional Brownian motion $(W_t)_{t\geqslant 0}$. Let $b\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon [0,T]\times\mathbb{R}^{d\times m}$ Borel measurable functions. Let ξ be an $\mathcal{F}_0$-random variable. Then a $\mathbb{R}^d$-valued process $(X_t)_{t\in[0,T]}$ is a solution of a SDE

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

if it satisfies

(i) $(X_t)_{t\in[0,T]}$ is continuous and $\mathbb{F}$-adapted,

(ii) $\int_0^T |b(t, X_t)| + |\sigma(t, X_t)|^2\, ds < \infty$,

(iii) X_t is as

$$X_t = \xi + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

1. Uniqueness and Existence:

   Theorem 15.1. *Suppose b, σ and ξ satisfies*

   (1) σ and b are Lipschitz continuous;

   (2) σ and b are linearly growth;

   (3) $\xi \in L^2$.

   the SDE

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

   with initial value $X_0 = \xi$ has a solution $(X_t)_{t\in[0,T]}$ satisfying $\mathbb{E}[\sup_{0\leqslant t\leqslant T}|X_t|^2] < \infty$. Moreover, it is unique in the sense of indistinguishability.

2. Let $(X_s^{t,\xi})_{t\leqslant s\leqslant T}$ be the unique solution of the above SDE with initial value $X_t = \xi$, *i.e.*

   $$X_s^{t,\xi} = \xi + \int_t^s b(u, X_u^{t,\xi})du + \int_t^s \sigma(u, X_u^{t,\xi})dW_u$$

   Note that it is a Markov process and has the strong Markov property.

3. For the following homogeneous SDE

   $$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

   consider the differential operator

   $$\mathcal{A}_t := \sum_{i=1}^d b_i(x)\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^m \sigma_{ik}(x)\sigma_{jk}(x)\partial^2_{x_i x_j}$$

   For $f(t,x)$, let $(\mathcal{A}_t f)(t,x) = (\mathcal{A}_t f(t,\cdot))(x)$.
   For $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ in $\mathcal{C}^{1,2}$,

   $$u(T, X_T) = u(0, X_0) + \int_0^T (\partial_t u + \mathcal{A}_t u)(t, X_t)dt + \sum_{i=1}^d \sum_{k=1}^m \int_0^T \partial_{x_i} u(t, X_t)\sigma_{ik}(t, X_t)dW_t^k$$

   So if u satisfies

   $$\begin{aligned} \partial_t u + \mathcal{A}_t u &= 0 \\ u(T,\cdot) &= g \end{aligned} \tag{$*_1$}$$

   then $(u(t, X_t))_{t\in[0,T]}$ is a martingale and $\mathbb{E}[g(X_T)] = u(0, X_0)$.

4. Let $X_t^x = X_t^{0,x}$.

$$u(t,x) := (\boldsymbol{P}_t f)(x) = \mathbb{E}[f(X_t^x)] = \mathbb{E}[f(X_t) \mid X_0 = x]$$

In fact, denote $L = \mathcal{A}_t$,

$$\boldsymbol{P}_t = e^{tL} \text{ by } Lf = \lim_{t\to 0} \frac{\boldsymbol{P}_t f - f}{t}$$

So

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial t}\boldsymbol{P}_t f = Le^{tL}f = Lu$$

with initial value $u(0,x) = f(x)$.

5. Dynkin's formula: Let $\mathcal{A}$ be the generator.

Theorem 15.2. *Suppose τ is a stopping time with $\mathbb{E}[\tau] < \infty$.*

$$\mathbb{E}[f(X_\tau^x)] = f(x) + \mathbb{E}\left[\int_0^\tau \mathcal{A}f(X_s^x)\right] ds$$

6. Kolmogorov's backward equation:

Theorem 15.3. *Let $f \in \mathcal{C}_0^2(\mathbb{R}^d)$. Define $u(t,x) = \mathbb{E}[f(X_t^x)]$. Then $u(t,\cdot) \in \mathcal{D}(\mathcal{A})$ and*

$$\begin{aligned} \partial_t u &= \mathcal{A}u, \\ u(0,x) &= f(x) \end{aligned}$$

Corollary 15.4 (Kolmogorov's forward equation)**.** *For the operator $\mathcal{A}$,*

$$\mathcal{A}f(y) = \sum_i b_i(y)\partial_i f(y) + \frac{1}{2}\sum_{i,j} a_{ij}(y)\partial^2_{x_i x_j} f(y)$$

where $a_{ij} = \sum_k \sigma_{ik}\sigma_{jk}$. The corresponding adjoint operator is

$$\mathcal{A}^*\phi(y) = -\sum_i \partial_i(b_i\phi) + \frac{1}{2}\sum_{i,j}\partial^2_{x_i x_j}(a_{ij}\phi)$$

If the transition density of X_t^x is $p(t,x)$, then

$$\frac{\partial}{\partial t}p(t,x) = \mathcal{A}^*p = -\sum_i \partial_i(b_i(x)p(t,x)) + \frac{1}{2}\sum_{i,j}\partial^2_{x_i x_j}(a_{ij}(x)p(t,x))$$

7. The Feynman-Kac formula:

Theorem 15.5. *Let $f \in \mathcal{C}_0^2(\mathbb{R}^d)$ and $q \in \mathcal{C}(\mathbb{R}^d)$ with lower bounding and*

$$v(t,x) = \mathbb{E}\left[\exp\left(-\int_0^t q(X_s^x)ds\right) f(X_t^x)\right]$$

Then

$$\begin{aligned} \partial_t v &= \mathcal{A}v - qv, \\ v(0,x) &= f(x) \end{aligned}$$

16 Other Topics Related Stochastic Process

- A process $(X_t)_{t\geqslant 0}$ is called (strictly) stationary if its finite-dimensional distribution is invariant *w.s.t.* time-moving, *i.e.*

$$\mu_{t_1+s,t_2+s,\cdots,t_n+s} = \mu_{t_1,t_2,\cdots,t_n}$$

  A process $(X_t)_{t\geqslant 0} \subset L^2$ is called secondary (or weakly) stationary if

$$\mathbb{E}[X_t] \equiv \mu, \ \mathbb{E}[(X_t-\mu)(X_s-\mu)] = C(t-s)$$

  for a constant $\mu \in \mathbb{R}$ and a function $C(t)$.

- Adjoint Semi-group: For a Markov process $(X_t)_{t\geqslant 0}$ with the transitive semi-group $(\boldsymbol{P}_t)_{t\geqslant 0}$, the corresponding adjoint semi-group $(\boldsymbol{P}_t^*)_{t\geqslant 0}$ acting on the space of probability measures is defined as

$$\boldsymbol{P}_t^*\mu(A) := \int_{\mathbb{R}^d} \boldsymbol{P}_t(x,A)d\mu(x) = \int_{\mathbb{R}^d} \mathbb{P}(X_t \in A \mid X_0 = x)d\mu(x)$$

  It is because

$$\int_{\mathbb{R}^d} \boldsymbol{P}_t f(x)d\mu(x) = \int_{\mathbb{R}^d} f(x)d(\boldsymbol{P}_t^*\mu(x))$$

  Also denote

$$\boldsymbol{P}_t^* = e^{tL^*}$$

  where L^* is the adjoint operator of L in L^2-sense, because

$$\int_{\mathbb{R}^d} (Lf)g dx = \int_{\mathbb{R}^d} f(L^*g)dx$$

  If $X_0 \sim \mu$, let $\mu_t := \boldsymbol{P}_t^*\mu$. By the definition, $X_t^x \sim \mu_t$. Moreover,

$$\frac{\partial \mu_t}{\partial t} = L^*\mu_t, \ \mu_0 = \mu$$

  It is the Kolmogorov's forward equation or called the Fokker-Plank equation.

- Consider the following SDE,

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t$$

  where $\beta > 0$, and its generator is

$$\mathcal{L}_V = -\nabla V(X_t)\cdot\nabla + \beta^{-1}\Delta$$

  and adjoint operator is

$$\mathcal{L}_V^* = \nabla\cdot(\nabla V(X_t)\ \cdot\ ) + \beta^{-1}\Delta$$

  So the corresponding Fokker-Plank equation is

$$\frac{\partial}{\partial t}\rho = \nabla\cdot(\nabla V(X_t)\rho) + \beta^{-1}\Delta\rho$$
$$\rho(0,x) = \rho_0(x)$$

  called Smoluchowski equation. Usually, assume potential V is confining, *i.e.* $\lim_{t\to\pm\infty} V(x) = 0$ and $\exp(-\beta V(x)) \in L^1$ for all $\beta > 0$.

  For a Markov process $(X_t)_{t\geqslant 0}$ with generator L, it is called ergodic if

$$\boldsymbol{P}_t g = g, \ \forall t \geqslant \quad \Leftrightarrow \quad g \equiv const.$$

  Then by the definition, there is an invariant measure μ, *i.e.*

$$\boldsymbol{P}_t^*\mu = \mu$$

Proposition 16.1. *Let $V(x)$ be a confining potential. Then the Markov process with the generator $\mathcal{L}_V$ is ergodic. And the unique invariant measure is Gibbs measure with the density function as*

$$\rho_V(x) = \frac{1}{Z} e^{-\beta V(x)}$$

Moreover, the generator $\mathcal{L}_V$ is self-adjoint on $L(\rho_V)$. And it is non-positive and with kernel consisting with constants.

Proposition 16.2. *If $\rho(t,x)$ is a solution of the Fokker-Plank equation defined as above with the initial value $\rho_0(x) \in L^2(\mathbb{R}^d; \rho_V^{-1})$ and V satisfies the Poincaré inequality with constant $\lambda > 0$* i.e. $\operatorname{Hess} V \geqslant \lambda I$, *then $\rho(t,x)$ converges $\rho_V(x)$ exponentially,* i.e.

$$\|\rho(t,\cdot) - \rho_V\|_{L^2(\rho_V^{-1})} \leqslant e^{-\lambda\beta^{-1}t} \|\rho_0 - \rho_V\|_{L^2(\rho_V^{-1})}$$

17 Diffusion Operator on Euclidean Space

- Markov triples on Euclidean space: A Markov operator $\boldsymbol{P}$ on $(E, \mathcal{F})$ is a linear operator on $B(E)$ satisfying

  (i) $\boldsymbol{P}1 = 1$,

  (ii) $\boldsymbol{P}f \geqslant f$ for $f \geqslant 0$.

  Remark. Note that for any Markov operator $\boldsymbol{P}$, there is a kernel $P\colon E \times \mathcal{F} \to [0,1]$ representing $\boldsymbol{P}$ as

  $$(\boldsymbol{P}f)(x) = \int f(y) P(x, dy)$$

  For a measure μ on $(E, \mathcal{F})$, μ is called

  (i) invariant for $\boldsymbol{P}$ if

  $$\int \boldsymbol{P}f d\mu = \int f d\mu, \ \forall\ f \in L^1(\mu) \cap L^\infty$$

  (ii) reversible for μ if

  $$\int f\boldsymbol{P}h d\mu = \int h\boldsymbol{P}f d\mu, \ \forall\ f, h \in L^2(\mu)$$

  Remark. Note that if μ is invariant, then for all $f \in L^1(\mu) \cap L^\infty$, $\|\boldsymbol{P}f\|_p \leqslant \|f\|_p$ for any $1 \leqslant p < \infty$. Since $L^1(\mu) \cap L^\infty$ is dense in $L^p(\mu)$, then $\boldsymbol{P}$ can be defined on $L^p(\mu)$. And so the definition of reversibility is well-defined because reversibility implies invariance.

  A symmetric Markov semi-group on $(E, \mathcal{F}, \mu)$ is a family of $\{\boldsymbol{P}_t\}_{t \geqslant 0}$ of operators satisfies

  (i) for all $t \geqslant 0$, $\boldsymbol{P}_t$ is a Markov operator,

  (ii) $\boldsymbol{P}_0 f = f$ for $f \in B(E)$,

  (iii) $\boldsymbol{P}_t \circ \boldsymbol{P}_s = \boldsymbol{P}_{t+s}$,

  (iv) for all $t \geqslant 0$, μ is reversible for $\boldsymbol{P}_t$,

  (v) for all $f \in L^2(\mu)$, $\boldsymbol{P}_t f \to f$ as $t \to 0$.

  For a symmetric Markov semi-group, its generator $(1 \leqslant p < \infty)$

  $$L\colon \mathcal{D}_p(L) \subset L^p(\mu) \to L^p(\mu)$$

is defined as

$$Lf := \lim_{t\to\infty} \frac{\boldsymbol{P}_t f - f}{t}$$

Note that the limits is independent with p so there is no ambiguity to denote L and in particular, $\mathcal{D}(L) := \mathcal{D}_2(L)$. For all $f, h \in \mathcal{D}(L)$,

$$\int fLh d\mu = \int hLf d\mu$$

i.e. L is symmetric. Moreover, $\int Lf d\mu = 0$ for $f \in \mathcal{D}_1(L)$. And for the generator L of a symmetric Markov semi-group, $\mathcal{D}(L) \subset L^2(\mu)$ is dense and L is closed and self-adjoint and non-positive.

Conversely, if L is a non-positive and self-adjoint operator on $L^2(\mu)$, then by the continuous functional calculus, for any $t \geqslant 0$, we define

$$\boldsymbol{P}_t := e^{tL}$$

Then such $\{\boldsymbol{P}_t\}_{t\geqslant 0}$ satisfies

(i) μ is reversible for $\boldsymbol{P}_t$,
(ii) $\boldsymbol{P}_t \boldsymbol{P}_s = \boldsymbol{P}_{t+s}$,
(iii) $\|\boldsymbol{P}_t f\|_2 \leqslant \|f\|_2$ for all $t \geqslant 0$ and $f \in L^2(\mu)$,
(iv) $\boldsymbol{P}_t f \to f$ as $t \to 0$ for $f \in L^2(\mu)$,
(v) for any $f \in L^2(\mu)$ and $k \in \mathbb{N}$, $\boldsymbol{P}_t f \in \mathcal{D}(L^k)$,
(vi) for any $f \in \mathcal{D}(L)$, $Lf = \lim_{t\to 0}(\boldsymbol{P}_t f - f)/t$.

Considering an algebra $\mathcal{A} \subset \mathcal{D}(L)$. Γ defined on $\mathcal{A} \times \mathcal{A}$ by

$$\Gamma(f, h) := \frac{1}{2}(L(fh) - fLh - hLf)$$

and $\Gamma(f) = \Gamma(f, f) \geqslant 0$. Moreover, there is the Cauchy-Schwartz inequality

$$|\Gamma(f, g)| \leqslant \sqrt{\Gamma(f)\Gamma(g)}$$

Considering an algebra $\mathcal{A}$ of function space on E such that for all $k \in \mathbb{N}$,

$$\forall\ f_1, \cdots, f_k \in \mathcal{A},\ \forall\ \Psi \in \mathcal{C}^\infty(\mathbb{R}^k),\ \Psi(f_1, \cdots, f_k) \in \mathcal{A}$$

A bilinear form $\Gamma\colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is called a diffusion carré du champ if for any $\Psi \in \mathcal{C}^\infty(\mathbb{R}^k)$

$$\Gamma(\Psi(f_1, \cdots, f_k), g) = \sum_{i=1}^k \partial_i \Psi(f_1, \cdots, f_k)\Gamma(f_i, g)$$

An operator L is called a corresponding diffusion operator if it satisfies

$$L\Psi(f_1, \cdots, f_k) = \sum_{i=1}^k \partial_i \Psi(f_1, \cdots, f_k) Lf_i + \sum_{i,j=1}^k \partial_i \partial_j \Psi(f_1, \cdots, f_k)\Gamma(f_i, f_j)$$

A symmetric Markov semi-group is called a diffusion semi-group if its generator is a diffusion operator.

A bilinear form $\mathcal{E}\colon \mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E}) \to \mathbb{R}$ is a Dirichlet form if

(i) $\mathcal{D}(\mathcal{E}) \subset L^2(\mu)$ dense,

(ii) $\mathcal{E}(f, g) = \mathcal{E}(g, f)$,

(iii) $\mathcal{E}(f) := \mathcal{E}(f, f) \geqslant 0$,

(iv) $\mathcal{D}(\mathcal{E})$ is complete *w.s.t.* the inner product

$$\langle f, g \rangle_{\mathcal{E}} := \int fg d\mu + \mathcal{E}(f, g)$$

(v) for every $f \in \mathcal{D}(\mathcal{E}), (0 \vee f \wedge 1) \in \mathcal{E}$ and $\mathcal{E}(0 \vee f \wedge 1) \leq \mathcal{E}(f)$.

Suppose Γ is a diffusion carré du champ defined on $\mathcal{A}$ that is dense in $L^2(\mu)$. Also, assume Γ is non-negative and symmetric. Dfine

$$\mathcal{E}(f, g) := \int \Gamma(f, g) d\mu$$

Assume it is closable and taking $\mathcal{D}(\mathcal{E})$ be the completion of $\mathcal{A}$. Then it is a Dirichlet form. Moreover, it induces a definition of a non-positive, self-adjoint operator L by

$$\int gLf d\mu = -\mathcal{E}(f, g)$$

on the domain

$$\mathcal{D}(L) = \{f \in \mathcal{D}(\mathcal{E}) \colon \exists\ C > 0\ s.t.\ |\mathcal{E}(f, g)| \leqslant C \|g\|_2\ \forall\ g \in \mathcal{D}(\mathcal{E})\}$$

Let $(E, \mathcal{F}, \mu)$ be a probability space. For $\mathcal{A} \subset L^2(\mu)$, let $\Gamma \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ be a symmetric and bilinear form. (E, μ, Γ) is called compact Markov diffusion triple if

(i) $\mathcal{A} \subset L^2(\mu)$ is dense,

(ii) $\mathcal{A}$ is closed under the composition of any $\Psi \in \mathcal{C}^\infty(\mathbb{R}^k)$ for all k,

(iii) $\Gamma(f) \geqslant 0$,

(iv) Γ is a diffusion carré du champ,

(v) $\Gamma(f) = 0$ implies f is constant, $\mathcal{E}(f, g) = \int \Gamma(f, g) d\mu$, that can be extended to a Dirichlet form, then

(vi) For all $f \in \mathcal{A}$, $\exists\ C > 0\ s.t.\ |\mathcal{E}(f, g)| \leqslant C \|g\|_2$ for all $g \in \mathcal{A}$. Define L by $\int gLf d\mu = -\mathcal{E}(f, g)$, then above implies $\mathcal{A} \subset \mathcal{D}(L)$, and let $\boldsymbol{P}_t = e^{tL}$,

(vii) $L\mathcal{A} \subset \mathcal{A}$,

(viii) $\boldsymbol{P}_t \mathcal{A} \subset \mathcal{A}$

If (E, μ, Γ) is a Markov diffusion triple, then $\boldsymbol{P}_t$ is a Markov semi-group. Moreover,

$$\boldsymbol{P}_t f \xrightarrow{t \to \infty} \int f d\mu,\ f \in L^2(\mu)$$

For (E, μ, Γ), define

$$\Gamma_2(f, g) = \frac{1}{2} \left(L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf) \right),\ \forall\ f, g \in \mathcal{A}$$

Then (E, μ, Γ) is said to satisfy the curvature-dimension condition $\mathrm{CD}(\rho, n)$ if

$$\Gamma_2(f) \geqslant \rho\Gamma(f) + \frac{1}{n}(Lf)^2,\ \forall\ f \in \mathcal{A}$$

Example 17.1 (Heat Semi-group)**.** For $\mathbb{R}^n$ with Lebesgue measure μ, define

$$p_t(x,y) := \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\|x-y\|^2}{4t}}$$

and the semi-group defined on

$$(\boldsymbol{P}_t f)(x) := \int_{\mathbb{R}^n} f(y) p_t(x,y) dy$$

Then by the Taylor's formula,

$$\frac{\boldsymbol{P}_t f - f}{t} = \Delta f$$

So $L = \Delta = \sum_i \frac{\partial^2}{\partial x_i^2}$ and $\mathcal{D}(L) = \mathcal{C}_c^2(\mathbb{R}^n) = \mathcal{A}$.

$$\Gamma(f,g) = \langle \nabla f, \nabla g \rangle$$

and

$$\Gamma_2(f,g) = \left\langle \nabla^2 f, \nabla^2 g \right\rangle_{\mathrm{F}}$$

A probability measure μ and a Dirichlet form $\mathcal{E}$ satisfy a Poincaré inequality with constant C if

$$\int f^2 d\mu - \left(\int f d\mu \right)^2 \leqslant C\mathcal{E}(f), \ \forall\ f \in \mathcal{D}(\mathcal{E})$$

For a compact Markov diffusion triple (E, μ, Γ), if it satisfies $\mathrm{CD}(\rho, n)$ $(\rho > 0, n > 1)$, then it satisfies the Poincaré inequality with constant $C = \frac{n-1}{\rho n}$. Then

$$\operatorname{Var}_\mu(\boldsymbol{P}_t f) \leqslant e^{-\frac{2t}{C}} \operatorname{Var}_\mu(f), \ \forall\ f \in L^2(\mu)$$

where

$$\operatorname{Var}_\mu f := \int f^2 d\mu - \left(\int f d\mu \right)^2$$

18 Markov Triples on Manifolds

- On Riemannian manifold $(\mathcal{M}, g)$, considering integration by Riemannian volume dV. For $f, h \in \mathcal{C}^\infty(\mathcal{M})$ with compact support,

$$\int_{\mathcal{M}} f \Delta h dV = \int_{\mathcal{M}} h \Delta f dV = -\int_{\mathcal{M}} \langle \nabla f, \nabla h \rangle dV$$

It is because of the divergence theorem,

$$\int_{\mathcal{M}} \operatorname{div}(X) dV = \int_{\partial \mathcal{M}} \langle X, \nu \rangle dS$$

and the relation

$$\operatorname{div}(fX) = \langle \nabla f, X \rangle + f \operatorname{div} X$$

- Let $(\mathcal{M}, g)$ be a compact Riemannian manifold. Consider a easy case, let $L = \Delta$ and it get the carré du champ

$$\Gamma(f,h) = \frac{1}{2}\left(L(fh) - fLh - hLf \right) = g(\nabla f, \nabla h)$$

Then $(\mathcal{M}, dV, \Gamma)$ is a Markov triple.

- Consider the weighted case. Taking $W \in \mathcal{C}^\infty(\mathcal{M})$ with $\int_{\mathcal{M}} e^{-W} dV = 1$, let

$$\mu(A) := \int_{\mathcal{M}} e^{-W} \mathbb{1}_A d\mu, \ \forall\ A \in \mathcal{B}(\mathcal{M})$$

be a probability measure.

Let $\mathcal{A} = \mathcal{C}^\infty(\mathcal{M})$. Define $\Gamma\colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ as

$$\Gamma(f,h) := \langle \nabla f, \nabla h \rangle = g^{ij} \partial_i f \partial_j h$$

Then

$$\begin{aligned}\int_{\mathcal{M}} f \Delta h d\mu &= \int_{\mathcal{M}} (e^{-W} f) \Delta h d\mu \\ &= -\int_{\mathcal{M}} \Gamma(e^{-W} f, h) dV \\ &= -\int_{\mathcal{M}} \Gamma(f,h) - f\Gamma(W,h) d\mu\end{aligned}$$

So if let $Lf = \Delta f - \langle \nabla W, \nabla f \rangle$ for $f \in \mathcal{C}^\infty(\mathcal{M})$, then

$$\int_{\mathcal{M}} f L h d\mu = -\int_{\mathcal{M}} \Gamma(f,h) d\mu$$

Considering the solution $u\colon \mathcal{M} \times \mathbb{R} \to \mathbb{R}$ of

$$\begin{aligned}\frac{\partial u(x,t)}{\partial t} &= \Delta u - \langle \nabla W, \nabla u \rangle \\ u(x,0) &= f(x)\end{aligned}$$

Then $\boldsymbol{P}_t f = u(\cdot, t) \in \mathcal{C}^\infty(\mathcal{M})$. So $(\mathcal{M}, \mu, \Gamma)$ is a compact Markov triple.

19 Langevin Diffusion

Consider the Langevin diffusion with potential V,

$$dZ_t = -\nabla V(Z_t) dt + \sqrt{2} dB_t$$

19.1 Markov Semi-group Theory

- Time-homogeneous Markov process: A Markov process $(X_t)_{t \geqslant 0}$ is called time-homogeneous if

$$\mathbb{P}(X_t = y \mid X_s = x) = \mathbb{P}(X_{t-s} = y \mid X_0 = x)$$

For time-homogeneous Markov process, it can define the transition probability $P(t,x,A) = \mathbb{P}(X_t \in A \mid X_0 \in x)$. Moreover, the Markov semi-group $(\boldsymbol{P}_t)_{t \geqslant 0}$ defined by

$$\boldsymbol{P}_t f := \mathbb{E}[f(X_t) \mid X_0 = x]$$

and satisfies $\boldsymbol{P}_t \boldsymbol{P}_s = \boldsymbol{P}_s \boldsymbol{P}_t = \boldsymbol{P}_{t+s}$. Also, the infinitesimal generator L is

$$Lf := \lim_{t \downarrow 0} \frac{\boldsymbol{P}_t f - f}{t}$$

acting on function subspace $\mathcal{D}(L)$. Because the dual space of function space is the space of measures, the adjoint operator L^* can act on densities π_0 of X_0, so are P_t^* (it is the push-forward of π_0 to t-step).

I. Kolmogorov's backward equation: $\partial_t \boldsymbol{P}_t f = L\boldsymbol{P}_t f = \boldsymbol{P}_t Lf$.

II. Kolmogorov's forward equation: $\partial_t \boldsymbol{P}_t^* \pi_0 = L^* \boldsymbol{P}_t^* \pi_0 = \boldsymbol{P}_t^* L^* \pi_0$

- The Markov semi-group $(\boldsymbol{P}_t)_{t\geqslant 0}$ is reversible *w.s.t.* π (stationary distribution) if

$$\int gLfd\pi = \int fLgd\pi, \ \forall\ f,g \in L^2(\pi)$$

  or equivalently, $\int g\boldsymbol{P}_t fd\pi = \int f\boldsymbol{P}_t gd\pi$.

Example 19.1. For the Langevin diffusion, the infinitesimal generator L is

$$Lf = \Delta f - \langle \nabla V, \nabla f \rangle$$

and the dual

$$L^* g = \Delta g + \operatorname{div}(g\Delta V)$$

And

$$\Gamma(f,g) = \langle \nabla f, \nabla g \rangle$$

Remark. It is an Euclidean case for weighted case of Markov triple on Riemannian manifold.

19.2 Optimal Transport

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of all probability measures on $\mathbb{R}^d$ and

$$\mathcal{P}_2(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) \colon \int \|x - x_0\|^2 \, d\mu(x) < \infty \right\}$$

Then $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is the Wasserstein 2-space, where

$$W_2(\mu,\nu) = \inf_{\gamma \in \mathcal{C}(\mu,\nu)} \int \|x-y\|^2 \, d\gamma(x,y)$$

and $\mathcal{C}(\mu,\nu)$ is the set of $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginal distribution μ and ν.

- Let $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ be the subset of $\mathcal{P}_2(\mathbb{R}^d)$ such that the elements are absolutely continuous *w.s.t.* Lebesgue measure. Considering a curve $t \mapsto \mu_t$ in $\mathcal{P}_{2,ac}(\mathbb{R}^d)$,

$$|\dot{\mu}|(t) := \lim_{s\to t} \frac{W_2(\mu_s,\mu_t)}{|s-t|}$$

  called the metric derivative.

  For the curve $t \mapsto \mu_t$ in $\mathcal{P}_{2,ac}(\mathbb{R}^d)$, considering the continuity equation,

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0$$

  the any vector field $\tilde{v}_t$ satisfies above equation with the property $|\dot{\mu}|(t) \leqslant \|\tilde{v}_t\|_{L^2(\mu_t)}$. Moreover, there is an unique vector field v_t satisfying the continuity equation and with the property $\|v_t\|_{L^2(\mu_t)} = |\dot{\mu}|(t)$. Such v_t has the form

$$v_t = \nabla \psi_t, \ \psi_t \colon \mathbb{R}^d \to \mathbb{R}$$

- Riemannian structure: For $\mu \in \mathcal{P}_{2,ac}$, define

$$T_\mu \mathcal{P}_{2,ac} := \overline{\{\nabla \psi \colon \psi \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}$$

And equipping $T_\mu \mathcal{P}_{2,ac}$ with $L^2(\mu)$-norm, which induces an inner product, *i.e.* Riemannian metric, but note that it is not a canonical Riemannian manifold. Besides,

$$T_\mu \mathcal{P}_{2,ac} = \overline{\{\lambda(T - id) \colon \lambda > 0,\ T \text{ is a optimal transport map}\}}^{L^2(\mu)}$$

For a curve $t \to \mu_t$ in $\mathcal{P}_{2,ac}$, by above, there is a unique such v_t satisfying the continuity equation and $\|v_t\|_{L^2(\mu_t)} = |\dot{\mu}|(t)$ and $v_t = \nabla \psi_t \in T_{\mu_t}\mathcal{P}_{2,ac}$. So we say v_t is the corresponding tangent vector of $t \to \mu_t$.

Theorem 19.2 (Wasserstein geodesics). *For $\mu_0, \mu_1 \in \mathcal{P}_{2,ac}$,*

$$W_2(\mu_0, \mu_1) = \inf\left\{\int_0^1 \|v_t\|_{L^2(\mu_t)} \,\mathrm{d}t \mid \partial_t \mu_t + \operatorname{div}(\mu_t v_t) = 0\right\}$$

The infimum is achieved as follows. Let $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$ be optimally coupled random variables, let $X_t := (1-t)X_0 + tX_1$, and let $\mu_t := \operatorname{law}(X_t)$. Then, $t \mapsto \mu_t$ is the unique constant-speed geodesic joining μ_0 to μ_1.

19.3 Langevin SDE as Gradient Flow

For a functional $F \colon \mathcal{P}_{2,ac} \to \mathbb{R} \cup \{\infty\}$ and $\mu \in \mathcal{P}_{2,ac}$, define $\nabla_{W_2} F(\mu) \in T_\mu \mathcal{P}_{2,ac}$ as for any $t \mapsto \mu_t$ with $\mu_0 = \mu$ and v_0 tangent vector at $t = 0$,

$$\left.\frac{d}{dt}\right|_{t=0} F(\mu_t) = \langle \nabla_{W_2} F(\mu), v_0 \rangle_\mu$$

Since $\mu_t \in \mathcal{P}_{2,ac}$, $\mu_t = \tilde{\mu}_t dm$, where m is the Lebesgue measure. Assume $\tilde{\mu}_t$ is differentiable *w.s.t.* t. The first variation of F is $\delta F(\mu) \colon \mathbb{R}^d \to \mathbb{R}$ *s.t.*

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} F(\mu_t) &= \int \delta F(\mu)\, \partial_t|_{t=0} \tilde{\mu}_t dm \\
&= -\int \delta F(\mu) \operatorname{div}(v_0 \tilde{\mu}) dm \\
&= \int \langle \nabla \delta F(\mu), v_0 \tilde{\mu} \rangle\, dm \\
&= \int \langle \nabla \delta F(\mu), v_0 \rangle\, d\mu
\end{aligned}$$

Therefore,

$$\nabla_{W_2} F(\mu) = \nabla \delta F(\mu)$$

For $F \colon \mathcal{P}_{2,ac} \to \mathbb{R} \cup \{\infty\}$, find a curve $t \mapsto \mu_t$ *s.t.* the corresponding vector field v_t of the curve satisfies

$$v_t = -\nabla_{W_2} F(\mu_t) = -\nabla \delta F(\mu_t)$$

So the Wasserstein gradient flow by the continuity equation is

$$\partial_t \mu_t = \operatorname{div}(\mu_t \nabla_{W_2} F(\mu_t)) = \operatorname{div}(\mu_t \nabla \delta F(\mu_t))$$

Example 19.3. Considering the Langevin dynamics.

(1) From the forward equation,

$$\partial_t \mu_t = L^* \mu_t = \Delta \mu_t + \operatorname{div}(\mu_t \nabla V) = \operatorname{div}(\mu_t \nabla(\log \mu_t + V))$$

(2) Let $\pi = \frac{1}{Z}\exp(-V(x))$, where $Z = \int \exp(-\beta V(x))dx$. It is known as the unique stationary solution of the Fokker-Plank equation

$$\partial_t \mu_t = L^* \mu_t = \Delta \mu_t + \operatorname{div}(\mu_t \nabla V)$$

which is induced by the Langevin dynamics

$$dZ_t = -\nabla V(Z_t)dt + \sqrt{2}dB_t$$

From the Wasserstein gradient flow, first let $F = \mathrm{KL}(\cdot\|\pi)$, then

$$F(\mu) = \int \mu \log \frac{\mu}{\pi} = \int V d\mu + \int \mu \log \mu$$

It is called the free energy functional, which serves as the Lyapunov function of the Fokker-Planck equation. Moreover,

$$\delta F(\mu) = V + \log \mu + const.$$

Therefore, $\nabla_{W_2} F(\mu) = \nabla(V + \log \mu)$, and the Wasserstein gradient flow

$$\partial_t \mu_t = \operatorname{div}(\mu_t \nabla(\log \mu_t + V))$$

The Langevin diffusion is the gradient flow of KL divergence.

20 Langevin on Manifold

First, fix a Riemannian manifold $(\mathcal{M}, g = (g_{ij}))$ with volume measure dx.

- If $(X_t)_{t\geqslant 0}$ on $\mathbb{R}^n$ is a standard Brownian motion, then its density evolve as

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{1}{2}\Delta\rho(x,t)$$

So to define the standard Brownian motion on $\mathcal{M}$, considering a $\mathcal{M}$-valued random process $(X_t)_{t\geqslant 0}$ whose densities satisfies

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{1}{2}\Delta_{\mathcal{M}}\rho(x,t),$$

where $\Delta_{\mathcal{M}}$ is the Laplace(-Beltrami) operator defined on the Riemannian manifold. In a local chart,

$$\Delta_{\mathcal{M}} = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^i}\left(g^{ij}\sqrt{|g|}\frac{\partial}{\partial x^j}\right) = g^{ij}\frac{\partial^2}{\partial x^i \partial x^j} + b^i \frac{\partial}{\partial x_i},$$

where

$$b^i = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x_j}\left(\sqrt{|g|}g^{ij}\right) = g^{jk}\Gamma^i_{jk}$$

Let $b = (b^i)$ and $\sigma = (\sigma_{ij})$ be the unique positive definite square root of g^{-1}. Then in the local chart,

$$dX_t = \frac{1}{2}b(X_t)dt + \sigma(X_t)dB_t$$

- Riemannian Langevin algorithm: First, consider the PDE on manifold $\mathcal{M}$,

$$\frac{\partial \rho_t}{\partial t}=\operatorname{div}(\rho_t \operatorname{grad} f+\operatorname{grad} \rho_t)$$

In a local chart, it can be expressed as

$$\begin{aligned}\frac{\partial \rho_t}{\partial t} =&\frac{1}{\sqrt{|g|}} \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\sqrt{|g|}\left(\sum_j g^{ij} \frac{\partial f}{\partial x_j}\right) \rho_t+\sqrt{|g|} \sum_j g^{ij} \frac{\partial \rho_t}{\partial x_j}\right) \\ =&\frac{1}{\sqrt{|g|}} \sum_i \frac{\partial}{\partial x_i}\left(\left(\sum_j g^{ij} \frac{\partial f}{\partial x_j}-\frac{1}{\sqrt{|g|}} \sum_j \frac{\partial}{\partial x_j}\left(\sqrt{|g|} g^{ij}\right)\right) \sqrt{|g|} \rho_t\right) \\ &+\frac{1}{\sqrt{|g|}} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}\left(g^{ij} \sqrt{|g|} \rho_t\right)\end{aligned}$$

Replacing ρ_t by $\tilde{\rho}_t=\sqrt{|g|}\rho_t$,

$$\frac{\partial \tilde{\rho}_t}{\partial t}=-\sum_i \frac{\partial}{\partial x_i}\left(\left(\frac{1}{\sqrt{|g|}} \sum_j \frac{\partial}{\partial x_j}\left(\sqrt{|g|} g^{ij}\right)-\sum_j g^{ij} \frac{\partial f}{\partial x_j}\right) \tilde{\rho}_t\right)+\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}\left(g^{ij} \tilde{\rho}_t\right)$$

Setting

$$F_i\left(x_t\right)=-\sum_j g^{ij} \frac{\partial f}{\partial x_j}+\frac{1}{\sqrt{|g|}} \sum_j \frac{\partial}{\partial x_j}\left(\sqrt{|g|} g^{ij}\right)$$

and $\sigma=(\sigma_{ij})$ be the unique positive definite square root of g^{-1} and $A=\frac{1}{2}\sigma\sigma^\top$. Then above equation is

$$\frac{\partial p(x, t)}{\partial t}=-\sum_{i=1}^n \frac{\partial}{\partial x_i}\left(F_i(x, t) p(x, t)\right)+\sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\left(A_{ij}(x, t) p(x, t)\right)$$

where p_t is the density of X_t in local chart (Euclidean). So So the local Langevin equation is

$$dX_t=F\left(X_t\right) dt+\sqrt{2 g^{-1}} dB_t$$

Therefore, by sampling on the tangent space $\epsilon F\left(x_0\right)+\sqrt{2 \epsilon g^{-1}\left(x_0\right)} \xi_0$,

$$x_1=\operatorname{Exp}_{x_0}\left(\epsilon F\left(x_0\right)+\sqrt{2 \epsilon g^{-1}\left(x_0\right)} \xi_0\right)$$

where $\xi_0 \sim \mathcal{N}(0, I)$. So

$$x_{k+1}=\operatorname{Exp}_{x_k}\left(\epsilon F\left(x_k\right)+\sqrt{2 \epsilon g^{-1}\left(x_k\right)} \xi_0\right)$$

- Log-Sobolev inequality: For a measure ν on $\mathcal{M}$, it is called to satisfy the logarithm Sobolev inequality if there is an $\alpha>0$ such that

$$\int_M g^2 \log g^2 d\nu-\left(\int_M g^2 d\nu\right) \log \left(\int_M g^2 d\nu\right) \leq \frac{2}{\alpha} \int_M\|\operatorname{grad} g\|^2 d\nu$$

for any $g \in \mathcal{C}^{\infty}(\mathcal{M})$ with $\int_{\mathcal{M}} g^2 d\nu<\infty$.

Remark. The left-hand side is denoted by $\operatorname{Ent}_\nu(g^2)$. Note that if $\mu \ll \nu$ with $d\mu=f d\nu$, then

$$H(\mu|\nu)=\operatorname{End}_\nu(f)$$

where H is the KL divergence. Moreover, if ρ_t is the densities of Langevin dynamics, then

$$\frac{d}{dt} H(\rho_t|\nu)=-\int_{\mathcal{M}} \rho_t(x)\left\|\operatorname{grad} \log \frac{\rho_t(x)}{\nu(x)}\right\|^2 dx \leqslant-2 \alpha H(\rho_t|\nu)$$

Therefore,

$$H(\rho_t|\nu) \leqslant e^{-2 \alpha t} H(\rho_0|\nu)$$

21 OU Process

Consider OU process,

$$dy_t = -y_t dt + \sqrt{2} dB_t, \quad y_0 \sim q_0,$$

which has the analytic solution

$$y_t \overset{d}{=} \lambda_t y_0 + \sigma_t W, \quad W \sim \mathcal{N}(0, I)$$

where

$$\lambda_t = e^{-t}, \quad \sigma_t = \sqrt{1 - e^{-2t}}$$

And the corresponding flow is

$$\dot{y}_t = -y_t - \nabla_x \log q_t(x)$$

OU process can be reparameterized from a Brownian motion, which is

$$dx_t = dB_t,\ x_0 \sim p_0$$

with density $p_t = p_0 \star \mathcal{N}(0, tI)$ that satisfies the heat equation*

$$\partial_t p_t(x) = \frac{1}{2} \Delta_x p_t(x)$$

and the heat flow

$$\dot{x}_t = -\frac{1}{2} \nabla_x \log p_t(x)$$

Proposition 21.1. *Suppose that $(x_t)_{t \geq 0}$ satisfies the probability flow ODE for Brownian motion starting at p_0; that is, letting $p_t = p_0 * \mathcal{N}(0, tI)$, we have $x_0 \sim p_0, \dot{x}_t = -\frac{1}{2} \nabla \ln p_t(x_t)$. Then, if we set*

$$y_t = e^{-t} x_{e^{2t}-1},$$

then $(y_t)_{t>0}$ satisfies the probability flow ODE for the OU process starting at p_0, that is

$$q_t(y) \propto p_{e^{2t}-1}\left(e^t y\right)$$

References

Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré (2008). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

Chen, Sitan, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang (2023). "Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions". In: *The Eleventh International Conference on Learning Representations*.

Le Gall, Jean-François (2016). *Brownian motion, martingales, and stochastic calculus*. Springer.

Liptser, Robert S and Albert N Shiryaev (2013). *Statistics of random processes: I. General theory*. Vol. 5. Springer Science & Business Media.

Revuz, Daniel and Marc Yor (2013). *Continuous martingales and Brownian motion*. Vol. 293. Springer Science & Business Media.