MAT1600/1601 Probability Theory

1 Basics

1.1 Probability Spaces

A probability space is a triple $(\Omega, \mathcal{F}, P)$, where:

1. Ω is the sample space: all possible outcomes of some random experiment.
2. $\mathcal{F}$ is the σ-field (σ-algebra, σ-ring) of events. It is a collection of subsets of Ω with the following properties. It represents the amount of information we have about the outcome of the experiment.
   (a) $\emptyset \in \mathcal{F}$
   (b) $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$
   (c) $A_1, A_2, ... \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
3. P is the probability measure that assigns numbers in $[0,1]$ to events, $\forall A \in \mathcal{F}$, $P(A) \in [0,1]$
   (a) $P(A) \geq P(\emptyset) = 0$, $\forall A \in \mathcal{F}$
   (b) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for A_i disjoint.
   (c) $P(\Omega) = 1$

***Proposition:* 1.1: Properties of σ-Fields**

1. If $\mathcal{F}_i$ are σ-fields, then $\cap \mathcal{F}_i$ is a σ-field
2. If G is any collection of subsets of Ω, then there exists a smallest σ-field containing G, called $\sigma(G)$.

Proof. The first statement follows from the definition. For the second statement, take intersection of all σ-field containing G. □

Example: Flip a coin $\Omega = \{0, 1\}$,
$\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$, $P(\emptyset) = 0$, $P(\{0,1\}) = 1$, $P(\{0\}) = p$, $P(\{1\}) = 1 - p$

Example: Flip 2 coins $\Omega = \{00, 01, 10, 11\}$, $\mathcal{F} = 2^{\Omega}$.
We can also define $\Omega = \{$zero heads, 1 head, 2 heads$\}$. Sample spaces are not unique for the same experiment.

Example: Flip a fair coin ∞ many times: $\Omega =$infinite binary strings.
Any $\omega \in (0,1)$ can be represented by $0.w_1 w_2 ....$
$A = [\frac{i}{2^n}, \frac{j}{2^n})$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \{\emptyset, [0, 0.5), [0.5, 1), \Omega\}$
$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$, $\mathcal{F} = \sigma(\bigcup \mathcal{F}_n)$ is the Borel sets.

Suppose $\Omega = [0, 1)$, why don't we use $\mathcal{F} = 2^{[0,1)}$?

Proof. Define equivalence $x \sim y$ if $x - y \in \mathbb{Q}$. There are uncountably many equivalent classes. Axiom of choice allows us to create a set B containing exactly one element of each equivalent class.

$[0,1) = \bigcup_{q\in\mathbb{Q}} \tau_q B$, where $\tau_q B = \{x+q : x \in B\}$ are all disjoint. But $P([0,1)) = \sum_q P(\tau_q B) = \sum_q P(B)$ which is either 0 or infinity. Contradiction. □

Proposition: 1.2: Monotonicity and Continuity of P

1. If $A \subset B$, then $P(A) \le P(B)$
2. If $A_i \in \mathcal{F}$, then $P(\bigcup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty P(A_i)$.
3. If $A_i \uparrow A$, *i.e.* $A_1 \subset A_2 \subset \cdots$, and $\bigcup_{i=1}^\infty A_i = A$, then $P(A_i) \uparrow P(A)$
4. If $A_i \downarrow A$, *i.e.* $A_1 \supset A_2 \supset \cdots$, and $bigcapi = 1^\infty A_i = A$, then $P(A_i) \downarrow P(A)$

Proof. (1) $B = A \cup (B \setminus A)$, $P(B) = P(A) + P(B\setminus A)$, and $P(B \setminus A) \ge 0$.

(2) Let $\tilde{A}_1 = A_1$, $\tilde{A}_2 = A_2 \setminus \tilde{A}_1$, $\tilde{A}_n = A_n \setminus \bigcup_{i=1}^{n-1} \tilde{A}_i$.
Then $P(\bigcup_{i=1}^\infty \tilde{A}_i) = \sum_{i=1}^\infty P(\tilde{A}_i) \le \sum_{i=1}^\infty P(A_i)$.

(3) $P(A) = P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(\tilde{A}_i) = \lim_{n\to\infty} \sum_{i=1}^n P(\tilde{A}_i) = \lim_{n\to\infty} P(A_n)$.

(4) Since $P(A_i) = 1 - P(A_i^C)$, $A_1^C \subset A_2^C \subset ...$ and $\bigcup A_i^C = A^C$. Therefore, by the previous argument, $P(A_i^C) \uparrow P(A^C)$, and $P(A_i) \downarrow 1 - P(A^C) = P(A)$. □

1.2 Distributions and Densities

Definition: 1.1: Probability Measures on $\mathbb{R}$

Define $F(x) = P(-\infty, x]$. $F(x)$ is a distribution function if it satisfies the following properties:

1. F is non-decreasing
2. F is right-continuous, *i.e.* $\lim_{y\downarrow x} F(y) = F(x)$, because if $y_n \downarrow x$, then $(-\infty, y_n) \downarrow (-\infty, x)$ and it then follows Prop 1.2.
3. $F(-\infty) = 0$, $F(\infty) = 1$

If $F(x) = \int_{-\infty}^x f(y)dy$, then $f(y)$ is the density, $\int_{-\infty}^\infty f(y)dy = 1$.

Probability measures on $\mathbb{R}$ has 1-1 correspondance with distribution functions.

Examples (Discrete):

1. Dirac: $P(A) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}$
2. Sum of Diracs: $P(\{x_i\}) = p_i$, $\sum p_i = 1$
3. Poisson: at $x_i = i$ for $i = 0, 1, 2$, $P(\{i\}) = \frac{\lambda^i}{i!} e^{-\lambda}$
4. Geometric: $P(\{i\}) = (1-p)^{i-1} p$

Examples (Continuous):

1. Uniform on $[0,1)$: $f(x) = \begin{cases} 1, x \in [0,1) \\ 0, \text{else} \end{cases}$
2. Uniform on $[a,b)$: $f(x) = \begin{cases} \frac{1}{b-a}, x \in [a,b) \\ 0, \text{else} \end{cases}$
3. Exponential: $f(x) = \lambda e^{-\lambda x} \mathbb{1}(x > 0)$

4. Gaussian/Normal: $f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

For any right continuous functins F, we can decompose

$$F(x) = \int_{-\infty}^{x} f(y)dy + \int_{-\infty}^{x} \sum \delta dy,$$

where the first part is absolutely continuous ($f \geq 0$ and $\int_{-\infty}^{\infty} f = 1$)

To estimate the normal distribution $F(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$, we can compute the complement:

$$1 - F(x) = \int_{x}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \leq \int_{x}^{\infty} \frac{y}{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = \frac{1}{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

1.3 Random Variables

***Definition:* 1.2: Random Variable**

A random variable is a function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ s.t. $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ for any $A \in \mathcal{B}$ (measurable) *i.e.* X is only using allowable information.
Similarly, $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n)$ is a random vector.

To check that X is measurable, we just need to check that $X^{-1}((-\infty, x])$ is measurable in $\mathcal{F}$. This defines a natural push-forward proability measure and distribution, $F(x) = P(X^{-1}(-\infty, x])$.

If $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, then it pushes foward P to a probability measure μ on $\mathbb{R}$ by $\mu(B) = P(X^{-1}(B))$.

Example: If $\mathcal{F} = \{\emptyset, \Omega\}$ (know nothing), then X must be constant.

Lemma 1. *If G generates $\mathcal{F}'$, then $X^{-1}(A) \in \mathcal{F}$ for $A \in G \Rightarrow X^{-1}(A) \in \mathcal{F}$ for $A \in \mathcal{F}'$*

Proof. Let $\mathcal{G} = \left\{A \in \mathcal{F}' : X^{-1}(A) \in \mathcal{F}\right\}$. For sure $G \subset \mathcal{G}$.

But $\mathcal{G}$ is a σ-field. Since G generates $\mathcal{F}'$, $\mathcal{G} = \mathcal{F}'$. □

***Proposition:* 1.3: Properties of Random Variables**

1. Composition: if $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ and $Y : (\Omega', \mathcal{F}') \to (\Omega'', \mathcal{F}'')$ are random variables, then $Y \circ X : (\Omega, \mathcal{F}) \to (\Omega'', \mathcal{F}'')$ is a random variable.
2. If $X_1, ..., X_n$ are random variables $(\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, then $X = (x_1, ..., x_n)$ is a random vector.
3. If $X_1, ..., X_n$ are random variables, then $X = X_1 + \cdots X_n$ is a random variable.
4. If $X_1, ..., X_n$ are random variables, then $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$, $\limsup_n X_n$ are random variables. Here we consider $\mathbb{R} \cup \{\pm\infty\}$ with Borel set generated by $[-\infty, x)$
5. $\{X_n \text{ converges}\} = \{\omega : X_n(\omega) \text{ converges}\}$ is measurable.

Proof. 1) $(Y \circ X)^{-1}(A'') = X^{-1}(Y^{-1}(A'')) = X^{-1}(A')$ measurable.

2) consider $X^{-1}\left(\prod_{i=1}^{n} [a_i, b_i)\right) = \bigcap_{i=1}^{n} X_i^{-1}([a_i, b_i)) \in \mathcal{F}$.

3) Need to show that $\{\omega : X_1(\omega) + \cdots X_n(\omega) \leq x\} \in \mathcal{F}$. This can be done by showing that $f(x_1, ..., x_n) = x_1 + \cdots + x_n$ is measurable $(\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}, \mathcal{B})$.

4) $\{\inf_n X_n \geq x\} = \bigcap \{X_n \geq x\}$, $\sup_n X_n = -\inf_n(-X_n)$, $\liminf_n X_n = \sup_n \inf_{m \geq n} X_m$.

5) $\{\omega : X_n(\omega) \text{ converges}\} = \{\liminf X_n - \limsup X_n \geq 0\}$ □

1.4 Expectation

***Definition:* 1.3: Expectation**

The expectation of a random variable X is defined as

$$E[X] = \int X(\omega)dP(\omega) = \int fd\mu$$

If X takes countably many values x_i with probabilities p_i, then $E[X] = \sum_i x_i p_i$, if $\sum_i |x_i| p_i < \infty$.

For continuous random variables, we need to start with simple functions:

Step 1: Simple functions: $\phi = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ for A_i disjoint, $\int \phi d\mu = \sum_{i=1}^n a_i \mu(A_i)$

***Proposition:* 1.4: Properties of Expectation**

1. If $\phi \geq 0$, then $\int \phi d\mu \geq 0$
2. $\int a\phi d\mu = a \int \phi d\mu$
3. $\int (\phi + \psi) d\mu = \int \phi d\mu + \int \psi d\mu$
4. If $\phi \leq \psi$, then $\int \phi d\mu \leq \int \psi d\mu$
5. If $\phi = \psi$, then $\int \phi d\mu = \int \psi d\mu$
6. $\left| \int \phi d\mu \right| \leq \int |\phi| \, d\mu$

Step 2: Bounded functions f, $\displaystyle\int fd\mu = \sup_{\phi \leq f} \int \phi d\mu = \inf_{\psi \geq f} \int \psi d\mu$ for ϕ, ψ simple.

Proof. We show why $\displaystyle\sup_{\phi \leq f} \int \phi d\mu = \inf_{\psi \geq f} \int \psi d\mu$:

1) $\leq$: obvious since $\phi \leq f \leq \psi$, then apply 4 in Proposition 1.4

2) $\geq$: Since f is bounded, $|f| \leq M$. Consider $E_k = \left\{ f^{-1} \left(\left[\frac{(k-1)M}{n}, \frac{kM}{n} \right) \right) \right\}$.

$$\psi_n = \sum_k \frac{kM}{n} 1_{E_k} \geq f \geq \sum_k \frac{(k-1)M}{n} 1_{E_k} = \phi_n$$

Then $\int (\psi_n - \phi_n) d\mu = \frac{M}{n} \to 0$ as $n \to \infty$. □

Step 3: $f \geq 0$, $\displaystyle\int fd\mu = \sup_{0 \leq g \leq f, |g| \leq M} \int gd\mu$.

Define $\min\{f, n\} = f \wedge n$, consider $\displaystyle\lim_{n\to\infty} \int (f \wedge n) d\mu$. The limit exists since it is increasing.

Also, $\displaystyle\lim_{n\to\infty} \int (f \wedge n) d\mu \leq \int fd\mu$. Even $g \leq n$ for some n, so we get the other way.

Step 4: For general f, separate the positive and negative parts. Define $f_+ = f \vee 0 = \max\{f, 0\}$, $f_- = (-f)_+$. $f = f_+ - f_-$. If $\int f_+ d\mu < \infty$ and $\int f_- d\mu < \infty$, we define $\int fd\mu = \int f_+ d\mu - \int f_- d\mu$.

Example: Cauchy distribution $f(x) = \frac{1}{\pi(1+x^2)}$. $f \geq 0$. $\int fdx = 1$, but $\int xf(x)dx \neq 0$.

Theorem: 1.1: Jensen's Inequality

If f is a convex function and X a random variable, then

$$\phi(E[X]) \leq E[\phi(X)]$$

1.5 Measure Theory

Definition: 1.4: Semi-Algebra

A semi-algebra $\mathcal{S}$ is a collection of sets with the following properties:

1. $S, T \in \mathcal{S} \Rightarrow S \cap T \in \mathcal{S}$
2. $S \in \mathcal{S} \Rightarrow S^C$ =finite disjoint union of sets $\in \mathcal{S}$

Definition: 1.5: Algebra (Field)

An algebra $\mathcal{A} \subset 2^S$ is a collection of sets with the following properties:

1. $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$
2. $A_1, ..., A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$

If in addition, $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$ under countable union, then it is a σ-algebra.

Lemma 2. *If $\mathcal{S}$ a semi-algebra, then $\mathcal{T} = \{$finite disjoint union of sets in $\mathcal{S}\}$ is algebra (algebra generated by $\mathcal{S}$)*

Theorem: 1.2: Caratheodory Extension Theorem

Let $\mathcal{S}$ be a semi-algebra, μ a measure on $\mathcal{S}$ with $\mu(\emptyset) = 0$. If the following 2 statements are true:

1. If $S, S_i \in \mathcal{S}$ and $S = \bigcup_{i=1}^n S_i$ finite disjoint union, Then $\mu(S) = \sum_{i=1}^n \mu(S_i)$
2. If $S, S_i \in \mathcal{S}$ and $S = \bigcup_{i=1}^\infty S_i$, then $\mu(S) \leq \sum_{i\geq 1} \mu(S_i)$.

Then μ has a unique extension $\overline{\mu}$ that is a measure on $\overline{\mathcal{S}}$ (algebra generated by $\mathcal{S}$).
If $\overline{\mu}$ is σ-finite, then there exists a unique extension ν that is a measure of $\sigma(\mathcal{S})$ (smallest σ-algebra containing $\mathcal{S}$)

Lemma 3. *If only 1 of Theorem 1.2 is true, then (1) if $A, B_i \in \overline{\mathcal{S}}$, $A = \bigcup_{i=1}^n B_i$ disjoint union, then $\overline{\mu}(A) = \sum_i \overline{\mu}(B_i)$; (2) if $A, B_i \in \overline{\mathcal{S}}$, $A \subset \bigcup_{i=1}^n B_i$, then $\overline{\mu}(A) \leq \sum_i \mu(B_i)$*

Definition: 1.6: π, λ Systems

$\mathcal{P}$ is a π-system if $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$. $\mathcal{L}$ is a λ-system if (1) $\Omega \in \mathcal{L}$; (2) $A, B \in \mathcal{L}$ and $A \in B$ $\Rightarrow B \setminus A \in \mathcal{L}$; (3) $A_n \in \mathcal{L}$ and $A_n \uparrow A \Rightarrow A \in \mathcal{L}$.

Theorem: 1.3: π-λ Theorem

If $\mathcal{P}$ is π-system and $\mathcal{L}$ is λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Definition: 1.7: Outer Measure

If $E \subset \Omega$, $\mu^*(E) = \inf\{\sum_i \mu(A_i) : A_i \in \mathcal{A}, E \subset \bigcup_i A_i\}$. μ^* is an outer measure with the following properties:

1. $\mu^*(\emptyset) = 0$
2. Monotonicity: If $E \subset F$, then $\mu^*(E) \leq \mu^*(F)$
3. σ-additivity: $F \subset \bigcup_{i=1}^{\infty} F_i$, then $\mu^*(F) \leq \sum_{i=1}^{\infty} \mu^*(F_i)$

Definition: 1.8: Measurable Sets

E is measurable if $\forall F \subset \Omega$, $\mu^*(F) = \mu^*(F \cap E) + \mu^*(F^C \cap E)$.

Lemma 4. *Let $A \in \mathcal{A}$. Then $\mu^*(A) = \mu(A)$ and A is measurable.*

Lemma 5. *The class $\mathcal{A}^*$ of measurable sets is a σ-field and $\mu^*|_{A^*}$ is a measure.*

1.6 Convergence of Random Variables

Definition: 1.9: Convergence of Random Variables

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables.

1. $X_n \overset{a.s.}{\to} X$ (almost surely convergence) if $X_n(\omega) \to X(\omega)$ for almost every ω, equivalently, $P(X_n \to X) = 1$.
2. $X_n \overset{p}{\to} X$ (X_n converges to X in probability) if $P(|X_n - X| > \epsilon) = P(\{\omega : |X_n(\omega) - X(\omega| > \epsilon\}) \to 0$ for all $\epsilon > 0$
3. $X_n \overset{d}{\to} X$ (X_n converges to X in distribution) if $F_n(x) = P(X_n \leq x) \to F(x) = P(X \leq x)$ for all continuity points of F.

Example: $X_n = \frac{1}{n}$, $X_n \to 0$, $F_n(0) = 0$, but $F(0) = 1$.

Proof. 3$\not\Rightarrow$1,2, because X_n in 3 may not live in the same probability spaces

1$\Rightarrow$2: $\{X_n \to X\} = \bigcap_m \bigcup_N \bigcap_{n \geq N} \{|X_n - X| < \frac{1}{m}\}$, *i.e.* $\forall m$, $\exists N$ s.t. $\forall n \geq N$, $|X_n - X| < \frac{1}{m}$.

If $P(\{X_n \to X\}) = 1$, then $P\left(\bigcup_N \bigcap_{n \geq N} \{|X_n - X| < \frac{1}{m}\}\right) = 1$, for all m.

Theerefore, $\lim_{N\to\infty} P\left(\bigcap_{n \geq N} \left\{|X_n - X| < \frac{1}{m}\right\}\right) = 1$, because the intersections are an increasing set.

Then $\lim_{n\to\infty} P\left(\left\{|X_n - X| < \frac{1}{m}\right\}\right) = 1$.

2$\Rightarrow$3,

$$\limsup_{n\to\infty} P(X_n \leq X) = \limsup_{n\to\infty} P(X_n \leq X, |X_n - x| < \epsilon) \leq \limsup_{n\to\infty} P(X \leq x + \epsilon),$$

so $\limsup_{n\to\infty} P(X_n \leq x) \leq \lim_{\epsilon \downarrow 0} P(X \leq x + \epsilon)$. Similarly,

$$\liminf_{n\to\infty} P(X_n \leq x) = \liminf_{n\to\infty} P(X_n \leq x, |X_n - X| < \epsilon) \geq \lim_{\epsilon \uparrow 0} P(X \leq x - \epsilon)$$

□

Lemma: 1.1: Portmanteau Lemma

If C is clsoed, then $\limsup_{n\to\infty} P_n(C) \le P(C)$
If O is open, then $\liminf_{n\to\infty} P_n(O) \ge P(O)$

Theorem: 1.4: Bounded Convergence Theorem

Let μ be a probability measure on $(\Omega, \mathcal{F})$, f_n be r.v.s on the space, $|f_n(\omega)| \le M$, $\forall \omega \in \Omega$, $M \in \mathbb{R}$, $f_n \xrightarrow{p} f$, then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

Proof.

$$\left| \int f_n d\mu - \int f d\mu \right| \le \int |f_n - f| \, d\mu = \int_{|f_n - f| > \epsilon} |f_n - f| \, d\mu + \int_{|f_n - f| < \epsilon} |f_n - f| \, d\mu,$$

By 2 of Definition 1.9, if $|f_n| \le M$, then $|f| \le M$ and

$$\int_{|f_n - f| > \epsilon} |f_n - f| \, d\mu + \int_{|f_n - f| < \epsilon} |f_n - f| \, d\mu \le 2M\mu \left(|f_n - f| > \epsilon\right) + \epsilon$$

Also, $\mu \left(|f_n - f| > \epsilon\right) \to 0$. □

Lemma: 1.2: Fatou's Lemma

Let μ be a probability measure on $(\Omega, \mathcal{F})$, f_n be r.v.s on the space. If $f_n \ge 0$, then

$$\liminf_{n\to\infty} \int f_n d\mu \ge \int \liminf_{n\to\infty} f_n d\mu$$

Proof.

$$f_n \ge \inf_{m \ge n} f_m \nearrow \liminf_{n\to\infty} f_n$$

Integrate both sides and take limits:

$$\liminf_{n\to\infty} \int f_n d\mu \ge \int \inf_{m \ge n} f_n d\mu \ge \int \inf_{m \ge n} f_m \wedge L d\mu$$

By Theorem 1.4, RHS$= \int \liminf f_n \cap L d\mu \nearrow \int \liminf f_n d\mu$. □

Theorem: 1.5: Monotone Convergence Theorem

If $f_n \ge 0$, $fn \uparrow f$, then $\int f_n d\mu \to \int f d\mu$

Proof. We need to show that $\limsup \int f_n d\mu \le \int f d\mu$ and $\liminf \int f_n d\mu \ge \inf f d\mu$.
The first is because $\int f_n d\mu \le \int f d\mu$, for all n. The second is by Lemma 1.2 □

Theorem: 1.6: Dominated Convergence Theorem

If $f_n \to f$ a.s., $|f_n| \le g$ and $\int g d\mu < \infty$, then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

Proof. Since $|f_n| \leq g$, $f_n + g \geq 0$, then by Lemma 1.2, $\liminf \int f_n + g d\mu \geq \int f + g d\mu$.

By linearity, $\liminf \int f_n d\mu \geq \int f d\mu$.

Apply the same proof to $-f_n$ to get $\limsup \int f_n d\mu \leq \int f d\mu$. □

Theorem: 1.7: Change of Variable

Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a r.v., $\phi : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$. If $E[|\phi(X)|] < \infty$, then $E[\phi(X)] = \int_{\mathbb{R}} \phi d\mu$, where $\mu(B) = P(X^{-1}(B))$. Since $F(x) = \mu((-\infty, x])$, we can also write as $\int_{\mathbb{R}} \phi dF$.

Proof. Case 1: $\phi = \mathbb{1}_B$,

$$E[\phi(X)] = E[(\mathbb{1}_B)] = P(X \in B) = P(X^{-1}(B)) = \mu(B) = \int_{\mathbb{R}} \mathbb{1}_B(x) d\mu$$

Case 2: $\phi = \sum_{i=1}^n c_i \mathbb{1}_{B_i}$,

$$E[\phi(X)] = \sum_{i=1}^n c_i P(X^{-1}(B_i)) = \sum_{i=1}^n c_i \mu(B_i) = \int \phi d\mu$$

Case 3: $\phi \geq 0$. Take ϕ_n simple, $\phi_n \uparrow \phi$. *e.g.* $\phi_n = \frac{\lfloor 2^n \phi \rfloor}{2^n} \wedge n$. By case 2, $E[\phi_n(X)] = \int \phi_n d\mu$. $E[\phi_n(X)] \uparrow E[\phi(X)]$ and $\int \phi_n d\mu \uparrow \int \phi d\mu$ by Theorem 1.5.

Case 4: General ϕ, $\phi = \phi_+ - \phi_-$. Then apply case 3. □

Definition: 1.10: Mean and Variance

Let X be an r.v. on $(\Omega, \mathcal{F}, P)$. The mean is

$$\mu = E[X] = \int x dF = \int x f(x) dx \text{ (if } F' = f),$$

It may not always exist

The variance of X is

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2 = \inf_y E[(X - y)^2],$$

$\text{Var}(aX) = a^2 \text{Var}(X)$, $\text{Var}(aX + b) = a^2 \text{Var}(X)$

Example: $X \sim N(0, 1)$, $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

Need to check $\int |x| \exp(-x^2/2) dx < \infty$, but $x \exp(-x^2/2)$ is a derivative. Therefore,

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0$$

The variance is:

$$\text{Var}(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx = x e^{-x^2/2}|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$$

The m-th moment is:

$$E[X^m] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^m e^{-x^2/2} dx = \begin{cases} 0, \text{if } m \text{ is odd} \\ (m-1)(m-3)\cdots(1), \text{if } m \text{ is even, by IBP} \end{cases}$$

Example: $X \sim \text{Exp}(1)$, $f(x) = e^{-x}\mathbb{1}(x > 0)$. $E[X^k] = k!$.

Skewness: $E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right]$.
Kurtosis: $E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right]$.

Example: For Gaussian, skewness is 0, kurtosis is 3.

1.7 Independence

Independence is equivalent to product measures.

Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two probability spaces. Define

- $\Omega = \Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$
- $\mathcal{F} = \sigma(\{A_1 \times A_2, A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\})$
- $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. This defines μ on finite unions of rectangles, and extends to μ on $(\Omega, \mathcal{F})$.

Caratheodory says all we ahve to do is check countable additivity on finite unions of rectangles. *i.e.* if $A_1 \times A_2 = \bigcup_i A_1^i \times A_2^i$ disjoint unions, we need to check $\mu(A_1^i \times A_2) = \sum_i \mu_1(A_1^i)\mu_2(A_2^i)$.
Note that $1_{A_1}(x_1)\mu_2(A_2) = \sum_i 1_{A_1^i}(x_1)\mu_2(A_2^i)$. Integrate both sides

$$\int 1_{A_1}(x_1)\mu_2(A_2)d\mu_1(x_1) = \sum_i \int 1_{A_1^i}(x_1)\mu_2(A_2^i)d\mu_1(x_1) \text{ by Theorem 1.5} \tag{1}$$

$$\Rightarrow \mu_1(A_1 \times A_2) = \sum_i \mu_1(A_1^i)\mu_2(A_2^i) \tag{2}$$

Let X_1, X_2 be r.v.s, $P(X_1 \in B_1) = \mu(B_1)$, $P(X_2 \in B_2) = \mu(B_2)$, $P((X_1, X_2) \in B) = \nu(B)$, where ν is a measure on $\mathbb{R}^2$. If $\nu = \mu_1 \times \mu_2$, then X_1 and X_2 are *independent*.

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2) \Leftrightarrow \nu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $A_1, A_2 \in \mathcal{F}$ are independent if 1_{A_1} and 1_{A_2} are *independent*:

$$P(1_{A_1} \in B_1, 1_{A_2} \in B_2) = P(1_{A_1} \in B_1)P(1_{A_2} \in B_2) \Rightarrow P(A_1 \cap A_2) = P(A_1)P(A_2)$$

Covariance: $\text{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2]$.
Correlation: $\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}$

X_1, X_2 are uncorrelated if $\text{Corr}(X_1, X_2) = 0$, but uncorrelated $\not\Rightarrow$ independent.

Example: $X \sim N(0, 1)$, $\text{Corr}(X, X^2) = 0$, but X and X^2 are not independent.

$X = (X_1, ..., X_n)$ is a Gaussian vector $X \sim N(\mu, C)$ if

$$P(X \in B) = \frac{1}{\sqrt{2\pi \det(C)}} \int_B \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)dx\right),$$

where $\{C_{ij}\}_{ij}$, $C_{ij} = \text{Cov}(X_i, X_j)$ is symmetric non-negative definite.

Remark 1. If X is Gaussian vector, then $\text{Corr}(X_i, X_j) = 0 \Rightarrow X_i, X_j$ are independent.

Proof. If we write $C = A^TA$, then $X = Az + \mu$, where $z \sim N(0, I)$, z_i are independent. $\square$

Theorem: 1.8: Maxwell

If X_1, X_2 are independent and $O(X_1, X_2)$ (the coordinates after rotation by O) are independent, then X_1, X_2 are Gaussian.

Definition: 1.11: Independence

Random variables $X_1, ..., X_n$ are independent if

$$P(X_1 \in B_1, ..., X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$$

$A_1, ..., A_n \in \mathcal{F}$ are independent if $1_{A_1}, ..., 1_{A_n}$ are independent or equivalently $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$ for all $I \subset \{1, ..., n\}$.
$A_1, ..., A_n$ are pairwise independent if $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i, j \in \{1, ..., n\}$.
For a probability space $(\Omega, \mathcal{F}, P)$, sub-σ-fields $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$ are independent if A_1, A_2 are independent for any $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$.

Independence $\Rightarrow$ pairwise independence, but pairwise independence $\not\Rightarrow$ independence.

Example: $X_1, X_2, X_2 \sim \text{Ber}(\frac{1}{2})$. $A_1 = \{X_1 = X_2\}$, $A_2 = \{X_1 = X_3\}$, $A_3 = \{X_2 = X_3\}$.

$P(A_i \cap A_j) = \frac{1}{4}$, $P(A_i) = \frac{1}{2}$, so they are pairwise independent. However, $P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$, so they are not independent.

Theorem: 1.9: Fubini

If $f \geq 0$ or $\int |f| d\mu < \infty$, then

$$\int_{X_1} \left[\int_{X_2} f(x_1, x_2) d\mu_1 \right] d\mu_2 = \int_{X_1 \times X_2} f d\mu$$

Proposition: 1.5: Sum of Independent Random Variables

If X_1, X_2 are independent, $X_1 \sim F_1$, $X_2 \sim F_2$, *i.e.* $P(X_1 \leq x) = F_1(x)$, then

$$P(X_1 + X_2 \leq x) = \int F_1(x - y) dF_2(y) = (F_1 * F_2)(x) \text{ (Convolution)}$$

Proof.

$$\begin{aligned} P(X_1 + X_2 \leq x) = E[1_{X_1 + X_2 \leq x}] = \int \int 1_{X_1 + X_2 \leq x} d\mu, &\text{ where } \mu(B) = P((X_1, X_2) \in B) \\ &= \int \int 1_{X_1 + X_2 \leq x} dF_1 dF_2 \\ &= \int dF_2(x_2) \int_{-\infty}^{x - x_2} dF_1(x_1) \\ &= \int F_1(x - x_2) dF_2(x_2) \text{ By Theorem 1.9} \end{aligned}$$

□

Remark 2. If one of X_1, X_2 has a density, then the convolution does $F_1' = f_1$, $P(X_1 + X_2 \leq x) = \int \int_{-\infty}^{x-y} f_1(z) dz dF_2(y) = \int_{-\infty}^{x} \int f_1(z - y) dF_2(y) dz$. Then the density of $X_1 + X_2$ is $\int f_1(z - y) dF_2(y)$.

Example: For n i.i.d. Exponential random variables, $X_1, ..., X_n \sim \text{Exp}(\lambda)$,

$$X_1 + X_2 + ... + X_n \sim \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} 1_{x>0}$$

Note that the Gamma function gives:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

$$\int_0^\infty \lambda^n x^{n-1} e^{-\lambda x} dx = (n-1)!$$

$$\int_0^1 (1-y)^{\alpha-1} y^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Example: $X_1 \sim \text{Gamma}(\alpha, \lambda)$, $X_2 \sim \text{Gamma}(\beta, \lambda)$, where the density of Gamma is $\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} 1_{x \geq 0}$. Then $X_1 + X_2$ has density $\frac{\lambda^{\alpha+\beta} x^{\alpha+\beta-1} e^{-\lambda x}}{\Gamma(\alpha+\beta)}$

Example: $X_1 \sim N(m_1, \sigma_1^2)$, $X_2 \sim N(m_2, \sigma_2^2)$, then $X_1 + X_2 \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

2 Law of Large Numbers

Theorem: 2.1: Chebyshev Inequality

$$P(|X_n - X| > \epsilon) < \frac{E\left[|X_n - X|^2\right]}{\epsilon^2}$$

Theorem: 2.2: Properties of Sum of Random Variables

1. If $X_1, ..., X_n$ are uncorrelated and $E[X_i^2] < \infty$, then $\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{Var}(X_i)$
2. If $E[X_i] = m$, $E[X_i^2] \leq C$, then $E\left[\left|\frac{X_1+\cdots+X_n}{n} - m\right|^2\right] \leq \frac{C}{n}$.
3. If $X_n \to X$ in L^2, then $X_n \to X$ in probability.

By 2, $\frac{X_1+\cdots+X_n}{n} \to m$ in L^2-norm, so $\frac{X_1+\cdots+X_n}{n} \xrightarrow{p} m$. This is the weak law of large numbers.

If $E[X_i] = m_i$, $\frac{\sum_{i=1}^n m_i}{n} \to m$, and $\frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} \to 0$, then $\frac{X_1+\cdots+X_n}{n} \xrightarrow{p} m$.

Lemma: 2.1: Borel-Cantelli

If A_n is a sequence of sets in Ω, $\limsup A_n = \{\omega : \omega \in A_n \ i.o.\} = \bigcap_m \bigcup_{n \geq m} A_n$, where $i.o.$ stands for infinitely often.
If $\sum_{n=1}^\infty P(A_n) < \infty$, then $P(\{A_n \ i.o.\}) = 0$.

Proof. $P(\bigcup_{n=m}^\infty A_n) \downarrow P(A_n \ i.o.)$
Also, $P(\bigcup_{n=m}^\infty A_n) \leq \sum_{n=m}^\infty P(A_n) \to 0$ □

Theorem: 2.3: Strong Law of Large Numbers

If $X_1, ..., X_n$ are identically distributed pairwise independent random variables, and $E[X_i] = m$, $E[|X_i|] < \infty$, then $\frac{X_1+\cdots+X_n}{n} \to m$ a.s.

Proof.

$$\sum_{k=1}^\infty P(X_1 > k) \leq \int_0^\infty P(X_1 > x)dx = E[X_1] < \infty$$

By Lemma 2.1, $P(X_1 > k i.o.) = 0$, so $\frac{S_n}{n} \to m$.

Define $T_n = \sum_{k=1}^n X_k 1_{X_k \leq k}$, $\frac{T_n}{n} \to m$ and $\frac{S_n - T_n}{n} \to 0$ a.s.

Let $\alpha > 1$, by Theorem 2.1,

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\left|\frac{T_{\lfloor \alpha^n \rfloor} - E[T_{\lfloor \alpha^n \rfloor}]}{\lfloor \alpha^n \rfloor}\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2}\sum_{n=1}^{\infty} \frac{Var(T_{\lfloor \alpha^n \rfloor})}{\lfloor \alpha^n \rfloor^2} \\
&= \frac{1}{\epsilon^2}\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} Var(X_k 1_{X_k \leq k})\alpha^{-2n} \\
&= \frac{1}{\epsilon^2}\sum_{k=1}^{n}\left(\sum_{n:\lfloor \alpha^n \rfloor > k} \alpha^{-2n}\right) Var(X_k 1_{X_k \leq k}) \\
&\leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} \frac{17k^{-2}}{1-\alpha^{-2}} E\left[X_1^2 1_{X_1 < k}\right] \text{ (Geometric series, and i.i.d.)} \\
&= C_1 \sum_{k=1}^{\infty} \frac{Var(X_k 1_{X_k < k})}{k^2} \\
&\leq \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} 1_{y<k} 2yP(|X_1| > y)dy = CE[|X_1|]
\end{aligned}$$

So, $\left|\frac{T_{\lfloor \alpha^n \rfloor} - E[T_{\lfloor \alpha^n \rfloor}]}{\lfloor \alpha^n \rfloor}\right| \to 0$ a.s.

Since $E[X_k 1_{X_k < k}] = E[X_1 1_{X_1 < k}] \to E[X_1]$ as $k \to \infty$, then $\frac{E[T_{\lfloor \alpha^n \rfloor}]}{\lfloor \alpha^n \rfloor} = \frac{1}{\lfloor \alpha^n \rfloor}\sum_{k=1}^{\lfloor \alpha^n \rfloor} E[X_k 1_{X_k < k}] \to m$. Therefore, $\frac{T_{\lfloor \alpha^n \rfloor}}{\lfloor \alpha^n \rfloor} \to m$ a.s.

Suppose $\lfloor \alpha^n \rfloor \leq k < \lfloor \alpha^{n+1} \rfloor$, we get

$$\frac{T_{\lfloor \alpha^n \rfloor}}{k} \leq \frac{T_k}{k} \leq \frac{T_{\lfloor \alpha^{n+1} \rfloor}}{k} \leq \frac{T_{\lfloor \alpha^{n+1} \rfloor}}{\lfloor \alpha^n \rfloor} \leq \alpha \frac{T_{\lfloor \alpha^{n+1} \rfloor}}{\lfloor \alpha^{n+1} \rfloor} \to \alpha m$$

Let $\alpha \to 1$ to get $\lim_{k\to\infty} \frac{T_k}{k} = m$. □

Example: Suppose $E[X_i^4] < \infty$, assume $E[X_i] = 0$ and X_is are independent. Then

$$E\left[(X_1 + \cdots + X_n)^4\right] = nE[X_i]^4 + \binom{4}{2}\binom{n}{2}\left(E[X_i^2]\right)^2$$

$$E\left[\left(\frac{X_1 + \cdots + X_n}{n}\right)^4\right] \leq \frac{C}{n^2}$$

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n}\right| > \epsilon\right) \leq \frac{E\left[\left(\frac{X_1+\cdots+X_n}{n}\right)^4\right]}{\epsilon^4} \leq \frac{C}{\epsilon^4 n^2} \text{ By Theorem 2.1}$$

Therefore, $P\left(\left|\frac{X_1+\cdots+X_n}{n}\right| > \epsilon \ i.o.\right) = 0$. Equivalently, $P\left(\left|\frac{X_1+\cdots+X_n}{n}\right| > \frac{1}{m} \ i.o. \text{ for some } m\right) = 0$.

***Theorem:* 2.4: Weak Law of Large Numbers in L^2**

Let $X_1, ..., X_n$ be uncorrelated r.v.s, $E[X_i] = \mu_i$, $S_n = X_1 + \cdots + X_n$. Assume $Var(X_i) \leq C < \infty$. Then $\frac{S_n}{n} - \mu \xrightarrow{p} 0$

Proof.

$$\begin{aligned}
P\left(\left|\frac{S_n}{n}-\mu\right|>\epsilon\right) &\le \epsilon^2 Var\left(\frac{S_n}{n}\right) \text{ By Theorem 2.1}\\
&= \frac{1}{n^2\epsilon^2}\sum_{i=1}^{n} Var(X_i) \le \frac{C}{n^2\epsilon^2} \to 0
\end{aligned}$$

□

Example: (Bernstein polynomials) Let $f \in C^0([0,1])$, $f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m)(1-x)^{n-m} f\left(\frac{m}{n}\right)$ (Expectation of $f\left(\frac{m}{n}\right)$ under binomial distribution). Then $\lim_{n\to\infty} \sup_{x\in[0,1]} |f_n(x) - f(x)| = 0$.

Proof. Let $X_1, ..., X_n$ be i.i.d. with $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$.
Let $S_n = \sum_{i=1}^{n} X_i$. $P(S_n = m) = \binom{n}{m} p^m (1-p)^{n-m}$. $E\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{m=0}^{n} P(S_n = m) f\left(\frac{m}{n}\right) = f_n(p)$ By Theorem 1.1,

$$\begin{aligned}
\left|E\left[f\left(\frac{S_n}{n}\right) - f(p)\right]\right| &\le E\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\right]\left(1\left(\left|\frac{S_n}{n} - p\right| \le \delta\right) + 1\left(\left|\frac{S_n}{n} - p\right| > \delta\right)\right)\\
&\le \epsilon + (2\max|f|)P\left(\left|\frac{S_n}{n} - p\right| > \delta\right)\\
&\le \epsilon + C\frac{1}{\delta^2} Var\left(\frac{S_n}{n}\right)\\
&\le \epsilon + \frac{\delta}{n^2\delta^2} p(1-p)n \to \epsilon \text{ as } n \to \infty
\end{aligned}$$

Therefore, $\limsup_n \sup_p |f_n(p) - f(p)| \le \epsilon$, $\forall \epsilon > 0$. □

***Lemma:* 2.2: Law of Large Numbers for Triangular Arrays**

Let $b_n > 0$, $b_n \to \infty$, $\bar{X}_{n,k} = X_{n,k} 1_{|X_{n,k}|<b_n}$, $\begin{matrix} X_{1,1} & & \\ X_{2,1} & X_{2,2} & \\ X_{2,1} & X_{2,2} & X_{2,3} \end{matrix}$ a triangle of random variables. Rows are independent. In each row, $X_{i,j}$ is independent of $X_{i,k}$ for $j \neq k$. Assume

1. $\sum_{k=1}^{n} P\left(|X_{n,k}| > b_n\right) \to 0$
2. $b_n^{-2} \sum_{k=1}^{n} E\left[\bar{X}_{n,k}^2\right] \to 0$

Then let $S_n = X_{n,1} + \cdots + X_{n,n}$, $a_n = \sum_{k=1}^{n} E\left[\bar{X}_{m,k}\right]$, we have $\frac{S_n - a_n}{b_n} \xrightarrow{p} 0$.

Proof. Let $\bar{S}_n = \bar{X}_{n,1} + \cdots + \bar{X}_{n,n}$

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right) \le P(S_n \ne \bar{S}_n) + P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right)$$

$$P(S_n \ne \bar{S}_n) \le P\left(\bigcup_{k=1}^{n} \{\bar{X}_{n,k} \ne X_{n,k}\}\right)$$

$$\le \sum_{k=1}^{n} P\left(\bar{X}_{n,k} \ne X_{n,k}\right) = \sum_{k=1}^{n} P\left(|X_{n,k}| > b_n\right) \to 0 \text{ by Assumption 1}$$

$$P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right) \le \epsilon^{-2} b_n^{-2} Var(\bar{S}_n) \text{ by Theorem 2.1}$$

$$= \epsilon^{-2} b_n^{-2} \sum_{k=1}^{n} Var(\bar{X}_{n,k}) \le \epsilon^{-2} b_n^{-2} \sum_{k=1}^{n} E[\bar{X}_{n,k}^2] \to 0 \text{ by Assumption 2}$$

□

Lemma: 2.3: Tail-Sum

Let $Y > 0$ be a r.v., then $\int_0^\infty p y^{p-1} P(Y = y) dy = E[Y^p]$.

Lemma: 2.4:

Let $X_1, X_2, ...$ be i.i.d. with $\lim_{x\to\infty} xP(|X_i| > x) = 0$. Let $S_n = X_1 + \cdots + X_n$, $\mu_n = E\left[X_i 1_{|X_i| \le n}\right]$. Then $\frac{S_n}{n} - \mu_n \xrightarrow{p} 0$

Proof. Let $b_n = n$, $X_{n,j} = X_j$, so we build a triangle of random variables like $\begin{matrix} X_1 & & \\ X_1 & X_2 & \\ X_1 & X_2 & X_3 \end{matrix}$. We check the condition for Lemma 2.2

Condition 1: $\sum_{k=1}^{n} P(|X_k| > n) = nP(|X_1| > n) \to 0$

Condition 2:

$$n^{-2} \sum_{k=1}^{n} E[X_k^2 1_{|X_k| < n}] = n^{-1} E[(X_k 1_{|X_k| < n})^2] = n^{-1} \int_0^\infty 2xP\left(\left|X_1 1_{|X_1| \le n}\right| \ge x\right) dx$$

$$= 2n^{-1} \int_0^n xP(|X_1| > x) dx$$

Let $g(x) = xP(|X_1| > x)$. $\lim_{x\to\infty} g(x) = 0$, $g_n(x) = g(nx) \to 0$ as $n \to \infty$.

Change of variable by $\tilde{x} = \frac{x}{n}$, and we get:

$$2n^{-1} \int_0^n xP(|X_1| > x) dx = 2 \int_0^1 n\tilde{x} P(|X_1| > n\tilde{x}) d\tilde{x} > 0$$

□

Theorem: 2.5: General Weak Law of Large Numbers

Let $X_1, X_2, ...$ be i.i.d. Assume $E[X_1] < \infty$. Let $\mu = E[X_1]$, $S_n = X_1 + \cdots + X_n$. Then $\frac{S_n}{n} - \mu \xrightarrow{p} 0$.

Remark 3. Strong law of large number replaces convergence in probability with convergence almost surely.

Proof. Since $E[x1_{|X_1|>x}] \leq E[X_1 1_{|X_1|>x}]$ and $E[|X_1|] < \infty$, we have $xP(|X_1| > x) \leq E[X_1 1_{|X_1|>x}] \to 0$. Thus Lemma 2.4 holds and $\frac{S_n}{n} \to \mu_n$ as $n \to \infty$, where $\mu_n = E[X_1 1_{|X_1|\leq n}] \to \mu$ by Theorem 1.6. □

Cauchy distribution does not satisfy Theorem 2.5, because it is heavy-tailed:

$$P(X_i \leq x) = \int_{-\infty}^{x} \frac{dt}{\pi(1+t^2)}, E[|X_i|] = \infty$$
$$xP(X_i > x) = x\int_{x}^{\infty} \frac{dt}{\pi(1+t^2)} \sim x\frac{2}{\pi x} = \frac{2}{\pi} \not\to 0$$

Example: Let $X_1, ..., X_n$ be i.i.d. random variables in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. By Theorem 2.5,

$$\frac{X_1^2 + \cdots + X_n^2}{n} \xrightarrow{p} E[X_1^2] = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$$

So $P\left(\left|\frac{X_1^2+\cdots+X_n^2}{n} - \frac{1}{12}\right| < \epsilon\right) \to 1$, *i.e.* $\int_A dx_1...dx_n = 1$, where $A = \left\{x : (1-\epsilon)\sqrt{\frac{n}{12}} \leq |x| \leq (1+\epsilon)\sqrt{\frac{n}{12}}\right\}$, and $|x| = \sqrt{X_1^2 + \cdots + X_n^2}$.
This means that the mass concentrates around the boundary/surface for high dimension.

St. Petersburg Paradox Let $X_1, X_2, ...$ be i.i.d. random variables, $P(X_i = 2^j) = \frac{1}{2^j}$ for $j = 1, ...$, $E[X_i] = \sum_{j=1}^{\infty} 2^j 2^{-j} = \infty$. Pay money to play the game, and each round, we have $\frac{1}{2^j}$ chance to win 2^j. This gives infinite reward for infinite plays. $\frac{X_1+\cdots+X_n}{n} \to \infty$.

Now, use triangular arrays. Let $X_{n,k} = X_k$, and choose $b_n \approx 2^{m_n}$.

$$\sum_{i=1}^{n} P\left(|X_{n,k}| > b_n\right) = nP(X_k > b_n) = n\sum_{j=m_n}^{\infty} \frac{1}{2^j} = 2n2^{-m_n} = \frac{2}{b_n} \to 0$$

Therefore, the first property in Lemma 2.2 holds.

$$b_n^{-2}\sum_{k=1}^{n} E\left[X_k^2 1_{X_k \leq b_n}\right] = b_n^{-2} n \sum_{j=1}^{n} 2^{2j}2^{-j} \sim b_n^{-2} 2n2^{m_n} = \frac{2n}{b_n} \to 0$$

The b_n are chosen to be large enough so we get the properties satisfied.

Now we compute a_n:

$$E\left[X_{n,k}1_{|X_{n,k}|\leq b_n}\right] = E[X_1 1_{X_1 \leq b_n}] = \sum_{j=1}^{m_n} 2^j 2^{-j} = m_n, a_n = \sum_{k=1}^{n} E\left[X_{n,k}1_{|X_{n,k}|\leq b_n}\right] = nm_n$$

Let $S_n = X_1 + \cdots + X_n$, $\frac{S_n - nm_n}{b_n} \to 0$ by Lemma 2.2. If $m_n \sim \log_2 n + k_n$, where $k_n = \log\log n$, $k_n \to \infty$, we have $\frac{S_n}{n \log n} \xrightarrow{p} 1$, so we are making $n \log n$ every n rounds.

2.1 Convergence of r.v.s

Theorem: 2.6:

If $X_n \xrightarrow{p} X$, then there is a subsequence n_k with $X_{n_k} \to X$ a.s.

Proof. If $X_n \xrightarrow{p} X$, then $P(|X_n - X| > \epsilon) \to 0$.

Choose n_k s.t. $\sum_{k=0}^{\infty} P\left(|X_{n_k} - X| > \frac{1}{k}\right) < \infty$. This means that $|X_{n_k} - X| < \frac{1}{k}$ for all but finitely many k w.p. 1. □

Proposition: 2.1:

If $X_n \xrightarrow{p} X$ and f is continuous, then $f(X_n) \xrightarrow{p} f(X)$.

Proof. Let $\gamma > 0$, we want N so that for $n \geq N$, $P(|f(X_n) - f(X)| > \epsilon) < \gamma$.

Inside some compact sets, f is uniformly continuous, and we just need $P(|X_n - X| > \delta) < \gamma$.

Outside the compact sets, there is a K so that $\forall n \geq N$, $P(|X_n|, |X| \leq K) < 1 - \gamma$. We can find that K' so that $P(|X| \leq K') > 1 - \frac{\gamma}{100}$. Then there is N so that $P(|X_n - X| > 1) < \frac{\gamma}{100}$. □

Theorem: 2.7: Weak Convergence of Probability Measures

If $X_n \xrightarrow{p} X$ and $X_n \sim \mu_n$, $X \sim \mu$, where $\mu_n(A) = P(X_n \in A)$, then $\mu_n \to \mu$, *i.e.* $\int f d\mu_n \to \int f d\mu$ for all bounded continuous functions f.

Proof. Since $E[f(X_n)] = \int f d\mu_n$ and $E[f(X)] = \int f d\mu$, we just need to show that $E[f(X_n)] \to E[f(X)]$. By Theorem 1.1,

$$\begin{aligned} |E[f(X_n)] - E[f(X)]| &\leq E[|f(X_n) - f(X)|] \\ &= E[|f(X_n) - f(X)|] 1_{|f(X_n)-f(X)|\geq\epsilon} + E[|f(X_n) - f(X)|] 1_{|f(X_n)-f(X)|<\epsilon} \\ &\leq 2BP(|f(X_n) - f(X)| \geq \epsilon) + \epsilon \end{aligned}$$

The last inequality is because f is bounded. □

Lemma: 2.5: 2nd Borel-Cantelli

If $A_1, A_2, ...$ are independent, and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(A_n \; i.o.) = 1$.

Proof. We want $P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) = 1$, which means that $P\left(\bigcup_{n=m}^{N} A_n\right) \uparrow 1$ for all m

and $P\left(\bigcap_{n=m}^N A_n^C\right) \downarrow 0$ for all n.

$$
\begin{aligned}
P\left(\bigcap_{n=m}^N A_n^C\right) &= \prod_{n=m}^N P(A_n^C) = \prod_{n=m}^N (1 - P(A_n^C)) \\
&\le \prod_{n=m}^N e^{-P(A_n)} = e^{-\sum_{n=m}^N P(A_n)} \to 0
\end{aligned}
$$

□

Example: $\Omega = [0,1]$, $P=$Lebesgue, $A_n = [0, 1/n)$.

Example (St. Peterburg game): Let $X_1, X_2, ...$ be i.i.d. s.t. $P(X_i = 2^k) = \frac{1}{2^k}$. We have showed that $\frac{X_1+\cdots+X_n}{n\log_2 n} \xrightarrow{p} 1$.

$$
P(X_i \ge x) = \sum_{k:2^k\ge x} \frac{1}{2^k} \sim \frac{1}{x} \quad P(X_n \ge cn\log n) \sim \frac{1}{cn\log n}
$$

By Lemma 2.5, since $\sum_{n=1}^\infty P(X_n \ge cn\log n) = \infty$, we have $P\left(\frac{X_n}{n\log n} \ge c \; i.o.\right) = 1$ for any c. Therefore, $P\left(\frac{X_1+\cdots+X_n}{n\log n} \ge c \; i.o.\right) = 1$, and $\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n\log n} = \infty$ a.s. Strong law of large number (Theorem 2.3) won't hold in this case.

Empirical Distribution Function Let $X_1, X_2, ...$ be i.i.d., $X_i \sim F$ and $Var(X_i) = \sigma^2 < \infty$. $F_n(x) = \frac{1}{n}\sum_{k=1}^n 1_{\{X_k \le x\}}$. As $n \to \infty$, $F_n(x) - F(x) \to 0$ a.s.

This is by Theorem 2.3, since $F_n(x) = \frac{Y_1+\cdots+Y_n}{n}$, where $E[Y_i] = P(X_i \le x) = F(x)$.

Gilvenko-Cantelli show $\sup_x |F_n(x) - F(x)| \to 0$ a.s.

Proof. If we just had $x_1, ..., x_k$, then $\max_{1\le i\le k} |F_n(x_i) - F(x_i)| \to 0$ as $n \to \infty$.

If F is continuous, take $x_{j,k} = F^{-1}\left(\frac{j}{k}\right)$, $|F_n(x_{j,k}) - F(x_{j,k})| < \frac{1}{k}$ for $n \ge N_k(\omega)$, $\omega \in \Omega$, $j = 0, ..., k$.

$$
\begin{aligned}
F_n(x) &\le F_n(x_{j,k}) \le F(x_{j,k}) + \frac{1}{k} \le F(x) + \frac{2}{k} \\
F_n(x) &\ge F_n(x_{j-1,k}) \ge F(x_{j-1,k}) - \frac{1}{k} \ge F(x) - \frac{2}{k}
\end{aligned}
$$

If F is not continuous, then $x_{j,k} = \inf\left\{y : F(y) \ge \frac{1}{k}\right\}$, $\left|F_n(x_{j,k}^-) - F(x_{j,k}^-)\right| < \frac{1}{k}$. We can do this because Theorem 2.3 also says $F_n(x^-) = \frac{1}{n}\sum_{k=1}^n 1_{X_k < x} \to F(x^-)$. Then the same argument applies. □

We know that $\sum \frac{1}{n^\alpha}$ converges for $\alpha > 1$, diverges for $\alpha \le 1$, and $\sum \frac{(-1)^n}{n}$ converges. Consider the random variables, Y_n, and $\sum \frac{Y_n}{n}$.

Assume $E[X_n] = 0$, $Var(X_n) < \infty$, X_n independent, when does $\sum X_n$ converge?

Since $Var\left(\sum_{n=1}^N X_n\right) = \sum_{n=1}^N Var(X_n)$, so if the variance is under control $\sum_{n=1}^\infty Var(X_n) < \infty$, then it is possible to have convergence. This is because for $S_m = X_1 + \cdots + X_m$, $S_N = X_1 + \cdots + X_N$, Theorem 2.1 gives

$$
P\left(|S_N - S_m| > \epsilon\right) \le \frac{1}{\epsilon^2} \sum_{k=m+1}^N \mathrm{Var}(X_k)
$$

Now we want $P(|S_n - S_m| > \epsilon)$ small for all $n, m \geq N$.
Note that $P\left(\sup_{m\geq N} |S_m - S_N| > \epsilon\right) \Rightarrow P\left(\sup_{m,n\geq N} |S_m - S_n| > \epsilon\right)$ by triangle inequality.

Let $X_1, ..., X_m, X_{m+1}, ..., X_n$ be independent r.v.s. $Y = f(X_1, ..., X_m)$, $Z = g(X_{m+1}, ..., X_n)$, where f, g are measurable. Then Y and Z are independent.

***Theorem:* 2.8: Kolmogorov Maximal Inequality**

Let $X_1, ..., X_n$ be i.i.d. $E[X_i] = 0$, $Var(X_i) = \sigma^2 < \infty$. $S_n = X_1 + \cdots + X_n$,

$$P\left(\max_{1\leq k\leq n} |S_k| \geq x\right) \leq \frac{1}{x^2} Var(S_n)$$

Proof. Break if by the first time that $|S_k| \geq x$. Let A_k be the event that the first time that $|S_k| \geq x$ is at k. $A_k = \{|S_k| \geq x, |S_j| < x, j = 1, ..., k-1\}$. $\bigcup_{k=1}^n A_k = \{\max_{1\leq k\leq n} |S_k| \geq x\}$. Since $E[X_i] = 0$,

$$\begin{aligned}
Var(S_n) &= E[S_n^2] \geq E[S_n^2 1_{|S_n|\geq x}] \geq x^2 P(|S_n| \geq x) \\
&\geq \sum_{k=1}^n \int 1_{A_k} S_n^2 dP \\
&= \sum_{k=1}^n \int 1_{A_k} S_k^2 + 2 1_{A_k}(S_n - S_k) + 1_{A_k}(S_n - S_k)^2 dP \\
&\geq \sum_{k=1}^n \int 1_{A_k} S_k^2 dP \\
&\geq x^2 \sum_{k=1}^n \int 1_{A_k} dP = x^2 P\left(\max_{1\leq k\leq n} |S_k| \geq x\right)
\end{aligned}$$

Note that $\int (S_n - S_k) dP = \int X_{k+1} + \cdots + X_n dP = 0$ and $\int (S_n - S_k)^2 dP > 0$. ☐

***Theorem:* 2.9:**

Let $X_1, X_2, ...$ be independent, $E[X_i] = 0$, $Var(X_i) = 0$. If $\sum_{n=1}^{\infty} Var(X_n) < \infty$, then S_n is Cauchy.

Proof.

$$P\left(\max_{M\leq m\leq N} |S_m - S_M| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{m=M+1}^{N} Var(X_m)$$

$$P\left(\max_{m\geq M} |S_m - S_M| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{m=M+1}^{\infty} Var(X_m)$$

$$\lim_{M\to\infty} P\left(\sup_{n,m\geq M} |S_m - S_n| \geq \epsilon\right) = 0$$

Also $T_M = \sup_{n,m\geq M} |S_m - S_n|$ is a decreasing sequence and it has a limit.

The limit must be 0 w.p. 1, or the limit cannot hold. This means that S_n is Cauchy w.p. 1. ☐

Lemma: 2.6: Kronecker's Lemma

Let $a_n \uparrow \infty$. If $\sum_{k=1}^{\infty} \frac{x_k}{a_k} < \infty$, then $\frac{1}{a_n} \sum_{k=1}^{n} x_k \to 0$.

Proof. Let $b_m = \sum_{k=1}^{m} \frac{x_k}{a_k}$, $b_m \to b$, and $x_m = a_m(b_m - b_{m-1})$.

$$\begin{aligned}\frac{1}{a_n}\sum_{k=1}^{n} x_k &= \frac{1}{a_n}\left(\sum_{m=1}^{n} a_m b_m - \sum_{m=1}^{n} a_m b_{m-1}\right)\\ &= \frac{1}{a_n}\left(a_n b_n + \sum_{m=2}^{n} a_{m-1} b_{m-1} - \sum_{m=1}^{n} a_m b_{m-1}\right)\\ &= b_n - \sum_{m=1}^{n} \frac{a_m - a_{m-1}}{a_n} b_{m-1}.\end{aligned}$$

Let $\epsilon > 0$, n large so that $\frac{a_m}{a_n} \le \frac{\epsilon}{4B}$, M large so that if $m \ge M$, $|b_m - b| < \frac{\epsilon}{2}$, $|b_m| \le B$. Then

$$\left|\sum_{m=1}^{n}\right| \frac{a_m - a_{m-1}}{a_n}(b_{m-1} - b) \le \frac{a_m}{a_n} 2B + \frac{a_n - a_M}{a_n}\,|b_M - b| < \epsilon$$

Therefore, we get $\frac{1}{a_n}\sum_{k=1}^{n} x_k \to 0$. □

Theorem: 2.10: Kolmogorov 3-Series

Let $X_1, X_2, \ldots$ be independent, $\sum_{n=1}^{\infty} X_n$ converges if and only if the following 3 series converge:

1. $\sum_{n=1}^{\infty} P(X_n > A) < \infty$ for each A
2. $\sum_{n=1}^{\infty} E[X_n 1_{|X_n| \le A}] < \infty$
3. $\sum_{n=1}^{\infty} Var(X_n 1_{|X_n| \le A}) < \infty$

Proof. ($\Rightarrow$) 3$\Rightarrow \sum_{n=1}^{\infty} X_n 1_{|X_n| \le A} - E[X_n 1_{|X_n| \le A}] < \infty$

2$\Rightarrow \sum_{n=1}^{\infty} X_n 1_{|X_n| \le A}$ converges

1$\Rightarrow P(|X_n| > A\ i.o.) = 0 \Rightarrow \sum_{n=1}^{\infty} X_n$ converges w.p. 1

($\Leftarrow$) $X_1, X_2, \ldots$ are i.i.d., $E[X_i] = 0$, $Var(X_i^2) = \sigma^2 < \infty$.

Theorem 2.8 and Lemma 2.6 implies that it is enough to show $\sum_{n=1}^{\infty} Var\left(\frac{X_n}{a_n}\right) < \infty$, which is equivalent to $\sum_{n=1}^{\infty} \frac{1}{a_n^2} < \infty$, since X_is are i.i.d. with zero mean. In particular, if $a_n = n$, we get Theorem 2.3.

Note that as long as $a_n \uparrow \infty$, $\frac{1}{a_n}\sum_{k=1}^{n} X_k \to 0$. □

Choose $a_n = n^{1/2} (\log n)^{1/2+\epsilon}$, $a_n \uparrow \infty$, $\frac{1}{a_n}\sum_{k=1}^{n} X_k \to 0$ a.s.

When $a_n = n^{1/2}$, we still have $\frac{1}{a_n}\sum_{k=1}^{n} X_k \to 0$ a.s.

When $a_n = \sqrt{2\sigma^2 n \log\log n}$, it has $\limsup = \liminf = 1$ a.s. This is the law of iterated logarithms by Khinchin. This is close to the central limit theorem.

Theorem: 2.11: Marcinkiewitz & Zygmund

Let $X_1, X_2 \ldots$ be i.i.d. If $E\left[|X_i|\right] < \infty$, $1 < p < 2$, but $E[X_i^2]$ can be infinity, then $\frac{1}{n^{1/p}} \sum_{k=1}^{\infty} X_k \to 0$ a.s.

Proof. Let $Y_k = X_k 1_{|X_k| \leq k^{1/p}}$, $T_n = Y_1 + \cdots + Y_n$.

$$\sum_{k=1}^{\infty} P(X_k \neq Y_k) = \sum_{k=1}^{\infty} P(|X_k|^p > k) \leq E\left[|X_k|^p\right] < \infty$$

Lemma 2.1 implies $P(X_k \neq Y_k i.o) = 0$, so it is enough to show $\frac{T_n}{n^{1/p}} \to 0$.

By Theorem 2.10 and Lemma 2.6, we just need to show that $\sum_{n=1}^{\infty} Var\left(\frac{Y_n}{n^{1/p}}\right)$ is finite. □

3 Central Limit Theorem

3.1 Weak Convergence

Recall the weak convergence of probability measures in Theorem 2.7. We can similarly define the weak convergence of distribution functions. Let F_n, F be distribution functions, $F_n \to F$ weakly if $F_n(x) \to F(x)$ for all continuity points x of F.

Recall Lemma 1.1. We have 4 equivalent statements

1. If A is closed, $\mu_n \to \mu$ weakly, then $\limsup_{n\to\infty} \mu_n(A) \le \mu(A)$
2. If B is open, then $\liminf_{n\to\infty} \mu_n(B) \ge \mu(B)$
3. If $\mu(\partial A) = 0$, where $\partial A = \overline{A} \setminus \text{int}(A)$ and $\mu_n \to \mu$ weakly, then $\mu_n(A) \to \mu(A)$.
4. If $\mu_n(A) \to \mu(A)$ for all A with $\mu(\partial A) = 0$, then $\mu_n \to \mu$ weakly.

Proof. 1. Let $f_\epsilon = \left(1 - \frac{d(x,A)}{\epsilon}\right)_+$.

$$\mu_n(A) = \int 1_A d\mu_n \le \int f_\epsilon d\mu_n \to \int f_\epsilon d\mu \le \mu(A^\epsilon)$$

Therefore, $\limsup_{n\to\infty} \mu_n(A) \le \mu(A^\epsilon)$ for all ϵ. By continuity of probability measures (Proposition 1.2), $\mu(A^\epsilon) \downarrow \mu(A)$ as $\epsilon \to 0$.

2. Apply 1 to the complement of B

3. By 1, $\limsup \mu_n(A) \le \limsup \mu_n(\bar{A}) \le \mu(\bar{A}) = \mu(A)$
By 2, $\liminf \mu_n(A) \ge \liminf \mu_n(\text{int}(A)) \ge \mu(\text{int}(A)) = \mu(A)$

4. Given a continuous function f s.t. $|f| \le M$, $D = \{t : \mu(\{f = t\}) \ne 0\}$ is countable.

Let $\delta > 0$, we want to show that $\left|\int f d\mu_n - \int f d\mu\right| < \delta$.

Choose $-M = t_0 < t_1 < \cdots < t_k = M$, $|t_{i+1} - t_i| < \delta$ s.t. none of t_i is in D. Approximate by simple functions:

$$\begin{aligned}\left|\int f d\mu_n - \int f d\mu\right| \le &\left|\int f d\mu_n - \sum_{i=0}^{k-1} t_i \mu_n(t_i \le f < t_{i+1})\right| < \delta \\ &+ \left|\sum_{i=0}^{k-1} t_i \left(\mu_n(t_i \le f < t_{i+1}) - \mu(t_i \le f < t_{i+1})\right)\right| \to 0 \\ &+ \left|\int f d\mu - \sum_{i=0}^{k-1} t_i \mu(t_i \le f < t_{i+1})\right| < \delta\end{aligned}$$

As $n \to \infty$, $\left|\int f d\mu_n - \int f d\mu\right| \to 0$. □

Theorem: 3.1: Helly's Selection Principle

If F_n are distribution functions, then there always exists a subsequence n_k and F so that $F_{n_k} \to F(x)$ at continuity points of F

Remark 4. F may be degenerate *e.g.* $F_n(x) = 1_{x \ge n}$, $F_n(x) = F(x) = 0$ for $x < n$. F is non-decreasing and right continuous.

Proof. Since $F_n(x) \in [0,1]$, there is a convergent subsequence by Heine-Borel.

Let $\mathbb{Q} = \{q_1, q_2, ...\}$ be the set of rational numbers, $F_{n_k^1}(q_1) \to F(q_1)$, $F_{n_k^2}(q_2) \to F(q_2)$, with $\{n_k^2\} \subset \{n_k^1\}$.

Then the diagonal sequence $F_{n_k^k}(q) \to F(q)$ for $q \in \mathbb{Q}$.

For $q < q'$, $F_{n_k^k}(q) \to F(q) \leq F_{n_k^k}(q') \to F(q')$ by property of distribution functions.

Define $\tilde{F}(x) = \inf_{q>x} F(q)$, $\tilde{F}$ is right continuous. We want to show that $F_{n_k^k}(x) \to \tilde{F}(x)$ at continuity points of $\tilde{F}$.

Let $\epsilon > 0$, assume x is s.t. $F_{n_k^k}(x) \to \tilde{F}(x)$, we can find $\underline{q} < x < \overline{q}$ with

$$F(x) - \frac{\epsilon}{2} < F(\underline{q}) \leq F(x) \leq F(\overline{q}) < F(x) + \frac{\epsilon}{2}$$

Choose N s.t. $\left|F_{n_k^k}(\underline{q}) - F(\underline{q})\right| \leq \frac{\epsilon}{2}$, $\left|F_{n_k^k}(\overline{q}) - F(\overline{q})\right| \leq \frac{\epsilon}{2}$ for $k \geq N$.

Then $F(x) - \epsilon \leq F_{n_k^k}(x) \leq F(x) + \epsilon$ □

Definition: 3.1: Tight

A set of probability measure μ_n is tight if $\forall \epsilon > 0$, there is compact set K s.t. $\mu_n(K^C) < \epsilon$ for all n.

Example: the set of probability measures defined by $F_n(x) = 1_{x \geq n}$ is not tight.

Theorem: 3.2: Prokhorov's Theorem

If μ_n are tight, then there exists μ probability measure and n_k with $\mu_{n_k} \to \mu$ weakly.
Conversely, if $\mu_n \to \mu$ weakly, then μ_n is tight if the metric space is separable and complete.

Proof. Tightness for F_n means there is a M s.t. $1 - F_n(M) + F_n(-M) < \epsilon$.

Choose M a continuity point of F. Take $n \to \infty$, $1 - F(M) + F(-M) < \epsilon$.

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

3.2 Characteristic Functions

Definition: 3.2: Characteristic Function

Let X be a random variable, the characteristic function of X is

$$\phi(t) = E[e^{itX}] = E[\cos tX] + iE[\sin tX] = \int e^{itX} d\mu$$

A closely related concept is moment generating function:

$$M(\lambda) = E[e^{\lambda X}] = \int e^{\lambda X} d\mu$$

Moment generating function is Laplace transform, while characteristic function is Fourier transform. Moment generating function may not exist, while characteristic function always exists, because the integrand is bounded.

Example: if $X \sim Exp(1)$, then $\int_0^\infty e^{\lambda x - x} dx$ is only defined if $\lambda < 1$.

Theorem: 3.3: Properties of Characteristic Functions

If X_1, X_2 are independent r.v.s with characteristic functions ϕ_1, ϕ_2, then

$$E[e^{it(X_1+X_2)}] = E[e^{itX_1}e^{itX_2}] = E[e^{itX_1}]E[e^{itX_2}] = \phi_1(t)\phi_2(t)$$

Other properties:

1. $\phi(0) = E[e^0] = 1$
2. Uniform continuity: $|\phi(t+h) - \phi(t)| =\leq E\left[\left|e^{ihX} - 1\right|\right] \to 0$ as $h \to 0$.
3. $E[e^{it(aX+b)}] = e^{itb}\phi(at)$

Proof. 2. $|\phi(t+h) - \phi(t)| = \left|E\left[e^{i(t+h)X} - e^{itX}\right]\right| = E[e^{itX}\left|e^{ihx} - 1\right|] \leq E\left[\left|e^{ihX} - 1\right|\right] \to 0$ □

Example: $X \sim Ber\left(\frac{1}{2}\right)$, $P(X=1) = P(X=-1) = \frac{1}{2}$.

$$E[e^{itX}] = \frac{1}{2}(e^{it} + e^{-it}) = \cos t$$
$$E[e^{\lambda X}] = \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \cosh t$$

Example: If X is symmetric ($X = -X$), then $\phi(t) = E[e^{itX}] = \int \cos tx d\mu + i \int \sin tx d\mu = \int \cos tx d\mu$ is real.

Example: $N \sim Poisson(\lambda)$,

$$E[e^{itN}] = \sum_{n=0}^{\infty} e^{itn}\frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^{it}\lambda)^n}{n!} = e^{-\lambda}e^{it\lambda} = e^{-\lambda(1-e^{it})}$$

This is by Taylor series of e^x.

Example: $X \sim N(0,1)$.

$$E[e^{itX}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2}dx$$

By change of variable and contour integration:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty-it}^{\infty-it} e^{-\frac{1}{2}y^2}dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}dy = 1$$

If $X \sim N(m, \sigma^2)$, then $\phi(t) = e^{itm - \frac{\sigma^2}{2}t^2}$

Example: $X \sim Unif[0,1]$,

$$E[e^{itX}] = \frac{1}{b-a}\int_a^b e^{itx}dx = \left.\frac{e^{itx}}{it(b-a)}\right|_a^b = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

Example: $X \sim Exp(1)$,

$$E[e^{itX}] = \int_0^{\infty} e^{(it-1)x}dx = \left.\frac{e^{(it-1)x}}{it-1}\right|_0^{\infty} = \frac{1}{1-it}$$

Example: $f(x) = (1 - |x|)_+$. Note that this is the distribution of $X + Y$ for $X, Y \sim Unif\left[-\frac{1}{2}, \frac{1}{2}\right]$.

$$\int_{-1}^{1} (1 - |x|)_+ e^{itx} d = \left(\frac{e^{it/2} - e^{-it/2}}{it}\right) = \left(\frac{2\sin(t/2)}{t}\right)^2 = \frac{2(1 - \cos t)}{t^2}$$

Example: $f(x) = \frac{1}{2}e^{-|x|}$

$$\frac{1}{2}\int_{-\infty}^{\infty} e^{itx-|x|}dx = \frac{1}{2}\left(\int_0^{\infty} e^{itx-x}dx + \int_0^{\infty} e^{-itx-x}dx\right) = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}$$

***Theorem:* 3.4: Inversion of Characteristic Functions**

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

Proof.

$$\begin{aligned}
\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt &= \frac{1}{2\pi}\int_{-T}^{T}\int \frac{e^{-it(a-x)} - e^{-it(b-x)}}{it}d\mu(x)dt \\
&= \frac{1}{2\pi}\int\int_{-T}^{T} \frac{e^{-it(a-x)} - e^{-it(b-x)}}{it}dtd\mu \text{ by Theorem 1.9} \\
&= \frac{1}{2\pi}\int\left[\int_{-T}^{T}\frac{\sin t(x-a)}{t}dt - \int_{-T}^{T}\frac{\sin t(x-b)}{t}dt\right]d\mu \text{ by symmetry in cos}
\end{aligned}$$

We can apply Theorem 1.9, because the integrals are bounded in both dt and $d\mu$.

Note that by symmetry:

$$\int_{-T}^{T}\frac{\sin t(x-a)}{t}dt = 2\int_0^T \frac{\sin t(x-a)}{t}dt = 2\int_0^{T|x-a|}\frac{\sin t}{t}dt \to \frac{\pi}{2}$$

Because, we can rewrite $\frac{1}{t} = \int_0^{\infty} e^{-xt}dx$ and

$$\int_0^T \frac{\sin t}{t}dt = \int_0^T\int_0^{\infty} e^{-xt}\sin t dx dt$$

$$\left|\int_0^T \frac{\sin t}{t}dt - \arctan T\right| \le \frac{2}{T}$$

Take $T \to \infty$, we get:

$$\int_{-T}^{T}\frac{\sin t(x-a)}{t}dt - \int_{-T}^{T}\frac{\sin t(x-b)}{t}dt = \begin{cases} 2\pi a < x < b \\ \pi, x = a, b \\ 0, \text{ otherwise} \end{cases}$$

Then the entire integral converges:

$$\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt \to \frac{1}{2\pi}\int 2\pi 1_{(a,b)}(x)d\mu + \frac{1}{2\pi}\pi\mu(\{a\}) + \frac{1}{2\pi}\pi\mu(\{b\})$$

□

Theorem: 3.5:

If $\int |\phi(t)|dt < \infty$, then there is a density function f s.t. $f(x) = \frac{1}{2\pi}\int e^{-itx}\phi(t)dt$ and f is bounded and continuous.

Remark 5. This is the Fourier transform of L^1 functions.

Proof. From Theorem 3.4, we have

$$\mu((a,b)) + \frac{1}{2}\mu(\{a\}) + \frac{1}{2}\mu(\{b\}) \leq \frac{b-a}{2\pi}\int |\phi(t)|dt$$

In particular, $\{a\}$ and $\{b\}$ do not contribute.

$$\begin{aligned}
\mu((x, x+h)) &= \frac{1}{2\pi}\int \frac{e^{-itx} - e^{-it(x+h)}}{it}\phi(t)dt \\
&= \frac{1}{2\pi}\int\int_x^{x+h} e^{ity}dy\phi(t)dt \\
&= \int_x^{x+h}\left(\frac{1}{2\pi}\int e^{-ity}\phi(t)dt\right)dy \text{ by Theorem 1.9}
\end{aligned}$$

Therefore, f is well-defined and bounded. □

Example: Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$. $E[X]$ is not defined. $\phi(t) = \int \frac{e^{itx}}{\pi(1+x^2)}dx = e^{-|t|}$ by the last example before Theorem 3.4.

Let X_1, X_2 be independent Cauchy, $\phi_{X_1+X_2}(t) = \phi_1(t)\phi_2(t) = e^{-2|t|}$, so $\phi_{\frac{X_1+X_2}{2}}(t) = e^{-|t|}$, $\frac{X_1+X_2}{2} = X_1$. Extending to n independent Cauchy, we have $\frac{X_1+\cdots+X_n}{n} = X_1$ in distribution.

Example: $X_1, X_2 \sim N(0,1)$ independent, then $\phi_{\frac{X_1+X_2}{\sqrt{2}}}(t) = e^{-t^2/2}$, and $\frac{X_1+X_2}{\sqrt{2}} = X_1$. Extending to n independent normal, $\frac{X_1+\cdots+X_n}{\sqrt{n}} = X_1$ in distribution.

Example: $\phi(t) = e^{-|t|^\alpha}$, $\frac{X_1+\cdots+X_n}{n^{1/\alpha}} = X_1$ in distribution, but is ϕ a characteristic function? Yes if $0 \leq \alpha \leq 2$, no if $\alpha > 2$.

Theorem: 3.6: Bochner's Theorem

ϕ is a characteristic function if and only if ϕ is positive definite, *i.e.* for all measurable complex function h, $\iint \phi(t-s)h(t)\bar{h}(s)dtds \geq 0$.

Proof. ($\Rightarrow$) Given ϕ a characteristic function.

$$\begin{aligned}
\iint \phi(t-s)h(t)\bar{h}(s)dtds &= \iiint e^{i(t-s)x}h(t)\overline{h(s)}dtdsd\mu(x) \\
&= \iiint e^{itx}h(t)\overline{e^{isx}h(s)}dtdsd\mu(x) \\
&= E\left[\left(\int e^{itx}h(t)dt\right)\overline{\left(\int e^{itx}h(t)dt\right)}\right] \\
&= E\left[\left|\int e^{itx}h(t)dt\right|^2\right] \geq 0
\end{aligned}$$

□

3.3 Central Limit Theorems

***Lemma:* 3.1:**

$$\mu\left(|x| \geq \frac{2}{u}\right) \leq \frac{1}{u}\int_{-u}^{u}(1-\phi(t))\,dt$$

Proof. By Definition 3.2 and Theorem 1.9, we have:

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt &= \frac{1}{u}\int\int_{-u}^{u}(1-e^{itx})dtd\mu \\ &= 2\int 1-\frac{\sin ux}{ux}d\mu \\ &\geq 2\int_{|x|\geq\frac{2}{u}} 1-\frac{1}{|ux|}d\mu \\ &\geq \int_{|x|\geq\frac{2}{u}} d\mu = \mu\left(|x|\geq\frac{2}{u}\right)\end{aligned}$$

The last two inequalities are by restricting the integration domain. □

***Theorem:* 3.7: Levy's Continuity Theorem**

Let μ_n, μ be probability measures with characteristic functions ϕ_n, ϕ.

1. If $\mu_n \to \mu$ weakly ($\int f d\mu_n \to \int f d\mu$ for bounded continuous f), then $\phi_n \to \phi$ pointwise
2. If $\phi_n \to \phi$ pointwise and ϕ is continuous at 0, then $\mu_n \to \mu$ weakly and μ has characteristic function ϕ.

Proof. 1. $\phi_n = \int e^{itx}d\mu_n \to \int e^{itx}d\mu = \phi$ since e^{itx} is a bounded continuous function.

2. Firstly, we show that μ_n is tight by using Lemma 3.1. Then if we take $\limsup_{n\to\infty}$ on both sides, and choose u s.t. $\limsup_{n\to\infty}\mu_n\left(|x|\geq\frac{2}{u}\right)$ is small, then μ_n is tight by Definition 3.1. Then there exists a subsequence n_k s.t. $\mu_{n_k} \to \tilde{\mu} = \mu$ weakly, and $\mu_n \to \mu$ weakly. □

Counter-example: when ϕ is not continous at 0, we can have $\mu_n \sim N(0,n)$, $\phi_n(t) = e^{-nt^2/2} \to \begin{cases} 1, t=0 \\ 0, \text{ else}\end{cases}$ not a characteristic function.

***Lemma:* 3.2:**

If $\int |X|^n d\mu < \infty$, then $\phi \in C^n$ and $\phi^{(n)}(t) = \int (ix)^n e^{itx}d\mu$. In particular, if $E[|X|^n] < \infty$, then $\phi^{(n)}(0) = i^n E[X^n]$.

Proof. Suppose $\int |X|d\mu < \infty$,

$$\phi'(t) = \frac{\phi(t+h)-\phi(t)}{h} = \int \frac{1}{h}\left(e^{i(t+h)x} - e^{itx}\right)d\mu = \int\left(\frac{e^{ihx}-1}{h}\right)e^{itx}dx$$

Note that $\left|\frac{e^{ihx}-1}{h}\right| \leq |x|$, so by Theorem 1.6, $\phi'(t) \to \int ixe^{itx}d\mu$.

Similarly, $\phi''(t) = \lim_{h\to 0}\int ix\left(\frac{e^{ihx}-1}{h}\right)e^{itx}d\mu$, so if $\int |X|^2 d\mu < \infty$ by Theorem 1.6, $\phi''(t) = \int (ix)^2 e^{itx}d\mu$.

Inductively, we prove for all n. $\square$

Theorem: 3.8: Central Limit Theorem

Let $X_1, X_2, ..., X_n$ be i.i.d. with $Var(X_m) = \sigma^2 < \infty$ and $\sigma \neq 0$. Then,

$$\frac{\sum_{m=1}^n X_m - E[X_m]}{\sigma\sqrt{n}} \to N(0,1)$$

The convergence is in distribution.

Proof. WLOG, assume $E[X_n] = 0$, if not we just define $Y_n = X_n - E[X_n]$. Consider the characteristic function of $\frac{1}{\sigma\sqrt{n}}\sum_{m=1}^n X_m$. Then

$$E\left[e^{i\frac{t}{\sigma\sqrt{n}}\sum_{m=1}^n X_m}\right] = E\left[e^{i\frac{t}{\sigma\sqrt{n}}X_n}\right]^m = \phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

Consider the Taylor expansion, $\phi(0) = 1$, $\phi'(t) = E[iXe^{itX}]$, $\phi'(0) = iE[X]$, $\phi''(t) = E[-X^2e^{itX}]$, $\phi''(0) = -E[X^2]$, $\phi^{(n)}(0) = i^nE[X^n]$.

If things are nice,

$$\phi(t) = E[e^{itX}] = E\left[\sum_{n=0}^{\infty}\frac{i^nt^n}{n!}X^n\right] = \sum_{n=0}^{\infty}\frac{i^nt^n}{n!}E[X^n]$$

Therefore,

$$\begin{aligned}\phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n &= \left(1 + \frac{t}{\sigma\sqrt{n}}\phi'(0) + \frac{t^2}{2\sigma^2 n}\phi''(0) + \cdots\right)^n \\ &= \left(1 - \frac{t^2}{2n} + \cdots\right)^n \to e^{-\frac{t^2}{2}}\end{aligned}$$

This shows that $\frac{\sum_{m=1}^n X_m}{\sigma\sqrt{n}} \to N(0,1)$ in distribution. Now we need to show the error bound is small, by proving the following inequality:

$$\left|e^{ix} - \sum_{m=0}^n \frac{(ix)^m}{m!}\right| \leq \min\left(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right)$$

Note that for a C^{n+1} function f, the integral form of the Taylor remainder at 0 is:

$$R_n(x) = f(x) - \sum_{m=0}^n \frac{f^{(m)}(0)}{m!}x^m = \frac{1}{n!}\int_0^x (x-s)^n f^{(n+1)}(s)ds$$

For $f(x) = e^{ix}$, $f^{(n+1)}(x) = i^{n+1}e^{ix}$. By recursive IBP,

$$\begin{aligned}\int_0^x (x-s)^n e^{is}ds &= \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n-1}e^{is}ds \\ e^{ix} - \sum_{m=0}^n \frac{(ix)^m}{m!} &= \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds\end{aligned}$$

Therefore,

$$\left|e^{ix} - \sum_{m=1}^n \frac{(ix)^m}{m!}\right| \leq \frac{1}{n!}\left|\int_0^x (x-s)^n e^{is}ds\right| \leq \frac{1}{n!}\int_0^x (x-s)^n ds = \frac{x^{n+1}}{(n+1)!}$$

Similarly,

$$\frac{i}{n}\int_0^x (x-s)^n e^{is}ds = -\frac{x^n}{n} + \int_0^x (x-s)^{n-1}e^{is}ds$$
$$\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)ds$$

Use $|e^{is}-1| \leq 2$ to get the remainder $\leq \frac{2|x|^n}{n!}$

So the error term $\leq E\left[\min\left(\frac{|t|^3}{6n^{3/2}\sigma^3}X_1^3, \frac{|t|^2}{\sigma^2 n}X_1^2\right)\right]$

$\phi(t) = E[e^{itX_1}] = 1 + itE[X_1] - \frac{t^2}{2}E[X^2] + R_3(t)$, where $|R_3(t)| \leq E\left[\min\left(\frac{t^3}{6}X^3, t^2X^2\right)\right]$. So we need this to be $o\left(\frac{1}{n}\right)$ if $E[X^2] < \infty$. *i.e.* $nE\left[\min\left(\frac{|t|^3}{6n^{3/2}\sigma^3}X_1^3, \frac{|t|^2}{\sigma^2 n}X_1^2\right)\right] \to 0$.

This can be achieved by splitting at $|X_1| = C$, and consider $Cn^{-1/2}|t|^3E[X_1^3]1_{|X_1|<C} + |t|^2E[X_1^2]1_{|X_1|\geq C}$. □

Example: Let $X_1, ..., X_n$ be i.i.d. $X_i \sim Ber(p)$, $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$, $E[X_i] = p$, $Var(X_i) = p(1-p)$.

Theorem 2.3 gives $\frac{X_1 + \cdots + X_n}{n} = p$.

Since X_i are i.i.d., $Var(X_1 + \cdots + X_n) = nVar(X_1)$, $Var\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n}Var(X_1)$, $Var\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right) = Var(X_1)$. Therefore, $P\left(\sqrt{n}\left(\frac{X_1+\cdots+X_n}{n} - p\right)\right) = Var(X_1)$.

Theorem 3.8 gives

$$\lim_{n\to\infty} P\left(\left|\frac{1}{\sqrt{p(1-p)}}\sqrt{n}\left(\frac{X_1 + \cdots + X_n}{n} - p\right)\right| \leq 2\right) = 0.95$$

which means that 95% of time, $X_1 + \cdots + X_n \in \left[np - 2\sqrt{np(1-p)}, np + 2\sqrt{np(1-p)}\right]$.

If we toss a fair coin 1000 times, we should expect $[468, 532]$ heads with 95% probability.

Remark 6. $2^{-n}\binom{n}{\frac{n}{2}} \sim \frac{1}{\sqrt{n}}$ by Stirling approximation $n! \sim \sqrt{2\pi n}n^ne^{-n}$

***Theorem:* 3.9:**

Let $z_1, ..., z_n, w_1, ..., w_n\mathbb{C}$, $|z_i|, |w_i| \leq C$, then

$$\left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m\right| \leq C^{n-1}\sum_{m-1}^n |z_m - w_m|$$

Proof. Proof by induction. Base case is $n = 1$. In the induction step, we consider the sum from 2 to n, which has $n-1$ terms:

$$\begin{aligned}
\left|\prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m\right| &= \left|z_1 \prod_{m=2}^{n} z_m - z_1 \prod_{m=2}^{n} w_m + (z_1 - w_1) \prod_{m=2}^{n} w_m\right| \\
&\leq |z_1| \left|\prod_{m=2}^{n} z_m - \prod_{m=2}^{n} w_m\right| + |z_1 - w_1| \left|\prod_{m=2}^{n} w_m\right| \text{ by Triangle inequality} \\
&\leq C \cdot C^{n-2} \sum_{m=2}^{m} |z_m - w_m| + |z_1 - w_1| C^{n-1} \text{ by IH} \\
&= C^{n-1} \sum_{m=1}^{n} |z_m - w_m|
\end{aligned}$$

□

Theorem: 3.10: Lindeberg-Feller Central Limit Theorem

Consider triangular array of r.v.s $X_{11}, X_{21}, X_{22}, \ldots$, $E[X_{n,m}] = 0$ independent in each row s.t.

1. $\sum_{m=1}^{n} E[X_{n,m}^2] \to \sigma^2 > 0$
2. $\sum_{m=1}^{n} E[X_{n,m}^2 1_{|X_{n,m}| > \epsilon}] \to 0$ for all $\epsilon > 0$ (No single r.v. has dominating contribution)

Then $X_{n,1} + \cdots + X_{n,n} \to N(0, \sigma^2)$ in distribution.

Proof. Consider the characteristic functions $\phi_{n,m}(t) = E[e^{itX_{n,m}}]$. We want to show $\prod_{m=1}^{n} \phi_{n,m}(t) \to e^{-\frac{\sigma^2 t^2}{2}}$. From the proof of Theorem 3.8, we have

$$\begin{aligned}
\left|\phi_{n,m}(t) - \left(1 - \frac{t^2 \sigma_{n,m}^2}{2}\right)\right| &\leq E\left[\min\left(\frac{|tX_{n,m}|^2}{6}, \frac{2|tX_{n,m}|^2}{2}\right)\right] \\
&\leq E\left[|tX_{n,m}|^3 1_{|X_{n,m}| \leq \epsilon}\right] + E\left[|tX_{n,m}|^2 1_{|X_{n,m}| > \epsilon}\right] \\
&\leq \epsilon t^3 E\left[X_{n,m}^2\right] + t^2 E\left[X_{n,m}^2 1_{|X_{n,m}| > \epsilon}\right]
\end{aligned}$$

Using the two properties, we get

$$\limsup_{n\to\infty} \sum_{m=1}^{n} \left|\phi_{n,m}(t) - \left(1 - \frac{t^2 \sigma_{n,m}^2}{2}\right)\right| \leq \epsilon \sigma^2 t^3 \to 0$$

Since $\sum_{m=1}^{n} \sigma_{n,m}^2 \to \sigma^2$, for $n > N_0$, we have $\sigma_{n,m}^2 = E[X_{n,m}^2] \leq \epsilon^2 + E[X_{n,m}^2 1_{|X_{n,m}| > \epsilon}]$ (split the expectation into 2 parts).

Then $|\phi_{n,m}(t)| \leq 1$, $\left|1 - \frac{t^2 \sigma_{n,m}^2}{2}\right| \leq 1$. Therefore, by Theorem 3.9,

$$\limsup_{n\to\infty} \left|\prod_{m=1}^{n} \phi_{n,m}(t) - \prod_{m=1}^{n} \left(1 - \frac{t^2 \sigma_{n,m}^2}{2}\right)\right| = 0$$

Similarly, we can show that

$$\left|\prod_{m=1}^{n} e^{-t^2 \sigma_{n,m}^2/2} - \prod_{m=1}^{n} \left(1 - \frac{t^2 \sigma_{m,n}^2}{2}\right)\right| \leq \sum_{m=1}^{n} \left|e^{-t^2 \sigma_{n,m}^2/2} - \left(1 - \frac{t^2 \sigma_{m,n}^2}{2}\right)\right| \to 0$$

□

Example: Let $X_1, ..., X_n$ be i.i.d. symmetric distribution $P(X_i > x) \sim \frac{1}{x^2}$, $E[X_i] = 0$, $E[X_i^2] = \infty$.

Define $Y_{n,m} = X_m 1_{|X_m| \le n^{1/2} \log\log n}$.

$$\begin{aligned}
E[X_{n,m}^2] &= \int_0^{n^{1/2}\log\log n} 2yP(|X_i| > y)dy \\
&\sim 2\int_0^{n^{1/2}\log\log n} \frac{1}{y}dy + \text{const} \\
&= \text{const} + 2\log(n^{1/2}\log\log n) \\
&\sim \text{const} + \log n
\end{aligned}$$

Define $X_{n,m} = \frac{Y_{n,m}}{\sqrt{n\log n}}$, $nE[X_{n,m}^2] \to 1$, since $E[X_{n,m}^2] \to \log n$.

Since $|X_{n,m}| > \epsilon$ is equivalent to $Y_{n,m} > \epsilon\sqrt{n\log n}$, but $\epsilon\sqrt{n\log n} \ge n^{1/2}\log\log n$ for large n, then $nE[X_{n,m}^2 1_{|X_{n,m}|>\epsilon}] = 0$ for large n.

By Theorem 3.10, $X_{n,1} + \cdots + X_{n,n} \to N(0,1)$ in distribution., equivalently $\frac{Y_{n,1}+\cdots+Y_{n,n}}{\sqrt{n\log n}} \to N(0,1)$ in distribution.

Then, $P\left(X_{n,m} \neq X_m\right) = P\left(|X_m| > n^{1/2}\log\log n\right) \sim \frac{1}{n(\log\log n)^2}$, which implies

$$P\left(\cup\left\{X_m \neq X_{n,m}\right\}\right) \le \sum_{m=1}^{n} P(X_m \neq Y_{n,m}) \le \frac{1}{(\log\log n)^2} \to 0$$

So $\frac{X_1+\cdots+X_n}{\sqrt{n\log n}} \to N(0,1)$ in distribution.

Consider more generally, $P(|X| \ge x) \sim x^{-\alpha}$

1. $\alpha > 2$: Regular Theorem 3.8
2. $\alpha = 2$: Example above
3. $0 < \alpha < 2$: Both Theorems will fail.

Consider the characteristic function $P(|X| \ge x) = \frac{x^{-\alpha}}{2}$.

The density function is $f(x) = \begin{cases} \alpha\frac{|x|^{-(\alpha+1)}}{2}, |x| > 1 \\ 0, \text{ else} \end{cases}$.

$$\begin{aligned}
\phi(t) &= \alpha\int_1^\infty e^{itx}\frac{dx}{2x^{\alpha+1}} + \alpha\int_{-\infty}^{-1} e^{itx}\frac{dx}{2|x|^{\alpha+1}} \\
&= \alpha\int_1^\infty \cos tx\frac{dx}{x^{\alpha+1}} \\
1-\phi(t) &= \alpha\int_1^\infty (1-\cos tx)\frac{dx}{x^{\alpha+1}} \\
&\text{Let } x = \frac{y}{t},\ t > 0 \\
&= \alpha t^\alpha\int_t^\infty (1-\cos y)\frac{dy}{y^{\alpha+1}} \\
&\to C\alpha t^\alpha \text{ as } t \to 0,
\end{aligned}$$

since the integral converges to $C < \infty$.

$$E\left[e^{it\frac{X_1+\cdots+X_n}{b_n}}\right] = \left(E\left[e^{itX_1/b_n}\right]\right)^n = \phi^n\left(\frac{t}{b_n}\right)$$
$$= \left(1 - \left(1 - \phi\left(\frac{t}{b_n}\right)\right)\right)^n$$
$$\sim \left(1 - C\frac{|t|^\alpha}{b_n^\alpha}\right)^n$$

If $b_n = e^{e^{e^n}}$, we get $E\left[e^{it\frac{X_1+\cdots+X_n}{b_n}}\right] = 1$ and $\frac{X_1+\cdots+X_n}{b_n} \to 0$, so we can have a characteristic function by properly choosing b_n.

If $b_n = n^{1/\alpha}$, $b_n^\alpha = n$, $E \sim \left(1 - C\frac{|t|^\alpha}{n}\right)^n$. The characteristic function is $\psi(t) = e^{-C|t|^\alpha}$, which is continuous at 0. This means that $\frac{X_1+\cdots+X_n}{n^{1/\alpha}} \to \psi'$, which has characteristic function $\psi(t)$. (If $\alpha = 1$, Cauchy; If $\alpha = 2$, Gaussian)

***Theorem:* 3.11: Poisson Limit Theorem**

Let $X_{11}, X_{21}, X_{12}, ...$ be triangle array of r.v.s independent in each row, $P(X_{n,m} = 1) = p_{n,m} = 1 - P(X_{n,m} = 0)$, where $p_{n,m} \in (0, 1)$. If the following holds

1. $\sum_{m=1}^n p_{n,m} \to \lambda \in (0, \infty)$
2. $\max_{m\in[1,n]} p_{n,m} \to 0$

then $X_{n,1} + \cdots + X_{n,n} \to Poisson(\lambda)$ in distribution.

Note the r.v.s are rarely 1. If we let $Y_{n,m} = X_{n,m} - p_{n,m}$ and check the conditions for Theorem 3.10:

- $\sum_{m=1}^n E[Y_{n,m}^2] = \sum_{m=1}^n p_{n,m} - p_{n,m}^2 \to \lambda$
- $\sum_{m=1}^n E[Y_{n,m}^2 1_{|Y_{n,m}|\geq\epsilon}] \not\to 0$.

Proof.

$$E\left[e^{it(X_{n,1}+\cdots+X_{n,n})}\right] = \prod_{m=1}^n (1 - p_{n,m} + p_{n,m}e^{it})$$

When $p_{n,m} = \frac{\lambda}{n}$, we get $(1 + \frac{\lambda}{n}(e^{it} - 1))^n$.

Since $\prod_{m=1}^n e^{p_{n,m}(e^{it}-1)} \to e^{\lambda(e^{it}-1)}$, we just need to show the error between products is small. By Theorem 3.9:

$$\left|\prod_{m=1}^n (1 - p_{n,m} + p_{n,m}e^{it}) - \prod_{m=1}^n e^{p_{n,m}(e^{it}-1)}\right| \leq \sum_{m=1}^n \left|1 - p_{n,m} + p_{n,m}e^{it} - e^{p_{n,m}(e^{it}-1)}\right|$$
$$\leq \sum_{m=1}^n \left|\frac{p_{n,m}^2|e^{it} - 1|^2}{2}\right|$$
$$\leq 2\sum_{m=1}^n p_{n,m}^2 \to 0$$

□

Example: Number of radioactive decays in $[0, 1]$ is Poisson

Proof. Let $X_{n,m}$ =number in $[\frac{m-1}{n}, \frac{m}{n})$, $X_{n,m}$ is i.i.d. if decay is slow.

If n is large enough, $X_{n,m}$ is 0 or 1, $P(X_{n,m} = 1) \sim \frac{\lambda}{n}$.

By Theorem 3.11, $X_{n,1} + \cdots + X_{n,n} \to Poisson(\lambda)$ in distribution. $\square$

Example: 10^6 people, typically 1 fatal accident/year, one year it happens to be 3 accidents

Proof. $p = 10^{-6}$, $n = 10^6$, $X_1, ..., X_n$ has $\sigma^2 = p(1-p) \sim p = 10^{-6}$, $\sigma = 10^{-3}$, $\sqrt{n} = 10^3$. mean $m = 1$.

By Theorem 3.8, $X_1 + \cdots + X_n \in [1 - 2\sqrt{n}\sigma, 1 + 2\sqrt{n}\sigma] = [-1, 3]$ 95% of time

Using Theorem 3.11, $\lambda = 1$, $P(\geq 3) = 1 - P(0, 1, 2) = 1 - (e^{-1} + e^{-1} + 0.5e^{-1}) \approx 0.08$ instead of 0.05. 3 accidents is more likely to happen than CLT approximation.

The issue with CLT is that σ is not fixed and is dependent on n. $\square$

What is the error in CLT?
Let F_1, F_2 be distribution functions, ϕ_1, ϕ_2 be characteristic functions. Assume that ϕ_1, ϕ_2 are integrable, and $E[|X_i|] < \infty$. Then F_1, F_2 have densities f_1, f_2. By Theorem 3.4:

$$\begin{aligned}
f_1(x) - f_2(x) &= \frac{1}{2\pi}\int e^{-itx}(\phi_1(t) - \phi_2(t))dt \\
F_1(x) - F_2(x) &= \frac{1}{2\pi}\int_{-\infty}^{x}\int e^{-ity}(\phi_1(t) - \phi_2(t))dtdy \\
F_1(x) - F_2(x) - (F_1(-L) - F_2(-L)) &= \frac{1}{2\pi}\int\int_{-L}^{x} e^{-ity}dy(\phi_1(t) - \phi_2(t))dt \\
&= \frac{1}{2\pi}\int \frac{e^{-itx} - e^{-itL}}{-it}(\phi_1(t) - \phi_2(t))dt
\end{aligned}$$

By Riemann-Lebesgue Lemma, $\int e^{-itL}\frac{\phi_1(t) - \phi_2(t)}{it}dt \to 0$ as $L \to \infty$, because $\frac{\phi_1(t) - \phi_2(t)}{t}$ is integrable. Note that $\phi(t) \sim 1 + itE[X]$ around $t = 0$. Then we get

$$\begin{aligned}
F_1(x) - F_2(x) &\to \frac{1}{2\pi}\int \frac{e^{-itx}}{-it}(\phi_1(t) - \phi_2(t))dt \\
|F_1(x) - F_2(x)| &\leq \frac{1}{2\pi}\int |\phi_1(t) - \phi_2(t)|\frac{dt}{|t|} \text{ by Triangle Inequality}
\end{aligned}$$

Assume $X_1, X_2, ...$ i.i.d. mean zero, variance one, if $|t| < \sqrt{n}$.

$$\phi(t) = 1 - \frac{t^2}{2} + \frac{(it)^3}{6}E[X^3] + \cdots$$

$$\left|\phi\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right)\right| \leq \frac{|t|^3}{6n^{3/2}}E[|X|^3]$$

$$\left|\phi\left(\frac{t}{\sqrt{n}}\right) - e^{-\frac{t^2}{2}}\right| = \left|\phi\left(\frac{t}{\sqrt{n}}\right) - \left(e^{-t^2/(2n)}\right)^n\right| \sim \left|\phi\left(\frac{t}{\sqrt{n}}\right)^n - \left(1 - \frac{t^2}{2n}\right)^n\right|$$

$$\text{By Theorem 3.9, } \leq C^n n|t^3|\frac{E[|X|^3]}{6n^{3/2}}$$

Substitute into the $|F_1(x) - F_2(x)|$ bound, cutoff by a characteristic function $\left(1 - \frac{|x|}{L}\right)_+$, which corresponds to a probability distribution $h_L(x) = \frac{1 - \cos Lx}{\pi L x^2}$

$$\sup_x |F_1(x) - F_2(x)| \leq 2\sup_x |F_1 * h_L(x) - F_2 * h_L(x)| + \frac{8\,\|F_2\|_\infty}{L}$$

Then we can bound $|F_1(x) - F_2(x)|$ by the following and take $L = \sqrt{n}$, $C = e^{-t^2/4}$, so we get the dersired bound:

$$\begin{aligned} |F_1(x) - F_2(x)| &\leq \frac{1}{2\pi} \int_{-L}^{L} |\phi_1(t) - \phi_2(t)| \frac{dt}{|t|} + \frac{8\,\|F_2\|_\infty}{L} \\ &\leq \frac{E[|X|^3]}{6\sqrt{n}} \int e^{-t^2/4} |t|^3 \frac{dt}{|t|} + \frac{8\,\|F_2\|_\infty}{L} \end{aligned}$$

Theorem: **3.12: Berry-Esseen**

Let F_n be the distribution function of $\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}$, where X_i are i.i.d. with mean zero and variance σ^2. Then

$$\sup_x \left| F_n(x) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \right| \leq C \frac{E[|X|^3]}{\sigma^3 \sqrt{n}}$$

In our example, $E[|X|^3] = p(1-p)^3 + (1-p)p^3 \sim p$, $\sigma^2 \sim p$.

$$\sup_x \left| F_n(x) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \right| \leq \frac{p}{p^{3/2}\sqrt{n}} = \frac{1}{\sqrt{np}}$$

Therefore, the actual difference between the Poisson and Gaussian for $n = 10^6, p = 10^{-3}$ is about $\frac{1}{32}$ (scaled by a non-negligible constant).

$X_1, X_2, ..., X_n$ i.i.d. Bernoulli with p, then $X_1 + X_2 + \cdots + X_n \sim Poisson(1)$.
$X_1, X_2, ..., X_n$ i.i.d. Poisson with $\lambda = 1$, with $m = \sigma = 1$, then $X_1 + X_2 + \cdots + X_n \sim Poisson(n)$.

$$P\left(\frac{Poisson(n) - n}{\sqrt{n}} \leq x \right) = P\left(Poisson(n) \leq \sqrt{n}x + n \right) = \sum_{k=0}^{\sqrt{n}x+n} \frac{n^k}{k!} e^{-n}$$

3.4 Poisson Process

Definition: **3.3: Total Variation**

If μ is a signed measure on $(\Omega, \mathcal{F})$, then its total variation is $\|\mu\|_{TV} = \sup_{A\in\mathcal{F}} |\mu(A)|$.
If μ is a probability measure, then $\|\mu\|_{TV} = 1$.

Hahn decomposition gives $\mu = \mu_+ - \mu_- = \mu|_A - \mu_{A^C}$, $\|\mu\|_{TV} = \mu_+(\Omega) + \mu_-(\Omega)$. Equivalently

$$\|\mu\|_{TV} = \sup_{A_1 \cup \cdots \cup A_n = \Omega} \sum_{i=1}^{n} |\mu(A_i)| = \sup \sum |F(x_{i+1} - F(x_i))| = \|F\|_{TV}$$

We are interested in defining the distance between two probability measures $d(\mu, \nu) = \|\mu - \nu\|_{TV}$.

If $\Omega = \{0, 1, ...\}$, $\sum |\mu(x) - \nu(x)| \geq |\mu(A) - \nu(A)| + \left|\mu(A^C) - \nu(A^C)\right|$. If $A = \{x : \mu(A) \geq \nu(A)\}$, then we have $\sum |\mu(x) - \nu(x)| = (\mu(A) - \nu(A)) - ((1 - \mu(A)) - (1 - \nu(A))) = 2(\mu(A) - \nu(A))$. Therefore,

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x=0}^{\infty} |\mu(x) - \nu(x)| = \sup_A |\mu(A) - \nu(A)|$$

We now show $d(\mu, \nu) = \|\mu - \nu\|_{TV}$ is a metric

1. $d(\mu,\nu) = d(\nu,\mu)$
2. $d(\mu,\nu) \geq 0$, $d(\mu,\nu) = 0$ if and only if $\mu = \nu$
3. $d(\mu,\nu) + d(\nu,\rho) \geq d(\mu,\rho)$, since $|\mu(x) - \rho(x)| \leq |\mu(x) - \nu(x)| + |\nu(x) - \rho(x)|$ and sum up.

If $\|\mu_n - \mu\|_{TV} \to 0$, then $\mu_n(x) \to \mu(x)$ for all x, $\mu_n \to \mu$ weakly.

Example: Let $X_i \sim Ber\left(\frac{1}{2}\right)$ i.i.d. $\mu_n \sim \frac{X_1 + \cdots + X_n - \frac{n}{2}}{\sqrt{n}}$. Let μ be a Gaussian, $\|\mu_n - \mu\|_{TV} = 2$.

Example: Product measures satisfy $\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|_{TV} \leq \|\mu_1 - \nu_1\|_{TV} + \|\mu_2 - \nu_2\|_{TV}$

$$\begin{aligned}
\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|_{TV} &= \frac{1}{2}\sum_{x,y} |\mu_1(x)\mu_2(y) - \nu_1(x)\nu_2(y)| \\
&\leq \frac{1}{2}\sum_{x,y} \mu_1(x)\,|\mu_2(y) - \nu_2(y)| + \nu_2(y)\,|\mu_1(x) - \nu_1(x)| \\
&= \frac{1}{2}\sum_{y} |\mu_2(y) - \nu_2(y)| + \frac{1}{2}\sum_{x} |\mu_1(x) - \nu_1(x)| \\
&= \|\mu_1 - \nu_1\|_{TV} + \|\mu_2 - \nu_2\|_{TV}
\end{aligned}$$

Example: Convolutions satisfy $\|\mu_1 * \mu_2 - \nu_1 * \nu_2\|_{TV} \leq \|\mu_1 - \nu_1\|_{TV} + \|\mu_2 - \nu_2\|_{TV}$

$$\begin{aligned}
\|\mu_1 * \mu_2 - \nu_1 * \nu_2\|_{TV} &= \frac{1}{2}\sum_{x} \left|\sum_{y} \mu_1(x-y)\mu_2(y) - \sum_{y} \nu_1(x-y)\nu_2(y)\right| \\
&\leq \frac{1}{2}\sum_{x,y} |\mu_1(x-y)\mu_2(y) - \nu_1(x-y)\nu_2(y)| \\
&= \frac{1}{2}\sum_{x,y} |\mu_1(x)\mu_2(y) - \nu_1(x)\nu_2(y)|
\end{aligned}$$

Consider $\mu \sim Ber\left(\frac{1}{2}\right)$, $\nu \sim Poisson(p)$, $\nu(n) = \frac{p^n}{n!}e^{-p}$, $n = 0, 1, 2, ...$

$$\begin{aligned}
2\,\|\mu - \nu\|_{TV} &= |\mu(0) - \nu(0)| + |\mu(1) - \nu(1)| + \sum_{n=2}^{\infty} |\mu(n) - \nu(n)| \\
&= \left|1 - p - e^{-p}\right| + \left|p - pe^{-p}\right| + 1 - e^{-p} - pe^{-p} \\
&= e^{-p} - 1 + p + p - pe^{-p} + 1 - e^{-p} - pe^{-p} \\
&= 2p(1 - e^{-p}) = 2p\left(p - \frac{p^2}{2} + \cdots\right) \\
&\leq 2p^2
\end{aligned}$$

Another proof for Theorem 3.11

Proof. Let $\mu_{n,m} \sim X_{n,m}$, $\nu_{n,m} \sim Poisson(p_{n,m})$.

$$Poisson\left(\sum_{m=1}^{n} p_{n,m}\right) = \nu_{n,1} * \cdots * \nu_{n,n} \to Poisson(\lambda)$$

Similarly, $X_{n,1} + \cdots X_{n,n} \sum \mu_{n,1} * \cdots * \nu_{n,n}$.

$$\begin{aligned}\|\mu_{n,1} * \cdots * \mu_{n,n} - \nu_{n,1} * \cdots * \nu_{n,n}\| &\leq \sum_{m=1}^{n} \|\mu_{n,m} - \nu_{n,m}\|_{TV} \\ &\leq 2\sum_{m=1}^{n} p_{n,m}^2 \leq 2\max p_{n,m} \sum_{m=1}^{n} p_{n,m} \to 0\end{aligned}$$

□

Theorem: 3.13: Generalized Poisson Limit Theorem

Let $X_{n,m}$, $m = 1, ..., n$ be independent with values in $\{0, 1, ...\}$ s.t. $P(X_{n,m} = 1) = p_{n,m}$, $P(X_{n,m} \geq 2) = \epsilon_{n,m}$. It satisfies:

1. $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$
2. $\max_{1\leq m\leq n} p_{n,m} \to 0$
3. $\sum_{m=1}^{n} \epsilon_{n,m} \to 0$

Then $X_{n,1} + \cdots + X_{n,n} \to Poisson(\lambda)$ in total variation

Proof. Rename $\tilde{X}_{n,m} = \begin{cases} X_{n,m}, X_{n,m} = 1 \\ 0, \text{else} \end{cases}$ Then

$$\begin{aligned}P(X_{n,1} + \cdots + X_{n,n} \neq \tilde{X}_{n,1} + \cdots + \tilde{X}_{n,n}) &\leq P(\text{some } X_{n,m} \neq \tilde{X}_{n,m}) \\ &\leq \sum_{m=1}^{n} P(X_{n,m} \neq \tilde{X}_{n,m}) = \sum_{m=1}^{n} \epsilon_{m,n} \to 0\end{aligned}$$

□

Let $N(s,t) = \#$ decays in $(s,t]$ s.t.

1. $N(s,t), N(s',t')$ are independent if $(s,t] \cap (s',t'] = \emptyset$
2. distribution of $N(s,t)$ only depends on $t - s$
3. $P(N(0,h) = 1) = \lambda h + o(h)$
4. $P(N(0,h) \geq 2) = o(h)$

Then $N(0,t) \sim Poisson(\lambda t)$

Definition: 3.4: Poisson Process

Let $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, $N_t = N(t) = N(t, \omega)$ is a Poisson process with rate λ if

1. $N_{t_k} - N_{t_{k-1}}$ are independent
2. $N_t - N_s \sim Poisson(\lambda(s - t))$

Let $\xi_1, \xi_2, ...$ be independent random variables s.t. $T_n = \xi_1 + \cdots + \xi_n$, then $P(\xi_t > t) = e^{-\lambda t}$ for $t \geq 0$. Note that ξ is memoryless: $P(\xi > t + s | \xi > t) = P(\xi > s)$. $N_t = \sup\{n : T_n \leq t\}$, where $T_0 = 0$, is a Poisson process.

Recall that $f_{T_n}(s) = \frac{\lambda^n s^{n-1}}{(n-1)!}e^{-\lambda s}$. Then, $P(N_t = 0) = P(T_1 > t) = e^{-\lambda t}$. For $n \geq 1$,

$$P(N_t = n) = P(T_n \leq t < T_{n+1}) = \int_0^t P(T_n = s)P(\xi_{n+1} > t - s)ds$$
$$= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!}e^{-\lambda s}e^{-\lambda(t-s)}ds = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

Therefore, N_t has a Poisson distribution with mean λt.

To check $N_{t_{k+1}} - N_{t_k}$ are independent, we can compute $P(T_{n+1} \geq u | N_t = n) = \frac{P(T_{n+1} > u, T_n \leq t)}{P(N_t = n)}$.

Let $T_1' = T_{N(t)+1} - t$, $T_k' = T_{N(t)+k} - T_{N(t)+k-1}$. $T_1', T_2', ...$ are i.i.d. and independent of N_t. If $0 = t_0 < t_1 < \cdots < t_n$, then $N(t_i) - N(t_{i-1})$ are independent.

Let $Y_1, Y_2, ...$ be i.i.d. r.v.s, N be an independent nonnegative integer valued random variable, and $S = Y_1 + \cdots + Y_N$ with $S = 0$ if $N = 0$. Then

1. If $E[|Y_i|] < \infty$ and $E[N] < \infty$, then $E[S] = E[N]E[Y]$
2. If $E[Y_1^2] < \infty$ and $E[N^2] < \infty$, then $Var(S) = E[N]Var(Y_1) + Var(N)E[Y_1^2]$
3. If $N \sim Poisson(\lambda)$, then $Var(S) = \lambda E[Y_1^2]$.

Let $N_j(t)$ be the number of $i \leq N(t)$ with $Y_i = j$. Then $N_j(t)$ are independent Poisson process with rate $\lambda P(Y_i = j)$.

Poisson Point Processes: Consider a measure space $(S, \mathcal{S}, \mu)$, where S is a set, $\mathcal{S}$ a σ-algebra, μ a σ-finite measure. Let $m : \mathcal{S} \to \{0, 1, ...\}$ be a random integer-valued measure. For $A_1, ..., A_n \in \mathcal{S}$, $m(A_1), ..., m(A_n)$ are independent Poisson with rate $\mu(A_i)$.

4 Martingales

4.1 Conditional Expectation

Consider X a random variable in $L^1(\Omega, \mathcal{F}, P)$, and a sub σ-algebra $\mathcal{G} \subset \mathcal{F}$. Note that X is not a random variable on $(\Omega, \mathcal{G}, P)$. We define the *conditional expectation* to be the random variable $E[X|\mathcal{G}] \in L^1(\Omega, \mathcal{G}, P)$ s.t.

$$\int_G E[X|\mathcal{G}]dP = \int_G XdP \; \forall G \in \mathcal{G}$$

> ***Theorem:* 4.1: Radon-Nikodym**
>
> Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a $\mathcal{F}$-measurable function f s.t. for all $A \in \mathcal{F}$, $\int_A f d\mu = \nu(A)$. f is denoted $d\nu/d\mu$ and called the Radon-Nikodym derivative.

Existence:

Proof. Suppose $X \geq 0$. Define $Q(A) = \int_A XdP$, for $Q \ll P$.

By Theorem 4.1, there is $\frac{dQ}{dP} = E[X|\mathcal{G}]$ s.t. $\int_G E[X|\mathcal{G}]dP = \int_G XdP$ for all $G \in \mathcal{G}$.

For general X, do $X = X_+ - X_-$, □

Uniqueness:

Proof. Suppose Y, Y' both satisfy the definition. Then $\int_A Y - Y'dP = 0$, for all $A \in \mathcal{G}$. If $A = \{Y' \geq Y\}$, then $Y' \leq Y$. If $A' = \{Y' \leq Y\}$, then $Y \geq Y'$, so we must have $Y = Y'$. □

Recall conditional probabilities $P(A|B) = \frac{P(A\cap B)}{P(B)}$.
We can define $P(A|\mathcal{G}) = E[1_A|\mathcal{G}]$, where $\mathcal{G} = \{\emptyset, B, B^C, \Omega\}$.

$$\int_B E[1_A|\mathcal{G}]dP = \int_B 1_A dP = P(A \cap B)$$
$$\int_{B^C} E[1_A|\mathcal{G}]dP = P(A \cap B^C)$$

Example: $\Omega = (0, 1)$. Consider the σ-field generated by $\mathcal{F}_i = \left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)$.

$$E[f|\mathcal{F}_i] = \sum \left(2^n \int_{i/2^n}^{(i+1)/2^n} f(\tilde{x})d\tilde{x}\right) 1_{\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)}(x)$$

Example: If we have a 2D strip $\mathcal{G} \times [0, 1]$, then $E[f|\mathcal{G}] = \int_0^1 f(x, y)dy$.

Example: If X is independent of $\mathcal{G}$, *i.e.* $P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \forall B \in \mathcal{G}$, then $E[X|\mathcal{G}] = E[X]$.

Proposition: 4.1: Properties of Conditional Expectation

1. If $\mathcal{G} \subset \mathcal{F}$, $Y \in \mathcal{G}$, and $E[XY] < \infty$, then $E[XY|\mathcal{G}] = YE[X|\mathcal{G}]$, *i.e.* $\int_A YE[X|\mathcal{G}]dP = \int_A XYdP$
2. If $\mathcal{G} = \{\emptyset, \Omega\}$ is trivial or X is independent of $\mathcal{G}$, then $E[X|\mathcal{G}] = E[X]$
3. If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$
4. If $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, then $E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1]$
5. If $X \in L^2(\Omega, \mathcal{F}, P)$, then $E[X|\mathcal{G}]$ is the orthogonal projection onto $L^2(\Omega, \mathcal{G}, P)$
6. If ϕ is convex, then $\phi(E[X|\mathcal{G}]) \leq E[\phi(X)|\mathcal{G}]$. (Jensen's inequality)

Proof. 1. Check for $Y = 1_B$, where $B \in \mathcal{G}$,

$$\int_A E[YX|\mathcal{G}]dP = \int_A YXdP$$
$$\int_A YE[X|\mathcal{G}]dP = \int_{A\cap B} E[X|\mathcal{G}]dP = \int_{A\cap B} XdP = \int_A YXdP$$

2. Check $X = 1_A$, X independent of $\mathcal{G}$ means that for all $B \in \mathcal{G}$, $P(A \cap B) = P(A)P(B)$.

$$\int_B E[X]dP = P(A)P(B) = P(A \cap B) = \int_B XdP$$

□

4.2 Martingales, Almost Sure Convergence

Definition: 4.1: Martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space. $\mathcal{F}_t, t \geq 0$ is a filtration if σ-fields $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $s \leq t$. Let $X_1, X_2, ...$ be r.v.s s.t. $X_n \in \mathcal{F}_n$ (adapted). If $E[|X_n|] < \infty$, and $E[X_{n+1}|\mathcal{F}_n] = X_n$ for all n, then X_n is a Martingale.

Example: Let $\xi_1, \xi_2, ...$ be i.i.d. with mean 0, $S_n = \xi_1 + \cdots + \xi_n$, $\mathcal{F}_n = \sigma(\xi_1, ..., \xi_n)$, then

$$E[S_{n+1}|\mathcal{F}_n] = E[S_n + \xi_{n+1}|\mathcal{F}_n] = S_n + E[\xi_{n+1}|\mathcal{F}_n] = S_n$$

Example: Now consider $S_n^2 - n\sigma^2$ with $E[\xi_i^2] = \sigma^2 < \infty$,

$$\begin{aligned} E[S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n] &= E[S_n^2 + n\sigma^2 + \xi_{n+1}^2 + 2\xi_{n+1}S_n - \sigma^2|calF_n] \\ &= S_n^2 - n\sigma^2 + E[\xi_{n+1}^2|\mathcal{F}_n] + 2S_nE[\xi_{n+1}\mathcal{F}_n] - \sigma^2 \\ &= S_n^2 - n\sigma^2 + \sigma^2 + 0 - \sigma^2 \\ &= S_n^2 - n\sigma^2 \end{aligned}$$

Example (moment generating functions): Let $M(\lambda) = E[e^{\lambda\xi_1}] < \infty$. Consider $e^{\lambda S_n - n\log M(\lambda)}$

$$\begin{aligned} E\left[e^{\lambda S_{n+1} - (n+1)\log M(\lambda)}|\mathcal{F}_n\right] &= E\left[e^{\lambda\lambda S_n + \lambda\xi_{n+1} - (n+1)\log M(\lambda)}|\mathcal{F}_n\right] \\ &= e^{\lambda S_n - (n+1)\log M(\lambda)}E\left[e^{\lambda\xi_{n+1}}|\mathcal{F}_n\right] \\ &= e^{\lambda S_n - (n+1)\log M(\lambda)}M(\lambda) \\ &= e^{\lambda S_n - n\log M(\lambda)} \end{aligned}$$

Example (characteristic functions): Let $\phi(\lambda) = E\left[e^{i\lambda\xi_1}\right]$, $e^{i\lambda S_n - n\log\phi(\lambda)}$ is also a martingale.

Note: $E\left[e^{i\lambda(S_{n+1}-S_n)}|\sigma(S_1,...,S_n)\right] = \phi(\lambda)$, so $e^{i\lambda(S_{n+1}-S_n)}$ is independent of $S_1,...,S_n$.

***Definition:* 4.2: Markov Chain**

A Markov chain on $(S,\mathcal{S})$ is a sequence of random variables X_i. $\forall B \in \mathcal{S}$, $P(X_{n+1} \in B|\sigma(X_1,...,X_n)) = P(X_{n+1} \in B|\sigma(X_n)) = P(X_n, B)$. For each $x \in S$, $P(x,B)$ is a probability measure, for each B, $P(x,B)$ is a measurable function.

If $\mathcal{F}$ is a σ-field on Ω, then w_1, w_2 can be distinguished by it if $\exists A_1, A_2 \in \mathcal{F}$ with $A_1 \cap A_2 = \emptyset$ and $w_1 \in A$, $w_2 \in A_2$. $\mathcal{F}_t$ can distinguish paths befor time t but not afterwards.

For ϕ a $\mathcal{S}$-measurable function, define $P\phi(x) = \int \phi(y)P(x,dy)$

Example: If $\phi(X_n) - \sum_{j=1}^{n-1}(1\phi)(X_j)$ is martingale,

$$E\left[\phi(X_{n+1}) - \sum_{j=1}^{n}(1\phi)(X_j)|\sigma(X_1,...,X_n)\right] = E[\phi(X_{n+1})|\sigma(X_1,...,X_n)] - \sum_{j=1}^{n}(1\phi)(X_j)$$
$$= \int \phi(y)P(X_n)dy - \sum_{j=1}^{n}(1\phi)(X_j)$$

then it is Markov chain.

If $E[|X|] < \infty$ and $\mathcal{F}_n$ is the filtration, then $X_n = E[X|\mathcal{F}_n]$ is martingale.

If $\mathcal{F}_n \uparrow \mathcal{F}$, then does $X_n \to X$?

Example: $\Omega = [0,1]$, $\mathcal{F}_n = \left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)$ divisions, $\mathcal{F}$ Borel, then it does converge.

Non-example: for general martingales, $S_n = \xi_1 + \cdots + \xi_n$, $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$. It does not converge, since $\frac{S_n}{\sqrt{n}} \to$Gaussian.

***Definition:* 4.3: Super/sub-martingale**

From Definition 4.1, if we replace $E[X_{n+1}|\mathcal{F}_n] = X_n$ to be $E[X_{n+1}|\mathcal{F}_n] \leq X_n$, then it is supermartingale. If we replace with $E[X_{n+1}|\mathcal{F}_n] \geq X_n$, then it is submartingale.

If X_n is martingale on $\mathcal{F}_n$ and ϕ is convex, then $\phi(X_n)$ is submartingale.

H_n is a *predictable sequence* if $H_n \in \mathcal{F}_{n-1}$ for $n \geq 1$. The *martingale transform* is

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

Gambler's martingale: $\xi_1, \xi_2, ...$ independent with $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$, $X_n = \xi_1 + \cdots + \xi_n$, $H_n = \begin{cases} 2H_{n-1}, \xi_1, ..., \xi_{n-1} = -1 \\ 0, \xi_{n-1} = 1 \end{cases}$ Double bet when we lose. If we lose k times and then win, the net weinnings will be 1. However, there is no system for beating an unfavorable game.

***Theorem:* 4.2:**

Let X_n, $n \geq 0$ be a supermartingale. If $H_n \geq 0$ is predictable and each H_n is bounded, then $(H \cdot X)_n$ is a supermartingale.

Proof.

$$
\begin{aligned}
E[(H\cdot X)_{n+1}|\mathcal{F}_n] &= E\left[\sum_{m=1}^{n+1} H_m(X_m - X_{m-1})|\mathcal{F}_n\right] \\
&= (H\cdot X)_n + H_{m+1}E[(X_{m+1}-X_m)|\mathcal{F}_n] \\
&= (H\cdot X)_n
\end{aligned}
$$

□

Definition: 4.4: Stopping Time

A r.v. N is a stopping time if $\{N = n\} \in \mathcal{F}_n$ for all $n < \infty$.

Example (optional stopping): If $H_n = 1_{N\geq n}$, then $\{N \geq n\} = bsN \leq n-1^c \in \mathcal{F}_{n-1}$, so H_n is predictable. Then by Theorem 4.2, $(H\cdot X)_n = X_{N\wedge n} - X_0$ is a supermartingale.

Example (random walk): the first time to get to a specific point 16 is stopping time; the last time to get to is not

Example: $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$, $X_n = \xi_1 + \cdots + \xi_n$. Let N be the first time $X_n = A$ or $X_n = -A$. $X_n^2 - n$, $X_{n\wedge N}^2 - (n\wedge N)$ are martingale, so $E[X_{n\wedge N}^2 - (n\wedge N)] = 0$. $E[N] = A^2$.

To get the distribution, use $e^{\lambda X_n - n\log M(\lambda)}$ is martingale, where $M(\lambda) = \frac{e^{\lambda}-e^{-\lambda}}{2}$. $E\left[e^{\lambda X_n - n\log M(\lambda)}\right] = 1 = \frac{e^{\lambda}-e^{-\lambda}}{2} E\left[e{-}n\log M(\lambda)\right]$

If we want to know the first time $X_n = A$. Apply the same method to get $E[X_{n\wedge N}^2] = E[N\wedge n] \to E[N]$ as $n\to\infty$. But $E[X_{n\wedge N}^2] \not\to A$.

Define instead $N =$first time $X_n = A$ or $-mA$. Then $E[N] = E[X_{Nm}^2] = A\frac{m}{m+1} + (Am)^2 + \frac{1}{m+1} \to \infty$.

Theorem: 4.3:

If N is a stopping time and X_n is a supermartingale, then $X_{N\wedge n}$ is a supermartingale.

Suppose X_n is a submartingale. Let $a < b$, $N_0 = -1$ and define $N_{2k-1} = \inf\{m > N_{2k-2} : X_m \leq a\}$ and $N_{2k} = \inf\{m > N_{2k-1} : X_m \geq b\}$. N_j are stopping times and $\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^C \in \mathcal{F}_{m-1}$, so $H_m = \begin{cases} 1, N_{2k-1} < m \leq N_{2k}, \\ 0,\ \text{else} \end{cases}$ defines a predictable sequence. $X(N_{2k-1}) \leq a$ and $X(N_{2k}) \geq b$, so between the two times, X_m crosses from below a to above b (upcrossing).

Everytime an upcrossing is completed, we make a profit of $\geq b-a$. Let $U_n = \sup\{k : N_{2k} \leq n\}$ be the number of upcrossings completed by time n.

Theorem: 4.4: Upcrossing Inequality

If X_m is a submartingale, then

$$(b-a)E[U_n] \leq E[(X_n - a)_+] - E[(X_0 - a)_+]$$

Proof. Let $Y_m = a + (X_m - a)_+$. Y_m is submartingale with the same number of upcrossing.

$(b-a)U_a \leq (H\cdot Y)_m$ and $Y_n - Y_0 = (H\cdot Y)_n + ((1-H)\cdot Y)_n$. Then $E[(1-H)\cdot Y_0] \geq 0$, $E[(H\cdot Y)_m] \leq E[Y_n - Y_0]$ □

Theorem: 4.5: Martingale Convergence Theorem

If X_n is a submartingale with $\sup_n E[(X_n)_+] < \infty$, then $X_n \to X$ a.s. Convergence may not be in L^1, but $X \in L^1$.

Proof. By Lemma 1.2,

$$\begin{aligned} E[X_+] &\le \liminf_{n\to\infty} E[(X_n)_+] < \infty \\ E[X_-] &\le \liminf_{n\to\infty} E[(X_n)_-] = \liminf_{n\to\infty} E[(X_n)_+] - E[X_n] \\ &= \liminf_{n\to\infty} E[(X_n)_+] - E[E[X_n|\mathcal{F}_0]] = \liminf_{n\to\infty} E[(X_n)_+] - E[X_0] < \infty \end{aligned}$$

□

Corollary 1. *If $X_n \ge 0$ is a supermartingale, then $X_n \to X$ a.s. and $E[X] \le E[X_0]$.*

Proof. Apply Theorem 4.5 to $Y_n = -X_n$.
The inequality follows from Lemma 1.2, $E[X] \le \liminf_{n\to\infty} E[X_n] \le E[X_0]$. □

Let $X_1, X_2, ...$ be martingale s.t. $|X_{n+1} - X_n| \le B$, $N_{-L} = \inf\{n : X_n \le -L\}$ is a *stopping time*. $X_{n\wedge N_{-L}}$ is a martingale and $X_{n\wedge N_{-L}} \ge -L - B$, so it converges a.s.

Let $L \to \infty$, then on $\{\liminf_{n\to\infty} X_n > -\infty\}$, X_n converges a.s.
By symmetry, on $\{\limsup_{n\to\infty} X_n < \infty\}$, X_n converges a.s.
Either X_n converges on some set A, or $B = \{\liminf X_n = -\infty \text{ and } \limsup X_n = \infty\}$. $\Omega = A \cup B \cup C$, where $\mu(C) = 0$.

Equivalently, let X_n be a martingale with bounded differences, *i.e.* $\exists M \in \mathbb{R}$ s.t. $|X_n - X_{n+1}| \le M$ a.s. for all n. $C = \{\lim_{n\to\infty} X_n < \infty\}$, $D = \{\lim_{n\to\infty} X_n = -\infty, \limsup_{n\to\infty} < \infty\}$. Then $P(C \cup D) = 1$.

4.3 Uniformly Integrable

Let X_n be martingales. For $m > n$, $X_n = E[X_m|\mathcal{F}_n]$. *i.e.* if $A \in \mathcal{F}_n$, then $\int_A X_m dP = \int_A X_n dP$.

If $X_n \to X$ a.s., does $\int_A X_m dP \to \int_A X dP$?

Definition: 4.5: Uniformly Integrable

Let X_n be r.v.s. X_n are uniformly integrable if $\displaystyle\lim_{L\to\infty} \sup_n \int_{|X_n|>L} |X_n| dP = 0$.

Theorem: 4.6:

Let $X^L = \begin{cases} X, |X| < L \\ 0, \text{ else} \end{cases}$. If X_ns are uniformly integrable, and $X_n \to X$ a.s., then $X \in L^1$.
Conversely, if $X_i \in L^1$, $X_n \to X$ a.s., and $X \in L^1$, then X_n are uniformly integrable.

Proof.

$$E\left[|X_n - X|\right] \le E\left[\left|X_n - X_n^L\right|\right] + E\left[\left|X_n^L - X^L\right|\right] + E\left[\left|X - X^L\right|\right]$$

By Theorem 1.6, $E\left[\left|X_n^L - X^L\right|\right] \to 0$.

By uniform integrability, $E\left[|X_n - X_n^L|\right] = \int_{|X_n|>L} |X_n|dP$, $E\left[|X - X^L|\right] = \int_{|X|>L} |X|dP$.

Note that $\int_{|X_n|>L} |X_n|dP \le 1$, so $\int |X_n|dP \le 1 + L$. Therefore, $X \in L^1$ by Lemma 1.2.

For the backwards part, note by triangle inequality and Theorem 1.1,

$$\left|\int |X_n|\,dP - \int |X|\,dP\right| \le \int ||X_n| - |X||\,dP \le \int |X_n - X|\,dP$$

Also,

$$\int_{|X_n|>L} |X_n|dP = \int |X_n|dP - \int |X_n^L|dP \to \int |X|dP - \int |X^L|dP$$

So, $\limsup_n \int_{|X_n|>L} |X_n|dP - \int_{|X|>L} |X|dP$.
Therefore, $\lim_{L\to\infty} \limsup_n \int_{|X_n|>L} |X_n|dP = \lim_{L\to\infty} \int_{|X|>L} |X|dP = 0$ ☐

Remark 7. It is also equivalent to $E[|X_n|] \to E[|X|]$ as $n \to \infty$.

Theorem: **4.7:**

Let X_n be a submartingale. Then the following are equivalent: X_n is uniformly integrable; $X_n \to X$ a.s. and in L^1; Convergence in L^1.

Theorem: **4.8: de la Valle Poussin**

There exists convex $\phi(x)$ s.t. $\phi(0) = 0$, $\frac{\phi(x)}{x} \to \infty$ and $\sup_n E[\phi(x_n)] < \infty$ if and only if X_n is uniformly integrable.

If $\phi(x) = |x|^{1+\delta}$, then

$$\int_{|X_n|>L} |X_n|dP = \int 1_{|X_n|>L}|X_n|dP \le \int \frac{|X_n|^\delta}{L^\delta}|X_n|dP = \frac{1}{L}\int |X_n|^{1+\delta}dP$$

Theorem: **4.9:**

If $X_n = E[X|\mathcal{F}_n]$, $\mathcal{F}_n$ may not necessarily be a filtration, then X_n is uniformly integrable.

Proof. Since $|E[X|\mathcal{F}_n]| \le E[|X||\mathcal{F}_n]$, we have $\int_{|X_n|\ge L} |X_n|dP \le \int_{|X_n|>L} |X|dP$. So $P(|X_n| > L) \le \frac{1}{L}E[|X_n|] \le \frac{E[|X|]}{L}$.

We just need that for each $\epsilon > 0$, there is $\delta > 0$ s.t. if $P(A) < \delta$, then $\int_A |X|dP < \epsilon$. Otherwise, there are sets A_L with $P(A_L) < \frac{1}{L}$ and $\int_{A_L} |X|dP \ge \epsilon$. ☐

Theorem: **4.10:**

If X_n is martingale, $X_n \to X$ a.s. and X_n are uniformly integrable, then $X_n \to X$ in L^1 and $X_n = E[X|\mathcal{F}_n]$.
Suppose $X \in L^1$ and $\mathcal{F}_n$ is a filtration, then $X_n = E[X|\mathcal{F}_n]$ is uniformly integrable martingale.
Therefore, $X_n \to X_\infty$ a.s. $\int_A XdP = \int_A X_\infty dP$, $\forall A \in \sigma(\cup_n \mathcal{F}_n)$, so $X_\infty = E[X|\sigma(\cup_n \mathcal{F}_n)]$

Remark 8. Note that $\sigma(\cup_n \mathcal{F}_n)$ may not be $\mathcal{F}$, $\mathcal{F}$ may contain more information; X_∞ may not be X.

Theorem: 4.11:

Let $(\mathcal{F}_n)_n$ be a filtration, $X \in L^1$. Then $E[X|\mathcal{F}_n] \to E[X|\mathcal{F}_\infty]$ a.s. and in L^1.

Proof. Let $Y_n = E[X|\mathcal{F}_n]$. It is a uniformly integrable martingale, converging to some $Y \in \mathcal{F}_\infty$.

If $A \in \mathcal{F}_n$, then

$$E[Y1_A] = \lim_{m\to\infty} E[Y_m 1_A] = E[Y_n 1_A] = E[E[X|\mathcal{F}_n]1_A] = E[X1_A]$$

so $E[Y1_A] = E[X1_A]$ for all $Y \in \cup_{n=1}^{\infty}\mathcal{F}_A$. □

Corollary 2. *If $A \in \mathcal{F}_\infty$, then $E[1_A|\mathcal{F}_n] \to 1_A$ a.s. and in L^1.*

Let $X_1, X_2, ...$ be r.v.s $T = \cap_n \sigma(X_n, X_{n+1}, ...)$ is the tail field.

Theorem: 4.12: Kolmogorov 0-1 Law

If $X_1, X_2, ...$ are independent, then T is trivial. *i.e.* if $A \in T$, then $P(A) = 0$ or 1.

Proof. Let $A \in T$, $\mathcal{F}_n = \sigma(X_1, ..., X_n)$. Then $X_n = E[1_A|\mathcal{F}_n] = E[1_A] = P(A)$.

As $n \to \infty$, $E[1_A|\mathcal{F}_n] \to E[1_A|\sigma(\cup\mathcal{F}_n)] = 1_A$. Since $P(A) = 1_A$, we must have $P(A) = 0$ or 1. □

4.4 Backwards Martingales

Let X_n be r.v.s. with $n = 0, -1, -2, ...$ The filtrations are $\cdots \subset \mathcal{F}_{-n-1} \subset \mathcal{F}_{-n} \subset \mathcal{F}_{-n+1} \subset \cdots \subset \mathcal{F}_0$. $X_n \in \mathcal{F}_n$. X_n is backwards martingale if $E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$.

- $X_{-1} = E[X_0|\mathcal{F}_{-1}]$
- $X_{-2} = E[X_{-1}|\mathcal{F}_{-2}] = E[X_0|\mathcal{F}_{-2}]$
- $X_{-n} = E[X_0|\mathcal{F}_{-n}]$

The backwards martingales satisfy:

- X_n, $n \le 0$ are uniformly integrable
- $X_n \to X_{-\infty}$ a.s.
- $\int_A X_{-\infty} dP = \int_A X_0 dP$ for $A \in \cap\mathcal{F}_n$
- $X_{-\infty} = E[X_0| \cap \mathcal{F}_n]$

If $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$, then $E[X|\mathcal{F}_n] \to E[X|\mathcal{F}_{-\infty}]$ a.s. and in L^1.

Upcrossing Inequality in Backwards Martingale: Assuming $E[X_0] < \infty$, then $(b-a)E[U_n] \le E[(X_0 - a)_+] < \infty$, where $U_n =$#of upcrosssings of $[a, b]$ in $[-n, 0]$. Therefore, $E[U_\infty] < \infty$, $U_\infty < \infty$ a.s. $X_{-n} \to X_{-\infty}$ a.s. and in L^1.

Since $\int_A X_{-n} dP = \int_A X_0 dP$ for all $A \in \mathcal{F}_{-n}$, and $\int_A X_{-n} dP \to \int_A X_{-\infty} dP$ for $A \in \mathcal{F}_{-\infty}$,
then $\int_A X_{-\infty} dP = \int_A X dP$ for all $A \in \mathcal{F}_{-\infty}$. In particular, $X_{-\infty} = E[X_0|\mathcal{F}_{-\infty}]$.

Example (Strong Law of Large Numbers): Let $\xi_1, \xi_2, ...$ be i.i.d. mean zero s.t. $E[|\xi_i|] < \infty$, $S_n = \xi_1 + \cdots + \xi_n$, $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \cdots)$. Define $X_n = \frac{S_n}{n}$, X_n is $\mathcal{F}_{-n}$-measurable. Notice that $E[\xi_k|\mathcal{F}_{-(n+1)}]$

are equal for all $k \in \{1,2,...,n+1\}$ and $E[\xi_k|\mathcal{F}_{-(n+1)}](n+1) = S_{n+1}$.

$$
\begin{aligned}
E[X_{-n}|\mathcal{F}_{-(n+1)}] &= E\left[\frac{S_{n+1}}{n} - \frac{\xi_{n+1}}{n}|\mathcal{F}_{-(n+1)}\right] = \frac{S_{n+1}}{n} - \frac{1}{n}E[\xi_{n+1}|\mathcal{F}_{-(n+1)}] \\
&= \frac{S_{n+1}}{n} - \frac{1}{n}\frac{S_{n+1}}{n+1} \\
&= \frac{S_{n+1}}{n+1} = X_{-(n+1)}
\end{aligned}
$$

Therefore, X_n is a backwards martingale; and $\frac{S_n}{n} \to E[\xi_1|\mathcal{F}_{-\infty}]$ a.s. and in L^1. To show that $E[\xi_1|\mathcal{F}_{-\infty}] = E[X_1]$, we neeed to apply the Hewitt-Savage -1 law. Since $\mathcal{F}_{-n} \subset \mathcal{E}_n$, $\mathcal{F}_{-\infty} = \cap_{n\geq 1}\mathcal{F}_{-n} \subset \mathcal{E}_1$. X_1 is independent of $\mathcal{F}_{-\infty}$, so $E[X_1|\mathcal{F}_{-\infty}] = E[X_1]$.

Let $X_1, X_2, ...$ be r.v.s. $\mathcal{E}_n$ be the σ-field of events invariant under permutations of $X_1, ..., X_n$. It is the smallest σ-field w.r.t. which all functions $\phi(X_1, ..., X_n)$ are measurable. ϕ are symmetric under permutations. *i.e.* $\phi(X_1, ..., X_n) = \phi(X_{\pi(1)}, ..., X_{\pi(n)})$, for $\pi \in S_n$. $\mathcal{E} = \cap_n \mathcal{E}_n$ is the exchangeable σ-field.

e.g.: $\phi(x_1, ..., x_n) = x_1 x_2 \cdots x_n$ is a symmetric function.

For $k \leq n$,

$$
\begin{aligned}
E[\phi(X_1, ..., X_n)|\mathcal{E}_n] &= \frac{1}{|S_n|}\sum_{\pi \in S_n} \phi(X_{\pi(1)}, ..., X_{\pi(n)}) \\
E[\phi(X_1, ..., X_k)|\mathcal{E}_n] &= \text{Average of } \phi(X_{i_1}, ..., X_{i_k}) \text{ for } i_1, ..., i_k \text{ disjoint in } \{1, ..., n\} \\
&= \frac{1}{n(n-1)\cdots(n-(k-1))} \sum_{i_1,...,i_k \text{ disjoint in } \{1,...,n\}} \phi(X_{i_1}, ..., X_{i_k})
\end{aligned}
$$

Let $n \to \infty$, $E[\phi(X_1, ..., X_n)|\mathcal{E}_n] \to E[\phi(X_1, ..., X_n)|\mathcal{E}]$

Claim: $E[\phi(X_1, ..., X_k)|\mathcal{E}] \in \sigma(X_L, X_{L+1}, X_{L+2}, ...)$ for any L.

Proof. X_1 only appears in $k(n-1)\cdots(n-(k-1))$, then $E[\phi(X_1, ..., X_n)|\mathcal{E}] \in \sigma(X_2, X_3, ...)$
X_1, X_2 only appears in $k(k+1)(n-2)\cdots(n-(k-1))$, then $E[\phi(X_1, ..., X_n)|\mathcal{E}] \in \sigma(X_3, X_4, ...)$
Inductively, we can prove the claim. □

Therefore, $E[\phi(X_1, ..., X_k)|\mathcal{E}] \in t$, where T is the tail field. Then $\mathcal{E} \subset T$, T is trivial, so $\mathcal{E}$ is trivial. This is **Hewitt-Savage 0-1 law**. Formally, let $\mathcal{E}_n$ =events generated by $(X_i)_{i=1}^{\infty}$ that are invariant under permutations of first n-coordinates. Then $\mathcal{E} = \bigcap_{n=1}^{\infty} \mathcal{E}_n$ is trivial. $P(A) = 0$ or 1 if $A \in \mathcal{E}$.

Write $\phi(X_1, ..., X_k) = f(X_1, ..., X_{k-1})g(X_k)$ by symmetry

$$
\begin{aligned}
&\frac{n}{n-(k-1)} \sum_{i_1,...,i_{k-1}} \frac{f(X_{i_1}, ..., X_{i_{k-1}})}{n(n-1)\cdots(n-(k-2))} \sum_{i_k} \frac{g(X_{i_k})}{n} \\
&= \sum_{i_1,...,i_k} \frac{f(X_{i_1}, ..., X_{i_{k-1}})g(X_{i_k})}{n(n-1)\cdots(n-(k-1))} \\
&+ \frac{k-1}{n-(k-1)} \sum_j \sum_{i_1,...,i_{k-1}} \frac{f(X_{i_1}, ..., X_{i_{k-1}})g(X_{i_j})}{k(k-1)n(n-1)\cdots(n-(k-2))}
\end{aligned}
$$

Therefore, $E[f(X_1, ..., X_{k-1})g(X_k)|\mathcal{E}] = E[f(X_1, ..., X_{k-1})|\mathcal{E}]E[g(X_k)|\mathcal{E}]$.
Inductively, $E\left[\prod_{j=1}^k f_j(X_j)|\mathcal{E}\right] = \prod_{j=1}^k E[f_j(X_j)|\mathcal{E}]$, so given $\mathcal{E}$, $X_1, ..., X_k$ are independent and identical. This is **de Finetti's Theorem**.

Assume $X_1, X_2, ...$ r.v.s with values in $[0,1]$. We don't really know the mean. Given $\mathcal{E}$, they are i.i.d., $\mathcal{E}$ encodes the mean.

$$P(X_1, ..., X_k|\mathcal{E}) = \binom{k}{j} p^j (1-p)^{k-j}$$
$$P(X_1, ..., X_k) = \int \binom{k}{j} p^j (1-p)^{k-j} d\mu(p)$$

if we have j 1s and $k-j$ 0s, for some probability measure on $[0,1]$.

***Lemma:* 4.1:**

A backwards martingale is uniformly integrable.

Proof. If X_n is a backwards martingale, then $|X_n|$ is a backwards submartingale, so $\int_{|X_n|\geq l} |X_n| dP \leq \int_{|X_n|\geq l} |X_0| dP$.

By Markov inequality, $P(|X_n| \geq l) \leq \frac{E[|X_n|]}{\lambda} \leq \frac{E[|X_0|]}{\lambda}$.

If $X_0 \in L^1$, then $\forall \epsilon > 0$, $\exists \delta > 0$ s.t. if $P(A) < \delta$, $\int_A |X_0| dP < \epsilon$. □

***Theorem:* 4.13: Ballot Theorem**

Let ξ_i be i.i.d. non-negative integer valued r.v.s. Let $S_k = \sum_{j=1}^k \xi_j$. Let G be the event $\{S_j < j : 1 \leq j \leq n\}$. Then $P(G|S_n) = \left(1 - \frac{S_n}{n}\right)^+$.

Proof. Define a backward martingale, $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, ...) = \sigma(S_n, \xi_{n+1}, ...)$, $E[X_{-j}|\mathcal{F}_{-j-1}] = X_{-j-1}$, where $X_j = S_{-j}/j$.

Let $T = \inf\{k \geq -n, X_n \geq 1\}$, $X_T = 1$ on $G^C \cap \{S_n < n\}$. On $G \cap \{S_n < n\}$, we have $T = -1$. Since $S_1 < 1$, it must be true that $S_1 = 0$, since x_js are non-negative integers. So $X_T = 0$ on $G \cap \{S_n < n\}$.

$$P(G^C \cap \{S_n < n\} | S_n) = E[X_T 1_{\{S_n<n\}} | \mathcal{F}_{-n}] = 1_{\{S_n<n\}} E[X_T | \mathcal{F}_{-n}] = 1_{S_n<n} X_{-n} = 1_{S_n<n} \frac{S_n}{n}$$

Recall that $Y_k = X_{T\wedge k}$ is a martingale, so $X_T = Y_{-1} \Rightarrow Y_{-n} = X_{-n}$. □

Example: Consider an election with two candiates A and B. Assume that everyone casts votes with probability 50-50, and A gets α votes, B gets β votes.

Then Theorem 4.13 says the probability that A wins ahead the entire time when counting votes, given that A wins is $\frac{\alpha-\beta}{\alpha+\beta}$.

Proof. Let $S_i = \begin{cases} 2, B \text{ gets votes} \\ 0, A \text{ gets votes} \end{cases}$, $S_j = 2$ times the number of B votes. $S_j < j$ is equivalent to number of B votes is $< \frac{j}{2}$. $1 - \frac{S_n}{n} = 1 - \frac{2\beta}{\alpha+\beta} = \frac{\alpha-\beta}{\alpha+\beta}$. □

4.5 Examples

Polya's Urn Scheme: Start with an urn that has r red balls and g green balls. Draw a ball from urn uniformly, replace it back in urn and add c balls of the same color. Let X_n be the fraction of green balls in the urn after n draws. Then X_n is a martingale:

$$X_{n+1} = \begin{cases} \frac{j+c}{i+j+c} \text{ w.p. } \frac{j}{i+j} \text{ if choose green} \\ \frac{j}{i+j+c} \text{ w.p. } \frac{i}{i+j} \text{ if choose red} \end{cases}$$

$$E[X_{n+1}|\mathcal{F}_n] = \frac{j+c}{i+j+c}\frac{j}{i+j} + \frac{j}{i+j+c}\frac{i}{i+j} = \frac{j(i+j+c)}{(i+j+c)(i+j)} = \frac{j}{i+j} = X_n$$

By Theorem 4.5, $\lim_{n\to\infty} X_n = X$ exists.

Computing the probability of getting green on first m draws, then red on the next $l = n - m$ draws:

$$\frac{g}{g+r}\frac{g+c}{g+r+c}\frac{g+2c}{g+r+2c}\cdots\frac{g+(m-1)c}{g+r+(m-1)c}\frac{r}{g+r+(m-1)c}\frac{r+c}{g+r+mc}\cdots\frac{r+(l-1)c}{g+r+(n-1)c}$$

Any other sequence of draws with m green balls and l red balls has the same probability. The sequence of colors of drawn balls is exchangeable.

Consider $c = g = r = 1$. Let $G_n = (n+2)X_n$ = number of green balls in urn after n draws. G_n is a r.v. taking values between 1 and $n+2$.

$$P(G_n = m+1) = \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1},$$

G_n is uniformly distributed on $\{1, 2, ..., n+1\}$. After scaling, it is a uniform distribution over $(0, 1)$.

Branching Process: $Z_n \geq 0$, $Z_n \in \mathbb{Z}$, $Z_0 = 1$.
Let ξ be a non-negative integer-valued r.v. $\left\{\xi_i^{(n)} : i \geq 1, n \geq 0\right\}$ be a family of r.v.s with distribution ξ. Given Z_n,

$$Z_{n+1} = \begin{cases} 0, Z_n = 0 \\ \xi_1^{(n+1)} + \xi_2^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}, Z_n > 0 \end{cases}$$

$\mathcal{F}_n = \sigma(\xi_j^{(i)}, 1 \leq i \leq n, j \geq 0)$ is a filtration, $Z_n \in \mathcal{F}_n$.

Assume $\mu = E[\xi] \in (0, \infty)$, $\frac{Z_n}{\mu^n}$ is a martingale:

$$E[Z_{n+1}|\mathcal{F}_n] = \sum_{j=1}^{Z_n} E[\xi_j^{(n+1)}|\mathcal{F}_n] = \sum_{j=1}^{Z_n} E[\xi_j^{(n+1)}] = Z_n\mu$$

If $\mu < 1$, then $Z_n \to 0$ a.s.

$$P(Z_n > 0) \leq E[Z_n] = \mu^n E[Z_0] = \mu^n \to 0 \text{ a.s.}$$

by Lemma 2.1

If $\mu = 1$, then Z_n is a non-negative martingale, so $\lim_{n\to\infty} Z_n = Z_\phi$ exists.

If $P(\xi = 1) < 1$ and $\mu = 1$, then $\forall N, K > 0$, $P(Z_n = K, \forall n \geq N) = 0$, $Z_n \to 0$ a.s.

Proof. For any $M > N$

$$P(Z_n = K, \forall n \geq N) \leq P(Z_n = K, \forall N \leq n \leq M) = P(Z_M = K | Z_{M-1} = K, ..., Z_N = K)$$
$$= P(Z_M = K | Z_{M-1} = K) = P(Z_2 = K | Z_1 = K) = c \in [0, 1)$$

Then inductively, $P(Z_n = K, \forall N \leq n \leq M) \leq cP(Z_n = K, \forall N \leq n \leq M-1) \leq C^{M-N} \to 0$ as $M \to \infty$. ☐

4.6 Maximal Inequalities

Theorem: 4.14: Doob's Inequality

Let X_n be a submartingale w.r.t. $\mathcal{F}_n$, $n = 1, 2, ...$, $X_n^+ = \begin{cases} X_n, X_n \geq 0 \\ 0, X_n < 0 \end{cases}$. Then

$$P\left(\max_{1 \leq k \leq n} X_k^+ \geq \lambda\right) \leq \frac{E[X_n^+]}{\lambda}$$

Proof. Let $A_k = \left\{X_k^+ \geq \lambda, but X_1^+, ..., X_{k-1}^+ < \lambda\right\}$. $\max_{1 \leq k \leq n} X_k \geq \lambda$ is the disjoint union of $A_1, ..., A_n$.

$$\begin{aligned} P\left(\max_{1 \leq k \leq n} X_k \geq \lambda\right) &= \sum_{k=1}^{n} P(A_k) \\ &\leq \sum_{k=1}^{n} \frac{1}{\lambda} \int_{A_k} X_k^+ dP \text{ (Markov Inequality)} \\ &\leq \sum_{k=1}^{n} \frac{1}{\lambda} \int_{A_k} X_n^+ dP \text{ (Submartingale)} \\ &\leq \frac{1}{\lambda} \int X_n dP \end{aligned}$$

☐

If X_n is a submartingale, then X_n^+ is a submartingale, by Jensen's inequality on martingales.

If X_n is a martingale, then $P\left(\max_{1 \leq k \leq n} |X_k| \geq \lambda\right) \leq \frac{E[X_n^2]}{\lambda^2}$

Theorem: 4.15: L^p-convergence

Let X_n be a submartingale, $\bar{X}_n = \max_{1 \leq m \leq n} X_m^+$. If $p > 1$, then

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

If Y_n is a martingale and $Y_n^* = \max_{1 \leq j \leq n} |Y_j|$, then

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

Corollary 3. *If $p > 1$, X_n is a martingale and $\sup_n E[|X_n|^p] < \infty$, then $\lim_{n \to \infty} X_n = X$ a.s. and $X_n \to X$ in L^p.*

Proof. Holder's inequality $\Rightarrow \sup_n E[|X_n|] < \infty$. Theorem 1.5 gives $\lim_{n\to\infty} X_n = X$ a.s.

$|X - X_n| = \lim_{m\to\infty} |X_m - X_n| \le 2\sup_n |X_n| \in L^p$. By Theorem 1.6, $E[|X_n - X|^p] \to 0$ as $n \to \infty$. $\square$

There is no L^1 maximal inequality of the form $E\left[\sup_{1\le j\le n} |X_j|\right] \le CE[|X_n|]$ for all martingales X_j and n.

Example: consider the branching process Z_n with $E[S] = \mu = 1$ and $P(\xi = 1) < 1$. Z_n is a positive martingale, $\lim_{n\to\infty} Z_n = Z$ exists and $Z = 0$, $\sup_n E[|Z_n|] = 1$.

If there was a L^1-maximal inequality, then $Z_n \to Z$ in L^1.
If this is true, then $0 = E[Z] = \lim_{n\to\infty} E[Z_n] = 1$.

When $\mu > 1$, $\phi(S) = \sum_{n=0}^{\infty} S_n P(\xi = K)$, $P(Z_n = 0 \text{ for large } n) = \rho$. ρ is the solution of $\rho = \phi(\rho)$.

The process $X_n = \frac{Z_n}{\mu^n}$is a nonnegative martinale s.t. $\lim_{n\to\infty} X_n = X$ a.s. If $P(X = 0) < 1$, then $P(X = 0) = \rho$. If X is nontrivial, then $\{X = 0\} = \{Z_n \to 0\}$ a.s.

Assume $E[\xi^2] < \infty$, then

$$\begin{aligned}
E[Z_{n+1}^2|\mathcal{F}_n] &= E\left[\left(\sum_{j=1}^{Z_n} \xi_j^{(n+1)}\right)^2 |\mathcal{F}_n\right] \\
&= \sum_{i=1}^{Z_n}\sum_{j=1}^{Z_n} E[\xi_i^{(n+1)}\xi_j^{(n+1)}|\mathcal{F}_n] = \sum_{i=1}^{Z_n}\sum_{j=1}^{Z_n} E[\xi_i^{(n+1)}\xi_j^{(n+1)}] \\
&= Z_n E[\xi^2] + Z_n(Z_n - 1)\mu^2 = Z_n Var(\xi) + Z_n^2\mu^2 \\
\Rightarrow E[X_{n+1}] &= E[X_n^2] + \frac{\sigma^2}{\mu^{n+2}}
\end{aligned}$$

If $E[X_0]^2 = 1$, $E[X_1^2] = 1 + \frac{\sigma^2}{\mu}$ and $E[X_n^2] = 1 + \sigma^2 \sum_{k=2}^{n+1} \mu^{-k}$. Therefore, $X_n \to X$ in L^2 and $E[X_n] \to E[X]$, X is not trivial.

Theorem: 4.16: Doob's Decomposition

Let X_n be a submartingale, then there exists a unique decomposition $X_n = M_n + A_n$, where M_n is a martingale, and A_n is an increasing predictable process s.t. $A_n = 0$ and $A_n \in \mathcal{F}_{n-1}$.

Proof. Suppose $X_n = M_n + A_n$, then $X_{n+1} = M_{n+1} + A_{n+1}$, $X_{n+1} - X_n = (M_{n+1} - M_n) - (A_{n+1} - A_n)$. Take conditional expectation on both sides w.r.t. $\mathcal{F}_n$.

$$\begin{aligned}
E[X_{n+1} - X_n|\mathcal{F}_n] &= E[M_{n+1} - M_n|\mathcal{F}_n] + E[A_{n+1} - A_n|\mathcal{F}_n] = 0 + A_{n+1} - A_n \\
&\Rightarrow A_{n+1} = A_n + E[X_{n+1}|\mathcal{F}_n] - X_n
\end{aligned}$$

Inductively, $A_n = \sum_{j=1}^n E[X_j|\mathcal{F}_{j-1}] - X_{j-1}$. Then $M_n = X_n - A_n$ is a martingale. $\square$

Example: If X_n is a martingale (with $X_n \in L^2$ and $X_0 = 0$), then $(X_n)^2$ is a submartingale, so there exists Doob decomposition $X_n^2 = M_n + A_n$,

$$A_n = \sum_{j=1}^n E[X_j^2|\mathcal{F}_{j-1}] - X_{j-1}^2 = \sum_{j=1}^n E[(X_j - X_{j-1})^2|\mathcal{F}_{j-1}]$$

This is related to quadratic variation.

Then Theorem 4.15 says

$$E\left[\sup_{1\leq m\leq n}|X_n|^2\right]\leq 4E[X_n^2]=4E[A_n]$$

Take limits to infinity:

$$E\left[\sup_m|X_m|^2\right]=\lim_{n\to\infty}E\left[\sup_{1\leq m\leq n}|X_n|^2\right]\leq\lim_{n\to\infty}4E[A_n]=4E[A_\infty]$$

Theorem: 4.17:

Let X_n be a square integrable martingale with $X_0=0$. Then $\lim_{n\to\infty}X_n$ exists and is finite on the event that $\{A_\infty<a\}$

Proof. Fix $0<a<\infty$. Let $N=\inf_n\{A_{n+1}>a\}$. If $N=n$, $A_n\leq a$. Since A_n is predictable, N is a stopping time.

By the above inequality, $E\left[\sup_k|X_{N\wedge k}|^2\right]\leq 4E[A_{N\wedge\infty}]\leq 4a$. $X_{N\wedge k}$ is bounded in L^2, so $\lim_{k\to\infty}X_{N\wedge k}$ exists and is finite.

Therefore, on event $\{A_\infty<a\}$, $\lim_{k\to\infty}X_k$ exists and is finite. □

Theorem: 4.18:

If X_j are independent, centered and $\sum_j Var(X_j)<\infty$, then $\sum_{j=1}^\infty X_j$ exists a.s.

4.7 Martingale Central Limit Theorem

For each n, let $\left\{\mathcal{F}_i^{(n)}\right\}_{i=1}^{m(n)}$ be a filtration, random variables $\left\{\xi_i^{(n)}\right\}_{i=1}^{m(n)}$ s.t. $E\left[\left|\xi_i^{(n)}\right|\right]<\infty$, $\xi_i^{(n)}\in\mathcal{F}_i^{(n)}$ and $E\left[\xi_i^{(n)}|\mathcal{F}_{i-1}^{(n)}\right]=0$. Then $X_k^{(n)}=\sum_{j=1}^k\xi_j^{(n)}$ is a martingale w.r.t. $\left\{\mathcal{F}_i^{(n)}\right\}_i$.

Theorem: 4.19: Martingale Central Limit Theorem

Assume (1) for all n, $\sum_{j=1}^{m(n)}E\left[\left|\xi_j^{(n)}\right|^2|\mathcal{F}_{j-1}\right]=1$ and (2) $\sum_{j=1}^{m(n)}E\left|\xi_j^{(n)}\right|^3\to 0$ as $n\to\infty$. Then $X_{m(n)}^n\to\mathcal{N}(0,1)$ in distribution as $n\to\infty$.

Proof. Drop superscripts. Define $X_k=\sum_{j=1}^k\xi_k$, $\sigma_j^2=E\left[|\xi_j|^2|\mathcal{F}_{j-1}\right]$.

Let $(Z_j)_j$ be a collection or i.i.d. $\mathcal{N}(0,1)$ independent of $\mathcal{F}_{m(n)}$. Define $Y_k=\sum_{j=k+1}^{m(n)}\sigma_jZ_j$.

Consider $Y_0=\sum_{j=1}^{m(n)}\sigma_jZ_j$. Condition on $\mathcal{F}_{m(n)}$, Y_0 is a Gaussian with variance $\sum_{j=1}^{m(n)}\sigma_j$, $Y_0\sim\mathcal{N}(0,1)$. This is because for any measurable function f,

$$E[f(Y_0)]=E\left[E\left[f(Y_0)|\mathcal{F}_{m(n)}\right]\right]=E\left[E\left[f(\mathcal{N}(0,1))\right]\right]=E\left[f(\mathcal{N}(0,1))\right]$$

Let f be a smooth function with bounded characteristic functions. We want to show that as $n\to\infty$, $E[f(X_{m(n)})]-E[f(Y_0)]\to 0$.

Expand out as telescoping series:

$$E[f(X_{m(n)})] - E[f(Y_0)] = \sum_{k=1}^{m(n)} E\left[f(X_k + Y_k) - f(X_{k-1} + Y_{k-1})\right]$$

Each term in the sum can be rewritten as:

$$E\left[f(X_k + Y_k) - f(X_{k-1} + Y_{k-1})\right] = E\left[f(X_{k-1} + \xi_k + Y_k) - f(X_{k-1} + \sigma_k Z_k + Y_k)\right]$$

By Taylor expansion,

$$= E\left[f'(X_{k-1} + Y_k)(\xi_k - \sigma_k Z_k)\right] + \frac{1}{2}E\left[f''(X_{k-1} + Y_k)(\xi_k^2 - \sigma_k^2 Z_k)\right] + \mathcal{O}\left(E\left[|\xi_k|^3\right] + E\left[\sigma_k^3 |Z_k|^3\right]\right)$$

Let $V_k = \sum_{j=1}^{k} \sigma_j^2 = 1 - \sum_{j=k+1}^{m(n)} \sigma_j^2$. Consider $\hat{Y} = \frac{Y_k}{\sqrt{1-V_k}} = \frac{\sum_{j=k+1}^{n} \sigma_j Z_j}{\sum_{j=k+1}^{n} \sigma_j^2}$. Conditional on $\mathcal{F}_{m(n)}$, $\hat{Y} \sim \mathcal{N}(0,1)$. $\hat{Y}$ is independent from $\mathcal{F}_{m(n)}$.

Write $Y_k = \sqrt{1 - V_k}\hat{Y}$.

$$\begin{aligned}
&E\left[f'(X_{k-1} + Y_k)(\xi_k - \sigma_k Z_k)|\mathcal{F}_{n-1}, \hat{Y}\right] \\
&= f'(X_{k-1} + \sqrt{1 - V_k}\hat{Y})\left(E\left[\xi_k|\mathcal{F}_{n-1}, \hat{Y}\right] - \sigma_k E\left[Z_k|\mathcal{F}_{n-1}, \hat{Y}\right]\right) \\
&E[\xi_k|\mathcal{F}_{n-1}, \hat{Y}] = E[\xi_k|\mathcal{F}_{n-1}] = 0 \\
&E[Z_k|\mathcal{F}_{n-1}, \hat{Y}] = E[Z_k] = 0 \\
\Rightarrow &E\left[f'(X_{k-1} + Y_k)(\xi_k - \sigma_k Z_k)|\mathcal{F}_{n-1}, \hat{Y}\right] = 0
\end{aligned}$$

Similarly, the second order term is zero. Therefore, $E[f(X_{m(n)})] - E[f(Y_0)] \to 0$. □

Example: Simple random walk on $\mathbb{Z}$, have $p = \frac{1}{8}$ to walk left or walk right, and $p = \frac{3}{4}$ to stay at the same position. Let X_n be the current position after n steps, then $\frac{X_n}{\sqrt{n}} \to \mathcal{N}(0,1)$.

Recall the basic definition of conditional probability $P(A|\mathcal{F})$. Suppose $A = \cup_i B_i$ disjoint union. Then $1_A = \sum_{i=1}^{\infty} 1_{B_i}$.

$$P(A|\mathcal{F}) = E[1_A|\mathcal{F}] = E\left[\sum_{i=1}^{\infty} 1_{B_i}|\mathcal{F}\right] = \sum_{i=1}^{\infty} E[1_{B_i}|\mathcal{F}] = \sum_{i=1}^{\infty}[(B_i|\mathcal{F})]$$

The equalities hold almost surely. The conditional probabilities behave like a measure. However, we may want to let it operator on some point $\omega \in \Omega$, so we need a more regular definition.

***Definition:* 4.6: Regular Conditional Probabilities**

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(S, \mathcal{S})$ be a measurable space, $X : \Omega \to S$ be a measurable map (S-valued r.v.). Let $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra. Then a regular conditional distribution (r.c.d.) of X given $\mathcal{G}$ is a map $\mu : \Omega \times S \to [0,1]$ s.t.

1. $\forall A \in \mathcal{S}$, $\omega \mapsto \mu(\omega, A)$ is a version of $P(X \in A|\mathcal{G})$
2. For a.e. $\omega \in \Omega$, $A \mapsto \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.

Example: Let $X, Y \in \mathbb{R}$ with joint density $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$ s.t. $f(x,y) > 0$ for all (x,y). Then the conditional distribution of X given Y ($\mathcal{G} = \sigma(Y)$) on A is $\mu(\omega, A) = \int_A f(x, Y(\omega))dx / \int_{\mathbb{R}} f(x, Y(\omega))dx$

***Theorem:* 4.20:**

If S is a separable metric space, and $\mathcal{S}$ is a Borel σ-algebra, then regular conditional distribution exists.

***Theorem:* 4.21: Conditional Expectation**

If μ is a r.c.d. of X given $\mathcal{G}$ and $f : S \to \mathbb{R}$ is a measurable with $E[|f(X)|] < \infty$, then the r.v. $w \mapsto \int f(x) d\mu(w, x)$ is a version of $E[f(X)|\mathcal{G}]$.

Proof. WLOG $f \geq 0$. If f is an indicator function, then holds by definition of μ. By linearity, it holds for simple functions. Then by Theorem 1.5, it holds for $f \geq 0$. □

Example: If (Y, Z) are independen and $Z \sim \mathcal{N}(0,1)$, and if $X = ZY$, then the distribution of X given Y is $\mathcal{N}(0, Y^2)$.

Example: If r.c.d. of X given $\mathcal{Y}$, is independent of $\omega \in \Omega$, *i.e.* $\mu(\omega, A) = \nu(A)$ for a.e. $\omega \in \Omega$, then X is independent of $\mathcal{Y}$.

Proof. Let $B \in \mathcal{Y}$, $A \in \mathcal{S}$, then

$$P(X \in A, B) = E[\mu(\omega, A)1_B] = E[\nu(A)1_B] = \nu(A)P(B) = P(X \in A)P(B)$$

□

4.8 Optional Stopping Theorem

A stopping time $T \geq 0, T \in \mathbb{Z} \cup \{\infty\}$ is s.t. $\{T \leq n\} \in \mathcal{F}_n$ for all n *i.e.* $\{T = n\} \in \mathcal{F}_n$.

Define $\mathcal{F}_T = \{A : A \cap \{T = n\} \in \mathcal{F}_n, \forall n\}$. This is a σ-algebra. $T \in \mathcal{F}_T$. If $X_n \in \mathcal{F}_n$ for all n, then $X_T 1_{\{T<\infty\}} \in \mathcal{F}_T$.

If S, T are stopping times, $S \wedge T$ is stopping time. $\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.

If T is a stopping time, $0 \leq T \leq K$ for some $K \in \mathbb{N}$. If X_n is a submartingale, then $E[X_0] \leq E[X_T] \leq E[X_K]$.

Example: If Z_n is a branching process with offspring distribution $\mu = 1$, then Z_n is martingale. $Z_n = 0$ for all large n a.s. $T = \inf\{n : Z_0 = 0\}$, $T < \infty$ a.s. $Z_T = 0$, $E[Z_T] = 0 \neq E[Z_0]$.

Corollary 4. *For stopping time T and submartingale X_n, $E[X_0] \leq E[X_{T \wedge N}] \leq E[X_n]$*

Lemma 6. *Let X_n be a uniformly integrable martingale and T a stopping time. Then $X_{T \wedge n}$ is a uniformly integrable submartingale.*

Proof. Since X_n is uniformly integrable, by Theorem 4.10, $\lim_{n\to\infty} X_n = X_\infty$ a.s. and in L^1.

We check that $X_T \in L^1$,

$$\begin{aligned} E\left[|X_{T\wedge n}|\right] &= 2E[(X_{T\wedge n})^+] - E[X_{T\wedge n}] \\ &\leq 2E[(X_n)^+] - E[X_0] \\ &\leq 2\sup_n E[|X_n|] - E[X_0] < \infty. \end{aligned}$$

By Jensen's inequality, $(X_n)^+$ is also a submartingale. By Lemma 1.2, $E[|X_T|] \leq \liminf_{n\to\infty} E[|X_{T\wedge n}|] < \infty$.

$$
\begin{aligned}
E\left[|X_{T\wedge n}|\, 1_{\{|X_{T\wedge n}|>K\}}\right] &= E\left[|X_T|\, 1_{\{|X_T|>K\}} 1_{\{T\leq n\}}\right] + E\left[|X_n|\, 1_{\{|X_n|>K\}} 1_{\{T>n\}}\right] \\
&\leq E\left[|X_T|\, 1_{\{|X_T|\}>K}\right] + \sup_n E\left[|X_n|\, 1_{\{|X_n|>k\}}\right]
\end{aligned}
$$

As $K\to\infty$, RHS$\to 0$. □

Corollary 5. *If X_n is a uniformly integrable submartingale, T is a stopping time, then $E[X_0] \leq E[X_T] \leq E[X_\infty]$.*

Proof. From previous corollary, $E[X_0] \leq E[X_{T\wedge n}] \leq E[X_n]$.

Take $n\to\infty$, $X_{T\wedge n}\to X_T$ a.s. and in L^1, $X_n \to X_\infty$ in L^1. □

***Theorem:* 4.22: Optional Stopping Theorem**

If $S\leq T$ are stopping times and $Y_{T\wedge n}$ is a uniformly integrable submartingale, then $E[Y_S]\leq E[Y_T]$ and $Y_S \leq E[Y_T|\mathcal{F}_S]$.

Proof. The first part follows from the corollary with $X_n = Y_{T\wedge n}$, $E[Y_S] = E[X_S] \leq E[X_\infty] = E[Y_T]$.

Let $A\in\mathcal{F}_S$, we want to show $E[1_A Y_S] \leq E[1_A Y_T]$. Let $N = \begin{cases} S \text{ on } A \\ T \text{ on } A^C \end{cases}$. We show that N is a stopping time.

$$\{N=n\} = (\{N=n\}\cap A) \cup \left(\{N=n\}\cap A^C\right) = (\{S=n\}\cap A)\cup\left(\{T=n\}\cap A^C \cap \{S\leq n\}\right)$$

By definition, $\{S=n\}\cap A \in \mathcal{F}_n$.

If $A\in\mathcal{F}_S$, since $\mathcal{F}_S$ is a σ-algebra, then $A^C\in\mathcal{F}_S$, and $A^C\cap\{S\leq n\} = \bigcup_{j=1}^n A^C\cap\{S=j\}\in\mathcal{F}_n$.

Apply the first part with $S=N$ $T=T$, we get $E[Y_N]\leq E[Y_T]$. Note that:

$$E[Y_N] = E[Y_S 1_A] + E[Y_T 1_{A^C}] \; E[Y_T] = E[Y_T 1_A] + E[Y_T 1_{A^C}]$$

Therefore, $E[Y_S 1_A] \leq E[Y_T 1_A]$ □

Example (Simple Random Walk): Let ξ_i i.i.d, $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$. Let $S_n = \sum_{j=1}^n \xi_i$. $T_0 = \inf\{n : S_n = 0\}$ the stopping time.

Let $a<0<b$, $T = T_a\wedge T_b$, $S_{T\wedge n}$ is uniformly integrable. Note $T<\infty$ a.s.

$$0 = E[S_0] = E[S_T] = aP(T_a<T_b) + bP(T_a>T_b) = ap + b(1-p),$$

so $P(T_a<T_b) = \frac{b}{b-a}$, $P(T_a>T_b) = \frac{-a}{b-a}$.

$X_n = S_n^2 - n$ is a martingale.

$$0 = E[X_0] = E[X_{T\wedge n}] = E[S_{T\wedge n}^2] - E[T\wedge n] \Rightarrow E[S_{T\wedge n}^2] = E[T\wedge n]$$

Take $n\to\infty$. By Theorem 1.5 and 1.6, $E[S_T^2] = E[T]$. Then

$$E[S_T^2] = a^2 P(T_a<T_b) + b^2 P(T_b<T_a) = \frac{a^2 b - ab^2}{b-a} = -ab$$

If $a \to -\infty$, $E[T_b] = \infty$.

Note: If not a simple random walk $P(\xi_i = 1) \neq P(\xi_i = -1)$, there is a bias and it is not a martingale.

Example: Let ξ_i i.i.d, $P(\xi_i = 1) = p, P(\xi_i = -1) = 1 - p$, $p > \frac{1}{2}$. Let $S_n = \sum_{j=1}^{n} \xi_i$.

Let $\theta \in (0, 1)$,

$$E[\theta^{S_n} | S_{n-1} = m] = \theta^{m+1} p + \theta^{m-1}(1 - p) = \theta^m \left(\theta p + \frac{1-p}{\theta} \right)$$

If $\theta = \frac{1-p}{p}$, then $\phi(S_n)$ is a martingale, where $\phi(x) = \theta^x$.

Let $T = T_a \wedge T_b$, $a < 0 < b$. Then

$$1 = E[\theta^{S_T}] = \phi(a) P(T_a < T_b) + \phi(b)(1 - P(T_a < T_b))$$

This gives $P(T_a < T_b) = \frac{1-\phi(b)}{\phi(a)-\phi(b)}$

If we set $b \to \infty$, $P(T_a < \infty) = \lim_{b\to\infty} P(T_a < T_b) = \frac{1}{\phi(a)} < 1$.

$$P\left(\min_n S_n \leq -m \right) = P(T_{-m} < \infty) = \frac{1}{\phi(-m)} = \theta^m \Rightarrow E\left[\left| \min_n S_n \right| \right] = \sum_{m=1}^{\infty} P\left(\left| \min_n S_n \right| \geq m \right) < \infty$$

$X_n = S_n - (2p - 1)n$ is a martingale, so $X_{T_b \wedge n}$ is also a martingale.
Therefore, $E[S_{T_b \wedge n}] = (2p - 1) E[T_b \wedge n]$.

By Theorem 1.5, RHS gives $\lim_{n\to\infty} (2p - 1) E[T_b \wedge n] = (2p - 1) E[T_b]$.

For the LHS, $T_b < \infty$ a.s. so $\lim_{n\to\infty} S_{T_b \wedge n} = b$ a.s. We have $\forall n$, $|S_{T_b \wedge n}| \leq b + |\min_n S_n|$. By Theorem 1.6, $\lim_{n\to\infty} E[S_{T_b \wedge n}] = b$, so $E[T_b] = \frac{b}{2p-1}$.

5 Markov Chains

Let I be a countable state space, λ a probability measure on I s.t. $\sum_{i\in I}\lambda_i = L$, and $P = (p_{ij})_{i,j\in I}$ a transition matrix (*i.e.* $p_{ij} \geq 0$ and $\sum_{j\in I} p_{ij} = 1$, for all $i \in I$).

A sequence of r.v.s $(X_m)_{m\geq 0}$ taking values in I is Markov(λ, P) if $X_0 \sim \lambda$, and all $i, j \in I$ satisfies the Markov property, $P(X_{n+1} = j|X_n = i, X_{n-1}, ..., X_1) = P(X_{n+1} = j|X_n = i) = p_{ij}$

***Theorem:* 5.1:**

$(X_n)_{0\leq n\leq N}$ is Markov(λ, P) if and only if $P(X_0 = i_0, ..., X_N = i_N) = \lambda_{i_0}p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{N-1}i_N}$

Proof. $(\Rightarrow)$ $N = 2$, if Markov, then

$$P(X_2 = i_2, X_1 = i_1, X_0 = i_0) = P(X_2 = i_2|X_1 = i_1)P(X_1 = i_1|X_0 = i_0)P(X_0 = i_0) = \lambda_{i_0}p_{i_0i_1}p_{i_1i_2}$$

And inductively, it will be satisfied for all N

$(\Leftarrow)$ if the equality holds for $N = 2$, compute

$$P(X_2 = i_2|X_1 = i_1, X_0 = i_0) = \frac{P(X_2 = i_2, X_1 = i_1, X_0 = i_0)}{P(X_1 = i_1, X_0 = i_0)} = \frac{\lambda_{i_0}p_{i_0i_1}p_{i_1i_2}}{\lambda_{i_0}p_{i_0i_1}} = p_{i_1i_2}$$

□

***Theorem:* 5.2: Markov Property**

Conditional on $X_m = i$, we have $(X_{n+m})_{n\geq 0}$ is Markov(δ, P) and is independent of $\sigma(X_0, ..., X_m)$.

Proof. If $A \in \sigma(X_0, ..., X_m)$,

$$P(X_m = i_0, ..., X_{m+n} = i_n, A|X_m = i) = \delta_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n}P(A|X_m = i)$$

WLOG, $A = \{X_0 = j_0, ..., X_m = j_m\}$, then

$$\delta_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n}P(A|X_m = i) = P(X_m = i_0, ..., X_{m+n} = i_n, X_0 = j_0, ..., X_m = j_m, X_m = i)/P(X_m = i)$$

This can be shown by expanding and rearranging the terms. □

***Definition:* 5.1: Lead/Communicate**

We say i leads to j if $P_i(X_n = j) > 0$ for some $n \geq 0$, and write $i \to j$. Here P_i is the probability of $(X_n)_{n\geq 0}$ started at $X_0 = i$.
Write $i \leftrightarrow j$ if $i \to j$ and $j \to i$, say i communicates with j.

By definition $i \leftrightarrow j$.

***Theorem:* 5.3:**

The following are equivalent

1. $i \to j$
2. there exists a sequence $i_1, ..., i_n$ s.t. $i_1 = i, ..., i_n = j$ and $p_{i_1i_2}\cdots p_{i_{n-1}i_n} > 0$
3. $(P^n)_{ij} := p^{(n)}_{ij} > 0$ for some n

Proof. Note $p_{ij}^{(n)} = P_i(X_n = j) \le P_i(X_m = j \text{for any } m \ge 0) = \sum_m P_i(X_m = j) = \sum_m p_{ij}^{(m)}$. This shows (3) $\Leftrightarrow$ (1)

(3) $\Leftrightarrow$ (2) beacause $p_{ij}^{(m)}$ =sum over terms of the type appearing in (2) ☐

Corollary 6. *If $i \to j$ and $j \to k$, then $i \to k$.*

***Definition:* 5.2: Communication Class**

$i \leftrightarrow j$ is an equivalence relation, $C = I/\leftrightarrow$ are called communicating classes.
A communicating class C is closed if whenever $i \in C$ and $j \in I$ s.t. $i \to j$, then $j \in C$.
A state $i \in I$ is absorbing if $\{i\}$ is a closed communicating class.

***Definition:* 5.3: Hitting Time**

For a set $A \subset I$, define hitting time of A as $H^A = \inf_n \{X_n \in A\}$, H^A is a stopping time w.r.t. $\mathcal{F}_n = \sigma(X_1, ..., X_n)$. Define the hitting probability $h_i^A = P_i(H^A < \infty)$ and expected hitting time $k_i^A = E[H^A]$.

***Theorem:* 5.4:**

h_i^A are the unique minimal non-negative solution to $h_i^A = \begin{cases} 1, i \in A \\ \sum_j p_{ij} h_j^A, i \notin A \end{cases}$. Any other solution has $X_i \ge h_i^A$ for all $i \in I$.

Proof. If $i \in A$, then $h_i^A = 1$. Otherwise

$$\begin{aligned} h_i^A &= P_i(H^A < \infty) = P_i(1 \le H^A < \infty) \\ &= \sum_j P_i(1 \le H^A < \infty | X_1 = j) P_i(X_1 = j) \\ &= \sum_j p_{ij} P_j(H^A < \infty) = \sum_j p_{ij} h_j^A \end{aligned}$$

Let X_i be another solution of the equations. For $i \notin A$:

$$\begin{aligned} X_i &= \sum_j p_{ij} X_j = \sum_{j \in A} p_{i,j} + \sum_{j \notin A} p_{ij} X_j \\ &= \sum_{j_1 \in A} p_{ij_1} + \sum_{j_1 \notin A} \sum_{j_2} p_{ij_1} p_{j_1 j_2} X_{j_2} \\ &= \sum_{j_1 \in A} p_{ij_1} + \sum_{j_1 \notin A} \sum_{j_2 \in A} p_{ij_1} p_{j_1 j_2} + \sum_{j_1 \notin A} \sum_{j_2 \notin A} p_{ij_1} p_{j_1 j_2} X_{j_2} \end{aligned}$$

Continue iterating and drop the final term. All the remaining terms are positive. This gives $X_i \ge P_i(H^A \le n)$ for all n.

Take $n \to \infty$, $X_i \ge \lim_{n \to \infty} P_i(H^A \le n) = P_i(H^A < \infty) = h_i^A$. ☐

Theorem: 5.5:

k_i^A is the minimal nonnegative solution to $k_i^A = \begin{cases} 0, i \in A \\ 1 + \sum_j p_{ij} k_j^A, i \notin A \end{cases}$

Example: Simple random walk. If $P(X_{n+1} = i+1 | X_n = i) = p$, $P(X_{n+1} = i-1 | X_n = i) = q = 1-p$, and $h_i = P_i(X_n = 0, n \geq 0)$, then:

$$h_i = ph_{i+1} + qh_{i-1} \Rightarrow ph_i + qh_i = ph_{i+1} + qh_{i-1} \Rightarrow p(h_i - h_{i+1}) = q(h_{i-1} - h_i)$$

If $p = q$, general solution is $h_i = A + B_i, i \geq 0$, $0 \leq h_i \leq 1$, $h_0 = 1$, so $h_i = 1$ for all i.

If $p \neq q$, general solution is $h_i = A + B\left(\frac{q}{p}\right)^i$.

If $q > p$, then $B = 0$, $h_i = A = h_0 = 1$.

If $p > q$, then $h_0 = A + B = 1$, $h_i = 1 - B\left(1 - \left(\frac{q}{p}\right)^i\right)$. By minimality, $B = 1$, $h_i = \left(\frac{q}{p}\right)^i$.

Define $u_{i+1} = h_i - h_{i+1}$, $u_{i+1} = \frac{q_i}{p_i} u_i = \prod_{j=1}^i \frac{q_j}{p_j} u_1$. Let $\gamma_i = \prod_{j=1}^i \frac{q_j}{p_j}$ and $\gamma_0 = 1$.
If $A = u_1$, then $h_i = 1 - A\left(\sum_{j=1}^i \gamma_j\right)$.
If $\sum_{j=1}^i \gamma_j = \infty$ a.s., then we must take $A = 0$ to enforece $0 \leq h_i \leq 1$, $h_i = 1$ for all i.
Otherwise, by minimality, $A = \frac{1}{\sum_j \gamma_j}$, $h_i = 1 - \frac{\sum_{j=0}^i \gamma_j}{\sum_{j=0}^\infty \gamma_j}$.

Check with the solutions given by optional stopping theorem.

Theorem: 5.6: Strong Markov Property

Let T be a stopping time. Conditional on $\{T < \infty\}$ and $X_T = i$, the process $\{X_{T+n}\}_{n \geq 0}$ is Markov(δ_1, P) and independent of $\mathcal{F}_T$.

Proof. If $B \subset \mathcal{F}_T$, we want to show:

$$\frac{P(X_T = j_0, ..., X_{T+n} = j_n, T < \infty, X_T = i, B)}{P(X_T = i, T < \infty)} = \frac{P(X_0 = j_0, ..., X_n = j_n) P(B, X_T = i, T < \infty)}{P(X_T = i, T < \infty)}$$

Consider $\{T = m\}$,

$$\begin{aligned} & P(X_T = j_0, ..., X_{T+n} = j_n, B \cap \{T = m\}, X_T = i) \\ =& P(X_m = j_0, ..., X_{m_n} = j_n, X_m = i, B \cap \{T = m\}) \\ & \text{By basic Markov property (5.2) at time } m \\ =& P_i(X_0 = j_0, ..., X_n = j_n) P(B \cap \{T = m\}, X_m = i) \end{aligned}$$

Summing over m gives the desired equality. □

Example: in simple random walk, we consider $H_j = \inf_m]bsX_m = j$.

For $0 < s < 1$, we want to compute $E_1[s^{H_0}]$.

Condition $E_2(s^{H_0})$ on $H_1 < \infty$, $H_0 = H_1 + \tilde{H}_0$, where $\tilde{H}_0$ has the same distribution as H_0 started from site 1 by Theorem 5.6. It is independent of H_1.

$$\begin{aligned}
E_2[s^{H_0}] &= E_2[s^{H_0} 1_{\{H_1 < \infty\}}] \\
&= E_2[s^{H_0} | H_1 < \infty] P_2(H_1 < \infty) \\
&= E_2[s^{H_1 + \tilde{H}_0} | H_1 < \infty] \\
&= E_2[s^{H_1}] E_1[s^{H_0}] = (E_1[s^{H_0}])^2
\end{aligned}$$

Then $E_1[s^{H_0}] = pE_1[s^{H_0}|X_1 = 2] + qE_1[s^{H_0}|X_1 = 0]$.

If $\phi(s) = E_1[s^{H_0}]$, then $\phi(s) = ps\phi(s)^2 + qs$. This gives $\phi(s) = \frac{1 \pm \sqrt{1-4pqs^2}}{2ps}$.
Since $\phi(s)$ should be continuous on $s \in (0,1)$ and $0 \le \phi(s) \le 1$, we take the negative root $\phi(s) = \frac{1-\sqrt{1-4pqs^2}}{2ps}$.

By Theorem 1.5, $\lim_{s \to 1} \phi(s) = P_1(H_1 < \infty) = \frac{1-\sqrt{1-4pq}}{2p} = \begin{cases} 1, p < q \\ \frac{q}{p}, p > q \end{cases}$

5.1 Recurrence and Transients

***Definition:* 5.4: Recurrent and Transient**

A state i is recurrent if $P_i(X_n = i \text{ infinitely many } n) = 1$. A state i is transient if $P_i(X_n = i \text{ infinitely many } n) = 0$.

Remark 9. A state can only be either recurrent or transient.

Let $T_i = \inf\{n \ge 1, X_n = i\}$. Consider $P_i(T_i < \infty)$. By Theorem 5.6, conditioned on $T_i < \infty$, $(X_{T_i+n})_{n \ge 0} \sim$ Markov(δ_i, P).

Let $V_i =\#$ visits of X_n to state i, $P_i(V_i \ge n | V_i \ge n-1) = P_i(T_i < \infty)$ by Theorem 5.6.

If $T_i^{(n-1)}$ =time of $(n-1)$th visit to state i, then by Theorem 5.6, conditioned on $T_i^{(n-1)} < \infty$. $(X_{n+T_i^{(n-1)}})_n \sim$Markov$(\delta_i, P)$.

$$P_i(V_i \ge n) = P_i(V_i \ge n | V_i \ge n-1) P_i(V_i \ge n-1) = P_i(T_i < \infty)^n$$

***Theorem:* 5.7:**

If $P_i(T_i < \infty) = 1$, then i is recurrent, and $E_i[V_i] = \infty$
If $P_i(T_i < \infty) < 1$, then i is transient and $E_i[V_i] < \infty$

Proof.

$$P_i(V_i = \infty) = \lim_{n \to \infty} P_i(V_i \ge n) = 1$$

$$E_i[V_i] = \sum_{n=1}^{a} P_i(V_i \ge n) = \sum_{n=1}^{\infty} P_i(T_i < \infty)^{n-1} = \frac{1}{1 - P_i(T_1 < \infty)}$$

□

***Theorem:* 5.8:**

Let C be a communicating class. Then all states in C are either transient or recurrent

Proof.

$$\sum_{n=0}^{\infty} E_i(1_{\{x_n=i\}}) = E_i[V_i] = \sum_{n=0}^{\infty} p_{ii}^{(n)}$$

Let $i \sim j$, there exists n, m s.t. $p_{ij}^{(n)} > 0$ and $p_{ji}^{(n)} > 0$. Then $p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{m} \leq p_{ii}^{(n+r+m)}$. This shows that $E_i[V_i] < \infty$ if and only if $E_j[V_j] < \infty$. □

Theorem: 5.9:

1. Every recurrent class is closed.
2. Every finite closed class is recurrent.

Proof. 1. Suppose it is not closed. Then there exists $i \in C$, $j \notin C$ and $m > 0$ s.t. $P_i(X_n = j) > 0$, but then $P_i(\{X_m = j\} \cap \{X_n = ii.o.\}) = 0$.

$P_i(X_n = j) + P_i(X_n = 1i.o.) \leq 1$, $P_i(X_n = ii.o.) < 1$ so it has to be zero.

2. Given any initial distribution λ s.t. $\lambda(C) = 1$. We have for some i, $P(X_n = ii.o.) > 0$. Apply Theorem 5.6 to $T_i = \inf\{n \geq 1, X_n = i\}$, we get $P(X_n = ii.o.) = P(X_n = i \text{ for some } n) P_i(X_n = ii.o.)$, so $P_i(X_n = ii.o.) = 1$. □

Definition: 5.5: Irreducible Markov Chain

P is irreducible if the full state space I is a communicating class

Theorem: 5.10:

If P is irreducible and recurrent, then for all initial distribution λ, we have $P(T_j < \infty) = 1$ for all j

Proof. It suffices to check $P_i(T_j < \infty) = 1$ for all i, j.

Choose m s.t. $p_{ji}^{(m)} > 0$.

Since j is recurrent,

$$\begin{aligned}
1 = P_j(X_n = j, n \geq m+1) &= \sum_{k \in I} P_j(X_n = j, n \geq m+1 | X_m = k) P_j(X_m = k) \\
&= \sum_{k \in I} P_j(X_n = j, n \geq m+1 | X_m = k) p_{jk}^{(m)}
\end{aligned}$$

Since $\sum_k p_{jk}^{(m)} = 1$, for all k with $p_{jk}^{(m)} > 0$, we must have $P_k(T_j < \infty) = 1$. □

Example: Random walks

In 1D, if $p = \frac{1}{2}$, then it is recurrent, otherwise, it is transient

In 2D, if $p = \frac{1}{4}$ for each direction, then it is recurrent.

Proof. If n is odd, then $p_{00}^{(n)} = 0$, compute $p_{00}^{(n)} = \left(\binom{2n}{n} \frac{1}{2^{2n}}\right)^2$. (decompose to 1D cases)

By Stirling approximation, $p_{00}^{(n)} \sim \frac{1}{n}$. □

In 3D, $p = \frac{1}{6}$ is transient. $\sum_n p_{00}^{(n)} < \infty$, because

$$\begin{aligned} p_{00}^{(n)} &= \sum_{i,j,k,i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} \left(\frac{1}{6}\right)^{2n} \\ &= \frac{(2n)!}{(n!)^2} \left(\frac{1}{2}\right)^{2n} \sum_{i,j,k} \left(\frac{n!}{i!j!k!}\right)^2 \left(\frac{1}{3}\right)^{2n} \\ &= \frac{(2n)!}{(n!)^2} \left(\frac{1}{2}\right)^{2n} \sum_{i,j,k} \binom{n}{i,j,k}^2 \left(\frac{1}{3}\right)^{2n} \end{aligned}$$

Note $\sum_{i,j,k} \binom{n}{i,j,k} \left(\frac{1}{3}\right)^n = 1$ and if $n = 3m$, then for all $i + j + k = n$, we have $\binom{n}{i,j,k} \le \binom{n}{m,m,m}$

Therefore, if $n = 3m$, we have

$$p_{00}^{(n)} \le \frac{(2n)!}{(n!)^2} \frac{1}{2^{2n}} \binom{n}{m,m,m} \frac{1}{3^n} \sim \frac{1}{n^{3/2}}$$

Therefore, $\sum_m p_{00}^{(6m)} \sim \sum_m \frac{1}{m^{3/2}} < \infty$. Also, $\frac{1}{6^2} p_{00}^{(6m-2)} \le p_{00}^{(6m)}$, $\frac{1}{6^4} p_{00}^{(6m-4)} \le p_{00}^{(6m)}$.

5.2 Invariance and Equilibrium

Definition: 5.6: Invaraint

A measure on I is $(\lambda_i)_{i \in I}$, $\lambda_i \ge 0$, and is distribution if $\sum_{i \in I} \lambda_i = 1$.
A measure λ is invariant for P if $\lambda P = \lambda$, *i.e.* $\sum_{i \in I} \lambda_i p_{ij} = \lambda_j$ for all j.

Lemma 7. *Let λ be an invariant distribution. If $X_n \sim$ Markov(λ, P), then*

1. *$X_n \sim \lambda$ for all n*
2. *$(X_{n+m})_{m \ge 0} \sim$ Markov(λ, P)*

Proof. 1. $P(X_n = j) = (\lambda P^n)_j = \lambda_j$. 2 is Theorem 5.2. □

In the finite-dimensional case, $|I| < \infty$, an invariant distribution exists by Perron-Frobenius Theorem.

Let $T_k = \inf\{n \ge 1 : X_n = k\}$. For states k, i, define $\gamma_i^k = E_k\left[\sum_{n=0}^{T_k - 1} 1_{X_n = i}\right]$, the expected number of visits to i before visiting k. Note that $\gamma_k^k = 1$ by definition.

Theorem: 5.11:

Let P be irreducible and recurrent. Then
1. $\gamma_k^k = 1$
2. $\gamma^k P = \gamma^k$
3. $0 < \gamma_i^k < \infty$ for all i

Proof. 1. by definition.

2. Note that $\{T_k \ge n\} \in \mathcal{F}_{n-1}$.

$$\begin{aligned} P_k(X_n = j, X_{n-1} = i, T_k \ge n) &= P_k(X_n = j, T_k \ge n | X_{n-1} = i) P(X_{n-1} = i) \\ &= P_k(X_n = j | X_{n-1} = i) P(T_k \ge n | X_{n-1} = i) P_k(X_{n-1} = i) \\ &= p_{ij} P(T_k \ge n, X_{n-1} = i) \end{aligned}$$

Since P is recurrent, $T_k \leq \infty$ a.s., so $\sum_{n=0}^{T_k-1} 1_{\{X_n=i\}} = \sum_{n=1}^{T_k} 1_{\{X_n=i\}}$

$$\begin{aligned}
\gamma_i^k &= E_n\left[\sum_{n=1}^{T_k} 1_{\{X_n=i\}}\right] = E_n\left[\sum_{n=1}^{\infty} 1_{\{X_n=i\}} 1_{\{T_k \geq n\}}\right] \\
&= \sum_{n=1}^{\infty} P_k(X_n = i, T_k \geq n) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} P_k(X_n = i, X_{n-1} = j, T_k \geq n) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} p_{ji} P_k(X_{n-1} = j, T_k \geq n) \\
&= \sum_{j \in I} \sum_{n=1}^{\infty} p_{ji} E_k\left[1_{\{X_{n-1}=j\}} 1_{\{T_k \geq n\}}\right] \\
&= \sum_{j \in I} p_{ji} E\left[\sum_{n=1}^{T_k} 1_{\{X_{n-1}=j\}}\right] \\
&= \sum_{j \in I} p_{ji} E_k\left[\sum_{n=0}^{T_k-1} 1_{\{X_n=j\}}\right] = \sum_{j \in I} \gamma_j^k p_{ji}
\end{aligned}$$

3. P is irreducible, so for all $i \in I$, there exists n, m s.t. $p_{ik}^{(n)} > 0$ and $p_{ki}^{(m)} > 0$.

$$1 = \gamma_k^k = \sum_j \gamma_j^k p_{jk}^{(n)} \geq \gamma_i^k p_{ik}^{(n)} \Rightarrow \gamma_i^k < \infty$$

Similarly, $\gamma_i^k = \sum_j \gamma_j^k p_{ji}^{(m)} \geq \gamma_k^k p_{ki}^{(m)} > 0$. □

***Theorem:* 5.12:**

Let P be irreducible, and let λ be an invariant measure with $\lambda_k = 1$ for some $k \in I$. Then

1. $\lambda_i \geq \gamma_i^k$, for all $i \in I$
2. If P is recurrent, then $\lambda = \gamma^k$

Proof. 1. $\lambda_j = \sum_i \lambda_i p_{ij} = \sum_{i \neq k} \lambda_i p_{ij} + p_{kj}$. Expand on the sum, to get

$$\begin{aligned}
\lambda_j &= P_k(X_n = j, T_k \geq N) + \cdots + P_k(X_2 = j, T_k \geq 2) + P_k(X_1 = j, T_k \geq 1) \\
&= E_k\left[\sum_{n=1}^{N} 1_{\{X_n=j\}} 1_{\{T_k \geq n\}}\right]
\end{aligned}$$

Take $N \to \infty$, $\lambda_j \geq E_k\left[\sum_{n=1}^{T_k} 1_{\{X_n=j\}}\right]$

If $j \neq k$, then $\lambda_j \geq E_k\left[\sum_{k=0}^{T_{k-1}} 1_{X_n=j}\right] \gamma_j^k$.

2. If P is irreducible and recurrent, $\mu = \lambda - \gamma^k$ is an invariant measure, $\mu_k = 0$. Let $i \in I$, we have for some n, $p_{ik}^{(n)} > 0$, then

$$0 = \mu_k = \sum_j \mu_j p_{jk}^{(n)} \geq \mu_i p_{ik}^{(n)},$$

so $\mu_i = 0$ for all i. □

Start from $\gamma_i^k = E_k\left[\sum_{n=0}^{T_k-1} 1_{\{X_n=i\}}\right]$.

$$\begin{aligned}\sum_i \gamma_i^k &= \sum_i E_k\left[\sum_{n=0}^{T_k-1} 1_{\{X_n=i\}}\right] \\ &= E_k\left[\sum_{n=0}^{T_k-1}\sum_i 1_{\{X_n=i\}}\right] \\ &= E_k\left[\sum_{n=0}^{T_k-1} 1\right] = E_k[T_k]\end{aligned}$$

***Definition:* 5.7: Positive Recurrent**

A state k is positive recurrent if $E_k[T_k] < \infty$. Otherwise if recurrent but not positive recurrent, call null recurrent.

***Theorem:* 5.13:**

Let P be irreducible. The following are equivalent

1. Every state is recurrent
2. There exists a positive recurrent state k
3. There exists an invariant distribution π for P.

Futhermore, $\pi_k = \frac{1}{E_k[T_k]}$

Proof. $1 \Rightarrow 2$ by definition

$2 \Rightarrow 3$, P is recurrent so γ^k is an invariant measure $\pi_j = \frac{\pi_j^k}{E_k[T_k]}$ is an invariant distribution

$3 \Rightarrow 1$ we need to check that $\pi_j > 0$ for all j.
Since P is irreducible, there exists n s.t. $p_{ij}^{(n)} > 0$. Then $\pi_j = \sum_l \pi_l p_{lj}^{(n)} \geq \pi_i p_{ij}^{(n)} > 0$.
Now, for any fixed state k, define $\lambda_j = \frac{\pi_j}{\pi_k}$. Then λ is an invariant distribution and $\lambda_k = 1$.
By Theorem 5.12, $\lambda_j \geq \gamma_j^k$, so $\sum_j \lambda_j = \sum_j \frac{\pi_j}{\pi_k} = \frac{1}{\pi_k} < \infty$. $E_k[T_k] = \sum_j \gamma_j^k < \infty$. k is positive recurrent, so P is recurrent.

If P is recurrent, then $\pi_k = \frac{1}{E_k[T_k]}$ from $\lambda_j = \gamma_j^k$. ☐

Example: 1D random walk, $\lambda P = \lambda$

If $p = q = \frac{1}{2}$, $\lambda_i = \frac{1}{2}\lambda_{i+1} + \frac{1}{2}\lambda_{i-1}$. The general solution is $\lambda_i = A + Bi$, $A, B \in \mathbb{R}$. Every state is null recurrent. Every invariant measure has $\lambda_1 = A$

If $p \neq \frac{1}{2}$. $\lambda_i = p\lambda_{i-1} + q\lambda_{i+1}$, $\lambda_i = A + B\left(\frac{p}{q}\right)^i$. If $B \geq 0$, $A \geq 0$, this is an invariant measure.

A distribution π is invariant if $\pi P = \pi$. If $n \to \infty$, will $X_n \to \pi$? The following lemma shows that invariant measure determines the long-term behavior.

Lemma 8. *Suppose $|I| < \infty$ and there exists $c \in I$ s.t. $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all j. Then π is invariant.*

Proof.

$$\sum_k \pi_k p_{kj} = \lim_{n\to\infty} \sum_k p_{ik}^{(n)} p_{kj} = \lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$$

□

Example: $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. $P^4 = P^2 = I$, $P^5 = P^3 = P$. In particular, $\lim_{n\to\infty} p_{ij}^{(n)}$ does not exist.

Definition: 5.8: Aperiodic State

A state is aperiodic if $p_{ii}^{(n)} > 0$ for all sufficiently large n

Lemma 9. *If P is irreducible and has an aperiodic state i, then for all j, k, we have $p_{jk}^{(n)} > 0$ for sufficiently large n, depending on j, k*

Proof. For r, s, $p_{jk}^{(n+r+s)} = p_{ji}^{(n)} p_{ii}^{(r)} p_{ik}^{(s)}$. Since P is invariant, $r, s > 0$ exist. □

Theorem: 5.14:

Let P be irreducible and aperiodic and have an invariant distribution π. Let λ be any distribution $X_n \sim \text{Markov}(\lambda, P)$. Then $\forall j \in I$, $\lim_{n\to\infty} P(X_n = j) = \pi_j$.

Proof. Let $Y_n \sim \text{Markov}(\pi, P)$, independent of X_n. Consider $W_n = (X_n, Y_n)$ on $I \times I$.
W_n is a Markov chain on $I \times I$ with transition matrix $\hat{P}((i,j) \to (k,l)) = p_{ik}p_{jl}$.

By Lemma 9, $p_{ik}^{(n)} > 0, p_{jl}^{(n)} > 0$, so $\hat{P}^{(n)}((i,j) \to (k,l)) = p_{ik}^{(n)} p_{jl}^{(n)} > 0$ for n large enough. $\hat{P}$ is irreducible.

$\hat{P}$ has invariant measure given by $\pi \times \pi(i,j) = \pi_i \pi_j$.

Fix state $b \in I$, let $T = \inf\{n \geq 0, (X_n, Y_n) = (b,b)\}$. Since $\hat{P}$ is irreducible and has recurrent states, $\hat{P}$ is recurrent and $P(T < \infty) = 1$.

Let $Z_n = \begin{cases} X_n, n \leq T \\ Y_n, n > T \end{cases}$.

Claim: $(X_n)_{n\geq 0} \sim (Z_n)_{n \geq 0}$.
By Theorem 5.6, $(W_{T+n})_{n\geq 0} \sim \text{Markov}(\delta_{(b,b)}, \hat{P})$ independent of $\mathcal{F}_T$.

$$(W_n)_{n\geq 0} \sim (\tilde{W}_n)_{n\geq 0}, \tilde{W}_n = W_n (n \leq T), \tilde{W}_n = (Y_n, X_n)(n > T)$$

W_n is the path up to time T and the path after time T.

Then, we have

$$\begin{aligned} P(X_n = j) = P(Z_n = j) &= P(Z_n = j, T < n) + P(Z_n = j, T \geq n) \\ &= P(Y_n = j, T < n) + P(Z_n = j, T \geq n) \\ & P(Y_n = j) - P(Y_n = j, T \geq n) + P(Z_n = j, T \geq n) \\ &= \pi_j \end{aligned}$$

as $n \to \infty$. □

Theorem: 5.15: Time Reversal

Let $X_n \sim \text{Markov}(\pi, P)$ with π invariant distribution for irreducible P. Then for all N, $Y_n = X_{N-n}$, $0 \leq n \leq N$, $Y_n \sim \text{Markov}(\pi, \hat{P})$, where $\hat{p}_{ji} = \frac{p_{ij}\pi_i}{\pi_j}$. Also π is invariant for $\hat{P}$ and $\hat{P}$ is irreducible.

Proof. $\sum_i \hat{p}_{ji} = \sum_i \frac{\pi_i p_{ij}}{\pi_j} = \frac{\pi_j}{\pi_j} = 1$, so $\hat{P}$ is stochastic

$\sum_j \pi_j \hat{p}_{ji} = \sum_j \pi_i p_{ij} = \pi_i$ is invariant for $\hat{P}$.

$$\begin{aligned} P(Y_0 = i_0, ..., Y_N = i_N) = P(X_0 = i_0, ..., X_N = i_0) \\ &= \pi_{i_N} p_{i_N i_{N-1}} \cdots p_{i_1 i_0} \\ &= \hat{p}_{i_{N-1} i_N} \pi_{i_{N-1}} p_{i_{N-1} i_{N-2}} \cdots \\ &= \pi_{i_0} \hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{N-1} i_N} \end{aligned}$$

□

Definition: 5.9: Detailed Balance

(λ, P) are in detailed balance if $\forall i, j$, $\lambda_i p_{ij} = \lambda_j p_{ji}$

Lemma 10. *If (λ, P) are in detailed balance, then λ is invariant for P.*

Proof. $\sum_j \lambda_j p_{ji} = \sum_j \lambda_i p_{ij} = \lambda_i$ as row eigen vectors. □

Definition: 5.10: Reversible Markov Chain

Let P be irreducible and $X_n \sim \text{Markov}(\lambda, P)$. X_n is reversible if $(X_{N-n})_{0 \le n \le N} \sim \text{Markov}(\lambda, P)$ for all N.

Theorem: 5.16:

If P is irreducible and $X_n \sim \text{Markov}(\lambda, P)$. Then the following are equivalent:

1. X_n is reversible
2. P and λ are in detailed balance.

Proof. ($\Rightarrow$) if X_n is reversible, then λ is invariant for P.

By Theorem 5.15, $(X_{N-n})_n \sim \text{Markov}(\lambda, \hat{P}) \sim \text{Markov}(\lambda, P)$. Then $\lambda_i \hat{p}_{ij} = \lambda_i p_{ij} = \lambda_j p_{ji}$ because $\hat{P} = P$. This shows that (λ, P) are in detailed balance.

($\Leftarrow$) λ is invariant for P. By Theorem 5.15, $(X_{N-n}) \sim \text{Markov}(\lambda, \hat{P})$, detailed balance $\Rightarrow \hat{P} = P$ reversible.

□

Remark 10. If (λ, P) are in detailed balance, then P is self-adjoint w.r.t. quadratic form given by $\langle v, u \rangle_\lambda = \sum_i v_i u_i \lambda_i$.

$$\langle v, Pu \rangle_\lambda = \sum_j v_j \sum_i p_{ji} u_i \lambda_j = \sum_j v_j u_i p_{ji} \lambda_i = \langle Pv, u \rangle_\lambda$$

5.3 Ergodic Theorem

Theorem: 5.17:

Let P be irreducible and let $V_i(n)$ be the number of visits to state i before time $n-1$, $V_i(n) = \sum_{j=0}^{n-1} 1_{\{X_j = i\}}$, where $X_n \sim M(\lambda, P)$ for any initial distribution λ. Then for all i, we have $\lim_{n\to\infty} \frac{V_i(n)}{n} = \frac{1}{m_i} = \frac{1}{E_i[T_i]}$ a.s.

Proof. In the transient case, $m_i = \infty$, and $\lim_{n\to\infty} \frac{V_i(n)}{n} = 0$

In the recurrent case, fix some state i, let $T^{(i)} = \inf_{n\geq 0}\{X_n = i\}$. Then by Theorem 5.6, $(X_{m+T^{(i)}}) \sim M(\delta, P)$. Since $T^{(i)} < \infty$ a.s., it suffices to show when $\lambda = S_i$. Let S_j be the length of jth excursion from state i.

If $T(k) = k$th time to visit state i, $T(k) - T(k-1) = S_{k+1}$. By Theorem 5.6, S_j are i.i.d. r.v.s with mean $m_i = E_i\left[\inf_{n\geq 0}\{X_n = i\}\right]$. Then $S_1 + \cdots + S_{V(n)-1} \leq n$, but $S_1 + \cdots + S_{V(n)} \geq n$.

$$1 \cdot m_i \leftarrow \frac{V(n)-1}{V(n)} \frac{S_1 + \cdots + S_{V(n)-1}}{V(n)-1} = \frac{S_1 + \cdots + S_{V(n)-1}}{V(n)} \leq \frac{n}{V(n)} \leq \frac{S_1 + \cdots + S_{V(n)}}{V(n)} \to m_i$$

By squeeze theorem, $\frac{n}{V(n)} \to m_i$. □

***Theorem:* 5.18:**

Let P be irreducible and positively recurrent, $\pi_k = \frac{1}{m_k}$ is an invariant measure, $X_n \sim M(\lambda, P)$ with any initial distribution λ. Let $f : I \to \mathbb{R}$ be bounded. Then almost surely,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(X_j) = \int f d\pi = \sum_{i\in I} f(i)\pi_i$$

Proof. By definition, $\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) = \sum_{i\in I} f(i)\frac{V_i(n)}{n}$. Assume for all i, $\frac{V_i(n)}{n} \to \pi_i$ as $n \to \infty$. Consider the difference:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(X_j) - \int f d\pi = \sum_{i\in I} f(i)\left(\frac{V_i(n)}{n} - \pi_i\right)$$

Let $\epsilon > 0$, there exists finite $J \subset I$ s.t. $\sum_{i\notin J} \pi_i < \epsilon$. WLOG, assume $|f(i)| \leq 1$ for all i.

$$\begin{aligned}
\left|\sum_{i\in I} f(i)\left(\frac{V_i(n)}{n} - \pi_i\right)\right| &\leq \sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + \sum_{i\notin J}\frac{V_i(n)}{n} + \sum_{i\notin J}\pi_i \\
&\leq 2\sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + 2\sum_{i\notin J}\pi_i \\
&< 2\epsilon
\end{aligned}$$

□

5.4 Continuous Time Markov Chains

Let the transition probabilities be $q_{ij} \geq 0$ for $j \neq i$. Suppose $X_t = i$, for any $j \neq i$, let $T_{ij} \sim Exp(q_{ij})$. Let $k = \arg\min_j T_{ij}$, then jump to state k at time $t + T_{ik}$. Recall that for exponentially distributed r.v.s, $T \sim Exp(\lambda)$, $P(T > t + s | T > s) = P(T > t)$.

Continuous Time Random Processes: Let $(\Omega, \mathcal{F}, P)$ be a probability space. For all $t \geq 0$, we have a r.v. $\omega \mapsto X_t(\omega)$. For each ω, we have a random function $t \mapsto X_t(\omega)$.

If we want to know $P(X_t = i \text{ for some } t \in [0, \infty))$, we need the smallest σ-alg for which X_t is measurable to be $\sigma(X_t : t \geq 0) \subset \mathcal{F}$. This makes sense for discrete cases as well.

If $t \to X_t(\omega)$ is right continuous for every $\omega \in \Omega$, then

$$\{X_t = i \text{ for some } t \in [0,\infty)\} = \bigcap_{n=1}^{\infty} \bigcup_{j_1,...,j_n \in I, j_1 \neq i, j_n \neq i} \{X_{q_1} = j_1, ..., X_{q_n} = j_n\},$$

where q_n is enumeration of rationals.

For right continuous process, specify X_t by jump times J_i and jump chain $Y_n = X_{J_n}$. The holding times are $S_n = J_n - J_{n-1}$. The explosion time is $\zeta = \sup_n J_n$. A process is minimal if $X_t = \{\infty\}$ for $t \geq \zeta$.

***Theorem:* 5.19:**

Let T_k be independent and $T_k \sim Exp(q)$. Assume $q = \sum_k q_k < \infty$. Let $T = \inf_k T_k$. Then the infimum is achieved at a unique index K, independent of the value T. $T \sim Exp(q)$ and $P(K = j) = \frac{q_j}{q}$.

Proof.

$$\begin{aligned} P(T > t, K = j) &= P(T_j > t, T_l > T_j, \forall l \neq j) \\ &= \int_0^{\infty} P(T_j > t, T_l > T_j, \forall l \neq j | T_j = s) P(T_j = s) ds \\ &= \int_t^{\infty} P(T_l > s, \forall l \neq j) q_j e^{-q_j s} ds \\ &= \int_t^{\infty} \prod_{l \neq j} e^{-q_l s} q_j e^{-q_j s} ds \\ &= q_j \int_t^{\infty} e^{-qs} ds = \frac{q_j}{q} e^{-qt} \end{aligned}$$

□

***Definition:* 5.11: Poisson Process**

X_t is a Poisson process with rate $\lambda > 0$ if jump chain $Y_n = n$ for all $n \geq 0$ and holding times are i.i.d. $S_n \sim Exp(\lambda)$.

If S_n are i.i.d. Exp(λ) and $J_n = \sum_{j=1}^n S_j$, then $X_t = n$ if and only if $J_n \leq t < J_{n+1}$ is a Poisson process. By Theorem 2.3, $J_n \to \infty$ as $n \to \infty$ almost surely, so $\zeta = \infty$ (no explosion).

***Theorem:* 5.20: Poisson Process Markov Property**

For all $s \geq 0$, the process $\tilde{X}_t = X_{t+s} - X_s$ is Poisson(λ) and is indepedent of $\sigma(X_r, r \leq s)$.

Proof. Condition on $\{X_s = i\} = \{J_i \leq s\} \cap \{J_{i+1} > s\} = \{J_i \leq s\} \cap \{S_{i+1} > s - J_i\}$.

For $r \leq s$, $X_r = \sum_{n=1}^i 1_{\{J_n \leq r\}} \in \sigma(S_1, ..., S_i)$.

Now compute the jump chain and holding times of $\tilde{X}_t$ conditional on $\{X_s = j\}$, $\tilde{Y_n} = n$, $\tilde{S_1} = J_{i+1} - s = S_{i+1} + J_i - s$. For $n \geq 2$, $\tilde{S_n} = S_{n+i}$.

$$P(\tilde{S_1} > u + (s - J_1) | J_i \leq s, S_{i+1} > s - J_i) = e^{-\lambda u}$$

Conditional on $\{X_s = i\}$, $\tilde{S_1} \sim Exp(\lambda)$, indepedent of $J_1, ..., J_i$. Therefore, for $n \geq 2$, $\tilde{S_n}$ are independent of $J_1, ..., J_i$ and are i.i.d. Exp(λ).

Let $B \in \sigma(X_r : r \leq s)$.

$$P(\tilde{S_k} > t_k, B) = \sum_i P(\tilde{S_k} > t_k, B | X_s = i) P(X_s = i)$$
$$= \sum_i P(\tilde{S_k} > t_k) P(B | X_s = i) P(X_s = i) = P(B) P(S_k > t_k)$$

□

Theorem: 5.21: Properties of Poisson Process

Let X_t be right continuous process that is integer valued $X_0 = 0$. The following are equivalent

1. Holding times are i.i.d. $\text{Exp}(\lambda)$, $Y_n = n$
2. $(X_t)_{t\geq 0}$ has independent increments, *i.e.* If $I_k = (a_k, b_k)$ are disjoint, then $(X_{t_k} - X_{t_{k-1}})_k$ are i.i.d. and as $h \to 0$, we have uniformly in t, $P(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h)$, $P(X_{t+h} - X_t = 1) = \lambda h + o(h)$.
3. Increments are independent and stationary, *i.e.* distribution of $X_{t+s} - X_s$ depends only on t and for all t, $X_t \sim$Poisson(λt), $P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

Proof. $1 \Rightarrow 2$: Suppose holding times are i.i.d. $\text{Exp}(\lambda)$.

$$P(X_{t+h} - X_t \geq 1) = P(X_h \geq 1) = P(J_1 \leq h) = 1 - e^{-\lambda h} = \lambda h + o(h)$$
$$P(X_{t+h} - X_t \geq 2) = P(X_h \geq 2) = P(S_1 + S_2 \leq h) \leq P(S_1 \leq h, S_2 \leq h) = (P(S_1 \leq h))^2$$
$$\leq Ch^2 = o(h)$$

By Theorem 5.20, $X_{b_3} - X_{a_3}$ is independent of $\sigma(X_s, s \leq a_3)$.

$2 \Rightarrow 3$: Set $p_j(t) = P(X_t = j)$. Assume $j \geq 1$.

$$p_j(t+h) = P(X_{t+h} = j) = \sum_{i=0}^{j} P(X_{t+h} = j, X_t = i)$$
$$= \sum_{i=0}^{j} P(X_{t+h} - X_t = j - i, X_t = i)$$
$$= \sum_{i=0}^{j} P(X_{t+h} - X_t = j - i) P(X_t = i)$$
$$= P(X_{t+h} - X_t = 0) P(X_t = j) + P(X_{t+h} - X_t = 1) P(X_t = j - 1) + o(h)$$
$$= (1 - \lambda h) p_j(t) + \lambda h p_{j-1}(t) + o(h)$$

Rearranging to get $\frac{p_j(t+h) - p_j(t)}{h} = -\lambda p_j(t) + \lambda p_{j-1}(t) + o(1)$.

Similarly, check the other side, $\frac{p_j(s) - p_j(s-h)}{h} = -\lambda p_j(s-h) + \lambda p_{j-1}(s-h) + o(1)$.

So $p_j(t)$ is Lipschitz (difference quotients are bounded), and $p_j(t)$ is differentiable a.e.

Solve the system $\begin{cases} p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t), j \geq 1 \\ p_0'(t) = -\lambda p_0(t) \\ p_0(0) = 1 \\ p_j(0) = 0, \forall j \geq 1 \end{cases}$ explicitly, we get $p_j(t) = P(X_t = j) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$.

With the same calculation, we get $\tilde{X}_t \sim X_t$. □

Theorem: 5.22:

Let $X_t \sim Poisson(\lambda)$, $Y_t \sim Poisson(\mu)$ be independent. Then $Z_t = X_t + Y_t \sim Poisson(\lambda + \mu)$.

Proof. Since X_t and Y_t have independent and stationary increments, so does Z_t. $Z_{t+s} - Z_s = (X_{t+s} - X_s) + (Y_{t+s} - Y_s)$. □

Theorem: 5.23:

Let $X_t \sim Poisson(\lambda)$.

1. Conditional on exactly one jump occuring in the interval $[s, s+t]$, the jump is uniformly distributed in the interval.
2. Conditional on $X_t = n$, then n jumps have density $f(t_1, ..., t_n) = \frac{n!}{t^n} 1_{\{0<t_1<\cdots<t_n<t\}}$

Proof. 1. By Theorem 5.2, it suffices to show for $s = 0$. Compute the density of first jump J_1.

$$P(J_1 \leq u | X_t = 1) = \frac{P(X_u \geq 1 \cap X_t = 1)}{X_t = 1} = \frac{P(X_u = 1, X_t - X_u = 0)}{P(X_t = 1)} = \frac{(\lambda u)e^{-\lambda u}e^{-\lambda(t-u)}}{(\lambda t)e^{-\lambda t}} = \frac{u}{t}$$

2. Note $\{X_t = n\} = \{J_n \leq t < J_{n+1}\}$.
Density of $J_1, ..., J_n$ conditional on $\{J_n \leq t < J_{n+1}\}$ are $S_1, ..., S_{n+1} \sim Exp(\lambda)$.
The density is $\lambda^{n+1} e^{-\lambda(S_1 + \cdots S_{n+1})} 1_{\{S_1>0,...,S_{n+1}>0\}}$.

Change of vairable from $(S_1, ..., S_{n+1})$ to $(J_1, ..., J_{n+1})$. Density of J is given by $\lambda^{n+1}e^{-\lambda J_{n+1}} 1_{\{0<J_1<\cdots<J_{n+1}\}}$.

Let $A \subset \mathbb{R}^n$,

$$P((J_1, ..., J_n) \subset A | \{J_n \leq t < J_{n+1}\}) = \frac{P((J_1, ..., J_n) \subset A \cap \{J_n \leq t < J_{n+1}\})}{P(J_n \leq t < J_{n+1})}$$

The numerator is

$$\lambda^{n+1} \int_{(t_1,...,t_n) \in A, t_n \leq t < t_{n+1}} e^{-\lambda t_{n+1}} 1_{\{t_1<t_2<\cdots<t_{n+1}\}} = \lambda^n e^{-\lambda t} \int 1_{\{t_1<\cdots<t_n<t\}}$$

□

Construction of Poisson processes Let I be state space, Q be transition matrix s.t.

1. $q_{ij} \geq 0$ for $i \neq j$
2. $\sum_{j \neq i} q_{ij} < \infty$
3. $q_{ii} = -\sum_{j \neq i} q_{ij}$

Define $q_i = -q_{ii} = \sum_{j \neq i} q_{ij}$, Π a $I \times I$ matrix s.t. for $i \neq j$, $\Pi_{ij} = \begin{cases} \frac{q_{ij}}{q_i}, q_i \neq 0 \\ 0, \text{ else} \end{cases}$, $\Pi_{ii} = \begin{cases} 0, q_i \neq 0 \\ 1, q_i = 0 \end{cases}$.

$X_t \sim Markov(\lambda, Q)$ if $Y_n \sim Markov(\lambda, \Pi)$ and if conditional on $Y_0, ..., Y_{n-1}$, the holding times $S_1, ..., S_n$ are independent $S_j \sim Exp(q_{Y_{j-1}})$. $X_t = \begin{cases} Y_n, J_n \leq t < J_{n+1} \\ \infty, t \geq \sup_n J_n \end{cases}$

First construction: Let $Y_n \sim Markov(\lambda, \Pi)$. Let T_i be i.i.d. Exp(1), independent of Y_n. Then the set $S_i = \frac{T_i}{q_{Y_{i-1}}}$ is a Poisson process.

Second construction: Let $(T_n^j)_{n\geq 1, j\in I}$ i.i.d. Exp(1). Set $X_0 \sim \lambda$. For $n \geq 1$, conditional on $\{Y_{n-1} = i\}$, for all $j \neq i$, set $S_n^j = \frac{T_n^j}{q_{ij}}$. Set $S_n = \inf_{j\neq i} S_n^j$ and let $Y_n = j$ s.t. $S_n^j = S_n$. We have $P(Y_n = j | Y_{n-1} = i) = \frac{q_{ij}}{q_i} = \Pi_{ij}$, and $S_n \sim Exp(q_{Y_{n-1}})$.

Third construction: $X_0 \sim \lambda$. For all states $i \neq j$. Let $N_t^{ij} \sim Poisson(q_{ij})$ with $N_t^{i,j}$ independent of $N_t^{k,l}$ if $(i,j) \neq (k,l)$.

***Theorem:* 5.24:**

Let $X_t \sim M(\lambda, Q)$. Then X_t does not explode if one of the following is satisfied:

1. I is finite
2. $\sup_i q_i < \infty$
3. $X_0 = i$ and i is recurrent for the jump chain

Proof. $1 \Rightarrow 2$.

2. Let $T_n = q_{Y_{n-1}} S_n$. Then $T_n \sim Exp(1)$ i.i.d. The explosion time $\zeta = \sup_n J_n = \sum_{n=1}^\infty S_n$.

$$\left(\sup_i q_i\right)\zeta = \sum_{n=1}^\infty \frac{\sup q_i T_n}{q_{Y_{n-1}}} \geq \sum_{n=1}^\infty T_n \to \infty \text{ a.s.}$$

Therefore $\zeta = \infty$, since $\sup_i q_i < \infty$ by assumption

3. Let $N_k \leq N_{k+1}$ s.t. $Y_{N_k} = i$. Then $T_{N_k} \sim Exp(1)$ i.i.d. $\zeta \geq \sum_{k=1}^\infty S_{N_k} = \sum_{k=1}^\infty \frac{T_{N_k}}{q_i} = \infty$. □

Remark 11. Explosion is related to the jump chain being transient.

***Theorem:* 5.25: Continuous Time Markov Property**

Let $X_t \sim M(\lambda, Q)$, $\zeta = \sup J_n$. Conditional on $t < \zeta$, $X_t = i$, $(X_{s+t})_s \sim M(\delta, Q)$ independent of $\mathcal{F}_t = \sigma(X_u, u \leq t)$.

$T : \Omega \to [0, \infty]$ is a stopping time if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. The σ-algebra associated to T is $\mathcal{F}_T = \{A : A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t\}$. If $S \leq T$ are stopping times, then $\mathcal{F}_S \subset \mathcal{F}_T$.

***Theorem:* 5.26: Continuous Time Strong Markov Property**

Let $X_t \sim M(\lambda, Q)$, T be a stopping time. Conditional on $T < \zeta$ and $X_T = i$. Then $(X_{T+s})_{s\geq 0} \sim M(\delta, Q)$, independent of $\mathcal{F}_T$.

Proof. Assume $\zeta = \infty$ and $T < \infty$ a.s. For all n, define $T_n = \frac{k}{2^n}$ if $T \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$. T_n is a stopping time. $T_n \to T$ as $n \to \infty$.

Since X_t is right continuous $X_{T_n} \to X_T$, $X_{T_{n+s}} \to X_{T+s}$ as $n \to \infty$.

Since $T_n \geq T$, $\mathcal{F}_T \subset \mathcal{F}_{T_n}$.

Let $A \in \mathcal{F}_T$,

$$\begin{aligned}
P(X_{T+s} = j, A, X_T = i) &= \lim_{n\to\infty} P(X_{T_n+s} = j, A, X_{T_n=i}) \\
&= \sum_k P\left(X_{T_n+s} = j, A, X_{T_n} = i, T_n = \frac{k}{2^n}\right) \\
&= \sum_k P\left(X_{\frac{k}{2^n}+s} = j, A \cap \left\{T_n = \frac{k}{2^n}\right\}, X_{\frac{k}{2^n}} = i\right) \\
&= \sum_k P_i(X_s = j) P\left(A \cap \left\{T_n = \frac{k}{2^n}\right\} | X_{\frac{k}{2^n}} = i\right) P\left(X_{\frac{k}{2^n}} = i\right) \\
&= \sum_k P_i(X_s = j) P\left(A \cap \left\{T_n = \frac{k}{2^n}\right\} \cap X_{\frac{k}{2^n}} = i\right) \\
&= P_i(X_s = j) P\left(A \cap \{X_{T_n} = i\}\right) \\
&\to P_i(X_s = j) P\left(A \cap \{X_T = i\}\right) = P(X_{T+s} = j, A, X_T = i)
\end{aligned}$$

as $n \to \infty$. □

Let $X_t \sim M(\lambda, Q)$, $P(t) = (P_{ij}(t))_{i,j\in I}$, $P_{ij}(t) = P_i(X_t = j)$. Define:

$$\begin{aligned}
\text{Forward equation} &: \frac{d}{dt}P(t) = P(t)Q \\
\text{Backward equation} &: \frac{d}{dt}P(t) = QP(t)
\end{aligned}$$

Finite state space:

$$P_i(\text{two jumps} \le h) = P_i\left(J_1 \le h \cap \left(J_1 + \frac{J_2}{q_{Y_1}}\right) \le h\right) \le P_i(J_1 \le h, J_2 \le Ch) \le Ch^2,$$

where $J_1 + \frac{J_2}{q_{Y_1}}$ is the second jump time.

If $j = i$, $P_i(X_n = i) = P_i(J_1 > h) + o(h) = e^{-q_i h} + o(h) = 1 - q_i h + o(h)$

If $j \neq i$, $P_i(X_n = j) = P_i(J_1 \le h, Y_1 = j) + o(h) = \pi_{ij}(1 - e^{-q_i h}) + o(h) = q_i \pi_{ij} + o(h) = q_{ij} + o(h)$

$$\begin{aligned}
P_{ij}(t+h) &= \sum_k P_i(X_t = k) P_k(X_h = j) = \sum_k P_{ik}(\delta_{kj} + q_{kj} h) + o(h) \\
\Rightarrow \frac{P_{ij}(t+h) - P_{ij}(t)}{h} &= \sum_k P_{ik}(t) Q_{kj} + o(1) \\
\Rightarrow \frac{d}{dt} Q &= PQ
\end{aligned}$$

Theorem: 5.27: Backward Equation

Let Q be a transition matrix, $(P(t))_{i,j} = P_i(X_t = j)$. Then P is the minimal nonnegative solution to $P' = QP$ (the backward equation). P is a semigroup $P(s+t) = P(s)P(t)$.

Proof.

$$\begin{aligned}
P_{ij}(t) &= P_i(X_t = j, t < J_1) + P_i(X_t = j, J_1 \le t) \\
&= e^{-q_i t}\delta_{ij} + \sum_k P_i(X_t = j, X_{J_1} = k, J_1 \le t) \\
&= e^{-q_i t}\delta_{ij} + \sum_k \int_0^t P_i(X_t = j, X_{J_1} = k | J_1 = s) e^{-q_i s} q_i ds \\
&= e^{-q_i t}\delta_{ij} + \sum_k \int_0^t P_k(X_{t-s} = j) P_i(X_{J_1} = k | J_1 = s) e^{-q_i s} q_i ds \\
&= e^{-q_i t}\delta_{ij} + \sum_k \int_0^t P_{kj}(t-s)\pi_{ik} q_i e^{-q_i s} ds
\end{aligned}$$

Let $u = t - s$.

$$e^{q_i t} P_{ij}(t) = \delta_{ij} \int_0^t \sum_{k \ne i} q_i e^{-q_i u} \pi_{ik} q_{kj}(u) du$$

p_{ij} is continuous in t and the integral of sum is uniformly convergent. Also, $q_i \pi_{ik} = q_{ik}$, we have $P' = QP$.

Minimality: suppose $\tilde{P}_{ij}(t)$ is another non-negative solution to the equation. Then

$$\tilde{P}_{ij} = e^{-q_i t}\delta_{ij} + \sum_{k \ne i} \int_0^t q_i e^{-q_i} \pi_{ik} \tilde{P}_{kj}(t-s)$$

Since also

$$P_i(X_t = j, t < J_{n+1}) = e^{-q_i t}\delta_{ij} + \sum_{k \ne i} \int_0^t q_i e^{-q_i s} \pi_{ik} P_k(X_{t-s} = j, t - s < J_n) ds,$$

we have $\tilde{P}_{ij}(t) \ge P_i(X_t = j, t < J_{n+1})$ □

***Theorem:* 5.28: Forward Equation**

P is also the minimal non-negative solution of the forward equation $P' = PQ$. *i.e.* $\frac{d}{dt} P_{ij}(t) = \sum_k P_{ik}(t) q_{kj}$.

For $P_{ij}(t-s) = P(X_t = j | X_s = i)$, we have the following forward/backward equations:

$$\begin{aligned}
&\frac{\partial}{\partial t} P_{ij}(t, s) = \sum_k P_{ik}(t, s) Q_{kj}(t), \text{ with } P_{ij}(s, s) = \delta_{ij} \\
&\frac{\partial}{\partial t} P_{ij}(t, s) = -\sum_k Q_{ik}(s) P_{kj}(s, t), \text{ with } P_{ij}(t, t) = \delta_{ij}
\end{aligned}$$

Example(Time-homogenous Markov Chain): Let $F : I \to \mathbb{R}$ be bounded,
$f(i, t) = E_i[F(X_t)] = \sum_j P_{ij}(t) F(j)$.

$$\partial_t f(i, t) = \sum_j \partial_t P_{ij}(t) F(j) = \sum_j q_{ik} P_{kj}(t) F(j) = (Qf(\cdot, t))_i$$

$\partial_t f(\cdot, t) = Qf(\cdot, t)$ solves the backwards equation with initial condition $f(\cdot, 0) = F$.

5.4.1 Properties

The properties of continuous-time Markov chain follows from the jump chain

Class structure:

1. Leads: $i \to j$ if $P_i(X_t = j, t \geq 0) > 0$
2. Communicates: $i \sim j$ if $i \to j$ and $j \to i$
3. Communicating class, irreducibility, absorbing states, closed classes are the same as in discrete settings

***Theorem:* 5.29:**

The following are equivalent

1. $i \to j$
2. $i \to j$ for jump chain (communicating class for X_t are the same as for jump chain Y_n)
3. $i = i_0, i_1, ..., i_n = j$, $q_{i_0 i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} > 0$
4. $P_{ij}(t) > 0$ for all $t > 0$
5. $P_{ij}(t) > 0$ for some t.

Proof. $2 \Rightarrow 3$: based on discrete case, there exists a sequence $i = i_0 \neq i_1 \neq \cdots \neq i_n = j$ s.t. $\pi_{i_0 i_1} \cdots \pi_{i_{n-1} i_n} > 0$, and $\pi_{ij} = \frac{q_{ij}}{q_i}$

$3 \Rightarrow 4$: suppose $q_{ab} > 0$ for $a \neq b$,

$$\begin{aligned} P_{ab}(t) &\geq P_a(X_t = b, J_1 < t, S_2 > t) = \pi_{ab}(1 - e^{-q_a t})e^{-q_b t} \\ &= \frac{q_{ab}}{q_a}(1 - e^{-q_a t})e^{-q_b t} > 0 \end{aligned}$$

If 3 holds, then $P_{ij}(t) \geq P_{i_0 i_1}(t/n) P_{i_1 i_2}(t/n) \cdots P_{i_{n-1} j}(t/n)$. ☐

Hitting times and hitting probabilities:
Given $A \subset I$, define $D^A = \inf\{t \geq 0 : X_t \in A\}$, $H^A = \inf\{n \geq 0, Y_n \in A\}$. $P_i(D^A < \infty) = P(H^A < \infty)$.

***Theorem:* 5.30:**

Let $h_i^A = P_i(D^A < \infty)$. Then h^A is the minimal non-negative solution to $\begin{cases} h_i^A = 1, i \in A \\ \sum_j q_{ij} h_j^A = 0, i \notin A \end{cases}$

Assume $q_i > 0$, $\forall i \notin A$, then $K_i^A = E_i[D^A]$ are minimal non-negative solution to
$\begin{cases} K_i^A = 0, i \in A \\ K_i^A = \frac{1}{q_i} + \frac{1}{q_i} \sum_{j \notin A} q_{ij} K_j^A, i \notin A \end{cases}$

Proof. For $i \notin A$, $h_i^A = \sum_{j \neq i} \pi_{ij} h_j^A = \sum_{j \neq i} \frac{q_{ij}}{q_i} h_j^A$

For $i \notin A$, $E_i[D^A - J_1 | Y_1 = j] = K_j^A$ by Theorem 5.26.

$$K_i^A = E_i[D^A - J_1] + E_i[J_1] = \frac{1}{q_i} + \sum_{j \neq i} E_i[D^A - J_1 | Y_1 = j]\pi_{ij} = \frac{1}{q_i} + \sum_{j \neq i, i \notin A} \frac{q_{ij}}{q_i} K_j^A$$

☐

Recurrence & Transience:

- i is recurrent if $P_i(\{t : X_t = i\}$ is unbounded$) = 1$
- i is transient if $P_i(\{t : X_t = i\}$ is unbounded$) = 0$

Remark 12. If $P_i(\zeta < \infty) > 0$, then it is not recurrent.

***Theorem:* 5.31:**

1. If i is recurrent for jump chain, then i is recurrent
2. If i is transient for jump chain, then i is transient
3. Every state is either recurrent or transient
4. Recurrence/transience are class properties

Proof. 1. If i is recurrent for jump chain, then $P_1(\zeta = \infty) = 1$, we return to i infinitely many times. Every bounded intervals has finitely many times, $\{t : X_t = i\}$ is bounded

2. similar to 1 □

***Theorem:* 5.32:**

Let $T_i = \inf\{t \geq J_i, X_t = i\}$.

1. If $q_i = 0$ or $P_i(T_i < \infty) = 1$, then i is recurrent and $\int_0^\infty P_{ii}(t)dt = \infty$
2. If $q_i > 0$ and $P_i(T_i < \infty) < 1$, then i is transient and $\int_0^\infty P_{ii}(t)dt < \infty$.

Proof. If $q_i > 0$, $N_1 = \inf\{n \geq 1, Y_n = i\}$, then $P_i(T_i < \infty) = P_i(N_1 < \infty)$.

$$\begin{aligned}
\int_0^\infty P_{ii}(t)dt &= \int_0^\infty E_i\left[1_{\{X_t=i\}}\right] = E_i\left[\int_0^\infty 1_{\{X_t=i\}}dt\right] \\
&= E_i\left[\sum_{n=0}^\infty S_{n+1}1_{\{Y_n=i\}}\right] \\
&= \sum_{n=0}^\infty E_i[S_{n+1}|Y_n = i]P_i[Y_n = i] \\
&= \sum_{n=0}^\infty \frac{1}{q_i}\pi_{ii}^{(n)} = \frac{1}{q_i}\sum_{n=0}^\infty \pi_{ii}^{(n)}
\end{aligned}$$

□

***Definition:* 5.12: Continuous-Time Invariant Measure**

A measure λ is invariant if $\lambda Q = 0$.

If I is finite, $P(t) = e^{tQ}$ and $\lambda P(t) = \lambda$.

If Q is irreducible, then $q_i > 0$ for all i.

***Theorem:* 5.33:**

If Q is irreducible and recurrent, then there exists a unique invariant measure λ up to scaler multiplication.

Proof. λ is invariant if and only if $(\mu\Pi)_i = \mu_i$ with $\mu_i = q_i\lambda_i$

$$(\mu\Pi)_i = \sum_{j\neq i} \mu_j \Pi_{ji} = \sum_{j\neq i} \lambda_j q_j \frac{q_{ji}}{q_j} = \sum_{j\neq i} \lambda_j q_{ji} = \lambda_i q_i$$

□

If Q is irreducible, then we say state i is positive recurrent if $m_i = E_i[T_i] < \infty$.

***Theorem:* 5.34:**

Let Q be irreducible. The following are equivalent

1. All states are positive recurrent
2. There exists a positive recurrent state
3. Q is non-explosive and has invariant distribution λ

In all cases, $m_i = \frac{1}{\lambda_i q_i}$

Proof. $1 \Rightarrow 2$ is trivial.

First, or any state i, define $\mu_j^i = E_i\left[\int_0^{T_i\wedge\zeta} 1_{\{X_s=j\}} ds\right]$, $\sum_j \mu_j^i = E_i[T_i \wedge \zeta]$.

Second, consider the jump chain Y_n, define $N_i = \inf\{n > 0, Y_n = i\}$.

$$\begin{aligned}
\mu_j^i &= E\left[\sum_{n=0}^{\infty} S_{n+1} 1_{\{Y_n=j,n<N_i\}}\right] \\
&= \sum_{n=0}^{\infty} E_i\left[S_{n+1}|Y_n = j, n < N_i\right] P_i(Y_n = j, n < N_i) \\
&= \sum_{n=0}^{\infty} \frac{1}{q_j} P_i(Y_n = j, n < N_i) \\
&= \frac{1}{q} E_i\left[\sum_{n=0}^{N_i-1} 1_{Y_n=j}\right] = \frac{\gamma_j^i}{q_j}
\end{aligned}$$

$2 \Rightarrow 3$: 2 means $\zeta = \infty$. Jump chain is irreducible and recurrent. γ_j^i is invariant measure. Therefore, by Theorem 5.33, μ_j^i is an invariant measure for Q. μ^i is a distribution.

$3 \Rightarrow 1$: Let λ be invariant distribution. Fix a state i and define $\nu_j = \frac{\lambda_j q_j}{\lambda_i q_i}$. Then by Theorem 5.33, ν is invariant for jump chain and $\nu_i = 1$. By discrete theory, $\nu_j \geq \gamma_j^i$ for all j.

$$\begin{aligned}
m_i &= \sum_j \mu_j^i = \sum_j \frac{\gamma_j^i}{q_j} \\
&\leq \sum_j \frac{\nu_j}{q_j} = \sum_j \frac{\lambda_j q_j}{q_j \lambda_i q_i} \\
&= \frac{1}{\lambda_i q_i} \sum_j \lambda_j = \frac{1}{\lambda_i q_i}
\end{aligned}$$

This also shows all the states are positive recurrent. □

Theorem: 5.35:

Let Q be invariant and recurrent. Let λ be the invariant measure. Then

1. $\forall s$, $\lambda P(s) = \lambda$, *i.e.* $(P(s))_{ij} = P_i(X_s = j)$
2. If μ is a measure s.t. $\mu P(s) = \mu$ for some s, then μ is proportional to λ.

Proof. Since Q is irreducible and recurrent, by Theorem 5.34, Q is non-explosive.

$\forall s > 0$, $P(s)$ is a stochastic matrix with strictly positive entries. $P(s)$ is irreducible.

Fix state i, $\mu_j = E_i\left[\int_0^{T_i} 1_{\{X_i=j\}} dt\right]$, we know $\mu Q = 0$ by Definition 5.12. We want to show $\mu P(s) = \mu$.

Theorem 5.26 implies $E_i\left[\int_0^s 1_{\{X_t=j\}} dt\right] = E_i\left[\int_{T_i}^{T_i+s} 1_{\{X_t=j\}} dt\right]$.

$$\begin{aligned}
\mu_j &= E_i\left[\int_0^{T_i} 1_{\{X_t=j\}} dt\right] \\
&= E_i\left[\int_s^{T_i} 1_{\{X_t=j\}} dt\right] + E_i\left[\int_0^{s} 1_{\{X_t=j\}} dt\right] \\
&= E_i\left[\int_s^{T_i} 1_{\{X_t=j\}} dt\right] + E_i\left[\int_{T_i}^{T_i+s} 1_{\{X_t=j\}} dt\right] \\
&= E_i\left[\int_s^{T_i+s} 1_{\{X_t=j\}} dt\right] \\
&= E_i\left[\int_0^{T_i} 1_{\{X_t=j\}} dt\right] \\
&= \int_0^{\infty} E_i\left[1_{\{X_{t+s}=j, t<T_i\}}\right] dt \\
&= \sum_k \int_0^{\infty} P_i(X_{s+t} = j, t < T_i, X_t = k) dt \\
&= \sum_k \int_0^{\infty} P_i(X_{s+t} = j | X_t = k, t < T_i) P_i(X_t = k, t < T_i) dt \\
&= \sum_k \int_0^{\infty} P_{kj}(s) E_i\left[1_{\{X_t=k, t<T_i\}}\right] dt \\
&= \sum_k P_{kj}(s) E_i\left[\int_0^{T_i} 1_{\{X_t=k\}} dt\right] \\
&= \sum_k \mu_k P_{kj}(s)
\end{aligned}$$

Therefore, $\mu P(s) = \mu$. □

6 Brownian Motions

Consider simple random walk on $\mathbb{Z}$, X_j is i.i.d. with $P(X_j = \pm 1) = \frac{1}{2}$, $S_n = \sum_{j=1}^{n} X_j$. We want to look at $B_n(t) = \frac{1}{\sqrt{n}} S_{nt}$.

By Theorem 3.8,

1. $\lim_{n\to\infty} B_n(t) = \mathcal{N}(0,t)$ in distribution.
2. For $r < s < t$, $B_n(t) - B_n(s) \to \mathcal{N}(0, t-s)$, $B_n(s) - B_n(r) \to \mathcal{N}(0, s-r)$. $(B_n(t) - B_n(s), B_n(s) - B_n(r)) \to (\mathcal{N}(0, t-s), \mathcal{N}(0, s-r))$, where the normal distributions are independent.

Definition: 6.1: Brownian Motion

A stochastic process $\{B(t)\}_{t\geq 0}$ is a collection of r.v.s, supported on a common probability space.
A stochastic process $B(t)$ is a Brownian motion started from $x \in \mathbb{R}$ if

1. $B(0) = x$
2. $B(t)$ has independent increments. $B(t) - B(s) \sim \mathcal{N}(0, t-s)$
3. Almost surely, $t \mapsto B(t)$ is a continuous function

A property P holds almost surely, if there exists event $\mathcal{A}$ with $P(\mathcal{A}) = 0$ s.t. all $\omega \in \Omega$ s.t. P does not hold is in $\mathcal{A}$.

Let $B(t)$ be a standard Brownian motion, started from $x = 0$

1. $-B(t)$ is a Brownian motion
2. $\forall \lambda > 0$, $t \mapsto \frac{1}{\lambda^2} B(\lambda t)$ is a Brownian motion
3. Markov property: Fix $s \geq 0$. Then $B^{(s)}(t) = B_{t+s} - B_s$ is a Brownian motion.
4. Time inversion: $X(t) = \begin{cases} 0, t = 0 \\ tB\left(\frac{1}{t}\right), t > 0 \end{cases}$ is a Brownian motion

Proof. 4. The two processes $(X(t_1), ..., X(t_p))$ and $(B(t_1), ..., B(t_p))$ have the same finite-dimensional distributions.

If $s < t$,

$$E[B(t)B(s)] = E[(B(t) - B(s))B(s)] + E[(B(s))^2] = s$$

Similarly,

$$E\left[X(t)X(s)\right] = tsE\left[B\left(\frac{1}{s}\right)B\left(\frac{1}{t}\right)\right] = ts\frac{1}{t} = s$$

Also, $\lim_{t\to 0^+} X(t) = 0$ a.s. □

Brownian motion exists

Proof. It suffices to construct it on $[0,1]$. The idea is to sum up Gaussian r.v.s.

Let $D_n = \left\{\frac{k}{2^n}, 0 \leq k \leq 2^n, k \in \mathbb{Z}\right\}$, $D = \cup_n D_n$.

A process $B(t)$ indexed by $t \in D_n$ is a Brownian motion on D_n if it has independent increments and $B(t) - B(s) \sim \mathcal{N}(0, t-s)$.

Let $Y(t)$, $t \in D$ be i.i.d. $Y(t) \sim \mathcal{N}(0,1)$. Then $B_0(t)$ on D_0 can be defined by $B_0(0) = 0, B_0(1) = Y(1)$. By induction, we construct a Brownian motion on D_n denoted $B_n(t)$ s.t. $B_n(t) = B_{n-1}(t)$ for $t \in D_{n-1}$.

Let $t \in D_n \setminus D_{n-1}$, $r = t - \frac{1}{2^n}$ and $s = t + \frac{1}{2^n}$. Set $Z_t = \frac{Y_t}{2^{\frac{n+1}{2}}}$. Define $B(t) = \frac{1}{2}(B(r) + B(s)) + Z_t$. Then

$$B(t) - B(r) = \frac{1}{2}(B(s) - B(r)) + Z_t$$
$$B(s) - B(t) = \frac{1}{2}(B(s) - B(r)) - Z_t$$

Then

$$Var(B(t) - B(r)) = \frac{1}{4}Var(B(s) - B(r)) + \frac{1}{2^{n+1}} = \frac{1}{4}(s - r) + \frac{1}{2^{n+1}} = \frac{1}{4}\frac{1}{2^{n-1}} + \frac{1}{2^{n+1}} = \frac{1}{2^n}$$

This is the same for $Var(B(s) - B(t))$. The covariance $Cov(B(t) - B(r), B(s) - B(t)) = \frac{1}{4}Var(B(s) - B(t)) - Var(Z_t) = 0$, so the increments are independent.

From the construction, $B(t)|_{D_n}$ is a Brownian motion.

Define $B^{(n)}(t)$ linear interpolation of $B(t)|_{B_n}$, and $M_n = \sup_{t\in[0,1]} \left|B^{(n)}(t) - B^{(n-1)}(t)\right|$. By Wierstrass M-test, it suffices to show $\sum_n M_n < \infty$ a.s.

Then $B^{(n)}(t)$ is Cauchy w.r.t. uniform continuity of functions. $\left|B^{(m)}(t) - B^{(n)}(t)\right| \le \sum_{j=n+1}^{m} M_j$.

$$M_n = \sup_{t\in D_n\setminus D_{n-1}} |Z_t| = \frac{1}{2^{\frac{n+1}{2}}} \sup_{t\in D_n\setminus D_{n-1}} |Y_t|$$
$$P\left(|M_n| > \frac{\lambda}{2^{\frac{n+1}{2}}}\right) \le 2^n P(|Y_1| > \lambda)$$
$$E\left[\left(2^{\frac{n+1}{2}} M_n\right)^p\right]^{1/p} \le (2^n)^{1/p} E\left[|Y_1|^p\right]^{1/p}$$
$$E\left[\sum_{n=1}^{\infty} M_n\right] = \sum_{n=1}^{\infty} E[M_n] \le \sum_{n=1}^{\infty} E[M_n^p]^{1/p} \le C_p \sum_{n=1}^{\infty} \frac{1}{2^{\frac{n+1}{2}}} 2^{\frac{n}{p}} < \infty$$

Therefore, almost surely, $B^{(n)}(t) \to W(t)$ uniformly, so $W(t)$ is continuous.

We want to show $W(t)$ has the same finite-dimensional distribution as Brownian motion. We show for $r < s < t$, $(W(t) - W(s), W(s) - W(r)) \sim (B(t) - B(s), B(s) - B(r))$.

Choose n large enough s.t. for $r_n, s_n, t_n \in D_n$, $r_n \to r$, $s_n \to s$, $t_n \to t$ as $n \to \infty$ and $r_n < s_n < t_n$.

Since W is a.s. continuous, $W(t) = \lim_{n\to\infty} W(t_n)$ a.s.

$$(W(t_n), W(s_n), W(r_n)) = (B(t_n)|_{D_n}, B(s_n)|_{D_n}, B(r_n)|_{D_n})$$

We have $B(t_n) - B(s_n) \sim \mathcal{N}(0, t_n - s_n)$ and $B(t_n) - B(s_n)$ independent of $B(s_n) - B(r_n)$ □

Theorem: 6.1: Blumenthal's 0-1 Law

Let $\mathcal{F}_t = \sigma(B_u, u \le t)$, $\mathcal{F}_{0+} = \cap_{s>0}\mathcal{F}_s$. If $A \in \mathcal{F}_{0+}$, then $P(A) \in \{0, 1\}$.

Proof. Let $A \in \mathcal{F}_{0+}$, $g : \mathbb{R}^n \to \mathbb{R}$ be bounded continuous.

$$\begin{aligned} E[1_A g(B_{t_1}, ..., B_{t_n})] &= \lim_{\epsilon \to 0} E[1_A g(B_{t_1} - B_\epsilon, ..., B_{t_n} - B_\epsilon)] \\ &= \lim_{\epsilon \to 0} E[1_A] E[g(B_{t_1} - B_\epsilon, ..., B_{t_n} - B_\epsilon)] \\ &= E[1_A] E[g(B_{t_1}, ..., B_{t_n})] \end{aligned}$$

A is independent of $\sigma(B_s : s > 0) = \sigma(B_s : s \geq 0)$. Since Brownian motion is continuous, $B_0 = \lim_{s \to 0} B_s \in \sigma(B_s : s > 0)$. □

Proposition: 6.1:

Let $B(t)$ be a Brownian motion started from 0.

1. Almost surely, $\forall \epsilon > 0$, $\sup_{0 \leq t \leq \epsilon} B(t) > 0$, $\inf_{0 \leq t \leq \epsilon} B(t) < 0$
2. If $T_a = \inf\{t \geq 0, B(t) = a\}$, then $T_a < \infty$ a.s.
3. $\limsup_{t \to \infty} B(t) = \infty$, $\liminf_{t \to \infty} B(t) = -\infty$

Proof. 1. Let $A = \bigcap_{n=1}^{\infty} \left\{ \sup_{0 \leq t \leq \frac{1}{n}} B(t) > 0 \right\}$. We want to show that $P(A) = 1$.

$A \in \mathcal{F}_{0+}$, by Theorem 6.1, we just need to check $P(A) = 0$

$$P(A) = \lim_{n \to \infty} P\left(\sup_{0 \leq t \leq \frac{1}{n}} B(t) > 0 \right) \geq \limsup_{n \to \infty} P\left(B\left(\frac{1}{n}\right) > 0 \right) = \frac{1}{2}$$

2. We want to show that $\sup_{t \geq 0} B(t) \geq M$ for all $M > 0$ a.s.

From 1, $P\left(\sup_{0 \leq t \leq 1} B(t) > 0\right)$. Let $\epsilon > 0$, $\exists \delta > 0$ s.t. $P\left(\sup_{0 \leq t \leq 1} B(t) > \delta\right) \geq 1 - \epsilon$ by continuity of probability measures. Using Brownian scaling

$$\begin{aligned} 1 - \epsilon \leq P\left(\sup_{0 \leq t \leq 1} B(t) > \delta \right) &= P\left(\sup_{0 \leq t \leq t} \frac{M}{\delta} B(t) \geq M \right) \\ &= P\left(\sup_{0 \leq \leq (M/\delta)^2} \frac{M}{\delta} B\left(t \left(\frac{\delta}{M} \right)^2 \right) \geq M \right) \\ &= P\left(\sup_{0 \leq \leq (M/\delta)^2} \frac{M}{\delta} \tilde{B}(t) \geq M \right) \\ &\leq P\left(\sup_{t \geq 0} \tilde{B}(t) \geq M \right) \end{aligned}$$

$2 \Rightarrow 3$ by deterministic □

- $t \mapsto B(t+s) - B(s)$ is a Brownian motion, independent of $\mathcal{F}_s$
- $\limsup_{t \to 0^+} \frac{B(t)}{\sqrt{t}} = \infty$ a.s. $\liminf_{t \to 0^-} \frac{B(t)}{\sqrt{t}} = -\infty$ a.s. Brownian motion is nowhere differentiable.

Let $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t)$. $T \in [0, \infty]$ os a stopping time if $\forall t \geq 0$, $\{T \leq t\} \in \mathcal{F}_t$. $\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}$.

Example: $T_a = \inf\{t \geq 0 : B(t) = a\}$ is a stopping time.

$$\{T_a \leq t\} = \bigcap_{n=1}^{\infty} \left\{ \inf_{s\in[0,t]} |B_s - a| \leq \frac{1}{n} \right\} = \bigcap_{n=1}^{\infty} \bigcup_{s\in[0,t]\cap\mathcal{Q}} \left\{ |B_s - a| \leq \frac{1}{n} \right\} \in \mathcal{F}_t$$

Example: $T_a = \inf\{t \geq 0 : B(t) > a\}$ is not a stopping time w.r.t. $\mathcal{F}_t$, but it is a stopping time w.r.t. $\mathcal{F}_{t^+} = \bigcap_{s>t} \mathcal{F}_s$.

Theorem: 6.2: Brownian Motion Strong Markov Property

Let T be a stopping time s.t. $T < \infty$ a.s. Then $B^{(T)}(s) = 1_{\{T<\infty\}}(B(T+s) - B(T))$ is a Brownian motion independent of $\mathcal{F}_T$.

Proof. Fix $A \in \mathcal{F}_T$, $t_1 < \cdots < t_p$, $g : \mathbb{R}^p \to \mathbb{R}$ is continuous and bounded.

$$E\left[1_A g(B^{(T)}(t_1), ...B^{(T)}(t_p))\right] = E[1_A]E[g(B(t_1), ..., B(t_p))]$$

Consider $p = 1$, set $T_n = \left\{ \frac{k}{2^n} : T \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) \right.$, $g(B^{(T)}(t)) = \lim_{n\to\infty} g(B^{(T_n)}(t))$, so

$$\begin{aligned}
E\left[1_A g(B^T)(t)\right] &= \lim_{n\to\infty} E\left[1_A g(B^{T_n}(t))\right] \\
&= \lim_{n\to\infty} \sum_{k=1}^{\infty} E\left[1_{A\cap\{\frac{k-1}{2^n}\leq T<\frac{k}{2^n}\}} g\left(B\left(\frac{k}{2^n}+t\right) - B\left(\frac{k}{2^n}\right)\right)\right] \\
&= \lim_{n\to\infty} \sum_{k=1}^{\infty} E[g(B(t))]P\left(A \cap \left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\}\right) \\
&= E[g(B(t))]P(A)
\end{aligned}$$

Because $A \cap \left\{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}\right\} \in \mathcal{F}_{k/2^n}$ □

Corollary 7. $\{t \geq 0 : B(t) = 0\}$ *has no isolated point.*

Theorem: 6.3: Reflection Principle

Fix $t > 0$, $S_t = \sup_{s\leq t} B_s$. Let $a \geq 0$, $b \in (-\infty, a]$. $P(S_t \geq a, B_t \leq b) = P(B_t \geq 2a - b)$.

Proof. Let $T_a = \inf\{t \geq 0 : B(t) = a\}$, $\{S_t \geq a\} = \{T_a \leq t\}$.

$$\begin{aligned}
P(S_t \geq a, B_t \leq b) &= P(T_a \leq t, B_t \leq b) \\
&= P(T_a \leq t, B(t) - B(T_a) \leq b - a) \\
&= P(T_a \leq t, \tilde{B}(t - T_a) \leq b - a) \\
&= P(T_a \leq t, -\tilde{B}(t - T_a) \leq b - a) \\
&= P(T_a \leq t, -B(t) + B(T_a) \leq b - a) \\
&= P(T_a \leq t, B(t) \geq 2a - b) = P(B(t) \geq 2a - b)
\end{aligned}$$

where $\tilde{B}(s) = B^{(T_a)}(s)$, $\tilde{B}$ is a Brownian motion independent of T_a because $T_a \in \mathcal{F}_{T_a}$. □

After hitting a, reflecting the Brownian motion around a gives another Brownian motion with reflected probability.

Let $S = C(\mathbb{R}, \mathbb{R})$ be the set of continuous functions, define the σ-algebra, $\mathcal{S} = \sigma(\omega(t), t \geq 0)$, where $\omega : S \to \mathbb{R}$ are Borel measurable functions as coordinate maps.

Let $(\Omega, \mathcal{F}, P)$ be a measure space that supports Brownian motion, $\omega \mapsto B(\cdot, \omega)$ is measurable.

Define the measure on $(S, \mathcal{S})$ given by $\mu(A) = P(B \in A)$. This is Wiener measure.

Consider Theorem 6.3 again,

$$P(T_a \leq t, B^{T_a}(t - T_a) \leq b - a) = P((T, \tilde{B}) \in H),$$

where T is an r.v. taking values in $\mathbb{R}_+$, $\tilde{B} \in S$, P is the law of stopping time T_as$\times$Wiener measure. $H = \{(s, \omega) \in \mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R}), \omega(t - s) \leq b - a\} \subset \mathbb{R}_+ \times S$.

Since $\tilde{B} = -\tilde{B}$ in distribution, $P((T, \tilde{B}) \in H) = P((T, -\tilde{B}) \in H)$.

***Definition:* 6.2: Continuous Martingale**

Given a stochastic process $(M_t)_{t \geq 0}$ and a filtration $(\mathcal{F}_t)_{t \geq 0}$. M_t is a continuous martingale if

1. $t \mapsto M_t$ is continuous
2. $M_t \in \mathcal{F}_t$
3. $E[|M_t|] < \infty$
4. $E[M_t | \mathcal{F}_s] = M_s$ a.s.

Example: B_t, $\mathcal{F}_t = \sigma(B_u : 0 \leq u \leq t)$. $E[B_t|\mathcal{F}_s] = E[B_t - B_s|\mathcal{F}_s] + B_s = B_s$

Example: $B_t^2 - t$ is a martingale.

Example: $Z_t = e^{\lambda B_t - \frac{\lambda^2}{2} t}$ is a martingale by Laplace transform.

***Theorem:* 6.4: Continuous Martingale Optional Stopping**

Let $S < T$ be bounded stopping time, M_t be a continuous martingale. Then $E[M_T|\mathcal{F}_S] = M_S$

Proof. Discretize the process. Fix N, define $X_n^{(N)} = M_{\frac{n}{2^N}}$, $\mathcal{G}_n^{(N)} = \mathcal{F}_{\frac{n}{2^N}}$. Then $(X_n^{(N)}, G_n^{(N)})$ is a discrete time martingale.

Let $T^{(N)} = k$ if $T \in \left[\frac{k-1}{2^N}, \frac{k}{2^N}\right)$, $S^{(N)} = k$ if $S \in \left[\frac{k-1}{2^N}, \frac{k}{2^N}\right)$.

By Theorem 4.22,$E[X_{T^{(N)}}^{(N)} | \mathcal{G}_{S^{(N)}}^{(N)}] = X_{S^{(N)}}^{(N)}$.

Let $A \in \mathcal{F}_S \subset \mathcal{F}_{\frac{S(N)}{2^N}} \subset \mathcal{G}_{S^{(N)}}^N$. We then get $E[1_A X_{T^{(N)}}^{((N))}] = E[1_A X_{S^{(N)}}^{((N))}]$.

Since $\frac{T^{(N)}}{2^N} \to T$ and $\frac{S^{(N)}}{2^N} \to S$, $X_{T^{(N)}}^{(N)} = M_{\frac{T^{(N)}}{2^N}} \to M_T$, $X_{S^{(N)}}^{(N)} \to M_S$ a.s.

If X is fixed, $E[X|\mathcal{G}]$ as $\mathcal{G}$ varies are uniformly integrable.

$$\lim_{N \to \infty} E[1_A X_{T(A)}^{(N)}] = E[\lim_{N \to \infty} X_{T^{(N)}}^{(N)} 1_A] = E[1_A M_T]$$

□

Example: Let $a < 0 < b$, $T = T_a \wedge T_b$. $E[B_{T\wedge t}] = 0$. As $t \to \infty$,

$$E[B_{T\wedge t}] = E[B_T] = aP(T_a < T_b) + bP(T_b < T_a) = 0$$

Therefore, $P(T_a < T_b) = \frac{b}{b-a}$.

6.1 Brownian Motion as Markov Chains

By definition $B(t) - B(s) \sim \mathcal{N}(0, t-s)$ independent of $B(s)$.

$$\begin{aligned} P(B(t) \in A|\mathcal{F}_s) &= \int_A e^{-\frac{-(x-B(s))^2}{2(t-s)}} \frac{dx}{\sqrt{2\pi}(t-s)} \\ &= p(t-s, B(s), A) \end{aligned}$$

$p : \mathbb{R}_+ \times \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+$ is transition kernel, mapping from time, state space, and σ-algebra of state space to probability.

If transition kernel satisfy the property $P(B(t) \in A|\mathcal{F}_s) = p(t-s, B(s), A)$, with p measurable, then it is a Markov chain, but we consider Brownian motion only here, and it is a Markov chain. The continuous Markov chain is related to PDEs.

Example: The transition kernel $p_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$ is associated with heat equation:

$$\partial_t p_t = \frac{1}{2}\partial_x^2 p_t(x,y)$$

Example: $Y_t = |B_t|$ is a Markov process w.r.t. $\mathcal{F}_t = \sigma(Y_u : 0 \le u \le t)$.

$$P(Y_t \in A|\mathcal{F}_s) = P\left(|(B_t - B_s) + B_s| \in A|\mathcal{F}_s\right)$$

Condition on B_s

$$\begin{aligned} &P\left(|(B_t - B_s) + B_s| \in A|\mathcal{F}_s\right) = P\left(\left|\sqrt{t-s}Z + |B_s|\right| \in A\right) \\ &P\left(|(B_t - B_s) + B_s| \in A|\mathcal{F}_s\right) = P(\left|\tilde{Z}\right| \in A), \end{aligned}$$

for $\tilde{Z} \in N(Y_s, t-s)$.

***Theorem:* 6.5: Levy**

Let $S_t = \sup_{0\le s\le t} B_s$, $t \mapsto S_t - B_t \sim Y_t = |B_t|$.

Remark 13. This is a different version of Theorem 6.3.

***Theorem:* 6.6:**

Let $f : \mathbb{R} \to \mathbb{R}$ be a nice function. Then

$$X(t) = f(B(t)) - \frac{1}{2}\int_0^t (\partial_x^2 f)(B(u))du$$

is a martingale.

Proof.

$$\begin{aligned}E[X(t)|\mathcal{F}_s] &= E[f(B(t))|\mathcal{F}_s] - \frac{1}{2}\int_s^t E[(\partial_x^2 f)(B(u))|\mathcal{F}_s]du - \frac{1}{2}\int_0^s (\partial_x^2 f)(B(u))du \\ &= E_{B(s)}[f(\tilde{B}(t-s))] - \frac{1}{2}\int_0^{t-s} E_{B(s)}[\partial_x^2 f(B(u))]du - \frac{1}{2}\int_0^s (\partial_x^2 f)(B(u))du\end{aligned}$$

We can compute the second term using heat kernel.

$$\begin{aligned}\frac{1}{2}E_y[\partial_x^2 f(B_u)] = \frac{1}{2}\int \partial_x^2 f(x)p_u(x,y)dx &= \frac{1}{2}\int f(x)\partial_x^2 p_u(x,y)dx \\ &= \int f(x)\partial_u p_u(x,y)dx = \partial_u \int f(x)p_u(x,y)dx \\ &= \partial_u E_y[f(B(u))]\end{aligned}$$

Replace E_y with $E_{B(s)}$, we have

$$\begin{aligned}-\frac{1}{2}\int_0^{t-s} E_{B(s)}[\partial_x^2 f(B(u))]du &= -\int_0^{t-s} \partial_u E_{B(s)}[f(\tilde{B}(u))]du \\ &= -E_{B(s)}[f(\tilde{B}(t-s))] + E_{B(s)}[f(\tilde{B}(0))] \\ &= -E_{B(s)}[f(\tilde{B}(t-s))] + f(B(s))\end{aligned}$$

Therefore,

$$E[X(t)|\mathcal{F}_s] = -\frac{1}{2}\int_0^s \partial_x^2 f(B(u))du + f(B(s)) = X(s)$$

$X(t)$ is a martingale. □

Remark 14. This is related to Ito's formula.

***Definition:* 6.3: Binary Splitting**

A martingale X_n is called binary splitting if each X_n takes only finitely many values. Condiitonal on $X_0 = x_0, ..., X_n = x_n$, X_{n+1} takes only two possible values.

Lemma 11. *Given X with $E[X^2] < \infty$, there exists a binary splitting X_n that converges to X a.s. and in L^2*

Proof. Let $X_0 = E[X]$, $\zeta_0 = \begin{cases} 1, X \geq X_0 \\ -1, X < X_0 \end{cases}$, $\mathcal{G}_1 = \sigma(\zeta_0)$.

Define $X_1 = E[X|\mathcal{G}_1] = \begin{cases} E[X1_{\{X \geq E[X]\}}]/P(X \geq E[X]), \zeta_0 = 1 \\ E[X1_{\{X < E[X]\}}]/P(X < E[X]), \zeta_0 = -1 \end{cases}$

Then for $n \geq 1$, recursively define $\mathcal{G}_n = \sigma(\zeta_0, ..., \zeta_{n-1})$, and $X_n = E[X|\mathcal{G}_n]$.

By Jensen's inequality, $E[X_n^2] \leq E[X^2] < \infty$, so X_n is an L^2-bounded martingale. $\lim_{n\to\infty} X_n = X_\infty$ a.s. and in L^2.

Now we show that $X_\infty = X$ by showing $\lim_{n\to\infty} \zeta_n(X - X_{n+1}) = |X - X_\infty|$.

Consider pointwise, if $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega) = X(\omega)$, then fine.

If $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega) > X(\omega)$, then for sufficiently large n, $\zeta_n = -1$,

$$\zeta_n(X - X_{n+1}) = X_{n+1}(\omega) - X(\omega) \to |X_\infty(\omega) - X(\omega)|$$

The case $X_\infty(\omega) < X(\omega)$ is the same.

$$\begin{aligned} E[\zeta(X - X_{n+1})] &= E[\zeta(X - E[X|\mathcal{G}_{n+1}])] = E[\zeta_n X] - E[\zeta_n E[X|\mathcal{G}_{n+1}]] \\ &= E[\zeta_n X] - E[E[\zeta_n X|\mathcal{G}_{n+1}]] = 0 \end{aligned}$$

Let $Y_n = \zeta_n(X - X_{n+1})$. Then $Y_n \to |X - X_\infty|$ a.s. and L^2-bounded. Therefore, $E[|X - X_\infty|] = E[\lim_{n\to\infty} Y_n] = \lim_{n\to\infty} E[Y_n] = 0$.

X_n is a binary splitting, because conditional on $\{X_1 = x_1, ..., X_n = x_n\}$, X_{n+1} can take only two different values, based on ζ_n. □

Theorem: 6.7: Skorokhod Embedding

Let X be any random variable s.t. $E[X] = 0, E[X^2] < \infty$. Let B_t be a Brownian motion. There exists a stopping time T s.t. $B(T) \sim X$ and $E[T] = E[X^2]$.
Any mean zero finite variance r.v. X can be found in a Brownian motion.

Proof. We construct an increasing sequence T_k s.t. $\forall n$, $(B(T_k))_{k=1}^n \sim (X_k)_{k=1}^n$, where X_k is the binary splitting converging to X.

Let $X_0 = E[X] = 0$, X_1 takes two values $a < 0 < b$, $E[X_1] = 0$. Let $T_1 = \inf\{B(t) \in \{a, b\}\}$, so $B(T_1) \sim X_1$.

Assume T_k, $k = 1, ..., n$ are constructed so that $(B(T_k))_{k=1}^n \sim (X_k)_{k=1}^n$.

Condition on $B(T_1) = x_1, ..., B(T_n) = x_n$, X_{n+1} takes two values a, b, and $E[X_{n+1}|\sigma(X_1, ..., X_n)] = X_n = x_n = B(T_n)$. Then $X_{n+1} - X_n$ is centered at 0. Then

$$T_{n+1} = \inf\{t > T_n : B_t - x_n \in \{a - x_n, b - x_n\}\}$$

Conditional on $B(T_1) = x_1, ..., B(T_n) = x_n$, we have $B(T_{n+1}) - B(T_n) \sim X_{n+1} - x_n$, so $B(T_{n+1}) \sim X_{n+1}$.

Since $\lim_{n\to\infty} T_n = T$, we have $B(T) = \lim_{n\to\infty} B(T_n)$ and $B(T) \sim X$.

By Theorem 1.5 and 6.4 applied to $M_t = B_t^2 - t$, we have:

$$E[T] = \lim_{n\to\infty} E[T_n] = \lim_{n\to\infty} E[B(T_n)^2] = \lim_{n\to\infty} E[X_n^2] = E[X^2],$$

since X converges in L^2. □

Donsker's invariance scaling: let Y_n be i.i.d. centered r.v.s s.t. $E[Y_n^2] = 1$, $S_n = \sum_{k=1}^n Y_k$. Define $S(t)$ the linear interpolation of S_n, $S_n^*(t) = \frac{S(nt)}{\sqrt{n}}$.

There exists a sequence of stopping times $T_1 < T_2 < \cdots < T_n$ s.t. $B(T_n) \sim \sum_{k=1}^n Y_k$.

Proof. By Theorem 6.7, we can construct T_1 s.t. $B(T_1) \sim Y_1$, T_2 s.t. $B(T_2) - B(T_1) \sim Y_2$.

By Theorem 5.6, $(T_n - T_{n-1})$ are i.i.d. with the same distribution as T_1. $\lim_{n\to\infty} \frac{T_n}{n} = E[T_1] = 1$. □

Construct $S_n^*(t)$ from $S_n = B(T_n)$, $\lim_{n\to\infty} P\left(\sup_{0\le t\le 1}\left|\frac{B(nt)}{\sqrt{n}} - S_n^*(t)\right| > \epsilon\right) = 0$.

Proof. Define $W(t) = \frac{B(nt)}{\sqrt{n}}$. Let $k = k(t)$ be the index s.t. $\frac{k-1}{n} \leq t < \frac{k}{n}$.
Let $A_n = \left\{\sup_{0\leq t\leq 1} \left|\frac{B(nt)}{\sqrt{n}} - S_n^*(t)\right| > \epsilon\right\}$

$$\begin{aligned}
A_n &\subset \left\{\exists t \in [0,1] : \left|\frac{S_k}{\sqrt{n}} - W_n(t)\right| > \epsilon, \left|\frac{S_k}{\sqrt{n}} - W_n(t)\right| > \epsilon\right\} \\
&\subset \{\exists s,t \in [0,2] : |s-t| < \delta \text{ and } |W_n(s) - W_n(t)| > \epsilon\} := B_n(s) \\
&\bigcup \left\{\exists t : \left|\frac{T_k}{n} - t\right| > \delta \text{ or } \left|\frac{T_{k-1}}{n} - t\right| > \delta\right\} := C_n(\delta)
\end{aligned}$$

$P(B_n(s)) = P(B_1(s)) \to 0$ as $s \to 0$. On the other hand, $\lim_{n\to\infty} P(C_n(\delta)) = 0$. By Theorem 2.3, $\frac{T_n}{n} \to 1$. □

Theorem: 6.8: Donsker's Theorem

Consider the space $C([0,1])$ with topology induced by $\|\cdot\|_\infty$, $S_n^* \to B$ in distribution. *i.e.* for all continuous functionals $F : C([0,1]) \to \mathbb{R}$, $\lim_{n\to\infty} E[F(S_n^*)] = E[F(B)]$.

Proof. By Lemma 1.1, it suffices to show for closed set $K \subset C([0,1])$ that $\limsup_{n\to\infty}(S_n^* \in K) \leq P(B \in K)$.

For $\epsilon > 0$, define $K_\epsilon = \{f \in C([0,1]) : \|f - g\|_\infty \leq \epsilon \text{ for some} g \in K\}$.

$$P(S_n^* \in K) \leq P(W_n \in K_\epsilon) + P(\|W_n - S_n^*\| > \epsilon)$$

Therefore by the previous proofs.

$$\begin{aligned}
\limsup_{n\to\infty} P(S_n^* \in K) &\leq P(B \in K_\epsilon) \\
&= \lim_{\epsilon\to 0} P(B \in K_\epsilon) = P\left(\bigcap_{n=1}^\infty B \in K_{1/n}\right) = P\left(B \in \bigcap_{n=1}^\infty K_{1/n}\right) \\
&= P(B \in K)
\end{aligned}$$

□

6.2 Brownian Motion in Higher Dimensions and Harmonic Functions

Let U be a bounded open connected set in $\mathbb{R}^d$. Given $\phi : \partial U \to \mathbb{R}$ continuous, $u(x)$ solves Dirichlet problem in U if $\Delta u(x) = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} u(x) = 0$ for $x \in U$ and $u(x) \to \phi(z)$ if $x \to z$.

Definition: 6.4: Harmonic

u is harmonic in U if $\Delta u(x) = 0$ for all $x \in U$.

Theorem: 6.9:

Let $u : U \to \mathbb{R}$ be measurable and locally bounded. Then the following are equivalent:
1. u is harmonic
2. $\forall \overline{B_r(x)} \subset U$, $u(x) = \frac{1}{|B_r(x)|}\int_{B_r(x)} u(y)dy$
3. $\forall \overline{B_r(x)} \subset U$, $u(x) = \frac{1}{|\partial B_r(x)|}\int_{\partial B_r(x)} u(y)d\sigma(y)$

Theorem: 6.10: Maximum Principle

1. If u is harmonic on U, then if it attains maximum in U, then u is constant.
2. If U extends to a continuous function on $\bar{U}$, then $\max_{x \in \bar{U}} u(x) = \max_{x \in \partial U} u(x)$
3. If u, v are two harmonic functions extended continuously to $\bar{U}$ s.t. $u = v$ on ∂U, then $u = v$ on U.

Definition: 6.5: Brownian Motion in $\mathbb{R}^d$

Let $W_1(t), ..., W_d(t)$ be standard independent Brownian motions. Then Brownian motion in $\mathbb{R}^d$ started at $x \in \mathbb{R}^d$ is $W(t) = (W_1(t), ..., W_d(t)) + x$.

All properties from one-dimensional Brownian motion still hold.

Let $\tau = \inf\{t > 0 : W(t) \in \partial U\}$ be a stopping time, $\phi : \partial U \to \mathbb{R}$ be continuous. Then $u(x) = E_x[\phi(W(\tau))]$.

Proof. Let $B_r\bar{(}x) \subset U$, $\hat{\tau} = \inf\{t > 0 : W(t) \in \partial B_r(x)\}$.

$$\begin{aligned} u(x) &= E_x[\phi(W(\tau))] = E_x[E_x[\phi(W(\tau))|\mathcal{F}_{\hat{\tau}}]] \\ &= E_x[E_{W(\hat{\tau})}[\phi(W(\tau))]] = E_x[u(W(\hat{\tau}))] \\ &= \frac{1}{|\partial B_r(x)|} \int_{\partial B_r(x)} u(y) d\sigma(y) \end{aligned}$$

□

Now we want to show $\forall z \in \partial U$, $x \in U$, $\lim_{x \to z} u(x) = \phi(z)$.

Definition: 6.6: Exterior Cone Condition

u satisfies exterior cone condition if $\forall z \in U$, there exists a cone $C_z(\alpha)$ s.t. $C_z(\alpha) \in U^C$ at least locally.

1. For all $\alpha > 0$, there exists $0 < a(\alpha) < 1$ s.t. $\forall x$ with $|x| \leq \frac{1}{4}$, we have $P_x(\tau(\partial B_1(0)) < \tau(C_0(\alpha))) \leq a$.
2. For all $k \geq 1$, all x with $|x| \leq 4 {-} k$, we have $P_x(\tau(\partial B_1(0)) < \tau(C_0(\alpha))) \leq a^k$.

Proof. If $|x| \leq 4^{-2}$, we have, $\tau(\partial B_{1/4}(0)) < \tau(\partial B_1(0)) < \tau(C_0(\alpha))$. Then recursively, if $|x| \leq 4^{-k}$, we have

$$P_x(\tau(\partial B_1(0)) < \tau(C_0(\alpha))) \leq \prod_{j=1}^{k} \sup_{x \in B_{4^{-k}}(0)} P(\tau(\partial B_{4^{-k+j+1}}(0)) < \tau(C_0(\alpha))) \leq a^k$$

□

Theorem: 6.11:

Let u obey exterior cone condition, $\phi : \partial U \to \mathbb{R}$ be continuous, then $u(x) = E_x[\phi(B(\tau))]$ solves Dirichlet problem.

Proof. Let $z \in \partial U$ and $\epsilon > 0$. Since ϕ is continuous, we have $\exists \delta > 0$, $|\phi(z) - \phi(x)| < \epsilon$ for all $|x - z| < \delta$, $x \in \partial U$. Let $C_z(\alpha)$ be exterior cone at z.

By previous two points, there exists n s.t. $\forall |x-z|<n$, $x \in U$ s.t. $P_x(\tau(\partial B_\delta(z)) < \tau(C_\alpha(z))) < \epsilon$. For such z,

$$\begin{aligned}
|u(x)-\phi(z)| &\leq E_x\left[|\phi(W(\tau))-\phi(z)|\right] \\
&= E_x\left[|\phi(W(\tau))-\phi(z)|\, 1_{\{\tau(\partial B_\delta(z))<\tau(C_z(\alpha))\}}\right] \\
&+ E_x\left[|\phi(W(z))-\phi(z)|\, 1_{\{\tau(C_z(\alpha))\leq\tau(\partial B_\delta(z))\}}\right] \\
&\leq 2\left\|\phi\right\|_\infty \epsilon + \epsilon
\end{aligned}$$

□

Let $A = \{r < |x| < R\}$, $u(x) = \begin{cases} \log|x|, d=2 \\ |x|^{2-d}, d\neq 2 \end{cases}$. Consider $T = \inf\{T_r, T_R\}$, where $T_r = \inf\{t : |W(t)| = r\}$, $T_R = \inf\{t : |W(t)| = R\}$. $\phi(x) : \partial A \to \mathbb{R}$ with $\phi(x) = u(x)$ on ∂A. Then $\forall x \in A$,

$$u(x) = E_x[\phi(W(T))] = u(r)P_x(T_r < T_R) + u(R)(1 - P_x(T_r < T_R))$$

This gives

$$P_x(T_r < T_R) = \frac{u(R)-u(x)}{u(R)-u(r)} = \begin{cases} \frac{\log R - \log|x|}{\log R - \log r}, d=2 \\ \frac{R^{2-d}-|x|^{2-d}}{R^{2-d}-r^{2-d}}, d\neq 2 \end{cases}$$

Take $R \to \infty$, $P_x(T_r < \infty) = \begin{cases} 1, d \leq 2 \\ \frac{r^{d-2}}{|x|^{d-2}}, d\geq 3 \end{cases}$.

If $n = 2$, Brownian motion is neighborhood recurrent.

i.e. for all open sets U, $P_x(\{t : W(t) \in U \text{ is unbounded}\}) = 1$,
but for all $y \neq x$, $P_x(\exists t : W(t) = y) = 0$, $d \geq 2$.

In $d \geq 3$, Brownian motion is transient, $P_0(\lim_{t\to\infty} |W(t)| = \infty) = 1$.

$$P_0\left(|B(t) \leq n| \text{ for some } t \geq \tau(\partial B_{n^3}(0))\right) \leq \frac{n^{d-2}}{(n^3)^{d-2}} = \frac{1}{n^{2(d-2)}}$$

7 Continuous Time Stochastic Processes

We start with Einstein's construction of Brownian motion. Let $B_t \in \mathbb{R}^3$ be the position of a particle at time t in $\mathbb{R}^3$. It should satisfy some evoluting probability density,

$$f(t,x)dx = P(B_t \in dx) = \lim_{h\to 0} \frac{P(B_t \in [x, x+h])}{h^3}, \quad i.e. \int_A f(t,x)dx = P(B_t \in A)$$

The transition probability: $P(s,x,t,y)dy = P(B_t \in dy | B_s = x)$, equivalently, $P(s,x,t,y) = P(B_t \in A | B_s \in x)$.

Assumption: Time and space homogeneity. $P(s,x,t,y) = P(t-s, y-x)$. It does not matter where the particle is in time and space. The probability density should not change.

$$\begin{aligned} f(t+\tau, x) &= \int f(t, x-y)p(\tau, y)dy \\ &= \int (f(t-x) - y\nabla f(t,x) + \frac{1}{2}y^2 D^2 f(t,x) + \cdots)p(\tau,y)dy \\ &= f(t,x)\int p(\tau,y)dy - \sum_i \partial_i f(t,x) \int y_i p(\tau,y)dy \\ &+ \frac{1}{2}\sum_{i,j}(\partial_i\partial_j f)(t,x)\int y_i y_j p(\tau,y)dy + \cdots \end{aligned}$$

Note that:

- $\int p(\tau,y)dy = 1$ as probability density
- $\int y_i p(\tau,y)dy = 0$ by antisymmetry (expectation)
- $\int y_i y_j p(\tau,y)dy = 0$ if $y_i \neq y_j$ by antisymmetry
- $\int y_i^2 p(\tau,y)dy$ is like sum of mean-zero i.i.d. r.v.s, $Var(X_1 + \cdots + X_n) = CN$, $\int y_i^2 p(\tau,y)dy \sim D\tau$, D is the diffusion coefficient.

Then, it gives the heat equation with solution being Gaussian:

$$\frac{d}{dt}f(t,x) = \frac{D}{2}\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f(t,x)$$

In 1D case, $d=1, D=1$, the solution is $p(t,x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$

For a stochastic process X_t, $t \in T = \mathbb{R}_+ = [0,\infty)$, we need to know the probability distributions $F_{t_1\ldots t_n}(X_{t_1} \le x_1, ..., X_{t_n} \le x_n)$, assuming $t_1 \le t_2 \le \cdots \le t_n$ ordered.

To handle the case $X_i = \infty$, we need to have a consistent family of finite dimensional distributions.

Theorem: 7.1: Kolmogorov Extension Theorem

There exists a unique probability measure μ on $\mathbb{R}^{[0,\infty)}$, $\mathcal{B}^{[0,\infty)}$ s.t. $\mu(\{X_t : X_{t_1} \le x_1, ..., X_{t_n} \le x_n\}) = F_{t_1\ldots t_n}(x_1, ..., x_n)$. $\mathcal{B}^{[0,\infty)}$ is the product σ-algebra (smallest σ-algebra containing the cylinder sets $\{X_t : X_{t_1} \le x_1, ..., X_{t_n} \le x_n\}$)

A **Markov process** is where we can decompose $P(X_{t_1} \in dy_1, ..., X_{t_n} \in dy_n)$ into each individual time intervals.

$$\begin{aligned}
&P(X_{t_1} \in dy_1, ..., X_{t_n} \in dy_n) \\
&= P(t_1, y_1)dy_1 P(t_2 - t_1, y_2 - y_1)dy_2 \cdots P(t_n - t_{n-1}, y_n - y_{n-1})dy_n \\
&\int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} P(X_{t_1} \in dy_1, ..., X_{t_n} \in dy_n) \\
&= \int_{-\infty}^{x_n} dy_n \cdots \int_{-\infty}^{x_1} dy_1 \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{y_1^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(y_2-y_1)^2}{2(t_2-t_1)}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(y_n - y_{n-1})^2}{2(t_n - t_{n-1})}}
\end{aligned}$$

A **Gaussian process** $X_t = (X_{t_1}, ..., X_{t_n})$ is a n-dimensional Gaussian for all $t_1, ..., t_n$. $a_1 X_{t_1} + \cdots + a_n X_{t_n}$ is Gaussian for all $a \in \mathbb{R}^n$. $E[X_t] = m_t$, $Cov(X_t, X_s) = c_{t,s}$.

For a **Brownian motion** B_t, $m_t = E[B_t] = 0$. $E[B_t B_s] = E[B_s^2] + E[(B_t - B_s)(B_s - B_0)] = s$ by independent increments, if $s < t$.

Let B_t be a Brownian motion, $u > 0$. The following are also Brownian motions

1. $\tilde{B}_t = B_{u+t} - B_u$
2. Let $\alpha > 0$, $\tilde{B}_t = \alpha B_{\alpha^{-2}t}$
3. $\tilde{B}_t = tB_{\frac{1}{t}}$

Proof. Firstly, all $E[\tilde{B}_t] = 0$, so we only need to consider the covariance. Assume $s < t$.

1. $E[\tilde{B}_t \tilde{B}_s] = E[(B_{u+t} - B_u)(B_{u+s} - B_u)] = s + u - u - u + u = 0$
2. $E[\alpha B_{\alpha^{-2}t} \alpha B_{\alpha^{-2}s}] = \alpha^2 \alpha^{-2} s = s$
3. $E\left[tB_{\frac{1}{t}} sB_{\frac{1}{s}}\right] = ts\frac{1}{t} = s$. □

Theorem: 7.2:

Brownian motion is continuous

Proof. Let $t \in [0, 1)$, define $B_n(t) = \begin{cases} B\left(\frac{i}{2^n}\right), t = \frac{i}{2^n}, i = 0, ..., 2^n \\ \text{linear in between} \end{cases}$

We want $\sum_{n=1}^{\infty} P\left(\sum_{m=n}^{\infty} \sup_{0 \le t \le 1} |B_{m+1}(t) - B_m(t)| > \epsilon\right) < \infty$, because Lemma 2.5 says if the probability of A_n is summable, then A_n cannot happen infinitely often.

Consider the Haar functions: $H_{n,j}(x) = 2^{\frac{n}{2}}\left(1_{[\frac{2j}{2^{n+1}}, \frac{2j+1}{2^{n+1}})} - 1_{[\frac{2j+1}{2^{n+1}}, \frac{2j+2}{2^{n+1}})}\right)$.
The Haar functions are orthonormal basis of $L^2[0, 1)$, $\int_0^1 H_{nj} H_{nj} = 1$, $\int_0^1 H_{nj} H_{\tilde{n}\tilde{j}} = 0$.

Let $S_{n,j}(x) = \int_0^x H_{n,j}(y)dy$, $B_{n+1}(t) - B_n(t) = \sum_{j=0}^{2^n-1} \xi_{n,j} S_{n,j}(t)$, where

$$\xi_{n,j} = 2^{\frac{n}{2}+1}\left(B\left(\frac{2j+1}{2^{n+1}}\right) - \frac{B\left(\frac{j+1}{2^{n+1}}\right) + B\left(\frac{j}{2^{n+1}}\right)}{2}\right) = -\int_0^1 \dot{H}_{n,j} B(t)dt = \int_0^1 H_{n,j}(t)\dot{B}(t)dt$$

Since $E[B(t)B(\tilde{t})] = t \wedge \tilde{t}$, $\frac{d}{d\tilde{t}}\frac{d}{dt} t \wedge \tilde{t} = \frac{d}{d\tilde{t}} 1_{t \leq \tilde{t}} = \delta_0(t - \tilde{t})$, then

$$\begin{aligned} E[\xi_{nj}\xi_{\tilde{n}\tilde{j}}] &= \int_0^1 \int_0^1 \dot{H}_{n,j}(t)\dot{H}_{\tilde{n},\tilde{j}}(\tilde{t}) E[B(t)B(\tilde{t})] dt d\tilde{t} \\ &= \int_0^1 \int_0^1 H_{n,j}(t) H_{\tilde{n},\tilde{j}}(\tilde{t}) \delta_0(t - \tilde{t}) dt d\tilde{t} \\ &= \int_0^1 H_{n,j}(t) H_{\tilde{n},\tilde{j}}(t) dt = \begin{cases} 1, n, j = \tilde{n}, \tilde{j} \\ 0 \end{cases} \end{aligned}$$

So $\xi_{n,j}$ are i.i.d. $\mathcal{N}(0,1)$. This is invariant under change of basis.

Suppose e_m are ONBs of $L^2[0,1)$, f_n are another, $\langle e_n, \xi \rangle$ i.i.d. $\mathcal{N}(0,1)$.

$$f_n = \sum_n a_{nm} e_m$$

$$\langle f_n, f_{\tilde{n}} \rangle = \sum_{m,\tilde{m}} a_{nm} a_{\tilde{n}\tilde{m}} \langle e_m, e_{\tilde{m}} \rangle = \sum_m a_{nm} a_{\tilde{n}m} = (AA^T)_{n\tilde{n}} = I_{n\tilde{n}}$$

$$E[\langle f_n, \xi \rangle \langle f_{\tilde{n}}, \xi \rangle] = \langle f_n, f_{\tilde{n}} \rangle = I,$$

so $\langle f_n, \xi \rangle$ are i.i.d. $\mathcal{N}(0,1)$.

Back to the major claim on probability

$$\|B_{m+1}(t) - B_m(t)\|_\infty = 2^{-\frac{m}{2}-1} \max_j |\xi_{m,j}|$$

$$P\left(-2^{-\frac{m}{2}-1} \max_j |\xi_{m,j}| > 2^{-\frac{m}{4}}\right) \leq 2^m P\left(|\xi_{m,j}| > 2^{\frac{m}{4}} 2\right) \leq 2^m e^{-2^{\frac{m}{4}+1}},$$

by Theorem 2.1, and the approximation of gaussian tail distribution.

$\sum_{m=n}^\infty 2^{-m/4} < \epsilon$ for n large enough.

$$\sum_{n=1}^\infty P\left(\sum_{m=n}^\infty \|B_{m+1} - B_m\|_\infty > \epsilon\right) \leq \sum_{n=1}^\infty \sum_{m=n}^\infty P(\|B_{m+1} - B_m\|_\infty > 2^{-m/4}) \leq \sum_{n=1}^\infty \sum_{m=n}^\infty 2^m e^{-2^{m/4}+1} < \infty$$

□

Theorem: 7.3: Kolmogorov Continuity Theorem

Let $X(t)$ be a stochastic process on $[0,1]$, if $E\left[|X(t) - X(s)|^\beta\right] \leq C|t-s|^{1+\alpha}$ for some $\alpha, \beta > 0$, then $X(t)$ is $\frac{\alpha}{\beta}$-Holder with probability 1, by changing $X(s)$ on a set of probability zero.

Proof.

$$E\left[|B(t) - B(s)|^\beta\right] = E\left[|t-s|^{\beta/2} Z^\beta\right] = |t-s|^{\frac{\beta}{2}} E[Z^\beta], \quad Z \sim \mathcal{N}(0,1)$$

So $\alpha = \frac{\beta}{2} - 1$, B is $\frac{\frac{\beta}{2}-1}{\beta}$-Holder. As β increases, we get up to $< \frac{1}{2}$-Holder. □

To see it is not $\frac{1}{2}$-Holder, check $Z_1 = \frac{B(\frac{1}{2}) - B(0)}{(\frac{1}{2})^{1/2}}$, $Z_2 = \frac{B(\frac{3}{4}) - B(\frac{1}{2})}{(\frac{1}{4})^{1/2}}$,... $Z_i \sim \mathcal{N}(0,1)$ i.i.d. $\|B\|_{1/2} \geq \max_i |Z_i| = \infty$.

$$P\left(\max_i Z_i > \lambda\right) = 1 - P\left(\max_i Z_i \leq \lambda\right) = 1 - P(Z_1 \leq \lambda)^n = 1 - (1 - P(Z_1 > \lambda))^n$$

If $\lambda \sim \sqrt{2\log n}$, $P(Z_1 \geq \lambda) \sim \frac{1}{n}$.

Definition: 7.1: Bounded Variation

The total variation of f is

$$\|f\|_{TV,[0,t]} = \sup_{0 \leq t_0 < t_1 < \cdots < t_n \leq t} \sum |f(t_{i+1}) - f(t_i)|$$

f is bounded variation if $\|f\|_{TV,[0,t]} < \infty$.

Theorem: 7.4:

1. f is bounded variation if and only if it is the difference of bounded monotone functions
2. Measures are equivalent to functions of bounded variation, $\mu([0,t]) = f(t)$
3. If f is bounded variation, then f is differentiable a.e.

Definition: 7.2: Quadratic Variation

A stochastic process X_t has quadratic variation $\langle X_t, X_t \rangle$ if $\sum_i \left|X_{t_{i+1}} - X_{t_i}\right|^2 \to \langle X_t, X_t \rangle$ in probability as the meshsize $\max_i |t_{i+1} - t_i| \to 0$.

Brownian motion is of bounded quadratic variation

Proof. Let $y_i = |B(t_{i+1}) - B(t_i)|^2 - (t_{i+1} - t_i)$, we want to show that $\sum_i y_i < \infty$. Consider the variance Since Brownian motions are i.i.d. $E[y_i y_j] = E[y_i]E[y_j] = 0$ if $i \neq j$.

$$\begin{aligned}
E\left[\left(\sum_i y_i\right)^2\right] &= \sum_{i,j} E[y_i y_j] = \sum_i E[y_i^2] \\
&= \sum_i E\left[\left(\left|\sqrt{t_{i+1} - t_i} Z\right|^2 - (t_{i+1} - t_i)\right)^2\right] \\
&= \sum_i (t_{i+1} - t_i)^2 E[(Z^2 - 1)^2] \\
&\leq C \max_i |t_{i+1} - t_i| \sum_i (t_{i+1} - t_i) \\
&= C \max_i |t_{i+1} - t_i| \, t \to 0 \text{ as } \max_i |t_{i+1} - t_i| \to 0
\end{aligned}$$

By Theorem 2.1,

$$P\left(\left|\sum_i |B(t_{i+1}) - B(t_i)|^2 - t\right| > \epsilon\right) \leq \frac{C \max_i |t_{i+1} - t_i| \, t}{\epsilon^2}$$

So $\sum_i |B(t_{i+1}) - B(t_i)|^2 \to t$ in probability as $\max_i |t_{i+1} - t_i| \to 0$. □

But Brownian motion is not of bounded variation nor α-Holder for $\alpha \geq \frac{1}{2}$.

Proof. Bounded variation:

$$\sum |B(t_{i+1}) - B(t_i)|^2 \leq \max |B(t_{i+1}) - B(t_i)| \sum_i |B(t_{i+1}) - B(t_i)|$$

LHS$\to t$, but $\max|B(t_{i+1})-B(t_i)|\to 0$, and it has to be $\sum_i |B(t_{i+1})-B(t_i)|\to\infty$.

Similar for α-Holder for $\alpha>\frac{1}{2}$:

$$\sum |B(t_{i+1})-B(t_i)|^2 \le \left(\max\frac{|B(t_{i+1})-B(t_i)|}{|t_{i+1}-t_i|^\alpha}\right)^2 \sum_i |t_{i+1}-t_i|^{2\alpha}$$

Still LHS$\to t$, but $\sum_i |t_{i+1}-t_i|^{2\alpha}\to 0$. □

Theorem: 7.5:

B_t is nowhere differentiable

Proof. Can show that $\dot{B}_t=\sum_k Z_k e^{2\pi ikt}$ (1933 Paley, Wiener, Zygmund)

Dvoredsky, Erdos, Kac: If B is differentiable at $s\in[0,1]$, then $\exists\epsilon>0$ and l s.t. $|B(s)-B(t)|\le l|t-s|$ for $0<t-s<\epsilon$.

Choose n large enough that $\frac{i-1}{n}<s\le\frac{i}{n}<\frac{i+1}{n}<\frac{i+2}{n}<\frac{i+3}{n}\le s+\epsilon$, and $\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}$ for $j=i+1,i+2,i+3$. Then

$$\bigcup_l\bigcap_m\bigcup_{n\ge m}\bigcup_{0\le i\le n+1}\bigcap_{i+1\le j\le i+3}\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}$$

it must have some point that's differentiable. (the union/intersection just means that there exists some l s.t. for all m, there is sufficiently large m and $0\le i\le n+1$ s.t. for all $i\le j\le i+1$ the event $\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}$ happens)

Note that $\bigcap_m\bigcup_{n\ge m}$ is equivalent to $\liminf_{n\to\infty}$, we want to show that for every l the probability $\to 0$.

$$\begin{aligned}
&P\left[\bigcap_m\bigcup_{n\ge m}\bigcup_{0\le i\le n+1}\bigcap_{i+1\le j\le i+3}\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}\right]\\
&\le\liminf_{n\to\infty}P\left[\bigcup_{0\le i\le n+1}\bigcap_{i+1\le j\le i+3}\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}\right]\\
&\le\liminf_{n\to\infty}\sum_{i=0}^{n+1}P\left[\bigcap_{i+1\le j\le i+3}\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}\right]\\
&\le\liminf_{n\to\infty}(n+2)P\left[\bigcap_{1\le j\le 3}\left\{\left|B\left(\frac{j}{n}\right)-B\left(\frac{j-1}{n}\right)\right|\le\frac{7l}{n}\right\}\right]\\
&=\liminf_{n\to\infty}(n+2)P\left[\left\{\left|B\left(\frac{1}{n}\right)\right|\le\frac{7l}{n}\right\}\right]^3=\liminf_{n\to\infty}(n+2)P\left[\frac{1}{\sqrt{n}}|Z|\le\frac{7l}{n}\right]^3\\
&=\liminf_{n\to\infty}(n+2)P\left[|Z|\le\frac{7l}{\sqrt{n}}\right]^3\le C\frac{n}{n^{3/2}}\to 0
\end{aligned}$$

□

For Brownian motion, its filtrations can be defined by $\mathcal{F}_t=\sigma(B_s,s\le t)$.

B_t is martingale w.r.t. $\mathcal{F}_t$.

$$E[B_t|\mathcal{F}_s] = E[B_t - B_s|\mathcal{F}_s] + E[B_s|\mathcal{F}_s] = 0 + B_s = B_s$$

B_t^2 is submartingale w.r.t. $\mathcal{F}_t$

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

This also shows that $B_t^2 - t$ is a martingale.

$e^{\lambda B_t - \frac{\lambda^2}{2}t}$ is a martingale for $\lambda \in \mathbb{R}$ if and only if B is a Brownian motion.

Proof. We want to show $E\left[e^{\lambda B_t - \frac{\lambda^2}{2}t}|\mathcal{F}_s\right] = e^{\lambda B_s - \frac{\lambda^2}{2}s}$

That is equivalent to $E\left[e^{\lambda(B_t - B_s) - \frac{\lambda^2}{2}(t-s)}|\mathcal{F}_s\right] = 1$ or $E[e^{\lambda(B_t - B_s)}|\mathcal{F}_s] = e^{\frac{\lambda^2}{2}(t-s)}$

This is true because $B_t - B_s$ is independent of $\mathcal{F}_s$ and it's the moment generating function. ☐

This is a generating function, and gives:

$$\begin{aligned} \left.\frac{d}{dt}\right|_{\lambda=0} e^{\lambda B_t - \frac{\lambda^2}{2}t} &= B_t \\ \left.\frac{d^2}{dt^2}\right|_{\lambda=0} e^{\lambda B_t - \frac{\lambda^2}{2}t} &= B_t^2 - t \end{aligned}$$

Theorem: 7.6: Continuous Time Doob's Inequality

Let X_t be a nonnegative submartingale w.r.t. $\mathcal{F}_t$, $t \geq 0$ with right continuous sample paths. Then

$$P\left(\sup_{0 \leq t \leq T} X_t \geq \lambda\right) \leq \frac{X_T}{\lambda}$$

Proof. Let $0 \leq t_0 < t_1 < \cdots < t_n = T$, $\tilde{X}_n = X_{t_n}$, $\tilde{F}_n = \mathcal{F}_{t_n}$.

On the finite grid, Theorem 4.14 give $P\left(\max_k X_{t_k} \geq \lambda\right) \leq \frac{E[X_T]}{\lambda}$. Take limit on partition mesh size and get the continuous version. ☐

A stopping time τ is a r.v. taking values in time $\mathbb{R}_+$ s.t. $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, \forall t \geq 0$.
$\mathcal{F}_\tau = \{A : A \cap \{t \geq \tau\} \in \mathcal{F}_t, \forall t \geq 0\}$ is the stopped σ-field

Theorem: 7.7: Optional Stopping

Let $\tau > \sigma$ be bounded stopping times and M_t a right continuous martingale, then $E[M_\tau|\mathcal{F}_\sigma] = M_\sigma$ a.s.

Proof. Firstly, consider the discrete case, we want $\int_A M_\tau dP = \int_A M_\sigma dP$ for all $A \in \mathcal{F}_\sigma$.

By definition, $A \cap \{\sigma = l\} \cap \{\tau = k\} \in \mathcal{F}_l$, so equivalently:

$$\int_{A\cap\{\sigma=l\}\cap\{\tau=k\}} M_k dP = \int_{A\cap\{\sigma=k\}\cap\{\tau=k\}} M_l dP$$

For LHS, it is $E[M_k|\mathcal{F}_\sigma]$, RHS is M_l. They are bounded so it is valid.

Continuous case: let $\sigma_n = 2^{-n}\left(\lfloor 2^n\sigma\rfloor + 1\right)$, $\tau_n = 2^{-n}\left(\lfloor 2^n\tau\rfloor + 1\right)$. They are still stopping times, and $\sigma \downarrow \sigma$, $\tau_n \downarrow \tau$.

We have from discrete case $\int_A M_{\tau_n} dP = \int_A M_{\sigma_n} dP$ for all $A \in \mathcal{F}_\sigma$, and we want to show that $\int_A M_\tau dP = \int_A M_\sigma dP$ for all $A \in \mathcal{F}_\sigma$. This requires taking limits in the integrals. We need to show that M_{τ_n} are uniformly integrable (Definition 4.5).

Let $X_n = M_{\tau_n}$, $\tilde{\mathcal{F}}_n = \mathcal{F}_{\tau_n}$, X_n is a backwards martingale, $E[X_{n-1}|\tilde{\mathcal{F}}_n] = E[M_{\tau_{n-1}}|\mathcal{F}_{\tau_n}] = M_{\tau_n}$. Then apply Lemma 4.1. ☐

Example: Let $B_t, t \geq 0$ be a Brownian motion with filtration $\mathcal{F}_t$, $\tau = \int \{t \geq 0, |B_t| \geq a\}$. We want to compute $E[\tau]$.

$\tau \wedge n$ is a bounded stopping time, $B_t^2 - t$ is a martingale. Apply Theorem 7.7 to $\tau \wedge n$ and $\sigma = 0$, we get $E[B_{\tau\wedge n}^2 - \tau \wedge n] = 0$, $E[B_{\tau\wedge n}^2] = E[\tau \wedge n] \to E[\tau]$ by Theorem 1.5. Take $n \to \infty$, $E[B_{\tau\wedge n}^2] \to E[B_\tau^2] = a^2$, so $E[\tau] = a^2$.

If $\tau = \inf\{t \geq 0, B_t \geq a\}$, the previous method cannot work, because it is a one-sided bound. Instead, we consider the generating function: $E[e^{\lambda B_{\tau\wedge n} - \frac{\lambda^2}{2}\tau\wedge n}] = 1$. This is fine because $0 < e^{\lambda B_{\tau\wedge n} - \frac{\lambda^2}{2}\tau\wedge n} \leq e^{\lambda a}$ is always bounded.
Take $n \to \infty$, $E\left[e^{\lambda a - \frac{\lambda^2}{2}\tau}\right] = 1$. Let $\lambda \downarrow 0$, $P(\tau < \infty) = 1$, we get $E[e^{-\gamma\tau}] = e^{-\sqrt{\gamma}a}$ for $\gamma = \frac{\lambda^2}{2}$.

Denote $\tau_a = \inf\{t \geq 0, B(t) \geq a\}$, $\tau_n = \tau_1^1 + \cdots + \tau_1^n$, where τ_1^i are independent copies of τ_1. Also, $B(t) = aB\left(\frac{t}{a^2}\right)$ in distribution. $\tau_a = a^2\tau_1$ in distribution.
This gives $\lim_{a\downarrow 0} P(\tau_a > \epsilon) = \lim_{a\downarrow 0} P(\tau_1 \geq \epsilon a^{-2}) = 0$.

$$E[\tau] = -\left.\frac{d}{d\gamma}\right|_{\gamma=0} e^{-\sqrt{2\gamma}a} = \frac{\sqrt{2}}{2}\gamma^{-1/2}a|_{\gamma=0} = \infty$$

Note that $\frac{\tau_1^1+\tau_1^2+\cdots+\tau_1^n}{n^2}$ in distribution is stable with parameter $\frac{1}{2}$.

Example: Given a Brownian motion starting at $x \in (a, b)$, what is $P(\tau_a < \tau_b)$?
$B_t - x$ is a martingale, $E[B_{\tau_a\wedge\tau_b\wedge n}] = x$. By Theorem 1.6,

$$E[B_{\tau_a\wedge\tau_b}] = aP(\tau_a < \tau b) + bP(\tau_a > \tau_b) = aP(\tau_a < \tau b) + b(1 - P(\tau_a < \tau_b))$$

Therefore, $P(\tau_a < \tau_b) = \frac{x-b}{a-b}$.

Example: Let $\tau = \{t \geq 0, B_t \geq a + bt\}$, what is $P(\tau < \infty)$?
Use $B\left[e^{\lambda B_{\tau\wedge n} - \frac{\lambda^2}{2}\tau\wedge n}\right] = 1$. If $\lambda b \leq \frac{1}{2}\lambda^2$, $B_{\tau\wedge n} \leq a + b\tau \wedge n$.
Let $n \to \infty$, $E\left[e^{\lambda a + \left(\lambda b - \frac{1}{2}\lambda^2\right)\tau} 1_{\tau<\infty}\right] = 1$.
Take $\lambda \downarrow 2b$, $E[2^{ba}1_{\tau<\infty}] = 1$, $P(\tau < \infty) = e^{-2ba}$

Definition: 7.3: (Strong) Markov Property

Let $s < t$, $P(B_t \in A|\mathcal{F}_s) = P(B_t \in A|B_s)$ a.s.. Given all previous information up to time s, the current position at time t only depends on B_s.
Strong Markov Property: If τ is a bounded stopping time, then $P(B_{t+\tau} \in A|\mathcal{F}_\tau) = P(B_{t+\tau} \in A|B_\tau)$ a.s.

$P(B_t \in A|B_s = x) = P(s, x, t, A) = P_x(t-s, A) = P_x(B_{t-s} \in A)$ are the same notation: starting at x at time s, the probability of final position in A at time t.

$P_x(t, A) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} dy.$

Feller function $f \mapsto \int p_x(t, dy) f(y)$ maps continuous bounded functions to continuous functions. Feller + Markov $\Rightarrow$ Strong Markov.

$\tilde{B}_t = B_{t+\tau} - B_\tau$ is a Brownian motion.

Proof. We want $E\left[e^{\lambda(B_{t+\tau}-B_\tau)-\frac{\lambda^2}{2}t}|\mathcal{F}_{\tau+s}\right] = e^{\lambda(B_{\tau+s}-B_t)-\frac{\lambda^2}{2}s}$.

By Theorem 7.7, $E\left[e^{\lambda B_{t+\tau}-\frac{\lambda^2}{2}(t+\tau)}|\mathcal{F}_{\tau+s}\right] = e^{\lambda B_{\tau+s}-\frac{\lambda^2}{2}(\tau+s)}$. Then multiply both sides by $e^{-\lambda B_\tau + \frac{\lambda^2}{2}\tau}$. This is valid since it is $\mathcal{F}_{\tau+s}$ measurable. □

Consider the hitting time example. Define $\tau_a = \inf\{t \geq 0, B(t) \geq a\}$, find $P(\tau_a < t)$.

By Reflection principle, $P(\tau_a \leq t) = P\left(\sup_{0\leq s\leq t} B_s \geq a\right) = 2P(B_t \geq a)$. This can be proved:

$$\begin{aligned}
P(\tau_a \leq t) = P\left(\sup_{0\leq s\leq t} B_s \geq a\right) &= P(B_t \geq a) + P\left(\sup_{0\leq s<t} B_s \geq a, B_t < a\right) \\
&= P(B_t \geq a) + P(\tau_a \leq t, B_{\tau_a+(t-\tau_a)} - B_{\tau_a} < 0) \text{ (increment is negative)} \\
&= P(B_t \geq a) + \frac{1}{2} P(\tau_a \leq t)
\end{aligned}$$

Then $P(\tau_a \leq t) = 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-y^2/(2t)} dy$.

Brownian motion in $\mathbb{R}^d$:

1. $B(t) = (B_1(t), ..., B_d(t))$, where $B_i(t)$ are independent one-dimensional Brownian motions
2. $B(t)$ is Markov with $P_x(t, A) = P(B_t \in A|B_s = x) = \int_A \frac{1}{(2\pi(t-s))^{d/2}} e^{-\frac{|y-x|^2}{2(t-s)}} dy$
3. $B(t)$ has stationary independent increements: Gaussian mean 0, $E[|B_i(t)|^2] = t - s$, $E[|B(t)|^2] = dt$.
4. $e^{\lambda B(t) - \frac{|\lambda|^2}{2}t}$ is a martingale for $\lambda \in \mathbb{R}^d$.

Lemma 12. *u is harmonic in open set G if and only if it satisfies the mean value property: $\forall r > 0$ s.t. $B_r(x) \subset G$, $\frac{1}{|\partial B(t,x)|} \int_{\partial B(r,x)} u(y) ds = u(x)$.*

If u is harmonic, $\Delta u = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} u = 0$, $u(B_t)$ is a martingale.

Theorem: 7.8:

Let G be a bounded open set on $\mathbb{R}^d$, f bounded measurable function on ∂G, then $u(x) = E_x[f(B_{\tau_G})]$ is harmonic in G, where $\tau_G = \inf\{t \geq 0, B(t) \notin G\}$, E_x means the expectation starting at x

Proof. Let H be a smaller ball contained in G.

$$\begin{aligned} u(x) &= E_x\left[E_x\left[f(B_{\tau_G})|\mathcal{F}_H\right]\right] = E_x[u(B_{\tau_H}) \\ &= \int u(y)\Pi_{\partial B}(x,dy) \\ &= \frac{1}{|\partial B(r,x)|}\int_{\partial B(r,x)} u(y)ds \end{aligned}$$

□

Optional stopping gives the solution to Dirichlet problem $\begin{cases}\Delta u = 0 x\in G, \\ u = f, x\in\partial G\end{cases}$. f is continuous at $a\in\partial G$, then $u(x)\to f(a)$ as $x\to a$ from inside G

We need to check that if $\lim\limits_{x\to a\in\partial G, x\in G} P_x\left(|B_{\tau_G}-x|<\delta\right)=1$,

then $\lim\limits_{x\to a\in\partial G, x\in G} P_x\left(|f(B_\tau)-f(a)|<\epsilon\right)=1$

Lemma 13. *If* $\lim\limits_{x\to a\in\partial G, x\in G} P_x\left(\tau_G>\epsilon\right)=0$*, then* $\lim\limits_{x\to a\in\partial G, x\in G} P_x\left(|B_{\tau_G}-x|<\delta\right)=1$

Proof.

$$\begin{aligned} P_x\left(|B_{\tau_G}-x|<\delta\right) &\geq P_x\left(\sup_{0\leq t\leq\epsilon}|B_t-x|<\delta, \tau_G<\epsilon\right), \text{ for any } \epsilon \\ &\geq P_x\left(\sup_{0\leq t\leq\epsilon}|B_t-x|<\delta\right) - P_x(\tau_G\geq\epsilon). \end{aligned}$$

Take $x\to a$, $P_x(\tau_G\geq\epsilon)\to 0$, and as $\epsilon\to 0$, $P_x\left(\sup_{0\leq t\leq\epsilon}|B_t-x|<\delta\right)\to 1$. □

Let $\sigma_G = \inf\{t>0 : B_t\notin G\}$, $\tau_G = \inf\{t\geq 0 : B_t\notin G\}$, $\sigma_G\geq\tau_G$. A boundary point a is regular if $P_a(\sigma_G=0)=1$.

Lemma 14. *A boundary point a is regular if and only if* $\lim_{x\to a\in\partial G, x\in G} E_x[f(B_{\tau_G})]=f(a)$ *for all bounded measurable f continuous at a.*

Proof. (⇒) It is enough to show that $P_x(G_a<\epsilon)$ is l.s.c., because then $\lim\limits_{x\to a\in\partial G, x\in G} P_x(\sigma_G<\epsilon)\geq P_a(G_a<\epsilon)=1$, and it gives $\lim\limits_{x\to a\in\partial G, x\in G} P_x(\tau_G<\epsilon)=1$, and we can use the previous lemma.

We show the lower semi-continuity. For an x, the probability that there is a time $s\in(\delta,\epsilon)$ with $B_s\notin G$ is

$$\int_{\mathbb{R}^d} dy\frac{1}{(2\pi\delta)^{d/2}}\exp\left(-\frac{|y-x|^2}{2\delta}\right)P_y(\exists s\in(0,\epsilon-\delta), B_s\notin G)$$

This increases to $P_x(\sigma_G<\epsilon)$ and is continuous. □

If ∂G is a $d-1$-dimensional smooth manifold at a, then a is regular, $P_a(\sigma_G=0)=1$.

Proof. Let $\tau_{c\delta^2}=\inf\left\{t\geq 0 : B_2(t)\geq c\delta^2\right\}$, $\sigma_{\pm\sigma}=\inf\{t\geq 0 : |B_1(t)|\geq\delta\}$,
we want to show $\lim_{\delta\to 0}P(\tau_{c\delta^2}\leq\sigma_{\pm\sigma})=1$ *i.e.* it hits the boundary $c\delta^2$ in small time.

$\tau_{c\delta^2}=c^2\delta^4\tau_1$, $\sigma_\pm=\delta^2\sigma_{\pm 1}$ by Brownian scaling, distance is just time squared, so $P(\tau_{c\delta^2}\leq\sigma_{\pm\delta})=P(\tau_1\leq c^{-2}\delta^{-2}\sigma_{\pm 1})$. Taking the limit $\delta\to 0$, it becomes $1-P(\sigma_{\pm 1}=0)=1$. □

Let $G = \left\{x \in \mathbb{R}^d, \delta < |x| < R\right\}$, $f = \begin{cases} 1, |x| = \delta, \\ 0, |x| = R \end{cases}$

$u(x) = E_x[f(\tau_G)] = P_x(\tau_\delta < \tau_R) = \begin{cases} \frac{\log R - \log |x|}{\log R - \log \delta}, d = 2 \\ \frac{|x|^{2-d} - R^{2-d}}{\delta^{2-d} - R^{2-d}}, d = 3, 4, ... \end{cases}$

Theorem: 7.9:

Let $d = 2$, Brownian motion visits a neighborhood of every point infinitely many times, but does not hit a point.

Proof. The probability fo hitting a point is $P_x(\tau_0 < \tau_R) = \lim_{\delta \to 0} \frac{\log R - \log |x|}{\log R - \log \delta} = 0$.

The probability of getting close to a point is $P_x(\tau_\delta < \infty) = \lim_{R \to \infty} P_x(\tau_\sigma < \tau_R) = \lim_{R \to \infty} \frac{\log R - \log |x|}{\log R - \log \delta} = 1$ □

8 Stochastic Integrals and Differential Equations

Let f be some deterministic function,

$$\int_0^t f(s)dB(s) = f(t)B(t) - \int_0^t f'(s)B(s)ds$$

Let $f = \sum c_i 1_{[a_i,b_i)}$, where $[a_i, b_i)$ are non-overlapping intervals.

$$E\left[\int_0^t fdB \int_0^t fdB\right] = \sum c_i c_j E\left[(B(b_i) - B(a_i))(B(b_j) - B(a_j))\right] = \sum c_i^2(b_i - a_i) = \int_0^t f^2 ds$$

If $f_n \in L^2[0,1]$, $f_n \to f$ in L^2 (f_n is Cauchy in L^2), then as $n, m \to \infty$

$$E\left[\left(\int_0^t f_n dB - \int_0^t f_m dB\right)^2\right] = \int_0^t (f_n - f_m)^2 ds \to 0$$

So $\int_0^t f_n dB$ are Cauchy in $L^2[0,1]$, so $\int fdB$ is defined for deterministic $f \in L^2$ (by Wiener)

What if $f = f(s,w)$ not deterministic? w is from some Brownian motion.

Simplest example: $\int_0^t B(s)dB(s)$.

Idea 1 (Stratonovich):

$$\int_0^t B(s)dB(s) = B^2(t) - \int_0^t B'(s)B(s)dS = B^2(t) - \int_0^t B(s)dB(s)$$

$$\Rightarrow \int_0^t B(s)dB(s) = \frac{1}{2}B^2(t)$$

Idea 2 (Riemann integration, Ito integral):

$$\int_0^t B(s)dB(s) = \lim_{n\to\infty}\sum_{j=0}^{2^n-1} B(t_j)\left(B\left(\frac{(j+1)t}{2^n}\right) - B\left(\frac{jt}{2^n}\right)\right), t_j \in \left[\frac{jt}{2^n}, \frac{(j+1)t}{2^n}\right)$$

$$R_t = \lim_{n\to\infty}\sum_{j=0}^{2^n-1} B\left(\frac{(j+1)t}{2^n}\right)\left(B\left(\frac{(j+1)t}{2^n}\right) - B\left(\frac{jt}{2^n}\right)\right)$$

$$L_t = \lim_{n\to\infty}\sum_{j=0}^{2^n-1} B\left(\frac{jt}{2^n}\right)\left(B\left(\frac{(j+1)t}{2^n}\right) - B\left(\frac{jt}{2^n}\right)\right)$$

$R_t - L_t = t$ (total variation), $R_t + L_t = B^2(t)$. This gives $\begin{cases} R_t = \frac{1}{2}B^2(t) + \frac{t}{2} \\ L_t = \frac{1}{2}B^2(t) - \frac{t}{2} \end{cases}$ They do not match, so we choose L_t to satisfy martingale property.

8.1 Lebesgue Integration and Progressively Measurable Functions

Let $(\Omega, \mathcal{F}, P)$ be a probability space, B_t a Brownian motion with filtration $\mathcal{F}_t = \sigma(B(s), s \le t)$. We want to compute $\int_0^t \sigma dB$, where $\sigma(s,\omega)$ is simple, non-anticipating, *i.e.* $\sigma(s,\omega) = \sigma_j(\omega)$ on $s_j \le s < s_{j+1}$, where $0 \le s_0 < s_1 < s_2 < \cdots$ is a partition of $[0,\infty)$, and $\sigma_j \in \mathcal{F}_{s_j}$. Define:

$$\int_0^t \sigma dB = \sum_{j=0}^{J(t)} \sigma_j\left(B(s_{j+1}) - B(s_j)\right) + \sigma_{J(t)}\left(B(t) - B(s_{J(t)})\right) = \int_0^\infty \sigma 1(s \le t)dB,$$

where $s_{J(t)} \le t \le s_{J(t+1)}$.

1. It is linear:

$$\int_0^t (c_1\sigma_1 + c_2\sigma_2)dB = c_1 \int_0^t \sigma_1 dB + c_2 \int_0^t \sigma_2 dB$$

2. It is a continuous martingale

   Proof. If $u < s_j$, preconditioning on a larger σ-field, gives

$$E\left[\sigma_j\left(B(s_{j+1}) - B(s_j)\right)|\mathcal{F}_u\right] = E\left[E\left[\sigma_j\left(B(s_{j+1}) - B(s_j)\right)|\mathcal{F}_j\right]|\mathcal{F}_u\right] = 0,$$

   because σ_j goes out and Brownian increments is mean zero for $u < s_j$

   If $u \geq s_j$, σ_j is $\mathcal{F}_u$ measurable,

$$E\left[\sigma_j\left(B(s_{j+1}) - B(s_j)\right)|\mathcal{F}_u\right] = \sigma_j E\left[\left(B(s_{j+1}) - B(s_j)\right)|\mathcal{F}_u\right] = \begin{cases} \sigma_j\left(B(s_{j+1}) - B(s_j)\right), u \geq s_{j+1} \\ \sigma_j\left(B(u) - B(s_j)\right), s_j \leq u < s_{j+1} \end{cases}$$

   □

3. Ito isometry:

$$E\left[\left(\int_0^t \sigma dB\right)^2\right] = E\left[\int_0^t \sigma^2 ds\right]$$

   Proof. Note that $\int_0^t \sigma dB = \int_0^\infty \sigma 1(s \leq t)dB$, so we consider to ∞

$$\begin{aligned} E\left[\left(\int_0^\infty \sigma dB\right)^2\right] &= E\left[\left(\sum_j \sigma_j(B(s_{j+1}) - B(s_j))\right)^2\right] \\ &= 2\sum_{j<\tilde{j}} E\left[\sigma_j\sigma_{\tilde{j}}(B(s_{j+1}) - B(s_j))(B(s_{\tilde{j}+1}) - B(s_{\tilde{j}}))\right] \\ &+ \sum_j E\left[\sigma_j^2\left(B(s_{j+1}) - B(s_j)\right)^2\right] \end{aligned}$$

   By conditional independence (preconditioning on $\mathcal{F}_{s_{\tilde{j}}}$) and measurability, the first term is zero.

   Again precondition on $\mathcal{F}_{s_j}$, we get $sum_j E\left[\sigma_j^2\left(B(s_{j+1}) - B(s_j)\right)^2\right] = \sum_j E\left[\sigma_j^2(s_{j+1} - s_j)\right]$.

   Since by definition, σ_j is constant on the interval, this is equivalent to $E\left[\int_0^t \sigma^2 ds\right]$. □

4. $e^{\int_0^t \sigma dB - \frac{1}{2}\int_0^t \sigma^2 dB}$ is a continuous martingale. We can also scale σ by $\lambda\sigma$ to get a general case. Note that this statement is already a stronger case of Ito isometry

   Proof. Suppose $s_{J(t)} \leq u < s_{J(t)+1}$

$$\begin{aligned} &E\left[e^{\left(\sum_{j=0}^{J(T)} \sigma_j(B(s_{j+1})-B(s_j))+\sigma_{J(t)}(B(t)-B(s_{J(t)}))-\frac{1}{2}\sum_{j=0}^{J(t)} \sigma_j^2(s_{j+1}-s_j)-\frac{1}{2}\sigma_{J(t)}^2(t-s_{J(t)})\right)}|\mathcal{F}_u\right] \\ &= e^{\sum_{j=0}^{J(T)} \sigma_j(B(s_{j+1})-B(s_j))-\frac{1}{2}\sum_{j=0}^{J(t)} \sigma_j^2(s_{j+1}-s_j)} E\left[e^{\sigma_{J(t)}(B(t)-B(s_{J(t)}))-\frac{1}{2}\sigma_{J(t)}^2(t-s_{J(t)})}|\mathcal{F}_u\right] \\ &= e^{\sum_{j=0}^{J(T)} \sigma_j(B(s_{j+1})-B(s_j))-\frac{1}{2}\sum_{j=0}^{J(t)} \sigma_j^2(s_{j+1}-s_j)} e^{\sigma_{J(t)}(B(u)-B(s_{J(t)}))-\frac{1}{2}\sigma_{J(t)}^2(u-s_{J(t)})} \end{aligned}$$

   □

Let $\mathcal{P}$ be the closure in $L^2([0,\infty), dt \times dP)$ of simple, non-anticipating functions. If $\sigma \in P$, then there exists a sequence of simple, non-anticipating $\sigma_n \to \sigma$, $E\left[\int_0^t |\sigma_n - \sigma|^2\, ds\right] \to 0$.

Let $X_n(t) = \int_0^t \sigma_n dB$, by Theorem 4.14,

$$P\left(\sup_{0\le t\le T} |X_n(T) - X_m(t)| > \epsilon\right) \le \frac{1}{\epsilon^2} E\left[|X_n(T) - X_m(T)|^2\right] = \frac{1}{\epsilon^2} E\left[\int_0^T (\sigma_n - \sigma_m)^2 ds\right] \to 0$$

as $n, m \to \infty$. Therefore, $X_n \to X$ continuous in t, $X(t) = \int_0^t \sigma dB$.

It doesn't depend on the choice of sequence $\sigma_n \to \sigma$. If $\sigma_n' \to \sigma$, then $E\left[\int_0^t |\sigma_n - \sigma_n'|^2\, ds\right] \to 0$, so $E\left[\sup_{0\le t\le T} |X_n(t) - X_n'(t)|^2\right] \to 0$

Properties:

1. Linear
2. Continuous martingale, because $E\left[\int_0^t \sigma_n dB | \mathcal{F}_u\right] = \int_0^u \sigma_n dB$ and L^2-convergence implies L^1 convergence.
3. Ito's isometry with the same argument
4. $e^{\int_0^t \sigma dB - \frac{1}{2}\int_0^t \sigma^2 dB}$ is a continuous martingale, assuming $|\sigma| \le C$.

   Proof. For each n, $Z_n(t) = e^{\int_0^t \sigma_n dB - \frac{1}{2}\int_0^t \sigma_n^2 dB}$ is a martingale. $E[Z_n(t)|\mathcal{F}_u] = Z_n(u)$.

   Take limit of n, $Z_n(u) \to Z(u)$. For the LHS, check uniform integrability.

   $$E[Z_n^2(t)] = E\left[e^{2\left(\int_0^t \sigma_n dB - \int_0^t \sigma_n^2 dB\right)} e^{\int_0^t \sigma_n^2 dB}\right] \le e^{C^2 t},$$

   because $E\left[e^{\left(\int_0^t \sigma_n dB - \int_0^t \sigma_n^2 dB\right)}\right] = 1$ and $E\left[e^{\int_0^t \sigma_n^2 dB}\right] \le e^{C^2 t}$ □

Lemma 15. *The progressively measurable functions $\mathcal{P}$ is the set of simple, non-anticipating functions $\sigma \in L^2([0,\infty), dt \times dP)$ s.t. $\forall t \ge 0$, the map $[0,t] \times \Omega \to \mathbb{R}$ given by $\sigma(s,\omega)$ is $B[0,t] \times \mathcal{F}_t$-measurable.*

Proof. Need to show the set equivalence. $P \subset \{\cdot\}$ is clear. We need to show $\{\cdot\} \subset P$.

Let $\sigma \in \{\cdot\}$.

1. $\sigma_N = \sigma 1_{|\sigma| \le N}$, $E\left[\int_0^t (\sigma - \sigma_N)^2 ds\right] \to 0$ by Theorem 1.6
2. σ bounded, $\sigma_h(t) = \frac{1}{h}\int_{t-h}^t \sigma(s) ds$, $E\left[\int_0^t (\sigma - \sigma_N)^2 ds\right] \to 0$ by Theorem 1.4
3. σ continuous, make it simple by $\sigma_n(s) = \sigma(2^{-n}\lfloor 2^n s \rfloor)$. by Theorem 1.4

□

Example: $\sigma(s,\omega) = f(B(s))$ s.t. $E\left[\int_0^t (f^2(B(s))) ds\right] < \infty$ is a progressively measurable function.

8.2 Stochastic Integrals and Ito's Formula

Definition: 8.1: Stochastic Integral

If σ, b are progressively measurable with $E\left[\int_0^\infty \sigma ds\right] < \infty$ and $\int_0^t b(s,\omega)ds < \infty$, then we can define the stochastic integral

$$X(t) = X_0 + \int_0^t b(s,\omega)ds + \int_0^t \sigma(s)dB_s$$

Often, we write as stochastic differential equation:

$$dX = bdt + \sigma dB$$

Example: $\int_0^t 2B_s dB_s = B_t^2 - t$. This is equivalent to $dB_t^2 = 2B_t dB_t - dt$.

Ito's formula: $df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt$.

Theorem: 8.1: Ito's Formula

If f is twice continuously differentiable, then

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$$

Proof. Let $0 = t_0 < t_1 < \cdots < t_n = t$ be a partition of $[0,t]$. By Taylor Expansion,

$$\begin{aligned} \text{LHS} &= \sum_{i=0}^{n-1} f(B_{t_{i+1}}) - f(B_{t_i}) \\ &= \sum_{i=0}^{n-1} f'(B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \frac{1}{2}f''(B_{\xi_i})(B_{t_{i+1}} - B_{t_i})^2 \end{aligned}$$

for $\xi_i \in (t_i, t_{i+1})$. As mesh size $\to 0$,

$$\sum_{i=0}^{n-1} f'(B_{t_i})(B_{t_{i+1}} - B_{t_i}) \to \int_0^t f'(B_s)dB_s$$

To remove ξ_i, consider the difference:

$$\left|\sum_{i=0}^{n-1}\left(f''(B_{\xi_i}) - f''(B_{t_i})\right)(B_{t_{i+1}} - B_{t_i})^2\right| \le \max\left|f''(B_{\xi_i}) - f''(B_{t_i})\right|\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2$$

Then we want to show that $\sum_{i=0}^{n-1} f''(B_{t_i})(B_{t_{i+1}} - B_{t_i})^2 \to \sum_{i=0}^{n-1} f''(B_{t_i})(t_{i+1} - t_i)$.
Let $X_i = (B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)$.

$$E\left[\left(\sum_{i=0}^{n-1} f''(B_{t_i})X_i\right)^2\right] = 2\sum_{i<j} E\left[f''(B_{t_i})f''(B_{t_j})X_iX_j\right] + \sum_i E\left[f''(B_{t_i})^2 X_i^2\right]$$

By preconditioning,

$$E\left[E\left[f''(B_{t_i})f''(B_{t_j})X_iX_j|\mathcal{F}_j\right]\right] = 0$$
$$E\left[E\left[f''(B_{t_i})^2X_i^2|\mathcal{F}_{t_i}\right]\right] = E\left[f''(B_{t_i})^2E\left[X_i^2\right]\right] = E\left[f''(B_{t_i})^2(t_{j+1} - t_j)^2E[Z^2-1]^2\right] = 0$$

As the mesh size $\to 0$, we get the integrals on the RHS. $\square$

It is essentially a Taylor expansion. Now consider if $f(t,x)$, $x = B_t$,

$$df(t,B_t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dt$$

If $B_t \in \mathbb{R}^d$ are independent/uncorrelated,

$$df(t,B_t) = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2 f}{\partial x_i^2}\right)dt + \sum_{i=1}^{d}\frac{\partial f}{\partial x_i}dB_t^i$$

If f is harmonic in $\mathbb{R}^d$, $df = \nabla f dB$, $f(B_t) = f(B_0) + \int_0^t \nabla f dB$, so $f(B_t)$ is martingale.

Example: Let $f(t,x) = e^{\lambda x - \frac{\lambda^2}{2}t}$. Since $\left(\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right)f = 0$, $e^{\lambda B_t - \frac{\lambda^2}{2}t}$ is a martingale.

Example: Let f be continuous on $[0,\infty)$, $L_t(x) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\left|\{0 \le s \le t, |f(x) - x| < \epsilon\}\right|$, where the outer $|\cdot|$ is the Lebesgue measure. $L_t(x) = \sum_{s_i\in[0,t],\phi(s_i)=x}|f'(s_i)|^{-1}$. This defines Local time of f at x. Then:

$$\int_0^t 1_A(f(s))ds = \int_A L_t(x)dx$$
$$\int_0^t \phi(f(s))ds = \int_A L_t(x)\phi(x)dx$$

Theorem: 8.2:

Brownian motion has an a.s. continuous local time.

Proof. Apply Theorem 8.1 to $f_\epsilon'' = \frac{1}{\epsilon}1_{[x-\epsilon,x+\epsilon]}$.

$$\int_0^t f_\epsilon''(B_s)ds = \frac{1}{2\epsilon}\left|\{0 \le s \le t, |f(x) - x| < \epsilon\}\right|$$

We want to show that it is equal to $f_\epsilon(B_t) - f_\epsilon(B_0) - \int_0^t \frac{\partial f_\epsilon}{\partial x}(B_s)dB_s$.

As $\epsilon \to 0$, continuation gives $L_t(x) = |B_t - x| - |B_0 - x| - \int_0^t \operatorname{sgn}(B_s - x)dB_s$ (Tanaka's formula).
Let $X_t = \int_0^t \operatorname{sgn}(B_s - x)dB_s$, $e^{\lambda X_t - \frac{\lambda^2}{2}\int_0^t \operatorname{sgn}^2(B_s - x)ds}$ is a martingale.

$$\int_0^t \operatorname{sgn}^2(B_s - x)ds = t - \int_0^t 1_{B_s - x = 0}ds = t \text{ a.s.}$$

Let $\hat{B}_t = |B_t - x|$, $\hat{B}_t$ is a Brownian motion starting at $-x$. Then $\left|\hat{B}_t\right|$ is a Brownian motion starting at $+x$ reflected at 0. $\square$

Example (Heat equation): If $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$, then $u(t-s, B_s)$ is a martingale. If $\int u_0(x)e^{-\frac{x^2}{2t}}dx < \infty$,

$$E_x[u_0(B_t)] = E[u(t-t, B_t)] = E_x[u(t, B_0)] = u(t,x)$$

Now if $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + Vu$. where $V = V(x)$, then $Z_s = u(t-s, B_s)e^{\int_0^s V(B_u)du}$ is a martingale:

$$dZ_s = -\frac{\partial u}{\partial s}e^{\int_0^s V(B_u)du} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}e^{\int_0^s V(B_u)du} + V(B_s)u(t-s, B_s)e^{\int_0^s V(B_u)du}$$

but $V(B_s)u(t-s,B_s)e^{\int_0^s V(B_u)du} = 0$, so

$$u(t,x) = E_x\left[u_0(B_t)e^{\int_0^t V(B_u)du}\right] = E_x[Z_t] = E_x[Z_0]$$

This is the Feynmann-Kac formula.

8.3 Stochastic Ordinary Differential Equations

Let X_t be progressively measurable, a stochastic differential equation is of the form

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \quad X_0 = x_0, \quad X_t, b \in \mathbb{R}^d, \sigma \in \mathbb{R}^{d\times k}, B_t \in \mathbb{R}^k$$

We want to search for an X_t that satisfy the equation, having the known drift $b(t,X_t)$ and diffusion $\sigma(t,X_t)$.

Conditions:

1. X_t is progressively measurable
2. $\int_0^T |b(t,X_t)|\, dt < \infty$
3. $\int_0^T \|\sigma(t,X_t)\|^2\, dt < \infty$.

Equivalently,

$$X_t = x_0 + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s$$

Langevin Equation:

$$dX_t = -\alpha t X_t dt + \sigma dB_t, \quad \alpha > 0$$

Notice that

$$\begin{aligned} d(e^{\alpha t}X_t) &= \alpha e^{\alpha t}X_t dt + e^{\alpha t}dX_t \\ &= \alpha e^{\alpha t}X_t dt + e^{\alpha t}(-\alpha X_t dt + \sigma dB_t) = \sigma e^{\alpha t}dB_t \end{aligned}$$

so $e^{\alpha t}X_t - X_0 = \sigma \int_0^t e^{\alpha s}dB_s$, *i.e.* $X_t = e^{-\alpha t}X_0 + \sigma\int_0^t e^{-\alpha(t-s)}dB_s$. This is the Ornstein-Uhlenbeck process.

For a formal proof, we need to check X_t satisfies $X_t = X_0 - \alpha\int_0^t X_s ds + \sigma(B_t - B_0)$, we can integrate X_t:

$$\begin{aligned} \int_0^t X_s ds &= X_0 \int_0^t e^{-\alpha s}ds + \sigma \int_0^s \int_0^t e^{-\alpha(s-u)}dB_u ds \\ &= \frac{X_0}{\alpha}(1-e^{-\alpha t}) + \sigma\int_0^t\int_u^t e^{-\alpha(s-u)}ds dB_u \\ &= \frac{X_0}{\alpha}(1-e^{-\alpha t}) + \sigma\int_0^t (1-e^{-\alpha(t-u)})dB_u \\ &= \frac{X_0}{\alpha}(1-e^{-\alpha t}) + \frac{\sigma}{\alpha}\left(B_t - B_u + \frac{1}{\sigma}(X_t - X_0)\right) \end{aligned}$$

Substitute into $X_t = X_0 - \alpha\int_0^t X_s ds + \sigma(B_t - B_0)$, we can verify the equality.

X_t is Gaussian, $m(t) = E[X_t] = e^{-\alpha t}E[X_0]$. Assume X_0 is independent of B_s:

$$Cov(X_t, X_s) = e^{-\alpha(t+s)}Var(X_0) + \frac{\sigma^2}{2\alpha}(e^{2\alpha(s\wedge t)} - 1)e^{-\alpha(t+s)}$$

If X_0 is Gaussian mean 0, var $\frac{\sigma^2}{2\alpha}$, then $Cov(X_t, X_s) = \frac{\sigma^2}{2\alpha}e^{-\alpha(t-s)}$ for $s < t$, $m(t) = 0$. $X_{t+\cdot} = X_0$ in distribution. It is a stationary process.

For a smooth function $f(x)$:

$$\begin{aligned} E\left[\left(\int f(s)X_s ds\right)^2\right] &= 2\iint_{t>s} f(s)f(t)E[X_sX_t]dsdt \\ &= 2\iint_{t>s} f(s)f(t)\frac{\sigma^2}{2\alpha}e^{-\alpha(t-s)}dtds \to \int f^2 dt \end{aligned}$$

X_t is the velocity process.

Geometric Brownian Motion:
Bachelier introduced B_t as a model of stock prices.

Samuelson proposed geometric Brownian motion

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_t,$$

σ is volatility, μ is drift. The relative price change is a Brownian motion.

In our language,

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad S_0 = s_0$$

Note that $\frac{dS_t}{S_t} \neq d\log S_t$ by Theorem 8.1.
Guess $S_t = f(t, B_t)$,

$$dS_t = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\right)dt + \frac{\partial f}{\partial x}dB_t$$

We want $\begin{cases} \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = \mu f \\ \frac{\partial f}{\partial x} = \sigma f \end{cases}$, so $S_t = f(t,x) = s_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma x}$. $-\frac{\sigma^2}{2}$ is the price of risk.

Bessel Process:
Consider $d = 2$, $B_t = (B_t^1, B_t^2)$,

$$r_t = |B_t| = \left((B_t^1)^2 + (B_t^2)^2\right)^{1/2}$$

By Theorem 8.1,

$$\begin{aligned} dr_t &= \frac{\partial f}{\partial x_1}dB_t^1 + \frac{\partial f}{\partial x_2}dB_t^2 + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right)dt \\ &= \frac{B_t^1}{r_t}dB_t^1 + \frac{B_t^2}{r_t}dB_t^2 + \frac{1}{2r_t}dt \end{aligned}$$

This is not a stochastic differential equation in the current form, due to the presence of B_t^1 and B_t^2.

Let $Y_t = \int_0^t \frac{B_t^1}{r_t}dB_t^1 + \frac{B_t^2}{r_t}dB_t^2$.
$e^{\lambda Y_t - \frac{\lambda^2}{2}\int_0^t \left(\frac{B_t^1}{r_t}\right)^2 + \left(\frac{B_t^2}{r_t}\right)^2 ds} = e^{\lambda Y_t - \frac{\lambda^2}{2}}$ is a Martingale, and Y_t is a Brownian motion.

Therefore, $dr_t = dY_t + \frac{1}{2r_t} dB_t$ is a stochastic differential equation.

Theorem: 8.3: Existence and Uniqueness of Solutions to SDEs

Assume $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$. If σ, b are measurable, and satisfy the Lipschitz condition:

$$\|\sigma(t,x) - \sigma(t,y)\| + |b(t,x) - b(t,y)| \le C\,|x-y|,$$

then there exists a unique solution X_t and $E\left[\int_0^t |X_s|^2\, ds\right] < \infty$.

Proof. Uniqueness: Suppose X_t^1 and X_t^2 are both solutions with $X_0^1 = x^1$, $X_0^2 = x^2$.

$$X_t^2 - X_t^1 = x^2 - x^1 + \int_0^t (b(s, X_s^2) - b(s, X_s^1))ds + \int_0^t (\sigma(s, X_s^2) - \sigma(s, X_s^1))dB_s$$

Then by $(A+B+C)^2 \le 4A^2 + 4B^2 + 4C^2$, we get

$$\begin{aligned} E\left[\left|X_t^2 - X_t^1\right|^2\right] &\le 4(x^2 - x^1)2 + 4E\left[\left(\int_0^t (b(s, X_s^2) - b(s, X_s^1))ds\right)^2\right] \\ &\quad + 4E\left[\left(\int_0^t (\sigma(s, X_s^2) - \sigma(s, X_s^1))dB_s\right)^2\right] \\ &\quad \text{By Cauchy/Jensen's Inequality and Ito's Isometry} \\ &\quad \le 4(x^2 - x^1)^2 + L\left|X_s^2 - X_s^1\right| + L\left|X_s^2 - X_s^1\right| \to 0 \end{aligned}$$

Note that with iteration, we get $f(t) = \left|X_t^2 - X_t^1\right|^2 \le C\left(f(0) + \int_0^t f(s)ds\right) \le e^{Ct} f(0)$.

Recall Theorem 7.6. Let $S = \sup_{0\le t\le T} X_t$ we can extend it to

$$\begin{aligned} E[S^2] &= 2\int_0^\infty \lambda P(S \ge \lambda d\lambda) \\ &\le 2\int_0^\infty \lambda \int \frac{|X_T|}{\lambda} 1_{S\ge\lambda} dP d\lambda \\ &= 2\int |X_T| \int_0^S d\lambda dP = 2\int |X_T| S dP \\ &\le 2\sqrt{\int |X_T|^2 dP}\sqrt{\int S^2 dP} \\ &\le 4E[X_T^2]. \end{aligned}$$

Existence: Let $X_0(t) = x_0$, $X_n(t) = x_0 + \int_0^t b(s, X_{n-1}(s))ds + \int_0^t \sigma(s, X_{n-1}(s))dB(s)$, we want to show that the limit X_∞ exists.

$$E\left[\sup_{0\le s\le t}\left|\int_0^s b(u, X_{n-1}(u)) - b(u, X_{n-2}(u))du\right|^2\right] \le \frac{C}{2}\int_0^t E\left[\sup_{0\le u\le s} |X_{n-1}(u) - X_{n-2}(u)|^2\, ds\right]$$

$$E\left[\sup_{0\le s\le t}\left|\int_0^s \sigma(u, X_{n-1}(u)) - \sigma(u, X_{n-2}(u))dB_u\right|^2\right] \le \frac{C}{2}\int_0^t E\left[\sup_{0\le u\le s} |X_{n-1}(u) - X_{n-2}(u)|^2\, ds\right]$$

$$f_n(t) = E\left[\sup_{0\le s\le t} |X_n(s) - X_{n-1}(s)|^2\right] \le C\int_0^t E\left[\sup_{0\le u\le s} |X_{n-1}(u) - X_{n-2}(u)|^2\right] du$$

From $\begin{cases} f_0(t) = A \\ f_n(t) = C\int_0^t f_{n-1}(s)ds \end{cases}$, we get $f_n(t) \le \frac{(CT)^n A}{n!}$, and $P\left(\sup_{0\le s\le t}|X_n(s) - X_{n-1}(s)| \ge \frac{1}{2^n}\right) \le \frac{(CT)^n A}{n!}2^{2n}$ is summable, so $|X_n(s) - X_{n-1}(s)| \le \frac{1}{2^n}$ except for finitely many n. $X_n(t) = \sum_{j=1}^n X_j(t) - X_{j-1}(t) \to X(t)$ uniformly with probability 1. □

Note: It does not matter how σ and b behave w.r.t. t. Also, this theorem does not cover all cases of existence and uniqueness. Examples in population genetics:

Feller's Diffusion: $dX_t = \sqrt{X_t}dB_t$
Fisher-Wright Diffusion: $dX_t = \sqrt{X_t(1-X_t)}dB_t$

$\sigma(x) = \sqrt{x}$ or $\sqrt{x(1-x)}$ are not Lipschitz, but those SDEs have solutions.

For $d = 1$, $|\sigma(t,x) - \sigma(t,y)| \le C|x-y|^{1/2}$ is $\frac{1}{2}$-Holder.

They are like square roots of the logistic models, $dX_t = \alpha X_t dt$, $dX_t = \alpha X_t(1-X_t)dt$.

More formally, **Feller's Diffusion** is a Markov process on $\mathbb{R}^d$ with transition probabilities $p(s,x,t,A) = p(X_t \in A|X_s = x)$, satisfying for $\delta > 0$, as $h \to 0$:

1. $\frac{1}{h}\int_{|y-x|>\delta} p(t,x,t+h,dy) \to 0$ (X_t is continuous)
2. $\frac{1}{h}\int_{|y-x|<\delta}(y_i - x_i)p(t,x,t+h,dy) \to b_i(t,x)$ (mean of jump)
3. $\frac{1}{h}\int_{|y-x|<\delta}(y_i - x_i)(y_j - x_j)p(t,x,t+h,dy) \to a_{ij}(t,x)$, where $a = \sigma\sigma^T$

If $b = 0$, $a = I$, X_t is a d-dim Brownian motion. If $a = 0$, $\dot{X}_u = b(u, x_u)$, $X_s = 0$, $p(s,x,t,A) = \begin{cases} 1, X_t \in A \\ 0, X_t \notin A \end{cases}$.

Ito's Diffusion: for every $\lambda \in \mathbb{R}^d$, $\exp\left[\lambda(X_t - \int_0^t b_s ds) - \frac{1}{2}\int_0^t \lambda^T a_s \lambda ds\right]$ is a martingale.

Assume $a = \sigma\sigma^T$ is non-negative definite, σ, b are constants, and $d = 1$; $X_t = X_0 + bt + \sigma B_t$.

1. $\frac{1}{h}\int_{|y-x|>\delta} \frac{1}{\sqrt{2\pi\sigma^2 h}}\exp\left(-\frac{(y-x-bh)^2}{2\sigma^2}\right)dy = \frac{1}{h}\int_{\left|z+\frac{b\sqrt{h}}{\sigma}\right|>\frac{\delta}{\sigma\sqrt{h}}} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \to 0$
2. $\frac{1}{h}\int_{|y-x|<\delta}(y_i - x_i)\exp\left(-\frac{(y-x-bh)^2}{2\sigma^2}\right)dy \to b_i(t,x)$ (mean of jump)
3. $\frac{1}{h}\int_{|y-x|<\delta}(y_i - x_i)(y_j - x_j)\exp\left(-\frac{(y-x-bh)^2}{2\sigma^2}\right)dy \to \sigma^2$.

Let $\sigma(x) = \text{sgn}(x)$, B be a Brownian motion, $d\tilde{B} = \text{sgn}(B)dB$, $\tilde{B(t)} = \int_0^t \text{sgn}(B)dB$ is another Brownian motion.

$dB = \text{sgn}(B)d\tilde{B}$ is a stochastic differential equation. $d(-B) = \text{sgn}(-B)d\tilde{B}$ is another SDE. But they are two different solutions starting at zero. Uniqueness is not guaranteed.

X_t is Markov from uniqueness (by construction), because we can solve $dX = bdt + \sigma dB$ on $[0,t]$, or we can solve it on $[0,s]$ and then solve on $[s,t]$ starting from X_s.

$$P(s,x,t,A) = P(X_t^{s,x} \in A), P(X_t^{0,x} \in A|\mathcal{F}_s) = P(s, X_s^{0,x}, t, A)$$

Let σ be progressively measurable and bounded, $dX_t = \sigma dB_t$. The quadratic variation is

$$\sum_{n=1}^N \left|X_{t_n} - X_{t_{n-1}}\right|^2 - \int_{t_{n-1}}^{t_n} \sigma^2 ds$$

Let $y_n = \left|X_{t_n} - X_{t_{n-1}}\right|^2 - \int_{t_{n-1}}^{t_n} \sigma^2 ds$, $E[y_n] = 0$.

$$E\left[\left(\sum_{n=1}^{N} y_n\right)^2 = 2\sum_{n>m} E[y_n y_m] + \sum E[y_n^2]\right],$$

but y_n and y_m are not independent.

Note $X_t^2 - \int_0^t \sigma^2 ds$ is a martingale.

$$E\left[X_{t_n}^2 - \int_0^{t_n} \sigma^2 ds|\mathcal{F}_{t_{n-1}}\right] = X_{t_{n-1}}^2 - \int_0^{t_{n-1}} \sigma^2 ds \Rightarrow E\left[X_{t_n}^2 - X_{t_{n-1}}^2 - \int_{t_{n-1}}^{t_n} \sigma^2 ds|\mathcal{F}_{t_{n-1}}\right] = 0$$

We have

$$E\left[(X_{t_n} - X_{t_{n-1}})^2 - \int_{t_{n-1}}^{t_n} \sigma^2 ds|\mathcal{F}_{t_{n-1}}\right] = E\left[X_{t_n}^2 + X_{t_{n-1}}^2 - 2X_{t_n}X_{t_{n-1}} - \int_{t_{n+1}}^{t_n} \sigma^2 ds|\mathcal{F}_{t_{n-1}}\right]$$

But $-2X_{t_n}X_{t_{n-1}} = -2X_{t_n}E[X_{t_n}|\mathcal{F}_{t_{n-1}}]$, so $E[y_n y_m] = 0$, even though they are not independent.

Now, we compute $E[y_n^2]$. We want to show that

$$E[y_n^2] = E\left[(X_h^2 - A_h)^2\right] \leq Ch^2,$$

where $X_h = \int_0^h \sigma dB$, $A_h = \int_0^h \sigma^2 dB$. Use the martingale generating function,

$$\begin{aligned}
1 = E[e^{\lambda X_h - \frac{\lambda^2}{2}A_h}] &= E\left[\left(1 + \lambda X + \frac{\lambda^2}{2}X^2\right)\left(1 - \frac{\lambda^2}{2}A + \frac{\lambda^4}{8}A^2 + \cdots\right)\right] \\
&= 1 + \cdots + \lambda^4 E\left[\frac{X^4}{4!} + \frac{A^2}{8} - \frac{1}{4}X^2 A\right], \\
\Rightarrow & E\left[\frac{X^4}{4!} + \frac{A^2}{8} - \frac{1}{4}X^2 A\right] = 0
\end{aligned}$$

$$E[X^4] = 6E[X^2 A] - 3E[A^2] \leq \frac{1}{2}E[X^4] + 72E[A^2] + 3E[A^2]$$

The last inequality is by $ab \leq \frac{1}{2}(a^2 + b^2)$. This shows that $E[y_n^2] \leq CE[A^2] \leq Ch^2$.

General Ito's Formula (1D): Consider

$$X_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$$

Let $f \in C^2$, $\xi_n \in [t_{n-1}, t_n]$.

$$\begin{aligned}
f(X_t) - f(X_0) &= \sum_{n=0}^{N} f(X_{t_n} - X_{t_{n-1}}) \\
&= \sum_{n=1}^{N} f'(X_{t_{n-1}})(X_{t_n} - X_{t_{n-1}}) + \frac{1}{2}f''(X_{\xi_n})(X_{t_n} - X_{t_{n-1}})^2 \\
&= \sum_{n=1}^{N} f'(X_{t_{n-1}}) \int_{t_{n-1}}^{t_n} b(s, X_s)ds + \sum_{n=1}^{N} f'(X_{t_{n-1}}) \int_{t_{n-1}}^{t_n} \sigma dB \\
&+ \sum_{n=1}^{N} \frac{1}{2}f''(X_{t_{n-1}})(X_{t_n} - X_{t_{n-1}})^2 + \sum_{n=1}^{N} \frac{1}{2}(f''(X_{\xi_n} - f''(X_{t_{n-1}}))(X_{t_n} - X_{t_{n-1}})^2 \\
&\to \int_0^t \left(b\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}\right) ds + \int_0^t \sigma \frac{\partial f}{\partial x} dB
\end{aligned}$$

More generally, if $f(t,X)$ for $X \in \mathbb{R}^d$,

$$f(t,X_t) - f(0,X_0) = \int_0^t (\partial_s + L) f(s,X_s) ds + \int_0^t \sum_{i,j} \frac{\partial f}{\partial x_i} \sigma_{ij} dB_j,$$

where $L = \frac{1}{2}\sum_{i,j=1}^d a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial}{\partial x_i}$ is the general second order linear elliptic operator, and $a = \sigma\sigma^T \geq 0$.

In differential form:

$$df(t,X_t) = \frac{\partial f}{\partial t} dt + \sum_i \frac{\partial f}{\partial x_i} dX_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} dX_i dX_j$$

***Theorem:* 8.4: Martingale Integration by Parts**

Let M_t be a continuous martingale, A_t be continuous, progressively measurable, with bounded variation. Assume $E\left[\sup_{0\leq s\leq t} |M_s| \, \|A_s\|_{TV}\right] < \infty$. Then $M_t A_t - \int_0^t M_s dA_s$ is a martingale.

Proof. Write $\int_0^t M_s dA_s = \sum_{j=1}^N M_{t_j}(A_{t_j} - A_{t_{j-1}})$. Take expectation conditional on $\mathcal{F}_s$.

$$E\left[\sum_{j=1}^N M_{t_j}(A_{t_j} - A_{t_{j-1}}) | \mathcal{F}_s\right] = \sum_{j, t_{j-1} > s} M_{t_j} A_{t_j} - M_{t_{j-1}} A_{t_{j-1}} \to M_t A_t - M_s A_s$$

□

Another proof of Ito's Formula: Consider $e^{i\lambda(X_t - \int_0^t b ds) + \frac{1}{2}\int_0^t \lambda^T a \lambda}$. It is a martingale.

Let $A_t = e^{i\lambda \int_0^t b ds - \frac{1}{2}\int_0^t \lambda^T a \lambda ds}$.

$$M_t A_t - M_0 A_0 - \int_0^t M_s dA_s = e^{i\lambda X_t} - e^{i\lambda X_0} - \int_0^t e^{i\lambda X_s (i\lambda b - \frac{1}{2}\lambda^T a \lambda) ds}$$

$$= e^{i\lambda X_t} - e^{i\lambda X_0} - \int_0^t L e^{i\lambda x} = \int \sigma \cdot \nabla f dB$$

is a martingale, and $f(x) = e^{i\lambda x}$. Now take linear combinations and use Fourier transform, so we can prove for smooth f with compact support.

Now we know that $f(t,X_t) - f(0,X_0) - \int_0^t (\partial_s + L) f(s,X_s) ds$ is a Martingale. Let $s < t$, $X_s = x$,

$$E\left[f(t,X_t) - f(s,X_s) - \int_s^t (\partial_u + L) f(u,X_u) du | \mathcal{F}_s\right] = 0$$

$$\int f(t,y) p(s,x,t,y) - f(s,x) - \int_s^t \int (\partial_u + L) f(u,y) p(s,x,u,dy) du = 0$$

This means that $p(s,x,t,A) = p(X_t \in A, X_s = s)$ is a weak solution of

$$\partial_t p = L_y^* p, t > s$$

$$L^* f = \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij} f) - \sum_i \frac{\partial}{\partial x_i}(b_i f)$$

with $p(s,x,t,dy) \to \delta_0(x-y) dy$ as $t \to s$. This is Fokker Planck forward equation. If $\lambda^T a \lambda \geq \delta |\lambda|^2$, strictly elliptic, then weak becomes normal solution.

Suppose $u(s,x)$, $0 \leq s \leq t$ solves $(\partial_s + L) u = 0$ and $u(t,y) = 1_A(y)$, then it gives the backward heat equation, $E_{x,s}[u(t,X_t)] = u(s,x)$. This is the backward Kolmogorov equation. $(\partial_s + L_x) p(s,x,t,A) = 0$, $p(s,x,t,A) = 1_A$ as $s \to t$.

Relationship with Measures Consider $e^{\lambda\left(X_t-\int_0^t bds\right)-\frac{\lambda^2}{2}\int_0^t ds}$, where $a=\sigma^2$ is a martingale for each λ w.r.t. $\Omega=C([0,T])$, $\mathcal{F}_t=\sigma(B_s,s\le t)$. It is a probability measure on $C([0,T])$, $P_x^{a,b}(A)=P(X\in A)$. The Brownian motion $P_0^{1,0}$ is Wiener measure.

From Ito's formula, the operator $L_t=a(t,x)\frac{\partial^2}{\partial x^2}+b(t,x)\frac{\partial}{\partial x}$ for $dx=bdt+\sigma dB$, $a=\sigma^2$ give transition probability $p(s,x,t,A)=P(X_t\in A|X_s=x)$. For the forward equation $-\partial_s\phi=L_sp$ (L acting on s), $\lim_{s\to t}p(s,x,t,A)=\begin{cases}1,x\in A\\0,x\notin A\end{cases}$. In case $a=1,b=0$, the Brownian motion, $p(s,x,t,A)=\int_A\frac{e^{-\frac{(y-x)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}dy$.

Relationship with Heat Equation Let $L_t=\frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^d b_i\frac{\partial}{\partial x_i}$, D be a domain with nice boundary. Consider the heat equation $\begin{cases}Lu=0,x\in D\\u=f,x\in\partial D\end{cases}$. We want to show that $u(x)=E_x[f(x_\tau)]$, where $\tau=\inf\left\{t\ge 0,B_t\in D^C\right\}$.

Let u solve the differential equation, $u(X_{t\wedge\tau})$ is a martingale. If $\lambda^T a\lambda>\delta|\lambda|^2$, $\delta>0$ (a is positive definite), then as $t\to\infty$, $u(x)=E_x[u(X_{t\wedge\tau})]\to E_x[u(X_\tau)]$.

Poisson Equation $\begin{cases}Lu=g,x\in D\\u=0,x\in\partial D\end{cases}$. $u(X_{t\wedge\tau})-\int_0^{t\wedge\tau}g(X_s)ds$ is a martingale. This is based on Ito's formula with the fact that $\int_0^t\sum_{i,j}\frac{\partial f}{\partial x_i}\sigma_{ij}dB_j$ is a martingale, and the function u has no dependency on t. Take $t\to\infty$, $u(x)=E_x\left[\int_0^\tau g(X_s)ds\right]$.

Feynman-Kac $\begin{cases}-\frac{\partial u}{\partial t}=Lu+Vu\\u(T,x)=f(x)\end{cases}$ (a time-reversed heat equation). Apply Ito's formula to $u(s,X_s)e^{\int_t^s V(u,X_u)du}$ with $s>t$. Note that $\left(\frac{\partial}{\partial t}+L_t\right)u=-Vu$, but when we differentiate $e^{\int_t^s V(u,X_u)du}$, we get another Vu, so by product rule, the first order term is zero, and $u(s,X_s)e^{\int_t^s V(u,X_u)du}$ is a martingale. Since $E_{t,x}[M_T]=E[M_T|\mathcal{F}_t]=M_t$, $E_{t,x}\left[u(T,X_T)e^{\int_t^T V(u,X_u)du}\right]=u(t,X)$. As long as f is nice and $V\le 0$, the solution makes sense.

Change of Measure Consider the measure $P_x^{a,b}$ on $C([0,T])$.
The total variation is $\lim_n\sum_n\left|X_{t_n}-X_{t_{n-1}}\right|^2=\int_s^t a(u,X_u)du$, so if X or a is changed, we get some singular measure. If we change b, it is generally not changed. If we apply Radon-Nikodym, we get the following formula.

***Theorem:* 8.5: Cameron-Martin-Girsanov Formula**

$$\frac{dP_x^{a,b}}{dP_x^{a,0}}=e^{\int_0^t a^{-1}bdX_s-\frac{1}{2}\int_0^t ba^{-1}bds}$$

Proof. We just need to show that $e^{\lambda\left(X_t-X_0-\int_0^t bds\right)-\frac{\lambda^2}{2}\int_0^t ads+\int_0^t a^{-1}bdX_s-\frac{1}{2}\int_0^t ba^{-1}bds}$ is a martingale w.r.t. $P_x^{a,0}$. Because $\int_0^t a^{-1}bdX_s-\frac{1}{2}\int_0^t ba^{-1}bds$ is the formula, and $\lambda\left(X_t-X_0-\int_0^t bds\right)-\frac{\lambda^2}{2}\int_0^t ads$ is $P_x^{a,b}$.

Let $y_t=\int_0^t(\lambda+a^{-1}b)dX=\int_0^t(\lambda+a^{-1}b)\sigma dB$ under $P_x^{a,0}$.

By Ito's formula, $e^{y_t - y_0 - \frac{1}{2}\int_0^t a(\lambda + a^{-1}b)^2 ds}$ is a martingale under $P_x^{a,0}$.

$$y_t - y_0 = \lambda(X_t - X_0) + \int_0^t a^{-1} b dX_s$$
$$\frac{1}{2}\int_0^t a(\lambda + a^{-1}b)^2 ds = \frac{\lambda^2}{2}\int_0^t a ds + \lambda \int_0^t b^2 ds + \frac{1}{2}\int_0^t b a^{-1} b ds$$

□

How physicists prove this formula?

For Brownian motion, assuming $t_n = 1$

$$P(B_{t_1} \in dx_1, ..., B_{t_n} \in dx_n) = \frac{e^{-\frac{1}{2}\sum_{i=1}^n \frac{(B_{t_i} - B_{t_{i-1}})^2}{2(t_i - t_{i-1})}}}{\prod_{i=1}^n (2\pi(t_i - t_{i-1}))^{1/2}}$$

We know it is Gaussian with covariance $E[B_t B_s] = t \wedge s := c(t,s)$. We want to prove this. Define the linear operator C s.t. $Cf(t) = \int_0^1 c(t,s) f(s) ds$.

If P is Gaussian, it should be of the form $\frac{e^{-\frac{1}{2}X^T C^{-1} X}}{(2\pi)^{d/2}\sqrt{\det(C)}}$. Note that $\frac{1}{2}X^T C^{-1} X = \sup_y X^T y - \frac{1}{2} y^T C y$. Substitude $y = f(s) \in L^2$, $X = B_s$. Then for some optimizer f, we have:

$$\begin{aligned} B(s) &= \int_0^1 c(s,t) f(t) dt = \int_0^1 s \wedge t f(t) dt \\ &= \int_0^s t f(t) dt + \int_s^1 s f(t) dt \\ &= -\int_0^s f(t) dt + sF(1) \end{aligned}$$

So $B(s)$ satisfies $\partial_s B = -F(s) + F(1)$, $\partial_s^2 B = -F'(s) = -f(s)$. This implies $-\int B_s \dot{B}_s ds = \int \left|\dot{B}_s\right|^2 ds$.

Now we want to know what $X = B + f$, where $\dot{f} = b$, is. B is translation invariant by any translation function f. Its density should be

$$e^{-\frac{1}{2}\int_0^1 (\dot{X}_s - \dot{f}_s)^2 ds} = e^{\int_0^1 \dot{f}_s \dot{X}_s - \frac{1}{2}\int_0^1 \left|\dot{f}_s\right|^2 - \frac{1}{2}\int_0^1 \left|\dot{X}_s\right|^2 ds}$$

The final part is a Brownian motion, and the first two parts are just $e^{\int_0^1 b_s dB_s - \frac{1}{2}\int_0^1 b_s^2 ds}$.

Consider a Brownian motion $B(t)$, potential $V(x)$, h_0, $\frac{e^{\int_0^t V(B_s) ds + h_0(B_t)}}{E_x\left[e^{\int_0^t V(B_s) ds + h_0(B_t)}\right]}$ is a measure on $C_x([0,t])$ starting at x.

Let $Z(t,x) = E_x\left[e^{\int_0^t V(B_s) ds + h_0(B_t)}\right]$, from Feynman-Kac, we have $\begin{cases} \partial_t Z = \frac{1}{2}\partial_x^2 Z + VZ \\ Z(0,x) = e^{h_0(x)} \end{cases}$.

Apply log transform $h(t,x) = \log Z(t,x)$, $Z = e^h$, $\partial_t Z = e^h \partial_t h$, $\partial_x Z = e^h \partial_x h$, $\partial_x^2 Z = e^h (\partial_x h)^2 + e^h (\partial_x^2 h)$. Then, we get the Hopf-Cole equation $\begin{cases} \partial_t h = \frac{1}{2}\partial_x^2 h + \frac{1}{2}(\partial_x h)^2 + V \\ h(0,x) = h_0(x) \end{cases}$

Apply Ito to $h(t-s, X_s)$,

$$h_0(X_t) = h(t,X) - \int_0^t \partial_x h(t-s,X_s)dX_s + \int_0^t \left(-\partial_s + \frac{1}{2}\partial_x^2\right) h(t-s,X_s)ds,$$
$$\Leftrightarrow d_s h(t-s,B_s) = \left(-\partial_s + \frac{1}{2}\partial_x^2\right) ds + \partial_x h dB_s$$

but $\left(-\partial_s + \frac{1}{2}\partial_x^2\right) h = -\frac{1}{2}(\partial_x h)^2 - V$, then

$$\int_0^t V(t-s,B_s)ds = -h(0,B_t) + h(t,B_0) + \int_0^t \partial_x h dB - \frac{1}{2}\int_0^t (\partial_x h)^2 ds$$
$$\int_0^t V(t-s,B_s)ds = -h(0,B_t) + h(t,x) + \int_0^t b dB - \frac{1}{2}\int_0^t b^2 ds$$

The numerator becomes:

$$e^{\int_0^t \partial_x h(t-s,X_s)dX_s - \frac{1}{2}\int_0^t (\partial_x h)^2(t-s,X_s) + h(t,x)}$$

The $h(t,x)$ cancels, so the new measure is just a measure derived from an SDE with $a = I$, $b = \partial_x h(t-s, X_s)$, where h satisfies $\begin{cases} \partial_t h = \frac{1}{2}\partial_x^2 h + \frac{1}{2}(\partial_x h)^2 + V \\ h(0,x) = h_0(x) \end{cases}$

Special Case (Harmonic oscillators): $V = -\frac{x^2}{2} + \frac{1}{2}$.

$$Z(t,x) = e^{-\frac{x^2}{2}+\frac{1}{2}}$$
$$\partial_x Z = -xe^{-\frac{x^2}{2}+\frac{1}{2}} \quad \frac{1}{2}\partial_x^2 Z = \frac{1}{2}x^2 e^{-\frac{x^2}{2}+\frac{1}{2}} \quad \partial_t Z = 0$$
$$\Rightarrow \partial_t Z = \frac{1}{2}\partial_x^2 Z + VZ$$

If we let $b = \partial_x \log Z = -x$, we get $dX = -Xdt + dB$ ($X_t = e^{-\frac{1}{2}\int_0^t B_s^2 ds + \frac{t}{2} - \frac{B_t^2}{2}}$) is the Ornstein-Uhlenbeck process.

8.4 Representation Theorems

***Theorem:* 8.6: Ito's Representation Theorem**

Let $\Omega = C([0,1])$, $\mathcal{F}$ Borel σ-algebra on Ω, P a Wiener measure (Brownian motion). If $F \in L^2(\Omega, \mathcal{F}, P)$, then there is a unique square integrable, progressively measurable $f(t,\omega)$ s.t. $F = E[F] + \int_0^1 f(t,\omega)dB_t$. *i.e.* Every square integrable function can be represented by a stochastic integral.

Proof. **Uniqueness**: Suppose f_1, f_2 both work. By Ito's isometry, $E\left[\int_0^1 (f_1 - f_2)^2 ds\right] = 0$. Therefore, $f_1 = f_2$ as elements of $L^2([0,1] \times \Omega, dx \times dP)$.

Existence: Consider the example $F = e^{\int_0^1 h_s dB_s - \frac{1}{2}\int_0^1 h_s^2 ds}$.

$$F_t = e^{\int_0^t h dB_t - \frac{1}{2}\int_0^t h^2 ds} \quad dF_t = h_t F_t dB_t$$
$$\Rightarrow F_t = 1 + \int_0^t h_t F_t dB_t \text{ and } F_1 = 1 + \int_0^1 h_t F_t dB_t$$

Define $f_t = h_t F_t$, f_t is square integrable and progressively measurable.

If $F = E[F] + \int_0^1 f dB$ and $G = E[G] + \int_0^1 g dB$, then $F + G = E[F+G] + \int_0^1 (f+g) dB$, so it is true for any finite linear combinations of $e^{\int_0^1 h dB - \frac{1}{2}\int_0^1 h^2 ds}$ for $h \in L^2([0,1])$.

Now consider the limit. Suppose $F_n = E[F_n] + \int_0^1 f_n dB$ and $F_n \to F$ in $L^2(\Omega, \mathcal{F}, P)$, then $E[F_n] \to E[F]$. As $n, m \to \infty$,

$$E\left[\left(\int_0^1 f_n dB - \int_0^1 f_m dB\right)^2\right] = E\left[(F_n - F_m)^2\right] + (E[F_n] - E[F_m])^2 \to 0$$

so $f_n \to f$ in $L^2([0,1] \times \Omega)$ and

$$F_n = E[F_n] + \int_0^1 f_n dB \to F = E[F] + \int_0^1 f dB$$

so limits have Ito's representation, and it is true for the closure of linear span of h.

Now we show this closure is all of $L^2(\Omega, \mathcal{F}, P)$. All we need to check is that if $g \perp e^{\int_0^1 h_s dB_s - \frac{1}{2}\int_0^1 h_s^2 ds}$, for all $h \in L^2([0,1])$, then $g = 0$.

We can carefully construct h so that $\int_0^1 h dB_s = \lambda_1 B_{t_1} + \cdots \lambda_n B_{t_n}$. Then we have for any $\lambda_1, ..., \lambda_n$ and $0 \le t_1 \le \cdots \le t_n \le 1$,

$$\phi(\lambda_1, ..., \lambda_n) = E\left[g e^{\lambda_1 B_{t_1} + \cdots + \lambda_n B_{t_n}}\right] = 0$$

$\phi(\lambda_1, ..., \lambda_n)$ is analytic in $\mathbb{C}^n$ and vanishes for λ_i real. $\phi \equiv 0$ means $E\left[g e^{i\lambda_1 B_{t_1} + \cdots + i\lambda_n B_{t_n}}\right] = 0$, and this implies $E[g\phi(B_{t_1}, ..., B_{t_n})] = 0$ and $g = 0$. $\square$

Theorem: 8.7: Martingale Representation Theorem

Let M_t be a square integrable martingale w.r.t. $\mathcal{F}_t = \sigma(\{B_s, s \le t\})$, then there exists a unique square integrable, progressively measurable $\sigma(t, \omega)$ s.t. $M_t = M_0 + \int_0^t \sigma_s dB_s$. *i.e.* any martingale can be represented as a stochastic integral.

Proof. By Theorem 8.6, $M_t = E[M_t] + \int_0^t \sigma_s^t dB_s = M_0 + \int_0^t \sigma_s^t dB_s$ for each t.

Take $t_2 > t_1$, $E[M_{t_2} | \mathcal{F}_{t_1}] = M_{t_1}$, so $\int_0^{t_1} \sigma_s^{t_2} dB_s = \int_0^{t_1} \sigma_s^{t_1} dB_s$, and therefore $\sigma_s^{t_2} = \sigma_s^{t_1}$ on $s \in [0, t_1]$, so $\sigma^{t_1}(s) = \sigma^{t_2}(s) = \sigma(s)$. $\square$

8.5 Wiener Chaos

Let f be a deterministic function, $\int_0^1 \int_0^t f(s,t) dB_s dB_t$ should be well-defined. Generalize it by letting $\Delta_n = \{0 \le t_1 \le t_2 \le \cdots \le t_n \le 1\}$ (Weyl chamber) be increasing sequence of time. $f_n \in L^2(\Delta_n, dt_1 \cdots dt_n)$. Define

$$I_n(f_n) = \int_0^1 \int_0^{t_n} \cdots \int_0^{t_2} \int_0^{t_1} f_n(t_1, ..., t_n) dB_{t_1} \cdots dB_{t_n} = \int_{\Delta_n} f_n dB_{t_1} \cdots dB_{t_n}$$

Let $K_n = \left\{I_n(f_n) : f_n \in L^2(\Delta_n)\right\}$. We know that $E\left[I_n(f_n)\right] = 0$ as a stochastic integral. What is $E\left[(I_n(f_n))^2\right]$? By Ito's isometry:

$$n = 1, \quad E\left[\left(\int_0^1 f(t_1)dB_{t_1}\right)^2\right] = \int_0^1 f^2(t_1)dt_1 = \|f\|_2^2$$

$$n = 2, \quad E\left[\left(\int_0^1 \int_0^{t_1} f_2(t_1,t_2)dB_{t_1}dB_{t_2}\right)^2\right] = \int_0^1 E\left[\left(\int_0^{t_1} f_2 dB_{t_1}\right)^2\right] dt_2$$

$$= \int_0^1 \int_0^{t_1} f_2 dt_1 dt_2 = \|f_2\|_{L^2(\Delta_2)}^2$$

When $m \neq n$, *e.g.* $m = 2$, $n = 1$

$$E\left[I_n(f_n)I_m(f_m)\right] = E\left[\int_0^1 \int_0^{t_2} f_2 dB_{t_1}dB_{t_2} \int_0^1 f_1 dB_{s_1}\right] = E\left[\int_0^1 \sigma(t)f_1(t)dt\right] = 0$$

In summary, $E\left[I_n(f_n)I_m(f_m)\right] = \begin{cases} 0, m \neq n \\ \|f_n\|_{L^2(\Delta_n)}^2 \, m = n \end{cases}$, K_n are orthogonal subspaces on $L^2(\Omega, \mathcal{F}, P)$, where P is the Wiener measure (Brownian motion)

Theorem: 8.8: Wiener Decomposition

$L^2(\Omega, \mathcal{F}, P) = \bigoplus_{n=0}^{\infty} K_n$, where $K_n \cong L^2(\Delta_n)$. It is also called the Fock space. We can represent any Brownian r.v.s with linear span of stochastic integrals $X = \{f_0, f_1, ...\}$.

- K_0: constants
- K_1: Gaussian r.v.
- K_2: chi-squared r.v.

The Hermite polynomials are:

$$H_n(x) = (-1)^n \frac{1}{n!} e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$$

By Taylor expansion,

$$e^{tx - t^2/2} = \sum_{k=0}^{\infty} t^k H_k(x)$$

Suppose $f \in L^2(\Delta_1) = L^2([0,1])$, denote $f^{\otimes n}(t_1, ..., t_n) = f(t_1)f(t_2)\cdots f(t_n)$.

Claim: $I_n(f^{\otimes n}) = \|f\|_{L^2(\Delta_1)}^2 H_n\left(\frac{\int_0^1 f dB}{\|f\|_{L^2(\Delta_1)}}\right)$.

Proof. Start from the generic martingale:

$$e^{\lambda \int_0^1 f_s dB_s - \frac{\lambda^2}{2}\int_0^1 f_s^2 ds} = \sum_{k=0}^{\infty} \lambda^k \|f_s\|_{L^2(\Delta_1)}^k H_k\left(\frac{\int_0^1 f dB}{\|f\|_{L^2(\Delta_1)}}\right)$$

Consider $M_t = e^{\lambda \int_0^t f_s dB_s - \frac{\lambda^2}{2}\int_0^t f_s^2 ds}$, $dM_t = \lambda f_t M_t dB_t$. Iteratively, we get

$$\begin{aligned} M_t &= 1 + \lambda \int_0^t f_s M_s dB_s \\ &= 1 + \lambda \int_0^t f_s dB_s + \lambda^2 f_s \int_0^s f_u M_u dB_u dB_s \\ &= \sum_{k=0}^{\infty} \lambda^k I_k(f^{\otimes k}) \end{aligned}$$

□

8.6 Stochastic Differential Equations

SDEs as limits of Markov Chains Let $X_1, X_2, ...$ be a Markov chain on $\mathbb{R}^d$. $P(X_{n+1} \in A | X_1, ..., X_n) = P(X_{n+1} \in A | X_n) = p(X_n, A)$ the one step transition probability. Rescale so time steps h, and space has corresponding rescaling.

Define A_h jump step operator, $A_h f(x) = \int (f(y) - f(x)) P_h(x, dy)$. $P_h(x, dy)$ is a transition probability if and only if $f(X_{kh}) - \sum_{j=0}^{k-1} (A_n f)(X_{jh})$ is a martingale w.r.t. $\mathcal{F}_{kh} = \sigma(X_{jh}, j \le k)$ for $f \in C_0^\infty(\mathbb{R}^d)$.

Define

$$\begin{aligned} b_h^i &= \frac{1}{h} \int_{|y-x| \le 1} (y_i - x_i) P_h(x, dy) \\ a_h^{ij}(x) &= \frac{1}{h} \int_{|y-x| \le 1} (y_i - x_i)(y_j - x_j) P_h(x, dy) \end{aligned}$$

***Theorem:* 8.9:**

Assume the following:

1. $\lim_{h \to 0} \sup_{|x| \le R} \frac{1}{h} P_h(x, B_\epsilon^C(x)) = 0$, for all $\epsilon > 0$
2. $\lim_{h \to 0} \sup_{|x| \le R} \left| b_h^i(x) - b^i(x) \right| = 0$
3. $\lim_{h \to 0} \sup_{|x| \le R} \left| a_h^{ij}(x) - a^{ij}(x) \right| = 0$
4. $a^{ij}(x), b^i(x)$ continuous in the definition of SDE $dX_t = bdt + \sigma dB_t$, $a = \sigma\sigma^T$.
5. $\delta |\lambda|^2 \le \lambda^T a \lambda \le \delta^{-1} |\lambda|^2$ (unifromly elliptic)

Then as $h \to 0$, $P_{h,x} \to P_x^{a,b}$ on $C[0, \infty)$.

Consider $L = \frac{1}{2} \sum_{i,j=1}^d a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial}{\partial x_i}$ and L_h the discrete approximation by a_h^{ij}, b_h^i. If $f \in C_0^\infty(\mathbb{R}^d)$, then $L_h f \to Lf$. $\frac{1}{h} A_h f - L_h f \to 0$ if and only if 1,2,3 in Theoerem 8.9 are true.

If the limit exists, then the limit $f(X(t)) - \int_0^t Lf(X(s)) ds$ is a martingale, and $(\partial_s + L_x) p(s, x, A) = 0$, so it identifies the limit.

$$\begin{aligned} \frac{1}{h} A_h f &= \frac{1}{h} \int (f(y) - f(x)) P_h(x, dy) \\ &= \frac{1}{h} \int_{|y-x|>1} (f(y) - f(x)) P_h(x, dy) + \frac{1}{h} \int_{|y-x| \le 1} (f(y) - f(x)) P_h(x, dy) \end{aligned}$$

By Taylor expansion,

$$f(y)-f(x)=\sum_{i=1}^{d}\frac{\partial f}{\partial x_i}(x)(y_i-x_i)+\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 f}{\partial x_i\partial x_j}(x)(y_i-x_i)(y_j-x_j)+D^3f(\xi)(y-x)^3$$

Subtracting L_hf from $\frac{1}{h}A_hf$, we get

$$\begin{aligned}\frac{1}{h}A_hf-L_hf &\le \|f\|_\infty \frac{1}{h}\int_{|y-x|>1}P_h(x,dy)\\ &\quad+\frac{C_\xi}{h}\int_{|y-x|\ge\epsilon}P_h(x,dy)+\frac{C_\xi}{h}\int_{|y-x|<\epsilon}|y-x|^3P_h(x,dy)\end{aligned}$$

But we don't know if it actually converges. With Theorem 3.8, X_is i.i.d. random variables, $E[X_i]=0$, $Var(X_i)=1$, $S_n=X_1+\cdots+X_n$, we have $\frac{S_n}{\sqrt{n}}\to N(0,1)$ in distribution. $\frac{S_{\lfloor nt\rfloor}}{\sqrt{n}}$ converges to Brownian motions in the sense of finite dimensional distribution. However, $P_n\to P$ (Wiener measure) weakly requires functional convergence.

Theorem: 8.10: Kolmogorov–Chentsov

Suppose $X_t\in[0,1]$ have $E\left[|X_t-X_s|^\gamma\right]\le C\left|t-s\right|^{1+\delta}$ for $\gamma,\delta>0$, then

$$E\left[\left(\sup_{0\le s<t\le 1}\frac{|X_t-X_s|}{|t-s|^\alpha}\right)^\gamma\right]\le\tilde{C}$$

for $0\le\alpha\le\frac{\delta}{\gamma}$, and $\tilde{C}$ only depends on C.

Remark 15. It is related to α-Holder. Also, if we have a sequence of stochastic processes X_t^n, and they have the same C, then $\tilde{C}$ is fixed. (uniform equicontinuity)

Proof. Let $D_m=\left\{0,\frac{1}{2^m},\frac{2}{2^m},...,1\right\}$, $D=\bigcup D_m$.
Note that $\sup_{0\le s<t\le 1}\frac{|X_t-X_s|}{|t-s|^\alpha}$ is equivalent to $\sup_{0\le s<t\le 1,s,t\in D}\frac{|X_t-X_s|}{|t-s|^\alpha}$ by continuity.

Let $K_m=\sup_{s,t\in\Delta_m}|X_t-X_s|$, where $\Delta_m=\left\{s<t,t-s=\frac{1}{2^m}\right\}$. Then

$$E[K_m^\gamma]\le\sum_{s,t\in\Delta_m}E\left[|X_t-X_s|^\gamma\right]\le C2^m2^{-m(1+\delta)}=C2^{-m\delta}$$

For $s,t\in D$, $s_m\le s_{m+1}\le\cdots\le s<t\le\cdots\le t_{m+1}\le t_m$.

If $s\le t\in D$, $t-s\le 2^{-m}$, then either $t_m=s_m$ or $t_m=s_m+\frac{1}{2^m}$.

$$\begin{aligned}X_t-X_s&=\sum_{i=m}^{\infty}(X_{t_{i+1}}-X_{t_i})+X_{t_m}-X_{s_m}+\sum_{i=m}^{\infty}(X_{s_i}-X_{s_{i+1}})\\ &\le\sum_{i=m}^{\infty}K_{i+1}+K_m+\sum_{i=m}^{\infty}K_{i+1}\le 2\sum_{i=m}^{\infty}K_i\end{aligned}$$

So $\sup_{0\le s<t\le 1}\frac{|X_t-X_s|}{|t-s|^\alpha}\le 4\sup_m 2^{m\alpha}\sum_{i=m}^\infty K_i$. Then

$$\begin{aligned}\left(E\left[\left(\sup_{0\le s<t\le 1}\frac{|X_t-X_s|}{|t-s|^\alpha}\right)^\gamma\right]\right)^{1/\gamma}&\le 4\sum_{i=0}^{\infty}2^{i\alpha}\left\|K_i\right\|_\gamma\\ &\le 4C^{1/\gamma}\sum_{i=0}^{\infty}2^{i\left(\alpha-\frac{\delta}{\gamma}\right)}\end{aligned}$$

where $\|X\|_\gamma = \left(E\left[|X|^\gamma\right]\right)^{1/\gamma}$, and $\tilde{C} = C^{1/\gamma}\sum_{i=0}^{\infty} 2^{i\left(\alpha - \frac{\delta}{\gamma}\right)}$ is finite with the condition $0 \leq \alpha \leq \frac{\delta}{\gamma}$.

For multi-dimension, we replace D_m by $\left\{0, \frac{1}{2^m}, \frac{2}{2^m}, ..., 1\right\}^d$ and the proof is the same. □

Now, back to random walks, $h = \frac{1}{n}$, $B_n(t) = \frac{S_{nt}}{\sqrt{n}}$, $t = \frac{k}{n}$.

$$b_n(x) = \frac{1}{h}E\left[B_n\left(\frac{k+1}{n}\right) - B_n\left(\frac{k}{n}\right)\middle| B_n\left(\frac{k}{n} = x\right)\right] = 0$$

$$a_n(x) = \frac{1}{h}E\left[\left(B_n\left(\frac{k+1}{n}\right) - B_n\left(\frac{k}{n}\right)\right)^2\middle| B_n\left(\frac{k}{n} = x\right)\right] = 1$$

$$E\left[\left(B_n(t) - B_n(s)\right)^4\right] = \left(\sum_{j=1}^{n}\frac{X_j}{\sqrt{n}}\right)^4 = (t-s)^2$$

becase $\sum_{j=1}^{n}\frac{X_j}{\sqrt{n}} \sim \sqrt{t-s}Z$.

***Theorem:* 8.11: Donsker's Theorem**

$$\max_{0\leq t\leq 1}\frac{S_{\lfloor nt\rfloor}}{\sqrt{n}} \to \max_{0\leq t\leq 1} B_t$$

Example: Ehrenfest's Chain on $\{0, 1, .., N\}$, $P_{i,i-1}\frac{i}{N}$, $P_{i,i+1} = \frac{N-i}{N}$.

If has an invariant measure $\pi_i = 2^{-N}\binom{N}{i}$.

$$\begin{aligned}\pi_{i-1}P_{i-1,i} + \pi_{i+1}P_{i+1,i} &= 2^{-N}\binom{N}{i-1}\frac{N-i+1}{N} + 2^{-N}\binom{N}{i+1}\frac{i+1}{N}\\ &= 2^{-N}\binom{N}{i}\left(\frac{i}{N-i+1}\frac{N-i+1}{N} + \frac{N-i}{i+1}\frac{i+1}{N}\right)\\ &= 2^{-N}\binom{N}{i} = \pi_i\end{aligned}$$

Let $Y^N(t) = Y_{\lfloor Nt\rfloor}$ be jumps in times $\frac{1}{N} \pm \frac{1}{\sqrt{N}}$, $y = \frac{i-\frac{N}{2}}{\sqrt{N}}$. $h = \frac{1}{N}$

$$\begin{aligned}E\left[Y^N\left(t+\frac{1}{N}\right) - Y^N(t)|Y^N(t) = y\right] &= \frac{1}{\sqrt{N}}\frac{N-i}{N} - \frac{1}{\sqrt{N}}\frac{i}{N}\\ &= \frac{1}{\sqrt{N}}\frac{N-\left(\sqrt{N}y + \frac{N}{2}\right)}{N} - \frac{1}{\sqrt{N}}\frac{\sqrt{N}y+\frac{N}{2}}{N} = -\frac{2Y}{N}\\ b_h(y) &= -2y\end{aligned}$$

$$\begin{aligned}E\left[\left(Y^N\left(t+\frac{1}{N}\right) - Y^N(t)\right)^2|Y^N(t) = y\right] &= \frac{1}{N}\left(\frac{1}{2} - \frac{y}{\sqrt{N}} + \frac{1}{2} + \frac{y}{\sqrt{N}}\right) = \frac{1}{N}\\ a_h(y) &= 1,\end{aligned}$$

so $Y^N \to dy = -2ydt + dB$ (Ornstein-Uhlenbeck)

The invariant distribution $\frac{i-\frac{N}{2}}{\sqrt{N}} \to \mathcal{N}\left(0, \sigma^2 = \frac{1}{4}\right)$.

Apply Ito's rule to $dy = -2ydt + dB$, we get the differential operator $L = \frac{1}{2}\frac{\partial^2}{\partial y^2} - 2y\frac{\partial}{\partial y}$. It should have an invariant measure $P_t(x, dy)$, which will give $\int \mu(dx)P_t(x, dy) = \mu(dy)$ and $\int \mu(dx)Lf(x) = 0$ for all $f \in C_0^2(\mathbb{R})$

Theorem: 8.12: Burkholder's Inequality

$$E\left[|M_n|^p\right] = C_p E\left[\left(\sum_{k=1}^{n-1}(M_{k+1} - M_k)^2\right)^{p/2}\right]$$

Proof. By Induction on p. □

Theorem: 8.13: Markov Chain Convergence Theorem

$X_t(nh)$ is a Markov chain taking steps every time h s.t.

1. $E\left[X((n+1)h) - X(nh)|\mathcal{F}_{nh}\right] = hb_h(X(nh))$, $\sup_{|x|\le R}|b_h(x) - b(x)| \to 0$ as $h \to 0$ for all R
2. $E\left[(X((n+1)h) - X(nh))^2|\mathcal{F}_{nh}\right] = ha_h(X(nh))$, $\sup_{|x|\le R}|a_h(x) - a(x)| \to 0$ as $h \to 0$ for all R
3. $E\left[(X((n+1)h) - X(nh))^4|\mathcal{F}_{nh}\right] \le C(X(nh))h^2$, $\sup_{|x|\le R}|C(x)| < \infty$ for all R

Then $X_t(t) \to X_t$ with distribution $P_x^{a,b}$, *i.e.* $dX_t = bdt + \sigma dB$ where $\sigma = \sqrt{a}$.
$f(X_h(t)) - \sum_j L_h f(X_j)$ is a martingale, $L_h f \to Lf$. $Lf = \int (f(y) - f(x))P_h(x, dy) = E_x[f(X(h)) - f(X)]$.

Proof. Let $t = kh, s = jh$ s.t. $0 \le s < t \le 1$. we want to show $E\left[(X(t) - X(s))^4\right] \le C|t-s|^2$.

Localize $\tau_R = \inf\{t \ge 0, |X(t)| \ge R\}$, $X_R(t) = X(t \wedge \tau_R)$. We will show that X_R is tight, *i.e.* $X_{hR} \to X_R$ as $h \to 0$. If we do this for each R, it is fine. Let $R \to \infty$, we can pretend a, b are bounded.

Assume $b_h = 0$, $X((n+1)h) - X(nh) = hb_h(X(nh)) + \tilde{X}((n+1)h) - \tilde{X}(nh)$.

$$X(kh) - X(jh) = \sum_{n=j}^{k} hb_h(X(nh)) + \tilde{X}(kh) - \tilde{X}(jh)$$

where $\sum_{n=j}^{k} hb_h(X(nh)) \to \int_s^t b(X(s))ds$.

$$\begin{aligned}
E\left[(X(kh) - X(jh))^2\right] &= E\left[\left(\sum_{n=j}^{k-1} X((n+1)h) - X(nh)\right)^2\right] \\
&= \sum_{n=j}^{k-1} E\left[(X((n+1)h) - X(nh))^2\right] \\
&\quad + 2\sum_{n_1<n_2=j}^{k-1} E\left[(X((n_2+1)h) - X(n_2h))(X((n_1+1)h) - X(n_1h))\right]
\end{aligned}$$

Precondition on $\mathcal{F}_{nh}$, the second term becomes zero.

$$\sum_{n=j}^{k-1} E\left[(X((n+1)h) - X(nh))^2\right] = \sum_{n=j}^{k-1} ha_h(X(kh)) \le C(k-j)h = C|t-s|$$

The same method won't work for the fourth order term, because the expansion contains more terms. Instead use Theorem 8.12

$$\begin{aligned}
E\left[(X(kh)-X(jh))^4\right] &= E\left[\sum_{n=j}^{k-1}(X((n+1)h)-X(jh))^4-(X(nh)-X(jh))^4\right]\\
&= E\left[\sum_{n=j}^{k-1}4(X((n+1)h)-X(nh))(X(nh)-X(jh))^3\right]\\
&+E\left[\sum_{n=j}^{k-1}6(X((n+1)h)-X(nh))^2(X(nh)-X(jh))^2\right]\\
&+E\left[\sum_{n=j}^{k-1}4(X((n+1)h)-X(nh))^3(X(nh)-X(jh))\right]\\
&+E\left[\sum_{n=j}^{k-1}(X((n+1)h)-X(nh))^4\right]
\end{aligned}$$

By $A^3B \le \frac{1}{2}A^4 + \frac{1}{2}A^2B^2$, we reduce the third term. By preconditioning, the first term is zero.

$$\begin{aligned}
&\le E\left[\sum_{n=j}^{k-1}8(X((n+1)h)-X(nh))^2(X(nh)-X(jh))^2\right]+3E\left[\sum_{n=j}^{k-1}(X((n+1)h)-X(nh))^4\right]\\
&\le \sum_{n=j}^{k-1}Ch(n-j)h+Ch^2\\
&= Ch^2(k-j)^2+Ch^2(k-j)\\
&= C\left|t-s\right|^2
\end{aligned}$$

If $t-s>h$, $(k-j)h^2 = h(t-s) \le (t-s)^2$.

If $t-s<h$, $\frac{1}{h^2} < \frac{1}{(t-s)^2}$.

$$E\left[(X(t)-X(s))^4\right] \le E\left[(t-s)\left(\frac{X((n+1)h)-X(nh)}{h}\right)^4\right] \le \frac{C(t-s)^4}{h^4}h^2 \le C\left|t-s\right|^2$$

□

Example (Random time change): Given

$$dX_t = bdt + \sigma dB_t$$

Let $\tau_t = \int_0^t \gamma(s,\omega)ds$, where γ is progressively measurable. Let $y_t = X_{\tau_t}$. Find the SDE for y_t.

$$\begin{aligned}
y(t+h)-y(t) &= X(\tau(t+h))-X(\tau(t)) = X(\tau(t)+\tau'(t)h+O(h^2))-X(\tau(t))\\
X(t+\epsilon)-X(t) &\sim \sigma(t,X_t)\sqrt{\epsilon}Z+b(t,X_t)\epsilon\\
y(t+h)-y(t) &\sim X(\tau(t)+\tau'(t)h)-X(\tau(t))\\
&\sim \sigma(\tau(t),y(t))\sqrt{\tau'(t)}\sqrt{h}Z+b(\tau(t),y(t))\tau'(t)h\\
dy_t &= \sigma(\tau_t,y_t)\sqrt{\tau'(t)}d\tilde{B}_t+b(\tau_t,y_t)\tau'(t)dt
\end{aligned}$$

The computation/proof is based on Theorem 8.13.

Population Model: Let $\xi_1, \xi_2, ...$ be i.i.d. values in $\{0, 1, 2, ...\}$ s.t. $E[\xi_i] = 1$, $Var(\xi_i) = \sigma^2 < \infty$, $E[\xi_i^4] < \infty$.

Let $X_n = \{0, 1, ...\}$ be size of population at time n. $X_{n+1} = \xi_1 + \xi_2 + \cdots + \xi_{X_n}$. Let $h = \frac{1}{N}$, $S_n h = c_N X_n$.

$$\begin{aligned} E\left[S_{(n+1)h} - S_{nh}|\mathcal{F}_{nh}\right] &= c_N E\left[\xi_1 + \cdots + \xi_{X_n} - X_n\right] = 0 \\ E\left[(S_{(n+1)h} - S_{nh})^2|\mathcal{F}_{nh}\right] &= c_N^2 E\left[(\xi_1 + \cdots + \xi_{X_n} - X_n)^2|\mathcal{F}_{nh}\right] \\ &= c_N^2\sigma^2 X_n = c_N^2\sigma^2\frac{S_{nh}}{c_N} = c_N\sigma^2 S_{nh} \end{aligned}$$

Therefore, $c_N = \frac{1}{N}$, $\frac{S_{\lfloor Nt \rfloor}}{N} \to y_t$, where $dy = \sigma\sqrt{y}dB + rydt$

8.7 Explosion

We know that ODEs can explode in finite time. For example, $\dot{x} = x^2$ with $x(0) = x_0 > 0$. Here the solution is $x(t) = (x_0^{-1} - t)^{-1}$. When $x_0 = 1$, explode at $t = 1$.

Equivalently, $dx = x^2 dt + dB$ will have trouble as well.

We want to find conditions s.t. if $\tau_l = \inf\{t \geq 0, x(t) \geq l\}$, then $P(\tau_l \leq T) \to 0$ as $l \to \infty$.

***Proposition:* 8.1:**

Suppose there exists smooth u s.t. $u(x) \geq 0$, $u(x) \to \infty$ as $|x| \to \infty$ and $(Lu)(x) \leq Cu(x)$ for $0 < C < \infty$, where L is the Ito's differential operator. Then there is no explosion, *i.e.* $P(\tau_l \leq T) \to 0$ as $l \to \infty$.

Proof. $Z_t = e^{-Ct}u(x_t)$ is a supermartingale.

$Z_t - Z_0 - \int_0^t (L - C)u(X_s)ds = M_t$ is a martingale. ($(L - C)u \leq 0$) By Theorem 7.7,

$$\inf_{|x| \geq l} E_x[e^{-C\tau_l}] \leq E_x[e^{-C\tau_l}u(X(\tau_l))] = E_x[Z_{\tau_l}] \leq E_x[Z_0] = u(x)$$

Therefore, by Theorem 2.1, as $l \to \infty$,

$$P(\tau_l \leq T) \leq \int_{\tau_l \leq T} e^{C(T - T_l)}dP \leq e^{CT}E_x[e^{-C\tau_l}] \leq e^{CT}\frac{u(x)}{\inf_{|x| \geq l} u(x)} \to 0$$

□

***Proposition:* 8.2:**

If $\|a(x)\| \leq C(1 + |x|^2)$ and $|L(x)| \leq C(1 + |x|)$, then there is no explosion

Proof. Try $u(x) = 1 + |x|^2$. Then

$$\begin{aligned} Lu &= \frac{1}{2}\sum_{i,j} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j}u + \sum_i b_i \frac{\partial}{\partial x_i}u \\ &\leq C\|a\| + C(1 + |x|)|x| \\ &\leq C(1 + |x|^2) \leq Cu(x) \end{aligned}$$

It goes the other way. Suppose $Lu \geq Cu$, then Z_t is a submartingale. $E_x[Z_{\tau_l}] \geq u(x)$, so $E[e^{-C\tau_l}] \geq \frac{u(x)}{\inf_{|x|\geq l}} u(x)$, so now if u is bounded, then there is an explosion. □

For the example $dx = x^2 dt + dB$, $L = \frac{1}{2}\frac{\partial^2}{\partial x^2} + x^2 \frac{\partial}{\partial x}$. We want to find bounded u s.t. $Lu \geq Cu$.

$u(x) = \begin{cases} e^{-2x^{-1}}, x > 0 \\ 0, x \leq 0 \end{cases}$ works.

$$\frac{\partial u}{\partial x} = 2x^{-2}e^{-2x^{-1}} \quad x^2\frac{\partial u}{\partial x} = 2u(x)$$
$$\frac{\partial^2 u}{\partial x^2} = (-4x^{-3} + 4x^{-4})e^{-2x^{-1}} = (-4x^{-3} + 4x^{-4})u(x)$$

By $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$ for $\frac{1}{p} + \frac{1}{q} = 1$, $p, q > 1$. We have $x^{-3} \leq \frac{3}{4}x^{-4} + \frac{1}{4}$.

Then $-4x^{-3} \geq -3x^{-4} - 1$, and $\frac{\partial^2 u}{\partial x^2} \geq u(x)$. Then $Lu(x) \geq Cu(x)$, and it explodes.

Example (Population model): $L = \sigma^2 y \frac{\partial^2}{\partial y^2} + ry\frac{\partial}{\partial y}$.

A trick for $dy = \sigma\sqrt{y}dB$: let $X(t) = e^{-y(t)\phi(T-t)}$, with ϕ deterministic. With Ito's formula,

$$dX = \left(\dot{\phi} + \frac{\sigma^2}{2}\phi^2\right) X dt - \phi\sigma X\sqrt{y}dB$$

Choose $\begin{cases} \dot{\phi} = -\frac{\sigma^2}{2}\phi^2 \\ \phi(0) = \lambda \end{cases}$, then X is a martingale. The solution is $\phi(t) = \left(\lambda^{-1} + \frac{\sigma^2}{2}t\right)^{-1}$. Then

$$E[e^{-\lambda y(t)}] = e^{-y(0)\phi(T)} = e^{-y_0\left(\lambda^{-1} + \frac{\sigma^2}{2}T\right)^{-1}}$$

This is weak uniqueness (uniqueness in distribution).

Now we want to solve $dx = x^{\alpha/2}dB$ $(0 < \alpha < 1)$. $x(t) = 0$ is a solution.

Time Change: From the generic SDE:

$$dx = \sigma(x)dB + b(x)dt$$

Step 1: remove b, *i.e.* $b = 0$, by Theorem 8.5.

Step 2: $dx = \sigma(x)dB$. Let $y_t = x(\tau_t)$,

$$dy_t = \sqrt{\tau'(t)}\sigma(x_t)d\tilde{B},$$

Take $\tau' = \frac{1}{a}$, where $a = \sigma^2$. Then y is a Brownian motion.

For $\sigma(x) = x^{\alpha/2}$, $a(x) = |x|^\alpha$, $x(t) = B(\tau_t)$, where $\int_0^{\tau_t} \frac{ds}{a(B(s))} = t$, so $dx = \sigma(x)d\tilde{B}$.

If $E\left[\int_0^t \frac{ds}{|B(s)|^\alpha}\right] < \infty$, then the solution is non-unique. $X(t) = B(\tau_t)$ is a well-defined solution.

$$E\left[\int_0^t \frac{ds}{|B(s)|^\alpha}\right] = \int_0^t \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s}} \frac{e^{-y^2/2s}}{|y|^\alpha} dy ds = \int_0^t s^{-\alpha/2} ds \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{e^{-y^2/2}}{|y|^\alpha} dy < \infty$$

for $0 < \alpha < 1$. The boundary for uniqueness is $\sigma(x) = x^{1/2}$ for $d = 1$.

Relevant example in ODE: $dx = 2\sqrt{x}dt$, $x(0) = 0$ has non-trivial solution $(x-a)^2 1_{x>a}$.

8.8 Markov Chain, Semigroup and Invariant Measures

Recall Markov chain.

$$P(s,x,t,A) = P(X(t) \in A | X(s) = x) = P(t-s,x,A)$$

The probability is time-homogenous. The Chapman-Kolmogorov equation is

$$\int P(s,x,dy)P(t,y,A) = P(s+t,x,A)$$

Consider the operator P_t defined by:

$$(P_t f)(x) = E_x[f(X(t))] = E[f(X(t))|X(0)=x] = \int P(t,x,dy)f(y)$$

$$\begin{aligned}(P_t(P_s f))(x) &= \int P(t,x,dy)(P_s f)(y) \\ &= \iint P(t,x,dy)P(s,y,dz)f(z) \\ &= \int P(t+s,x,dz)f(z) \\ &= (P_{t+s}f)(x)\end{aligned}$$

so as an operator, $P_t \circ P_s = P_{t+s}$, where $P_0 = Id$. This is a *semigroup* of operators. The inverse P_{-t} does not exist, so it is not a group.

Let $\mathcal{L} = \frac{d}{dt}P_t|_{t=0}$ (infinitesmal generator of Markov chain), $P_t = e^{t\mathcal{L}}$. This igves an infinitesimal generator. For $(\mathcal{L}f)(x) = \lim_{t\to 0}\frac{P_t f(x) - f(x)}{t}$ to be well defined, we need some domain.

In our case,

$$\mathcal{L} = \frac{1}{2}\sum_{i,j} a_{ij}(X)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_i b_i(x)\frac{\partial}{\partial x_i}$$

$$E_x\left[f(X(t)) - f(X(0)) - \int_0^t \mathcal{L}f(X(s))ds\right] = 0$$

$$\lim_{t\to 0}\frac{E_x[f(X(t))] - f(x)}{t} = \lim_{t\to 0}\frac{1}{t}\int_0^t E_x[\mathcal{L}f(X(s))]ds = \mathcal{L}f(x)$$

Examples:

- Heat equation:

$$(e^{t\mathcal{L}}f)(x) = u(t,x) \Leftrightarrow \begin{cases}\partial_t u = \mathcal{L}u \\ u(0,x) = f(x)\end{cases} \quad \Leftrightarrow u(t,x) = \int P(t,x,dy)f(y)$$

- $P_t 1 = 1$, $\mathcal{L}1 = 0$
- If $f \geq 0$, $P_t f \geq 0$, but this is not necessarily true for all semigroups
- Let ϕ be convex,

$$\mathcal{L}\phi(f) = \lim_{t\to 0}\frac{E_x[\phi(f(X(t)))] - \phi(f(x))}{t} \geq \lim_{t\to 0}\frac{\phi\left(E_x[f(X(t))]\right) - \phi(f(x))}{t} = \phi'(f(x))\mathcal{L}f(x)$$

- Carre du champ: $\mathcal{L}f^2 - 2f\mathcal{L}f \geq 0$

For the last one, note that $f(x_t) - \int_0^t \mathcal{L}f(x_s)ds = M_t$ is a Martingale.

$$\begin{aligned}
dM_t^2 &= df^2(x_t) - 2f(x_t)\int_0^t \mathcal{L}f + \left(\int_0^t \mathcal{L}f\right)^2 \\
&= \mathcal{L}f^2 - 2\mathcal{L}f\int_0^t \mathcal{L}f - 2\mathcal{L}f\int_0^t \mathcal{L}f + dN_t \\
E[M_t^2] &= \int_0^t E[\mathcal{L}f^2 - 2f\mathcal{L}f](s)ds
\end{aligned}$$

Example: Half-Laplacian $\mathcal{L} = \frac{1}{2}\Delta$.

$$\frac{1}{2}\Delta f^2 - f\Delta f = \frac{1}{2}\nabla(2f\nabla f) - f\Delta f = |\nabla f|^2$$

Invariant measure Let $\mu_0(x)$ be distribution of X_0, X_t has distribution

$$\mu_t(A) = \int P_t(y, A)\mu_0(dy) = P_t^*\mu_0,$$

where P_t^* is adjoint of P_t. μ is an *invariant measure* if $\mu_0 = \mu_t$.

$$\begin{aligned}
\mu(A) &= \int P_t(y, A)\mu(dy) \\
\int_A f(x)\mu(dx) &= \int_A\int_Y P_t(y, dx)\mu(dy)f(x) = \int P_t f(y)\mu(dy)
\end{aligned}$$

If μ is invariant, for nice fs, $\int(P_t f - f)d\mu = 0$

$$\lim_{t\to 0}\int \frac{1}{t}(P_t f - f)d\mu = 0, \int \mathcal{L}f d\mu = 0,$$

Typically, write as $\mathcal{L}^*\mu = 0$.

Example: For $\mathcal{L}$ the diffusion operator from Ito, $d = 1, b = 0$,

$$\mathcal{L}^*g = \frac{1}{2}\frac{\partial^2}{\partial x^2}(a(x)g(x)) = 0 \Rightarrow g(x) = \frac{Ax + B}{a(x)},$$

but we need $\int \frac{Ax+B}{a(x)}dx = 1$, and A, B may not exist, so there may not be an invariant measure.

Example: In $\mathbb{R}^d$, let $d\mu = gdx = \frac{e^{-H(x)}}{z}dx$, where z is a number. Let f be some nice function. If we can solve the Dirichlet form below, then the $\mathcal{L}$ in the equation will have μ as the invariant measure.

$$\begin{aligned}
D(f) &= \frac{1}{2}\int |\nabla f|^2 d\mu = \frac{1}{2}\int |\nabla f|^2 \frac{e^{-H(x)}}{z}dx = \int f(-\mathcal{L})f\frac{e^{-H(x)}}{z}dx \\
\text{LHS} &= -\frac{1}{2}\int f\nabla(\nabla f e^{-H})\frac{dx}{z} \\
&= -\int f\left(\frac{1}{2}\Delta f - \frac{1}{2}\nabla H\nabla f\right)\frac{e^{-H}}{z}dx
\end{aligned}$$

Define $\mathcal{L}f = \frac{1}{2}\Delta f - \frac{1}{2}\nabla H\nabla f$. It will have μ as the invariant measure. Consider $f = 1$. ($\frac{1}{2}\int \nabla\tilde{f}\nabla f d\mu = \int \tilde{f}(-\mathcal{L})f d\mu$, replace $\tilde{f} = 1$, and we get $\mathcal{L}^*f = 0$)

When we have $H = -\frac{|x|^2}{2}$ the harmonic oscillator, $\frac{1}{2}\nabla H = -\frac{1}{2}x$, we get the Ornstein-Ulhenbeck process in $\mathbb{R}^d$, $\frac{1}{2}\Delta f - \frac{1}{2}x\nabla f$.

Jump process Let x be start point, y_i be targets, $\lambda_i(x)$ be rates, $\lambda = \sum \lambda_i$.

$$P(n \text{ jumps in time } t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$P(n=0) = e^{-\lambda t} \sim 1 - \lambda t \quad P(n=1) = \lambda t e^{-\lambda t} \sim \lambda t \quad P(n=2) = O(t^2)$$

$$\lim_{t\to 0} \frac{E_x[f(X(t))] - f(x)}{t} = \lim_{t\to 0} \frac{\sum_i P_x(\text{jump to } i \text{ before others in time } t)(f(y) - f(x))}{t}$$

$$= \lim_{t\to 0} \frac{\sum_i (1 - e^{-\lambda_i t}) \prod_{j\neq i} e^{-\lambda_j t}(f(y) - f(x))}{t}$$

$$\mathcal{L}f(x) = \sum_i \lambda_i(x)(f(y_i) - f(x))$$

8.9 Stochastic PDEs

Let $t \geq 0$, $x \in \mathbb{R}$, consider $u(t,x)$. Some examples:

$$du_{\frac{i}{N}} = \frac{N^2}{2}\left(u_{\frac{i+1}{N}} - 2u_{\frac{i}{N}} + u_{\frac{i-1}{N}}\right) dt + c_N dB_{\frac{i}{N}}$$

$$du_{\frac{i}{N}} = \frac{N^2}{2}\left(u_{\frac{i+1}{N}} - 2u_{\frac{i}{N}} + u_{\frac{i-1}{N}}\right) dt + \frac{\epsilon}{N}\sqrt{u_{\frac{i}{N}}}dB_{\frac{i}{N}} \text{ (population growth with shifts among cities)}$$

$$du_{\frac{i}{N}} = \frac{N^2}{2}\left(u_{\frac{i+1}{N}} - 2u_{\frac{i}{N}} + u_{\frac{i-1}{N}}\right) dt + \frac{\epsilon}{N}u_{\frac{i}{N}}dB_{\frac{i}{N}} \text{ (investment in different assets)}$$

As $N \to \infty$, let ξ be a space function representing white noise:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + \xi \text{ (Stochastic heat equation, O-U process)}$$

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + \sqrt{u}\xi \text{ (Super Brownian, Dawson-Watanabe process)}$$

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + u\xi \text{ (Multiplicity stochastic heat equation)}$$

Gaussian white noise is a random distribution on ξ. Let f be a function of compact support. $\xi(f)$, ξ_f, $\langle f, \xi\rangle$, $\iint f\xi dtdx$ all means ξ acts on f. It is a Gaussian family with mean zero, and $E[\xi(f)\xi(\tilde{f})] = \left\langle f, \tilde{f}\right\rangle = \iint f\tilde{f}dtdx$. If $f = 1_A$, $\tilde{f} = 1_{\tilde{A}}$, $E[\xi(f)\xi(\tilde{f})] = 0$ if $A \cap \tilde{A} = \emptyset$. Essentially, it shows the independence of white noise.

$$E\left[\left(\frac{1}{N}\sum_i \int_0^1 f\left(s, \frac{i}{N}\right) c_N dB_i\right)^2\right] \to \iint f^2 dtdx$$

$$= \frac{c_N^2}{N^2}\sum_i \int_0^1 f\left(s, \frac{i}{N}\right) ds$$

This gives $c_N = \sqrt{N}$.

The stochastic heat equation $\partial_t u = \frac{1}{2}\partial_x^2 u + \xi$ is equivalent to

$$u(t,x) = \int p(t, x-y)u_0(y)dy + \int_0^t \int p(t-s, x-y)\xi(s,y)dyds,$$

where $p(t,x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$ is the heat kernel.

The discrete version has

$$u_{\frac{i}{N}}(t) = \frac{1}{N}\sum_j p_N\left(t, \frac{i}{N}-\frac{j}{N}\right) u\left(0,\frac{j}{N}\right) + \int_0^t \frac{1}{N}\sum_j p_N\left(t-s, \frac{i}{N}-\frac{j}{N}\right)\sqrt{N} dB_{\frac{j}{N}},$$

where $p_N\left(t, \frac{i}{N}\right) = \left(e^{\frac{1}{2}t\Delta_N}\right)\left(\frac{i}{N}\right)$, $\Delta_N u_{\frac{i}{N}} = N^2\left(u_{\frac{i+1}{N}} - 2u_{\frac{i}{N}} + u_{\frac{i-1}{N}}\right)$ is the Laplacian.

Proof. Consider the derivative in time

$$\begin{aligned}
& ds\left(\frac{1}{N}\sum_j p_N\left(t, \frac{i}{N}-\frac{j}{N}\right) u_{\frac{i}{N}}(s)\right) \\
=& \frac{1}{N}\sum_j \left(-\frac{1}{2}\Delta_N\right) p_N\left(t-s, \frac{i}{N}-\frac{j}{N}\right) u_{\frac{i}{N}}(s) ds + \frac{1}{N}\sum_j p_N\left(t, \frac{i}{N}-\frac{j}{N}\right)\left(-\frac{1}{2}\Delta_N u\right)_{\frac{i}{N}}(s) ds \\
& + \frac{1}{N}\sum_j p_N\left(t-s, \frac{i}{N}-\frac{j}{N}\right)\sqrt{N} dB_{\frac{j}{N}}
\end{aligned}$$

The first two parts cancels out by summation by parts. Integrate both sides in time dimension s, we get the equality in discrete version. □

When $d = 1$, If $\int_0^t \int p^2(t-s, x-y) dy ds < \infty$ for each $t > 0$, $x \in \mathbb{R}$, then $\int_0^t \int p(t-s, x-y)\xi(s,y) dy ds$ makes sense.

$$\begin{aligned}
& \int p^2(t-s, x-y) dy = \int p(t-s, y) p(t-s, y) dy = p(2(t-s), 0) = \frac{1}{\sqrt{2\pi 2(t-s)}} ds \\
\Rightarrow & \int_0^t \int p^2(t-s, x-y) dy ds = \int_0^t p(2(t-s), 0) ds < \infty
\end{aligned}$$

For $d \geq 2$, $p(t, 0) \sim \frac{1}{t^{d/2}}$, and it does not work directly. Need to check with test function.

Now, we want to know the behavior of the solution to $\partial_t u = \frac{1}{2}\partial_x^2 u + \xi$.

Claim: Brownian motion is invariant $\int \mathcal{L} f d\mu = 0$.

Proof. The continuous version cannot be checked. Instead we check the analogue for the discretization (Gaussian random walk). It has product measure

$$d\mu = \prod_j \frac{1}{\sqrt{2\pi \frac{i}{N}}} e^{-\frac{1}{2}\frac{\left(u_{\frac{j+1}{N}} - u_{\frac{j}{N}}\right)^2}{1/N}} du_{\frac{j}{N}} = e^{-\frac{1}{2}\sum_j \frac{\left(u_{\frac{j+1}{N}} - u_{\frac{j}{N}}\right)^2}{1/N}} \prod_j \frac{1}{\sqrt{2\pi \frac{i}{N}}} du_{\frac{j}{N}}$$

Then

$$D(f) = \int \frac{N}{2}\sum_i \left(\frac{\partial f}{\partial u_{\frac{i}{N}}}\right)^2 d\mu = -\int f \sum_i N \frac{\partial^2 f}{\partial {u_{\frac{i}{N}}}^2} - N^2\left(u_{\frac{i+1}{N}} - 2u_{\frac{i}{N}} + u_{\frac{i-1}{N}}\right)\frac{\partial f}{\partial u_{\frac{i}{N}}} d\mu$$

The integrand is the discrete version $\mathcal{L}_N f$. Then we have $D(f) = \int f(-\mathcal{L} f) d\mu$, so it is invariant. □

The invariance is modulo initial position. If $\partial_t u = v$, $\partial_t v = \frac{1}{2}\partial_x^2 v + \partial_x \xi$ has white noise as its invariant measure.

Check the increments in both dimensions:

$$u(t, x+\delta) - u(t, x) \sim \mathcal{N}(0, \delta)$$

Increment in time dimension:

$$\begin{aligned}
&E\left[\left(\int_0^{t+h}\int p(t+h-s, x-y)dyds + \int_0^t \int p^2(t-s, x-y)\xi(s,y)dyds\right)^2\right] \\
&= \int_0^{t+h}\int p^2(t+h-s, x-y)dyds + \int_0^t\int p^2(t-s, x-y)dyds \\
&- 2\int_0^t\int p(t+h-s, x-y)p(t-s, x-y)dyds \\
&= \int_0^{t+h}\frac{1}{\sqrt{2\pi 2(t+h-s)}}ds + \int_0^t \frac{1}{\sqrt{2\pi 2(t-s)}}ds - 2\int_0^t \frac{1}{\sqrt{2\pi(2(t-s)+h)}}ds \\
&= \frac{1}{4\sqrt{\pi}}\left(\sqrt{t+h} + \sqrt{t} - 2\left(\sqrt{t+\frac{h}{2}} - \sqrt{\frac{h}{2}}\right)\right)
\end{aligned}$$

Expand in first order of h, we get that the expectation is

$$= \frac{1}{4\sqrt{\pi}}\left(\sqrt{t} + \frac{h}{2\sqrt{t}} + \sqrt{t} - 2\left(\sqrt{t} + \frac{h}{4\sqrt{t}} - 2\sqrt{\frac{h}{2}}\right)\right) \sim c\sqrt{h}$$

Therefore,

$$u(t+h, x) - u(t, x) \sim \mathcal{N}(0, \sqrt{h})$$

Continuity: Check if u is continuous in t, x. By Theorem 8.10, we just need

$$E\left[(u(t+h, x+\delta) - u(t, x))^\gamma\right] \leq C\left(h^{2+\epsilon} + \delta^{2+\epsilon}\right)$$

From the previous estimation, we know that $u(t+h, x) - u(t, x) \sim h^{1/4}Z_1$, $u(t, x+\delta) - u(t, x) \sim h^{1/2}Z_2$, where Z_1 and Z_2 are Gaussian. For the inequality to hold, we just need $\gamma \geq 8+\epsilon$.

Scaling: Given $\xi(t, x)$, what does $\xi(At, Bx)$ do?

$$\begin{aligned}
&\iint f(t, x)\xi(At, Bx)dxdt = \iint \frac{f\left(\frac{t}{A}, \frac{x}{B}\right)}{AB}\xi(t, x)dxdt \\
&E\left[\left(\frac{f\left(\frac{t}{A}, \frac{x}{B}\right)}{AB}\xi(t, x)\right)^2\right] = \iint \frac{1}{A^2B^2}f^2\left(\frac{t}{A}, \frac{x}{B}\right)dxdt = \frac{1}{AB}\iint f^2(t, x)dxdt
\end{aligned}$$

Therefore, $\xi(At, Bx) = A^{-1/2}B^{-1/2}\tilde{\xi}(t, x)$ in distribution. This gives a transformed equation $u(t, x) = C\tilde{u}(At, Bx)$, $\partial_t\tilde{u} = \frac{1}{2}\partial_x^2\tilde{u} + \frac{B^{1/2}}{A^{1/2}C}\tilde{\xi}$. It is invariant under $u(t, x) \to \epsilon^{\frac{1}{2}}u(\epsilon^{-2}t, \epsilon^{-1}x)$.

Now consider the general version

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u^\alpha \xi, \quad \alpha = 0, \frac{1}{2}, 1$$

Discretize it, $du_{\frac{i}{N}} = \frac{1}{2}\Delta_N u_{\frac{i}{N}}dt + u_{\frac{i}{N}}^\alpha\sqrt{N}dB_{\frac{i}{N}}$. Apply Ito's formula:

$$df(u_{\frac{i}{N}}) = \left(f'(u_{\frac{i}{N}})\frac{1}{2}\Delta_N u_{\frac{i}{N}} + \frac{1}{2}f''(u_{\frac{i}{N}})N\right)dt + f'(u_{\frac{i}{N}})u_{\frac{i}{N}}^\alpha\sqrt{N}dB_{\frac{i}{N}}$$

The term $\frac{1}{2}f''(u_{\frac{i}{N}})N$ is an issue, because we cannot take N to infinity now.

Add a test function ϕ,

$$d\frac{1}{N}\sum_i \phi\left(\frac{i}{N}\right)u_{\frac{i}{N}} = \frac{1}{N}\sum_i \frac{1}{2}\Delta_N\phi\left(\frac{i}{N}\right)u_{\frac{i}{N}} + \frac{1}{N}\sum_i \phi\left(\frac{i}{N}\right)u^{\alpha}_{\frac{i}{N}}\sqrt{N}dB_{\frac{i}{N}}$$

Take limit as $N \to \infty$, we get:

$$d\langle \phi, u\rangle = \left\langle \frac{1}{2}\Delta\phi, u\right\rangle dt + \langle \phi, u^{\alpha}\xi\rangle$$

Note that $\langle \phi, u^{\alpha}\xi\rangle = \int \phi(x)u^{\alpha}(t,x)\xi(t,x)dx$ is a Martingale. The integral form is

$$\langle \phi, u(t)\rangle - \langle \phi, u(0)\rangle = \int_0^t \left\langle \frac{1}{2}\Delta\phi, u\right\rangle ds + M_t,$$

where $M_t = \int_0^t \int \phi(x)u^{\alpha}(s,x)\xi(s,x)dxds$. $\int M_t^2 - \int_0^t \int \phi^2 u^{2\alpha}dxds = N_t$ is the martingale for stochastic PDEs.

Super Brownian (Dawson-Watanabe) Process ($\alpha = \frac{1}{2}$):

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u^{\frac{1}{2}}\xi$$
$$du_{\frac{i}{N}} = \frac{1}{2}\Delta_N u_{\frac{i}{N}}dt + \sqrt{u_{\frac{i}{N}}}\sqrt{N}dB_{\frac{i}{N}}$$

This also has strong uniqueness in discrete version, but we don't know about the continuous version even in $d = 1$, when we assume positivity (for non-negative solutions, there is counter example)

Weak uniqueness (uniqueness in distribution) is guaranteed.

Let $X(t) = e^{-\frac{1}{N}\sum_i \phi(\frac{i}{N}, T-t)u_{\frac{i}{N}}(t)}$. Differentiate it,

$$dX = \frac{1}{N}\sum_i \left[-\frac{1}{2}\Delta_N\phi\left(\frac{i}{N}, T-t\right) + \frac{\partial\phi}{\partial t}\left(\frac{i}{N}, T-t\right) + \frac{1}{2}\phi^2\left(\frac{i}{N}, T-t\right)\right]u_{\frac{i}{N}}dt + dM_t$$

Choose ϕ s.t. $\partial_t\phi = \frac{1}{2}\Delta\phi - \frac{1}{2}\phi^2$, we get uniqueness,

$$E\left[e^{-\int \phi(0,x)u(T,x)dx}\right] = e^{-\int \phi(T,x)u_0(x)dx}$$

Example: $\phi(0,x) = c$, $\dot{\phi} = -\frac{1}{2}\phi^2$, $\langle \phi, u\rangle = \phi(t)\langle 1, u\rangle$

$$d\langle 1, u\rangle = \langle \Delta 1, u\rangle dt + \left\langle 1, u^{1/2}\xi\right\rangle = \langle 1, u\rangle^{1/2} dB$$

This is equivalently, $dy = \sqrt{y}dB$. $y(t) = \langle 1, u(t)\rangle$ is a non-negative martingale. It converges to zero, as $t \to \infty$. Then

$$E\left[e^{-a\langle 1, u(t)\rangle}\right] = e^{-\left(a^{-1}+\frac{1}{2}t\right)^{-1}\langle 1, u(0)\rangle}$$

As $a \to \infty$. LHS becomes expectation of an indicator function and is equivalent to $P(\langle 1, u(t)\rangle = 0)$. RHS becomes $e^{-\frac{2\langle 1, u(0)\rangle}{t}}$. The population dies.

Multiplicative Stochastic Heat Equation:

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u\xi$$
$$du_{\frac{i}{N}} = \frac{1}{2}\Delta_N u_{\frac{i}{N}} dt + u_{\frac{i}{N}}\sqrt{N} dB_{\frac{i}{N}}$$

The general form is $du = Audt + udB$. If we swap the index from $\frac{i}{N}$ to $\frac{i+1}{N}$, $Audt$ won't be affected.

$$\begin{aligned} udB &= \sum_i u_{\frac{i+1}{N}}\left(B_{\frac{i+1}{N}} - B_{\frac{i}{N}}\right) \\ &= \sum_i u_{\frac{i}{N}}\left(B_{\frac{i+1}{N}} - B_{\frac{i}{N}}\right) + \sum_i \left(u_{\frac{i+1}{N}} - u_{\frac{i}{N}}\right)\left(B_{\frac{i+1}{N}} - B_{\frac{i}{N}}\right) \end{aligned}$$

The second term is $\sim u_{\frac{i}{N}}\left(B_{\frac{i+1}{N}} - B_{\frac{i}{N}}\right)^2$. It just becomes $du = Audt + udB + cudt$. Therefore, we can consider the perturbation

$$\partial_t u_\epsilon = \frac{1}{2}\Delta u_\epsilon + u_\epsilon(\xi_\epsilon - c_\epsilon)$$

Note $E\left[\frac{1}{N}\sum_i u_{\frac{i}{N}}(t)\right] = E\left[\frac{1}{N}\sum_i u_{\frac{i}{N}}(0)\right]$, so $E\left[\int u_\epsilon(t,x)dx\right] = E\left[\int u_\epsilon(0,x)dx\right]$. This identifies c_ϵ.

Now $\xi_\epsilon - c_\epsilon$ becomes V in the Feymann-Kac's formula, we can solve it with

$$u_\epsilon(t,x) = E_x\left[e^{\int_0^t \xi_\epsilon(t-s,B_s)ds - c_\epsilon t} u_\epsilon(0,B_t)\right]$$

Might as well take $u_\epsilon(0,x) = \delta_y(x)$.

$$E_{xy}\left[e^{\int_0^t \xi_\epsilon(s,B_s)ds}\right] = E\left[e^{\langle \xi_\epsilon, \nu\rangle}\right],$$

where $\nu(du,dx) = \int_0^t \delta_0(s-u)\delta_0(B_s - x)ds$.

Using IBP, we can push ϵ to ν and get $E\left[e^{\langle \xi,\nu_\epsilon\rangle}\right]$, where $\nu_\epsilon = \int_0^t \delta_\epsilon(s-u)\delta_\epsilon(B_s - x)ds$. Then $E\left[E\left[e^{\langle \xi,\nu_\epsilon\rangle}\right]\right] = E\left[e^{c\epsilon^{-1}t}\right]$.

$$\begin{aligned} &E\left[u_\epsilon(t,x_1)\cdots u_\epsilon(t,x_n)\right] \\ &= E\left[E_{x_1 y}\left[e^{\int_0^t \xi_\epsilon(t-s,B_s^1)ds - c_\epsilon t}\right] p_t(x_1,y)\cdots E_{x_n y}\left[e^{\int_0^t \xi_\epsilon(t-s,B_s^n)ds - c_\epsilon t}\right] p_t(x_n,y)\right] \\ &= E\left[E_{x_1,\dots,x_n\to y,\dots,y}\left[e^{\sum_{i=1}^m \int_0^t \xi_\epsilon(t-s,B_s^i)ds - c_\epsilon t}\right]\right]\prod_{i=1}^m p_t(x_i,y) \\ &= E\left[E\left[e^{\langle \xi, \sum_i \nu_\epsilon^i\rangle - \frac{1}{2}\sum_i \langle \nu_\epsilon^i,\nu_\epsilon^i\rangle}\right]\right]\prod_{i=1}^m p_t(x_i,y) \\ &= E\left[e^{\sum_{i<j}\langle v_\epsilon^i, v_\epsilon^j\rangle}\right]\prod_{i=1}^m p_t(x_i,y) \end{aligned}$$

Now we compute

$$\begin{aligned} \lim_{\epsilon\to 0}\langle \nu_\epsilon^i,\nu_\epsilon^j\rangle &= \int\int\int_0^t ds_1 \int_0^t ds_2 \delta_{\epsilon_1}(s_1-u)\delta_{\epsilon_2}(s_2-u)\delta_{\epsilon_1}(B_{s_1}^i - x)\delta_{\epsilon_2}(B_{s_2}^j - x)dudx \\ &= \int\int\int_0^t ds_1 \int_0^t ds_2 \delta_0(s_1-u)\delta_0(s_2-u)\delta_0(B_{s_1}^i - x)\delta_0(B_{s_2}^j - x)dudx \\ &= \int dx \int_0^t ds \delta_0(B_s^i - x)\delta_0(B_s^j - x) \\ &= \int_0^t \delta_0(B_s^i - B_s^j)ds \end{aligned}$$

Then

$$v(t, x_1, ..., x_n) = E\left[\prod_{i=1}^{n} u(t, x_i)\right] = E_{x_1,...,x_n}\left[e^{\sum_{i<j}\int_0^t \delta_0(B_s^i - B_s^j)ds}\delta_y(x_1)\cdots\delta_y(x_n)\right]$$

This gives the following PDE system

$$\partial_t v = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}v + \left(\sum_{1\le i<j\neq n}\delta_0(x_i - x_j)\right)v$$
$$v(0, x_1, ..., x_n) = \delta_{y,...,y}(x_1, ..., x_n)$$

This is a PDE with a well defined operator $\mathcal{L} = \frac{1}{2}\Delta + \sum_{i<j}\delta_0(x_i - x_j)$. This is explicitly diagonalizable.