A GRADUATE COURSE

IN

STOCHASTIC PROCESSES

DDA6001 Notebook

Lecturer

Prof. Masakiyo Miyazawa

Institute for Data and Decision Analytics

Tex Written By

Mr. Jie Wang

The Chinese University of Hongkong, Shenzhen

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments

This book is taken notes from the DDA6001 in spring semester, 2020. These lecture notes were taken and compiled in LaTeX by Jie Wang, an undergraduate student in Spring 2020. Prof. Masakiyo Miyazawa has not edited this document. Students taking this course may use the notes as part of their reading and reference materials. This version of the lecture notes were revised and extended for many times, but may still contain many mistakes and typos, including English grammatical and spelling errors, in the notes. It would be greatly appreciated if those students, who will use the notes as their reading or reference material, tell any mistakes and typos to Jie Wang for improving this notebook.

Notations and Conventions

sup	least upper bound
inf	greatest lower bound
$\overline{E}$	closure of E
$f \circ g$	composition
$\limsup(\liminf)$	upper (lower) limit
$L(\mathcal{P},f),U(\mathcal{P},f)$	Riemann sums
$\mathcal{R}[a,b]$	classes of Riemann integrable functions on $[a,b]$
$\underline{\int_a^b} f(x)\,\mathrm{d}x, \overline{\int_a^b} f(x)\,\mathrm{d}x$	Riemann integrals
$\langle \boldsymbol{x},\boldsymbol{y} \rangle$	inner product
$\omega(f;E)$	oscillation of f over set E
$\Vert \cdot \Vert$	norm
∇f	gradient
$\frac{\partial f}{\partial x_i}, f_{x_i}, f_i, \partial_i f, D_i f$	partial derivatives
$D_{\boldsymbol{v}} f$	directional derivative at direction $\boldsymbol{v}$
$\frac{\partial(y_1,\ldots,y_m)}{\partial(x_1,\ldots,x_n)}$	Jacobian
$\mathbb{S}^n$	set of real symmetric $n \times n$ matrices
$\succ$ $(\succeq)$	positive (semi)-definite
$\mathcal{C}^m$	classes of m-th order continuously differentiable functions
$\mathcal{C}(E;\mathbb{R}^m)$	set of $\mathcal{C}^1$ mapping from E to $\mathbb{R}^m$
$(\mathcal{H},d)$	metric space

Chapter 1

Week1

1.1. Monday

1.1.1. Introduction to Stochastic Process

Course Information.

- Instructor: Prof. Masakiyo Miyazawa
- Course Venue: Chengdao 208
- Office Hour: 3-5pm, Monday, at Daoyuan 514

Motivation of Stochastic Process.

- capture randomness;
- characterize randomness influences in real systems;
- improve the design and operation of such systems.

■ **Example 1.1** [Coin Tossing Example] Consider a scenario where someone bets money by coin tossing. Assume that

- Head and tail occur likely and independently;
- Get or loss money by head or tail, respectively;
- There is no limit for money to bet.

The goal of this people is to double the money he initialy obtains. Therefore, the best policy is that he should stop the betting when the money becomes double of the initial. The question is that is this policy **realizable**?

- By probability theory, we could argue that such event will eventually happen with probability one.
- However, this policy is actually not realizable. This is because the expected time for the realization of this event is infinite! See the Figure. 1.1 for three simulation cases of betting.

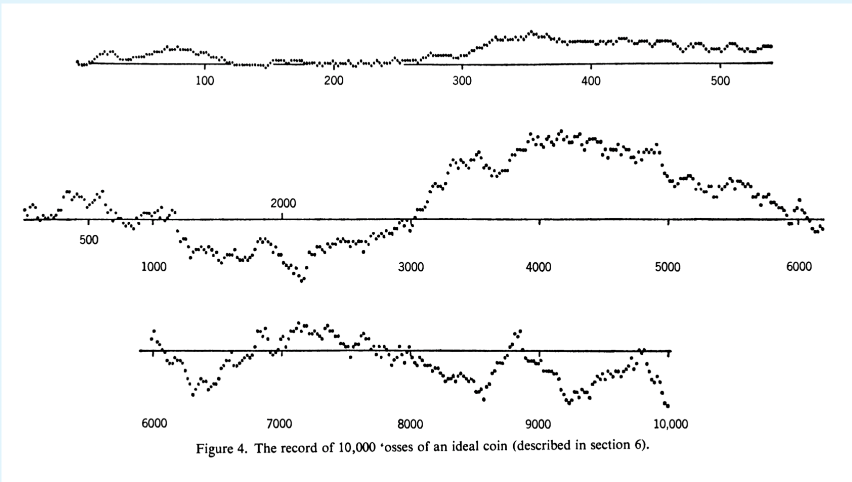

Figure 1.1: Simulation of coin tossing for 10000 times

Course Outline. Mathematical Model for random phenomena: probability space.

-

Two ways to constructor a probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

- Measure theory based: $\mathcal{F} = \mathcal{B}(\Omega)$.

  Standard for stochastic processes. (indexed by time)

- Functional Analysis based: Define the probability $\mathbb{P}$ by functional mapping:

$$\begin{aligned} \phi: &\quad \mathcal{C}_b(\Omega) \to \mathbb{R}_+ \\ \text{with} &\quad f \in \mathcal{C}_b(\Omega) : \Omega \to S \text{ (state space)} \end{aligned}$$

  Also we could define the norm for f and ϕ:

$$\|f\| = \sup_{x\in\Omega} f(x), \quad \|\phi\| = \sup_{\|f\|\le 1} \|f\|$$

  Suitable for Gaussian Processes (indexed by vector).

Discrete time stochastic processes (DTSP):

- random walk and limit theorems
- Conditional expectation, martingale, stopping time
- DTMC and applications;

Continuous time stochastic processes (CTSP):

- Possion process, Brownian motion, martingale
- Stochastic analysis; predictability, semi-martingale, stochastic integral, Ito's formula
- CTMC and applications.

W. Rudin Real and Complex Analysis.

1.1.2. Mathematical Background

Logics. Get familar with the operations $\wedge, \vee, \neg, \equiv, \Rightarrow$:

- $P \Rightarrow Q$ is T iff $\neg P \vee Q$ is T;
- $\neg(P \Rightarrow Q) \equiv P \wedge (\neg Q)$;
- $\neg(\forall x \in A, P(x)) \equiv \exists x \in A, \neg P(x)$

Limits. Get familar with $\sup, \inf, \limsup, \liminf, \lim$:

- $\sup\limits_{n\geq 1} a_n = \bar{a}$ iff
$$\begin{cases} \forall \varepsilon > 0, \exists n_0, \forall n \geq n_0, a_n < \bar{a} + \varepsilon \\ \forall \varepsilon > 0, \forall n \geq 1, \exists n_1 \geq n, \bar{a} - \varepsilon < a_{n_1} \end{cases}$$
and $\inf\limits_{n\geq 1} a_n = -\sup\limits_{n\geq 1}(-a_n)$.

- $\limsup\limits_{n\to\infty} a_n$ and $\liminf\limits_{n\to\infty} a_n$.

- $\lim\limits_{n\to\infty} a_n = a$ iff $\limsup\limits_{n\to\infty} a_n = a = \liminf\limits_{n\to\infty} a_n$. Or equivalently,
$$\forall \varepsilon > 0, \exists n_0 \geq 1, \forall n \geq n_0, \qquad |a_n - a| < \varepsilon.$$

Set Theory. Given mapping $f : A \to B$

- Onto mapping: $f(A) = B$;

- One-to-one mapping: $f(x) = f(y)$ implies $x = y$

- Berstein's theorem: if there are both one-to-one mappings from A to B and from B to A, then $\|A\| = \|B\|$.

1.1.3. Random Phenomena

The key feature for random phenomena is **multiple outcomes**. Define possible outcomes as follows.

Definition 1.1 [Sample Space] An outcome is called a sample. The set of all possible samples is called a **sample space**, denoted by Ω. ■

■ **Example 1.2** Toss a coin for n times. Define H for head and T for tail. The possible outcome can be denoted as $\omega = (\omega_i)_{i=1}^n$, with $\omega_i \in \{H, T\}$. The sample space Ω is the set of all these ω, with 2^n elements. ■

In order to characterize behaviours such as coin tossing, we need to define how to choose a sample from Ω. The statistical way is to assign a number to each sample which represents the corrresponding likelihood to choose it, called a **probability**. The formal definition of probability relies on the distribution function:

Definition 1.2 [Distribution] Assume Ω to be finite or countably infinite set. Define a function $p:\Omega\to[0,1]$ as follows:

- $p(\omega)\ge 0, \forall\omega\in\Omega$
- $\sum_{\omega\in\Omega}p(\omega)=1$

then $p(\cdot)$ is called a **distribution** over Ω. Moreover, $(\Omega, p(\cdot))$ is called a **discrete probability space**, and $p(\omega)$ denotes the probability that the event ω occurs.

R This definition cannot be directly applied into the uncountable sample space Ω. For instance, consider the unlimited coin tossing behaviour. The uncountable Ω defined as follows:

$$\Omega=\{0,1\}^{\infty}\supseteq[0,1].$$

Since any event ω happens with the same probability as the coin is fair, one can argue that $p(\omega)=0, \forall\omega\in\Omega$, which contradicts to the second condition in Definition 1.2.

Ideas to define a probability for uncountable sample space. Use limiting operation when faced uncountable.

Definition 1.3 [σ-field] A set $\mathcal{F}$ containing subsets of Ω is called a σ-field if:

1. $\Omega\in\mathcal{F}$;
2. $\mathcal{F}$ is closed under complement, i.e., $A\in\mathcal{F}$ implies $\Omega\setminus A\in\mathcal{F}$;
3. $\mathcal{F}$ is closed under infinite union operation, i.e., $A_i\in\mathcal{F}, i\ge 1$ implies $\cup_{i=1}^{\infty}A_i\in\mathcal{F}$.

Definition 1.4 [Probability Measure] A function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called a **probability measure** on $(\Omega, \mathcal{F})$ if

- $\mathbb{P}(\Omega) = 1$;
- $\mathbb{P}(A) \geq 0, \forall A \in \mathcal{F}$;
- $\mathbb{P}$ is σ-additive, i.e., when $A_i \in \mathcal{F}, i \geq 1$ and $A_i \cap A_j = \emptyset, \forall i \neq j$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

where $\mathbb{P}(A)$ is called the **probability of the event** A.

Definition 1.5 [Probability Space] A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is a set of samples, called the sample space;
- $\mathcal{F}$ is a σ-field, a collection of events;
- $\mathbb{P}$ is a probability measure, which assigns probability to events.

R Here we give an explicit construction of the **canonical** probability space, as the σ-field induced by Ω may not be unique.

- When Ω is discrete, we can construct **the** probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathcal{F} = 2^{\Omega}, \quad \mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \quad A \in \mathcal{F}.$$

- Otherwise, we define $\mathcal{F}$ with no "redundancy". When Ω is a **topological space**, with $\mathcal{O}$ being the **topology** of Ω, define

$$\mathcal{B}(\Omega) = \sigma(\mathcal{O}) \equiv \text{the minimal } \sigma\text{-field on } \Omega \text{ containing } \mathcal{O}.$$

Let $\mathcal{F} = \mathcal{B}(\Omega)$ and construct the associated probability measure $\mathbb{P}$. Then

$(\Omega, \mathcal{F}, \mathbb{P})$ is the **canonical** probability space.

Next we give some reviewings on the topological space:

Definition 1.6 A **topological space** $(X, \mathcal{T})$ consists of a (non-empty) set X, and a family of subsets of X ("open sets" $\mathcal{T}$) such that

1. $\emptyset, X \in \mathcal{T}$
2. $U, V \in \mathcal{T}$ implies $U \bigcap V \in \mathcal{T}$
3. If $U_\alpha \in \mathcal{T}$ for all $\alpha \in \mathcal{A}$, then $\bigcup_{\alpha \in \mathcal{A}} U_\alpha \in \mathcal{T}$.

When $A \in \mathcal{T}$, A is called the open subset of X. The $\mathcal{T}$ is called a **topology** on X. ■

■ **Example 1.3** Let (X, d) be any metric space, and define

$$\mathcal{T} = \{O \subseteq X : \forall x \in O, \exists \varepsilon > 0, \mathbb{B}_\varepsilon(x) \subseteq O\}$$

It's clear that $(X, \mathcal{T})$ is a topological space. ■

Definition 1.7 [Topological Space] A is open when

$$\forall x \in A, \exists B \in \mathcal{O}, x \in B \subseteq A$$

■

Define the distribution function for $\Omega = \mathbb{R}$:

Define the probability of an event $A \in \mathcal{B}(\mathbb{R})$.

1.1.4. Random variable and test functions

$$\mathbb{E} f(X), \qquad f \in \mathcal{C},$$

where f is called a **test function** of X.

1.2. Thursday

Reviewing. Probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω: collection of data;
- $\mathcal{F}$: collection of events;
- $\mathbb{P}$: probability measure $\mathcal{F} \to [0,1]$ satisfying:
  - $\mathbb{P}(\Omega) = 1$;
  - $A_i \in \mathcal{F}, A_i \cap A_j = \emptyset$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

The random variable X is a mapping from a sample into a real number $\Omega \ni \omega \to X(\omega) \in \mathbb{R}$. Therefore, a random variable should be **measurable**, i.e.,

$$\{\omega \in \Omega; X(\omega) \in B\} := \{X \in B\} := X^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Here is the formal definition of aa random variable:

Definition 1.8 [$\mathcal{F}$-Measurable / Random Variable]

1. A function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is called $\mathcal{F}$-**measurable** if

$$f^{-1}(\boldsymbol{B}) = \{w \mid f(w) \in \mathcal{B}\} \in \mathcal{F},$$

   for any $\boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)$.
2. A random variable X is a function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and is $\mathcal{F}$-measurable.

R By applying the definition of $\mathcal{B}(\mathbb{R}^n)$,

$$\{\omega \in \Omega; X(\omega) \in B\} \in \mathcal{F}, \; \forall B \in \mathcal{B}(\mathbb{R}) \iff \{X \le a\} \in \mathcal{F}, \forall a \in \mathbb{R}.$$

1.2.1. Expectation of random variable X

We define the representative value of a random variable by the following three steps:

- Consider first that X is simple, i.e., taking finitely many values. Assume that X only takes $x_1, x_2, \dots, x_n \in \mathbb{R}$, then define the expectation of X as:

$$\mathbb{E}[X] := \sum_{i=1}^{m} x_i \mathbb{P}(X = x_i),$$

  with $\mathbb{P}(X = x_i) = \mathbb{P}(\{X = x_i\}) = \mathbb{P}\{\omega \in \Omega; X(\omega) = x_i\}$.

- Then consider the case where X is non-negative, which is approximated by simple random variables. For each $n \geq 1$,

$$X_n(\omega) = \begin{cases} (i-1)2^{-n}, & \text{if } (i-1)2^{-n} \leq X(\omega) < i2^{-n}, i = 1,2,\dots,n2^n+1 \\ n, & \text{if } X(\omega) \geq n \end{cases}$$

  Taking $n \to \infty$, $X_n(\omega) \uparrow X(\omega)$. Define $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

- Finally consider the general X where the output is on the whole real line. Define

$$X_+(\omega) = \max(X(\omega), 0), \quad X_-(\omega) = \max(-X(\omega), 0).$$

  Then define the expectation of X by

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

R The expectation $\mathbb{E}[X]$ does not fully characterizes the information about X. Instead, the distribution function defined below has full information.

$$F(x) = \mathbb{P}(\{X \leq x\}), \quad x \in \mathbb{R}$$

We may view $F(x)$ as the set of the expectations $\mathbb{E}[f_x(X)]$ with the test function defined as

$$f_x(u) := \mathbf{1}\{u \leq x\}.$$

■ **Example 1.4** Consider the news vender problem: ■

The random variable X may take other types of values than $(-\infty,\infty)$, such as vector values.

$$X: \quad \omega \in \Omega \to X(\omega) \in S,$$

with S being a topological space. A speical case is the metric space. The advantage of topological space is that we can use it to define borel set conveniently:

$$\mathcal{B}(S) := \sigma(\{\text{topology of } S\})$$

X is S-valued random variable if $X : \Omega \to S$ and $\{X \in B\} \in \mathcal{F}, \forall B \in \mathcal{B}(S)$.

It may be difficult to evaluate the expectation of a random variable, and therefore we pick the test function

$$f : S \to \mathbb{R}$$

satisfying measurability:

$$f^{-1}(B') \in \mathcal{B}(S), \forall B' \in \mathcal{B}(\mathbb{R})$$

Define $f(X)(\omega) = f(X(\omega)), \forall \omega \in \Omega$:

$$\Omega \xrightarrow{X} S \xrightarrow{f} \mathbb{R}.$$

We use a family of test functions, denoted by $\mathcal{C}$. If we take a sufficiently large $\mathcal{C}$, we can determine the distribution of X.

- When $S = \mathbb{R}$,

$$\mathcal{C} = \{f_\theta : \quad f_\theta(x) = e^{\theta x}, \theta \in \mathbb{R}\}$$

and $\mathbb{E}[f_\theta(x)]$ is the moment generating function.

Chapter 2

Week 2

2.1. Monday

2.1.1. Stochastic Process

Suppose that we have some random events that change in time. We obtain some observations, say outcome, from random events. They take values in a set S, where S is a **topological space**.

Recap about Topological Space.

Definition 2.1 [Convergence in Metric Space] Let (X,d) be a metric space, then $\{x_n\} \to x$ means that

$$\forall \varepsilon > 0, \quad \exists N \text{ such that } d(x_n,x) < \varepsilon, \quad \forall n \geq N.$$

■

Definition 2.2 [Convergence in Topological Space] Let $(X,\mathcal{T})$ be a topological space, then $\{x_n\} \to x$ means that

$$\forall U \ni x \text{ that is open }, \exists N \text{ such that } x_n \in U, \forall n \geq N.$$

■

Definition 2.3 [Complete] A topological space $(X,\mathcal{T})$ is said to be complete if each Cauchy sequence in X converges in X. ■

Definition 2.4 [Separable] A topological space $(X, \mathcal{T})$ is said to be **Hausdorff** (second separable) if for all $x \neq y \in X$, there exists $U, V \in \mathcal{T}$ such that

$$x \in U, \quad y \in V, U \cap V = \emptyset.$$

Definition 2.5 [Polish Space] A complete separable metric space is called a polish space.

Given a topological space $(S, \mathcal{O})$, which is also called a state space, we define the Borel field of S as:

$$\mathcal{B}(S) = \sigma(\mathcal{O}) := \{\text{minimal } \sigma\text{-field including all elements in } \mathcal{O}\}$$

The state at time $n \in \mathbb{N}$ is denoted by S-valued random variable X_n: [1]

$$\begin{aligned} X_n : &\quad \Omega \to S \\ \text{with} &\quad X_n(\omega) \in S \end{aligned}$$

The collection of random variables X_n for $n = 0, 1, 2, \ldots$ is denoted by $\{X_n : \; n \geq 0\}$. This is called a stochastic process with state space S. Collection of all observations (information of events) up to time n is denoted by $\mathcal{F}_n$:

- $\mathcal{F}_n$ is a subset of $\mathcal{F}$;
- $\mathcal{F}_n$ is a σ-field on Ω;
- For $0 \leq m < n$, $\mathcal{F}_m \subseteq \mathcal{F}_n$, i.e., $F_n \uparrow$.

Define $\mathbb{F} := \{\mathcal{F}_n : \; n \geq 0\}$ is called a **filtration**.

$$X_n^{-1}(B) = \{\omega \in \Omega \;\; X_n(\omega) \in B\} = \{X \in B\}$$

[1] if time is continuous, then use $t \in \mathbb{R}_+$ to denote the time index

Definition 2.6 [measurable] If $X_n^{-1}(B) \in \mathcal{F}_n$ for any $B \in \mathcal{B}(S)$, then X_n is said to be **measurable** w.r.t. $\mathcal{F}_n$. For simplicity, X_n is $\mathcal{F}_n$-measurable. ■

Definition 2.7 [adapted] If X_n is $\mathcal{F}_n$-measurable for any n, then $X. = \{X_n : n \geq 0\}$ is said to be **adapted** to filtration $\mathbb{F}$. ■

■ **Example 2.1** Define

$$\mathcal{F}_n^X = \sigma(X_1, \ldots, X_n), \quad n \geq 0$$

Define $\mathbb{F}^X = \{F_m^X : n \geq 0\}$. Then $X.$ is adapted to $\mathbb{F}^X$. ■

We finally need the probability measure for $X.$, called a distribution of $X.$.

$$\begin{aligned} X. : \quad & \Omega \to \otimes_{i=1}^{\infty} S_i \\ \text{with} \quad & \omega \mapsto X.(\omega) = \{X_n(\omega) : n \geq 0\} \end{aligned}$$

Define $\mathcal{F}_\infty = \cup_{n=1}^{\infty} \mathcal{F}_n \subseteq \mathcal{F}$. The distribution of X is defined on $(\Omega, \mathcal{F}_\infty)$.

General Process.

- $S, \mathcal{B}(S)$
- $X. = \{X_n : n \geq 0\}$
- $\mathbb{F} = \{\mathcal{F}_n : n \geq 0\}$
- $\mathbb{P}$ on $(\Omega, \mathcal{F})$
- $(\Omega, \mathcal{F}, \mathbb{P})$ is called probability space; and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a stochastic basis.

Question: How to define $\mathbb{P}$ or distribution of $X.$ on $(\Omega, \mathcal{F}_\infty)$? How to define $X_n \to X$?

By Joint Distribution. The joint distribution of $X_0, X_1, \ldots, X_n$ is obtained as $\mathbb{P}_n$, where $\mathbb{P}_n$ is defined on $(S^{n+1}, \mathcal{B}(S^{n+1}))$

$$\mathbb{P}_{n+1}(B_n \times S) = \mathbb{P}_n(B_n), \quad B_n \in \mathcal{B}(S^{n+1}),$$

called comptibility. How can the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}_\infty := \mathcal{B}(S^\infty))$ be constructed?

Theorem 2.1 If S is a polish space, then

■ **Example 2.2** Consider a Markov chain, with S countable.

$$\mathcal{B}(S) = \{\text{all subsets of } S\}$$

- For $i, j \in S, p_{i,j} \geq 0$ such that $\sum_{j \in S} p_{i,j} = 1$.
- Define the distribution $\{a_i\}_{i \in S}$ such that $a_i \geq 0, \sum_{i \in S} a_i = 1$.
- For any $n \geq 0, i_0, i_1, \ldots, i_n \in S$, define

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_m = i_m) = a_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}.$$

To check $\mathbb{P}$ a probability distribution,

$$\sum_{i_0,\ldots,i_n \in S} \mathbb{P}(X_0 = i_0, \ldots, X_m = i_m) = 1.$$

We can also check the comptibility condition.

It's easy to check that

$$\mathbb{P}\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i\} = \mathbb{P}\{X_{n+1} = j \mid X_n = i\} = p_{i,j}$$

This is called the Markov property, and $\{X_n : n \geq 0\}$ is called a discrete time Markov chain. ■

2.1.2. Random Walk

Example of a discrete time stochastic process. $S = \mathbb{R}$.

Definition 2.8

- $A, B \in \mathcal{F}$ are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

- S-valued random variables X, Y are independent if $\{X \in B_1\}$ and $\{Y \in B_2\}$ are independent for any $B_1, B_2 \in \mathcal{B}(S)$.

- $A_1, A_2, \ldots, A_n \in \mathcal{F}$ are independent if for any $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, \ldots, n\}$

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

- S-valued random variables $X_1, \ldots, X_n$ are independent if $\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ are independent for any $B_i \in \mathcal{B}(S), i = 1, \ldots, n$.

- $\{X_n : \ n \geq 1\}$ is independent if for each $n \geq 1$, $X_1, \ldots, X_n$ are independent.

Suppose that $S = \mathbb{R}$ and $\mathbb{R}$-valued random variables $U_0, U_1, \ldots,$ are independent, and the distribution of U_n is independent of n, i.e., $U_1, U_2, \ldots$ have the same distribution. Define

$$X_n := U_1 + U_2 + \cdots + U_n, \qquad n \geq 0.$$

Then $\{X_n : \ n \geq 0\}$ is called a random walk on $\mathbb{R}$.

Special case: $U_0 = 0$ and

$$U_n = \begin{cases} 1, & \text{w.p. } 0.5 \\ -1, & \text{w.p. } 0.5 \end{cases}$$

U_n represents coin tossing.

$$T_n = \inf\{n \geq 1 : \ X_n(\omega) = 0\}.$$

2.2. Thursday

We have defined what is a stochastic process. Now we study several special stochastic processes. The first example is a random walk, a discrete time stochastic process:

2.2.1. Random Walk

Firstly, we need to define the independence among random variables:

Definition 2.9 [Independence]

- Events $A, B \in \mathcal{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$;
- Random variables $X, Y : \Omega \to S$ are independent if events $\{X \in B_1\}$ and $\{X \in B_2\}$ are independent for any $B_1, B_2 \in \mathcal{B}(S)$;
- A sequence of random variables $\{X_n\}_{n \geq 0}$ is independent if for each finite n, $\{X_i \in B_i\}$ for $i = 1, 2, \dots,$ are independent for any $B_i \in \mathcal{B}(S)$.

R If random variables X_1, X_2 are independent and $\mathbb{E}[f_k(X_k)]$ are finite for $k = 1, 2$, then

$$\mathbb{E}[f_1(X_1)f_2(X_2)] = \mathbb{E}[f_1(X_1)]\mathbb{E}[f_2(X_2)].$$

Now we start defining a random walk based on coin tossing. Let $\{U_n\}_{n \geq 1}$ be the unlimited coin tossing, i.e., U_n is i.i.d. with $\mathbb{P}(U_n = 1) = \mathbb{P}(U_n = -1) = 1/2$. Then define

$$X_0 = 0, \ X_n = \sum_{i=1}^{n} U_i.$$

Here the stochastic process $\{X_n : \ n \geq 0\}$ is called a simple symmetric random walk with the initial state 0. Define the random variable

$$T(\omega) = \inf\{n \geq 1 : \quad X_n = 0\}, \quad \omega \in \Omega.$$

The mapping $T : \Omega \to \mathbb{Z}_+$ is called the **first return time** at the origin. Next, we study the distribution of T and compute $\mathbb{E}[T]$. The calculation of the distribution is by means of the generating function $g(z) \equiv \sum_{n=0}^{\infty} z^n a_n$ for a non-negative bounded sequence $\{a_n\}_{n\geq 0}$ and $|z| < 1$.

Determine the Distribution of T. Define the generating functions:

$$g(z) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = 0) z^n$$
$$h(z) = \sum_{n=0}^{\infty} \mathbb{P}(T = n) z^n$$

Here

$$\mathbb{P}(X_n = 0) = \begin{cases} 1, & \text{if } n = 0 \\ \sum_{\ell=1}^{n} \mathbb{P}(T = \ell)\mathbb{P}(X_{n-\ell} = 0), & \text{if } n \geq 1 \end{cases}$$

Substituting $\mathbb{P}(X_n = 0)$ into $g(z)$, we imply $g(z) = 1 + h(z)g(z)$.

Moreover, the direct calculation of $\mathbb{P}(X_n = 0)$ gives

$$\mathbb{P}(X_{2n} = 0) = \binom{2n}{n} 2^{-2n}$$
$$\mathbb{P}(X_{2n+1} = 0) = 0$$

Substituting $\mathbb{P}(X_n = 0)$ into $g(z)$, we imply

$$g(z) = \sum_{n=0}^{\infty} \binom{2n}{n} 2^{-2n} z^{2n} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{2^{2n}} \frac{(2n)!}{n!} z^{2n} = (1 - z^2)^{-1/2}.$$

As a result,

$$h(z) = 1 - \frac{1}{g(z)} = 1 - (1 - z^2)^{1/2}.$$

- The distribution of T can be obtained by taking derivative of h:

$$h^{(k)}(z)\Big|_{z=0} = k!\mathbb{P}(T = k)$$

- An alternative approach for determining the distribution of T is the following.

Consider the generating function induced by $\mathbb{P}(T > 2n)$:

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbb{P}(T > 2n) z^{2n} &= \sum_{n=0}^{\infty} \sum_{\ell=n+1}^{\infty} \mathbb{P}(T = 2\ell) z^{2n} \\ &= \sum_{\ell=1}^{\infty} \sum_{n=0}^{\ell-1} \mathbb{P}(T = 2\ell) z^{2n} \\ &= \sum_{\ell=1}^{\infty} \mathbb{P}(T = 2\ell) \frac{1 - z^{2\ell}}{1 - z^2}\end{aligned}$$

Also observe that

$$h(1) - h(z) = \sum_{n=0}^{\infty} \mathbb{P}(T = n) \cdot (1 - z^n) = \sum_{n=1}^{\infty} \mathbb{P}(T = 2n) \cdot (1 - z^{2n})$$

Therefore,

$$\sum_{n=0}^{\infty} \mathbb{P}(T > 2n) z^{2n} = \sum_{\ell=1}^{\infty} \mathbb{P}(T = 2\ell) \frac{1 - z^{2\ell}}{1 - z^2} = \frac{1 - h(z)}{1 - z^2} = (1 - z^2)^{-1/2} = g(z).$$

As a result,

$$\mathbb{P}(T > 2n) = \mathbb{P}(X_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n}.$$

Apply the Stirling's formula that

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n} \implies \mathbb{P}(T > 2n) \sim \frac{1}{\sqrt{n\pi}}.$$

The expectation $\mathbb{E}[T]$ can be obtained as the following:

$$\mathbb{E}[T] = \sum_{n=0}^{\infty} n \mathbb{P}(T = n) = \sum_{n=1}^{\infty} n[\mathbb{P}(T > n) - \mathbb{P}(T > n - 1)] = -\sum_{n=0}^{\infty} \mathbb{P}(T > n) = \infty.$$

The key observation of the deviation above is

$$\mathbb{P}(T > 2n) = \mathbb{P}(X_{2n} = 0),$$

which can also be explained by the reflection principle.

2.2.2. Basic Tips on Probability Theory

We are going to study two important limit theorems:

- Law of large numbers (LLN);
- Central Limit Theorem (CLT);

Firstly we need to review fundamentals of probability theory:

- Limiting operations on events;
- Continuity of probabilities;
- Relation between distribution and random variable.

Limiting operations on events. The difficulty for defining limit on events is as follows. Given two sets A, B, they may be ordered by inclusion, but may not be always. A rescue is by the union and intersection. Define

$$\underline{A}_n = \bigcap_{\ell=n}^{\infty} A_\ell, \qquad \overline{A}_n = \bigcup_{\ell=n}^{\infty} A_\ell.$$

and

$$\liminf_{m\to\infty} A_m := \bigcup_{n=1}^{\infty}\bigcap_{\ell=n}^{\infty} A_\ell, \qquad \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{\ell=n}^{\infty} A_\ell$$

By definition, $\underline{A}_m \subseteq A_m \subseteq \overline{A}_m$. If $\liminf A_m = \limsup A_m = A$, then $\{A_n;\ n \ge 1\}$ is said to converge to A, denoted as

$$\lim_{m\to\infty} A_m = A.$$

Theorem 2.2 — Continuity of Probability. If A_n converges to A as $n \to \infty$, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$$

In other words,

$$\lim_{n\to\infty} \int \mathbf{1}\{A_n\}\,\mathrm{d}m = \int \lim_{n\to\infty} \mathbf{1}\{A_n\}\,\mathrm{d}m$$

Proof. 1. Firstly, assume that $A_n \uparrow A$ and show $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

2. Secondly, when $A_n \downarrow A$, $A_n^c \uparrow A^c$. Thus the same result can be shown.
3. The general A_n can also be shown by considering

$$\mathbb{P}\left(\cap_{\ell=n}^{\infty} A_\ell\right) \le \mathbb{P}(A_n) \le \mathbb{P}\left(\cup_{\ell=n}^{\infty} A_\ell\right)$$

■

Definition 2.10 [Distribution Function] A **distribution function** $F : \mathbb{R} \to [0,1]$ is defined as follows:

1. Nondecreasing: $F(x) \le F(y)$ for all $x < y$;
2. Right-continuous: $F(x) = F(x+) \equiv \lim_{\varepsilon \to 0} F(x+\varepsilon)$;
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

Proposition 2.1 For distribution function F, there uniquely exists a probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mathbb{P}((a,b]) = F(b) - F(a), \qquad \forall a,b \text{ with } a \le b.$$

The random variable X associated with distribution function F does not necessarily exist on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The answer is yes when a random variable U with uniform distribution on $[0,1]$ exists on it. The explicit construction of X is the following. Define the inverse function $h : [0,1] \to \mathbb{R}$ such that

$$h(y) = \inf\{x \in \mathbb{R} : \ y \le F(x)\}, \qquad y \in [0,1].$$

Then define $X = h(U)$, which follows that

$$\mathbb{P}(X \le x) = \mathbb{P}(h(U) \le x) = \mathbb{P}(U \le F(x)) = F(x).$$

Chapter 3

Week 3

3.1. Monday

3.1.1. Weak law of large numbers

Firstly, review two modes of convergence for random variables:

Definition 3.1 [Convergence of random variables] For a sequence of $X_n, n \geq 1$ and X,

1. we say $X_n \to X$ a.s. if $\mathbb{P}(\Omega_0) = 1$ for

$$\Omega_0 = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}.$$

2. we say $X_n \to X$ in probability if

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| > \varepsilon\Big) = 0.$$

Almost sure convergence is defined by the sample-path based condition, while the convergence in probability is defined based on probability. To answer the relation between these two convergence modes, we reformulate the condition for almost sure convergence into a probability-based one.

Proposition 3.1 The sequence of random variables $X_n \to X$ a.s. if and only if

$$\forall k \geq 1, \quad \lim_{\ell\to\infty} \mathbb{P}\Big\{ \cup_{n=\ell}^{\infty} \{|X_n - X| > 1/k\}\Big\} = 0.$$

Proof. Starting from the a.s. convergence definition, we find an equivanent definition for $\omega \in \Omega_0$:

$$\forall \varepsilon > 0,\ \exists n_0 \geq 1 \text{ s.t. } \forall n \geq n_0,\ |X_n(\omega) - X(\omega)| < \varepsilon.$$

Or equivalently,

$$\forall k \geq 1,\ \exists n_0 \geq 1 \text{ s.t. } \forall n \geq n_0,\ |X_n(\omega) - X(\omega)| \leq \frac{1}{k}.$$

Therefore, we have an equivalent definition for Ω_0:

$$\Omega_0 = \bigcap_{k=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \left\{\omega :\ |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}$$

Taking the complement of Ω_0, we say $X_n \to X$ a.s. if and only if

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} \bigcap_{n_0=1}^{\infty} \bigcup_{n=n_0}^{\infty} \left\{\omega :\ |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0.$$

Since $\mathbb{P}(\cup_{k=1}^{\infty} A_k) = 0$ is equivalent to $\mathbb{P}(A_k) = 0, \forall k \geq 1$, we imply

$$\mathbb{P}\left(\lim_{n_0 \to \infty} \bigcup_{n=n_0}^{\infty} \left\{\omega :\ |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0, \forall k \geq 1.$$

By the continuity of probability,

$$\forall k \geq 1, \lim_{n_0 \to \infty} \mathbb{P}\left(\bigcup_{n=n_0}^{\infty} \left\{\omega :\ |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0.$$

The proof is complete. ∎

Based on this proposition, we imply the relation between convergence a.s. and convergence in probability:

Corollary 3.1 If $X_n \to X$ a.s., then $X_n \to X$ in probability.

Proof. This is because for any $\varepsilon > 0$,

$$0 \leq \mathbb{P}(|X_n - X| < \varepsilon) \leq \mathbb{P}(\cup_{n=\ell}^{\infty} \{|X_n - X| < \varepsilon\}).$$

■

These facts above shows that we need to show the law of large numbers (LLN) based on different modes of convergence. We will first show the weak LLN, assuming the finiteness of moments of X_n.

Definition 3.2 [Moment of a random variable] Consider a random variable X and $a \geq 0, c \in \mathbb{R}$.

- $\mathbb{E}(X^a)$ denotes the a-th moment of X;
- $\mathbb{E}(|X|^a)$ denotes the a-th absolute moment of X;
- $\mathbb{E}(|X-c|^a)$ denotes the a-th absolute moment of X, centered at c.

■

Proposition 3.2 Consider a random variable X,

- For a, b satisfying $0 \leq a < b$,

$$\mathbb{E}(|X|^b) < \infty \implies \mathbb{E}(|X|^a) < \infty.$$

- For $a > 0$ and $c \in \mathbb{R}$,

$$\mathbb{E}(|X|^a) < \infty \Longleftrightarrow \mathbb{E}(|X-c|^a) < \infty.$$

Proof. The proof of the first and the second part is based on the following inequalities:

$$|x|^a \leq 1 + |x|^b,$$
$$|x-c| \leq |x| + |c| \leq 2\max\{|x|, |c|\}.$$

■

An important tool for showing the weak LLN is Markov inequality:

Proposition 3.3 — **Markov Inequality.** For a random variable X and $\mathcal{B}(\mathbb{R})$-measurable function $h : \mathbb{R} \to \mathbb{R}_+$,

$$\mathbb{P}\Big(h(X) > t\Big) \le \frac{1}{t}\mathbb{E}[h(X)], \qquad t > 0.$$

Proof. Obseve the inequality holds: $t1(h(X) > t) \le h(X)$. Taking the expectation both sides leads to the desired result. ■

> **Corollary 3.2** [Chebyshev's inequality] For a random variable X, we have the following concentration-type inequality:
>
> $$\mathbb{P}\Big(|X - \mathbb{E}[X]| > \varepsilon\Big) \le \frac{1}{\varepsilon^2}\mathbb{E}[|X - \mathbb{E}[X]|^2].$$

Proof. Apply Markov inequality with $h(x) = |x - \mathbb{E}[X]|^2$ and $t = \varepsilon^2$. ■

Let $\{X_n\}_{n\ge 0}$ be a sequence of random variables, and define the finite sum and sample mean by

$$S_n = X_1 + X_2 + \cdots + X_n, \qquad \overline{X}_n = S_n/n, \; n \ge 1.$$

- If $\overline{X}_n \to m$ in probability and $\mathbb{E}[\overline{X}_n] \to m$, the sequence $\{X_n\}_{n\ge 0}$ is said to obey the **weak law of large numbers**.
- If there are constants a_n, b_n such that $\frac{S_n - b_n}{a_n} \to 0$ in probability and $\mathbb{E}[\overline{X}_n] \to m$, the sequence $\{X_n\}_{n\ge 0}$ is said to obey the **extended weak law of large numbers**.
- If $\overline{X}_n \to m$ a.s. and $\mathbb{E}[\overline{X}_n] \to m$, the sequence $\{X_n\}_{n\ge 0}$ is said to obey the **strong law of large numbers**.

Here we give a general sufficient condition for constructing the weak LLN:

Proposition 3.4 If

$$\lim_{n\to\infty} \mathbb{E}[(\overline{X}_n - m)^2] = 0, \tag{3.1}$$

then $\overline{X}_n \to m$ in probability.

Proof. Apply the Chebyshev's inequality while replacing X with $\overline{X}_n$:

$$\mathbb{P}\left(|\overline{X}_n - m| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}[(\overline{X}_n - m)^2] \to 0.$$

The proof is complete. ∎

Note that once (3.1) is satisfied, the weak LLN holds, no matter whether $\{X_n\}$ is i.i.d. or not. Now we provide the sufficient condition of $\{X_n\}$ for satisfying (3.1).

Theorem 3.1 — A baby version of weak LLN. Assume that

1. If $\{X_n\}$ is i.i.d.
2. $\mathbb{E}[X_n^2] < \infty$, or equivalently, $\sigma_X^2 < \infty$,

then condition (3.1) for the weak LLN is satisfied, with $m = \mathbb{E}[X_n]$.

Proof. Observe the following equality holds:

$$\begin{aligned}
\mathbb{E}[(\overline{X}_n - m)^2] &= \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{\ell=1}^{n}(X_\ell - m)\right)^2\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\sum_{\ell=1}^{n}(X_\ell - m)^2 + \sum_{\ell\neq k:\ell,k\in[1:n]}(X_\ell - m)(X_k - m)\right] \\
&= \frac{1}{n^2}\sum_{\ell=1}^{n}\mathbb{E}\left[(X_\ell - m)^2\right] \triangleq \frac{1}{n^2}\sum_{\ell=1}^{n}\sigma_X^2 \\
&= \frac{1}{n}\sigma_X^2 \to 0.
\end{aligned}$$

∎

Let's see an interesting application of weak LLN in polynomial fitting:

Proposition 3.5 For a continuous function $f : [0,1] \to \mathbb{R}$, let

$$f_n(x) \triangleq \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(k/n),$$

then $\sup_{x\in[0,1]}|f_n(x) - f(x)| \to 0$ as $n \to \infty$.

Proof. The idea is to construct $\{X_n\}$ such that $f_n(x)$ relates to the expectation of some function of $\{X_n\}$. Construct $S_n = X_1 + X_2 + \cdots + X_n$ with i.i.d. $\{X_n\}$ and $\mathbb{P}(X_n = 1) = p, \mathbb{P}(X_n = 0) = 1 - p$. By weak LLN, $S_n/n \to p$ in probability.

Moreover, observe that

$$\mathbb{E}[f(S_n/n)] = f_n(p)$$

Apply the uniform continuity of f, i.e., for $\forall \varepsilon > 0$, there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Also, f is bounded, i.e., $\sup_{x\in[0,1]} |f(x)| < M$ for some $M \geq 0$. Now we begin to upper bound $|f_n(p) - f(p)|$:

$$\begin{aligned}
|f_n(p) - f(p)| &= |\mathbb{E}[f(S_n/n) - f(p)]| \\
&\leq \mathbb{E}[|f(S_n/n) - f(p)|] \\
&= \mathbb{E}[|f(S_n/n) - f(p)|1(|S_n/n - p| \leq \delta)] + \mathbb{E}[|f(S_n/n) - f(p)|1(|S_n/n - p| > \delta)]] \\
&\leq \varepsilon + 2M\mathbb{P}(|S_n/n - p| > \delta)
\end{aligned}$$

Since $S_n/n \to p$ in prob., $\lim_{n\to\infty} \mathbb{P}(|S_n/n - p| > \delta) = 0$ holds uniformly for p. Therefore, taking $n \to \infty$ both sides, $|f_n(p) - f(p)| \leq (2M + 1)\varepsilon$.

■

Now we study a stronger version of weak LLN:

Theorem 3.2 — Stronger version of weak LLN. Assume that

1. If $\{X_n\}$ is i.i.d.
2. $\mathbb{E}[|X_n|] < \infty$,

then $\overline{X}_n \to m$ in probability, with $m = \mathbb{E}[X_n]$.

This theorem can be proved by the truncation of X_n:

1. First define $X_\ell^{(n)} = X_\ell 1(|X_\ell \leq n|)$ and show $\overline{X}_n \to \overline{m}_n^{(n)}$ with $\overline{m}_n^{(n)} = \mathbb{E}(\overline{X}_n^{(n)})$;
2. Then show that $\overline{m}_n^{(n)} \to m$.

Proof for the first part. Define $X_\ell^{(n)} = X_\ell 1(|X_\ell \leq n|)$, $S_n^{(n)} = \sum_{\ell=1}^n X_\ell^{(n)}$, $\overline{X}_n^{(n)} = S_n^{(n)}/n$, and

$\overline{m}_n^{(n)} = \mathbb{E}[\overline{X}_n^{(n)}]$. Firstly show that $n \cdot \mathbb{P}(X > n) \to 0$:

$$
\begin{aligned}
0 \le n \cdot \mathbb{P}(X > n) &\le n \cdot \mathbb{P}(|X| > n) \\
&= \mathbb{E}[n \cdot 1(|X| > n)] \le \mathbb{E}[|X| \cdot 1(|X| > n)] \\
&= \mathbb{E}[|X|] - \mathbb{E}[|X| \cdot 1(|X| \le n)] \to 0.
\end{aligned}
$$

For fixed $\varepsilon > 0$, we need to upper bound $\mathbb{P}(|\overline{X}_n - \overline{m}_n^{(n)}| > \varepsilon)$ as the following:

$$
\begin{aligned}
\mathbb{P}(|\overline{X}_n - \overline{m}_n^{(n)}| > \varepsilon) &= \mathbb{P}(|\overline{X}_n - \overline{m}_n^{(n)}| > \varepsilon, S_n^{(n)} \neq S_n) + \mathbb{P}(|\overline{X}_n - \overline{m}_n^{(n)}| > \varepsilon, S_n^{(n)} = S_n) \\
&\le \mathbb{P}(S_n^{(n)} \neq S_n) + \mathbb{P}(|\overline{X}_n^{(n)} - \overline{m}_n^{(n)}| > \varepsilon)
\end{aligned}
$$

where

$$
\mathbb{P}(S_n^{(n)} \neq S_n) \le \sum_{\ell=1}^n \mathbb{P}(X_\ell^{(n)} \neq X_\ell) = n \cdot \mathbb{P}(|X_1| > n) \to 0,
$$

and

$$
\mathbb{P}(|\overline{X}_n^{(n)} - \overline{m}_n^{(n)}| > \varepsilon) \le \frac{1}{\varepsilon^2} \mathbb{E}[|\overline{X}_n^{(n)} - \overline{m}_n^{(n)}|^2] \to 0.
$$

■

Proof for the second part. Note that

$$
\overline{m}_n^{(n)} - m = \frac{1}{n} \sum_{\ell=1}^n m_\ell^{(n)} - m = m_1^{(n)} - m = -\mathbb{E}[X_1 1(|X_1| > n)] \to 0.
$$

■

3.1.2. Strong Law of Large Numbers

The strong law of large numbers says that the sample average $\overline{X}_n$ of $X_1, X_2, \ldots, X_n$ almost surely converges to their mean $m = \mathbb{E}[X_1]$. The a.s. convergence result allows us to estimate m by the sample average. Since the strong LLN requires $\mathbb{P}(|\overline{X}_n - m| > \varepsilon)$ converges to zero **faster** than the weak LLN, we may need a higher moment of X_n to be finite.

In order to show the strong LLN, first prepare for two lemmas:

Proposition 3.6 For $A_\ell \in \mathcal{F}, \ell = 1,2,\ldots$, if $\sum_{\ell=1}^{\infty} \mathbb{P}(A_\ell) < \infty$, then

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{\ell=n}^{\infty} A_\ell\right) = 0.$$

Proof. This is because

$$\mathbb{P}\left(\bigcup_{\ell=n}^{\infty} A_\ell\right) \leq \sum_{\ell=n}^{\infty} \mathbb{P}(A_\ell) \to 0.$$

■

Proposition 3.7 The sequence of random variables $X_n \to X$ a.s. if

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty.$$

The baby version of strong LLN is stated in the following:

Theorem 3.3 — Baby version of strong LLN. Assume that

- $\{X_n\}$ is i.i.d.;
- $\mathbb{E}[X_1^4] < \infty$;

then $\overline{X}_n \to m$ a.s. with $m = \mathbb{E}[X_1]$.

By proposition 3.7, it suffices to show $\sum_{n=1}^{\infty} \mathbb{P}(|\overline{X}_n - m| > \varepsilon) < \infty, \forall \varepsilon > 0$. Actually, we need to show that for $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}_n - m| > \varepsilon) < \text{constant} \cdot 1/n^2.$$

By Markov inequality,

$$\mathbb{P}(|\overline{X}_n - m| > \varepsilon) \leq \frac{\mathbb{E}[(\overline{X}_n - m)^4]}{\varepsilon^4}$$

Therefore, we need to show the following result:

$$\mathbb{E}[(\overline{X}_n - m)^4] = \frac{1}{n^4}\left(n \cdot \beta + 3n(n-1)\sigma^4\right), \qquad \beta = \mathbb{E}[(X_1 - m)^4] < \infty.$$

The 4-th moment condition can be weakened as follows:

Theorem 3.4 — Strong LLN. The strong LLN holds if $\{X_n\}_{n\geq 1}$ is i.i.d. and $\mathbb{E}[|X_1|] < \infty$.

R The i.i.d. assumption in the strong LLN can also be weaked into:

- $\{X_n\}$ is **stationary**, and **ergodic**.

3.2. Thursday

3.2.1. Central Limit Theorem

The strong LLN indicates that $\overline{X}_n \to m$ a.s. It can be applied to real data $\{X_n(\omega), n \geq 1\}$ to estimate m. We need to see how big n is sufficient. This may be answered by computing for a given error level $\varepsilon > 0$, $\mathbb{P}(|\overline{X}_n - m| > \varepsilon)$. Therefore, we need to know the general distribution of $\overline{X}_n$, which is independent of the distribution of X_n.

Theorem 3.5 — Central Limit Theorem. Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with finite $m = \mathbb{E}[X_1]$ and finite $\sigma = \sqrt{V(X_1)}$, then the sample average $\overline{X}_n$ satisfies

$$\mathbb{E}[\overline{X}_n] = m, \qquad V(\overline{X}_n) = \frac{\sigma^2}{n}$$

Define the normalized random variable of $\overline{X}_n$ as

$$Z_n = \frac{\overline{X}_n - m}{\sqrt{V(\overline{X}_n)}} = \frac{S_n - mn}{\sqrt{n}\sigma}.$$

Then

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \Phi(x), \qquad x \in \mathbb{R},$$

where Φ is the standard normal distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du.$$

As a result, we have the following approximation:

$$\mathbb{P}(|\overline{X}_n - m| > \varepsilon) = \mathbb{P}\left(|Z_n| > \frac{\sqrt{n}}{\sigma}\varepsilon\right) = 2\left(1 - \Phi(\frac{\sqrt{n}}{\sigma}\varepsilon)\right).$$

We have the following questions about central limit theorem:

- Is the Φ a distribution function?
- How to define convergence for distribution functions?

- How to show the central limit theorem?

Proposition 3.8 The function Φ is a distribution function, where

$$\Phi(x) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u.$$

Proof. It suffices to show that $\lim_{x\to\infty} \Phi(x) = 1$. In other words, it suffices to show

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2 - \frac{1}{2}v^2\right) \mathrm{d}u\,\mathrm{d}v = 2\pi.$$

By the change of variable $\mathrm{d}u\,\mathrm{d}v = r\,\mathrm{d}r\,\mathrm{d}\theta$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2 - \frac{1}{2}v^2\right) \mathrm{d}u\,\mathrm{d}v = \int_{0}^{2\pi}\int_{0}^{\infty} \exp\left(-\frac{1}{2}r^2\right) r\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi.$$

∎

Definition 3.3 [Convergence for distribution functions] For a distribution function F, define C_F as the set of all continuous points of F:

$$C_F = \{x \in \mathbb{R} : F(x-) = F(x)\}.$$

1. For distributions $F_n, n \geq 1$ and F, we say $F_n \xrightarrow{w} F$ if

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

2. For random variables $X_n, n \geq 1$ and X, we say $X_n \xrightarrow{d} X$ if the distribution of X_n weakly converges to that of X.

The discontinuous points are excluded for defining weak convergence, since otherwise the convergence mode will be too limited.

■ **Example 3.1** [Examples of weak convergence] For a constant a, define distribution functions F_n, F as

$$F_n(x) = \begin{cases} 0, & \text{if } x < a - 1/n \\ \frac{n}{2}(x - a + \frac{1}{n}), & \text{if } a - 1/n \le x < a + 1/n \\ 1, & \text{if } a + 1/n \le x \end{cases} \qquad F(x) = \begin{cases} 0, & \text{if } x < a \\ 1, & \text{if } a \le x \end{cases}$$

As a result,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{1}{2}, & \text{if } x = a \\ 1, & \text{if } x > a \end{cases}$$

Therefore, $F_n(x)$ converges to $F(x)$ except for $x = a$. ■

General Guidance for Proving Central Limit Theorem.

1. Firstly define the characteristic function $\phi : \mathbb{R} \to \mathbb{C}$:

$$\phi(\theta) \triangleq \int_0^\infty e^{-i\theta x} F(\mathrm{d}x)$$

   Then showing the following facts.

2. ϕ uniquely determines distribution function F.
3. $F_n \xrightarrow{w} F$ if and only if $\phi_n(\theta) \to \phi(\theta), \forall \theta \in \mathbb{R}$.
4. $\phi_n(\theta)$ for the distribution of $\frac{1}{\sigma\sqrt{n}}(S_n - mn)$ converges to $\exp(-\frac{1}{2}\theta^2)$ for all $\theta \in \mathbb{R}$.

3.2.2. Complex Numbers and Functions

Definition 3.4 [Exponential Function] Define the **exponential funciton** with variable $z \in \mathbb{C}$ as

$$e^z = \sum_{n=0}^\infty \frac{1}{n!} z^n.$$

Exponential function admits the Euler's formula:

$$e^{iz} = \sum_{n=0}^{\infty} \frac{1}{n!}(iz)^n = \cos z + i \sin z.$$

Chapter 4

Week 4

4.1. Monday

4.1.1. Basic Convergence Results

In order to show CLT, first study basic convergence results for $\mathbb{E}[X_n] \to \mathbb{E}[X]$. The reason is that convergence of expectations has wide applications. We may apply those convergence results for the sequence of $f(X_n)$. For instance, if using the test function $f(x) = e^{i\theta x}$, then $\mathbb{E}[f(X_n)]$ is the characteristic function of the r.v. X_n. Thus the proof of CLT is reduced to show that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

The expectation for a random variable $X : \Omega \to \mathbb{R}$ is defined as follows:

1. When X is simple, i.e., $X : \Omega \to \{x_i\}_{i=1}^{\ell}$, define

$$\mathbb{E}[X] = \sum_{i=1}^{\ell} x_i \mathbb{P}(X = x_i)$$

2. When $X \geq 0$, define an approximation of X:

$$X_n = \begin{cases} (k-1)2^{-n}, & \text{if } (k-1)2^{-n} \leq X < k2^{-n}, k = 1, 2, \ldots, n2^n + 1 \\ n, & \text{if } X \geq n \end{cases}$$

   Then X_n is simple and $X_n \uparrow X$. Then define $\mathbb{E}[X] \triangleq \lim_{n\to\infty} \mathbb{E}[X_n]$.

3. When X is general, define $X = X^+ - X^-$ and take $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

Proposition 4.1 Suppose that $X \geq 0$ and consider a sequence of simple non-negative r.v.'s $\{Y_n\}_{n\geq 1}$. If $Y_n \uparrow X$ as $n \to \infty$, then $\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[X]$.

Proof. Define $X_k = \eta_k(X)$ and $Y_{n,k} = \eta_k(Y_n)$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}[Y_n] = \lim_{n\to\infty}\lim_{k\to\infty} \mathbb{E}[Y_{n,k}] \le \lim_{k\to\infty} \mathbb{E}[X_k] = \mathbb{E}[X],$$

where the inequality is because that $Y_n \le X$ implies $\eta_k(Y_n) \le \eta_k(X)$, i.e., $Y_{n,k} \le X_k$.

For the reverse direction, let Z be a simple r.v. such that $0 \le Z \le X$. Define $b = \max_{\omega\in\Omega} Z(\omega)$. Note that

$$\mathbb{E}[Z] = \mathbb{E}[Z1(Y_n + \varepsilon > Z)] + \mathbb{E}[Z1(Y_n + \varepsilon \le Z)]$$

Therefore, for any $\varepsilon > 0$,

$$\mathbb{E}[Z] \le \mathbb{E}[Y_n + \varepsilon] + \mathbb{E}[Z1(Y_n + \varepsilon \le X)] \le \mathbb{E}[Y_n] + \varepsilon + b\mathbb{P}(\varepsilon \le X - Y_n).$$

Taking $n \to \infty$, $\mathbb{E}[Z] \le \lim_{n\to\infty} \mathbb{E}[Y_n] + \varepsilon$. Thus $\mathbb{E}[Z] \le \lim_{n\to\infty} \mathbb{E}[Y_n]$. Take $Z = X_k$, we have

$$\mathbb{E}[X] = \lim_{k\to\infty} \mathbb{E}[X_k] \le \lim_{n\to\infty} \mathbb{E}[Y_n].$$

■

The linearity of expectation can be shown by applying this proposition.

Proposition 4.2 — Expectations of almost surely identical r.v.'s. If $X = Y$ a.s., and their expectations are finite, then $\mathbb{E}[X] = \mathbb{E}[Y]$.

Proof. It suffices to show that $\mathbb{P}(X = 0) = 1$ implies $\mathbb{E}[X] = 0$. Let $X = X^+ - X^-$, and it suffices to show $\mathbb{E}[X^+] = 0$ and $\mathbb{E}[X^-] = 0$. Construct $X_n^+ = \eta_n(X^+)$, which is a simple r.v. and $0 \le X_n^+ \uparrow X^+$. As a result,

$$\mathbb{E}[X^+] = \lim_{n\to\infty} \mathbb{E}[X_n^+]$$

Note that $\mathbb{P}(X_n^+ > 0) = \mathbb{P}(X^+ \ge 2^{-n}) = \mathbb{P}(X \ge 2^{-n}) \le 1 - \mathbb{P}(X = 0) = 0$, we imply $\mathbb{P}(X_n^+ = 0) = 1, \forall n$, i.e., $\mathbb{E}[X^+] = 0$. ■

Proposition 4.3 Suppose that $X \geq 0$ a.s. and consider a sequence of a.s. simple non-negative r.v.'s $\{Y_n\}_{n\geq 1}$. If $Y_n \uparrow X$ a.s. as $n \to \infty$, then $\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[X]$.

Proof. Define

$$\Omega_0 = \{X \geq 0, Y_n \text{ is simple for all } n,\, Y_n \uparrow X\}.$$

Since $\mathbb{P}(\cap_{\ell=1}^{\infty} A_\ell) = 1$ when $\mathbb{P}(A_\ell) = 1$, we imply $\mathbb{P}(\Omega_0) = 1$. Repeat the same proof as in Proposition 4.1 and then apply Proposition 4.2.

■

> **Theorem 4.1 — Bounded Convergence Theorem.** Suppose that $X_n \xrightarrow{p} X$ as $n \to \infty$, and $|X_n| \leq b$ a.s., then
>
> $$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proof. Firstly argue that there exists a sub-sequence of X_n that converges to X a.s. Construct a subsequence X_{n_k} of X_n such that $\mathbb{P}(|X_{n_k} - X| > \varepsilon) < 2^{-k}$, which implies

$$\sum_{k\geq 1} \mathbb{P}(|X_{n_k} - X| > \varepsilon) < 1 \implies X_{n_k} \xrightarrow{a.s.} X.$$

Therefore, $|X| \leq b$ a.s. Then

$$\begin{aligned}
|\mathbb{E}[X_n] - \mathbb{E}[X]| &\leq \mathbb{E}[|X_n - X|] \\
&= \mathbb{E}[|X_n - X| 1(|X_n - X| > \varepsilon)] + \mathbb{E}[|X_n - X| 1(|X_n - X| \leq \varepsilon)] \\
&\leq 2b\mathbb{P}(|X_{n_k} - X| > \varepsilon) + \varepsilon
\end{aligned}$$

Take $n \to \infty$ both sides, and take $\varepsilon \to 0$, the desired result holds. ■

> **Theorem 4.2 — Monotone Convergence Theorem.** Suppose that $X_n \geq 0$ a.s., and $X_n \uparrow X$ a.s., then
>
> $$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proof. The condition $X_n \leq X$ implies $\mathbb{E}[X_n] \leq \mathbb{E}[X]$, i.e., $\limsup_{n\to\infty} \mathbb{E}[X_n] \leq \mathbb{E}[X]$. It suffices to show that

$$\mathbb{E}[X] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$

It suffices to construct simple r.v.'s $Y_k \geq 0$ such that

$$Y_k \leq X_k, \forall k \geq 1, \quad \text{and } Y_k \uparrow X \text{ as } k \to \infty.$$

Therefore, $\mathbb{E}[Y_k] \to \mathbb{E}[X] \leq \liminf_{k\to\infty} \mathbb{E}[X_k]$.

Construct $X_{n,k} = \eta_k(X_n)$ and $X_{n,k} \uparrow X_n$. Define $Y_k = \max(X_{1,k}, X_{2,k}, \ldots, X_{k,k})$, which implies $X_{n,k} \leq Y_k \leq X_k$.

■

Theorem 4.3 — **Fatou's Lemma.** Suppose that $X_n \geq 0$ a.s., then

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$

Proof. Take $Y_n = \inf_{\ell \geq n} X_\ell$, then $Y_n \uparrow Y \triangleq \liminf_{n\to\infty} X_n$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[Y].$$

Considering that $Y_n \leq X_n$, we imply $\mathbb{E}[Y_n] \leq \mathbb{E}[X_n]$. Taking liminf both sides gives the desired result. ■

Theorem 4.4 — **Dominated Convergence Theorem.** Suppose that $X_n \xrightarrow{a.s.} X$ as $n \to \infty$, and $|X_n| \leq Y$ a.s. for some Y with $\mathbb{E}[Y] < \infty$, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Proof. Apply Fatou's lemma on $Y + X_n$ and $Y - X_n$ gives the desired result. ■

4.2. Thursday

4.2.1. Characteristic Function

Define the complex valued r.v. Z as

$$Z(\omega) = X(\omega) + iY(\omega), \quad \omega \in \Omega.$$

When X, Y have finite expectations, define the expectation of Z as

$$\mathbb{E}(Z) = \mathbb{E}(X) + i\mathbb{E}(Y)$$

Definition 4.1 [Characteristic Function] Given a distribution function F, define the characteristic function ϕ as

$$\begin{aligned}\phi(\theta) &= \int_{-\infty}^{+\infty} e^{i\theta x} F(\mathrm{d}x)\\ &= \int_{-\infty}^{+\infty} \cos(\theta x) F(\mathrm{d}x) + i \int_{-\infty}^{+\infty} \sin(\theta x) F(\mathrm{d}x)\end{aligned}$$

For random variable X subject to F, we have

$$\phi(\theta) = \mathbb{E}(\cos(\theta X)) + i\mathbb{E}(\sin(\theta X)).$$

Proposition 4.4 If complex valued random variable Z has a finite expectation, then

$$|\mathbb{E}(Z)| \le \mathbb{E}(|Z|).$$

Proof. The definition for real valued random variable has three steps. Now we only check the first step for complex valued random variable. Suppose that Z takes finite many values $z_1, z_2, \ldots, z_m$, and $z_j = x_j + iy_j$, then

$$\mathbb{E}[Z] = \sum_{j=1}^{m} \sum_{k=1}^{m} (x_j + iy_k)\mathbb{P}(X = x_j, Y = y_k).$$

Apply the triangle inequality,

$$\begin{aligned}
|\mathbb{E}(Z)| &\le \sum_{j=1}^{m} \sum_{k=1}^{m} |x_j + iy_k| \mathbb{P}(X = x_j, Y = y_k) \\
&= \sum_{\ell=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} |x_j + iy_k| \mathbb{P}(Z = z_\ell, X = x_j, Y = y_k) \\
&= \sum_{\ell=1}^{m} |z_\ell| \mathbb{P}(Z = z_\ell) = \mathbb{E}(|Z|).
\end{aligned}$$

■

Corollary 4.1 For a characteristic function ϕ,

$$|\phi(\theta)| \le 1, \quad \forall \theta \in \mathbb{R}.$$

Therefore, $\phi(\theta)$ is well-defined and finite for any θ.

Proposition 4.5 Suppose that X, Y are independent, then

$$\phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta).$$

Proof. It suffices to verify that $\mathbb{E}(e^{i\theta(X+Y)}) = \mathbb{E}(e^{i\theta X})\mathbb{E}(e^{i\theta Y})$:

$$\mathbb{E}(e^{i\theta X})\mathbb{E}(e^{i\theta Y}) = (\mathbb{E}[\cos\theta X] + i\mathbb{E}[\sin\theta X])\mathbb{E}(e^{i\theta Y})$$

Verify that

$$\mathbb{E}\left(\cos\theta X \cdot (\cos\theta Y + i\sin\theta Y)\right) = \mathbb{E}[\cos\theta X]\mathbb{E}(e^{i\theta Y})$$

and

$$\mathbb{E}\left(\sin\theta X \cdot (\cos\theta Y + i\sin\theta Y)\right) = \mathbb{E}[\sin\theta X]\mathbb{E}(e^{i\theta Y})$$

■

We may compute the characteristic function $\phi(\theta)$ for normal distribution. Suppose that f is the density of $\mathcal{N}(0,1)$, then

$$\phi(\theta) = \int \cos(\theta x) f(x)\,\mathrm{d}x + i \int \sin(\theta x) f(x)\,\mathrm{d}x = \int \cos(\theta x) f(x)\,\mathrm{d}x,$$

where the last equality is because that f is an even function. We study the derivative of $\phi'(\theta)$:

$$\begin{aligned}\phi'(\theta) &= \int (-x\sin(\theta x)) f(x)\,\mathrm{d}x = \int \sin(\theta x) f'(x)\,\mathrm{d}x \\ &= [\sin(\theta x) f(x)]_{-\infty}^{\infty} - \int (\sin(\theta x))' f(x)\,\mathrm{d}x = -\theta\phi(\theta).\end{aligned}$$

Solving this differential equation with $\phi(0) = 1$, we imply

$$\phi(\theta) = \exp\left(-\frac{1}{2}\theta^2\right).$$

4.2.2. Inversion Formula

Based on the distribution function F, we can define the characterisic fucntion. The inverse is also possible. Let C_F be the set of all continuous points of F.

Theorem 4.5 — **Inversion Formula.** For $x, y \in C_F$ satisfying $x < y$,

$$\lim_{a\downarrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{i\theta}(e^{-i\theta x} - e^{-i\theta y})\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\,\mathrm{d}\theta = F(y) - F(x).$$

Hence, F is determined by ϕ.

The formula above is slightly different from the typical inversion formula:

$$\lim_{n\to\infty} \frac{1}{2\pi} \int_{-n}^{n} \frac{1}{i\theta}(e^{-i\theta x} - e^{-i\theta y})\phi(\theta)\,\mathrm{d}\theta = F(y) - F(x).$$

Theorem 4.6 — Parseval's Lemma. For $a > 0$ and $u, \theta \in \mathbb{R}$,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta u}\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}a}\exp\left(-\frac{(u-s)^2}{2a^2}\right)F(\mathrm{d}s) \tag{4.1}$$

Proof. Direct calculation on the LHS gives

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta u}\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta u}\int_{-\infty}^{\infty} e^{i\theta s}F(\mathrm{d}s)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(s-u)\theta}F(\mathrm{d}s)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta \\ &= \frac{1}{\sqrt{2\pi}a}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(s-u)\theta}\frac{a}{\sqrt{2\pi}}\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta F(\mathrm{d}s)\end{aligned}$$

Since $\frac{a}{\sqrt{2\pi}}\exp\left(-\frac{a^2\theta^2}{2}\right)$ is the pdf of $\mathcal{N}(0, 1/a^2)$,

$$\int_{-\infty}^{\infty} e^{i(s-u)\theta}\frac{a}{\sqrt{2\pi}}\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta = \exp\left(-\frac{1}{2a^2}(s-u)^2\right).$$

As a result,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta u}\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta = \frac{1}{\sqrt{2\pi}a}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2a^2}(s-u)^2\right)F(\mathrm{d}s).$$

■

R Setting g as the pdf of $\mathcal{N}(u, a^2)$, then the characteristic function of g us $e^{iu\theta - \frac{1}{a^2}\theta^2}$. Therefore, the Parseval's formula can be written as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(\theta)\overline{\psi(\theta)}\,\mathrm{d}\theta = \int_{-\infty}^{\infty} g(s)F(\mathrm{d}s).$$

This formula holds for any continous pdf g under an integrability condition on $|\phi(\theta)\psi(\theta)|$. If F has a density, this fact can be interpreted that the inner products are proportional for the Hilbert spaces of characteristic functions and that of probability density functions, i.e., two Hilbert spaces are isometric.

Based on the Parseval's formula, we may begin to show the inversion formula:

Proof. Integrate both sides of (4.1) within the interval $[x,y]$ by $\mathrm{d}u$. The LHS becomes:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{i\theta}[e^{-i\theta x}-e^{-i\theta y}]\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta.$$

he RHS of (4.1) becomes:

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{x}^{y}\frac{1}{\sqrt{2\pi}a}\exp\left(-\frac{(u-s)^2}{2a^2}\right)\mathrm{d}uF(\mathrm{d}s) &= \int_{-\infty}^{\infty}\int_{(x-s)/a}^{(y-s)/a}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right)\mathrm{d}vF(\mathrm{d}s)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}1(x-av<s\leq y-av)F(\mathrm{d}s)\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}\,\mathrm{d}v\\ &= \int_{-\infty}^{\infty}[F(y-av)-F(x-av)]\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}\,\mathrm{d}v\end{aligned}$$

Taking $a\to 0$ both sides,

$$\lim_{a\downarrow 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{i\theta}(e^{-i\theta x}-e^{-i\theta y})\phi(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta = F(y)-F(x)$$

$\blacksquare$

Chapter 5

Week 5

5.1. Monday

5.1.1. Proof Outline of CLT

The key for showing hte CLT is about the convergence of characteristic functions:

Theorem 5.1 For distributions $F_n, n \geq 1$ and F, their characteristic functions $\phi_n, n \geq 1$ and ϕ, the following four conditions are equivalent:

1. $F_n \xrightarrow{w} F$ as $n \to \infty$
2. There exists $X_n, n \geq 1$ and X such that $X_n \overset{d}{\sim} F_n$ for $n \geq 1$ and $X \overset{d}{\sim} F$, while $X_n \xrightarrow{a.s.} X$.
3. For any bounded and continuous function $g : \mathbb{R} \to \mathbb{R}$,
$$\lim_{n\to\infty} \int_0^\infty g(x) F_n(\mathrm{d}x) = \int_0^\infty g(x) F(\mathrm{d}x).$$
4. For any $\theta \in \mathbb{R}$, $\lim_{n\to\infty} \phi_n(\theta) = \phi(\theta)$.

Outline for Showing CLT. Recall that $X_n, n \geq 1$ are i.i.d., and $S_n = \sum_{\ell=1}^n X_\ell$. Define

$$Z_n = \frac{1}{\sqrt{n}\sigma}(S_n - mn), \qquad n \geq 1,$$

and F_n, Φ_n are distribution and characteristic functions of Z_n. If we can show that

$$\lim_{n\to\infty} \Phi_n(\theta) = \exp\left(-\frac{1}{2}\theta^2\right), \quad \theta \in \mathbb{R}, \tag{5.1}$$

then by theorem 5.1 and the uniqueness of characteristc functions, the CLT is proved.

Now we give an intuitive proof of (5.1). Define $Y_n \triangleq \frac{1}{\sigma}(X_n - m)$, which is normalized X_n. Thus $Z_n = \frac{1}{\sqrt{n}}\sum_{\ell=1}^{n} Y_\ell$, which implies

$$\Phi_n(\theta) = \mathbb{E}(e^{i\theta Z_n}) = \prod_{\ell=1}^{n}\mathbb{E}(e^{\frac{i\theta}{\sqrt{n}}Y_\ell}) = \phi^n(\theta/\sqrt{n}),$$

where ϕ is the characteristic function of Y_ℓ. By the Taylor expansion of $\phi(\theta/\sqrt{n})$ around the origin, since $\phi^{(n)}(0) = i^n\mathbb{E}[Y_\ell^n]$,

$$\phi(\theta/\sqrt{n}) = 1 - \frac{1}{2}\phi''(0)\frac{\theta^2}{n} + o(\theta^2/n) = 1 - \frac{\theta^2}{2n} + o(\theta^2/n).$$

As a result,

$$\lim_{n\to\infty}\Phi_n(\theta) = \lim_{n\to\infty}\left[1 - \frac{\theta^2}{2n} + o(\theta^2/n).\right]^n = \exp\left(-\frac{1}{2}\theta^2\right).$$

Now we give a formal proof that $\phi^n(\theta/\sqrt{n}) \to e^{-1/2\theta^2}$.

Proof. Firstly, inductively derive these inequalities:

$$\begin{aligned}
|e^{ix} - 1| \le 2, \qquad\qquad |e^{ix} - 1| &= \left|\int_0^x ie^{iu}\,\mathrm{d}u\right| \le \int_0^{|x|}|ie^{iu}| = |x| \\
|e^{ix} - 1 - ix| &= \left|\int_0^x i(e^{iu} - 1)\,\mathrm{d}u\right| \le \int_0^{|x|}|e^{iu} - 1|\,\mathrm{d}u \le \min\left(2|x|, \frac{1}{2}x^2\right), \\
\left|e^{ix} - 1 - ix + \frac{1}{2}x^2\right| &= \left|\int_0^x i(e^{iu} - 1 - iu)\,\mathrm{d}u\right| \\
&\le \int_0^{|x|}|e^{iu} - 1 - iu|\,\mathrm{d}u \le \min\left(x^2, \frac{1}{6}|x|^3\right).
\end{aligned}$$

As a result, set $x = \theta/\sqrt{n}Y_\ell$,

$$\begin{aligned}
\left|\phi(\theta/\sqrt{n}) - 1 + \frac{1}{2n}\theta^2\right| &= \left|\mathbb{E}[e^{ix}] - 1 + \frac{1}{2n}\theta^2\right| \\
&\le \mathbb{E}\left|e^{ix} - 1 - ix + \frac{1}{2}x^2\right| \\
&\le \frac{\theta^2}{n}\mathbb{E}\min\left(Y^2, \frac{1}{6}|Y|^3\right) = o(\theta^2/n)
\end{aligned}$$

Next, for $z,w \in \mathbb{C}$, define $a = \max(|z|,|w|)$, we have the following inequality:

$$|z^n - w^n| = \left|(z-w)\sum_{\ell=0}^{n-1} z^\ell w^{n-1-\ell}\right| \leq |z-w|na^{n-1}.$$

Apply this inequality with $z \triangleq \phi(\theta/\sqrt{n})$ and $w = e^{-\theta^2/2n}$,

$$\begin{aligned}\left|\phi^n(\theta/\sqrt{n}) - \exp\left(-\frac{1}{2}\theta^2\right)\right| &\leq \left|\phi(\theta/\sqrt{n}) - \exp\left(-\frac{1}{2n}\theta^2\right)\right| n \\ &\leq \left|\phi(\theta/\sqrt{n}) - 1 + \frac{\theta^2}{2n}\right| n + \left|1 - \frac{\theta^2}{2n} - \exp\left(-\frac{1}{2n}\theta^2\right)\right| n \to 0.\end{aligned}$$

■

Now we give a proof for Theoerm 5.1:

1) implies 2).

- Construct $X \overset{w}{\sim} F$ and $X_n \overset{d}{\sim} F_n$ as the following. Define h,h_n such that

  $$h(y) = \inf\{u:\ y \leq F(u)\}, \quad h_n(y) = \inf\{u:\ y \leq F_n(u)\}.$$

  Let $U \sim \mathcal{U}[0,1]$, and define $X = h(U), X_n = h_n(U)$. It suffices to show that $X_n \xrightarrow{a.s.} X$. In other words, it suffices to show that $h_n(y) \to h(y), \forall y \in C_h$.

- Since C_h, C_F are dense in $\mathbb{R}$, for any $y \in C_h$ and any $\varepsilon > 0$,

  $$\exists x \in C_F, \text{ such that } h(y) - \varepsilon < x < h(y). \tag{5.2}$$

  Then $x < h(y)$ is equivalent to $F(x) < y$. Because $F_n \xrightarrow{w} F$, $F_n(x) < y$ for large n. As a result, $x < h_n(y)$, and together with (5.2),

  $$h(y) - \varepsilon < x < h_n(y).$$

  Taking liminf for $n \to \infty$, and $\varepsilon \to 0$,

  $$h(y) \leq \liminf h_n(y).$$

- Then consider lower bonuding $h(y)$ for $y \in C_h$:

$$\forall y' > y, \forall \varepsilon > 0, \exists x \in C_F, \text{ such that } h(y') < x < h(y') + \varepsilon. \tag{5.3}$$

  Then $h(y') < x$ implies $y' \leq F(x)$. Hence, $y < y' \leq F(x)$, and $y < F_n(x)$ for large n. Thus $y \leq F_n(x)$, i.e, $h_n(y) \leq x$. Together with (5.3),

$$h_n(y) \leq x < h(y') + \varepsilon.$$

  Taking limsup for $n \to \infty$ and $\varepsilon \to 0$,

$$\forall y' > y, \limsup h_n(y) \leq h(y').$$

  Taking $y' \downarrow y$,

$$\limsup h_n(y) \leq h(y).$$

■

2) implies 3). It suffices to show that

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)].$$

Since $X_n \xrightarrow{a.s.} X$, and g is continuous, we imply $g(X_n) \xrightarrow{a.s.} g(X)$. Considering that g is bounded, applying dominated convergence theorem gives the desired result. ■

3) implies 4). It suffices to show that

$$\lim_{n\to\infty} \phi_n(\theta) = \lim_{n\to\infty} \mathbb{E}[\cos(\theta X_n) + i\sin(\theta X_n)] = \phi(\theta),$$

which is true by applying 3) and take $g(x)$ to be $\cos(\theta x)$ and $\sin(\theta x)$. ■

4) implies 1). Assume that 4) holds but 1) does not. Then there exists a subsequence $\{F_{n_k}\}$ such that

$$\exists x_0 \in C_F, \quad \liminf |F_{n_k}(x_0) - F(x_0)| > 0.$$

Apply Helly Theorem, there exists a subsequence of $\{F_{n_k}\}$, denoted by $\{F_{n'_k}\}$, such that $F_{n'_k} \to G$ vaguely for some non-decreasing function G. It suffices to show that $G = F$ to construct a contradiction.

Apply Perseval's Lemma for $F_{n'_k}$ and $\phi_{n'_k}$:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\theta u}\phi_{n'_k}(\theta)\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}a}\exp\left(-\frac{(u-s)^2}{2a^2}\right)F_{n'_k}(\mathrm{d}s) \tag{5.4}$$

Integrable both sides w.r.t u on the interval $[x,y]$, and taking $k \to \infty$,

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{1}{i\theta}(\exp(-i\theta x) - \exp(-i\theta y))\right]\cdot\phi(\theta)\cdot\exp\left(-\frac{a^2\theta^2}{2}\right)\mathrm{d}\theta\\ &= \int_{-\infty}^{\infty}[G(y-av) - G(x-av)]\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right)\mathrm{d}v\end{aligned}$$

By Perseval's Lemma, the LHS becomes

$$\int_{-\infty}^{\infty}[F(y-av) - F(x-av)]\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right)\mathrm{d}v$$

Taking $a \to 0$, we have for $x, y \in C_F \cap C - G$,

$$F(y) - F(x) = G(y) - G(x).$$

The proof is complete. ■

Chapter 6

Week 6

6.1. Monday

6.1.1. Conditional Expectation

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where the σ-field $\mathcal{F}$ represents the collection of all information to be measured.

In practice, we often meet the situation that the information is partially available. This can be represented by $\mathcal{G}$, another σ-field but a subset of $\mathcal{F}$. We call $\mathcal{G}$ a sub-σ-field.

Definition 6.1 [Partition] We call $\mathcal{A} \triangleq \{A_i \in \mathcal{F}; i \geq 1\}$ a partition of Ω if

$$A_i \cap A_j = \emptyset, \forall i \neq j, \quad \text{and} \quad \bigcup_{i=1}^{\infty} A_i = \Omega.$$

Denote the minimal σ-field on Ω containing $\mathcal{A}$ by $\sigma(\mathcal{A})$, which is a sub-σ-field, and it is easy to show that

$$\sigma(\mathcal{A}) = \left\{ \bigcup_{i \in B} A_i : \ B \subseteq \mathbb{N} \right\}.$$

The partition $\{A_i \in \mathcal{F}; i \geq 1\}$ denotes the set of minimal information units for $\sigma(\mathcal{A})$.

The expectation $\mathbb{E}[X]$ can be considered as the prediction of $X(\omega)$ under no information on the sample $\omega \in \Omega$. Instead, if we have some information about ω, e.g., $\omega \in A_1$ for $A_1 \in \mathcal{F}$ and $\mathbb{P}(A_1) > 0$, then the prediction of X given A_1 is defined as

$$\mathbb{E}[X \mid A_1] \triangleq \frac{\mathbb{E}[X1_{A_1}]}{\mathbb{P}(A_1)}.$$

We call it the **conditional expectation of X given A_1**.

Based on this definition, we can define conditional expectation given a simple σ-field. For $A_1 \in \mathcal{F}, A_2 = A_1^c$, set $\mathcal{G} = \{\emptyset, A_1, A_2, \Omega\}$. Then define the **conditional expectation of X given $\mathcal{G}$** as

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \begin{cases} \mathbb{E}[X \mid A_1] 1(\mathbb{P}(A_1) > 0), & \text{for } \omega \in A_1 \\ \mathbb{E}[X \mid A_2] 1(\mathbb{P}(A_2) > 0), & \text{for } \omega \in A_2 \end{cases} \tag{6.1}$$

The following presents an image of conditional expectation given $\sigma(\mathcal{A})$, where $\mathcal{A}$ is a partition:

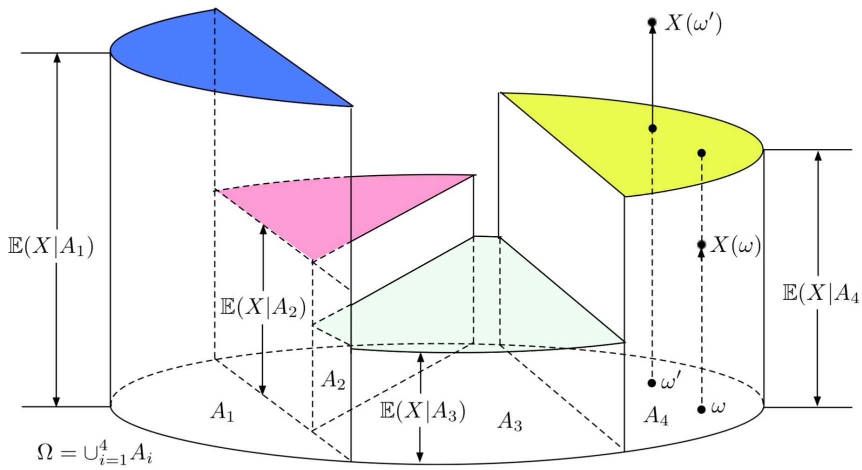

Figure 6.1: Illustration of $\mathbb{E}[X \mid \sigma(\mathcal{A})]$, where $\mathcal{A} = \{A_i \in \mathcal{F};\ \cup_i A_i = \Omega\}$.

We extend the definition in (6.1) to general σ-field:

Definition 6.2 Given an integrable random variable and a sub σ-field $\mathcal{G}$, we say Z is the conditional expectation of X given $\mathcal{G}$, denoted by $\mathbb{E}[X \mid \mathcal{G}]$, if

1. Z is $\mathcal{G}$-measurable;
2. For any $A \in \mathcal{G}$, $\mathbb{E}[Z 1_A] = \mathbb{E}[X 1_A]$.

The existence of such Z is because of Radon-Nikodym theorem. Intuitively, this can be seen by approximating $\mathcal{G}$ from partitions.

Theorem 6.1 — Uniqueness of Conditional Expectation. Under the setting of Definition 6.2, suppose that Z_1, Z_2 satisfies the condition 1) and 2), then $Z_1 = Z_2$ a.s.

Proof. Let $A = \{Z_1 - Z_2 > 0\} \in \mathcal{G}$, then by condition 2) in Definition 6.2,

$$\mathbb{E}[(Z_1 - Z_2)1_A] = \mathbb{E}[Z_1 1_A] - \mathbb{E}[Z_2 1_A] = \mathbb{E}[X 1_A] - \mathbb{E}[X 1_A] = 0.$$

This equality implies that $\mathbb{P}(A) = 0$. Similarly, $\mathbb{P}(Z_2 - Z_1 > 0) = 0$. The proof is complete. ■

R Since $\mathbb{E}[X \mid \mathcal{G}]$ is uniquely determined a.s., in what follows, we omit a.s. unless otherwise stated.

Proposition 6.1 — Basic properties of the conditional expectation.

1. $$\mathbb{E}(\mathbb{E}[X \mid \mathcal{G}]) = \mathbb{E}[X]$$

2. $$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$$

3. If X and $\mathcal{G}$ are independent, i.e., for any $B \in \mathcal{B}(\mathbb{R})$ and $A \in \mathcal{G}$, the events $\{X \in B\}$ and A are independent, then

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

4. If Y is $\mathcal{G}$-measurable and $\mathbb{E}|X|, \mathbb{E}|XY| < \infty$, then

$$\mathbb{E}[XY \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}]Y.$$

Now we can also define the conditional expectation given a random variable as the following:

Definition 6.3 [Conditional Expectation given a Random Variable] Given a random variable Y, define the generated σ-field

$$\sigma(Y) \triangleq \{\{Y \in B\} : B \in \mathcal{B}(\mathbb{R})\}.$$

The conditional expectation of X given Y, denoted as $\mathbb{E}[X \mid Y]$, is defined as

$$\mathbb{E}[X \mid Y] \triangleq \mathbb{E}[X \mid \sigma(Y)].$$

Definition 6.4 [Conditional Probability] Given $A \in \mathcal{F}$ and a sub-σ-field $\mathcal{G}$, define the **conditional probability of** A **given** $\mathcal{G}$ as

$$\mathbb{P}(A \mid \mathcal{G}) \triangleq \mathbb{E}[1_A \mid \mathcal{G}].$$

In particular, for a random variable Y,

$$\mathbb{P}(A \mid Y) \triangleq \mathbb{E}[1_A \mid \sigma(Y)]$$

is called the **conditional probability of** A **given** Y.

The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ minimizes the expected squared error to estimate X among all $\mathcal{G}$-measurable random variables, when $\mathbb{E}[X^2] < \infty$.

Theorem 6.2 — **Minimizing Squared Error.** For a random variable and sub σ-field $\mathcal{G}$, assume that $\mathbb{E}[X^2] < \infty$. Then, for $\mathcal{G}$-measurable random variable Z, $\mathbb{E}[(X-Z)^2]$ is minimized by $Z = \mathbb{E}[X \mid \mathcal{G}]$.

Proof. Expand $\mathbb{E}[(X-Z)^2]$ as the following:

$$\begin{aligned}\mathbb{E}[(X-Z)^2] &= \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[X \mid \mathcal{G}] - Z)^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])^2] + \mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}] - Z)^2] + 2\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])(\mathbb{E}[X \mid \mathcal{G}] - Z)]\end{aligned}$$

In particular, for $\mathcal{G}$-measurable random variable H,

$$H\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[HX \mid \mathcal{G}] \implies \mathbb{E}[H\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[HX]$$

As a result, $\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])H] = 0$. Hence

$$\mathbb{E}[(X - Z)^2] = \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])^2] + \mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}] - Z)^2].$$

This objective is minimized by $Z = \mathbb{E}[X \mid \mathcal{G}]$. ■

6.1.2. Discrete Time Stochastic Process

Definition 6.5 [Stochastic Process]

- Consider the topological space $(S, \mathcal{O})$, define

$$\mathcal{B}(S^n) = \sigma(\mathcal{O}^n).$$

- The discrete time stochastic process with state space S is defined as

$$X. = \{X_n : \Omega \to S : \ n \geq 0\}.$$

  For each $\omega \in \Omega$, $X.(\omega) \triangleq \{X_n(\omega) : \ n \geq 0\}$ is called a sample path.

- For $n \geq 0$, let $\mathcal{F}_n^X \triangleq \{\{(X_0, X_1, \dots, X_n) \in \boldsymbol{B}\} : \ \boldsymbol{B} \in \mathcal{B}(S^n)\}$, which is a sub σ-field of $\mathcal{F}$, which presents all information generated by $X_0, X_1, \dots, X_n$.

A collection of sample paths is called a history of $X.$. The sub σ-field $\mathcal{F}_n^X$ can be viewed as the observation on the history of $X.$ up to n.

Definition 6.6 [Filtration] A sequence of sub σ-fields, $\mathbb{F} \triangleq \{\mathcal{F}_n : \ n \geq \}$, is called a **filtration** if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}, \forall n \geq 0$. We say the stochastic process is **adapted** to filtration

$\mathbb{F}$ if $\mathcal{F}_n^X \subseteq \mathcal{F}_n, \forall n \geq 0$. The special **filtration**, $\mathbb{F}^X \triangleq \{\mathcal{F}_n^X : n \geq 0\}$, is called the **natural filtration** of $X.$.

Given a stochastic process $X.$, one may prdict the value of $f(X_{n+1})$ given the information up to time n by

$$\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n].$$

We can also show that

$$\mathbb{E}[\mathbb{E}[f(X_{n+2}) \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[f(X_{n+2}) \mid \mathcal{F}_n],$$

which means that smaller information makes the conditional expectation smoother as a function of sample $\omega \in \Omega$.

■ **Example 6.1** [Random Walk] Suppose that $U_1, U_2, \ldots$ are i.i.d. with $\mathbb{P}(U_n = 1) = p, \mathbb{P}(U_n = -1) = 1 - p$. Then define the stochastic process $X.$ with $X_0 = 0$,

$$X_n = U_1 + U_2 + \cdots + U_n, \quad n \geq 1.$$

Then $X.$ is called a simple random walk. We can see that $X.$ is $\mathbb{F}^X$-adapted, and

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n^X] = X_n + \mathbb{E}[U_{n+1} \mid \mathcal{F}_n^X] = X_n + (p - (1-p)), \quad n \geq 0.$$

6.2. Thursday

6.2.1. Markov Process

Now we study one of the most popular stochastic processes, the Markov processes. Denote $D_b(S)$ as the set of all bounded measurable functions from S to $\mathbb{R}$.

Definition 6.7 [Markov Process] A stochastic process $X.$ with state space S is called a **discrete-time Markov process** (DTMC) for filtration $\mathbb{F}$ if

1. $X.$ is adapted to $\mathbb{F}$;
2. $\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \mathbb{E}[f(X_{n+1}) \mid X_n]$ for any $f \in D_b(S)$ and $n \geq 0$.

In particular, if S is countable, then it is called a Markov Chain.

It is clear that for a Markov Chain $X.$, the second condition is equivalent to

$$\mathbb{P}[X_{n+1} = j \mid \mathcal{F}_n] = \mathbb{P}[X_{n+1} = j \mid X_n], \quad j \in S, n \geq 0.$$

Definition 6.8 [Time-Homogeneous] A DTMC $X.$ is said to have **time-homogeneous** transitions if the second condition in Definition 6.7 is independent of n.

Definition 6.9 [Transition of Markov Chain] Given a Markov Chain $X.$ with state space S, define the transition probability

$$p_{i,j}(n) = \mathbb{P}[X_{n+1} = j \mid X_n = i], \quad i,j \in S, n \geq 0.$$

Define the transition operator $P_n : D_b(S) \to D_b(S)$ as

$$P_n[f(i)] = \sum_{j \in S} f(j) p_{i,j}(n), \quad i \in S, f \in D_b(S).$$

In particular, if the Markov Chain $X.$ has time-homogeneous transitions, the matrix $P = (p_{i,j})_{i,j}$ is called a **transition matrix**. A directed graph (V, E) with

$$V = S, \quad E = \{e_{i,j} : \ p_{i,j} > 0, i,j \in S\}$$

is called a **transition diagram**.

Given a time-homogeneous Markov Chain, define the higher order transitons as the following:

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i), \quad i,j \in S$$
$$P^{(n)}[f(i)] = \mathbb{E}[f(X_n) \mid X_0 = i], \quad i \in S, f \in D_b(S).$$

Because of the time-homogeous transition assumption, for $n, \ell \geq 0$,

$$p_{i,j}^{(n)} = \mathbb{P}(X_{n+\ell} = j \mid X_\ell = i), \quad i,j \in S$$
$$P^{(n)}[f(i)] = \mathbb{E}[f(X_{n+\ell}) \mid X_\ell = i], \quad i \in S, f \in D_b(S).$$

Define $P^n[f(i)] = \sum_{i \in S} [P^n]_{i,j} f(j)$, then we can show that

$$P^{(n)}[f(i)] = P^n[f(i)], \quad i \in S, f \in D_b(S).$$

■ **Example 6.2** The simple random walk is a Markov chain with state space $\mathbb{Z}$ and natural filtration $\mathbb{F}^X$ since

$$\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n^X] = \mathbb{E}[f(X_n + U_{n+1}) \mid X_n]$$

The transition probability $p_{i,j}$ is

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = p \cdot 1(j = i+1) + (1-p) \cdot 1(j = i-1).$$

■

■ **Example 6.3** [Stochastic Process generated by i.i.d. Random Variables] Consider an i.i.d. sequence of random variables U_n, define $X.$ with

$$X_n = g(X_{n-1}, U_n).$$

We can see that $X.$ is a Markov Chain with time-homogeneous transitions for filtration

$\mathbb{F}^X$ when U_n is independent of $\mathcal{F}_n^X$ ■

■ **Example 6.4** Suppose that X_n takes values in $\mathbb{Z}_+$ with $X_0 = 1$,

$$X_{n+1} = \sum_{k=1}^{X_n} U_{n,k}, \quad n \geq 0,$$

where $U_{n,k}$'s are i.i.d. random variables. Then $X.$ is called a **branching process**. We can see that such a process is a Markov process:

$$\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n^X] = \mathbb{E}[f(\sum_{k=1}^{X_n} U_{n,k}) \mid X_n]$$

The transition probability is given by

$$p_{i,j} = \mathbb{P}\left(\sum_{k=1}^{i} U_{n,k} \middle| X_n = i\right).$$

Take $\lambda = \mathbb{E}[U_{n,k}]$, then

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[\sum_{k=1}^{X_n} \mathbb{E}[U_{n,k} \mid X_n]] = \lambda \mathbb{E}[X_n].$$

Therefore, $\mathbb{E}[X_n] = \lambda^n$. ■

6.2.2. Martingale and Semi-martingale

Definition 6.10 [Predictable] The stochastic process $X. \triangleq \{X_n : n \geq 0\}$ is said to be $\mathbb{F}$-predictable for filtration $\mathbb{F} \triangleq \{\mathcal{F}_n : n \geq 0\}$ if

1. X_0 is $\mathcal{F}_0$-measurable;
2. X_n is $\mathcal{F}_{n-1}$-measurable, $\forall n \geq 1$.

■

Define $E_0 = 0$ and

$$E_n = X_n - \mathbb{E}[X_n \mid \mathcal{F}_{n-1}], \quad n \geq 1,$$

which is an estimation error of X_n given $\mathbb{F}_{n-1}$, the observation by time $n-1$. Then X_n can be written as

$$X_n = \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] + E_n, \quad n \geq 1.$$

Then a stochastic process can be decomposed as a predictable process and a martingale:

$$Y_n = X_0 + \sum_{\ell=1}^{n} \mathbb{E}[X_\ell - X_{\ell-1} \mid \mathcal{F}_{\ell-1}], \qquad M_n = \sum_{\ell=1}^{n} E_\ell.$$

It is easy to see that

$$X_n = Y_n + M_n, n \geq 0. \tag{6.2}$$

We can verify that $\{Y_n : \ n \geq 0\}$ is predictable, while

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

Definition 6.11 [Martingale] A real valued stochastic prcoess $M. \triangleq \{M_n : \ n \geq 0\}$ is called an $\mathbb{F}$-martingale if

1. M_n is integrable for all n;
2. $M.$ is adapted to the filtration $\mathbb{F}$;
3. $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n, \forall n \geq 0$.

If the equality is replaced by "$\leq$" or "$\geq$", then $M.$ is called an $\mathbb{F}$ super-(or sub-) martingale.

■

Theorem 6.3 — **Semi-martingale Representation due to Doob.** An $\mathbb{F}$-adapted process $X.$ with integrable $X_n, \forall n \geq 0$ is unquely decomposed in the form of (6.2) almost surely by an $\mathbb{F}$-predictable process $\{Y_n : \ n \geq 0\}$ and an $\mathbb{F}$-martingale $\{M_n : \ n \geq 0\}$ with $M_n = 0$. The stochastic process $X.$ having this decomposition is called an $\mathbb{F}$-semimartingale.

Proof. Assume on the contrary that there is another decomposition $X_n = Y_n' + M_n'$, then

$$Y_{n+1} - Y_{n+1}' = \mathbb{E}[Y_{n+1} - Y_{n+1}' \mid \mathcal{F}_n] = \mathbb{E}[M_{n+1}' - M_{n+1} \mid \mathcal{F}_n] = M_n' - M_n.$$

Since $M_0' - M_0 = 0$, we imply $Y_1 - Y_n' = 0$, which further implies $M_1 = M_1'$. Inductively, the uniqueness is proved.

■

Corollary 6.1 If $X.$ is an $\mathbb{F}$-subimartingale, then the $\mathbb{F}$-predictable component Y_n is non-decreasing in n a.s. for the semi-martingale representation $X_n = Y_n + M_n$.

Proof. By the definition of submartingale,

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n = Y_n + M_n.$$

Substituting X_{n+1} with $Y_{n+1} + M_{n+1}$ gives

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = Y_{n+1} + M_n.$$

Thus $Y_{n+1} - Y_n \geq 0$. ■

In (6.2), $\{E_n : n \geq 1\}$ may be considered as a sequence of white noises, and $\mathbb{F}$-martingale $M.$ may be interpreted as an accumulated white noise. We need M_n to be a martingale because:

1. It uniquely determines the decomposition;
2. It extracts the noise component from ongoing history;
3. It is analytically tractable, enabling us to study various asymptotic problems.

Any discrete-time stochastic process can be viewed as a semi-martingale, but its martingale component may not be easily computed. Here gives several simple cases:

■ **Example 6.5** • For simple random walk,

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n = \mathbb{E}[U_{n+1} \mid \mathcal{F}_n] = a$$

- Assume that $\mathbb{E}[U_n^2] = b < \infty$, then

$$\mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] - X_n^2 = 2aX_n + b.$$

Chapter 7

Week 7

7.1. Monday

7.1.1. Stopping Time and Martingale

Definition 7.1 [Discrete Time Stochastic Basis] For probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and filtration $\mathbb{F}$ on $(\Omega, \mathcal{F})$, the set $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a **stochastic basis** (or a discrete-time filtered probability space).

Consider to bet money by fair coin tossing. We are interested when to stop this coin tossing to maximize the assert, based on all outcomes up to the time. We define this time as a random variable as the following.

Definition 7.2 [Stopping Time] Let $\mathbb{F} \triangleq \{\mathcal{F}_n : n \geq 0\}$ be a filtration. A random variable τ valued in $\overline{\mathbb{Z}}_+ \equiv \mathbb{Z}_+ \cup \{\infty\}$ is called a stopping time if $\{\tau \leq n\} \in \mathcal{F}_n, \forall n \geq 0$.

R $\{\tau \leq n\} \in \mathcal{F}_n, \forall n \geq 0$ if and only if $\{\tau = n\} \in \mathcal{F}_n, \forall n \geq 0$

Definition 7.3 [Stopped σ-field] For filtration $\mathbb{F} \triangleq \{\mathcal{F}_n : n \geq 0\}$ and stopping time τ,

$$\mathcal{F}_\tau \triangleq \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n, \forall n \geq 0\}$$

is called a σ-field stopped by τ.

■ **Example 7.1** Let $X_\cdot \triangleq \{X_n: \ n \geq 0\}$ with $X_0 = 0$ be a real valued process. Let $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ and $\mathbb{F} \triangleq \{\mathcal{F}_n: \ n \geq 0\}$.

- Define τ_a as
$$\tau_a = \inf\{\ell \geq 0: \ X_\ell \geq a\}, \quad a > 0$$
and $\inf \emptyset = +\infty$. Then τ_a is a stopping time since
$$\tau_a = n \Longleftrightarrow \max_{0 \leq \ell \leq n-1} X_\ell < a \leq X_n.$$

- Define $\tau_{\max}$ as
$$\tau_{\max} = \inf\{\ell \geq 0: \ \max_{k \geq 0} X_k = X_\ell\}.$$
Then $\tau_{\max}$ is not a stopping time since
$$\tau_{\max} = n \Longleftrightarrow X_\ell < X_n, \ell \leq n-1, X_\ell \leq X_n, \ell \geq n+1.$$

■

(R) If τ, η are stopping times, then
$$\tau + \eta, \quad \tau \wedge \eta, \quad \tau \vee \eta$$
are stopping times, but $0 \vee (\tau - \eta)$ may not be.

Proposition 7.1 Let $X_\cdot \triangleq \{X_n: \ n \geq 0\}$ be a martingale (sub or super-martingale). If τ is a stopping time, then $\{X_{\tau \wedge n}: \ n \geq 0\}$ is an $\mathbb{F}$-martingale (sub or super-martingale).

Proof. We prove the case where $X_\cdot$ is a sub-martingale.

- Write $X_{\tau \wedge n}$ as
$$X_{\tau \wedge n} = X_0 + \sum_{\ell=1}^{\tau \wedge n} (X_\ell - X_{\ell-1}) = X_0 + \sum_{\ell=1}^{n} 1(\ell \leq \tau)(X_\ell - X_{\ell-1})$$

Since $1(\ell \le \tau)$ is $\mathcal{F}_n$-measurable for $\ell = 1,\dots,n$, we imply that $X_{\tau\wedge n}$ is also $\mathcal{F}_n$-measurable.

- Moreover,

$$X_{\tau\wedge(n+1)} - X_{\tau\wedge n} = 1(n+1 \le \tau)(X_{n+1} - X_n).$$

which implies that

$$\mathbb{E}[X_{\tau\wedge(n+1)} \mid \mathcal{F}_n] - X_{\tau\wedge n} = 1(n+1 \le \tau)(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n) \ge 0.$$

■

We can define a stopped process by a stopping time:

$$X^\tau_\cdot \triangleq \{X^\tau_n : n \ge 0\}, \quad \text{where } X^\tau_n = X_{\tau\wedge n}.$$

Then $X^\tau_\cdot$ is called a τ-stopped process. It is adapted to $\mathbb{F}$ as well as to $\mathbb{F}^\tau \triangleq \{\mathcal{F}_{\tau\wedge n} : n \ge 0\}$, if X is adapted to $\mathbb{F}$.

For a sequence of finite stopping times $\{\tau_n : n \ge 1\}$ such that $\tau_n \to \infty$ as $n \to \infty$, a sequnece of stopped processes $X^{\tau_n}_\cdot \triangleq \{X^{\tau_n}_\ell : \ell \ge 0\}$ for $n = 1,2,\dots$ is called a **localization** of $X_\cdot$. Since $X^{\tau_n}_\ell \to X_\ell$ a.s. as $n \to \infty$, it may be possible to consider $X^{\tau_n}_\ell$ instead of X_ℓ. This localization does not much earn for a discrete-time process, but it does for a continuous-time process as we will see.

7.1.2. Optional Sampling Theorem

Theorem 7.1 — **Optional Inequality 1.** Suppose that $X_\cdot$ is a submartingale and τ is a stopping time taking values in $\{0,1,\dots,m\}$ for a positive integer m, then we have

$$\mathbb{E}[X_0] \le \mathbb{E}[X_\tau] \le \mathbb{E}[X_m]$$

Proof. By proposition 7.1, $\{X_{\tau\wedge n} : n \geq 0\}$ is a submartingale, which implies

$$X_{\tau\wedge 0} \leq \mathbb{E}[X_{\tau\wedge m} \mid \mathcal{F}_0] \implies \mathbb{E}[X_0] = \mathbb{E}[X_{\tau\wedge 0}] \leq \mathbb{E}[X_{\tau\wedge m}] = \mathbb{E}[X_\tau].$$

Moreover, $\{X_n - X_{\tau\wedge n} : n \geq 0\}$ is a submartingale:

$$\begin{aligned}
\mathbb{E}[X_n - X_{\tau\wedge n} \mid \mathcal{F}_{n-1}] &= \mathbb{E}[X_n - (X_\tau 1(\tau \leq n-1) + X_n 1(\tau \geq n)) \mid \mathcal{F}_{n-1}] \\
&= \mathbb{E}[X_n 1(\tau \leq n-1) - X_\tau 1(\tau \leq n-1) \mid \mathcal{F}_{n-1}] \\
&= \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] 1(\tau \leq n-1) - X_\tau 1(\tau \leq n-1) \\
&\geq (X_{n-1} - X_\tau) 1(\tau \leq n-1) \\
&= X_{n-1} - X_{\tau\wedge(n-1)}
\end{aligned}$$

Therefore,

$$\mathbb{E}[X_m - X_\tau] = \mathbb{E}[X_m - X_{\tau\wedge m}] \geq \mathbb{E}[X_0 - X_{\tau\wedge 0}] = 0.$$

■

Corollary 7.1 Suppose that $X.$ is a submartingale and τ, η are stopping times such that $\tau \leq \eta \leq m$ for a constant time m, then we have

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X_\eta] \leq \mathbb{E}[X_m]$$

Proof. Take $Y_n = X_{\eta\wedge n}$, then Y_n is a sub-martingale. Applying Theorem 7.1 into the stopping time τ and $Y.$ gives

$$\mathbb{E}[Y_\tau] \leq \mathbb{E}[Y_m] = \mathbb{E}[X_\eta]$$

In other words, $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_\eta]$. ■

R In Theorem 7.1 and Corollary 7.1, "$\leq$" is replaced by "$=$" or "$\geq$" if $X.$ is a martingale or super-martingale.

Theorem 7.2 — **Optional Sampling Theorem.** If $X.$ is a sub-martingale and τ, η are stopping times such that $\tau \le \eta \le m$ for a constant time m, then

$$X_\tau \le \mathbb{E}[X_\eta \mid \mathcal{F}_\tau]$$

almost surely. Here $\ge$ is replaced by $=$ if $X.$ is a martingale.

Proof. For any $A \in \mathcal{F}_\tau$, define

$$\zeta = \tau 1(A) + \eta 1(A^c).$$

Then ζ is a stopping time with $\tau \le \zeta \le \eta \le m$. Applying Corollary 7.1 gives

$$\mathbb{E}[X_\tau 1_A] + \mathbb{E}[X_\eta 1_{A^c}] = \mathbb{E}[\zeta] \le \mathbb{E}[X_\eta].$$

As a result,

$$\mathbb{E}[X_\tau 1_A] \le \mathbb{E}[X_\eta 1_A] = \mathbb{E}[\mathbb{E}[X_\eta \mid \mathcal{F}_\tau] 1_A], \quad A \in \mathcal{F}_\tau.$$

This means that $\mathbb{P}(X_\tau \le \mathbb{E}[X_\eta \mid \mathcal{F}_\tau]) = 1$. ■

▪ **Example 7.2** [Bet by Coin Tossing with initial asset a] Let U_n's be i.i.d. with $\mathbb{P}(U_n = 1) = \mathbb{P}(U_n = -1) = \frac{1}{2}$. Define $X_0 = a$ and $X_n = X_0 + U_1 + \cdots + U_n$ for $n \ge 1$.

Let $M_n = X_n - X_0$, then $M. = \{M_n : \ n \ge 0\}$ is a martingale for the filtration $\mathbb{F}^X$ since $\mathbb{E}[U_n] = 0$. We are interetsed in the probability that X_n hits b before to hit 0. Define stopping times

$$\tau_0 = \inf\{n \ge 1 : \ X_n = 0\}, \quad \tau_b = \inf\{n \ge 1 : \ X_n = b\}.$$

Then we aim to compute $\mathbb{P}(\tau_0 < \tau_b)$. Apply Theorem 7.1 for $\tau_n := \tau_0 \wedge \tau_b \wedge n$ gives

$$\mathbb{E}[M_{\tau_n}] = \mathbb{E}[M_0] = 0.$$

Or equivalently,

$$(b-a)\mathbb{P}(\tau_b < \tau_0 \wedge n) - a\mathbb{P}(\tau_0 < \tau_b \wedge n) + \mathbb{E}[M_n 1(n \le \tau_b \wedge \tau_0)] = 0$$

Considering that $\liminf M_n = -\infty, \limsup M_n = \infty$, by the law of iterated logarithms, we can assert that $\mathbb{P}(n \le \tau_b \wedge \tau_0) \to 0$. Together with the fact that $|M_n| \le b$ under the event $\{n \le \tau_b \wedge \tau_0\}$, we imply

$$(b-a)\mathbb{P}(\tau_b < \tau_0 \wedge n) - a\mathbb{P}(\tau_0 < \tau_b \wedge n) = 0.$$

On the other hand, $\mathbb{P}(\tau_b \vee \tau_0 < \infty) = 1$ gives

$$\mathbb{P}(\tau_b < \tau_0) + \mathbb{P}(\tau_0 < \tau_b) = 1.$$

Thus $\mathbb{P}(\tau_b < \tau_0) = b/a$. ■

This deviation has the following problems:

1. It requires the extra result: the law of iterated logarithms;
2. It cannot be applied for an unfair coin such that $\mathbb{P}(U_n = 1) \neq \frac{1}{2}$.

We will overcome these problems by considering another type of martingale, in the next week.

7.1.3. Strong and Stopped Markov Processes

Recall the definition of discrete Markov process:

1. $X_\cdot$ is $\mathbb{F}$-adapted;
2. $\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] = P[f(X_n)]$

For a stopping time τ, we want to ask

(a) Does the second condition still holds for $n = \tau$?

(b) Is the stopped prcess $X_\cdot^\tau$ again Markov?

We first introduce a notion for (a):

Definition 7.4 [Strong Markov Property] A discrete-time Markov process $X_\cdot$ is called **strong Markov** if

$$\mathbb{E}[f(X_{\tau+1}) \mid \mathcal{F}_\tau] = P[f(X_\tau)], \quad f \in D_b(S)$$

for all finite stopping time τ (i.e., $\mathbb{P}(\tau < \infty) = 1$).

Theorem 7.3 A discrete-time Markov process $X_\cdot$ is strong Markov.

Proof. Let τ be an arbitrary finite stopping time. Then the strong Markov property holds iff

$$\mathbb{E}[f(X_{\tau+1})1_A] = \mathbb{E}[P[f(X_\tau)]1_A], \quad A \in \mathcal{F}_\tau.$$

Or equivalently, for any $n \geq 1$,

$$\mathbb{E}[f(X_{\tau+1})1_{A\cap\{\tau=n\}}] = \mathbb{E}[P[f(X_\tau)]1_{A\cap\{\tau=n\}}], \quad A \in \mathcal{F}_\tau.$$

The RHS is equivalent to

$$\mathbb{E}[\mathbb{E}[f(X_{n+1} \mid X_n)]1_{A\cap\{\tau=n\}}] = \mathbb{E}[\mathbb{E}[f(X_{n+1} \mid \mathcal{F}_n)]1_{A\cap\{\tau=n\}}] = \mathbb{E}[f(X_{n+1})1_{A\cap\{\tau=n\}}].$$

The proof is completed. ■

In some applications, we start to observe a discrete-time Markov process $X_\cdot$ from time τ, which is a finite stopping time. In this case, is $X_{\tau+\cdot} = \{X_{\tau+n} : n \geq 0\}$ again Markov? It must be. To see this, we need to show that

$$\mathbb{E}[f(X_{\tau+n+1}) \mid \mathcal{F}_{\tau+n}] = \mathbb{E}[f(X_{\tau+n+1}) \mid X_{\tau+n}]$$

This is true since $\tau + n$ is a stopping time.

Proposition 7.2 For a discrete-time Markov process $X_\cdot$ and stopping time τ, the stopped process $X_\cdot^\tau$ is $\mathbb{F}^\tau$-Markov, where $\mathbb{F}^\tau = \{\mathcal{F}_{\tau\wedge n} : n \geq 0\}$.

Proof. It suffices to show that

$$\mathbb{E}[f(X_{\tau\wedge(n+1)}) \mid \mathcal{F}_{\tau\wedge n}] = \mathbb{E}[f(X_{\tau\wedge(n+1)}) \mid X_{\tau\wedge n}]$$

By definition of conditional expectation, it suffices to show that

$$\mathbb{E}[f(X_{\tau\wedge(n+1)})1_A] = \mathbb{E}[\mathbb{E}[f(X_{\tau\wedge(n+1)}) \mid X_{\tau\wedge n}]1_A], \forall A \in \mathcal{F}_{\tau\wedge n}. \tag{7.1}$$

Note that

$$\sigma(X_{\tau\wedge n}) = \sigma\Big(\cup_{\ell=0}^{n} \left(\{\tau=\ell\} \cap \sigma(X_\ell)\right) \cup \left(\{\tau>n\} \cap \sigma(X_n)\right)\Big)$$

Thus $\{\tau \le n\}$ and $\{\tau > n\}$ are in $\sigma(X_{\tau\wedge n})$. Then we can simplify the RHS of (7.1):

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[f(X_{\tau\wedge(n+1)}) \mid X_{\tau\wedge n}]1_A] &= \mathbb{E}[\mathbb{E}[f(X_{\tau\wedge(n+1)})(1\{\tau\le n\}+1\{\tau>n\}) \mid X_{\tau\wedge n}]1_A] \\
&= \mathbb{E}[f(X_{\tau\wedge(n)})1\{\tau\le n\}1_A] + \mathbb{E}[\mathbb{E}[f(X_{\tau\wedge(n+1)})1\{\tau>n\} \mid X_{\tau\wedge n}]1_A] \\
&= \mathbb{E}[f(X_{\tau\wedge(n)})1\{\tau\le n\}1_A] + \mathbb{E}[\mathbb{E}[f(X_{\tau\wedge(n+1)}) \mid \mathcal{F}_{\tau\wedge n}]1\{A\cap\{\tau>n\}\}] \\
&= \mathbb{E}[f(X_{\tau\wedge(n)})1\{\tau\le n\}1_A] + \mathbb{E}[f(X_{\tau\wedge(n+1)})1\{A\cap\{\tau>n\}\}] \\
&= \mathbb{E}[f(X_{\tau\wedge(n+1)})1_A]
\end{aligned}$$

■

R Note that the stopped process $X_\cdot^\tau$ does not have time-homogeneous transition in general. For instance, let $\tau \equiv k$, then $X_n^\tau = X_{k\wedge n}$, i.e., the transition may change before or after time k.

Chapter 8

Week 8

8.1. Monday

8.1.1. Convex Function

Convex functions are very useful to geometrically understand analytical results and to derive various inequalities.

Definition 8.1 [Convex] Function $f : G \to \mathbb{R}$ is said to be convex on the open interval $G \subseteq \mathbb{R}$ if for $p, q > 0$ satisfying $p + q = 1$,

$$f(px + qy) \le pf(x) + qf(y), \quad \forall x, y \in G.$$

Proposition 8.1 — **Linear Envelop of a Convex Function.** f is convex on G if and only if there exists $c_x \in \mathbb{R}$ such that for any $x \in G$,

$$c_x(y - x) \le f(y) - f(x), \quad \forall y \in G.$$

Further, for convex function f on G, a) f is continuous, b) c_x is non-decreasing in $x \in G$, and c) $c_x = f'(x)$ if f is differentiale at x.

Proposition 8.2 — **Jensen's Inequality.** For a random variable X, a sub σ-field $\mathcal{G}$, and convex function f on $\mathbb{R}$, if $\mathbb{E}[|X|]$ and $\mathbb{E}[|f(X)|]$ are finite, then

$$f(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[f(X) \mid \mathcal{G}]$$

Proof. Applying Proposition 8.1 with $y = X, x = \mathbb{E}[X \mid \mathcal{G}]$ gives

$$f(X) - f(\mathbb{E}[X \mid \mathcal{G}]) \geq c_{\mathbb{E}[X|\mathcal{G}]}(X - \mathbb{E}[X \mid \mathcal{G}]).$$

Taking the conditional expectations of both sides given $\mathcal{G}$, we have

$$\mathbb{E}\left(f(X) - f(\mathbb{E}[X \mid \mathcal{G}])\middle|\mathcal{G}\right) \geq c_{\mathbb{E}[X|\mathcal{G}]}\mathbb{E}[X - \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}] = 0.$$

■

Theorem 8.1 Let f be a convex function on $\mathbb{R}$. If $X_{\cdot}$ is $\mathbb{F}$-martingale and if $\mathbb{E}[|X|]$ and $\mathbb{E}[|f(X_n)|]$ are finite for all $n \geq 0$, then $\{f(X_n) : \ n \geq 0\}$ is an $\mathbb{F}$-submartingale.

Proof. Applying proposition 8.2 with $X = X_n, \mathcal{G} = \mathcal{F}_{n-1}$ gives

$$\mathbb{E}[f(X_n) \mid \mathcal{F}_{n-1}] - f(\mathbb{E}[X_n \mid \mathcal{F}_{n-1}]) \geq 0.$$

Therefore,

$$\mathbb{E}[f(X_n) \mid \mathcal{F}_{n-1}] \geq f(X_{n-1}).$$

■

Chapter 9

Week 9

9.1. Monday

9.1.1. Classification of States

Motivation. Let $X.$ denote a Markov chain. We are interested in the transition probability $p_{i,j}^{(n)}$ for $i,j \in S$. We want to ask the following questions:

1. Is the limit $p_{i,j}^{(n)}$ exists as $n \to \infty$?
2. If so, how to compute it?

Before answering this question, we classify states in S according to:

- Go to state j from state i with positive probability?
- Is it any period to return to the starting state?
- With probability 1 to return to the same state? The expectation time for return is finite?

Definition 9.1
- We say i,j commute to each other if $i \Leftrightarrow j$;
- We say C is irreducible if $i \Leftrightarrow j$ for any $i,j \in C$;
- We say C is closed if $p_{i,j} = 0$ for $\forall i \in C, j \in S \setminus C$.

The notion $\Rightarrow$ is an equivalence relation on C: when $i \Rightarrow j$ and $j \Rightarrow k$, i.e., $p_{i,j}^{(\ell)} > 0, p_{j,k}^{(m)} >$

0, then

$$p_{i,k}^{(\ell+m)} = \sum_{j' \in S} p_{i,j'}^{(\ell)} p_{j',k}^{(m)} \geq p_{i,j}^{(\ell)} p_{j,k}^{(m)} > 0,$$

which means that $i \Rightarrow k$.

Proposition 9.1 The state space S of a Markov chain can be partitioned into mutually exclusive irreducible closed sets and the other set which does not contain any irreducible closed set.

Proof. Construction of such set: Let $C(i) = \{j \in S : i \Leftrightarrow j\}$ if it is closed, otherwise $C(i) = \emptyset$. Finally, construct $T = S \setminus (\cup_{i \in S} C(i))$. ■

This lemma suggests that we only need to consider an irreducible closed set to ask the question at the beginning.

Definition 9.2 [Stopping Time]

- Define the first time to get j from i as $\tau_j = \inf\{n \geq 1 : X_n = j\}$, where $\tau_j = \infty$ if $X_n \neq j, \forall n$;
- Define the hitting probability $f_{i,j}^{(n)} = \mathbb{P}(\tau_j = n \mid X_0 = i)$.

Proposition 9.2 — **First Time Decomposition.** When $n = 0$, $p_{i,j}^{(0)} = 1(i = j)$. When $n \geq 1$,

$$p_{i,j}^{(n)} = \sum_{\ell=1}^{n} f_{i,j}^{(\ell)} p_{j,j}^{(n-\ell)}.$$

Proof.

$$\begin{aligned}
\mathbb{P}(X_n = j \mid X_0 = i) &= \sum_{\ell=1}^{n} \mathbb{P}(X_n = j, \tau_j = \ell \mid X_0 = i) \\
&= \sum_{\ell=1}^{n} \mathbb{P}(\tau_j = \ell \mid X_0 = i)\mathbb{P}(X_n = j \mid \tau_j = \ell, X_0 = i) \\
&= \sum_{\ell=1}^{n} \mathbb{P}(\tau_j = \ell \mid X_0 = i)\mathbb{P}(X_n = j \mid X_\ell = j) = \sum_{\ell=1}^{n} f_{i,j}^{(\ell)} p_{j,j}^{(n-\ell)}.
\end{aligned}$$

■

Definition 9.3 [Period] Let $\mathbb{N} = \{k \in \mathbb{Z}, k \geq 1\}$, For $i \in S$, define the period of i as

$$d(i) = \text{GCD}\{n : \ p_{i,i}^{(n)} > 0\}$$

We denote $d(i) = \infty$ if $p_{i,i}^{(n)} = 0, \forall n$.

Proposition 9.3 For $i, j \in S$ with $i \neq j$, $i \Leftrightarrow j$ implies $d(i) = d(j) < \infty$.

Proof. It suffices to show that $d(j)$ is a multiple of $d(i)$. Since $i \Leftrightarrow j$, there exists ℓ, m such that $p_{i,j}^{(\ell)} > 0, p_{j,i}^{(m)} > 0$. For n such that $p_{j,j}^{(n)} > 0$,

$$p_{i,i}^{(\ell+n+m)} \geq p_{i,j}^{(\ell)} p_{j,j}^{(n)} p_{j,i}^{(m)} > 0,$$

which means that $\ell + n + m$ is a multiple of $d(i)$. Since $\ell + m$ does so, n is a multiple of $d(i)$. ■

Definition 9.4 [Aperiodic] An irreducible Markov chain is said to be aperiodic if some state has period 1.

The following gives an alternative for finding the period of a state.

Proposition 9.4 For $j \in S$, let $f(j) = \text{GCD}\{n : \ f_{j,j}^{(n)} > 0\}$. Then $f(j) = d(j)$.

Proof. Note that $f(j)$ is a multiple of $d(j)$ since $p_{j,j}^{(n)} > 0$. Conversely, pick any n so that $p_{j,j}^{(n)} > 0$. Then there exists ℓ_1 so that $f_{j,j}^{(\ell_1)} p_{j,j}^{(n-\ell_1)} > 0$. Inductively, $n = \ell_1 + \cdots + \ell_k$, i.e., n is a multiple of $f(j)$. Hence, $d(j)$ is a multiple of $f(j)$. ■

Definition 9.5 [Return-Time Classes] Define the following notation for $i, j \in S$:

$$f_{i,j}^{(*)} = \mathbb{P}(\tau_j < \infty \mid X_0 = i), \quad \mu_{i,j} = \begin{cases} \mathbb{E}[\tau_j \mid X_0 = i], & \text{if } f_{i,j}^{(*)} = 1 \\ \infty, & \text{if } f_{i,j}^{(*)} < 1 \end{cases}$$

1. The state i is transient if $f_{i,j}^{(*)} < 1$;

2. The state i is recurrent if $f_{i,j}^{(*)} = 1$;
3. The state is positive recurrent if $f_{i,j}^{(*)} = 1$ and $\mu_{i,i} < \infty$;
4. The state is null recurrent if $f_{i,j}^{(*)} = 1$ and $\mu_{i,i} = \infty$;

Define the following series to characterize the relation between $\{f_{i,j}^{(n)}\}$ and $\{p_{i,j}^{(n)}\}$:

$$F_{i,j}(z) = \sum_{n=1}^{\infty} f_{i,j}^{(n)} z^n, \qquad U_{i,j}(z) = \sum_{n=0}^{\infty} p_{i,j}^{(n)} z^n.$$

Proposition 9.5 For $i, j \in S$ and $0 < z < 1$,

$$U_{i,j}(z) - 1(i = j) = F_{i,j}(z) U_{j,j}(z).$$

Moreover, $\lim_{z \to 1-} F_{i,i}(z) = f_{i,i}^{(*)}$ and $\lim_{z \to 1-} \frac{1 - F_{i,i}(z)}{1 - z} = \mu_{i,i}$ if $f_{i,i}^{(*)} = 1$.

Proof.

$$U_{i,j}(z) - 1(i = j) = \sum_{n=1}^{\infty} z^n \sum_{\ell=1}^{n} f_{i,j}^{(\ell)} p_{j,j}^{(n-\ell)} = F_{i,j}(z) U_{j,j}(z).$$

■

Theorem 9.1 If $i \Leftrightarrow j$, then i and j are in the same return-time class.

Proof. The key is to show the identity

$$(1 - F_{j,j}(z)) \geq c_{i,j}^{(\ell,m)}(z)(1 - F_{i,i}(z)) \geq 0.$$

■

Proposition 9.6

$$f_{i,i}^{(*)} = p_{i,i} + \sum_{j \in S \setminus i} p_{i,j} f_{j,i}^{(*)}.$$

Proof.

$$\mathbb{P}(\tau_i < \infty \mid X_0 = i) = p_{i,i} + \sum_{j \in S \setminus i} \mathbb{P}(\tau_i < \infty \mid X_1 = j)\mathbb{P}(X_1 = j \mid X_0 = i).$$

■

If there is some j so that $p_{i,j} > 0$ but $f_{j,i}^{(*)} = 0$, then immediately $f_{i,i}^{(*)} < 1$.

9.1.2. Renewal Equation

For a probability distribution $\{f_n : n \geq 1\}$, define the renewal equation by $u_0 = 1$,

$$u_n = \sum_{\ell=1}^{n} f_\ell u_{n-\ell}.$$

Specifically, we take $f_\ell = f_{j,j}^{(\ell)}$. Then $\{f_\ell\}$ is a distribution, and $p_{j,j}^{(n)}$ satisfies the renewal equation. The existence of the limit u_n is answered by the following theorem:

Theorem 9.2 For the renewal equation, if $\mathrm{GCD}\{n \geq 1 : f_n > 0\} = 1$, then the solution $\{u_n : n \geq 1\}$ admits the limit

$$\lim_{n \to \infty} u_n = \frac{1}{\mu}$$

with $\mu = \sum_{\ell=1}^{\infty} \ell f_\ell$.

9.1.3. Limit of State Distribution

Theorem 9.3 If state j is transient, then

$$\lim_{n \to \infty} p_{i,j}^{(n)} = 0, \quad \forall i \in S.$$

Proof. First consider $i = j$. Since j is transient, $F_{j,j}(1) < 1$, and

$$\sum_{n=0}^{\infty} p_{j,j}^{(n)} = \lim_{z \to 1-} U_{j,j}(z) = \frac{1}{1 - F_{j,j}(1)} < \infty.$$

Thus $p_{j,j}^{(n)} \to 0$.

For $i \neq j$,

$$p_{i,j}^{(n)} = \sum_{\ell=1}^{\infty} f_{i,j}^{(\ell)} 1(\ell \leq n) p_{j,j}^{(n-\ell)}$$

Taking $n \to \infty$ gives the desired result. ■

Theorem 9.4 If j is aperiodic and recurrent, then

$$\lim_{n\to\infty} p_{j,j}^{(n)} = \frac{1}{\mu_{j,j}}, \quad \lim_{n\to\infty} p_{i,j}^{(n)} = f_{i,j}^{(*)} \frac{1}{\mu_{j,j}},$$

where $\mu_{j,j} = \mathbb{E}[\tau_j \mid X_0 = j]$.

Proof. The first limit is by renewal theorem:

$$\mu_{j,j} = \sum_{j=1}^{\infty} \ell f_{j,j}^{(\ell)} = \mathbb{E}[\tau_j \mid X_0 = j].$$

The second limit is based on the first-time decomposition:

$$p_{i,j}^{(n)} = \sum_{\ell=1}^{n} f_{i,j}^{(\ell)} p_{j,j}^{(n-\ell)}$$

■

Theorem 9.5 If j is aperiodic and has period d, then

$$\lim_{n\to\infty} p_{i,j}^{(nd)} = f_{i,j}^{(*)} \frac{d}{\mu_{j,j}}$$

Proof. Construct the 1-step transition Markov process $Y_n = X_{nd}$, then

$$\lim_{n\to\infty} p_{j,j}^{(nd)} = \frac{1}{\mathbb{E}[\tau_j^{(d)} \mid X_0 = j]}$$

■

Theorem 9.6 If the Markov chain is aperiodic and irreducible, then

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \frac{1}{\mu_{j,j}}.$$

Proof. It suffices to show $f_{i,j}^{(*)} = 1$. ■

Remaining problems

- What is the role of $p_{i,j}^{(n)}$ for large n in application?
- In the positive recurrent case, the distribution X_n weakly converges to a distribution?
- Is there any sufficient condition for $\mu_j < \infty$?
- Is there any simpler way to compute μ_j?

Chapter 10

Week 10

10.1. Monday

10.1.1. Stationary Meansure and Distribution

Definition 10.1 [stationary measure]

- We call $\{\pi(i): i \in S\}$ a measure on S if $\pi(i) \geq 0$ for any $i \in S$.
- Furthermore, it is called a stationary measure if

$$\pi(j) = \sum_{i \in S} \pi(i) p_{i,j}, \quad j \in S,$$

  and the equation above is called a stationary equation.
- In particular, $\{\pi(i): i \in S\}$ is called a stationary equation if

$$\sum_{i \in S} \pi(i) = 1.$$

■

■ **Example 10.1** Transition matrix P is said to be doubly stochastic if $p_{i,j} = p_{j,i}$. A Markov chain with doubly stochastic transitions always has the stationary measure : $\pi(i) \equiv C$. ■

R Inductively, we can show that $\pi(k) = \sum_{i \in S} \pi(i) p_{i,k}^{(n)}$:

$$\begin{aligned}\pi(k) &= \sum_{i \in S} \pi(i) p_{i,k} = \sum_{i \in S} \sum_{j \in S} \pi(j) p_{j,i} p_{i,k} \\ &= \sum_{j \in S} \pi(j) \left[\sum_{i \in S} p_{j,i} p_{i,k} \right] = \sum_{j \in S} \pi(j) \pi_{j,k}^{(2)}.\end{aligned}$$

In other words, if the initial measure is π, then this measure is invariant of time. That's why π is called a stationary measure.

Theorem 10.1 For a Markov chain with state space S and transition probabilities $\{p_{i,j}\}$, if it is irreducible and aperiodic, then the following two conditions are equivalent:

1. It has a stationary equation;
2. It is positive recurrent.

If either one of the conditions holds, then the stationary distribution can be uniquely obtained as

$$\lim_{n \to \infty} p_{i,j}^{(n)} = \pi(j) = \frac{1}{\mu_{j,j}}.$$

The aperiodic condition is not essential for the equivalence, but for the uniqueness of stationary equation.

Proof.

- Assume 1) holds but 2) does not. Then the Markov chain is transient or null recurrent by the irreducibility. By Theorem 9.3 and Theorem 9.4, $p_{i,j}^{(n)} \to 0$ as $n \to \infty$. Then by the dominated convergence theorem,

$$\pi(j) = \lim_{n \to \infty} \sum_{i \in S} \pi(i) p_{i,j}^{(n)} = \sum_{i \in S} \pi(i) \lim_{n \to \infty} p_{i,j}^{(n)} = 0.$$

  This contradicts to 1).

- Assume 2) holds, then by Theorem 9.4, $p_{i,j}^{(n)} \to \frac{1}{\mu_{j,j}}$. We construct the stationary measure as $\pi(j) = \frac{1}{\mu_{j,j}}$. Then by Fatou's lemma,

$$\sum_{j\in S}\frac{1}{\mu_{j,j}} = \sum_{j\in S}\lim_{n\to\infty} p_{i,j}^{(n)} \le \liminf_{n\to\infty}\sum_{j\in S} p_{i,j}^{(n)} = 1.$$

Moreover, by $p_{i,k}^{(n+1)} = \sum_{j\in S} p_{i,j}^{(n)} p_{j,k}$ we have

$$\pi(k) = \liminf_{n\to\infty} p_{i,k}^{(n+1)} = \liminf_{n\to\infty}\sum_{j\in S} p_{i,j}^{(n)} p_{j,k} \ge \sum_{j\in S}\liminf_{n\to\infty} p_{i,j}^{(n)} p_{j,k} = \sum_{j\in S}\pi(j)p_{j,k}.$$

Summing up both sides for $k\in S$,

$$\sum_{k\in S}\pi(k) \ge \sum_{k\in S}\sum_{j\in S}\pi(j)p_{j,k} = \sum_{j\in S}\pi(j).$$

Hence the inequaality must be equality, i.e., the statarion equation holds.

- Finally, we show that the stationary measure π is unique. Let ν be an arbitrary stationary distribution, then $\nu(j) = \sum_{i\in S}\nu(i)p_{i,j}^{(n)}$. Taking $n\to\infty$ gives

$$\nu(j) = \sum_{i\in S}\nu(i)\lim_{n\to\infty} p_{i,j}^{(n)} = \sum_{i\in S}\nu(i)\frac{1}{\mu_{j,j}} = \frac{1}{\mu_{j,j}}.$$

The proof is completed.

■

■ **Example 10.2** The stationary equation for birth and death process is

$$\begin{cases}\pi(0) = q\pi(0) + q\pi(1)\\ \pi(i) = p\pi(i-1) + q\pi(i+1), \quad i\ge 1.\end{cases}$$

From the second equation, $q\pi(i) - p\pi(i-1) = q\pi(i+1) - p\pi(i)$, which implies

$$q\pi(i) - p\pi(i-1) = q\pi(1) - p\pi(0) = 0.$$

Take $\rho = p/q$, then

$$\pi(j) = \rho\pi(j-1) = \cdots = \rho^j \pi(0).$$

Thus π is a stationary distribution iff $\rho < 1$. In this case, $\pi(j) = (1-\rho)\rho^j$. ■

Then we study the stationary distribution for null recurrent case:

Theorem 10.2 Assume that $\{X_n\}$ is irreducible and recurrent.

1. For each fixed $k \in S$, define

$$\nu_k(j) = \mathbb{E}\left(\sum_{\ell=1}^{\infty} 1(X_\ell = j, \ell \le \tau_k) \middle| X_0 = k\right), \quad j \in S,$$

   then ν_k is a stationary measure.

2. A stationary measure is unique up to constant multiple of ν_k
3. If it is positive recurrent, then the stationary distribution is defined as

$$\pi(j) = \frac{1}{\mu_{k,k}} \nu_k(j).$$

R $\sum_{\ell=1}^{\infty} 1(X_\ell = j, \ell \le \tau_k)$ denotes the number of visits to state j before returning to k. The collection of its expectations is called an **occupation measure** before returning to k.

Proof. 1. Assume that $\nu_k(j) < \infty$, then we start to show the stationary equation:

$$\begin{aligned}
\nu_k(j) &= \mathbb{E}\left(\sum_{\ell=1}^{\infty} 1(X_\ell = j, \ell \le \tau_k) \middle| X_0 = k\right) \\
&= p_{k,j} + \sum_{\ell=2}^{\infty} \sum_{i \in S \setminus k} \mathbb{P}\left(X_\ell = j, \ell \le \tau_k, X_{\ell-1} = i | X_0 = k\right) \\
&= p_{k,j} + \sum_{\ell=2}^{\infty} \sum_{i \in S \setminus k} \mathbb{P}\left(X_{\ell-1} = i, \ell \le \tau_k | X_0 = k\right) \\
&\qquad \times \mathbb{P}\left(X_\ell = j | X_0 = k, X_{\ell-1} = i, \ell \le \tau_k\right) \\
&= p_{k,j} + \sum_{\ell=1}^{\infty} \sum_{i \in S \setminus k} \mathbb{P}\left(X_\ell = i, \ell \le \tau_k - 1 | X_0 = k\right) p_{i,j}
\end{aligned}$$

Since $\sum_{i\in S\backslash k}\mathbb{P}\left(X_\ell=i,\ell=\tau_k|X_0=k\right)=0$, we have

$$\nu_k(j)=\sum_{\ell=1}^{\infty}\sum_{i\in S}\mathbb{P}\left(X_\ell=i,\ell\le\tau_k|X_0=k\right)p_{i,j}=\sum_{i\in S}\nu_k(i)p_{i,j}.$$

Next, we show that $\nu_k(j)<\infty,\forall j\in S$. We can see that $\nu_k(k)=1$. By the stationary equation and substitute j with k, we can see $\nu_k(i)<\infty$ when $p_{i,j}>0$. Furthermoe, by irreducibility result, there exists $i_0=j,i_1,\ldots,i_\ell=k$ so that $p_{i_m,i_{m+1}}>0$, and thus $\nu_k(i_m)>0$ for $m=\ell-1,\ldots,0$.

2. Let η be an arbitrary measure, then

$$\begin{aligned}
\eta(j)&=\sum_{i\in S}\eta(i)p_{i,j}=\eta(k)p_{k,j}+\sum_{i\in S\backslash k}\eta(i)p_{i,j}\\
&=\eta(k)p_{k,j}+\sum_{i\in S\backslash k}\eta(k)p_{k,i}p_{i,j}+\sum_{i_1\in S\backslash k,i_2\in S\backslash k}\eta(i_2)p_{i_2,i_1}p_{i_1,j}\\
&=\eta(k)\sum_{\ell=1}^{n}\mathbb{P}(\ell<\tau_k,X_\ell=j\mid X_0=k)+\sum_{i_n\in S\backslash k}\mathbb{P}(n<\tau_k,X_{n+1}=j\mid X_0=i_n)\eta(i_n)
\end{aligned}$$

Taking $n\to\infty$ gives $\eta(j)\ge\eta(k)\nu_k(j)$. Hence

$$\eta(k)=\sum_{j\in S}\eta(j)p_{j,k}\ge\sum_{j\in S}\eta(k)\nu_k(j)p_{j,k}=\eta(k)\nu_k(k)=\eta(k).$$

The inequality must be inequality.

3. 3) is obvious since $\sum_{j\in S}\nu_k(j)=\mathbb{E}[\tau_k\mid X_0=k]=\mu_{k,k}$.

■

Theorem 10.3 Let $X.$ be an irreducible and positive recurrent Markov chain, and π be the stationary distribution. Then for function $f:\ S\to\mathbb{R}_+$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}f(X_\ell)=\sum_{j\in S}f(j)\pi(j)$$

The RHS can be denoted as $\mathbb{E}_\pi[f(X)]$, which can be infinite.

This formula means that the time average equals the state space average, i.e., **ergodic**.

Proof. We show this result under the conditional distribution given $X_0 = i$. Take $\tau_i(0) = 0$ and $\tau_i(n) = \inf\{\ell > \tau_i(n-1) : X_\ell = i\}$. By strong Markov property, $T_i(n) = \tau_i(n) - \tau_{i-1}(n)$ are i.i.d. random variables. By positive recurrence, $\mu_{ii} = \mathbb{E}[T_i(n) \mid X_0 = i] < \infty$. Take $Y_n = \sum_{\ell=\tau_i(n-1)+1}^{\tau_i(n)} f(X_\ell)$.

- We first consider $\mathbb{E}[Y_1] < \infty$, then $\frac{1}{n}\sum_{\ell=1}^{m} Y_\ell \to \mathbb{E}[Y_1]$. There exists N_n so that $\tau_i(N_n) \leq n < \tau_i(N_{n+1})$. By LLN,

$$\lim_{n\to\infty} \frac{N_n}{n} = \lim_{n\to\infty} \frac{N_n}{\tau_i(N_n)} = \frac{1}{\mu_{ii}}$$

  On the other hand,

$$\frac{N_n}{n}\frac{1}{N_n}\sum_{m=1}^{N_n} Y_m \leq \frac{1}{n}\sum_{\ell=1}^{n} f(X_\ell) \leq \frac{N_{n+1}}{n}\frac{1}{N_{n+1}}\sum_{m=1}^{N_{n+1}} Y_m$$

  Taking $n \to \infty$ gives $\frac{1}{n}\sum_{\ell=1}^{n} f(X_\ell \to \frac{1}{\mu_{ii}}\mathbb{E}[Y_1]$. It suffices to show that

$$\frac{1}{\mu_{ii}}\mathbb{E}[Y_1] = \sum_{j\in S} f(j)\pi(j).$$

  The detailed computation is the following:

$$\begin{aligned}
\frac{1}{\mu_{ii}}\mathbb{E}[Y_1] &= \frac{1}{\mu_{ii}}\mathbb{E}\left[\sum_{\ell=1}^{\tau_i} f(X_\ell)\middle| X_0 = i\right] \\
&= \frac{1}{\mu_{ii}}\mathbb{E}\left[\sum_{\ell=1}^{\infty} f(X_\ell)1(\ell \leq \tau_i)\middle| X_0 = i\right] \\
&= \frac{1}{\mu_{ii}}\mathbb{E}\left[\sum_{\ell=1}^{\infty}\sum_{j\in S} f(X_\ell)1(X_\ell = j, \ell \leq \tau_i)\middle| X_0 = i\right] \\
&= \frac{1}{\mu_{ii}}\sum_{j\in S} f(X_j)\mathbb{E}\left[\sum_{\ell=1}^{\infty} 1(X_\ell = j, \ell \leq \tau_i)\middle| X_0 = i\right] \\
&= \sum_{j\in S} f(X_j)\pi(j)
\end{aligned}$$

- Now we consider $\mathbb{E}[Y_1] = \infty$. Define the truncated function $f^{(k)}(i) = \min(k, f(i))$.

Thus $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} f^{(k)}(X_\ell) = \sum_{j\in S} f^{(k)}(X_j)\pi(j)$. Moreover,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} f(X_\ell) \geq \liminf_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} f^{(k)}(X_\ell) = \sum_{j\in S} f^{(k)}(X_j)\pi(j).$$

By monotone convergence theorem, $\sum_{j\in S} f^{(k)}(X_j)\pi(j) \to \mathbb{E}[Y_1] = \infty$. The proof is completed.

■

10.1.2. Stability Conditions

It is important to determine whether an irreducible Markov chain is positive recurrent or not, since $p_{i,j}^{(n)}$ vanishes as $n \to \infty$ otherwise. We call this problem as stability one since hte positive recurrence implies that the limiting state distribution is stationary. We consider the problem in two cases:

- $|S| < \infty$: all states are positive recurrent;
- $|S| = \infty$: if we can solve the stationary equation, then we can answer. However, it is difficult to do so for general case.

Here we use a test function $f: S \to \mathbb{R}$. By the irreducibility, it suffices to find $\exists i \in S$ so that $\mu_{i,i} < \infty$ to prove the positive recurrence. This means that the Markov chain returns to some finite set of states sufficiently fast, which can be realized by a drift condition towards the set:

Theorem 10.4 Let $X.$ be a Markov chain that is irreducible and aperiodic. If there is a test function $f: S \to \mathbb{R}_+$ and a finite set $S_0 \subseteq S$ so that

- $\mathbb{E}[f(X_n)]$ is finite for all n;
- there exists $\varepsilon > 0$ and $\forall n \geq 0$,

$$\mathbb{E}[f(X_{n+1}) \mid X_n] - f(X_n) \leq -\varepsilon, \quad X_n \notin S_0,$$

then $X_\cdot$ is positive recurrent.

■ **Example 10.3** Let $X_\cdot$ be a simple random walk on $\mathbb{Z}^2$. We modify the random walk so that it is restricted in $\mathbb{Z}_+^2$. Define the Lyapunov function

$$f(\boldsymbol{x}) = \sqrt{ax_1^2 + bx_2^2 + cx_1x_2}, \quad 4ab > c^2.$$

Then by calculation,

$$\begin{aligned}
f(\boldsymbol{x}+U(\boldsymbol{x})) &= f(\boldsymbol{x})\left(1+\frac{f^2(U(\boldsymbol{x}))+(2ax_1+cx_2)U_1(\boldsymbol{x})+(2bx_2+cx_1)U_2(\boldsymbol{x})}{f^2(\boldsymbol{x})}\right)^{1/2} \\
&= f(\boldsymbol{x})\left(1+\frac{f^2(U(\boldsymbol{x}))+(2ax_1+cx_2)U_1(\boldsymbol{x})+(2bx_2+cx_1)U_2(\boldsymbol{x})}{2f^2(\boldsymbol{x})}\right) \\
&= f(\boldsymbol{x})+\frac{f^2(U(\boldsymbol{x}))+(2ax_1+cx_2)U_1(\boldsymbol{x})+(2bx_2+cx_1)U_2(\boldsymbol{x})}{2f(\boldsymbol{x})}
\end{aligned}$$

It follows that

$$\mathbb{E}[f(\boldsymbol{X}_{n+1}) \mid \boldsymbol{X}_n = \boldsymbol{x}] - f(\boldsymbol{x}) = \frac{(2a\mu_1(\boldsymbol{x})+c\mu_2(\boldsymbol{x}))x_1 + (2b\mu_2(\boldsymbol{x})+c\mu_1(\boldsymbol{x}))x_2}{2f(\boldsymbol{x})}$$

■

10.1.3. Markov Modulated Process in Discrete Time

A random walk on a line has the following features:

- Skip free in one-step movements;
- Independent increments.

A discrete time Markov chain relaxes these conditions, but too general as an extension. Now we use the following ideas:

- Keep the state space and the skip free movements of the original process, denoted as $X_\cdot$;
- Add a background state, described by a finite state Markov Chain, denoted as $J_\cdot$.

Definition 10.2 [Markov modulated random walk] We say $Y_\cdot$ is a **Markov modulated random walk** with $Y_n = (X_n, J_n) \in S \equiv S_L \times S_B$ if the following conditions hold:

1. $J_\cdot$ is a Markov chain with transition matrix $A \equiv \{a_{i,j} : i,j \in S_B\}$, where $a_{i,j}$ can be decomposed as $a_{i,j} = a^-_{i,j} + a^0_{i,j} + a^+_{i,j}$ for $a^-_{i,j}, a^0_{i,j}, a^+_{i,j} \geq 0$;
2. Markov chain $Y_\cdot$ has the following transition probabilities:

$$p_{(k,i),(\ell,j)} = \begin{cases} a^-_{i,j} & \ell = k-1 \\ a^0_{i,j} & \ell = k \\ a^+_{i,j} & \ell = k+1 \\ 0 & \text{otherwise} \end{cases}$$

Here $X_\cdot$ is called a level process, and $J_\cdot$ is called a phase process. The transition of phase J_n to J_{n+1} controls the level increment $\Delta X_n = X_{n+1} - X_n$, i.e., $J_\cdot$ modulates the skip free process $X_\cdot$. ■

The most simple case is $S_B = \{1\}$, and $a^+_{1,1} = p, a^0_{1,1} = q, a^-_{1,1} = r$ with $p + q + r = 1$. Consider the occupation measure ν^+ on $S^+_L = \{\ell \in S_L : \ell \geq 0\}$ before returning to $\{k \in S_L : k \leq 0\}$. Now we have

$$\nu^+(\ell) = \begin{cases} (q+r) \cdot \nu^+(0) + r \cdot \nu^+(1), & \ell = 0 \\ p \cdot \nu^+(\ell-1) + q \cdot \nu^+(\ell) + r \cdot \nu^+(\ell+1), & \ell \geq 1. \end{cases}$$

Then we have $\nu^+(\ell) = (p/r)^\ell \nu^+(0)$ for $\ell \geq 1$. We are interested in the following questions:

1. The occupation measure of a Markov modulated random walk has a similar geometric form?
2. How such a result can be used in application?

Definition 10.3 Let $\tau_{0-} = \inf\{n \geq 1 : X_n - X_0 \leq 0\}$ and $D_b(S)$ be the set of all bounded functions from S to $\mathbb{R}$. Define $S_B \times S_B$ matrices $R^{(\ell)}$ for $\ell \geq 1$ as

$$R^{(\ell)}_{i,j} = \mathbb{E}\left(\sum_{n=1}^{\tau_{0-}-1} 1(X_n = \ell, J_n = j) \middle| Y_0 = (0,i) \right), \quad i,j \in S_B$$

Here $\{R^{(\ell)}_{i,j} : j \in S_B\}$ is the occupation measure at $\ell \geq 1$ before returning to the set $A = \{(n,k) \in S : n \leq 0\}$.

Proposition 10.1 The matrix $R^{(\ell)}$ admits the following product form, while R is not a probability matrix in general:

$$R^{(\ell)} = R^{\ell}, \qquad \ell \geq 1.$$

Proof.

$$\begin{aligned}
R^{(\ell+1)}_{i,j} &= \sum_{n=1}^{\infty} \mathbb{P}\left(Y_n = (\ell+1, j), n \leq \tau_{0-} - 1 \middle| Y_0 = (0,i) \right) \\
&= \sum_{n=1}^{\infty} \sum_{m=1}^{n-1} \sum_{i' \in S_B} \mathbb{P}\left(Y_m = (\ell, i'), X_k > \ell, k = m+1, \dots, n, Y_n = (\ell+1, j), n \leq \tau_{0-} - 1 \middle| Y_0 = (0,i) \right) \\
&= \sum_{i' \in S_B} \sum_{m=1}^{\infty} \mathbb{P}\left(Y_m = (\ell, i'), m \leq \tau_{0-} - 1 \middle| Y_0 = (0,i) \right) \\
&\qquad\qquad \times \sum_{n=m+1}^{\infty} \mathbb{P}\left(Y_n = (\ell+1, j), n \leq \tau_{0-} - 1 \middle| Y_m = (\ell, i') \right) \\
&= \sum_{i' \in S_B} R^{(\ell)}_{i,i'} R_{i',j}.
\end{aligned}$$

■

Proposition 10.2 The rate matrix R is a minimal non-negative matrix solution to the equation

$$T = A^{+} + TA^{0} + T^{2}A^{-}.$$

Proof. We first show that $R^{(1)}$ satisfies the stationary equation:

$$
\begin{aligned}
R_{i,j} &= \sum_{n=1}^{\infty} \mathbb{P}\left(Y_n = (1,j), n < \tau_{0-} | Y_0 = (0,i)\right) \\
&= \mathbb{P}\left(Y_1 = (1,j) | Y_0 = (0,i)\right) + \sum_{n=2}^{\infty} \mathbb{P}\left(Y_n = (1,j), n < \tau_{0-} | Y_0 = (0,i)\right) \\
&= a^{+}_{i,j} + \sum_{i' \in S_B} \sum_{n=2}^{\infty} \mathbb{P}\left(Y_{n-1} = (1,i'), Y_n = (1,j), n < \tau_{0-} \middle| Y_0 = (0,i)\right) \\
&+ \sum_{i' \in S_B} \sum_{n=2}^{\infty} \mathbb{P}\left(Y_{n-1} = (2,i'), Y_n = (1,j), n < \tau_{0-} \middle| Y_0 = (0,i)\right)
\end{aligned}
$$

Then we compute the last two double summations:

$$
\begin{aligned}
&\mathbb{P}\left(Y_{n-1} = (1,i'), Y_n = (1,j), n < \tau_{0-} \middle| Y_0 = (0,i)\right) \\
=&\mathbb{P}\left(Y_{n-1} = (1,i'), n-1 < \tau_{0-} \middle| Y_0 = (0,i)\right) \\
&\qquad \mathbb{P}\left(Y_n = (1,j), n < \tau_{0-} \middle| Y_0 = (0,i), Y_{n-1} = (1,i'), n-1 < \tau_{0-}\right) \\
&= \mathbb{P}\left(Y_{n-1} = (1,i'), n-1 < \tau_{0-} \middle| Y_0 = (0,i)\right) a^{0}_{i',j}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{P}\left(Y_{n-1} = (2,i'), Y_n = (1,j), n < \tau_{0-} \middle| Y_0 = (0,i)\right) \\
&= \mathbb{P}\left(Y_{n-1} = (2,i'), n-1 < \tau_{0-} \middle| Y_0 = (0,i)\right) a^{-}_{i',j}.
\end{aligned}
$$

Therefore, $R = A^{+} + RA^{0} + R^{2}A^{-}$.

Next, we show that R is a minimal non-negative solution. Let $T^{(0)} = O$ and inductively define $T^{(n)}$ by

$$
T^{(n)} = A^{+} + T^{(n-1)}A^{0} + (T^{(n-1)})^{2}A^{-}
$$

Note that $T^{(n)}$ satisfies

$$
(T^{(n)} - T^{(n-1)}) = (T^{(n-1)} - T^{(n-2)})A^{0} + ((T^{(n-1)})^{2} - (T^{(n-2)})^{2})A^{-}.
$$

Hence, $T^{(\infty)}$ exists and it must be the minimal solution since $T^{(0)} = O$, In particular, $T^{(\infty)} \leq R$. It suffices to show that $R \leq T^{(\infty)}$. Define $R(n)$ by

$$[R(n)]_{i,j} = \sum_{m=1}^{\infty} \mathbb{P}\left(Y_m = (1,j),\ m < \tau_{0-} \wedge n \middle| Y_0 = (0,i) \right).$$

We can show that $T^{(n-1)} = R(n)$ since $T^{(0)} = O = R(1)$ and $R(n)$ satisfies the inductive equation

$$R(n) = A^+ + R(n-1)A^0 + (R(n-1))^2 A^-$$

■

Definition 10.4 [Markov modulated reflected simple random walk] We put a reflecting boundary at state $0 \in S_L$ in such a way that

$$\tilde{P} = \begin{pmatrix} B^{00} & B^{01} & 0 & 0 & 0 & \cdots \\ B^{10} & A^0 & A^+ & 0 & 0 & \cdots \\ 0 & A^- & A^0 & A^+ & 0 & \vdots \\ 0 & 0 & A^- & A^0 & A^+ & \vdots \\ 0 & 0 & 0 & A^- & A^0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

The modified process $\tilde{Y}_\cdot = (\tilde{X}_\cdot, \tilde{J}_\cdot)$ is called a Markov modulated reflected simple random walk.

Let π be the unique stationary distribution for transition matrix A, and denote the mean drift of the random walk $Y_\cdot$ as

$$\gamma = -\pi A^- + \pi A^+.$$

It is easy to see $\tilde{Y}_n$ has a stationary distribution iff $\gamma < 0$. Denote the stationary distribution as $\tilde{\nu}$, characterized by $\tilde{\nu} = \tilde{\nu}\tilde{P}$. Partition $\tilde{\nu}$ by level and $\tilde{\nu}_n = \mathbb{P}(\tilde{X}_\infty = n)$.

Then we have the stationary equation

$$\tilde{v}_0 = \tilde{v}_0 B^{00} + \tilde{v}_1 B^{10}$$
$$\tilde{v}_1 = \tilde{v}_0 B^{01} + \tilde{v}_1 A^0 + \tilde{v}_2 A^-$$
$$\tilde{v}_n = \tilde{v}_{n-1} A^- + \tilde{v}_n A^0 + \tilde{v}_{n+1} A^-, \quad n \geq 2$$

From the third equation we have $\tilde{v}_n = \tilde{v}_1 R^{n-1}$. From the first two equation we have

$$\begin{cases} \tilde{v}_0 = \tilde{v}_1 B^{10}(I - B^{00})^{-1} \\ \tilde{v}_1 = \tilde{v}_0 B^{01}(I - A^0 - RA^-)^{-1} \end{cases}$$

Here we ask the following questions:

- $\tilde{v}_n$ has a nice expression. Is it possible to use it for asymptotic analysis?
- Can we apply this method to a queueing network?

The first question is yes. For the second question, the corresponding random walk is multi-dimensional, though we only discuss for the one-dimensional case. When S_L, S_B are uncountable, the matrices need to be in the function space, which is a big paradigm beyond the present approach.

Chapter 11

Week 11

11.1. Monday

11.1.1. Continuous Time Stochastic Processes

The big diference between discrete and continuous time stochastic processes is that the time index is uncountable.

- Thus it is natural to consider a continuous time stochastic process $X.$ as a function from $\mathbb{R}_+ \times \Omega \to S$;
- Then we need to define the σ-algebra on $\mathbb{R}_+ \times \Omega$; We also need the notion of null set and completeness.

Definition 11.1 [Null Set and Complete Set] For a σ-algebra $\mathcal{F}$, we say $N \subseteq \Omega$ is a null set if there exists $A \in \mathcal{F}$ so that $N \subseteq A$ and $\mathbb{P}(A) = 0$. We say $\mathcal{F}$ and all its sub σ-fields are complete, if they contains all null sets. ∎

Definition 11.2 [Product σ-algebra] The product σ-algebra, $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$ defined on $\mathbb{R}_+ \times \Omega$, is defined as the minimal σ-field on $\mathbb{R}_+ \times \Omega$ containing

$$\{B \times A : \ B \in \mathcal{B}(\mathbb{R}_+), A \in \mathcal{F}\}.$$

Then we have the measurable space $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F})$. We can define a measure on this space as the following. For a measure μ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ and probability measure

$\mathbb{P}$ on $(\Omega, \mathcal{F})$, the product measure $\mu \otimes \mathbb{P}$ is defined as

$$(\mu \otimes \mathbb{P})(B \times A) = \mu(B)\mathbb{P}(A).$$

Definition 11.3 [Measurable Process] A continuous time stochastic process $X(\cdot)$ is a mapping

$$X(\cdot): \ (t, \omega) \in \mathbb{R}_+ \times \Omega \to X(t, \omega) \equiv X(t)(\omega) \in S.$$

We say $X(\cdot)$ is a measurable process, or $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$-measurable, if

$$\{(t, \omega) \in \mathbb{R}_+ \times \Omega: \ X(t, \omega) \in B\} \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}, \quad \forall t \geq 0, B \in \mathcal{B}(S).$$

Definition 11.4 [Filtration] Let $\mathcal{F}_t$ be a sub σ-algebra of $\mathcal{F}$ so that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $0 \leq s < t$. Then $\mathbb{F} \equiv \{\mathcal{F}_t: \ t \geq 0\}$ is called a filtration. For $t \geq 0$, define

$$\mathcal{F}_{t-} = \sigma\Big(\cup_{0 \leq s < t} \mathcal{F}_s\Big), \quad \mathcal{F}_{t+} = \cap_{t<s} \mathcal{F}_s, \quad \mathcal{F}_\infty = \sigma\Big(\cup_{s \geq 0} \mathcal{F}_s\Big),$$

with $\mathcal{F}_{0-} = \{\emptyset, \Omega\}$. Assume that the filtration $\mathbb{F}$ is right-continuous, i.e., $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$. Then $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a stochastic basis.

Definition 11.5 [Stopping Time] A random variable τ valued in $[0, \infty]$ is called a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \geq 0$. Define the τ-stopped σ-field as

$$\mathcal{F}_\tau \equiv \{A \in \mathcal{F}: \ A \cap \{\tau \leq t\} \in \mathcal{F}_t, \forall t \in [0, \infty]\}.$$

We say $X(\cdot)$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is adapted if $X(t)$ is $\mathcal{F}_t$-measurable. We extend this notion on the product measurable space $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F})$:

Definition 11.6 [Progressively Measurable] We say $X(\cdot)$ is progressively measurable if

$$\forall B \in \mathcal{B}(S), \forall t \geq 0, \quad \{(u,\omega) \in [0,t] \times \Omega : \ X(u,\omega) \in B\} \in \mathcal{B}_t \otimes \mathcal{F}_t$$

is progressively measurable.

Proposition 11.1 Let τ be a stopping time, and $\mathbb{F}$ be right-continuous. Suppose that $X(\cdot)$ is adapted to $\mathbb{F}$ and right-continuous in time, then

1. $X(\tau)$ is $\mathcal{F}_\tau$-measurable;
2. $X(\cdot)$ is progressively measurable;
3. τ-stopped process $X^\tau(\cdot)$ is progressively measurable.

Proof. Here we only present the proof for b) implies c), i.e., suffices to show that for any $B \in \mathcal{B}(S)$ and $t \geq 0$, $A(\tau,t,B) \in \mathcal{B}_t \otimes \mathcal{F}_t$, where

$$A(\tau,t,B) = \{(u,\omega) \in [0,t] \times \Omega : \ X(u \wedge \tau,\omega) \in B\}.$$

Construct the stochastic interval

$$[\sigma,\tau] = \{(u,\omega) \in [0,t] \times \Omega : \ \sigma(\omega) \leq u \leq \tau(\omega)\},$$

then we have

$$[0,\tau] \cap A(\infty,t,B) \in \mathcal{B}_t \otimes \mathcal{F}_t, \quad (\tau,t] \cap A(\tau,t,B) \in \mathcal{B}_t \otimes \mathcal{F}_t.$$

■

The progressive measurability is wearker than the right-continuity of a sample path. But the latter is easy to check. Therefore, we make the following assumptions:

- $X.$ is adapted;
- For each ω, the sample path $X(t)(\omega)$ is right-continuous, and has limit from the

left at each $t \geq 0$:

$$X(t)(\omega) = X(t+)(\omega), \qquad X(t-)(\omega) \text{ exists.}$$

- Filtration $\mathbb{F}$ is right-continuous, i.e., $\mathcal{F} = \mathcal{F}_{t+} = \sigma(\cap_{t<u}\mathcal{F}_u)$.
- Filtration $\mathbb{F}$ is complete.

Definition 11.7 [Predictable] Let $\mathcal{U}_-(S)$ be the set of all left-continuous and adapted measurable process $\mathbb{R}_+ \times \Omega$ to S. Define

$$\mathcal{P} = \sigma\left(\bigcup_{f\in\mathcal{U}_-(S)} \bigcup_{B\in\mathcal{B}(S)} \{(t,\omega) \in \mathbb{R}_+ \times \Omega : \ f(t,\omega) \in B\}\right),$$

which is called a predictable σ-field. A stochastic process $X(\cdot)$ is said to be predictable if it is $\mathcal{P}$-measurable. A random variable $\tau \geq 0$ is said to be a predictable time if $[0,\tau)$ is $\mathcal{P}$-measurable. ■

Proposition 11.2 The predictable σ-algebra $\mathcal{P}$ is generated by either one of the sets:

- $\{0\} \times A$ for $A \in \mathcal{F}_0$, and $[t,\tau]$ for all stopping time τ;
- $\{0\} \times A$ for $A \in \mathcal{F}_0$, and $(s,t] \times A$ for all $s < t, A \in \mathcal{F}_s$.

Proof. Let $\mathcal{P}_1, \mathcal{P}_2$ be the σ-algebra generated by i), ii). Define $f(t,\omega) = 1(0 \leq t \leq \tau(\omega))$, which is predictable, i.e., $\mathcal{P}_1 \subseteq \mathcal{P}$. For $A \in \mathcal{F}_s$, define the notion $u_A = u1_A + \infty 1_{A^c}$ for $u \geq s$. For the set of rational numbers Q,

$$(s,t] \times A = (s_A, t_A] = [0,t_A] \setminus [0,s_A] \in \mathcal{P}_1.$$

Thus $\mathcal{P}_2 \subseteq \mathcal{P}_1$. Note that $X(t)$ is the limit of the discrete process $X^n(t) \equiv \sum_{i\geq 0} X(i2^{-n})1(i2^{-n} < t \leq (i+1)2^{-n})$, which is $\mathcal{P}_2$-measurable. So $\mathcal{P} \subseteq \mathcal{P}_2$.

■

R If $X(\cdot)$ is predictable, then $X(t)$ is $\mathcal{F}_{t-}$-measurable. But the converse is not true.

11.1.2. Continuous time martingale

For discrete case, we find any real-valued process can be expressed as a semi-martingale. We wish to study the similar topic for continuous case. In general this result does not hold, but for most processes we encounter in applications, the answer is yes. We will focus on a semi-martingale.

Definition 11.8 [Martingale] The process $M(\cdot)$ is called a martingale if

$$\mathbb{E}[M(t) \mid \mathcal{F}_s] = M(s), \quad 0 \le s < t.$$

Now we define the semi-martingale as

$$X(t) = A(t) + M(t),$$

where $A(\cdot)$ is an adapted process of finite variations:

$$|A|(t) = \sup_{\delta>0, t_0=0} \sup_{\Delta t \ge \delta} \sum_{n=1}^{\infty} 1(t_n \le t)|A(t_n) - A(t_{n-1})| M\infty, \quad \forall t \ge 0,$$

and $M(\cdot)$ is a martingale. The process $A(\cdot)$ can be expressed as the difference of non-decreasing functions. If $A(\cdot)$ is predictable, then $X(\cdot)$ is called a special semi-martingale.

■

Definition 11.9 [Uniform Integrable] For a set K, a collection of real-valued random variable $\{X_u : u \in K\}$ is said to be uniformly integrable if

$$\lim_{a\to\infty} \sup_{u\in K} \mathbb{E}[|X_u| 1(|X_u| > a)] = 0.$$

A martingale $X(\cdot)$ is said to be uniform integrable if $\{X(t) : t \ge 0\}$ is uniform integrable.

■

Proposition 11.3 — UI for Conditional Expectations. Let Y be a random variable with $\mathbb{E}[|Y|] < \infty$, and $\mathcal{H}$ be a collection of all σ-fields on Ω, then $\{\mathbb{E}[Y \mid \mathcal{G}] : \mathcal{G} \in \mathcal{H}\}$ is uniform integrable.

Proof. Define $Y_{\mathcal{G}} = \mathbb{E}[Y \mid \mathcal{G}]$, then

$$\begin{aligned}
\mathbb{E}[|Y_{\mathcal{G}}|1(|Y_{\mathcal{G}}| > a)] &= \mathbb{E}[|\mathbb{E}[Y \mid \mathcal{G}]|1(|Y_{\mathcal{G}}| > a)] \\
&\leq \mathbb{E}[\mathbb{E}[|Y| \mid \mathcal{G}]1(|Y_{\mathcal{G}}| > a)] \\
&= \mathbb{E}[|Y|1(|Y_{\mathcal{G}}| > a)],
\end{aligned}$$

where

$$\mathbb{E}[|Y|1(|Y_{\mathcal{G}}| > a)] \leq \mathbb{E}[|Y|1(|Y| > b)] + b\mathbb{E}[1(|Y_{\mathcal{G}}| > a)]$$

■

Then we define $X(t) = \mathbb{E}[Y \mid \mathcal{F}_t]$ for $t \geq 0$. By the proposition above, $X(\cdot)$ is UI. Moreover, we can show that $X(\cdot)$ is a martingale.

Definition 11.10 [Closed Martingale] A martingale $X_u(\cdot) = \{X_u(t) : 0 \leq t < u\}$ for a constaint $u \in [0, \infty]$ is said to be closed if $X_u(u)$ exists and $\{X_u(t) : 0 \leq t \leq u\}$ is a martingale.

If we define $X(\infty) = \mathbb{E}[Y \mid \mathcal{F}_\infty]$, then $X(\cdot)$ is closed. Moreover, it is UI. This suggests the following lemma:

Proposition 11.4 Let $X_u(\cdot)$ be a martingae for $u \in [0, \infty]$. Then $X_u(\cdot)$ is UI iff it is closed.

Proof. It suffices to show the closeness result. Construct $u_n \uparrow u$ and $Y_n = X_u(u_n)$, then $Y.$ is UI. There exists Y_∞ so that $Y_n \to Y_\infty$ almost surely. Therefore, $X_u(\cdot)$ can be expressed as the conditional expectation form, which is closed:

$$X_u(t) = \mathbb{E}[X_u(u_n) \mid \mathcal{F}_t] = \mathbb{E}[\mathbb{E}[Y_\infty \mid \mathcal{F}_{u_n}] \mid \mathcal{F}_t] = \mathbb{E}[Y_\infty \mid \mathcal{F}_t].$$

■

Definition 11.11 [Local Martingale] A process $X(\cdot)$ is called a **local martingale** if there is a sequence of finite stopping times $\tau_n \uparrow \infty$ so that $X^{\tau_n} \equiv \{X(\tau_n \wedge t) : t \geq 0\}$ is a martingale for each $n \geq 1$. ■

Definition 11.12 [Class D] A process $X(\cdot)$ is called of class D if $\{X(\tau) : \tau \in \mathcal{P}\}$ is uniformly integrable for $\mathcal{P}$ be the set of all finite stopping times. ■

Proposition 11.5
1. A martingale $X(\cdot)$ is a local martingale;
2. If $X(\cdot)$ is an uniformly integrable martingale, then it is of class D;
3. A local martingale $X(\cdot)$ is uniformly integrable iff it is of class D.

Proof.
1. It is clear if we take $\tau_n \equiv n$.
2. By proposition 11.4, $X(\cdot)$ is closed, and therefore $X(t) = \mathbb{E}[Y \mid \mathcal{F}_t]$ for a random variable Y with finite $\mathbb{E}[Y]$. Therefore, $X(\tau) = \mathbb{E}[Y \mid \mathcal{F}_\tau]$ for any stopping time τ. Hence, $\{X(\tau) : \tau \in \mathcal{P}\}$ is uniformly integrable.
3. Let $\{\tau_n : n \geq 1\}$ be localizing sequence of finite stopping times for $X(\cdot)$. Assume that it is of class D, then $\{X(\tau_n \wedge t) : t \geq 0, n \geq 1\}$ are uniformly integrable. The uniform integrability of $X(\cdot)$ is because

$$\sup_{n \geq 1, t \geq 0} \mathbb{E}[|X^{\tau_n}(t)| 1(|X^{\tau_n}(t)| > a)] \geq \sup_{t \geq 0} \mathbb{E}[|X(t)| 1(|X(t)| > a)].$$

Conversely, assume that $\{X(\tau_n \wedge t) : t \geq 0, n \geq 1\}$ are uniformly integrable, then $\{X(\tau_n \wedge u) : t \geq 0, n \geq 1\}$ are uniformly integrable and $X^{\tau_n}(u) \to X(u)$ almost surely. Hence, $X^{\tau_n}(u) \xrightarrow{L^1} X(u)$ and $X(s) = \mathbb{E}[X(t) \mid \mathcal{F}_s]$. Thus it is of class D.

■

A non-decreasing process $X(\cdot)$ is a sub-martingale since

$$\mathbb{E}[X(t) \mid \mathcal{F}_s] = X(s) + \mathbb{E}[X(t) - X(s) \mid \mathcal{F}_s] \geq X(s), \quad 0 \leq s < t.$$

The following tells what a sub-martingale is.

Theorem 11.1 — Doob Decomposition. If $X(\cdot)$ is a sub-martingale of class D, then there uniquely exists a predictable non-decreasing process $A(\cdot)$ with $A(0) = 0$ and a uniformly integrable martingale $M(\cdot)$ so that

$$X(t) = A(t) + M(t), \quad t \geq 0.$$

Proposition 11.6 If a martingale $X(\cdot)$ is predictable and has finite variations, then $X(t) = 0$ for all $t \geq 0$ except for all the null sets.

Proof. It suffices to consider the case where $X(\cdot)$ is non-decreasing. Then we have $X(\cdot) = A(\cdot)$ and $X(\cdot) = M(\cdot)$. By the uniqueness of decomposition, $X(\cdot) = 0$. ■

Theorem 11.2 If $X(\cdot)$ is a special martingale, i.e., $A(\cdot)$ is predictable, then the semi-martingale decomposition is unique.

Proof. Suppose that $X(t) = A'(t) + M'(t)$ is another semi-martingale form, then $A(t) - A'(t) = M'(t) - M(t)$ is a predictable martingale of finite variations, so it must vanish.

■

11.1.3. From discrete to continuous time

Assume that $X(\cdot)$ is a cadlag process. We transfer results from its discrete time counterpart in the following way:

1. For $n, i \geq 0$, define the mesh $t_{n,i} = i2^{-n}$ and $Y_i^{(n)} = X(t_{n,i}), \mathcal{F}_i^{(n)} = \mathcal{F}_{n,i}$. Consider the discrete time process $Y_\cdot^{(n)} = \{Y_i^{(n)} : i \geq 0\}$ under the filtration $\mathbb{F}^{(n)} \equiv \{\mathcal{F}_i^{(n)} : i \geq 0\}$.
2. For $t > 0, n \geq 0$, let $Y_\cdot^{(n)}(t) = \{Y_i^{(n)} : 0 \leq i \leq n_t - 1\} \cup \{X(t)\}$, and define $\mathbb{F}^{(n)}(t) \equiv \{\mathcal{F}_i^{(n)} : 0 \leq i \leq n_t - 1\} \cup \{\mathcal{F}_t\}$, where

   $$n_t = \min\{\ell \in \mathbb{Z}_+ : l \geq t2^n\}.$$

   Approximate $\{X(u) : 0 \leq u \leq t\}$ by $Y_\cdot^{(n)}(t)$;
3. Derive continuous time results from those for $Y_\cdot^{(n)}$ by taking $n \to \infty$. Here we need a topology on state space S of $X(\cdot)$. Usually we take $S = \mathbb{R}$ and the topology by Euclid norm.

This procedure involves at most countable events, so the limiting result holds with probability 1 if each result holds with probability 1.

Proposition 11.7 — Observation at Stopping Time. Let τ be a stopping time, then

1. $X(\tau)1(\tau < \infty)$ is $\mathcal{F}_\tau$-measurable;
2. If $X(\cdot)$ is predictable, then $X(\tau)1(\tau < \infty)$ is $\mathcal{F}_{\tau-}$-measurable, with

$$\mathcal{F}_{\tau-} = \sigma(\mathcal{F}_0, \{A \in \mathcal{F} : A \cap \{t < \tau\} \in \mathcal{F}_t, t \geq 0\}).$$

Proof. It suffices to show that $\{X(\tau) \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$ for any t. On the event $\{\tau < \infty\}$, we approximate τ by:

$$\tau_n^+ = \sum_{i=0}^{\infty} t_{n,i} 1(t_{n,i-1} \leq \tau < t_{n,i}), \quad n \geq 0,$$

with $t_{n,i} = i2^{-n}$ and $\mathbb{F}^{(n)} \equiv \{\mathcal{F}_{t_{n,i}} : i \geq 0\}$. It's clear that τ_n^+ is $\mathbb{F}^{(n)}$-stopping time and $\tau_n^+ \downarrow \tau$ as $n \to \infty$. Note that $X(t_{n,i})1(\tau_n^+ = t_{n,i})$ is $\mathcal{F}_{t_{n,i}}$-measurable, which implies

$$X(\tau_n^+)1(\tau_n^+ \leq t_{n,j}) = \sum_{i=0}^{j} X(t_{n,i})1(\tau_n^+ = t_{n,i}) \text{ is } \mathcal{F}_{t_{n,j}}\text{-measurable.}$$

Thus $X(\tau_n^+)$ is $\mathcal{F}_{\tau_n^+}$-measurable for each $n \geq 0$. Moreover, $X(\tau_n^+) \to X(\tau)$ on $\{\tau < \infty\}$ as $n \to \infty$. Therefore, $X(\tau)1\{\tau < \infty\}$ is $\mathcal{F}_\tau$-measurable.

For part 2), we construct

$$\tau_n^- = \sum_{i=0}^{\infty} t_{n,i-1} 1(t_{n,i-1} \leq \tau < t_{n,i}), \quad n \geq 0,$$

then $\tau_n^- \uparrow \tau$. We can see that $X(\tau_n^-)$ is $\mathcal{F}_{\tau_n^-}$-measurable. By the predictability of $X(\cdot)$, $X(\tau)1\{\tau < \infty\}$ is $\mathcal{F}_{\tau-}$-measurable since $t < \tau_n^-$ implies $t < \tau$. ∎

We introduce some new notions to study stopped martingales.

Definition 11.13 [Reverse Filtration] The filtration $\mathbb{F} \equiv \{\mathcal{F}_n : 0 \leq n \leq \infty\}$ is called a reverse filtration if $\mathcal{F}_n \downarrow F_\infty = \cap_{\ell=1}^{\infty} \mathcal{F}_\ell$. An adapted process $X_\cdot$ is called a reverse martingale if $\mathbb{E}[X_m \mid \mathcal{F}_n] = X_n$ for $m < n$.

Proposition 11.8 A reverse sub-martingale $X_\cdot$ is uniformly integrable if $\mathbb{E}[X_n] > b$ for any $n \geq 0$ and some $b \in \mathbb{R}$.

Proof. By the reverse sub-martingale assumption, $\mathbb{E}[X_n]$ is decreasing in n and converges to a finite α. Moreover, $X_n \leq \mathbb{E}[X_0 \mid \mathcal{F}_n]$. Hence, $X_\cdot^+$ is uniformly integrable by conditional expectation. It remains to check whether $X_\cdot^-$ is uniformly integrable. For $a > 0$,

$$\begin{aligned}\mathbb{E}[X_n^- 1(X_n^- > a)] &= \mathbb{E}[-X_n 1(-X_n > a)] = \mathbb{E}[-X_n(1 - 1(-X_n \leq a))] \\ &= -\mathbb{E}[X_n] + \mathbb{E}[X_n 1(-X_n \leq a)]\end{aligned}$$

By the convergence of X_n, there exists an $m \geq 0$ so that $|\mathbb{E}[X_n] - \mathbb{E}[X_m]| < \varepsilon$ as long as $n \geq m$. As a result,

$$\begin{aligned}\mathbb{E}[X_n^- 1(X_n^- > a)] &\leq -\mathbb{E}[X_m] + \mathbb{E}[X_n 1(-X_n \leq a)] + \varepsilon \\ &\leq -\mathbb{E}[X_m] + \mathbb{E}[X_m 1(-X_n \leq a)] + \varepsilon \\ &= -\mathbb{E}[X_m 1(-X_n > a)] + \varepsilon \leq \mathbb{E}[|X_m| 1(-X_n > a)] + \varepsilon.\end{aligned}$$

It remains to show that

$$\lim_{a \to \infty} \sup_{n \geq m} \mathbb{E}[|X_m| 1(-X_n > a)] = 0.$$

Observe that $\mathbb{E}[|X_n|] \leq -\alpha + \varepsilon + 2\mathbb{E}[X_n^+] \leq -\alpha + \varepsilon + 2\mathbb{E}[X_0^+] \equiv \beta$. It follows that

$$\mathbb{E}[|X_m| 1(-X_n > a)] \leq \mathbb{E}[|X_m| 1(|X_n| > a)] \to 0$$

■

Recall that $X^\tau(t) = X(\tau \wedge t)$ for $t \geq 0$, and $X^\tau(\cdot)$ is a τ-stopped process.

Proposition 11.9 If $X(\cdot)$ is a martingale and τ is a stopping time, then $X^\tau(\cdot)$ is also a martingale.

In discrete time case, we obtain the optional sampling theorem under the uniform integrability condition in the following steps:

1. Assume that the stopping time τ is bounded, and derive the theorem;
2. Under the condition 1), show the almost sure convergence of a sub-martingale assuming $\sup_{n\geq 0}\mathbb{E}[X_n^+] < \infty$. The key is applying Doob's upcrossing inequality;
3. For a general stopping time τ, show that a stopped process $X_\cdot^\tau$ is a uniformly integrable martingale if $X_\cdot$ is so.

All those steps can be carried for the continuous process $X(\cdot)$ in such a way that in step 2), the condition is replaced by $\sup_{n\geq 0}\mathbb{E}[X^+(t)] < \infty$, and Doob's inequality is applied into the discrete time process $Y_\cdot^{(n)}(t)$.

Consider $Y_\cdot^{(n)}(t)$ defined as

$$Y_\cdot^{(n)}(t) = \{X(t_{n,i}) : \ 0 \leq i \leq n_t - 1\} \cup \{X(t)\},$$

which is a sub-martingale when $X(\cdot)$ is so. For $a < b$, define $N_0 = -1$, and define the down and up crossings:

$$N_{2k-1}^{(n)} = \inf\{i > N_{2k-2}^{(n)} : \ Y_i^{(n)}(t) \leq a\}, \quad N_{2k}^{(n)} = \inf\{i > N_{2k-1}^{(n)} : \ Y_i^{(n)}(t) \geq b\}.$$

Take $U^{(n)}(t) = \sup\{k \geq 1 : \ N_{2k}^{(n)} \leq n_t\}$, where $n_t = \min\{\ell \geq 0 : \ \ell 2^{-n} \geq t\}$. The following lemma holds immediately:

Proposition 11.10 If $X(\cdot)$ is a sub-martingale, then for any $a, b \in \mathbb{R}$ with $a < b$, and any $t > 0, n \geq 0$,

$$(b-a)\mathbb{E}[U^{(n)}(t)] \leq \mathbb{E}[(X(t)-a)^+] - \mathbb{E}[(X(0)-a)^+]$$

Theorem 11.3 Suppose that $X(\cdot)$ is a sub-martingale and $\sup_{n\geq 0}\mathbb{E}[X^+(t)] < \infty$, then

$$X(t) \xrightarrow{a.s.} X(\infty)$$

for some random variable $X(\infty)$ with $\mathbb{E}[|X(\infty)|] < \infty$.

Theorem 11.4 Suppose that $X(t) \xrightarrow{P} X$ as $t \to \infty$, then the followings are equivalent:

1. $X(\cdot)$ is UI;
2. $X(t) \xrightarrow{L^1} X$;
3. $\mathbb{E}[|X(t)|] \to \mathbb{E}[|X|] < \infty$.

Theorem 11.5 — **Uniformly Integrable Submartingale.** If $X(\cdot)$ is a uniformly integrable sub-martingale, then there exists $X(\infty)$ with $X(t) \to X(\infty)$ almost surely and L^1. Furthermore, $\mathbb{E}[|X(\infty)|] < \infty$, $X(t) \le \mathbb{E}[X(\infty) \mid \mathcal{F}_t]$.

Theorem 11.6 — **Optional Sampling Theorem.** If $X(\cdot)$ is a uniformly integrable sub-martingale, and τ, η are stopping times with $\tau \le \eta$, then

$$\mathbb{E}[X(0)] \le \mathbb{E}[X(\tau)] \le \mathbb{E}[X(\eta)] \le \mathbb{E}[X(\infty)],$$

and $X(\tau) \le \mathbb{E}[X(\eta) \mid \mathcal{F}_\tau]$.

Chapter 12

Week 12

12.1. Monday

We will introduce three continuous time stochastic processes, the first two of which have **independent increments**. They naturally arise in many applications, but their sample paths are quite different from the discrete time case, since the sample paths of continuous time processes are functions on $\mathbb{R}_+$. The two extremes of these sample paths are:

- Only changes discontinuously, i.e., piecewise step function;
- Only changes continuously, i.e., continuous function of time.

A typical example of i) is a Possion process, while the example of ii) is a Brownian motion.

In general, a real-valued process with independent increments is called a Levy process. In this week, we will discuss another process, the continuous time Markov chain, It relaxes the independent increment condition, and it is a natural extension of discrete time Markov chain. The Possion process is a special case of CTMC. We will study this process first, and it will motivate us to study a general continuous time Markoc chain, or further continuous time Markov process.

12.1.1. Poisson Process

Definition 12.1 [Simple Counting Process] A non-negative integer-valued process $N(\cdot) \equiv \{N(t): t \geq 0\}$ is called a simple counting process, if $N(0) = 0$, and $N(t)$ is piecewise constant in $t \geq 0$, and $\Delta N(t) \equiv N(t) - N(t-)$ is either 0 or 1 for all $t \geq 0$. ■

Definition 12.2 [Poisson Process] If a simple counting process $N(\cdot)$ adapted to $\mathbb{F}$ satisfies

1. For all $s, t > 0$, $N(s+t) - N(s)$ is independent of $\mathcal{F}_s$;
2. For any $s, t > 0$, $N(t) \overset{d}{\sim} N(s+t) - N(s)$;
3. $\lambda \equiv \mathbb{E}[N(1)] < \infty$ and $\lambda > 0$,

then $N(\cdot)$ is called a time-homogeneous Poisson process with intensity λ. ■

Theorem 12.1 — Martingale Characterization of Poisson Process. Let $N(\cdot)$ be a simple counting process and define $M(\cdot)$ by

$$M(t) = N(t) - \lambda t, \quad t \geq 0,$$

then $N(\cdot)$ is a Poisson process with intensity λ if and only if $M(\cdot)$ is a martingale.

We will make use of the following fact for showing this theorem:

> Suppose that f is a right-continuous function from $\mathbb{R}_+$ to $\mathbb{R}$. Then for $\forall s, t \geq 0$, $f(s+t) = f(s) + f(t)$ holds if and only if $f(t) = f(1)t$.

Proof. Assume that $M(\cdot)$ is a Poisson process, then $\mathbb{E}[N(s+t)] = \mathbb{E}[N(s)] + \mathbb{E}[N(t)]$, which implies $\mathbb{E}[N(t)] = \lambda t$. Furthermore, for $0 \leq s < t$,

$$\begin{aligned}
\mathbb{E}[M(t) \mid \mathcal{F}_s] &= \mathbb{E}[N(t) \mid \mathcal{F}_s] - \lambda t \\
&= N(s) + \mathbb{E}[N(t) - N(s) \mid \mathcal{F}_s] - \lambda t \\
&= N(s) + \mathbb{E}[N(t-s)] - \lambda t = N(s) - \lambda s = M(s).
\end{aligned}$$

Moreover, $\mathbb{E}[|M(t)|] \leq 2\lambda t$, which implies $M(\cdot)$ is a martingale.

Conversely, suppose that $M(\cdot)$ is a martingale. Then $\mathbb{E}[N(t)] = \lambda t < \infty$, so 3) for Poisson process holds. We use the test function $e^{\theta x}, \theta < 0$ to study the distribution of

$N(t)$. Denote $\Delta N(t) = N(t) - N(t-)$, then for $s < t$,

$$\begin{aligned} e^{\theta N(t)} - e^{\theta N(s)} &= \sum_{\Delta N(u) > 0} (e^{\theta N(u)} - e^{\theta N(u-)}) 1(s < u \le t) \\ &= (e^{\theta} - 1) \int_s^t e^{\theta N(u-)} \, \mathrm{d}N(u) \\ &= (e^{\theta} - 1) \left[\int_s^t e^{\theta N(u)} \lambda \, \mathrm{d}u + \int_s^t e^{\theta N(u-)} \, \mathrm{d}M(u) \right] \end{aligned}$$

Define $Y(t) = \int_0^t e^{\theta N(u-)} \, \mathrm{d}M(u)$ and $\sigma_n = \inf\{t \ge 0 : \ N(t) = n\}$. Then we have

$$\int_{\sigma_{n-1}}^{\sigma_n} e^{\theta N(u-)} \, \mathrm{d}M(u) = e^{\theta(n-1)} [M(\sigma_n) - M(\sigma_{n-1})].$$

Therefore,

$$\begin{aligned} \int_0^t e^{\theta N(u-)} \, \mathrm{d}M(u) &= \sum_{n=1}^{\infty} \int_{\sigma_{n-1} \wedge t}^{\sigma_n \wedge t} e^{\theta N(u-)} \, \mathrm{d}M(u) \\ &= \sum_{n=1}^{\infty} e^{\theta(n-1)} [M(\sigma_n \wedge t) - M(\sigma_{n-1} \wedge t)], \end{aligned}$$

and thus

$$\int_s^t e^{\theta N(u-)} \, \mathrm{d}M(u) = Y(t) - Y(s) = \sum_{n=1}^{\infty} e^{\theta(n-1)} [M((s \vee \sigma_n) \wedge t) - M((s \vee \sigma_{n-1}) \wedge t)].$$

It follows that

$$\begin{aligned} e^{\theta N(t)} - e^{\theta N(s)} = (e^{\theta} - 1) \bigg[& \int_s^t e^{\theta N(u)} \lambda \, \mathrm{d}u \\ & + \sum_{n=1}^{\infty} e^{\theta(n-1)} [M((s \vee \sigma_n) \wedge t) - M((s \vee \sigma_{n-1}) \wedge t)] \bigg] \end{aligned}$$

By optional sampling theorem, $\mathbb{E}[M((s \vee \sigma_n) \wedge t) \mid \mathcal{F}_s] = \mathbb{E}[M((s \vee \sigma_{n-1}) \wedge t) \mid \mathcal{F}_s] = M(s)$, and thus

$$\mathbb{E}[e^{\theta N(t)} - e^{\theta N(s)} \mid \mathcal{F}_s] = (e^{\theta} - 1) \lambda \mathbb{E} \left[\int_s^t e^{\theta N(u)} \, \mathrm{d}u \middle| \mathcal{F}_s \right]$$

Dividing both sides by $e^{\theta N(s)}$ gives

$$\mathbb{E}[\exp(\theta(N(t)-N(s)))\mid\mathcal{F}_s]=\lambda(e^\theta-1)\int_s^t\mathbb{E}[\exp(\theta(N(u)-N(s)))\mid\mathcal{F}_s]\,\mathrm{d}u$$

Differentiate both sides and by initial value,

$$\mathbb{E}[\exp(\theta(N(t)-N(s)))\mid\mathcal{F}_s]=e^{-\lambda(t-s)(1-e^\theta)}=\mathbb{E}[e^{\theta N(t-s)}]. \tag{12.1}$$

This shows the property 2 for Poisson process. From the equation above, we can also see that $N(t)-N(s)$ is independent of $\mathcal{F}_s$, so property 1 is also shown. ■

From the proof we can see that $\mathbb{E}[e^{\theta N(t)}]=\exp(-\lambda t(1-e^\theta))$, which implies the following corollary:

Corollary 12.1 If $N(\cdot)$ is the Poisson process with intensity λ, then $N(t)$ has the Poisson distribution with mean λ:

$$\mathbb{P}(N(t)=k)=\frac{1}{k!}(\lambda t)^k e^{-\lambda t}$$

Therefore, a probability measure exists on which the Poisson process $N(\cdot)$ is uniquely defined. The next theorem answers it in another perspective:

Theorem 12.2 — **Inter-Counting Time Characterization of Poisson Process.** A simple counting process $N(\cdot)$ is a Poisson process with intensity $\lambda>0$, if and only if, for $\sigma_n=\inf\{t\geq 0:\ N(t)\geq n\}$ we have

- $T_n=\sigma_n-\sigma_{n-1}$ are i.i.d. with $\mathbb{P}(T_n\leq x)=1-e^{-\lambda x}, x\geq 0$.

Proof.
- Assume that $N(\cdot)$ is Poisson with intensity λ, and $X(t)=\exp(\theta N(t)+\lambda t(1-e^\theta))$. Then $X(\cdot)$ is a martingale. By the optinal sampling theorem for stopping times $\sigma_n^{(k)}=\sigma_n\wedge k,\sigma_{n-1}^{(k)}=\sigma_{n-1}\wedge k$, we have

$$\mathbb{E}\left[e^{\theta N(\sigma_n^{(k)})+\lambda\sigma_n^{(k)}(1-e^\theta)}\middle|\mathcal{F}_{\sigma_{n-1}^{(k)}}\right]=X(\sigma_{n-1}^{(k)})=e^{\theta N(\sigma_{n-1}^{(k)})+\lambda\sigma_{n-1}^{(k)}(1-e^\theta)}$$

Taking $k \to \infty$ gives

$$\mathbb{E}[e^{\theta+\lambda(\sigma_n-\sigma_{n-1})(1-e^{\theta})} \mid \mathcal{F}_{\sigma_{n-1}}] = 1.$$

Define $u = -\lambda(1-e^{\theta})$. It follows that

$$\mathbb{E}[e^{-T_n u} \mid \mathcal{F}_{\sigma_{n-1}}] = e^{-\theta} = \frac{\lambda}{\lambda+u}.$$

- Conversely, assume the property holds, then we first show the following equality by induction:

$$\mathbb{P}(N(t) \le \ell) = \mathbb{P}(\sigma_{\ell+1} > t) = \sum_{k=0}^{\ell} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This holds for $\ell = 0$. Assume that it holds for a given $\ell \ge 0$, then

$$\begin{aligned}
\mathbb{P}(\sigma_{\ell+2} > t) &= \mathbb{P}(\sigma_1 > t) + \mathbb{P}(\sigma_1 \le t, \sigma_{\ell+2} - \sigma_1 + \sigma_1 > t) \\
&= e^{-\lambda t} + \int_0^t \mathbb{P}(\sigma_{\ell+2} - \sigma_1 + u > t)\, \mathrm{d}F(u) \\
&= e^{-\lambda t} + \int_0^t \mathbb{P}(\sigma_{\ell+1} + u > t) \lambda e^{-\lambda u}\, \mathrm{d}u \\
&= \sum_{k=0}^{\ell+1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.
\end{aligned}$$

It follows that $\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{P}(\sigma_n \le u) = \frac{\lambda^n u^{n-1}}{(n-1)!} e^{-\lambda u}$. Hence,

$$\begin{aligned}
&\mathbb{P}(N(s) = k, N(t) - N(s) \le \ell) \\
&= \mathbb{P}(\sigma_k \le s < \sigma_{k+1}, \sigma_{k+1} > t) + \mathbb{P}(\sigma_k \le s < \sigma_{k+1}, \sigma_{k+1} \le t, \sigma_{k+\ell+1} > t) \\
&= \mathbb{P}(\sigma_k \le s, \sigma_{k+1} > t) + \mathbb{P}(\sigma_k \le s < \sigma_{k+1} \le t, \sigma_{k+\ell+1} - \sigma_{k+1} + \sigma_{k+1} > t) \\
&= \int_0^s \mathbb{P}(T_k > t - u)\, \mathrm{d}_u \mathbb{P}(\sigma_k \le u) \\
&+ \int_0^s \int_{s-u}^{t-u} \mathbb{P}(\sigma_\ell > t - (u+v))\, \mathrm{d}_v \mathbb{P}(T_k \le v)\, \mathrm{d}_u \mathbb{P}(\sigma_k \le u) \\
&= \mathbb{P}(N(s) = k)\mathbb{P}(N(t-s) \le \ell)
\end{aligned}$$

From this equality we have

$$\mathbb{P}(N(s) = k, N(t) - N(s) \le \ell) = \mathbb{P}(N(s) = k)\mathbb{P}(N(t) - N(s) \le \ell)$$

Similarly, $\mathbb{P}(A \cap \{N(s) = k, N(t) - N(s) \le \ell\}) = \mathbb{P}(A \cap \{N(s) = k\})\mathbb{P}(N(t) - N(s) \le \ell), \quad A \in \mathcal{F}_s$. As a result,

$$\mathbb{E}[N(t) - N(s) \mid \mathcal{F}_s] = \mathbb{E}[N(t) - N(s)] = \lambda(t - s).$$

Hence, $N(t) - \lambda t$ is a martingale, which implies $N(\cdot)$ is a Poisson process.

■

12.1.2. Brownian Motion

A Poisson process has independent and stationary increments, but its sample path is purely discontinous. The Brownian motion has independent and stationary increments, but its sample path is continuous.

Definition 12.3 [Standard Brownian Motion] An adapted and continuous stochastic process $B(\cdot)$ with $B(0) = 0$ is called a standard Brownian motion if

(B1) For $0 \le s < t$, $B(t) - B(s)$ is independent of $\mathcal{F}_s$;

(B2) For $0 \le s < t$, $B(t) - B(s) \overset{d}{\sim} B(t - s)$;

(B3) $B(t) \sim \mathcal{N}(0, t)$ for $t > 0$.

Proposition 12.1 For the standard Brownian motion $B(\cdot)$, let $M(t) = B^2(t) - t$, then $B(\cdot)$ and $M(\cdot)$ are continuous martingales.

Proof. Note that $\mathbb{E}[|B(t)|] < \infty$ since $\mathbb{E}[B^2(t)] = t < \infty$. Moreover,

$$\mathbb{E}[B(t) \mid \mathcal{F}_s] = B(s), \quad 0 \le s < t.$$

Similarly,

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

Thus $M(\cdot)$ is a martingale. ■

Proposition 12.2 For the standard BM $B(\cdot)$,

$$\mathbb{P}\left(\sup_{t\geq 0} B(t) = \infty,\ \inf_{t\geq 0} B(t) = -\infty\right) = 1.$$

Proof. We can see that $\{cB(t/c^2):\ t \geq 0\}$ is the standard BM. Take $Z_s = \sup_{t\geq s} B(t)$, then for $c > 0$,

$$cZ_0 = \sup_{t\geq 0} cB(t) = \sup_{t\geq 0} cB(t/c^2) \sim Z_0.$$

Hence, $Z_0 \in \{0, \infty\}$ with probability 1. Take $p = \mathbb{P}(Z_0 = 0)$, then by $Z_1 - B(1) \sim Z_0$,

$$p \leq \mathbb{P}(B(1) \leq 0, Z_1 \leq 0) \leq \mathbb{P}(B(1) \leq 0, Z_1 - B(1) = 0) = \frac{1}{2}p.$$

Hence, $p = 0$ and $\mathbb{P}(Z_0 = \infty) = 1$. Since $\{-B(t):\ t \geq 0\}$ is also a standard BM, $\mathbb{P}\left(\inf_{t\geq 0} B(t) = -\infty\right) = 1$. The proof is completed. ■

Proposition 12.3 — Quadratic Variation of the standard BM. For the standard BM $B(\cdot)$, let $t_{n,i} = i2^{-n}$ and $n_t = \min\{\ell \geq t2^n\}$,

$$[B]_n^t = \sum_{i=1}^{n_t} (\Delta_i B^t(t_{n,i-1}))^2, \quad t \geq 0, n \geq 0,$$

with $\Delta_i B^t(t_{n,i-1}) = B(t \wedge t_{n,i}) - B(t \wedge t_{n,i-1})$. Then $[B]_n^t \to t$ almost surely as $n \to \infty$.

Proof. We first start to show that $[B]_n^t \to t$ in L^2:

$$[B]_n^t - t = \sum_{i=1}^{n_t} (\Delta_i B^t(t_{n,i-1}))^2 - (t_{n,i} - t_{n,i-1})$$

It follows that

$$\begin{aligned}\mathbb{E}\left[[B]_n^t - t\right]^2 &= \sum_{i=1}^{n_t} (\Delta_i B^t(t_{n,i-1}))^4 - 2\sum_{i=1}^{n_t} (\Delta_i B^t(t_{n,i-1}))^2 (t_{n,i} - t_{n,i-1}) + \sum_{i=1}^{n_t} (t_{n,i} - t_{n,i-1})^2 \\ &= 3\sum_{i=1}^{n_t} (t_{n,i} - t_{n,i-1})^2 - 2\sum_{i=1}^{n_t} (t_{n,i} - t_{n,i-1})^2 + \sum_{i=1}^{n_t} (t_{n,i} - t_{n,i-1})^2 \leq (t2^n + 1)2^{-2n+1}\end{aligned}$$

Now by Marokv indquality,

$$\sum_{n=1}^{\infty} \mathbb{P}(|[B]_n^t - t| > \varepsilon) \le \varepsilon^{-2} 2(t + 2^{-n}) \sum_{n=1}^{\infty} 2^{-n} < \infty.$$

The proof is completed.

■

Recall the variation of $B(\cdot)$ defined as

$$|B|(t) = \sup_{\delta > 0, t_0 = 0} \sup_{t_n - t_{n-1} \ge \delta} \sum_{n=1}^{\infty} 1(t_n \le t) |B(t_n) - B(t_{n-1})|, \qquad \forall t \ge 0.$$

Proposition 12.4 For the standard BM $B(\cdot)$, the variation $|B|(t) = \infty$ with probability 1 for each $t > 0$.

Proof. Assume that $\mathbb{P}(|B|(t) < \infty) > 0$ for some $t \ge 0$, then

$$[B]_n^t \le \sup_{i=1,2,\ldots,n_t} |\Delta_i B^t(t_{n,i-1})| |B|(t)$$

By the uniform continutiy of $B(\cdot)$ on $[0, t]$,

$$\lim_{n \to \infty} \sup_{i=1,2,\ldots,n_t} |\Delta_i B^t(t_{n,i-1})| = 0,$$

which implies $\lim_{n \to \infty} [B]_n^t = 0$ for $\omega \in \Omega$ for which $|B|(t) < \infty$, which contracits to previous proposition. ■

Now we begin to define the stochastic integral as the following. The time evolution of the standard BM $B(\cdot)$ can be charactized by some test function $f \in C_b^1(\mathbb{R})$:

$$f(B(t)) = f(B(0)) + \int_0^t f'(B(u))\, \mathrm{d}B(u).$$

However, since $B(\cdot)$ has infinite variations, it is unclear how we can define the above integral. We need to re-consider the integration by $\mathrm{d}B(u)$. We will do it by slotting in time. Let $s_{i,n}^{(t)} = i2^{-n} \wedge t$ and $\Delta_n f(B(s_{i,n}^{(t)})) = f(B(s_{i+1,n}^{(t)})) - f(B(s_{i,n}^{(t)}))$, then

$$f(B(t)) = f(B(0)) + \sum_{i=0}^{\infty} \Delta_n f(B(s_{i,n}^{(t)})).$$

We apply the Taylor expansion of $f(x)$ at $x = B(s_{i,n}^{(t)})$ with $\delta = \Delta_n B(s_{i,n}^{(t)})$ so that

$$f(x+\delta) = f(x) + f'(\delta)\delta + \frac{1}{2}f''(x)\delta^2 + O(\delta^3).$$

So we have

$$\begin{aligned} f(B(t)) = f(B(0)) &+ \sum_{i=0}^{\infty} f'(B(s_{i,n}^{(t)}))\Delta_n B(s_{i,n}^{(t)}) \\ &+ \frac{1}{2}\int_0^t f''(u)\,\mathrm{d}[B]_n^u + O\left(\max_{0\le i\le n_t} |\Delta_n B(s_{i,n}^{(t)})| \cdot [B]_n^t\right). \end{aligned}$$

Provided that $\int_0^t f'(B(u))\,\mathrm{d}B(u)$ exists, and take $n \to \infty$ on the integral, we obtain

$$f(B(t)) = f(B(0)) + \int_0^t f'(B(u))\,\mathrm{d}B(u) + \frac{1}{2}\int_0^t f''(u)\,\mathrm{d}u.$$

Now we make use of the Ito's integration formula to show $M(\cdot)$ is a martingale implies $B(\cdot)$ is the standard B<. Let $f(x) = e^{i\theta x}$, and we have

$$\mathbb{E}[e^{i\theta B(t)} \mid \mathcal{F}_s] = e^{i\theta B(s)} - \frac{1}{2}\theta^2 \int_s^t \mathbb{E}[e^{i\theta B(u)} \mid \mathcal{F}_s]\,\mathrm{d}u.$$

Differentiaing both sides after dividing by $e^{i\theta B(s)}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[e^{i\theta(B(t)-B(s))} \mid \mathcal{F}_s] = -\frac{1}{2}\theta^2\mathbb{E}[e^{i\theta(B(t)-B(s))} \mid \mathcal{F}_s],$$

which implies that $\mathbb{E}[e^{i\theta(B(t)-B(s))} \mid \mathcal{F}_s] = e^{-\frac{1}{2}\theta^2(t-s)}$. As a result,

$$\begin{aligned} \mathbb{E}\big[e^{i\theta(B(t)-B(s))}\big] &= e^{-\frac{1}{2}\theta^2(t-s)} = \mathbb{E}\big[e^{i\theta(B(t-s)}\big] \\ \mathbb{E}\big[e^{i\theta(B(t)-B(s))} \mid \mathcal{F}_s\big] &= e^{-\frac{1}{2}\theta^2(t-s)} = \mathbb{E}\big[e^{i\theta(B(t-s)}\big] \end{aligned}$$

This completes the claim.

The following martingale is useful in applications:

Proposition 12.5 For the standard BM $B(\cdot)$ and $\theta \in \mathbb{R}$, let $Y(t) = e^{\theta B(t) - \frac{1}{2}\theta^2 t}$, then $Y(\cdot)$ is a martingale.

Proof. Note that

$$\mathbb{E}[Y(t)/Y(s) \mid \mathcal{F}_s] = \mathbb{E}[e^{\theta(B(t)-B(s)) - \frac{1}{2}\theta^2(t-s)}] = 1.$$

As a result,

$$\mathbb{E}[Y(t) \mid \mathcal{F}_s] = Y_s \mathbb{E}[Y(t)/Y(s) \mid \mathcal{F}_s] = Y_s.$$

■

For the standard BM $B(\cdot)$, we consider the hitting time $\tau_x = \inf\{t \geq 0 : \ B(t) = x\}$ fr $x \neq 0$. We first consider the distribution of τ_a for $a > 0$. By the optional sampling theorem,

$$\mathbb{E}[Y^{\tau_a}(\tau_a)] = \mathbb{E}[Y^{\tau_a}(0)] = 1 \implies \mathbb{E}[e^{\theta a - \frac{1}{2}\theta^2 \tau_a}] = 1.$$

Hence, we take $u = \frac{1}{2}\theta^2$, and have

$$\mathbb{E}[e^{-u\tau_a}] = e^{-a\sqrt{2u}}, \quad u > 0.$$

Example 1.3. *A 2-D Random Walk with Reflection*

The reader is refered to the monograph (Fayolle, Malyshev, and Menshikov, 1995) for a general theory of reflected random walks and for bibliographical details.

The state space is $E = \mathbb{N}^2$ and therefore $X_n = (X_n^1, X_n^2)$. The process $\{X_n\}_{n\geq 0}$ evolves according to

$$X_{n+1}^\ell = X_n^\ell + Z_{n+1}^\ell, \tag{1.5}$$

($\ell \in \{1, 2\}$), where Z_{n+1}^ℓ is a *bounded* random variable such that $\{X_n\}_{n\geq 0}$ remains in the first quadrant, and with a conditional distribution given the past of the chain up to time n depending only on the last state X_n. In particular, with $Z_n = (Z_n^1, Z_n^2)$,

$$E[Z_{n+1} \mid X_n = i] = d(i), \tag{1.6}$$

where

$$d(i) = (\alpha(i), \beta(i)). \tag{1.7}$$

The vector $d(i)$ is the *drift* at position $i = (i_1, i_2)$. It will be assumed that

$$(\alpha(i), \beta(i)) = \begin{cases} (\alpha, \beta) & \text{if } i_1 > 0, i_2 > 0 \\ (\alpha_1, \beta_1) & \text{if } i_1 > 0, i_2 = 0 \\ (\alpha_2, \beta_2) & \text{if } i_1 = 0, i_2 > 0 \\ (\alpha_0, \beta_0) & \text{if } i_1 = i_2 = 0 \end{cases} \tag{1.8}$$

Of course, since the chain must remain in the nonnegative quadrant, we have the conditions $\alpha_0 \geq 0$, $\beta_0 \geq 0$, and $\beta_1 \geq 0$, $\alpha_2 \geq 0$. Actually we shall assume slightly more:

$$\beta_1 > 0, \ \alpha_2 > 0, \tag{1.9}$$

just to prevent the chain to be absorbed by a boundary.

It will be shown that for positive recurrence, it suffices to have

$$\alpha < 0, \ \beta < 0, \ \alpha\beta_1 - \alpha_1\beta < 0, \ \beta\alpha_2 - \alpha\beta_2 < 0. \tag{1.10}$$

For this, Foster's theorem will be applied with a quadratic Lyapunov function

$$h(i) = h(i_1, i_2) = \frac{1}{2}(ui_1^2 + vi_2^2) + wi_1i_2, \tag{1.11}$$

for suitable, u, v, w.

First u, v, w will be selected in such a way that h is a positive definite quadratic form. This is guaranteed by

$$u > 0, v > 0, w^2 > uv. \tag{1.12}$$

Next a further condition on u, v, w must be found in order to guarantee that

$$E[h(X_{n+1}) - h(X_n) \mid X_n = i] \leq -\epsilon.$$

Elementary computations yield

$$\begin{aligned} h(X_{n+1}) - h(X_n) = h(X_{n+1} - X_n) &+ [u(X_{n+1}^1 - X_n^1) + w(X_{n+1}^2 - X_n^2)]X_n^1 \\ &+ [v(X_{n+1}^2 - X_n^2) + w(X_{n+1}^1 - X_n^1)]X_n^2. \end{aligned}$$

Therefore for all $i = (i_1, i_2)$

$$\begin{aligned} E[h(X_{n+1}) - h(X_n) \mid X_n = i] = E[h(X_{n+1} - X_n) \mid X_n = i] &+ [u\alpha(i) + w\beta(i)]i_1 \\ &+ [v\beta(i) + w\alpha(i)]i_2. \end{aligned}$$

By the boundedness hypothesis on (Z_{n+1}^1, Z_{n+1}^2), the first term in the right hand side of the above equality is bounded. Therefore, if one can choose u, v, w such that (1.12) is verified and the left-hand side can be made smaller than $-\epsilon$ where ϵ is an arbitrary positive number as long as i_1 and i_2 are sufficiently large, and the proof will be completed.

We therefore have to show the existence of u, v, w such that

(C1) $u\alpha + w\beta < 0$

(C2) $u\alpha_1 + w\beta_1 < 0$

(C3) $v\beta + w\alpha < 0$

(C4) $v\beta_2 + w\alpha_2 < 0$

with the conditions

(C5) $u > 0$

(C6) $v > 0$

(C7) $w^2 > uv$

For this, one has the hypotheses

(h1) $\beta_1 > 0$

(h2) $\alpha_2 > 0$

(h3) $\alpha < 0$

(h4) $\beta < 0$

(h5) $\alpha\beta_1 - \alpha_1\beta < 0$

(h6) $\beta\alpha_2 - \alpha\beta_2 < 0$

Since $\beta < 0$ and $\beta_1 > 0$, (C1) and (C2) are respecitively equivalent to

$$w > -u\alpha/\beta,$$

and

$$-u\alpha_1/\beta_1 > w,$$

that is

(C′1) $-u\alpha/\beta < w < -u\alpha_1/\beta_1$

Similarly, (C3) and (C4) are equivalent to

(C′3) $-v\beta/\alpha < w < -v\beta_2/\alpha_2$

So far the existence of u, v, w verifying (C′1), (C′3), (C5), (C6) and (C7) remains to be proven. But since $\alpha < 0$ and $\beta < 0$, (C′1) and (C′3) imply (C7) provided $u > 0$, $v > 0$, so that we have to find u, v, w verifying (C′1), (C′3) and $u > 0$, $v > 0$. For this we select

$u > 0$ and $v > 0$ arbitrarily, and then observe that in view of (h5) and $\beta < 0$, $\beta_1 > 0$,

$$-u\alpha/\beta < -u\alpha_1/\beta_1,$$

and in view of (h6) and $\alpha < 0$, $\alpha_2 > 0$,

$$-v\beta/\alpha < -v\beta_2/\alpha_2.$$

These two inequalities show that either one of (C′1) and (C′3) is feasible, for any $u > 0$, $v > 0$. They will be simultaneously satisfied by some w if

$$(-u\alpha/\beta, -u\alpha_1/\beta_1) \cap (-v\beta/\alpha, -v\beta_2/\alpha_2) \neq \emptyset.$$

But for this, it suffices to take $u = \beta/\alpha$ and $v = \alpha/\beta$. ◇