# Contents

# Acknowledgments

This book is from the MAT4001 in fall semester, 2018.

CUHK(SZ)

# Notations and Conventions

| | |
|---|---|
| $\mathbb{R}^n$ | $n$-dimensional real space |
| $\mathbb{C}^n$ | $n$-dimensional complex space |
| $\mathbb{R}^{m\times n}$ | set of all $m \times n$ real-valued matrices |
| $\mathbb{C}^{m\times n}$ | set of all $m \times n$ complex-valued matrices |
| $x_i$ | $i$th entry of column vector $\boldsymbol{x}$ |
| $a_{ij}$ | $(i,j)$th entry of matrix $\boldsymbol{A}$ |
| $\boldsymbol{a}_i$ | $i$th column of matrix $\boldsymbol{A}$ |
| $\boldsymbol{a}_i^{\mathrm{T}}$ | $i$th row of matrix $\boldsymbol{A}$ |
| $\mathbb{S}^n$ | set of all $n \times n$ real symmetric matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all $i,j$ |
| $\mathbb{H}^n$ | set of all $n \times n$ complex Hermitian matrices, i.e., $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $\bar{a}_{ij} = a_{ji}$ for all $i,j$ |
| $\boldsymbol{A}^{\mathrm{T}}$ | transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$ means $b_{ji} = a_{ij}$ for all $i,j$ |
| $\boldsymbol{A}^{\mathrm{H}}$ | Hermitian transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all $i,j$ |
| $\mathrm{trace}(\boldsymbol{A})$ | sum of diagonal entries of square matrix $\boldsymbol{A}$ |
| $\mathbf{1}$ | A vector with all 1 entries |
| $\mathbf{0}$ | either a vector of all zeros, or a matrix of all zeros |
| $\boldsymbol{e}_i$ | a unit vector with the nonzero element at the $i$th entry |
| $\mathcal{C}(\boldsymbol{A})$ | the column space of $\boldsymbol{A}$ |
| $\mathcal{R}(\boldsymbol{A})$ | the row space of $\boldsymbol{A}$ |
| $\mathcal{N}(\boldsymbol{A})$ | the null space of $\boldsymbol{A}$ |
| $\mathrm{Proj}_{\mathcal{M}}(\boldsymbol{A})$ | the projection of $\boldsymbol{A}$ onto the set $\mathcal{M}$ |

# Chapter 1

# Week1

## 1.1. Tuesday

### 1.1.1. Difference between ODE and SDE

We first discuss the difference between deterministic differential equations and stochastic ones by considering several real-life problems.

**Problem 1: Population Growth Model.** Consider the first-order ODE

$$\begin{cases} \dfrac{\mathrm{d}N(t)}{\mathrm{d}t} = a(t)N(t) \\ N(0) = N_0 \end{cases}$$

where $N(t)$ denotes the **size** of the population at time $t$; $a(t)$ is the given (deterministic) function describing the **rate** of growth of population at time $t$; and $N_0$ is a given constant.

If $a(t)$ is not completely known, e.g.,

$$a(t) = r(t) \cdot \text{noise}, \text{ or } r(t) + \text{noise},$$

with $r(t)$ being a deterministic function of $t$, and the "noise" term models something random. The question arises: How to *rigorously* describe the "noise" term and solve it?

**Problem 2: Electric Circuit.** Let $Q(t)$ denote the charge at time $t$ in an electrical circuit, which admits the following ODE:

$$\begin{cases} LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = F(t), \\ \qquad Q(0) = Q_0, \quad Q'(0) = Q'_0 \end{cases}$$

where $L$ denotes the inductance, $R$ denotes the resistance, $C$ denotes the capacity, and $F(t)$ denotes the potential source.

Now consider the scenario where $F(t)$ is not completely known, e.g.,

$$F(t) = G(t) + \text{noise}$$

where $G(t)$ is deterministic. The question is how to solve the problem.

R The differential equations above involving non-deterministic coefficients are called the **stochastic differential equations** (SDEs). Clearly, the solution to an SDE should involve the randomness.

### 1.1.2. Applications of SDE

Now we discuss some applications of SDE shown in the finance area.

**Problem 3: Optimal Stopping Problem.** Suppose someone holds an asset (e.g., stock, house). He plans to sell it at some future time. Denote $X(t)$ as the price of the asset at time $t$, satisfying the following dynamics:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = rX(t) + \alpha X(t) \cdot \text{noise}$$

where $r, \alpha$ are given constants. The goal of this person is to maximize the expected selling price:

$$\sup_{\tau \geq 0} \quad \mathbb{E}[X(\tau)]$$

where the optimal solution $\tau^*$ is the optimal stopping time.

**Problem 4: Portfolio Selection Problem.** Suppose a person is interested in two types of assets:

- A risk-free asset which generates a deterministic return $\rho$, whose price $X_1(t)$ follows a deterministic dynamics

$$\frac{\mathrm{d}X_1(t)}{\mathrm{d}t} = \rho X_1(t),$$

- A risky asset whose price $X_2(t)$ satisfies the following SDE:

$$\frac{\mathrm{d}X_2(t)}{\mathrm{d}t} = \mu X_2(t) + \sigma X_2(t) \cdot \text{noise}$$

  where $\mu, \sigma > 0$ are given constants.

The policy of the investment is as follows. The wealth at time $t$ is denoted as $v(t)$. This person decides to invest the fraction $u(t)$ of his wealth into the risky asset, with the remaining $1 - u(t)$ part to be invested into the safe asset. Suppose that the utility function for this person is $U(\cdot)$, and his goal is to maximize the expected total wealth at the terminal time $T$:

$$\max_{u(t), 0 \leq t \leq T} \mathbb{E}[U(v^u(T))]$$

where the decision variable is the portfolio function $u(t)$ along whole horizon $[0, T]$.

**Problem 5: Option Pricing Problem.** The financial derivates are products in the market whose value depends on the underlying asset. The European call option is a typical financial derivative. Suppose that the underlying asset is stock $A$, whose price at time $t$ is $X(t)$. Then the call option gives the option holder the right (not the obligation) to buy one unit of stock A at a specified price (strike price) $K$ at maturity date $T$. The task is to inference the fair price of the option at the current time. The formula for the price of the option is the following:

$$c_0 = \mathbb{E}[(X(T) - K)^+]$$

which is the famous Black-Sholes-Merton Formula.

### 1.1.3. Reviewing for Probability Space

Firstly, we review some basic concepts in real analysis.

**Definition 1.1** [$\sigma$-Algebra] A set $\mathcal{F}$ containing subsets of $\Omega$ is called a $\sigma$-algebra 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 countably union operation, i.e., $A_i \in \mathcal{F}, i \geq 1$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

**Definition 1.2** [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 $\sigma$-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 1.3** [Probability Space] A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ defined as follows:

1. $\Omega$ denotes the **sample space**, and a point $\omega \in \Omega$ is called a sample point;
2. $\mathcal{F}$ is a $\sigma$-algebra of $\Omega$, which is a collection of subsets in $\Omega$. The element $A \in \mathcal{F}$ is called an "event"; and
3. $\mathbb{P}$ is a probability measure defined in the space $(\Omega, \mathcal{F})$.

**Definition 1.4** [Almost Surely True] A statement $S$ is said to be **almost surely (a.s.) true** or **true with probability 1**, if

- $\mathfrak{B} := \{w : S(w) \text{ is true}\} \in \mathcal{F}$
- $\mathbb{P}(F) = 1$.

**Definition 1.5** [Topological Space] 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 $X$.

**Definition 1.6** [Borel $\sigma$-Algebra] Consider a topological space $\Omega$, with $\mathcal{U}$ being the topology of $\Omega$. The **Borel $\sigma$-Algebra** $\mathcal{B}(\Omega)$ on $\Omega$ is defined to be the *minimal* $\sigma$-algebra containing $\mathcal{U}$:

$$\mathcal{B}(\Omega) \triangleq \sigma(\mathcal{U}).$$

Any element $B \in \mathcal{B}(\Omega)$ is called the **Borel set**.

**Definition 1.7** [$\mathcal{F}$-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.

**Definition 1.8** [Generated $\sigma$-Algebra] Suppose $X$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the $\sigma$-algebra generated by $X$, say $\mathcal{H}_X$ is defined to be the **minimal $\sigma$-algebra** on $\Omega$ to make $X$ measurable.

**Proposition 1.1** $\mathcal{H}_X = \{X^{-1}(\boldsymbol{B}) : \boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)\}$.

*Proof.* Since $X$ is $\mathcal{H}_X$-measurable, for any $\boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)$, $X^{-1}(\boldsymbol{B}) \in \mathcal{H}_X$. Thus $\mathcal{H}_X \supseteq \{X^{-1}(\boldsymbol{B}) : \boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)\}$. It suffices to show that $\{X^{-1}(\boldsymbol{B}) : \boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)\}$ is a $\sigma$-algebra to finish the proof, which is true since $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{U})$, with $\mathcal{U}$ being the topology of $\mathbb{R}^n$.

∎

## 1.2. Thursday

**Reviewing for Probability Space.**

- $(\Omega, \mathcal{F}, \mathbb{P})$;
- Random variable;
- Generated $\sigma$-algebra;

### 1.2.1. More on Probability Theory

**Definition 1.9** [Distribution] A probability measure $\mu_X$ on $\mathbb{R}^n$ induced by the random variabe $X$ isdefined as

$$\mu_X(\boldsymbol{B}) = \mathbb{P}(X^{-1}(\boldsymbol{B})),$$

where $\boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)$. The $\mu_X$ is called the **distribution** of $X$.

**Definition 1.10** [Expectation] The expectation of $X$ is given by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

When $\Omega = \mathbb{R}^n$, the expectation can be written in terms of distribution function:

$$\mathbb{E}[X] = \int_{\mathbb{R}^n} y\,\mathrm{d}\mu_X(y)$$

Note that the expectation of the random variable $X$ is well-defined when $X$ is integrable:

**Definition 1.11** [Integrable] The random variable $X$ is **integrable**, if

$$\int_\Omega |X(w)|\,\mathrm{d}\mathbb{P}(w) < \infty.$$

In other words, $X$ is said to be $\mathcal{L}^1$-integrable, denoted as $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

■ **Example 1.1** If $f:\mathbb{R}^n \to \mathbb{R}$ is Borel measurable, and $\int_\Omega |f(X(\omega))|\,\mathrm{d}\mathbb{P}(\omega) < \infty$, then

$$\mathbb{E}[f(x)] = \int_\Omega f(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(y)\,\mathrm{d}\mu_X(y).$$

■

**Definition 1.12** [$L^p$ space] Suppose $X: \Omega \to \mathbb{R}$ is a random variable and $p \geq 1$.

- Define $L^p$-norm of $X$ as

$$\|X\|_p = \left(\int_\Omega |X(\omega)|^p\,\mathrm{d}\mathbb{P}\right)^{1/p}$$

  If $p = \infty$, define

$$\|X\|_\infty = \inf\{N \in \mathbb{R} \mid |X(w)| \leq N, \text{ a.s.}\}$$

- A random variable $X$ is said to be in the $L^p$ space ($p$-th integrable) if

$$\int_\Omega |X(\omega)|^p\,\mathrm{d}\mathbb{P}(\omega) < \infty,$$

  denoted as $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$.

■

**Proposition 1.2** If $p \geq q$, then $\|X\|_p \geq \|X\|_q$. Thus $\mathcal{L}^p(\Omega,\mathcal{F},\mathbb{P}) \subseteq \mathcal{L}^q(\Omega,\mathcal{F},\mathbb{P})$.

*Proof.* The inequality is shown by using Holder's inequality:

$$\|X\|_q^q = \int_\Omega |X|^q\,\mathrm{d}\mathbb{P} \leq \left(\int_\Omega (|X|^q)^{p/q}\,\mathrm{d}\mathbb{P}\right)^{q/p} = \left(\int_\Omega |X|^p\,\mathrm{d}\mathbb{P}\right)^{\frac{1}{p}\cdot q} = \|X\|_p^q.$$

■

Then we discuss how to define independence between two random variables, by the following three steps:

**Definition 1.13** [Independence]

1. Two events $A_1, A_2 \in \mathcal{F}$ are said to be **independent** if $\mathbb{P}(A_1 \bigcap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$.
2. Two $\sigma$-algebras $\mathcal{F}_1, \mathcal{F}_2$ are said to be **independent** if $F_1, F_2$ are independent events for $\forall F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2$
3. Two random variables $X, Y$ are said to be **independent** if $\mathcal{H}_X, \mathcal{H}_Y$, the $\sigma$-algebra generated by $X$ and $Y$, respectively, are independent.

R The independence defined above can be generalized from two events into finite number of events.

**Proposition 1.3** If $X$ and $Y$ are two independent random variables, and $\mathbb{E}[|X|] < \infty, \mathbb{E}[|Y|] < \infty$, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] < \infty.$$

*Proof.* The first step is to simplify the probability distribution for the product random variable $(X, Y)$, i.e., $\mu_{X,Y}$.

R From now on, we also write the event $\{X^{-1}(\boldsymbol{B})\}$ as $\{X \in \boldsymbol{B}\}$ for $\boldsymbol{B} \in \mathcal{B}(\mathbb{R}^n)$.

By the definition of independence, we have the following:

$$\begin{aligned}\mu_{X,Y}(A_1 \times A_2) &\triangleq \mathbb{P}(\{(X,Y) \in (A_1 \times A_2)\}) = \mathbb{P}(\{X \in A_1, Y \in A_2\}) \\ &= \mathbb{P}(\{X \in A_1\})\mathbb{P}(\{Y \in A_2\}) = \mu_X(A_1)\mu_Y(A_2).\end{aligned}$$

Now we begin to simplify the expectation of product:

$$\begin{aligned}\mathbb{E}[XY] &= \int xy \, \mathrm{d}\mu_{X,Y}(x,y) = \iint xy \, \mathrm{d}\mu_X(x)\mu_Y(y) \\ &= \int y \left[\int x \, \mathrm{d}\mu_X(x)\right] \mu_Y(y) = \int \mathbb{E}[X] y \mu_Y(y) = \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

■

### 1.2.2. Stochastic Process

Consider a set $T$ of time index, e.g., a non-negative integer set or a time interval $[0,\infty)$. We will discuss a discrete/continuous time stochastic process.

**Definition 1.14** [Stochastic Process] A collection of random variables $\{X_t\}_{t\in T}$, defined on $(\Omega,\mathcal{F},\mathbb{P})$ and taking values in $\mathbb{R}^n$, is called a **stochastic process**. ■

R A stochastic process $\{X_t\}_{t\in T}$ can also be viewed as a random function, since it is a mapping $\Omega\times T\to\mathbb{R}^n$. Sometimes we omit the subscript to denote a stochastic process $\{X_t\}$.

**Definition 1.15** [Sample Path] Fixing $\omega\in\Omega$, then $\{X_t(\omega)\}_{t\in T}$ (denoted as $X_\cdot(\omega)$) is called a **sample path**, or **trajectory**. ■

**Definition 1.16** [Continuous] A stochastic process $\{X_t\}$ is said to be **continuous** (right-cot, left-cot, resp.) a.s., if $t\to X_t(\omega)$ is **continuous** (right-cot, left-cot, resp.) a.s., i.e.,

$$\mathbb{P}\left(\{\omega : t\to X_t(\omega) \text{ is continuous (right-cot, left-cot, resp.)}\}\right)=1.$$

■

■ **Example 1.2** [Poisson Process] Consider $(\xi_j, j=1,2,\dots)$ a sequence of i.i.d. random variables with Possion distribution with intensity $\lambda>0$. Let $T_0=0$, and $T_n=\sum_{j=1}^n\xi_j$. Define $X_t=n$ if $T_n\le t<T_{n+1}$. Verify that $\{X_t\}$ is a stochastic process with right-continuity and left-limit exists. Instead of giving a mathematical proof, we provide a numerical simulation of $\{X_t\}$ plotted in Figure. 1.1. [a] ■

[a] The corresponding matlab code can be found in

https://github.com/WalterBabyRudin/Courseware/tree/master/MAT4500/week1

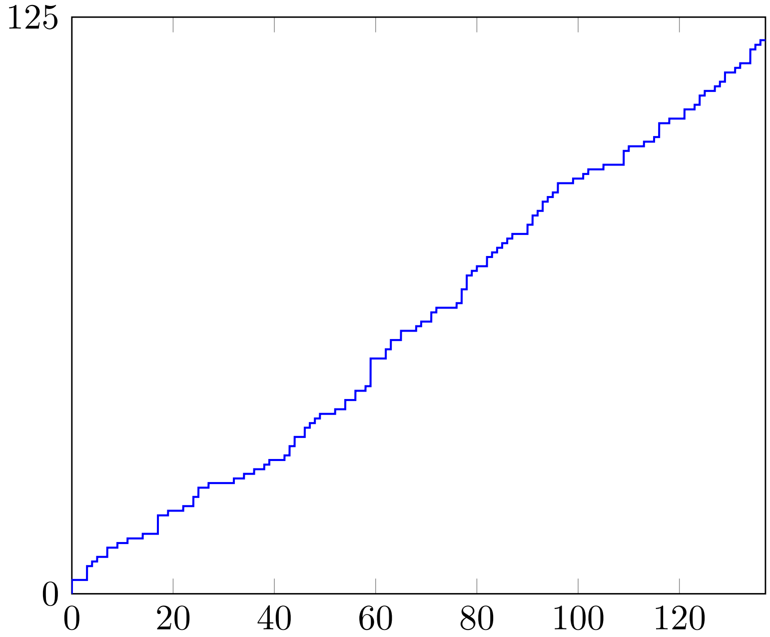


Figure 1.1: One simulation of $\{X_t\}$ with intensity $\lambda = 1.2$ and 500 samples

# Chapter 2

# Week2

## 2.1. Tuesday

### 2.1.1. More on Stochastic Process

For simplicity of notation, we write

$$\{X \in F\} \triangleq \{\omega : X(\omega) \in F\} = X^{-1}(F).$$

**Definition 2.1** [Joint Distribution of a Stochastic Process] Let $\{X_t\}$ be a stochastic process. Let $0 = t_0 \le t_1 \le \cdots \le t_k$. The joint distribution of random variables $X_{t_1}, \dots, X_{t_k}$ is defined as

$$\mu_{t_1,t_2,\dots,t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}(X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k),$$

where $F_1, \dots, F_k$ are all Borel sets in $\mathbb{R}^n$. ■

R The measure $\mu_{t_1,t_2,\dots,t_k}$ is the **finite-dimensional distribution**. In particular, $\mu_{t_1,t_2,\dots,t_k}$ is a probability measure on the product space $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$.

■ **Example 2.1** [Brownian Motion] Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the function

$$\mathsf{P}(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right), \qquad x,y \in \mathbb{R}, t > 0$$

The Brownian motion [a] is denoted by $\{B_t\}_{t \ge 0}$. Then the joint distribution of $\{B_t\}$ at time

$t_1, t_2, \ldots, t_k$ is given by:

$$\mathbb{P}(B_{t_1} \in F_1, \ldots, B_{t_k} \in F_k) =$$
$$\int_{F_1 \times \cdots \times F_k} \mathsf{P}(t_1, 0, x_1)\mathsf{P}(t_2 - t_1, x_1, x_2) \cdots \mathsf{P}(t_l - t_{k-1}, x_{k-1}, x_k)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_k$$

[a] now consider only the Brownian motion with independent, normally distributed increment.

**Definition 2.2** [Measurable Set] Let $(S, \mathcal{F})$ be a pair, with $S$ being a set and $\mathcal{F}$ is a $\sigma$-algebra on $S$. Then the set $\mathcal{F}$ is called a **measurable space**, and an element of $\mathcal{F}$ is called a $\mathcal{F}$-**measurable** subset of $S$.

R Cosnider a stochastic process $\{X_t\}$ in continuous time, e.g., a Brownian motion. Consider the space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$, and define the collection of outcomes

$$F = \{\omega \in \Omega \mid X_t(\omega) \in [0,1], \forall t \leq 1\}$$

The issue is that this event $F$ is not necessarily $\mathcal{B}(\mathbb{R}^{[0,\infty)})$-measurable. Sometimes we need some extra conditions on the stochastic process to make $F$ measurable. The significance of $F$ will also discussed in the future.

**Proposition 2.1** Suppose that $\{X_t\}$ is a continuous-time stochastic process. Let $\mathcal{T}$ be a countable subset of $[0, \infty)$, then given $B \in \mathcal{B}(\mathbb{R}^n)$,

- The set $\{\omega : X_t(\omega) \in B \text{ for any } t \in \mathcal{T}\}$ is measurable;
- The function $h = \sup_{t \in \mathcal{T}} |X_t|$ is $\mathcal{F}$-measurable.

*Proof.* For fixed $t \in \mathcal{T}$, because of the $\mathcal{F}$-measurability of $X_t$, the set

$$\{X_t \in B\} := \{\omega : X_t(\omega) \in B\} \text{ is measurable.}$$

It is easy to see that the countably intersection $\cap_{t \in \mathcal{T}} \{X_t \in B\}$ is measurable as well. For the second assertion, it suffices to check that $h^{-1}([-\infty, a)) = \cap_{t \in \mathcal{T}} \{X_t < a\}$ is measurable. ∎

However, when $\mathcal{T}$ is uncountable, it is problematic to show the measurability. It is even difficult to show that for almost all $\omega$, $t \mapsto X_t(\omega)$ is continuous. In order to obtain a "continuous" process, we need the following important concept:

**Definition 2.3** [Equivalent random variables] Let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be two stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\{Y_t\}$ is called an **equivalent** (a **version**) of $\{X_t\}$ if

$$\mathbb{P}(\{w \mid X_t(\omega) = Y_t(\omega)\}) = 1, \qquad \text{for any time } t.$$

R It is easy to see that when $\{X_t\}_{t\geq 0}$ is a version of $\{Y_t\}_{t\geq 0}$, they have the same finite-dimensional distributions, but their path properties may be different, e.g., for almost all $\omega$, $t \mapsto X_t(\omega)$ may be continuous while $t \mapsto Y_t(\omega)$ may not.

### 2.1.2. Conditional Expectation

**Definition 2.4** [Conditional Expectation] Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $\mathcal{G}$ is a sub $\sigma$-algebra of $\mathcal{F}$, i.e., $\mathcal{G} \subseteq \mathcal{F}$. Let $X : \Omega \to \mathbb{R}^n$ be an *integrable* random variable, and the **conditional expectation** $X$ given $\mathcal{G}$, denoted as $\mathbb{E}[X \mid \mathcal{G}]$, is a random variable satisfying the following conditions:

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

$$\int_A \mathbb{E}[X \mid \mathcal{G}] \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P}$$

   In other words,

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

R Let $X$ be an integrable random variable. Then for each sub $\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ exists and is unique up to $\mathcal{V}$-measurable sets of probability zero. The proof is based on the Radon-Nikodym theorem.

In other words, suppose that $Y$ is another random variable satisfying the condition mentioned in Definition 2.4, i.e.,

- $Y$ is $\mathcal{G}$-measurable;
- $\mathbb{E}[Y1_A] = \mathbb{E}[X1_A]$ for any $A \in \mathcal{G}$;

then we can assert that $Y = \mathbb{E}[X \mid \mathcal{G}]$ a.s., and $Y$ is called a **version** of $\mathbb{E}[X \mid \mathcal{G}]$.

Conditional expectation has many of the same properties that ordinary expectation does:

**Theorem 2.1** — **Properties of Conditional Expectation.** Let $X$ be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}$ is a sub $\sigma$-algebra of $\mathcal{F}$, then the following holds:

1. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$
2. If $X$ is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$ a.s..
3. (Linearity) For any $a_1, a_2 \in \mathbb{R}$,
$$\mathbb{E}[a_1 X_1 + a_2 X_2 \mid \mathcal{G}] = a_1 \mathbb{E}[X_1 \mid \mathcal{G}] + a_2 \mathbb{E}[X_2 \mid \mathcal{G}], \quad \text{a.s.}$$
4. (Positivity) If $X \geq 0$, then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$.
5. (Jensen Inequality) If $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function, then
$$\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \phi(\mathbb{E}[X \mid \mathcal{G}]).$$
6. (Tower Property) Let $\mathcal{H}$ be a sub $\sigma$-algebra of $\mathcal{G}$. Then
$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right] = \mathbb{E}[X \mid \mathcal{H}], \quad \text{a.s.}$$
7. (Conditional Independence) Suppose that $\mathcal{H}$ is a $\sigma$-algebra independent of $\sigma(\sigma(X), \mathcal{G})$, then
$$\mathbb{E}\left[X \middle| \sigma(\mathcal{G}, \mathcal{H})\right] = \mathbb{E}[X \mid \mathcal{G}].$$
In particular, $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$ if $\mathcal{H}$ is independent of $X$.

*Proof.* 1. Recall the definition of $\mathbb{E}[X \mid \mathcal{G}]$ and take $A = \Omega$,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1_{\Omega}] = \mathbb{E}[X].$$

2. It suffices to verify that $X$ satisfies 1) and 2) in Definition 2.4, and the result holds by the uniqueness of conditional expectation.
3. Again, verify the RHS satisfies 1) and 2) in Definition 2.4, and the result holds by the uniqueness of conditional expectation.
4. For fixed $\omega \in \Omega$,

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1_{\{\omega\}}] = \mathbb{E}[X1_{\{\omega\}}] = X(\omega) \geq 0.$$

5. • Assume that we can construct a collection of affine functions $\mathcal{L} = \{L(x) : L(x) = ax + b\}$, such that $\phi(x) = \sup_{L \in \mathcal{L}} L(x)$. As a result, for any $L \in \mathcal{L}$,

$$\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \mathbb{E}[L(X) \mid \mathcal{G}] = L(\mathbb{E}[X \mid \mathcal{G}])$$

Taking the supermum over all $L \in \mathcal{L}$, the desired result holds.

• Here we give an explicit construction of $\mathcal{L}$:

$$\mathcal{L} = \{x \mapsto \phi(x_0) + g^{\mathrm{T}}(x - x_0) \mid x_0 \in \mathrm{dom}(\phi), g \in \partial\phi(x_0)\}$$

Note that $L(x) \leq \phi(x)$ for any $L \in \mathcal{L}$ since the subgradient inequality holds for convex functions. Reversely, $[\phi(x_0) + g^{\mathrm{T}}(x - x_0)] \mid_{x=x_0} = \phi(x_0)$. Therefore, $\phi(x) = \sup_{L \in \mathcal{L}} L(x)$.

6. It suffices to show that $\mathbb{E}[X \mid \mathcal{H}]$ is a version of $\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right]$. The key is to show that for all $A \in \mathcal{H}$,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]1_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1_A].$$

Verify that both sides equal to $\mathbb{E}[X1_A]$.

7. It suffices to show that $\mathbb{E}[X \mid \mathcal{G}]$ is a version of $\mathbb{E}\left[X \middle| \sigma(\mathcal{G},\mathcal{H})\right]$, i.e., for any $A \in \sigma(\mathcal{G},\mathcal{H})$,
$$\mathbb{E}[X1_A] = \mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}]1_A\Big].$$

■

### 2.1.3. Tips about Probability Theory

Suppose that $\{E_n\}$ is a seuqence of events. We aim to define the limit of this sequence. A key issue is that two sets may loss orders. For instance, it is possible that neither $A \subseteq B$ nor $B \subseteq A$. Therefore, based on a seuqence of events, we first define monotone increasing/decreasing sequence of events as follows:

$$\overline{E}_m = \bigcup_{n \ge m} E_n, \qquad \underline{E}_m = \bigcap_{n \ge m} E_n$$

Then $\{\overline{E}_m\}$ and $\{\underline{E}_m\}$ are montone decreasing/increasing, and it is easy to define their limits:

$$\limsup_{n\to\infty} E_n = \cap_m \overline{E}_m, \qquad \liminf_{n\to\infty} E_n = \cup_m \underline{E}_m.$$

According to this definition, we have:

$$\limsup_{n\to\infty} E_n \triangleq \{\omega : \quad \omega \in E_n \text{ for infinitely many } n\}$$
$$\liminf_{n\to\infty} E_n \triangleq \{\omega : \quad \omega \in E_n \text{ for all large enough } n\}$$

**Theorem 2.2** — **Borel-Cantelli Lemma.** If $\{E_n\}$ is a sequence of events satisfying $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$, then
$$\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0.$$

*Proof.* Define $\overline{E}_m$ as above, and thus $\limsup_{n\to\infty} E_n = \cap_m \overline{E}_m$. As a result, for any $m$,

$$\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = \mathbb{P}(\cap_m \overline{E}_m) \le \mathbb{P}(\overline{E}_m) \le \sum_{n=m}^{\infty} \mathbb{P}(E_n).$$

Because of the condition $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$, as $m \to \infty$,

$$\sum_{n=m}^{\infty} \mathbb{P}(E_n) \to 0 \implies \mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0.$$

■

### 2.1.4. Reviewing on Real Analysis

**Theorem 2.3** — **Monotone Convergence Theorem.** Let $\{f_n\}$ be a sequence of non-negative measurable functions on $(S, \Sigma, \mu)$ satisfying

- $f_1(x) \le f_2(x) \le \cdots$ for almost all $x \in S$;
- $f_n(x) \to f(x)$ for almost all $x \in S$, for some measurable function $f$.

Then

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

The proof for the monotone convergence theorem (MCT) can be found in the website

> Daniel Wong, Jie Wang. (2019) Lecture Notes for MAT3006: Real Analysis, Lecture 21. Available at the link
>
> https://walterbabyrudin.github.io/information/Updates/Updates.html

We can apply MCT to show the Fatou's lemma, in which the required condition is weaker:

**Theorem 2.4** — **Fatou's Lemma.** Suppose that $\{f_n\}$ is a sequence of measurable, non-negative functions. Then

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

*Proof.* Define the function $g_n = \inf_{k\ge n} f_k$. Then $\{g_n\}$ is a non-decreasing sequence of non-negative functions. Then

$$\begin{aligned}\int \liminf_{n\to\infty} f_n \,\mathrm{d}\mu = \int \lim_{n\to\infty} g_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu \\ = \liminf_{n\to\infty} \int g_n \,\mathrm{d}\mu \\ \le \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu\end{aligned}$$

where the second equality is by MCT, and the last equality is because that $g_n \le f_n, \forall n$. ∎

■ **Example 2.2** In general the integral of the limit-inf on a sequence of functions is smaller. For instance, consider a sequence of functions on $\mathbb{R}$:

$$f_n(x) = \begin{cases} \mathbf{1}_{[0,1/2]}(x), & \text{when } n \text{ is odd} \\ \mathbf{1}_{[1/2,1]}(x), & \text{when } n \text{ is even} \end{cases}$$

Then

$$\liminf_{n\to\infty} f_n = 1_{\{1/2\}} \implies \int \liminf_{n\to\infty} f_n \,\mathrm{d}m = 0,$$

while $\int_{[0,1]} f_n \,\mathrm{d}m = 1/2$ for each $n$. ■

R We also have the reversed fatou's lemma, saying that in general the integral of the limit-sup on a sequence of functions is bigger:

$$\int_S \limsup_{n\to\infty} f_n \,\mathrm{d}\mu \ge \limsup_{n\to\infty} \int_S f_n \,\mathrm{d}\mu.$$

**Theorem 2.5** — **Dominated Convergence Theorem.** Let $\{f_n\}$ be a sequence of measurable functions on $(S, \Sigma, \mu)$ satisfying

1. $f_n$ is dominated by an integrable function $g$, i.e.,

$$|f_n(x)| \le g(x)$$

for almost all $x \in S$, with $\int_S |g| \, d\mu < \infty$.

2. $f_n$ converges to $f$ almost everywhere for some measurable function $f$.

Then $f$ is integrable and $f_n \to f$ in $L^1$, i.e., $\lim_{n\to\infty} \int_S |f_n - f| \, d\mu = 0$, which implies that

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

*Proof.*

- The integrability of $f$ is because that $|f| \leq g$ a.e.;
- The $L^1$-convergence for $f_n$ is by the reversed fatou's lemma:

$$\limsup \int |f_n - f| \, dm \leq \int \limsup |f_n - f| \, dm = 0.$$

- The remaining part is by applying Fatou's lemma on a sequence of functions $\{g + f_n\}$ and $\{g - f_n\}$. The details are in the reference

  Daniel Wong, Jie Wang. (2019) Lecture Notes for MAT3006: Real Analysis, Lecture 23. Available at the link

  https://walterbabyrudin.github.io/information/Updates/Updates.html

■

## 2.2. Thursday

### 2.2.1. Uniform Integrability

In this lecture, we discuss the uniform integrability, which is an useful tool to handle the convergence of random variables in $L^1$.

**Definition 2.5** [$L_1$-convergence] Given a sequence of functions $\{f_n\}$, we say $f_n \to f$ in $L^1$ if

$$\lim_{n\to\infty} \int_S |f_n - f| \, \mathrm{d}\mu = 0.$$

**Proposition 2.2** Suppose that a random variable $X$ is integrable, denoted as $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $F \in \mathcal{F}$ with $\mathbb{P}(F) < \delta$, we have

$$\mathbb{E}[|X|; F] \triangleq \mathbb{E}[|X| 1_F] = \int_F |X| \, \mathrm{d}\mathbb{P} < \varepsilon$$

*Proof.* Suppose on the contrary that there exists some $\varepsilon_0 > 0$, and a sequence of events $\{F_n\}$ with each $F_n \in \mathcal{F}$ such that

$$\mathbb{P}(F_n) < \frac{1}{2^n}, \qquad \text{but } \mathbb{E}[|X|; F_n] \ge \varepsilon_0.$$

As a result, $\sum_{n=1}^{\infty} \mathbb{P}(F_n) < \infty$. By applying theorem 2.2,

$$\mathbb{P}(H) = 0, \qquad \text{where } H \triangleq \limsup_{n\to\infty} F_n.$$

On the other hand, by the reversed Fatou's lemma,

$$\mathbb{E}[|X|; H] = \int |X| 1_H \, \mathrm{d}\mathbb{P} \ge \limsup_{n\to\infty} \int |X| 1_{F_n} \, \mathrm{d}\mathbb{P} = \limsup_{n\to\infty} \mathbb{E}[|X|; F_n] \ge \varepsilon_0$$

which contradicts to the fact that $\mathbb{P}(H) = 0$. ■

**Corollary 2.1** Suppose that $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then for any $\varepsilon > 0$, there exists $K > 0$, such that

$$\mathbb{E}[|X|;|X| > K] := \int_{\{|X|>K\}} |X| \, d\mathbb{P} < \varepsilon.$$

*Proof.* The idea is to construct $K$ such that $\{|X| > K\}$ happens with small probability.

- Firstly we have the Markov inequality $\mathbb{P}(\{|X| > K\}) \le \frac{1}{K}\mathbb{E}[|X|]$, since the following inequality holds:

$$\begin{aligned}\mathbb{E}[|X|] &= \mathbb{E}[|X|;|X| > K] + \mathbb{E}[|X|;|X| \le K] \\ &\ge \mathbb{E}[K;|X| > K] = K\mathbb{E}[1_{|X|>K}] = K\mathbb{P}(|X| > K)\end{aligned}$$

- Applying Proposition (3.1), we choose $K$ large enough such that $\frac{\mathbb{E}|X|}{K} < \delta$, which implies $\mathbb{P}(|X| > K) < \delta$. The desired result follows immediately.

■

**Definition 2.6** A collection $\mathcal{C}$ of random variables are said to be **uniform integrable** if and only if for any given $\varepsilon > 0$, there exists **a** $K \ge 0$ such that

$$\mathbb{E}[|X|;|X| > K] < \varepsilon, \qquad \forall X \in \mathcal{C}.$$

R An uniform integrable (UI) class $\mathcal{C}$ is also $L^1$-bounded:

*Proof.* Choose $\varepsilon = 1$, then there exists $K > 0$ such that for any $X \in \mathcal{C}$,

$$\mathbb{E}[|X|] = \mathbb{E}[|X|;|X| > K] + \mathbb{E}[|X|;X \le K] \le \varepsilon + K = 1 + K,$$

■

However, the converse of this statement is not necessarily true. See Example 2.3 for a counter-example.

■ **Example 2.3** Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \text{Leb})$, and the collection $\mathcal{C} = \{X_n\}$, with $X_n = n \cdot 1_{E_n}$ and $E_n = (0, 1/n)$.

- It is easy to show that $\mathbb{E}[X_n] = 1, \forall n$, which means that $\mathcal{C}$ is $L^1$-bounded.
- However, $\mathcal{C}$ is not UI. Take $\varepsilon = 1$, and for any $K > 0$, as long as $n > K$,

$$\mathbb{E}[|X_n|; |X_n| > K] = 1$$

- Moreover, $L^1$-boundedness does not mean $L^1$-convergence. Observe that $X_n \to 0$ a.s., but

$$\int |X_n - 0| \, \mathrm{d}\mathbb{P} = 1, \quad \forall n.$$

■

Although $L^1$-boundedness does not imply UI, the $L^p$-boundness for $p > 1$ does.

**Theorem 2.6** Let $p > 1$. Suppose that a class $\mathcal{C}$ of random variables are uniformly bounded in $L^p$, i.e.,

$$\mathbb{E}[|X|^p] = \int_\Omega |X|^p \, \mathrm{d}\mathbb{P} < M < \infty, \quad \forall X \in \mathcal{C},$$

where $M$ is some finite constant. Then the class $\mathcal{C}$ is uniformly integrable (UI).

*Proof.* Choose some $K > 0$, the idea is to bound the term $\mathbb{E}[|X|; |X| > K]$, for any $X \in \mathcal{C}$:

$$\begin{aligned}
\int_{\{|X|>K\}} |X| \, \mathrm{d}\mathbb{P} &= \int_{\{|X|>K\}} \frac{|X|^p}{|X|^{p-1}} \, \mathrm{d}\mathbb{P} \\
&\le \int_{\{|X|>K\}} \frac{|X|^p}{K^{p-1}} \, \mathrm{d}\mathbb{P} = \frac{1}{K^{p-1}} \int_{\{|X|>K\}} |X|^p \, \mathrm{d}\mathbb{P} \\
&\le \frac{1}{K^{p-1}} \int_\Omega |X|^p \, \mathrm{d}\mathbb{P} \\
&\le \frac{M}{K^{p-1}}.
\end{aligned}$$

where the last inequality is by the $L^p$-boundedness. Therefore, for any given $\varepsilon > 0$, the desired result holds by choosing $K$ large enough such that $\frac{M}{K^{p-1}} \leq \varepsilon$.

■

The uniform integrability also has the dominance property:

**Theorem 2.7** Suppose that a class $\mathcal{C}$ of random variables are dominated by an integrable random variable $Y$, i.e., $\forall X \in \mathcal{C}$,

$$|X(\omega)| \leq Y(\omega), \quad \forall \omega \in \Omega, \ \mathbb{E}|Y| < \infty,$$

then the class $\mathcal{C}$ is UI.

*Proof.* The idea is to bound the term $\mathbb{E}[|X|; \ |X| > K]$ to show the UI:

$$\int_{\{|X|>K\}} |X| \, d\mathbb{P} \leq \int_{\{|Y|>K\}} |X| \, d\mathbb{P} \leq \int_{\{|Y|>K\}} |Y| \, d\mathbb{P}$$

where the first inequality is because that $\{|X| > K\} \subseteq \{|Y| > K\}$, and the second is because that $|X| < Y$. The desired result holds by applying Corollary 3.1 such that

$$\int_{\{|Y|>K\}} |Y| \, d\mathbb{P} < \varepsilon.$$

■

# Chapter 3

# Week3

## 3.1. Tuesday

### 3.1.1. Reviewing

**Definition 3.1** For $p \geq 1$, we say a random variable $X \in \mathcal{L}^p$ if

$$\|X\|_p^p \triangleq \mathbb{E}[|X|^p] < \infty.$$

Particularly, when $X \in \mathcal{L}^1$, the random variable $X$ is said to be **integrable**.

A useful propoerty of integrability is the following:

**Proposition 3.1** Suppose that a random variable $X$ is integrable, then for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $F \in \mathcal{F}$ with $\mathbb{P}(F) < \delta$, we have

$$\mathbb{E}[|X|;F] \triangleq \mathbb{E}[|X|1_F] = \int_F |X|\,\mathrm{d}\mathbb{P} < \varepsilon$$

Since $\{|X| > K\}$ happens with small probability, we have the following corollary:

**Corollary 3.1** Suppose that $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then for any $\varepsilon > 0$, there exists $K > 0$, such that

$$\mathbb{E}[|X|;|X| > K] := \int_{\{|X|>K\}} |X|\,\mathrm{d}\mathbb{P} < \varepsilon.$$

**Definition 3.2** Consider a collcetion of random variables instead, denoted as $\mathcal{C}$:

- $\mathcal{C}$ is said to be $L^p$-**bounded** if there exists **a** finite $M$ such that

$$\mathbb{E}[|X|^p] < M, \quad \forall X \in \mathcal{C}.$$

- $\mathcal{C}$ is said to be **uniformly integrable** if any given $\varepsilon > 0$, there exists **a** $K \geq 0$ such that

$$\mathbb{E}[|X|1_{\{|X|>K\}}] < \varepsilon, \qquad \forall X \in \mathcal{C}.$$

R UI implies $L^1$-boundedness: Try to upper bound $\mathbb{E}[|X|]$. However, the converse is not true: One counter-example is $\mathcal{C} = \{X_n\}_n$ with $X_n = n \cdot 1_{(0,1/n)}$.

**Proposition 3.2**

- $L^p$-boundedness for $p > 1$ will imply UI;
- The class of random variables dominated by an integrable random variable is UI.

Recall the proof stated in Theorem 2.6 and Theorem 2.7 in detail.

*Proof Outline.*

1. The first statement is by applying the $L^p$-boundedness on the following formula:

$$\mathbb{E}[|X|1_{\{|X|>K\}}] = \int_{\{|X|>K\}} |X| \, \mathrm{d}\mathbb{P} \leq \frac{1}{K^{p-1}} \int_{\{|X|>K\}} |X|^p \, \mathrm{d}\mathbb{P}.$$

2. Firstly show that

$$\mathbb{E}[|X|1_{\{|X|>K\}}] = \int_{\{|X|>K\}} |X| \, \mathrm{d}\mathbb{P} \leq \int_{\{|Y|>K\}} |Y| \, \mathrm{d}\mathbb{P}.$$

Apply Corollary 3.1 concludes the proof.

■

### 3.1.2. Necessary and Sufficient Conditions for UI

Our first result is about sufficient conditions for the UI on a collection of conditional expectations:

**Theorem 3.1** Suppose that $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{G}_\alpha\}_{\alpha \in \mathcal{A}}$ is a collection of $\sigma$-algebras such that $\mathcal{G}_\alpha \subseteq \mathcal{F}$. Then the collection of random variables

$$\mathcal{C} = \left\{ \mathbb{E}[X \mid \mathcal{G}_\alpha] : \ \alpha \in \mathcal{A} \right\}$$

is uniformly integrable.

*Proof.*

- Apply proposition 3.1 on $X$: For given $\varepsilon > 0$, there exists $\delta > 0$ such that when $\mathbb{P}(F) < \delta$ with $F \in \mathcal{F}$, $\mathbb{E}[|X| \cdot 1_F] < \varepsilon$.

- Define $Y_\alpha = \mathbb{E}[X \mid \mathcal{G}_\alpha]$. By Jensen's inequality, $|Y_\alpha| \le \mathbb{E}[|X| \mid \mathcal{G}_\alpha]$, which motivates us to upper bound the following integral:

$$\begin{aligned}
\mathbb{E}\left[ |\mathbb{E}[X \mid \mathcal{G}_\alpha]|; \ |\mathbb{E}[X \mid \mathcal{G}_\alpha]| > K \right] &= \int_{\{|Y_\alpha| > K\}} |Y_\alpha| \, \mathrm{d}\mathbb{P} \\
&\le \int_{\{|Y_\alpha| > K\}} \mathbb{E}[|X| \mid \mathcal{G}_\alpha] \, \mathrm{d}\mathbb{P} \\
&\le \int_{\{\mathbb{E}[|X||\mathcal{G}_\alpha] > K\}} \mathbb{E}[|X| \mid \mathcal{G}_\alpha] \, \mathrm{d}\mathbb{P} \\
&= \int_{\{\mathbb{E}[|X||\mathcal{G}_\alpha] > K\}} |X| \, \mathrm{d}\mathbb{P}
\end{aligned}$$

  where the last equality is because of the definition for conditional expectation and that $\{\mathbb{E}[|X| \mid \mathcal{G}_\alpha] > K\} \in \mathcal{G}_\alpha$.

- Then consider upper bounding $\mathbb{P}\{\mathbb{E}[|X| \mid \mathcal{G}_\alpha] > K\}$ using Markov inequality:

$$\mathbb{P}\{\mathbb{E}[|X| \mid \mathcal{G}_\alpha] > K\} \le \frac{\mathbb{E}[\mathbb{E}[|X| \mid \mathcal{G}_\alpha]]}{K} = \frac{\mathbb{E}[|X|]}{K},$$

  where the equality is by the tower property of conditional expectation. Here we choose $K$ such that $\frac{\mathbb{E}[|X|]}{K} < \delta$, which implies $\mathbb{P}\{\mathbb{E}[|X| \mid \mathcal{G}_\alpha] > K\} < \delta$. By applying the result on the first part, we have

$$\mathbb{E}\left[|\mathbb{E}[X \mid \mathcal{G}_\alpha]|;\ |\mathbb{E}[X \mid \mathcal{G}_\alpha]| > K\right] \le \int_{\{\mathbb{E}[|X||\mathcal{G}_\alpha]>K\}} |X| \,\mathrm{d}\mathbb{P} \le \varepsilon.$$

■

R A class $\mathcal{C}$ of random variables is uniformly integrable if and only if

$$\lim_{k\to\infty} \sup_{X\in\mathcal{C}} \int_{\{|X|>K\}} |X| \,\mathrm{d}\mathbb{P} = 0.$$

### 3.1.3. Convergence of random variables

In the following part we study several convergence versions shown in probability theory.

**Definition 3.3** [Convergence in probability] Let $\{X_n\}$ be a sequence of random variables.

- We call $\{X_n\}$ converges to a random variable $X$ in probability, denoted as $X_n \to X$ in prob., if for any $\varepsilon > 0$,

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

- We call $\{X_n\}$ converges to a random variable $X$ a.s., denoted as $X_n \to X$ a.s.., if

$$\mathbb{P}\Big(\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

- We call $\{X_n\}$ converges to a random variable $X$ in $L^1$, denoted as $X_n \to X$ in $L^1$, if

$$\lim_{n\to\infty} \|X_n - X\|_1 = 0.$$

R

- $X_n \to X$ a.s. implies $X_n \to X$ in prob.;
- $X_n \to X$ in $L^1$ implies $X_n \to X$ in prob.;
- A natural question is what is the connection between convergence a.s. and convergence in $L^1$. The dominated convergence theorem provides the following characterization:

$$\left.\begin{matrix} X_n \xrightarrow{\text{a.s.}} X \\ |X_n| < Y \\ \mathrm{E}(Y) < \infty \end{matrix}\right\} \quad \Rightarrow \quad X_n \xrightarrow{L^1} X$$

Then we provide sufficient conditions for convergence in probability to imply convergence in $L^1$:

**Theorem 3.2 — Bounded Convergence Theorem.** Let $\{X_n\}$ be a sequence of random variables converging to $X$ in probability. Suppose that $\{X_n\}$ is bounded by $M$, i.e., $|X_n(\omega)| \le M, \forall \omega \in \Omega, n \ge 1$. Then $\{X_n\}$ converges to $X$ in $L^1$:

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

R Note that it is a stronger version of bounded convergence theorem compared with the one studied in MAT3006. In the theorem above, we only require convergence in probability rather than convergence a.s.

The relevance between uniform integrability and convergence of random variables is explained by the following theorem:

**Theorem 3.3** Let $\{X_n\}$ be a sequence of random variables with $X_n \in \mathcal{L}^1$, and let $X \in \mathcal{L}^1$. The sequence $\{X_n\}$ converges to $X$ in $L^1$ if and only if

1. $X_n \to X$ in probability, and
2. $\{X_n\}$ is uniformly integrable.

*Proof for the Reverse Direction.* For $K > 0$, construct a function $\phi_K : \mathbb{R} \to [-K, K]$:

$$\phi_K(x) = \begin{cases} K, & \text{if } x > K \\ x, & \text{if } |x| \le K \\ -K, & \text{if } x < -K \end{cases}$$

By the triangle inequality,

$$|X_n - X| \le |X_n - \phi_K(X_n)| + |\phi_K(X_n) - \phi_K(X)| + |\phi_K(X) - X|.$$

It suffices to upper bound three terms on the RHS for the following formula:

$$\begin{aligned}
\mathbb{E}[|X_n - X|] &\le \mathbb{E}[|X_n - \phi_K(X_n)|] + \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] + \mathbb{E}[|\phi_K(X) - X|] \\
&= \int_{\{|X|>K\}} [|X| - K] \, \mathrm{d}\mathbb{P} + \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] + \int_{\{|X_n|>K\}} [|X_n| - K] \, \mathrm{d}\mathbb{P} \\
&\le \int_{\{|X|>K\}} [|X|] \, \mathrm{d}\mathbb{P} + \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] + \int_{\{|X_n|>K\}} [|X_n|] \, \mathrm{d}\mathbb{P} \qquad (3.1)
\end{aligned}$$

- For the first term, by choosing sufficiently large $K$, by Corollary 3.1, it can be upper bounded by $\varepsilon/3$;

- For the third term, when $K$ is large enough, by the uniform integrability of $\{X_n\}$, it can be upper bounded by $\varepsilon/3$;

- Observe the following inequality holds:

  $$|\phi_K(x) - \phi_K(y)| \le |x - y|, \forall x, y \implies \{|\phi_K(X_n) - \phi_K(X)| > \varepsilon\} \subseteq \{|X_n - X| > \varepsilon\},$$

  which means that $\mathbb{P}(\{|\phi_K(X_n) - \phi_K(X)| > \varepsilon\}) \le \mathbb{P}(\{|X_n - X| > \varepsilon\})$. As a result, $X_n \to X$ in prob. implies $\phi_K(X_n) \to \phi_K(X)$ in prob.[1]

  By the Bounded Convergence Theorem 3.2, $\lim_{n\to\infty} \mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] = 0$.

[1]Following the similar method, we can show that as long as $f$ is continuous and $X_n \to X$ in prob., we have $f(X_n) \to f(X)$ in prob.

Thus for sufficiently large $n$,

$$\mathbb{E}[|\phi_K(X_n) - \phi_K(X)|] < \frac{\varepsilon}{3}$$

Combining these three bounds above, for fixed $\varepsilon > 0$, we can pick $K > 0$ and there exists sufficiently large $n$ such that

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

■

*Proof for the Forward Direction.*

- Firstly we show that $\{X_n\}$ is $L^1$-bounded, which suffices to show that $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$, which is because of the following observation:

$$\Big|\mathbb{E}[|X_n|] - \mathbb{E}[|X|]\Big| \leq \mathbb{E}||X_n| - |X|| \leq \mathbb{E}|X_n - X| \to 0.$$

- Then we show the uniform integrability result. By the $L^1$-convergence, for fixed $\varepsilon > 0$, there exists $N_0 > 0$ such that

$$\mathbb{E}\Big[|X_n - X|\Big] < \frac{\varepsilon}{2}, \quad \forall n > N_0.$$

  Similar as the previous proof for the uniform integrability results, we should apply proposition 3.1 on *finitely many* random variables: for fixed $\varepsilon > 0$, there exists **a** $\delta > 0$ such that whenever $\mathbb{P}(F) < \delta, F \in \mathcal{F}$,

$$\mathbb{E}[|X|1_F] < \frac{\varepsilon}{2} \tag{3.2a}$$

$$\mathbb{E}[|X_n|1_F] < \frac{\varepsilon}{2}, \quad \forall n \leq N_0 \tag{3.2b}$$

- Construct **a** $K$ such that $\mathbb{P}(|X_n| > K)$ is small for any $n$:

$$\mathbb{P}(|X_n| > K) \leq \frac{\mathbb{E}|X_n|}{K} \leq \frac{\sup_n \mathbb{E}|X_n|}{K}$$

Therefore, we choose $K$ such that $\frac{\sup_n \mathbb{E}|X_n|}{K} < \delta$, and then $\mathbb{P}(|X_n| > K) < \delta$.

- Now we can conclude the uniform integrability result: For $n \le N_0$, by the construction of $K$ and (3.2b),

$$\mathbb{E}[|X_n| 1_{\{|X_n|>K\}}] < \varepsilon.$$

For $n > N_0$,

$$\begin{aligned}
\mathbb{E}\Big[|X_n| 1_{\{|X_n|>K\}}\Big] &\le \mathbb{E}\Big[|X - X_n| 1_{\{|X_n|>K\}}\Big] + \mathbb{E}\Big[|X| 1_{\{|X_n|>K\}}\Big] \\
&\le \mathbb{E}\Big[|X - X_n|\Big] + \mathbb{E}\Big[|X| 1_{\{|X_n|>K\}}\Big] \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

where the last inequality is because of the $L^1$-convergence and (3.2a).

- Finally, the convergence in probability can be shown by the Markov inequality:

$$\mathbb{P}\Big(|X_n - X| > \varepsilon\Big) \le \frac{\mathbb{E}[|X_n - X|]}{\varepsilon} \to 0, \text{ as } n \to \infty.$$

■

## 3.1.4. Martingales in Discrete Time

**Definition 3.4** [Stochastic Process] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We describe random phenomena in discrete time by a collection of random variables $\{X_n : n \ge 1\}$ and increasing sequence of sub $\sigma$-fields $\{\mathcal{F}_n : \mathcal{F}_n \subseteq \mathcal{F}\}$.

- $X(\cdot) \triangleq \{X_n : n \ge 1\}$ is called a **stochastic process**;
- $\mathbb{F} \triangleq \{\mathcal{F}_n : \mathcal{F}_n \subseteq \mathcal{F}\}$ is called a **filtration**.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated with a filtration $\mathbb{F}$ is called a **filtered** probability space, written as $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ ■

R A typical example of $\mathcal{F}$ is defined by generated $\sigma$-algebra:

$$\mathcal{F}_n^X \triangleq \sigma(X_t : t \le n), \qquad \forall n \ge 0.$$

This natural filteration is the sequence of smallest $\sigma$-algebras such that $X_n$ is $\mathcal{F}_n$-measurable for all $n$.

**Definition 3.5** [Predictable process]

- A stochastic process $X(\cdot) \triangleq \{X_n : n \ge 1\}$ is said to be **adapted** to the filtration $\mathbb{F} \triangleq \{\mathcal{F}_n : n \ge 1\}$ if $X_n$ is $\mathcal{F}_n$-measurable for each $n$. We call $X$ an **adapted process** with respect to $\mathbb{F}$.
- If $X_n$ is $\mathcal{F}_{n-1}$-measurable for each $n$ and $X_0$ is $\mathcal{F}_0$-measurable, $X$ is said to be a **predictable process**.

## 3.2. Thursday

### 3.2.1. Stopping Time

**Definition 3.6** [Stopping Time] A mapping $T : \Omega \to \{0,1,2,\dots,\infty\}$ is called a stopping time with respect to the filteration $\{\mathcal{F}_n\}_{n\ge 0}$ if

$$\{T \le n\} \triangleq \{\omega \in \Omega : T(\omega) \le n\} \in \mathcal{F}_n, \qquad \forall n.$$

R

1. $T$ can take the infinite value
2. An equivalent definition for a stopping time $T$ is $\{T = n\} \in \mathcal{F}_n, \forall n$.

*Proof.* (a) Suppose that $\{T \leq n\} \in \mathcal{F}_n, \forall n$, then

$$\{T \leq n-1\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n \implies \{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n.$$

(b) Suppose that $\{T = n\} \in \mathcal{F}_n, \forall n$, then

$$\{T = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n, \forall k \leq n \implies \{T = n\} = \bigcup_{k \leq n} \{T = k\} \in \mathcal{F}_n.$$

■

3. A constant mapping $T \equiv N$ with $N \in \mathbb{Z}_+$ is always a stopping time.

■ **Example 3.1** Let $\{X_n\}_{n\geq 0}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathbb{P})$. Let $B \in \mathcal{B}(\mathbb{R})$ be a Borel set. Define

$$T(\omega) \triangleq \inf \{n \geq 0 : \; X_n(\omega) \in B\}.$$

Here $T$ denotes the first time that $\{X_n\}_{n\geq 0}$ enters into set $B$. Define $\inf(\emptyset) = \infty$ by default, i.e., $T = \infty$ when $\{X_n\}_{n\geq 0}$ never enters into $V$. To check that $T$ is a stopping time, observe that

$$\begin{aligned}\{T = n\} &= \{X_0 \in B^c, X_1 \in B^c, X_2 \in B^c, \dots, X_{n-1} \in B^c, X_n \in B\} \\ &= \{X_n \in B\} \cup (\cup_{0\leq k\leq n-1} \{X_k \in B^c\})\end{aligned}$$

Since $\{X_n\}$ is adapted, $\{X_k \in B^c\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$ for $0 \leq k \leq n-1$. Moreover, $\{X_n \in B\} \in \mathcal{F}_n$. Therefore, $\{T = n\} \in \mathcal{F}_n$ for each $n$.

■

**Definition 3.7** [Stopping Time $\sigma$-algebra] Define the stopping time $\sigma$-algebra for a given stopping time $T$ as the following:

$$\mathcal{F}_T = \{A \in \mathcal{F} : \; A \cap \{T \leq n\} \in \mathcal{F}_n, \forall n\}.$$

Here $\mathcal{F}_T$ represents the information available up to a random time $T$.

**Proposition 3.3**

1. $\mathcal{F}_T$ is a $\sigma$-algebra;
2. $T$ is $\mathcal{F}_T$-measurable;
3. When $T_1, T_2$ are two stopping times with $T_1 \le T_2$ a.s., $\mathcal{F}_{T_1} \subseteq \mathcal{F}_{T_2}$.

*Proof.*

1. It is trivial that $\emptyset \in \mathcal{F}_T$. Suppose that $A \in \mathcal{F}_T$, then $(A \cap \{T \le n\})^c \in \mathcal{F}_n$, which implies that

$$A^c \cap \{T \le n\} = \Big(A \cap \{T \le n\}\Big)^c \cap \{T \le n\} \in \mathcal{F}_n.$$

Suppose that $A_k \in \mathcal{F}_T, k \ge 1$, then

$$\left(\bigcup_{k \ge 1} A_k\right) \cap \{T \le n\} = \bigcup_{k \ge 1} (A_k \cap \{T \le n\}) \in \mathcal{F}_n.$$

2. It suffices to show that $\{T \le m\} \in \mathcal{F}_T$ for any $m$. This is true because for any $n$,

$$\{T \le m\} \cap \{T \le n\} = \{T \le m \wedge n\} \in \mathcal{F}_{m \wedge n} \subseteq \mathcal{F}_n.$$

3. Consider any $A \in \mathcal{F}_{T_1}$, then $A \cap \{T_1 \le n\} \in \mathcal{F}_n$ for any $n$. Moreover,

$$\{T_2 \le n\} \subseteq \{T_1 \le n\} \implies A \cap \{T_2 \le n\} \subseteq A \cap \{T_1 \le n\} \in \mathcal{F}_n,$$

which implies the desired result.

■

**Theorem 3.4** Let $\{X_n\}_{n \ge 0}$ be an adapted process on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \ge 0}, \mathbb{P})$. Let $T$ be a stopping time w.r.t. $\{\mathcal{F}_n\}_{n \ge 0}$. Define a random variable $X_T$:

$$X_T(\omega) \triangleq X_{T(\omega)}(\omega), \qquad \forall \omega \in \Omega.$$

Then $X_T$ is $\mathcal{F}_T$-measurable.

*Proof.* It suffices to check $\{X_T \le a\} \in \mathcal{F}_T, \forall a \in \mathbb{R}$. By definition of the stopping time

$\sigma$-algebra, it suffices to check

$$\{X_T \le a\} \cap \{T \le n\} \in \mathcal{F}_n, \forall n \quad \Longleftarrow \quad \bigcup_{0 \le k \le n} \{X_k \le a\} \cap \{T = k\} \in \mathcal{F}_n, \forall n.$$

Since $\{X_n\}$ is adapted, $\{X_k \le a\} \in \mathcal{F}_k, \forall k$. By definition of the stopping time, $\{T = k\} \in \mathcal{F}_k, \forall k$. Therefore,

$$\{X_k \le a\} \cap \{T = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n.$$

The proof is complete. ■

**Definition 3.8** [Martingale] Let $\{X_n\}_{n\ge 0}$ be an adapted process on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\ge 0}, \mathbb{P})$. A stochastic process $\{X_n\}_{n\ge 0}$ is called a **martingale** if

1. $X_n \in \mathcal{L}^1, \forall n$;
2. $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ a.s., for all $n$.

If in the last definition, "$=$" is replaced by "$\le$" or "$\ge$", then $\{X_n\}_{n\ge 0}$ is said to be a **supermartingale** or **submartingale**, respectively.

R

- A supermartingale goes downward on average, and a submartigale goes upward on average.
- $\{X_n\}_{n\ge 0}$ is a supermartingale if and only if $\{-X_n\}_{n\ge 0}$ is a submartingale.
- $\{X_n\}_{n\ge 0}$ is a martingale if and only if it is both a supermartingale and a submartingale.

■ **Example 3.2** Let $\{Y_n\}_{n\ge 1}$ be a sequence of independent random variables with $\mathbb{E}[|Y_k|] < \infty$ and $\mathbb{E}[Y_k] = 0, \forall k$. Define $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$ for $n \ge 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Define $X_n = Y_1 + Y_2 + \cdots + Y_n, \forall n \ge 1$ and $X_0 = 0$. Then $\{X_n\}_{n\ge 0}$ is a martingale:

1. $\mathbb{E}[|X_n|] \le \sum_{i=1}^n \mathbb{E}[|X_i|] < \infty$, which means that $X_n$ is integrable;

2. Check that

$$
\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_n + Y_{n+1} \mid \mathcal{F}_n] \\
&= \mathbb{E}[X_n \mid \mathcal{F}_n] + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \\
&= X_n + \mathbb{E}[Y_{n+1}] = X_n,
\end{aligned}
$$

where the third equality is because that $X_n$ is $\mathcal{F}_n$-measurable and $Y_{n+1}$ is independent of $\mathcal{F}_n$.

■

■ **Example 3.3** Let $\{Y_n\}_{n\geq 1}$ be a sequence of independent random variables with $Y_k \geq 0$ a.s. and $\mathbb{E}[Y_k] = 1, \forall k$. Define $\mathcal{F}_n = \sigma(Y_1, Y_2, \ldots, Y_n)$ for $n \geq 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Define $X_n = Y_1 \cdot Y_2 \cdots Y_n, \forall n \geq 1$ and $X_0 = 1$. Then $\{X_n\}_{n\geq 0}$ is a martingale:

1. $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = \prod_{k=1}^{n} \mathbb{E}[Y_k] = 1 < \infty$, which means that $X_n$ is integrable;
2. Check that

$$
\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_n \cdot Y_{n+1} \mid \mathcal{F}_n] \\
&= X_n \cdot \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \\
&= X_n \cdot \mathbb{E}[Y_{n+1}] = X_n,
\end{aligned}
$$

where the second equality is because that $X_n$ is $\mathcal{F}_n$-measurable; the third equality is because that $Y_{n+1}$ is independent of $\mathcal{F}_n$.

■

# Chapter 4

# Week4

## 4.1. Tuesday

### 4.1.1. Martingales in Discrete Time

■ **Example 4.1** Let $\mathbb{F} \triangleq \{\mathcal{F}_n\}_{n\ge 0}$ be a filteration and consider a random variable $\zeta \in \mathcal{L}^1$. Define $X_n \triangleq \mathbb{E}[\zeta \mid \mathcal{F}_n]$, and we can check that $\{X_n\}_{n\ge}$ is a martingale with respect to $\mathbb{F}$:

- Firstly we need to show the integrability of $X_n$ for any $n$:

$$\begin{aligned}\mathbb{E}[|X_n|] &= \mathbb{E}[|\mathbb{E}[\zeta \mid \mathcal{F}_n]|] \\ &\le \mathbb{E}[\mathbb{E}[|\zeta| \mid \mathcal{F}_n]] \\ &= \mathbb{E}[|\zeta|] < \infty\end{aligned}$$

  where the first inequality is by the Jensen's inequality.

- Then we check that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ for any $n$:

$$\begin{aligned}\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[\mathbb{E}[\zeta \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] \\ &= \mathbb{E}[\zeta \mid \mathcal{F}_n] = X_n.\end{aligned}$$

■

■ **Example 4.2** [Martingale Transform] Let $C_n$ be the stake to be bet on game $n$, and $X_n - X_{n-1}$ be the net with per stake in game $n$, with $n \geq 1$. Suppose that the process $\{C_n\}_{n\geq 1}$ is predictable, and the total win up to time $n$ is $Y_n = \sum_{1\leq k\leq n} C_k(X_k - X_{k-1})$. Define $Y_0 := 0$. If $\{X_n\}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$ and $\{C_n\}$ is bounded a.s.[a], then we can show that $\{Y_n\}$ is also a martingale:

Firstly note that $Y_n$ is $\mathcal{F}_n$-measurable since $X_k, C_k$ are all $\mathcal{F}_n$-measurable for $1 \leq k \leq n$. Then we check $\{Y_n\}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$:

- For any $n$, we have

$$\begin{aligned}
\mathbb{E}[|Y_n|] &\leq \sum_{1\leq k\leq n} \mathbb{E}[|C_k(X_k - X_{k-1})|] \\
&\leq M \cdot \sum_{1\leq k\leq n} \mathbb{E}[|X_k - X_{k-1}|] \\
&\leq M \cdot \sum_{1\leq k\leq n} \mathbb{E}[|X_k|] + \mathbb{E}[|X_{k-1}|] < \infty,
\end{aligned}$$

  where the second inequality is by the boundedness of $\{C_n\}$.

- Moreover,

$$\begin{aligned}
\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[\sum_{1\leq k\leq n+1} C_k(X_k - X_{k-1}) \middle| \mathcal{F}_n\right] \\
&= \mathbb{E}[Y_n + C_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] \\
&= \mathbb{E}[Y_n \mid \mathcal{F}_n] + C_{n+1}\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] \\
&= Y_n + C_{n+1}\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = Y_n,
\end{aligned}$$

  where the third equality is by the $\mathcal{F}_n$-measurability of $C_{n+1}$; the fourth equality is by the $\mathcal{F}_n$-measurability of $Y_n$, and the last equality is by $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$.

■

[a] Here the boundedness means that $|C_n(\omega)| \leq M$ for some $M > 0$ and almost all $\omega$

**Theorem 4.1** Suppose that $\{X_n\}$ is a martingale with respect to $\mathcal{F}_n$, and $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\phi(X_n)$ is integrable for all $n$, then $\{\phi(X_n)\}$ is a sub-martingale with respect to $\mathcal{F}_n$.

*Proof.* By the Jensen's inequality and $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$, we have

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

■

By similar proof, we can show the following theorem:

**Theorem 4.2** Suppose that $\{X_n\}$ is a sub-martingale with respect to $\mathcal{F}_n$, and $\phi$ is an increasing convex function such that $\phi(X_n)$ is integrable for all $n$, then $\{\phi(X_n)\}$ is a sub-martingale with respect to $\mathcal{F}_n$.

A direct example is the following:

■ **Example 4.3** Suppose that $\{X_n\}$ is a sub-martingale, define the convex function $X^+ := \max(X, 0)$, then $\{X_n^+\}$ is also a sub-martingale. ■

**Theorem 4.3** Let $\{X_n\}$ be a martingale and $T$ be a stopping time. Define the stopped process $\{X_{n\wedge T}\}$ as

$$X_{n\wedge T}(\omega) \triangleq X_{n\wedge T(\omega)}(\omega), \qquad \forall \omega \in \Omega, \forall n.$$

Then $\{X_{n\wedge T}\}$ is a martingale. In particular,

$$\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0], \quad \forall n.$$

*Proof.* We will show this result by applying the Martingale transform technique mentioned in Example 4.2. Define the stake process $\{C_n^T\}_{n\geq 0}$ as

$$C_n^T(\omega) = 1\{n \leq T(\omega)\}(\omega), \qquad \forall \omega \in \Omega, \forall n.$$

Note that $\{C_n^T\}_{n\geq 0}$ is predictable since $\{C_n^T = 0\} = \{T(\omega) \leq n-1\} \in \mathcal{F}_{n-1}$. Now we begin to simplify $\sum_{1\leq k\leq n} C_k^T(X_k - X_{k-1})$:

$$\begin{aligned}
\sum_{1\leq k\leq n} C_k^T(X_k - X_{k-1}) &= 1\{T\geq 1\}(X_1 - X_0) + 1\{T\geq 2\}(X_2 - X_1) + \cdots + 1\{T\geq n\}(X_n - X_{n-1}) \\
&= -1\{T\geq 1\}X_0 + (1\{T\geq 1\} - 1\{T\geq 2\})X_1 + (1\{T\geq 2\} - 1\{T\geq 3\})X_2 \\
&+ \cdots + (1\{T\geq n-1\} - 1\{T\geq n\})X_{n-1} + 1\{T\geq n\}X_n \\
&= 1\{T\geq n\}X_n - 1\{T\geq 1\}X_0 + \sum_{i=1}^{n-1} 1\{T=i\}X_i \\
&= \left(1\{T\geq n\}X_n + \sum_{i=0}^{n-1} 1\{T=i\}X_i\right) - \left(1\{T\geq 1\}X_0 - 1\{T=0\}X_0\right) \\
&= X_{n\wedge T} - X_0.
\end{aligned}$$

By the boundedness of $\{C_n^T\}$, and the result in Example 4.2, we can show that $\{X_{n\wedge T} - X_0\}$ is a martingale, i.e., $\{X_{n\wedge T}\}$ is a martingale. Therefore,

$$\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[\mathbb{E}[X_{n\wedge T} \mid \mathcal{F}_{n-1}]] = \mathbb{E}[X_{(n-1)\wedge T}] = \cdots = \mathbb{E}[X_{0\wedge T}] = \mathbb{E}[X_0].$$

■

Note that $\mathbb{E}[X_T]$ does not equal to $\lim_{n\to\infty}\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0]$. The following provides a counter-example:

▪ **Example 4.4** Let $\{X_n\}$ be a simple symmetric random walk on integers and $X_0 = 0$. Then $\{X_n\}$ is a martingale.Define the stopping time

$$T \triangleq \inf\{n\geq 0 \ X_n = 1\}.$$

Then $\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0] = 0, \forall n$. Since the random walk is recurrent, $\mathbb{P}(T<\infty)$, and $X_T = 1$ a.s., which implies that

$$1 = \mathbb{E}[X_T] \neq \lim_{n\to\infty}\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0] = 0.$$

The Doob's optional stopping theorem provides sufficient conditions for $\mathbb{E}[X_T] = \mathbb{E}[X_0]$:

**Theorem 4.4 — Doob's Optional Stopping Theorem.** Let $\{X_n\}$ be a martingale and $T$ be a stopping time. Then $X_T$ is integrable and $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ if any of the following conditions hold:

1. $T$ is bounded a.s.;
2. $\{X_n\}$ is bounded and $T$ is finite ($\mathbb{P}(T < \infty) = 1$)[a], a.s.;
3. $\mathbb{E}[T] < \infty$[b] and $\{|X_n - X_{n-1}|\}$ is bounded.

[a] The finiteness is a weaker condition, which does not imply boundedness

[b] The $\mathbb{E}[T] < \infty$ is a little bit stronger condition than finiteness, but still does not imply boundedness.

*Proof.*

1. Suppose that $T$ is bounded a.s., which means that there exists $K$ such that $\mathbb{P}\{\omega : T(\omega) \le K\} = 1$. Therefore, $X_{K\wedge T} = X_T$ a.s. By Theorem 4.3, $X_T$ is integrable with $\mathbb{E}[X_T] = \mathbb{E}[X_{K\wedge T}] = \mathbb{E}[X_0]$.
2. Suppose that $T$ is finite a.s., then we can show that $X_{n\wedge T} \to X_T$ a.s.: Note that $\mathbb{P}\{T < \infty\} = 1$, and
$$\omega \in \{T < \infty\} \implies \lim_{n\to\infty} X_{n\wedge T}(\omega) = X_T(\omega).$$
Since $\{X_n\}$ is bounded a.s., $\{X_{n\wedge T}\}$ is bounded a.s. as well. By the Bounded Convergence Theorem, $X_T$ is integrable and $X_{n\wedge T} \to X_T$ in $L^1$, which implies that
$$\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0].$$
3. We first show that $\{X_{T\wedge n} - X_0\}$ is dominated by an integrable random variable:
$$|X_{T\wedge n} - X_0| = \left|\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})\right| \le \sum_{k=1}^{T\wedge n}|X_k - X_{k-1}| \le M\cdot(T\wedge n) \le MT,$$
where the second inequality is by the boundedness of $\{|X_n - X_{n-1}|\}$, i.e., for any $n$ and $\omega \in \Omega$, $|X_n(\omega) - X_{n-1}(\omega)| \le M$. Considering that $\mathbb{E}[T] < \infty$, $T$ is finite a.s., which implies that $X_{n\wedge T} \to X_T$. Applying the dominated convergence theorem,

$X_{T\wedge n} \to X_T$ in $L^1$, and

$$\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0].$$

■

### 4.1.2. Doob's Inequalities

**Theorem 4.5 — Doob's Optional Sampling Theorem.** Let $\{X_n\}$ be a martingale and $S, T$ be two bounded stopping times, with $S \le T$ a.s., then $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s.

Moreover, if instead $\{X_n\}$ is assumed to be a sub-martingale or super-martingale, then the equality in the result is replaced by $\ge$ or $\le$, respectively.

*Proof.* We only show the result based on the assumption that $\{X_n\}$ is a sub-martingale, since the remaining part follows the similar logic. Since $S, T$ are bounded a.s., random variables $X_T$ and $X_S$ are integrable. In order to simplify $\mathbb{E}[X_T \mid \mathcal{F}_S]$, we need to study the structure of $\mathcal{F}_S$: For any $A \in \mathcal{F}_S$, by the definition of stopping-time $\sigma$-algebra, $A \cap \{S \le j\} \in \mathcal{F}_j$. Considering that $\{S > j-1\} = \{S \le j-1\}^c \in \mathcal{F}_{j-1} \subseteq \mathcal{F}_j$ and $\{T > j\} = \{T \le j\}^c \in \mathcal{F}_j$,

$$A \cap \{S \le j\} \cap \{S > j-1\} \cap \{T > j\} = A \cap \{S = j\} \cap \{T > j\} \in \mathcal{F}_j, \quad \forall j.$$

- Assume that $0 \le T - S \le 1$, a.s., it follows that

$$\int_A (X_T - X_S)\,\mathrm{d}\mathbb{P} = \sum_{j=0}^{N} \int_{A\cap\{S=j\}} (X_T - X_S)\,\mathrm{d}\mathbb{P} \tag{4.1a}$$

$$= \sum_{j=0}^{N} \int_{A\cap\{S=j\}\cap\{T>j\}} (X_T - X_S)\,\mathrm{d}\mathbb{P} + \sum_{j=0}^{N} \int_{A\cap\{S=j\}\cap\{T=j\}} (X_T - X_S)\,\mathrm{d}\mathbb{P} \tag{4.1b}$$

$$= \sum_{j=0}^{N} \int_{A\cap\{S=j\}\cap\{T>j\}} (X_T - X_S)\,\mathrm{d}\mathbb{P} \tag{4.1c}$$

where (4.1a) is by the assumption that $S$ is bounded a.s., i.e., $|S| \le N$ a.s.; (4.1b) is by the assumption that $T \geq S$; (4.1c) is because $1\{S=j, T=j\} \cdot (X_T - X_S) = 0$. Since $\mathbb{E}[X_{j+1} \mid \mathcal{F}_j] \leq X_j$ a.s. for any $j$,

$$\int_{A\cap\{S=j\}\cap\{T>j\}} (X_j - X_{j+1})\,\mathrm{d}\mathbb{P} \geq 0 \implies \int_A (X_S - X_T)\,\mathrm{d}\mathbb{P} \geq 0, \forall A \in \mathcal{F}_S. \quad (4.2)$$

For two $\mathcal{F}$-measurable random variables, if $\int_A (X-Y)\,\mathrm{d}\mathbb{P} \leq 0, \forall A \in \mathcal{F}$, then one can assert that $X \leq Y$ a.s. Therefore, (4.2) implies $\mathbb{E}[X_T \mid \mathcal{F}_S] \leq X_S$ a.s.

- Now suppose that $T - S \geq 0$ a.s., and construct intermediate variables $R_j = T \wedge (S+j), j = 1, 2, \ldots, N$. It follows that $R_j$ is a stopping time and $S \leq R_1 \leq R_2 \leq \cdots \leq R_N \leq T$ a.s., with

$$0 \leq R_1 - S \leq 1, \quad 0 \leq R_j - R_{j-1} \leq 1, \forall j \quad 0 \leq T - R_N \leq 1.$$

Consider any $A \in \mathcal{F}_S$, and since $0 \leq R_1 - S \leq 1$ a.s.,

$$\int_A (X_S - X_{R_1})\,\mathrm{d}\mathbb{P} \geq 0.$$

By definition of stopping time $\sigma$-algebra, $A \cap \{S \leq j\} \in \mathcal{F}_j$, which implies that

$$A \cap \{S \leq j\} \cap \{R_1 \leq j\} = A \cap \{R_1 \leq j\} \in \mathcal{F}_j \implies A \in \mathcal{F}_{R_1}.$$

Considering that $0 \leq R_2 - R_1 \leq 1$ a.s.,

$$\int_A (X_{R_1} - X_{R_2})\,\mathrm{d}\mathbb{P} \geq 0.$$

Similarly,

$$\int_A (X_{R_{j-1}} - X_{R_j})\,\mathrm{d}\mathbb{P} \geq 0, j = 2, \ldots, N, \quad \int_A (X_{R_N} - X_T)\,\mathrm{d}\mathbb{P} \geq 0.$$

Adding those integrals above, $\int_A (X_S - X_T)\,\mathrm{d}\mathbb{P} \geq 0, \forall A \in \mathcal{F}_S$, i.e.,

$$\mathbb{E}[X_T \mid \mathcal{F}_S] \leq X_S, \quad \text{a.s.}$$

■

**R** We can assume the uniform integrability of $\{X_n\}$ and the conclusion still holds, without assuming that $T, S$ are bounded.

## 4.2. Thursday

### 4.2.1. Doob's Maximal Inequality

**Theorem 4.6 — Doob's Maximal Inequality.** Let $\{X_n\}$ be a super-martingale. Choose some $N > 0$ , then for any $\lambda > 0$,

1. $\lambda \cdot \mathbb{P}\left(\sup_{k\le N} X_k \ge \lambda\right) \le \mathbb{E}[X_0] - \mathbb{E}\left[X_N \cdot 1\left\{\sup_{k\le N} X_k < \lambda\right\}\right]$;
2. $\lambda \cdot \mathbb{P}\left(\inf_{k\le N} X_k \le -\lambda\right) \le \mathbb{E}\left[(-X_N) \cdot 1\left\{\inf_{k\le N} X_k \le -\lambda\right\}\right]$.

*Proof.* 1. Define a stopping time $R$:

$$R(\omega) = \inf\{k \ge 0 : \ X_k(\omega) \ge \lambda\}, \quad \forall \omega \in \Omega.$$

Take $T = R \wedge N$, which is a bounded stopping time. Apply the Optional Sampling Theorem 4.5,

$$\begin{aligned}
\mathbb{E}[X_0] &\ge \mathbb{E}\Big[\mathbb{E}[X_T \mid \mathcal{F}_0]\Big] = \mathbb{E}[X_T] \\
&= \int 1\left\{\sup_{k\le N} X_k \ge \lambda\right\} X_T \,\mathrm{d}\mathbb{P} + \int 1\left\{\sup_{k\le N} X_k < \lambda\right\} X_T \,\mathrm{d}\mathbb{P} \\
&\ge \lambda \cdot \mathbb{P}\left(\sup_{k\le N} X_k \ge \lambda\right) + \int 1\left\{\sup_{k\le N} X_k < \lambda\right\} X_N \,\mathrm{d}\mathbb{P}
\end{aligned}$$

where the first inequality is because that $X_0 \ge \mathbb{E}[X_T \mid \mathcal{F}_0]$ a.s.; and the last inequality is because that conditioned on the event $\left\{\sup_{k\le N} X_k < \lambda\right\}$, $X_T \equiv X_N$. Thus the desired result holds.

2. Let $Y_n = -X_n$, and $\{Y_n\}$ is a sub-martingale. Define the stopping time

$$R(\omega) = \inf\{k \geq 0 : \; Y_k(\omega) \geq \lambda\}, \quad \forall \omega \in \Omega.$$

Take $T = R \wedge N$, which is a bounded stopping time. Apply the Optional Sampling Theorem 4.5,

$$\mathbb{E}[Y_N] \geq \mathbb{E}[Y_T] \geq \lambda \mathbb{P}\left(\sup_{k \leq N} Y_k \geq \lambda\right) + \mathbb{E}\left[Y_N 1\left(\sup_{k \leq N} Y_k < \lambda\right)\right]$$

It follows that

$$\begin{aligned}\lambda \cdot \mathbb{P}\left(\inf_{k \leq N} X_k \leq -\lambda\right) &= \lambda \cdot \mathbb{P}\left(\sup_{k \leq N} Y_k \geq \lambda\right) \leq \mathbb{E}[Y_N] - \mathbb{E}\left[Y_N 1\left(\sup_{k \leq N} Y_k < \lambda\right)\right] \\ &= \mathbb{E}\left[(-X_N) 1\left(\inf_{k \leq N} X_k \leq -\lambda\right)\right].\end{aligned}$$

■

R Summing up these two results in Theorem 4.6, we imply

$$\begin{aligned}\lambda \cdot \mathbb{P}\left(\sup_{k \leq N} |X_k| \geq \lambda\right) &\leq \mathbb{E}[X_0] - \mathbb{E}\left[X_N \cdot 1\left\{\sup_{k \leq N} X_k < \lambda\right\}\right] - \mathbb{E}\left[X_N \cdot 1\left\{\inf_{k \leq N} X_k \leq -\lambda\right\}\right] \\ &\leq \mathbb{E}[X_0] + 2\mathbb{E}[X_N^-],\end{aligned}$$

where $X^- \triangleq \max(-X, 0)$.

**Theorem 4.7** Let $\{X_n\}$ be a martingale. Choose some $N > 0$ and let $X_N \in \mathcal{L}^2$, i.e., $\mathbb{E}[X_N^2] < \infty$. Then for any $\lambda > 0$,

$$\mathbb{P}\left(\sup_{k \leq N} |X_k| > \lambda\right) \leq \frac{1}{\lambda^2} \mathbb{E}[X_N^2].$$

*Proof.* We can show that $\{X_n^2\}$ is a sub-martingale by applying Jensen's inequality:

$$\mathbb{E}[X_{n+1}^2 \mid \mathcal{F}_n] \geq (\mathbb{E}[X_{n+1} \mid \mathcal{F}_n])^2 = X_n^2.$$

As a result, $\mathbb{E}[X_k^2] \le \mathbb{E}[X_N^2] < \infty, \forall k \le N$, i.e., $\{-X_k^2\}_{k\le N}$ is a super-martingale. Apply the second part in Theorem 4.6 completes the proof:

$$\begin{aligned}\lambda^2 \cdot \mathbb{P}\left(\inf_{k\le N}(-X_k^2) \le -\lambda^2\right) &= \lambda^2 \cdot \mathbb{P}\left(\sup_{k\le N}|X_k| \ge \lambda\right)\\ &\le \mathbb{E}\left[X_N^2 \cdot 1\left\{\inf_{k\le N}(-X_k^2) \le -\lambda^2\right\}\right]\\ &\le \mathbb{E}[X_N^2].\end{aligned}$$

■

**Theorem 4.8** — **Doob's $L^p$-inequality.**

1. Suppose that $\{X_n\}$ is a sub-martingale, then for any $p > 1$,

$$\mathbb{E}\left[\left(\sup_{k\le n} X_k^+\right)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n^+)^p].$$

2. Suppose that $\{X_n\}$ is a martingale, then for any $p > 1$,

$$\mathbb{E}\left[\sup_{k\le n}|X_k|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

*Proof.* 1. W.l.o.g., assume that $\{X_n\}$ is non-negative, and we may replace $X_n^+$ by $X_n$. Consider a continuous increasing function $\phi : \mathbb{R}_+ \to [0, +\infty)$ with $\phi(0) = 0$, and we evaluate the expectation for $\phi(Z)$, where $Z$ is a given random variable:

$$\begin{aligned}\mathbb{E}[\phi(Z)] &= \int_\Omega \phi(Z(\omega))\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_\Omega \int_0^{Z(\omega)} \mathrm{d}\phi(y)\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_\Omega \int_{[0,\infty)} 1\{y \le Z(\omega)\}\,\mathrm{d}\phi(y)\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_{[0,\infty)} \int_\Omega 1\{y \le Z(\omega)\}\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\phi(y) \qquad (4.3a)\\ &= \int_{[0,\infty)} \mathbb{P}(Z \ge y)\,\mathrm{d}\phi(y)\end{aligned}$$

where (4.3a) is by the Fubini's theorem.

Take $\phi(y) \equiv y^p$ and define $X_n^* = \sup_{k\le n} X_n$[1] for notation simplifcation. As a result, using a little bit calculus gives

$$
\begin{aligned}
\mathbb{E}[(X_n^*)^p] &= \int_{[0,\infty)} \mathbb{P}(X_n^* \ge \lambda)\,\mathrm{d}\lambda^p \\
&\le \int_{[0,\infty)} \frac{1}{\lambda}\mathbb{E}\left[|X_n| 1\left\{\sup_{k\le n}|X_k| \ge \lambda\right\}\right]\mathrm{d}\lambda^p && \text{(4.3b)} \\
&= \int_0^\infty \frac{1}{\lambda}\int_\Omega X_n(\omega) 1\{X_n^*(\omega)\ge\lambda\}\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\lambda^p && \text{(4.3c)} \\
&= \int_\Omega X_n(\omega)\int_0^\infty \frac{1}{\lambda} 1\{X_n^*(\omega)\ge\lambda\}\,\mathrm{d}\lambda^p\,\mathrm{d}\mathbb{P}(\omega) && \text{(4.3d)} \\
&= \int_\Omega X_n(\omega)\int_0^{X_n^*(\omega)} p\lambda^{p-2}\,\mathrm{d}\lambda\,\mathrm{d}\mathbb{P}(\omega) \\
&= \int_\Omega X_n(\omega)\frac{p}{p-1}[X_n^*(\omega)]^{p-1}\,\mathrm{d}\mathbb{P}(\omega) \\
&\le \frac{p}{p-1}\left(\mathbb{E}[(X_n)^p]\right)^{1/p}\left(\mathbb{E}[(X_n^*)^{(p-1)q}]\right)^{1/q}, \quad \text{with } 1/q = 1-1/p && \text{(4.3e)} \\
&= \frac{p}{p-1}\left(\mathbb{E}[(X_n)^p]\right)^{1/p}\left(\mathbb{E}[(X_n^*)^p]\right)^{(p-1)/p} && \text{(4.3f)}
\end{aligned}
$$

where (4.3b) is by the Doob's maximal inequality; (4.3c) is by the assumption that $X_n \ge 0$; (4.3d) is by Fubini's theorem; (4.3e) is by Holder's inequality. If dividing both sides in (4.3f) by $(\mathbb{E}[(X_n^*)^p])^{(p-1)/p}$, we get the desired result.

2. The second inequality follows by applying the first and replacing $\{X_n\}$ with $\{|X_n|\}$.

■

■ **Example 4.5** Let $\{X_n\}$ be a non-negative sub-martingale. Then we can apply the similar proceed to show the following upper bound holds:

$$\mathbb{E}\left[\sup_{k\le n} X_k\right] \le \frac{e}{e-1}\left(1 + \sup_{k\le n}\mathbb{E}[X_k \log^+ X_k]\right),$$

where $\log^+ X \triangleq (\log X)1\{X \ge 1\}$.

[1] called the running maximal of $\{X_n\}$

Tale $\phi(y) = (y-1)^+$, then

$$\begin{aligned}
\mathbb{E}[(X_n^* - 1)^+] &= \int_0^\infty \mathbb{P}(X_n^* \geq \lambda)\,\mathrm{d}\phi(\lambda) \\
&\leq \int_0^\infty \frac{1}{\lambda}\mathbb{E}[X_n 1\{X_n^* \geq \lambda\}]\,\mathrm{d}\phi(\lambda) && \text{(4.4a)} \\
&= \int_\Omega X_n \int_0^{X_n^*} \frac{1}{\lambda}\,\mathrm{d}\phi(\lambda)\,\mathrm{d}\mathbb{P} && \text{(4.4b)} \\
&= \int_\Omega X_n \int_0^{X_n^*} \frac{1}{\lambda} 1\{\lambda \geq 1\}\,\mathrm{d}\lambda\,\mathrm{d}\mathbb{P} \\
&= \int_\Omega X_n 1\{X_n^* \geq 1\} \log X_n^*\,\mathrm{d}\mathbb{P} \\
&= \mathbb{E}[X_n \log^+ X_n^*].
\end{aligned}$$

where (4.4a) is by the Doob's maximal inequality, and (4.4b) is by Fubini's theorem. As a result,

$$\mathbb{E}[X_n^*] - 1 \leq \mathbb{E}[(X_n^* - 1)^+] \leq \mathbb{E}[X_n \log^+ X_n^*].$$

We can use a bit calculus to show that

$$a \log^+ b \leq a \log^+ a + \frac{b}{e}, \tag{4.4c}$$

which implies that

$$\mathbb{E}[X_n^*] - 1 \leq \mathbb{E}[X_n \log^+ X_n] + \frac{1}{e}\mathbb{E}[X_n^*].$$

The proof is complete. ■

# Chapter 5

# Week5

## 5.1. Tuesday

### 5.1.1. Convergence of Martingales

Let $\{X_n\}_{n\geq 0}$ be an adapted process on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathbb{P})$, and $[a,b]$ be a closed interval. Define $T_0 = \inf\{n \geq 0, X_n \leq a\}$, and

$$T_{2k-1} = \inf\{n > T_{2k-2} : X_n \geq b\}, \qquad T_{2k} = \inf\{n > T_{2k-1} : X_n \leq a\}$$

See the Figure for a illustration of $T_k, k \geq 0$.

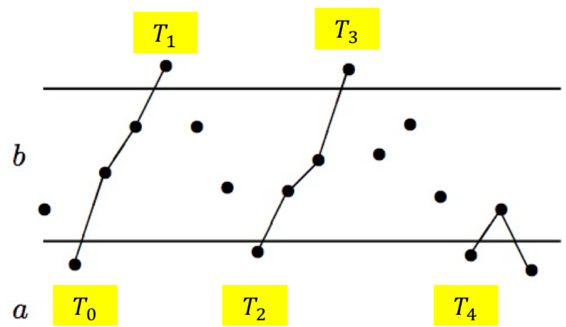


Figure 5.1: Upcrossings of $[a,b]$

We may check that $\{T_k\}_{k\geq 0}$ is a sequence of stopping times and is increasing:

*Proof.* The increasing property is trivial. To check $T_k$ is a stopping time, observe that

$$\{T_{2k-1}=m\}=\left(\bigcap_{t=T_{2k-2}+1}^{m-1}\{X_t<b\}\right)\cap\{X_m\geq b\}\in\mathcal{F}_m,$$
$$\{T_{2k}=m\}=\left(\bigcap_{t=T_{2k-1}+1}^{m-1}\{X_t>a\}\right)\cap\{X_m\leq a\}\in\mathcal{F}_m.$$

■

**Definition 5.1** [Upcrossing]

- If $T_{2k-1}<\infty$ a.s., then the sequence $X_{T_0},X_{T_1},\dots,X_{T_{2k-1}}$ is said to **upcrosses** the interval $[a,b]$ by $k$ times.
- Define $U_a^b[X;n]$ to be the number of upcrossing the interval $[a,b]$ by the process $X\triangleq\{X_k\}_{k\geq 0}$ up to time $n$. We can check that $U_a^b[X;n]$ is also a stopping time:

$$\{U_a^b[X;n]=j\}=\{T_{2j-1}\leq n<T_{2j+1}\}=\{T_{2j-1}\leq n\}\cap\{T_{2j+1}\leq n\}^c\in\mathcal{F}_n.$$

We can also assert that $X_{T_{2j}}\leq a$ if $T_{2j}<\infty$ a.s.; and $X_{T_{2j+1}}\geq b$ if $T_{2j+1}<\infty$ a.s.

**Theorem 5.1** — **Doob's Upcrossing Theorem.**

1. Suppose that $\{X_n\}_{n\geq 0}$ is a super-martingale, then for any $n\geq 1,k\geq 0$,

$$\mathbb{P}\left(U_a^b[X;n]\geq k+1\right)\leq\frac{1}{b-a}\mathbb{E}\left[(X_n-a)^-1\{U_a^b[X;n]=k\}\right].$$

As a result, $\mathbb{E}[U_a^b[X;n]]\leq\frac{1}{b-a}\mathbb{E}[(X_n-a)^-]$.

2. Suppose that $\{X_n\}_{n\geq 0}$ is a sub-martingale, then for any $n\geq 1,k\geq 0$,

$$\mathbb{P}\left(U_a^b[X;n]\geq k\right)\leq\frac{1}{b-a}\mathbb{E}\left[(X_n-a)^+1\{U_a^b[X;n]=k\}\right].$$

As a result, $\mathbb{E}[U_a^b[X;n]]\leq\frac{1}{b-a}\mathbb{E}[(X_n-a)^+]$.

*Proof.* 1. Considering that $\{X_n\}$ is a super-martingale and $T_{(2k+1)\wedge n}, T_{2k\wedge n}$ are two bounded stopping times, by Doob's Optional Sampling Theorem 4.5,

$$\begin{aligned}
&0 \geq \mathbb{E}[X_{T_{(2k+1)\wedge n}} - X_{T_{2k\wedge n}}] \\
&= \mathbb{E}[(X_{T_{(2k+1)\wedge n}} - X_{T_{2k\wedge n}})1\{n < T_{2k}\}] + \mathbb{E}[(X_{T_{(2k+1)\wedge n}} - X_{T_{2k\wedge n}})1\{T_{2k} \leq n < T_{2k+1}\}] \\
&+ \mathbb{E}[(X_{T_{(2k+1)\wedge n}} - X_{T_{2k\wedge n}})1\{n \geq T_{2k+1}\}] \\
&= \mathbb{E}[(X_n - X_{T_{2k}})1\{T_{2k} \leq n < T_{2k+1}\}] + \mathbb{E}[(X_{T_{2k+1}} - X_{T_{2k}})1\{n \geq T_{2k+1}\}] \\
&\geq \mathbb{E}[(X_n - a)1\{T_{2k} \leq n < T_{2k+1}\}] + \mathbb{E}[(b-a)1\{n \geq T_{2k+1}\}] \qquad (5.1a) \\
&\geq -\mathbb{E}[(X_n - a)^- 1\{T_{2k} \leq n < T_{2k+1}\}] + (b-a)\mathbb{P}\{n \geq T_{2k+1}\} \\
&\geq -\mathbb{E}[(X_n - a)^- 1\{T_{2k-1} \leq n < T_{2k+1}\}] + (b-a)\mathbb{P}\{n \geq T_{2k+1}\}
\end{aligned}$$

where (5.1a) is by the fact that $X_{T_{2k}} \leq a, X_{T_{2k+1}} \geq b$, a.s. Therefore,

$$\mathbb{P}\{n \geq T_{2k+1}\} \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)^- 1\{T_{2k-1} \leq n < T_{2k+1}\}]. \qquad (5.1b)$$

Note that $\{U_a^b[X;n] = k\} = \{T_{2k-1} \leq n < T_{2k+1}\}$ and

$$\{U_a^b[X;n] \geq k+1\} = \cup_{j \geq k+1}\{U_a^b[X;n] = j\} \subseteq \{T_{2k+1} \leq n\}.$$

Therefore, applying these two conditions on (5.1b) gives the desired inequality:

$$\mathbb{P}\{U_a^b[X;n] \geq k+1\} \leq \mathbb{P}\{n \geq T_{2k+1}\} \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)^- 1\{U_a^b[X;n] = k\}].$$

Summing up the inequality above for $k \geq 0$, we imply

$$\sum_{k \geq 0}\mathbb{P}\{U_a^b[X;n] \geq k+1\} \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)^-].$$

The LHS is essentially $\mathbb{E}[U_a^b[X;n]]$:

$$\sum_{k\geq 0}\mathbb{P}\{U_a^b[X;n]\geq k+1\}=\sum_{k\geq 0}\sum_{j=k+1}^{\infty}\mathbb{P}\{U_a^b[X;n]=j\}=\sum_{j=1}^{\infty}\sum_{k=0}^{j-1}\mathbb{P}\{U_a^b[X;n]=j\}$$
$$=\sum_{j=1}^{\infty}j\mathbb{P}\{U_a^b[X;n]=j\}=\mathbb{E}[U_a^b[X;n]].$$

The proof is complete.

2. We may use the similar technique to finish the proof on the second part. Apply Doob's Optional Sampling Theorem 4.5 on $T_{(2k-1)\wedge n}, T_{2k\wedge n}$ gives

$$\begin{aligned}
0 &\geq \mathbb{E}[X_{T_{(2k-1)\wedge n}}-X_{T_{2k\wedge n}}]\\
&=\mathbb{E}[(X_{T_{(2k-1)\wedge n}}-X_{T_{2k\wedge n}})1\{n<T_{2k-1}\}]+\mathbb{E}[(X_{T_{(2k-1)\wedge n}}-X_{T_{2k\wedge n}})1\{T_{2k-1}\leq n<T_{2k}\}]\\
&+\mathbb{E}[(X_{T_{(2k-1)\wedge n}}-X_{T_{2k\wedge n}})1\{n\geq T_{2k}\}]\\
&=\mathbb{E}[(X_{T_{2k-1}}-X_n)1\{T_{2k-1}\leq n<T_{2k}\}]+\mathbb{E}[(X_{T_{2k-1}}-X_{T_{2k}})1\{n\geq T_{2k}\}]\\
&\geq \mathbb{E}[(b-X_n)1\{T_{2k-1}\leq n<T_{2k}\}]+\mathbb{E}[(b-a)1\{n\geq T_{2k}\}]\\
&=\mathbb{E}[(a-X_n)1\{T_{2k-1}\leq n<T_{2k}\}]+\mathbb{E}[(b-a)1\{T_{2k-1}\leq n<T_{2k}\}]+\mathbb{E}[(b-a)1\{n\geq T_{2k}\}]\\
&=\mathbb{E}[(a-X_n)1\{T_{2k-1}\leq n<T_{2k}\}]+\mathbb{E}[(b-a)1\{n\geq T_{2k-1}\}]\\
&\geq -\mathbb{E}[(X_n-a)^+1\{T_{2k-1}\leq n<T_{2k}\}]+(b-a)\mathbb{P}\{n\geq T_{2k-1}\}
\end{aligned}$$

Or equivalently,

$$\mathbb{P}\{n\geq T_{2k-1}\}\leq\frac{1}{b-a}\mathbb{E}[(X_n-a)^+1\{T_{2k-1}\leq n<T_{2k}\}]$$

Considering that $\{U_a^b[X;n]=k\}=\{T_{2k-1}\leq n<T_{2k}\}$ and $\{U_a^b[X;n]\geq k\}\subseteq\{n\geq T_{2k-1}\}$, we imply

$$\mathbb{P}\{U_a^b[X;n]\geq k\}\leq\frac{1}{b-a}\mathbb{E}[(X_n-a)^+1\{U_a^b[X;n]=k\}]$$

Summing up for $k\geq 1$ both sides, we conclude the desired result.

■

**R** The upcrossing of $\{X_n\}$ on the interval is the same as teh upcrossing of $\{-X_n\}$ on $[-b,-a]$. Using this fact, we can assert that

- If $\{X_n\}$ is a super-martingale, for any $n \geq 1, k \geq 1$,

$$\mathbb{P}\left(U_a^b[X;n] \geq k\right) \leq \frac{1}{b-a}\mathbb{E}\left[(X_n - b)^{-}1\{U_a^b[X;n] = k\}\right].$$

- If $\{X_n\}$ is a sub-martingale, for any $n \geq 1, k \geq 1$,

$$\mathbb{P}\left(U_a^b[X;n] \geq k+1\right) \leq \frac{1}{b-a}\mathbb{E}\left[(X_n - b)^{+}1\{U_a^b[X;n] = k\}\right].$$

From the upcrossing inequality, we can easily get the result for the convergence of a martingale.

**Theorem 5.2 — Martingale Convergence Theorem.** Suppose that $\{X_n\}$ is a super-martingale which is $L^1$-bounded, i.e., $\sup_n \mathbb{E}[|X_n|] < \infty$. Then there exists a random variable $X_\infty$. such that $X_\infty \in \mathcal{L}^1$, and $X_n \to X_\infty$ a.s.

If we further assume that $\{X_n\}$ is lower bounded by zero, then $\mathbb{E}[X_\infty \mid \mathcal{F}_n] \leq X_n$ a.s., for any $n$.

*Proof.*
- Firstly, we study the limit of $\{U_a^b[X;n]\}_{n\geq 1}$, which is guaranteed to exist since $U_a^b[X;n]$ is increasing in $n$. Define $U_a^b[X] \triangleq \lim_{n\to\infty} U_a^b[X;n]$, then

$$\begin{aligned}
\mathbb{E}[U_a^b[X]] &\leq \liminf_{n\to\infty} \mathbb{E}[U_a^b[X;n]] && (5.2a)\\
&\leq \frac{1}{b-a}\liminf_{n\to\infty} \mathbb{E}[(X_n - a)^{-}] && (5.2b)\\
&\leq \frac{1}{b-a}\sup_n \mathbb{E}[(X_n - a)^{-}] \leq \frac{1}{b-a}\left(\sup_n \mathbb{E}[|X_n|] + |a|\right) < \infty.
\end{aligned}$$

where (5.2a) is by the Fatou's lemma and (5.2b) is by the Doob's upcrossing Theorem 5.1. As a result, $U_a^b[X] < \infty$ a.s.

- Note that the result in the first part holds for any rational $a, b$ with $a < b$, which means that $\mathbb{P}(U_a^b[x] < \infty) = 1, \forall a, b \in \mathbb{Q}, a < b$. Therefore, we can show that

$$\mathbb{P}[W] = 0, \quad \text{where } W = \bigcup_{a,b\in\mathbb{Q},a<b} \{\liminf X_n < a < b < \limsup X_n\}.$$

- Then we construct $X_\infty$ as follows. Because of the denseness of $\mathbb{Q}$, for $\omega \notin W$, $X_n(\omega)$ is convergent and define $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$. Otherwise, define $X_n(\omega) = 0$. Then the almost sure convergence of $X_n \to X_\infty$ is obtained.

  Also, we can check that $X_\infty \in \mathcal{L}^1$:

$$\mathbb{E}[|X_\infty|] \le \liminf_{n\to\infty} \mathbb{E}[|X_n|] \le \sup_n \mathbb{E}[|X_n|] < \infty.$$

- Given that $\{X_n\}$ is lower bounded by zero, the remaining result can be shown by upper bounding the integral $\int_A \mathbb{E}[X_\infty \mid \mathcal{F}_n]$ for any $A \in \mathcal{F}_n$:

$$\begin{aligned}
\int_A \mathbb{E}[X_\infty \mid \mathcal{F}_n] &= \int_A X_\infty \,\mathrm{d}\mathbb{P} \\
&\le \liminf_{m\to\infty} \int_A X_m \,\mathrm{d}\mathbb{P} && (5.3a) \\
&= \liminf_{m\to\infty} \int_A \mathbb{E}[X_m \mid \mathcal{F}_n] \,\mathrm{d}\mathbb{P} && (5.3b) \\
&\le \liminf_{m\to\infty} \int_A X_n \,\mathrm{d}\mathbb{P} = \int_A X_n \,\mathrm{d}\mathbb{P}. && (5.3c)
\end{aligned}$$

  where (5.3a) is by Fatou's lemma, (5.3b) is by the definition of conditional expectation, and (5.3c) is by the definition of super-martingale.

■

## 5.1.2. Continuous-time Martingales

Now we discuss the concepts of martingales, super-martingales, sub-martingales for continuous-time.

**Definition 5.2** [Martingale] Let $\{X_t\}_{t\ge0}$ be an adapted process on filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0}, \mathbb{P})$. A stochastic process $\{X_t\}_{t\ge0}$ is called a **martingale** if

1. $X_t \in \mathcal{L}^1, \forall t$;
2. $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ a.s., for all $0 \leq s \leq t$.

If in the last definition, "$=$" is replaced by "$\leq$" or "$\geq$", then $\{X_t\}_{t \geq 0}$ is said to be a **supermartingale** or **submartingale**, respectively.

**Definition 5.3** [Optional Time]

1. A mapping $T : \Omega \to [0, \infty]$ is called an $\{\mathcal{F}_t\}$-stopping time if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.
2. A mapping $T : \Omega \to [0, \infty]$ is called an $\{\mathcal{F}_t\}$-optional time if $\{T < t\} \in \mathcal{F}_t$ for all $t > 0$.

It is easy to check that a stopping time is always an optional time. Now we discuss an example about a specific optional time.

**Example 5.1** Let $T$ be an optional time. For each $n \geq 1$, define the step-function mapping

$$T_n = \begin{cases} \dfrac{k}{2^n}, & \text{if } (k-1)/2^n \leq T < k/2^n, k = 1, 2, \dots \\ \infty, & \text{if } T = \infty \end{cases}$$

Then $\{T_n\}_{n \geq 1}$ are stopping times with $T_n \downarrow T$:

- To show $T_n$ is a stopping time, study the set

$$\begin{aligned} \{T_n \leq t\} &= \bigcup_{k \geq 1} \left[ \{T_n \leq t\} \cap \left\{ \frac{k-1}{2^n} \leq T < \frac{k}{2^n} \right\} \right] \\ &= \bigcup_{k=1}^{\lfloor t \cdot 2^n \rfloor} \left[ \{T_n \leq t\} \cap \left\{ \frac{k-1}{2^n} \leq T < \frac{k}{2^n} \right\} \right] = \bigcup_{k=1}^{\lfloor t \cdot 2^n \rfloor} \left\{ \frac{k-1}{2^n} \leq T < \frac{k}{2^n} \right\}. \end{aligned}$$

  Since $T$ is an optional time,

$$\left\{ \frac{k-1}{2^n} \leq T < \frac{k}{2^n} \right\} \in \mathcal{F}_{k/2^n} \subseteq \mathcal{F}_t, \quad \forall k \leq t \cdot 2^n \implies \{T_n \leq t\} \in \mathcal{F}_t, \forall t.$$

- The result for $T_n \downarrow T$ can be found in MAT3006 knowledge:

  Daniel Wong, Jie Wang. (2019) Lecture Notes for MAT3006: Real Analysis, Lecture 19, Proposition 10.4. Available at the link

  https://walterbabyrudin.github.io/information/Updates/Updates.html

**Definition 5.4**
1. Let $\{\mathcal{F}_t\}_{t\ge 0}$ be a filteration. Define $\mathcal{F}_{t+} \triangleq \bigcap_{s>t} \mathcal{F}_s$. Then $\{\mathcal{F}_{t+}\}_{t\ge 0}$ is also a filteration. A filteration $\{\mathcal{F}_t\}_{t\ge 0}$ is said to be **right-continuous** if $\mathcal{F}_t = \mathcal{F}_{t+}, \forall t$.

   R $\mathcal{F}_{t+}$ can be interpreted as the information available immediately after time $t$. We can show that $\mathcal{F}_{t+}$ is a $\sigma$-algebra and $\mathcal{F}_{t+} \supseteq \mathcal{F}_t$.
   When the filteration is right-continuous, a stopping time is the same as an optional time.

2. A filteration $\{\mathcal{F}_t\}_{t\ge 0}$ is said to be **complete** if each $\mathcal{F}_t$ contains all $\mathbb{P}$-null sets in $\mathcal{F}$.
3. A filteration $\{\mathcal{F}_t\}_{t\ge 0}$ is called an **augmented filteration**, or said to satisfy the **usual conditions**, if it is complete and right-continuous.

# 5.2. Thursday

## 5.2.1. Theorems for Continuous Time Martingales

**Definition 5.5** [Stopping Time $\sigma$-algebra] Let $T$ be an $\{\mathcal{F}_t\}$-stopping time. Define

$$\mathcal{F}_T \triangleq \{A \in \mathcal{F} : \ A \cap \{T \le t\} \in \mathcal{F}_t, \forall t \ge 0\}.$$

Then $\mathcal{F}_T$ is the $\sigma$-algebra for $T$, containing the information available up to time $T$.

**Proposition 5.1** Let $\{\mathcal{F}_t\}$ be a filtration. Define

$$\mathcal{F}_{T+} \triangleq \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_{t+}, \forall t \geq 0\}$$
$$\mathcal{G}_T \triangleq \{A \in \mathcal{F} : A \cap \{T < t\} \in \mathcal{F}_t, \forall t > 0\}$$

We can show that $\mathcal{F}_{T+} = \mathcal{G}_T$.

**Theorem 5.3**

1. Let $\{X_t\}_{t\geq 0}$ be a martingale and the filteration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfies the usual condition. Then there exists a version of $\{X_t\}_{t\geq 0}$, which is right-continuous with left limits, denoted as $\{\tilde{X}_t\}_{t\geq 0}$. Then $\{\tilde{X}_t\}_{t\geq 0}$ is a right-continuous martingales w.r.t. $\{\mathcal{F}_t\}_{t\geq 0}$.

   R The martingales we encountered are basically right-continuous by construction.

2. (Maximal Inequality) Denote $X_t^* \triangleq \sup_{s\leq t} |X_s|$ as the running maximal. Then for any $t > 0$ and $\lambda > 0$,

$$\mathbb{P}(X_t^* \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}[|X_t|].$$

3. (Convergence Theorem) Let $\{X_t\}_{t\geq 0}$ be a right-continuous super-martingale, which is $L^1$-bounded. Then there exists a random variable $X_\infty$ such that $X_\infty \in \mathcal{L}^1$ and $X_n \to X_\infty$ a.s.

**Theorem 5.4** — **Doob's Optional Sampling Theorem for Continuous-time Martingales.** Let $\{X_t\}_{t\geq 0}$ be a right-continuous martingale with the last element $X_\infty$, i.e., $X_\infty \in \mathcal{L}^1$ and $\mathbb{E}[X_\infty \mid \mathcal{F}_t] = X_t$ a.s. for any $t \geq 0$. Let $S \leq T$ be two $\{\mathcal{F}_t\}$-optional times. Then $\mathbb{E}[X_T \mid \mathcal{F}_{S+}] = X_S$ a.s. Specifically, if $S$ is a stopping time, we replace $\mathcal{F}_{S+}$ by $\mathcal{F}_S$. In particular, $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Let's first show a necessary and sufficient condition for the existence of $X_\infty$:

**Proposition 5.2** A last element $X_\infty$ exists if and only if $\{X_t\}_{t\geq 0}$ is uniformly integrable.

*Proof.* Assume that $\{X_t\}_{t\geq 0}$ is uniformly integrable, which implies that $\{X_t\}_{t\geq 0}$ is $L^1$-bounded. By the martingale convergence theorem 5.3, there exists a random variable $X_\infty$ such that $X_t \to X_\infty$ a.s. and $X_\infty \in \mathcal{L}^1$. In particular, $X_t \xrightarrow{P} X_\infty$. Together with the uniform integrability of $\{X_t\}_{t\geq 0}$, we can assert that $X_t \to X_\infty$ in $L^1$. Finally, we check that $\mathbb{E}[X_\infty \mid \mathcal{F}_u] = X_u$, a.s., for any $u \geq 0$, i.e., for any $A \in \mathcal{F}_u$, we have

$$\begin{aligned}\int_A X_\infty \,\mathrm{d}\mathbb{P} &= \lim_{t\to\infty}\int_A X_t \,\mathrm{d}\mathbb{P} \\ &= \lim_{t\to\infty}\int_A \mathbb{E}[X_t \mid \mathcal{F}_u]\,\mathrm{d}\mathbb{P} = \lim_{t\to\infty}\int_A X_u \,\mathrm{d}\mathbb{P} = \int_A X_u \,\mathrm{d}\mathbb{P}.\end{aligned}$$

Now we assume that the last element exists. Then $\{X_t\}$ is a collection of conditional expectations of $X_\infty$. Applying Theorem 3.1 gives the desired result.

∎

Now we begin to prove Theorem 5.4.

*Proof.*
- Firstly, construct the approximation of $S,T$ and argue the similar optional sampling results hold:

$$S_n = \begin{cases} \dfrac{k}{2^n}, & \text{if } (k-1)/2^n \leq S < k/2^n, k = 1,2,\ldots \\ \infty, & \text{if } S = \infty \end{cases}$$
$$T_n = \begin{cases} \dfrac{k}{2^n}, & \text{if } (k-1)/2^n \leq T < k/2^n, k = 1,2,\ldots \\ \infty, & \text{if } T = \infty \end{cases}$$

  By Example 5.1, $\{S_n\},\{T_n\}$ are two sequences of stopping times with $S_n \downarrow S, T_n \downarrow T$. Moroever, for each $n \geq 1$, $S_n \leq T_n$ a.s., taking values in a countable set. Since $\{X_t\}_{t\geq 0}$ is uniformly integrable, applying the discrete-time optional sampling theorem,

$$\mathbb{E}[X_{T_n} \mid \mathcal{F}_{S_n}] = X_{S_n}, \quad \text{a.s.}$$

  Therefore, for any $A \in \mathcal{F}_{S_n}$, $\int_A X_{T_n}\,\mathrm{d}\mathbb{P} = \int_A X_{S_n}\,\mathrm{d}\mathbb{P}$.

- We claim that $\mathcal{F}_{S+} = \cap_{n\geq 1}\mathcal{F}_{S_n}$. Therefore, for any $A \in \mathcal{F}_{S+}$,

$$\int_A X_{T_n} \,\mathrm{d}\mathbb{P} = \int_A X_{S_n} \,\mathrm{d}\mathbb{P}. \tag{5.4}$$

  Here $\{X_{S_n}\}_{n\geq 1}$ is called a backward (discrete) martingale w.r.t. $\{\mathcal{F}_{S_n}\}_{n=1}^{\infty}$, i.e., $\mathbb{E}[X_{S_n} \mid \mathcal{F}_{S_{n+1}}] = X_{S_{n+1}}$. Therefore, for any $A \in \mathcal{F}_{S_{n+1}}$,

$$\int_A X_{S_n} \,\mathrm{d}\mathbb{P} = \int_A X_{S_{n+1}} \,\mathrm{d}\mathbb{P}.$$

  Thus $\mathbb{E}[X_{S_{n+1}}] = \mathbb{E}[X_{S_n}] = \mathbb{E}[X_0] > -\infty$ and $\lim_{n\to\infty} \mathbb{E}[X_{S_n}] > -\infty$.

- We also claim that $\{X_{S_n}\}_{n\geq 1}$ is uniformly integrable, which implies $\{X_{T_n}\}_{n\geq 1}$ is uniformly integrable. Since $\{X_t\}$ is right-continuous and $T_n \downarrow T, S_n \downarrow S$, the limit of $\{X_{T_n}\}$ and $\{X_{S_n}\}$ always exist:

$$X_T \triangleq \lim_{n\to\infty} X_{T_n} \text{ a.s.,} \qquad X_S \triangleq \lim_{n\to\infty} X_{S_n} \text{ a.s.}$$

  In particular, $X_{T_n} \to X_T$ in prob. and $X_{S_n} \to X_S$ in prob. By Theorem 3.3, $X_{T_n} \to X_T$ in $L^1$ and $X_{S_n} \to X_S$ in $L^1$.

- Then we can show that $\mathbb{E}[X_T \mid \mathcal{F}_{S+}] = X_S$ a.s. as the follwoing. For any $A \in \mathcal{F}_{S+}$,

$$\begin{aligned}
\int_A \mathbb{E}[X_T \mid \mathcal{F}_{S+}] \,\mathrm{d}\mathbb{P} &= \int_A X_T \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_A X_{T_n} \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_A X_{S_n} \,\mathrm{d}\mathbb{P} \\
&= \int_A X_S \,\mathrm{d}\mathbb{P}
\end{aligned}$$

  where the second and the last equality is because of the $L^1$ convergence, and the third equality is because of (5.4).

- Provided that $S$ is a stopping time, $S \leq S_n$ implies $\mathcal{F}_S \subseteq \mathcal{F}_{S_n}$. Therefore, for any $A \in \mathcal{F}_S$, $\int_A X_{T_n} \,\mathrm{d}\mathbb{P} = \int_A X_{S_n} \,\mathrm{d}\mathbb{P}$. The proof is complete.

■

# Chapter 6

# Week6

## 6.1. Tuesday

At the beginning of this lecture, let's fill the gap for the Theorem 5.3.

**Proposition 6.1** Suppose that $T_n$ is a positive stopping time and $T < T_n$ conditioned on the event $\{T < \infty\}, \forall n \geq 1$, where $T \triangleq \inf_n T_n$. Then $\mathcal{F}_{T+} = \cap_{n=1}^{\infty} \mathcal{F}_{T_n}$, where

$$\begin{aligned}\mathcal{F}_{T+} &\triangleq \{A \in \mathcal{F} : \ A \cap \{T \leq t\} \in \mathcal{F}_{t+}, \ \forall t \geq 0\} \\ &\triangleq \{A \in \mathcal{F} : \ A \cap \{T < t\} \in \mathcal{F}_{t}, \ \forall t > 0\}\end{aligned}$$

*Proof.* Firstly, we show that $T$ is an optional time:

$$\begin{aligned}\{T < t\} = \{T \geq t\}^c &= \{\cap_n \{T_n > t\}\}^c \\ &= \cup_n \{T_n > t\}^c = \cup_n \{T_n \leq t\}\end{aligned}$$

Since $T_n$ is a stopping time, $\{T_n \leq t\} \in \mathcal{F}_t, \forall t$, which implies that $T$ is an optional time.

- Suppose that $A \in \cap_{n=1}^{\infty} \mathcal{F}_{T_n}$, then $A \cap \{T_n \leq t\} \in \mathcal{F}_t, \forall t, \forall n \geq 1$. As a result,

  $$\begin{aligned}\mathcal{F}_t \ni \bigcup_n \{A \cap \{T_n \leq t\}\} &= A \cap [\cup_n \{T_n \leq t\}] \\ &= A \cap \{T < t\}, \quad \forall t.\end{aligned}$$

  In other words, $A \in \mathcal{F}_{T+}$.

- Suppose that $A \in \mathcal{F}_{T+}$, then $A \cap \{T < t\} \in \mathcal{F}_t, \ \forall t$. Moreover, $\{T_n \leq t\} \in \mathcal{F}_{T_n}$.

Therefore,

$$\mathcal{F}_t \ni (A \cap \{T < t\}) \cup (A \cap \{T_n \leq t\}) = A \cap \{T_n \leq t\}, \, \forall t > 0.$$

In other words, $A \in \cap_{n=1}^{\infty} F_{T_n}$.

■

**Proposition 6.2** Let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be a decreasing sequence of sub-$\sigma$-algebras of $\mathcal{F}$:

$$\mathcal{F} \supseteq \mathcal{F}_1 \supseteq \cdots \supseteq \mathcal{F}_n \supseteq \cdots.$$

Suppose that $\{X_n\}_{n=1}^{\infty}$ is a backward submartingale w.r.t $\{\mathcal{F}_n\}_{n=1}^{\infty}$, i.e., i) $X_n$ is $\mathcal{F}_n$-measurable, ii) $\mathbb{E}[|X_n|] < \infty$, iii) $\mathbb{E}[X_n \mid \mathcal{F}_{n+1}] \geq X_{n+1}$, a.s. If $\lim_{n\to\infty} \mathbb{E}[X_n] > -\infty$, then the sequence $\{X_n\}_{n=1}^{\infty}$ is UI.

*Proof.* Note that the limit of $\mathbb{E}[X_n]$ exists since it is decreasing. Denote $c \triangleq \lim_{n\to\infty} \mathbb{E}[X_n]$. We can argue the uniform convergence of $\mathbb{P}(|X_n| > \lambda)$ as the following:

$$\begin{aligned}
\forall \lambda > 0, \, \mathbb{P}(|X_n| > \lambda) &\leq \frac{\mathbb{E}[|X_n|]}{\lambda} \\
&= \frac{1}{\lambda}\left(2\mathbb{E}[X_n^+] - \mathbb{E}[X_n]\right) \\
&\leq \frac{1}{\lambda}\left(2\mathbb{E}[X_1^+] - c\right) < \infty.
\end{aligned}$$

where the first inequality is because that $\mathbb{E}[X_n] \downarrow c$ and $\{X_n^+\}$ is a backward submartingale. In other words, $\mathbb{P}(|X_n| > \lambda)$ uniformly converges to 0 as $\lambda \to \infty$.

Now we begin to show the UI of $\{X_n\}$. Applying the useful proposition 3.1 on $X_1$, for any $\varepsilon > 0$, there exists $\delta$ such that for any $A \in \mathcal{F}_1, \mathbb{P}(A) < \delta$,

$$\int_A |X_1| \, \mathrm{d}\mathbb{P} < \varepsilon. \tag{6.1}$$

We can choose $\lambda$ sufficiently large such that $\mathbb{P}(|X_n| > \lambda) < \delta, \forall n$. As a result,

$$\begin{aligned}\int_{\{|X_n|>\lambda\}} |X_n| \, d\mathbb{P} &\le \int_{\{|X_n|>\lambda\}} |X_{n-1}| \, d\mathbb{P} \le \cdots \\ &\le \int_{\{|X_n|>\lambda\}} |X_1| \, d\mathbb{P} < \varepsilon\end{aligned}$$

where the first inequality is because that $\mathbb{E}[|X_{n-1}|\mathcal{F}_n] \ge |X_n|$, and the last inequality is because of (6.1). Therefore, $\int_{\{|X_n|>\lambda\}} |X_1| \, d\mathbb{P} < \varepsilon$ for any $n$. The proof is complete. ■

### 6.1.1. Localization

The concepts of stopping times provide a tool of "localizing" quantities.

**Definition 6.1** [Stopped Process] Suppose that $\{X_t\}_{t\ge 0}$ is an $\{\mathcal{F}_t\}$-adapted process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, and $T$ is a stopping time. Define the stopped process $\{X_{t\wedge T}\}_{t\ge 0}$ such that

$$X_{t\wedge T}(\omega) = X_{t\wedge T(\omega)}(\omega), \quad \forall \omega \in \Omega.$$

Note that $\{X_{t\wedge T}\}_{t\ge 0}$ is also an $\{\mathcal{F}_t\}$-adapted process.

**Definition 6.2** [Local Martingale] An $\{\mathcal{F}_t\}$-adapted process $\{X_t\}_{t\ge 0}$ is called a **local martingale** if there is an increasing sequence of stopping times $\{T_n\}_{n\ge 0}$ and $T_n \uparrow \infty$ a.s., such that $\{X_{t\wedge T_n}\}_{t\ge 0}$ is a martingale for each $n$, w.r.t. $\{\mathcal{F}_t\}$.

Note that a martingale is a local martingale. Now we give a sufficient condition such that local martingale can be a martingale.

**Theorem 6.1** Suppose that $\{X_t\}_{t\ge 0}$ is an $\{\mathcal{F}_t\}$-adapted local martingale, and there is a sequence $\{T_n\}_{n\ge 0}$ that reduces $\{X_t\}_{t\ge 0}$:

$$\{X_{t\wedge T_n}\}_{t\ge 0} \text{ is a martingale for each } n.$$

Suppose that $\mathbb{E}[\sup_n |X_{t\wedge T}|] < \infty$ for each $t$, then $\{X_t\}_{t\ge 0}$ is a martingale.

*Proof.* Considering that i) $X_{t\wedge T_n} \to X_t$ a.s. because $T_n \uparrow \infty$, ii) $|X_{t\wedge T_n}| \leq \sup_n |X_{t\wedge T_n}|$, with the random variable $\sup_n |X_{t\wedge T_n}|$ integrable, we can apply the dominated convergence theorem to show that $X_{t\wedge T_n} \xrightarrow{L^1} X_t, \forall t$, as $n \to \infty$.

Now we check that $\{X_t\}_{t\geq 0}$ is a martingale, i.e., for any $A \in \mathcal{F}_s, 0 \leq s \leq t$, we have

$$
\begin{aligned}
\int_A \mathbb{E}[X_t \mid \mathcal{F}_s] \,\mathrm{d}\mathbb{P} &= \int_A X_t \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_A X_{t\wedge T_n} \,\mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_A X_{s\wedge T_n} \,\mathrm{d}\mathbb{P} \\
&= \int_A X_s \,\mathrm{d}\mathbb{P}
\end{aligned}
$$

where the third inequality is because that $\mathbb{E}[X_{t\wedge T_n} \mid \mathcal{F}_s] = X_{s\wedge T_n}$. ■

## 6.1.2. Introduction to Brownian Motion

Brownian motion is a mathematical model of random movements observed by botanist Robert Brown. Now we give a way for constructing the Brownian motion.

**Definition 6.3** [Brownian Motion] A stochastic process $B = \{B_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in $\mathbb{R}$, is called a Brownian motion if:

1. $\mathbb{P}(B_0 = 0) = 1$;
2. (Independent Increments) For every $0 \leq t_1 < \cdots < t_k < \infty$ and $x_1, x_2, \dots, x_{k-1} \in \mathbb{R}$,

$$
\mathbb{P}\left(B_{t_2} - B_{t_1} \leq x_1, \dots, B_{t_k} - B_{t_{k-1}} \leq x_{k-1}\right) = \prod_{2\leq j\leq k} \mathbb{P}(B_{t_j} - B_{t_{j-1}} \leq x_{j-1})
$$

3. (Normal Distribution)For each $0 \leq s < t$, $B_t - B_s$ follows normal distribution with mean 0 and variance $\sigma^2(t-s)$, where $\sigma > 0$.
4. Almost all the sample paths of $\{B_t\}_{t\geq 0}$ are continuous. In particular, when $\sigma = 1$, we call it the standard Brownian motion.

■

R In some situations, the first condition may not be satisfied. Instead, the process may start at a non-zero point $x$. Then we write such a process $\{x + B_t\}$.

**Definition 6.4** [Canonical Wiener Measure] Let the sample space be $\Omega = C[0,\infty)$, and its associated topology is $\mathcal{T}$. Define the Borel $\sigma$-algebra $\mathcal{B} = \sigma(\mathcal{T})$. Thus $\omega \in \Omega$ is a continuos function with support $[0,\infty)$. Define $B_t(\omega) = \omega(t)$. A probability measure $\mathbb{P}$ on $(C[0,\infty),\mathcal{B})$ is called a Wiener measure if conditions (1)-(3) in Definition 6.3 are satisfied. With such a probability measure, $\{B_t\}_{t\ge 0}$ is said to be a Brownian motion on $(C[0,\infty),\mathcal{B},\mathbb{P})$. ■

**Theorem 6.2** — **Existence and Uniqueness of Wiener Measure.** For each $\sigma > 0$, there exists a unique Wiener measure in Definition 6.4.

# 6.2. Thursday

## 6.2.1. Properties of Brownian Motion

**Proposition 6.3** Suppose that $\{B_t\}_{t\ge 0}$ is a standard Brownian motion, then it satisfies the following properties:

1. Joint distribution: Fix $0 \le t_1 < t_2 < \cdots < t_k$. Given $x_1, x_2, \dots, x_k \in \mathbb{R}$, the joint density of $(B_{t_1}, B_{t_2}, \dots, B_{t_k})$ in $(x_1, x_2, \dots, x_k)$ is equal to the joint density of $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}})$ in $(x_1, x_2 - x_1, \dots, x_k - x_{k-1})$, which is
$$\prod_{j=2}^{k} \frac{1}{\sqrt{2\pi(t_j - t_{j-1})}} \exp\left(-\frac{(x_j - x_{j-1})^2}{2(t_j - t_{j-1})}\right).$$
2. Stationary: For any $s > 0$, define $B_t^s = B_{t+s} - B_s, t \ge 0$. Then $\{B_t^s\}_{t\ge 0}$ is a Brownian motion.
3. Scaling:
   - For each $c \neq 0$, $\{cB_t\}_{t\ge 0}$ is a Brownian motion with variance $c^2$;

- For each $c > 0$, $\{B_{t/c}\}_{t\geq 0}$ is a Brownian motion with variance $1/c$;
- (Scaling invariance / self-similarity) By previous two properties, $\{\sqrt{c}B_{t/c}\}_{t\geq 0}$ is a standard Brownian motion, $c > 0$.

4. Covariance: for fixed $0 \leq s \leq t$, $\mathrm{cov}(B_t, B_s) = s$.
5. Time reversal: Given a standard Brownian motion $\{B_t\}$, define a new process $\{\hat{B}_t\}$ with $\hat{B}_t = tB_{1/t}$ for $t > 0$, and $\hat{B}_0 = 0$. Then $\{\hat{B}_t\}$ is a standard Brownian motion.

*Proof on the first four parts.* 1) can be shown by the indepedent increments and normal distribution properties of Brownian motion; 2), 3) can be shown by checking the definition of Browninan motion; 4) can be shown by directly computing the covariance:

$$\begin{aligned}
\mathrm{cov}(B_t, B_s) &= \mathbb{E}[B_t B_s] - \mathbb{E}[B_t]\mathbb{E}[B_s] \\
&= \mathbb{E}[(B_t - B_s + B_s)B_s] \\
&= \mathbb{E}[(B_t - B_s)B_s] + \mathbb{E}[B_s^2] \\
&= \mathbb{E}[B_t - B_s]\mathbb{E}[B_s] + \mathbb{E}[B_s^2] \\
&= s.
\end{aligned}$$

■

*Proof on the time reversal part.* We need to check those four conditions in Definition 6.3 are satisfied. The condition (1) is trivial.

- Now check condition (3). Fix $0 < s < t$, then

$$\hat{B}_t - \hat{B}_s = tB_{1/t} - sB_{1/s} = (t-s)B_{1/t} + s(B_{1/t} - B_{1/s}).$$

  Since $B_{1/t} - B_{1/s} \sim \mathcal{N}(0, 1/s - 1/t)$, we imply $s(B_{1/t} - B_{1/s}) \sim \mathcal{N}(0, s^2(1/s - 1/t))$. Moreover, $(t-s)B_{1/t} \sim \mathcal{N}(0, (t-s)^2/t)$. By the increment independent property, this term is independent with $-s(B_{1/s} - B_{1/t})$. Therefore, $(t-s)B_{1/t} - s(B_{1/s} - B_{1/t})$ is normally distributed with mean 0 and variance $(t-s)^2/t + s^2(1/s - 1/t) = t - s$.

- In order to check condition (2), fix $t_1 < t_2 < t_3$. It suffices to check $\hat{B}_{t_3} - \hat{B}_{t_2}$ and $\hat{B}_{t_2} - \hat{B}_{t_1}$ are independent. Considering that these two r.v.'s are jointly normal, it suffices to verify their covariance is zero:

$$\begin{aligned}
t_3 - t_1 &= \mathrm{Var}(\hat{B}_{t_3} - \hat{B}_{t_1}) \\
&= \mathrm{Var}(\hat{B}_{t_3} - \hat{B}_{t_2} + \hat{B}_{t_2} - \hat{B}_{t_1}) \\
&= \mathrm{Var}(\hat{B}_{t_3} - \hat{B}_{t_2}) + \mathrm{Var}(\hat{B}_{t_2} - \hat{B}_{t_1}) + 2\mathrm{Cov}(\hat{B}_{t_3} - \hat{B}_{t_2}, \hat{B}_{t_2} - \hat{B}_{t_1}) \\
&= t_3 - t_2 + t_2 - t_1 + 2\mathrm{Cov}(\hat{B}_{t_3} - \hat{B}_{t_2}, \hat{B}_{t_2} - \hat{B}_{t_1})
\end{aligned}$$

  which implies the desired result.

- Finally we check the condition (4). Since the continuity of $\{\hat{B}_t\}$ holds at any $t > 0$, it suffices to check $t = 0$ is also a continuous point, i.e., almost surely $\lim_{t\to 0} \hat{B}_t = \lim_{t\to 0} tB_{1/t}(\omega) = 0$.

  - Firstly show that $\hat{B}_t \to 0$ when $t = 1/n, n \to \infty$. For fixed $n \in \mathbb{Z}_+$, $B_n = \sum_{j=1}^n (B_j - B_{j-1})$, i.e., $B_n$ is a sum of i.i.d. random variables with standard normal distribution. By strong law of large numbers, $B_n/n \to 0$ a.s. as $n \to \infty$.

  - Then we show that $\hat{B}_t \to 0$ for other values of $t$. Fix any $s \in (n, n+1)$, note that

$$\begin{aligned}
\left|\frac{B_s}{s} - \frac{B_n}{n}\right| &\le \left|\frac{B_s}{s} - \frac{B_n}{s}\right| + \left|\frac{B_n}{s} - \frac{B_n}{n}\right| \\
&= \frac{1}{s}|B_s - B_n| + |B_n|\left|\frac{1}{s} - \frac{1}{n}\right| \\
&\le \frac{1}{n}\sup_{n\le s\le n+1}|B_s - B_n| + \frac{1}{n^2}|B_n|
\end{aligned}$$

    Since $B_n/n \to 0$ a.s., we have $B_n/n^2 \to 0$ a.s. Define $Z_n \triangleq \sup_{n\le s\le n+1}|B_s - B_n|$, then

$$\sup_{n<s<n+1}\left|\frac{B_s}{s} - \frac{B_n}{n}\right| \le \frac{Z_n}{n} + \frac{1}{n^2}|B_n|$$

We claim that for any $\varepsilon > 0$,

$$\mathbb{P}\left\{\omega \in \Omega : \frac{Z_n(\omega)}{n} > \varepsilon, \text{ infinitely often}\right\} = 0.$$

Then $\frac{Z_n(\omega)}{n} \to 0$ for almost all $\omega \in \Omega$. As a result,

$$\sup_{n<s<n+1} \left| \frac{B_s}{s} - \frac{B_n}{n} \right| \to 0, \text{ a.s.,}$$

which implies the desired continuity result.

– Then we need to show the correctness of our claim. By the stationary and independent increments of Brownian motion, $\sup_{n \leq s \leq n+1} |B_s - B_n|$ has the same distribution as $\sup_{0 \leq s \leq 1} |B_s|$, then

$$\begin{aligned}
\mathbb{E}[Z_0] = \mathbb{E}[\sup_{0\leq s\leq 1} |B_s|] &= \int_0^\infty \mathbb{P}(Z_0 > x)\,\mathrm{d}x \\
&= \sum_{n=0}^{\infty} \int_{n\varepsilon}^{(n+1)\varepsilon} \mathbb{P}(Z_0 > x)\,\mathrm{d}x \\
&\geq \sum_{n=0}^{\infty} \int_{n\varepsilon}^{(n+1)\varepsilon} \mathbb{P}(Z_0 > (n+1)\varepsilon)\,\mathrm{d}x \\
&= \sum_{n=0}^{\infty} \varepsilon \mathbb{P}(Z_0 > (n+1)\varepsilon) = \varepsilon \sum_{n=1}^{\infty} \mathbb{P}(Z_0 > n\varepsilon) \\
&= \varepsilon \sum_{n=1}^{\infty} \mathbb{P}(Z_0/n > \varepsilon) \\
&= \varepsilon \sum_{n=1}^{\infty} \mathbb{P}(Z_n/n > \varepsilon)
\end{aligned}$$

We claim that $\mathbb{E}[Z_0] < \infty$ (which will be shown in the next lecture), which implies that

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n/n > \varepsilon) < \infty.$$

Applying the Borel-Cantelli Lemma gives the desired result.

■

# Chapter 7

# Week7

## 7.1. Tuesday

### 7.1.1. Reflection Principle

Consider a Brownian motion $\{B_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let the filteration $\{\mathcal{F}_t\}_{t\geq 0}$ be the natural filteration, i.e., $\mathcal{F}_t = \sigma(B_u : u \leq t)$. Suppose that $a > 0$, define

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

Then $T_a$ is the first time that Brownian motion hits level $a$. By convention, $\inf \emptyset = +\infty$.

**Theorem 7.1** The hitting time is finite almost surely:

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

*Proof.* Based on $B_t$, define new stochastic process $\{Z_t^\theta\}$ with

$$Z_t^\theta = \exp\left(\theta B_t - \frac{\theta^2 t}{2}\right), \quad t \geq 0.$$

As a result,

- Since $\mathbb{E}[e^{\theta X}] = e^{\theta^2\sigma^2/2}$ for $X \sim \mathcal{N}(0, \sigma^2)$,

$$\mathbb{E}[|Z_t^\theta|] = \mathbb{E}\exp\left(\theta B_t - \frac{\theta^2 t}{2}\right) = 1, \forall t$$

- For any $0 \le u < t$, we have

$$\begin{aligned}
\mathbb{E}[Z_t^\theta \mid \mathcal{F}_u] &= \mathbb{E}\left[\exp\left(\theta B_t - \frac{\theta^2}{t}\right)\middle|\mathcal{F}_u\right] \\
&= \mathbb{E}\left[\exp\left(\theta(B_t - B_u) - \frac{\theta^2(t-u)}{2}\right)\exp\left(\theta B_u - \frac{\theta^2 u}{2}\right)\middle|\mathcal{F}_u\right] \\
&= \exp\left(\theta B_u - \frac{\theta^2 u}{2}\right)\cdot\mathbb{E}\left[\exp\left(\theta(B_t - B_u) - \frac{\theta^2(t-u)}{2}\right)\middle|\mathcal{F}_u\right] \\
&= \exp\left(\theta B_u - \frac{\theta^2 u}{2}\right)\cdot\mathbb{E}\left[\exp\left(\theta B_{t-u} - \frac{\theta^2(t-u)}{2}\right)\right] = Z_u^\theta
\end{aligned}$$

Therefore, $\{Z_t^\theta\}$ is a martingale w.r.t. $\{\mathcal{F}_t\}_{t\ge 0}$. Now we compute $\lim_{t\to\infty}\mathbb{E}[Z_{T_a}^\theta 1\{T_a < \infty\}]$ as the following:

- Since i) $Z_{t\wedge T_a}^\theta 1\{T_a < \infty\} \xrightarrow{a.s.} Z_{T_a}^\theta 1\{T_a < \infty\}$ as $t \to \infty$, ii) $|Z_{t\wedge T_a}^\theta| \le e^{\theta a}$ for any $t$, by bounded convergence theorem,

$$\lim_{t\to\infty}\mathbb{E}[Z_{t\wedge T_a}^\theta 1\{T_a < \infty\}] = \mathbb{E}[Z_{T_a}^\theta 1\{T_a < \infty\}]$$

- Since $Z_{t\wedge T_a}^\theta 1\{T_a = \infty\} = Z_t^\theta 1\{T_a = \infty\}$ for any $t$, and

$$Z_t^\theta 1\{T_a = \infty\} \le e^{\theta a - \theta^2 t/2} \to 0,$$

  by bounded convergence theorem,

$$\lim_{t\to\infty}\mathbb{E}[Z_{t\wedge T_a}^\theta 1\{T_a = \infty\}] = \lim_{t\to\infty}\mathbb{E}[Z_t^\theta 1\{T_a = \infty\}] = 0.$$

On the other hand, since the stopped process $\{Z_{t\wedge T_a}^\theta\}$ is a martingale,

$$\mathbb{E}[Z_{t\wedge T_a}^\theta] = \mathbb{E}[Z_0^\theta] = 1, \quad \forall t.$$

It follows that

$$\begin{aligned}
1 &= \lim_{t\to\infty}\mathbb{E}[Z_{t\wedge T_a}^\theta 1\{T_a < \infty\}] + \lim_{t\to\infty}\mathbb{E}[Z_{t\wedge T_a}^\theta 1\{T_a = \infty\}] \\
&= \mathbb{E}[Z_{T_a}^\theta 1\{T_a < \infty\}] = \mathbb{E}[e^{\theta a - \theta^2 T_a/2} 1\{T_a < \infty\}].
\end{aligned}$$

Therefore, $\mathbb{E}[e^{-\theta^2 T_a/2}1\{T_a < \infty\}] = e^{-\theta a}$. Since $e^{-\theta^2 T_a/2}1\{T_a < \infty\}$ is increasing, by monotone convergence theorem,

$$1 = \lim_{\theta \to 0} e^{-\theta a} = \lim_{\theta \to 0} \mathbb{E}[e^{-\theta^2 T_a/2}1\{T_a < \infty\}] = \mathbb{E}[1\{T_a < \infty\}] = \mathbb{P}[1\{T_a < \infty\}].$$

■

R The stationary property shows that $\{B_{t+s} - B_s\}_{t\geq 0}$ is also a Brownian motion for any $s > 0$. Given that $T_a$ is a stopping time and finite a.s., we can assert that $\{B_{t+T_a} - B_{T_a}\}_{t\geq 0}$ is also a Brownian motion, independent of $\mathcal{F}_{T_a}$.

**Theorem 7.2** Let $\{B_t\}_{t\geq 0}$ be a standard Brownian motion. Let $M_t = \sup_{0\leq u\leq t} B_u$ be the running maximal of Brownian motion. For any $a \geq 0$,

$$\mathbb{P}(M_t \geq a) = 2\mathbb{P}(B_t \geq a) = \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} \, \mathrm{d}x.$$

*Proof.* Firstly simplify $\mathbb{P}(B_t \geq a)$ as the following:

$$\begin{aligned}
\mathbb{P}(B_t \geq a) &= \mathbb{P}(B_t \geq a, M_t \geq a) + \mathbb{P}(B_t \geq a, M_t < a) \\
&= \mathbb{P}(B_t \geq a, M_t \geq a) \\
&= \mathbb{P}(B_t \geq a \mid M_t \geq a)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t \geq a \mid T_a \leq t)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t - B_{T_a} \geq 0 \mid T_a \leq t)\mathbb{P}(M_t \geq a)
\end{aligned}$$

Since conditioned on $\{T_a \leq t\}$, $B_t - B_{T_a}$ is normally distributed with mean 0, we imply $\mathbb{P}(B_t - B_{T_a} \geq 0 \mid T_a \leq t) = \frac{1}{2}$. Therefore,

$$\mathbb{P}(B_t \geq a) = \frac{1}{2}\mathbb{P}(M_t \geq a) = \frac{1}{\sqrt{2\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} \, \mathrm{d}x.$$

■

## 7.1.2. Distributions of Brownian Motion

**Theorem 7.3** The joint distribution of the Brownian motion and the running maximal of Brownian motion is:

$$\mathbb{P}(M_t \geq a, B_t \leq a - y) = \mathbb{P}(B_t \geq a + y) = \frac{1}{\sqrt{2\pi t}} \int_{a+y}^{\infty} e^{-\frac{x^2}{2t}} \, \mathrm{d}x.$$

*Proof.* Simplify $\mathbb{P}(B_t \geq a + y)$ as the following:

$$\begin{aligned}
\mathbb{P}(B_t \geq a + y) &= \mathbb{P}(B_t \geq a + y \mid M_t \geq a)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t \geq a + y \mid T_a \leq t)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t - B_{T_a} \geq y \mid T_a \leq t)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t - B_{T_a} \leq -y \mid T_a \leq t)\mathbb{P}(M_t \geq a) \\
&= \mathbb{P}(B_t \leq a - y \mid M_t \geq a)\mathbb{P}(M_t \geq a).
\end{aligned}$$

■

**Theorem 7.4** For any $\lambda > 0$,

$$\mathbb{E}[e^{-\lambda T_a}] = e^{-\sqrt{2\lambda} a}.$$

*Proof.* By Theorem 7.2, the density of $T_a$ is

$$f_{T_a}(t) = \frac{a}{t\sqrt{t}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2t}\right).$$

Computing the integral $\int_0^\infty f_{T_a}(t) e^{-\lambda t} \, \mathrm{d}t$ gives the desired result. ■

*Another Quick Proof.* Since $\mathbb{P}(T_a < \infty) = 1$, substituting $\theta$ with $\sqrt{2\lambda}$ for $\mathbb{E}[e^{-\theta^2 T_a/2} 1\{T_a < \infty\}] = e^{-\theta a}$ gives the desired result. ■

**Theorem 7.5** Consider the Brownian motion with drift:

$$X_t \triangleq \mu t + \sigma B_t,$$

where $\mu \neq 0, \sigma > 0$.

1. For any $0 \le s < t$, $X_t - X_s$ is normally distributed with mean $\mu(t-s)$ and variance $\sigma^2(t-s)$. Independent incremental property also holds.
2. Time reversal: $\lim_{t\to\infty} \frac{X_t}{t} = \mu$.
3. For $\mu < 0$, define $M_\infty = \sup_{t\ge 0} X_t$ as the running maximal of drifted Brownian motion to infinity. Then $M_\infty$ is exponentially distributed with parameter $\frac{2|\mu|}{\sigma^2}$:

$$\mathbb{P}(M_\infty > y) = \exp\left(-\frac{2|\mu|}{\sigma^2} y\right), \quad y \ge 0.$$

The first two parts are trivial, and we give a proof for the last part:

*Proof.* Choose some $\theta \neq 0$ and define the random process $\{V_t^\theta\}_{t\ge 0}$ such that

$$V_t^\theta = \exp\left(\theta X_t - \mu\theta t - \frac{\sigma^2\theta^2}{2} t\right).$$

It follows that $\{V_t^\theta\}_{t\ge 0}$ is a martingale:

- Since $X_t \sim \mathcal{N}(\mu t, \sigma^2 t)$, $\mathbb{E}[|V_t^\theta|] = 1$;
- For any $0 \le u < t$, we have

$$\begin{aligned}
\mathbb{E}[V_t^\theta \mid \mathcal{F}_u] &= \mathbb{E}\left[\exp\left(\theta X_t - \mu\theta t - \frac{\sigma^2\theta^2}{2} t\right)\middle| \mathcal{F}_u\right] \\
&= \mathbb{E}\left[\exp\left(\theta(X_t - X_u) - (\mu\theta + \frac{\sigma^2\theta^2}{2})(t-u)\right) e^{(\theta X_u - (\mu\theta + \frac{\sigma^2\theta^2}{2})u)}\middle| \mathcal{F}_u\right] \\
&= e^{(\theta X_u - (\mu\theta + \frac{\sigma^2\theta^2}{2})u)} \mathbb{E}\left[\exp\left(\theta(X_t - X_u) - (\mu\theta + \frac{\sigma^2\theta^2}{2})(t-u)\right)\right] \\
&= V_u^\theta.
\end{aligned}$$

For $a < 0 < b$, define $T_{a,b} = \inf\{t \ge 0 : X_t = a \text{ or } X_t = b\}$. Define $\theta = -\frac{2\mu}{\sigma^2}$ such that $\mu\theta + \frac{1}{2}\sigma^2\theta^2 = 0$. Considering that $T_{a,b} \le T_a$ with $T_a$ being finite a.s., we imply $V_{t\wedge T_{a,b}}^\theta \xrightarrow{a.s.}$

$V^{\theta}_{T_{a,b}}$. Moreover, $|V^{\theta}_{t\wedge T_{a,b}}| \le \max(e^{-\theta a}, e^{\theta b})$. By dominated convergence theorem,

$$\mathbb{E}[V^{\theta}_{T_{a,b}}] = \lim_{t\to\infty} \mathbb{E}[V^{\theta}_{t\wedge T_{a,b}}] = \lim_{t\to\infty} \mathbb{E}[V^{\theta}_{0}] = 1.$$

Therefore,

$$\begin{aligned}1 &= \mathbb{E}[V^{\theta}_{T_{a,b}}] = \mathbb{E}[V^{\theta}_{T_{a,b}} 1\{T_{a,b} = a\}] + \mathbb{E}[V^{\theta}_{T_{a,b}} 1\{T_{a,b} = b\}] \\ &= e^{\theta a}\mathbb{P}(T_{a,b} = a) + e^{\theta b}\mathbb{P}(T_{a,b} = b).\end{aligned}$$

Together with the fact that $\mathbb{P}(T_{a,b} = a) + \mathbb{P}(T_{a,b} = b) = 1$, we assert that

$$\mathbb{P}(T_{a,b} = a) = \frac{e^{\theta b} - 1}{e^{\theta b} - e^{\theta a}}.$$

Now define the set $A_a = \{T_{a,b} = T_a\}$, then the sequence $\{A_a\}$ is monotone in $a$:

$$a_2 < a_1 < 0 \implies A_{a_2} \subseteq A_{a_1}.$$

Define the set

$$\Lambda \triangleq \{M_\infty < b\} = \{\omega \in \Omega : \; X_t(\omega) \text{ never hits level } b\} = \bigcap_{a<0, a\in\mathbb{Q}} A_a.$$

Thus

$$\mathbb{P}\Bigg(\bigcap_{a<0, a\in\mathbb{Q}} A_a\Bigg) = \lim_{a\to\infty} \mathbb{P}(A_a) = \lim_{a\to\infty} \mathbb{P}(T_{a,b} = a) = 1 - e^{-\theta b}.$$

Then we conclude that

$$\mathbb{P}(M_\infty < b) = \mathbb{P}(\Lambda) = 1 - e^{-2|\mu|/\sigma^2} b.$$

In particular, we imply $\mathbb{P}(M_\infty < \infty) = 1$. ■

## 7.2. Thursday

### 7.2.1. Unbounded Variation of Brownian Motion

**Definition 7.1** [Partition] Consider a closed interval $[a,b]$. A sequence

$$a = t_0 < t_1 < \cdots < t_n = b$$

is called a partition of $[a,b]$, denoted as $\Pi = \Pi(t_0, t_1, \ldots, t_n)$.

**Definition 7.2** [Total Variation] Let $f$ be a continuous function on $[a,b]$:

$$f: \ [a,b] \mapsto \mathbb{R}.$$

Then the total variation of $f$ is defined as

$$TV(f)[a,b] = \sup_{\Pi} \sum_k |f(t_k) - f(t_{k-1})|$$

where the supremum is taken over all the possible partitions $\Pi$ on the interval $[a,b]$. Since $\sum_k |f(t_k) - f(t_{k-1})|$ is increasing as the partition being smaller,

$$TV(f)[a,b] = \lim_{\|\Pi\| \to 0} \sum_k |f(t_k) - f(t_{k-1})|$$

with $\|\Pi\| = \max_k |t_k - t_{k-1}|$.

R The real analysis shows that a bounded variation function, i.e., the function whose total variation is bounded, is differentiable almost everywhere. Then if the function is nowhere differentiable, it is not of bounded variation.

**Theorem 7.6** Brownian motion is not of bounded variation almost surely, i.e.,

$$\mathbb{P}(\{\omega \in \Omega: \ TV(B.(\omega))[0,t] = \infty\}) = 1, \qquad \forall t > 0.$$

This result is based on the fact that the Brownian motion is nowhere differentiable alomst surely.

**Theorem 7.7** Brownian motion is nowhere differentiable almost surely. In particular,

$$\mathbb{P}\left(\left\{\omega \in \Omega : \limsup_{n\to\infty}\left|\frac{B_{t+h}(\omega)-B_t(\omega)}{h}\right| = \infty, t \in [0,\infty)\right\}\right) = 1$$

(R) If a stochastic process $\{A_t\}_{t\geq 0}$ has bounded variation, then the integral $\int_a^b f(t)\,\mathrm{d}A_t(\omega)$ can be defined in Riemann integraion sense $\omega$-wisely. However, Brownian motion is not of the bounded variaiton. The stochastic integral $\int_a^b f(t)\,\mathrm{d}B_t(\omega)$ shall be defined in a new manner.

*Proof.* Choose any $T>0$ and $M>0$, define the set

$$A^{(M)} \triangleq \left\{\omega \in \Omega : \ \exists t \in [0,T] \text{ such that } \limsup_{h\to 0}\left|\frac{B_{t+h}(\omega)-B_t(\omega)}{h}\right| \leq M\right\}.$$

It suffices to show that $\mathbb{P}(A^{(M)}) = 0$. If $\omega \in A^{(M)}$, there exists $t \in [0,T]$ and $n_0$ such that when $n \geq n_0$,

$$\left|\frac{B_u(\omega)-B_t(\omega)}{u-t}\right| \leq 2M, \qquad \forall u \in (t-2/n, t+2/n).$$

Decompose $A^{(M)}$ into many smaller sets. Define the set

$$A_n^{(M)} \triangleq \left\{\omega \in \Omega : \ \exists t \in [0,T] \text{ such that } |B_u(\omega)-B_t(\omega)| \leq 2M|u-t|, \quad \forall u \in (t-2/n, t+2/n)\right\}. \tag{7.1}$$

Then i) $A^{(M)} \subseteq \cup_n A_n^{(M)}$, and ii) $\{A_n^{(M)}\}$ is monotone: $A_n \subseteq A_{n+1}$.

Suppose that $\omega \in A_n$ with $t$ having such a property in (7.1). Let $k = \sup\{j \in \mathbb{Z} : \ j/n \leq t\}$, which means $k$ is close enough to $t$. Define $Y_k$ as the maximal of three independent increments:

$$Y_k = \max\left\{\left|B_{(k+2)/n} - B_{(k+1)/n}\right|, \left|B_{(k+1)/n} - B_{k/n}\right|, \left|B_{k/n} - B_{(k-1)/n}\right|\right\}.$$

We can show that $Y_k(\omega) \le 6M/n, \forall \omega \in A_n^{(M)}$ as follows. Firstly,

$$\begin{aligned}
\left|B_{(k+2)/n}(\omega) - B_{(k+1)/n}(\omega)\right| &\le \left|B_{(k+2)/n}(\omega) - B_t(\omega)\right| + \left|B_t(\omega) - B_{(k+1)/n}(\omega)\right| \\
&\le 2M\left|\frac{k+2}{n} - t\right| + 2M\left|\frac{k+1}{n} - t\right| \\
&\le 2M \cdot \frac{2}{n} + 2M \cdot \frac{1}{n} = \frac{6M}{n}
\end{aligned}$$

where the last inequality is because that $k/n \le t < (k+1)/n$. Following the similar technique, we can show that

$$\left|B_{(k+1)/n}(\omega) - B_{k/n}(\omega)\right|, \left|B_{k/n}(\omega) - B_{(k-1)/n}(\omega)\right| \le \frac{6M}{n} \implies Y_k(\omega) \le \frac{6M}{n}.$$

Now define the new set based on the consequence of the claim about $A_n^{(M)}$:

$$E_n^{(M)} \triangleq \left\{\omega \in \Omega : \exists j \in [1, T_n] \cap \mathbb{Z} \text{ such that } Y_j(\omega) \le \frac{6M}{n}\right\}.$$

with

$$Y_j = \max\left\{\left|B_{(j+2)/n} - B_{(j+1)/n}\right|, \left|B_{(j+1)/n} - B_{j/n}\right|, \left|B_{j/n} - B_{(j-1)/n}\right|\right\}.$$

Directly $A_n^{(M)} \subseteq E_n^{(M)}$ for each $n$. Now we begin to upper bound $\mathbb{P}(E_n^{(M)})$:

$$\begin{aligned}
\mathbb{P}(E_n^{(M)}) &\le \sum_{1 \le j \le T_n} \mathbb{P}(Y_j \le \frac{6M}{n}) \\
&\le T_n \mathbb{P}\left(\max\left\{\left|B_{(j+2)/n} - B_{(j+1)/n}\right|, \left|B_{(j+1)/n} - B_{j/n}\right|, \left|B_{j/n} - B_{(j-1)/n}\right|\right\} \le \frac{6M}{n}\right) \\
&= T_n \cdot \prod_{i=j-1:j+1} \mathbb{P}\left(\left|B_{(i+1)/n} - B_{(i)/n}\right| \le \frac{6M}{n}\right) && \text{(7.2a)} \\
&= T_n \cdot \left[\mathbb{P}\left(|B_{1/n}| \le \frac{6M}{n}\right)\right]^3, && \text{(7.2b)}
\end{aligned}$$

where (7.2a) is because of the independent increments of Brownian motion, and (7.2b) is because of its stationary increment property. In particular,

$$\begin{aligned}
\mathbb{P}\Big(|B_{1/n}| \leq \frac{6M}{n}\Big) &= \mathbb{P}\Big(-\frac{6M}{n} \leq B_{1/n} \leq \frac{6M}{n}\Big) \\
&= \mathbb{P}\Big(-\frac{6M}{\sqrt{n}} \leq B_1 \leq \frac{6M}{\sqrt{n}}\Big) \qquad (7.2c) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\frac{6M}{\sqrt{n}}}^{\frac{6M}{\sqrt{n}}} e^{-x^2/2} \mathrm{d}x \\
&\leq \frac{2}{\sqrt{2\pi}} \frac{6M}{\sqrt{n}} \qquad (7.2d)
\end{aligned}$$

where (7.2c) is by the scaling property, and (7.2d) is by upper bounding $e^{-x^2/2} \leq 1$. It follows that

$$\mathbb{P}(E_n^{(M)}) \leq T_n(\frac{2}{\sqrt{2\pi}} \frac{6M}{\sqrt{n}})^3 \to 0.$$

Since $A_n^{(M)} \subseteq E_n^{(M)}$, $\mathbb{P}(A_n^{(M)}) \to 0$. Since $A^{(M)} \subseteq \cup_n A_n^{(M)}$,

$$\mathbb{P}(A^{(M)}) \leq \mathbb{P}(\cup_n A_n^{(M)}) = \lim_{n\to\infty} \mathbb{P}(A_n^{(M)}) = 0.$$

The proof is complete. ■

# Chapter 8

# Weak 8

## 8.1. Thursday

### 8.1.1. Quadratic Variation

**Definition 8.1** [Quadratic Variation] Consider a partition $\Pi$ on the interval $[0,T]$. The **quadratic variation** of $\{B_t(\omega)\}_{0\le t\le T}$ over the partition $\Pi$ is defined as

$$Q(\Pi,\omega)=\sum_k |B_{t_k}(\omega)-B_{t_{k-1}}(\omega)|^2.$$

**Theorem 8.1** Consider a sequence of partitions $\{\Pi^{(n)}\}$ with $\|\Pi^{(n)}\|\to 0$, where $\|\Pi\|\triangleq \max_k |t_k-t_{k-1}|$. Then

$$\lim_{n\to\infty}\mathbb{E}\left[(Q(\Pi^{(n)})-T)^2\right]=0.$$

*Proof.* Given a partition $\Pi$ on the interval $[0,T]$, define

$$\theta_k=(B_{t_k}-B_{t_{k-1}})^2-(t_k-t_{k-1}) \implies Q(\Pi)=T+\sum_k \theta_k$$

We claim that $\theta_j, \theta_k$ are uncorrelated for $j \neq k$:

$$\begin{aligned}
\mathbb{E}[\theta_j\theta_k] &= \mathbb{E}\Big[\Big((B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})\Big)\left((B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})\right)\Big] \\
&= \mathbb{E}\left[(B_{t_j} - B_{t_{j-1}})^2(B_{t_k} - B_{t_{k-1}})^2\right] - (t_j - t_{j-1})\mathbb{E}\left[(B_{t_k} - B_{t_{k-1}})^2\right] \\
&\quad - (t_k - t_{k-1})\mathbb{E}\left[(B_{t_j} - B_{t_{j-1}})^2\right] + (t_j - t_{j-1})(t_k - t_{k-1}) \\
&= \mathbb{E}\left[(B_{t_j} - B_{t_{j-1}})^2\right]\mathbb{E}\left[(B_{t_k} - B_{t_{k-1}})^2\right] - (t_j - t_{j-1})\mathbb{E}\left[(B_{t_k} - B_{t_{k-1}})^2\right] \\
&\quad - (t_k - t_{k-1})\mathbb{E}\left[(B_{t_j} - B_{t_{j-1}})^2\right] + (t_j - t_{j-1})(t_k - t_{k-1}) \\
&= (t_j - t_{j-1})(t_k - t_{k-1}) - (t_j - t_{j-1})(t_k - t_{k-1}) \\
&\quad - (t_k - t_{k-1})(t_j - t_{j-1}) + (t_j - t_{j-1})(t_k - t_{k-1}) = 0.
\end{aligned}$$

Then we begin to simplify $\mathbb{E}[(Q(\Pi^{(n)}) - T)^2]$:

$$\begin{aligned}
\mathbb{E}\left[(Q(\Pi^{(n)}) - T)^2\right] &= \mathbb{E}\left[(\sum_k \theta_k)^2\right] \\
&= \sum_k \mathbb{E}\left[\theta_k^2\right] + \sum_{j\neq k}\mathbb{E}\left[\theta_j\theta_k\right] \\
&= \sum_k \mathbb{E}\left[\left((B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})\right)^2\right] \\
&= \sum_k \mathbb{E}\left[(B_{t_k} - B_{t_{k-1}})^4\right] \\
&\quad - 2\sum_k (t_k - t_{k-1})\mathbb{E}\left[(B_{t_k} - B_{t_{k-1}})^2\right] + \sum_k (t_k - t_{k-1})^2 \\
&= 3\sum_k (t_k - t_{k-1})^2 - 2\sum_k (t_k - t_{k-1})^2 + \sum_k (t_k - t_{k-1})^2 \\
&= 2\sum_k (t_k - t_{k-1})^2 \le 2\|\Pi^{(n)}\| \sum_k (t_k - t_{k-1}) \\
&= 2T \cdot \|\Pi^{(n)}\| \to 0.
\end{aligned}$$

■

**R** The Theorem 8.1 shows that the quadratic variation of Brownian motion on the interval $[0, T]$ converges to $T$ in $L^2$ for any $T$. This implies that $Q(\Pi^{(n)}) \to T$ in probability as $\|\Pi^{(n)}\| \to 0$. Then there exists a subsequence of $\Pi^{(n)}$, such that $Q(\Pi^{(n)}) \to T$ almost surely.

**Theorem 8.2** If $\|\Pi^{(n)}\| \to 0$ faster than $1/n^2$, i.e.,

$$\lim_{n\to\infty} n^2 \cdot \|\Pi^{(n)}\| = 0,$$

then $Q(\Pi^{(n)}) \to T$ almost surely.

*Proof.* Take $\delta_n \triangleq n^2 \cdot \|\Pi^{(n)}\|$, then by Markov inequality,

$$\mathbb{P}\left((Q(\Pi^{(n)}) - T)^2 > 2\delta_n\right) \leq \frac{\mathbb{E}\left[(Q(\Pi^{(n)}) - T)^2\right]}{2\delta_n} \leq \frac{2T\|\Pi^{(n)}\|}{2\delta_n} = \frac{T}{n^2}.$$

Considering that $\sum_n \frac{T}{n^2} < \infty$,

$$\sum_n \mathbb{P}\left((Q(\Pi^{(n)}) - T)^2 > 2\delta_n\right) < \infty.$$

By Borel-Cantelli Lemma,

$$\mathbb{P}\left((Q(\Pi^{(n)}) - T)^2 > 2\delta_n, \quad \text{infinitely often}\right) = 0.$$

Therefore, for almost all $\omega \in \Omega$,

$$|Q(\Pi^{(n)}, \omega) - T| > \sqrt{2\delta_n}, \quad \text{for finite } n \implies |Q(\Pi^{(n)}, \omega) - T| \leq \sqrt{2\delta_n},\ n \to \infty.$$

By the assumption that $\delta_n \to 0$ as $n \to \infty$, we conclude that

$$|Q(\Pi^{(n)}, \omega) - T| \to 0.$$

The proof is complete. ■

The Brownian motion is nowhere differentiable almost surely and does not have bounded variation. However, it turns to have finite quadratic variation limit. Using this quadratic variation property, we can go back to show that the Brownian motion does not have bounded variation.

*A Direct Proof for Theorem 7.6.* Since the Brownian motion is almost surely continuous, on the closed interval $[0, T]$, $B_t(\omega)$ is uniformly continuous for almost all $\omega \in \Omega$:

$$\max_k |B_{t_k}(\omega) - B_{t_{k-1}}(\omega)| \to 0, \quad \text{as } \max_k |t_k - t_{k-1}| \to 0$$

Assume on the contrary that there exists $t > 0$ such that

$$\mathbb{P}(TV(B)[0,t] = \infty) < 1 \implies \mathbb{P}(TV(B)[0,t] < \infty) > 0.$$

Define the set $\Lambda = \{\omega \in \Omega : \; TV(B)[0,t] < \infty\}$. Then for any partition $\Pi$, if $\omega \in \Lambda$,

$$\sum_k |B_{t_k}(\omega) - B_{t_{k-1}}(\omega)| \le TV(B.(\omega))[0,t] < \infty$$

As a result, for $\omega \in \Lambda$, the quadratic variation converges to 0 as $\|\Pi\| \to 0$:

$$Q(\Pi, \omega) = \sum_k (B_{t_k}(\omega) - B_{t_{k-1}}(\omega))^2 \le \|\Pi\| \sum_k |B_{t_k}(\omega) - B_{t_{k-1}}(\omega)| \to 0.$$

Then $\mathbb{P}(\lim_{n\to\infty} Q(\Pi^{(n)}) = 0) \ge \mathbb{P}(\Lambda) > 0$, where $\|\Pi^{(n)}\| \to 0$. Choose some $\epsilon \in (0, t)$. If $\omega \in \{\lim_{n\to\infty} Q(\Pi^{(n)}) = 0\}$, there exists $n_0$ such that for $n \ge n_0$,

$$\omega \in \{|Q(\Pi^{(n)}) - t| > \epsilon\} \implies \lim_{n\to\infty} \mathbb{P}(|Q(\Pi^{(n)}) - t| > \epsilon) > 0,$$

which contradicts to the fact that $Q(\Pi^{(n)}) \to t$ in probability. ∎

# Chapter 9

# Week9

## 9.1. Tuesday

### 9.1.1. Introduction to Ito Calculus

Throughout this chapter, we consider a **complete** probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., for any $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ and $\forall B \subseteq A$, we have $B \in \mathcal{F}$. Let $\{B_t\}_{t\ge 0}$ be a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\mathcal{F}_t\}_{t\ge 0}$ be the natural filtration, i.e., $\mathcal{F}_t \triangleq \sigma(\{B_u :\ u \le t\})$.

Suppose that $\{X_t\}_{t\ge 0}$ is an $\{\mathcal{F}_t\}_{t\ge 0}$-adapted stochastic process. One typical example for such a process is $X_t = f(B_t)$ for some Borel measurable function $f : \mathbb{R} \to \mathbb{R}$. In this chapter, we aim to define the integral of the following form:

$$\int_0^t X_s \,\mathrm{d}B_s, \qquad t \ge 0. \tag{9.1}$$

A naive idea is to define this integral using the Riemann-sum approach:

$$\int_0^t X_s \,\mathrm{d}B_s \triangleq \lim_{\|\pi\|\to 0} \sum_k X_{t_{k-1}} \cdot (B_{t_k} - B_{t_{k-1}}), \tag{9.2}$$

where the limit is taken in $L^2$ sense along the partition $\pi$ defined on the interval $[0, t]$, and $\|\pi\| \triangleq \max_k |t_k - t_{k-1}|$.

It is reasonable to study how the limit for (9.2) looks like, by considering a simple stochastic process $\{X_t\}_{t\ge 0}$, and then extend by some approximation procedure.

**Definition 9.1** [Simple Stochastic Process]

1. Let $\mathcal{L}^2$ be the space of all adapted stochastic process $\{X_t\}_{t\ge 0}$ satisfying

$$\mathbb{E}\left[\int_0^T X_t^2 \,\mathrm{d}t\right] < \infty, \quad \forall T > 0.$$

2. An adapted stochastic process $\{X_t\}_{t\ge 0} \in \mathcal{L}^2$ is called **simple** if for any $\omega \in \Omega$,

$$X_t(\omega) = X_{t_j}(\omega), \qquad t \in [t_j, t_{j+1}),\ j = 0, 1, \dots,$$

where $0 = t_0 < t_1 < \cdots < t_n < \cdots$ is an increasing sequence with $\lim_{n\to\infty} t_n = \infty$. Denote $\mathcal{L}_0^2$ be the class of all simple processes.

R Not every piece-wise constant process is a simple process. Consider an $\{\mathcal{F}_t\}_{t\ge 0}$-adapted stochastic process $\{X_t\}_{t\ge 0}$ and define

$$Y_t(\omega) = X_{t_{j+1}}(\omega), \quad t \in (t_j, t_{j+1}].$$

Then $Y_t$ is not $\mathcal{F}_t$-measurable, and thus $\{Y_t\}_{t\ge 0}$ is not an $\{\mathcal{F}_t\}_{t\ge 0}$-adapted process.

**Definition 9.2** [Ito Integral for Simple Stochastic Process] Suppose that $\{X_t\}_{t\ge 0}$ is a simple process. Given $T > 0$, define, for each $\omega \in \Omega$,

$$\int_0^T X_t(\omega)\,\mathrm{d}B_t(\omega) = \sum_{k=0}^{n-1} X_{t_k}(\omega)\cdot[B_{t_{k+1}}(\omega) - B_{t_k}(\omega)] + X_{t_n}(\omega)\cdot[B_T(\omega) - B_{t_n}(\omega)] \tag{9.3}$$

where $n \triangleq \max\{j \in \mathbb{N} : t_j \le T\}$.

**Proposition 9.1** The Ito Integral for Simple Stochastic Process admits the following properties:

1. Linearity: Suppose that $\{X_t\}_{t\geq 0},\{Y_t\}_{t\geq 0}\in\mathcal{L}_0^2$ and $\alpha,\beta\in\mathbb{R}$, then

$$\int_0^T(\alpha X_t+\beta Y_t)\,\mathrm{d}B_t=\alpha\cdot\int_0^T X_t\,\mathrm{d}B_t+\beta\cdot\int_0^T Y_t\,\mathrm{d}B_t$$

2. Ito Isometry: for $\{X_t\}_{t\geq 0}\in\mathcal{L}_0^2$,

$$\mathbb{E}\left[\left(\int_0^T X_t\,\mathrm{d}B_t\right)^2\right]=\mathbb{E}\left[\int_0^T X_t^2\,\mathrm{d}t\right]$$

3. Define the new random variable $I_t[X]\triangleq\int_0^t X_u\,\mathrm{d}B_u$, then the process $\{I_t[X]\}_{t\geq 0}$ is an almost sure continuous martingale, with square integrability:

$$\mathbb{E}[(I_t[X])^2]<\infty,\quad\forall t\geq 0.$$

*Proof for Part 1).* Denote $\{t_k^{(1)}\}_{k\geq 0}$ and $\{t_k^{(2)}\}_{k\geq 0}$ as the paritions corresponding to simple processes $\{X_t\}_{t\geq 0},\{Y_t\}_{t\geq 0}$. Consider a partition $\{t_k\}_{k\geq 0}$ obtained as a union of these two partitions. With respect to this new sequence $\{t_k\}$, processes $\{X_t\}_{t\geq 0},\{Y_t\}_{t\geq 0}$ are still simple. Then $\{\alpha X_t+\beta Y_t\}$ is also a simple process corresponding to $\{t_k\}$. The linearity property follows by checking the definition in (9.3). ■

*Proof for Part 2).* Note that the left side in this part can be expanded as the following:

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T X_t\,\mathrm{d}B_t\right)^2\right]&=\mathbb{E}\left[\left(\sum_{k=0}^{n-1}X_{t_k}\cdot[B_{t_{k+1}}-B_{t_k}]+X_{t_n}\cdot[B_T-B_{t_n}]\right)^2\right]\\&=\sum_{0\leq k_1,k_2\leq n-1}\mathbb{E}\left[X_{t_{k_1}}X_{t_{k_2}}(B_{t_{k_1+1}}-B_{t_{k_1}})(B_{t_{k_2+1}}-B_{t_{k_2}})\right]\\&\quad+\mathbb{E}\left[X_{t_n}^2\cdot(B_T-B_{t_n})^2\right]+2\sum_{k=0}^{n-1}\mathbb{E}\left[X_{t_k}X_{t_n}\cdot(B_{t_{k+1}}-B_{t_k})(B_T-B_{t_n})\right]\end{aligned}$$

where the second term equals $\mathbb{E}\left[X_{t_n}^2\cdot(T-t_n)\right]$ by the independent incremental property of Brownian motion, and the third term vanishes since

$$\begin{aligned}\mathbb{E}\left[X_{t_k}(B_{t_{k+1}}-B_{t_k})\right]&=\mathbb{E}\left[\mathbb{E}[X_{t_k}(B_{t_{k+1}}-B_{t_k})\mid\mathcal{F}_k]\right]\\&=\mathbb{E}\left[X_{t_k}\mathbb{E}[(B_{t_{k+1}}-B_{t_k})\mid\mathcal{F}_k]\right]=\mathbb{E}\left[X_{t_k}\mathbb{E}[(B_{t_{k+1}}-B_{t_k})]\right]=0.\end{aligned}$$

Following the similar trick, for $0 \le k_1 < k_2 \le n-1$, we have

$$\begin{aligned}\mathbb{E}\left[X_{t_{k_1}} X_{t_{k_2}} (B_{t_{k_1+1}} - B_{t_{k_1}})(B_{t_{k_2+1}} - B_{t_{k_2}})\right] &= \mathbb{E}\left[\mathbb{E}[X_{t_{k_1}} X_{t_{k_2}} (B_{t_{k_1+1}} - B_{t_{k_1}})(B_{t_{k_2+1}} - B_{t_{k_2}}) \mid \mathcal{F}_{t_{k_2}}]\right] \\ &= \mathbb{E}\left[X_{t_{k_1}} (B_{t_{k_1+1}} - B_{t_{k_1}}) X_{t_{k_2}} \mathbb{E}[(B_{t_{k_2+1}} - B_{t_{k_2}})]\right] = 0,\end{aligned}$$

where the second equality is because that all random variables except $B_{t_{k_2+1}}$ are $\mathcal{F}_{t_{k_2}}$-measurable. Now when $k_1 = k_2 \equiv k$, we have

$$\begin{aligned}\mathbb{E}\left[X_{t_k}^2 (B_{t_{k+1}} - B_{t_k})^2\right] &= \mathbb{E}\left[X_{t_k}^2 \mathbb{E}[(B_{t_{k+1}} - B_{t_k})^2 \mid \mathcal{F}_{t_k}]\right] \\ &= \mathbb{E}\left[X_{t_k}^2 (t_{k+1} - t_k)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2\right] &= \sum_{k=0}^{n-1} \mathbb{E}\left[X_{t_k}^2 (t_{k+1} - t_k)\right] + \mathbb{E}\left[X_{t_n}^2 \cdot (T - t_n)\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{n-1} X_{t_k}^2 (t_{k+1} - t_k) + X_{t_n}^2 \cdot (T - t_n)\right] = \mathbb{E}\left[\int_0^T X_t^2 \,\mathrm{d}t\right].\end{aligned}$$

■

*Proof of Part 3).* Since $\{B_t\}_{t \ge 0}$ is almost surely continuous, by definition, $I_t[X]$ is also a.s. continuous. To show the square integrability, by Ito isometry property,

$$\mathbb{E}[(I_t[X])^2] = \mathbb{E}[\left(\int_0^t X_u \,\mathrm{d}B_u\right)^2] = \mathbb{E}\left[\int_0^t X_u^2 \,\mathrm{d}u\right].$$

Since $\{X_t\}_{t \ge 0} \in \mathcal{L}_0^2$, $\mathbb{E}[(I_t[X])^2] < \infty$ for all $t \ge 0$.

Now we begin to show that $\{I_t[X]\}_{t \ge 0}$ is a martingale with respect to $\{\mathcal{F}_t\}_{t \ge 0}$. For any $0 \le s < t$ and take $n' \triangleq \max\{j \in \mathbb{N} : \ t_{n'} \le s\}$, we have

$$\mathbb{E}[I_t[X] \mid \mathcal{F}_s] = \mathbb{E}\left[\sum_{k=0}^{n-1} X_{t_k} (B_{t_{k+1}} - B_{t_k}) + X_{t_n} (B_t - B_{t_n}) \middle| \mathcal{F}_s\right].$$

Considering separate the summation from 0 to $n-1$ into two parts, we further have

$$\begin{aligned}\mathbb{E}[I_t[X] \mid \mathcal{F}_s] = &\ \mathbb{E}\left[\sum_{k=0}^{n'-1} X_{t_k}(B_{t_{k+1}} - B_{t_k}) + X_{t_{n'}}(B_s - B_{t_{n'}})\middle| \mathcal{F}_s\right] \\ &+ \mathbb{E}\left[X_{t_{n'}}(B_{t_{n'+1}} - B_s)\middle| \mathcal{F}_s\right] \\ &+ \mathbb{E}\left[\sum_{k=n'+1}^{n'-1} X_{t_k}(B_{t_{k+1}} - B_{t_k}) + X_{t_n}(B_t - B_{t_n})\middle| \mathcal{F}_s\right]\end{aligned}$$

where the first term equals

$$\sum_{k=0}^{n'-1} X_{t_k}(B_{t_{k+1}} - B_{t_k}) + X_{t_{n'}}(B_s - B_{t_{n'}}),$$

the second term equals

$$X_{t_{n'}}\mathbb{E}\left[(B_{t_{n'+1}} - B_s)\right] = 0,$$

and the third term equals

$$\sum_{k=n'+1}^{n'-1} \mathbb{E}\left[X_{t_k}\mathbb{E}[B_{t_{k+1}} - B_{t_k} \mid \mathcal{F}_{t_k}]\middle|\mathcal{F}_s\right] + \mathbb{E}\left[X_{t_n}\mathbb{E}[B_t - B_{t_n} \mid \mathcal{F}_{t_n}]|\mathcal{F}_s\right] = 0.$$

As a result,

$$\begin{aligned}\mathbb{E}[I_t[X] \mid \mathcal{F}_s] &= \sum_{k=0}^{n'-1} X_{t_k}(B_{t_{k+1}} - B_{t_k}) + X_{t_{n'}}(B_s - B_{t_{n'}}) \\ &= \int_0^s X_u \, \mathrm{d}B_u.\end{aligned}$$

Since $\{I_t[X]\}$ is square integrable, it is $\mathcal{L}^1$-integrable. Therefore, $\{I_t[X]\}$ is a martingale. The proof is completed. ■

## 9.2. Thursday

### 9.2.1. Approxiation by simple processes

R Before defining the Ito integral for general adapted process $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$, we show that any such $\{X_t\}_{t\geq 0}$ can be approximated by a sequence of simple processes $\{X_t^{(n)}\}_{t\geq 0} \in \mathcal{L}_0^2, n = 1, 2, \ldots$.

**Theorem 9.1** 1. Let $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$ be an almost surely bounded and continuous process, i.e.,

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{t\geq 0} |X_t(\omega)| \leq M \text{ and } X_t(\omega) \text{ continuous on } t \geq 0\right\}\right) = 1.$$

Then for given $T > 0$, there exists a sequence of simple processes $\{X_t^{(n)}\}_{t\geq 0} \in \mathcal{L}_0^2$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t\right] = 0.$$

2. Let $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$ be an almost surely bounded (but not necessarily continuous) process. Then for given $T > 0$, there exists a sequence of almost surely bounded and continuous process $\{X_t^{(n)}\}_{t\geq 0} \in \mathcal{L}^2, \forall n \geq 1$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t\right] = 0.$$

3. Let $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$. Then for given $T > 0$, there exists a sequence of almost surely bounded processes $\{X_t^{(n)}\}_{t\geq 0} \in \mathcal{L}^2, \forall n \geq 1$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t\right] = 0.$$

*Proof for Part 1).* Construct $\{X_t^{(n)}\}_{t\geq 0}$ as the following. Pick a sequence of partitions $\{\Pi^{(n)}\}$ on the interval $[0, T]$ with $\|\Pi^{(n)}\| \to 0$. For each $n$, define the stochastic process $\{X_t^{(n)}\}$ with

$$X_t^{(n)}(\omega) = X_{t_j}^{(n)}(\omega), \quad \text{for } t \in [t_j^{(n)}, t_{j+1}^{(n)}),$$

where $\Pi^{(n)} \triangleq \{0 = t_0^{(n)} < t_1^{(n)} < \cdots < T\}$. It is clear that $\{X_t^{(n)}\}_{t\geq 0}$ is a simple process.

Define the set

$$\Lambda = \left\{\omega \in \Omega : \sup_{t\geq 0} |X_t(\omega)| \leq M \text{ and } X_t(\omega) \text{ continuous on } t \geq 0\right\}$$

For each $\omega \in \Omega$, $X_t(\omega)$ is continuous (and thus uniformly continuous) on $[0, T]$: for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|X_s(\omega) - X_t(\omega)| < \sqrt{\frac{\varepsilon}{T}}, \quad \text{for any } |s - t| < \delta.$$

Choose large $n$ such that $\|\Pi^{(n)}\| < \delta$, which implies that

$$\forall t \in [0, T], \quad |X_t^{(n)}(\omega) - X_t(\omega)| < \sqrt{\frac{\varepsilon}{T}} \implies \int_0^T [X_t^{(n)}(\omega) - X_t(\omega)]^2 \, \mathrm{d}t < \varepsilon.$$

Therefore, the random variable $\int_0^T [X_t^{(n)} - X_t]^2 \, \mathrm{d}t \to 0$ almost surely. For $\omega \in \Lambda$, $\{X_t(\omega)\}$ is bounded by $M$, and thus $\{X_t^{(n)}(\omega)\}$ is bounded by $M$ as well. Thus the random variable $\int_0^T [X_t^{(n)} - X_t]^2 \, \mathrm{d}t$ is upper bounded by $(2M)^2 \cdot T$ almost surely. By the bounded convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t\right] = 0.$$

■

*Proof for Part 2).* Construct $\{X_t^{(n)}\}_{t\geq 0}$ as the following. For each $n$, pick a non-negative continuous funciton $\phi_n : \mathbb{R} \to \mathbb{R}$ satisfying

1. $\phi_n(x) = 0$ for $x \in (-\infty, -\frac{1}{n}] \cup [0, \infty)$;
2. $\int_{-\infty}^{\infty} \phi_n(x) \, \mathrm{d}x = 1$.

Then define the process $\{X_t^{(n)}\}_{t\geq 0}$ with

$$X_t^{(n)}(\omega) \equiv \phi_n * X_{\cdot}(\omega) \mid_0^t \triangleq \int_0^t \phi_n(s - t) X_s(\omega) \, \mathrm{d}s, \quad \forall \omega \in \Omega.$$

Define the set

$$\Lambda = \left\{\omega \in \Omega : \sup_{t\geq 0} |X_t(\omega)| \leq M\right\}.$$

Then for each $\omega \in \Lambda$,

$$|X_t^{(n)}(\omega)| \leq \int_0^t \phi_n(s-t)|X_s(\omega)|\,\mathrm{d}s \leq M \cdot \int_0^t \phi_n(s-t)\,\mathrm{d}s \leq M, \quad \forall t \geq 0, \forall n.$$

Therefore, the process $\{X_t^{(n)}\}_{t\geq 0}$ is almost surely bounded. Moreover, by definition, $\{X_t^{(n)}\}_{t\geq 0}$ is continuous a.s. and $\{\mathcal{F}_t\}$ adapted.

Now we begin to show that $\int_0^T [X_t^{(n)} - X_t]^2\,\mathrm{d}t \to 0$ almost surely. Take $\omega \in \Lambda$, then

$$\int_0^T [X_t^{(n)}(\omega) - X_t(\omega)]^2\,\mathrm{d}t \leq 2M \cdot \int_0^T |X_t^{(n)}(\omega) - X_t(\omega)|\,\mathrm{d}t,$$

where the integral on the RHS can be upper bounded as the following:

$$\begin{aligned}
\int_0^T |X_t^{(n)}(\omega) - X_t(\omega)|\,\mathrm{d}t &= \int_0^T \left| \int_0^t \phi_n(s-t)X_s(\omega)\,\mathrm{d}s - X_t(\omega) \right| \mathrm{d}t && \text{(9.4a)} \\
&= \int_0^T \left| \int_0^\infty \phi_n(-s)X_{t-s}(\omega)\,\mathrm{d}s - X_t(\omega) \right| \mathrm{d}t && \text{(9.4b)} \\
&= \int_0^T \left| \int_0^\infty \phi_n(-s)X_{t-s}(\omega)\,\mathrm{d}s - \int_0^\infty \phi_n(-s)X_t(\omega)\,\mathrm{d}s \right| \mathrm{d}t && \text{(9.4c)} \\
&\leq \int_0^T \int_0^\infty \phi_n(-s)|X_t(\omega) - X_{t-s}(\omega)|\,\mathrm{d}s\,\mathrm{d}t && \text{(9.4d)} \\
&= \int_0^\infty \phi_n(-s) \int_0^T |X_t(\omega) - X_{t-s}(\omega)|\,\mathrm{d}t\,\mathrm{d}s && \text{(9.4e)}
\end{aligned}$$

where (9.4b) is by the change of variable $s' = t - s$; (9.4c) is because that $\int_{-\infty}^\infty \phi_n(s)\,\mathrm{d}s = \int_{-\infty}^0 \phi_n(s)\,\mathrm{d}s = 1$; (9.4e) is by the Fubini's theorem. We claim that the term $\int_0^T |X_t(\omega) - X_{t-s}(\omega)|\,\mathrm{d}t$ is small when $s$ is small, i.e., for any $\varepsilon > 0$, there exists $\delta$ such that when $s < \delta$,

$$\int_0^T |X_t(\omega) - X_{t-s}(\omega)|\,\mathrm{d}t < \varepsilon. \tag{9.5}$$

We can further apply (9.5) to upper bound the term (9.4e):

$$
\begin{aligned}
&\int_0^\infty \phi_n(-s) \int_0^T |X_t(\omega) - X_{t-s}(\omega)| \,\mathrm{d}t \,\mathrm{d}s \\
=&\int_0^\delta \phi_n(-s) \int_0^T |X_t(\omega) - X_{t-s}(\omega)| \,\mathrm{d}t \,\mathrm{d}s + \int_\delta^\infty \phi_n(-s) \int_0^T |X_t(\omega) - X_{t-s}(\omega)| \,\mathrm{d}t \,\mathrm{d}s \\
\leq&\varepsilon \int_\delta^\infty \phi_n(-s) \,\mathrm{d}s + 2MT \cdot \int_\delta^\infty \phi_n(-s) \,\mathrm{d}s = \varepsilon.
\end{aligned}
$$

where the last equality holds when we choose $n$ large enough such that $-\delta \leq -\frac{1}{n}$. Thus

$$
\int_0^T |X_t^{(n)}(\omega) - X_t(\omega)| \,\mathrm{d}t \to 0 \quad \text{for } \omega \in \Lambda.
$$

The remaining part follows the similar logic as in part 1).

Finally, we show the claim in (9.5) by discussing cases for continuous and discontinuous $X_t$:

- When $X_t$ is continuous on $t \in [0, T]$, we show that for any $\varepsilon > 0$, there exists $\delta$ such that

$$
\forall s < \delta, \quad \int_0^T |X_t - X_{t-s}| \,\mathrm{d}t < \varepsilon.
$$

  Since $X_t$ is continuous (and thus uniformly continuous) on $t \in [0, T]$, for any $\varepsilon > 0$, there exists $\delta < \frac{\varepsilon}{2M}$ such that for any $s < \delta$,

$$
|X_t - X_{t-s}| < \frac{\varepsilon}{2T}, \quad \forall t \in [s, T].
$$

  As a result,

$$
\begin{aligned}
\int_0^T |X_t - X_{t-s}| \,\mathrm{d}t &= \int_0^s |X_t - X_{t-s}| \,\mathrm{d}t + \int_s^T |X_t - X_{t-s}| \,\mathrm{d}t \\
&\leq M\delta + T \cdot \frac{\varepsilon}{2T} < \varepsilon.
\end{aligned}
$$

- When $X_t$ is not continuous on $t \in [0, T]$, we can also show the same result:

  Because the continuous functions are dense in $\mathcal{L}^p (1 \leq p < \infty)$, for any $\varepsilon > 0$, there exists a continuous function $\hat{X}_t$ such that

$$\int_0^T |X_t - \hat{X}_t| \, \mathrm{d}t < \frac{\varepsilon}{3}. \tag{9.7}$$

As a result,

$$\begin{aligned}\int_0^T |X_t - X_{t-s}| \, \mathrm{d}t &\leq \int_0^T |X_t - \hat{X}_t| \, \mathrm{d}t + \int_0^T |X_{t-s} - \hat{X}_{t-s}| \, \mathrm{d}t + \int_0^T |\hat{X}_{t-s} - \hat{X}_t| \, \mathrm{d}t \\ &< \varepsilon\end{aligned}$$

where the first two terms are all bounded by $\varepsilon/3$ because of (9.7), and the last term is also bounded by $\varepsilon/3$ since $\hat{X}_t$ is continuous on $t \in [0, T]$.

■

*Proof of Part 3).* For each $n$, we construct the almost surely bounded process $\{X_t^{(n)}\}_{t\geq 0}$ using the truncation method:

$$X_t^{(n)}(\omega) = \begin{cases} n, & \text{if } X_t(\omega) \geq n \\ X_t(\omega), & -n < X_t(\omega) < n \\ -n, & \text{if } X_t(\omega) \leq -n \end{cases}$$

Therefore, $|X_t^{(n)}(\omega)| \leq |X_t(\omega)|, \forall \omega \in \Omega$. Together with the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t \leq 2 \int_0^T (X_t^{(n)})^2 \, \mathrm{d}t + 2 \int_0^T (X_t)^2 \, \mathrm{d}t \leq 4 \int_0^T (X_t)^2 \, \mathrm{d}t < \infty, \forall t.$$

Therefore, $\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t$ is dominated by an integrable random variable. Moreover, substituting the form of $X_t^{(n)}$ and considering $\int_0^T (X_t)^2 \, \mathrm{d}t < \infty$ gives

$$\int_0^T (X_t^{(n)} - X_t)^2 \, \mathrm{d}t \leq \int_0^T (X_t)^2 1\{(X_t)^2 \geq n\} \, \mathrm{d}t + \int_0^T (X_t)^2 1\{(X_t)^2 \leq -n\} \, \mathrm{d}t \to 0, \quad \text{as } n \to \infty.$$

Applying the dominated convergence theorem gives the desired result.

■

Combining Theorem 9.1 part 1) to 3), we conclude that for $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$, there exists a sequence of simple processes $\{X_t^{(n)}\}_{t\geq 0} \in \mathcal{L}_0^2$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \,\mathrm{d}t\right] = 0.$$

# Chapter 10

# Week10

## 10.1. Tuesday

Recall the property 3) in proposition 9.1, in which we show that when $\{X_u\}_u$ is a simple process, $\left\{\int_0^t X_u \,\mathrm{d}B_u\right\}_t$ is a square integrable, almost surely continuous martingale. In this lecture, we first review some basic ideas about square integrability.

### 10.1.1. Square Integrable Process

**Definition 10.1** [Square Integrable Martingales]

1. A stochastic process $\{X_t\}_{t\ge 0}$ is said to be square integrable if $\mathbb{E}[X_t^2] < \infty, \forall t \ge 0$.
2. Let $\mathcal{U}^2$ be the class of square integrable, right-continuous, with left limit exist martingales.
3. Let $\mathcal{U}_c^2$ be the class of square integrable, almost surely continuous martingales.

In particular, we denote $\mathcal{L}^2$ as the set of square integrable random variables.

**Definition 10.2** [Norm on $\mathcal{U}^2$] For given $T > 0$, define a norm $\|\cdot\|$ on $\mathcal{U}^2$:

$$\|X\| \triangleq \left(\mathbb{E}[X_T^2]\right)^{1/2}, \quad \{X_t\}_{t\ge 0} \in \mathcal{U}^2.$$

**Theorem 10.1 — Completeness of Square Integrable Martingales.** With respect to the norm $\|\cdot\|$,

1. $\mathcal{U}^2$ is a complete metric space;
2. $\mathcal{U}_c^2$ is a closed subspace of $\mathcal{U}^2$.

*Proof.* 1. It is easy to see that $(\mathcal{U}^2, \|\cdot\|)$ is a metric space. To show the completeness, it suffices to show that for any Cauchy sequence $\{X_t^{(n)}\}_{t\ge 0} \in \mathcal{U}^2$ where $n = 1,2,\ldots$, there exists a process $\{X_t\}_{t\ge 0} \in \mathcal{U}^2$ such that

$$\|X^{(n)} - X\| \to 0, \quad \text{as } n \to \infty.$$

We first give a construction on such $\{X_t\}_{t\ge 0}$:

- Since $\{X^{(n)}\}_n$ is a Cauchy sequence, for any $\varepsilon > 0$, there exists $N$ such that as long as $m,n > N$,

$$\|X^{(m)} - X^{(n)}\| < \varepsilon.$$

Since $\{X_t^{(m)} - X_t^{(n)}\}_{t\ge 0}$ is a martingale, by convexity of quadratic function, $\{(X_t^{(m)} - X_t^{(n)})^2\}_{t\ge 0}$ is a sub-martingale, which implies

$$\forall t \le T, \quad \mathbb{E}[(X_t^{(m)} - X_t^{(n)})^2] \le \mathbb{E}[(X_T^{(m)} - X_T^{(n)})^2] = \|X^{(m)} - X^{(n)}\|^2 < \varepsilon^2.$$

This means that for fixed $t \le T$, $\{X_t^{(n)}\}_n$ is a Cauchy sequence in $\mathcal{L}^2$, with the metric defined as $\|X\|_{\mathcal{L}^2} \triangleq \left(\mathbb{E}[X^2]\right)^{1/2}$. By the completeness of $\mathcal{L}^2$, there exists a random variable $X_t \in \mathcal{L}^2$ such that $X_t^{(n)} \xrightarrow{L^2} X_t$ as $n \to \infty$.

Next, we show that $\{X_t\}_{t\ge 0} \in \mathcal{U}^2$:

(a) In order to show $\{X_t\}_{t\ge 0}$ is $\{\mathcal{F}_t\}$-adapted, we need the almost sure convergence of some process into $\{X_t\}$, as the definition of almost sure convergence is based on sample-path. The details are given as follows. The $L^2$ convergence implies $X_t^{(n)} \xrightarrow{P} X_t$, and thus there exist s a subsequence $\{X_t^{(n_k)}\}_k$ such that

$$X_t^{(n_k)} \xrightarrow{a.s.} X_t \quad \text{as } k \to \infty.$$

Then define the sample-path-based set

$$\Lambda = \left\{ \omega \in \Omega : \lim_{k\to\infty} X_t^{(n_k)}(\omega) = X_t(\omega) \right\}.$$

Re-define $X_t(\omega) = 0$ for $\omega \in \Lambda^c$. Note that $\Lambda^c \in \mathcal{F}_t$ since we assume $\{\mathcal{F}_t\}$ satisfies the usual condition (see Definition 5.4). It follows that for $a < 0$,

$$\left\{ \omega \in \Omega : X_t(\omega) \le a \right\} = \bigcap_j \bigcup_m \bigcup_{k>m} \left\{ \omega \in \Omega : X_t^{(n_k)}(\omega) < a + \frac{1}{j} \right\} \in \mathcal{F}_t.$$

For $a \ge 0$, since $\Lambda^c \in \mathcal{F}_t$,

$$\left\{ \omega \in \Omega : X_t(\omega) \le a \right\} = \Lambda^c \bigcup \left[ \bigcap_j \bigcup_m \bigcup_{k>m} \left\{ \omega \in \Omega : X_t^{(n_k)}(\omega) < a + \frac{1}{j} \right\} \right] \in \mathcal{F}_t.$$

(b) The integrability of $X_t$ is because $X_t \in \mathcal{L}^2$. Then we show $\{X_t\}_{t\ge0}$ satisfies the martingale property, i.e, for fixed $0 \le s < t$, we need to show $\int_A X_t \,\mathrm{d}\mathbb{P} = \int_A X_s \,\mathrm{d}\mathbb{P}$, $\forall A \in \mathcal{F}_s$. Direct calculation, together with the martingability of $\{X_t^{(n)}\}_{t\ge0}$ gives

$$\int_A X_t \,\mathrm{d}\mathbb{P} - \int_A X_s \,\mathrm{d}\mathbb{P} = \int_A (X_t - X_t^{(n)}) \,\mathrm{d}\mathbb{P} - \int_A (X_s - X_s^{(n)}) \,\mathrm{d}\mathbb{P}, \quad \forall n.$$

By the $L^2$ convergence of $X_t^{(n)}$,

$$\int_A |X_t - X_t^{(n)}| \,\mathrm{d}\mathbb{P} \le \mathbb{E}[|X_t - X_t^{(n)}|] \le \left( \mathbb{E}[(X_t - X_t^{(n)})^2] \right)^{1/2} \to 0, \quad n \to \infty.$$

Similarly, $\int_A |X_t - X_t^{(n)}| \,\mathrm{d}\mathbb{P} \to 0$. The martingability of $\{X_t\}_{t\ge0}$ follows since

$$\begin{aligned} \left| \int_A X_t \,\mathrm{d}\mathbb{P} - \int_A X_s \,\mathrm{d}\mathbb{P} \right| &= \lim_{n\to\infty} \left| \int_A (X_t - X_t^{(n)}) \,\mathrm{d}\mathbb{P} - \int_A (X_s - X_s^{(n)}) \,\mathrm{d}\mathbb{P} \right| \\ &\le \lim_{n\to\infty} \int_A |X_t - X_t^{(n)}| \,\mathrm{d}\mathbb{P} + \int_A |X_t - X_t^{(n)}| \,\mathrm{d}\mathbb{P} = 0. \end{aligned}$$

(c) To make $\{X_t\}_{t\geq 0}$ right-continuous with left limit exists, apply part 1) in Theorem 5.3.

2. Consider a sequence $\{X_t^{(n)}\}_t \in \mathcal{U}_c^2$ for $n = 1,2,\ldots$. By result in part 1), there exists $\{X_t\}_t \in \mathcal{U}^2$ as a limt. It suffices to show that $\{X_t\}_t \in \mathcal{U}_c^2$. In order to show the continuity result, we first construct the (almost sure) uniform convergence of some sequence in $\mathcal{U}_c^2$ with the limit $\{X_t\}_t$.

- Note that $\{X_t^{(n)} - X_t\}_t$ is a martingale for all $n$, then by Doob's inequality,

$$\mathbb{P}\left(\sup_{t\leq T}|X_t^{(n)} - X_t| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[(X_T^{(n)} - X_T)^2\right] \to 0, \quad \text{as } n \to \infty.$$

Since $\sup_{t\leq T}|X_t^{(n)} - X_t| \xrightarrow{P} 0$, there exists a subsequence $\{\sup_{t\leq T}|X_t^{(n_k)} - X_t|\}_k$ that converges to 0 almost surely. Define

$$\begin{aligned}
W_1 &= \bigcap_k \left\{\omega \in \Omega : \; X_t^{(n_k)}(\omega) \text{ is continuous}\right\} \\
W_2 &= \left\{\omega \in \Omega : \; \lim_{k\to\infty}\sup_{t\leq T}|X_t^{(n_k)} - X_t| = 0\right\} \\
W &= W_1 \cap W_2 \implies \mathbb{P}(W) = 1.
\end{aligned}$$

Take $\omega \in W$, then for any $\varepsilon > 0$, there exists $N > 0$ such that as long as $k > N$ and $t \leq T$,

$$|X_t^{(n_k)}(\omega) - X_t(\omega)| \leq \sup_{t\leq T}|X_t^{(n_k)}(\omega) - X_t(\omega)| < \frac{\varepsilon}{3}.$$

In other words, $X_t^{(n_k)}(\omega)$ uniformly converges to $X_t(\omega)$ for $\omega \in W$.

The continuity of $X_t^{(n_k)}(\omega)$ (and therefore uniform continuity) implies there exists $\delta$ such that as long as $h < \delta$,

$$|X_{t+h}^{(n_k)}(\omega) - X_t^{(n_k)}(\omega)| < \frac{\varepsilon}{3}, \quad \forall t \leq T.$$

Hence we conclude the continuity of $X_t(\omega)$ for $\omega \in W$:

$$|X_{t+h}(\omega) - X_t(\omega)| \le |X_{t+h}(\omega) - X_{t+h}^{(n_k)}(\omega)| + |X_{t+h}^{(n_k)}(\omega) - X_t^{(n_k)}(\omega)| + |X_t^{(n_k)}(\omega) - X_t(\omega)| < \varepsilon.$$

The proof is completed.

■

## 10.2. Thursday

### 10.2.1. Introduction to Ito Integral

By the conclusion in Section 9.2.1, for any process $\{X_t\}_{t\ge0} \in \mathcal{L}^2$, when fixing $T > 0$, there exists a sequence of simple processes $\{X_t^{(n)}\}_{t\ge0}, n = 1, 2, \ldots$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \,\mathrm{d}t\right] = 0. \tag{10.1}$$

The Ito integral for a simple process $\{\tilde{X}_t\}_{t\ge0}$ is well-defined as in (9.3):

$$\int_0^T \tilde{X}_t(\omega)\,\mathrm{d}B_t(\omega) = \sum_{k=0}^{n-1} \tilde{X}_{t_k}(\omega)\cdot[B_{t_{k+1}}(\omega) - B_{t_k}(\omega)] + \tilde{X}_{t_n}(\omega)\cdot[B_T(\omega) - B_{t_n}(\omega)]$$

Denote $I_t(\tilde{X}) \triangleq \{\int_0^t \tilde{X}_s(\omega)\,\mathrm{d}B_s(\omega)\}_{\omega\in\Omega}$.

**Theorem 10.2** Given a process $\{X_t\}_{t\ge0} \in \mathcal{L}^2$ and fixing $T > 0$, there exists an **unique** process $\{Z_t\}_{t\ge0} \in \mathcal{U}_c^2$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[(I_t(X^{(n)}) - Z_t)^2\right] = 0, \quad \forall t \in [0, T]. \tag{10.2}$$

R The process $\{Z_t\}_{t\ge0}$ is unique in the following sense: if there is another sequence of simple process, $\{\tilde{X}_t^{(n)}\}_{t\ge0}$, approximating $\{X_t\}_{t\ge0}$, and there exists $\{\tilde{Z}_t\}_{t\geq 0} \in \mathcal{U}_c^2$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[(I_t(\tilde{X}^{(n)}) - \tilde{Z}_t)^2\right] = 0, \quad \forall t \in [0,T],$$

then

$$\mathbb{P}\Big(Z_t = \tilde{Z}_t \text{ for all } t \in [0,T]\Big) = 1.$$

*Proof.* The existence of $\{Z_t\}_{t\geq 0}$ can be argued if we show $\{I_t(X^{(n)})\}_{t\geq 0}, n = 1,2,\ldots$ is a Cauchy sequence:

- We first show that $I_T(X^{(n)}), n = 1,2,\ldots$ is Cauchy (with respect to $L^2$ measure):

$$\mathbb{E}\left[(I_T(X^{(j)}) - I_T(X^{(k)}))^2\right] = \mathbb{E}\left[\left(\int_0^T X_t^{(j)}\,\mathrm{d}B_t - \int_0^T X_t^{(k)}\,\mathrm{d}B_t\right)^2\right] \tag{10.3a}$$

$$= \mathbb{E}\left[\int_0^T (X_t^{(j)} - X_t^{(k)})^2\,\mathrm{d}t\right] \tag{10.3b}$$

$$= \mathbb{E}\left[\int_0^T (X_t^{(j)} - X_t + X_t - X_t^{(k)})^2\,\mathrm{d}t\right] \tag{10.3c}$$

$$\leq 2\mathbb{E}\left[\int_0^T (X_t^{(j)} - X_t)^2\,\mathrm{d}t\right] + 2\mathbb{E}\left[\int_0^T (X_t - X_t^{(k)})^2\,\mathrm{d}t\right] \tag{10.3d}$$

  where (10.3b) is by the linearity and isometry property in Proposition 9.1; (10.3d) is by the inequality $(a+b)^2 \leq 2a^2 + 2b^2$.

  Recall (10.1), i.e., for any $\varepsilon > 0$, there exists $N$ such that as long as $j,k > N$,

$$\mathbb{E}\left[\int_0^T (X_t^{(j)} - X_t)^2\,\mathrm{d}t\right] < \frac{\varepsilon}{4}, \quad \mathbb{E}\left[\int_0^T (X_t^{(k)} - X_t)^2\,\mathrm{d}t\right] < \frac{\varepsilon}{4}.$$

  Thus $\mathbb{E}\left[(I_T(X^{(j)}) - I_T(X^{(k)}))^2\right] < \varepsilon$ for large $j,k$.

- Then we show that $I_t(X^{(n)}), n = 1,2,\ldots$ is Cauchy for $t < T$. Since $\{I_t(X^{(j)})\}_{t\geq 0}$ and $\{I_t(X^{(k)})\}_{t\geq 0}$ are martingales, together with the convexity of quadratic function, we can assert that $\{(I_t(X^{(j)}) - I_t(X^{(k)}))^2\}_{t\geq 0}$ is a sub-martingale, which means for large $j,k$,

$$\mathbb{E}\left[(I_t(X^{(j)})-I_t(X^{(k)}))^2\right]\le\mathbb{E}\left[(I_T(X^{(j)})-I_T(X^{(k)}))^2\right]<\varepsilon,\forall t<T.$$

By part 3) in Proposition 9.1, the process $\{I_t(X^{(n)})\}_{t\ge0}\in\mathcal{U}_c^2$ for each $n$. By the Cauchy property for $\{I_t(X^{(n)})\}_{t\ge0}, n=1,2,\ldots$ and the closedness of $\mathcal{U}_c^2$ with respect to the norm $\|X\|=(\mathbb{E}[X_T{}^2])^{1/2}$, there exists a limit $\{Z_t\}_{t\ge0}\in\mathcal{U}_c^2$ such that

$$\lim_{n\to\infty}\mathbb{E}\left[(I_T(X^{(n)})-Z_T)^2\right]=0.$$

We can further show (10.2) holds by the sub-martingability of $\{(I_t(X^{(n)})-Z_t)^2\}_{t\ge0}$.

Now we begin to show the uniqueness of $\{Z_t\}_{t\ge0}$. Suppose there is another simple process $\{\tilde{X}_t^{(n)}\}_{t\ge0}$ approximating $\{X_t\}_{t\ge0}$, and we have $\{\tilde{Z}_t\}_{t\ge0}$ such that $\lim_{n\to\infty}\mathbb{E}\left[(I_t(\tilde{X}_t^{(n)})-\tilde{Z}_t)^2\right]=0,\forall t\in[0,T]$. Observe that $Z_T-\tilde{Z}_T\in\mathcal{L}^2$. By doob's inequality,

$$\forall\varepsilon,\quad \mathbb{P}\left(\sup_{t\le T}|Z_t-\tilde{Z}_t|>\varepsilon\right)\le\frac{1}{\varepsilon^2}\mathbb{E}[(Z_T-\tilde{Z}_T)^2].$$

This suggests that in order to show that $\{\tilde{Z}_t\}_{t\ge0}$ is a version of $\{Z_t\}_{t\ge0}$, we can start with bounding their $L^2$ distance:

$$\begin{aligned}\mathbb{E}[(Z_T-\tilde{Z}_T)^2]&=\mathbb{E}[(Z_T-I_T(X_t^{(n)})+I_T(X_t^{(n)})-I_T(\tilde{X}_t^{(n)})+I_T(\tilde{X}_t^{(n)})-\tilde{Z}_T)^2]\\&\le3\mathbb{E}\left[(Z_T-I_T(X_t^{(n)}))^2\right]+3\mathbb{E}\left[(I_T(X_t^{(n)})-I_T(\tilde{X}_t^{(n)}))^2\right]\\&\quad+3\mathbb{E}\left[(I_T(\tilde{X}_t^{(n)})-\tilde{Z}_T)^2\right]\end{aligned}$$

where the inequality is by applying $(\sum_{i=1}^n a_i)^2\le n\sum_{i=1}^n a_i^2$. The first and last term vanishes as $n\to\infty$. Moreover, the second term can be upper bounded as

$$\begin{aligned}\mathbb{E}\left[(I_T(X_t^{(n)})-I_T(\tilde{X}_t^{(n)}))^2\right]&=\mathbb{E}\left[(\int_0^T(X_t^{(n)}-\tilde{X}_t^{(n)})\,\mathrm{d}B_t)^2\right]\\&=\mathbb{E}\left[\int_0^T(X_t^{(n)}-\tilde{X}_t^{(n)})^2\,\mathrm{d}t\right]\\&\le2\mathbb{E}\left[\int_0^T(X_t^{(n)}-X_t)^2\,\mathrm{d}t\right]+2\mathbb{E}\left[\int_0^T(\tilde{X}_t^{(n)}-X_t)^2\,\mathrm{d}t\right]\end{aligned}$$

and thus $\mathbb{E}\left[(I_T(X_t^{(n)}) - I_T(\tilde{X}_t^{(n)}))^2\right] \to 0$ as $n \to \infty$. Put things together, we can assert that

$$\mathbb{E}[(Z_T - \tilde{Z}_T)^2] = 0 \implies \mathbb{P}\left(\sup_{t\le T} |Z_t - \tilde{Z}_t| > \varepsilon\right) = 0, \forall \varepsilon > 0.$$

Define $\Lambda = \{\omega : Z_t(\omega) - \tilde{Z}_t(\omega) \neq 0 \text{ for some } t \in [0,T]\}$, then

$$\Lambda \subseteq \bigcup_{n=1}^{\infty} \left(\sup_{t\le T} |Z_t - \tilde{Z}_t| > \frac{1}{n}\right),$$

which implies $\mathbb{P}(\Lambda) = 0$, i.e.,

$$\mathbb{P}\left(Z_t = \tilde{Z}_t \text{ for all } t \in [0,T]\right) = \mathbb{P}(\Lambda^c) = 1.$$

■

**Definition 10.3** [Ito Integral for Square Integrable Process] For any process $\{X_t\}_{t\ge 0} \in \mathcal{L}^2$, define the Ito integral

$$I_t(X) = Z_t, \quad t \in [0,T], \tag{10.4}$$

where $\{Z_t\}_{t\ge 0}$ is defined in Theorem 10.2.

**Proposition 10.1** We find the Ito integral

$$\int_0^T B_t \,\mathrm{d}B_t = \frac{1}{2}\left[B_T^2 - T\right].$$

*Proof.* Before the computation, we need to show $\{B_t\}_{t\ge 0}$ satisfies the assumption for Ito integral, i.e., $\{B_t\}_{t\ge 0} \in \mathcal{L}^2$, which is trivial.

Firstly, we need to figure out the simple process $\{B_t^{(n)}\}_{t\ge 0}$ that approximating the argument in the integral, say $\{B_t\}_{t\ge 0}$. Define $\Pi^{(n)} = \{0, t_1^{(n)}, t_2^{(n)}, \dots, T\}$ the partition on $[0,T]$, and construct

$$B_t^{(n)} = B_{t_j^{(n)}}, \quad \forall t \in [t_j^{(n)}, t_{j+1}^{(n)}).$$

As a consequence,

$$\begin{aligned}\mathbb{E}\left[\int_0^T (B_t^{(n)} - B_t)^2 \, \mathrm{d}t\right] &= \mathbb{E}\left[\sum_j \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (B_t^{(n)} - B_t)^2 \, \mathrm{d}t\right] = \sum_i \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} \mathbb{E}\left[(B_t^{(n)} - B_t)^2\right] \mathrm{d}t \\ &= \sum_j \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} \mathbb{E}\left[(B_{t_j^{(n)}} - B_t)^2\right] \mathrm{d}t = \sum_j \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (t - t_j^{(n)}) \, \mathrm{d}t \\ &= \frac{1}{2}\sum_j (t_{j+1}^{(n)} - t_j^{(n)})^2 \le \frac{1}{2}\|\Pi^{(n)}\| \sum_j (t_{j+1}^{(n)} - t_j^{(n)}) = \frac{1}{2}\|\Pi^{(n)}\| T,\end{aligned}$$

which indicates that as long as we construct $\{B_t^{(n)}\}_{t\ge 0}$ such that $\|\Pi^{(n)}\| \to 0$, $\mathbb{E}\left[\int_0^T (B_t^{(n)} - B_t)^2 \, \mathrm{d}t\right] \to 0$ as $n \to \infty$.

2. The Ito integral of the simple process $\{B_t^{(n)}\}_{t\ge 0}$ is

$$I_T(B^{(n)}) = \sum_j B_{t_j^{(n)}} (B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}})$$

We will show this term converges to $\frac{1}{2}[B_T^2 - T]$ in $L^2$. Observe that

$$B_{t_j^{(n)}} (B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}) = \frac{1}{2}\left[B_{t_{j+1}^{(n)}}^2 - B_{t_j^{(n)}}^2 - \left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)^2\right]$$

which implies

$$\begin{aligned}I_T(B^{(n)}) &= \frac{1}{2}\left[\sum_j (B_{t_{j+1}^{(n)}}^2 - B_{t_j^{(n)}}^2) - \sum_j \left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)^2\right] \\ &= \frac{1}{2}\left[(B_T^2 - B_0^2) - Q(\Pi^{(n)})\right] = \frac{1}{2}\left[B_T^2 - Q(\Pi^{(n)})\right]\end{aligned}$$

where $Q(\Pi^{(n)})$ is the quadratic variation of Brownian motion $\{B_t\}_{t\ge 0}$ over the partition $\Pi^{(n)}$. Recall that $Q(\Pi^{(n)}) \xrightarrow{L^2} T$, which implies $I_T(B^{(n)}) \xrightarrow{L^2} \frac{1}{2}\left[B_T^2 - T\right]$.

So we conclude that

$$I_T(B) \triangleq \int_0^T B_t \, \mathrm{d}B_t = \frac{1}{2}\left[B_T^2 - T\right].$$

■

## 10.2.2. Properties of Ito Integral

**Proposition 10.2** For any $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$, the Ito integral $\int_0^T X_t \, \mathrm{d}B_t$ has the following properties:

1. Linearity: let $\{X_t\}_{t\geq 0}, \{Y_t\}_{t\geq 0} \in \mathcal{L}^2$, for any $\alpha, \beta$,
$$\int_0^T (\alpha X_t + \beta Y_t) \, \mathrm{d}B_t = \alpha \int_0^T X_t \, \mathrm{d}B_t + \beta \int_0^T Y_t \, \mathrm{d}B_t.$$
2. Ito isometry: for any $0 \leq s < t < T$,
$$\mathbb{E}\left[\left(\int_s^t X_u \, \mathrm{d}B_u\right)^2 \middle| \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t X_u^2 \, \mathrm{d}u \middle| \mathcal{F}_s\right].$$
3. $\{I_t(X)\}_{t\geq 0} \in \mathcal{U}_c^2$.

*Proof.* The linearity property is trivial to show, and the third property comes from Theorem 10.2. It suffices to show the second propperty. Let $\{X_t^{(n)}\}_{t\geq 0}, n = 1, 2, \ldots$ be the sequence of simple processes approximating $\{X_t\}_{t\geq 0}$, and $A \in \mathcal{F}_S$. It follows that

$$\begin{aligned}
\mathbb{E}\left[\left(\int_s^t X_u \, \mathrm{d}B_u\right)^2 1_A\right] &= \mathbb{E}\left[(I_t(X) - I_s(X))^2 1_A\right] \\
&= \mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)}) + I_t(X^{(n)}) - I_s(X^{(n)}) + I_s(X^{(n)}) - I_s(X)\right)^2 1_A\right] \\
&= \underbrace{\mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)^2 1_A\right]}_{(a)} + \underbrace{\mathbb{E}\left[\left(I_t(X^{(n)}) - I_s(X^{(n)})\right)^2 1_A\right]}_{(b)} \\
&\quad + \underbrace{\mathbb{E}\left[\left(I_s(X^{(n)}) - I_s(X)\right)^2 1_A\right]}_{(c)} \\
&\quad + \underbrace{2\mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)\left(I_t(X^{(n)}) - I_s(X^{(n)})\right) 1_A\right]}_{(d)} \\
&\quad + \underbrace{2\mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)\left(I_s(X^{(n)}) - I_s(X)\right) 1_A\right]}_{(e)} \\
&\quad + \underbrace{2\mathbb{E}\left[\left(I_t(X^{(n)}) - I_s(X^{(n)})\right)\left(I_s(X^{(n)}) - I_s(X)\right) 1_A\right]}_{(f)}
\end{aligned}$$

It is easy to show that $(a), (c)$ vanishes as $n \to \infty$:

$$\mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)^2 1_A\right] \leq \mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)^2\right] \to 0.$$

It is also easy to show that $(d), (e), (f)$ vanishes as $n \to \infty$. For instance,

$$\mathbb{E}\left[\left(I_t(X) - I_t(X^{(n)})\right)\left(I_t(X^{(n)}) - I_s(X^{(n)})\right) 1_A\right] \leq \mathbb{E}\left[\left|I_t(X) - I_t(X^{(n)})\right|\left|I_t(X^{(n)}) - I_s(X^{(n)})\right|\right]$$
$$\leq \left(\mathbb{E}\left[(I_t(X) - I_t(X^{(n)}))^2\right]\right)^{1/2}\left(\mathbb{E}\left[(I_t(X^{(n)}) - I_s(X^{(n)}))^2\right]\right)^{1/2} \to 0,$$

as $\mathbb{E}\left[(I_t(X) - I_t(X^{(n)}))^2\right] \to 0$ as $n \to \infty$.

It remains to show that the term (b) in fact converges to $\mathbb{E}\left[\left(\int_s^t X_u^2 \, \mathrm{d}u\right) 1_A\right]$ as $n \to \infty$:

$$\mathbb{E}\left[\left(I_t(X^{(n)}) - I_s(X^{(n)})\right)^2 1_A\right] = \mathbb{E}\left[\left(\int_s^t X_u^{(n)} \, \mathrm{d}B_u\right)^2 1_A\right] = \mathbb{E}\left[\left(\int_s^t (X_u^{(n)})^2 \, \mathrm{d}u\right) 1_A\right],$$

and

$$\begin{aligned}
&\left|\mathbb{E}\left[\left(\int_s^t (X_u^{(n)})^2 \, \mathrm{d}u\right) 1_A\right] - \mathbb{E}\left[\left(\int_s^t (X_u)^2 \, \mathrm{d}u\right) 1_A\right]\right| \\
=&\left|\mathbb{E}\left[\left(\int_s^t (X_u^{(n)} + X_u)(X_u^{(n)} - X_u) \, \mathrm{d}u\right) 1_A\right]\right| \\
\leq&\left|\mathbb{E}\left[\left(\int_s^t |X_u^{(n)} + X_u||X_u^{(n)} - X_u| \, \mathrm{d}u\right)\right]\right| \\
\leq&\mathbb{E}\left[\int_s^t (X_u^{(n)} + X_u)^2 \, \mathrm{d}u\right]^{1/2} \mathbb{E}\left[\int_s^t (X_u^{(n)} - X_u)^2 \, \mathrm{d}u\right]^{1/2}
\end{aligned}$$

Note that the first term is bounded:

$$\begin{aligned}
\mathbb{E}\left[\int_s^t (X_u^{(n)} + X_u)^2 \, \mathrm{d}u\right] &= \mathbb{E}\left[\int_s^t (X_u^{(n)} - X_u + 2X_u)^2 \, \mathrm{d}u\right] \\
&\leq 2\mathbb{E}\left[\int_s^t (X_u^{(n)} - X_u)^2 \, \mathrm{d}u\right] + 2\mathbb{E}\left[\int_s^t (2X_u)^2 \, \mathrm{d}u\right] < \infty.
\end{aligned}$$

Together with the fact that the second term vanishes as $n \to \infty$, we assert that

$$\left|\mathbb{E}\left[\left(\int_s^t (X_u^{(n)})^2 \, \mathrm{d}u\right) 1_A\right] - \mathbb{E}\left[\left(\int_s^t (X_u)^2 \, \mathrm{d}u\right) 1_A\right]\right| \to 0.$$

Or equivalently,

$$\lim_{n\to\infty} \mathbb{E}\left[\left(I_t(X^{(n)}) - I_s(X^{(n)})\right)^2 1_A\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\int_s^t (X_u^{(n)})^2 \,\mathrm{d}u\right) 1_A\right] = \mathbb{E}\left[\left(\int_s^t (X_u)^2 \,\mathrm{d}u\right) 1_A\right].$$

The proof is completed.

■

# Chapter 11

# Week11

## 11.1. Tuesday

### 11.1.1. Quadratic Variation of Ito Integral

Recall that the Ito integral of the process $\{X_t\}_{t\ge 0} \in \mathcal{L}^2$ is denoted as

$$I_t(X) = \int_0^t X_u \, \mathrm{d}B_u.$$

It is a random function of $t$, and it is continuous and adapted. However, compute $I_t(X)$ directly from definition is sophisticated. The quadratic variation formula of Ito integral plays a central role for simplifying the calculation of Ito integral. In this lecture, we will give a introduction on this topic.

**Definition 11.1** [Qudaratic Variation] Suppose that $X$ is a function of time $t$, define its qudaratic variation over the interval $[0,t]$ as the limit (when it exists)[a]

$$Q(\Pi^{(n)};X)[0,t] = \sum_{i=1}^{n}[X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}], \tag{11.1}$$

where the limit is taken over the partitions

$$\Pi^{(n)} := \{0 = t_0^{(n)}, t_1^{(n)}, \dots, t = t_n^{(n)}\},$$

with $\|\Pi^{(n)}\| = \max_j (t_j^{(n)} - (t_{j-1}^{(n)}) \to 0$. ■

[a] To simplify the notation, sometimes we write $X[t]$ for $X_t$, and $X[t](\omega)$ for $X_t(\omega)$.

We first discuss how to compute the quadratic variation for $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$, by applying the **Doob-Meyer** Decomposition trick:

**Definition 11.2** [Doob-Meyer Decomposition] Consider $\{X_t\}_{t\geq 0} \in \mathcal{U}^2$, then $\{X_t^2\}_{t\geq 0}$ is a non-negative sub-martinalge. The process $X_\cdot^2$ admits the unique Doob-Meyer decomposition:

$$X_t^2 = M_t + A_t, \quad 0 \leq t < \infty,$$

with $\{M_t\}_{t\geq 0}$ being a right-continuous martingale and $\{A_t\}_{t\geq 0}$ being an increasing predictable process. In particular, when $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$, the processes $\{M_t\}_{t\geq 0}$ and $\{A_t\}_{t\geq 0}$ are continuous, and denote

$$\langle X \rangle \triangleq A, \quad \langle X \rangle_0 = 0.$$

■

■ **Example 11.1** Consider the Brownian motion $\{B_t\}_{t\geq 0} \in \mathcal{U}_c^2$, then the Doob-Meyer Decomposition is

$$B_t^2 = (B_t^2 - t) + t,$$

where $M_t := B_t^2 - t$ is a martingale, and $\langle B \rangle_t = t$ is increasing. ■

The following theorem illustrates $\langle X \rangle$ can be constructed by taking the limit of quadratic variation:

**Theorem 11.1** Let $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$ and $\{\Pi^{(n)}\}_{n=1}^{\infty}$ be a sequence of partitions on $[0,\infty)$ with $\|\Pi^{(n)}\| \to 0$. Then for any $t > 0$, then the quadratic variation

$$Q(\Pi^{(n)}; X)[0,t] = \sum_j [X_{t_{j+1}^{(n)} \wedge t} - X_{t_j^{(n)} \wedge t}]^2$$

converge to $\langle X \rangle_t$ in probability. Here we call $\{\langle X \rangle_t\}_{t\geq 0}$ the **quadratic variational process** of $X$.

We first consider the bounded process $X_\bullet$ and $\langle X \rangle_\bullet$, and then extend the results into general case by the localization techinique for martingales.

*Proof.* 1. We first consider the bounded process $X_\cdot$ on $[0,t]$, i.e., $|X_s| \leq K$ almost surely for any $s \in [0,t]$, and $\langle X \rangle_s \leq K$ almost surely for all $s \in [0,t]$. In order to show the desired result, we will show the stronger result in the sense that $\mathbb{E}\left[Q(\Pi^{(n)};X)[0,t] - \langle X \rangle_t\right]^2 \to 0$.

$$\begin{aligned}
&\mathbb{E}\left[Q(\Pi^{(n)};X)[0,t] - \langle X \rangle_t\right]^2 = \mathbb{E}\left[\sum_j \left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \langle X \rangle_t\right]^2 \\
=&\mathbb{E}\left[\sum_j \left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \sum_j \left(\langle X \rangle[t_{j+1}^{(n)} \wedge t] - \langle X \rangle[t_j^{(n)} \wedge t]\right)\right]^2 \\
=&\mathbb{E}\left[\sum_j \left\{\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \left(\langle X \rangle[t_{j+1}^{(n)} \wedge t] - \langle X \rangle[t_j^{(n)} \wedge t]\right)\right\}\right]^2 \\
=&\sum_j \mathbb{E}\left\{\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \left(\langle X \rangle[t_{j+1}^{(n)} \wedge t] - \langle X \rangle[t_j^{(n)} \wedge t]\right)\right\}^2 \\
&+ 2\sum_{j<k} \mathbb{E}\left\{\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \left(\langle X \rangle[t_{j+1}^{(n)} \wedge t] - \langle X \rangle[t_j^{(n)} \wedge t]\right)\right\} \\
&\times \left\{\left(X[t_{k+1}^{(n)} \wedge t] - X[t_k^{(n)} \wedge t]\right)^2 - \left(\langle X \rangle[t_{k+1}^{(n)} \wedge t] - \langle X \rangle[t_k^{(n)} \wedge t]\right)\right\}
\end{aligned} \tag{11.2}$$

Now we show that the second part actually vanishes. For any $0 \leq s < t \leq u < v$,

$$\begin{aligned}
&\mathbb{E}\left[(X_t - X_s)^2 - (\langle X \rangle_t - \langle X \rangle_s)\right]\left[(X_v - X_u)^2 - (\langle X \rangle_v - \langle X \rangle_u)\right] \\
=&\mathbb{E}\left\{\mathbb{E}\left(\left[(X_t - X_s)^2 - (\langle X \rangle_t - \langle X \rangle_s)\right]\left[(X_v - X_u)^2 - (\langle X \rangle_v - \langle X \rangle_u)\right]\middle|\mathcal{F}_t\right)\right\} \\
=&\mathbb{E}\left\{\left[(X_t - X_s)^2 - (\langle X \rangle_t - \langle X \rangle_s)\right]\mathbb{E}\left(\left[(X_v - X_u)^2 - (\langle X \rangle_v - \langle X \rangle_u)\right]\middle|\mathcal{F}_t\right)\right\}
\end{aligned} \tag{11.3}$$

Moreover,

$$\begin{aligned}
\mathbb{E}\left[(X_v - X_u)^2 \mid \mathcal{F}_t\right] &= \mathbb{E}\left[X_v^2 + X_u^2 - 2X_vX_u \mid \mathcal{F}_t\right] \\
&= \mathbb{E}\left[X_v^2 + X_u^2 - 2\mathbb{E}[X_vX_u \mid \mathcal{F}_u] \mid \mathcal{F}_t\right] \\
&= \mathbb{E}\left[X_v^2 + X_u^2 - 2X_u^2 \mid \mathcal{F}_t\right] = \mathbb{E}\left[X_v^2 - X_u^2 \mid \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\langle X \rangle_v - \langle X \rangle_u \mid \mathcal{F}_t\right],
\end{aligned}$$

where the last equality is because $\{X^2 - \langle X \rangle\}$ is a martingale. This implies the term in (11.3) vanishes, and therefore the second part in (11.2) vanishes.

By the elementary inequality $(a-b)^2 \le 2a^2 + 2b^2$, the first part in (11.2) can be upper bounded as

$$\begin{aligned}
&\sum_j \mathbb{E}\left\{\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \left(\langle X\rangle[t_{j+1}^{(n)} \wedge t] - \langle X\rangle[t_j^{(n)} \wedge t]\right)\right\}^2 \\
\le& 2\sum_j \mathbb{E}\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^4 + 2\sum_j \mathbb{E}\left(\langle X\rangle[t_{j+1}^{(n)} \wedge t] - \langle X\rangle[t_j^{(n)} \wedge t]\right)^2
\end{aligned} \tag{11.4}$$

We claim that:

(a) $\sum_j \mathbb{E}\left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^4 \to 0$ as $n \to \infty$;

(b) $\sum_j \mathbb{E}\left(\langle X\rangle[t_{j+1}^{(n)} \wedge t] - \langle X\rangle[t_j^{(n)} \wedge t]\right)^2 \to 0$ as $n \to \infty$.

Then $\mathbb{E}\left[Q(\Pi^{(n)};X)[0,t] - \langle X\rangle_t\right]^2 \to 0$ as $n \to \infty$, i.e., $Q(\Pi^{(n)};X)[0,t] \xrightarrow{L^2} \langle X\rangle_t$. The desired result holds since $L^2$ convergence implies convergence in probability.

2. Then consider the case where $\{X_t\}_{t\ge 0}, \langle X\rangle$ are unbounded. We argue for this case by the technique of localization. Define a sequence of stopping time

$$T_k = \inf\{t \ge 0 : \ |X_t| \ge K \text{ or } \langle X\rangle_t \ge K\}.$$

Thus the stopped process $\{X[t \wedge T_k]\}_{t\ge 0}$ is a bounded martingale, denoted as $X^{(k)}[t] \equiv X[t \wedge T_k]$. By the Doob-Meyer decomposition,

$$(X^{(k)})^2 = \left\{(X_t^{(k)})^2 - \langle X^{(k)}\rangle_t\right\} + \langle X^{(k)}\rangle_t.$$

Now we begin to simplify the formula presented above by the uniqueness of Doob-Meyer decomposition (Inituitively we should have $\langle X^{(k)}\rangle[t] = \langle X\rangle[t \wedge T_k]$).

- The stopped process $\{X[t \wedge T_k]^2 - \langle X\rangle[t \wedge T_k]\}$ is a martingale because $\{X_t^2 - \langle X\rangle_t\}$ is a martingale and $T_k$ is a stopping time.

- Moreover, $\langle X\rangle[t \wedge T_k]$ is increasing and predicatable.

Hence $X^{(k)} \equiv X[t \wedge T_k]$ admits the Doob-Meyer decomposition

$$(X^{(k)})^2 = \{X[t \wedge T_k]^2 - \langle X\rangle[t \wedge T_k]\} + \langle X\rangle[t \wedge T_k].$$

By the uniqueness of Doob-Meyer decomposition, $\langle X^{(k)}\rangle[t] = \langle X\rangle[t \wedge T_k]$. Applying the result in part 1) into the bounded process $X^{(k)}_\bullet$, for any $\varepsilon > 0, \eta > 0$ there exists $N$ such that as long as $n > N$,

$$\mathbb{P}\left(|Q(\Pi^{(n)}; X^{(k)})[0,t] - \langle X^{(k)}\rangle[t]| > \varepsilon\right) = \mathbb{P}\left(|Q(\Pi^{(n)}; X^{(k)})[0,t] - \langle X\rangle[t \wedge T_k]| > \varepsilon\right) < \frac{\eta}{2}. \tag{11.5}$$

Now we begin to show that $Q(\Pi^{(n)}; X)[0,t] \xrightarrow{P} \langle X\rangle_t$:

$$\begin{aligned}
&\mathbb{P}\left(|Q(\Pi^{(n)}; X)[0,t] - \langle X\rangle[t]| > \varepsilon\right) \\
=&\mathbb{P}\left(|Q(\Pi^{(n)}; X)[0,t] - \langle X\rangle[t]| > \varepsilon, T_k \geq t\right) + \mathbb{P}\left(|Q(\Pi^{(n)}; X)[0,t] - \langle X\rangle[t]| > \varepsilon, T_k < t\right) \\
\leq&\mathbb{P}\left(|Q(\Pi^{(n)}; X)[0,t] - \langle X\rangle[t]| > \varepsilon, T_k \geq t\right) + \mathbb{P}(T_k < t) \qquad (11.6a) \\
=&\mathbb{P}\left(\left|\sum_j \left(X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t]\right)^2 - \langle X\rangle[t \wedge T_k]\right| > \varepsilon, T_k \geq t\right) + \mathbb{P}(T_k < t) \\
=&\mathbb{P}\left(\left|\sum_j \left(X[t_{j+1}^{(n)} \wedge T_k \wedge t] - X[t_j^{(n)} \wedge T_k \wedge t]\right)^2 - \langle X\rangle[t \wedge T_k]\right| > \varepsilon, T_k \geq t\right) + \mathbb{P}(T_k < t) \\
\leq&\mathbb{P}\left(\left|\sum_j \left(X[t_{j+1}^{(n)} \wedge T_k \wedge t] - X[t_j^{(n)} \wedge T_k \wedge t]\right)^2 - \langle X\rangle[t \wedge T_k]\right| > \varepsilon\right) + \mathbb{P}(T_k < t) \\
=&\mathbb{P}\left(\left|Q(\Pi^{(n)}; X^{(k)})[0,t] - \langle X\rangle[t \wedge T_k]\right| > \varepsilon\right) + \mathbb{P}(T_k < t) \\
\leq&\frac{\eta}{2} + \frac{\eta}{2} = \eta. \qquad (11.6b)
\end{aligned}$$

where the upper bound presented (11.6a) is because $\mathbb{P}(T_k < \infty) \to 0$, and (11.6b) makes use of the upper bound (11.5).

■

In order to complete the proof, it remains to show the following two claims are correct.

**Proposition 11.1** Let $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$ with $|X_s| \leq K$ almost surely for any $s \in [0,t]$. Let $\{\Pi^{(n)}\}_{n=1}^{\infty}$ be a sequence of partitions on $[0,t]$ with $\|\Pi^{(n)}\| \to 0$. Then

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_j (X[t_{j+1}^{(n)}] - X[t_j^{(n)}])^4\right] = 0.$$

*Proof.* 1. We first show that $\mathbb{E}\left[Q(\Pi;X)[0,t]\right]^2$ is bounded for any partition $\Pi$:

$$\begin{aligned}
\mathbb{E}\Big[Q(\Pi;X)[0,t]\Big]^2 &= \mathbb{E}\left[\sum_j (X[t_{j+1}] - X[t_j])^2\right]^2 \\
&= \mathbb{E}\left[\sum_j (X[t_{j+1}] - X[t_j])^4\right] + 2\mathbb{E}\left[\sum_k\sum_{j>k}(X[t_{j+1}] - X[t_j])^2(X[t_{k+1}] - X[t_k])^2\right]
\end{aligned} \tag{11.7}$$

In particular,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j>k}(X[t_{j+1}] - X[t_j])^2 \middle| \mathcal{F}_{t_{k+1}}\right] = \mathbb{E}\left[\sum_{j>k}(X[t_{j+1}]^2 + X[t_j]^2 - 2X[t_{j+1}]X[t_j]) \middle| \mathcal{F}_{t_{k+1}}\right] \\
=&\mathbb{E}\left[\sum_{j>k} X[t_{j+1}]^2 + X[t_j]^2 - 2\mathbb{E}\left(X[t_{j+1}]X[t_j]\middle|\mathcal{F}_{t_j}\right)\middle|\mathcal{F}_{t_{k+1}}\right] \\
=&\mathbb{E}\left[\sum_{j>k} X[t_{j+1}]^2 - X[t_j]^2 \middle| \mathcal{F}_{t_{k+1}}\right] \leq \mathbb{E}\left[X[t_{k+2}]|\mathcal{F}_{t_{k+1}}\right] \leq K^2.
\end{aligned}$$

Taking $k = 0$, we have $\mathbb{E}\left[\sum_{j\geq 1}(X[t_{j+1}] - X[t_j])^2\right] \leq K^2$. We can apply these inequalities to upper bound the second term in (11.7):

$$\begin{aligned}
&\mathbb{E}\left[\sum_k\sum_{j>k}(X[t_{j+1}] - X[t_j])^2(X[t_{k+1}] - X[t_k])^2\right] \\
=&\mathbb{E}\left[\sum_k \mathbb{E}\left\{\sum_{j>k}(X[t_{j+1}] - X[t_j])^2(X[t_{k+1}] - X[t_k])^2\middle|\mathcal{F}_{t_{k+1}}\right\}\right] \\
=&\mathbb{E}\left[\sum_k (X[t_{k+1}] - X[t_k])^2\mathbb{E}\left\{\sum_{j>k}(X[t_{j+1}] - X[t_j])^2\middle|\mathcal{F}_{t_{k+1}}\right\}\right] \\
\leq&\mathbb{E}\left[\sum_k (X[t_{k+1}] - X[t_k])^2 K^2\right] \leq K^4.
\end{aligned}$$

Then we upper bound the first term in (11.7):

$$\begin{aligned}
&\mathbb{E}\left[\sum_j (X[t_{j+1}] - X[t_j])^4\right] \\
=&\mathbb{E}\left[\sum_j (X[t_{j+1}] - X[t_j])^2 (X[t_{j+1}] - X[t_j])^2\right] \\
\le&\mathbb{E}\left[\sum_j 4K^2 (X[t_{j+1}] - X[t_j])^2\right] \le 4K^4,
\end{aligned}$$

where the first inequality is because $\mathbb{E}\left[\sum_{j\ge 1}(X[t_{j+1}] - X[t_j])^2\right] \le K^2$. Then we can assert that $\mathbb{E}\left[Q(\Pi;X)[0,t]\right]^2 \le 6K^4$.

2. In order to show the desired result, we start with simplifying $\sum_j (X[t_{j+1}] - X[t_j])^4$:

$$\begin{aligned}
\sum_j (X[t_{j+1}] - X[t_j])^4 &= \sum_j (X[t_{j+1}] - X[t_j])^2 (X[t_{j+1}] - X[t_j])^2 \\
&\le \left(\sup_j |X[t_{j+1}] - X[t_j]|\right)^2 \sum_j (X[t_{j+1}] - X[t_j])^2 \\
&\le A_\Pi^2 \sum_j (X[t_{j+1}] - X[t_j])^2
\end{aligned}$$

in which we define $A_\Pi = \sup\{|X_y - X_s| :\ 0 \le s < y \le t, |y-s| \le \|\Pi\|\}$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\sum_j (X[t_{j+1}] - X[t_j])^4 \\
&\le \mathbb{E}\left[A_\Pi^2 \sum_j (X[t_{j+1}] - X[t_j])^2\right]^2 \\
&\le \left\{\mathbb{E}A_\Pi^4\right\}^{1/2} \left\{\mathbb{E}\left[\sum_j (X[t_{j+1}] - X[t_j])^2\right]^2\right\}^{1/2} \\
&= \left\{\mathbb{E}A_\Pi^4\right\}^{1/2} \left\{\mathbb{E}\left[Q(\Pi;X)[0,t]\right]^2\right\}^{1/2} \\
&\le \left\{\mathbb{E}A_\Pi^4\right\}^{1/2} \left\{6K^4\right\}^{1/2}
\end{aligned}$$

Since $X$ is continuous and thus uniformly continuous on $[0,t]$, $A_\Pi \to 0$ as $\|\Pi\| \to 0$. Applying bounded convergence theorem gives $\mathbb{E}A_\Pi^4 \to 0$. The proof is completed.

$\blacksquare$

**Proposition 11.2** Let $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$ and $\langle X\rangle_s \leq K$ almost surely for any $s \in [0,t]$. Suppose that $\{\Pi^{(n)}\}$ is a sequence of partitions on $[0,t]$ such that $\|\Pi^{(n)}\| \to 0$, then we have

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_j \left(\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right)^2\right] = 0.$$

*Proof.* The proof is by direct computing and upper bounding the term $\mathbb{E}\left[\sum_j \left(\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right)^2\right]$:

$$\begin{aligned}
&\mathbb{E}\left[\sum_j \left(\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right)^2\right] \\
\leq&\mathbb{E}\left[\sup_j \left|\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right| \cdot \sum_j \left(\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right)\right] \\
=&\mathbb{E}\left[\sup_j \left|\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right| \cdot \langle X\rangle_t\right] \leq \mathbb{E}[\hat{A}_\Pi \cdot \langle X\rangle_t],
\end{aligned}$$

where we define $\hat{A}_\Pi = \sup\{|\langle X\rangle_y - \langle X\rangle_s| : \ 0 \leq s < y \leq t, |y-s| \leq \|\Pi\|\}$. Since $\langle X\rangle$ is continous, and therefore uniformly continuous on $[0,t]$, when $\|\Pi\| \to 0$, $\hat{A}_\Pi \to 0$. It is easy to see that $|\hat{A}_\Pi \cdot \langle X\rangle_t| \leq 2K^2$. By bounded convergence theorem, as $\|\Pi\| \to 0$,

$$\mathbb{E}\left[\sum_j \left(\langle X\rangle[t_{j+1}^{(n)}] - \langle X\rangle[t_j^{(n)}]\right)^2\right] \leq \mathbb{E}[\hat{A}_\Pi \cdot \langle X\rangle_t] \to 0.$$

The proof is completed. $\blacksquare$

Now we begin to characterize the quadratic variation of Ito's integral $I_t(X)$ for $X_\bullet \in \mathcal{L}^2$. However, we don't prove this result by taking the limit of the formula presented in Theorem 11.1 (because it is a little bit complicated), but making use of the Doob-Meyer decomposition.

**Theorem 11.2** The quadratic variation of the Ito integral $I_t(X)$ for $\{X_t\}_{t\geq 0} \in \mathcal{L}^2$ on the interval $[0,T]$ is

$$\int_0^T X_t^2 \, \mathrm{d}t.$$

*Proof.* The Doob-Meyer decomposition for $\{I_t(X)\}_{t\geq 0} \in \mathcal{U}_c^2$ is

$$I_t^2(X) \triangleq (I_t^2(X) - \langle I(X)\rangle_t) + \langle I(X)\rangle_t.$$

We will characterize this decomposition form as follows. On the one hand,

$$\begin{aligned}
\mathbb{E}[(I_t(X) - I_s(X))^2 \mid \mathcal{F}_s] &= \mathbb{E}[I_t^2(X) - 2I_t(X)I_s(X) + I_s^2(X) \mid \mathcal{F}_s] \\
&= \mathbb{E}[I_t^2(X) \mid \mathcal{F}_s] - 2I_s(X)\mathbb{E}[I_t(X) \mid \mathcal{F}_s] + I_s^2(X) \\
&= \mathbb{E}[I_t^2(X) \mid \mathcal{F}_s] - I_s^2(X).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}[(I_t(X) - I_s(X))^2 \mid \mathcal{F}_s] &= \mathbb{E}\left[\int_s^t X_u^2 \, \mathrm{d}u \mid \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\int_0^t X_u^2 \, \mathrm{d}u \mid \mathcal{F}_s\right] - \int_0^s X_u^2 \, \mathrm{d}u.
\end{aligned}$$

Equating those two terms aand re-arranging gives

$$\mathbb{E}\left[[I_t^2(X) - \int_0^t X_u^2 \, \mathrm{d}u \middle| \mathcal{F}_s\right] = I_s^2(X) - \int_0^s X_u^2 \, \mathrm{d}u.$$

Hence, $\{I_t^2(X) - \int_0^t X_u^2 \, \mathrm{d}u\}_{t\geq 0}$ is a martingale. Moreover, $\int_0^t X_u^2 \, \mathrm{d}u$ is a predictable increasing process in $t$. By the uniqueness of Doob-Mayer decomposition,

$$\langle I(X)\rangle_t = \int_0^t X_u^2 \, \mathrm{d}u.$$

∎

## 11.2. Thursday

### 11.2.1. Quadratic Covariation

Note that quadratic variation characterizes the sum of the quadratic of jumps for a single process. Now we aim to quantify the sum of product of jumps for two processes $\{X_t\}_{t\ge 0}, \{Y_t\}_{t\ge 0} \in \mathcal{U}_c^2$:

**Definition 11.3** [Quadratic Covariation] Define the quadratic covariation of $X_\bullet, Y_\bullet \in \mathcal{U}_c^2$ as

$$Q_c(\Pi; X, Y)[0,t] = \sum_j \left( X[t_{j+1}^{(n)} \wedge t] - X[t_j^{(n)} \wedge t] \right) \left( Y[t_{j+1}^{(n)} \wedge t] - Y[t_j^{(n)} \wedge t] \right)$$

Define the quadratic covariational process of $X$ and $Y$ as

$$\langle X, Y \rangle = \frac{1}{4}(\langle X+Y \rangle - \langle X-Y \rangle),$$

which is well-defined since $X+Y, X-Y \in \mathcal{U}_c^2$.

Similar as in Theorem 11.1, we can show that the quadratic covariational process is the limit of the quadratic covariation as $\|\Pi\| \to 0$.

**Theorem 11.3** Let $\{\Pi^{(n)}\}_{n=1}^\infty$ be a sequence of partitions with $\|\Pi^{(n)}\| \to 0$. Then the quadratic covariation $Q_c(\Pi^{(n)}; X, Y)[0,t]$ converges to $\langle X, Y \rangle_t$ in probability.

*Proof.* It suffices to show that

$$Q_c(\Pi; X, Y)[0,t] = \frac{1}{4}\left(Q(\Pi; X+Y)[0,t] + Q(\Pi; X-Y)[0,t]\right), \tag{11.8}$$

and the remainings can be shown by taking the limit of the quadratic variation $Q(\Pi^{(n)}; X+Y)[0,t]$ and $Q(\Pi^{(n)}; X-Y)[0,t]$. Here we begin to simplify the RHS in

(11.8):

$$Q(\Pi; X+Y)[0,t] = \sum_j \left(X[t_{j+1}\wedge t] + Y[t_{j+1}\wedge t] - X[t_j \wedge t] - Y[t_{j+1}\wedge t]\right)^2$$
$$= \sum_j \left(X[t_{j+1}\wedge t] - X[t_j\wedge t]\right)^2 + \sum_j \left(Y[t_{j+1}\wedge t] - Y[t_{j+1}\wedge t]\right)^2$$
$$+ 2\sum_j \left(X[t_{j+1}\wedge t] - X[t_j\wedge t]\right)\left(Y[t_{j+1}\wedge t] - Y[t_{j+1}\wedge t]\right).$$

Similarly,

$$Q(\Pi; X-Y)[0,t] = \sum_j \left(X[t_{j+1}\wedge t] - X[t_j\wedge t]\right)^2 + \sum_j \left(Y[t_{j+1}\wedge t] - Y[t_{j+1}\wedge t]\right)^2$$
$$- 2\sum_j \left(X[t_{j+1}\wedge t] - X[t_j\wedge t]\right)\left(Y[t_{j+1}\wedge t] - Y[t_{j+1}\wedge t]\right).$$

Then it is clear that (11.8) holds. The proof is completed. ■

**Proposition 11.3 — Properties about Quadratic Covariational Process.** Let stochastic processes $X_\bullet, Y_\bullet, X^{(1)}_\bullet, X^{(2)}_\bullet \in \mathcal{U}_c^2$, then the following properties hold:

1. Symmetry: $\langle X, Y\rangle = \langle Y, X\rangle$;
2. For any $a, b \in \mathbb{R}$, $\langle aX^{(1)} + bX^{(2)}, Y\rangle = a\langle X^{(1)}, Y\rangle + b\langle X^{(2)}, Y\rangle$;
3. Cauchy-Schwarz inequality: $|\langle X, Y\rangle_t|^2 \le \langle X\rangle_t \langle Y\rangle_t$.

*Proof.* Part 1) and part 2) can be shown by definition. For part 3), we first apply Cauchy-scharz inequality on the quadratic covariation of $X$ and $Y$:

$$(Q_c(\Pi; X, Y)[0,t])^2 = \left(\sum_j \left(X[t_{j+1}^{(n)}\wedge t] - X[t_j^{(n)}\wedge t]\right)\left(Y[t_{j+1}^{(n)}\wedge t] - Y[t_j^{(n)}\wedge t]\right)\right)^2$$
$$\le \left(\sum_j \left(X[t_{j+1}^{(n)}\wedge t] - X[t_j^{(n)}\wedge t]\right)^2\right)\left(\sum_j \left(Y[t_{j+1}^{(n)}\wedge t] - Y[t_j^{(n)}\wedge t]\right)^2\right)$$
$$= Q(\Pi; X)[0,t]\, Q(\Pi; Y)[0,t]$$

Then taking the limit on both sides gives the desired result. ■

R Recall that a monotone function has the bounded variation over any finite intervals. Then $\langle X \rangle$ has the bounded variation since it is increasing. It follows that the quadratic covariational process $\langle X, Y \rangle$ has the bounded variation since it is a linear combination of two quadratic variational processes.

Next, we show that the product process $XY$ admits the decomposition that is similar to Doob-Meyer decomposition, though in this case $\langle X, Y \rangle$ may not be increasing:

■ **Example 11.2** We can show that $XY - \langle X, Y \rangle$ is a martingale as follows. Since stochastic processes $X + Y, X - Y \in \mathcal{U}_c^2$, the Doob-Meyer decomposition implies that $(X+Y)^2 - \langle X + Y \rangle, (X-Y)^2 - \langle X - Y \rangle$ are martingales. Then we can express the term $XY - \langle X, Y \rangle$ as the linear combination of two martingales, which is a martingale as well:

$$
\begin{aligned}
&\frac{1}{4}\left[[(X+Y)^2 - \langle X+Y \rangle] - [(X-Y)^2 - \langle X-Y \rangle]\right] \\
=&\frac{1}{4}\left[[X^2 + 2XY + Y^2 - X^2 + 2XY - Y^2] - [\langle X+Y \rangle - \langle X-Y \rangle]\right] \\
=&XY - \frac{1}{4}[\langle X+Y \rangle - \langle X-Y \rangle].
\end{aligned}
$$

■

## 11.2.2. Ito Integral for General Processes

**Ito Integral for Simple Processes.** Let $M_\bullet \in \mathcal{U}_c^2$, and $\{X_t\}_{t \ge 0}$ be a simple process with the corresponding partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n \le T\}$. Then the Ito integral of $X_t$ w.r.t. $M_t$ is defined as

$$
\int_0^T X_t \, \mathrm{d}M_t = \sum_{j=1}^{n-1} X[t_j] \cdot (M[t_{j+1}] - M[t_j]) + X[t_n](M[T] - N[t_n]),
$$

denoted as $I_T(X)$. Immediately we have the following properties:

1. $\{I_t(X)\}_{t \ge 0} \in \mathcal{U}_c^2$;

2. The quadratic variational process of $I_\bullet(X)$ is

$$\langle I(X)\rangle_t = \int_0^t X_u^2 \,\mathrm{d}\langle M\rangle_u.$$

The process $\left\{I_t^2(X) - \int_0^t X_u^2 \,\mathrm{d}\langle M\rangle_u\right\}_{t\ge 0}$ is a martingale;

3. Ito isometry: for any $t \ge 0$ we have

$$\mathbb{E}\left[\int_0^t X_u \,\mathrm{d}M_u\right]^2 = \mathbb{E}\left[\int_0^t X_u^2 \,\mathrm{d}\langle M\rangle_u\right].$$

We need some conditions, such as integrability results about $X$, to define the Ito integral $\int_0^T X_t \,\mathrm{d}M_t$ for general stochastic processes $X_\bullet$.

**Definition 11.4** [Ito Integrable Space] The set $\mathcal{L}^2(M)$ denotes the space containing all adapted stochastic processes $\{X_t\}_{t\ge 0}$, such that, there exists a sequence of simple processes $\{X_t^{(n)}\}_{t\ge 0}$ satisfying

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (X_t^{(n)} - X_t)^2 \,\mathrm{d}\langle M\rangle_t\right] = 0, \quad \forall T > 0.$$

■

**Ito's Integral on $\mathcal{L}^2(M)$.** For any $X_\bullet \in \mathcal{L}^2(M)$ and associated simple processes $X_\bullet^{(n)}$ appproximating $X_\bullet$, there exists a limit of the Ito integral $\{I_t(X^{(n)})\}_{t\ge 0}$ in the closed set $\mathcal{U}_c^2$, Then we take

$$\int_0^T X_t \,\mathrm{d}M_t = \lim_{n\to\infty} \int_0^T X_t^{(n)} \,\mathrm{d}M_t.$$

**Ito's Integral on Locally Bounded Processes.** The question is whether we could extend the Ito's integral to other processes, such as locally bounded process $\{X_t\}_{t\ge 0}$ and continuous square-integrable local martingale $\{M_t\}_{t\ge 0}$.

**Definition 11.5** [Local Martingale] A process $X_\bullet$ is called a local martingale if there is a sequence of finite stopping times $\{\tau_n\}$ with $\tau_n \uparrow \infty$ so that $X^{\tau_n} \equiv \{X(\tau_n \wedge t)\}_{t\ge 0}$ is a martingale for each $n \ge 1$. ■

**Definition 11.6** [Local Bounded] A process $X_\bullet$ is said to be locally bounded if there is a sequence of finite stopping times $\{\sigma_n\}$ with $\sigma_n \uparrow \infty$ so that $X^{\sigma_n}$ is bounded for each $n \ge 1$.

The answer to the question above is yes. We take $\tilde{\tau}_n = \tau_n \wedge \sigma_n$, then $\tilde{\tau}_n \uparrow \infty$, and $M^{\tilde{\tau}_n} \in \mathcal{U}_c^2$. Construct $X_t^{(n)} = X_t 1\{t \le \tilde{\tau}_n\}$, then $X_\bullet^{(n)} \in \mathcal{L}^2(M^{\tilde{\tau}_n})$. Then we define the Ito's integral

$$\int_0^T X_t \,\mathrm{d}M_t = \int_0^T X_t^{(n)} \,\mathrm{d}M^{\tilde{\tau}_n}, \quad T \in [0, \tilde{\tau}_n].$$

Note that this definition is consistent, i.e., does not depend on the particular choice of the sequence $\{\tilde{\tau}_n\}$.

**Ito's Integral on the class of semi-martingales.** Finally, we are wondering whether it is possible to define the stochastic integral on the class of semi-martingales. We define the semi-martingale for continuous case as the following:

**Definition 11.7** [Semi-martingale] We say $X_\bullet$ is a semi-martingale if it admits the decomposition

$$X_t = A_t + M_t,$$

where $M_\bullet$ is a continuous local martingale, and $A_\bullet$ is an adapted process of finite variations:

$$|A|(t) \equiv \sup_{\delta>0, t_0=0} \sup_{t_n - t_{n-1} \ge \delta} \sum_{n=1}^{\infty} 1(t_n \le t) |A[t_n] - A[t_{n-1}]| < \infty, \quad \forall t \ge 0.$$

Let $\{X_t\}_{t\ge 0}$ be a left-continuous adapted process, and $\{Y_t\}_{t\ge 0}$ be a continuous semi-martingale with the decomposition $Y_t = A_t + M_t$. Then define the stochastic integral

$$\int_0^T X_t \,\mathrm{d}Y_t = \int_0^T X_t \,\mathrm{d}M_t + \int_0^T X_t \,\mathrm{d}A_t.$$

Note tht a left-continuous adapted process is locally bounded. Hence the first term $\int_0^T X_t \,\mathrm{d}M_t$ is the Ito integral w.r.t. a continuous local martingale. The second term is the Riemann integation defined for each $\omega \in \Omega$.

# Chapter 12

# Week12

## 12.1. Tuesday

### 12.1.1. Ito's Formula

In this lecture, we will study the Ito's formula, which is very useful for evaluating Ito's integrals. The following elementary identity will be used frequently:

$$X[t_{j+1}]^2 - X[t_j]^2 = (X[t_{j+1}] - X[t_j])^2 + 2X[t_j](X[t_{j+1}] - X[t_j]) \tag{12.1}$$

Now we show how to compute the Ito integral for $X_t^2$ based on this identity:

▪ **Example 12.1** Suppose that $\{X_t\}_{t\geq 0} \in \mathcal{U}_c^2$, and $\Pi = \{t_0 < t_1 < \cdots < t\}$ is a partition on the interval $[0,t]$. Then take the summation on 12.1 both sides yields

$$X_t^2 - X_0^2 = 2\sum_j X[t_j](X[t_{j+1}] - X[t_j]) + \sum_j (X[t_{j+1}] - X[t_j])^2.$$

Taking the limit both sides as $\|\Pi\| \to 0$, we have

$$X_t^2 - X_0^2 = 2\int_0^t X_s \, \mathrm{d}X_s + \langle X \rangle_t.$$

It is the Ito's formula applied to $X_t^2$. ▪

**Theorem 12.1** Let $\{B_t\}_{t\geq 0}$ be a Brownian motion and $f \in \mathcal{C}^2(\mathbb{R})$. Then

$$f(B_t) = f(0) + \int_0^t f'(B_u)\,\mathrm{d}B_u + \frac{1}{2}\int_0^t f''(B_u)\,\mathrm{d}u,$$

almost surely for any $t \geq 0$.

*Proof.* Let $\Pi = \{t_0 < t_1 < \cdots < t_n = t\}$ be a partition on $[0,t]$, then $f(B_t)$ admits the expansion

$$f(B_t) = f(0) + \sum_{j=0}^{n-1}[f(B[t_{j+1}]) - f(B[t_j])]. \tag{12.2}$$

By Taylor expansion on the RHS, there exists $\theta_{t_j}(\omega) \in [B[t_j](\omega), B[t_{j+1}](\omega)]$ such that

$$f(B[t_{j+1}]) - f(B[t_j]) = f'(B[t_j])(B[t_{j+1}] - B[t_j]) + \frac{1}{2}f''(\theta_{t_j}) \cdot (B[t_{j+1}] - B[t_j])^2. \tag{12.3}$$

Substituting (12.3) into (12.2) yields

$$f(B_t) - f(0) = \sum_j f'(B[t_j])(B[t_{j+1}] - B[t_j]) + \frac{1}{2}\sum_j f''(\theta_{t_j}) \cdot (B[t_{j+1}] - B[t_j])^2.$$

As $\|\Pi\| \to 0$, the term $\sum_j f'(B[t_j])(B[t_{j+1}] - B[t_j]) \to \int_0^t f'(B_u)\,\mathrm{d}B_u$ in probability. Then we begin to compute the limit for the second term on RHS.

- We first consider the case where $\theta_{t_j}(\omega) \equiv B_{t_j}$.

$$\begin{aligned}
&\mathbb{E}\left[\sum_j f''(B[t_j]) \cdot (B[t_{j+1}] - B[t_j])^2 - \sum_j f''(B[t_j]) \cdot (t_{j+1} - t_j)\right]^2 \\
=&\mathbb{E}\left[\sum_j f''(B[t_j]) \cdot \left((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right)\right]^2 \\
=&\sum_j \mathbb{E}\left[f''(B[t_j]) \cdot \left((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right)\right]^2 \\
&\quad + 2\sum_{j<k} \mathbb{E}\left[f''(B[t_j]) \cdot \left((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right)\right] \\
&\qquad \times \left[f''(B[t_k]) \cdot \left((B[t_{k+1}] - B[t_k])^2 - (t_{k+1} - t_k)\right)\right]
\end{aligned}$$

The second term vanishes because for any $j < k$,

$$
\begin{aligned}
&\mathbb{E}\Big[f''(B[t_j]) \cdot \big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\Big] \\
&\qquad\qquad \times \Big[f''(B[t_k]) \cdot \big((B[t_{k+1}] - B[t_k])^2 - (t_{k+1} - t_k)\big)\Big] \\
=&\mathbb{E}\mathbb{E}\Big[f''(B[t_j]) \cdot \big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\Big|\mathcal{F}_{t_j}\Big] \\
&\qquad\qquad \times \Big[f''(B[t_k]) \cdot \big((B[t_{k+1}] - B[t_k])^2 - (t_{k+1} - t_k)\big)\Big] \\
=&\mathbb{E}f''(B[t_j])\mathbb{E}\Big[\big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\Big|\mathcal{F}_{t_j}\Big] \\
&\qquad\qquad \times \Big[f''(B[t_k]) \cdot \big((B[t_{k+1}] - B[t_k])^2 - (t_{k+1} - t_k)\big)\Big] = 0,
\end{aligned}
$$

where the first equality is by tower property and the last equality is because $\{B_t^2 - t\}_{t\geq 0}$ is a martingale. To simplify the first term, observe that

$$
\begin{aligned}
&\mathbb{E}\left[f''(B[t_j]) \cdot \big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\right]^2 \\
=&\mathbb{E}\left[f''(B[t_j]) \cdot \mathbb{E}\left((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\Big|\mathcal{F}_{t_j}\right)\right]^2 \\
=&\mathbb{E}\left[f''(B[t_j]) \cdot 2(t_{j+1} - t_j)^2\right]^2 = 2(t_{j+1} - t_j)^2\mathbb{E}\left[f''(B[t_j])\right]^2
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sum_j \mathbb{E}\left[f''(B[t_j]) \cdot \big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\right]^2 \\
=&2\sum_j (t_{j+1} - t_j)^2\mathbb{E}\left[f''(B[t_j])\right]^2 \\
\leq&2\|\Pi\|\sum_j (t_{j+1} - t_j)\mathbb{E}\left[f''(B[t_j])\right]^2 = 2t\|\Pi\|\mathbb{E}\left[f''(B[t_j])\right]^2.
\end{aligned}
$$

(a) Suppose that $f''$ is bounded, i.e., $|f''(x)| \leq K$ for any $x \in \mathbb{R}$, then as $\|\Pi\| \to 0$,

$$
\sum_j \mathbb{E}\left[f''(B[t_j]) \cdot \big((B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\big)\right]^2 \leq 2t\|\Pi\|K^2 \to 0,
$$

which means that $\sum_j f''(B[t_j]) \cdot (B[t_{j+1}] - B[t_j])^2 \xrightarrow{L^2} \sum_j f''(B[t_j]) \cdot (t_{j+1} - t_j)$

and therefore in probability. Also, by the Lebesgue integration knowledge, since $f''$ is continuous and $B_t$ is almost surely continuous, the term $\sum_j f''(B[t_j]) \cdot (t_{j+1} - t_j)$ converges to $\int_0^t f''(B_u)\,\mathrm{d}u$ almost surely. As a result,

$$\sum_j f''(B[t_j]) \cdot (B[t_{j+1}] - B[t_j])^2 \xrightarrow{P} \int_0^t f''(B_u)\,\mathrm{d}u.$$

(b) Now consider the case where $f''$ is unbounded, then for any $K$, we wish to show that

$$\sum_j f''(B[t_j]) \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right] \xrightarrow{P} 0.$$

We apply the truncation technique so that

$$\begin{aligned}
&f''(B[t_j]) \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right] \\
=&f''(B[t_j])1\{|f''(B[t_j])| \le K\} \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right] \\
&\quad + f''(B[t_j])1\{|f''(B[t_j])| > K\} \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right]
\end{aligned}$$

By applying the result in the part (a), as $\|\Pi\| \to 0$,

$$\sum_j f''(B[t_j])1\{|f''(B[t_j])| \le K\} \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right] \to 0.$$

The remainder term can be upper bounded as the following:

$$\begin{aligned}
&\left|\sum_j f''(B[t_j])1\{|f''(B[t_j])| > K\} \cdot \left[(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right]\right| \\
\le&\max_j \left|f''(B[t_j])1\{|f''(B[t_j])| > K\}\right| \cdot \sum_j \left|(B[t_{j+1}] - B[t_j])^2 - (t_{j+1} - t_j)\right| \\
\le&\max_{0\le u\le t} \left|f''(B[t_j])1\{|f''(B[t_j])| > K\}\right| \cdot \left[\sum_j (B[t_{j+1}] - B[t_j])^2 + t\right]
\end{aligned}$$

Applying Theorem 8.1 gives $\sum_j (B[t_{j+1}] - B[t_j])^2 \to t$ in probability. By the continuity of $f''$ and $B_t$ on the interval $[0,t]$,

$$\max_{0\leq u\leq t}\left|f''(B[t_j])1\{|f''(B[t_j])|>K\}\right|\to 0,\qquad \text{as } K\to\infty.$$

Hence,

$$\sum_j f''(B[t_j])1\{|f''(B[t_j])|\leq K\}\left[(B[t_{j+1}]-B[t_j])^2-(t_{j+1}-t_j)\right]\xrightarrow{P}0,\quad \text{as } \|\Pi\|\to 0.$$

As a result,

$$\sum_j f''(B[t_j])\cdot(B[t_{j+1}]-B[t_j])^2\xrightarrow{P}\sum_j f''(B[t_j])\cdot(t_{j+1}-t_j)\xrightarrow{P}\int_0^t f''(B_u)\,\mathrm{d}u.$$

- Now consider the case where $\theta_{t_j}\neq B_{t_j}$. It remains to show that $\sum_j f''(\theta[t_j])(B[t_{j+1}]-B[t_j])^2$ converges to $\int_0^t f''(B_u)\,\mathrm{d}u$ in probability:

$$\begin{aligned}&\left|\sum_j f''(\theta[t_j])(B[t_{j+1}]-B[t_j])^2-\sum_j f''(B[t_j])(B[t_{j+1}]-B[t_j])^2\right|\\ =&\left|\sum_j[f''(\theta[t_j])-f''(B[t_j])](B[t_{j+1}]-B[t_j])^2\right|\\ \leq&\max_j\left|f''(\theta[t_j])-f''(B[t_j])\right|\cdot\sum_j(B[t_{j+1}]-B[t_j])^2\end{aligned}$$

By Theorem 8.1 we can see that $\sum_j(B[t_{j+1}]-B[t_j])^2\xrightarrow{P}t$. For almost all $\omega\in\Omega$, by the continuity and thus uniform continuity of $f''(B_t)$,

$$\max_j\left|f''(\theta[t_j])-f''(B[t_j])\right|\to 0.$$

Hence,

$$\sum_j f''(\theta[t_j])(B[t_{j+1}]-B[t_j])^2\xrightarrow{P}\sum_j f''(B[t_j])(B[t_{j+1}]-B[t_j])^2\xrightarrow{P}\int_0^t f''(B_u)\,\mathrm{d}u.$$

■

**Definition 12.1** [Ito Processes] An Ito process is a stochastic process $X_\bullet$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ of the form

$$X_t = X_0 + \int_0^t \mu[u]\,\mathrm{d}u + \int_0^t \sigma[u]\,\mathrm{d}B[u], \quad 0 \le t \le T, \tag{12.4}$$

where $X_0$ is a $\mathcal{F}_0$-measurable random variable; $\{\mu_t\}_{t\ge 0}$ is $\{\mathcal{F}_t\}$-adapted so that

$$\mathbb{P}\left[\int_0^t |\mu_u|\,\mathrm{d}u < \infty, \quad \text{for all } t \ge 0\right] = 1;$$

and $\{\sigma_t\}_{t\ge 0}$ is $\{\mathcal{F}_t\}$-adapted so that

$$\mathbb{P}\left[\int_0^t \sigma_u^2\,\mathrm{d}u < \infty, \quad \text{for all } t \ge 0\right] = 1.$$

Sometimes we also write (12.4) as the stochastic differential form

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t.$$

Now we are ready to show the main result of this lecture:

**Theorem 12.2** — **Ito Formula for 1-dimension.** Let $\{X_t\}_{t\ge 0}$ be an Ito process with stochastic differential

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t.$$

Let $f \in \mathcal{C}^2(\mathbb{R})$, then $f(X_t)$ is again an Ito process, with the following equality holds almost surely:

$$f(X_t) = f(X_0) + \int_0^t f'(X_u)\,\mathrm{d}X_u + \frac{1}{2}\int_0^t f''(X_u)(\mathrm{d}X_u)^2.$$

In particular, $(\mathrm{d}X_t)^2$ can be computed according to the rules

$$(\mathrm{d}t)^2 = (\mathrm{d}t)(\mathrm{d}B_t) = (\mathrm{d}B_t)(\mathrm{d}t) = 0, \quad (\mathrm{d}B_t)^2 = \mathrm{d}t.$$

So we further have the representation

$$f(X_t) = f(X_0) + \int_0^t f'(X_u)\mu_u \,\mathrm{d}u + \frac{1}{2}\int_0^t f''(X_u)\sigma_u^2 \,\mathrm{d}u + \int_0^t f'(X_u)\sigma_u \,\mathrm{d}B_u, \quad \forall t \ge 0$$

**R** The Ito process $\{f(X_t)\}_{t\ge 0}$ has the stochastic differential form

$$\mathrm{d}f(X_t) = \left[f'(X_t)\mu_t + \frac{1}{2}f''(X_t)\sigma_t^2\right]\mathrm{d}t + f'(X_t)\sigma_t \,\mathrm{d}B_t.$$

The proof follows the similar idea as stated in Theorem 12.1. Here we only provide a sketrch of the proof:

*Outline of Proof.* Let $\Pi = \{t_0 < t_1 < \cdots < t_n = t\}$ be a partition on $[0, t]$, then $f(X_t)$ admits the expansion

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_j \left[f(X[t_{j+1}]) - f(X[t_j])\right] \\ &= f(X_0) + \sum_j \left[f'(X[t_j])(X[t_{j+1}] - X[t_j]) + \frac{1}{2}f''(\theta[t_j])(X[t_{j+1}] - X[t_j])^2\right] \end{aligned} \tag{12.5}$$

(a) First consider the case where $\{\mu_t\}$ and $\{\sigma_t\}$ are simple processes, then

$$X[t_{j+1}] - X[t_j] = \mu[t_j](t_{j+1} - t_j) + \sigma[t_j](B[t_{j+1}] - B[t_j])$$

Substituting this into (12.5) gives

$$\begin{aligned} f(X_t) = f(X_0) &+ \sum_j f'(X[t_j])\mu[t_j](t_{j+1} - t_j) + \sum_j f'(X[t_j])\sigma[t_j](B[t_{j+1}] - B[t_j]) \\ &+ \frac{1}{2}\sum_j f''(\theta[t_j])(X[t_{j+1}] - X[t_j])^2. \end{aligned}$$

Assume that $\theta[t_j] = X[t_j]$, then we first show $\sum_j f''(X[t_j])(X[t_{j+1}] - X[t_j])^2 \xrightarrow{P} \sum_j f''(X[t_j])\sigma[t_j]^2(t_{j+1} - t_j)$, and then show that

$$\sum_j f''(X[t_j])\sigma[t_j]^2(t_{j+1} - t_j) \xrightarrow{a.s.} \frac{1}{2}\int_0^t f''(X_u)\sigma_u^2 \,\mathrm{d}u.$$

For the case where $\theta[t_j] \neq X[t_j]$, we will show that

$$\left| \sum_j f''(\theta[t_j])(X[t_{j+1}] - X[t_j])^2 - \sum_j f''(X[t_j])(X[t_{j+1}] - X[t_j])^2 \right| \xrightarrow{a.s.} 0.$$

(b) For general processes $\{\mu_t\}_{t\geq 0}$ and $\{\sigma_t\}_{t\geq 0}$, we will use the approximation of simple processes.

■

■ **Example 12.2** Let $\{X_t\}_{t\geq 0}$ be the drifted Brownian motion:

$$X_t = \mu t + B_t, \quad \mu \in \mathbb{R}.$$

Then we apply the Ito's formula to compute the stochastic differential form of $X^2$:

- Take $f(x) = x^2$ and $f'(x) = 2x, f''(x) = 2$. Hence,

$$\begin{aligned} \mathrm{d}f(X_t) &= \mathrm{d}X_t^2 \\ &= (2\mu X_t + 1)\,\mathrm{d}t + 2X_t\,\mathrm{d}B_t. \end{aligned}$$

■

The Ito's formula can also be generalized into multiple processes:

**Theorem 12.3 — Ito's Formula.** Let $\{X_t^{(1)}\}_{t\geq 0}, \dots, \{X_t^{(d)}\}_{t\geq 0}$ be continuous semimartingales, and $f \in \mathcal{C}^2(\mathbb{R})$, then

$$\begin{aligned} f(X_t^{(1)}, \dots, X_t^{(d)}) = f(X_0^{(1)}, \dots, X_0^{(d)}) & \\ + \sum_{j=1}^d \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \dots, X_u^{(d)})\,\mathrm{d}X_u^{(j)} & \\ + \frac{1}{2}\sum_{j,k=1}^d \int_0^t \frac{\partial^2 f}{\partial X_j \partial X_k}(X_u^{(1)}, \dots, X_u^{(d)})\,\mathrm{d}\langle X^{(j)}, X^{(k)}\rangle_u & \end{aligned}$$

In particular, $\mathrm{d}\langle B_i, B_j\rangle = 1(i = j)\,\mathrm{d}t$ and $\mathrm{d}\langle B_i, t\rangle = \mathrm{d}\langle t, B_i\rangle = 0$.

Recall that in discrete case, any real-valued process is a semi-martingale, while it is not true for continuous case. We define the semi-martingale for continuous case as the following:

**Definition 12.2** [Semi-martingale] We say $X_\bullet$ is a semi-martingale if it admits the decomposition

$$X_t = A_t + M_t,$$

where $M_\bullet$ is a continuous local martingale, and $A_\bullet$ is an adapted process of finite variations:

$$|A|(t) \equiv \sup_{\delta>0, t_0=0} \sup_{t_n - t_{n-1} \geq \delta} \sum_{n=1}^{\infty} 1(t_n \leq t)|A[t_n] - A[t_{n-1}]| < \infty, \quad \forall t \geq 0.$$

R Suppose that $X^{(j)}$ admits the decomposition $X_t^{(j)} = A_t^{(j)} + M_t^{(j)}$, then

$$\begin{aligned}\sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}X_u^{(j)} &= \sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}M_u^{(j)} \\ &\quad + \sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}A_u^{(j)}\end{aligned}$$

Moreover, we can show that $\langle X^{(j)}, X^{(k)} \rangle = \langle M^{(j)}, M^{(k)} \rangle$. Then the process $\{f(X_t^{(1)}, \ldots, X_t^{(d)})\}_{t\geq 0}$ also admits the semi-martingale decomposition:

$$f(X_t^{(1)}, \ldots, X_t^{(d)}) = A_t^f + M_t^f,$$

where

$$\begin{aligned}M_t^f &= f(X_0^{(1)}, \ldots, X_0^{(d)}) + \sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}M_u^{(j)}, \\ A_t^f &= \sum_{j=1}^{d} \int_0^t \frac{\partial f}{\partial X_j}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}A_u^{(j)} \\ &\quad + \frac{1}{2} \sum_{j,k=1}^{d} \int_0^t \frac{\partial^2 f}{\partial X_j \partial X_k}(X_u^{(1)}, \ldots, X_u^{(d)}) \,\mathrm{d}\langle M^{(j)}, M^{(k)} \rangle_u\end{aligned}$$

■ **Example 12.3** We can also apply Theorem 12.3 to obtain an "integration by parts" formula. Suppose that $X_\bullet$ and $Y_\bullet$ are semi-martingales, then by direct computation with $f(X_t, Y_t) = X_t Y_t$,

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s \, \mathrm{d}Y_s + \int_0^t Y_s \, \mathrm{d}X_s + \langle X, Y \rangle_t$$

Suppose that $X_t$ and $Y_t$ admits the semi-martingale decomposition

$$X_t = M_t + A_t, \quad Y_t = N_t + W_t,$$

then $XY$ also admits the semi-martingale decomposition

$$X_t Y_t = \left[ X_0 Y_0 + \int_0^t X_s \, \mathrm{d}N_s + \int_0^t Y_s \, \mathrm{d}M_s \right] + \left[ \int_0^t X_s \, \mathrm{d}W_s + \int_0^t Y_s \, \mathrm{d}A_s + \langle M, N \rangle_t \right]$$

■

## 12.1.2. Applications of Ito's Formula

Here we presents some examples for how to use Ito's formula.

■ **Example 12.4** We aim to sovle the following stochastic differential equation with $X_t = M_t + A_t$ being a continuous semi-martingale;

$$\mathrm{d}Z_t = Z_t \, \mathrm{d}X_t, \quad Z_0 = 1. \tag{12.6}$$

This equation is called the stochastic exponential of $X$, which can be re-written as the integral form:

$$Z_t = 1 + \int_0^t Z_u \, \mathrm{d}X_u, \tag{12.7}$$

where the integration refers to the Ito's integral. We guess the solution should be $Z_t = \exp(X_t + V_t)$, with $V_t$ to be determined. Applying Ito's formula on $Z_t$ with $f(\zeta) = \exp(\zeta)$

$(\zeta = X + V)$ gives

$$\begin{aligned} Z_t &= 1 + \int_0^t Z_u \, \mathrm{d}(X_u + V_u) + \frac{1}{2} \int_0^t Z_u \, \mathrm{d}\langle X + V \rangle_u \\ &= 1 + \int_0^t Z_u \, \mathrm{d}X_u + \int_0^t Z_u \, \mathrm{d}V_u + \frac{1}{2} \int_0^t Z_u \, \mathrm{d}\langle M + V \rangle_u \end{aligned}$$

In order to satisfy (12.7), we take $V_t = -\frac{1}{2}\langle M \rangle_t$. As a result, $\exp\left(X_t - \frac{1}{2}\langle M \rangle_t\right)$ is the solution to (12.7), which is called the stochastic exponential.

**Example 12.5** [Levy's characterization of Brownian Motion] Levy states that the quadratic variational process of a continuous local martingale will characterize the Brownian motion uniquely:

**Theorem 12.4 — Levy's Theorem.** Consider the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the usual condition. Let $\{M_t\}_{t \geq 0}$ be a stochastic process on this filtered probability space, with $M_0 = 0$ almost surely. Then the process $\{M_t\}_{t \geq 0}$ is a Brownian motion if and only if:

1. $\{M_t\}_{t \geq 0}$ is a continuous local martingale with respect to $\mathbb{F}$;
2. The quadratic variation $\langle M \rangle_t = t$ almost surely.

*Proof.* In order to show the sufficiency, we construct $Z_t^{(\zeta)} = \exp\left(\zeta M_t - \frac{\zeta^2}{2} t\right)$. By Ito's formula on $Z_t^{(\zeta)}$ and $f(x) = e^x$, for $0 \leq s < t$, we have

$$\mathrm{d}Z_t^{(\zeta)} = \zeta Z_t^{(\zeta)} \, \mathrm{d}M_t.$$

It indicates that $Z_t^{(\zeta)}$ is a martingale, by the uniqueness of semi-martingale decomposition. Re-arranging the term $\mathbb{E}[Z_t^{(\zeta)} \mid \mathcal{F}_s] = Z_s^{(\zeta)}$ yields

$$\mathbb{E}[\exp(\zeta(M_t - M_s)) \mid \mathcal{F}_s] = \exp\left(\frac{\zeta^2}{2}(t - s)\right),$$

which, together with the uniqueness of characterization function, explains that $M_t$ is a Brownian motion. ∎

A third application indicates that when a continuous local martingale is applied time-change, it is a Brownian motion:

**Theorem 12.5** Consider a stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, where $\{M_t\}_{t\geq 0}$ is a continuous local martingale with initial value equal to 0 and $\langle M\rangle_\infty = \infty$. Let $\tau_t = \inf\{u : \langle M\rangle_u > t\}$. Then for $\forall t \geq 0$, $\tau_t$ is a stopping time, and $B_t \equiv M_{\tau_t}$ is a $\mathcal{F}_{\tau_t}$-Brownian motion with $M_t = B_{\langle M\rangle_t}$.

*Proof Outline.* We call the increasing sequence of stopping time $\{\tau_t\}_{t\geq 0}$ the time-change. Since $\langle M\rangle_\infty = \infty$ almost surely, we can assert that each $\tau_t$ is finite almost surely. By the continuity of $\{M_t\}_{t\geq 0}$, we have $\langle M\rangle_{\tau_t} = t$. Applying the optional samping theorem on $\{M_{u\wedge\tau_t}\}_{u\geq 0}$ gives

$$\mathbb{E}[M_{\tau_t} \mid \mathcal{F}_{\tau_u}] = M_{\tau_u},$$

which implies that $B_t$ is a $\mathcal{F}_{\tau_t}$-local martingale. Applying the optional samping theorem on $\{M^2_{u\wedge\tau_t} - \langle M\rangle_{u\wedge\tau_t}\}_{u\geq 0}$ gives

$$\mathbb{E}[M^2_{\tau_t} - \langle M\rangle_{\tau_t} \mid \mathcal{F}_{\tau_u}] = M^2_{\tau_u} - \langle M\rangle_{\tau_u}.$$

This implies that $\{B_t^2 - t\}$ is a $\mathcal{F}_{\tau_t}$-local martingale. By Levy's characterization, together with the continuity of $B_t$, we conclude that $\{B_t\}_{t\geq 0}$ is a $\mathcal{F}_{\tau_t}$-Brownian motion. ∎

**Example 12.6** Consider a probability space $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$. Let $T > 0$ and $Q$ be a probability measure on $(\Omega,\mathcal{F}_T)$ that is absolutely continuous with respect to $\mathbb{P}$. Denote $\zeta = \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_T}$. Then for any $\mathcal{F}_T$-measurable bounded random variable $X$, we have

$$\int_\Omega X(\omega)\,\mathrm{d}Q(\omega) = \int_\Omega X(\omega)\zeta(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

Or equivalently, $\mathbb{E}_{X\sim Q}[X] = \mathbb{E}_{X\sim\mathbb{P}}[\zeta X]$.

Conversely, let $T > 0$ and $\{Z_t\}_{t\geq 0}$ be a continuous martingale on $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ with $Z_0 \equiv 1$ and $Z_t(\omega) > 0, \forall (t,\omega) \in [0,T]\times\Omega$. Define a probability measure $Q$ on $(\Omega,\mathcal{F}_T,\mathbb{P})$

to be $Q(A) = \mathbb{E}[Z_T 1_A], A \in \mathcal{F}_T$. Then we have $\left.\frac{dQ}{d\mathbb{P}}\right|_{\mathcal{F}_T} = Z_T$. Because $\{Z_t\}_{t\geq 0}$ is a martingale, then for any $t \leq T$, we have

$$\left.\frac{dQ}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t.$$

**Theorem 12.6** — **Girsanov.** Let $M_\bullet$ be a continuous local martingale on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $0 \leq t \leq T$, and $Z_\bullet$ be a continuous martingale, strictly positive, and initial value equal to 1. Then the process $X_t \equiv M_t - \int_0^t Z_u \, d\langle M, Z\rangle_u$ is a continuous local martingale on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$.

*Proof Outline.* By the technique of localization, we assume that $M, Z, \frac{1}{Z}$ are all bounded. It suffices to show that $X$ is a martingale w.r.t. the probability measure $Q$, i.e., for any $0 \leq u < t \leq T$, $\mathbb{E}^{\mathbb{Q}}[1_A(X_t - X_u)] = 0, \forall A \in \mathcal{F}_u$. By definition,

$$\mathbb{E}^{\mathbb{Q}}[1_A(X_t - X_u)] = \mathbb{E}^{\mathbb{P}}[1_A(Z_t X_t - Z_u X_u)].$$

It remains to show that $\{Z_t X_t\}$ is a martingale w.r.t. the probability measure $\mathbb{P}$. Applying the integration by parts gives

$$\begin{aligned}
Z_t X_t &= Z_0 X_0 + \int_0^t Z_s \, dX_s + \int_0^t X_s \, dZ_s + \langle X, Z\rangle_t \\
&= Z_0 X_0 + \int_0^t Z_s \left( dM_s - \frac{1}{Z_s} d\langle M, Z\rangle_s \right) + \int_0^t X_s \, dZ_s + \langle X, Z\rangle_t \\
&= Z_0 X_0 + \int_0^t Z_s \, dM_s + \int_0^t X_s \, dZ_s.
\end{aligned}$$

By the uniqueness of semi-martingale decomposition, we can see that $\{Z_t X_t\}$ is a martingale. ■

**Motivation for Martingale Representation Theorem.** The last application we will discuss is the martingale representation theorem. Previously we noticed that if the stochastic process $\{f_t\}_{t\geq 0} \in \mathcal{L}^2$, i.e., is square-integrable, then the process $X_t =$

$X_0 + \int_0^t f_s \, \mathrm{d}B_s$ is always a martingale w.r.t. $\mathcal{F}_t$.[1] The martingale representation theorem states that the converse is also true: any $\mathcal{F}_t$-martingale can be represented as an Ito's integral.

**Theorem 12.7 — Martingale Representation Theorem.** Let $B_\bullet$ be a standard Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{F}^0 \equiv \{\mathcal{F}_t^0\}_{t\geq 0}$ be the natural filtration generated by $B_\bullet$, with $\mathcal{F}_\infty^0 = \sigma\left(\cup_{t\geq 0}\mathcal{F}_t^0\right)$. Let $\mathcal{F}_\infty, \mathcal{F}$ be the completion of $\mathcal{F}_\infty^0, \mathcal{F}^0$, respectively, and denote $\mathbb{F} \equiv \{\mathcal{F}_t\}_{t\geq 0}$. Then consider a square-integrable martingale $M_\bullet$ on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. There exists a stocastic process $\{f_t\}_{t\geq 0} \in \mathcal{L}^2$ so that

$$M_t = \mathbb{E}[M_0] + \int_0^t f_u \, \mathrm{d}B_u.$$

This theorem relies on an important auxiliary result:

> Let $\zeta \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, then there exists $f_\bullet \in \mathcal{L}^2$ so that
>
> $$\zeta = \mathbb{E}[\zeta] + \int_0^T f_t \, \mathrm{d}B_t.$$

*Proof for Theorem 12.7.* Assume the claim is true, then for the martingale $M_\bullet$ with $M_T \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, there exists $\{f_t\} \in \mathcal{L}^2$ so that

$$M_T = \mathbb{E}[M_T] + \int_0^T f_t \, \mathrm{d}B_t = \mathbb{E}[M_0] + \int_0^T f_t \, \mathrm{d}B_t.$$

Hence for any $t \in [0, T]$, by the martingability of $M_\bullet$,

$$M_t = \mathbb{E}[M_T \mid \mathcal{F}_t] = \mathbb{E}[M_0] + \int_0^t f_u \, \mathrm{d}B_u.$$

∎

This key auxiliary result can be shown by applying Ito's formula:

[1] This can be shown either by directly applying definition or by the uniqueness of semi-martingale decomposition.

*Proof on the Claim.* Let $T > 0$ and define the stochastic exponential martingale for any $f \in \mathcal{L}^2([0,T])$:

$$\mathcal{E}[f]_t = \exp\left\{\int_0^t f(u)\,\mathrm{d}B_u - \frac{1}{2}\int_0^t f^2(u)\,\mathrm{d}u\right\}, \quad t \in [0,T].$$

Then we can show that $\ell \equiv \mathrm{span}\{\mathcal{E}[f]_t : f \in \mathcal{L}^2([0,T])\}$ is dense in the space $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$. It remains to verify this claim when $\zeta = \mathbb{E}[f]_t$ for some $f \in \mathcal{L}^2([0,T])$. Directly applying Ito's formula gives

$$\mathrm{d}\mathcal{E}[f]_t = f(t)\mathcal{E}[f]_t\,\mathrm{d}B_t \implies \mathbb{E}[f]_t = 1 + \int_0^T f(t)\mathcal{E}[f]_t\,\mathrm{d}B_t.$$

■

# 12.2. Thursday

## 12.2.1. Introduction to SDE

In this week, we will introduce some concept of stochastic differential equations (SDEs). We will talk about how to solve some simple SDEs. We will also study some important SDEs in applications that cannot find explicit solutions, and establish the theorem of existence and uniqueness of solutions.

**Motivation.** SDEs are usually regarded as an ODE plus a stocahstic perturbation driven by Brownian motion, which is also called "noises". The following ODE characterizes the population growth in practice:

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = r\,\mathrm{d}t.$$

This simple ODE admits a deterministic solution $S_t$. The corresponding SDE is $\frac{\mathrm{d}S_t}{\mathrm{d}t} = r\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, called Black-Sholes equation, in which the increasing rate is a constant plus a random perturbation.

In this lecture, we will conside the SDE of the Markovian type, i.e., the parameter $\mu$ and $\sigma$ at time index $t$ only depends on $t, X_t$ instead of $\{X_u\}_{u<t}$:

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \tag{12.8}$$

where $B_\bullet$ is a standard Brownian motion, and $X_\bullet$ is the unknown continuous process. The differential for SDE has no meaning, while only the integration has the meaning. Therefore, the above equation refers to the follwoing equation in integral form:

$$X_t - X_0 = \int_0^t \mu(u, X_u)\,\mathrm{d}u + \int_0^t \sigma(u, X_u)\,\mathrm{d}B_u. \tag{12.9}$$

where $\mu: [0,\infty) \times \mathbb{R} \to \mathbb{R}$ and $\sigma: [0,\infty) \times \mathbb{R} \to \mathbb{R}$. A particular situation is when $\mu(t, X_t) = \mu(X_t)$ and $\sigma(t, X_t) = \sigma(X_t)$, i.e., $\mu: [0,\infty) \to \mathbb{R}$ and $\sigma: [0,\infty) \to \mathbb{R}$. We call the SDE in this case the type of **time-homogeneous Markovian** or **Ito**.

**Definition 12.3** [Solution to SDE] Given the functions $\mu, \sigma$ defined above, the solution to (12.8) is the process $(X, B)$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying

- $B_\bullet$ is a standard $\mathbb{F}$-Brownian motion;
- The equation (12.9) holds.

■

**R** In addition to the unknown process $X_\bullet$, the probability space and Brownian motion are also parts of the solution. Only the coefficients $\mu$ and $\sigma$ are given.

There are two interpretations for the uniqueness of an SDE:

**Definition 12.4** [Pathwise Uniqueness] The solution to (12.8) has pathwise uniqueness if given the initial value $X_0$ and the Brownian motion $\{B_t\}_{t\geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, there is a unique stochastic process $X_\bullet$, pathwisely, satisfying the equation. In other words, if two solutions $(X, B)$ and $(X', B')$ have the same initial value, i.e., $X_0 = X'_0$ and $B_0 = B'_0$, then $X_t = X'_t$ almost surely for any $t \geq 0$. ■

**Definition 12.5** [Uniqueness in Law] The solution to (12.8) has uniqueness in law if two solutions $X, X'$ with the same initial distribution are equivalent, i.e., $X$ and $X'$ have the same finite dimensional distribution. ■

**Definition 12.6** [Strong/Weak Solution] A solution $(X, B)$ to equation (12.8) is called a **strong solution** if $X$ is adapted to the filteration $\mathbb{F}^B$ that is generated by the Brownian motion $B_\bullet$ with completion. A solution that is not a strong solution is called a **weak solution**. ■

**R** If a strong solution exists, then the SDE always has a solution for any given probability space and Brownian motion. In other words, when a strong solution exists, *any initial value and Brownian motion corresponds to at least one solution*. If one has the pathwise uniqueness together with the existence of a strong solution, then *any initial value and Brownian motion will correspond to a unique solution.*

The following example shows an important concept in SDE: sometimes the solution exists but there is no strong solution; sometimes we have no pathwise uniqueness but only uniqueness in law.

■ **Example 12.7** [The Tanaka Equation] Consider the 1-dimensional equation

$$X_t = \int_0^t \text{sign}(X_u)\,\mathrm{d}B_u, \quad 0 \le t < \infty, \tag{12.10}$$

where $\text{sign}(X) = 1\{X \ge 0\} - 1\{X < 0\}$ denotes the sign function. It corresponds to the SDE

$$\mathrm{d}X_t = \text{sign}(X_u)\,\mathrm{d}B_t, \quad X_0 = 0. \tag{12.11}$$

We have the following conclusions on this SDE:

1. The solution to (12.11) has uniqueness in law: Take $X$ be a standard Brownian motion. For any other solution $X'$ satisfying (12.10), notice that $X'$ is a continuouis local martingale and the quadratic variation $\langle X' \rangle_t = \int_0^t \text{sign}(X'_u)^2\,\mathrm{d}u = \int_0^t \mathrm{d}u = t$. By the Levy's theorem, $X$ and $X'$ share the same distribution.
2. A weak solution exists: choose $X$ be any Brownian motion $\{W_t\}_{t\ge 0}$, and define $\tilde{B}_t$ by $\tilde{B}_t = \int_0^t \text{sign}(W_s)\,\mathrm{d}W_s$, i.e., $\mathrm{d}\tilde{B}_t = \text{sign}(X_t)\,\mathrm{d}X_t$. Then $\tilde{B}_\bullet$ is also a Brownian motion. Moreover,
$$\mathrm{d}X_t = \text{sign}(X_t)\,\mathrm{d}\tilde{B}_t.$$
Hence, the pair $(W, \tilde{B})$ is a weak solution.
3. Pathwise uniqueness does not hold: when $(X, B)$ is a solution, then $(-X, B)$ is also a solution.
4. There is no strong solution.

■

Next we will discuss how to solve some simple SDEs:

■ **Example 12.8** [Ornstein-Uhlenbeck Process] Consider solving the SDE

$$\mathrm{d}X_t = -\alpha X_t \,\mathrm{d}t + \sigma \,\mathrm{d}B_t, \quad X_0 = 1,$$

where $\alpha, \sigma$ are some non-negative constants.

- Take $Y_t = e^{\alpha t} X_t$, where $e^{\alpha t}$ can be viewed as the **integrating factor**. Applying Ito's formula on $Y_t$ with $f(x,t) = e^{\alpha t} x$ gives

$$\begin{aligned}
\mathrm{d}Y_t &= \alpha e^{\alpha t} X_t \,\mathrm{d}t + e^{\alpha t} \,\mathrm{d}X_t \\
&= \alpha e^{\alpha t} X_t \,\mathrm{d}t + e^{\alpha t}(-\alpha X_t) \,\mathrm{d}t + e^{\alpha t} \sigma \,\mathrm{d}B_t \\
&= \sigma e^{\alpha t} \,\mathrm{d}B_t
\end{aligned}$$

It follows that

$$Y_t = Y_0 + \sigma \int_0^t e^{\alpha u} \,\mathrm{d}B_u \implies X_t = e^{-\alpha t} Y_t = e^{-\alpha t} + \sigma \int_0^t e^{\alpha(u-t)} \,\mathrm{d}B_u.$$

■

■ **Example 12.9** [Geometric Brownian Motion] Consider the SDE

$$\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}B_t, \quad X_0 = 1,$$

where $\mu, \sigma$ are constants. We claim that there is a unique strong solution $X_t = e^{(\mu - 1/2\sigma^2)t + \sigma B_t}$.

- To check it is indeed a solution, apply the Ito's formula on $X_t$ with $f(t,B) =$

$e^{(\mu-1/2\sigma^2)t+\sigma B}$ gives

$$\begin{aligned}
\mathrm{d}X_t &= \left(\mu - \frac{1}{2}\sigma^2\right) e^{(\mu-1/2\sigma^2)t+\sigma B_t}\,\mathrm{d}t + \sigma e^{(\mu-1/2\sigma^2)t+\sigma B_t}\,\mathrm{d}B_t \\
&\quad + \frac{1}{2}\sigma^2 e^{(\mu-1/2\sigma^2)t+\sigma B_t}\,\mathrm{d}t \\
&= \mu e^{(\mu-1/2\sigma^2)t+\sigma B_t}\,\mathrm{d}t + \sigma e^{(\mu-1/2\sigma^2)t+\sigma B_t}\,\mathrm{d}B_t \\
&= \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t.
\end{aligned}$$

■

■ **Example 12.10** Consider the simple SDE

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \mathrm{d}B_t,$$

where $b(t,x) : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ is a bounded Borel measurable function.

- We can apply the change of probability measure trick to solve this SDE. Let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion on a filtered probabiltiy space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Define a new probability measure $Q$ with $\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} = Z_t, t \geq 0$, where $Z_t$ is the stochastic exponential of $N_t \equiv \int_0^t b(u, W_u)\,\mathrm{d}W_u$:

$$Z_t = \exp\left\{\int_0^t b(u, W_u)\,\mathrm{d}W_u - \frac{1}{2}\int_0^t b^2(u, W_u)\,\mathrm{d}u\right\}.$$

Since $N_\bullet$ is a martingale w.r.t. $\mathbb{P}$, we can check that $Z_\bullet$ is also a martingale. Applying Ito's formula on $Z_t$ with $f(\zeta) = e^\zeta$ and $\zeta = N_t - \frac{1}{2}\langle N\rangle_t$ gives

$$\begin{aligned}
\mathrm{d}Z_t &= \mathrm{d}\exp\left(N_t - \frac{1}{2}\langle N\rangle_t\right) \\
&= e^{N_t - \frac{1}{2}\langle N\rangle_t}\,\mathrm{d}N_t - \frac{1}{2}e^{N_t - \frac{1}{2}\langle N\rangle_t}\,\mathrm{d}\langle N\rangle_t + \frac{1}{2}e^{N_t - \frac{1}{2}\langle N\rangle_t}\,\mathrm{d}\langle N\rangle_t = Z_t\,\mathrm{d}N_t.
\end{aligned}$$

Hence, the process $Z_t$ admits the integral equation

$$Z_t = 1 + \int_0^t Z_u\,\mathrm{d}N_u.$$

- We recover the original solution by Girsanov theorem. Define $\tilde{B}_t = W_t - W_0 - \int_0^t \frac{1}{Z_u} \mathrm{d}\langle W, Z\rangle_u$. In particular,

$$\langle W, Z\rangle_u = \langle \int_0^t \mathrm{d}W_u, \int_0^t Z_u \,\mathrm{d}N_u \rangle = \int_0^t Z_u \,\mathrm{d}\langle W, N\rangle_u,$$

which implies that $\int_0^t \frac{1}{Z_u} \mathrm{d}\langle W, Z\rangle_u = \langle W, N\rangle_t$. Therefore, $\tilde{B}_t = W_t - W_0 - \langle W, N\rangle_t$ is a martingale w.r.t. the probability measure $Q$. Moreover, $\langle \tilde{B}\rangle_t = \langle W\rangle_t = t$. By Levy's theorem, $\tilde{B}_\bullet$ is a standard Brownian motion w.r.t. $Q$. Furthermore,

$$\langle W, N\rangle_t = \langle \int_0^t \mathrm{d}W_u, \int_0^t b(u, W_u) \,\mathrm{d}W_u \rangle = \int_0^t b(u, W_u) \,\mathrm{d}u.$$

Substituting this form into $\tilde{B}_t$ yields

$$W_t = W_0 + \int_0^t b(u, W_u) \,\mathrm{d}u + \tilde{B}_t = \int_0^t b(u, W_u) \,\mathrm{d}u + \tilde{B}_t.$$

As a result, $(W, \tilde{B})$ is a solution on the probability space $(\Omega, \mathcal{F}, Q)$. This solution is a weak solution.

■

# Chapter 13

# Week13

## 13.1. Thursday

### 13.1.1. Fundamental Theorems in SDE

The solution to an SDE could be either a strong solution or a weak solution. If we concern about the Brownian motion to be **pre-determined**, then we need to consider the strong solution; otherwise the weak solution is enough, if we only concern the distribution of a process or the construction of a process. Only when we need to construct different solutions w.r.t. the same Brownian motion, the strong solution is needed. The first theorem indicates that the existence of solution, together with the pathwise uniqueness, implies the existence and uniqueness of a strong solution.

**Theorem 13.1** — **Existence and Uniqueness for SDEs.** Suppose that the SDE in (12.8) has pathwise uniqueness, then

1. It also has uniqueness in law;
2. The existence of solution implies the existence of strong solution, i.e., there exists a functional $F : \mathbb{R} \times \mathcal{C}[0,\infty) \to \mathcal{C}[0,\infty)$ so that $x = F(x_0, B)$.

The pathwise uniqueness is difficult to verify. The following theorem gives a different way to check the existence and uniqueness:

**Theorem 13.2** Suppose that the SDE in (12.8) satisfies

1. Coefficients $\mu, \sigma$ are Lipschitz in $x$ uniformly in $t \in [0, T]$, i.e., there exists

constant $L$ so that

$$|\mu(t,x)-\mu(t,y)| \le L|x-y|, \quad |\sigma(t,x)-\sigma(t,x)| \le L|x-y|,\ \forall x,y,t \in [0,T].$$

2. Coefficients $\mu, \sigma$ satisfy the linear growth condition in $x$ uniformly in $t$, i.e., there exists a constant $C$ so that

$$|\mu(t,x)| \le C(1+|x|), |\sigma(t,x)| \le C(1+|x|), \quad \forall x,t \in [0,T].$$

3. Let $Z$ be a random variable on $(\Omega, \mathcal{F}_0, \mathbb{P})$, independent of the Brownian motion $\{B_t\}_{t\ge0}$, satisfying $\mathbb{E}[Z^2] < \infty$.

Then the SDE with the initial condition $X_0 = Z$ admits a strong solution $\{X_t\}_{t\ge0}$ adapted to the filtration $\{\mathcal{F}_t\}_{t\ge0}$ and $X \in \mathcal{L}^2$. Furthermore, the solution has pathwise uniqueness.

*Proof.* The uniqueness is by Ito's isometry and Lipschitz condition. The existence of strong solution can be shown by Picard iteration method used in ODE. ■

We can use Ito's formula to solve the time-homogeneous SDE:

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t. \tag{13.1}$$

Suppose that $\mu, \sigma \in \mathcal{C}^\infty(\mathbb{R})$ are smooth functions with bounded derivative. Let $X_\bullet$ be the unique strong solution. Take $f \in \mathcal{C}^2(\mathbb{R})$, and by Ito's formula,

$$f(X_t) - f(X_0) = \int_0^t \frac{\partial f}{\partial X}(X_u)\,\mathrm{d}X_u + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial X^2}(X_u)\,\mathrm{d}\langle X\rangle_u.$$

The SDE in (13.1) can also be written as the integral form:

$$X_t = X_0 + \int_0^t \mu(X_u)\,\mathrm{d}u + \int_0^t \sigma(X_u)\,\mathrm{d}B_u \implies \langle X\rangle_t = \int_0^t \sigma^2(X_u)\,\mathrm{d}u.$$

It follows that

$$f(X_t) - f(X_0) = \int_0^t \left[ \frac{\partial f}{\partial X}(X_u)\mu(X_u) + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(X_u)\sigma^2(X_u) \right] \mathrm{d}u + \int_0^t \frac{\partial f}{\partial X}(X_u)\sigma(X_u)\,\mathrm{d}B_u.$$

**Theorem 13.3** Let $\mu, \sigma$ be bounded Borel-measurable functions, and assume there exists a constant $\lambda > 0$ so that $\frac{1}{\lambda} < \sigma(\cdot) < \lambda$. Define the operator

$$\mathcal{L} = \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \mu(x)\frac{\mathrm{d}}{\mathrm{d}x}.$$

If $X_\bullet$ is a continuous process on $(\Omega, \mathcal{F}, \mathbb{P})$ such taht for any $f \in \mathcal{C}^2(\mathbb{R})$, the process

$$W_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{L}f)(X_s)\,\mathrm{d}s$$

is a continuous local martingale, then $X_\bullet$ solves the SDE (13.1) on the space $(\Omega, \mathcal{F}, \mathbb{P})$.

*Proof.* In order to show $X_\bullet$ is a weak solution on $(\Omega, \mathcal{F}, \mathbb{P})$, it suffices to construct a Brownian motion $B_\bullet$ so that

$$X_t = X_0 + \int_0^t \sigma(X_u)\,\mathrm{d}B_u + \int_0^t \mu(X_u)\,\mathrm{d}u.$$

We set $f(x) = x$ and $B_t = \int_0^t \frac{1}{\sigma(X_u)}\,\mathrm{d}W_u^f$. Then we show it is a Brownian motion by Levy's characterization. ∎