Contents

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:

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

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, α 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 τ^* 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 ρ, 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:

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

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 [σ-Algebra] A set $\mathcal{F}$ containing subsets of Ω is called a σ-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 σ-additive, i.e., when $A_i \in \mathcal{F}, i \geq 1$ and $A_i \cap A_j = \emptyset, \forall i \neq j$,

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

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

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

1. Ω denotes the **sample space**, and a point $\omega \in \Omega$ is called a sample point;
2. $\mathcal{F}$ is a σ-algebra of Ω, which is a collection of subsets in Ω. 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 σ-Algebra] Consider a topological space Ω, with $\mathcal{U}$ being the topology of Ω. The **Borel σ-Algebra** $\mathcal{B}(\Omega)$ on Ω is defined to be the *minimal* σ-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 σ-Algebra] Suppose X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the σ-algebra generated by X, say $\mathcal{H}_X$ is defined to be the **minimal σ-algebra** on Ω 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 σ-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 σ-algebra;

1.2.1. More on Probability Theory

Definition 1.9 [Distribution] A probability measure μ_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 μ_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 σ-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 σ-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.,

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

■

■ **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

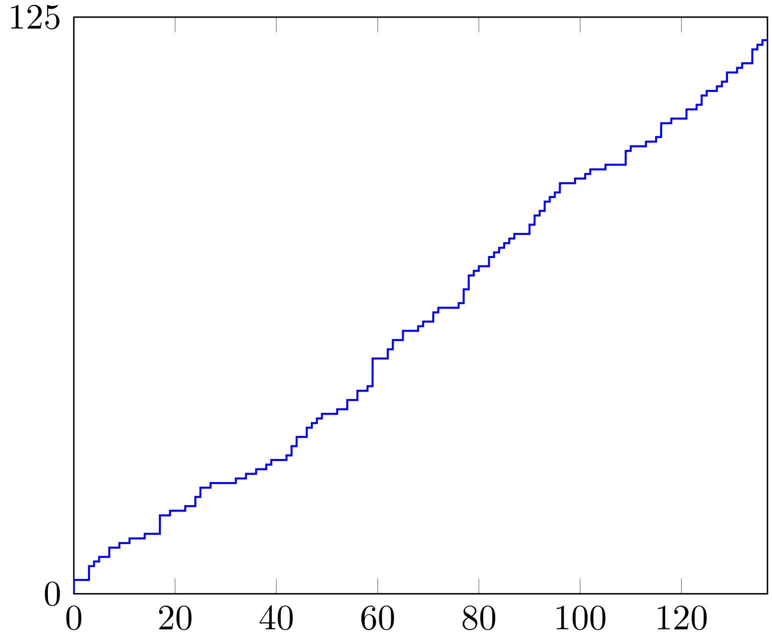

Figure 1.1: One simulation of $\{X_t\}$ with intensity $\lambda = 1.2$ and 500 samples

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 σ-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 ω, $t \mapsto X_t(\omega)$ 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 ω, $t \mapsto X_t(\omega)$ 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 σ-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 σ-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 σ-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 σ-algebra of $\mathcal{G}$. Then
$$\mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\Big] = \mathbb{E}[X \mid \mathcal{H}], \quad \text{a.s.}$$
7. (Conditional Independence) Suppose that $\mathcal{H}$ is a σ-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 $A \in \sigma(\mathcal{G}, \mathcal{H})$,

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

■

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:

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

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:

$$\limsup_{n \to \infty} E_n \triangleq \{\omega : \;\; \omega \in E_n \text{ for infinitely many } n\}$$

$$\liminf_{n \to \infty} E_n \triangleq \{\omega : \;\; \omega \in E_n \text{ for all large enough } n\}$$

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,

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

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, Σ, μ) 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, Σ, μ) 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

$\blacksquare$

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

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

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,

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

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 $\mathbb{P}(\{|X| > K\}) \leq \frac{1}{K}\mathbb{E}[|X|]$, since the following inequality holds:

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

- 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** $K \geq 0$ 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 $X \in \mathcal{C}$,

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

■

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:

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

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

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

■

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 σ-algebras such that $\mathcal{G}_\alpha \subseteq \mathcal{F}$. Then the collection of random variables

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

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:

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

  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 $\varepsilon > 0$,

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

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

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

- 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:

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

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 $n > N_0$,

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

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:

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

■

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 σ-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 σ-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 σ-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

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

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 σ-algebra] Define the stopping time σ-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 σ-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

σ-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 ω

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 ϕ 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 σ-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,

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

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 σ-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,

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

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:

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

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

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

$$\mathbb{E}[(X_n^*)^p] = \int_{[0,\infty)} \mathbb{P}(X_n^* \geq \lambda)\,\mathrm{d}\lambda^p$$

$$\leq \int_{[0,\infty)} \frac{1}{\lambda}\mathbb{E}\left[|X_n| 1\left\{\sup_{k\leq n}|X_k| \geq \lambda\right\}\right]\mathrm{d}\lambda^p \tag{4.3b}$$

$$= \int_0^\infty \frac{1}{\lambda}\int_\Omega X_n(\omega) 1\left\{X_n^*(\omega) \geq \lambda\right\}\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}\lambda^p \tag{4.3c}$$

$$= \int_\Omega X_n(\omega)\int_0^\infty \frac{1}{\lambda} 1\left\{X_n^*(\omega) \geq \lambda\right\}\mathrm{d}\lambda^p\,\mathrm{d}\mathbb{P}(\omega) \tag{4.3d}$$

$$= \int_\Omega X_n(\omega)\int_0^{X_n^*(\omega)} p\lambda^{p-2}\,\mathrm{d}\lambda\,\mathrm{d}\mathbb{P}(\omega)$$

$$= \int_\Omega X_n(\omega)\frac{p}{p-1}\left[X_n^*(\omega)\right]^{p-1}\mathrm{d}\mathbb{P}(\omega)$$

$$\leq \frac{p}{p-1}\left(\mathbb{E}[(X_n)^p]\right)^{1/p}\left(\mathbb{E}[(X_n^*)^{(p-1)q}]\right)^{1/q}, \quad \text{with } 1/q = 1 - 1/p \tag{4.3e}$$

$$= \frac{p}{p-1}\left(\mathbb{E}[(X_n)^p]\right)^{1/p}\left(\mathbb{E}[(X_n^*)^p]\right)^{(p-1)/p} \tag{4.3f}$$

where (4.3b) is by the Doob's maximal inequality; (4.3c) is by the assumption that $X_n \geq 0$; (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:

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

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

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

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

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

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 $T_k, k \geq 0$.

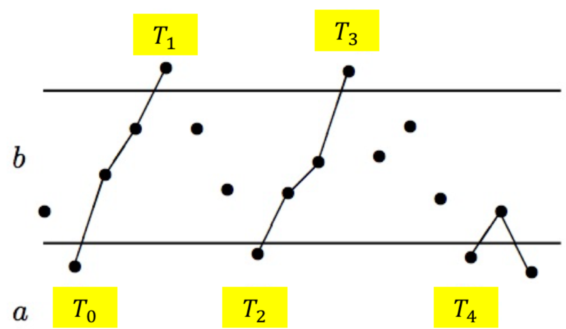

Figure 5.1: Upcrossings of $[a,b]$

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

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

■

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 $X\triangleq\{X_k\}_{k\ge0}$ up to time n. We can check that $U_a^b[X;n]$ is also a stopping time:

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

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

Theorem 5.1 — Doob's Upcrossing Theorem.

1. Suppose that $\{X_n\}_{n\ge0}$ is a super-martingale, then for any $n\ge1,k\ge0$,

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

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

2. Suppose that $\{X_n\}_{n\ge0}$ is a sub-martingale, then for any $n\ge1,k\ge0$,

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

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

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.

$\blacksquare$

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_∞. 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_∞ 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 $A \in \mathcal{F}_n$:

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

  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 $\{X_t\}_{t\ge 0}$ be an adapted process on filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, \mathbb{P})$. A stochastic process $\{X_t\}_{t\ge 0}$ 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 $0 \le s \le t$.

If in the last definition, "$=$" is replaced by "$\le$" or "$\ge$", then $\{X_t\}_{t\ge 0}$ 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 $\{T \le t\} \in \mathcal{F}_t$ for all $t \ge 0$.
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 $n \ge 1$, define the step-function mapping

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

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

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

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

  Since T is an optional time,

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

- 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 σ-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 σ-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 σ-algebra for T, containing the information available up to time T.

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

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

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

Theorem 5.3 1. Let $\{X_t\}_{t\ge 0}$ be a martingale and the filteration $\{\mathcal{F}_t\}_{t\ge 0}$ satisfies the usual condition. Then there exists a version of $\{X_t\}_{t\ge 0}$, which is right-continuous with left limits, denoted as $\{\tilde{X}_t\}_{t\ge 0}$. Then $\{\tilde{X}_t\}_{t\ge 0}$ is a right-continuous martingales w.r.t. $\{\mathcal{F}_t\}_{t\ge 0}$.

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

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

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

3. (Convergence Theorem) Let $\{X_t\}_{t\ge 0}$ be a right-continuous super-martingale, which is L^1-bounded. Then there exists a random variable X_∞ 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 $\{X_t\}_{t\ge 0}$ be a right-continuous martingale with the last element X_∞, i.e., $X_\infty \in \mathcal{L}^1$ and $\mathbb{E}[X_\infty \mid \mathcal{F}_t] = X_t$ a.s. for any $t \ge 0$. Let $S \le T$ 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_∞:

Proposition 5.2 A last element X_∞ exists if and only if $\{X_t\}_{t\ge 0}$ 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_∞ 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_∞. 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 (Draft Notes)

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 $\{T < \infty\}, \forall n \ge 1$, where $T \triangleq \inf_n T_n$. Then $\mathcal{F}_{T+} = \cap_{n=1}^{\infty} \mathcal{F}_{T_n}$, where

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

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

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

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

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

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

  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, $\{T_n \le t\} \in \mathcal{F}_{T_n}$.

Therefore,

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

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-σ-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) $\mathbb{E}[X_n \mid \mathcal{F}_{n+1}] \ge X_{n+1}$, 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:

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

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 δ such that for any $A \in \mathcal{F}_1, \mathbb{P}(A) < \delta$,

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

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

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

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, $\int_{\{|X_n|>\lambda\}} |X_1| \, \mathrm{d}\mathbb{P} < \varepsilon$ 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) $|X_{t\wedge T_n}| \le \sup_n |X_{t\wedge T_n}|$, 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 $\{X_t\}_{t\ge 0}$ is a martingale, i.e., for any $A \in \mathcal{F}_s, 0 \le s \le t$, 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 $B = \{B_t\}_{t\ge 0}$ 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 $0 \le t_1 < \cdots < t_k < \infty$ and $x_1, x_2, \dots, x_{k-1} \in \mathbb{R}$,

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

3. (Normal Distribution)For each $0 \le s < t$, $B_t - B_s$ follows normal distribution with mean 0 and variance $\sigma^2(t-s)$, where $\sigma > 0$.
4. Almost all the sample paths of $\{B_t\}_{t\ge 0}$ 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 σ-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, $\{B_t\}_{t\geq 0}$ 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.

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

1. Joint distribution: Fix $0 \leq t_1 < t_2 < \cdots < t_k$. Given $x_1, x_2, \ldots, x_k \in \mathbb{R}$, the joint density of $(B_{t_1}, B_{t_2}, \ldots, B_{t_k})$ in $(x_1, x_2, \ldots, x_k)$ is equal to the joint density of $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}})$ in $(x_1, x_2 - x_1, \ldots, x_k - x_{k-1})$, 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 $B_t^s = B_{t+s} - B_s, t \geq 0$. Then $\{B_t^s\}_{t\geq 0}$ is a Brownian motion.
3. Scaling:
   - For each $c \neq 0$, $\{cB_t\}_{t\geq 0}$ 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, $\sup_{n\le s\le n+1} |B_s - B_n|$ has the same distribution as $\sup_{0\le s\le 1} |B_s|$, then

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

We also claim that $\mathbb{E}[Z_0] < \infty$, which implies

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

Applying the Borel-Cantelli Lemma gives the desired result.

■