Contents

Acknowledgments vii

Notations ix

1 Week1 1

1.1 Tuesday 1

1.1.1 Difference between ODE and SDE 1

1.1.2 Applications of SDE 2

1.1.3 Reviewing for Probability Space 4

1.2 Thursday 7

1.2.1 More on Probability Theory 7

1.2.2 Stochastic Process 10

2 Week2 13

2.1 Tuesday 13

2.1.1 More on Stochastic Process 13

2.1.2 Conditional Expectation 15

2.1.3 Tips about Probability Theory 18

2.1.4 Reviewing on Real Analysis 19

2.2 Thursday 22

2.2.1 Uniform Integrability 22

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:

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

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

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

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

which is the famous Black-Sholes-Merton Formula.

1.1.3. Reviewing for Probability Space

Firstly, we review some basic concepts in real analysis.

Definition 1.1 [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 X. ∎

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.,

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

■

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

[a] The corresponding matlab code can be found in

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

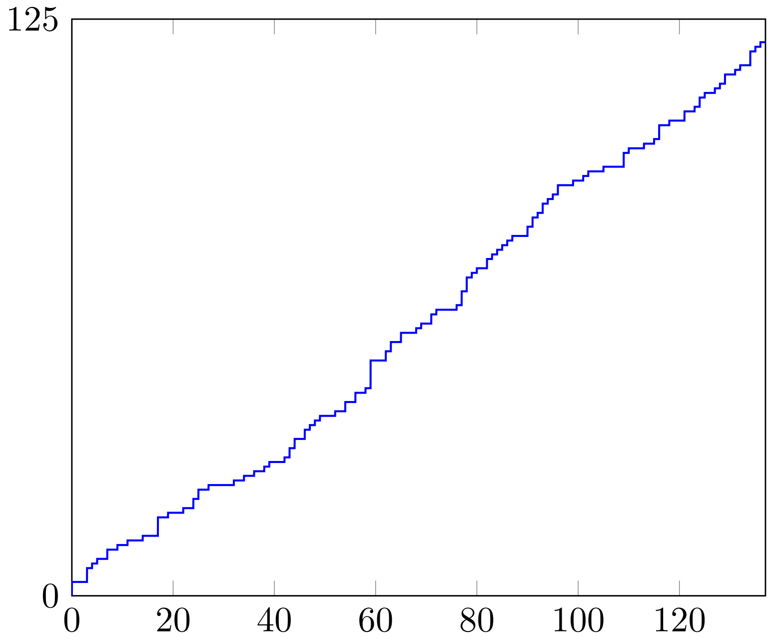

Figure 1.1: One simulation of $\{X_t\}$ with intensity $\lambda = 1.2$ and 500 samples

Chapter 2

Week2

2.1. Tuesday

2.1.1. More on Stochastic Process

For simplicity of notation, we write

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

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

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

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

R The measure $\mu_{t_1,t_2,\ldots,t_k}$ is the **finite-dimensional distribution**. In particular, $\mu_{t_1,t_2,\ldots,t_k}$ 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, if $\tilde{Y}$ is another random variable satisfying 1) and 2) in Definition 2.4, then

$$\mathbb{E}[X \mid \mathcal{G}] = \tilde{Y}, \quad \text{a.s.}$$

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 $a_1, a_2 \in \mathbb{R}$,

$$\mathbb{E}[a_1 X_1 + a_2 X_2 \mid \mathcal{G}] = a_1 \mathbb{E}[X_1 \mid \mathcal{G}] + a_2 \mathbb{E}[X_2 \mid \mathcal{G}].$$

   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

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

   a.s.

7. (Conditional Independence) Suppose that $\mathcal{H}$ is a σ-algebra independent of $\sigma(\sigma(X), \mathcal{G})$, then

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

   In particular, $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$ if $\mathcal{H}$ is independent of X.

Proof.

1. It relies on the definition on $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]]$:

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \int_{\Omega} \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P}.$$

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. 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.

6. Notice that the LHS is $\mathcal{H}$-measurable, and for any $A \in \mathcal{H}$, similar as in property (a), argue that

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]1_A\right] = \mathbb{E}[X1_A].$$

Moreover, $\mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]1_A] = \mathbb{E}[X1_A]$, which implies

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

By the uniqueness of conditional expectation, the desired result holds.

7. It suffices to show that for any $A \in \sigma(\mathcal{G}, \mathcal{H})$,

$$\mathbb{E}[X1_A] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]1_A\right].$$

■

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:

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

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

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

$$\mathbb{P}(\limsup_{n\to\infty} E_n) = 0.$$

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,

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

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

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

■

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:

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

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.,

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

   for almost all $x \in S$, with $\int_S |g| \, \mathrm{d}\mu < \infty$.
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., $\lim_{n\to\infty} \int_S |f_n - f| \, \mathrm{d}\mu = 0$, which implies that

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

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:

$$\limsup \int |f_n - f| \, \mathrm{d}m \le \int \limsup |f_n - f| \, \mathrm{d}m = 0.$$

- 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

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

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

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

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,

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

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

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

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 (2.2), 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

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

■

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.,

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

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 $X \in \mathcal{C}$:

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

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 2.1 such that

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

■