Lecture 1: Introduction

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
March 27th, 2024

1 Introduction

The notion of being measurable or belonging to a measure space hinges on set theory. Some set examples are $\{0\}, \{''cat'', “fox”, “dog”\}, \mathbb{R}^p$. Some typical set operations are union, intersection, complement, belong to, and subset.

2 σ-algebra/field

Definition 2.1 (σ-algebra/field)**.** Let Ω be a set. Then a σ-algebra $\mathcal{F}$ on Ω is a class/collection of subsets of Ω such that

1. $\emptyset \in \mathcal{F}$,
2. (closed to complement) $A \in \mathcal{F} \implies A^{\mathsf{c}} \in \mathcal{F}$,
3. (closed to countable union) Given a countable collection of elements in $\mathcal{F}$, i.e., $A_1, A_2, \dots \in \mathcal{F}$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Remark 2.2.

1. A σ-algebra is also closed under countable intersection.
2. intersection of any arbitrary collection of σ- algebras on a set is a σ-algebra.

If we replace the countability condition with finiteness, then a σ-algebra becomes an *algebra*.

Definition 2.3 (Algebra/field)**.** Let Ω be a set. Then an algebra $\mathcal{F}$ on Ω is a class/collection of subsets of Ω such that

1. $\emptyset \in \mathcal{F}$,
2. (closed to complement) $A \in \mathcal{F} \implies A^{\mathsf{c}} \in \mathcal{F}$,
3. (closed to finite union) Given a finite collection of elements in $\mathcal{F}$, i.e., $A_1, A_2, \dots, A_m \in \mathcal{F}$, $\bigcup_{n=1}^{m} A_n \in \mathcal{F}$.

We can generate a σ-algebra from a set as follows:

Definition 2.4 (σ-algebra generated by $\mathcal{A}$)**.** Let Ω be a set and $\mathcal{A}$ be a collection of subsets of Ω. The *smallest* σ-algebra containing $\mathcal{A}$ is called the σ-algebra generated by $\mathcal{A}$, and is denoted by $\sigma(\mathcal{A})$.

An important class of σ-algebra is the *Borel σ-algebra*.

Definition 2.5 (Borel σ-algebra)**.** Let Ω be a set endowed with a topology. The Borel σ-algebra on Ω, $\mathcal{B}(\Omega)$, is the σ-algebra generated by the collection of open subsets of Ω.

Remark 2.6.

1. $\mathcal{B}(\mathbb{R})$ is also generated by the set of all open intervals, half-open intervals, or closed intervals.
2. $\mathcal{B}(\mathbb{R})$ is also generated by the set of the intervals of the form (x, ∞) or $(-\infty, x)$.

3 Measurable/measure space

Definition 3.1 (Measurable space)**.** A pair $(\Omega, \mathcal{F})$ is said to be a *measurable space* if $\mathcal{F}$ is a σ-algebra.

Definition 3.2 (Measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A *measure* μ on this space is a set function such that

1. $\mu : \mathcal{F} \to [0, \infty]$,
2. $\mu(\emptyset) = 0$,
3. (Countable additivity) If a countable collection of *disjoint* elements in $\mathcal{F}$, $A_1, A_2, \cdots \in \mathcal{F}$, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

Definition 3.3 (Measure space)**.** The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

Remark 3.4. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

1. (Monotonicity) $\forall\, A, B \in \mathcal{F}$ such that $A \subseteq B$, we have $\mu(A) \leq \mu(B)$.
2. (Countable subadditivity) $\forall\, A_1, A_2, \cdots \in \mathcal{F}$, we have $\mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Now, think about how you would define a measure on $\mathbb{R}^3$. It would require a measurable space (e.g., $(\mathbb{R}^3, \mathcal{B}(R^3))$) and *Lebesgue measure* which we will cover later in the course.

Why do we need measurable spaces to define measures? Can't we just define measures on 2^{Ω}? By Banach-Tarski paradox, if we define a measure μ on $\mathbb{R}^3$, then we can show that $\mu(\cdot)$ is either 0 or ∞ or even undefined! See the figure 1 for more details..

References

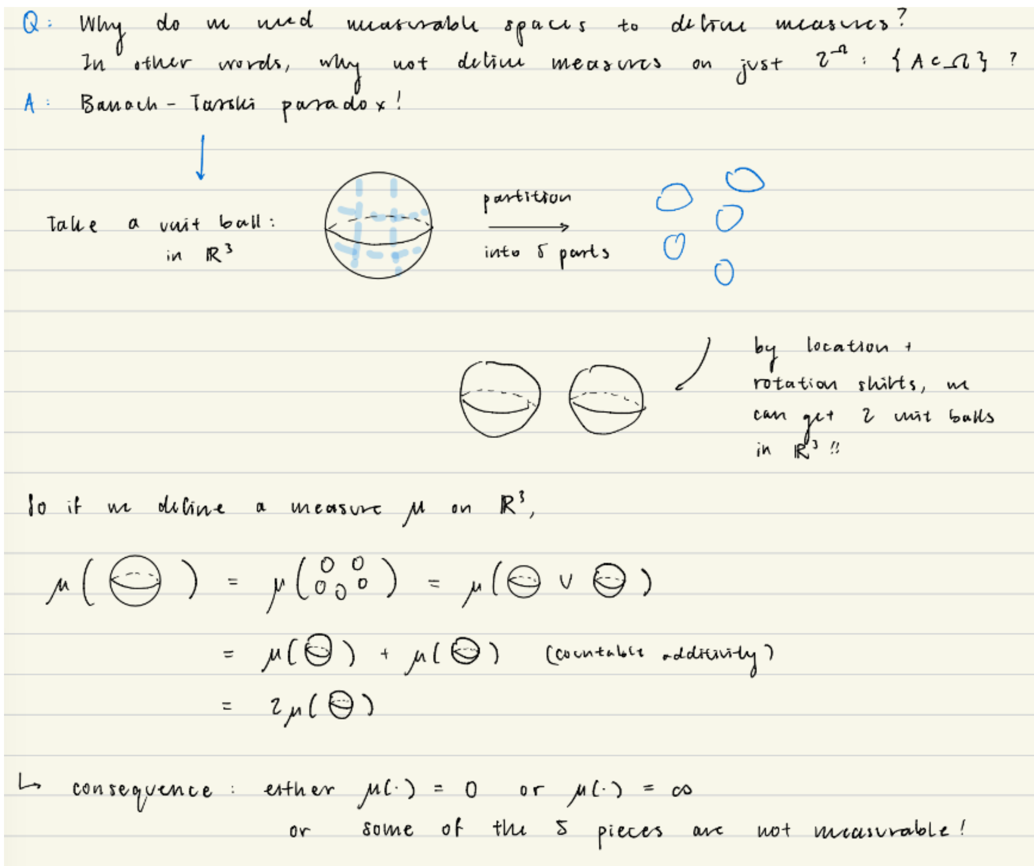

Figure 1 Visualization of Banach-Tarski paradox (credit: Danielle Tsao).

Lecture 2: σ-algebra and measure spaces continued

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
March 29th, 2024

1 Properties of σ-algebra

Last time we introduced σ-algebra. Let's remark on some important properties of it.

Remark 1.1.
1. $\Omega \in \mathcal{F}$.
2. A σ-algebra is also closed under countable intersection.
3. intersection of any arbitrary collection of σ- algebras on a set is a σ-algebra.

We can generate a σ-algebra from a set as follows:

Definition 1.2 (σ-algebra generated by $\mathcal{A}$)**.** Let Ω be a set and $\mathcal{A}$ be a collection of subsets of Ω. The *smallest* σ-algebra containing $\mathcal{A}$ is called the σ-algebra generated by $\mathcal{A}$, and is denoted by $\sigma(\mathcal{A})$.

Proof. The proof of existence follows three steps:

1. There exists at least one σ-algebra containing $\mathcal{A}$: the power set 2^{Ω}.
2. We can denote all σ-algebra $\mathcal{B}_i$ containing $\mathcal{A}$ as $\{\mathcal{B}_i : i \in \mathcal{I}\}$.
3. We know that $\bigcap_{i\in\mathcal{I}} \mathcal{B}_i$ (i) is still a σ-algebra, (ii) still contains $\mathcal{A}$, and (iii) the smallest σ-algebra containing $\mathcal{A}$.

□

Definition 1.3 (Topological space)**.** In point-set topology, a topological space (Ω, τ) is one such that

1. Ω is a set
2. τ is a class of subsets in Ω such that (a) $\emptyset, \Omega \in \tau$, (b) closed to arbitrary union, and (c) closed to finite intersection.

An important class of σ-algebra is the *Borel σ-algebra*. The definition in the textbook is

Definition 1.4 (Borel σ-algebra)**.** Let Ω be a set endowed with a topology. The Borel σ-algebra on Ω, $\mathcal{B}(\Omega)$, is the σ-algebra generated by the collection of open subsets of Ω.

And the definition in the lecture is

Definition 1.5 (Borel σ-algebra)**.** Let (Ω, τ) be a topological space, then Borel σ-algebra of (Ω, τ) is $\sigma(\tau)$.

Remark 1.6.
1. $\mathcal{B}(\mathbb{R})$ is also generated by the set of all open intervals, half-open intervals, or closed intervals.
2. $\mathcal{B}(\mathbb{R})$ is also generated by the set of the intervals of the form (x, ∞) or $(-\infty, x)$.

2 Properties of measurable/measure space

Last time we also introduced measurable space and measure. Let's remark on some important properties of them.

Definition 2.1 (Probability measure and Probability space)**.** If a measure μ on a measurable space $(\Omega, \mathcal{F})$ satisfies $\mu(\Omega) = 1$, then it is called a *probability measure*. The triple $(\Omega, \mathcal{F}, \mu)$ is then called a *probability space*. In this case, elements of $\mathcal{F}$ are usually called *events*.

Remark 2.2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

1. (Monotonicity) $\forall\, A, B \in \mathcal{F}$ such that $A \subseteq B$, we have $\mu(A) \leq \mu(B)$.
2. (Countable subadditivity) $\forall\, A_1, A_2, \cdots \in \mathcal{F}$, we have $\mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.
3. (Continuity I) Let $A_1, A_2, \cdots \in \mathcal{F}$ be a collection of increasing events in $\mathcal{F}$ (i.e., $A_1 \subseteq A_2 \subseteq \cdots$), then $\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.
4. (Continuity II) Let $A_1, A_2, \cdots \in \mathcal{F}$ be a collection of decreasing events in $\mathcal{F}$ (i.e., $A_1 \supseteq A_2 \supseteq \cdots$) and $\mu(\Omega) < \infty$ (e.g., probability measure), then $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

3 Well-approximation theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If μ is a probability measure, then any element of $\mathcal{F}$ can be arbitrarily well-approximated by elements of any generating algebra.

Theorem 3.1 (Well approximation of a σ-algebra that generates it)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let $\mathcal{A}$ be an algebra that generates $\mathcal{F}$. Then for any $A \in \mathcal{F}$ and any $\epsilon > 0$, there is some $B \in \mathcal{A}$ such that $\mu(A \Delta B) < \epsilon$, where $A \Delta B$ is the symmetric difference of A and B.*

Proof. Define $\mathcal{G} := \{A \in \mathcal{F} : A \text{ can be } \epsilon\text{-approximated by } \mathcal{A}\}$. Thus, by definition, $\mathcal{G} \subseteq \mathcal{F}$ and $\mathcal{A} \subseteq \mathcal{G}$. We can show that $\mathcal{G}$ is a σ-algebra (see the proof in the textbook). Thus, $\mathcal{G} \supseteq \sigma(\mathcal{A}) = \mathcal{F}$. Thus, $\mathcal{G} = \mathcal{F}$. □

4 Motivating Caratheodory's extension theorem

How can we define "measure" of $\mathbb{R}$? We have seen why Banach-Tarski paradox says we *cannot* define μ over $2^{\mathbb{R}}$. Consider $\mathcal{P} := \{(a, b] : a \leq b \in \mathbb{R} \cup \{-\infty, \infty\}\}$. Let $\mu((a, b]) := b - a$. If we want to extend from $(\mathbb{R}, \mathcal{P}, \mu)$ to $(\mathbb{R}, \sigma(\mathcal{P}), \tilde{\mu})$, this will be *Caratheodory's extension theorem*. Specifically, we want

- $\forall A \in \mathcal{P}$, $\tilde{\mu}(A) = \mu(A)$.
- $\tilde{\mu}$ is a measure of support $\sigma(\mathcal{P})$.
- $\tilde{\mu}$ is the *unique* extension from $(\mathbb{R}, \mathcal{P}, \mu)$.

References

Lecture 3: Dynkin's π-λ theorem, Outer measure, Starting Caratheodory extension theorem

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 1st, 2024

1 Big picture

We aim to define a measure on $\mathbb{R}$. Given the Banach-Tarski paradox, we CANNOT define a measure over $2^{\mathbb{R}}$. Instead, we aim to define a measure over a subset of $2^{\mathbb{R}}$, which are called the *class of measurable sets*. To achieve this, we need Caratheodory extension theorem. It has *five* steps.

2 Step 1: Dynkin's π-λ theorem

The first step is to show that if there exists such an extension from $(\mathcal{P}, \mu)$ to $(\sigma(\mathcal{P}), \tilde{\mu})$, it must be *unique*. Intuitively speaking, if we want to measure something on $\sigma(\mathcal{P})$, there should be only one way to measure it. To show this, we use Dynkin's π-λ theorem.

Definition 2.1 (π-system)**.** Let Ω be a set. A class $\mathcal{P}$ of subsets of Ω is a π-*system* if it is closed under finite intersection, i.e., $\forall A, B \in \mathcal{P}$, we have $A \cap B \in \mathcal{P}$.

Example 2.2. $\{(a, b) : a \leq b \in \mathbb{R}^*\}$ is a π-system.

Definition 2.3 (λ-system)**.** Let Ω be a set. A class $\mathcal{L}$ of subsets of Ω is a λ-*system* if

1. $\Omega \in \mathcal{L}$,
2. closed under complement,
3. closed under countable disjoint union

Lemma 2.4. *If a λ-system is also a π-system, then it is a σ-algebra.*

Proof. Skipped in the lecture. See the textbook. □

Theorem 2.5 (Dynkin's π-λ theorem)**.** *Let Ω be a set. Let $\mathcal{P}$ be a π-system of subsets of Ω, and let $\mathcal{L} \supseteq \mathcal{P}$ be a λ-system of subsets of Ω. Then $\mathcal{L} \supseteq \sigma(\mathcal{P})$.*

Proof. Skipped in the lecture. See the textbook. □

Theorem 2.6 (Unique extension)**.** *Consider a π-system $\mathcal{P}$ and two measures μ_1, μ_2 on the same measurable space $(\Omega, \sigma(\mathcal{P}))$. Then, as long as,*

1. *μ_1, μ_2 agree on $\mathcal{P}$, i.e., $\forall A \in \mathcal{P}, \mu_1(A) = \mu_2(A)$.*
2. *(σ-finiteness) $\exists$ a sequence $A_1, A_2, \dots \in \mathcal{P}$ such that A_n increases to Ω and μ_1, μ_2 are both finite on every A_n,*

μ_1, μ_2 agree on $\sigma(\mathcal{P})$.

Proof. Consider any $A \in \mathcal{P}$ such that $\mu_1(A) = \mu_2(A) < \infty$. Let

$$\mathcal{L} := \{B \in \sigma(\mathcal{P}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}.$$

We claim that $\mathcal{L}$ is a π-system containing $\mathcal{P}$. Clearly, $\Omega \in \mathcal{L}$. To show the closeness under complement, consider any $B \in \mathcal{L}$, then

$$\begin{aligned}
\mu_1(A \cap B^{\mathsf{c}}) &= \mu_1(A) - \mu_1(A \cap B) \qquad \text{if } \mu_1(A) < \infty \\
&= \mu_2(A) - \mu_2(A \cap B) \\
&= \mu_2(A \cap B^{\mathsf{c}}) \qquad \text{if } \mu_1(A) < \infty.
\end{aligned}$$

So $B^{\mathsf{c}} \in \mathcal{L}$. Finally, to show the closeness under countable disjoint union, consider a countable disjoint collection of elements $B_1, B_2, \dots$ in $\mathcal{L}$. Let $B := \bigcup_{n=1}^{\infty} B_n$. Then,

$$\begin{aligned}
\mu_1(A \cap B) &= \mu_1\left(A \cap \bigcup_{n=1}^{\infty} B_n\right) \\
&= \mu_1\left(\bigcup_{n=1}^{\infty} (A \cap B_n)\right) \\
&= \sum_{n=1}^{\infty} \mu_1(A \cap B_n) \qquad \because \text{countable additivity} \\
&= \sum_{n=1}^{\infty} \mu_2(A \cap B_n) \\
&= \mu_2(A \cap B) \qquad \because \text{countable additivity.}
\end{aligned}$$

By Dynkin's π-λ theorem, $\sigma(\mathcal{P}) \subseteq mcL$. By construction of $\mathcal{L}$, $\mathcal{L} \subseteq \sigma(\mathcal{P})$. So $\sigma(\mathcal{P}) = \mathcal{L}$.

So far, we have proven for every $A \in \mathcal{P}$ and $B \in \sigma(\mathcal{P})$ such that $\mu_1(A) < \infty$ and $\mu_1(A \cap B) = \mu_2(A \cap B)$. By the given conditions, there exists an increasing sequence $A_n \to \Omega$ such that $\mu_1(A_n) < \infty$. So for any $B \in \sigma(\mathcal{P})$,

$$\mu_1(B) = \lim_{n\to\infty} \mu_1(A \cap B) = \lim_{n\to\infty} \mu_2(A \cap B) = \mu_2(B).$$

$\square$

We continue to prove Caratheodory extension theorem. We first want to extend $(\mathcal{P}, \mu)$ to $\left(2^{\Omega}, \mu^*\right)$ for some μ^* and then restrict it to $\left(\mathcal{F}^{\mu^*}, \mu^*\right)$ where $\mathcal{F}^{\mu^*}$ is defined by μ^*. Caratheodory showed that (1) $\mathcal{F}^{\mu^*}$ is a σ-algebra and (2) μ^* is a measure on $\mathcal{F}^{\mu^*}$. Then, $\left(\Omega, \mathcal{F}^{\mu^*}, \mu^*\right)$ is a measure space that extends $(\mathcal{P}, \mu)$.

We will show how to construct $\mu*$ over 2^{Ω} in the next lecture. This time, we will show $\left(\Omega, \mathcal{F}^{\mu^*}, \mu^*\right)$ is a measure space by as long as μ^* is an *outer measure*, then $\left(\Omega, \mathcal{F}^{\mu^*}, \mu^*\right)$ is a measure space.

3 Step 2: Outer measure

Definition 3.1 (Outer measure)**.** Let Ω be a set and to 2^{Ω} is the power set. A set function $\phi : 2^{\Omega} \to [0, \infty]$ is called an outer measure if

1. $\phi(\emptyset) = 0$,
2. (monotonicity) If $A \subseteq B$, then $\phi(A) \leq \phi(B)$,
3. (subadditivity) If $A_1, A_2, \cdots \in \Omega$, $\phi(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \phi(A_n)$.

How to define the σ-algebra induced by ϕ, $\mathcal{F}^{\phi}$? Let's use the definition of ϕ-measurability.

Definition 3.2 (ϕ-measurable)**.** If ϕ is an outer measure on a set Ω, a subset $A \subseteq \Omega$ is called ϕ-*meausrable* if for all $B \in \Omega$,

$$\phi(B) = \phi(B \cap A) + \phi(B \cap A^{\mathsf{c}}).$$

Remark 3.3. A subset A is ϕ-measurable if and only if $\phi(B) \geq \phi(B \cap A) + \phi(B \cap A^{\mathsf{c}})$.

Theorem 3.4 (Caratheodory)**.** *Let Ω be a set and ϕ be an outer measure on Ω. Let $\mathcal{F}^{\phi}$ be the collection of all ϕ-measurable subsets of Ω. Then, $\mathcal{F}^{\phi}$ is a σ-algebra and ϕ is a measure on $\mathcal{F}^{\phi}$, i.e., $\left(\Omega, \mathcal{F}^{\phi}, \phi\right)$ is a measure space.*

Proof. The *first step* is to show $\mathcal{F}^{\phi}$ is an algebra. To show $\emptyset \in \mathcal{F}^{\phi}$, we want to show $\emptyset$ is ϕ-measurable. Since $\forall E \subseteq \Omega$, $\phi(E \cap \emptyset) = \phi(\emptyset) = 0$ and $\phi(E \cap \emptyset^{\mathsf{c}}) = \phi(E \cap \Omega) = \phi(E)$, we have $\phi(E) = 0 + \phi(E) = \phi(E \cap \emptyset) + \phi(E \cap \emptyset^{\mathsf{c}})$. Next, to show $\mathcal{F}^{\phi}$ is closed to complements, consider any $A \in \mathcal{F}^{\phi}$. Then, for any $E \subseteq \Omega$, $\phi(E) = \phi(E \cap A) + \phi(E \cap A^{\mathsf{c}}) = \phi(E \cap (A^{\mathsf{c}})^{\mathsf{c}}) + \phi(E \cap A^{\mathsf{c}})$, implying that $A^{\mathsf{c}} \in \mathcal{F}^{\phi}$. Finally, to show $\mathcal{F}^{\phi}$ is closed to finite union, consider any $A, B \in \mathcal{F}^{\phi}$. Let $D := A \cup B$. Then,

$$\begin{aligned}
\phi(E \cap D) + \phi(E \cap D^{\mathsf{c}}) &= \phi(E \cap (A \cup B)) + \phi(E \cap (A \cup B)^{\mathsf{c}}) \\
&= \phi(E \cap (A \cup (B - A))) + \phi(E \cap A^{\mathsf{c}} \cap B^{\mathsf{c}}) \\
&= \phi(E \cap (A \cup (B \cap A^{\mathsf{c}}))) + \phi(E \cap A^{\mathsf{c}} \cap B^{\mathsf{c}}) \\
&= \phi((E \cap A) \cup (E \cap B \cap A^{\mathsf{c}})) + \phi(E \cap A^{\mathsf{c}} \cap B^{\mathsf{c}}) \\
&\leq \phi(E \cap A) + \phi(E \cap B \cap A^{\mathsf{c}}) + \phi(E \cap A^{\mathsf{c}} \cap B^{\mathsf{c}}) \\
&= \phi(E \cap A) + \phi(E \cap A^{\mathsf{c}}) \qquad \because \text{B is } \phi\text{-measurable} \\
&= \phi(E).
\end{aligned}$$

On the other hand, by monotonicity,

$$\phi(E) = \phi((E \cap D) \cup (E \cap D^{\mathsf{c}})) \leq \phi(E \cap D) + \phi(E \cap D^{\mathsf{c}}).$$

So $\phi(E) = \phi(E \cap D) + \phi(E \cap D^{\mathsf{c}})$.

We will continue the rest of the proof in the next lecture. ∎

References

Lecture 4: Caratheodory extension theorem continued

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 3rd, 2024

1 Continuing step 2

Recall the Caratheodory theorem from last time,

Theorem 1.1 (Caratheodory)**.** *Let Ω be a set and ϕ be an outer measure on Ω. Let $\mathcal{F}^\phi$ be the collection of all ϕ-measurable subsets of Ω. Then, $\mathcal{F}^\phi$ is a σ-algebra and ϕ is a measure on $\mathcal{F}^\phi$, i.e., $\left(\Omega, \mathcal{F}^\phi, \phi\right)$ is a measure space.*

Proof. **Step 1** is show that $\mathcal{F}^\phi$ is an algebra. We have done this last time. **Step 2** is to show ϕ is finite-additive over $\mathcal{F}^\phi$. In other words, for any $E \subseteq \Omega$ and any disjoint $A_1, A_2, \ldots, A_n \in \mathcal{F}^\phi$, $\phi(E \cap (A_1 \cup A_2 \cup \cdots \cup A_n)) = \sum_{i=1}^n \phi(E \cap A_i)$. Define $B_n := \bigcup_{i=1}^n A_i$. Then,

$$\begin{aligned}
\phi(E \cap B_n) &= \phi(E \cap B_n \cap A_n) + \phi(E \cap B_n \cap A_n^{\mathsf{c}}) && \because A_n \text{ is } \phi\text{-measurable} \\
&= \phi(E \cap A_n) + \phi(E \cap B_{n-1}) \\
&= \phi(E \cap A_n) + \phi(E \cap B_{n-1} \cap A_{n-1}) + \phi(E \cap B_{n-1} \cap A_{n-1}^{\mathsf{c}}) && \because A_{n-1} \text{ is } \phi\text{-measurable} \\
&= \phi(E \cap A_n) + \phi(E \cap A_{n-1}) + \phi(E \cap B_{n-2}) \\
&\vdots \\
&= \sum_{i=1}^n \phi(E \cap A_i)
\end{aligned}$$

Step 3 is to show that if $A_1, A_2, \ldots$ is a increasing sequence of ϕ-measurable sets to $A \subseteq \Omega$ (A is not necessarily ϕ-measurable), then for any $E \subseteq \Omega$, $\phi(E \cap A) \leq \lim_{n\to\infty} \phi(E \cap A_n)$. Define $B_n := A_n - A_{n-1} = A_n \cap (A_{n-1})^{\mathsf{c}}$. We have the following observations:

1. B_n's are disjoint and in $\mathcal{F}^\phi$.
2. $A_n = \bigcup_{i=1}^n B_i$.
3. ϕ is finite additive over $\mathcal{F}^\phi$ so for any $E \subseteq \Omega$

$$\phi(E \cap A_n) = \phi\left(E \cap \left(\bigcup_{i=1}^n B_i\right)\right) = \sum_{i=1}^n \phi(E \cap B_i).$$

Now, $\lim_{n\to\infty}\phi(E\cap A_n)=\lim_{n\to\infty}\sum_{i=1}^n\phi(E\cap B_i)=\sum_{i=1}^\infty\phi(E\cap B_i)$. Thus,

$$\begin{aligned}\phi(E\cap A) &= \phi\left(E\cap\left(\bigcup_{i=1}^\infty B_i\right)\right)\\ &= \phi\left(\bigcup_{i=1}^\infty(E\cap B_i)\right)\\ &\leq \sum_{i=1}^\infty \phi(E\cap B_i)\\ &= \lim_{n\to\infty}\phi(E\cap A_n).\end{aligned}$$

Step 4 is to show that $\mathcal{F}^\phi$ is a σ-algebra (i.e., $\mathcal{F}^\phi$ is closed to countable union). Let $A_1, A_2, \cdots \in \mathcal{F}^\phi$, we want to show $A := \bigcup_{n=1}^\infty A_n$ is ϕ-measurable. For any given n, define $B_n := \bigcup_{i=1}^n A_i$. Then, for any $E\subseteq\Omega$, since $B_n\in\mathcal{F}^\phi$ from step 1 ($\mathcal{F}^\phi$ is closed to finite union), $\phi(E)=\phi(E\cap B_n)+\phi(E\cap B_n^{\mathsf{c}})$. Since $B_n=\bigcup_{i=1}^n A_i\subseteq A$, $E\cap B_n^{\mathsf{c}}\supseteq E\cap A^{\mathsf{c}}$, and $\phi(E\cap B_n)\geq\phi(E\cap A^{\mathsf{c}})$ by monotonicity. So $\phi(E)\geq\phi(E\cap B_n)+\phi(E\cap A^{\mathsf{c}})$.

To finish proving $\mathcal{F}^\phi$ is a σ-algebra, by applying step 3 to B_n, $\phi(E\cap A)\leq\lim_{n\to\infty}\phi(E\cap B_n)$. So, $\phi(E)\geq\phi(E\cap A)+\phi(E\cap A^{\mathsf{c}})$ by taking the limit on both sides. On the other hand, $\phi(E)=\phi((E\cap A)\cup(E\cap A^{\mathsf{c}}))\leq\phi(E\cap A)+\phi(E\cap A^{\mathsf{c}})$. Thus, $\phi(E)=\phi(E\cap A)+\phi(E\cap A^{\mathsf{c}})$, implying A is ϕ-measurable, and thus in $\mathcal{F}^\phi$. Thus, $\mathcal{F}^\phi$ is a σ-algebra.

Step 5 is to show ϕ is a measure over $\mathcal{F}^\phi$, i.e. ϕ is countable additive. Consider a countable disjoint collection $A_1, A_2, \cdots\in\mathcal{F}^\phi$. Define $B_n := \bigcup_{i=1}^n A_i$ and $B := \bigcup_{i=1}^\infty A_i$ such that B_n's increase to B which is in $\mathcal{F}^\phi$ from step 4. Thus,

$$\phi(B)\geq\phi(B_n)=\sum_{i=1}^n\phi(A_i),$$

implying that $\phi(B)\geq\sum_{i=1}^\infty\phi(A_i)$ by taking the limit on both sides. The first inequality follows from monotonicity. The first equality follows from finite additivity (set $E=\Omega$ in step 2). On the other hand, the countable subadditivity implies $\phi(B)=\phi(\bigcup_{i=1}^\infty A_i)\leq\sum_{i=1}^\infty\phi(A_i)$. So $\phi(B)=\sum_{i=1}^\infty\phi(A_i)$. □

2 Step 3: Construction of μ^*

The final step to prove Caratheodory extension theorem is to construct an outer measure μ^* on 2^Ω. Let's also fully state the theorem here.

Theorem 2.1 (Caratheodory extension theorem). *If $(\mathcal{A},\mu)$ is a tuple such that*

1. *$\mathcal{A}$ is an algebra of subsets of Ω.*
2. *μ is a pre-measure over $\mathcal{A}$, i.e.*
   (a) *$\mu:\mathcal{A}\to[0,\infty]$,*
   (b) *$\mu(\emptyset)=0$,*

(c) *(countable additivity over an algebra) For any disjoint* $A_1, A_2, \cdots \in \mathcal{A}$*, as long as* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$*, then* $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$*,*

then there exists an extension of $(\Omega, \mathcal{A}, \mu)$ *to* $(\Omega, \sigma(\mathcal{A}), \tilde{\mu})$ *such that the latter is a measure space.*

In addition, if there exists a sequence $A_i \in \mathcal{A}$ *such that* $\Omega = \bigcup_{i=1}^{\infty} A_i$ *and* $\mu(A_i) < \infty$ *(i.e.,* μ *is* σ*-finite), then the extension above is unique.*

Proof. As stated before, it remains to construct an outer measure μ^* such that $\mu^* = \mu$ over $\mathcal{A}$. For any $A \in 2^{\Omega}$, define

$$\mu^*(A) := \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) : A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i\right\}.$$

So $\tilde{\mu}$ in the theorem statement is μ^* restricted to $\sigma(\mathcal{A})$.

Step 1 is to show μ^* is an outer measure.

- Since $\mu^*(\emptyset) \le \mu(\emptyset) = 0$ and $\mu^*(\emptyset) \ge 0$ trivially, $\mu^*(\emptyset) = 0$.
- For any $A \subseteq B \subseteq \Omega$, $\mu^*(A) \le \mu^*(B)$ since any cover of B is a cover of A.
- For any $A_1, A_2, \cdots \subseteq \Omega$, let $A := \bigcup_{i=1}^{\infty} A_i$. By the definition of infimum, fix an $\epsilon > 0$ and for each i, let $\{A_{ij}\}_{j=1}^{\infty}$ be a collection of elements in $\mathcal{A}$ such that $A_i \subseteq \cup_j A_{ij}$ and $\sum_j \mu(A_{ij}) \le \mu^*(A_i) + \epsilon 2^{-i}$. Then, $A = \cup_i A_i \subseteq \cup_i \cup_j A_{ij}$, implying that

$$\begin{aligned} \mu^*(A) &\le \sum_i^{\infty} \sum_j^{\infty} \mu(A_{ij}) \\ &\le \sum_i^{\infty} \Big(\mu^*(A_i) + \epsilon 2^{-i}\Big) \\ &= \sum_i^{\infty} \mu^*(A_i) + \epsilon. \end{aligned}$$

  Since ϵ is arbitrary, $\mu^*(A) \le \sum_i^{\infty} \mu^*(A_i)$.

Step 2 is to show μ^* is an extension of μ: i.e., for any $A \in \mathcal{A}$, $\mu^*(A) = \mu(A)$. Check Lemma 1.5.3. in the textbook.

Step 3 is to show $\mathcal{F}^{\mu^*} \supseteq \mathcal{A}$. For any $A \in \mathcal{A}$ and any $E \subseteq \Omega$, let $A_1, A_2, \dots$ be any sequence of elements of $\mathcal{A}$ that cover E. Then $\{A_i \cap A\}_{i=1}^{\infty}$ covers $E \cap A$ and $\{A_i \cap A^{\mathsf{c}}\}_{i=1}^{\infty}$ covers $E \cap A^{\mathsf{c}}$. Thus,

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}}) &\le \sum_{i=1}^{\infty} \mu(A_i \cap A) + \sum_{i=1}^{\infty} \mu(A_i \cap A^{\mathsf{c}}) \\ &= \left\{\sum_{i=1}^{\infty} \mu(A_i \cap A) + \mu(A_i \cap A^{\mathsf{c}})\right\} \\ &= \sum_{i=1}^{\infty} \mu(A_i) \qquad \because \text{countable additivity of } \mu. \end{aligned}$$

By the definition of infimum, $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}})$. On the other hand, by the subadditivity of μ^*, $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}})$, so $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}})$. Therefore, $A \in \mathcal{F}^{\mu^*}$ by definition, and $\mathcal{A} \subseteq \mathcal{F}^{\mu^*}$.

The addition part of the theorem is proved by the Dynkin's π-λ system and the unique extension theorem in the last lecture. □

Remark 2.2. In the theorem statement, $\sigma(A) = \mathcal{F}^{\mu^*}$ and $\tilde{\mu} = \mu^*$.

References

Lecture 5: Lebesgue measure, completion of σ-algebra, measurable function

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 5th, 2024

1 Lebesgue measure space on R^1

Let $\mathcal{A}$ be the set of all subsets in $\mathbb{R}^1$ such that they are finite disjoint unions of half-open intervals of $(a, b] \cap \mathbb{R}$ where $-\infty \leq a \leq b \leq \infty$. Then, we know $\mathcal{A}$ is an algebra and α generates the Borel σ-algebra of $\mathbb{R}$.

Define $\lambda : \mathcal{A} \to \mathbb{R}$ as the measurement of the length of an element in $\mathcal{A}$. Then λ is a σ-finite measure on $\mathcal{A}$.

2 Completion of σ-algebra in a measure space

In probability theory, we wish to have: as long as $A \in \mathcal{F}$ such that $\mathbb{P}(A) = 0$, then all the subsets of A are in $\mathcal{F}$. A measure/probability space is *complete* if it satisfies the property above.

For example, $\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda|_{\mathcal{B}(\mathbb{R})}\right)$ is not complete but $\left(\mathbb{R}, \mathcal{F}^{\lambda^*}, \lambda|_{\mathcal{F}^{\lambda^*}}\right)$ is complete.

Proposition 2.1. *Given any measure space $(\Omega, \mathcal{F}, \mu)$, we can extend it to a complete measure space $(\Omega, \mathcal{F}', \mu')$ by applying Caratheodory extension theorem to $(\mathcal{F}, \mu)$. Note that $\mathcal{F} \subseteq \mathcal{F}'$.*

3 Measurable function

Recall from mathematical analysis that the (Riemann) integration is defined as some sort of area under the function. Now we introduce the Lebesgue integration which is more general than the Riemann integration and defined based on a measure space $(\Omega, \mathcal{F}, \mu)$. It is written down as

$$\int_\Omega f(\omega) d\mu(\omega).$$

Lebesgue integration requires a *measurable* function, so let's talk about this idea.

Definition 3.1 (Measurable function)**.** For two measurable spaces $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$, a function or mapping $f : \Omega \to \Omega'$ is said to be $\mathcal{F}$-$\mathcal{F}'$ measurable if for any $A \in \mathcal{F}'$, $f^{-1}(A) := \{\omega \in \Omega : f(\omega) \in A\} \in \mathcal{F}$ (i.e., $f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$).

Lemma 3.2. *Let $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ be two measurable spaces and $f : \Omega \to \Omega'$ be a function. Suppose there is a set $\mathcal{A} \subseteq \mathcal{F}'$ that generates $\mathcal{F}'$, and suppose that $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$ (i.e., $f^{-1}(\mathcal{A}) \subseteq \mathcal{F}$). Then f is $\mathcal{F}$-$\mathcal{F}'$ measurable.*

Proof. Define $\mathcal{B} := \{B \subseteq \Omega' : f^{-1}(B) \in \mathcal{F}\}$. It's easy to verify that $\mathcal{B}$ is a σ-algebra. Then, we know $\mathcal{A} \subseteq \mathcal{B}$, so $\mathcal{F}' = \sigma(\mathcal{A}) \subseteq \mathcal{B}$ (i.e., for any $A \in \mathcal{F}'$, $f^{-1}(A) \in \mathcal{F}$. $\square$

Proposition 3.3. *Suppose* $(\Omega, \tau), (\Omega', \tau')$ *are two topological spaces and* $\mathcal{F}, \mathcal{F}'$ *are their Borel* σ*-algebras. Then any continuous function from* Ω *into* Ω' *is* $\mathcal{F}$-$\mathcal{F}'$ *measurable.*

Proof. By the definition of continuity, $f^{-1}(\tau') \subseteq \tau \subseteq \sigma\ of \tau$ and $\sigma(\tau') = \mathcal{F}'$. The lemma shows f is $\mathcal{F}$-$\mathcal{F}'$ measurable. □

3.1 Rules of measurability

How do we easily verify the measurability of a function? Here are some important propositions.

Proposition 3.4 (Composition of measurable functions)**.** *Consider three measurable spaces* $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}'), (\Omega'', \mathcal{F}'')$*. If two functions* f, g *are* $\mathcal{F}$-$\mathcal{F}'$ *and* $\mathcal{F}'$-$\mathcal{F}''$ *measurable, then* $g \circ f$ *is* $\mathcal{F}$-$\mathcal{F}''$ *measurable.*

Proposition 3.5 (Sum and product of measurable functions)**.** *Consider two measurable spaces* $(\Omega, \mathcal{F}), (\mathbb{R}, \mathcal{B}(R))$*. If two functions* $f, g : \Omega \to \mathbb{R}$ *are* $\mathcal{F}$-$\mathcal{B}(R)$ *measurable, then* $f + g, f - g, f * g$ *are all measurable.*

Proposition 3.6 (Left-continuous and right-continuous)**.** *Any right-continuous or left-continuous function* $f : \mathbb{R} \to \mathbb{R}$ *is measurable.*

Proposition 3.7 (Monotone)**.** *Any monotone function* $f : \mathbb{R} \to \mathbb{R}$ *is measurable.*

Proposition 3.8 (Infimum and supremum)**.** *For two measurable spaces* $(\Omega, \mathcal{F})$ *and* $(\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$*, let* $\{f_n\}_{n \geq 1}$ *be a sequence of measurable functions from* Ω *to* $\mathbb{R}^*$*. Then, we have the following:*

- $g := \inf_{n \geq 1} f_n$ *and* $h := \sup_{n \geq 1} f_n$ *are measurable,*
- $\liminf_{n \to \infty} f_n$ *and* $\limsup_{n \to \infty} f_n$ *are measurable,*
- *If* $f_n \to f$ *pointwise, then* f *is measurable. This can be generalized to any general function* f *from* $(\Omega, \mathcal{F})$ *to another measurable space* $(S, \mathcal{B}(S))$*.*

References

Lecture 6: Lebesgue integration, Linearity

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 15th, 2024

1 Lebesgue integration

1.1 Setup

Consider a measurable function $f : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$. We want to define the Lebesgue integration

$$\int_\Omega f(\omega) d\mu(\omega).$$

1.2 Indicator function

Let $f := \mathbb{1}_A(\cdot)$ for some $A \in \mathcal{F}$. Then,

$$\int_\Omega f(\omega) d\mu(\omega) = \mu(A).$$

1.3 Nonnegative simple function (NSF)

Let $f := \sum_{i=1} a_i \mathbb{1}_{A_i}(\cdot)$ for some $n < \infty, a_i \geq 0, A_i$'s are disjoint elements in $\mathcal{F}$. Then

$$\int_\Omega f(\omega) d\mu(\omega) = \sum_{i=1}^n a_i \mu(A_i).$$

Remark 1.1.

- If f is an NSF and $a \geq 0$, then $a \cdot f$ is an NSF and so

$$\int af d\mu = a \int f d\mu,$$

- For any NSFs f and g, $f + g$ is still an NSF and

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu$$

1.4 Nonnegative measurable function

Let $f : \Omega \to [0, \infty]$. Define $\mathrm{SF}^+(f) := \{\text{all NSF bounded by } f\}$. Then

$$\int_\Omega f(\omega) d\mu(\omega) = \sup_{g \in \mathrm{SF}^+(f)} \int g d\mu.$$

1.5 General measurable function

Let $f : \Omega \to \mathbb{R}^*$. Define $f^+ := \max\{f, 0\}$ and $f^- := -\min\{f, 0\}$. Observe that $f = f^+ - f^-$ where f^+, f^- are measurable and nonnegative. Then

$$\int_\Omega f(\omega) d\mu(\omega) = \int f^+ d\mu - \int f^- d\mu.$$

Definition 1.2 (Well-defined integral)**.** Consider any measurable function $f : \Omega \to \mathbb{R}^*$. $\int f d\mu$ is said to be *well-defined* if one of $\int f^+ d\mu, \int f^- d\mu$ is finite (in order to avoid $\infty - \infty$).

Definition 1.3 (Integrable)**.** A measurable function $f : \Omega \to \mathbb{R}^*$ is said to be *integrable* if both $\int f^+ d\mu, \int f^- d\mu$ are finite.

In practice, we usually do not integrate over the entire Ω but a subset $S \subseteq \Omega$. We define

$$\int_S f d\mu = \int_S f \mathbb{1}_S(\cdot) d\mu.$$

2 Linearity of Lebesgue integration

We want to show the following proposition that Lebesgue integration is a linear operator.

Proposition 2.1. *If f, g are measurable and nonnegative from Ω to $[0, \infty]$, then*

- $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ *and*
- *For any $\alpha \in \mathbb{R}$, $\int (\alpha f) d\mu = \alpha \int f d\mu$.*

If f, g are measurable and integrable, then for any $\alpha, \beta \in \mathbb{R}$,

- *$\alpha f + \beta g$ is integrable and*
- $(\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$.

To prove this proposition, we need the following two propositions or theorems:

Proposition 2.2. *Given any negative measurable function f, there exists a sequence of NSFs $\{f_n\}_{n \geq 1}$ increasing pointwisely to f.*

Proof. See the textbook Proposition 2.3.6. □

Theorem 2.3 (Monotone convergence theorem (MCT))**.** *Consider a sequence of measurable, nonnegative functions $\{f_n\}_{n \geq 1}$ increasing to f pointwisely. Then,*

- *f is nonnegative,*
- *f is measurable, and*
- $\lim_{n \to \infty} \int f_n d\mu = \int f d\mu$.

Proof. See the next lecture note. □

Now let's prove the first part of proposition 2.1.

Proof. By proposition 2.2, there exists increasing sequences of NSFs $\{f_n\}_{n \geq 1}$ and $\{g_n\}_{n \geq 1}$ to f and g. By MCT,

1. $\lim \int f_n d\mu = \int f d\mu$ and $\lim \int g_n d\mu = \int g d\mu$,
2. $f_n + g_n$ is an NSF such that $\int (f_n + g_n) d\mu = \int f_n d\mu + \int g_n d\mu$ and $f_n + g_n$ increasing pointwise to $f + g$,

3.

$$\begin{aligned}\int (f+g)d\mu &= \int \lim_{n\to\infty}(f_n+g_n)d\mu \\ &= \lim_{n\to\infty}(f_n+g_n)d\mu \\ &= \lim_{n\to\infty}\left(\int f_n d\mu + \int g_n d\mu\right) \\ &= \lim_{n\to\infty}\int f_n d\mu + \lim_{n\to\infty}\int g_n d\mu \\ &= \int f d\mu + \int g d\mu.\end{aligned}$$

□

References

Lecture 7: Monotone convergence theorem

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 17th, 2024

1 Recap

There are some nice properties on Lebesgue integration

1. In many simple cases (e.g., f is continuous, bounded, and has bounded domain), if f is Riemann integrable, then it is also Lebesgue integrable and two integrals are equal.

2. Lebesgue integration is a linear operator.

2 Monotone convergence theorem (MCT)

Before we prove the MCT, we need two lemmas.

Lemma 2.1. *If f, g are two measurable, nonnegative functions such that $f \leq g$, then $\int f d\mu \leq \int g d\mu$*

Proof. Since $f \leq g$, $\mathrm{SF}^+(f) \subseteq \mathrm{SF}^+(g)$. So

$$\begin{aligned}\int f d\mu &= \sup_{s \in \mathrm{SF}^+(f)} \int s d\mu \\ &\leq \sup_{s \in \mathrm{SF}^+(g)} \int s d\mu \\ &= \int g d\mu.\end{aligned}$$

□

Lemma 2.2. *Let s be an NSF on Ω. We can define $\nu : \mathcal{F} \to \mathbb{R}^*$ such that $\nu(S) := \int_S s d\mu$ for each $S \in \mathcal{F}$. Then ν is a measure over $(\Omega, \mathcal{F})$*

Proof. By construction, $\nu \geq 0$ and $\nu(\emptyset) = 0$. Let $S_1, S_2, \ldots$ be a sequence of disjoint sets in $\mathcal{F}$, and let $S := \bigcup_{n=1}^{\infty} S_n$. Then,

$$\begin{aligned}
\nu(S) &= \int_S s(\omega) d\mu(\omega) \\
&= \int_S \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega) d\mu(\omega) \\
&= \int_\Omega \left(\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega) \right) \mathbb{1}_S(\omega) d\mu(\omega) \\
&= \int_\Omega \sum_{i=1}^{n} a_i \mathbb{1}_{A_i \cap S}(\omega) d\mu(\omega) \\
&= \sum_{i=1}^{n} a_i \mu(A_i \cap S) \\
&= \sum_{i=1}^{n} a_i \mu\Big(A \cap \cup_{j=1}^{\infty} S_j\Big) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{\infty} \mu(A_i \cap S_j) \\
&= \sum_{j=1}^{\infty} \sum_{i=1}^{n} \mu(A_i \cap S_j) \\
&= \sum_{j=1}^{\infty} \nu(S_j)
\end{aligned}$$

□

Recall the MCT.

Theorem 2.3 (Monotone convergence theorem (MCT)). *Consider a sequence of measurable, nonnegative functions $\{f_n\}_{n\geq 1}$ increasing to f pointwisely. Then,*

- *f is nonnegative,*
- *f is measurable, and*
- $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

Proof. Because $f_n \geq f_m$ for all $n > m$, by the first lemma, $\int f_n d\mu \geq \int f_m d\mu$. Thus, $\int f_n d\mu$ is increasing, implying that $\lim \int f_n d\mu$ exists and can be ∞.

Because $f \geq f_n$, by the first lemma again, for all n $\int f d\mu \geq \int f_n d\mu$, implying that $\int f d\mu \geq \lim_{n\to\infty} \int f_n d\mu$.

Take any $s \in \mathrm{SF}^+(f)$ and fix arbitrary $\alpha \in (0, 1)$. By the second claim, we can define $\nu(S) = \int_S s d\mu$ for any $S \in \mathcal{F}$. Define $S_n := \{\omega \in \Omega : \alpha s(\omega) \leq f_n(\omega)\}$. Since $s \in \mathrm{SF}^+(f)$ and f_n increases to f, S_n increases to Ω. So $\nu(S_n)$ increases to $\nu(\Omega)$. Now we have $\int_\Omega s d\mu = \nu(\Omega) = \lim_{n\to\infty} \nu(S_n) = \lim_{n\to\infty} \int_{S_n} s d\mu$. This implies that $\alpha \int_\Omega s d\mu = \alpha\nu(\Omega) = \lim_{n\to\infty} \alpha\nu(S_n) = \lim_{n\to\infty} \int_{S_n} \alpha s d\mu$. Thus, $\int \alpha s d\mu = \lim_{n\to\infty} \int_{S_n} \alpha s d\mu = \lim_{n\to\infty} \int_\Omega \alpha s \mathbb{1}_{S_n}(\cdot) d\mu \leq \lim_{n\to\infty} \int_\Omega f_n d\mu$. The last equality follows the first claim.

Now we have for any $s \in \mathrm{SF}^+(f)$ and any $\alpha \in (0,1)$, $\alpha \cdot \int s d\mu \leq \lim_{n\to\infty} \int f_n d\mu$, implying that $\int f d\mu = \sup_{s\in\mathrm{SF}^+(f),\alpha\in(0,1)} \alpha \int s d\mu \leq \lim \int f_n d\mu$. □

References

Lecture 8: Fatou's lemma, DCT, a.e., a.s.

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 19th, 2024

1 Fatou's lemma

Lemma 1.1 (Fatou's lemma). *For a sequence of measurable nonnegative functions f_n,*

$$\int (\liminf f_n) d\mu \leq \liminf \int f_n d\mu. \tag{1}$$

Proof. Define $g_n := \inf_{m \geq n} f_m$. We can see that $g_n \geq 0$, g_n is increasing to $\liminf f_n$, and $g_n \leq f_n$. The LHS of the equation (1) is $\int \lim g_n d\mu = \lim \int g_n d\mu \leq \liminf \int f_n d\mu$ which is the RHS of the equation (1). The first equality follows from MCT, the first inequality follows from that $\int g_n d\mu \leq \int t f_n d\mu$. □

2 DCT

Theorem 2.1 (DCT). *Suppose the following:*

1. $f_n : \Omega \to \mathbb{R}^*$ *are measurable.*
2. $f := \lim f_n$ *exists pointwisely.*
3. f_n*'s are dominated by an integrable function $h(\cdot)$, i.e., $\exists h : \Omega \to \mathbb{R}^*$ measurable such that $\forall n, \forall w \in \Omega, |f_n(w)| \leq h(w)$ and $\int h d\mu < \infty$.*

Then, We claim that

1. $\int f d\mu = \int \lim f_n d\mu = \lim \int f_n d\mu$.
2. $\lim \int |f_n - f| d\mu = 0$.

Proof. The proof is based on Fatou's lemma. Observe that following:

1. From the homework, we know f is measurable.
2. Since $\forall n, |f_n| \leq h$, we have $|f| \leq h$.
3. Since $|f_n| \leq h$ and $|f| \leq h$, we have $f_n + h$ and $f + h$ are measurable, nonnegative, and integrable.

4.

$$
\begin{aligned}
f_n \to f &\implies f_n + h \to f + h \\
&\implies \int \liminf (f_n + h) d\mu \le \liminf \int (f_n + h) d\mu \qquad \because \text{Fatou's lemma} \\
&\implies \int (f + h) d\mu \le \liminf \int (f_n + h) d\mu \\
&\implies \int (f + h) d\mu = \int f d\mu + \int h d\mu \le \liminf \int f_n d\mu + \int h d\mu \\
&\qquad = \liminf \int f_n d\mu + \int h \mathrm{d}\mu \\
&\qquad \because f, h, f_n, f + h, f_n + h \text{ are integrable and } \int h d\mu < \infty \\
&\implies \int f d\mu \le \liminf \int f_n d\mu.
\end{aligned}
$$

5. It remains to show that $\int f d\mu \ge \limsup \int f_n d\mu$. Apply the argument in (4) to $-f_n, -f$. In particular, we have

$$
\begin{aligned}
\int (-f) d\mu \le \liminf \int (-f_n) d\mu &\implies -\int f d\mu = \int (-f) d\mu \le \liminf \int (-f_n) d\mu \\
&\qquad = -\limsup \int (f_n) d\mu \\
&\implies \int f d\mu \ge \limsup \int (f_n) d\mu
\end{aligned}
$$

□

Remark 2.2 (Remarks for DCT).

- The *dominated condition*, which is the third condition, CANNOT be dropped. i.e., if it doesn't hold, then *probably* $\int \lim f_n d\mu \neq \lim \int f_n d\mu$.
- Notice that the nonnegativity is need for MCT and Fatou but *not* needed for DCT.
- The second claim implies the first claim as

$$
\begin{aligned}
\left| \int f d\mu - \lim \int f_n d\mu \right| &= \lim \left| \int (f - f_n) d\mu \right| \\
&\le \lim \int |f - f_n| d\mu \\
&= 0.
\end{aligned}
$$

- The first claim implies the first claim as applying the first claim to the sequence $g_n := |f_n - f|$.

Here's an example of DCT theorem.

Example 2.3. Given $x_1, x_2, \ldots, x_n \overset{\text{i.i.d.}}{\sim} f_{\theta_0}(\cdot)$, the goal is to estimate or infer θ_0 based only on $x_1, \ldots, x_n$. The MLE method is to find $\hat{\theta}_n$ such that $\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^n \log f_\theta(x_i)$, implying that $\hat{\theta}$ is the root of $\frac{\partial}{\partial\theta}[\sum_{i=1}^n \log f_\theta(x_i)]$. Fisher claimed that $\hat{\theta}$ should be close to θ_0, as θ_0 should make $\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(x_i)\big|_{\theta=\theta_0}$ close to 0. By the law of large numbers, $\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(x_i)\big|_{\theta=\theta_0}$ should converge to $\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f_\theta(X)\big|_{\theta=\theta_0}\right]$, and we claim that $\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f_\theta(X)\big|_{\theta=\theta_0}\right]$ is 0. The proof from 513 goes as follows:

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f_\theta(X)\big|_{\theta=\theta_0}\right] &= \int \frac{\partial}{\partial\theta} \log f_\theta(X)\big|_{\theta=\theta_0} f_{\theta_0}(x)dx \\
&= \int \frac{\frac{\partial}{\partial\theta} f_\theta(X)\big|_{\theta=\theta_0}}{f_{\theta_0}(x)} f_{\theta_0}(x)dx \\
&= \int \frac{\partial}{\partial\theta} f_\theta(X)\big|_{\theta=\theta_0} dx \\
&= \frac{\partial}{\partial\theta}\int f_{\theta_0}(x)dx \qquad \because \text{DCT} \\
&= \frac{\partial}{\partial\theta} 1 \\
&= 0.
\end{aligned}$$

When can we switch $\int$ and $\frac{\partial}{\partial\theta}$?

Proposition 2.4. *Let $f : \mathcal{I} \times \Omega \to \mathbb{R}$ where $\mathcal{I}$ is an open set in $\mathbb{R}$. Then, under certain conditions, we have $\forall x \in \mathcal{I}$, $\frac{d}{dx}[\int f(x,w)d\mu(w)] = \int \left[\frac{d}{dx} f(x,w)d\mu(w)\right]$.*

3 "Almost everywhere" and "Almost surely"

Definition 3.1 (almost everywhere)**.** Consider a measure space $(\Omega, \mathcal{F}, \mu)$. A set $A \in \mathcal{F}$ is aid to *happen almost everywhere (a.e.)* if and only if $\mu(A^{\mathsf{c}}) = 0$. In this case, we say A is μ-a.e., or A is a.e.

Definition 3.2 (almost surely)**.** It is a.e. when a measure space $(\Omega, \mathcal{F}, \mu)$ is a probability space.

Example 3.3. We say $f = g$ a.e. if, except for a μ-measure 0 set, $f = g$. i.e., $\exists E \in \mathcal{F}$ such that $\mu(E) = 0$ and $f(w) = g(w), \forall w \in E^{\mathsf{c}}$.

Remark 3.4. If $(\Omega, \mathcal{F}, \mu)$ is complete, then $f = g$ a.e. if and only if $\mu(\{w : f(w) \neq g(w)\}) = 0$.

Example 3.5. We can say that $f_n \to f$ a.e. if, except for a set E of $\mu(E) = 0$, $f_n(w) \to f(w), \forall w \in E^{\mathsf{c}}$.

Proposition 3.6. *Assume $f : \Omega \to [0, \infty]$ to be measurable, we claim that $\int f d\mu = 0$ if and only if $f = 0$ a.e..*

Proof. To prove the backward direction, suppose $f = 0$ a.e., then $\forall g \in \mathrm{SF}^+(f)$, we have $\int g d\mu = 0$.

To prove the forward direction, we prove by contradiction. If $f : \Omega \to [0, \infty]$ and $f \neq 0$ a.e., then

$$\begin{aligned}
\mu(\{w : f(w) + 0\}) > 0 &\implies \mu(\{w : f(w) + 0\}) > 0 = \mu\left(\bigcup_{n=1}^{\infty} \{w : f(w) > \frac{1}{n}\}\right) \because \text{set theory} \\
&\implies \mu(\{w : f(w) + 0\}) > 0 = \lim \mu\left(\{w : f(w) > \frac{1}{n}\}\right) > 0 \\
&\implies \exists n, \mu\left(\{w : f(w) > \frac{1}{n}\}\right) > 0 \\
&\implies \int f d\mu \geq \int f \mathbb{1}(A_n) d\mu \geq \int \frac{1}{n} \mathbb{1}(A_n) d\mu = \frac{\mu(A_n)}{n} > 0
\end{aligned}$$

where $A_n := \{w : f(w) > \frac{1}{n}\}$. □

References

Lecture 9: Product Space

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 24th, 2024

1 Famous Quote from Fang today

"I don't like mathematical induction. It's unintuitive to me." – Fang Han

2 Review and introduction

So far we've covered

- (Chap 1) Measure space $(\Omega, \mathcal{F}, \mu)$, Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^n$ based on Caratheodory extension theorem. From $(\mathcal{A}, \mu_{\mathcal{A}})$ to $(\sigma(\mathcal{A}), \mu)$ and to $\left(\mathcal{F}^{\mu^*}, \mu\right)$.
- (Chap 2) Lebesgue integral of $f : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}^*, \mathcal{B}(\mathbb{R}*))$ where f is a SS, NSF, nonnegative, or general function. Also we learned linear operator, MCT, Fatou, and DCT.

Next, in Chapter 3, we will learn product space, which generalizes $\mathbb{R}^n$ and Fubini-Tornelli theorem. The motivation

1. The homework gives a specific way to define λ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$; but it CANNOT be easily generalized to a general "n-dimensional" space. For example, given two manifolds $(\mathcal{M}_1, \mathcal{F}_1, \mu_1)$ and $(\mathcal{M}_2, \mathcal{F}_2, \mu_2)$, how to define on $(\mathcal{M}_1 \times \mathcal{M}_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu)$?
2. The Fubini-Torelli theorem. From mathematical analysis,

$$\int f(x,y)d(x,y) = \int \left(\int f(x,y)dx \right) dy$$

where the inside integral is Riemann and the equality is under some conditions. Also, the notation of $d(x,y)$ is confusing. From measure theory, we have

$$\int f(x,y)d\mu(x,y)$$

where the integral is Lebesgue and $d\mu(x,y)$ will be defined on product measure in a product measure space w.r.t. product measurable space.

3 Product space

Definition 3.1 (Product set). Given *finitely many general measure spaces*: $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), \ldots, (\Omega_n, \mathcal{F}_n)$ with $n < \infty$. Then we can define the *product set* as

$$\Omega := \Omega_1 \times \Omega_2 \times \cdots \Omega_n.$$

i.e, $w = (w_1, w_2, \ldots, w_n) \in \Omega$ if and only if $w_1 \in \Omega_1, w_2 \in \Omega_2, \ldots, w_n \in \Omega_n$.

Example 3.2. If each $\Omega_i = \mathbb{R}$, then $\Omega = \mathbb{R}^n$. If $n = 2$, Ω can be a rectangle.

Definition 3.3 (Product σ-algebra)**.** Given a product set Ω. The *product σ-algebra* over Ω is

$$\mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n = \sigma(\{A_1 \times A_2 \times \cdots \times A_n : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \cdots, A_n \in \mathcal{F}_n\}).$$

Example 3.4. If $\Omega_1 = \{0,1\}, \Omega_2 = \{0,1,2\}$, what is $\mathcal{F}_1 \times \mathcal{F}_2$?

Remark 3.5.

- $(\Omega, \mathcal{F})$ is called the *the product measurable space from* $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ $, \ldots, (\Omega_n, \mathcal{F}_n)$.
- If $(\Omega_i, \mathcal{F}_i) = (\mathbb{R}, \mathcal{B}(R))$ for each $i = 1, \ldots, n$, then $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Given $\Omega_1 \times \Omega_2 \times \cdots \Omega_n$ and $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \mathcal{F}_n$, suppose we have n measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2), \ldots, (\Omega_n, \mathcal{F}_n, \mu_n)$. What would be the corresponding product measure on $(\Omega, \mathcal{F})$? In particular, when $(\Omega, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, I wish the product measure to be Lebesgue measure. To answer this question, we start with if μ is a product measure, then μ should satisfy for any $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$

$$\mu(A_1 \times A_2 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i). \tag{1}$$

  Next, define $\mu = \mu_1 \times \cdots \times \mu_n$ over $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$ following Chap 1's method: $(\mathcal{A}, \mu_{\mathcal{A}})$ to be set up such that (1) $\mathcal{A}$ is a algebra, (2) $\mu_{\mathcal{A}}$ is an pre-measure over $\mathcal{A}$, and (3) Equation (1) holds true over $(\mathcal{A}, \mu_{\mathcal{A}})$. Then, Caratheodory gives us a measure space $(\sigma(\mathcal{A}), \mu)$. It remains to construct a "good" $(\mathcal{A}, \mu_{\mathcal{A}})$.

Theorem 3.6 (Product measure space)**.** *Given*

1. $(\Omega, \mathcal{F})$ *is a product measurable space;*
2. *Each* $(\Omega_i, \mathcal{F}_i)$ *has a σ-finite measure* μ_i *(uniqueness of the extension).*

Then there exists a unique measure $\mu = \mu_1 \times \cdots \times \mu_n$ *over* $(\Omega, \mathcal{F})$ *such that for any* $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$*, equation* (1) *is true.*

Proof. It remains to construct $(\mathcal{A}, \mu_{\mathcal{A}})$ and show it satisfies the previous conditions. Let $\mathcal{A} := \{\text{finite disjoint union of rectangles (i.e., } A_1 \times \cdots \times A_n \text{ with } A_i \in \mathcal{F}_i)\}$. Then, for any $A \in \mathcal{A}$,

$$\begin{aligned} \mu_{\mathcal{A}}(A) &= \mu_{\mathcal{A}}\left(\bigcup_{i=1}^{m} A_{i1} \times \cdots \times A_{in}\right) \\ &:= \sum_{i=1}^{m} \mu_1(A_i 1)\mu_2(A_{i2}) \times \mu_n(A_{in}). \end{aligned}$$

It is immediate that μ_A satisfies the equation (1). We claim that $\mathcal{A}$ is an algebra and $\mu_{\mathcal{A}}$ is a pre-measure on $\mathcal{A}$. To see this, follow these steps:

1. $\mathcal{A}$ is an algebra. It is proved by applying induction w.r.t. the dimension n in $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$.

2. $\mu_{\mathcal{A}}$ is a pre-measure. We will only prove $\mu_{\mathcal{A}}$ is countable additive over $\mathcal{A}$.

Now, we proceed with the proof by induction. The base case $n = 1$ is trivial. Assume that μ is indeed a product measure over dimension $n - 1$. Using a similar argument of the homework (prop. 1.6.4.), we just show as long as

$$A_1 \times \cdots \times A_n = \bigcup_{i=1}^{\infty} (A_{i1} \times \cdots \times A_{in})$$

then

$$\mu(A_1 \times \cdots \times A_n) = \sum_{i=1}^{\infty} \mu_1(A_{i1}) \times \cdots \times \mu_n(A_{in}).$$

To show this, we use dimension reduction.

1. We fix an $x \in A_1 \times \cdots \times A_{n-1}$. Then introduce an index set $\mathcal{I} = \mathcal{I}_x$ to index all $i = 1, 2, \ldots$ such that $x \in A_{i1} \times A_{i2} \times \cdots \times A_{i,n-1}$.
2. We can show, since $A_1 \times \cdots A_n = \bigcup_{i=1}^n A_{i1} \times \cdots \times A_{in}$, we have $A_n = \bigcup_{i \in \mathcal{I}} A_{i,n}$.
3. observe that $\{A_{i,n} : i \in \mathcal{I}\}$ is a disjoint sequence.
4. $\forall x \in \Omega_1 \times \cdots \times \Omega_{n-1}$,

$$\mathbb{1}_{A_1 \times \cdots \times A_{n-1}}(x) \cdot \mu_n(A_n) = \sum_{i=1}^{\infty} \mathbb{1}_{A_{i,1} \times \cdots \times A_{i,n-1}}(x) \mu_n(A_{i,n}) \tag{2}$$

. To see this, if $x \in A_1 \times \cdots A_{n-1}$, LHS of the equation (2) is $\mu_n(A_n)$ and the RHS of the equation (2) is $\sum_{i \in \mathcal{I}} \mu_n(A_{in}) = \mu_n(A_n)$. Otherwise, both LHS and RHS of the equation (2) are 0.

5. By induction, $(\Omega_1 \times \cdots \times \Omega_{n-1}, \mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1}, \mu_1 \times \cdots \times \mu_{n-1})$ is a product measure space. Let $\mu' := \mu_1 \times \cdots \times \mu_{n-1}$.

$$\int \mathbb{1}_{A_1 \times \cdots \times A_{n-1}}(x) \mu_n(A_n) d\mu' = \int \sum_{i=1}^{\infty} \mathbb{1}_{A_{i,1} \times \cdots \times A_{i,n-1}}(x) \mu_n(A_{i,n}) d\mu'$$

from the last step. Then LHS of the equation (2) is $\mu_n(A_n)\mu'(A_1 \times \cdots \times A_{n-1}) = \mu_n(A_n)\mu_1(A_1) \times \mu_{n-1}(A_{n-1})$ and RHS $= \sum_{i=1}^{\infty} \mu_n(A_{i,n}) \cdot \mu'(A_{i,1} \times \cdots \times A_{i,n-1})$. Since LHS is equal to RHS from the last step, we have $\prod_i \mu_i(A_i) = \sum_{i=1}^{\infty} \prod_{j=1}^{n} \mu_i(A_{ij}) = \sum_{i=1}^{\infty} \mu(A_{i1} \times \cdots \times A_{in})$, implying the countable additivity.

□

References

Lecture 10: Fubini Theorem

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
April 24th, 2024

1 Famous quote from Fang today

"You will see it in 30 seconds. No, 15 seconds. Wait, just 1 second!!!" – Fang Han

2 Review

Given product space on $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2), \ldots, (\Omega_n, \mathcal{F}_n, \mu_n)$, we can construct $\Omega := \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} := \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, and $\mu := \mu_1 \times \cdots \times \mu_n$. $(\Omega, \mathcal{F})$ is called the *product measurable space*, $(\Omega, \mathcal{F}, \mu)$ is called the *product measure space*, and μ is called the product measure over $\mathcal{F}$.

We claimed that μ is the unique measure over $(\Omega, \mathcal{F})$ such that $\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mu_i(A_i)$ for any $A_i \in \mathcal{F}_i$.

3 Fubini Theorem

What would be the corresponding $\int f(w_1, w_2, \ldots, w_n) d\mu$ where $f : \Omega \to \mathbb{R}^*$ is measurable w.r.t. $\mathcal{F}$. Fubini's claim is about *how and when*

$$\int f(w_1, \ldots, w_n) d(\mu_1 \times \cdots \times \mu_n) = \int \int \cdots \int f(w_1, \ldots, w_n) d\mu_1 d\mu_2 \cdots d\mu_n.$$

Theorem 3.1 (Fubini-Tonelli theorem for Lebesgue measure)**.** *Given a finite collection of measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2), \ldots, (\Omega_n, \mathcal{F}_n, \mu_n)$ and $\Omega := \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} := \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, and $\mu := \mu_1 \times \cdots \times \mu_n$. Then,*

1. *It suffices to consider $n = 2$.*
2. *Assume $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ to be two σ-finite measure space.*
3. *Consider $(\Omega, \mathcal{F}, \mu)$ to be the product measure space.*
4. *Consider $f : \Omega_1 \times \Omega_2 \to \mathbb{R}^*$ to be measurable.*

We claim that if f is either nonnegative or integrable, then

1. *The following two functions $x \to \int_{\Omega_2} f(x, y) d\mu_2(y)$ and $y \to \int_{\Omega_1} f(x, y) d\mu_1(x)$ are well-defined and measurable,*
2. *(Fubini)*
$$\int_\Omega f(x, y) d\mu(x, y) = \int_{\Omega_1} \int_{\Omega_2} f(x, y) d\mu_2(y) d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} f(x, y) d\mu_1(x) d\mu_2(y). \tag{1}$$
3. *(Tonelli) If either $\int_{\Omega_1} \left[\int_{\Omega_2} |f(x, y)| d\mu_2(y)\right] d\mu_1(x)$ or $\int_{\Omega_2} \left[\int_{\Omega_1} |f(x, y)| d\mu_1(x)\right] d\mu_2(y)$ is finite, then f is integrable w.r.t. μ and we can apply Fubini to f.*

Proof. To prove the third claim above, assume the first and the second claims to be held. Then because $|f(x,y)| \geq 0$, the first and the second claims imply $\int_\Omega |f(x,y)| d\mu(x,y) = \int_{\Omega_1} [int_{\Omega_2} |f(x,y)| d\mu_2(y)] d\mu_1(x) < \infty$. This yields f is integrable.

The proof of the first claim is left as an exercise.

Step 1 To prove the second claim, we start to assume that f is a naive simple function: $f = \mathbb{1}_{A_1 \times A_2}(w)$ where $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, w \in \Omega_1 \times \Omega_2$. We assume $A_1 \times A_2$ is a rectangle in $\mathcal{F}$ here. The LHS of the equation (1) is $\int_{\Omega_1 \times \Omega_2} \mathbb{1}_{A_1 \times A_2}((x,y)) d\mu(x,y) = \mu(A_1 \times A_2) = \mu(A_1)\mu(A_2)$ by the definition of Lebesgue integration. The RHS of the equation (1) is

$$\begin{aligned}
\int_{\Omega_1} \left[\int_{\Omega_2} \mathbb{1}_{A_1 \times A_2}((x,y)) d\mu_2(y) \right] d\mu_1(x) &= \int_{\Omega_1} \left[\int_{\Omega_2} \mathbb{1}_{A_1}(x) \mathbb{1}_{A_2}(y) d\mu_2(y) \right] d\mu_1(x) \\
&= \int_{\Omega_1} [\mathbb{1}_{A_1}(x) \mu(A_2)] d\mu_1(x) \\
&= \mu_2(A_2) \int_{\Omega_1} \mathbb{1}_{A_1}(x) d\mu_1(x) \\
&= \mu_2(A_2)\mu_1(A_1) \\
&= \text{LHS}.
\end{aligned}$$

Step 2, consider an advanced simple function $f = \mathbb{1}_A(w)$ where A is an arbitrary element in $\mathcal{F}$. This can be proved through Dynkin's $\pi - \lambda$ theorem. Construct a class of sets $\mathcal{B} := \{A \in \mathcal{F} : \text{Fubini's claim is true for } \mathbb{1}_A((\cdot))\}$. Observe that $\mathcal{B} \subseteq \mathcal{F}$ and $\{\text{all rectangles in } \mathcal{F}\} \subseteq \mathcal{B}$ from step 1.

Next, we want to show that $\mathcal{B}$ is a λ-system. Hereafter, instead of assuming σ-finite, I'll assume μ_1 and μ_2 to be finite. To show the closeness under countable disjoint union, consider disjoint elements $B_1, B_2, \times \in \mathcal{B}$. The LHS is $\int_\Omega \mathbb{1}_B(\cdot) d\mu = \mu(B)$. The RHS is $\int_{\Omega_1} \left[\int_{\Omega_2} \mathbb{1}_B(\cdot) d\mu_2 \right] d\mu_1 = \int_{\Omega_1} \mu_2(B_x) d\mu_1$ where $B_x := \{y \in \Omega_2 : (x,y) \in \mathcal{B}\}$. We claim that $B_x = \bigcup_{n=1}^\infty (B_n)_x$ and $(B_n)_x$ for $n = 1, 2, \ldots$ are disjoint. So

$$\begin{aligned}
\int_{\Omega_1} \mu_2(B_x) d\mu_1 &= \int_{\Omega_1} \mu_2 \left(\bigcup_{n=1}^\infty (B_n)_x \right) d\mu_1 \\
&= \int_{\Omega_1} \sum_{n=1}^\infty \mu_2((B_n)_x) d\mu_1 \\
&= \sum_{n=1}^\infty \int_{\Omega_1} \mu_2((B_n)_x) d\mu_1 \\
&= \sum_{n=1}^\infty \int_{\Omega_1} \mathbb{1}_{B_n}(\cdot) d\mu \\
&= \sum_{n=1}^\infty \mu(B_n) \\
&= \mu(B) \\
&= \text{LHS}.
\end{aligned}$$

To show the closeness under complement, claim that $(B^{\mathsf{c}})_x = (B_x)^{\mathsf{c}}$, implying that $B \in \mathcal{B}$. Then the RHS of $\mathbb{1}_{B^{\mathsf{c}}}(\cdot)$ is

$$\begin{aligned}
\int_{\Omega_1} \mu_2((B^{\mathsf{c}})_x) d\mu_1 &= \int_{\Omega_1} \mu_2((B_x)^{\mathsf{c}}) d\mu_1 \\
&= \int_{\Omega_1} (\mu_2(\Omega_2) - \mu_2(B_x)) d\mu_1 \\
&= \mu_2(\Omega_2) \int_{\Omega_1} d\mu_1 - \int_{\Omega_1} \mu_2(B_x) d\mu_1(x) \\
&= \mu_2(\Omega_2) \int_{\Omega_1} d\mu_1 - \int_{\Omega} \mathbb{1}_B(\cdot) d\mu \\
&= \mu_2(\Omega_2) \int_{\Omega_1} d\mu_1 - \mu(B) \\
&= \mu(\Omega) - \mu(B) \\
&= \mu(B^{\mathsf{c}}) \\
&= \int_{\Omega} \mathbb{1}_{B^{\mathsf{c}}}(\cdot) d\mu \\
&= \text{LHS}.
\end{aligned}$$

Then, define $\mathcal{D} := \{\text{all rectanges in } \Omega\}$. Observe that $\mathcal{D}$ is a π-system, $\mathcal{D} \subseteq \mathcal{B}$, and $\sigma(\mathcal{D}) = \mathcal{F}$.

Finally, Dynkin's $\pi - \lambda$ theorem confirms that $\mathcal{F} \supseteq \mathcal{B} \supseteq \sigma(\mathcal{D}) = \mathcal{F}$, implying that $\mathcal{B} = \mathcal{F}$.

Step 3, consider a nonnegative simple function $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(w)$ w.r.t. $(\Omega, \mathcal{F})$. Assume the step 2 holds. We have

$$\begin{aligned}
\text{LHS} &= \int_{\Omega} \left(\sum_{i=1}^n a_i \mathbb{1}_{A_i}(\cdot) \right) d\mu \\
&= \sum_{i=1}^n \left(a_i \int_{\Omega} \mathbb{1}_{A_i}(\cdot) d\mu \right) \because \text{LI is linear} \\
&= \sum_{i=1}^n a_i \int_{\Omega_1} \left[\int_{\Omega_2} \mathbb{1}_{A_i}(\cdot) d\mu_2 \right] d\mu_1 \because \text{Step 2} \\
&= \int_{\Omega_1} \left[\int_{\Omega_2} \sum_{i=1}^n a_i \mathbb{1}_{A_i}(\cdot) d\mu_2 \right] d\mu_1 \because \text{LI is linear} \\
&= \int_{\Omega_1} \left[\int_{\Omega_2} f d\mu_2 \right] d\mu_1 \\
&= \text{RHS}.
\end{aligned}$$

Step 4, consider a nonnegative measurable function f. With MCT and Prop 2.3.6.,

$$\begin{aligned}\text{LHS} &= \int_\Omega f(x,y)d\mu \\ &= \lim_{n\to\infty}\int_\Omega f_n d\mu \because \text{MCT and Prop 2.3.6.} \\ &= \lim_{n\to\infty}\left[\int_{\Omega_1}\left[\int_{\Omega_2} f_n d\mu_2\right]d\mu_1\right] \because \text{step 3} \\ &= \int_{\Omega_1}\left[\int_{\Omega_2}\lim f_n d\mu_2\right]d\mu_1 \because \text{MC} \\ &= \text{RHS.}\end{aligned}$$

Step 5, consider an integrable function f. We have

$$\begin{aligned}\text{LHS} &= \int_\Omega f d\mu \\ &= \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu \because \text{definition of LI} \\ &= \int_{\Omega_1}\int_{\Omega_2} f^+ d\mu_2 d\mu_1 - \int_{\Omega_1}\int_{\Omega_2} f^- d\mu_2 d\mu_1 \because \text{step 4} \\ &= \int_{\Omega_1}\int_{\Omega_2} f^+ - f^- d\mu_2 d\mu_1 \\ &= \text{RHS.}\end{aligned}$$

□

References

Lecture 11: L^p space I

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 1st, 2024

1 Inequalities

Theorem 1.1 (Markov inequality). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and a measurable function $f : \Omega \to [0, \infty]$. Then for any $t > 0$,*

$$\mu(\{\omega \in \Omega : f(\omega) \geq t\}) \leq \frac{1}{t} \int_\Omega f d\mu.$$

Remark 1.2. In probability theory, Markov inequality is stated as $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ for any $t > 0$.

Theorem 1.3 (Jensen's inequality). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}^*$ be measurable and integrable. Let $I \supseteq Range(f)$ be a interval in $\mathbb{R}^*$ such that $\phi : I \to \mathbb{R}$ is convex. Then*

$$\phi\left(\int_\Omega f d\,\mathbb{P}\right) \leq \int_\Omega (\phi \circ f) d\,\mathbb{P}\,.$$

Theorem 1.4 (Young's inequality). *If $x, y > 0$ and (p, q) is a natural couple, i.e., $p, q < 1$, $p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Theorem 1.5 (Holder's inequality). *For any measurable function $f, g : \Omega \to \mathbb{R}^*$ and any $p \in [1, \infty]$,*

$$||f \cdot g||_{L^1} \leq ||f||_{L^p} \cdot ||g||_{L^q}$$

whenever $\frac{1}{p} + \frac{1}{q} = 1$.

Theorem 1.6 (Minkowski inequality). *For any $p \in [1, \infty]$ and any f, g satisfy $f, g : \Omega \to \mathbb{R}^*$, $||f||_{L^p}, ||g||_{L^p} < \infty$, and measurable,*

$$||f + g||_{L^p} \leq ||f||_{L^p} + ||g||_{L^p}.$$

Proof. It needs Jensen's and Holder's inequality.

Suppose $p = 1$. The LHS is $\int |f + g| d\mu \leq \int (|f| + |g|) d\mu = \int |f| d\mu + \int |g| d\mu$ which is RHS.

Suppose $p = \infty$,then $\operatorname{ess\,sup} |f(\omega) + f(\omega)| \leq \operatorname{ess\,sup} |f(\omega)| + \operatorname{ess\,sup} |g(\omega)|$.

Suppose $p \in (1, \infty)$. Then,

$$
\begin{aligned}
(\text{LHS})^p &= \int |f+g|^p d\mu \\
&= \int |f+g| \cdot |f+g|^{p-1} d\mu \\
&\le \int |f| \cdot |f+g|^{p-1} d\mu + \int |g| \cdot |f+g|^{p-1} d\mu \\
&\le ||f||_{L^p} \cdot \left|\left||f+g|^{p-1}\right|\right|_{L^q} + ||g||_{L^p} \cdot \left|\left||f+g|^{p-1}\right|\right|_{L^q} && \because \text{Holder's inequality} \\
&= ||f||_{L^p} \cdot ||f+g||_{L^p}^{\frac{p}{q}} + ||g||_{L^p} \cdot ||f+g||_{L^p}^{\frac{p}{q}} && \because \frac{1}{p} + \frac{1}{q} = 1.
\end{aligned}
$$

Thus, $||f+g||_{L^p}^{p-\frac{p}{q}} \le ||f||_{L^p} + ||g||_{L^p}$. Since $p - \frac{p}{q} = 1$, we have the desired result. □

2 L^p space

Definition 2.1 (L^p norm)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}^*$ be a measurable function. Then for any $p \in [1, \infty]$, the L^p norm of $f(\cdot)$ is defined to be

1. If $p \in [1, \infty)$, $||f||_{L^p} := (\int |f|^p d\mu)^{\frac{1}{p}}$,
2. If $p = \infty$, $||f||_{L^p} := \inf\{K \in [0, \infty] : |f| \le K \; a.e.\}$. This is also called "essential supremum".

Definition 2.2 (L^p space of measurable function)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. The space, $L^p(\Omega, \mathcal{F}, \mu)$, is defined to be $\{f : \Omega \to \mathbb{R}^* : f$ is meausrable and $||f||_{L^p} < \infty\}$.

Theorem 2.3 (L^p space)**.** *We have the following claims:*

1. *Any L^p function space is a vector space equipped with a norm, $||\cdot||_{L^p}$, i.e., L^p is a normed vector space.*
2. *This L^p space is a complete function space w.r.t. $||\cdot||_{L^p}$. i.e.,*

   (a) (Triangular inequality) L^p norm is indeed a norm; in particular, for any $f, g \in L^p(\Omega, \mathcal{F}, \mu)$, $||f+g||_{L^p} \le ||f||_{L^p} + ||g||_{L^p}$.

   (b) (Completeness) Any sequence of functions that is Cauchy w.r.t. L^p norm will converge to limit w.r.t. L^p norm. Cauchy means a sequence of functions $\{f_n\}$ is so that for any $\epsilon > 0$, there exists N_ϵ such that for any $n, m > N_\epsilon$, $||f_n - f_m||_{L^p} \le \epsilon$.

If these two claims are true, then $L^p(\Omega, \mathcal{F}, \mu)$ is a complete normed vector space, i.e., a Banach space.

Proof. The first claim is proved by Minkowski's inequality.

We will prove the second claim in the next lecture. □

References

Lecture 12: L^p space II

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 8th, 2024

1 Review

Recall the definition

Definition 1.1 (L^p space)**.** The space $L^p(\Omega, \mathcal{F}, \mu)$ is $\{f : \Omega \to \mathbb{R}^*; \text{measurable and } ||f||_{L^p} < \infty\}$.

Theorem 1.2. *Any L^p space, coupled with $||\cdot||_{L^p}$, is a normed complete space, i.e., Banach space.*

Theorem 1.3 (Minkowski inequality)**.** *For any $p \in [1, \infty]$ and any f, g satisfy $f, g : \Omega \to \mathbb{R}^*$, $||f||_{L^p}, ||g||_{L^p} < \infty$, and measurable,*

$$||f + g||_{L^p} \leq ||f||_{L^p} + ||g||_{L^p}.$$

Theorem 1.4 (Holder's inequality)**.** *For any measurable function $f, g : \Omega \to \mathbb{R}^*$ and any $p \in [1, \infty]$,*

$$||f \cdot g||_{L^1} \leq ||f||_{L^p} \cdot ||g||_{L^q}$$

whenever $\frac{1}{p} + \frac{1}{q} = 1$.

Remark 1.5. When $p = q = 2$, we obtain Cauchy-Schwartz

$$||f \cdot g||_{L^1} \leq ||f||_{L^2} \cdot ||g||_{L^2}.$$

For example, $\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$.

Proof. If $p = 1, q = \infty$, then

$$\begin{aligned}
\text{LHS} &= \int |f \cdot g| d\mu \\
&= \int |f||g| d\mu \\
&\leq \int |f| ||g||_{L^\infty} d\mu \\
&= ||g||_{L^\infty} \int |f| d\mu \\
&= \text{RHS}
\end{aligned}$$

If $p = \infty, q = 1$, we have the symmetric argument.

If $p, q \in (1, \infty)$, we will use Young's inequality. Let $u := \frac{f}{||f||_{L^p}}$ and $v := \frac{g}{||g||_{L^q}}$ assuming that $||f||_{L^p} < \infty$ and $||g||_{L^q} < \infty$. Notice that $||u||_{L^p} = ||v||_{L^q} = 1$. Also, for any $\omega \in \Omega$,

$$|u(\omega)v(\omega)| = ||u(\omega)| \cdot |v(\omega)|| \leq \frac{|u(\omega)|^p}{p} + \frac{|v(\omega)|^q}{q}$$

by Young's inequality. This implies that

$$\begin{aligned}\int |uv| d\mu &\leq \frac{\int |u(\omega)|^p d\mu(\omega)}{p} + \frac{\int |v(\omega)|^p d\mu(\omega)}{q} \\ &\leq \frac{||u||_{L^p}^p}{p} + \frac{||v||_{L^q}^q}{q} \\ &= \frac{1}{p} + \frac{1}{q} \\ &= 1,\end{aligned}$$

so $\int |fg| d\mu \leq ||f||_{L^p} ||g||_{L^q}$. □

Now, we want to show the completeness of L^p norm. That is, any sequence of functions that is Cauchy in the L^p norm converges to a limit in L^p space. We start by showing the following lemma.

Lemma 1.6. *Consider $\{f_n\}$ to be a Cauchy sequence w.r.t. L^p norm:*

$$\forall \epsilon > 0, \quad \exists N_q > 0 \quad s.t. \quad \forall n, m \geq N_q, \quad ||f_n - f_m||_{L^p} \leq \epsilon.$$

We claim that $\exists$ some $f \in L^p(\Omega, \mathcal{F}, \mu)$ and $\exists$ a subsequence $\{f_{n_k}\}_{k=1,2,\ldots}$ such that $f_{n_k} \to f$ a.e. as $k \to \infty$.

Proof. The idea is to construct an $\{n_k\}$ and use Borel-Caretalli 1st lemma. We follows these steps:

1. Since $f_{n_k} \to f$ a.e. as $k \to \infty$, $\exists B \in \mathcal{F}$ such that $\mu(B^c) = 0$ and $\forall \omega \in B$, $f_{n_k}(\omega) \to f(\omega)$ as $k \to \infty$.
2. By the Cauchy sequence condition, for any $k = 1, 2, 3, 4, \ldots$, pick $\epsilon_k := 2^{-k}$, $N_k = N_{\epsilon_k}$, $n_k = N_k + 1$. Then by definition, for $k = 1, 2, \ldots$, $||f_{n_k} - f_{n_{k+1}}||_{L^p} \leq \epsilon_k \Big(= 2^{-k}\Big)$.
3. Defien $A_k := \Big\{\omega \in \Omega : \Big|f_{n_k(\omega)} - f_{n_{k+1}}(\omega)\Big| \geq 2^{-\frac{k}{2}}\Big\}$. Then by Markov inequality,

$$\begin{aligned}\mu(A_k) &\leq \frac{\int |f_{n_k} - f_{n_{k+1}}|^p d\mu}{2^{-\frac{kp}{2}}} \\ &= 2^{\frac{kp}{2}} ||f_{n_k} - f_{n_{k+1}}||_{L^p}^p \\ &\leq 2^{\frac{kp}{2}} \epsilon_k^p \\ &= 2^{\frac{kp}{2}} 2^{-kp} \\ &= 2^{-\frac{kp}{2}}.\end{aligned}$$

   This implies that $\sum_{k=1}^{\infty} \mu(A_k) \leq \sum_{k=1}^{\infty} 2^{-\frac{kp}{2}} < \infty$.
4. 1st (B-C) lemma: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $A_1, A_2, \cdots \in \mathcal{F}$ satisfy $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. Then, $\mu(\{\omega \in \Omega : \omega$ is in infinitely many A_n's$\}) = 0$, i.e., $\{\omega \in \Omega : \omega$ is in infinitely many A_n's$\}$ is dominating Ω.

5. Define $B := \{\omega \in \Omega : \omega \text{ is only in finitely many } A_k\text{'s}\}$ where $A_k = \left\{\omega : |f_{n_k}(\omega) - f_{n_{k+1}}(\omega)| \geq 2^{-\frac{k}{2}}\right\}$. Using 1st (B-C) lemma, we have $\mu(B^{\mathsf{c}}) = 0$. In other words, if $\omega \in B$, for all sufficiently large k, we have

$$|f_{n_k}(\omega) - f_{n_{k+1}}(\omega)| \leq 2^{-\frac{k}{2}}$$

implying that

$$\forall \omega \in B, \quad \{f_{n_k}(\omega)\} \text{ is a Cauchy Real sequence.}$$

implying that

$$\text{Sicne the real space is complete, we have} \quad \forall \omega \in B, f_{n_k}(\omega) \to \text{a limit as } k \to \infty$$

implying that

$$\exists \text{ a measurable function } f \quad s.t. \quad f_{n_k} \to f a.e.$$

6. It remains to show $f \in L^p(\Omega, \mathcal{F}, \mu)$. i.e., $||f||_{L^p} < \infty$. Using Fatou's lemma,

$$\begin{aligned} \int \liminf |f_{n_k}|^p d\mu \leq \liminf \int |f_{n_k}|^p d\mu &\implies \int |f|^p d\mu \leq \liminf \int |f_{n_k}|^p d\mu \\ &\implies ||f||_{L^p}^p \leq \liminf ||f_{n_k}||_{L^p}^p < \infty. \end{aligned}$$

□

Theorem 1.7 (Riesz-Fischer). *$L^p(\Omega, \mathcal{F}, \mu)$ is complete.*

Proof. Given $p \in [1, \infty)$. Then, fixing an arbitrary $\epsilon > 0$, there exists N_ϵ such that for any $n, m \geq N_\epsilon$, $||f_n - f_m||_{L^p} \leq \epsilon$. The previous lemma then states there exists $\{n_k\}_{k=1,2,\ldots}$ and there exists a measurable $f \in L^p(\Omega, \mathcal{F}, \mu)$ such that $f_{n_k} \overset{k\to\infty}{\to} f$ a.e. Then, for any $n \geq N_\epsilon$,

$$\begin{aligned} \int |f_n - f|^p d\mu &= \int \lim_{k\to\infty} |f_n - f_{n_k}|^p d\mu \\ &\leq \liminf \int |f_n - f_{n_k}|^p d\mu \qquad \because \text{Fatou's lemma} \quad = \liminf_{k\to\infty} ||f_n - f_{n_k}||_{L^p}^p \\ &\leq \epsilon^p \end{aligned}$$

for any k large enough. This means LHS $= ||f_n - f||_{L^p}^p \leq \epsilon^p \implies ||f_n - f||_{L^p} \to 0$ as $n \to \infty \implies f_n \to f$ a.e. The case when $p = \infty$ is left to prove. In conclusion, L^p space is complete. □

Theorem 1.8 (Lyapunov theorem). *Consider $(\Omega, \mathcal{F}, \mathbb{P})$ to be a probability space. Let $f : \Omega \to \mathbb{R}^*$ be measurable. Then as long as $p \leq q \in [1, \infty]$, then $||f||_{L^p} \leq ||f||_{L^q}$. i.e., $||f||_{L^p}$ increasing w.r.t. p.*

Proof. It is based on Jensen by realizing

$$\phi : x \to |x|^{\frac{q}{p}}$$

is convex when $1 \le p \le q < \infty$. Then

$$\begin{aligned} ||f||_{L^p}^q &= \left(\int |f|^p d\mu \right)^{\frac{q}{p}} \\ &= \phi\left(\int |f|^p d\mu \right) \\ &\le \int \phi(|f|^p) d\mu \\ &= \int |f|^{p \cdot \frac{q}{p}} d\mu \\ &= \int |f|^q d\mu \\ &\le ||f||_{L^q}^q. \end{aligned}$$

So $||f||_{L^p} \le ||f||_{L^q}$. If, on the other hand, $q = \infty$, $||f||_{L^p} = (\int |f|^p d\mu)^{\frac{1}{p}} \le (||f||_{L^\infty}^p \int 1 d\mu)^{\frac{1}{p}} = ||f||_{L^\infty}$. If $p = q = \infty$, there is nothing to prove. □

References

Lecture 13: Random variable, CDF and pdf, expectation, independence

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 10th, 2024

Starting today, we will talk about probability theory built on measure theory.

1 Random variable

Definition 1.1 (Random variable)**.** A *random variable* is a measurable function $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\Omega', \mathcal{F}')$ where $\mathbb{P}(\Omega) = 1$ and Ω' is an object space.

Remark 1.2.

- In this class, we focused on real-valued random variables $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- $\mathbb{P}(X \in A)$ for some $A \in \mathcal{B}(\mathbb{R})$ is defined to be $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$.
- If there are two random variables X, Y defined on some $(\Omega, \mathcal{F}, \mathbb{P})$ then

$$\mathbb{P}(\{X \in A, Y \in B\}) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A, Y(\omega) \in B\}).$$

- For any $f : \mathbb{R} \to \mathbb{R}$, $f(X) := f \circ X$. e.g., if $f(x) = x^2$, then $f(X) = X^2$ i.e., $f(X(\omega)) = (X(\omega))^2$.
- The "σ-algebra generated by a random variable X" is defined to be $\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R}))$. i.e. $\sigma(X)$ is *the smallest σ-algebra* in Ω such that X is still measurable.
- If $\{X_i\}_{i \in \mathcal{I}}$ ($\mathcal{I}$ is a general index set NOT necessarily countable) is a collection of RVs defined on the same probability space: $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then the σ-algebra generated by $\{X_i\}_{i \in \mathcal{I}}$ is defined to be $\sigma\big(\{X_i\}_{i \in \mathcal{I}}\big) := \sigma(\bigcup_{i \in \mathcal{I}} \sigma(X_i))$.

1.1 Cumulative distribution function

Definition 1.3 (Cumulative distribution function (CDF))**.** *Cumulative distribution function (CDF)* w.r.t. a RV X is defined as $F_X : \mathbb{R} \to [0, 1]$ such that $F_X(t) := \mathbb{P}(X \le t) = \mathbb{P}(\{\omega : X(\omega) \le t\})$.

Theorem 1.4. *Suppose $F : \mathbb{R} \to [0, 1]$ such that it is non-decreasing, right-continuous, and $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$. Then $\exists$ a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a RV X over it such that F is the CDF of X. Conversely, if X is a RV on $(\Omega, \mathcal{F}, \mathbb{P})$, then F_X satisfies the properties above.*

Proof. Define $(\Omega, \mathcal{F}, \mathbb{P}) := \Big([0, 1], \mathcal{B}(\mathbb{R}) \cap [0, 1], \lambda|_{[0,1]}\Big)$ and $X(\omega) := \inf\{t \in \mathbb{R} : F(t) \ge \omega\}$. The X is called the generalized inverse of F (F^{-}) or quantile transformation. It can be verified that X indeed has CDF = F. Then by 512 knowledge, we can complete the proof. $\square$

1.2 The law of a random variable

Definition 1.5 (The law of random variable)**.** The *law* of a random variable X is the induced probability measure, denoted as $\mu_X(\mathbb{P}_X)$, over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu_X(A) = \mathbb{P}(X \in A)$

Remark 1.6. If X and Y have the same law, and $g : \mathbb{R} \to \mathbb{R}$ is a measurable function, then $g(X)$ and $g(Y)$ also have the same law.

Theorem 1.7. *Two random variables have the same CDF if and only if they have the same law.*

Proof. The backward direction is trivial. To prove the forward direction, suppose $F_X = F_Y$. Then $\forall t \in \mathbb{R}$, $\mathbb{P}(X \leq t) = \mathbb{P}(Y \leq t)$ or $\mu_X((-\infty, t]) = \mu_Y((-\infty, t])$. Then, by the uniqueness extension theorem, $\mu_X = \mu_Y$ over $\sigma(\{(-\infty, t]; t \in \mathbb{R}\})$. □

1.3 Probability density function

The probability density function is usually introduced using Radon-Nykodym theorem plus Lebesgue decomposition theorem, and is the Radon-Nykodym derivative between two laws.

Definition 1.8 (Probability density function (PDF))**.** Suppose we have a function $f : \mathbb{R} \to \mathbb{R}$ such that it is nonnegative, integrable, $\int_{\mathbb{R}} f(x) d\lambda(x) = 1$. Then it defines a probability measure over $(\mathbb{R}, \mathcal{B}(R))$:

$$\nu(A) := \int_A f(x) d\lambda(x).$$

The function f is said to be the pdf of $\nu(\cdot)$, which is the law of a continuous RV.

Theorem 1.9. *A function f is the pdf of a RV X (i.e., f is the pdf of μ_X) if and only if $\forall A = [a, b]$ with a, b to be any continuity points of F_X, we have $\mu_X(A) = \int_A f(x) dx$.*

Proof. The proof is left as a homework problem. □

Theorem 1.10. *If f, g corresponds to the same law, then $f = g$ a.e.*

Theorem 1.11. *If f is the pdf, then $F(t) := \int_{-\infty}^{t} f(y) d\lambda(y)$ is the CDF of the law that f corresponds to. Conversely, if F is a CDF on $\mathbb{R}$ for which there exists a nonnegative measurable function f satisfying $F(t) := \int_{-\infty}^{t} f(y) d\lambda(y)$ for all t, then f is a pdf generating the probability measure corresponding to F.*

2 Expectation

Definition 2.1 (Expectation)**.** For random variblae $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the expectation of X is defined as

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

provided that the integral is well-defined.

Theorem 2.2 (The unconscious statistician theorem). *If the RV X has the law μ_X, then, for any measurable function $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We claim that*

$$\begin{aligned}\mathbb{E}[g(X)] &= \int_\Omega g(X(\omega))d\mathbb{P}(\omega) \\ &= \int_\mathbb{R} g(x)d\mu_X(x) \\ &= \int_\mathbb{R} g(x)f(x)d\lambda(x).\end{aligned}$$

Proof. We divide the proof by cases: super simple function, NSF, nonnegative function, and general function. □

Remark 2.3. If $X \geq 0$, then $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq t)d\lambda(t)$.

The definitions of *variance, covariance, moment generating function, characteristic function* can be defined based on the expectation. See details in the textbook.

3 Independence

Definition 3.1 (Independence of σ-algebras). Consider $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n \subseteq \mathcal{F}$ to be a *sub*-σ-algebras of $\mathcal{F}$. Then we say $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if for any $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$, we have

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i).$$

Definition 3.2 (General definition of independence). Any collection $\{\mathcal{F}_i\}_{i\in\mathcal{I}}$ with $\mathcal{F}_i \subseteq \mathcal{F}$ is said to be independent if *any finitely many* of them is independent.

Definition 3.3 (Independence of sets). Any collection $\{A_i\}_{i\in\mathcal{I}}$ with $A_i \in \mathcal{F}$ *is said to be independent* if *ANY FINITELY* many $\sigma(\{A_i\})$ are independent.

Definition 3.4 (Definition of independent of RVs). Any collection $\{X_i\}_{i\in\mathcal{I}}$ is said to be independent if $\{\sigma(X_i)\}_{i\in\mathcal{I}}$ are independent.

References

Lecture 14: Four notions of convergence I

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 15th, 2024

1 Independence

Definition 1.1 (Independence of σ-algebras)**.** Consider $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n \subseteq \mathcal{F}$ to be a *sub*-σ-algebras of $\mathcal{F}$. Then we say $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if for any $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$, we have

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

Definition 1.2 (General definition of independence)**.** Any collection $\{\mathcal{F}_i\}_{i\in\mathcal{I}}$ with $\mathcal{F}_i \subseteq \mathcal{F}$ is said to be independent if *any finitely many* of them is independent.

Definition 1.3 (Independence of sets)**.** Any collection $\{A_i\}_{i\in\mathcal{I}}$ with $A_i \in \mathcal{F}$ *is said to be independent* if $ANY\ FINITELY$ many $\sigma(\{A_i\})$ are independent.

Definition 1.4 (Definition of independent of RVs)**.** Any collection $\{X_i\}_{i\in\mathcal{I}}$ is said to be independent if $\{\sigma(X_i)\}_{i\in\mathcal{I}}$ are independent.

Lemma 1.5 (Borel-Contalli)**.**

1. *As long as $\{A_n\}_{n=1}^{\infty}$ satisfies $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ (no conditions on $\mathbb{P}$ or A_i's), then $\mathbb{P}(\{w : w \text{ is in infinitely many } A_n\text{'s}\}) = 0$.*
2. *If $\{A_n\}_{n=1}^{\infty}$ are independent, (i.e., any finitely many $\sigma(A_i)$ are independent), then as long as $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\{w : w \text{ is in infinitely many } A_n\text{'s}\}) = 1$*

2 Four notions of convergence of RVs

Definition 2.1 (Almost surely convergence)**.** A sequence of RVs $\{X_n\}_{n=1}^{\infty}$ is said to be *converging almost surely* to another RV X ($X_n \to X$ a.e.) if

1. All RVs, $\{X_n\}$ and X, are over the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$;
2. $\lim_{n\to\infty} X_n = X$ a.e. ($\lim_{n\to\infty} f_n = f, a.e.$).

Definition 2.2 (Converging in probability)**.** A sequence of RVs $\{X_n\}_{n=1}^{\infty}$ is said to be *converging in probability* to another RV X ($X_n \xrightarrow{p} X$) if

1. All RVs, $\{X_n\}$ and X, are over the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$;
2. For any $\epsilon > 0$, we have $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$.

Definition 2.3 (Converging in L^p norm)**.** Consider $p \in [1, \infty]$. A sequence of RVs $\{X_n\}_{n=1}^{\infty}$ is said to be *converging in L^p norm* to another RV X ($X_n \xrightarrow{L^p} X$) if

1. All RVs, $\{X_n\}$ and X, are over the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$;

2. $||X_n - X||_{L^p} \to 0$ as $n \to \infty$.

Definition 2.4 (Converging in distribution/weak convergence)**.** A sequence of RVs $\{X_n\}_{n=1}^{\infty}$, each of a CDF $\{F_n\}$, is said to be *converging in distribution or weakly convergent* to another RV X, of a CDF F, ($X_n \xrightarrow{d} X$) if for any continuity points t of F, we have $\lim_{n\to\infty} F_n(t) = F(t)$.

3 Relation between the 4 notes of convergence

Theorem 3.1. *Convergence almost surely implies convergence in probability.*

Proof. Given $X_n \xrightarrow{a.s.} X$, we have

$$\mathbb{P}(\lim X_n = X) = 1 \iff \mathbb{P}(\forall \epsilon > 0, \quad \exists \quad n_\epsilon, \quad s.t. \quad \forall n > n_\epsilon, \quad |X_n - X| \leq \epsilon) = 1.$$

By fixing $\epsilon > 0$, we obtain

$$\begin{aligned}
1 &= \mathbb{P}(\exists n, \quad s.t. \quad \forall k \geq n, \quad |X_k - X| \leq \epsilon) \\
&= \mathbb{P}(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k - X| \leq \epsilon\}) \\
&= \mathbb{P}(\cup_{n=1}^{\infty} A_n) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n) \qquad \because A_n \text{ is non-decreasing} \\
&= \lim_{n\to\infty} \mathbb{P}(\cap_{k=n}^{\infty}\{|X_k - X| \leq \epsilon\}) \\
&\leq \lim_{n\to\infty} \mathbb{P}(|X_n - X| \leq \epsilon) \\
&\leq 1.
\end{aligned}$$

So $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \leq \epsilon) = 1$. □

Theorem 3.2 (Riesz)**.** *Given $X_n \xrightarrow{p} X$, there exists a subsequence $\{n_k\}_{k\geq 1}$ such that $X_{n_k} \xrightarrow{a.s.} X$.*

Proof. It is similar to the proof of Riesz-Fischer theorem (i.e., L^p space is complete. Given $X_n \xrightarrow{p} X$, we have for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ which means for any $\epsilon > 0$ and any $\delta > 0$, there exists n such that for any $K > n$, $\mathbb{P}(|X_k - X| > \delta)$. Now choose $\epsilon = 2^{-k-1}$ and $\delta = 2^{-k}$, then there exists n_k such that for any $m > n_k$, $\mathbb{P}\Big(|X_m - X| > 2^{-k-1}\Big) \leq 2^{-k}$. We can similarly pick n_k and n_{k+1} such that $\mathbb{P}\Big(|X_{n_k} - X| > 2^{-K-1}\Big) < 2^{-k}$ and $\mathbb{P}\Big(|X_{n_{k+1}} - X| > 2^{-k-1}\Big) < 2^{-k}$. It means $\mathbb{P}\Big(|X_{n_{k+1}} - X_{n_k}| > 2^{-k-1}2\Big) \leq 2 \cdot 2^{-k}$. By the proof of L^p completeness theorem, we established $\{X_{n_k}(\omega)\}_{k=1}^{\infty}$ is a Cauchy sequence for almost all ω. Thus, there exists some Y such that $X_{n_k}(\omega) \to Y(\omega)$ a.e. (i.e. $X_{n_k} \xrightarrow{a.e.}$ some RV Y. It remains to show $X = Y$ a.e. If $X_n \xrightarrow{p} X$ and $X_{n_k} \xrightarrow{a.s.} Y$, then $X = Y$ a.s., completing the proof. □

Theorem 3.3. *Convergence in probability implies convergence in distribution.*

Proof. Fix an $\epsilon > 0$. Then $F_{X_n}(t) := \mathbb{P}(X_n \leq t) = \mathbb{P}(X_n - X + X \leq t)$. Observe that $\{X_n - X + X \leq t\} \subseteq \{X \leq t + \epsilon\} \cup \{|X_n - X| > \epsilon\}$. So

$$\mathbb{P}(X_n - X + X \leq t) \leq \mathbb{P}(X \leq t + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

and $F_{X_n}(t) \leq F_X(t + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon) \implies \limsup F_{X_n}(t) \leq \limsup F_X(t + \epsilon) + \limsup \mathbb{P}(|X_n - X| > \epsilon) = \limsup F_X(t + \epsilon) = F_X(t + \epsilon)$. By taking $\epsilon \to 0$, we have $\limsup F_{X_n}(t) \leq F_X(t)$. Symmetrically, picking $\epsilon < 0$, we have $\liminf F_{X_n}(t) \geq F_X(t)$, implying that for any continuous point t of F_X, $\lim F_{X_n}(t) = F_X(t)$ (i.e., $X_n \xrightarrow{d} X$.). □

Remark 3.4. The theorem is not correct if we ask for all t including discontinuity points of F_X.

Theorem 3.5. *If $X_n \xrightarrow{d} c$ where c is a constant, then $X_n \xrightarrow{p} c$.*

References

Lecture 15: Four notions of convergence II

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 17th, 2024

1 Finishing convergence theorems

Theorem 1.1 (DCT)**.** *Assume*

- $\{X_n\}, X$ *from* $(\Omega, \mathcal{F}, \mathbb{P})$ *to* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,
- $|X_n| \leq Y$ *a.s. (domination condition),*
- $\mathbb{E}[Y] < \infty$ *(*Y *is in* $L^1(\mathbb{P})$ *space,*
- $X_n \xrightarrow{p} X$.

Then $\mathbb{E}[|X_n - X|] \to 0$ *as* $n \to \infty$ *implies* $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Proof. Suppose $X_n \xrightarrow{p} X$. Then take $Z_n := |X_n - X|$ and $a := \limsup_{n\to\infty} \mathbb{E}[Z_n] \geq 0$. Because $a = \limsup \mathbb{E}[Z_n]$ by mathematical analysis, there exists a subsequence $\{n_k\}$ such that $\mathbb{E}[Z_{n_k}] \to a$. Since $Z_n \xrightarrow{p} 0$ as $X_n \xrightarrow{p} X$, $Z_{n_k} \xrightarrow{p} 0$, and there exists a further subsequence $n_{k'}$ of n_k such that $Z_{n_{k'}} \xrightarrow{a.s.} 0$ and $\mathbb{E}\left[Z_{n_{k'}}\right] \to a$. However, by DCT $\mathbb{E}\left[Z_{n_{k'}}\right] \to 0$, so $a = 0$. Thus $0 = \limsup \mathbb{E}[Z_n] \geq \liminf \mathbb{E}[Z_n] \geq 0$, so $\lim \mathbb{E}[Z_n] = 0$. □

Theorem 1.2 (Slutsky's theorem)**.** *If* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$ *for some constant* c*, then* $X_n + Y_n \xrightarrow{d} X + c$ *and* $X_n Y_n \xrightarrow{d} cX$.

Theorem 1.3. *Convergence in* L^p *norm implies convergence in probability.*

Proof. For any $\epsilon > 0$, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) &= \lim_{n\to\infty} \mathbb{P}(|X_n - X|^p > \epsilon^p) \\
&\leq \lim_{n\to\infty} \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p} \\
&= \lim_{n\to\infty} \frac{||X_n - X||_{L^p}^p}{\epsilon^p} \\
&= 0
\end{aligned}$$

□

Theorem 1.4. *Convergence in probability implies convergence in* L^p *norm provided that* $\sup|X_n| \leq c < \infty$ *a.s.*

Proof. Use the "truncation trick". Observe that

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}[|X_n-X|^p] &= \lim_{n\to\infty}\mathbb{E}\Big[|X_n-X|^p\mathbb{1}_{(\epsilon,\infty)}(|X_n-X|^p)\Big]+\mathbb{E}\Big[|X_n-X|^p\mathbb{1}_{(-\infty,\epsilon)}(|X_n-X|^p)\Big]\\ &\le \lim_{n\to\infty}(2c)^p\cdot\mathbb{P}(|X_n-X|^p>\epsilon)+\epsilon^p\\ &\le \epsilon^p.\end{aligned}$$

Thus, $\lim_{n\to\infty}\mathbb{E}[|X_n-X|^p]=0$. □

Remark 1.5. This bound is loose and the uniform boundedness assumption is strong.

2 Uniform Integrability

A much better condition than the *uniform boundeness one* is Vitalli's uniform integrability (u.i.)

Definition 2.1 (u.i.)**.** A sequence of RVs $\{X_n\}$ is said to u.i. if for any $\epsilon>0$, there exists $K=K_\epsilon>0$ such that $\sup_n\mathbb{E}[|X_n|\mathbb{1}(|X_n|>K_\epsilon)]\le\epsilon$.

Theorem 2.2 (Vitalli's L^1 convergence theorem)**.** *Given uniform integrable and convergences a.s. or convergence in probability, we have* $X_n\xrightarrow{L^1}X$.

Proof. Consider another type of truncation function

$$\phi_K(x)=\begin{cases}-K & x\in(-\infty,-K)\\ x & x\in[-K,-K]\\ K & x\in(K,\infty)\end{cases}.$$

Observe that

$$\begin{aligned}\mathbb{E}[|X_n-X|] &= \mathbb{E}[X_n-\phi_K(X_n)+\phi_K(X_n)-\phi_K(X)+\phi_K(X)-X]\\ &\le \mathbb{E}[|X_n-\phi_K(X_n)|]+\mathbb{E}[|\phi_K(X_n)-\phi_K(X)|]+\mathbb{E}[|\phi_K(X)-X|]\end{aligned}$$

The first term is

$$\begin{aligned}&\mathbb{E}\Big[(X_n-K)\mathbb{1}_{(K,\infty)}(X_n)+0\mathbb{1}_{(-\infty,K)}(|X_n|)+(X_n-(-K))\mathbb{1}_{(-\infty,-K)}(X_n)\Big]\\ &\le\sup_n\mathbb{E}\Big[|X_n|\cdot\mathbb{1}_{(K,\infty)}(|X_n|)\Big]\\ &\le\epsilon\end{aligned}$$

by choosing $K>K_\epsilon$. The second term has the property that $\phi_K(X_n)\xrightarrow{L^1}\phi_K(X)$ as $X_n\xrightarrow{p}X$ and ϕ_K is continuous implying $\phi_K(X_n)\xrightarrow{p}\Phi_K(X)$. For the third term, by u.i. and $X_n\xrightarrow{p}X$, $\mathbb{E}[|X|]<\infty$ and $|\phi_K(X)-X|\le|X|$. By DCT, $\lim_{K\to\infty}\mathbb{E}[|\phi_K(X)-X|]=\mathbb{E}[0]=0$ as $\lim_{K\to\infty}|\phi_K(X)-X|=0$. Thus, $\mathbb{E}[|X_n-X|]\to 0$. □

2.1 Sufficiency condition

How to verify uniform integrability?

Proposition 2.3 (Sufficient condition of u.i.). *If there exists* $p \in (1, \infty)$ *such that* $\sup \mathbb{E}[|X_n|^p] < \infty$*, then* $\{X_n\}$ *is u.i.*

Proof. Observe that

$$\begin{aligned}
\sup_n \mathbb{E}\left[|X_n| \mathbb{1}_{(K,\infty)}(|X_n|)\right] &= \sup_n \mathbb{E}\left[\frac{|X_n|^p}{|X_n|^{p-1}} \mathbb{1}_{(K,\infty)}(|X_n|)\right] \\
&\leq \sup_n \mathbb{E}\left[\frac{|X_n|^p}{K^{p-1}} \mathbb{1}_{(K,\infty)}(|X_n|)\right] \\
&= \sup_n \frac{\mathbb{E}\left[|X_n|^p \mathbb{1}_{(K,\infty)}(|X_n|)\right]}{K^{p-1}} \\
&= \sup_n \frac{\mathbb{E}[|X_n|^p]}{K^{p-1}} \\
&\leq \frac{C}{K^{p-1}} \\
&\to 0
\end{aligned}$$

as $K \to \infty$. In other words, for any $\epsilon > 0$, there exists $K = K_\epsilon$ such that $\sup_n \mathbb{E}\left[|X_n| \mathbb{1}_{(K_\epsilon,\infty)}(|X_n|)\right] \leq \epsilon$. □

References

Lecture 16: Uniform integrability cont., Strong/Weak law of large numbers

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 22nd, 2024

1 Review

Recall that we have the following implications between notions of convergence:

- Convergence almost surely $\implies$ convergence in probability,
- Convergence in probability $\implies$ there exists a subsequence converging almost surely,
- Convergence in probability $\implies$ convergence in distribution,
- Convergence in L^p norm $\implies$ convergence in probability,
- Convergence in probability $\implies$ convergence in L^p norm with Vitalli uniform integrability.

This lecture we will prove the last implication.

2 Uniform integrability

Definition 2.1 (u.i.). A sequence of RVs $\{X_n\}$ is said to u.i. if for any $\epsilon > 0$, there exists $K = K_\epsilon > 0$ such that $\sup \mathbb{E}[|X_n| \mathbb{1}(|X_n| > K_\epsilon)] \leq \epsilon$.

Proposition 2.2 (Sufficient condition of u.i.). *If there exists $P \in (1, \infty)$ such that $\sup \mathbb{E}[|X_n|^p] < \infty$, then $\{X_n\}$ is u.i.*

Proposition 2.3 (Necessary + sufficient for u.i.). *$\{X_n\}_{n\geq 1}$ is u.i. if and only if*

- $\sup_n \mathbb{E}[|X_n|] \leq \infty$,
- *For any $\epsilon > 0$, there exists $\delta > 0$, such that for any $A \in \mathcal{F}$ satisfying $\mathbb{P}(A) < \delta$, we have $\sup_n \int_A |X_n| d\,\mathbb{P} < \epsilon$.*

Proof. **Forward direction:** If $\{X_n\}$ is u.i., then picking $\epsilon = 1$, there exists a constant $K < \infty$ such that $\sup_n \mathbb{E}\left[|X_n| \mathbb{1}_{(K,\infty)}(|X_n|)\right] \leq 1$. Thus,

$$\begin{aligned}
\sup_n \mathbb{E}[|X_n|] &= \sup_n \mathbb{E}\left[|X_n| \mathbb{1}_{(K,\infty)}(|X_n|)\right] + \sup_n \mathbb{E}\left[|X_n| \mathbb{1}_{(-\infty,K)}(|X_n|)\right] \\
&\leq 1 + K \\
&< \infty.
\end{aligned}$$

On the other hand, for any $A \in \mathcal{F}$ and any $a > 0$, we have, for any $n = 1, 2, \dots,$

$$\begin{aligned}\int_A |X_n| d\,\mathbb{P} &= \int_{A \cap \{|X_n| \le a\}} |X_n| d\,\mathbb{P} + \int_{A \cap \{|X_n| > a\}} |X_n| d\,\mathbb{P} \\ &\le \int_A a d\,\mathbb{P} + \int_{\{|X_n| > a\}} |X_n| d\,\mathbb{P} \\ &= a\mathbb{P}(A) + \mathbb{E}\Big[|X_n| \mathbb{1}_{(a,\infty)}(|X_n|)\Big].\end{aligned}$$

The first term is small by picking $\mathbb{P}(A)$ to be small enough. The second term is small by u.i. and putting a to be large.

Backward direction: Using Markov's inequality, for any $a > 0$, $\sup_{n \ge 1} \mathbb{P}(|X_n| > a) \le \frac{\sup_{n \ge 1} \mathbb{E}[|X_n|]}{a}$, implying that $\{|X_n| > a\}$ is converging to 0 as $a \to \infty$. By choosing $A = \{\omega : |X_n(\omega)| > a\}$ in the second condition, it implies u.i. □

Corollary 2.4. *For any integrable random variable X (i.e., $\mathbb{E}[|X|] < \infty$), we have $\forall \epsilon > 0, \exists \delta > 0,$ s.t. $\forall A \in \mathcal{F}$ satisfying $\mathbb{P}(A) < \delta$, it is true that $\int_A |X| d\,\mathbb{P} < \epsilon$.*

Proof. By picking $X_1 = X_2 = \cdots = X_n = \cdots = X$. □

3 Strong/Weak law of large numbers

3.1 Weak law of large numbers

Theorem 3.1 (Weak law of large numbers (WLLN)). *Consider $X_1, \dots, X_n$ to be*

- *L^2 bounded (i.e., $\mathbb{E}\Big[|X_i|^2\Big] < \infty$,*
- *over the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

Let $\mu_i := \mathbb{E}[X_i]$, $\sigma_{ij} := \mathrm{Cov}(X_i, X_j)$. We claim that

1. *For any $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i\right| > \epsilon\right) \le \frac{\sum_i \sum_j \sigma_{ij}}{n^2 \epsilon^2};$$

2. *As long as*

$$\lim_{n \to \infty} \frac{\sum_i \sum_j \sigma_{ij}}{n^2} = 0,$$

we have

$$\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i \xrightarrow{p} 0;$$

3. *In particular, if*
   (a) $\{X_i\}$ are pairwise uncorrelated, i.e., $\sigma_{ij} = 0$ if $i \ne j$;
   (b) $\mu_i = \mu_j$ for any i, j;

(c) $\sup_i \sigma_{ii} < \infty$. *Then* $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{p} \mu = \mu_1$.

Notice that if (a) is true, then

$$\lim_{n\to\infty} \frac{\sum_i \sum_j \sigma_{ij}}{n^2} = \lim_{n\to\infty} \frac{\sum_i \sigma_{ii}}{n^2} \le \lim_{n\to\infty} \frac{n \sup_i \sigma_{ii}}{n^2} \le \frac{\sup_i \sigma_{ii}}{n} \to 0.$$

Proof. We only prove the first claim here. LHS is equal to

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (X_i - \mu_i)\right|^2 \ge \epsilon^2\right) &\le \frac{\mathbb{E}[\sum_{i=1}^n (X_i - \mu_i)]^2}{n^2\epsilon^2} \quad \because \text{Markov's inequality} \\ &= \frac{\mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}\right]}{n^2\epsilon^2}\end{aligned}$$

□

3.2 Strong law of large numbers

Theorem 3.2 (Strong law of large numbers (SLLN), Etemadi 1981)**.** *Assume*

1. $\{X_n\}_{n\ge 1}$ *is pairwise independent and identically distributed;*
2. $\mathbb{E}[|X_n|] < \infty$ *(i.e., the mean exists).*

We claim $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} \mu =: \mathbb{E}[X_1]$.

Proof. **Step 0:** We claim that we only have to consider these X_i's that are nonnegative by separately discussing X_i^+ and X_i^-.

Step 1: Starting from here, we assume $X_i \ge 0$. Introduce $Y_i := X_i \cdot \mathbb{1}(X_i < i)$. We claim that as long as we can show $\frac{1}{n}\sum_i Y_i \xrightarrow{a.s.} \mu$, then $\frac{1}{n}\sum_i X_i \xrightarrow{a.s.} \mu$. To see this, by 1st B-C lemma,

$$\begin{aligned}\sum_{i=1}^\infty \mathbb{P}(X_i \ne Y_i) &= \sum_{i=1}^\infty \mathbb{P}(X_i \ne X_i \mathbb{1}(X_i < i)) \\ &\le \sum_{i=1}^\infty \mathbb{P}(X_i \ge i) \\ &= \sum_{i=1}^\infty \mathbb{P}(X_1 \ge i) \quad \because X_i\text{'s are identically distributed} \\ &\le \mathbb{E}[X_i] \\ &< \infty.\end{aligned}$$

This implies $\sum_{i=1}^\infty \mathbb{P}(X_i \ne Y_i) < \infty \implies \mathbb{P}(X_i \ne Y_i \quad i.o.) = 0$ by 1st B-C lemma. This implies $\frac{1}{n}\sum_{i=1}^n Y_i - \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} 0$. Thus, the claim is true.

Step 2: We claim that as long as $Z_n := \frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) \xrightarrow{a.s.} 0$, then $\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{a.s.} \mu$. To see that, it suffices to show $\frac{1}{n}\sum_i \mathbb{E}[Y_i] \to \mu =: \mathbb{E}[X_1]$. Notice that $\mathbb{E}[Y_i] - \mathbb{E}[X_i] = -\mathbb{E}\left[X_i \mathbb{1}_{[i,\infty)}(X_i)\right]$. It is clear that $X_i \mathbb{1}_{[i,\infty)}(i) \le X_i$ which has a finite expectation. By

DCT, $\lim_i\{\mathbb{E}[Y_i] - \mathbb{E}[X_i]\} = \mathbb{E}\Big[\lim_i -X_i \mathbb{1}_{[i,\infty)}(X_i)\Big] = 0$. Then, mathematical analysis confirms $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i] - \frac{1}{n}\sum_{i=1}^{n}\mu \to 0$ as $n \to \infty$.

Step 3: We claim that there exists a subsequence $\{K_n\}_{n\geq 1}$ of $\{n\}$ s.t. $Z_{K_n} \overset{a.s.}{\to} 0$. To see this, fix an arbitrary $\alpha > 1$ and let $K_n := [\alpha^n]$ where $[\cdot]$ takes the integer part of the input. Then we can show $Z_{K_n} \overset{a.s.}{\to} 0$.

Step 4: $Z_n \overset{a.s.}{\to} 0$. Denote $T_n := \sum_{i=1}^{n} Y_i$. Then, for any $m \in (K_n, K_{n+1})$, $\frac{T_m}{m} \leq \frac{T_{K_{n+1}}}{K_n} = \frac{T_{K_{n+1}}}{K_{n+1}} \cdot \frac{K_{n+1}}{K_n}$. Also, $\frac{T_m}{m} \geq \frac{T_{K_n}}{K_{n+1}} = \frac{T_{K_n}}{K_n}\frac{K_n}{K_{n+1}}$. From step 3, $\frac{T_{K_n}}{K_n}, \frac{T_{K_{n+1}}}{K_{n+1}} \overset{a.s.}{\to} \mu$ and $\frac{K_n}{K_{n+1}} = \frac{1}{\alpha}$ and $\frac{K_{n+1}}{K_n} = \alpha$. This means $\frac{\mu}{\alpha} \leq \liminf \frac{T_m}{m} \leq \limsup \frac{T_m}{m} \leq \alpha\mu$. Finally, pushing α to 1 yields $\lim_{m\to\infty} \frac{T_m}{m} = \mu$. □

References

Lecture 17: Central Limit Theorem I

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 24th, 2024

1 Famous Quote from Fang today

"Gauss probably knows more than me at his age 10." – Fang Han

2 Strong law of large numbers, CLT

2.1 Strong law of large numbers

Theorem 2.1 (Strong law of large numbers (SLLN), Etemadi 1981)**.** *Assume*

1. $\{X_n\}_{n\geq 1}$ *is pairwise independent and identically distributed;*
2. $\mathbb{E}[|X_n|] < \infty$ *(i.e., the mean exists).*

We claim $\frac{1}{n}\sum_{i=1}^n X_i \overset{a.s.}{\to} \mu =: \mathbb{E}[X_1]$.

Proof. More detailed proof was in the last lecture. Here we briefly review the proof and focus on step 3. **Step 0:** Only need to consider $X_i \geq 0$.

Step 1: Define $Y_i := X_i \cdot \mathbb{1}(X_i \leq i)$

Step 2: Define $Z_n := \frac{1}{n}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])$.

Step 3: For $K_n = [\alpha^n]$ for some $\alpha > 1$, it holds true that $Z_{K_n} \overset{a.s.}{\to} 0$.To see this, fix any $\epsilon > 0$, and check

$$\begin{aligned}
\mathbb{P}(|Z_{K_n}| > \epsilon) &= \mathbb{P}\left(Z_{K_n}^2 > \epsilon^2\right) \\
&\leq \frac{\mathbb{E}\left[Z_{K_n}^2\right]}{\epsilon^2} \\
&= \frac{\mathbb{E}[(\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]))]^2}{K_n^2\epsilon^2} \\
&= \frac{\sum_{i=1}^{K_n} \mathrm{Var}(Y_i)}{K_n^2\epsilon^2}.
\end{aligned}$$

The second step follows by $\mathbb{E}\left[X_i^2\right]$ can be infinity but $\mathbb{E}\left[Z_{K_n}^2\right]$ can't. The last step follows by X_i's are pairwise independent implying that Y_i's are pairwise independent, further implying that $\mathrm{Cov}(X_i, X_j) = 0$ for any $i \neq j$ provided that $\mathbb{E}\left[Y_i^2\right] < \infty$. Continuing the proof, by 1st B-C Lemma,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(|Z_{K_n}| > \epsilon) &\leq \sum_{n=1}^{\infty} \frac{\sum_{k_n \geq i} \mathrm{Var}(Y_i)}{K_n^2 \epsilon^2} \\
&= \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} \left\{ \mathrm{Var}(Y_i) \sum_{n: K_n \geq i} \frac{1}{K_n^2} \right\} \quad \because \text{Fubini's theorem} \\
&\leq \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} \left\{ \mathrm{Var}(Y_i) \cdot \frac{C}{i^2} \right\} \quad \because \exists C = C_\alpha \text{ s.t. } \sum_{K_n = i}^{\infty} \frac{1}{K_n^2} \leq \frac{C}{i^2} \\
&= \frac{C}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\mathrm{Var}(Y_i)}{i^2} \\
&\leq \frac{C}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2} \\
&= \frac{C}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[X_1^2 \mathbb{1}(X_1 \leq i)]}{i^2} \quad \because \text{identical distributed} \\
&= \frac{C}{\epsilon^2} \sum_{i=1}^{\infty} \mathbb{E}\left[X_1^2 \frac{\mathbb{1}(X_i \leq i)}{i^2} \right] \\
&= \frac{C}{\epsilon^2} \mathbb{E}\left[X_1^2 \sum_{i=1}^{\infty} \frac{\mathbb{1}(X_i \leq i)}{i^2} \right] \\
&= \sum_{i \geq X_i} \frac{1}{i^2} \\
&\leq \frac{C'}{X_1} \quad \because \exists C' \text{ s.t. the inequality holds} \\
&\leq \frac{C}{\epsilon^2} \mathbb{E}[C' \cdot X_1] \\
&= \frac{C \cdot C'}{\epsilon^2} \mathbb{E}[X_1] \\
&< \infty.
\end{aligned}$$

This implies that $\sum_{n=1}^{\infty} \mathbb{P}(|Z_{K_n}| > \epsilon) < \infty$. By 1st BC lemma, $\mathbb{P}(|Z_{K_n}| > \epsilon, i.o.) = 0$, implying that $Z_{K_n} \overset{a.s.}{\to} 0$.

Step 4: $Z_n \overset{a.s.}{\to} 0$. □

3 Central limit theorem (LLT)

Theorem 3.1 (CLT, Lyapunov). *Let $X_1, X_2, \ldots$ be i.i.d. with mean $\mu < \infty$ and $\sigma^2 < \infty$. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \overset{d}{\to} \mathcal{N}(0, 1).$$

The approach to proving CLT is by Paul Levy using characteristic function.

Theorem 3.2 (Levy continuity theorem)**.** *$X_n \xrightarrow{d} X$ if and only if $\phi_{X_n}(t) \to \phi_X(t)$ for any $t \in \mathbb{R}$ where $\phi_{X_n}(t) = \mathbb{E}\left[e^{i\cdot tX_n}\right]$ and $\phi_X(t) = e^{i\cdot tX}$.*

Why do we use the characteristic function? First, Cauchy develops a form of CLT:

- $\mathbb{E}[X_n] \to \mathbb{E}[X]$,
- $\mathbb{E}\left[X_n^2\right] \to \mathbb{E}\left[X^2\right]$,
- $\vdots$
- $\implies X_n \xrightarrow{d} X$

This argument requires all moments of X_n's existing. Instead, the c.f. approach does NOT require all but ONLY 1st and 2nd moment existing.

Second, if X is a R.V. with c.f. $\phi_X(\cdot)$, for each $\theta > 0$, define $f_\theta(x) := \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\cdot tx-\theta t^2}\phi_X(t)dt$. Then, for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[g(X)] = \lim_{\theta\to 0}\int_{\mathbb{R}} g(x)f_\theta(x)d_x$ which dependes only on $\phi(X_i)$.

Theorem 3.3 (Portmanteau Lemma)**.** *A sequence of RVs $\{X_n\}$ converges in distribution to a RV X if and only if for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, $\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$. In particular, by choosing $X_1 = X_2 = \cdots = X_n = \cdots = Y$, Portmanteau lemma says $Y \stackrel{d}{=} X$ if and only if for any bounded continuous g, $\mathbb{E}[g(Y)] = \mathbb{E}[g(X)]$. In particular, with inversion formula and Portmanteau, $X \stackrel{d}{=} Y$ if and only if $\phi_X(\cdot) = \phi_Y(\cdot)$.*

Proof. If $X \stackrel{d}{=} Y$, then $\phi_X(\cdot) = \phi_Y(\cdot)$.

If $\phi_X(\cdot) = \phi_Y(\cdot)$, then for any $\theta > 0$, $f_\theta^X(\cdot) = \frac{1}{2\pi}\int\cdots\phi_X(t)d_t$ and $f_\theta^Y(\cdot) = \frac{1}{2\pi}\int\cdots\phi_Y(t)d_t$. By inversion formula, for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[g(X)] = \lim_{\theta\to 0}\int\cdots f_\theta^x(t)dt = \lim_{\theta\to 0}\int\cdots f_\theta^Y(t)d_t$. By Portmanteau, $X \stackrel{d}{=} Y$. □

References

Lecture 18: Central Limit Theorem II

Scribe: Wenhao Pan

STAT 559. Spring 2024
Prof. Fang Han
May 29th, 2024

1 Famous Quote from Fang Today

"I'm doing the European style bowing as the people we talked about in this class are all Europeans." – Fang

2 Final

Here are important topics that might show up in the final exam:

- almost everywhere
- four notions of convergence
- strong/weak law of large numbers
- Fubini-Tonelli
- Random variables: law, CDF, pdf, σ-algebra generated by RVs, independence, expectation, theorem of unconscious statisticians.

3 Inervsion formula

Theorem 3.1 (Inversion formula)**.**

$$\mathbb{E}[g(X)] = \lim_{\theta \to 0} \int_{-\infty}^{\infty} g(x) f_\theta(x) dx$$

with $f_\theta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx - \theta t^2} \phi_X(t) dt.$

Proof. Let $\phi(t) = \phi_x(t)$ and μ be the law of X. Then

1. $\phi(t) = \mathbb{E}\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{ity} d\mu(y)$
2. Fubini then applies to $f_\theta(\cdot)$, giving

$$\begin{aligned}
& f_\theta(x) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{i(y-x)t - \theta t^2} dt \right] d\mu(y) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\sqrt{\frac{\pi}{\theta}} \int_{-\infty}^{\infty} e^{i(2\theta)^{-\frac{1}{2}}(y-x)s} \cdot \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}} ds \right] d\mu(y) \quad \because \pi\text{-parameterization} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\sqrt{\frac{\pi}{\theta}} e^{-\frac{(y-x)^2}{4\theta}} \right] d\mu(y) \\
&= \int_{-\infty}^{\infty} \frac{e^{-\frac{(y-x)^2}{4\theta}}}{\sqrt{4\pi\theta}} d\mu(y)
\end{aligned}$$

which is the pdf of $X + Z_\theta$ where Z_θ is independent of X and $Z_\theta \sim \mathcal{N}(0, 2\theta)$. Then Slutsky theorem proves since $Z_\theta \xrightarrow{p} 0$ as $\theta \to 0$, then $X + Z_\theta \xrightarrow{d} X$, then Portmanteau lemma shows for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[g(X + Z_\theta)] \overset{\theta \to 0}{\to} \mathbb{E}[g(X)]$.

□

4 CLT

Lemma 4.1 (Telescoping inequality). *Consider $a_1, \cdots, a_n$ and $b_1, \ldots, b_n$ s.t. $|a_j| \le 1$, $|b_i| \le 1$, for any $i, j = 1, 2, \cdots, n$. Then $|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i| \le \sum_{i=1}^n |a_i - b_i|$.*

Lemma 4.2 (Taylor expansion). *For any $x \in \mathbb{R}$,*

$$\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \le \min\left\{ |x|^2, \frac{|x|^3}{\sigma} \right\}.$$

Theorem 4.3 (CLT, Lyapunov). *Let $X_1, \ldots, X_n, \ldots$ be i.i.d. sequence of variables with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. The proof is by Paul Levy that related $\xrightarrow{d}$ to the pointwise converge of $\phi_{X_n}(\cdot) \to \phi_X(\cdot)$ as $n \to \infty$.

Levys continuity theorem shows $X_n \xrightarrow{d} X$ if and only if $\phi_{X_n}(\cdot) \to \phi_X(\cdot)$ pointwise for any $t \in \mathbb{R}$. To prove CLT, defining $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$. Then it suffices to prove for any $t \in \mathbb{R}$, $\phi_{S_n}(t) \to \phi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$.

The LHS is

$$\begin{aligned} \mathbb{E}\left[e^{it \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i} \right] &= \mathbb{E}\left[e^{i \frac{t}{\sqrt{n}} Z_1} e^{i \frac{t}{\sqrt{n}} Z_2} \cdots e^{i \frac{t}{\sqrt{n}} Z_n} \right] \\ &= \mathbb{E}\left[e^{i \frac{t}{\sqrt{n}} Z_1} \right] \mathbb{E}\left[e^{i \frac{t}{\sqrt{n}} Z_2} \right] \cdots \mathbb{E}\left[e^{i \frac{t}{\sqrt{n}} Z_n} \right] \\ &= \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}} \right) \right]^n. \end{aligned}$$

It remains to show $\left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}} \right) \right]^n \to e^{-\frac{t^2}{2}} = \lim_{n \to \infty} \left(1 - \frac{t^2}{2n} \right)^n$. So we only have to show for any $t \in \mathbb{R}$,

$$\left| \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}} \right) \right]^n - \left(1 - \frac{t^2}{2n} \right)^n \right| \to 0.$$

By Telescoping inequality lemma,

$$\begin{aligned} \left| \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}} \right) \right]^n - \left(1 - \frac{t^2}{2n} \right)^n \right| &\le n \cdot \left| \phi_{Z_1}\left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n} \right) \right| \\ &= n \cdot \left| \mathbb{E}\left[e^{it \frac{z}{\sqrt{n}}} - 1 - \frac{itZ_1}{\sqrt{n}} + \frac{t^2 z^2}{2n} \right] \right|. \end{aligned}$$

By Taylor expansion lemma,

$$n\cdot\left|\mathbb{E}\left[e^{it\frac{z}{\sqrt{n}}}-1-\frac{itZ_1}{\sqrt{n}}+\frac{t^2z^2}{2n}\right]\right| \leq \mathbb{E}\left[\min\left\{t^2Z_1^2, \frac{|t|^3|Z_1|^3}{\sigma\sqrt{n}}\right\}\right] \quad \because \text{picking } x = tz/\sqrt{n}$$
$$\leq \mathbb{E}\Big[t^2Z_1^2\Big].$$

The second line follows by picking $x=\frac{tZ}{\sqrt{n}}$. Since $t^2Z_1^2$ is integrable, by DCT,

$$\begin{aligned}\lim_{n\to\infty}\left|\left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n-\left(1-\frac{t^2}{2n}\right)^n\right| &\leq \lim_{n\to\infty}\mathbb{E}\left[\min\left\{t^2Z_1^2, \frac{|t|^3|Z_1|^3}{\sigma\sqrt{n}}\right\}\right]\\ &= \mathbb{E}\left[\lim_{n\to\infty}\min\left\{t^2Z_1^2, \frac{|t|^3|Z_1|^3}{\sigma\sqrt{n}}\right\}\right]\\ &= 0.\end{aligned}$$

□

References