STAT 501: Introduction to Probabilities

Lecture Notes

Yanchao Yang

Fall 2024

These notes were taken in **STAT 501: Introduction to Probabilities** at the University of Illinois Chicago in Fall 2024, taught by Prof. Cheng Ouyang. The course follows Sidney I. Resnick's *A Probability Path* [Res99], and the chapter/section numbering of these notes matches the book. All errors are mine.

Contents

1 Sets and Events

1.1 Preliminaries

1.1.1 What is this Chapter about?

When we start learning numbers in our previous studies, we learned two things

(i) Operations of numbers.

For example, we learned $+, -, \times, \div$ and a^{-1} to help us better understand numbers.

(ii) Closure properties of those algebraic operations.

This means, for example, we see that if we add two real numbers a, b, the result we formally denoted as $a + b$ is still a real number. Then we say real number is closed under addition. Similarly, if we multiply two real numbers a, b, the result we formally denoted as $a \times b$ is still a real number. Then we say real number is closed under multiplication. Likewise, if we take inverse of a real number a, the result which we formally denoted as a^{-1} is still a real number.

Because we have so many algebraic operations, sometimes we want to focus on some number of operations, and this motivates the notion of algebraic structure. For example, suppose we only care about addition and multiplication, then we look at a set that is closed under both addition and multiplication, we call it has *ring* structure. For example, real number $\mathbb{R}$ has a ring structure. In addition, we have groups, fields, and algebras. The meaning of these structures is they allow us to talk about the "closure of operations". For example, we want after we multiply two real numbers, the result is still a real number. If it is not, then we may have to invent a new structure.

However, these two things are dealing with the relationship between two points. But when we work on probability theory, we are no longer only interested in the relationship between points. Rather, we are interested in relationship between sets. Therefore, we have to extend all above notions to *Set* first. Likewise, we used (a_n) to denote a sequence of real numbers, and we can talk about limit of this sequence. We will use (A_n) to denote a sequence of sets, and we can define the notion of "limit of sets".

1.1.2 Least Upper Bound & Greatest Lower Bound of Sets

A subset of $\mathbb{R}$ is "bounded" if it does not stretch off to infinity. So "boundedness" is a finite condition.

Definition 1.1 (Upper Bound/Lower Bound)**.** *Let E be a non-empty subset of $\mathbb{R}$. An **upper bound** for E is a number β such that*

$$x \leqslant \beta \qquad \textit{for all} \quad x \in E$$

*If an upper bound exists for E, then E is said to be **bounded above**. Similarly, a **lower bound** for E is a number β such that*

$$x \geqslant \beta \qquad \textit{for all} \quad x \in E$$

*If a lower bound exists for E, then E is said to be **bounded below**. If E is bounded both above and below, then E is said to be **bounded**.*

A "least upper bound" for a set is then an upper bound that is as small as possible.

Definition 1.2 (Least Upper Bound/Supremum)**.** *Given a non-empty subset $E \subseteq \mathbb{R}$ that is bounded above. A **least upper bound** λ for E is the smallest upper bound of E, i.e. λ has two properties*

(i) λ is an upper bound of E,

(ii) If $\gamma \in \mathbb{R}$ is another upper bound ($x \leqslant \gamma$ for all $x \in E$), then $\lambda \leqslant \gamma$.

*Least upper bound is also called **Supremum** of E, and is denoted as*

$$\lambda := \sup E$$

In the case that E does not have an upper bound (i.e. not bounded above) or E is empty, we can still define Supremum for E. The convention is $\sup E = +\infty$ if E does not have an upper bound, and $\sup \varnothing = -\infty$.

There is another characterization of supremum/least upper bound which is not intuitive but useful for computation purpose. In the rest of this notes, whenever we have to show something is the supremum, we will use the following characterization.

Theorem 1.3 (Least Upper Bound)**.** *Let E be a non-empty subset of $\mathbb{R}$ that is bounded above, and let λ be an upper bound for E. Then λ is the least upper bound for E if and only if for every $\epsilon > 0$, there is a $y \in E$ such that $y > \lambda - \epsilon$. Consequently, the least upper bound is defined via the following two properties:*

(i) λ is an upper bound of E.

(ii) For all $\epsilon > 0$, there is a $y \in E$ such that $y > \lambda - \epsilon$.

Proof. For the forward direction.

- Suppose that λ is the least upper bound of E.
- Let $\epsilon > 0$.
- Since $\lambda - \epsilon < \lambda$, and λ is the smallest upper bound of E, it follows that $\lambda - \epsilon$ is not an upper bound for E.
- This means that E must contain some element of $y \in E$ such that $y > \lambda - \epsilon$.

Conversely.

- Suppose λ is an upper bound for E and for every $\epsilon > 0$, there is a $y \in E$ such that $y > \lambda - \epsilon$.
- We want to show that λ is the least upper bound for E.
- Suppose that γ is some upper bound of E. We need to show $\lambda \leqslant \gamma$.

- Suppose for contradiction that $\lambda > \gamma$. Then $0 < \lambda - \gamma =: \epsilon$.
- By assumption, there exists a $y \in E$ such that $y > \lambda - \epsilon = \lambda - (\lambda - \gamma) = \gamma$. This contradicts that γ being an upper bound.

□

Remark 1.4. *The idea of this new characterization can be summarized into following two properties*

(i) λ *is an upper bound of* E.

(ii) There is no number smaller than λ *that is an upper bound.*

A "greatest lower bound" for a set is then an lower bound that is as big as possible.

Definition 1.5 (Greatest Lower Bound/Infimum)**.** *Given a non-empty subset* $E \subseteq \mathbb{R}$ *that is bounded below. A **greatest lower bound*** μ *for* E *is the greatest lower bound of* E*, i.e.* μ *has two properties*

(i) μ *is an lower bound of* E,

(ii) If $\gamma \in \mathbb{R}$ *is another lower bound* ($x \geqslant \gamma$ *for all* $x \in E$)*, then* $\mu \geqslant \gamma$.

*Greatest lower bound is also called **Infimum** of* E*, and is denoted as*

$$\mu := \inf E$$

In the case that E does not have an lower bound (i.e. not bounded below) or E is empty, we can still define Infimum for E. The convention is $\inf E = -\infty$ if E does not have a lower bound, and $\inf \varnothing = +\infty$.

Similar to supremum, there is another characterization of infimum/greatest lower bound which is not intuitive but useful for computation purpose.

Theorem 1.6 (Greatest Lower Bound)**.** *Let* E *be a non-empty subset of* $\mathbb{R}$ *that is bounded below, and let* λ *be a lower bound for* E*. Then* λ *is the greatest lower bound for* E *if and only if for every* $\epsilon > 0$*, there is a* $y \in E$ *such that* $y < \lambda + \epsilon$*. Consequently, the greatest lower bound is defined via the following two properties:*

(i) λ *is a lower bound of* E.

(ii) For all $\epsilon > 0$*, there is a* $y \in E$ *such that* $y < \lambda + \epsilon$.

Proof. The forward direction is easy. The backward direction again use proof by contradiction. I will not verify the details. □

Remark 1.7. *The idea of this new characterization can be summarized into following two properties*

(i) λ *is a lower bound of* E.

(ii) There is no number bigger than λ *that is a lower bound.*

Lastly, if we have a sequence of real numbers $\{a_n\}_{n=1}^{\infty}$. Then we define

$$\liminf_{n\to\infty} a_n := \sup_{N\in\mathbb{Z}_{\geqslant 1}} \inf_{n\geqslant N} a_n, \qquad \limsup_{n\to\infty} a_n := \inf_{n\in\mathbb{Z}_{\geqslant 1}} \sup_{n\geqslant N} a_n \tag{1}$$

1.2 Naive Set Theory

1.2.1 Set Builder Notation

Measure-theoretical Probability is build from measure theory, which is a field deal with theories of sets. So it is important for us to know all basics of sets.

Recall that a **Set** is a collection of unique objects.

- If A is a set, the notation $x \in A$ means element x is in the set A. This reads as
  - "x is an element of A" or,
  - "x is in A" or
  - A contains x
- Express the fact that the statement $x \in A$ is FALSE by writing $x \notin A$. This reads as
  - "x is not an element of A" or,
  - "x is not in A" or
  - A does not contain x

A particular set is defined exactly by its members. We can define a set by the following ways:

1. Explicitly listing out the elements: for example $A = \{\text{dog, cat, panda}\}$, $B = \{0, 2, 4, 6, \ldots\}$.
2. Giving a description of the set: for example let B be the set of non-negative even numbers.
3. Using **Set Builder Notation**: for example, writing

$$A = \underset{\text{The set of}}{\{} \quad \underset{\text{All real numbers}}{x \in \mathbb{R}} \quad \underset{\text{Such that}}{\mid} \quad \underset{\text{This statement is true}}{x^2 + 1 > 37} \quad \}$$

Sometimes we can abbreviate set builder notation. For example, for the set

$$A = \{x \in \mathbb{R} \mid \text{there exists } k \in \mathbb{Z} \text{ such that } x = k\pi\}$$

we can write $A = \{k\pi \mid k \in \mathbb{Z}\}$. This is called **Abbreviated Set Builder Notation**. We can read this abbreviated notation part by part:

$$A = \underset{\text{The set of}}{\{} \quad \underset{\text{real numbers of this form}}{k\pi} \quad \underset{\text{such that}}{\mid} \quad \underset{\text{this statement is true}}{k \in \mathbb{Z}} \quad \}$$

1.2.2 Inclusion of sets

There are two simple relations between sets.

1. We say A is a subset of B when every element of A is an element of B. This is written as $A \subseteq B$, and is read as

   - "A is a subset of B" or
   - "A is contained in B" or
   - "B contains A"

   By convention, $A \subset B$ means the same thing as $A \subseteq B$. (This is unlike $<$ and $\leqslant$). For proper containment it is denoted as $A \subsetneq B$. In terms of logic, $A \subseteq B$ means that

   $$x \in A \implies x \in B$$

   (the quantifier $\forall x$ is understood); that is, "if x is an element of A, then x is an element of B; that is, all elements of A are elements of B.

2. We say two sets A and B are equal when every element of A is an element of B and every element of B is an element of A. This is written as $A = B$. Equivalently, $A = B$ iff $A \subseteq B$ and $B \subseteq A$. In terms of logic, $A = B$ means that

   $$x \in A \iff x \in B$$

The symbol $|S|$ for a set S means the *number of elements* of S, it is also called **cardinality** of S. If sets A and B are finite, then

$$A \subseteq B \implies |A| \leqslant |B|$$

The subsets of a set S forms a set, called the **power set**, or the **set of parts of** S. Power set is denoted as $\mathscr{P}(S)$; a popular alternative is 2^S. This notation comes from the fact that the cardinatlity of power set for a finite set S is $|\mathscr{P}(S)| = 2^{|S|}$.

1.2.3 Basic operations between sets

Previously we have "operations of numbers". Now we want to know what are the "operations of sets" just like $+, -, \times, \div$ for numbers. Here are the basic set operations that we should know for this subject

- $\cup$: the *union*
- $\cap$: the *intersection*
- $\setminus$: the *set difference* (or *set complement*)
- $\triangle$: the *symmetric difference*

For those who has not seen symmetric difference, it is defined as

$$A \triangle B = (A \backslash B) \cup (B \backslash A)$$

And for *union* and *intersection*, it is taken over an **index set** which we will denoted as $\mathbb{T}$. The *union* and *intersection* over the *index set* $\mathbb{T}$ are defined as

$$\bigcup_{t \in \mathbb{T}} A_t := \{\omega \mid \omega \in A_t;\ \text{for some } t \in \mathbb{T}\}$$
$$\bigcap_{t \in \mathbb{T}} A_t := \{\omega \mid \omega \in A_t;\ \text{for all } t \in \mathbb{T}\}$$

However, there are different versions of *unions* depends on different version of *index set* $\mathbb{T}$. There are 3 different cases of index set:

1. $\mathbb{T}$ finite: (e.g. $\mathbb{T} = \{1, \ldots, n\}$) Then we have **finite union**.
2. $\mathbb{T}$ countable: (e.g. $\mathbb{T} = \mathbb{N}, \mathbb{Z}, \mathbb{Q}$) Then we have **countable union**.
3. $\mathbb{T}$ uncountable: (e.g. $\mathbb{T} = \mathbb{R}, [0, 1]$) Then we have **arbitrary union**.

Likewise for intersection. Depends on different *index set* $\mathbb{T}$, we have 1) **finite intersection**, 2) **countable intersection**, 3) **arbitrary intersection**.

1.2.4 Indicator functions

There is a very nice and useful duality between sets and functions which emphasizes the algebraic properties of sets. That is, given a set A, we can define a function called **Indicator Function** of A as

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \tag{2}$$

This is also called **characteristic function** by some measure-theorist. Here are two simple properties of *indicator functions*

- $\mathbf{1}_A \leqslant \mathbf{1}_B$ if and only if $A \subseteq B$
- $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$

The big deal of this duality is that, rather than studying sets directly, we can study function instead. And we will have many tools to study functions.

1.3 Limits of Sets

Lecture 2024-08-31

Just like we can define infimum, supremum and limits for a sequence of numbers, we can also define the same notion for a sequence of sets. Let A_n, $n \geqslant 1$ be a sequence of sets. We define

- **infimum of sets**: $\inf_{k \geqslant n} A_k := \bigcap_{k=n}^{\infty} A_k$

- **supremum of sets**: $\sup_{k \geqslant n} A_k := \bigcup_{k=n}^{\infty} A_k$
- **limit infimum of sets**: $\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \sup_{n=1}^{\infty} \inf_{k=n}^{\infty} A_k$
- **limit supremum of sets**: $\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \inf_{n=1}^{\infty} \sup_{k=n}^{\infty} A_k$

The **limit** of a sequence of sets is defined as follows: if for some sequence B_n of subsets has

$$\limsup_{n\to\infty} B_n = \liminf_{n\to\infty} B_n = B$$

Then the set B is called the limit of B_n, and we write

$$\lim_{n\to\infty} B_n = B \qquad \text{or} \qquad B_n \longrightarrow B$$

Next we give the interpretation of limsup and liminf.

limsup of sets

Let $\{A_n \mid n \geqslant 1\}$ be a sequence of subsets of Ω. For limsup we have the interpretation

$$\begin{aligned}
\limsup_{n\to\infty} A_n &= \{\omega : \omega \text{ is in infinitely many of } A_n' s\} \\
&= \{\omega : \omega \in A_{n_k}, k = 1, 2, \ldots \quad \text{for some subsequence } n_k \text{ depending on } \omega\} \\
&= \left\{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = +\infty\right\} \\
&= \{A_n \quad \text{i.o.}\}
\end{aligned}$$

Proof. The LHS is $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, and the RHS is any one of the four. Show LHS in RHS. Pick $x \in LHS$, WTS $x \in RHS$.

1. $n = 1$, then $x \in \cup_{k\geqslant 1} A_k$, so there exists n_1 such that $x \in A_{n_1}$.
2. If $n = n_1 + 1$, then $x \in \bigcup_{k\geqslant n_1+1} A_k$. So there exists $n_2 > n$ such that $x \in A_{n_2}$.
3. Let $n = n_2 + 1$, then $x \in \bigcup_{n_2+1} A_k$, so there exists $n_3 > n_2$ such that $x \in A_{n_3}$.
4. And so on.

Show RHS in LHS. Pick $\omega \in RHS$. WTS, $\omega \in LHS$. Let $\omega \in A_{n_1}, A_{n_2}, \ldots$ where $n_1 < n_2 < n_3 < \ldots$.

1. For $n = 1$, $\omega \in \bigcup_{k=1}^{+\infty} A_k = B_1$.
2. For $n = 2$, $\omega \in \bigcup_{k=2}^{+\infty} A_k = B_2$.
3. For $n = 3$, $\omega \in \bigcup_{k=3}^{+\infty} A_k = B_3$.
4. And so on.
5. If we keep doing this, $\omega \in B_n$ for all n, so $\omega \in \cap_{n=1}^{+\infty} B_n$ which is limsup.

$\square$

liminf of sets

Let $\{A_n \mid n \geqslant 1\}$ be a sequence of subsets of Ω. For liminf we have the interpretation

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \{\omega : \omega \text{ is in all the } A_n's \text{ except finite many of them}\}\\ &= \left\{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(\omega) < +\infty\right\}\\ &= \{\omega : \text{ there is a } n_0(\omega) \text{ s.t. } \omega \in A_n \text{ for all } n > n_0\}\end{aligned}$$

Proof. Show LHS in RHS Pick $\omega \in LHS = \cup_{n=1}^{+\infty}(\cap_{k=n}^{\infty} A_k)$, WTS $\omega \in RHS$.

- There exists n_0 such that $\omega \in C_{n_0} = \cap_{k=n_0}^{+\infty} A_k$. So $\omega \in A_k$ for all $k \geqslant n_0$.

Show RHS in LHS

- Pick $\omega \in RHS$. WTS, $\omega \in LHS$. $\omega \in RHS$ means there exists n_0 such that $\omega \in A_n$ for all $n \geqslant n_0$.
- This implies $\omega \in C_{n_0} = \bigcap_{k=n_0}^{+\infty} A_k$.
- This again implies $\omega \in \bigcup_{n=1}^{+\infty} C_n = \liminf_{n\to\infty} A_n$

□

Example 1.1. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions, and f is also a function. Consider the set of x where $f_n(x)$ converge to $f(x)$ (note that only some of x makes $f_n(x)$ converge to $f(x)$).*

$$\{x : f_n(x) \to f(x) \text{ as } n \to \infty\}$$

But this set if too vague. So we consider

$$A_{n,k} := \left\{x : |f_n(x) - f(x)| \leqslant \frac{1}{k}\right\}$$

Then the above set can be re-written as

$$\{x : f_n(x) \to f(x) \text{ as } n \to \infty\} = \bigcap_{k=1}^{\infty} \liminf_{n\to\infty} A_{n,k}$$

1.4 Monotone Sequences

Recall that for real numbers, we have monotone non-decreasing sequence of numbers $(a_n)_{n=1}^{\infty}$ if $a_n \leqslant a_{n+1}$ for all $n \in \mathbb{Z}_{\geqslant 1}$, and monotone non-increasing sequence of numbers $(a_n)_{n=1}^{\infty}$ if $a_n \geqslant a_{n+1}$ for all $n \in \mathbb{Z}_{\geqslant 1}$. They may or may not have a limit. In this section, we will define the same notion for sets. Let $(A_n)_{n=1}^{\infty}$ be a sequence. Then $(A_n)_{n=1}^{\infty}$ is called

1. **Monotone non-decreasing sequence of sets**: if $A_1 \subseteq A_2 \subseteq \ldots$, we denote it as $A_n \nearrow$ or $A_n \uparrow$

2. **Monotone non-increasing sequence of sets**: if $A_1 \supseteq A_2 \supseteq \dots$, we denote it as $A_n \searrow$ or $A_n \downarrow$

For a monotone sequence of points, the limit may or may not exist. But for a monotone sequence of sets, the limit always exists. Next proposition tells us the explicit form of these limits.

Proposition 1.8 (Limit of Monotone Sequence of Sets)**.** *Let $(A_n)_{n\geqslant 1}$ be a monotone sequence of subsets.*

1. *If $A_n \nearrow$, then* $\lim_{n\to\infty} A_n := \cup_{n=1}^{\infty} A_n$.
2. *If $A_n \searrow$, then* $\lim_{n\to\infty} A_n := \cap_{n=1}^{\infty} A_n$.

Proof. For Case (1), the goal is to show

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = \limsup_{n\to\infty} A_n$$

- Since $A_k \subseteq A_{k+1}$, then $A_n = \cap_{k\geqslant n} A_k$.
- Hence,

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty} A_k\right) = \bigcup_{n=1}^{\infty} A_n$$

- Likewise,

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\left(\bigcup_{k=n}^{\infty} A_k\right) \subseteq \bigcup_{k=1}^{\infty} A_k = \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$$

- By sandwich theorem, we have

$$\liminf_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} A_k = \limsup_{n\to\infty} A_n$$

- Hence, we conclude that

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = \limsup_{n\to\infty} A_n$$

For Case (2), the goal is to show

$$\liminf_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \limsup_{n\to\infty} A_n$$

By taking the complement, we have $A_n^c \nearrow$. Then using the consequence of above result, we have

$$\liminf_{n\to\infty} A_n^c = \bigcup_{n=1}^{\infty} A_n^c = \limsup_{n\to\infty} A_n^c$$

Taking the complement, we have

$$\liminf_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \limsup_{n\to\infty} A_n$$

This completes the proof. □

Corollary 1.9. *For any sequences* (A_n)*, we have*

- $\inf_{k\geqslant n} A_k \nearrow$;
- $\sup_{k\geqslant n} A_k \searrow$;

Then it follows that

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty}\left(\inf_{k\geqslant n} A_n\right) = \lim_{n\to\infty}\left(\bigcap_{k\geqslant n} A_n\right) = \bigcup_{n\to\infty}\bigcap_{k\geqslant n} A_n$$
$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty}\left(\sup_{k\geqslant n} A_n\right) = \lim_{n\to\infty}\left(\bigcup_{k\geqslant n} A_n\right) = \bigcap_{n\to\infty}\bigcup_{k\geqslant n} A_n$$

1.5 Set Operations and Closures

Lecture 2024-09-07

Remember that the Second motivation of this Chapter is to define the notion of "closure" of operation for set as an analogue for "closure" of operations of numbers (e.g. real numbers $\mathbb{R}$ is closed under $+, \times$).

In Section 1.2.3 and Section 1.4, we introduced several set operations. One can understand this as an analog of $+, -, \times, \div$ and taking limits for numbers (although taking limits of numbers is not considered as an algebraic operation for numbers). These are

1. *Set Complementation*: if $A \subseteq \Omega$, then the *complement* is written as $\Omega\backslash A$. When the ambient set Ω is clear, we can simply use A^c to denote $\Omega\backslash A$.

2. *Finite Union*: $A \cup B$

   Countable Union: $\bigcup_{n\in\mathbb{Z}_{\geqslant 1}} A_n = \bigcup_{n=1}^{\infty} A_n = A_1 \cup \cdots \cup A_n \cup \ldots$

   Arbitrary Union: $\bigcup_{t\in\mathbb{T}} A_t$ for $\mathbb{T}$ arbitrary (means can be uncountable).

3. *Finite Intersection*: $A \cap B$

   Countable Intersection: $\bigcap_{n\in\mathbb{Z}_{\geqslant 1}} A_n = \bigcap_{n=1}^{\infty} A_n = A_1 \cap \cdots \cap A_n \cap \ldots$

   Arbitrary Intersection: $\bigcap_{t\in\mathbb{T}} A_t$ for $\mathbb{T}$ arbitrary (means can be uncountable).

4. *Taking Monotone Limits*: if (A_n) is a monotone sequence then

$$\lim_{n\to\infty} A_n := \begin{cases} \bigcup_{n=1}^{\infty} A_n & \text{if } A_n \nearrow \\ \bigcap_{n=1}^{\infty} A_n & \text{if } A_n \searrow \end{cases}$$

Having all these set operations, we can start to define the "algebraic structure" for sets.

To understand what does it mean by "algebraic structure for sets", let's first recall that, for example, $\mathbb{R}$ is a collection of numbers, but it is really a set together with two operations: $(\mathbb{R}, +, \times)$ (the terminology for this is: $\mathbb{R}$ has a *field* structure, some people will also use the term *algebra*). This means $+, \times$ are well-defined so that we don't have to worry about adding two real numbers will return us a non-real number (people call this $\mathbb{R}$ is closed $+, \times$).

Just like we can have a "collection of numbers" with certain closure structure, we can also have a "collection of sets" with certain closure structures. The main task of this section is to introduce us some of those structures. Next is a list of structures we need:

1. Field (Algebra);
2. σ-Field (σ-Algebra);
3. Semi-algebra;
4. Semi-ring;
5. σ-ring;
6. Monotone Class;
7. π-system;
8. λ-system

Some structures that are more important in this subject. These are 1) Field, 2) σ-Field (they are core of measure space), 6) Monotone Class (it defines limits of sets), 7) π-system and 8) λ-system (they are important because of Dynkin's $\pi - \lambda$ theorem). We will introduce Field and σ-field here first.

One last remark is that these collections of subsets are under an ambient set which we denote as Ω.

Before formalizing these structures, here is a motivating example of a collection of sets with only weak closure properties.

Example 1.2. *Consider $\Omega = (0, 1]$ and the set*

$$\mathcal{C} := \{(a, b] : 0 \leqslant a \leqslant b \leqslant 1\}$$

The convention is $(a, a] = \varnothing$.

- ***Claim 1**: $\mathcal{C}$ is closed under finite intersection.*
- ***Claim 2**: $\mathcal{C}$ is NOT closed under countable intersection.*
  e.g. Consider $C_n = \left(\frac{1}{2} - \frac{1}{n+3}, 1\right]$. Then $\cap_{n=1}^{\infty} C_n = \left[\frac{1}{2}, 1\right] \notin \mathcal{C}$.
- ***Claim 3**: $\mathcal{C}$ is NOT closed under complement. Obvious.*

Definition 1.10 (Field/Algebra)**.** *A **Field** (a synonym is **Algebra**), denoted as $\mathscr{A}$, is a non-empty collection of subsets of Ω closed under*

1. *Complements;*

2. *Finite union;*

3. *Finite intersection.*

A minimal set of axioms for $\mathscr{A}$ to be a field is

(A1) $\Omega \in \mathscr{A}$;

(A2) if $A \in \mathscr{A}$, then $A^c \in \mathscr{A}$;

(AF3) if $A, B \in \mathscr{A}$, then $A \cup B \in \mathscr{A}$.

Remark 1.11. *From all the above axioms, we can deduce the following consequences:*

- *If $A_1, A_2, A_3 \in \mathscr{A}$, then we can first deduce $A_1 \cup A_2 \in \mathscr{A}$, then we further deduce $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3 \in \mathscr{A}$ inductively.*

- *By the same inductive procedure, we can deduce if $A_1, \ldots, A_n \in \mathscr{A}$, then $A_1 \cup \cdots \cup A_n \in \mathscr{A}$.*

- *By taking complement, we have $A_1 \cap \cdots \cap A_n \in \mathscr{A}$. The chain of deduction is as follows:*

  - *$A_i \in \mathscr{A}$ implies $A_i^c \in \mathscr{A}$ (using axiom (2));*
  - *$A_i^c \in \mathscr{A}$ implies $\cup_{i=1}^{\infty} A_i^c \in \mathscr{A}$ (using axiom (3));*
  - *$\cup_{i=1}^{\infty} A_i^c \in \mathscr{A}$ implies $(\cup_{i=1}^{\infty} A_i^c)^c \in \mathscr{A}$ (using axiom (2));*
  - *$\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c \in \mathscr{A}$ (using de Morgan's Law);*

Next let's introduce another structure:

Definition 1.12 (σ-Field/σ-Algebra)**.** *A σ-**Field** (a synonym is σ-**Algebra**), denoted as $\mathscr{B}$, is a non-empty collection of subsets of Ω closed under*

1. *Complements;*

2. *Countable union;*

3. *Countable intersection.*

A minimal set of axioms for $\mathscr{B}$ to be a field is

(A1) $\Omega \in \mathscr{B}$;

(A2) if $A \in \mathscr{B}$, then $A^c \in \mathscr{B}$;

(A3) if $A_n \in \mathscr{B}$ for all $n \in \mathbb{Z}_{\geqslant 1}$, then $\cup_{n \in \mathbb{Z}_{\geqslant 1}} A_n \in \mathscr{B}$.

Remark 1.13. *Then by symmetry proof, we can show* $\cap_{n\geqslant 1}A_n \in \mathscr{B}$.

Example 1.3 (Field, but NOT σ-Field)**.** *Let* $\Omega = (0,1]$. *Suppose* $\mathscr{A}$ *consists of empty set* $\varnothing$ *and finite unions of disjoint intervals, i.e.*

$$\mathscr{A} := \left\{ \bigcup_{i=1}^{n} (a_i, b_i] : \text{ for some } n \in \mathbb{Z}_{\geqslant 1}, 0 \leqslant a_i \leqslant b_i \leqslant 1 \right\} \cup \varnothing$$

Then the claim is $\mathscr{A}$ *is a field, but not a* σ*-field. To show* $\mathscr{A}$ *is a field, DIY. To see* $\mathscr{A}$ *is not a* σ*-field, consider the sequence of subsets*

$$\left\{ \left(0, \frac{1}{2}\right] \right\} \cup \left\{ \left(\sum_{m=1}^{n} \frac{1}{2^m}, \sum_{m=1}^{n+1} \frac{1}{2^m} \right] \mid n \geqslant 1 \right\}$$

Each one of these subsets belong to $\mathscr{A}$. *But if we consider the countable union of these subsets*

$$\left(0, \frac{1}{2}\right] \cup \left(\frac{1}{2}, \frac{1}{2} + \frac{1}{2^2}\right] \cup \left(\frac{1}{2} + \frac{1}{2^2}, \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3}\right] \cup \cdots \notin \mathscr{A}$$

Also, one can argue that $A_n = (1/2 - 1/(n+3), 1] \in \mathscr{A}$, *but* $\cap_{n\geqslant 1}A_n = [1/2, 1] \notin \mathscr{A}$.

Therefore, $\mathscr{A}$ *is NOT a* σ*-field.*

Example 1.4. *The following are some examples of* σ*-fields.*

- $\mathscr{B} = \mathscr{P}(\Omega)$ *The largest* σ*-field over* Ω. *Any* σ*-field should be contained in power set.*
- $\mathscr{B} = \{\varnothing, \Omega\}$ *The smallest* σ*-field over* Ω. *Because any* σ*-field should contain* $\varnothing, \Omega$.
- *Let* $\Omega = \mathbb{R}$. *Define*

  $$\mathscr{B} = \{A \subset \mathbb{R} : \text{ either } A \text{ is a countable set, or } A^c \text{ is a countable set}\}$$

  (A1) $\Omega \in \mathscr{B}$. *Since* $\Omega^c = \varnothing$ *is countable*

  (A2) If $A \in \mathscr{B}$. *Then two cases. (1) A is countable, (2) A is uncountable. Case (1) means* $A^c \in \mathscr{B}$. *Case (2) means* A^c *countable, hence* $A^c \in \mathscr{B}$.

  (A3) If $A_1, \ldots A_n \in \mathscr{B}$. *Then two cases. (1) All* A_n *is countable. (2) At least one of* A_n *is uncountable. Case (1) is trivial. Case (2), let's say* A_1 *is uncountable. Then* A_1^c *is countable. Consider* $(\cup_{n\geqslant 1}A_n)^c = \cap_{n\geqslant 1}A_n^c \subset A_1^c$, *subset of countable set is countable, hence* $(\cup_{n\geqslant 1}A_n)^c$ *is countable. Hence* $\bigcup_{n\geqslant 1} A_n \in \mathscr{B}$.
- *Consider* σ*-field generated by a subset of the power set* $\mathscr{C} \subseteq \mathscr{P}(\Omega)$, *denoted as* $\sigma(\mathscr{C})$.
- *Borel* σ*-field and Borel sets on* $\mathbb{R}$, *denoted as* $\mathscr{B}(\mathbb{R})$.

1.6 The σ-algebra generated by a Given Class $\mathscr{C}$

Lecture 2024-09-10

Lemma 1.14. *Intersection of σ-fields is still a σ-field.*

Definition 1.15 (σ-field generated by $\mathcal{C} \subset \mathcal{P}(\Omega)$)**.** *The σ-field generated by $\mathcal{C} \subset \mathcal{P}(\Omega)$ is the smallest σ-field containing $\mathcal{C}$, denoted as $\sigma(\mathcal{C})$, defined as*

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathscr{B}\ \sigma\text{-field}, \\ \mathscr{B} \supset \mathcal{C}}} \mathscr{B}$$

1.7 Borel Sets on the real line

Lecture 2024-09-10

There is an important σ-field called **Borel σ-field**. The reason that it is important is that it connects topological space to measure space, and having a topological space allows us to talk about continuous function.

Definition 1.16 (Borel σ-field)**.** *Let $\Omega = \mathbb{R}$. Let $\mathscr{C} = \{(a, b] : -\infty < a < b < +\infty\}$, and we use this collection of sets $\mathscr{C}$ to generate the σ-field. Hence*

$$\sigma(\mathscr{C}) =: \mathscr{B}(\mathbb{R})$$

*is called **Borel σ-field**. Elements (which are sets) in $\sigma(\mathscr{C})$ is called a **Borel set**.*

There are other definitions of Borel σ-field. All the following definitions are equivalent:

$$\begin{aligned} \mathscr{B}(\mathbb{R}) &= \sigma(\{(a, b] : -\infty < a < b < \infty\}) =: \mathscr{C}^{(]} \\ &= \sigma(\{[a, b] : -\infty < a < b < \infty\}) =: \mathscr{C}^{[]} \\ &= \sigma(\{[a, b) : -\infty < a < b < \infty\}) =: \mathscr{C}^{[)} \\ &= \sigma(\{(a, b) : -\infty < a < b < \infty\}) =: \mathscr{C}^{()} \\ &= \sigma(\text{open sets of } \mathbb{R}) \end{aligned}$$

Proof. We will only show $\sigma(\mathscr{C}^{()}) = \sigma(\mathscr{C}^{(]})$.
Show $\subseteq$:

- Pick any $(a, b) \in \mathscr{C}^{()}$.
- Note
$$(a, b) = \bigcup_{n=1}^{\infty} \underbrace{\left(a, b - \frac{1}{n}\right]}_{\in \mathscr{C}^{(]}}$$
- Hence, $(a, b) \in \sigma(\mathscr{C}^{(]})$ because σ-field is closed under countable union. Therefore $\sigma(\mathscr{C}^{()}) \subseteq \sigma(\mathscr{C}^{(]})$.

Show $\supseteq$:

- Pick any $(a, b] \in \mathscr{C}^{(]}$.
- Note
$$(a, b] = \bigcap_{n=1}^{\infty} \underbrace{\left(a, b + \frac{1}{n}\right)}_{\in \mathscr{C}^{()}}$$
- Hence, $(a, b] \in \sigma(\mathscr{C}^{()})$ because σ-field is closed under countable intersection.

□

The following lemma is only a technical tool for the proof of next lemma.

Lemma 1.17. *Any open subset of $O \subset \mathbb{R}$, O can be expressed as a countable union of open intervals. $O = \bigcup_{i=1}^{\infty}(a_i, b_i)$.*

Theorem 1.18 (Open sets generates Borel σ-algebra)**.** *Denote*
$$\mathcal{O} = \{\textit{the collection of open subsets of } \mathbb{R}\}$$
Then
$$\sigma(\{(a, b) : -\infty < a < b < \infty\}) = \sigma(\mathcal{O})$$

Proof. We will show $\sigma(\mathcal{O}) = \sigma(\mathscr{C}^{()})$.
Show $\subseteq$

- Pick any $(a, b) \in \mathcal{O} \subseteq \sigma(\mathcal{O})$.
- So $\sigma(\mathcal{O})$ is a σ-field containing $\mathscr{C}^{()}$.
- Hence $\sigma(\mathcal{O}) \supseteq \sigma(\mathscr{C}^{()})$.

Show $\supseteq$

- Pick any $O \in \mathcal{O}$.
- By lemma 1.17, we have
$$O = \bigcup_{i=1}^{\infty}(a_i, b_i) \in \mathscr{C}^{()} \subset \sigma(\mathscr{C}^{()})$$
- So $\sigma(\mathscr{C}^{()})$ is a σ-field containing $\mathcal{O}$.
- Hence, $\sigma(\mathscr{C}^{()}) \supset \sigma(\mathcal{O})$.

□

Remark 1.19. *If Ω is only a topological space (i.e. not $\mathbb{R}$ any more), then the Borel σ-field of Ω can be defined as*
$$\mathscr{B}(\Omega) := \sigma(\textit{open sets})$$
For example, $\Omega = \mathbb{R}^n$. Then
$$\mathscr{B}(\Omega) := \sigma(\{(a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n)\})$$

1.8 Comparing Borel Sets

Lecture 2024-09-12

Motivation

We already know $\mathscr{B}(\mathbb{R})$ for $\Omega = \mathbb{R}$. If we take $\Omega = (0,1)$, we can say $(0,1)$ is a topological space. So we can talk about what is the set generated by all open sets of $(0,1)$:

$$\mathscr{B}((0,1)) := \sigma(\text{open subsets of } (0,1))$$

There are several questions we can ask:

1. Is $\mathscr{B}((0,1))$ also a Borel σ-field?
2. What's the relation between $\mathscr{B}((0,1))$ and $\mathscr{B}(\mathbb{R})$? Do we have $\mathscr{B}((0,1)) \subset \mathscr{B}(\mathbb{R})$?

Main theorem

Theorem 1.20. *Let $\Omega_0 \subset \Omega$ be a subset. Then we can claim the following*

1. *If $\mathscr{B}$ is a σ-field of subsets of Ω (i.e. $\mathscr{B}$ is a σ-field over Ω). Define*

$$\mathscr{B}_0 := \{A \cap \Omega_0 : \text{ for all } A \in \mathscr{B}\} =: \Omega_0 \cap \mathscr{B}$$

   Here $\mathscr{B}_0$ is a collection of subsets of Ω_0. The claim is $\mathscr{B}_0$ is a σ-field of subsets of Ω_0. (i.e. $\mathscr{B}_0$ is a σ-field over Ω_0).

2. *If $\mathcal{C}$ is a collection of subsets of Ω. Then we can use $\mathcal{C}$ to generate the σ-field and intersect it with Ω_0, i.e.*

$$\sigma(\mathcal{C}) \cap \Omega_0 := \{A \cap \Omega_0 : A \in \sigma(\mathscr{C})\}$$

   Here $\sigma(\mathcal{C}) = \mathscr{B}$ in part (1). By part (1), we know $\sigma(\mathcal{C}) \cap \Omega_0$ is a σ-field. On the other hand, we could consider the notion $\sigma(\mathcal{C} \cap \Omega_0)$. This is also a σ-field, but the whole space is not Ω any more, but is Ω_0, because every element in $\sigma(\mathcal{C} \cap \Omega_0)$ are subsets of Ω_0. The claim is

$$\sigma(\mathcal{C}) \cap \Omega_0 = \sigma(\mathcal{C} \cap \Omega_0)$$

Proof. The strategy of showing a collection $\mathscr{B}_0$ is a σ-field is to check 3 axioms of σ-field.
Proof of (1): Need to show $\mathscr{B}_0$ is a σ-field over Ω_0.

(A1) Check $\Omega_0 \in \mathscr{B}_0$. See that $\Omega_0 = \Omega \cap \Omega_0 \in \mathscr{B}_0$.

(A2) If $B \in \mathscr{B}_0$, i.e. there exists $A \in \Omega$ such that $B = A \cap \Omega_0$. Need to show $B^c = \Omega_0 \backslash B \in \mathscr{B}_0$. But $B^c = \Omega_0 \backslash A = \Omega_0 \cap A^c \in \mathscr{B}$.

(A3) If $B_1, \ldots, B_n, \ldots \in \mathscr{B}_0$. Need to show $\cup_{n\geqslant 1} B_n \in \mathscr{B}_0$. Since $B_n \in \mathscr{B}_0$, we have $B_n = A_n \cap \Omega_0$ for some $A_n \subseteq \Omega$. Hence, $\cup_{n\geqslant 1} B_n = \cup_{n\geqslant 1}(A_n \cap \Omega_0) = \Omega_0 \cap (\cup_{n\geqslant 1} A_n)$.

Proof of (2): We want to show $\sigma(\mathcal{C}) \cap \Omega_0 = \sigma(\mathcal{C} \cap \Omega_0)$.

- Show $\supseteq$. Using Part (1), we know $\sigma(\mathcal{C}) \cap \Omega_0$ is a σ-field. And see that $\sigma(\mathcal{C}) \cap \Omega_0 \supseteq \mathcal{C} \cap \Omega_0$. Hence, the σ-field $\sigma(\mathcal{C}) \cap \Omega_0$ contains the smallest σ-field containing $\mathcal{C} \cap \Omega_0$. Hence $\sigma(\mathcal{C}) \cap \Omega_0 \supseteq \sigma(\mathcal{C} \cap \Omega_0)$.
- Show $\subseteq$. We have to use $\sigma(\mathcal{C})$ being the smallest σ-field containing $\mathcal{C}$. The strategy is as follows:
  - We define
  $$\mathcal{G} := \{A \subseteq \Omega : A \cap \Omega_0 \in \sigma(\mathscr{C} \cap \Omega_0)\}$$
  - Suppose we can prove
  $$\mathcal{G} \supseteq \sigma(\mathscr{C})$$
  - Then for any $A \in \sigma(\mathscr{C}) \subseteq \mathcal{G}$.
  - Then by definition of $\mathcal{G}$, we have
  $$A \cap \Omega_0 \in \sigma(\mathscr{C} \cap \Omega_0)$$
  - This means $\sigma(\mathscr{C}) \cap \Omega_0 \subseteq \sigma(\mathcal{C} \cap \Omega_0)$. Because
  $$\sigma(\mathscr{C}) \cap \Omega_0 = \{A \cap \Omega_0 : A \in \sigma(\mathscr{C})\}$$
  - Hence, completes the proof.
- So we only left to prove $\mathcal{G} \supseteq \sigma(\mathcal{C})$.
  - First, it is obvious that $\mathcal{G} \supset \mathcal{C}$,
  - Because for any $A \in \mathcal{C}$, we have $A \cap \Omega_0 \in \mathcal{C} \cap \Omega_0 \subseteq \sigma(\mathcal{C} \cap \Omega_0)$., * hence $A \in \mathcal{G}$ by definition of $\mathcal{G}$.
  - Second, we show $\mathcal{G}$ is a σ-field of Ω.

  (A1) $\Omega \in \mathcal{G}$. Because $\Omega \cap \Omega_0 = \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)$.

  (A2) If $A \in \mathcal{G}$, i.e. $A \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)$. Then for $A^c = \Omega \backslash A$, using the fact that $\sigma(\mathcal{C} \cap \Omega_0)$ being a σ-field, we have
  $$A^c \cap \Omega_0 = (\Omega \backslash A) \cap \Omega_0 = \Omega_0 \backslash (A \cap \Omega_0) \in \sigma(\mathcal{C} \cap \Omega_0)$$
  Hence, $A^c \in \mathcal{G}$.

  (A3) If $A_1, \ldots, A_n, \ldots \in \mathcal{G}$. WTS $\cup_{n\geqslant 1} A_n \in \mathcal{G}$. We have $A_n \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)$. Then using the fact that $\sigma(\mathcal{C} \cap \Omega_0)$ being a σ-field, we have
  $$\left(\bigcup_{n\geqslant 1} A_n\right) \cap \Omega_0 = \bigcup_{n\geqslant 1} (A_n \cap \Omega_0) \in \sigma(\mathcal{C} \cap \Omega_0)$$
  Hence, we conclude $\cup_{n\geqslant 1} A_n \in \mathcal{G}$.

- Since $\mathcal{G}$ is a σ-field, and $\mathcal{G} \supseteq \mathcal{C}$. By construction of $\sigma(\mathcal{C})$, which is the smallest σ-field containing $\mathcal{C}$, we have $\mathcal{G} \supseteq \sigma(\mathcal{C})$.

$\square$

Corollary 1.21. *If $\Omega_0 \in \sigma(\mathcal{C})$, then*

$$\sigma(\mathcal{C} \cap \Omega_0) = \{A \subseteq \Omega : A \subseteq \Omega_0, A \in \sigma(\mathcal{C})\}$$

Remark 1.22. *Now, take $\Omega = \mathbb{R}$, let $\mathcal{C} = \{(a,b) : -\infty < a < b < \infty\}$ be the generating set, and let $\Omega_0 = (0,1)$. We have*

$$\mathscr{B}((0,1)) = \sigma(\mathcal{C} \cap \Omega_0) = \{A : A \subset (0,1), A \in \mathscr{B}(\mathbb{R})\}$$

2 Probability Spaces

2.1 Basic Definitions and Properties

Definition 2.1 (Measurable space)**.** *A measurable space is a couple* $(\Omega, \mathscr{B})$ *where* Ω *is the whole space, and* $\mathscr{B}$ *is a* σ*-field over* Ω*.*

Definition 2.2 (Measure space)**.** *A measure space is a triple* $(\Omega, \mathscr{B}, \mu)$ *such that*

- $(\Omega, \mathscr{B})$ *is a measurable space.*
- μ *is a measure on* $(\Omega, \mathscr{B})$*. That is,* μ *is a set function with domain* $\mathscr{B}$ *such that*

  (M1) $\mu(\varnothing) = 0$

  (M2) $\mu(A) \geqslant 0$ *for all* $A \in \mathscr{B}$*.*

  (M3) μ *is* σ*-additive: for* $A_n \in \mathscr{B}_n$ *disjoint, we have* $\mu(\cup_{n \geqslant 1} A_n) = \sum_{n \geqslant 1} \mu(A_n)$

Definition 2.3 (Probability space)**.** *A probability space is a triple* $(\Omega, \mathscr{B}, \mathbf{P})$ *such that* $(\Omega, \mathscr{B}, \mathbf{P})$ *is a measure space and*

(P1) $\mathbf{P}(\Omega) = 1$*.*

Remark 2.4. *Most of the time in this course, we deal with probability space. The only exception is the Lebesgue measure.*

Example 2.1. *Flip a coin twice. Then*

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

And $A = \{(H, H)\}$ *is the event that 1st flip is a head, and 2nd flip is also a head.*$\mathscr{B} = \mathcal{P}(\Omega)$*. Then* $\mathbf{P}(A) = 1/4$*.*

Here are some simple consequences of the definition of a probability measure:

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
2. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
3. (Monotonicity): If $A \subseteq B$, then $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.
4. (Subadditivity): $\mathbf{P}(\bigcup_{n \geqslant 1} A_n) \leqslant \sum_{n \geqslant 1} \mathbf{P}(A_n)$
5. (Continuity)
   - If $A_n \uparrow A$ with $A_n, A \in \mathscr{B}$. Then $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$.
   - If $A_n \downarrow A$ with $A_n, A \in \mathscr{B}$. Then $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.

Proof of (1):

- Let $A_1 = A$, $A_2 = A^c$, and $A_3 = A_4 = \ldots = \varnothing$.
- Then using countable additivity

$$
\begin{aligned}
\mathbf{P}(\Omega) &= \mathbf{P}\left(\bigcup_{n\geqslant 1} A_n\right) \\
&= \sum_{n\geqslant 1} \mathbf{P}(A_n) && \text{(countable additivity)} \\
&= \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) + \dots \\
&= \mathbf{P}(A) + \mathbf{P}(A^c) + 0 + \dots \\
&= \mathbf{P}(A) + \mathbf{P}(A^c)
\end{aligned}
$$

Proof of (2):

- Observe that: $A \cup B = (A \backslash B) \cup (A \cap B) \cup (B \backslash A)$.
- Then by finite additivity: $\mathbf{P}(A \cup B) = \mathbf{P}(A \backslash B) + \mathbf{P}(A \cap B) + \mathbf{P}(B \backslash A)$
- Lastly, use $\mathbf{P}(A \backslash B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$.

Proof of (3):

- Because $B = (B \backslash A) \cup A$
- Then $\mathbf{P}(B) = \mathbf{P}(B \backslash A) + \mathbf{P}(A) \geqslant \mathbf{P}(A)$

Proof of (4):

- We can decompose arbitrary union $\cup_{n\geqslant 1} A_n$ into disjoint unions

$$
\bigcup_{n\geqslant 1} A_n = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash (A_1 \cup A_2)) \cup (A_4 \backslash (A_1 \cup A_2 \cup A_3)) \cup \dots
$$

- Then by axiom (3) of a measure, we have

$$
\mathbf{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \mathbf{P}(A_1) + \underbrace{\mathbf{P}(A_2 \backslash A_1)}_{\leqslant \mathbf{P}(A_2)} + \underbrace{\mathbf{P}(A_3 \backslash (A_1 \cup A_2)}_{\leqslant \mathbf{P}(A_3)} + \dots \leqslant \sum_{n\geqslant 1} \mathbf{P}(A_n)
$$

Proof of (5.1):

- Set

$$
B_1 = A_1, \quad B_2 = A_2 \backslash A_1, \quad B_3 = A_3 \backslash A_2, \dots
$$

- These B_n are disjoint, and $B_n \in \mathscr{B}$.
- $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$.
- Also $\cup_{n=1}^{k} B_n = \cup_{n=1}^{k} A_n$ for every $k \geqslant 1$.
- So $\mathbf{P}(A) = \mathbf{P}(\cup_{n\geqslant 1} A_n) = \mathbf{P}(\cup_{n\geqslant 1} B_n) = \sum_{n\geqslant 1} \mathbf{P}(B_n)$.

- Then see that
$$\sum_{n=1}^{\infty} \mathbf{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{P}(B_n) = \lim_{N\to\infty} \mathbf{P}(\bigcup_{n=1}^{N} B_n) = \lim_{N\to\infty} \mathbf{P}(\bigcup_{n=1}^{N} A_n) = \lim_{N\to\infty} \mathbf{P}(A_N)$$
- Hence, $\mathbf{P}(A) = \lim_{N\to\infty} \mathbf{P}(A_N)$.

Proof of (5.2):

- If $A_n \downarrow A$, then $A_n^c \uparrow A^c$.
- Hence $1 - \mathbf{P}(A) = \mathbf{P}(A^c) \uparrow \mathbf{P}(A_n^c) = 1 - \mathbf{P}(A_n)$.
- So $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.

Next we are going to discuss inclusion-exclusion principle. Recall from combinatorics that the techniue to count number of elements of union of two finite sets is

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Consider n houses. And a mailman is trying to deliver n mails to these n houses. Question: What is the probability that this mailman deliver at least one letter correctly?

Define the sample space as

$$\Omega = \{\text{all permutations of } 1, \ldots, n\}$$

Hence, $|\Omega| = n!$.

Let A be the event that at least one letter is correctly delivered.

$$A = \{(\cdot, \ldots, \cdot, i, \ldots, \cdot) : i = 1, \ldots, n\}$$

Then
$$\mathbf{P}(A) = \frac{\#\text{ of elements in } A}{n!}$$
Set $A_k = \{(*, \ldots, k, \ldots, *)\}$ where k is at the k-th spot. Then

$$A = \cup_{k=1}^{n} A_k$$

Thus

$$\begin{aligned}
\mathbf{P}(A_k) &= \frac{(n-1)!}{n!} \\
\mathbf{P}(A_i \cap A_j) &= \frac{(n-2)!}{n!} \\
&\vdots \\
\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_m}) &= \frac{(n-m)!}{n!} \\
&\vdots \\
\mathbf{P}(A_1 \cap \ldots \cap A_n) &= \frac{1}{n!}
\end{aligned}$$

$$
\begin{aligned}
\mathbf{P}(\cup_{k=1}^n A_k) = & \sum_{k=1}^n \mathbf{P}(A_k) - \sum_{1\leqslant i<j\leqslant n} \mathbf{P}(A_i \cap A_j) + \sum_{1\leqslant i<j<k\leqslant n} \mathbf{P}(A_i \cap A_j \cap A_k) - \\
& \ldots + (-1)^{n-1} \sum_{1\leqslant i_1<\ldots<i_n\leqslant n} \mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_n}) \\
= & \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} - \binom{n}{3}\frac{(n-3)!}{n!} + \ldots + \frac{1}{n!}
\end{aligned}
$$

Lecture 2024-09-17

Theorem 2.5 (Fatou's lemma). *1. For any sequence of sets A_n we have*

$$\mathbf{P}(\liminf_{n\to\infty} A_n) \leqslant \liminf_{n\to\infty} \mathbf{P}(A_n) \leqslant \limsup_{n\to\infty} \mathbf{P}(A_n) \leqslant \mathbf{P}(\limsup_{n\to\infty} A_n)$$

2. If A_n has a limit, then

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

Proof of (2): See that if A_n has a limit, then (1) implies (2). Because having a limit means

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \lim_{n\to\infty} A_n$$

Hence, we have

$$\mathbf{P}(\liminf_{n\to\infty} A_n) = \mathbf{P}(\limsup_{n\to\infty} A_n) = \mathbf{P}(\lim_{n\to\infty} A_n)$$

By (1) we have a sequence of equalities

$$\mathbf{P}(\liminf_{n\to\infty} A_n) = \liminf_{n\to\infty} \mathbf{P}(A_n) = \limsup_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(\limsup_{n\to\infty} A_n)$$

And if limsup equals liminf, then

$$\liminf_{n\to\infty} \mathbf{P}(A_n) = \limsup_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

Hence, we conclude

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

Proof of (1): See that

$$
\begin{aligned}
\mathbf{P}(\liminf_{n\to\infty} A_n) &= \mathbf{P}(\cup_{n\geqslant 1} \underbrace{\cap_{k\geqslant n} A_k}_{=:B_n}) \\
&= \mathbf{P}(\cup_{n\geqslant 1} B_n) = \mathbf{P}(\lim_{n\to\infty} B_n) \\
&= \lim_{n\to\infty} \mathbf{P}(B_n), \quad \text{by } B_n \text{ increasing sequence} \\
&= \lim_{n\to\infty} \mathbf{P}(\cap_{k\geqslant n} A_k) \\
&\leqslant \liminf_{n\to\infty} \mathbf{P}(A_n), \quad \text{since } \cap_{k\geqslant n} A_k \subseteq A_n
\end{aligned}
$$

Because $\cap_{k\geqslant n}A_k \subseteq A_n$ for all $n \geqslant 1$. So

$$\mathbf{P}(\cap_{k\geqslant n}A_k) \leqslant \mathbf{P}(A_n) \qquad \forall n \geqslant 1$$

Because the results holds for all $n \geqslant 1$, so it will also holds when n becomes large.

$$\lim_{n\to\infty} \mathbf{P}(\cap_{k\geqslant n}A_n) \leqslant \liminf_{n\to\infty} \mathbf{P}(A_n)$$

The RHS is lim inf is because we we don't know if $\mathbf{P}(A_n)$ has a limit or no.

And

$$\begin{aligned}
\mathbf{P}(\limsup_{n\to\infty} A_n) &= \mathbf{P}(\cap_{n\geqslant 1} \underbrace{\cup_{k\geqslant n}A_k}_{=:C_n}) \\
&= \mathbf{P}(\cap_{n\geqslant 1}C_n) = \mathbf{P}(\lim_{n\to\infty} C_n) \\
&= \lim_{n\to\infty} \mathbf{P}(C_n), \quad \text{ by } C_n \text{ decreasing sequence} \\
&= \lim_{n\to\infty} \mathbf{P}(\cup_{k\geqslant n}A_k) \\
&\geqslant \limsup_{n\to\infty} \mathbf{P}(A_n), \quad \text{ since } \cup_{k\geqslant n} A_k \supseteq A_n
\end{aligned}$$

Example 2.2. *Look at* $(\Omega, \mathscr{B}, \mathbf{P})$*. where* $\Omega = \mathbb{R}$ *and* $\mathscr{B} = \mathscr{B}(\mathbb{R})$*. Set*

$$F(x) := \mathbf{P}((-\infty, x]), \quad x \in \mathbb{R}$$

Properties of $F(x)$ *are*

1. $F(x)$ *is right-continuous. Hint is:* $\lim_{x_n\downarrow x} F(x_n) = F(x)$.
2. $F(x)$ *is non-decreasing*
3. *If you look at* $\lim_{x\uparrow+\infty} F(x) = 1$ *and* $\lim_{x\downarrow-\infty} F(x) = 0$

Proof of (1): Note that because $A_n := (-\infty, x_n]$ is a decreasing sequence of sets such that $A_n \downarrow A$ for $A = (-\infty, x]$. Apply the continuity of probability, we have

$$\begin{aligned}
\lim_{x_n\downarrow x} F(x_n) &= \lim_{n\to\infty} \mathbf{P}((-\infty, x_n]) = \mathbf{P}(\lim_{n\to\infty}(-\infty, x_n]) \\
&= \mathbf{P}((-\infty, x]) = F(x)
\end{aligned}$$

Remark 2.6. $F(x)$ *is in general not left continuous. Counter-example is because*

$$\cup_{n\geqslant 1}(-\infty, x_n] = (-\infty, x)$$

here we have $(-\infty, x)$ *not* $(-\infty, x]$.

Example 2.3 (Non-example)**.** *Consider*

$$A_n = \begin{cases} [0, \frac{1}{2}] & \textit{when } n \textit{ odd} \\ [\frac{1}{2}, 1] & \textit{when } n \textit{ even} \end{cases}$$

and let $\mathbf{P}$ *to be Lebesgue measure. Then*

$$\mathbf{P}(A_n) = \frac{1}{2} \longrightarrow \frac{1}{2}$$

But A_n *does not have a limit.*

2.2 Dynkin's $\pi - \lambda$ Theorem

Lectures 2024-09-19, 2024-09-21

Recall from the previous section the function attached to a probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$,

$$F(x) := \mathbf{P}((-\infty, x])$$

and its three properties: $F(x)$ is right-continuous, $F(x)$ is non-decreasing, and $F(x) \uparrow$ if $x \uparrow \infty$ and $F(x) \downarrow$ if $x \downarrow -\infty$.

Definition 2.7 (Distribution Function). *If $F(x)$ is a function satisfying the above three properties. Then we say $F(x)$ is a distribution function (DF).*

Last time, we have shown starting from a probability measure $\mathbf{P}$ on a measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, we can always get a distribution function by setting

$$F(x) := \mathbf{P}((-\infty, x])$$

Later we will see that starting from a DF $F(x)$, we can always construct a probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $F(x) = \mathbf{P}((-\infty, x])$.

This is a one-to-one correspondence between the following

$$\mathbf{P} \quad \longleftrightarrow \quad F(x)$$

Recall that $\sigma(\mathcal{C})$ the σ-field generated by $\mathcal{C}$ is the smallest (minimal) σ-field containing (or generated by) $\mathcal{C}$.

Definition 2.8 (Minimal structure). *The minimal structure $\mathcal{S}$ generated by a collection of sets $\mathcal{C}$ is a non-empty structure such that*

1) $\mathcal{S}$ will contain $\mathcal{C}$, i.e. $\mathcal{S} \supseteq \mathcal{C}$

2) If there is another structure $\mathcal{S}'$ containing $\mathcal{C}$, then $\mathcal{S}'$ also contains $\mathcal{S}$, i.e. $\mathcal{S}' \supseteq \mathcal{S}$

Remark 2.9. *The minimal structure generated by $\mathcal{C}$ is obtained by taking intersection of all structures*

$$\bigcap_{\substack{\mathcal{S}' \text{ is a structure,} \\ \mathcal{S}' \supset \mathcal{C}}} \mathcal{S}'$$

Definition 2.10 (π-system). *A π-system $\mathcal{P}$ is a collection of subsets of Ω that is closed under finite intersections. That is*

$$\text{if } A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$$

Definition 2.11 (λ-system). *A λ system $\mathcal{L}$ is a collection of subsets of Ω such that*

	old	*new*
(1)	$\Omega \in \mathcal{L}$	$\Omega \in \mathcal{L}$
(2)	*if* $A, B \in \mathcal{L}$ *and* $A \subset B$ *then* $B \backslash A \in \mathcal{L}$	*if* $A \in \mathcal{L}$ *then* $A^c \in \mathcal{L}$
(3)	*if* $A_n \subset \mathcal{L}$ *and* $A_n \uparrow$, *then* $\cup_{n=1}^{\infty} A_n \in \mathcal{L}$	*suppose* $A_n \in \mathcal{L}$ *and* $A_i \cap A_j = \varnothing$ *for* $i \neq j$ *then* $\cup_{n=1}^{\infty} A_n \in \mathcal{L}$

Remark 2.12.

- *old* $\iff$ *new*
- *A σ-field is always a λ-system*

Theorem 2.13 (Dynkin's $\pi - \lambda$ theorem)**.** *(a) If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subseteq \mathcal{L}$. Then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

(b) If $\mathcal{P}$ is a π-system. Then $\sigma(\mathcal{P}) = \lambda(\mathcal{P})$.

If $\mathcal{P}$ is a π-system, then we can use this π-system to generated a σ-field. We can also use this π-system to generated a λ-system. I.e. $\sigma(\mathcal{P})$ is a σ-field, and $\lambda(\mathcal{P})$ is a λ-system. And we have

$$\sigma(\mathcal{P}) = \lambda(\mathcal{P})$$

Remark 2.14. *This theorem means what keep λ-system from being a σ-field is closure under finite intersection.*

Application of Dynkin

Proposition 2.15. *Suppose $\mathbf{P}_1$ and $\mathbf{P}_2$ are two probability measures on $(\Omega, \mathscr{B})$. Let*

$$\mathcal{L} := \{B \in \mathscr{B} : \mathbf{P}_1(B) = \mathbf{P}_2(B)\}$$

Then this is a λ-system.

Proof.

1. $\Omega \in \mathcal{L}$ because $\mathbf{P}_1(\Omega) = \mathbf{P}_2(\Omega) = 1$
2. If $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$. Because $\mathbf{P}_1(A^c) = 1 - \mathbf{P}_1(A) = 1 - \mathbf{P}_2(A) = \mathbf{P}_2(A^c)$
3. If $A_n \in \mathcal{L}$ and pairwise disjoint. Then $\cup_n A_n \in \mathcal{L}$. Because by σ-additivity of probability measure and $\mathbf{P}_1(A_n) = \mathbf{P}_2(A_n)$, we have

$$\mathbf{P}_1(\cup_n A_n) = \sum_{n \geqslant 1} \mathbf{P}_1(A_n) = \sum_{n \geqslant 1} \mathbf{P}_2(A_n) = \mathbf{P}_2(\cup_n A_n)$$

□

Corollary 2.16 (Corollary of Dynkin)**.** *Suppose we have two probability measures $\mathbf{P}_1$ and $\mathbf{P}_2$ on the same measurable space $(\Omega, \mathcal{F})$. Suppose $\mathcal{P}$ is a π-system such that*

$$\mathbf{P}_1(B) = \mathbf{P}_2(B), \qquad \text{for all } B \in \mathcal{P}$$

Another way to say above statement is $\mathbf{P}_1 = \mathbf{P}_2$ on $\mathcal{P}$. Then $\mathbf{P}_1 = \mathbf{P}_2$ on $\sigma(\mathcal{P})$.

Proof. Set

$$\mathcal{L} := \{B \in \mathcal{F} : \mathbf{P}_1(B) = \mathbf{P}_2(B)\}$$

By proposition, $\mathcal{L}$ is a λ-system. By the assumption, $\mathcal{L}$ contains $\mathcal{P}$, i.e. $\mathcal{L} \supset \mathcal{P}$. By Dynkin's theorem, $\mathcal{L} \supset \sigma(\mathcal{P})$. □

Corollary 2.17 (Corollary of Corollary)**.** *Suppose we have two probability measures* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are two probability measures on* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *such that they corresponds to the same distribution function. That is, for following two DFs*

$$F_1(x) = \mathbf{P}_1((-\infty, x]), \qquad F_2(x) = \mathbf{P}_2((-\infty, x])$$

we have $F_1(x) = F_2(x)$*. Then*

$$\mathbf{P}_1 \equiv \mathbf{P}_2 \qquad \text{on } \mathscr{B}(\mathbb{R})$$

Proof. Set

$$\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}$$

Claim: $\mathcal{P}$ **is a** π**-system.** Obvious.

Claim: Two probability measures $\mathbf{P}_1$ **and** $\mathbf{P}_2$ **are the same on this** π**-system** $\mathcal{P}$**.** This is by assumption.

Hence, $\mathbf{P}_1 = \mathbf{P}_2$ on $\sigma(\mathcal{P}) = \mathscr{B}(\mathbb{R})$. □

Proof of Dynkin's π**-**λ **theorem**

Proof (a) $\implies$ **(b):** Assume part (a) is true. First note that

- $\sigma(\mathcal{P})$ is a λ-system
- $\sigma(\mathcal{P}) \supset \mathcal{P}$

Then $\sigma(\mathcal{P}) \supseteq \lambda(\mathcal{P})$.

Second,

- $\lambda(\mathcal{P})$ is a λ-system
- $\lambda(\mathcal{P}) \supseteq \mathcal{P}$

Hence, we have $\lambda(\mathcal{P}) \supseteq \sigma(\mathcal{P})$.

In order to prove Dynkin's theorem, we only need to prove (a).

Lemma 2.18. *If* $\mathcal{L}$ *is both a* λ*-system and a* π*-system, then it is a* σ*-field.*

Proof. All we have to show is, for $A_n \in \mathcal{L}$ an arbitrary sequence of sets (not necessarily disjoint), we have $\cup_n A_n \in \mathcal{L}$.

The strategy is to use disjointification. Set

$$\begin{aligned}
&B_1 := A_1\\
&B_2 := A_2 \backslash A_1 = A_2 \cap A_1^c\\
&B_3 := A_3 \backslash (B_1 \cup B_2) = (A_3 \backslash B_1) \backslash B_2 = A_3 \cap B_2^c \cap B_1^c\\
&\quad \vdots\\
&B_n := A_n \backslash (\cup_{k=1}^{n-1} B_k) = A_n \cap (\cap_{k=1}^{n-1} B_k^c)\\
&\quad \vdots
\end{aligned}$$

Then under this construction, all B_n's are pairwise disjoint. And we have

$$\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k$$

We have $\mathcal{L}$ is a π-system, hence $B_n \in \mathcal{L}$ for each n because each B_n is a finite intersection of sets in $\mathcal{L}$. Lastly, by definition of λ-system, $\cup_{k=1}^{\infty} B_k \in \mathcal{L}$. Hence, $\mathcal{L}$ is a σ-field. □

Remark 2.19. *This lemma tells us what prevent a λ-system from being a σ-algebra. The thing is λ-system is not closed under finite intersection.*

Definition 2.20. *For each $A \in \lambda(\mathcal{P})$, define*

$$\mathcal{G}_A := \{B \in \lambda(\mathcal{P}) : A \cap B \in \lambda(\mathcal{P})\}$$

to be the collection of sets in λ-system $\lambda(\mathcal{P})$ that is closed under finite intersection.

Proof of (a): It suffices to show that $\lambda(\mathcal{P})$ is a π-system. Because if it is true, then by the Lemma, $\lambda(\mathcal{P})$ is a σ-field. Hence,

$$\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P}) \subseteq \mathcal{L}$$

Prove $\lambda(\mathcal{P})$ is a π-system.

Claim 2.21. *If $A \in \lambda(\mathcal{P})$, then $\mathcal{G}_A$ is a λ-system.*

We check 3 axioms of λ-system

1. $\Omega \in \mathcal{G}_A$. Because $A \cap \Omega = A \in \lambda(\mathcal{P})$

2. If $B \in \mathcal{G}_A$, i.e. $A \cap B \in \lambda(\mathcal{P})$. We need to show $A \cap B^c \in \lambda(\mathcal{P})$. See that $A \cap B^c = A \backslash B = A \backslash (A \cap B) \in \lambda(\mathcal{P})$, because $A, A \cap B \in \lambda(\mathcal{P})$ and $A \cap B \subseteq A$, and we use axiom 2 of the old version of λ-system.

3. If $B_1, \ldots, B_n, \ldots \in \mathcal{G}_A$ and all pairwise disjoint. We want to show $\cup_n B_n \in \mathcal{G}_A$. Which means $(\cup_n B_n) \cap A \in \lambda(\mathcal{P})$. See that

$$\left(\bigcup_n B_n\right) \cap A = \bigcup_{n=1}^{\infty} (B_n \cap A) \in \lambda(\mathcal{P})$$

because $B_n \cap A$ are all pairwise disjoint. Hence by axiom 3 of new version of λ-system, we conclude this set is in $\lambda(\mathcal{P})$.

Claim 2.22. *If $A \in \mathcal{P}$, then $\mathcal{G}_A \supseteq \lambda(\mathcal{P})$.*

This claim is equivalent to the statement: For any $A \in \mathcal{P}$ and $B \in \lambda(\mathcal{P})$, we have $A \cap B \in \lambda(\mathcal{P})$.

First note that if $A \in \mathcal{P}$, then $\mathcal{G}_A \supseteq \mathcal{P}$. Because for any $B \in \mathcal{P}$, and we consider $A \cap B \in \mathcal{P}$ by definition of π-system. Hence $A \cap B \in \lambda(\mathcal{P})$. By claim 1, $\mathcal{G}_A$ is a λ-system. Hence, we conclude $\mathcal{G}_A \supseteq \lambda(\mathcal{P})$.

Claim 2.23. *If $A \in \lambda(\mathcal{P})$, then $\mathcal{G}_A \supseteq \lambda(\mathcal{P})$.*

This claim is equivalent to the statement: For any $A \in \lambda(\mathcal{P})$ and $B \in \lambda(\mathcal{P})$, we have $A \cap B \in \lambda(\mathcal{P})$. Which is equivalent to say $\lambda(\mathcal{P})$ is a π-system.

- Take $A \in \lambda(\mathcal{P})$. Note by claim 2, for any $B \in \mathcal{P}$, we have $A \cap B \in \lambda(\mathcal{P})$.
- So $\mathcal{G}_A \supseteq \mathcal{P}$.
- By claim 1, $\mathcal{G}_A$ is a λ-system.
- By definition of generated λ-system of $\mathcal{P}$ being the smallest λ-system contains $\mathcal{P}$, we have $\mathcal{G}_A \supseteq \lambda(\mathcal{P})$.

2.3 Two Constructions

Lecture 2024-09-24

There are two cases.

Case 1

Suppose that the sample space $\Omega = \{x_n : n \in \mathbb{Z}_{\geqslant 1}\}$ is discrete/countable. In this case, we don't have to worry about σ-field, because we can always set it to be the power set $\mathcal{P}(\Omega)$, i.e. Set

$$\mathscr{B} := \mathcal{P}(\Omega)$$

We can assign measures on $(\Omega, \mathscr{B})$ as follows: For any sequence $P_1, P_2, \ldots, P_n, \ldots$ such that

1. $P_i \geqslant 0$ for all i.
2. $\sum_{i=1}^{\infty} P_i = 1$

We can assign P_i to x_i. Then for any $A \subset \Omega$, we have

$$\mathbf{P}(A) = \sum_{x_i \in A} P_i$$

Example 2.4. *Let $A = \{x_5, x_6, x_{10}\}$. Then $\mathbf{P}(A) = P_5 + P_6 + P_{10}$. This gives us a probability measure on $(\Omega, \mathscr{B})$.*

Check $\mathbf{P}$ is a probability measure. We only check σ-additivity. Let $A_n \subset \Omega$ be disjoint sequence of sets. Then we want to show $\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$. Let's check

$$\begin{aligned}
\mathbf{P}(\cup_{n=1}^{\infty} A_n) &= \sum_{x_i \in \cup_{n=1}^{\infty} A_n} P_i \\
&= \sum_{n=1}^{\infty} \sum_{x_i \in A_n} P_i, \qquad \text{can change order of summation because all } P_i \geqslant 0 \\
&= \sum_{n=1}^{\infty} \mathbf{P}(A_n)
\end{aligned}$$

Example 2.5 (Change of the order of summation matters). *Consider summation of harmonic series*

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots=\log_e(2)$$

But the sum of positive part sums to ∞, and sum of negative parts sums to $-\infty$.

Example 2.6. *Consider flip a coin n-times. The the sample space will be*

$$\Omega := \{\underbrace{(*,\ldots,*)}_{n \text{ of them}} : * = \ H \text{ or } T\}$$

We assign probability as follows

$$(\omega_1,\omega_2,\ldots,\omega_n) \longleftrightarrow (p^{\omega_1}q^{1-\omega_1})\ldots(p^{\omega_n}q^{1-\omega_n}) = \prod_{k=1}^{n} p^{\omega_k}q^{1-\omega_k}$$

where $\omega = 0$ or 1, where $1 \leftrightarrow\ H$ and $0 \leftrightarrow\ T$. We want

$$\mathbf{P}(H)=p, \qquad \mathbf{P}(T)=1-p=q, \qquad 0<p<1$$

To see this indeed gives us a probability measure, we only need to show

$$\sum_{(\omega_1,\ldots,\omega_n)\in\Omega}\prod_{k=1}^{n} p^{\omega_k}q^{1-\omega_k} = 1$$

Compute this, we have

$$\begin{aligned}
\sum_{(\omega_1,\ldots,\omega_n)\in\Omega}\prod_{k=1}^{n} p^{\omega_k}q^{1-\omega_k} &= \sum_{\omega_k=0 \text{ or } 1; k=1,\ldots,n}\prod_{k=1}^{n} p^{\omega_k}q^{1-\omega_k}\\
&= \sum_{\omega_k=0 \text{ or } 1; k=1,\ldots,n-1}\ \sum_{\omega_n=0 \text{ or } 1}\left(\prod_{k=1}^{n-1} p^{\omega_k}q^{1-\omega_k}\right)(p^{\omega_n}q^{1-\omega_n})\\
&= \sum_{\omega_k=0 \text{ or } 1; k=1,\ldots,n-1}\prod_{k=1}^{n-1} p^{\omega_k}q^{1-\omega_k}(q+p)\\
&= \sum_{\omega_k=0 \text{ or } 1; k=1,\ldots,n-1}\prod_{k=1}^{n-1} p^{\omega_k}q^{1-\omega_k}\\
&= \vdots \qquad \textit{(repeat this procedure)}\\
&= 1
\end{aligned}$$

Case 2

Suppose that the sample space Ω is continuous/uncountable.

Example 2.7. *Consider the sample space to be $\Omega = \mathbb{R}$. Suppose $F(x)$ is a DF. Then how to construct a probability measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ such that*

$$\mathbf{P}((-\infty,x]) = F(x)$$

The question is, starting from the DF $F(x)$, how do we construct a probability measure $\mathbf{P}$?

Example 2.8. *Flip a coin infinitely many times. Then the sample space is*

$$\Omega := \{(\omega_1, \ldots, \omega_n, \ldots) : \omega_i = 0 \text{ or } 1\}$$

we want

$$\mathbf{P}(H) = p, \qquad \mathbf{P}(T) = q = 1 - p$$

Then see that

$$(H, H, \ldots) \longleftrightarrow p \cdot p \cdot \ldots = 0$$

We need some need way to assign probability measures.

Example 2.9. *Consider $\Omega = (0, 1]$. This probability $\mathbf{P}(*) = 0$ for any $* \in (0, 1]$.*

2.4 Constructions of Probability Spaces

Lectures 2024-09-24, 2024-09-26, 2024-09-28, 2024-10-01, 2024-10-03, 2024-10-05

Definition 2.24 (σ-additive set function)**.** *Suppose $\mathcal{G} \subset \mathcal{P}(\Omega)$ is a collection of subsets of Ω. Consider a set function*

$$\mathbf{P} : \mathcal{G} \longrightarrow [0, 1]$$

We say that $\mathbf{P}$ is σ-additive if for any sequence of disjoint subsets $A_n \in \mathcal{G}$ and $\cup_{n=1}^{\infty} A_n \in \mathcal{G}$. We have

$$\mathbf{P}(\cup_{n \geqslant 1} A_n) = \sum_{n \geqslant 1} \mathbf{P}(A_n)$$

Remark 2.25. *Suppose $A_n \subset \Omega$ is a sequence of subsets. We will use*

$$\sum_{n \geqslant 1} A_n \text{ for } A_n \text{ are disjoint}$$

Definition 2.26 (Extension of set functions)**.** *Consider two collections of subsets $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{P}(\Omega)$, where $\mathcal{G}_1 \subset \mathcal{G}_2$. Also consider two set functions*

$$\begin{array}{rcl} \mathbf{P}_1 : & \mathcal{G}_1 & \longrightarrow & [0, 1] \\ \mathbf{P}_2 : & \mathcal{G}_2 & \longrightarrow & [0, 1] \end{array}$$

We say $\mathbf{P}_2$ is an extension of $\mathbf{P}_1$ and write

$$\mathbf{P}_2 |_{\mathcal{G}_1} = \mathbf{P}_1$$

if

$$\mathbf{P}_2(A) = \mathbf{P}_1(A) \quad \text{for all } A \in \mathcal{G}_1$$

General construction of measures

Definition 2.27 (Semi-algebra)**.** *Let $\mathcal{S} \subset \mathcal{P}(\Omega)$ be a collection of subsets of Ω. We say $\mathcal{S}$ is a semi-algebra if*

1. $\varnothing, \Omega \in \mathcal{S}$
2. *Closed under finite intersection.*
3. *If $S \in \mathcal{S}$. Then $S^c = \sum_{i=1}^n S_i$ which means S^c can be written as a finite disjoint union of $S_i \in \mathcal{S}$.*

Example 2.10. *Consider $\mathcal{S} = \{(a,b] : -\infty \leqslant a \leqslant b \leqslant \infty\}$. There are several conventions*

- $(a,a] = \varnothing$
- $(a,\infty] = (a,\infty)$

Then

$$(a,b]^c = (-\infty,a] \cup (b,\infty) = (-\infty,a] \cup (b,\infty]$$

Example 2.11. *As an example of the general construction. Consider a DF $F(x)$. We want to construct a Probability Measure on the measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that*

$$\mathbf{P}((-\infty,x]) = F(x)$$

Then

$$\mathbf{P}((a,b]) = F(b) - F(a)$$

Lemma 2.28. *Suppose $\mathcal{S}$ is a semi-algebra. Then the Algebra generated by this semi-algebra $\mathcal{S}$, which is denoted as $\mathcal{A}(\mathcal{S})$ is defined as*

$$\mathcal{A}(\mathcal{S}) := \left\{ \sum_{i\in I} S_i : S_i \in \mathcal{S} \textit{ and } I \textit{ is a finite index set} \right\}$$

Note that the sets in $\mathcal{A}(\mathcal{S})$ are finite disjoint unions of sets in $\mathcal{S}$.

Proof. Denote the RHS of above by Λ, i.e.

$$\Lambda = \left\{ \sum_{i\in I} S_i : S_i \in \mathcal{S} \text{ and } I \text{ is a finite index set} \right\}$$

We check this is an Algebra/Field.

1. $\varnothing \in \Lambda$ because $\varnothing \in \mathcal{S} \subset \Lambda$.
2. Closed under finite intersection. Because $A_1, A_2 \in \Lambda$ where $A_i = \sum_k S_k^i$, we have

$$A_1 \cap A_2 = \left(\sum_{k=1}^n S_k^1\right) \cap \left(\sum_{l=1}^m S_l^2\right) = \sum_k \sum_l S_k^1 \cap S_l^2 \in \Lambda$$

3. If $A \in \Lambda$. Then $A^c \in \Lambda$ because if consider $A = \sum_i S_i$. Then

$$A^c = \left(\sum_{k=1}^{n} S_k\right)^c = \bigcap_{i=1}^{n} S_i^c = \bigcap_{i=1}^{n}\left(\sum_{j=1}^{m} S_{ij}\right) \in \Lambda$$

   because we know $S_i \in \mathcal{S}$, hence $S_i^c = \sum_{j=1}^{m} S_{ij}$ for some $S_{ij} \in \mathcal{S}$. Also see that $S_i^c \in \Lambda$ because it is a disjoint union of sets in $\mathcal{S}$. Using (2), Λ is closed under finite intersection, and we conclude $A^c \in \Lambda$.

Show $\Lambda \supseteq \mathcal{A}(\mathcal{S})$. Hence, we showed and know

- Λ is a field (showed)
- $\Lambda \supset \mathcal{S}$ (know)

Hence, $\Lambda \supseteq \mathcal{A}(\mathcal{S})$ by definition of the smallest Algebra/Field.
Show $\Lambda \subseteq \mathcal{A}(\mathcal{S})$. Next, we want to show $\Lambda \subseteq \mathcal{A}(\mathcal{S})$. For any $\sum_i S_i \in \Lambda$. It is trivial that $\sum_i S_i \in \mathcal{A}(\mathcal{S})$. Because each $S_i \in \mathcal{S} \subset \mathcal{A}(\mathcal{S})$ and $\mathcal{A}(\mathcal{S})$ is a field. Hence it is closed under finite union. □

Definition 2.29 (Probability measure)**.** *Let $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ be a collection of subsets of Ω. We say $\mathbf{P}$ is a Probability Measure on $\mathcal{C}$ if $\mathbf{P}$ is a set function of $\mathcal{C}$, i.e. $\mathbf{P} : \mathcal{C} \to [0, 1]$ such that the following holds*

1) $\mathbf{P}(\Omega) = 1$

2) $\mathbf{P}(C) \geqslant 0$ *for all* $C \in \mathcal{C}$

3) (σ-additivity) For disjoint $C_1, \ldots, C_n, \ldots \in \mathcal{C}$ with $\cup_{n\geqslant 1} C_n \in \mathcal{C}$, then

$$\mathbf{P}(\cup_{i\geqslant 1} C_i) = \sum_{i\geqslant 1} \mathbf{P}(C_i)$$

Theorem 2.30 (First extension)**.** *Suppose $\mathbf{P}$ is a probability measure on a Semi-algebra $\mathcal{S}$. There is a unique extension of $\mathbf{P}$ to a probability measure $\mathbf{P}'$ on a bigger domain $\mathcal{A}(\mathcal{S})$. And for any*

$$A = \sum_{i\in I} S_i \in \mathcal{A}(\mathcal{S})$$

we define $\mathbf{P}'$ as

$$\mathbf{P}'(A) = \sum_{i\in I} \mathbf{P}(S_i)$$

Theorem 2.31 (Second Extension)**.** *Suppose $\mathbf{P}$ is a Probability Measure on an Algebra/Field $\mathcal{A}$. Then there is a unique extension of $\mathbf{P}$ to a probability measure $\mathbf{P}'$ on the σ-algebra (σ-field) generated by $\mathcal{A}$, which is denoted as $\sigma(\mathcal{A})$.*

Example 2.12. *Consider $(\Omega, \mathscr{B}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Consider the semi-algebra*

$$\mathcal{S} = \{(a, b] : -\infty \leqslant a \leqslant b \leqslant +\infty\}$$

For any DF $F(x)$. Set

$$\mathbf{P}((a,b]) := F(b) - F(a)$$

By First extension, we can extend $\mathbf{P}$ to a probability measure $\mathbf{P}'$ on $\mathcal{A}(\mathcal{S})$.

By Second extension, we can further extend $\mathbf{P}'$ on $\mathcal{A}(\mathcal{S})$ to a probability measure $\mathbf{P}''$ on $\sigma(\mathcal{A}(\mathcal{S})) = \sigma(\mathcal{S}) = \mathscr{B}(\mathbb{R})$.

In particular, $\mathbf{P}''$ is an extension of $\mathbf{P}$, so

$$\mathbf{P}''((a,b]) = \mathbf{P}((a,b]) := F(b) - F(a)$$

Take $a = -\infty$, we have

$$\mathbf{P}''((-\infty,b]) = F(b)$$

Proof of First Extension: Mostly we need to show $\mathbf{P}'$ is a Probability measure on $\mathcal{A}(\mathcal{S})$.

First thing we need to show is $\mathbf{P}'$ is a well-defined set function, i.e. we need to show that no matter which disjoint union decomposition of a set A we use, we will always have the equality. So suppose we have two different disjoint union decomposition of A

$$A = \sum_j S'_j = \sum_i S_i$$

then we want

$$\sum_j \mathbf{P}(S'_j) = \sum_i \mathbf{P}(S_i)$$

See that

$$\begin{aligned}
\sum_j \mathbf{P}(S'_j) &= \sum_j \mathbf{P}(S'_j \cap A) \\
&= \sum_j \mathbf{P}(S'_j \cap A) \\
&= \sum_j \mathbf{P}(S'_j \cap (\sum_i S_i)) \\
&= \sum_j \mathbf{P}(\sum_i \underbrace{(S'_j \cap S_i)}_{\in \mathcal{S}}) \\
&= \sum_j \sum_i \mathbf{P}(S'_j \cap S_i)
\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_i \mathbf{P}(S_i) &= \sum_i \mathbf{P}(S_i \cap A) \\ &= \sum_i \mathbf{P}(S_i \cap A) \\ &= \sum_i \mathbf{P}(S_i \cap (\sum_j S_j^{'})) \\ &= \sum_i \mathbf{P}(\sum_j \underbrace{(S_j^{'} \cap S_i)}_{\in \mathcal{S}}) \\ &= \sum_i \sum_j \mathbf{P}(S_j^{'} \cap S_i)\end{aligned}$$

Hence, we conclude

$$\sum_j \mathbf{P}(S_j^{'}) = \sum_i \mathbf{P}(S_i) = \sum_i \sum_j \mathbf{P}(S_j^{'} \cap S_i)$$

We need to show $\mathbf{P}^{'}$ is a probability measure, i.e. need to check 3 axioms of probability measure. The first two are trivial, we only need to check σ-additivity. Suppose $A_n \in \mathcal{A}(\mathcal{S})$ and disjoint, and $\cup_{n \geqslant 1} A_n \in \mathcal{A}(\mathcal{S})$. Need to show

$$\mathbf{P}^{'}(\cup_{n \geqslant 1} A_n) = \sum_{n \geqslant 1} \mathbf{P}^{'}(A_n)$$

Since $A_n \in \mathcal{A}(\mathcal{S})$, thus

$$A_n = \sum_{j=1}^{k} S_{nj} \qquad \text{for some } S_{nj} \in \mathcal{S}$$

Denote $A = \cup_{n \geqslant 1} A_n \in \mathcal{A}(\mathcal{S})$. Then A has the decomposition

$$A = \sum_k S_k$$

Let's first do the LHS,

$$\mathbf{P}^{'}(A) = \sum_k \mathbf{P}(S_k)$$

Here

$$\begin{aligned}S_k = S_k \cap A &= S_k \cap \left(\sum_{n \geqslant 1} A_n \right) \\ &= S_k \cap \left(\sum_{n \geqslant 1} \sum_{j \text{ finitely many}} S_{nj} \right) \\ &= \sum_{n \geqslant 1} \sum_j \underbrace{(S_{nj} \cap S_k)}_{\in \mathcal{S}}\end{aligned}$$

Hence, by σ-additivity of $\mathbf{P}$, we have

$$\begin{aligned}\mathbf{P}^{'}(A) &= \sum_k \mathbf{P}(S_k)\\ &= \sum_k \sum_{n\geqslant 1} \sum_j \mathbf{P}(S_{nj} \cap S_k)\\ &= \sum_{n\geqslant 1} \sum_j \sum_k \mathbf{P}(S_{nj} \cap S_k)\end{aligned}$$

Also, see that

$$\sum_k S_{nj} \cap S_k = S_{nj} \cap \sum_k S_k = S_{nj} \cap A = S_{nj}$$

Hence,

$$\begin{aligned}\mathbf{P}^{'}(A) &= \sum_{n\geqslant 1} \sum_j \sum_k \mathbf{P}(S_{nj} \cap S_k)\\ &= \sum_{n\geqslant 1} \sum_j \mathbf{P}(S_{nj})\\ &= \sum_{n\geqslant 1} \mathbf{P}(A_n)\end{aligned}$$

The last equality is by definition of $\mathbf{P}^{'}$. This completes the proof.

Lemma 2.32. *If two sequences A_n, B_n are both monotone increasing sequences in $\mathcal{A}$, and $\cup_{n\geqslant 1} A_n \subset \cup_{n\geqslant 1} B_n$. Then*

$$\lim_{n\to\infty} \mathbf{P}(A_n) \leqslant \lim_{n\to\infty} \mathbf{P}(B_n)$$

Also, if $\cup_{n\geqslant 1} A_n \supset \cup_{n\geqslant 1} B_n$, then we have

$$\lim_{n\to\infty} \mathbf{P}(A_n) \geqslant \lim_{n\to\infty} \mathbf{P}(B_n)$$

Proof. In the first case, note $A_n \cap B_m \uparrow A_n$ as $m \to \infty$. Then

$$\lim_{m\to\infty} \mathbf{P}(A_n \cap B_m) = \mathbf{P}(A_n)$$

On the other hand, $B_m \supset A_n \cap B_m$ for all $n \in \mathbb{Z}_{\geqslant 1}$. Then

$$\lim_{n\to\infty} \mathbf{P}(A_n \cap B_m) \leqslant \mathbf{P}(B_m)$$

Hence, if we send n to ∞, we will have

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \lim_{m\to\infty} \mathbf{P}(A_n \cap B_m) \leqslant \lim_{m\to\infty} \mathbf{P}(B_m)$$

This completes the proof $\square$

Proof of Second Extension:
Step 1. We extend $\mathbf{P}$ to a set function π on $\mathcal{G}$, so that π is a probability measure on $\mathcal{G}$. (Here, $\mathcal{G}$ is not a σ-field/algebra).
Define $\mathcal{G}$. Set

$$\mathcal{G} := \{\cup_{i=1}^{\infty} A_i : A_i \in \mathcal{A}\} \equiv \left\{\lim_{n\to\infty} \uparrow B_n : B_n \in \mathcal{A}, \uparrow \text{ means increasing limit}\right\}$$

By the first definition, we have $\mathcal{G} \supset \mathcal{A}$. Because of the second definition, the sequence B_n will be called an approximating sequence of an element in $\mathcal{G}$.

For any $G \in \mathcal{G}$, define the set function π as

$$\pi(G) = \lim_{n\to\infty} \mathbf{P}(B_n)$$

Using Lemma above, π is a well-defined set function.
Properties of π on $\mathcal{G}$:

1. If $G_1 \subset G_2$. Then $\pi(G_1) \leqslant \pi(G_2)$.
2. π is an extension of $\mathbf{P}$. This means

$$\pi(A) = \mathbf{P}(A), \qquad \text{for any } A \in \mathcal{A}$$

   In particular, $\pi(\varnothing) = 0$ and $\pi(\Omega) = 1$.
3. If $G_1, G_2 \in \mathcal{G}$, then $G_1 \cap G_2, G_1 \cup G_2 \in \mathcal{G}$. Moreover,

$$\pi(G_1 \cap G_2) + \pi(G_1 \cup G_2) = \pi(G_1) + \pi(G_2)$$

Proof of 3. Define

$$\begin{aligned} G_1 &:= \lim_{n\to\infty} \uparrow A_n, \qquad A_n \in \mathcal{A} \\ G_2 &:= \lim_{n\to\infty} \uparrow B_n, \qquad B_n \in \mathcal{A} \end{aligned}$$

By the definition of Algebra $\mathcal{A}$, we have $A_n \cap B_n \in \mathcal{A}$ for all $n \in \mathbb{Z}_{\geqslant 1}$. And

$$A_n \cap B_n \uparrow G_1 \cap G_2 \in \mathcal{G}$$

Also, we have $A_n \cup B_n \in \mathcal{A}$ for all $n \in \mathbb{Z}_{\geqslant 1}$. And

$$A_n \cup B_n \uparrow G_1 \cup G_2 \in \mathcal{G}$$

Next, note

$$\mathbf{P}(A_n \cup B_n) + \mathbf{P}(A_n \cap B_n) = \mathbf{P}(A_n) + \mathbf{P}(B_n)$$

Send n to ∞, we have

$$\pi(G_1 \cup G_2) + \pi(G_1 \cap G_2) = \pi(G_1) + \pi(G_2)$$

□

The next lecture adds two more properties of π and $\mathcal{G}$:

4) Since $G_1 \cap G_2 = \varnothing$, then

$$\pi(G_1 \cup G_2) = \pi(G_1) + \pi(G_2)$$

5) If $G_n \in \mathcal{G}$ and $G_n \uparrow G$, Then $G \in \mathcal{G}$ and $\pi(G) = \lim_{n\to\infty} \pi(G_n)$

Proof of the last property. Consider an increasing sequence $G_1., G_2, \ldots, G_n, \ldots \in \mathcal{G}$ where $G_n \uparrow G$. Then for each G_i, it has its own approximating sequence $B_{i,k} \uparrow G_i$ as $k \to \infty$.

$$\begin{array}{cccccccc}
G_1 & G_2 & G_3 & \ldots & G_n & \ldots & G_m & \ldots \\
\hline
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
B_{n,1} & B_{n,2} & B_{n,3} & \ldots & B_{n,n} & \ldots & B_{m,n} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
B_{3,1} & B_{3,2} & B_{3,3} & \ldots & B_{3,n} & \ldots & B_{3,n} & \ldots \\
B_{2,1} & B_{2,2} & B_{2,3} & \ldots & B_{2,n} & \ldots & B_{2,n} & \ldots \\
B_{1,1} & B_{1,2} & B_{1,3} & \ldots & B_{1,n} & \ldots & B_{1,n} & \ldots
\end{array}$$

Define

$$\begin{aligned}
D_m &:= \cup_{i=1}^{m} B_{i,m} \\
&\vdots \\
D_n &:= \cup_{i=1}^{n} B_{i,n} \\
&\vdots \\
D_3 &:= \cup_{i=1}^{3} B_{i,3} \\
D_2 &:= \cup_{i=1}^{2} B_{i,2} \\
D_1 &:= B_{1,1}
\end{aligned}$$

From this definition, we know

- $D_n \in \mathcal{A}$ for each n
- D_n is an increasing sequence because $(B_{i,n})_{n\geqslant 1}$ is an increasing sequence in n for all i.
- $B_{n,m} \subset D_m \subset G_m$ for $m \geqslant n$
- Send m to ∞, we have $G_n \subset \lim_{m\to\infty} D_m \subset G$
- Send n to ∞, we have $G \subset \lim_{m\to\infty} D_m \subset G$
- Thus, $\lim_{m\to\infty} D_m = G$.
- Hence, $G \in \mathcal{G}$ by definition of G.

From $B_{n,m} \subset D_m \subset G_m$, since π and $\mathbf{P}$ agree on $\mathcal{A}$ and π is monotone on $\mathcal{G}$, we have

$$\mathbf{P}(B_{n,m}) \leqslant \mathbf{P}(D_m) \leqslant \pi(G_m)$$

Send m to ∞, we have the following inequalites for each $n \in \mathbb{Z}_{\geqslant 1}$:

$$\pi(G_n) \leqslant \lim_{m\to\infty} \mathbf{P}(D_m) \leqslant \lim_{m\to\infty} \pi(G_m)$$

Send n to ∞, we have

$$\pi(G) = \lim_{n\to\infty} \pi(G_n) \leqslant \lim_{m\to\infty} \mathbf{P}(D_m) \leqslant \lim_{m\to\infty} \pi(G_m) = \pi(G)$$

Thus,

$$\lim_{m\to\infty} \mathbf{P}(D_m) = \pi(G)$$

□

Remark 2.33. *π is σ-additive on $\mathcal{G}$ because for any sequence $G_n \in \mathcal{G}$ disjoint, by $\mathcal{G}$ being closed under finite union, we have for $G_n' = \cup_{k=1}^n G_k \in \mathcal{G}$ where $(G_n')_{n\geqslant 1}$ is an increasing sequence of sets. By the continuity of π on $\mathcal{G}$, we have*

$$\begin{aligned}
\pi(\cup_{n=1}^{\infty} G_n) &= \pi(\lim_{n\to\infty} G') = \lim_{n\to\infty} \pi(G_n') \\
&= \lim_{n\to\infty} \pi(\cup_{k=1}^n G_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \pi(G_k) \\
&= \sum_{k=1}^{\infty} \pi(G_k)
\end{aligned}$$

So π is a probability measure on $\mathcal{G}$. But $\mathcal{G}$ is not a σ-field, because $\mathcal{G}$ is not necessarily closed under complement.

Step 2. Further extend the set function π to another set function π^* that is defined on the power set $\mathcal{P}(\Omega)$. (Here, the problem is π^* is NOT a probability measure on $\mathcal{P}(\Omega)$).

For any $A \subset \Omega$ (i.e. $A \in \mathcal{P}(\Omega)$). Define a new set function π^* on the power set $\mathcal{P}(\Omega)$ as

$$\pi^*(A) := \inf\{\pi(G) : A \subset G \in \mathcal{G}\}$$

Properties of π^*:

1) $\pi^*|_{\mathcal{G}} = \pi$

2) For any $A_1, A_2 \in \mathcal{P}(\Omega)$, we have

$$\pi^*(A_1 \cup A_2) + \pi^*(A_1 \cap A_2) \leqslant \pi^*(A_1) + \pi^*(A_2)$$

3) If $A_1 \subset A_2$. Then $\pi^*(A_1) \leqslant \pi^*(A_2)$.

4) If the sequence of sets $A_n \uparrow A$. Then we have continuity of π^*

$$\pi^*(A) = \pi^*(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \pi^*(A_n)$$

Remark 2.34. *Property (2) is actually the issue of π^* that prevent it from being a probability measure on $\mathcal{P}(\Omega)$. It is not finite additive.*

Proof of (2). For any $\epsilon > 0$, we can find $G_1, G_2 \in \mathcal{G}$ such that

$$\pi(G_1) \leqslant \pi^*(A_1) + \epsilon, \qquad \pi(G_2) \leqslant \pi^*(A_2) + \epsilon$$

[Board figure omitted: two overlapping sets $G_1 \supset A_1$ and $G_2 \supset A_2$, with $G_1 \cup G_2$ and $G_1 \cap G_2$ indicated]

By additivity of π, we have

$$\pi(G_1 \cup G_2) + \pi(G_1 \cap G_2) = \pi(G_1) + \pi(G_2)$$

Because $G_1 \cup G_2 \supset G_1, G_2$ resp. $\supset A_1, A_2$ resp., we have

$$\pi(G_1 \cup G_2) \geqslant \pi^*(A_1 \cup A_2), \qquad \pi(G_1 \cap G_2) \geqslant \pi^*(A_1 \cap A_2)$$

Thus, we have

$$\begin{aligned}\pi^*(A_1 \cup A_2) + \pi^*(A_1 \cap A_2) &\leqslant \pi(G_1 \cup G_2) + \pi(G_1 \cap G_2) = \pi(G_1) + \pi(G_2)\\ &\leqslant \pi^*(A_1) + \pi^*(A_2) + 2\epsilon\end{aligned}$$

Since $\epsilon > 0$ is arbitrary and the left-hand side does not depend on ϵ, we have

$$\pi^*(A_1 \cup A_2) + \pi^*(A_1 \cap A_2) \leqslant \pi^*(A_1) + \pi^*(A_2)$$

□

Proof of (3). See that by definition of $\pi^*(A_1), \pi^*(A_2)$,

$$\begin{aligned}\pi^*(A_1) &= \inf\{\pi(G) : A_1 \subset G \in \mathcal{G}\}\\ \pi^*(A_2) &= \inf\{\pi(G) : A_2 \subset G \in \mathcal{G}\}\end{aligned}$$

If $A_1 \subset A_2$, then $G \supset A_2$ that contains A_2 will also contains A_1, i.e. $G \supset A_2 \supset A_1$. Hence every G admissible in the infimum defining $\pi^*(A_2)$ is also admissible in the infimum defining $\pi^*(A_1)$; the infimum over a larger collection can only be smaller, so $\pi^*(A_1) \leqslant \pi^*(A_2)$. □

The obstruction that motivates the next step is the following. From property (2), if $A_1 \cap A_2 = \varnothing$, then

$$\pi^*(A_1 \cup A_2) \leqslant \pi^*(A_1) + \pi^*(A_2)$$

And if $A_1 = \Omega$ and $A_2 = \varnothing$, we have

$$1 = \mathbf{P}(\Omega) = \pi^*(\Omega) = \pi^*(A \cup A^c) \leqslant \pi^*(A) + \pi^*(A^c)$$

So we try to restrict π^* to a smaller collection of sets.

Step 3. Restrict π^* to $\mathcal{D}$ where $\sigma(\mathcal{A}) \subset \mathcal{D} \subset \mathcal{P}(\Omega)$, so that $\mathcal{D}$ is a σ-field (σ-algebra) and $\pi^*|_{\mathcal{D}}$ is a probability measure on $\mathcal{D}$. If we are successful, then $\pi^*|_{\mathcal{D}}$ will also be a probability measure on $\sigma(\mathcal{A})$.

The question is how to define this set $\mathcal{D}$. The subquestion is, how to make

$$\pi^*(A) + \pi^*(A^c) = 1$$

So we define

$$\mathcal{D} := \{A : A \subset \Omega \text{ such that } \pi^*(A) + \pi^*(A^c) = 1\}$$

Here, we have to show two things:

1) $\mathcal{D}$ is a σ-field.

2) $\pi^*|_{\mathcal{D}}$ is a probability measure on $\mathcal{D}$.

Claim 2.35. *$\mathcal{D}$ is a σ-field.*

Step 3.1.1. First, we show $\mathcal{D}$ is a field.

1) $\varnothing \in \mathcal{D}$. Because $\pi^*(\varnothing) + \pi^*(\Omega) = \mathbf{P}(\varnothing) + \mathbf{P}(\Omega) = 1$

2) If $A \in \mathcal{D}$, then by definition of $\mathcal{D}$, $A^c \in \mathcal{D}$. Because by the choice of A, $\pi^*(A) + \pi^*(A^c) = 1$. Want to show $\pi^*(A^c) + \pi^*((A^c)^c) = 1$, but $(A^c)^c = A$. Hence we are done.

3) Let $D_1, D_2 \in \mathcal{D}$. Then by property of π^*, we have $\pi^*(D_1 \cup D_2) + \pi^*(D_1 \cap D_2) \leqslant \pi^*(D_1) + \pi^*(D_2)$. By closure under complement, we have $D_1^c, D_2^c \in \mathcal{D}$. Hence $\pi^*(D_1^c \cup D_2^c) + \pi^*(D_1^c \cap D_2^c) \leqslant \pi^*(D_1^c) + \pi^*(D_2^c)$. We add LHS & RHS of these two equalities together.

$$\begin{aligned} 2 &\geqslant \pi^*(D_1) + \pi^*(D_2) + \pi^*(D_1^c) + \pi^*(D_2^c) \\ &\geqslant \pi^*(D_1 \cup D_2) + \pi^*(D_1 \cap D_2) + \pi^*(D_1^c \cup D_2^c) + \pi^*(D_1^c \cap D_2^c) \\ &= \underbrace{\pi^*(D_1 \cup D_2) + \pi^*((D_1 \cup D_2)^c)}_{\geqslant 1 \text{ by a property of } \pi^*} + \underbrace{\pi^*(D_1 \cap D_2) + \pi^*((D_1 \cap D_2)^c)}_{\geqslant 1 \text{ by a property of } \pi^*} \\ &\geqslant 2 \end{aligned}$$

Hence, by sandwich theorem, every inequality we have used should be an equality. In particular,

$$\begin{aligned} \pi^*(D_1 \cup D_2) + \pi^*((D_1 \cup D_2)^c) &= 1 \\ \pi^*(D_1 \cap D_2) + \pi^*((D_1 \cap D_2)^c) &= 1 \\ \pi^*(D_1 \cup D_2) + \pi^*(D_1 \cap D_2) &= \pi^*(D_1) + \pi^*(D_2) \end{aligned}$$

Hence, we showed $D_1 \cup D_2 \in \mathcal{D}$ and $D_1 \cap D_2 \in \mathcal{D}$. And, π^* is additive on $\mathcal{D}$.

Step 3.1.2. To finish the proof of Claim 1, we show $\mathcal{D}$ is a monotone class. Since $\mathcal{D}$ is closed under complement, we only need to show $\mathcal{D}$ is closed under increasing limit. That is, $D_n \in \mathcal{D}$ and $D_n \uparrow$. Then $\lim_{n\to\infty} D_n = \cup_{n\geqslant 1} D_n \in \mathcal{D}$.

We want

$$1 \leqslant \pi^*(\cup_{n\geqslant 1} D_n) + \pi^*(\cap_{n\geqslant 1} D_n^c) \leqslant 1$$

Here,

$$\pi^*(\cup_{n\geqslant 1} D_n) + \pi^*(\cap_{n\geqslant 1} D_n^c) \leqslant 1$$

is not obvious. First recall by continuity of π^*, we have

$$\pi^*(\lim_{n\to\infty} \uparrow D_n) = \lim_{n\to\infty} \pi^*(D_n)$$

For the other term, note that since D_n is an increasing sequence, thus D_n^c is a decreasing sequence. Hence

$$\pi^*(\cap_{n\geqslant 1} D_n^c) \leqslant \pi^*(D_n^c) \qquad \text{for any } n$$

Thus,

$$\pi^*(\cap_{n\geqslant 1} D_n^c) \leqslant \lim_{n\to\infty} \pi^*(D_n^c)$$

Hence,

$$\begin{aligned}
\pi^*(\cup_{n\geqslant 1} D_n) + \pi^*(\cap_{n\geqslant 1} D_n^c) &\leqslant \lim_{n\to\infty} \pi^*(D_n) + \lim_{n\to\infty} \pi^*(D_n^c) \\
&= \lim_{n\to\infty} \underbrace{(\pi^*(D_n) + \pi^*(D_n^c))}_{=1} \\
&= 1
\end{aligned}$$

Hence, we have

$$1 \leqslant \pi^*(\cup_{n\geqslant 1} D_n) + \pi^*(\cap_{n\geqslant 1} D_n^c) \leqslant 1$$

Thus, by sandwich theorem, we have

$$\pi^*(\cup_{n\geqslant 1} D_n) + \pi^*(\cap_{n\geqslant 1} D_n^c) = 1$$

Thus, $\lim_{n\to\infty} D_n = \cup_{n\geqslant 1} D_n \in \mathcal{D}$. Thus, we completes the proof that $\mathcal{D}$ is a monotone class.

Step 3.1.3. Now we have $\mathcal{D}$ is a field and it is a monotone class. Hence, $\mathcal{D}$ is a σ-field.

Claim 2.36. *$\pi^*|_{\mathcal{D}}$ is a probability measure on $\mathcal{D}$.*

Step 3.2.1. Show $\pi^*|_{\mathcal{D}}(\Omega) = 1$ and $\pi^*|_{\mathcal{D}}(D) \geqslant 0$ for all $D \in \mathcal{D}$. DIY

Step 3.2.2. We focus on showing π^* is σ-additive on $\mathcal{D}$, that is if we have a disjoint sequence of sets $D_n \in \mathcal{D}$, then

$$\pi^*(\cup_{n=1}^{\infty} D_n) = \sum_{n=1}^{\infty} \pi^*(D_n)$$

Consider

$$\begin{aligned}
\pi^*(\cup_{n\geqslant 1} D_n) &= \pi^*(\lim_{N\to\infty} \cup_{n=1}^N D_n) \\
&= \lim_{N\to\infty} \pi^*(\cup_{n=1}^N D_n), && \text{by Monotone Convergence Thm} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \pi^*(D_n), && \text{by finite additivity of } \pi^* \\
&= \sum_{n=1}^{\infty} \pi^*(D_n)
\end{aligned}$$

Step 3.2.3. To conclude, since $\pi^*|_{\mathcal{D}}$ satisfies all 3 axioms of probability measure on $\mathcal{D}$, it is a probability measure on $\mathcal{D}$.
Wrap up. Note that $\sigma(\mathcal{A}) \subset \mathcal{D}$ because $\mathcal{A} \subset \mathcal{D}$, and $\mathcal{D}$ is a σ-field. Thus, $\sigma(\mathcal{A}) \subset \mathcal{D}$. Thus, $\pi^*|_{\mathcal{D}}$ is a probability measure on the σ-field $\sigma(\mathcal{A})$.

Overall, we extended the probability measure $\mathbf{P}$ on $\mathcal{A}$ to a probability measure $\pi^*|_{\mathcal{D}}$ on $\sigma(\mathcal{A})$.
Proof of uniqueness: Suppose we have two extensions $\mathbf{P}_1$ and $\mathbf{P}_2$ of $\mathbf{P}$, i.e.

$$\mathbf{P}_1 |_{\mathcal{A}} = \mathbf{P}_2 |_{\mathcal{A}}$$

which means $\mathbf{P}_1$ ad $\mathbf{P}_2$ agrees on $\mathcal{A}$.

Recall the corollary of Dynkin's $\pi - \lambda$ theorem, we have proved that if $\mathcal{P}$ is a π-system and $\mathbf{P}_1 = \mathbf{P}_2$ on $\mathcal{P}$. Then $\mathbf{P}_1 = \mathbf{P}_2$ on $\sigma(\mathcal{P})$.

Apply this corollary to $\mathbf{P}_1$ and $\mathbf{P}_2$ and $\mathcal{P} = \mathcal{A}$, we conclude that $\mathbf{P}_1 = \mathbf{P}_2$ on $\sigma(\mathcal{A})$.

2.5 Measure Constructions

Lectures 2024-10-05, 2024-10-02

Two important examples

Example 2.13. *Consider the measurable space* $(\Omega = (0, 1], \mathscr{B} = \mathscr{B}((0, 1]))$. *Consider the Lebesgue measure* $\mathbf{P}((a, b]) = b - a$ *which measure the length of the interval.*

Let the set function $\mathbf{P}$ *defined on a semi-algebra*

$$\mathcal{S} = \{(a, b] : 0 \leqslant a \leqslant b \leqslant 1\}$$

by

$$\mathbf{P}((a, b]) = b - a$$

If we can show $\mathbf{P}$ *is a probability measure on* $\mathcal{S}$, *then we can use two extension theorems to obtain a probability measure in the following way:*

- *By first extension, we get a probability measure on* $\mathcal{A}(\mathcal{S})$
- *By second extension, we get a probability measure on* $\sigma(\mathcal{A}(\mathcal{S})) = \sigma(\mathcal{S}) = \mathscr{B}((0, 1])$.

Now we show $\mathbf{P}$ is a probability measure. Usually the first two axioms are trivial. The only non-trivial axiom is the σ-additivity. That is for the event

$$(a, b] = \sum_{i \geqslant 1} (a_i, b_i]$$

Want to show

$$b - a = \mathbf{P}((a, b]) = \sum_{i \geqslant 1} \mathbf{P}((a_i, b_i]) = \sum_{i \geqslant 1} (b_i - a_i)$$

The problem of doing a countable summation of positive numbers, things can really go wrong. For example, this summation may goes to infinity. You don't know if this countable sum will converge or not.

Step 1: show $b - a \leqslant \sum_{i=1}^{\infty} (b_i - a_i)$. What we will do is to fix a small $\epsilon > 0$. And we will make the interval $(a, b]$ a bit smaller by looking at the interval $[a + \epsilon, b]$. In addition, we will make the intervals $(a_i, b_i]$ slightly bigger by looking at $(a_i, b_i + \frac{1}{2^i}\epsilon)$. Then we have

$$[a + \epsilon, b] \subset (a, b] = \sum_{i \geqslant 1} (a_i, b_i] = \bigcup_{i=1}^{\infty} \left(a_i, b_i + \frac{1}{2^i}\epsilon \right)$$

By compactness of the interval $[a+\epsilon, b]$, every open cover has a finite subcover. Hence, there exists N such that

$$[a + \epsilon, b] \subset \bigcup_{i=1}^{N} \left(a_i, b_i + \frac{1}{2^i}\epsilon \right)$$

So by monotonicity of probability measure

$$\begin{aligned} b - (a + \epsilon) = \mathbf{P}([a + \epsilon, b]) \leqslant \mathbf{P}\left(\bigcup_{i=1}^{N} \left(a_i, b_i + \frac{1}{2^i}\epsilon \right) \right) &= \sum_{i=1}^{N} (b_i + \frac{1}{2^i}\epsilon - a_i) \\ &= \sum_{i=1}^{N} (b_i - a_i) + \epsilon \sum_{i=1}^{N} \frac{1}{2^i} \\ &\leqslant \sum_{i=1}^{\infty} (b_i - a_i) + \epsilon \sum_{i=1}^{\infty} \frac{1}{2^i} \\ &= \sum_{i=1}^{\infty} (b_i - a_i) + \epsilon \end{aligned}$$

Thus,

$$b - a \leqslant \sum_{i=1}^{\infty} (b_i - a_i) + 2\epsilon$$

Since ϵ is arbitrary, we send ϵ to 0, we get

$$b - a \leqslant \sum_{i=1}^{\infty} (b_i - a_i)$$

Step 2: show $b - a \geqslant \sum_{i=1}^{\infty}(b_i - a_i)$. Next we show

$$b - a \geqslant \sum_{i=1}^{\infty}(b_i - a_i)$$

Note

$$(a, b] = \sum_{i=1}^{\infty}(a_i, b_i] \supset \sum_{i=1}^{N}(a_i, b_i], \quad \text{for any } N \in \mathbb{Z}_{\geqslant 1}$$

Then by the monotonicity of probability measure, we have

$$b - a = \mathbf{P}((a, b]) \geqslant \mathbf{P}(\sum_{i=1}^{N}(a_i, b_i]) = \sum_{i=1}^{N}(b_i - a_i)$$

Send N to infinity, we have

$$b - a \geqslant \sum_{i=1}^{\infty}(b_i - a_i)$$

Example 2.14. *Consider the set of rational numbers $\mathbb{Q}$ on $(0, 1]$, i.e. $\mathbb{Q} \cap (0, 1]$. The Lebesgue measure on $\mathbb{Q} \cap (0, 1]$ will be zero, i.e.*

$$\mathbf{P}(\mathbb{Q} \cap (0, 1]) = 0$$

Because $\mathbb{Q} \cap (0, 1]$ is a countable collection of numbers, i.e. we can enumerate its elements as $\mathbb{Q} \cap (0, 1] = \{a_1, a_2, \ldots\}$. Fix $\epsilon > 0$ and look at

$$I_i := \left(a_i - \frac{\epsilon}{2^{i+1}}, a_i + \frac{\epsilon}{2^{i+1}}\right] \cap (0, 1]$$

so that $a_i \in I_i$, and I_i is an interval of length at most $\frac{\epsilon}{2^i}$. Then see that

$$\mathbb{Q} \cap (0, 1] \subset \bigcup_{i=1}^{\infty} I_i$$

by subadditivity of probability measure, we have

$$\begin{aligned}
\mathbf{P}(\mathbb{Q} \cap (0, 1]) &\leqslant \mathbf{P}(\cup_{i=1}^{\infty} I_i) \\
&\leqslant \sum_{i=1}^{\infty} \mathbf{P}(I_i) \\
&\leqslant \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} \\
&= \epsilon
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we conclude $\mathbf{P}(\mathbb{Q} \cap (0, 1]) = 0$.

Example 2.15. *Consider the measurable space $(\Omega, \mathscr{B}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Start from a DF $F(x)$. How to construct a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ so that*

$$F(x) = \mathbf{P}((-\infty, x])$$

So for semi-algebra

$$\mathcal{S} = \{(a, b] : -\infty \leqslant a \leqslant b \leqslant \infty\}$$

Set

$$\mathbf{P}((a, b]) = F(b) - F(a)$$

If we can show $\mathbf{P}$ is a probability measure on $\mathcal{S}$, then by two extension theorems, we get a probability measure on $\sigma(\mathcal{S}) = \mathscr{B}(\mathbb{R})$. The σ-additivity of $\mathbf{P}$ on $\mathcal{S}$ is verified right below this example.

Verification of σ-additivity for this example *(From the midterm review, Lecture 2024-10-02.)*

(3.1.2) Show $\mathbf{P}$ satisfies monotone class theorem. Let $\{A_n\} \subset \mathcal{S}$ be such that $A_n \downarrow \varnothing$ as $n \to \infty$. (For example, $A_n = (a, a + \frac{1}{n}]$).

Step (3.1.2.1). Let's first assume that there exists $n < \infty$ such that A_n is bounded. WLOG, we can assume A_1 is bounded (otherwise, we discard all A_i before n and can only consider the sequence starting from n). A_1 is bounded means there exists $N \in \mathbb{Z}_{\geqslant 1}$ such that for all $n \geqslant 1$,

$$A_n \subset [-N, N]$$

The goal is to show $\mathbf{P}(A_n) \downarrow 0$ as $n \to \infty$. Fix an $\epsilon > 0$.

Observe that for all $n \geqslant 1$, there exists some $B_n \in \mathcal{S}$ such that $\overline{B_n} \subset A_n$ (here $\overline{B_n}$ means closure of B_n) and

$$\mathbf{P}(A_n \backslash B_n) = \mathbf{P}(A_n) - \mathbf{P}(B_n) \leqslant \epsilon \frac{1}{2^n}$$

To see such B_n exists, suppose $A_n = (a_n, b_n]$ and $B_n = (a_{n,k}, b_n]$ where $a_{n,k} \downarrow a_n$ as $k \to \infty$. Then

$$\mathbf{P}((a_{n,k}, b_n]) = F(b_n) - F(a_{n,k}) \longrightarrow F(b_n) - F(a_n) = \mathbf{P}((a_n, b_n]), \quad \text{as } k \to \infty$$

Because $\overline{B_n} \subset A_n$ for all $n \geqslant 1$, hence

$$\bigcap_{n=1}^{\infty} \overline{B_n} \subset \bigcap_{n=1}^{\infty} A_n = \lim_{n \to \infty} A_n = \varnothing$$

Take the complement, we get

$$\bigcup_{n=1}^{\infty} \overline{B_n}^c = \left(\bigcap_{n=1}^{\infty} \overline{B_n} \right)^c = \mathbb{R}$$

Here for each $n \geqslant 1$, $\overline{B_n}^c$ are complement of closed sets, hence are open sets. Thus, $\{\overline{B_n}^c\}_{n \geqslant 1}$ is an open cover of $\mathbb{R}$. In particular, $\{\overline{B_n}^c\}_{n \geqslant 1}$ is an open cover of the compact set $[-N, N]$, i.e.

$$\bigcup_{n=1}^{\infty} \overline{B_n}^c \supset [-N, N]$$

(Remark: the compactness comes from the fact that $[-N, N]$ is closed and bounded, and Hausdorff. Thus, by Heine–Borel theorem, this is equivalent to (quasi-)compactness). By (quasi-)compactness, every open cover has a finite sub-cover. Hence, there exists some n_0 such that

$$\bigcup_{n=1}^{n_0} \overline{B_n}^c \supset [-N, N]$$

On the other hand, $\bigcap_{n=1}^{n_0} \overline{B_n} \subset \bigcap_{n=1}^{n_0} A_n = A_{n_0} \subset [-N, N]$; taking complements,

$$\bigcup_{n=1}^{n_0} \overline{B_n}^c \supset [-N, N]^c$$

Combining this with the finite subcover $\bigcup_{n=1}^{n_0} \overline{B_n}^c \supset [-N, N]$ obtained above, we get $\bigcup_{n=1}^{n_0} \overline{B_n}^c = \mathbb{R}$. Equivalently, $\cap_{n=1}^{n_0} \overline{B_n} = \varnothing$. Hence, we get

$$\tilde{B}_{n_0} := \bigcap_{n=1}^{n_0} B_n = \bigcap_{n=1}^{n_0} \overline{B_n} = \varnothing$$

Now, consider $\mathbf{P}(A_{n_0})$.

$$\begin{aligned}
\mathbf{P}(A_{n_0}) &= \mathbf{P}(A_{n_0} \cap \tilde{B}_{n_0}^c) + \underbrace{\mathbf{P}(A_{n_0} \cap \tilde{B}_{n_0})}_{=\mathbf{P}(\varnothing)=0} \\
&= \mathbf{P}(A_{n_0} \cap \tilde{B}_{n_0}^c) = \mathbf{P}\left(A_{n_0} \cap \left(\bigcup_{n=1}^{n_0} B_n^c\right)\right) \\
&= \mathbf{P}\left(\bigcup_{n=1}^{n_0} (A_{n_0} \cap B_n^c)\right) \\
&\leqslant \mathbf{P}\left(\bigcup_{n=1}^{n_0} (A_n \cap B_n^c)\right), \qquad \text{because } A_{n_0} \subset A_n \\
&\leqslant \sum_{n=1}^{n_0} \mathbf{P}\underbrace{\left(A_n \cap B_n^c\right)}_{=\mathbf{P}(A_n \setminus B_n) \leqslant \frac{\epsilon}{2^n}}, \qquad \text{sub-additivity} \\
&= \sum_{n=1}^{n_0} \epsilon \frac{1}{2^n} = \epsilon \sum_{n=1}^{n_0} \frac{1}{2^n} \\
&\leqslant \epsilon \sum_{n=1}^{\infty} \frac{1}{2^n} = \epsilon
\end{aligned}$$

Lastly, since $\epsilon > 0$ is arbitrary and we know for all $k \geqslant n_0$, $\mathbf{P}(A_k) \leqslant \mathbf{P}(A_{n_0})$. We conclude

$$\mathbf{P}(A_n) \longrightarrow 0 \qquad \text{as } n \to \infty$$

Step (3.1.2.2). If we lift the assumption of A_1 being bounded, we can still truncate A_n by choosing a $N > 0$ such that $\mathbf{P}((-N, N]^c) < \epsilon$, and define $C_n = A_n \cap (-N, N]$.

Then

$$\mathbf{P}(A_n) = \underbrace{\mathbf{P}(A_n \cap (-N, N])}_{\leqslant \mathbf{P}(A_n \cap [-N,N])} + \underbrace{\mathbf{P}(A_n \cap (-N, N]^c)}_{\leqslant \mathbf{P}((-N,N]^c)} \leqslant \mathbf{P}(C_n) + \epsilon$$

Apply Step (3.1.2.1) to $\mathbf{P}(C_n)$, we see that there exists some $n_0 \in \mathbb{Z}_{\geqslant 1}$ such that $\mathbf{P}(C_{n_0}) < \epsilon$. Hence, we get

$$\mathbf{P}(A_n) < 2\epsilon$$

Again, since $\epsilon > 0$ is arbitrary, we obtain the claim.

Remark 2.37. *The measure on $\mathscr{B}((0,1])$ we get from two extension theorems is called Lebesgue measure on $(0,1]$.*

Remark 2.38. *Most of the time, we are dealing with Lebesgue measure on $\mathbb{R}$ instead of just an interval.*

The way to construct Lebesgue measure on $\mathbb{R}$ from Lebesgue measure on interval is by taking the translate $(0,1]$ to $(1,2]$ and so on.

[Board figure omitted: the real line cut into consecutive unit intervals $(n, n+1]$, $n \in \mathbb{Z}$, obtained by translating $(0,1]$]

So we get Lebesgue measure on all the unit intervals. $(n, n+1]$ for $n \in \mathbb{Z}$. On the point $n \in \mathbb{Z}$, we can measure the length on two intervals separately and add them up.

3 Random Variables, Elements, and Measurable Maps

3.1 Inverse Maps

Lecture 2024-10-08

For this map $X : \Omega \to \Omega'$, we can define for every $A' \subset \Omega'$,

$$X^{-1}(A') := \{\omega \in \Omega : X(\omega) \in A'\}$$

This is called the preimage of X.

From this definition of preimage, we can define a map

$$\begin{array}{rccc} X^{-1}: & \mathcal{P}(\Omega') & \longrightarrow & \mathcal{P}(\Omega) \\ & A' & \longmapsto & X^{-1}(A') \end{array}$$

Several important properties of this map X^{-1} are, it commutes with all set operations:

1. $X^{-1}(\varnothing) = \varnothing$.
2. $X^{-1}(\Omega') = \Omega$
3. $X^{-1}(\Omega' \backslash A') = \Omega \backslash X^{-1}(A')$
4. $X^{-1}(\cup_{t \in T} A'_t) = \cup_{t \in T} X^{-1}(A'_t)$
5. $X^{-1}(\cap_{t \in T} A'_t) = \cap_{t \in T} X^{-1}(A'_t)$

Some notation

For a $\mathcal{C}' \subset \mathcal{P}(\Omega')$. Its pre-image is

$$X^{-1}(\mathcal{C}') = \{X^{-1}(C') : C' \in \mathcal{C}'\}$$

Theorem 3.1. *If $\mathscr{B}'$ is a σ-field of Ω', then $X^{-1}(\mathscr{B}')$ is also a σ-field.*

Proof. Because $X^{-1}(\varnothing) = \varnothing$. Hence $\varnothing \in X^{-1}(\mathscr{B}')$.

If $B \in X^{-1}(\mathscr{B}')$, then $\Omega \backslash B \in X^{-1}(\mathscr{B}')$. Because $B = X^{-1}(B')$ for $B' \in \mathscr{B}'$, then $\Omega \backslash B = \Omega \backslash X^{-1}(B') = X^{-1}(\Omega' \backslash B')$.

If $B_1, B_2, \ldots \in X^{-1}(\mathscr{B}')$ then $\cup_{n=1}^{\infty} B_n \in X^{-1}(\mathscr{B}')$. Because $B_n = X^{-1}(B'_n)$ where $B'_n \in \mathscr{B}'$ And $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} X^{-1}(B'_n) = X^{-1}(\cup_{n=1}^{\infty} B'_n)$ and $\cup_{n=1}^{\infty} B'_n \in \mathscr{B}'$. □

Theorem 3.2. *Let $\mathcal{C}' \subset \mathcal{P}(\Omega')$. Then*

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}'))$$

Proof. By the previous theorem, we have $X^{-1}(\sigma(\mathcal{C}'))$ is a σ-field. Hence, we conclude

$$X^{-1}(\sigma(\mathcal{C}')) \supseteq X^{-1}(\mathcal{C}')$$

Hence, by the definition of $\sigma(X^{-1}(\mathcal{C}'))$ being the smallest σ-field contains $X^{-1}(\mathcal{C}')$, we conclude

$$X^{-1}(\sigma(\mathcal{C}')) \supseteq \sigma(X^{-1}(\mathcal{C}'))$$

For the other direction, we consider all sets in $A' \in \sigma(\mathcal{C}')$ such that $X^{-1}(A') \in \sigma(X^{-1}(\mathcal{C}'))$.

$$\mathcal{G} := \{A' \in \sigma(\mathcal{C}') : X^{-1}(A') \in \sigma(X^{-1}(\mathcal{C}'))\}$$

Then it suffices to show

1. $\mathcal{G}$ is a σ-field
2. $\mathcal{G} \supset \mathcal{C}'$.

For the 2nd one, it is trivial. Because if $C' \in \mathcal{C}'$, we have $X^{-1}(C') \in X^{-1}(\mathcal{C}') \subset \sigma(X^{-1}(\mathcal{C}'))$.

We focus on proving $\mathcal{G}$ is a σ-field.

1. $\varnothing \in \mathcal{G}$ because $X^{-1}(\varnothing) = \varnothing \in \sigma(X^{-1}(\mathcal{C}'))$.
2. If $A' \in \mathcal{G}$, then $\Omega' \backslash A' \in \mathcal{G}$, because
$$X^{-1}(\Omega' \backslash A') = \Omega \backslash X^{-1}(A')$$
and $X^{-1}(A') \in \sigma(X^{-1}(\mathcal{C}'))$. Hence, $\Omega \backslash X^{-1}(A') \in \sigma(X^{-1}(\mathcal{C}'))$.
3. If $A_n' \in \mathcal{G}$, then $\cup_{n \geqslant 1} A_n' \in \mathcal{G}$ Because
$$X^{-1}(\cup_{n \geqslant 1} A_n') = \cup_{n \geqslant 1} X^{-1}(A_n')$$
and $X^{-1}(A_n') \in \sigma(X^{-1}(\mathcal{C}'))$ for all $n \geqslant 1$. Hence $\cup_{n \geqslant 1} X^{-1}(A_n') \in \sigma(X^{-1}(\mathcal{C}'))$.

□

Remark 3.3. *This theorem says X^{-1} and the operation of taking σ commutes.*

3.2 Measurable Maps, Random Elements, Induced Probability Measures

Lectures 2024-10-08, 2024-10-10, 2024-10-12, 2024-10-15

Definition 3.4 (Measurable map)**.** *Let $(\Omega, \mathscr{B})$ and $(\Omega', \mathscr{B}')$ are two measurable spaces. Let $X : \Omega \to \Omega'$ is a map. If $X^{-1}(\mathscr{B}') \subseteq \mathscr{B}$, then we say X is a measurable map with respect to $\mathscr{B}$ and $\mathscr{B}'$, write $X \in \mathscr{B}/\mathscr{B}'$.*

Sometimes, we just say X is a measurable map.

Definition 3.5 (Random Variable)**.** *In particular, let $\Omega' = \mathbb{R}, \mathscr{B}' = \mathscr{B}(\mathbb{R})$, and $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$, then we say X is a random variable.*

Remark 3.6. *Why measurability $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$ is required to define a random variable X where*

$$X : \quad (\Omega, \mathscr{B}, \mathbf{P}) \quad \longrightarrow \quad (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

We are interested in compute $\mathbf{P}(X \in A)$ for some $A \in \mathscr{B}(\mathbb{R})$. But what does $\mathbf{P}(X \in A)$ mean? It means

$$\mathbf{P}(X \in A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

Induced measure by a measurable map

Suppose we have two measurable spaces $(\Omega, \mathscr{B})$ and $(\Omega^{'}, \mathscr{B}^{'})$. We also have a measurable map

$$X : \Omega \longrightarrow \Omega^{'}$$

which we will denote this map as $X \in \mathscr{B}/\mathscr{B}^{'}$. Here X induces a probability measure on $(\Omega^{'}, \mathscr{B}^{'})$ as follows: for any $B^{'} \in \mathscr{B}^{'}$ define a set function $\mathbf{P}^{'}$ as

$$\mathbf{P}^{'}(B^{'}) = \mathbf{P}(X^{-1}(B^{'}))$$

This gives us a probability measure on $(\Omega^{'}, \mathscr{B}^{'})$.

The only hard axiom of probability measure to check is σ-additivity, that is for disjoint sets $B_1^{'}, \ldots, B_n^{'}, \ldots \in \mathscr{B}^{'}$ such that $\cup_{n\geqslant 1} B_n^{'} \in \mathscr{B}^{'}$ we want

$$\mathbf{P}^{'}(\cup_{n\geqslant 1} B_n^{'}) = \sum_{n\geqslant 1} \mathbf{P}^{'}(B_n^{'})$$

Let's do it

$$\begin{aligned}
\mathbf{P}^{'}\left(\bigcup_{n\geqslant 1} B_n^{'}\right) &= \mathbf{P}\left(X^{-1}(\bigcup_{n\geqslant 1} B_n^{'})\right) \\
&= \mathbf{P}\left(\bigcup_{n\geqslant 1} X^{-1}(B_n^{'})\right) \\
&= \sum_{n\geqslant 1} \mathbf{P}(X^{-1}(B_n^{'})) \\
&= \sum_{n\geqslant 1} \mathbf{P}^{'}(B_n^{'})
\end{aligned}$$

Definition 3.7 (Random Variable)**.** *In particular, if we consider a RV* $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$ *which is the measurable map*

$$X : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbf{P}^{'} = \mathbf{P} \circ X^{-1})$$

For any Borel set $B^{'} \in \mathscr{B}(\mathbb{R})$*, evaluate* $\mathbf{P}^{'}$ *at* $B^{'}$

$$\mathbf{P}^{'}(B^{'}) = \mathbf{P}(X^{-1}(B^{'})) = \mathbf{P}(\{\omega \in \Omega : X(\omega) \in B^{'}\})$$

The event $\{\omega \in \Omega : X(\omega) \in B^{'}\}$ *means the random variable* X *will take values in the set* $B^{'}$*.*

Example 3.1 (Distribution Function)**.** *For example, if* $B^{'} = (-\infty, x]$*, then*

$$\mathbf{P}^{'}((-\infty, x]) = \mathbf{P}(\{\omega \in \Omega : X(\omega) \in (-\infty, x]\}) = \mathbf{P}(X \leqslant x)$$

Hence, $\mathbf{P}^{'}$ *gives the distribution function of the RV* X*.*

Example 3.2 (Toss a dice twice)**.** *The sample space will be*

$$\Omega := \{(i, j) : 1 \leqslant i, j \leqslant 6\}$$

Then the probability of $(1, 2) \in \mathscr{B} = \mathcal{P}(\Omega)$ *will be* $\mathbf{P}((1, 2)) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ *where the* σ*-algebra is defined as* $\mathscr{B} = \mathcal{P}(\Omega)$*.*

Suppose X *means the sum of the two outcomes.*

[Board figure omitted: the map X sending the 6×6 grid of outcomes (i, j) to the set of possible sums $\{2, 3, \ldots, 12\}$]

Then the $\Omega' = \{2, 3, \ldots, 12\}$, *and* $\mathbf{P}' = \mathbf{P} \circ X^{-1}$. *For example,*

$$\mathbf{P}'(\{2\}) = \mathbf{P}(X = 2) = \mathbf{P}(X^{-1}(\{2\})) = \frac{1}{36}$$

Another example, we have

$$\mathbf{P}(X = 4) = \mathbf{P}'(\{4\}) = \mathbf{P}(X^{-1}(\{4\})) = \mathbf{P}(\{(1,3), (2,2), (3,1)\}) = \frac{3}{36}$$

Another example,

$$\mathbf{P}(X \in [0, 1]) = \mathbf{P}'([0, 1]) = \mathbf{P}(X^{-1}([0, 1])) = \mathbf{P}(\varnothing) = 0$$

Another example,

$$\mathbf{P}(X \in [1, 2.5]) = \mathbf{P}'([1, 2.5]) = \mathbf{P}(X^{-1}([1, 2.5])) = \mathbf{P}(\{(1, 1)\}) = \frac{1}{36}$$

Example 3.3 (Construction of Gaussian Random Variable)**.** *Given a DF* $F(x)$ *with*

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

How do we know we have a RV X *with DF* $F(x)$*? We mean the induced measure* $\mathbf{P}' = \mathbf{P} \circ X^{-1}$ *evaluate at the set* $(-\infty, x]$ *where* $\mathbf{P}'((-\infty, x]) = F(x)$ *with* $F(x)$ *will be of the form of above.*

We need to construct a probability space $(\Omega, \mathscr{B}, \mathbf{P})$ *and a random variable* X *such that*

$$\mathbf{P}'((-\infty, x]) = F(x)$$

To start, we just choose

$$(\Omega, \mathscr{B}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

By the one-to-one correspondence between $\mathbf{P}$ *and* $F(x)$*, for any given DF* $F(x)$*, we can construct a probability measure* $\mathbf{P}$ *on* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *such that*

$$\mathbf{P}((-\infty, x]) = F(x)$$

But we still need to construct a map

$$X : (\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbf{P}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

We choose $X = \mathrm{id}$ *to be the identity map, so that we define* $\mathbf{P}' = \mathbf{P}$. *Under this choice of* $\mathbf{P}'$*, we have*

$$\mathbf{P}((-\infty, x]) = F(x)$$

Test measurability

Proposition 3.8. *Suppose we have a map*

$$X : (\Omega, \mathscr{B}) \longrightarrow (\Omega', \mathscr{B}')$$

Suppose $\mathscr{B}' = \sigma(\mathcal{C}')$ *for some generating set* $\mathcal{C}' \subset \mathscr{B}'$*, and* $X^{-1}(\mathcal{C}') \subset \mathscr{B}$*. Then* $X \in \mathscr{B}/\mathscr{B}'$*.*

Proof. Consider the inverse map X^{-1} on $\mathscr{B}'$

$$X^{-1}(\mathscr{B}') = X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}'))$$

where $X^{-1}(\mathcal{C}') \subset \mathscr{B}$, and use $\sigma(X^{-1}(\mathcal{C}'))$ being the smallest σ-field that contains $X^{-1}(\mathcal{C}')$. So we will have

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}')) \subset \mathscr{B}$$

which proves measurability of X. □

Corollary 3.9. *If we have a map*

$$X : (\Omega, \mathscr{B}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

If for any $x \in \mathbb{R}$*, we have* $X^{-1}((-\infty, x]) \in \mathscr{B}$*, then* X *is a random variable.*

Proof. Choose $\mathcal{C}' = \{(-\infty, x] : x \in \mathbb{R}\}$, then apply the previous proposition, we have $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$. □

Proposition 3.10 (Composition of measurable maps)**.** *Consider the following composition of measurable maps*

$$(\Omega, \mathscr{B}) \xrightarrow{X} (\Omega', \mathscr{B}') \xrightarrow{Y} (\Omega'', \mathscr{B}'')$$

where $X \in \mathscr{B}/\mathscr{B}'$ *and* $Y \in \mathscr{B}'/\mathscr{B}''$*. The conclusion is* $Y \circ X \in \mathscr{B}/\mathscr{B}''$*.*

Proof. Observe that

$$(Y \circ X)^{-1}(\mathscr{B}'') = X^{-1} \circ Y^{-1}(\mathscr{B}'') = X^{-1}(\underbrace{Y^{-1}(\mathscr{B}'')}_{\subset \mathscr{B}'}) \subset \mathscr{B}$$

Hence, we have the result. □

Measurability and continuity

Example 3.4 (Euclidean Space)**.** *Consider* $\mathbb{R}^n$*. Let* $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$*. Their (Euclidean) distance is defined as*

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$$

Definition 3.11 (Metric space)**.** *In general, for a metric space* S*, the distance function* $d(x, y)$ *is a function satisfies the following axioms*

1. *(Positivity)* $d(x, y) \geqslant 0$ *for all* $x, y \in S$*, and* $d(x, y) = 0$ *if and only if* $x = y$.

2. *(Symmetric)* $d(x, y) = d(y, x)$

3. *(Triangle inequality)* $d(x, z) \leqslant d(x, y) + d(y, z)$

Definition 3.12 (Continuous function, convergence)**.** *On a metric space S, a function is said to be continuous if $x_n \to x$ (pointwise convergence), which is equivalent to $d(x_n, x) \to 0$, then*

$$f(x_n) \to f(x), \qquad \text{for all } x \in S$$

Definition 3.13 (Continuous function)**.** *Let (S, d) be a metric space. Another equivalent definition of continuity of a function $f : S \to \mathbb{R}^n$ is that preimage $f^{-1}(O)$ is an open subset of S for any open subset $O \subset \mathbb{R}^n$, i.e.*

$$f^{-1}(O) \subset S \text{ is open} \qquad \forall O \subset \mathbb{R}^n \text{ open}$$

Definition 3.14 (Borel measurable function)**.** *Recall for any topological space, the Borel σ-field is the σ-field generated by the collection of open sets. Now we look at a continuous function*

$$f : S \longrightarrow \mathbb{R}^n$$

We have

1. *Preimage of open set is open, i.e. $f^{-1}(O) \subset S$ is open for any open set $O \subset \mathbb{R}^n$*

2. *The collection of open sets of S and $\mathbb{R}^n$ generate the corresponding Borel σ-fields, i.e. $\mathscr{B}(S) = \sigma(\text{open sets of } S)$ and $\mathscr{B}(\mathbb{R}^n) = \sigma(\text{open sets of } \mathbb{R}^n)$. Hence, f is measurable with respect to $f \in \mathscr{B}(S)/\mathscr{B}(\mathbb{R}^n)$.*

Remark 3.15. *Again the conclusion is, a continuous function is always a measurable map with respect to the corresponding Borel σ-field.*

Corollary 3.16. *Suppose X is a random vector, that is $X \in \mathscr{B}/\mathscr{B}(\mathbb{R}^n)$, i.e.*

$$X : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$$

We can denote random vector X as $X = (X_1, \ldots, X_n)$ where each X_i is a vector component of X, which is a random variable.

Proposition 3.17. *Suppose g is a continuous function of n variables, i.e.*

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad \text{continuous}$$

Then $g(X)$ is a random variable.

$$\begin{tikzcd}
(\Omega, \mathscr{B}, \mathbf{P}) \arrow[r, "X"] \arrow[dr, dashed, "g(X)"'] & (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n)) \arrow[d, "g"] \\
& (\mathbb{R}, \mathscr{B}(\mathbb{R}))
\end{tikzcd}$$

Example 3.5. *Consider the function*

$$\begin{array}{rccc} g: & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & (x_1, \dots, x_n) & \longmapsto & g(x_1, \dots, x_n) = x_k \end{array}$$

This is the projection map, hence continuous. Then

$$g(X) = g(X_1, X_2, \dots, X_n) = X_k$$

Hence X_k is a random variable.

Proposition 3.18. *Suppose $X_1, X_2, \dots, X_n$ are n random variables. Then*

$$X = (X_1, \dots, X_n) : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$$

is a random vector, that is

$$X \in \mathscr{B}/\mathscr{B}(\mathbb{R}^n)$$

Proof. Consider collection of sets

$$\mathcal{C} := \{(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n) : a_i, b_i \in \mathbb{R}; a_i \leqslant b_i\}$$

What special about this collection of sets is it has $\sigma(\mathcal{C}) = \mathscr{B}(\mathbb{R}^n)$.

Then for any element I from $\mathcal{C}$, its preimage under X will be

$$\begin{aligned} X^{-1}(I) &= \{\omega \in \Omega : X(\omega) \in I\} \\ &= \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in I\} \\ &= \{\omega \in \Omega : X_i(\omega) \in (a_i, b_i); i = 1, \dots, n\} \\ &= \bigcap_{k=1}^{n} \{\omega \in \Omega : X_k(\omega) \in (a_k, b_k)\} \\ &= \bigcap_{k=1}^{n} \underbrace{X_k^{-1}((a_k, b_k))}_{\in \mathscr{B}} \\ &\in \mathscr{B} \end{aligned}$$

So $X^{-1}(\mathcal{C}) \subset \mathscr{B}$. Hence,

$$X^{-1}(\mathscr{B}(\mathbb{R}^n)) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})) \subset \mathscr{B}$$

□

Example 3.6. *If $X_1, \dots, X_n$ are n random variables. Then*

$$X = (X_1, \dots, X_n)$$

is a random vector. Then for a continuous function g, the composition of g and X

$$g(X) = g(X_1, \dots, X_n)$$

is a random variable. More explicitly,

$$(g \circ X)(\omega) = g(X(\omega)) = g(X_1(\omega), \dots, X_n(\omega))$$

is a random variable.

To summarize, this means for g being a continuous function and f being a measurable function. The composition function of $g \circ f$ is a measurable function.

Measurability and limit

Proposition 3.19. *Suppose we have a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$. We will have the following results.*

1. *Suppose* $(\sup_{n\geqslant 1} X_n)(\omega)$ *is finite for all* $\omega \in \Omega$. *Then the* $\sup_{n\geqslant 1} X_n$ *will be a random variable.*
2. *Suppose* $(\inf_{n\geqslant 1} X_n)(\omega)$ *is finite for all* $\omega \in \Omega$. *Then the* $\inf_{n\geqslant 1} X_n$ *will be a random variable.*
3. *Suppose* $(\limsup_{n\to\infty} X_n)(\omega)$ *is finite for all* $\omega \in \Omega$. *Then the* $\limsup_{n\to\infty} X_n$ *will be a random variable.*
4. *Suppose* $(\liminf_{n\to\infty} X_n)(\omega)$ *is finite for all* $\omega \in \Omega$. *Then the* $\liminf_{n\to\infty} X_n$ *will be a random variable.*
5. *Suppose* $(\lim_{n\to\infty} X_n)(\omega)$ *exists, for all* $\omega \in \Omega$. *Then* $\lim_{n\to\infty} X_n$ *is a random variable.*
6. *The set*
$$\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\}$$
is a measurable set, i.e. it lies in $\mathscr{B}$.

Proof. **(1) Consider $\sup_n X_n$:** Recall that each $X_k : (\Omega, \mathscr{B}, \mathbf{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is a measurable function into Borel measurable space, and Borel measurable space is generated by the sets of the form $(-\infty, x]$. Hence, we consider the pre-image

$$\begin{aligned}
\left(\sup_n X_n\right)^{-1}((-\infty, x]) &= \left\{\omega \in \Omega : \sup_n X_n(\omega) \leqslant x\right\} \\
&= \{\omega \in \Omega : X_n(\omega) \leqslant x \text{ for all } n \in \mathbb{Z}_{\geqslant 1}\} \\
&= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : X_n(\omega) \leqslant x\} \\
&= \bigcap_{n=1}^{\infty} \underbrace{X_n^{-1}((-\infty, x])}_{\in \mathscr{B}} \\
&\in \mathscr{B}
\end{aligned}$$

(2): This time consider

$$\begin{aligned}
\left(\inf_n X_n\right)^{-1}([x, +\infty)) &= \left\{\omega \in \Omega : \inf_n X_n(\omega) \geqslant x\right\} \\
&= \{\omega \in \Omega : X_n(\omega) \geqslant x \text{ for all } n \in \mathbb{Z}_{\geqslant 1}\} \\
&= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : X_n(\omega) \geqslant x\} \\
&= \bigcap_{n=1}^{\infty} \underbrace{X_n^{-1}([x, +\infty))}_{\in \mathscr{B}} \\
&\in \mathscr{B}
\end{aligned}$$

Proof of (3), (4): Use the definition of limsup, we have

$$\limsup_{n\to\infty} X_n = \inf_{n\geqslant 1} \sup_{m\geqslant n} X_m$$

We have proved $\sup_{m\geqslant n} X_m$ is a random variable. Let $Y_n = \sup_{m\geqslant n} X_m$. Also, we have proved $\inf_{n\geqslant 1} Y_n$ is a random variable. Hence, $\limsup_{n\to\infty} X_n$ is a random variable.

By symmetry,

$$\liminf_{n\to\infty} X_n = \sup_{n\geqslant 1} \inf_{m\geqslant n} X_m$$

is a random variable. Here we define $Z_n = \inf_{m\geqslant n} X_m$. Then use the above argument.

Proof of (5): Consider the set

$$\begin{aligned}
\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\right\} &= \left\{\omega \in \Omega : \limsup_{n\to\infty} X_n(\omega) = \liminf_{n\to\infty} X_n(\omega)\right\} \\
&= \left\{\omega \in \Omega : \underbrace{\limsup_{n\to\infty} X_n(\omega)}_{=Y} - \underbrace{\liminf_{n\to\infty} X_n(\omega)}_{=Z} = 0\right\} \\
&= \{\omega \in \Omega : X(\omega) = Y(\omega) - Z(\omega) = 0\} \\
&= \{\omega \in \Omega : X(\omega) = 0\} \\
&= X^{-1}(\{0\}) \\
&\in \mathscr{B}
\end{aligned}$$

Because Y, Z are both random variables, their difference $X = Y - Z$ is also a random variable. More precisely, it is a random variable $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$. □

Remark 3.20. *Consider the extended real line, we have*

$$\mathscr{B}(\overline{\mathbb{R}}) \supset \mathscr{B}(\mathbb{R})$$

Later, when we look at random variables, we actually look at the map

$$X : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$$

which means $X \in \mathscr{B}/\mathscr{B}(\overline{\mathbb{R}})$. That means, we do allow random variables to take $+\infty$ and $-\infty$ as their values.

If $X \in \mathscr{B}/\mathscr{B}(\overline{\mathbb{R}})$ and $X \neq \pm\infty$ (so that X takes values in $\mathbb{R}$), then $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$.

In probability, we can not avoid a random variable taking $\pm\infty$ as its value.

Example 3.7. *Let $X_n = n$. Then $\lim_{n\to\infty} X_n = +\infty$.*

3.3 σ-algebras Generated by Maps

Lectures 2024-10-15, 2024-10-17

Setup 1

Consider a measurable map

$$X : (\Omega, \mathscr{B}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

which we mean $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$.

Definition 3.21 (σ-field generated by X)**.** *The σ-field generated by X is defined as*

$$\sigma(X) := X^{-1}(\mathscr{B}(\mathbb{R}))$$

Here, $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$ means $X^{-1}(\mathscr{B}(\mathbb{R})) \subset \mathscr{B}$.

Remark 3.22. *If X, Y are both random variables, then*

$$\sigma(X) = X^{-1}(\mathscr{B}(\mathbb{R})) \subset \mathscr{B}$$
$$\sigma(Y) = Y^{-1}(\mathscr{B}(\mathbb{R})) \subset \mathscr{B}$$

This means if the domain of a random variable is the measurable space $(\Omega, \mathscr{B})$, then all information of this random variable comes from this σ-algebra $\mathscr{B}$. Thus, $\mathscr{B}$ is an "ambient event set" of preimage of $\mathscr{B}(\mathbb{R})$ under the random variable.

Hence, the σ-algebra $\sigma(X)$ contains all the information relates to X.

Example 3.8. *Consider the set*

$$\{\omega \in \Omega : X(\omega) \in (1, 2)\} = X^{-1}((1, 2)) \in \sigma(X)$$

In general, for any $B \in \mathscr{B}(\mathbb{R})$, we have

$$\{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B) \in \sigma(X)$$

Remark 3.23. *In probability theory, σ-algebra will change all the time as the random variable changes. In contrast, in real analysis, once we define the measurable map, the σ-algebra is usually not important any more.*

Setup 2

Consider the measurable map

$$X : (\Omega, \mathscr{B}) \longrightarrow (\Omega', \mathscr{B}')$$

which we mean $X \in \mathscr{B}/\mathscr{B}'$.

Definition 3.24. *The σ-field generated by X is defined as*

$$\sigma(X) := X^{-1}(\mathscr{B}')$$

Here, $X \in \mathscr{B}/\mathscr{B}'$ means $X^{-1}(\mathscr{B}') \subset \mathscr{B}$.

Definition 3.25 (measurable with respect to a sub-σ-algebra $\mathcal{F} \subset \mathscr{B}$)**.** *For a sub σ-field $\mathcal{F}$ of $\mathscr{B}$, we say X is measurable with respect to $\mathcal{F}$ and writes $X \in \mathcal{F}$ if $\sigma(X) \subset \mathcal{F}$.*

Example 3.9. *Let $(\Omega, \mathscr{B}, \mathbf{P})$ be a probability space. Let $A \in \mathscr{B}$. Look at the indicator function $X = \mathbf{1}_A$. Then what is $\sigma(X)$? There are only four situations.*

[Board figure omitted: the indicator $\mathbf{1}_A$ on Ω, with Ω split into A and A^c, and the four possible preimages $X^{-1}(B)$ marked]

Consider the preimage $X^{-1}(B)$, there are only four situations.

$$X^{-1}(B) = \begin{cases} A^c & \text{if } 0 \in B, 1 \notin B \\ A & \text{if } 0 \notin B, 1 \in B \\ \varnothing & \text{if } 0 \notin B, 1 \notin B \\ \Omega & \text{if } 0 \in B, 1 \in B \end{cases}$$

Thus,

$$\sigma(X) = X^{-1}(\mathscr{B}(\mathbb{R})) = \{\varnothing, \Omega, A, A^c\}$$

Definition 3.26 (Simple random variable)**.** *A random variable is said to be a simple random variable if the values of X can take is a finite set $\{a_1, a_2, \ldots, a_n\}$. Thus,*

$$X = a_1 \mathbf{1}_{A_1} + \ldots + a_n \mathbf{1}_{A_n}$$

where $A_k \in \mathscr{B}$. In particular, A_k is a partition of Ω, meaning

- $\cup_{k=1}^n A_k = \Omega$
- $A_i \cap A_j = \varnothing$ *for $i \neq j$.*

Thus, the σ-algebra generated by X is of the form

$$\sigma(X) = \sigma(\{A_1, \ldots, A_n\}) = \{\cup_{i \in I} A_i : I \subset \{1, 2, \ldots, n\}\}$$

which is all possible combination (i.e. union) of different atomic pieces.

Proposition 3.27. *If X is a random variable. Let $\mathscr{C}$ be a collection of subsets of $\mathbb{R}$ such that*

$$\sigma(\mathscr{C}) = \mathscr{B}(\mathbb{R})$$

Then the σ-field generated by X is

$$X^{-1}(\mathscr{B}(\mathbb{R})) = \sigma(X) = \sigma(X^{-1}(\mathscr{C})) = X^{-1}(\sigma(\mathscr{C}))$$

Remark 3.28 (about σ-field)**.** *The σ-field generated by a random variable X is the collection of information related to X. Later in STAT 502 will look at a (discrete time) stochastic process $(X_n)_{n \geqslant 1}$ which can be thought as a random system evolve dynamically.*

Example 3.10. *Let X_n be the number of customer goes to a store at time n. We will denote*

$$\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n) = \sigma\left(\bigcup_{i=1}^n \sigma(X_i)\right)$$

That is σ-field generated by n random variables. It can be interpreted as the collection of information relates to $X_1, \ldots, X_n$. (This is all information up to time n).

Remark 3.29 (Fair game (Martingale))**.** *A fair game means given all information up until now, we play one more game, then we are not supposed to win or lose. In terms of conditional expectation, we will have*

$$\mathbf{E}(X_{n+1} | \mathcal{F}_n) = X_n$$

4 Independence

4.1 Basic Definitions

Lectures 2024-10-17, 2024-10-19

Definition 4.1 (Independence of events)**.** *Two events A and B are independent if*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$$

Remark 4.2. *If A and B are independent, then the question is do we have A^c and B^c independent? As well as A and B^c, A^c and B. The answer is yes. First let's consider do we have*

$$\mathbf{P}(A \cap B^c) = \mathbf{P}(A)\,\mathbf{P}(B^c)$$

Because

$$\begin{aligned}\mathbf{P}(A \cap B^c) &= \mathbf{P}(A \backslash B) = \mathbf{P}(A) - \mathbf{P}(A \cap B) \\ &= \mathbf{P}(A) - \mathbf{P}(A)\,\mathbf{P}(B) = \mathbf{P}(A)(1 - \mathbf{P}(B)) \\ &= \mathbf{P}(A)\,\mathbf{P}(B^c)\end{aligned}$$

Similarly, A^c, B are also independent. Lastly, for A^c and B^c:

$$\begin{aligned}\mathbf{P}(A^c \cap B^c) &= \mathbf{P}((A \cup B)^c) = 1 - \mathbf{P}(A \cup B) \\ &= 1 - \mathbf{P}(A) - \mathbf{P}(B) + \mathbf{P}(A \cap B) \\ &= 1 - \mathbf{P}(A) - \mathbf{P}(B) + \mathbf{P}(A)\,\mathbf{P}(B) \\ &= 1 - \mathbf{P}(A)(1 - \mathbf{P}(B)) - \mathbf{P}(B) \\ &= (1 - \mathbf{P}(B))(1 - \mathbf{P}(A)) \\ &= \mathbf{P}(A^c)\,\mathbf{P}(B^c)\end{aligned}$$

Definition 4.3. *1) Let $A_1, \ldots, A_n$ are n events. They are independent if*

$$\mathbf{P}\left(\bigcap_{i=1}^{n} C_i\right) = \prod_{i=1}^{n} \mathbf{P}(C_i)$$

where $C_i = A_i$ or A_i^c.

2) The above statement is equivalent to

$$\mathbf{P}\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} \mathbf{P}(B_i)$$

where

$$B_i = A_i \text{ or } \Omega$$

Another way to phrase this is the following

$$\mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i)$$

for all $I \subset \{1, \ldots, n\}$.

Definition 4.4 (Pairwise independence)**.** *If $A_1, \dots, A_n$ are events. We say they are pairwise independent if*

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\,\mathbf{P}(A_j)$$

for all $i \neq j$.

Remark 4.5. *But pairwise independent does not implies independence.*

Example 4.1 (Non-example)**.** *Consider four pairwise disjoint events C_1, C_2, C_3, C_4 (e.g. a partition of Ω into four equally likely pieces) where*

$$\mathbf{P}(C_i) = \frac{1}{4}$$

Define

$$\begin{aligned} A_1 &= C_1 \cup C_2 \\ A_2 &= C_2 \cup C_3 \\ A_3 &= C_1 \cup C_3 \end{aligned}$$

See that A_1, A_2, A_3 are pair-wise independent. But

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) \neq \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3)$$

Remark 4.6. *If two events A, B are disjoint, they must be dependent with each other. Because for $\mathbf{P}(A) > 0$ and $\mathbf{P}(B) > 0$, we will have*

$$\mathbf{P}(A \cap B) = 0$$

but

$$\mathbf{P}(A)\,\mathbf{P}(B) \neq 0$$

Definition 4.7 (Independence of collection of sets)**.** *If we have n collection of subsets of Ω, $\mathscr{C}_1, \dots, \mathscr{C}_n \subset \mathcal{P}(\Omega)$. We say that they are independent if any choice of $C_1 \in \mathscr{C}_1, \dots, C_n \in \mathscr{C}_n$, we have $C_1, \dots, C_n$ are independent.*

Definition 4.8 (Independence of an arbitrary collection of events)**.** *Suppose T is an arbitrary index set. For each $t \in T$, $\mathscr{C}_t$ is a collection of events. We say $\{\mathscr{C}_t : t \in T\}$ are independent if for any finite set $I \subset T$, $\{\mathscr{C}_t : t \in I\}$ are independent.*

Remark 4.9 (FDD)**.** *The name FDD refers to the* ***finite dimensional distributions*** *of a stochastic process. A* ***stochastic process*** *(random process) is a collection of random variables $\{X_t : t \in T\}$ indexed by an arbitrary set T (e.g. $T = \mathbb{Z}_{\geqslant 1}$ or $T = [0, \infty)$), all defined on the same probability space. For a finite subset $I = \{t_1, \dots, t_n\} \subset T$, the joint distribution of the random vector $(X_{t_1}, \dots, X_{t_n})$ is called a finite dimensional distribution of the process. The definitions above follow the same pattern: independence of an arbitrary collection is defined through its finite subcollections, i.e. through finite dimensional objects.*

4.2 Independent Random Variables

Lectures 2024-10-17, 2024-10-19

Definition 4.10 (Independence of random variable)**.** *Let $X_1, \ldots, X_n$ be random variables. We say that they are independent if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.*

Lemma 4.11. *Let $\mathscr{C}_1, \ldots, \mathscr{C}_n \subset \mathcal{P}(\Omega)$. Suppose*

1) $\mathscr{C}_i$ is a π-system for $i = 1, \ldots, n$.

2) $\mathscr{C}_1, \ldots, \mathscr{C}_n \subset \mathcal{P}(\Omega)$ are independent

Then $\sigma(\mathscr{C}_1), \ldots, \sigma(\mathscr{C}_n)$ are independent.

Proof. Read the book. □

Proposition 4.12. *Let $X_1, \ldots, X_n$ be random variables. They are independent if and only if*

$$\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i)$$

for all $x_i \in \mathbb{R}$ for $i = 1, \ldots, n$.

Proof. **Forward direction $\Longrightarrow$.** If $X_1, \ldots, X_n$ are independent. Then the following will hold

$$\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i)$$

Because

$$\begin{aligned}\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) &= \mathbf{P}\left(\bigcap_{i=1}^{n} X_i^{-1}((-\infty, x_i])\right) \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i^{-1}((-\infty, x_i])) \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i)\end{aligned}$$

Backward direction $\Longleftarrow$. Suppose now the converse holds, i.e. we have

$$\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i)$$

We are trying to show $X_1, \ldots, X_n$ are independent. Set

$$\mathscr{C}_i := \{\{X_i \leqslant x_i\} : x_i \in \mathbb{R}\}$$

Obviously, $\mathscr{C}_i$ is a π-system, because

$$\{X_i \leqslant x_i\} \cap \{X_i \leqslant \tilde{x}_i\} = \{X_i \leqslant \min\{x_i, \tilde{x}_i\}\} \in \mathscr{C}_i$$

Hence, the assumption

$$\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i)$$

just means $\mathscr{C}_i$'s are independent. So the previous lemma implies

$$\sigma(\mathscr{C}_1), \ldots, \sigma(\mathscr{C}_n)$$

are independent.

The last thing to show is that $\sigma(\mathscr{C}_i) = \sigma(X_i)$. See that

$$\begin{aligned}
\sigma(\mathscr{C}_i) &= \sigma(\{X_i^{-1}((-\infty, x_i]) : x_i \in \mathbb{R}\}) \\
&= \sigma(X_i^{-1}(\{(-\infty, x_i] : x_i \in \mathbb{R}\})) \\
&= X_i^{-1}(\sigma(\{(-\infty, x_i] : x_i \in \mathbb{R}\})) \\
&= X_i^{-1}(\mathscr{B}(\mathbb{R})) \\
&= \sigma(X_i)
\end{aligned}$$

□

Corollary 4.13. *In the case of discrete random variables $X_1, \ldots, X_n$ are discrete, that is the values it can take is a countable set (e.g. $X_i \in \mathbb{Z}$). Then they are independent if and only if the following is true.*

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall x_i \in \mathbb{R}$$

This is joint PMF and marginal PMFs.

Proof. If independent, then the condition

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall x_i \in \mathbb{R}$$

will hold.

All we need to show is, if the condition

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall x_i \in \mathbb{R}$$

holds, then we will have independence.

So basically we are trying to show, if we have

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall x_i \in \mathbb{R}$$

then we will have

$$\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \leqslant x_i), \qquad \forall x_i \in \mathbb{R}$$

Since each X_i is discrete, write its possible values as $\{a_{ij} : j \in \mathbb{Z}_{\geqslant 1}\}$. Let's look $\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$

$$\begin{aligned}
\mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) &= \sum_{\substack{j_1,\ldots,j_n:\\ a_{ij_i}\leqslant x_i \text{ for } i=1,\ldots,n}} \mathbf{P}(X_1 = a_{1j_1}, \ldots, X_n = a_{nj_n}) \\
&= \sum_{\substack{j_1,\ldots,j_n:\\ a_{ij_i}\leqslant x_i \text{ for } i=1,\ldots,n}} \prod_{i=1}^n \mathbf{P}(X_i = a_{ij_i}) \\
&= \sum_{\substack{j_2,\ldots,j_n:\\ a_{ij_i}\leqslant x_i,\ i\geqslant 2}} \sum_{j_1:\ a_{1j_1}\leqslant x_1} \left(\prod_{i=2}^n \mathbf{P}(X_i = a_{ij_i})\right) \mathbf{P}(X_1 = a_{1j_1}) \\
&= \left(\sum_{\substack{j_2,\ldots,j_n:\\ a_{ij_i}\leqslant x_i,\ i\geqslant 2}} \prod_{i=2}^n \mathbf{P}(X_i = a_{ij_i})\right) \underbrace{\left(\sum_{j_1:\ a_{1j_1}\leqslant x_1} \mathbf{P}(X_1 = a_{1j_1})\right)}_{\mathbf{P}(X_1\leqslant x_1)} \\
&= \left(\sum_{\substack{j_2,\ldots,j_n:\\ a_{ij_i}\leqslant x_i,\ i\geqslant 2}} \prod_{i=2}^n \mathbf{P}(X_i = a_{ij_i})\right) \mathbf{P}(X_1 \leqslant x_1) \\
&= \vdots \\
&= \prod_{i=1}^n \mathbf{P}(X_i \leqslant x_i)
\end{aligned}$$

where in the middle, we just keep repeating the same trick. This finishes the proof. □

Definition 4.14 (Independence of an arbitrary collection of random variables)**.** *Suppose T is an arbitrary index set. For each $t \in T$, X_t is a random variable. We say $\{X_t : t \in T\}$ are independent if for any finite $I \subset T$, the collection of random variables $\{X_t : t \in I\}$ are independent.*

Example 4.2. *Consider a sequence of random variables $\{X_i : i \in \mathbb{N}\} = \{X_1, \ldots, X_n, \ldots\}$. We say they are independent if any finite collection of them are independent.*

Remark 4.15. *If $X_1, \ldots, X_n$ are independent. Then any sub collection of them are independent. For example X_1, X_2 are independent; X_2, X_3, X_4, X_n are independent, and etc.*

4.3 Two Examples of Independence

Lectures 2024-10-19, 2024-10-22

Example 4.3 (10-nary expansion)**.** *Consider* 0.91*. This means we divide an unit interval* $(0, 1]$ *evenly into ten parts, and look at where does* 0.91 *lies in. Then further divide that interval into another ten parts. So it is*

$$0.91 = \frac{9}{10} + \frac{1}{10^2}$$

Another example

$$0.891 = \frac{8}{10} + \frac{9}{10^2} + \frac{1}{10^3}$$

Thus in general, we will have

$$\sum_{i=1}^{\infty} \frac{a_i}{10^i}$$

Example 4.4 (Dyadic expansion of uniform random variables, Binary expansion)**.** *Consider a probability space* $(\Omega, \mathscr{B}, \mathbf{P}) = ((0,1], \mathscr{B}((0,1]), \lambda)$ *where* λ *is Lebesgue measure. Then for any* $\omega \in (0,1]$ *we have the so-called dyadic expansion of* ω *which is defined as following:*

$$\omega = \sum_{i=1}^{\infty} \frac{a_i}{2^i} = 0.a_1a_2a_3\ldots, \qquad a_i = 0 \text{ or } 1$$

This is just a binary expansion of a number in $(0,1]$*.*

[Board figure omitted: the unit interval $(0,1]$ *repeatedly halved, showing how the digits* $a_1, a_2, a_3, \ldots$ *locate* ω*]*

Example 4.5. *For* $\frac{1}{2}$*, the convention for finite expansion is*

$$\frac{1}{2} = 0.1$$

and the convention for infinite expansion is

$$\frac{1}{2} = 0.011111\ldots$$

Observation

From this expansion

$$\omega = \sum_{i=1}^{\infty} \frac{a_i}{2^i} = 0.a_1a_2a_3\ldots, \qquad a_i = 0 \text{ or } 1$$

we get a sequence of functions on Ω,

$$a_1(\omega), \quad a_2(\omega), \qquad \ldots$$

We then have the following claims:

- Claim I: Each $a_i(\omega)$ is a random variable.
- Claim II: $\mathbf{P}(a_i = 0) = \frac{1}{2}$, and $\mathbf{P}(a_i = 1) = \frac{1}{2}$, which means all $a_i(\omega)$ are identically distributed
- Claim III: they are independent.

So we just construct a sequence of maps in a special way, and it turns out that they are IID sequence.

Proof of Claim I. *Several observations.* Let's look at the map $a_1(\omega)$. Consider the event

$$\{\omega \in \Omega : a_1(\omega) = 0\} = \{0.0 * * \ldots : * = 0 \text{ or } 1\} = \{0.000\ldots, \ldots, 0.01111\ldots\} = (0, \tfrac{1}{2}]$$

Consider a different event

$$\{\omega \in \Omega : a_1(\omega) = 1\} = \{0.1 * * \ldots : * = 0 \text{ or } 1\} = \{0.100\ldots, \ldots, 0.11111\ldots\} = (\tfrac{1}{2}, 1]$$

Since a_1 takes only the two values 0 and 1, for any $B \in \mathscr{B}(\mathbb{R})$ the preimage $a_1^{-1}(B)$ is one of the four sets

$$\varnothing, \qquad (0, \tfrac{1}{2}], \qquad (\tfrac{1}{2}, 1], \qquad \Omega$$

according to whether $0 \in B$ and whether $1 \in B$; all four lie in $\mathscr{B}((0,1])$. Hence a_1 is a random variable. The same argument works for every a_n: the sets $a_n^{-1}(0)$ and $a_n^{-1}(1)$ are finite unions of dyadic intervals (written out under Claim II below), hence lie in $\mathscr{B}((0,1])$, and a_n is a random variable.

The expansion is unique if we pick the expansion that has infinitely many 1's.

Proof of Claim II.
Look at a_2.

[Board figure omitted: the four dyadic subintervals of $(0,1]$ of length $1/4$, marked according to whether $a_2 = 0$ or $a_2 = 1$ on them]

Consider the event

$$\begin{aligned} a_2^{-1}(1) &= \{\omega \in \Omega : a_2(\omega) = 1\} \\ &= \{(0.*100\ldots, 0.*111\ldots]\} \\ &= (0.0100\ldots, 0.0111\ldots] \cup (0.1100\ldots, 0.1111\ldots] \\ &\in \mathscr{B}((0,1]) \end{aligned}$$

Also,

$$\begin{aligned} a_2^{-1}(0) &= \{\omega \in \Omega : a_2(\omega) = 0\} \\ &= \{(0.*000\ldots, 0.*011\ldots]\} \\ &= (0.0000\ldots, 0.0011\ldots] \cup (0.1000\ldots, 0.1011\ldots] \\ &\in \mathscr{B}((0,1]) \end{aligned}$$

Look at a_n in general. Consider the event

$$\begin{aligned} a_n^{-1}(0) &= \{\omega \in \Omega : a_n(\omega) = 0\} \\ &= \text{the union of } 2^{n-1} \text{ disjoint intervals of length } \frac{1}{2^n} \end{aligned}$$

and similarly,

$$\begin{aligned} a_n^{-1}(1) &= \{\omega \in \Omega : a_n(\omega) = 1\} \\ &= \text{the union of } 2^{n-1} \text{ disjoint intervals of length } \frac{1}{2^n} \end{aligned}$$

Thus,

$$\mathbf{P}(a_n = 0) = \frac{1}{2^n} \times 2^{n-1} = \frac{1}{2}, \qquad \mathbf{P}(a_n = 1) = \frac{1}{2}$$

Proof of Claim III. We are trying to prove a_i's are independent. This gives us an IID sequence of Bernoulli random variables. We need to show

$$\mathbf{P}(a_1 = u_1, \ldots, a_n = u_n) = \prod_{i=1}^{n} \mathbf{P}(a_i = u_i) = \frac{1}{2^n}$$

for all $u_i = 0$ or 1 and all $n \geqslant 1$. Consider the event

$$\begin{aligned} \{\omega \in \Omega : a_1(\omega) = u_1, \ldots, a_n(\omega) = u_n\} &= (0.u_1u_2 \ldots u_n00 \ldots, 0.u_1u_2 \ldots u_n11 \ldots] \\ &= \left(\sum_{i=1}^{n} \frac{u_i}{2^i}, \sum_{i=1}^{n} \frac{u_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} \right] \end{aligned}$$

Since the probability of an interval is its length,

$$\begin{aligned} \mathbf{P}(a_1 = u_1, \ldots, a_n = u_n) &= \left(\sum_{i=1}^{n} \frac{u_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} \right) - \sum_{i=1}^{n} \frac{u_i}{2^i} \\ &= \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n} = \prod_{i=1}^{n} \mathbf{P}(a_i = u_i) \end{aligned}$$

which is exactly the required factorization.

4.4 More on Independence: Groupings

Lecture 2024-10-22

Lemma 4.16 (Grouping). *Suppose $\{\mathscr{B}_t : t \in T\}$ is an independent family of σ-fields, where T is an index set. Let S be an index set, and suppose for each $s \in S$ we are given a subset $T_s \subset T$ such that the collection $\{T_s : s \in S\}$ is pairwise disjoint. Set*

$$\mathscr{B}_{T_s} := \sigma(\mathscr{B}_t : t \in T_s) = \sigma\left(\bigcup_{t \in T_s} \mathscr{B}_t \right)$$

Then $\{\mathscr{B}_{T_s} : s \in S\}$ are independent.

Example 4.6. *Consider a sequence of independent σ-fields*

$$\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_n, \ldots$$

Let $S = \{1, 2, 3, \ldots\}$. Consider the σ-fields generated by finitely many of σ-fields above

1. $\sigma(\mathscr{B}_1, \mathscr{B}_2)$ *where* $T_1 = \{1, 2\}$
2. $\sigma(\mathscr{B}_3, \mathscr{B}_4, \mathscr{B}_7)$ *where* $T_2 = \{3, 4, 7\}$
3. $\sigma(\mathscr{B}_6, \mathscr{B}_8)$ *where* $T_3 = \{6, 8\}$

Because T_1, T_2, T_3 *are all disjoint. Thus, the above three* σ*-fields are independent.*

Example 4.7. *Consider a sequence of independent random variables*

$$X_1, X_2, X_3, \ldots$$

Then consider the following σ*-fields*

1. $\sigma(X_1, X_5)$
2. $\sigma(X_2, X_7, X_{100})$
3. $\sigma(X_{13}, X_{14})$
4. $\sigma(X_{101}, X_{102}, \ldots)$

The claim is, applying grouping lemma, these σ*-fields are independent. First,*

$$\mathscr{B}_k = \sigma(X_k)$$

Let $T = \{1, 2, 3, \ldots\}$ *and* $S = \{1, 2, 3, 4\}$*. Then*

1. $T_1 = \{1, 5\} \subset T$
2. $T_2 = \{2, 7, 100\} \subset T$
3. $T_3 = \{13, 14\} \subset T$
4. $T_4 = \{101, 102, \ldots\} \subset T$

Under this new setting, we have

1. $\mathscr{B}_{T_1} = \sigma(X_1, X_5)$
2. $\mathscr{B}_{T_2} = \sigma(X_2, X_7, X_{100})$
3. $\mathscr{B}_{T_3} = \sigma(X_{13}, X_{14})$
4. $\mathscr{B}_{T_4} = \sigma(X_{101}, X_{102}, \ldots)$

Because index sets T_1, T_2, T_3, T_4 *are all disjoint. Thus by the grouping lemma, these* σ*-fields are independent. Which means random variables are independent.*

Question. *Are*

$$X_1 + X_5, \quad (X_2^2 + X_7^3)X_{100}, \quad X_{13}\log(X_{14}), \quad \limsup_{n\to\infty} X_n$$

are independent?

The claim is continuous manipulation of independent random variables are still going to be independent.

Because

$$X_1 \in \sigma(X_1) \subset \sigma(X_1, X_5), \quad X_5 \in \sigma(X_5) \subset \sigma(X_1, X_5)$$

Hence apply a continuous manipulation to X_1, X_5, we have

$$X_1 + X_5 \in \sigma(X_1, X_5)$$

In addition, we define

1. $\mathscr{B}_{T_1} = \sigma(X_1, X_5)$
2. $\mathscr{B}_{T_2} = \sigma(X_2, X_7, X_{100})$
3. $\mathscr{B}_{T_3} = \sigma(X_{13}, X_{14})$
4. $\mathscr{B}_{T_4} = \sigma(X_n : n \geqslant 101)$

Each of the four random variables in the question is measurable with respect to the corresponding σ-field: $X_1 + X_5 \in \mathscr{B}_{T_1}$ as shown above, and in the same way $(X_2^2 + X_7^3)X_{100} \in \mathscr{B}_{T_2}$ and $X_{13}\log(X_{14}) \in \mathscr{B}_{T_3}$, being continuous manipulations of the generating variables; and $\limsup_{n\to\infty} X_n \in \mathscr{B}_{T_4}$, because the limsup does not depend on any finite initial segment of the sequence, so it is measurable with respect to $\sigma(X_n : n \geqslant 101)$. The index sets $T_1 = \{1, 5\}$, $T_2 = \{2, 7, 100\}$, $T_3 = \{13, 14\}$, $T_4 = \{101, 102, \ldots\}$ are pairwise disjoint, so by the grouping lemma the four σ-fields are independent, and therefore so are the four random variables. This answers the question affirmatively.

4.5 Independence, Zero-One Laws, Borel-Cantelli Lemma

Lectures 2024-10-22, 2024-10-24, 2024-10-26, 2024-10-29

Lemma 4.17 (Borel-Cantelli Lemma)**.** *Suppose A_n's is a sequence of events such that*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$$

Then

$$\mathbf{P}(A_n \textit{ happens i.o.}) = \mathbf{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = 0$$

Example 4.8. *Suppose A_n is the event that the light breaks on day n. Let's say*

$$\mathbf{P}(A_n) = \frac{1}{2^n}$$

This means when time evolves, the technology gets better and better so that the probability of the light breaks is getting lower and lower. Then

$$\mathbf{P}(A_n \textit{ i.o. }) = 0$$

This means with 0 *probability that* A_n *happens infinitely often. If we look at the complement where*

$$\{A_n \ i.o.\}^c = \left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty} A_n^c$$

Its probability is going to be

$$\mathbf{P}(\{A_n \ i.o.\}^c) = 1$$

That means for some n_0, A_n^c *will keep happening for* $n \geqslant n_0$. *That means, take any* $\omega \in \liminf_{n\to\infty} A_n^c$, *there exists* n_0 *such that* $\omega \in A_n^c$ *for all* $n \geqslant n_0$. *(It means there is a certain day such that after that day, the light does not break any more).*

Proof. Consider the following probability

$$\begin{aligned}
\mathbf{P}(A_n \text{ i.o. }) = \mathbf{P}\left(\limsup_{n\to\infty} A_n\right) &= \mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) \\
&= \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\
&\leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbf{P}(A_k) \\
&= 0
\end{aligned}$$

where we use continuity of probability and countable sub-additivity of measure. The

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{n-1} a_k + \sum_{k=n}^{\infty} a_k$$

Send $n \to \infty$, then

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} a_k + \lim_{n\to\infty} \sum_{k=n}^{\infty} a_k$$

□

Theorem 4.18 (Borel zero-one law)**.** *If* A_n *is a sequence of independent sets. Then*

$$\mathbf{P}(A_n \ i.o.) = \begin{cases} 0, & \text{if } \sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty \\ 1, & \text{if } \sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty \end{cases}$$

Proof. The first part is by Borel Cantelli lemma. The only part we need to show is the second part, i.e. if $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$, then we have

$$\mathbf{P}(A_n \text{ i.o.}) = 1$$

Observe that

$$\begin{aligned}
\mathbf{P}(A_n \text{ i.o. }) &= \mathbf{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) \\
&= 1 - \mathbf{P}\left(\liminf_{n\to\infty} A_n^c\right) \\
&= 1 - \mathbf{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) \\
&= 1 - \lim_{n\to\infty} \mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right), \qquad \text{by continuity of probability} \\
&= 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} \mathbf{P}(A_k^c), \qquad \text{by independence} \\
&= 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} \underbrace{(1-\mathbf{P}(A_k))}_{\leqslant e^{-\mathbf{P}(A_k)}}, \qquad \text{by the fact } 1+x \leqslant e^x \text{ with } x = -\mathbf{P}(A_k) \\
&\geqslant 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} e^{-\mathbf{P}(A_k)} \\
&= 1 - \lim_{n\to\infty} \underbrace{e^{-\sum_{k=n}^{\infty}\mathbf{P}(A_k)}}_{=e^{-\infty}}, \qquad \text{by assumption} \\
&= 1 - \lim_{n\to\infty} 0 = 1
\end{aligned}$$

Since a probability is never larger than 1, the inequality forces $\mathbf{P}(A_n \text{ i.o.}) = 1$. This completes the proof. $\square$

Example 4.9 (Borel-Cantelli)**.** *Consider X_n are IID unit exponential random variables $X_n \sim Exp(1)$. That is*

$$\mathbf{P}(X_n > x) = e^{-x}, \qquad x \geqslant 0$$

Then

$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{X_n}{\log_e(n)} = 1\right) = 1$$

Mathematically, we get a new sequence of random variables $Y_n = \frac{X_n}{\log_e(n)}$. Taking limsup of this new sequence and set it equal to 1.

The heuristics is, we sampling this X_n's. This get a sequence of dots on the plane.

[Board figure omitted: a scatter of the sampled values X_n against n, with the curve $\log n$ traced through the highest dots]

We want to see what is the curve from starting point to the highest dots. The statement says that this curve is almost surely (means with probability 1) a log function $\log(n)$.

This is a consequence of Borel zero-one law.

Proof. What is the meaning of the event

$$\left\{\omega \in \Omega : \limsup_{n\to\infty} \frac{X_n(\omega)}{\log(n)} = 1\right\}$$

It means for any $\epsilon > 0$, we should have the following two cases

1) If we look at the point $1 + \epsilon$ on the real line $\mathbb{R}$, there will be only finitely many points $\frac{X_n(\omega)}{\log(n)}$ to the right of this point $1 + \epsilon$. This means

$$\frac{X_n(\omega)}{\log n} \leqslant 1 + \epsilon \qquad \text{for all large } n$$

2) If we look at the point $1 - \epsilon$ on the real line $\mathbb{R}$, there will be infinitely many points $\frac{X_n(\omega)}{\log(n)}$ to the right of this point $1 - \epsilon$. This means

$$\frac{X_n(\omega)}{\log n} > 1 - \epsilon \qquad \text{for infinitely many } n$$

For fixed ϵ, $\omega \in \left\{\omega \in \Omega : \limsup_{n\to\infty} \frac{X_n(\omega)}{\log(n)} = 1\right\}$. Then ω satisfies (1) and (2). If we define a sequence of sets (depend on ϵ) as

$$A_n^\epsilon := \left\{\omega \in \Omega : \frac{X_n(\omega)}{\log(n)} \leqslant 1 + \epsilon\right\}$$

The collection of ω's satisfying (1) is defined as

$$\liminf_{n\to\infty} A_n^\epsilon$$

Similarly, if we define a sequence of sets (depend on ϵ) as

$$B_n^\epsilon := \left\{\omega \in \Omega : \frac{X_n(\omega)}{\log(n)} > 1 - \epsilon\right\}$$

The collection of ω's satisfying (2) is defined as

$$\limsup_{n\to\infty} B_n^\epsilon$$

Next, for fixed $\epsilon > 0$, the collection of ω's satisfying (1) and (2) will be in the intersection of limsup and liminf we defined above, i.e.

$$\left(\liminf_{n\to\infty} A_n^\epsilon\right) \cap \left(\limsup_{n\to\infty} B_n^\epsilon\right)$$

We pick a sequence of ϵ_k such that $\epsilon_k \downarrow 0$. Then if

$$\omega \in \left\{\omega \in \Omega : \limsup_{n\to\infty} \frac{X_n(\omega)}{\log(n)} = 1\right\}$$

Then for any ϵ_k, (1) and (2) should be satisfied, i.e.

$$\omega \in \left(\liminf_{n\to\infty} A_n^{\epsilon_k}\right) \cap \left(\limsup_{n\to\infty} B_n^{\epsilon_k}\right), \qquad \forall \epsilon_k$$

This means

$$\left\{\omega \in \Omega : \limsup_{n\to\infty} \frac{X_n(\omega)}{\log(n)} = 1\right\} = \bigcap_{k=1}^{\infty}\left(\left(\liminf_{n\to\infty} A_n^{\epsilon_k}\right) \cap \left(\limsup_{n\to\infty} B_n^{\epsilon_k}\right)\right)$$

Let's further define

$$C_k := \liminf_{n\to\infty} A_n^{\epsilon_k}$$

and

$$D_k = \limsup_{n\to\infty} B_n^{\epsilon_k}$$

If we can show $\mathbf{P}(C_k) = \mathbf{P}(D_k) = 1$ for all $k \geqslant 1$. Then by Exercise 2.6.11, we will have

$$\mathbf{P}\left(\bigcap_{k=1}^{\infty}(C_k \cap D_k)\right) = 1$$

First, look at

$$\mathbf{P}(D_k) = \mathbf{P}\left(\limsup_{n\to\infty} B_n^{\epsilon_k}\right)$$

where

$$\begin{aligned}\mathbf{P}(B_n^{\epsilon_k}) &= \mathbf{P}\left(\frac{X_n}{\log n} > 1 - \epsilon_k\right)\\ &= \mathbf{P}(X_n > (1-\epsilon_k)\log n)\\ &= e^{-(1-\epsilon_k)\log n}\\ &= \frac{1}{n^{1-\epsilon_k}}\end{aligned}$$

Since $1 - \epsilon_k < 1$. Then

$$\sum_{n=1}^{\infty} \mathbf{P}(B_n^{\epsilon_k}) = \sum_{n=1}^{\infty} \frac{1}{n^{1-\epsilon_k}} = \infty$$

Moreover, $B_n^{\epsilon_k}$ for $n = 1, 2, \ldots$ are independent. Hence, by Borel zero-one law, we have

$$\mathbf{P}(B_n^{\epsilon_k} \text{ i.o.}) = 1$$

Next, look at $\mathbf{P}(C_k)$.

$$\mathbf{P}(C_k) = \mathbf{P}\left(\liminf_{n\to\infty} A_n^{\epsilon_k}\right) = 1 - \mathbf{P}\left(\limsup_{n\to\infty} (A_n^{\epsilon_k})^c\right)$$

Consider

$$\mathbf{P}((A_n^{\epsilon_k})^c) = \mathbf{P}\left(\frac{X_n}{\log n} > 1 + \epsilon_k\right) = \frac{1}{n^{1+\epsilon_k}}$$

Hence, we have

$$\sum_{n=1}^{\infty} \mathbf{P}((A_n^{\epsilon_k})^c) = \sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon_k}} < \infty$$

Moreover, $(A_n^{\epsilon_k})^c$ for $n = 1, 2, \ldots$ are independent. Hence, by Borel zero-one law, we have

$$\mathbf{P}((A_n^{\epsilon_k})^c \text{ i.o.}) = 0$$

□

Definition 4.19 (Kolmogorov zero-one law — set up). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. We define*

1) $\mathscr{F}_n := \sigma(X_1, \ldots, X_n)$
2) $\mathscr{F}_n' := \sigma(X_{n+1}, X_{n+2}, \ldots)$
3) *The tail σ-field $\mathscr{T}$ to be*

$$\mathscr{T} := \bigcap_{n=1}^{\infty} \mathscr{F}_n'$$

Note that $\mathscr{F}_1' \supset \mathscr{F}_2' \supset \mathscr{F}_3' \ldots$.

Example 4.10. *Consider the following event*

$$A = \left\{\omega \in \Omega : \sum_{k=1}^{\infty} X_k^2(\omega) \text{ converges}\right\}$$

Then $A \in \mathscr{T}$. That is because the event

$$\sum_{k=1}^{\infty} X_k^2(\omega) \qquad \text{converges}$$

is equivalent to the following sum converges

$$\sum_{k=n+1}^{\infty} X_k^2(\omega) \qquad \text{converges for all } n \geqslant 1$$

Thus,

$$\left\{\omega \in \Omega : \sum_{k=1}^{\infty} X_k^2(\omega) \text{ converges}\right\} = \bigcap_{n=1}^{\infty}\left\{\omega \in \Omega : \sum_{k=n+1}^{\infty} X_k^2(\omega) \text{ converges}\right\}$$

And see that the following event belongs to $\mathscr{F}_n'$.

$$\left\{\omega \in \Omega : \sum_{k=n+1}^{\infty} X_k^2(\omega) \text{ converges}\right\} \in \mathscr{F}_n'$$

because $\sigma(X_k) \subset \mathscr{F}_n'$. *Hence, the event* A *is measurable with respect to* $\mathscr{F}_n'$ *for all* $n \geqslant 1$, *i.e.*

$$\left\{\omega \in \Omega : \sum_{k=1}^{\infty} X_k^2(\omega) \text{ converges}\right\} \in \mathscr{F}_n', \qquad \forall n \geqslant 1$$

Hence,

$$A \in \bigcap_{n=1}^{\infty} \mathscr{F}_n'$$

See that the event A *has nothing to do with respect to first n random variables* $X_1, \ldots, X_n$. *Because whether the series converges or not depend only on the random variables in the later part of the series.*

Example 4.11. *Let* X_n *be bounded random variables. (e.g. let* X_n *be uniform RV on* $[0,1]$*). Then*

$$\limsup_{n\to\infty} X_n \in \mathscr{T}$$

Still because $\limsup_{n\to\infty} X_n$ *has nothing to do with first n random variables.*

Definition 4.20 (Almost trivial σ-field)**.** *A* σ*-field* $\mathscr{G}$ *is said to be almost trivial (with respect to* $\mathbf{P}$*) if for all* $G \in \mathscr{G}$, *we have*

$$\mathbf{P}(G) = 0 \text{ or } 1$$

Lemma 4.21. *Suppose* X *is a real-valued random variable (which means* X *does not take* ∞ *as its values). And* $\mathscr{G}$ *is an almost trivial* σ*-field such that* X *is* $\mathscr{G}$*-measurable, i.e.* $X \in \mathscr{G}$. *Then in this case*

$$X = \text{ constant with probability } 1$$

That is, there exists $c \in \mathbb{R}$ *such that*

$$\mathbf{P}(X = c) = 1$$

Proof. Consider the distribution function

$$F(x) = \mathbf{P}(X \leqslant x) = \mathbf{P}(X^{-1}((-\infty, x]))$$

Because $X \in \mathscr{G}$, we have

$$X^{-1}((-\infty, x]) \in \mathscr{G}$$

But we assumed $\mathscr{G}$ is almost trivial, hence

$$F(x) = \mathbf{P}(X \leqslant x) = \mathbf{P}(X^{-1}((-\infty, x])) = 0 \text{ or } 1$$

This means

$$\lim_{x\to\infty} F(x) = 1$$

$$\lim_{x\to-\infty} F(x) = 0$$

The only possibility that this can happen is, at some value $c \in \mathbb{R}$, $F(x)$ suddenly jump to 1, i.e.

$$\mathbf{P}(X = c) = 1$$

The reason is, suppose the jump happens at c. If we consider the event $X \in (c-\epsilon, c+\epsilon]$ for some small $\epsilon > 0$,

$$\begin{aligned}\mathbf{P}(X \in (c-\epsilon, c+\epsilon]) &= \mathbf{P}(c-\epsilon < X \leqslant c+\epsilon)\\ &= F(c+\epsilon) - F(c-\epsilon)\\ &= 1 - 0\\ &= 1\end{aligned}$$

Then we let $\epsilon \to 0$. We have

$$\mathbf{P}(X = c) = \lim_{\epsilon \to 0} \mathbf{P}(X \in (c-\epsilon, c+\epsilon]) = 1$$

□

Theorem 4.22 (Kolmogorov zero-one law)**.** *Suppose X_n is a sequence of independent random variables. Then the tail σ-field of X_n is almost trivial.*

Example 4.12. *Let X_n be a sequence of bounded and independent random variables. Then*

$$\limsup_{n \to \infty} X_n \in \mathscr{T}$$

This has nothing to do with independence yet. But if we add the assumption of X_n being independent. Then $\mathscr{T}$ will be almost trivial under this assumption. Because by the last lemma, then $\limsup_{n\to\infty} X_n$ will be constant with probability 1.

Proof of the Kolmogorov zero-one law. Pick any $\Lambda \in \mathscr{T}$. Want to show

$$\mathbf{P}(\Lambda) = 0 \text{ or } 1$$

Need to show Λ is independent of itself. We want

$$\mathbf{P}(\Lambda \cap \Lambda) = \mathbf{P}(\Lambda)\,\mathbf{P}(\Lambda)$$

There is only two possible solutions

$$\mathbf{P}(\Lambda) = 0 \text{ or } 1$$

Recall

$$\mathscr{F}_n = \sigma(X_1, X_2, \ldots, X_n)$$

Then

$$\mathscr{F}_n' = \sigma(X_{n+1}, X_{n+2}, \ldots)$$

Introduce

$$\mathscr{F}_\infty := \sigma(X_1, X_2, \ldots)$$

Since $\Lambda \in \mathscr{T} = \bigcap_{n=1}^\infty \mathscr{F}_n' \subset \mathscr{F}_n'$ for all $n \geqslant 1$.

On the other hand, by grouping lemma

$$\mathscr{F}_n' \text{ and } \mathscr{F}_n \text{ are independent}$$

Because $\Lambda \in \mathscr{F}_n'$ is independent of $\mathscr{F}_n$ for all $n \geqslant 1$.

Set $\mathscr{C}_1 = \{\Lambda\}$ and $\mathscr{C}_2 = \bigcup_{n=1}^\infty \mathscr{F}_n$. We have

1) $\mathscr{C}_1$ and $\mathscr{C}_2$ are independent

2) $\mathscr{C}_1$ and $\mathscr{C}_2$ are both π-systems.

Then one of the lemma tells us $\sigma(\mathscr{C}_1)$ is independent to $\sigma(\mathscr{C}_2)$.

Finally, we know

$$\sigma(\mathscr{C}_2) = \mathscr{F}_\infty$$

On the other hand, we also know

$$\Lambda \in \mathscr{T} = \bigcap_{n=1}^{\infty} \underbrace{\mathscr{F}_n'}_{\subset \mathscr{F}_\infty} \subset \mathscr{F}_\infty$$

To summarize, we have $\Lambda \in \sigma(\mathscr{C}_1)$ by construction of $\mathscr{C}_1$, and also $\Lambda \in \sigma(\mathscr{C}_2)$ because $\Lambda \in \mathscr{F}_\infty$. Because $\sigma(\mathscr{C}_1)$ and $\sigma(\mathscr{C}_2)$ are independent. Hence, $\Lambda \in \sigma(\mathscr{C}_1)$ is independent to $\Lambda \in \sigma(\mathscr{C}_2)$. This completes the proof. $\square$

Remark 4.23. *Self-independence gives* $\mathbf{P}(\Lambda) = \mathbf{P}(\Lambda \cap \Lambda) = \mathbf{P}(\Lambda)^2$*, and the only solutions of* $p = p^2$ *are* $p = 0$ *and* $p = 1$*. This is a statement about the* probability *of* Λ*, not about the set itself: it does not force* $\Lambda = \varnothing$ *or* $\Lambda = \Omega$*. For example, on* $((0,1], \mathscr{B}((0,1]), \lambda)$ *the event* $\mathbb{Q} \cap (0,1]$ *has probability* 0 *but is not empty, and its complement has probability* 1 *without being all of* Ω*. This is why the conclusion of the theorem is that the tail* σ*-field is* almost *trivial, not trivial.*

5 Integration and Expectation

5.1 Standard Constructions of Integrations

Lectures 2024-10-29, 2024-10-31

Simple random variable

Definition 5.1 (Simple random variable). *A random variable X is called a simple random variable if the random variable is a finite sum of indicator functions of disjoint events*

$$X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

where $a_i \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathscr{B}$ are disjoint events.

Denote $\mathcal{E}$ be the collection of all simple random variables. (here $\mathcal{E}$ stands for elementary. But do remember some people define an elementary random variable as finite sum of indicator functions on rectangles).

1) $\mathcal{E}$ is a vector space. That is, for any $X, Y \in \mathcal{E}$, and $a, b \in \mathbb{R}$. We have $aX + bY \in \mathcal{E}$. This is because

$$X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}, \qquad Y = \sum_{j=1}^{m} b_j \mathbf{1}_{B_j}$$

Then

$$\begin{aligned} aX + bY &= a \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} + b \sum_{j=1}^{m} b_j \mathbf{1}_{B_j} \\ &= a \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mathbf{1}_{A_i \cap B_j} + b \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mathbf{1}_{A_i \cap B_j} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a a_i + b b_j) \mathbf{1}_{A_i \cap B_j} \\ &\in \mathcal{E} \end{aligned}$$

Similarly, if $X, Y \in \mathcal{E}$. Then $XY \in \mathcal{E}$. Because

$$X \cdot Y = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathbf{1}_{A_i \cap B_j}$$

If $X, Y \in \mathcal{E}$. Then $\max\{X, Y\}, \min\{X, Y\} \in \mathcal{E}$. Because

$$\max\{X, Y\} = \sum_{i=1}^{n} \sum_{j=1}^{m} \max\{a_i, b_j\} \mathbf{1}_{A_i \cap B_j}$$

and similarly,

$$\min\{X, Y\} = \sum_{i=1}^{n} \sum_{j=1}^{m} \min\{a_i, b_j\} \mathbf{1}_{A_i \cap B_j}$$

Theorem 5.2. *Consider a map*

$$X : \Omega \longrightarrow \mathbb{R}$$

where $X \geqslant 0$. Then X is a random variable if and only if there is a sequence of simple random variables $X_n \geqslant 0$ such that $X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$.

Proof. $\Longleftarrow$ **:** We have proved that if we have a sequence of random variables X_n converges to some map. Then that map will be measurable, hence is a random variable.
$\Longrightarrow$ **:** If any non-negative random variable $X \geqslant 0$, we can always find a sequence of simple random variables X_n that goes to X. In this case, we say X_n is an approximating sequence of X. (Note, we only say there exists an approximating sequence. We didn't say this sequence is unique).

Now let's see how to construct one approximating sequence. Define

$$X_n := \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}} + n \mathbf{1}_{B_n}$$

where

$$A_{n,k} := \left\{ \omega \in \Omega : \frac{k-1}{2^n} \leqslant X(\omega) < \frac{k}{2^n} \right\}$$

and

$$B_n := \{\omega \in \Omega : X(\omega) \geqslant n\}$$

[Board figure omitted: the graph of X over Ω with the range cut into the levels $\frac{k-1}{2^n}$, showing the horizontal slabs $A_{n,k}$ and the top region $B_n = \{X \geqslant n\}$]

So that when we consider the events $A_{n,k}$, we are going to have

- $X_n(\omega) \leqslant X(\omega)$ and
- $X_{n+1}(\omega) \geqslant X_n(\omega)$ and
- $X_n(\omega) \uparrow X(\omega)$.
- When $|X(\omega)| < \infty$, then $|X_n(\omega) - X(\omega)| \leqslant \frac{1}{2^n}$ for large enough n. So $X_n(\omega) \to X(\omega)$.

□

Remark 5.3. *the proofs allow X to be $\overline{R}$-valued random variable. That is, it is ok for X to take $+\infty$ as its values. Because if $X(\omega) = \infty$, then $\omega \in B_n$. Hence, $X_n(\omega) = n$. Then $X_n(\omega) \uparrow X(\omega)$.*

$\overline{R}$-valued random variables

Let's start from $\mathbb{R}$-valued random variables. Consider the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Then a random variable is a measurable mapping

$$X : (\Omega, \mathscr{F}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

We write $X \in \mathscr{B}/\mathscr{B}(\mathbb{R})$.

[Board figure omitted: the mapping X from Ω to the real line, with the preimage of a Borel set marked in Ω]

For $\overline{\mathbb{R}}$-valued random variable. Still consider the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Then we are considering a measurable mapping

$$X : (\Omega, \mathscr{F}) \longrightarrow (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$$

We write $X \in \mathscr{B}/\mathscr{B}(\overline{\mathbb{R}})$. The claim is $+\infty$ and $-\infty$ are also points that should be considered. Rigorously speaking, we have the sets

$$D_1 := X^{-1}(\{-\infty\}) \in \mathscr{F}$$
$$D_2 := X^{-1}(\{\infty\}) \in \mathscr{F}$$

[Board figure omitted: the extended real line $\overline{\mathbb{R}}$ with the two extra points $\pm\infty$ attached at its ends, and the two preimages $D_1, D_2 \subseteq \Omega$]

Remark 5.4. *If we have a $\mathbb{R}$-valued RV, we can consider it as a $\overline{\mathbb{R}}$-valued RV. We just set $D_1 = D_2 = \varnothing$.*

Example 5.1. *Compare to a RV taking values in $(0, 1)$, i.e.*

$$Y : (\Omega, \mathscr{F}) \longrightarrow ((0, 1), \mathscr{B}((0, 1)))$$

where we write $Y \in \mathscr{B}/\mathscr{B}((0, 1))$. Then a RV taking values in $[0, 1]$ is

$$Y : (\Omega, \mathscr{F}) \longrightarrow ([0, 1], \mathscr{B}([0, 1]))$$

5.2 Expectation and Integration

Lectures 2024-10-31, 2024-11-02, 2024-11-05

Step 1: Define expectation for Simple RVs

Recall $\mathcal{E}$ is the collection for simple RVs. Define expectation for $X \in \mathcal{E}$, i.e. for

$$X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

where A_i is a partition of Ω. Then we define

$$\mathbf{E}[X] = \sum_{i=1}^{n} a_i \, \mathbf{P}(A_i) = \int X d\, \mathbf{P}$$

Properties of $\mathbf{E}$ on $\mathcal{E}$:

1) $\mathbf{E}[a] = \mathbf{E}[a\mathbf{1}_\Omega] = a\, \mathbf{P}(\Omega) = a$. And $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$ for $A \in \mathscr{F}$.

2) If $X \in \mathcal{E}$ and $X \geqslant 0$. Then $\mathbf{E}\, X \geqslant 0$. Because $X \geqslant 0$ means for $X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$, all $a_i \geqslant 0$.

3) The operator $\mathbf{E}$ is a linear operator on vector space $L^1(\Omega, \mathscr{F}, \mathbf{P})$, i.e. $\mathbf{E}[aX + bY] = a\,\mathbf{E}\, X + b\,\mathbf{E}\, Y$. Because for $X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$ and $Y = \sum_{j=1}^{m} b_j \mathbf{1}_{B_j}$

$$aX + bY = \sum_{i=1}^{n}\sum_{j=1}^{m}(aa_i + bb_j)\mathbf{1}_{A_i \cap B_j}$$

Thus,

$$\begin{aligned}
\mathbf{E}[aX + bY] &= \sum_{i=1}^{n}\sum_{j=1}^{m}(aa_i + bb_j)\,\mathbf{P}(A_i \cap B_j) \\
&= a\sum_{i=1}^{n}\sum_{j=1}^{m} a_i\,\mathbf{P}(A_i \cap B_j) + b\sum_{i=1}^{n}\sum_{j=1}^{m} b_j\,\mathbf{P}(A_i \cap B_j) \\
&= a\sum_{i=1}^{n} a_i\,\mathbf{P}\left(A_i \cap \left(\bigcup_{j=1}^{m} B_j\right)\right) + b\sum_{j=1}^{m} b_j\,\mathbf{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap B_j\right) \\
&= a\sum_{i=1}^{n} a_i\,\mathbf{P}(A_i) + b\sum_{j=1}^{m} b_j\,\mathbf{P}(B_j) \\
&= a\,\mathbf{E}\, X + b\,\mathbf{E}\, Y
\end{aligned}$$

4) If $X, Y \in \mathcal{E}$ and $X \leqslant Y$. Then $\mathbf{E}\, X \leqslant \mathbf{E}\, Y$. Because if $Y - X \geqslant 0$, by Property (3) and (2), we have

$$\mathbf{E}\, Y - \mathbf{E}\, X = \mathbf{E}[Y - X] \geqslant 0$$

5) Let $X \in \mathcal{E}$. Suppose we have a sequence of simple random variables $X_n \in \mathcal{E}$ such that $X_n \uparrow X$ or $X_n \downarrow X$ (i.e. they are monotone). Then $\mathbf{E}\, X_n \uparrow \mathbf{E}\, X$ or $(\mathbf{E}\, X_n \downarrow \mathbf{E}\, X)$, i.e.

$$\mathbf{E} \lim_{n\to\infty} X_n = \lim_{n\to\infty} \mathbf{E}\, X_n$$

WLOG, we only need to show if $X_n \in \mathcal{E}$ such that $X_n \downarrow 0$, then $\mathbf{E}\, X_n \downarrow 0$. (Otherwise, if $X_n \uparrow X$, we can consider $(X - X_n) \downarrow 0$). If $X_n \downarrow X$, can consider $(X_n - X) \downarrow 0$.

Because X_n is a decreasing sequence, and using the fact that simple RVs are bounded, then

$$0 \leqslant X_n \leqslant X_1 \leqslant K \qquad \text{for some constant } K$$

For any $\epsilon > 0$, we have

$$\begin{aligned}
X_n &= \underbrace{X_n}_{\leqslant \epsilon \text{ under } \{X_n \leqslant \epsilon\}} \mathbf{1}_{X_n \leqslant \epsilon} + \underbrace{X_n}_{\leqslant K} \mathbf{1}_{X_n > \epsilon} \\
&\leqslant \epsilon + K\mathbf{1}_{X_n > \epsilon}
\end{aligned}$$

If we take expectation both sides

$$\mathbf{E}\, X_n \leqslant \epsilon + K\,\mathbf{P}(X_n > \epsilon)$$

Use the assumption $X_n \downarrow 0$. Then

$$\lim_{n\to\infty} \{X_n > \epsilon\} = \varnothing$$

Hence, by continuity of probability, we have

$$\lim_{n\to\infty} \mathbf{P}(X_n > \epsilon) = 0$$

Hence,

$$\lim_{n\to\infty} \mathbf{E}\, X_n \leqslant \epsilon + K \lim_{n\to\infty} \mathbf{P}(X_n > \epsilon) = \epsilon$$

Because this holds for all $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbf{E}\, X_n = 0$$

Step 2: Extend $\mathbf{E}\, X$ to non-negative RVs

Set

$$\overline{\mathcal{E}}_+ = \left\{X \geqslant 0, X \text{ is } \overline{R}\text{-valued RV}\right\}$$

We will define expectation for $X \in \overline{\mathcal{E}}_+$ For $X \in \overline{\mathcal{E}}_+$, by the lemma, we can find an approximating sequence $X_n \in \mathcal{E}$, where $0 \leqslant X_n \uparrow X$. Define

$$\mathbf{E}\, X := \lim_{n\to\infty} \mathbf{E}\, X_n$$

The only issue is we have to show this limit is well-defined. That is, this limit is independent of the choice of approximating sequence of simple random variables. If we have two approximating sequence of random variable $X \geqslant 0$ where the two sequences are $X_n \uparrow, Y_n \uparrow \in \mathcal{E}$. Then we should have and

$$\lim_{n\to\infty} X_n \leqslant \lim_{n\to\infty} Y_n \qquad \text{and} \qquad \lim_{n\to\infty} X_n \geqslant \lim_{n\to\infty} Y_n$$

Then by Property (5) we will have

$$\lim_{n\to\infty} \mathbf{E}\, X_n \leqslant \lim_{n\to\infty} \mathbf{E}\, Y_n \qquad \text{and} \qquad \lim_{n\to\infty} \mathbf{E}\, X_n \geqslant \lim_{n\to\infty} \mathbf{E}\, Y_n$$

Proof. We have

- $X_n \wedge Y_m \uparrow X_n$
- $X_n \wedge Y_m \leqslant X_n$

Then by Property (5), we have

- $\lim_{m\to\infty} \mathbf{E}[X_n \wedge Y_m] = \mathbf{E}\, X_n$
- $\lim_{m\to\infty} \mathbf{E}[X_n \wedge Y_m] \leqslant \lim_{m\to\infty} \mathbf{E}\, Y_m$

Hence, $\lim_{m\to\infty} \mathbf{E}\, Y_m \geqslant \mathbf{E}\, X_n$ for all n. Send $n \to \infty$, we get

$$\lim_{m\to\infty} \mathbf{E}\, Y_m \geqslant \lim_{n\to\infty} \mathbf{E}\, X_n$$

□

Written with the notation $\mathcal{E}_+$, this is the extension of $\mathbf{E}\,X$ to

$$\mathcal{E}_+ := \big\{X \geqslant 0 : X \text{ is a } \overline{\mathbb{R}}\text{-valued RV}\big\}$$

by

$$\mathbf{E}\,X := \lim_{n\to\infty} \mathbf{E}\,X_n$$

where $(X_n)_{n\geqslant 1}$ is an approximating sequence of X.
Properties of $\mathbf{E}\,X$ for $X \geqslant 0$:

1) $\mathbf{E}\,X$ is well-defined.
2) Linearity: for any $X, Y \in \mathcal{E}_+$ and $a, b \geqslant 0$ are constants. Then $\mathbf{E}[aX + bY] = a\,\mathbf{E}\,X + b\,\mathbf{E}\,Y$.

Proof of (2). Let $X_n \in \mathcal{E}$ where $X_n \uparrow X$, and $Y_n \in \mathcal{E}$ where $Y_n \uparrow Y$. Then

$$aX_n + bY_n \uparrow aX + bY$$

Then

$$\begin{aligned}
\mathbf{E}[aX + bY] &:= \lim_{n\to\infty} \mathbf{E}[aX_n + bY_n] \\
&= \lim_{n\to\infty} (a\,\mathbf{E}\,X_n + b\,\mathbf{E}\,Y_n) \\
&= a \lim_{n\to\infty} \mathbf{E}\,X_n + b \lim_{n\to\infty} \mathbf{E}\,Y_n \\
&= \mathbf{E}[aX] + \mathbf{E}[bY]
\end{aligned}$$

□

3) (Monotone Convergence theorem) If $X_n \geqslant 0$ and $X_n \uparrow X$. Then $\mathbf{E}\,X_n \uparrow \mathbf{E}\,X$, i.e.

$$\lim_{n\to\infty} \mathbf{E}\,X_n = \mathbf{E}\Big[\lim_{n\to\infty} X_n\Big]$$

Proof of (3). If we consider an approximating sequence $X_1, \ldots, X_n, \ldots \uparrow X$. Each one individual X_n has an approximating sequence.

$$\begin{array}{ccccccccc}
X_1 & X_2 & X_3 & \ldots & X_n & \ldots & X_m & \ldots & \uparrow X \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_m^1 & Y_m^2 & Y_m^3 & \ldots & Y_m^n & \ldots & Y_m^m & \ldots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_n^1 & Y_n^2 & Y_n^3 & \ldots & Y_n^n & \ldots & Y_n^m & \ldots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_3^1 & Y_3^2 & Y_3^3 & \ldots & Y_3^n & \ldots & Y_3^m & \ldots & \\
Y_2^1 & Y_2^2 & Y_2^3 & \ldots & Y_2^n & \ldots & Y_2^m & \ldots & \\
Y_1^1 & Y_1^2 & Y_1^3 & \ldots & Y_1^n & \ldots & Y_1^m & \ldots &
\end{array}$$

Define

$$Z_m := \max\{Y_m^i : 1 \leqslant i \leqslant m\}$$

Then we have

- $Z_m \in \mathcal{E}$.
- $Z_m \uparrow$
- $Y_m^n \leqslant Z_m \leqslant X_m$.

Send m to $+\infty$. Then we have

$$X_n \leqslant \lim_{m\to\infty} Z_m \leqslant X$$

Send n to $+\infty$, then we have

$$X \leqslant \lim_{m\to\infty} Z_m \leqslant X$$

So Z_m is an approximating sequence of X. In particular:

$$\mathbf{E}\, X = \lim_{m\to\infty} \mathbf{E}\, Z_m$$

On the other hand, consider the following inequalities

$$\mathbf{E}\, Y_m^n \leqslant \mathbf{E}\, Z_m \leqslant \mathbf{E}\, X_m$$

Send m to $+\infty$, we have

$$\begin{array}{ccccc} \lim_{m\to\infty} \mathbf{E}\, Y_m^n & \leqslant & \lim_{m\to\infty} \mathbf{E}\, Z_m & \leqslant & \lim_{m\to\infty} \mathbf{E}\, X_m \\ \mathbf{E}\, X_n & \leqslant & \mathbf{E}\, X & \leqslant & \lim_{m\to\infty} \mathbf{E}\, X_m \end{array}$$

Send n to ∞, we have

$$\lim_{n\to\infty} \mathbf{E}\, X_n \leqslant \mathbf{E}\, X \leqslant \lim_{m\to\infty} \mathbf{E}\, X_m$$

This completes the proof. □

Remark 5.5 (Remark on MCT). *If X is a $\overline{\mathbb{R}}$-valued RV, such that $X \geqslant 0$. Suppose*

$$\mathbf{P}(\{X = +\infty\}) > 0$$

Then $\mathbf{E}\, X = +\infty$. The reason is, if we considering the approximating sequence $X_n \in \mathcal{E}$ that we constructed for any non-negative RV X, i.e. $X \geqslant 0$ and $X_n \uparrow X$ as $n \to \infty$.

$$X_n = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbf{1}_{A_k} + n \cdot \mathbf{1}_{B_n}$$

where

$$A_k := \left\{ \frac{k-1}{2^n} \leqslant X < \frac{k}{2^n} \right\}$$
$$B_n := \{X \geqslant n\}$$

Then

$$\mathbf{E}\, X_n \geqslant n\, \mathbf{P}(X \geqslant n)$$

Send n to ∞, then

$$\mathbf{E}\, X \overset{MCT}{=} \lim_{n\to\infty} \mathbf{E}\, X_n \geqslant \lim_{n\to\infty} n\, \underbrace{\mathbf{P}(X \geqslant n)}_{>0} = +\infty$$

This means $\mathbf{E}\, X = +\infty$.

Step 3: Define $\mathbf{E}\,X$ for general RV X

Set

$$X^+(\omega) := \max\{X(\omega), 0\}$$
$$X^-(\omega) := -\min\{X(\omega), 0\}$$

[Board figure omitted: the graph of a sign-changing X together with its positive part X^+ (the graph truncated below at 0) and negative part X^- (the reflected lower half)]

X^+ is called positive part of X, and X^- is called negative part of X. Then observe that

$$X(\omega) = X^+(\omega) - X^-(\omega)$$
$$|X(\omega)| = X^+(\omega) + X^-(\omega)$$

Define expectation on a general random variable X as

$$\mathbf{E}\,X := \mathbf{E}\,X^+ - \mathbf{E}\,X^-$$

We say $\mathbf{E}\,X$ is well-defined if at least one of these two numbers $\mathbf{E}\,X^+$ and $\mathbf{E}\,X^-$ is finite. Equivalently, $\mathbf{E}\,X$ is defined provided at least one of $\mathbf{E}\,X^+, \mathbf{E}\,X^- < \infty$ (means at least one of them exists).

Definition 5.6 ($L^1(\mathbf{P})$ space)**.** *Define*

$$L^1(\Omega, \mathscr{F}, \mathbf{P}) = L^1(\mathbf{P}) := \left\{X : \Omega \to \overline{\mathbb{R}} : X \text{ is a } \overline{\mathbb{R}}\text{-valued RV, s.t. } \mathbf{E}|X| < \infty\right\}$$

and

$$\mathbf{E}|X| = \mathbf{E}[X^+ + X^-] = \mathbf{E}\,X^+ + \mathbf{E}\,X^-$$

Properties:

1) If $X \in L^1(\mathbf{P})$, then $\mathbf{P}(X = \pm\infty) = 0$. Because of the Remark of MCT above.

2) If $\mathbf{E}\,X$ is well-defined, and $c \in \mathbb{R}$. Then

$$\mathbf{E}[cX] = c\,\mathbf{E}\,X$$

   This property also implicitly implies that if $\mathbf{E}\,X$ is well-defined, then $\mathbf{E}[cX]$ is well-defined.

Proof of (2). Suppose $c < 0$. Consider the positive part and negative part of cX, i.e.

$$(cX)^+ := (-c)X^-$$
$$(cX)^- := (-c)X^+$$

Then

$$\begin{aligned}
\mathbf{E}[cX] &= \mathbf{E}(cX)^+ - \mathbf{E}(cX)^- \\
&= \mathbf{E}[(-c)X^-] - \mathbf{E}[(-c)X^+] \\
&= (-c)\,\mathbf{E}[X^-] - (-c)\,\mathbf{E}[X^+] \\
&= (-c)(\mathbf{E}[X^-] - \mathbf{E}[X^+]) \\
&= c\,\mathbf{E}\,X
\end{aligned}$$

For $c > 0$, follows from the $X \in \mathcal{E}_+$ case. □

3) If $X \in L^1(\mathbf{P})$ and $Y \in L^1(\mathbf{P})$. Then $aX + bY \in L^1(\mathbf{P})$ for any $a, b \in \mathbb{R}$.

Proof of (3).

$$\begin{aligned}
\mathbf{E}|aX + bY| &\leqslant \mathbf{E}(|a||X| + |b||Y|) \\
&= |a|\,\mathbf{E}|X| + |b|\,\mathbf{E}|Y| \\
&< \infty
\end{aligned}$$

□

Properties of $\mathbf{E}\,X$ for general X

1) If $\mathbf{E}\,X$ is well-defined/exists, and $c \in \mathbb{R}$. Then $\mathbf{E}[cX] = c\,\mathbf{E}\,X$.
2) If $X, Y \in L^1(\mathbf{P}) = \{X : \mathbf{E}|X| < \infty\}$. Then $\mathbf{E}(X + Y) = \mathbf{E}\,X + \mathbf{E}\,Y$.
3) (1) and (2) together implies $\mathbf{E}(aX + bY) = a\,\mathbf{E}\,X + b\,\mathbf{E}\,Y$.
4) If $X, Y \in L^1(\mathbf{P})$ and $Y \geqslant X$. Then $\mathbf{E}\,Y \geqslant \mathbf{E}\,X$. Because $Y - X \geqslant 0$. So $0 \leqslant \mathbf{E}(Y - X) = \mathbf{E}\,Y - \mathbf{E}\,X$.
5) Suppose a sequence of RVs $X_n \in L^1(\mathbf{P})$, and $X_n \uparrow X$ or $X_n \downarrow X$. Here X may or may not in $L^1(\mathbf{P})$. But $\mathbf{E}\,X_n \to \mathbf{E}\,X$.

Proof of Property (2). See that

$$(X + Y) = (X + Y)^+ - (X + Y)^-$$

In addition,

$$(X + Y) = X^+ - X^- + Y^+ - Y^-$$

Equate them and re-arrange,

$$\begin{aligned}
(X + Y) = (X + Y)^+ - (X + Y)^- &= X^+ - X^- + Y^+ - Y^- \\
X^+ + Y^+ + (X + Y)^- &= (X + Y)^+ + X^- + Y^-
\end{aligned}$$

Taking expectation of both sides (which are all exists),

$$\begin{aligned}
\mathbf{E}\,X^+ + \mathbf{E}\,Y^+ + \mathbf{E}(X + Y)^- &= \mathbf{E}(X + Y)^+ + \mathbf{E}\,X^- + \mathbf{E}\,Y^- \\
\underbrace{\mathbf{E}\,X^+ - \mathbf{E}\,X^-}_{=\mathbf{E}\,X} + \underbrace{\mathbf{E}\,Y^+ - \mathbf{E}\,Y^-}_{=\mathbf{E}\,Y} &= \underbrace{\mathbf{E}(X + Y)^+ - \mathbf{E}(X + Y)^-}_{=\mathbf{E}(X+Y)} \\
\mathbf{E}\,X + \mathbf{E}\,Y &= \mathbf{E}(X + Y)
\end{aligned}$$

□

Remark 5.7. *We can move numbers between equalities. But usually we cannot move $\infty, -\infty$ around. In above, we move $\mathbf{E}\, X$ and $\mathbf{E}\, Y$ around because we know $X, Y \in L^1(\mathbf{P})$. So*

$$\mathbf{E}|X| = \mathbf{E}\, X^+ + \mathbf{E}\, X^- < \infty$$

Because $\mathbf{E}\, X^+, \mathbf{E}\, X^- > 0$, thus, $\mathbf{E}\, X^+, \mathbf{E}\, X^- < \infty$.

But we can relax the condition $X, Y \in L^1(\mathbf{P})$ to at least one of X and Y is in $L^1(\mathbf{P})$.

Example 5.2 (Non-example of $X \notin L^1(\mathbf{P})$)**.** *Let $X_n = n$. $X_n \to \infty \notin L^1(\mathbf{P})$.*

Proof of Property (5). WLOG, assume $X_n \uparrow X$. Otherwise, just look at $-X_n$ and $-X$. See that

$$X_n^+ - X_n^- = X_n \uparrow X = X^+ - X^-$$

This is equivalent to

$$X_n^+ \uparrow X^+, \qquad X_n^- \downarrow X^-$$

and

$$0 \leqslant X_n^- \leqslant X_1^-$$

So

$$0 \leqslant X_n + X_1^- \uparrow X + X_1^-$$

But $X_n + X_1^- = X_n^+ - X_n^- + X_1^- \geqslant 0$. So by MCT and

$$\lim_{n\to\infty} \mathbf{E}(X_n + X_1^-) = \mathbf{E}\left(\lim_{n\to\infty} X_n + X_1^-\right) = \mathbf{E}(X + X_1^-)$$

By the remark, we have

$$\mathbf{E}(X + X_1^-) = \mathbf{E}\, X + \mathbf{E}\, X_1^-$$

Hence, we have

$$\mathbf{E}\, X + \mathbf{E}\, X_1^- = \lim_{n\to\infty} \mathbf{E}(X_n + X_1^-) = \lim_{n\to\infty} \mathbf{E}\, X_n + \lim_{n\to\infty} \mathbf{E}\, X_1^- = \lim_{n\to\infty} \mathbf{E}\, X_n + \mathbf{E}\, X_1^-$$

Cancel $\mathbf{E}\, X_1^-$ both sides, we have

$$\mathbf{E}\, X = \lim_{n\to\infty} \mathbf{E}\, X_n$$

This completes the proof. □

Remark 5.8. *If we didn't use the remark,*

$$\begin{aligned} \mathbf{E}(X + X_1^-) &= \mathbf{E}(X^+ \underbrace{-X^- + X_1^-}_{\geqslant 0}) \\ &= \mathbf{E}\, X^+ + \mathbf{E}(X_1^- - X^-) \\ &= \mathbf{E}\, X^+ + \mathbf{E}\, X_1^- - \mathbf{E}\, X^- \\ &= \mathbf{E}\, X + \mathbf{E}\, X_1^- \end{aligned}$$

Lemma 5.9 (Markov Inequality)**.** *If $X \geqslant 0$. Then for any $\lambda > 0$.*

$$\mathbf{P}(X \geqslant \lambda) \leqslant \frac{\mathbf{E}\,X}{\lambda}$$

Proof. See that

$$\begin{aligned}
\mathbf{E}\,\mathbf{1}_{X\geqslant\lambda} &= \int \mathbf{1}_{X\geqslant\lambda} d\,\mathbf{P} \\
&\leqslant \int \frac{X}{\lambda}\mathbf{1}_{X\geqslant\lambda} d\,\mathbf{P} \\
&= \frac{1}{\lambda}\int X\mathbf{1}_{X\geqslant\lambda} d\,\mathbf{P} \\
&\leqslant \frac{1}{\lambda}\int X d\,\mathbf{P} = \frac{1}{\lambda}\mathbf{E}\,X, \qquad \text{using } X \geqslant 0
\end{aligned}$$

Since $\mathbf{E}\,\mathbf{1}_{X\geqslant\lambda} = \mathbf{P}(X \geqslant \lambda)$, this is exactly the claim. $\square$

Remark 5.10. *The meaning of Markov inequality is probability of $X \geqslant 0$ is bigger than a very large number λ is controlled by some number. In particular, it is controlled by a rate of c/λ. This means the probability of X goes to infinity is controlled.*

Corollary 5.11. *If $\mathbf{E}|X|^p < +\infty$ with $p \geqslant 1$.*

$$\mathbf{P}(|X| \geqslant \lambda) \leqslant \frac{\mathbf{E}|X|^p}{\lambda^p}$$

Proof.

$$\mathbf{P}(|X| \geqslant \lambda) = \mathbf{P}(|X|^p \geqslant \lambda^p) \leqslant \frac{\mathbf{E}|X|^p}{\lambda^p}$$

where the last inequality is the Markov inequality applied to the non-negative random variable $|X|^p$ with threshold λ^p. $\square$

Remark 5.12. *The meaning of this corollary is that the probability of higher moments of X, i.e. $|X|^p$ taking larger values is even smaller.*

5.3 Limits and Integrals

Lectures 2024-11-05, 2024-11-07

Theorem 5.13 (Monotone Convergence theorem)**.** ***Version 1:*** *If $X_n \geqslant 0$ and $X_n \uparrow X$. Then $\lim_{n\to\infty} \mathbf{E}\,X_n = \mathbf{E}\lim_{n\to\infty} X_n = \mathbf{E}\,X$.*
Version 2: *If $X_n \in L^1(\mathbf{P})$ and $X_n \uparrow X$ or $X_n \downarrow X$. Then*

$$\lim_{n\to\infty} \mathbf{E}\,X_n = \mathbf{E}\lim_{n\to\infty} X_n = \mathbf{E}\,X$$

Corollary 5.14. *If ξ_i are non-negative RVs. Then*

$$\mathbf{E}\sum_{i=1}^{\infty} \xi_i = \sum_{i=1}^{\infty} \mathbf{E}\,\xi_i$$

Proof. Set $X_n = \sum_{i=1}^n \xi_i$. Then X_n satisfies the assumption in MCT version 1. So

$$\sum_{i=1}^{\infty} \mathbf{E}\,\xi_i = \lim_{n\to\infty} \mathbf{E}\,X_n = \mathbf{E}\lim_{n\to\infty} X_n = \mathbf{E}\sum_{i=1}^{\infty}\xi_i$$

□

Remark 5.15. *In Version (1), one can relax the condition $X_n \geqslant 0$ to $X_n \geqslant Z \in L^1(\mathbf{P})$ because otherwise consider $X_n - Z \geqslant 0$ and apply version (1).*

Lemma 5.16 (Fatou's lemma)**.** *Suppose $X_n \geqslant 0$. Then*

$$\mathbf{E}\left[\liminf_{n\to\infty} X_n\right] \leqslant \liminf_{n\to\infty} \mathbf{E}\,X_n$$

Proof. Start from LHS.

$$\begin{aligned}
\mathbf{E}\left[\liminf_{n\to\infty} X_n\right] &= \mathbf{E}\Big[\lim_{n\to\infty} \underbrace{\inf_{k\geqslant n} X_k}_{=:Y_n\uparrow}\Big] \\
&= \mathbf{E}\left[\lim_{n\to\infty} Y_n\right] \\
&= \lim_{n\to\infty} \mathbf{E}\,Y_n, \qquad \text{by MCT} \\
&= \lim_{n\to\infty} \mathbf{E}\Big[\underbrace{\inf_{k\geqslant n} X_k}_{\leqslant X_n}\Big] \\
&\leqslant \liminf_{n\to\infty} \mathbf{E}\,X_n
\end{aligned}$$

□

Remark 5.17. *In Fatou's lemma, we can relax the condition $X_n \geqslant 0$ to the condition $X_n \geqslant Z \in L^1(\mathbf{P})$. Because otherwise consider $X_n - Z \geqslant 0$ and apply previous version of Fatou's lemma.*

Corollary 5.18. *If $X_n \leqslant Z \in L^1(\mathbf{P})$. Then*

$$\mathbf{E}\left[\limsup_{n\to\infty} X_n\right] \geqslant \limsup_{n\to\infty} \mathbf{E}\,X_n$$

Proof. Consider $Z - X_n \geqslant 0$. Apply Fatou's lemma. So we have

$$\mathbf{E}\left[\liminf_{n\to\infty}(Z - X_n)\right] \leqslant \liminf_{n\to\infty} \mathbf{E}(Z - X_n)$$

Then the LHS becomes

$$\mathbf{E}\left(Z + \liminf_{n\to\infty}(-X_n)\right) = \mathbf{E}\left(Z - \limsup_{n\to\infty} X_n\right) = \mathbf{E}\,Z - \mathbf{E}\left(\limsup_{n\to\infty} X_n\right)$$

The RHS becomes

$$\liminf_{n\to\infty}(\mathbf{E}\,Z - \mathbf{E}\,X_n) = \mathbf{E}\,Z - \limsup_{n\to\infty} \mathbf{E}\,X_n$$

Hence, we have the desired result. □

Theorem 5.19 (Dominated Convergence Theorem)**.** *Suppose X_n is a sequence of random variables such that*

- $|X_n| \leqslant Y \in L^1(\mathbf{P})$.
- $X_n \to X$ *pointwise for almost every* $\omega \in \Omega$.

Then

$$\lim_{n\to\infty} \mathbf{E}\, X_n = \mathbf{E}\, X = \mathbf{E} \lim_{n\to\infty} X_n$$

Proof. From $|X_n| \leqslant Y$, we have $-Y \leqslant X_n \leqslant Y$. So $Y - X_n \geqslant 0$. By Fatou's lemma and its dual, we have

$$\mathbf{E} \liminf_{n\to\infty} X_n \leqslant \liminf_{n\to\infty} \mathbf{E}\, X_n \leqslant \limsup_{n\to\infty} \mathbf{E}\, X_n \leqslant \mathbf{E} \limsup_{n\to\infty} X_n$$

But we assumed $X_n \to X$ a.e., so

$$X = \liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$$

Hence,

$$\mathbf{E}\, X = \mathbf{E} \liminf_{n\to\infty} X_n = \mathbf{E} \limsup_{n\to\infty} X_n$$

Then by sandwich theorem, we have

$$\mathbf{E} \liminf_{n\to\infty} X_n = \liminf_{n\to\infty} \mathbf{E}\, X_n = \limsup_{n\to\infty} \mathbf{E}\, X_n = \mathbf{E} \limsup_{n\to\infty} X_n$$

□

Remark 5.20. *In DCT, $X \in L^1(\mathbf{P})$. Because $\mathbf{E}\, X = \mathbf{E} \lim_{n\to\infty} |X_n|$. Then by Fatou's lemma, we have*

$$\mathbf{E}|X| = \mathbf{E} \lim_{n\to\infty} |X_n| = \mathbf{E} \liminf_{n\to\infty} |X_n| \leqslant \liminf_{n\to\infty} \mathbf{E}|X_n| \leqslant \mathbf{E}|Y| < \infty$$

Example 5.3. *Consider $([0,1], \mathscr{B}([0,1]), \mathbf{P})$ where $\mathbf{P}$ is the Lebesgue measure. Suppose $X_n = n^2 \mathbf{1}_{(0,1/n]}$. Then $\lim_{n\to\infty} X_n = 0$, and $\mathbf{E}\, X_n = n^2 \frac{1}{n} = n$. So*

$$\mathbf{E} \lim_{n\to\infty} X_n = \mathbf{E}\, 0 = 0 < \infty = \lim_{n\to\infty} n = \lim_{n\to\infty} \mathbf{E}\, X_n$$

So the conclusion of DCT fails for this sequence; the point is that no dominating $Z \in L^1(\mathbf{P})$ exists here, so the domination hypothesis in DCT cannot be dropped.

5.4 Indefinite Integrals

Lectures 2024-11-07, 2024-11-09

Consider a probability space $(\Omega, \mathscr{B}, \mathbf{P})$. Let X be a RV. For any $A \in \mathscr{B}$, we define

$$\int_A X \, d\,\mathbf{P} := \int X \cdot \mathbf{1}_A \, d\,\mathbf{P} = \mathbf{E}(X \mathbf{1}_A)$$

Definition 5.21 (Almost surely). *We say a property holds almost surely (a.s.) if there is a null set $N \in \mathscr{B}$ where $\mathbf{P}(N) = 0$ such that the property holds for all $\omega \in N^c$.*

Example 5.4. *Consider the probability space $([0,1], \mathscr{B}([0,1]), \mathbf{P})$ where $\mathbf{P}$ is the Lebesgue measure. Define*

$$X := 0$$
$$Y := \mathbf{1}_{\mathbb{Q}\cap[0,1]}$$

Then

$$X = Y \text{ a.s. }, \qquad N = \mathbb{Q} \cap [0,1]$$

This means for all $\omega \in N^c$, $X(\omega) = Y(\omega)$.

Example 5.5. *Still consider $X_n = n^2 \mathbf{1}_{(0,1/n]}$. Then $N = \{0\}$. Then $X_n \to 0$ almost surely, because $X_n(\omega) \to 0$ for all $\omega \notin N$.*

$$X_n(\omega) \to 0 \qquad \text{for all } \omega \neq 0$$

And $X_n(0) = n^2 \uparrow \infty$.

Example 5.6. *Recall we had X_n IID unit exponential RVs, where*

$$\mathbf{P}(X > x) = e^{-x}$$

Then

$$\mathbf{P}\left(\limsup \frac{X_n}{\log n} = 1\right) = 1 \iff \limsup_{n\to\infty} \frac{X_n}{\log n} = 1 \text{ a.s.}$$

where $N = \{\omega : \limsup_{n\to\infty} \frac{X_n}{\log n} \neq 1\}$.

Example 5.7. *Let $A \in \mathscr{B}(\mathbb{R})$. Then using*

$$X^{-1}(A) = (X^{-1}(A) \cap N) \cup (X^{-1}(A) \cap N^c)$$

we have

$$\begin{aligned}
\mathbf{P}(X \in A) &= \mathbf{P}(\{\omega : X(\omega) \in A\}) \\
&= \mathbf{P}(\{\omega : X(\omega) \in A\} \cap N^c) + \underbrace{\mathbf{P}(\{\omega : X(\omega) \in A\} \cap N)}_{\leqslant \mathbf{P}(N)=0} \\
&= \mathbf{P}(\{\omega : X(\omega) \in A\} \cap N^c) \\
&= \mathbf{P}(\{\omega : Y(\omega) \in A\} \cap N^c) \\
&= \mathbf{P}(Y \in A)
\end{aligned}$$

If $X = Y$ almost surely, we have

$$\mathbf{P}(X \in A) = \mathbf{P}(Y \in A)$$

Properties of indefinite integral

Suppose $X \geqslant 0$.

1) $\int_A X d\mathbf{P} = 0$ if and only if $\mathbf{P}(\{X > 0\} \cap A) = 0$, that is, $X = 0$ almost surely on A. Here, the null set is $N = \{X > 0\} \cap A$.

2) Let $X \geqslant 0$. Let $A_n \in \mathscr{B}$ be disjoint subsets of Ω. Then
$$\int_{\cup_{n=1}^{\infty} A_n} X d\mathbf{P} = \sum_{n=1}^{\infty} \int_{A_n} X d\mathbf{P}$$

3) Let $X \geqslant 0$, and $A_n \uparrow A$. Then by MCT, we have
$$\int_{A_n} X d\mathbf{P} \longrightarrow \int_A X d\mathbf{P}$$

4) Let $X \in L^1(\mathbf{P})$, and $A_n \uparrow A$. Then by DCT, we have
$$\int_{A_n} X d\mathbf{P} \longrightarrow \int_A X d\mathbf{P}$$

Proof of (1). $\Longleftarrow$**:** Suppose $X = 0$ almost surely on A, i.e. $P(\{X > 0\} \cap A) = 0$. Then
$$\int_A X d\mathbf{P} = \int_\Omega X \mathbf{1}_A d\mathbf{P}$$
Let X_n be an approximating sequence of X, i.e.
$$X_n = \sum_{k=1}^{n \cdot 2^n} \frac{k}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} < X \leqslant \frac{k+1}{2^n}\right\}} + n\mathbf{1}_{\{X > n\}} \uparrow X$$

Then
$$X_n \cdot \mathbf{1}_A = \sum_{k=1}^{n \cdot 2^n} \left(\frac{k}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} < X \leqslant \frac{k+1}{2^n}\right\}} + n\mathbf{1}_{\{X > n\}}\right) \mathbf{1}_A = \sum_{k=1}^{n \cdot 2^n} \frac{k}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} < X \leqslant \frac{k+1}{2^n}\right\} \cap A} + n\mathbf{1}_{\{X > n\} \cap A} \geqslant 0$$

is an approximating sequence of $X \cdot \mathbf{1}_A$. So by the definition of expectation of non-negative random variable, we have $\mathbf{E}\, X_n \cdot \mathbf{1}_A \uparrow \mathbf{E}\, X \cdot \mathbf{1}_A$. On the other hand,
$$\mathbf{E}\, X_n \cdot \mathbf{1}_A = \sum_{k=1}^{n \cdot 2^n} \frac{k}{2^n} \underbrace{\mathbf{P}\left(\left\{\frac{k}{2^n} < X \leqslant \frac{k+1}{2^n}\right\} \cap A\right)}_{=0} + n \underbrace{\mathbf{P}(\{X > n\} \cap A)}_{=0} = 0$$

Here
$$\mathbf{P}\left(\left\{\frac{k}{2^n} < X \leqslant \frac{k+1}{2^n}\right\} \cap A\right) = 0, \qquad \mathbf{P}(\{X > n\} \cap A) = 0$$
because we assumed $X = 0$ a.s. on A, so X cannot take nonzero values on A.

Because we have $\mathbf{E}\, X_n \cdot \mathbf{1}_A \uparrow \mathbf{E}\, X \cdot \mathbf{1}_A$, thus $0 \uparrow \mathbf{E}\, X \cdot \mathbf{1}_A$. Hence,

$$\mathbf{E}\, X \cdot \mathbf{1}_A = 0$$

$\Longrightarrow$**:** Now suppose $\int_A X d\,\mathbf{P} = 0$. Want to show $\mathbf{P}(\{X > 0\} \cap A) = 0$.
Proof by contradiction. Suppose $\mathbf{P}(\{X > 0\} \cap A) > 0$. Look at

$$\mathbf{P}(\underbrace{\{X > \frac{1}{n}\} \cap A}_{=:B_n})$$

Then we have

$$\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n = \{X > 0\} \cap A$$

By continuity of probability, we have

$$\lim_{n\to\infty} \mathbf{P}\Bigg(\left\{X > \frac{1}{n}\right\} \cap A\Bigg) = \mathbf{P}\Bigg(\{X > 0\} \cap A\Bigg) > 0$$

Hence, there exists some n_0 such that

$$\mathbf{P}\Bigg(\left\{X > \frac{1}{n_0}\right\} \cap A\Bigg) > 0$$

Then

$$\begin{aligned}
\int_A X d\,\mathbf{P} &= \int X \cdot \mathbf{1}_A d\,\mathbf{P} \\
&\geqslant \int X \cdot \mathbf{1}_{A\cap\left\{X > \frac{1}{n_0}\right\}} dP \\
&\geqslant \int \frac{1}{n_0} \cdot \mathbf{1}_{A\cap\left\{X > \frac{1}{n_0}\right\}} d\,\mathbf{P} \\
&= \frac{1}{n_0} \mathbf{P}\Bigg(A \cap \left\{X > \frac{1}{n_0}\right\}\Bigg) \\
&> 0
\end{aligned}$$

Hence, a contradiction. $\square$

Remark 5.22. *If $X = 0$ almost surely on A. Then $\int_A X = 0$ because $X = 0$ almost surely on A. This implies*

$$\begin{aligned}
X^+ &= 0, \qquad && a.s. \text{ on } A \\
X^- &= 0, \qquad && a.s. \text{ on } A
\end{aligned}$$

Then

$$\int_A X d\,\mathbf{P} = \int_A X^+ d\,\mathbf{P} - \int_A X^- d\,\mathbf{P} = 0 - 0 = 0$$

Remark 5.23. *If $X = Y$ almost surely on A. Then*

$$\int_A X d\mathbf{P} = \int_A Y d\mathbf{P}$$

Because $X - Y = 0$ almost surely on A. Then

$$0 = \int_A X d\mathbf{P} - \int_A Y d\mathbf{P} = \int_A (X - Y) d\mathbf{P}$$

Proof of (2). By definition and MCT,

$$LHS = \int X \mathbf{1}_{\cup_{n=1}^{\infty} A_n} d\mathbf{P} = \int X \cdot \sum_{n=1}^{\infty} \mathbf{1}_{A_n} d\mathbf{P} \overset{MCT}{=} \sum_{n=1}^{\infty} \int X \cdot \mathbf{1}_{A_n} d\mathbf{P} = \sum_{n=1}^{\infty} \int_{A_n} X \cdot d\mathbf{P} = RHS$$

□

Proof of (4). If $X \in L^1(\mathbf{P})$. Then $|X \cdot \mathbf{1}_{A_n}| \leqslant |X| \in L^1(\mathbf{P})$. We have $X \cdot \mathbf{1}_{A_m} \to X \cdot \mathbf{1}_A$. Hence, by DCT, we have

$$LHS = \int X \cdot \mathbf{1}_{A_n} d\mathbf{P} \xrightarrow{DCT} \int X \cdot \mathbf{1}_A d\mathbf{P}$$

□

5.5 The Transformation Theorem and Densities

Lecture 2024-11-12

Setup

Consider a measurable map

$$T : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\Omega', \mathscr{B}')$$

So we denote $T \in \mathscr{B}/\mathscr{B}'$. So the image space $(\Omega', \mathscr{B}')$ becomes a probability space $(\Omega', \mathscr{B}', \mathbf{P}')$ where $\mathbf{P}' = \mathbf{P} \circ T^{-1}$. Then consider another random variable

$$X' : (\Omega', \mathscr{B}', \mathbf{P}') \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

Overall, we are considering the composition

$$X : (\Omega, \mathscr{B}, \mathbf{P}) \xrightarrow{\quad T \quad} (\Omega', \mathscr{B}', \mathbf{P}') \xrightarrow{\quad X' \quad} (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

So we are considering the composition $X = X' \circ T$.

Theorem 5.24. *1) We have $X \geqslant 0 \iff X' \geqslant 0$, and*

$$\int X d\mathbf{P} = \int X' d\mathbf{P}'$$

2) We have $X \in L^1(\mathbf{P}) \iff X' \in L^1(\mathbf{P}')$, and

$$\int X d\,\mathbf{P} = \int X' d\,\mathbf{P}'$$

Proof. **Step 1: Suppose $X' = \mathbf{1}_{A'}$ for some $A' \in \mathscr{B}(\mathbb{R})$.** Observe that

$$X(\omega) = X' \circ T(\omega) = \mathbf{1}_{A'}(T(\omega)) = \mathbf{1}_{T^{-1}(A')}(\omega)$$

Starting from the LHS is

$$LHS = \int X d\,\mathbf{P} = \int \mathbf{1}_{T^{-1}(A')} d\,\mathbf{P} = \mathbf{P}(T^{-1}(A')) = \mathbf{P}'(A')$$

And the RHS is

$$RHS = \int X' d\,\mathbf{P}' = \int \mathbf{1}_{A'} d\,\mathbf{P}' = \mathbf{P}'(A')$$

So we have LHS equals to RHS.

The key point here is, $\mathbf{P}'$ is not an arbitrary measure. It is the induced measure by a measurable function T.

Step 2: Suppose X' is a simple RV. Since integration/expectation is linear, so if the identity holds for indicator function, it holds for simple RVs.

Step 3: Suppose X' is a non-negative RV. Suppose $X' \geqslant 0$, then $X \geqslant 0$. We can find a sequence of simple random variables X'_n such that $0 \leqslant X'_n \uparrow X'$. Hence, $0 \leqslant X_n = X'_n \circ T$ is an approximating sequence of $X = X' \circ T$, i.e. $X_n \uparrow X$. By step (2), we have

$$\int X_n d\,\mathbf{P} = \int X'_n d\,\mathbf{P}'$$

Send n to ∞. By MCT, we have

$$\int X d\,\mathbf{P} = \int X' d\,\mathbf{P}'$$

This finishes the proof for (1).

For (2), note $X = X^+ - X^-$. □

Example 5.8. *Suppose $X \sim N(0,1)$ is a standard Gaussian. We try to compute*

$$\mathbf{E}\, f(X) = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

for some measurable function $f : \mathbb{R} \to \mathbb{R}$. Use the language of STAT 501, a standard Gaussian RV X is defined as a measurable function

$$X : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

such that the induced measure $\mathbf{P}' = \mathbf{P} \circ X^{-1}$ is given by

$$\mathbf{P}'(X \in dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Or one can say applying measure $\mathbf{P}'$ on the set $(-\infty, x]$, it will be defined as

$$\mathbf{P}'((-\infty, x]) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Observe that

$$\mathbf{P}'((-\infty, x]) = \mathbf{E}' \mathbf{1}_{(-\infty,x]} = \int_{-\infty}^{\infty} \mathbf{1}_{(-\infty,x]}(z) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

So you see that the expectation holds for $f(z) = \mathbf{1}_{(-\infty,x]}(z)$. Hence, it will holds for general measurable function $f \in \mathscr{B}(\mathbb{R})/\mathscr{B}(\mathbb{R})$. So in general, we have

$$\mathbf{E}' f(X) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

This equal sign comes from the definition of what we mean by an abstract RV has a given distribution.

Lastly, we use the theorem above. So the composition we are dealing with is

$$(\Omega, \mathscr{B}, \mathbf{P}) \xrightarrow{X} (\mathbb{R}, \mathscr{B}(\mathbb{R})) \xrightarrow{f} (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

But we are actually considering

$$(\Omega, \mathscr{B}, \mathbf{P}) \xrightarrow{T=X} (\Omega', \mathscr{B}', \mathbf{P}') \xrightarrow{X'=f} (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

where $\Omega' = \mathbb{R}$, $\mathscr{B}' = \mathscr{B}(\mathbb{R})$ and $\mathbf{P}' = \mathbf{P} \circ X^{-1}$. The transformation is $T = X$ and $X' = f$, so that $f(X) = X' \circ T$. And we have

$$\mathbf{E} f(X) = \mathbf{E}' f = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Remark 5.25 (Expectation)**.** *If we are computing $\mathbf{E} X$. Then we are considering the following composition*

$$(\Omega, \mathscr{B}, \mathbf{P}) \xrightarrow{X} (\mathbb{R}, \mathscr{B}(\mathbb{R})) \xrightarrow{\text{id}} (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

Using the previous framework of $X = X' \circ T$, we have $T = X$ and $X' = \text{id}$. Hence, this expectation becomes

$$\mathbf{E} X = \mathbf{E}' \text{id} = \int_{-\infty}^{+\infty} \text{id}(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

5.6 The Riemann vs Lebesgue Integral

Lectures 2024-11-12, 2024-11-14

Consider the measure space $([0,1], \mathscr{B}([0,1]), \mathbf{P})$ where $\mathbf{P}$ is the Lebesgue measure.
Suppose f is a bounded continuous function on $[0,1]$.
A Riemann integral is defined as $\int_0^1 f(x)dx$.

[Board figure omitted: the region under the graph of f over $[0,1]$ sliced by vertical strips over a partition of the domain]

Here we divide the domain into rectangles.

A Lebesgue integral is defined as $\int f d\mathbf{P}$. Here we divide the range into rectangles. (Dividing the range of RVs means to construct an approximating sequence).

Riemann integration is like when counting total values of coins in a bag, you pick each coin one by one and read the values, and add them together.

Lebesgue integration is like pour the coin out, and separate different types of coin (5 cents, 50 cent, 1 dollar, etc.). Count each pack of them individually, and finally add them all up.

Example 5.9. *Consider $X = \mathbf{1}_{\mathbb{Q}\cap[0,1]}$. We can define the Lebesgue integration as*

$$\int X d\mathbf{P} = \mathbf{P}(Q \cap [0,1])$$

However, the Riemann integral of $X = \mathbf{1}_{\mathbb{Q}\cap[0,1]}$ does not exist. Because the Riemann is defined as limit of Riemann sum

$$\sum_{i=1}^{n} f(x_i)\frac{1}{n} \longrightarrow \int_0^1 f(x)dx$$

The choice of points does not matter. But here, the choice of end points does matter.

Theorem 5.26. *If $f(x)$ is a continuous function, then the Lebesgue integral and Riemann integral are the same.*

Proof. Consider the probability space $([0,1], \mathscr{B}([0,1]), \mathbf{P})$ where $\mathbf{P}$ is the Lebesgue measure. Chop $[0,1]$ into finitely many intervals $I_1, \ldots, I_n$ each with length $\frac{1}{n}$. And $[0,1] = \bigcup_{i=1}^{n} I_i$.

[Board figure omitted: the interval $[0,1]$ chopped into n equal subintervals $I_1, \ldots, I_n$, with the graph of f and its sup and inf over one subinterval marked]

Construction of Riemann integral. On an arbitrary interval I_i, define

$$\overline{f_i} = \sup_{x\in I_i} f(x), \qquad \underline{f_i} := \inf_{x\in I_i} f(x)$$

Define the upper Riemann sum as

$$\overline{\sigma}_n := \sum_{i=1}^{n} \overline{f_i}\frac{1}{n}$$

and the lower Riemann sum as

$$\underline{\sigma}_n := \sum_{i=1}^{n} \underline{f_i}\frac{1}{n}$$

Recall that we say $f(x)$ is Riemann integrable if and only if

$$\lim_{n\to\infty} \overline{\sigma}_n = \lim_{n\to\infty} \underline{\sigma}_n = \int_0^1 f(x)dx$$

where $\int_0^1 f(x)dx$ is the Riemann integral.

Construction of Lebesgue integral. On the same partition of intervals, define

$$\overline{\mathcal{F}_n}(x) = \sum_{i=1}^{n} \overline{f_i}\mathbf{1}_{I_i}(x)$$

and

$$\underline{\mathcal{F}_n}(x) = \sum_{i=1}^{n} \underline{f_i}\mathbf{1}_{I_i}(x)$$

Then

$$\underline{\mathcal{F}_n}(x) \leqslant f(x) \leqslant \overline{\mathcal{F}_n}(x)$$

Taking the Lebesgue integral, see that

$$\int_{[0,1]} \underline{\mathcal{F}_n}(x)d\,\mathbf{P} = \sum_{i=1}^{n} \underline{f_i}\frac{1}{n} = \underline{\sigma}_n, \qquad \int_{[0,1]} \overline{\mathcal{F}_n}(x)d\,\mathbf{P} = \sum_{i=1}^{n} \overline{f_i}\frac{1}{n} = \overline{\sigma}_n$$

Hence, we have

$$\underline{\sigma}_n = \int_{[0,1]} \underline{F_n}(x)d\,\mathbf{P} \leqslant \int_{[0,1]} f(x)d\,\mathbf{P} \leqslant \int_{[0,1]} \overline{\mathcal{F}_n}(x)d\,\mathbf{P} = \overline{\sigma}_n$$

Now, send $n \to \infty$, then we have

$$\int_0^1 f(x)dx \leqslant \int_{[0,1]} f(x)d\,\mathbf{P} \leqslant \int_0^1 f(x)dx$$

□

Remark 5.27. *1) In general, for a function defined in any finite interval $[a,b]$, if it is Riemann integrable and measurable. Then it is also Lebesgue integrable and the two integrals are the same.*

2) Lebesgue integral on $[a,b]$. So we are considering the probability space $([a,b], \mathscr{B}([a,b]), \frac{\lambda}{b-a})$ instead, where $\mathbf{P} = \frac{\lambda}{b-a}$ and λ is the Lebesgue measure on $\mathscr{B}([a,b])$. We can do

$$(b-a)\int_{[a,b]} f(x)d\,\mathbf{P} = (b-a)\int_{[a,b]} f(x)\frac{\lambda}{b-a} = \int_{[a,b]} f(x)d\lambda$$

5.7 Product Spaces, Independence, Fubini Theorem

Lecture 2024-11-14

Definition 5.28 (Product Space). *Suppose we have two measurable spaces $(\Omega_1, \mathscr{B}_1)$ and $(\Omega_2, \mathscr{B}_2)$. Now we can define the product of sample spaces to be*

$$\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$$

[Board figure omitted: the rectangle $\Omega_1 \times \Omega_2$ drawn with Ω_1 on the horizontal axis and Ω_2 on the vertical axis, and a point (ω_1, ω_2) inside]

Projection map

On this product space, there is a natural projection map

$$\begin{array}{rccc} \pi_i: & \Omega_1 \times \Omega_2 & \longrightarrow & \Omega_i \\ & (\omega_1, \omega_2) & \longmapsto & \omega_i \end{array}$$

where $i = 1, 2$. For a set $A \subset \Omega_1 \times \Omega_2$, we can define a so-called "section" of this set A. If we fix a $\omega_1 \in \Omega_1$, define the ω_1-section of A as

$$A_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$$

Similarly, if we fix a $\omega_2 \in \Omega_2$, define the ω_2-section of A as

$$A_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$$

Properties:

1) If $A \subset \Omega_1 \times \Omega_2$. The ω_i-section is compatible with the complement

$$(A^c)_{\omega_i} = (A_{\omega_i})^c$$

2) $A_\alpha \subset \Omega_1 \times \Omega_2$ where $\alpha \in T$, and T is some index set. Then taking ω_i-section is compatible with taking arbitrary union and arbitrary intersection

$$\left(\bigcup_{\alpha \in T} A_\alpha\right)_{\omega_i} = \bigcup_{\alpha \in T} (A_\alpha)_{\omega_i}$$
$$\left(\bigcap_{\alpha \in T} A_\alpha\right)_{\omega_i} = \bigcap_{\alpha \in T} (A_\alpha)_{\omega_i}$$

Definition 5.29 (ω_i-section of a measurable map on the Product space)**.** *Consider a measurable map*

$$\begin{array}{rccc} X: & \Omega_1 \times \Omega_2 & \longrightarrow & \mathbb{R}^n \\ & (\omega_1, \omega_2) & \longmapsto & X(\omega_1, \omega_2) \end{array}$$

Fix a $\omega_1 \in \Omega_1$. Look at

$$X_{\omega_1}(\omega_2) := X(\omega_1, \omega_2)$$

The map X_{ω_1} is called the ω_1 section of X.

Similarly, we can define ω_2 section of X as

$$X_{\omega_2}(\omega_1) = X(\omega_1, \omega_2)$$

Properties: Let $A \subset \Omega_1 \times \Omega_2$.

1) $(\mathbf{1}_A)_{\omega_i} = \mathbf{1}_{A_{\omega_i}}$

2) $(X_1 + X_2)_{\omega_i} = (X_1)_{\omega_i} + (X_2)_{\omega_i}$

3) If $X_n \to X$ for all $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$. Then $(X_n)_{\omega_i} \to X_{\omega_i}$.

Definition 5.30 (Rectangle). *A rectangle in $\Omega_1 \times \Omega_2$ is a subset of the form $A_1 \times A_2$ where $A_1 \subset \Omega_1$ and $A_2 \subset \Omega_2$. We say $A_1 \times A_2$ is a measurable rectangle if $A_1 \in \mathscr{B}_1$ and $A_2 \in \mathscr{B}_2$.*

[Board figure omitted: a measurable rectangle $A_1 \times A_2$ drawn as a box inside $\Omega_1 \times \Omega_2$, sitting over A_1 on the horizontal axis and A_2 on the vertical axis]

Lemma 5.31. *Denote $RECT$ as the collection of measurable rectangles. Then $RECT$ is a semi-algebra.*

Proof. 1) $\varnothing, \Omega_1 \times \Omega_2 \in RECT$. Here, $\varnothing = \varnothing_1 \times \varnothing_2$.

2) If we have $A_1 \times A_2 \in RECT$ and $B_1 \times B_2 \in RECT$. Then $(A_1 \times A_2) \cap (B_1 \times B_2) \in RECT$. Because

$$(A_1 \times A_2) \cap (B_1 \times B_2) = \underbrace{(A_1 \cap B_1)}_{\in \mathscr{B}_1} \times \underbrace{(A_2 \cap B_2)}_{\in \mathscr{B}_2} \in RECT$$

3) If $A_1 \times A_2 \in RECT$. Then $(A_1 \times A_2)^c$ is a finite disjoint union of members of $RECT$. Observe that

$$(A_1 \times A_2)^c = (A_1^c \times A_2) \cup (A_1 \times A_2^c) \cup (A_1^c \times A_2^c)$$

The three rectangles on the right are pairwise disjoint, so $(A_1 \times A_2)^c$ is exhibited as a finite disjoint union of members of $RECT$ — which is exactly what axiom 3) of a semi-algebra requires.

□

Definition 5.32. *Consider two measurable spaces $(\Omega_1, \mathscr{B}_1)$ and $(\Omega_2, \mathscr{B}_2)$. Define the product σ-algebra as*

$$\mathscr{B}_1 \otimes \mathscr{B}_2 = \sigma(RECT)$$

This is indeed a σ-algebra over $\Omega_1 \times \Omega_2$. So we have a product space $(\Omega_1 \times \Omega_2, \mathscr{B}_1 \otimes \mathscr{B}_2)$.

5.8 Probability Measures on Product Spaces

Lectures 2024-11-16, 2024-11-19

The setting is, given two measurable space $(\Omega_1, \mathscr{B}_1)$ and $(\Omega_2, \mathscr{B}_2)$. Then the product space is

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

Then the product σ-algebra is defined as

$$\mathscr{B}_1 \otimes \mathscr{B}_2 = \sigma(RECT)$$

where $RECT$ is the collection of measurable rectangles. So we have a product measurable space $(\Omega_1 \times \Omega_2, \mathscr{B}_1 \otimes \mathscr{B}_2)$.
Goal. Want to construct the so-called product measure on the product measurable space.

Definition 5.33. *Define a set function* $\mathbf{P}_1 \times \mathbf{P}_2$ *on* $RECT$ *by*

$$(\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times A_2) = \mathbf{P}_1(A_1)\,\mathbf{P}_2(A_2)$$

[Board figure omitted: the rectangle $A_1 \times A_2$ inside $\Omega_1 \times \Omega_2$, with its measure read off as the product of the side measures $\mathbf{P}_1(A_1)$ and $\mathbf{P}_2(A_2)$]

Recall that $RECT$ is a semi-algebra.
Strategy. If we can show $\mathbf{P}_1 \times \mathbf{P}_2$ is a probability measure on the semi-algebra $RECT$, then we can use the two extension theorems to extend $\mathbf{P}_1 \times \mathbf{P}_2$ to a probability measure (still denoted as $\mathbf{P}_1 \times \mathbf{P}_2$), on $\sigma(RECT) = \mathscr{B}_1 \otimes \mathscr{B}_2$.
Show σ-additivity. Now we show $\mathbf{P}_1 \times \mathbf{P}_2$ is a probability measure on $RECT$. The only non-trivial part is σ-additivity. That is, if $A_1 \times A_2$ is a measurable rectangle, and we have a sequence of disjoint measurable rectangles $A_1^n \times A_2^n$ such that

$$A_1 \times A_2 = \bigcup_{n=1}^{\infty} A_1^n \times A_2^n$$

Then

$$(\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times A_2) = \sum_{n=1}^{\infty} (\mathbf{P}_1 \times \mathbf{P}_2)(A_1^n \times A_2^n)$$

Then the LHS is

$$\begin{aligned}
LHS &= (\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times A_2) = \mathbf{P}_1(A_1)\,\mathbf{P}_2(A_2) \\
&= \left(\int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) d\,\mathbf{P}_1\right) \cdot \left(\int_{\Omega_2} \mathbf{1}_{A_2}(\omega_2) d\,\mathbf{P}_2\right) \\
&= \left(\int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) d\,\mathbf{P}_1\right) \mathbf{1}_{A_2}(\omega_2) d\,\mathbf{P}_2\right) = \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) \mathbf{1}_{A_2}(\omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1 \times A_2}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{\cup_{n=1}^{\infty}(A_1^n \times A_2^n)}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} \sum_{n=1}^{\infty} \mathbf{1}_{A_1^n \times A_2^n}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 = \int_{\Omega_2} \left(\int_{\Omega_1} \lim_{N \to \infty} \sum_{n=1}^{N} \mathbf{1}_{A_1^n \times A_2^n}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&\overset{MCT \times 2}{=} \lim_{N \to \infty} \int_{\Omega_2} \left(\int_{\Omega_1} \sum_{n=1}^{N} \mathbf{1}_{A_1^n \times A_2^n}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&= \lim_{N \to \infty} \sum_{n=1}^{N} \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1^n \times A_2^n}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 = \sum_{n=1}^{\infty} \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1^n \times A_2^n}(\omega_1, \omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 \\
&= \sum_{n=1}^{\infty} \int_{\Omega_2} \left(\int_{\Omega_1} \mathbf{1}_{A_1^n}(\omega_1) \mathbf{1}_{A_2^n}(\omega_2) d\,\mathbf{P}_1\right) d\,\mathbf{P}_2 = \sum_{n=1}^{\infty} \left(\int_{\Omega_1} \mathbf{1}_{A_1^n}(\omega_1) d\,\mathbf{P}_1\right) \cdot \left(\int_{\Omega_2} \mathbf{1}_{A_2^n}(\omega_2) d\,\mathbf{P}_2\right) \\
&= \sum_{n=1}^{\infty} \mathbf{P}_1(A_1^n)\,\mathbf{P}_2(A_2^n) = \sum_{n=1}^{\infty} (\mathbf{P}_1 \times \mathbf{P}_2)(A_1^n \times A_2^n)
\end{aligned}$$

Example 5.10. *Consider* $(\Omega_1, \mathscr{B}_1) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *and* $(\Omega_2, \mathscr{B}_2) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. *Then the product space is* $(\mathbb{R} \times \mathbb{R}, \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}))$. *The claim is, this is same as* $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$, *i.e.*

$$\mathbb{R} \times \mathbb{R} \cong \mathbb{R}^2, \qquad \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^2)$$

Because $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) = \sigma(RECT)$. *But* $RECT \subset \mathscr{B}(\mathbb{R}^2)$. *So*

$$\mathscr{B}(\mathbb{R}^2) = \sigma(\{I_1 \times I_2 : I_1, I_2 \textit{ are open intervals of } \mathbb{R}\})$$

So $\mathscr{B}(\mathbb{R}^2) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$.

Similarly,

$$\mathscr{B}(\mathbb{R}^4) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}^3) = \mathscr{B}(\mathbb{R}^2) \otimes \mathscr{B}(\mathbb{R}^2)$$

Meaning of $\mathbf{P}_1 \times \mathbf{P}_2$

If we have two independent RVs X, Y on a probability space $(\Omega, \mathscr{B}, \mathbf{P})$. Then we can put two RVs together to form a random vector, i.e.

$$\begin{pmatrix} X \\ Y \end{pmatrix} : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$$

Then the induced measure by this random vector is $\mathbf{P}' := \mathbf{P} \circ \begin{pmatrix} X \\ Y \end{pmatrix}^{-1}$. This is the joint distribution of the random vector $\begin{pmatrix} X \\ Y \end{pmatrix}$. And for the RVs

$$X, Y : (\Omega, \mathscr{B}, \mathbf{P}) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

Their induced measure $\mathbf{P}_1 = \mathbf{P} \circ X^{-1}$ and $\mathbf{P}_2 = \mathbf{P} \circ Y^{-1}$ are the marginal distributions.

What is the connection between this and the product measure? Now we have two probability measures on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$:

- $\mathbf{P}_1 \times \mathbf{P}_2$ is a probability measure on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$
- $\mathbf{P}'$ is a probability measure on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$

The claim is : $\mathbf{P}' = \mathbf{P}_1 \times \mathbf{P}_2$.

Because for any $A_1 \times A_2 \in RECT$, we have

$$\begin{aligned} \mathbf{P}'(A_1 \times A_2) = \mathbf{P} \circ \begin{pmatrix} X \\ Y \end{pmatrix}^{-1} (A_1 \times A_2) &= \mathbf{P}\left(\begin{pmatrix} X \\ Y \end{pmatrix} \in A_1 \times A_2\right) = \mathbf{P}(X \in A_1, Y \in A_2) \\ &= \mathbf{P}(X \in A_1)\, \mathbf{P}(Y \in A_2) = \mathbf{P}_1(A_1)\, \mathbf{P}_2(A_2) = (\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times A_2) \end{aligned}$$

So $\mathbf{P}' = \mathbf{P}_1 \times \mathbf{P}_2$ on $RECT$. But $RECT$ is a π-system. Then by Dynkin's $\pi - \lambda$-theorem, if two probability measures agree on a π-system, they agree on the σ-algebra generated by that π-system. Hence,

$$\mathbf{P}' = \mathbf{P}_1 \times \mathbf{P}_2 \qquad \text{on } \sigma(RECT) = \mathscr{B}(\mathbb{R}^2)$$

Remark 5.34. *Independence corresponds to product measure.*

Definition 5.35 (Strips). *Define the collection of strips of Ω_1 as*

$$\mathscr{B}_1^{\#} := \{A_1 \times \Omega_2 : A_1 \in \mathscr{B}_1\}$$

as well as the collection of strips of Ω_2 as

$$\mathscr{B}_2^{\#} := \{\Omega_1 \times A_2 : A_2 \in \mathscr{B}_2\}$$

[Board figure omitted: a vertical strip $A_1 \times \Omega_2$ and a horizontal strip $\Omega_1 \times A_2$ drawn across the box $\Omega_1 \times \Omega_2$, crossing in the rectangle $A_1 \times A_2$]

Claim 5.36. *Thus, $\mathscr{B}_1^{\#}$ and $\mathscr{B}_2^{\#}$ are independent on $\mathbf{P}_1 \times \mathbf{P}_2$.*

Because

$$(\mathbf{P}_1 \times \mathbf{P}_2)(A_1 \times \Omega_2 \cap \Omega_1 \times A_2) = \mathbf{P}_1 \times \mathbf{P}_2(A_1 \times A_2) = \mathbf{P}_1(A_1)\,\mathbf{P}_2(A_2)$$

On the other hand, we have

$$\begin{aligned}\mathbf{P}_1 \times \mathbf{P}_2(A_1 \times \Omega_2) &= \mathbf{P}_1(A_1)\,\mathbf{P}_2(\Omega_2) = \mathbf{P}_1(A_1)\\ \mathbf{P}_1 \times \mathbf{P}_2(\Omega_1 \times A_2) &= \mathbf{P}_1(\Omega_1)\,\mathbf{P}_2(A_2) = \mathbf{P}_2(A_2)\end{aligned}$$

Next, let us see how we can construct two independent random variables X_1, X_2 with given distribution functions $F_1(x)$ and $F_2(x)$.

We have already learned to construct

$$X_1 : (\Omega_1, \mathscr{B}_1, \mathbf{P}_1) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

such that X_1 has $F_1(x)$ as its DF. And we also have already learned to construct

$$X_2 : (\Omega_2, \mathscr{B}_2, \mathbf{P}_2) \longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

such that X_2 has $F_2(x)$ as its DF.

Construct independent RVs from Product space

Suppose we have two probability spaces $(\Omega_1, \mathscr{B}_1, \mathbf{P}_1)$ and $(\Omega_2, \mathscr{B}_2, \mathbf{P}_2)$. Then we can construct the product space $(\Omega_1 \times \Omega_2, \mathscr{B}_1 \otimes \mathscr{B}_2, \mathbf{P}_1 \times \mathbf{P}_2)$, and $\mathscr{B}_1^{\#}$ and $\mathscr{B}_2^{\#}$ are independent.

Question. *How to construct a probability space with two RVs X_1 and X_2 such that*

1) $X_1 \sim F_1(x)$ and $X_2 \sim F_2(x)$.

2) X_1 and X_2 are independent.

Last time we have quickly mentioned that we can construct $(\Omega_1, \mathscr{B}_1, \mathbf{P}_1)$ with X_1 such that X_1 has DF $F_1(x)$. Similarly, we can construct $(\Omega_2, \mathscr{B}_2, \mathbf{P}_2)$ with X_2 such that X_2 has DF $F_2(x)$.

Right now, X_1 and X_2 are two RVs on different probability spaces. So we can't talk about their independence. The solution is really to consider the product space.

Construct. Set $(\Omega_1 \times \Omega_2, \mathscr{B}_1 \otimes \mathscr{B}_2, \mathbf{P}_1 \times \mathbf{P}_2)$. Fix a $\omega_2 \in \Omega_2$. Define a map on $\Omega_1 \times \Omega_2$

$$\bar{X}_1(\omega_1, \omega_2) := X_1(\omega_1)$$

Similarly, fix a $\omega_1 \in \Omega_1$, define

$$\bar{X}_2(\omega_1, \omega_2) := X_2(\omega_2)$$

Then the σ-algebra generated by $\bar{X}_1$ will be

$$\begin{aligned}
\sigma(\bar{X}_1) &= \{\bar{X}_1^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{\{(\omega_1, \omega_2) : \bar{X}_1(\omega_1, \omega_2) \in B\} : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{\{(\omega_1, \omega_2) : X_1(\omega_1) \in B\} : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{X_1^{-1}(B) \times \Omega_2 : B \in \mathscr{B}(\mathbb{R})\} \\
&\subset \mathscr{B}_1^{\#}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sigma(\bar{X}_2) &= \{\bar{X}_2^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{\{(\omega_1, \omega_2) : \bar{X}_2(\omega_1, \omega_2) \in B\} : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{\{(\omega_1, \omega_2) : X_2(\omega_2) \in B\} : B \in \mathscr{B}(\mathbb{R})\} \\
&= \{\Omega_1 \times X_2^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\} \\
&\subset \mathscr{B}_2^{\#}
\end{aligned}$$

So $\sigma(\bar{X}_1)$ and $\sigma(\bar{X}_2)$ are independent. Hence $\bar{X}_1$ and $\bar{X}_2$ are independent.

The DF for $\bar{X}_1$ is

$$\begin{aligned}
F_{\bar{X}_1}(x) &= (\mathbf{P}_1 \times \mathbf{P}_2)(\bar{X}_1 \leqslant x) = \mathbf{P}_1 \times \mathbf{P}_2(\{(\omega_1, \omega_2) : \bar{X}_1(\omega_1, \omega_2) \leqslant x\}) \\
&= (\mathbf{P}_1 \times \mathbf{P}_2)(\bar{X}_1 \leqslant x) = \mathbf{P}_1 \times \mathbf{P}_2(\{(\omega_1, \omega_2) : X_1(\omega_1) \leqslant x\}) \\
&= (\mathbf{P}_1 \times \mathbf{P}_2)(X_1^{-1}((-\infty, x]) \times \Omega_2) \\
&= \mathbf{P}_1(X_1^{-1}((-\infty, x]))\, \mathbf{P}_2(\Omega_2) \\
&= \mathbf{P}_1(X_1^{-1}((-\infty, x])) = F_{X_1}(x) = F_1(x)
\end{aligned}$$

Hence, DF of $\bar{X}_1$ is equal to DF of X_1.

5.9 Fubini's theorem

Lecture 2024-11-19

Motivation

If you look at Riemann integration, and you have a function of two variables $f(x,y)$. You are trying to integrate it

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy$$

We can define the so-called double integration of this function. The way to do it is still divide the domain $\mathbb{R}^2$ into grid, and consider the Riemann sum. Then shrink the size of the grid to 0.

But the way we really do the computation is to first compute integration wrt x (or y), then integrate another, i.e.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)dx\right)dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)dy\right)dx$$

We first do dx if it is easier to compute. But the important remark is, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy$ is not defined to be the latter two. They are just happened to be equal. In fact, prove they are equal is a non-trivial task.

Key question

In Lebesgue integral, we consider the product space $(\Omega_1 \times \Omega_2, \mathscr{B}_1 \otimes \mathscr{B}_2, \mathbf{P}_1 \times \mathbf{P}_2)$. Consider a RV $X(\omega_1,\omega_2)$ on the product space. The key question is, do we have

$$\int_{\Omega_1\times\Omega_2} X d\,\mathbf{P}_1\times\mathbf{P}_2 \stackrel{?}{=} \int_{\Omega_1}\left(\int_{\Omega_2} X(\omega_1,\omega_2)d\,\mathbf{P}_2\right)d\,\mathbf{P}_1 \stackrel{?}{=} \int_{\Omega_2}\left(\int_{\Omega_1} X(\omega_1,\omega_2)d\,\mathbf{P}_1\right)d\,\mathbf{P}_2$$

Define

$$Y(\omega_1) = \int_{\Omega_2} X(\omega_1,\omega_2)d\,\mathbf{P}_2$$

Theorem 5.37 (Fubini's Theorem)**.** *1) $Y(\omega_1)$ is well-defined and $Y(\omega_1) \in \mathscr{B}_1$.*

2) If $X \geqslant 0$, then $Y(\omega_1) \geqslant 0$ and the following holds:

$$\int_{\Omega_1\times\Omega_2} X d\,\mathbf{P}_1\times\mathbf{P}_2 = \int_{\Omega_1}\left(\int_{\Omega_2} X(\omega_1,\omega_2)d\,\mathbf{P}_2\right)d\,\mathbf{P}_1 = \int_{\Omega_2}\left(\int_{\Omega_1} X(\omega_1,\omega_2)d\,\mathbf{P}_1\right)d\,\mathbf{P}_2$$

Call this property $(*)$.

3) If $X \in L^1(\mathbf{P}_1\times\mathbf{P}_2)$, then $Y(\omega_1) \in L^1(\mathbf{P}_1)$ and $(*)$ holds as well.

Proof. **Part (1).** Remember that when we do the Lebesgue integration, we need the function under the integral to be a measurable function. Here we are integrating Y, so we need Y to be a measurable function.

We will skip the proof.

Part (2). Assume $X \geqslant 0$.

Set

$$\mathscr{C} := \{A \in \mathscr{B}_1 \otimes \mathscr{B}_2 : \text{ s.t. } (*) \text{ holds for } \mathbf{1}_A\}$$

First goal is to show $\mathscr{B}_1 \otimes \mathscr{B}_2 \subset \mathscr{C}$. Note that

1) $(*)$ holds when $A = A_1 \times A_2$ is a measurable rectangle, i.e. $A \in RECT$, equivalently, this means $RECT \subset \mathscr{C}$.

2) $\mathscr{C}$ is a λ-system.

Proof of (1). To see (1), start from LHS,

$$LHS = \int_{\Omega_1 \times \Omega_2} \mathbf{1}_{A_1 \times A_2} d\,\mathbf{P}_1 \times \mathbf{P}_2 = \mathbf{P}_1 \times \mathbf{P}_2(A_1 \times A_2) = \mathbf{P}_1(A_1)\,\mathbf{P}_2(A_2)$$

And RHS is

$$\begin{aligned}
RHS &= \int_{\Omega_1} \left(\int_{\Omega_2} \mathbf{1}_{A_1 \times A_2}(\omega_1, \omega_2) d\,\mathbf{P}_2 \right) d\,\mathbf{P}_1 \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} \mathbf{1}_{A_1}(\omega_1) \mathbf{1}_{A_2}(\omega_2) d\,\mathbf{P}_2 \right) d\,\mathbf{P}_1 \\
&= \int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) \underbrace{\left(\int_{\Omega_2} \mathbf{1}_{A_2}(\omega_2) d\,\mathbf{P}_2 \right)}_{=\mathbf{P}_2(A_2)} d\,\mathbf{P}_1 \\
&= \int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1)\,\mathbf{P}_2(A_2) d\,\mathbf{P}_1 \\
&= \left(\int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) d\,\mathbf{P}_1 \right) \mathbf{P}_2(A_2) \\
&= \mathbf{P}_1(A_1)\,\mathbf{P}_2(A_2)
\end{aligned}$$

So LHS equals to RHS.
Proof of (2). We need to check 3 axioms of the λ-system.

1) $\Omega_1 \times \Omega_2 \in \mathscr{C}$ because $\Omega_1 \times \Omega_2 \in RECT$.

2) If $A \in \mathscr{C}$. Then $A^c \in \mathscr{C}$.

$$\begin{aligned}
LHS &= \int_{\Omega_1 \times \Omega_2} \mathbf{1}_{A^c} d\,\mathbf{P}_1 \times d\,\mathbf{P}_2 = 1 - \int_{\Omega_1 \times \Omega_2} \mathbf{1}_{A} d\,\mathbf{P}_1 \times d\,\mathbf{P}_2 \\
&= 1 - \int_{\Omega_1} \left(\int_{\Omega_2} \mathbf{1}_{A}(\omega_1, \omega_2) d\,\mathbf{P}_2 \right) d\,\mathbf{P}_1 \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} (1 - \mathbf{1}_{A}(\omega_1, \omega_2)) d\,\mathbf{P}_2 \right) d\,\mathbf{P}_1 \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} (\mathbf{1}_{A^c}(\omega_1, \omega_2)) d\,\mathbf{P}_2 \right) d\,\mathbf{P}_1 \\
&= RHS
\end{aligned}$$

3) If $A_1, A_2, \ldots \in \mathscr{C}$, and disjoint. Then $\cup_{n=1}^{\infty} A_n \in \mathscr{C}$.

Then

$$
\begin{aligned}
LHS &= \int_{\Omega_1\times\Omega_2} \mathbf{1}_{\cup_{n=1}^\infty A_n} d\,\mathbf{P}_1\times\mathbf{P}_2 = \int_{\Omega_1\times\Omega_2}\sum_{n=1}^\infty \mathbf{1}_{A_n} d\,\mathbf{P}_1\times\mathbf{P}_2 = \sum_{n=1}^\infty \int_{\Omega_1\times\Omega_2} \mathbf{1}_{A_n} d\,\mathbf{P}_1\times\mathbf{P}_2 \\
&= \sum_{n=1}^\infty \int_{\Omega_1}\left(\int_{\Omega_2} \mathbf{1}_{A_n} d\,\mathbf{P}_2\right) d\,\mathbf{P}_1 \\
&\overset{MCT\times 2}{=} \int_{\Omega_1}\left(\int_{\Omega_2}\sum_{n=1}^\infty \mathbf{1}_{A_n} d\,\mathbf{P}_2\right) d\,\mathbf{P}_1 \\
&= \int_{\Omega_1}\left(\int_{\Omega_2} \mathbf{1}_{\cup_{n=1}^\infty A_n} d\,\mathbf{P}_2\right) d\,\mathbf{P}_1 \\
&= RHS
\end{aligned}
$$

Hence, we proved $\mathscr{C}$ is a λ-system.

Then by Dynkin's $\pi-\lambda$ theorem, we know

1. $RECT$ is a π-system,
2. $\mathscr{C}$ is a λ-system (from (2)) such that $RECT \subset \mathscr{C}$ (from (1))

Hence, $\mathscr{C} \supset \sigma(RECT) = \mathscr{B}_1 \otimes \mathscr{B}_2$.

We have proved $(*)$ holds for $\mathbf{1}_A$ for all $A \in \mathscr{B}_1\otimes\mathscr{B}_2$. Hence $(*)$ holds for all simple RVs. Then approximation, $(*)$ holds for all $X \geqslant 0$.

Part (3). Now let $X \in L^1(\mathbf{P}_1\times\mathbf{P}_2)$. Write $X = X^+ - X^-$ with $X^+, X^- \geqslant 0$ as in the construction of the expectation. By Part (2), $(*)$ holds for X^+ and for X^-, and both integrals

$$\int_{\Omega_1\times\Omega_2} X^\pm\, d\,\mathbf{P}_1\times\mathbf{P}_2 \leqslant \int_{\Omega_1\times\Omega_2} |X|\, d\,\mathbf{P}_1\times\mathbf{P}_2 < \infty$$

are finite. In particular the functions

$$Y^\pm(\omega_1) := \int_{\Omega_2} X^\pm(\omega_1,\omega_2)\, d\,\mathbf{P}_2$$

satisfy $\int_{\Omega_1} Y^\pm\, d\,\mathbf{P}_1 < \infty$, so $Y^\pm \in L^1(\mathbf{P}_1)$, and $Y^\pm(\omega_1) < \infty$ for $\mathbf{P}_1$-almost every ω_1. For every such ω_1 the inner integral

$$Y(\omega_1) = \int_{\Omega_2} X(\omega_1,\omega_2)\, d\,\mathbf{P}_2 = Y^+(\omega_1) - Y^-(\omega_1)$$

is well defined and finite, and $Y = Y^+ - Y^- \in L^1(\mathbf{P}_1)$. Subtracting the two instances of $(*)$, all quantities being finite,

$$\int_{\Omega_1\times\Omega_2} X\, d\,\mathbf{P}_1\times\mathbf{P}_2 = \int_{\Omega_1} Y^+\, d\,\mathbf{P}_1 - \int_{\Omega_1} Y^-\, d\,\mathbf{P}_1 = \int_{\Omega_1}\left(\int_{\Omega_2} X(\omega_1,\omega_2)\, d\,\mathbf{P}_2\right) d\,\mathbf{P}_1$$

and the same argument with the roles of Ω_1 and Ω_2 exchanged gives the second iterated integral. This proves $(*)$ for all $X \in L^1(\mathbf{P}_1\times\mathbf{P}_2)$. $\square$

6 Convergence Concepts

6.1 Convergence Almost Surely

Lecture 2024-11-21

Convergence of Numbers

A sequence of numbers (a_n) converges to a. That means, the sequence has a limit. For example, if a_n is a Cauchy sequence, that is for any $\epsilon > 0$, we can find a N_0 such that

$$|a_n - a_m| < \epsilon, \qquad \text{for any } n, m \geqslant N_0$$

Definition 6.1 (Convergence almost surely)**.** *Given a probability space $(\Omega, \mathscr{B}, \mathbf{P})$. A sequence of random variables X_n converges to a target X almost surely if there is a null set $N \in \mathscr{B}$ and $\mathbf{P}(N) = 0$ such that*

$$X_n(\omega) \longrightarrow X(\omega), \qquad \text{as } n \to +\infty, \quad \text{for all } \omega \notin N$$

Remark 6.2. *This is really point-wise convergence.*

Example 6.1. *Consider the probability space $([0,1], \mathscr{B}([0,1]), \lambda)$ where λ is the Lebesgue measure. Then*

$$X_n = n\mathbf{1}_{[0,\frac{1}{n}]}$$

Then

$$X_n(\omega) \longrightarrow 0 \text{ if } \omega \neq 0$$

And only when

$$X_n(0) = n \longrightarrow +\infty$$

Here we say X_n converge to 0 almost surely, with $N = \{0\}$.

Example 6.2. *Let X_n be a sequence of IID RVs with common DF $F(x)$ such that $F(x) < 1$ for all $x \in \mathbb{R}$. Set*

$$M_n := \max\{X_1, \ldots, X_n\}$$

The claim is: $M_n \to +\infty$ almost surely.

Proof. Look at for any fixed $M > 0$. Denote $B_n = \{M_n \leqslant M\}$. Let

$$A_M = \{\{\omega \in \Omega : M_n(\omega) \leqslant M\} \text{ i.o.}\} = \limsup_{n\to\infty}\{M_n \leqslant M\} = \limsup_{n\to\infty} B_n$$

If $\omega \in A_M$, there are infinitely many $M_n(\omega)$ such that $M_n(\omega) \leqslant M$. Observe that

$$\begin{aligned}
\mathbf{P}(B_n) &= \mathbf{P}(M_n \leqslant M) \\
&= \mathbf{P}(X_1 \leqslant M, X_2 \leqslant M, \ldots, X_n \leqslant M) \\
&= \prod_{k=1}^{n} \mathbf{P}(X_k \leqslant M) = F(M)^n
\end{aligned}$$

Hence,
$$\sum_{n=1}^{\infty} \mathbf{P}(B_n) = \sum_{n=1}^{\infty} F(M)^n < +\infty$$
By Borel-Cantelli, we will have $\mathbf{P}(A_M) = 0$.

Because for each $M \in \mathbb{Z}_{\geqslant 1}$, we have $\mathbf{P}(A_M) = 0$. Hence, if we consider the union of all A_M, its probability will be

$$\mathbf{P}\left(\bigcup_{M=1}^{\infty} A_M\right) = 0$$

Here our null set will be
$$N = \bigcup_{M=1}^{\infty} A_M$$

- If $\omega \notin N$, then $\omega \in \cap_{M=1}^{\infty} A_M^c$.
- That is, for any $M \geqslant 1$, $\omega \in A_M^c = \liminf_{n\to\infty}\{M_n > M\}$.
- That is, there exists N_0 such that $M_n(\omega) > M$ for all $n \geqslant N_0$.

Summarise, we have $M_n(\omega) \to +\infty$ if $\omega \notin N$. □

Definition 6.3 (Cauchy sequence almost surely)**.** *A sequence of RVs X_n is said to be Cauchy almost surely, if there is a null set $N \in \mathscr{B}$ with $\mathbf{P}(N) = 0$ such that $X_n(\omega)$ is Cauchy sequence of numbers for all $\omega \in N^c$.*

Remark 6.4. *By definition, X_n is a Cauchy sequence almost surely, if and only if X_n converges almost surely. Because after we fix a $\omega \in N^c$, $X_n(\omega)$ is a number. So convergence almost surely is really convergence of numbers.*

6.2 Convergence in Probability

Lectures 2024-11-21, 2024-11-23

Definition 6.5 (Convergence in probability)**.** *A sequence of RVs X_n converges to X in probability if for any $\delta > 0$, we have*

$$\mathbf{P}(|X_n - X| > \delta) \longrightarrow 0 \qquad \text{as} \quad n \to +\infty$$

Example 6.3. *Consider the probability space $([0,1], \mathscr{B}([0,1]), \lambda)$ where λ is the Lebesgue* *measure. We define a sequence of RVs in the following way:*

$$\begin{aligned}
X_1 &:= 1\\
X_2 &:= \mathbf{1}_{[0,\frac{1}{2}]}\\
X_3 &:= \mathbf{1}_{[\frac{1}{2},1]}\\
X_4 &:= \mathbf{1}_{[0,\frac{1}{4}]}\\
X_5 &:= \mathbf{1}_{[\frac{1}{4},\frac{1}{2}]}\\
X_6 &:= \mathbf{1}_{[\frac{1}{2},\frac{3}{4}]}\\
X_7 &:= \mathbf{1}_{[\frac{3}{4},1]}\\
X_8 &:= \mathbf{1}_{[0,\frac{1}{8}]}\\
&\vdots
\end{aligned}$$

Claim 6.6. *$X_n(\omega)$ does not converge for all $\omega \in [0,1]$.*

Because $X_n(\omega)$ is a sequence of 0's and 1's with infinitely many 0's and infinitely many 1's.

So X_n does not converge almost surely.

Claim 6.7. *$X_n(\omega)$ does converge in probability.*

The intuition is, because the bar gets narrower and narrower, so the measure of these bars will shrink to zero. (In the picture.)

[Board figure omitted: the graphs of the indicator random variables X_n over $[0,1]$, whose bars become narrower as n grows]

$$\lambda(|X_n - 0| > \delta) = \lambda(A_n) \longrightarrow 0$$

Definition 6.8. *X_n is said to be a Cauchy sequence in probability if: for any $\delta > 0$, and $\epsilon > 0$, there is a N_0 such that*

$$\mathbf{P}(|X_n - X_m| > \delta) < \epsilon \qquad \forall n, m \geqslant N_0$$

Lemma 6.9 (EXAM)**.** *Suppose X_n is a Cauchy sequence in probability, then there is a subsequence that converges almost surely.*

Proof. Want to find a sub-sequence X_{n_k} such that X_{n_k} converges almost surely.

Set $n_0 = 1$. Because X_n is a Cauchy sequence, so for all $n, m > N$, we will have $\mathbf{P}(|X_n - X_m| > \frac{1}{2}) < \frac{1}{2}$. We define n_1 by picking the smallest of such N.

$$n_1 := \min\left\{ N \geqslant n_0 : \ \mathbf{P}(|X_n - X_m| > \frac{1}{2}) < \frac{1}{2} \quad \forall n, m \geqslant N \right\}$$

Similarly, define

$$n_2 := \min\left\{ N \geqslant n_1 : \ \mathbf{P}\left(|X_n - X_m| > \left(\frac{1}{2}\right)^2 \right) < \left(\frac{1}{2}\right)^2 \quad \forall n, m \geqslant N \right\}$$

Keep going, we define

$$n_k := \min\left\{ N \geqslant n_{k-1} : \ \mathbf{P}\left(|X_n - X_m| > \left(\frac{1}{2}\right)^k \right) < \left(\frac{1}{2}\right)^k \quad \forall n, m \geqslant N \right\}$$

Next, we show X_{n_k} is a Cauchy sequence almost surely. Because we are working in a complete space. A sequence being Cauchy is equivalent to a sequence is convergent. Hence, X_{n_k} converges almost surely. Let

$$A_k := \left\{ |X_{n_{k+1}} - X_{n_k}| > \left(\frac{1}{2}\right)^k \right\}$$

by our choice of n_k's, we will have

$$\mathbf{P}(A_k) < \left(\frac{1}{2}\right)^k$$

Hence,

$$\sum_{k=1}^{\infty} \mathbf{P}(A_k) < +\infty$$

Hence, by Borel-Cantelli, this implies

$$\mathbf{P}(\{A_k \text{ i.o.}\}) = 0$$

Thus, our null set is $N = \{A_k \text{ i.o.}\}$.

Finally, we showed for every $\omega \notin N$, the sequence $X_{n_k}(\omega)$ is a Cauchy sequence of numbers.

See that if

$$\omega \in N^c = \liminf_{k \to \infty} A_k^c$$

That is, for all $k \geqslant M_0(\omega)$, we have

$$\omega \in A_k^c = \left\{ |X_{n_{k+1}} - X_{n_k}| < \left(\frac{1}{2}\right)^k \right\}$$

This implies for that fixed ω, we have

$$|X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| < \left(\frac{1}{2}\right)^k$$

To see this is a Cauchy sequence, consider WLOG $r > l$, then using telescoping sum

$$\begin{aligned}
|X_{n_r}(\omega) - X_{n_l}(\omega)| &\leqslant \underbrace{|X_{n_r}(\omega) - X_{n_{r-1}}(\omega)|}_{\leqslant (\frac{1}{2})^{r-1}} + \underbrace{|X_{n_{r-1}}(\omega) - X_{n_{r-2}}(\omega)|}_{\leqslant (\frac{1}{2})^{r-2}} + \ldots + \underbrace{|X_{n_{l+1}}(\omega) - X_{n_l}(\omega)|}_{\leqslant (\frac{1}{2})^{l}} \\
&\leqslant \sum_{k=l}^{r-1} \left(\frac{1}{2}\right)^k \\
&\leqslant \sum_{k=l}^{\infty} \left(\frac{1}{2}\right)^k \\
&= \left(\frac{1}{2}\right)^{l-1}
\end{aligned}$$

□

Theorem 6.10. *Let X_n be a Cauchy sequence in probability is equivalent to X_n converges to X in probability.*

Proof. ($\Longleftarrow$) If $X_n \to X$ in probability. Then

$$\begin{aligned}\mathbf{P}(|X_n - X_m| > \delta) &= \mathbf{P}(|X_n - X + X - X_m| > \delta)\\ &\leqslant \mathbf{P}(|X_n - X| + |X - X_m| > \delta)\\ &\leqslant \mathbf{P}\left(|X_n - X| > \frac{\delta}{2}\right) + \mathbf{P}\left(|X - X_m| > \frac{\delta}{2}\right)\end{aligned}$$

($\Longrightarrow$) Suppose X_n is a Cauchy sequence in probability. Then by the lemma we have a subsequence X_{n_k} that converges to X almost surely. This further implies $X_{n_k} \to X$ in probability.

Next, we want to show $X_n \to X$ in probability

$$\begin{aligned}\mathbf{P}(|X_n - X| > \delta) &= \mathbf{P}(|X_n - X_{n_k} + X_{n_k} - X| > \delta)\\ &\leqslant \mathbf{P}(|X_n - X_{n_k}| > \frac{\delta}{2}) + \mathbf{P}(|X_{n_k} - X| > \frac{\delta}{2})\end{aligned}$$

But because X_n is a Cauchy sequence in probability, so when n is large, the probabilities

$$\mathbf{P}(|X_n - X_{n_k}| > \frac{\delta}{2}), \qquad \mathbf{P}(|X_{n_k} - X| > \frac{\delta}{2})$$

are both small. This finishes the proof. ☐

6.3 Connections Between a.s. and i.p. Convergence

Lectures 2024-11-21, 2024-11-23, 2024-11-26

Lemma 6.11. *Convergence almost surely implies convergence in probability.*

Proof. If we have $X_n \to X$ almost surely. Then we have

$$\begin{aligned}0 &= \mathbf{P}(|X_n - X| > \delta \text{ i.o.})\\ &= \mathbf{P}\left(\limsup_{n\to\infty}\{|X_n - X| > \delta\}\right)\\ &= \mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k - X| > \delta\}\right)\\ &= \lim_{n\to\infty}\mathbf{P}\underbrace{\left(\bigcup_{k=n}^{\infty}\{|X_k - X| > \delta\}\right)}_{\supset\{|X_n - X| > \delta\}}\\ &\geqslant \limsup_{n\to\infty}\mathbf{P}(\{|X_n - X| > \delta\})\\ &\geqslant 0\end{aligned}$$

This implies

$$\lim_{n\to\infty}\mathbf{P}(\{|X_n - X| > \delta\}) = 0$$

☐

Theorem 6.12. *X_n converges to X in probability is equivalent to for any subsequence X_{n_k} of X_n, there is a further subsequence $X_{n_{k_i}}$ of X_{n_k} which converges to X almost surely.*

Proof. ($\Longrightarrow$) Suppose X_n converges to X in probability. Then we pick any subsequence X_{n_k}, this subsequence also converges in probability. Then X_{n_k} is a Cauchy sequence in probability. Hence, there is a further subsequence $X_{n_{k_i}}$ converges to X almost surely by the lemma (EXAM).
($\Longleftarrow$) Suppose any subsequence X_{n_k} of X_n, there is a further subsequence $X_{n_{k_i}}$ of X_{n_k} which converges to X almost surely.

We prove by contradiction. Suppose X_n does not converges to X in probability. Then there exists some $\delta > 0$ and $\epsilon > 0$ and a sequence n_k such that

$$\mathbf{P}(|X_{n_k} - X| > \delta) > \epsilon$$

We look at X_{n_k}, then there is a further subsequence $X_{n_{k_i}}$ of X_{n_k} such that

$$X_{n_{k_i}} \longrightarrow X \qquad \text{almost surely}$$

This implies

$$\mathbf{P}(|X_{n_{k_i}} - X| > \delta) \to 0 \qquad \text{as } i \to \infty$$

This contradicts there exists some $\delta > 0$ and $\epsilon > 0$ and for all n_k we will have

$$\mathbf{P}(|X_{n_k} - X| > \delta) > \epsilon$$

□

Corollary 6.13. *Let X_n be a sequence of RVs. Let $f(x)$ be a continuous function. Then*

1) $X_n \longrightarrow X$ almost surely implies $f(X_n) \longrightarrow f(X)$ almost surely.

2) $X_n \longrightarrow X$ in probability implies $f(X_n) \longrightarrow f(X)$ in probability

Proof. **Proof of (1):** Suppose $X_n \to X$ almost surely. Then there exists a null set $N \in \mathscr{B}$ such that $\mathbf{P}(N) = 0$, and $X_n(\omega) \to X(\omega)$ for all $\omega \in N^c$.

Then

$$f(X_n(\omega)) \longrightarrow f(X(\omega)) \qquad \text{for all } \omega \in N^c$$

Proof of (2): In order to show

$$f(X_n) \longrightarrow f(X) \qquad \text{in probability}$$

is equivalent to show for any subsequence $f(X_{n_k})$ of $f(X_n)$, there is a further subsequence $f(X_{n_{k_i}})$ of $f(X_{n_k})$ converge almost surely.

By assumption, $X_n \to X$ in probability. According to the theorem, for any subsequence X_{n_k}, there is a further subsequence $X_{n_{k_i}}$ of X_{n_k} such that $X_{n_{k_i}} \to X$ almost surely. Then by (1), then $f(X_{n_{k_i}}) \to f(X)$ almost surely. Then we use converge almost surely implies converge in probability. □

Proposition 6.14. *Let $X_n \to X$ in probability and $Y_n \to Y$ in probability. Then*

$$X_n + Y_n \to X + Y \qquad \textit{in probability}$$

and

$$X_n \cdot Y_n \to X \cdot Y \qquad \textit{in probability}$$

Proof. Want to show for any subsequence $X_{n_k} + Y_{n_k}$, there is a further subsequence which converges to $X + Y$ almost surely.

The question is how to find this "further subsequence".

Since $X_n \to X$ in probability, by the theorem, we can find a further subsequence $X_{n_{k_i}}$ of X_{n_k} such that $X_{n_{k_i}} \to X$ almost surely.

Now the problem is, if we consider the subsubsequence $Y_{n_{k_i}}$, it may not converge almost surely. (It is just a random subsequence of Y_n. It is not constructed using the theorem we were using).

But by the theorem, there is a further subsequence $Y_{n_{k_{i_j}}}$ of $Y_{n_{k_i}}$ such that $Y_{n_{k_{i_j}}} \to Y$ almost surely.

Now, consider the subsequence $X_{n_{k_{i_j}}}$ of $X_{n_{k_i}}$. Any subsequence of a sequence converge almost surely also converge almost surely. Hence, $X_{n_{k_{i_j}}} \to X$ almost surely.

At the end,

$$X_{n_{k_{i_j}}} + Y_{n_{k_{i_j}}} \longrightarrow X + Y \qquad \text{almost surely}$$

By the theorem, this proves $X_n + Y_n \to X + Y$ in probability.

The product. The same construction proves the second statement; no new extraction is needed. Fix a subsequence $X_{n_k} Y_{n_k}$ of $X_n Y_n$. Running the argument above on the indices n_k produces indices $n_{k_{i_j}}$ along which

$$X_{n_{k_{i_j}}} \longrightarrow X \qquad \text{and} \qquad Y_{n_{k_{i_j}}} \longrightarrow Y \qquad \text{almost surely}$$

Let N be the union of the two null sets off which these two convergences hold; then $\mathbf{P}(N) = 0$, and for every $\omega \in N^c$ the two sequences of real numbers $X_{n_{k_{i_j}}}(\omega)$ and $Y_{n_{k_{i_j}}}(\omega)$ converge to $X(\omega)$ and $Y(\omega)$. The product of two convergent sequences of real numbers converges to the product of the limits, so

$$X_{n_{k_{i_j}}} \cdot Y_{n_{k_{i_j}}} \longrightarrow X \cdot Y \qquad \text{almost surely}$$

Thus every subsequence of $X_n Y_n$ has a further subsequence converging to $X \cdot Y$ almost surely, and the theorem gives $X_n \cdot Y_n \to X \cdot Y$ in probability. $\square$

Corollary 6.15 (DCT – relaxed version)**.** *If $X_n \to X$ in probability, and $|X_n| \leqslant Y \in L^1(\mathbf{P})$. Then*

$$\lim_{n\to\infty} \mathbf{E}\, X_n = \mathbf{E}\, X$$

Proof. Prove by contradiction. Suppose the limit of $\mathbf{E}\, X_n$ is not $\mathbf{E}\, X$. Then there exists a subsequence $\mathbf{E}\, X_{n_k}$ such that

$$|\mathbf{E}\, X_{n_k} - \mathbf{E}\, X| > \epsilon$$

for some small $\epsilon > 0$.

On the other hand, for X_{n_k}, we can find a further subsequence $X_{n_{k_i}}$ of X_{n_k} such that $X_{n_{k_i}} \to X$ almost surely.

By the usual DCT, we have

$$\mathbf{E}\, X_{n_{k_i}} \longrightarrow \mathbf{E}\, X \qquad \text{as } i \to \infty$$

This contradicts the $|\mathbf{E}\, X_{n_k} - \mathbf{E}\, X| > \epsilon$ for some small $\epsilon > 0$. □

6.5 L^p Convergence

Lectures 2024-11-26, 2024-12-03, 2024-12-05

Definition 6.16 ($L^p(\mathbf{P})$ space)**.** *For any $p \geqslant 1$, define*

$$L^p(\mathbf{P}) := \{X : \mathbf{E}|X|^p < \infty\}$$

Proposition 6.17 (Properties of $L^p(\mathbf{P})$ space)**.** *1) $L^p(\mathbf{P})$ is a $\mathbb{R}$-linear space. That is, if $X, Y \in L^p(\mathbf{P})$, and $a, b \in \mathbb{R}$. Then $aX + bY \in L^p(\mathbf{P})$.*

Proof. **Proof of (1):** It is trivial that $aX \in L^p(\mathbf{P})$ and $bY \in L^p(\mathbf{P})$. So, we only need to show $X + Y \in L^p(\mathbf{P})$ if $X, Y \in L^p(\mathbf{P})$. Because

$$\begin{aligned}\mathbf{E}|X+Y|^p &\leqslant \mathbf{E}\big(2(|X| \vee |Y|)\big)^p = 2^p\, \mathbf{E}(|X|^p \vee |Y|^p) \\ &\leqslant 2^p\, \mathbf{E}(|X|^p + |Y|^p) = 2^p(\underbrace{\mathbf{E}(|X|^p)}_{<\infty} + \underbrace{\mathbf{E}(|Y|^p)}_{<\infty}) \\ &< \infty\end{aligned}$$

This completes the proof. □

Definition 6.18 ($L^p(\mathbf{P})$-norm)**.** *$L^p(\mathbf{P})$ norm of $X \in L^p(\mathbf{P})$ is defined by*

$$\|X\|_p := \big(\mathbf{E}|X|^p\big)^{\frac{1}{p}}$$

This is a norm of the $\mathbb{R}$-vector space $L^p(\mathbf{P})$, because it satisfies the axioms of norm. That is:

1) $\|X\|_p \geqslant 0$, and $\|X\|_p = 0$ if and only if $X = 0$ almost surely.

2) $\|aX\|_p = |a|\|X\|_p$.

3) Triangular inequality.

$$\|X+Y\|_p \leqslant \|X\|_p + \|Y\|_p$$

Remark 6.19. *Triangular inequality for $L^p(\mathbf{P})$-norm is called Minkowski inequality.*

Definition 6.20 ($L^p(\mathbf{P})$ convergence)**.** *Let $X_n, X \in L^p(\mathbf{P})$ space. We say $X_n \to X$ in $L^p(\mathbf{P})$ if*

$$\|X_n - X\|_p \to 0$$

Lemma 6.21 (Holder's Inequality)**.** *If $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$. (In literature, people say p is conjugate to q). Then*

$$\|XY\|_1 = \mathbf{E}|XY| \leqslant \|X\|_p \cdot \|Y\|_q$$

Meaning of Holder Inequality

If $p = q = 2$, then we have

$$\mathbf{E}|XY| \leqslant \|X\|_2 \|Y\|_2$$

If $X = Y$, then we have

$$\mathbf{E}|X \cdot X| = \mathbf{E}\, X^2 < \infty$$

This means, if we want $X \in L^1(\mathbf{P})$, then a sufficient condition is $X \in L^2(\mathbf{P})$.

There is a balance between integrability of X and integrability of Y.

Important Inequalities

1. (Holder's inequality): for $p, q \geqslant 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then $\mathbf{E}|XY| \leqslant \|X\|_p \|Y\|_q$
2. (Minkowski inequality): for any $p \geqslant 1$, we have $\|X + Y\|_p \leqslant \|X\|_p + \|Y\|_p$.
3. (Jensen's inequality): If $f(x)$ is a convex function. Then $f(\mathbf{E}\, Y) \leqslant \mathbf{E}\, f(Y)$.

Definition 6.22 (Convex function)**.** *A function $f(x)$ is called a convex function if it satisfies the following property*

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y), \qquad \forall\ t \in (0,1)$$

[Board figure omitted: the graph of a convex function together with the chord joining two of its points, the chord lying above the graph]

Another way to define a convex function is, to write $f(x)$ as

$$f(x) = \sup_{l(x) \in \mathrm{Lin}(\mathbb{R},\mathbb{R}),\ l(x) \leqslant f(x)\ \forall x} l(x)$$

Proof of Holder's Inequality. Note, for $a, b > 0$, we have

$$ab \leqslant \frac{1}{p} a^p + \frac{1}{q} b^q, \qquad (*)$$

because if we look at $f(s) = e^s$, and we pick s, t such that $a = e^{\frac{s}{p}}$ and $b = e^{\frac{t}{q}}$. Thus, we get

$$\begin{aligned} ab = e^{\frac{s}{p} + \frac{t}{q}} &= f(\frac{1}{p}s + \frac{1}{q}t) \\ &\leqslant \frac{1}{p} f(s) + \frac{1}{q} f(t) \\ &= \frac{1}{p} e^s + \frac{1}{q} e^t = \frac{1}{p} a^p + \frac{1}{q} b^q \end{aligned}$$

Take

$$a = \frac{|X|}{\|X\|_p}, \qquad b = \frac{|Y|}{\|Y\|_q}$$

By $(*)$, we have

$$\frac{|X|}{\|X\|_p} \cdot \frac{|Y|}{\|Y\|_q} \leqslant \frac{1}{p}\left(\frac{|X|}{\|X\|_p}\right)^p + \frac{1}{q}\left(\frac{|Y|}{\|Y\|_q}\right)^q$$

Taking expectation both sides, we have

$$\frac{\mathbf{E}|X||Y|}{\|X\|_p\|Y\|_q} \leqslant \frac{1}{p}\underbrace{\frac{\mathbf{E}|X|^p}{\|X\|_p^p}}_{=1} + \frac{1}{q}\underbrace{\frac{\mathbf{E}|Y|^q}{\|Y\|_q^q}}_{=1} = \frac{1}{p} + \frac{1}{q} = 1$$

Hence, we get

$$\mathbf{E}|XY| \leqslant \|X\|_p\|Y\|_q$$

This completes the proof. □

Proof of Minkowski's Inequality. If $p = 1$, then we get the usual triangular inequality for real numbers, i.e.

$$\mathbf{E}|X+Y| \leqslant \mathbf{E}(|X| + |Y|) = \mathbf{E}|X| + \mathbf{E}|Y|$$

For $p > 1$. Note first $p - 1 = \frac{p}{q}$. So

$$\begin{aligned}
\|X+Y\|_p^p &= \mathbf{E}|X+Y|^p = \mathbf{E}|X+Y|^{1+\frac{p}{q}} = \mathbf{E}|X+Y||X+Y|^{\frac{p}{q}} \\
&\leqslant \mathbf{E}[(|X|+|Y|)|X+Y|^{\frac{p}{q}}] = \mathbf{E}|X||X+Y|^{\frac{p}{q}} + \mathbf{E}|Y||X+Y|^{\frac{p}{q}}, \quad \Delta\text{-ineq for } \mathbb{R} \\
&\leqslant \|X\|_p\||X+Y|^{\frac{p}{q}}\|_q + \|Y\|_p\||X+Y|^{\frac{p}{q}}\|_q, \qquad \text{Holder's ineq} \\
&= \|X\|_p(\mathbf{E}|X+Y|^p)^{\frac{1}{q}} + \|Y\|_p(\mathbf{E}|X+Y|^p)^{\frac{1}{q}} \\
&= \|X\|_p((\mathbf{E}|X+Y|^p)^{\frac{1}{p}})^{\frac{p}{q}} + \|Y\|_p((\mathbf{E}|X+Y|^p)^{\frac{1}{p}})^{\frac{p}{q}} \\
&= \|X\|_p \cdot \||X+Y|\|_p^{\frac{p}{q}} + \|Y\|_p \cdot \||X+Y|\|_p^{\frac{p}{q}} \\
&= (\|X\|_p + \|Y\|_p) \cdot \||X+Y|\|_p^{\frac{p}{q}} = (\|X\|_p + \|Y\|_p) \cdot \||X+Y|\|_p^{p-1}
\end{aligned}$$

Factor out $\||X+Y|\|_p^{p-1}$ both sides, we have

$$\|X+Y\|_p \leqslant \|X\|_p + \|Y\|_p$$

This completes the proof. □

Example 6.4 (Jensen's inequality)**.** *Consider $f(x) = x^2$. By Jensen's inequality, we have*

$$(\mathbf{E}\,X)^2 \leqslant \mathbf{E}\,X^2$$

Recall the variance of the random variable is always non-negative, and it is of the form

$$0 \leqslant \operatorname{Var}(X) = \mathbf{E}(X - \mathbf{E}\,X)^2 = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2$$

Hence, we get the Jensen's inequality.

Proof of Jensen's inequality. Recall that a convex function is defined as

$$f(\mathbf{E}\,X) = \sup_{l \in \mathrm{Lin}(\mathbb{R},\mathbb{R}),\ l(z) \leqslant f(z)\ \forall z} l(\mathbf{E}\,X)$$

Recall that a linear function in $\mathbb{R}$ has the form $l(x) = ax + b$ for some $a, b \in \mathbb{R}$. Hence,

$$l(\mathbf{E}\,X) = a\,\mathbf{E}\,X + b = \mathbf{E}(aX + b) = \mathbf{E}\,l(X)$$

Hence, the convex function can be expressed as

$$f(\mathbf{E}\,X) = \sup_{l \in \mathrm{Lin}(\mathbb{R},\mathbb{R}),\ l(z) \leqslant f(z)\ \forall z} l(\mathbf{E}\,X) = \sup_{l \in \mathrm{Lin}(\mathbb{R},\mathbb{R}),\ l(z) \leqslant f(z)\ \forall z} \mathbf{E}\,l(X)$$

But because $l(x) \leqslant f(x)$ for all $x \in \mathbb{R}$, hence we have

$$f(\mathbf{E}\,X) = \sup_{l \in \mathrm{Lin}(\mathbb{R},\mathbb{R}),\ l(z) \leqslant f(z)\ \forall z} \mathbf{E}\,\underbrace{l(X)}_{\leqslant f(X)} \leqslant \mathbf{E}\,f(X)$$

This completes the proof. □

Remark 6.23. *If $m > n$ for some $m, n \geqslant 1$. Then if $X \in L^m(\mathbf{P})$, then $X \in L^n(\mathbf{P})$.*
Given $\mathbf{E}|X|^m < \infty$. We want to show $\mathbf{E}|X|^n < \infty$.
By Holder's inequality and choose $p = \frac{m}{n}$, we have

$$\mathbf{E}|X|^n = \mathbf{E}|X|^n \cdot 1 \leqslant \||X|^n\|_p \|1\|_q = (\mathbf{E}|X|^m)^{\frac{n}{m}} = (\mathbf{E}|X|^m)^{\frac{n}{m}} < \infty$$

Recap

We have three form of convergence

1. Convergence almost surely
2. Convergence in probability
3. Convergence in $L^p(\mathbf{P})$ for $p \geqslant 1$.

We have discussed the first two form of convergence. This time we will discuss convergence in $L^p(\mathbf{P})$.

[Board figure omitted: finite dimensional vector convergence]

In general, $L^p(\mathbf{P})$ is an infinite-dimensional vector space.

Remark 6.24. *The relation of these three form of convergence is*

- *Convergence in $L^p(\mathbf{P})$ implies convergence in probability*
- *No relation between convergence in $L^p(\mathbf{P})$ and convergence almost surely*

Theorem 6.25. *Convergence in $L^p(\mathbf{P})$ implies convergence in probability.*

Proof. We want to show

$$\mathbf{P}(|X_n - X| > \delta) \longrightarrow 0 \qquad \text{as } n \to \infty$$

We will apply Chebyshev's inequality. See that

$$\{|X_n - X| > \delta\} = \{|X_n - X|^p > \delta^p\}$$

Then by Chebyshev's inequality, we have

$$\begin{aligned}\mathbf{P}(|X_n - X| > \delta) &= \mathbf{P}(|X_n - X|^p > \delta^p) \\ &\leqslant \frac{\mathbf{E}|X_n - X|^p}{\delta^p} = \frac{\|X_n - X\|_p^p}{\delta^p} \\ &\to 0 \qquad \text{by assumption}\end{aligned}$$

□

Definition 6.26 (Cauchy sequence)**.** *We say X_n is a Cauchy sequence in $L^p(\mathbf{P})$ if*

- $X_n \in L^p(\mathbf{P})$
- $\|X_n - X_m\|_p \longrightarrow 0$ *as* $n, m \to \infty$.

Question. *Are we going to have following equivalence? X_n converges to X in $L^p(\mathbf{P})$ if and only if X_n is a Cauchy sequence in $L^p(\mathbf{P})$.*

Question. *Suppose X_n converges to X in probability. What can we say for convergence in $L^p(\mathbf{P})$?*

These two questions is going to be answered in one theorem. By before that, we need a definition.

Definition 6.27 (Uniform Integrability – 1)**.** *A family of RVs $\{X_t : t \in \mathbb{T}\}$ is said to be uniformly integrable (UI) if the following holds:*

1. *$X_t \in L^1(\mathbf{P})$ for all $t \in \mathbb{T}$.*
2. *Because $X_t \in L^1(\mathbf{P})$, so the set $\{X_t > N\} \downarrow \varnothing$ as N becomes large. Thus,*

$$\sup_{t \in \mathbb{T}} \int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} \longrightarrow 0 \qquad \textit{as } N \to +\infty$$

Example 6.5. *Consider a single random variable $\{X\}$ where $X \in L^p(\mathbf{P})$. Then this family is uniformly integrable.*

Because since $X \in L^1(\mathbf{P})$. So

$$\mathbf{P}(|X| \geqslant N) \longrightarrow 0$$

So

$$Y_N := X \cdot \mathbf{1}_{\{|X| > N\}} \longrightarrow 0 \qquad \textit{almost surely}$$

In addition, the random variable Y_N is bounded above by $|X| \in L^1(\mathbf{P})$, i.e. $|Y_N| \leqslant |X| \in L^1(\mathbf{P})$. So by DCT, we have

$$\int_{\{|X| > N\}} |X| d\,\mathbf{P} = \int Y_N d\,\mathbf{P} \longrightarrow 0 \qquad \textit{as } N \to \infty$$

Example 6.6 (Dominated family is uniformly integrable). *A family of random variables $\{X_t : t \in \mathbb{T}\}$ is called a dominated family if there exists a $Y \in L^1(\mathbf{P})$ such that*

$$|X_t| \leqslant Y \in L^1(\mathbf{P})$$

Then $\{X_t : t \in \mathbb{T}\}$ is uniformly integrable.

This is because

$$\sup_{t \in \mathbb{T}} \int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} \leqslant \int_{\{|Y| > N\}} |Y| d\,\mathbf{P}$$

Example 6.7 (Finite family is uniformly integrable). *If we have a finite family of random variables $\{X_1, \ldots, X_n\}$ where each $X_i \in L^1(\mathbf{P})$. Then this family is uniformly integrable.*

This is because if we let $Y = |X_1| + \ldots + |X_n|$. Then

- *Y is integrable, i.e. $Y \in L^1(\mathbf{P})$*
- *All X_n is dominated by Y, i.e. $|X_n| \leqslant Y$ for all n.*

Thus, the uniformly integrability of this family follows from the last example.

Example 6.8 (A family that is dominated by a uniformly integrable family is also uniformly integrable). *If $\{X_t : t \in \mathbb{T}\}$ is dominated by $\{Y_s : s \in \mathbb{S}\}$, i.e. $|X_t| \leqslant |Y_t|$. And $\{Y_s : s \in \mathbb{S}\}$ is uniformly integrable. Then X_t is uniformly integrable.*

Example 6.9. *Suppose for some $\varepsilon > 0$, we have*

$$\sup_{t \in \mathbb{T}} \mathbf{E}|X_t|^{1+\varepsilon} \leqslant M < +\infty$$

Then $\{X_t : t \in \mathbb{T}\}$ is uniformly integrable.

Proof. Consider the targeted integral. Note that under the event $\{|X_t| > N\}$, we have the following inequality.

$$\begin{aligned}
\int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} &\leqslant \int_{\{|X_t| > N\}} |X_t| \left(\frac{|X_t|}{N}\right)^{\epsilon} d\,\mathbf{P} \\
&= \frac{1}{N^{\varepsilon}} \int_{\{|X_t| > N\}} |X_t|^{1+\varepsilon} d\,\mathbf{P} \\
&\leqslant \frac{1}{N^{\varepsilon}} \sup_t \mathbf{E}|X_t|^{1+\varepsilon} \\
&\leqslant \frac{M}{N^{\varepsilon}} \\
&\to 0 \qquad \text{as } N \to +\infty
\end{aligned}$$

□

Remark 6.28. *There is an equivalent definition of uniformly integrability.*

Definition 6.29 (Uniform integrability – 2). *A family of random variables $\{X_t : t \in \mathbb{T}\}$ is said to be uniformly integrable if*

1) $\sup_t \mathbf{E}|X_t| \leqslant M < +\infty$

2) For any $\epsilon > 0$*, there exists a* $\delta > 0$ *such that*

$$\sup_t \int_B |X_t| d\,\mathbf{P} < \epsilon$$

if $\mathbf{P}(B) < \delta$*.*

Theorem 6.30. *Two definitions of uniformly integrability are equivalent.*

Proof. ($\implies$) Consider the integral.

$$\int |X_t| d\,\mathbf{P} = \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} + \int_{\{|X_t|\leqslant N\}} |X_t| d\,\mathbf{P}$$

Taking the supremum both sides,

$$\sup_t \int |X_t| d\,\mathbf{P} \leqslant \sup_t \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} + \sup_t \int_{\{|X_t|\leqslant N\}} |X_t| d\,\mathbf{P}$$

From previously showed results, if N is large enough, then we have

$$\int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} \to 0 \implies \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} < 1$$

Hence, we can pick N large enough such that

$$\begin{aligned}
\sup_t \int |X_t| d\,\mathbf{P} &= \underbrace{\sup_t \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P}}_{\leqslant 1} + \underbrace{\sup_t \int_{\{|X_t|\leqslant N\}} |X_t| d\,\mathbf{P}}_{\leqslant \int_{\{|X_t|\leqslant N\}} N d\,\mathbf{P}} \\
&\leqslant 1 + \int_{\{|X_t|\leqslant N\}} N d\,\mathbf{P} \\
&\leqslant 1 + \int N d\,\mathbf{P} = 1 + \mathbf{E}\, N = 1 + N
\end{aligned}$$

Hence, if we further integrate over B, then we will have

$$\begin{aligned}
\int_B |X_t| d\,\mathbf{P} &= \int_{B\cap\{|X_t|>N\}} |X_t| d\,\mathbf{P} + \int_{B\cap\{|X_t|\leqslant N\}} |X_t| d\,\mathbf{P} \\
&\leqslant \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} + \int_B N d\,\mathbf{P} \\
&= \int_{\{|X_t|>N\}} |X_t| d\,\mathbf{P} + N\,\mathbf{P}(B)
\end{aligned}$$

By the first definition of uniform integrability we may fix an N_1, depending only on ϵ, with

$$\sup_t \int_{\{|X_t|>N_1\}} |X_t| d\,\mathbf{P} < \frac{\epsilon}{2}$$

Having fixed N_1, set $\delta := \frac{\epsilon}{2N_1}$. Then every B with $\mathbf{P}(B) < \delta$ satisfies $N_1 \mathbf{P}(B) < \frac{\epsilon}{2}$, and since the derivation of the display above is valid for any cutoff in place of N, applying it with N_1 gives

$$\sup_t \int_B |X_t| d\,\mathbf{P} \leqslant \sup_t \int_{\{|X_t| > N_1\}} |X_t| d\,\mathbf{P} + N_1 \mathbf{P}(B) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which is condition 2) of the second definition. Condition 1) is the bound $\sup_t \mathbf{E}|X_t| \leqslant 1 + N$ obtained above.

($\Longleftarrow$) Suppose 1) and 2) of the second definition hold. Condition 1) gives $\mathbf{E}|X_t| \leqslant M < +\infty$ for every t, so each $X_t \in L^1(\mathbf{P})$, which is condition 1. of the first definition.

Fix $\epsilon > 0$ and let $\delta > 0$ be the number supplied by 2). By Chebyshev's inequality and 1),

$$\sup_t \mathbf{P}(|X_t| > N) \leqslant \sup_t \frac{\mathbf{E}|X_t|}{N} \leqslant \frac{M}{N}$$

So if $N_0 > \frac{M}{\delta}$, then for every $N \geqslant N_0$ and every t the event $B_t := \{|X_t| > N\}$ has $\mathbf{P}(B_t) \leqslant \frac{M}{N} < \delta$. Condition 2) is a statement about *every* event of probability less than δ, so we are free to apply it to this t-dependent event:

$$\int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} = \int_{B_t} |X_t| d\,\mathbf{P} \leqslant \sup_s \int_{B_t} |X_s| d\,\mathbf{P} < \epsilon$$

Taking the supremum over t,

$$\sup_t \int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} \leqslant \epsilon \qquad \text{for every } N \geqslant N_0$$

Since $\epsilon > 0$ was arbitrary, this says exactly

$$\sup_t \int_{\{|X_t| > N\}} |X_t| d\,\mathbf{P} \longrightarrow 0 \qquad \text{as } N \to +\infty$$

which is condition 2. of the first definition. □

6.6 More on L^p Convergence

Lecture 2024-12-07

Theorem 6.31. *Let (X_n) be a sequence of RVs and $p \geqslant 1$. The following are equivalent*

1) *(X_n) converges in $L^p(\mathbf{P})$.*

2) *(X_n) is a Cauchy sequence in $L^p(\mathbf{P})$.*

3) *The sequence $(|X_n|^p)$ is uniformly integrable and (X_n) converges in probability.*

Proof. We prove for $p = 1$. The proof for general $p > 1$ will be similar.

(1) to (2): We know X_n converges to X in $L^p(\mathbf{P})$, i.e.

$$\|X_n - X\|_p \longrightarrow 0 \quad \text{as } n \to \infty$$

What we want is

$$\begin{aligned}\|X_n - X_m\|_p &= \|X_n - X + X - X_m\|_p \\ &\leqslant \underbrace{\|X_n - X\|_p}_{\to 0} + \underbrace{\|X_m - X\|_p}_{\to 0}, \qquad \text{triangle ineq} \\ &\to 0 \quad \text{as } n, m \to \infty\end{aligned}$$

(2) to (3): Suppose $(X_n) \subset L^1(\mathbf{P})$ is Cauchy. Two things have to be shown: that the family $\{X_n : n \geqslant 1\}$ is uniformly integrable, and that (X_n) converges in probability. We verify uniform integrability in the form of the second definition, that is

1) $\sup_{n\geqslant 1} \mathbf{E}|X_n| < +\infty$, and

2) for every $\epsilon > 0$ there is a $\delta > 0$ with $\sup_{n\geqslant 1} \int_B |X_n| d\,\mathbf{P} < \epsilon$ whenever $\mathbf{P}(B) < \delta$.

We begin with 1). We pick an arbitrary $n \geqslant 1$ and look at

$$\begin{aligned}\mathbf{E}|X_n| &= \|X_n\|_1 = \|X_n - X_N + X_N\|_1 \\ &\leqslant \|X_n - X_N\|_1 + \|X_N\|_1\end{aligned}$$

Now we pick N so that $\|X_n - X_N\|_1 \leqslant 1$ for all $n \geqslant N$. Hence, we get

$$\mathbf{E}|X_n| \leqslant \|X_n - X_N\|_1 + \|X_N\|_1 \leqslant 1 + \|X_N\|_1$$

Hence, taking the supremum over all $n \geqslant N$, we will get

$$\sup_{n\geqslant N} \mathbf{E}|X_n| \leqslant 1 + \|X_N\|_1$$

This again implies

$$\sup_{n\geqslant 1} \mathbf{E}|X_n| \leqslant \max\{\|X_1\|_1, \|X_2\|_1, \ldots, \|X_{N-1}\|_1, 1 + \|X_N\|_1\} < \infty$$

Consider

$$\begin{aligned}\int_B |X_n| d\,\mathbf{P} &= \int_B |X_n - X_N + X_N| d\,\mathbf{P} \\ &\leqslant \int_B |X_n - X_N| + |X_N| d\,\mathbf{P} \\ &\leqslant \int_\Omega |X_n - X_N| d\,\mathbf{P} + \int_B |X_N| d\,\mathbf{P} \\ &= \|X_n - X_N\|_1 + \int_B |X_N| d\,\mathbf{P}\end{aligned}$$

We now combine the two estimates. Given $\epsilon > 0$, use the Cauchy property to pick N with

$$\|X_n - X_N\|_1 < \frac{\epsilon}{2} \qquad \text{for all } n \geqslant N$$

Also, recall from the last midterm question that for a single L^1 random variable the integral is absolutely continuous: for each of the finitely many variables $X_1, \ldots, X_N$ there is a $\delta_m > 0$ such that $\int_B |X_m| d\,\mathbf{P} < \frac{\epsilon}{2}$ whenever $\mathbf{P}(B) < \delta_m$. Put

$$\delta := \min_{m \leqslant N} \delta_m > 0$$

which is positive because the minimum is over finitely many positive numbers. Let B be any event with $\mathbf{P}(B) < \delta$. For $n \geqslant N$ the estimate above gives

$$\int_B |X_n| d\,\mathbf{P} \leqslant \|X_n - X_N\|_1 + \int_B |X_N| d\,\mathbf{P} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and for $n \leqslant N$ we get $\int_B |X_n| d\,\mathbf{P} < \frac{\epsilon}{2} < \epsilon$ directly from the choice of δ. Hence

$$\sup_{n \geqslant 1} \int_B |X_n| d\,\mathbf{P} \leqslant \epsilon \qquad \text{whenever } \mathbf{P}(B) < \delta$$

which is 2). Together with 1) this shows $\{X_n : n \geqslant 1\}$ is uniformly integrable.

It remains to check that (X_n) converges in probability. By Chebyshev's inequality,

$$\mathbf{P}(|X_n - X_m| > \delta') \leqslant \frac{\mathbf{E}|X_n - X_m|}{\delta'} = \frac{\|X_n - X_m\|_1}{\delta'} \longrightarrow 0 \qquad \text{as } n, m \to \infty$$

for every $\delta' > 0$, so (X_n) is a Cauchy sequence in probability, and therefore converges in probability, by the theorem on Cauchy sequences in probability from the subsection on convergence in probability.

(3) to (1): Assume X_n converge in probability. Then there is a subsequence X_{n_k} such that

$$X_{n_k} \longrightarrow X \quad \text{almost surely}$$

Then

$$\mathbf{E}|X| = \mathbf{E} \liminf_{k \to \infty} |X_{n_k}| \leqslant \liminf_{k \to \infty} \mathbf{E}|X_{n_k}| \leqslant \sup_{n \geqslant 1} \mathbf{E}|X_n| < \infty$$

where the first equality holds because $|X_{n_k}| \to |X|$ almost surely along the subsequence, the first inequality is Fatou's lemma, and the final bound is condition 1) of uniform integrability. In particular $X \in L^1(\mathbf{P})$.

Next, note that

$$|X_n - X| \leqslant |X_n| + \underbrace{|X|}_{\in L^1(\mathbf{P})}$$

Because $|X_n|$ is uniformly integrable, then $|X_n| + |X|$ is also uniformly integrable because $|X| \in L^1(\mathbf{P})$. Indeed, $\{|X_n| : n \geqslant 1\}$ is uniformly integrable by hypothesis, and the single random variable $|X|$ is uniformly integrable because $X \in L^1(\mathbf{P})$, as shown above. Both conditions of the second definition are preserved under sums: the first moments add, so $\sup_{n \geqslant 1} \mathbf{E}(|X_n| + |X|) \leqslant \sup_{n \geqslant 1} \mathbf{E}|X_n| + \mathbf{E}|X| < \infty$, and given $\epsilon > 0$, a δ that works for both families with $\frac{\epsilon}{2}$ in place of ϵ works for the sum. Hence, $|X_n - X|$ which is dominated by $|X_n| + |X|$ is uniformly integrable. To summarize, we have

- $|X_n - X|$ is uniformly integrable.

- $|X_n - X|$ converges to 0 in probability, i.e.

$$\mathbf{P}(|X_n - X| > \epsilon) \to 0 \quad \text{as } n \to \infty$$

For $\|X_n - X\|_1$,

$$\|X_n - X\|_1 = \int |X_n - X| d\,\mathbf{P} = \int_{\{|X_n - X| > \epsilon\}} |X_n - X| d\,\mathbf{P} + \int_{\{|X_n - X| \leqslant \epsilon\}} |X_n - X| d\,\mathbf{P}$$

For the second integral term, under the event

$$\int_{\{|X_n - X| \leqslant \epsilon\}} |X_n - X| d\,\mathbf{P} \leqslant \int_{\{|X_n - X| \leqslant \epsilon\}} \epsilon d\,\mathbf{P} \leqslant \int_{\Omega} \epsilon d\,\mathbf{P} = \epsilon$$

For the first integral term, we use uniformly integrability. From convergence in probability, we have the measure of the set $\{|X_n - X| > \epsilon\}$ is small. Then by uniformly integrability, the integral over this set will also be small, i.e.

$$\int_{\{|X_n - X| > \epsilon\}} |X_n - X| d\,\mathbf{P} \leqslant \epsilon \qquad \text{for all large } n$$

Combining the two estimates, for all n large enough

$$\|X_n - X\|_1 \leqslant \epsilon + \epsilon = 2\epsilon$$

Since $\epsilon > 0$ was arbitrary, $\|X_n - X\|_1 \to 0$, that is, $X_n \to X$ in $L^1(\mathbf{P})$. This completes the proof. □

Remark 6.32. *If (X_n) converges in $L^p(\mathbf{P})$, that is, there is a random variable X such that $\|X_n - X\|_p \to 0$. Then $X \in L^p(\mathbf{P})$.*

- *If (X_n) is a Cauchy sequence in $L^p(\mathbf{P})$, then the limit is in $L^p(\mathbf{P})$.*

For a Normed Vector Space, if any Cauchy sequence has a limit in the same space. We call it a Banach space.

References

[Res99] Sidney I. Resnick. *A Probability Path.* Birkhäuser, Boston, 1999.