Advanced Probability Theory

Math5411
HKUST

Kani Chen (Instructor)

Chapter 1. Law of Large Numbers

§ 1.1. σ-algebra, measure, probability space and random variables.

This section lays the necessary rigorous foundation for probability as a mathematical theory. It begins with sets, relations among sets, measurement of sets and functions defined on the sets.

Example 1.1. (A PROTOTYPE OF PROBABILITY SPACE.) Drop a needle blindly on the interval $[0, 1]$. The needle hits interval $[a, b]$, a sub-interval of $[0, 1]$ with chance $b - a$. Suppose A is any subset of $[0, 1]$. What's the chance or length of A?

Here, we might interprete the largest set $\Omega = [0, 1]$ as the "universe". Note that not all subsets are "nice" in the sense that their volume/length can be properly assigned. So we first focus our attention on certain class of "nice" subsets.

To begin with, the "Basic" subsets are all the sub-intervals of $[0, 1]$, which may be denoted as $[a, b]$, with $0 \leq a \leq b \leq 1$. Denote $\mathcal{B}$ as the collection of all subsets of $[0, 1]$, which are generated by all basic sets after $finite$ set operations. $\mathcal{B}$ is called an $algebra$ of Ω.

It can be proved that any set in $\mathcal{B}$ is a finite union of disjoint intervals (closed, open or half-closed).

Still, $\mathcal{B}$ is not rich enough. For example, it does not contain the set of all rational numbers. More importantly, the limits of sets in $\mathcal{B}$ are often not in $\mathcal{B}$. This is serious restrictions of mathematical analysis.

Let $\mathcal{A}$ be the collection of all subsets of $[0, 1]$, which are generated by all "basic" sets after *countably* many set operations. $\mathcal{A}$ is called Borel σ-algebra of Ω. Sets in $\mathcal{A}$ are called *Borel sets.* Limits of sets in $\mathcal{A}$ are still in $\mathcal{A}$. $(\Omega, \mathcal{A})$ is a *measurable space.*

Borel measure: any set A in $\mathcal{A}$ can be assigned a volume, denoted as $\mu(A)$, such that

(i). $\mu([a, b]) = b - a$.

(ii). $\mu(A) = \lim \mu(A_n)$ for any sequence of Borel sets $A_n \uparrow A$.

Lebesgue measure (1901): Completion of Borel σ-algebra by adding all subsets of Borel measure 0 sets, denoted as $\mathcal{F}$. Sets with measure 0 are called null sets.

Why should Borel measure or Lebesgue measure exist in general?

Caratheodory's extension theorem: extending a (σ-finite) measure on an algebra $\mathcal{B}$ to the σ-algebra $\mathcal{A} = \sigma(\mathcal{B})$.

$\Omega = [0, 1]$ (the universe).

$\mathcal{B}$: an algebra (finite set operations) generated by subintervals.

$\mathcal{A}$: the Borel σ-algebra, is a σ-algebra, generated by subintervals.

$\mathcal{F}$: completion of $\mathcal{A}$, a σ-algebra, generated by $\mathcal{A}$ and null sets.

$(\Omega, \mathcal{B}, \mu)$ does not form a probability space,

$(\Omega, \mathcal{A}, \mu)$ forms a probability space.

$(\Omega, \mathcal{F}, \mu)$ forms a probability space.

Sets and set operations:

Consider Ω as the "universe", (*Beyond which is nothing.*) Write $\Omega = \{\omega\}$, ω denotes an member of the set, called *element*. Let A and B: be two subsets of Ω, called "events".

The set operations are:

intersection: $\cap$, $A \cap B$: both A and B (happens).

union: $\cup$, $A \cup B$: either A or B (happens).

complement: $A^c = \Omega \setminus A$: everything except for A, or A does not happen.

minus: $A \setminus B = A \cap B^c$: A but not B.

An elementary theorem about set operation is

DeMorgan's identity:

$$\left(\cup_{j=1}^{\infty} A_j\right)^c = \cap_{j=1}^{\infty} A_j^c, \qquad \left(\cap_{j=1}^{\infty} A_j\right)^c = \cup_{j=1}^{\infty} A_j^c.$$

In particular, $(A \cup B)^c = (A^c \cap B^c)$, i.e., $(A \cap B)^c = (A^c \cup B^c)$.

Remark. Intersection can be generated by complement and union; and union can be generated by complement and intersection.

Relation: $A \subset B$, if $\omega \in A$ ensures $\omega \in B$.

A sequence of sets $\{A_n : n \geq 1\}$ is called *increasing (decreasing)* if $A_n \subset A_{n+1}$ ($A_n \supset A_{n+1}$.)

$A = B$ if and only if $A \subset B$ and $B \subset A$.

Indicator functions. (A very useful tool to translate set operation into numerical operation)

The relation and operation of sets are equivalent to the indication set functions. For any subset $A \subset \Omega$, define its indicator function as

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{Otherwise.} \end{cases}$$

The indicator function is a function defined on Ω.

Set operations vs. function operations:

$$\begin{aligned} A \subset B &\iff 1_A \leq 1_B. \\ A \cap B &\iff 1_A \times 1_B = 1_{A \cap B} = \min(1_A, 1_B). \\ A^c = \Omega \setminus A &\iff 1 - 1_A = 1_{A^c}. \\ A \cup B &\iff 1_{A \cup B} = 1_A + 1_B, \quad \text{if } A \cap B = \emptyset \\ &\iff 1_{A \cup B} = \max(1_A, 1_B). \end{aligned}$$

Set limits.

There are two limits of sets: upper limit and low limit.

$$\begin{aligned} &\limsup A_n \equiv \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k = \{A_n \text{ infinitely occurs.}\} \\ &1_{\limsup A_n} = \limsup 1_{A_n} \end{aligned}$$

$\omega \in \limsup A_n$ if and only if ω belongs to infinitely many A_n.

Lower limit.

$$\begin{aligned} &\liminf A_n \equiv \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k \\ = \; &\{A_n \text{ always occurs except for finite number of times.}\} \\ &1_{\liminf A_n} = \liminf 1_{A_n} \end{aligned}$$

$\omega \in \liminf A_n$ if and only if ω belongs to all but finitely many A_n.

We say the set limit of $A_1, A_2, \ldots$ exists if their lower limit is the same as the upper limit.

Algebra and σ-algebra

$\mathcal{A}$ is a non-empty collection (set) of subsets of Ω.

Definition. $\mathcal{A}$ is called an *algebra* if

(i). $A^c \in \mathcal{A}$ if $A \in \mathcal{A}$;

(ii). $A \cup B \in \mathcal{A}$ if $A, B \in \mathcal{A}$.

$\mathcal{A}$ is called an σ-algebra if, (ii) is strengthened as,

(iii). $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$ if $A_n \in \mathcal{A}$ for $n \geq 1$.

An algebra is closed for (finite) set operations. $\Omega \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$.

A σ-algebra is closed for *countable* operations.

$(\Omega, \mathcal{A})$ is called a measurable space, if $\mathcal{A}$ is a σ-algebra of Ω.

Measure, measure space and probability space.

$\mathcal{A}$, containing $\emptyset$, is a non-empty collection (set) of subsets of Ω. μ is a nonnegative set function on $\mathcal{A}$.

μ is called a *measure*, if

(i). $\mu(\emptyset) = 0$.

(ii). $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$ if $A, A_1, A_2, \ldots$ are all in $\mathcal{A}$ and $A_1, A_2, \ldots$ are disjoint.

$(\Omega, \mathcal{A}, \mu)$ is called a measure space, if μ is a measure on $\mathcal{A}$ and $\mathcal{A}$ is a σ-algebra of Ω.

$(\Omega, \mathcal{A}, P)$ is called a probability space if $(\Omega, \mathcal{A}, P)$ is a measure space and $P(\Omega) = 1$.

For probability space $(\Omega, \mathcal{A}, P)$, Ω is called sample space, every A in $\mathcal{A}$ is an event, and $P(A)$ is the probability of the event, the chance that it happens.

Random variable (r.v.).

Loosely speaking, given a probability space $(\Omega, \mathcal{F}, P)$, a random variable (r.v.) X is defined as a real-valued function of Ω, satisfying certain measurability condition. Loosely speaking, viewing $X = X(\omega)$ as a mapping from Ω to R, the real line, then $X^{-1}(B)$ must be in $\mathcal{F}$ for all Borel sets B. (Borel sets on real line are the σ-algebra generated by intervals, i.e., the sets generated by countable operations on intervals).

A random variable X defined on a probability space $(\Omega, \mathcal{A}, P)$ is a function defined on Ω, such that $X^{-1}(B) \in \mathcal{A}$ for every interval B on $[-\infty, \infty]$, where $X^{-1}(B) = \{\omega : X(\omega) \in B\}$. (*We need to identify its probability.*)

$X^{-1}(B)$ is called the inverse image of B.

$X = X(\cdot)$ can be viewed as a map or transformation from $(\Omega, \mathcal{A})$ to $(R, \mathcal{B})$, where $R = [-\infty, \infty]$ and $\mathcal{B}$ is the σ-algebra generated by the intervals in R.

X is a *measurable map/transformation* since $X^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$ (DIY.)

Because $\mathcal{A}$ is a σ-algebra, the upper and lower limits of X_n is a r.v. if X_n are r.v.s., and the algebraic operations: $+, -, \times, /$, of r.v.s are still r.v.s.

Measurable map and random vectors.

$f(\cdot)$ is called a *measurable map/transformation/function* from a measurable space $(\Omega, \mathcal{A})$ to another measurable space $(S, \mathcal{S})$, if $f^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{S}$. i.e. $\{w : f(w) \in B\} \in \mathcal{A}$.

X is called a random vector of p dimension if it is a measurable map from a probability space $(\Omega, \mathcal{A}, P)$ to $(R^p, \mathcal{B}^p)$, where $\mathcal{B}^p$ is the Borel σ-algebra in p dimensional real space, $R^p = [-\infty, \infty]^p$.

Proposition 1.1 ((2.3) in the textbook.) If $X = (X_1, ..., X_p)$ is a random vector of p dimension on a probability space $(\Omega, \mathcal{A}, P)$, and $f(\cdot)$ is measurable function from $(R^p, \mathcal{B}^p)$ to $(R, \mathcal{B})$, then $f(X)$ is a random variable.

Proof. For any Borel set $B \in \mathcal{B}$,

$$\{\omega : f(X(\omega)) \in B\} = \{\omega : X(\omega) \in f^{-1}(B)\} \in \mathcal{A}$$

since $f^{-1}(B) \in \mathcal{B}^p$. □

Proposition 1.2 ((2.5) in the textbook.) If $X_1, X_2, ...$ are r.v.s. So are

$$\inf_n X_n, \quad \sup_n X_n \quad \liminf_n X_n \quad \text{and} \quad \limsup_n X_n.$$

Proof. Let the probability space be $(\Omega, \mathcal{A}, P)$. For any x,

$$\{\omega : \inf_n X_n(\omega) \geq x\} = \cap_n \{\omega : X_n(\omega) \geq x\} \in \mathcal{A};$$

$$\{\omega : \sup_n X_n(\omega) \leq x\} = \cap_n \{\omega : X_n(\omega) \leq x\} \in \mathcal{A};$$

$$\{\liminf_n X_n > x\} = \cup_n \{\inf_{k \geq n} X_k > x\} \in \mathcal{A};$$

$$\{\limsup_n X_n < x\} = \cup_n \{\sup_{k \geq n} X_k < x\} \in \mathcal{A}.$$

Therefore, $\inf_n X_n, \sup_n X_n, \liminf_n X_n$ and $\limsup_n X_n$ are r.v.s. □

Proposition 1.3 Suppose X is a map from a measurable space $(\Omega, \mathcal{A})$ to another measurable space $(\mathbf{S}, \mathcal{S})$. If $X^{-1}(C) \in \mathcal{A}$ for every $C \in \mathcal{C}$ and $\mathcal{S} = \sigma(\mathcal{C})$. Then, X is a measurable map, i.e., $X^{-1}(S) \in \mathcal{A}$ for every $S \in \mathcal{S}$. In particular, when $(\mathbf{S}, \mathcal{S}) = ([-\infty, \infty], \mathcal{B})$, $X^{-1}([-\infty, x]) \in \mathcal{A}$ for every x is enough to ensure X is a r.v..

Proof. Note that $\sigma(\mathcal{C})$, the σ-algebra generated by $\mathcal{C}$, is defined mathematically as the smallest σ-algebra containing $\mathcal{C}$.

Set $\mathcal{B}^* = \{B \in \mathcal{S} : X^{-1}(B) \in \mathcal{A}\}$.

We first show $\mathcal{B}^*$ is a σ-algebra. Observe that

(i). for any $B \in \mathcal{B}^*$, $X^{-1}(B) \in \mathcal{A}$ and, therefore, $X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{A}$;

(ii). for any $B_n \in \mathcal{B}^*$, $X^{-1}(B_n) \in \mathcal{A}$ and $X^{-1}(\cup_n B_n) = \cup_n X^{-1}(B_n) \in \mathcal{A}$.

Consequently, $\mathcal{B}^*$ is a σ-algebra. Since $\mathcal{C} \subset \mathcal{B}^* \subset \mathcal{S}$, it follows that $\mathcal{B}^* = \mathcal{S}$. □

Summary of Section 1.1

σ-algebra: collection of sets which is closed under countably many set operations.

Probability space: The trio $(\Omega, \mathcal{A}, P)$ with $\mathcal{A}$ as a σ-algebra of Ω and P a set function such that

(i) $0 \leq P(A) \leq 1$ for any $A \in \mathcal{A}$ and $P(\Omega) = 1$.

(ii). $P(\cup_n A_n) = \sum_n P(A_n)$ for countable disjoint $A_n \in \mathcal{A}$.

A *random variable* X is a function/map on Ω with value in $[-\infty, \infty]$ such that $\{X \in [-\infty, x]\} \in \mathcal{A}$.

$F(x) \equiv P(X \leq x)$ is called (cumulative) distribution function of X.

The moral is to ensure calibration of the distribution of r.v.s and validity of algebraic operation and limits of r.v.s.

indicator function as a useful tool.

DIY Exercises:

EXERCISE 1.1 Show $1_{\liminf A_n} = \liminf 1_{A_n}$ and DeMorgen's identity.

EXERCISE 1.2 Show that, the so called "countable additivity" or "σ-additivity", ($P(\cup_n A_n) = \sum_n P(A_n)$ for countable disjoint $A_n \in \mathcal{A}$), is equivalent to "finite additivity" plus "continuity" (if $A_n \downarrow \emptyset$, then $P(A_n) \to 0$.)

EXERCISE 1.3 *(Completion of a Probability space)* Let $(\Omega, \mathcal{F}, P)$ be a probability space. Define

$$\bar{\mathcal{F}} = \{A : P(A \setminus B) + P(B \setminus A) = 0, \text{ for some} B \in \mathcal{F}\},$$

And for each $A \in \bar{\mathcal{F}}$, $P(A)$ is defined as $P(B)$ for the B given above. Prove that $(\Omega, \bar{\mathcal{F}}, P)$ is also a probability space. (Hint: need to show that $\mathcal{F}$ is a σ-algebra and that P is a probability measure.)

EXERCISE 1.4 If X_1 and X_2 are two r.v.s, so is $X_1 + X_2$. (Hint: cite Propositions 1.1 and 1.3)