ST5215 Advanced Statistical Theory, Lecture 1

HUANG Dongming

National University of Singapore

8 Aug 2020

Pre-Requisite: ST2131 and ST2132 or Departmental approval
Modes: 1.5h lecture + 0.5h tutorial (tentative)
Online teaching:

- Disable your webcam to release network demand
- For questions: "raise hands" or text chat in Zoom
- Keep your microphone muted unless speaking

Office Hours: Tue/Thur 16:00 - 17:00PM via Zoom (tentative)
Textbook:

- Jun Shao, Mathematical Statistics, Springer, 2nd Edition, 2003.

References:

- L. Wasserman, *All of Statistics*.
- G. Casella and R. L. Berger, *Statistical Inference* 2nd Edition.
- P. J. Bickel and K. A. Doksum, *Mathematical statistics: basic ideas and selected topics*.
- R.W. Keener, *Theoretical Statistics: Topics for a Core Course*.
- E.L. Lehmann G. Casella, *Theory of Point Estimation*.

Grading Policy: 4 homework ($5\% \times 4 = 20\%$), 1 midterm (30%) and 1 final (50%).

Homework:

- Submitted to **LumniNUS/Files/Submissions/HWXX** by 2pm on the due date.
- Late submission will never be accepted
- Homework submission must be typeset (LaTeX, Mathpix ...)
- You can discuss with your classmates on homework but you **must** write up your solutions independently in your own words
- You will get zero for a submission if someone's is identical or very similar to yours.

Exam:

- Midterm: Thur, 01 Oct 2020, 14:00-16:00
- Final: Wed, 02 Dec 2020, 13:00 - 15:00
- Closed-book (No books, notes, or electronics)
- Online (you will be supervised via Zoom)

Topics and chapters:

- Probability theory: Ch 1.1 – 1.4.3
- Models and statistics: Ch 2.1 – 2.2
- Point estimators (the method of moments, the method of maximum likelihood)
- Decision theory: Ch 2.3
- UMVUE: Ch 3.1.1 – 3.1.3
- Bayes estimation and risk: Ch 4.1, 4.3
- Asymptotics: Ch 1.5, 2.5, 4.5
- Other topics (Linear model: Ch 3.3; U-statistics: Ch 3.2)

Measure and σ-fields

- Let Ω be a set of objects. *outcome space*; *sample space*.
- Want to measure the "size" of subsets of Ω
  - If $\Omega = \mathcal{R}$ and $A = [a, b]$, then the size of A is naturally its length $b - a$.
  - If $\Omega = \mathcal{R}^2$ and A is a polygon, its size is defined as its area
- For a general Ω, we need to specify which subsets are measurable (i.e., define what can be an "event")

Definition

A collection $\mathcal{F}$ of subsets of a set Ω is called a *σ-field* (or *σ-algebra*) if

1. $\emptyset \in \mathcal{F}$,
2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
3. if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots,$then $\bigcup A_i \in \mathcal{F}$.

A pair $(\Omega, \mathcal{F})$ of a set Ω and a σ-field $\mathcal{F}$ on Ω is called a *measurable space*.

- In probability, a σ-field is a collection of events of interest
- There can be many σ-fields on Ω
- e.g.: $\Omega = \{1, 2\}$, $\mathcal{F}_1 = \{\emptyset, \Omega\}$, $\mathcal{F}_2 = \{\emptyset, \{1\}, \{2\}, \Omega\}$

Exercise: Suppose $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measure spaces, is $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_i \in \mathcal{F}_i\}$ a σ-field on $\Omega_1 \times \Omega_2$? Why?
Counter example:

- $\Omega_1 = \{a, b\}, \Omega_2 = \{c, d\}, \mathcal{F}_i = 2^{\Omega_i}$ (the power set)
- $K = \{a\} \times \{c\}$ and $L = \{b\} \times \{d\}$.
- Note that by definition of the Cartesian product, $K = \{a\} \times \{c\} = \{(a, c)\}$.
- $K, L \in \mathcal{F}_1 \times \mathcal{F}_2$
- But $K \cup L = \{(a, c), (b, d)\} \notin \mathcal{F}_1 \times \mathcal{F}_2$

The smallest σ-field

- Given a collection $\mathcal{C}$ of subsets, there exists a σ-field $\mathcal{F}$ such that $\mathcal{C} \subset \mathcal{F}$ and if $\mathcal{E}$ is a σ-field that also contains $\mathcal{C}$, then $\mathcal{F} \subset \mathcal{E}$. (See Ex 1.6.2)
  Such $\mathcal{F}$ is the *smallest* σ-field that contains $\mathcal{C}$. It is denoted by $\sigma(\mathcal{C})$ and is called the σ-field generated by $\mathcal{C}$
- Particularly, when $\Omega = \mathcal{R}$ and $\mathcal{O}$ is all open sets, we call $\sigma(\mathcal{O})$ the Borel σ-field, denoted by $\mathcal{B}$. Elements in $\mathcal{B}$ are called Borel sets.
- Similarly, when $\Omega = \mathcal{R}^d$ and $\mathcal{O}^d$ is all open sets in $\mathcal{R}^d$, $\sigma(\mathcal{O}^d)$ is the d-dimensional Borel σ-field, denoted by $\mathcal{B}^d$

Exercise

Show that $\mathcal{B}$ is also the σ-field generated by all closed subset of $\mathcal{R}$.

- Let $\mathcal{C}$ be the collection of all closed subset of $\mathcal{R}$
- For any $B \in \mathcal{O}$, we have $B^c \in \mathcal{C}$, so $B^c \in \sigma(\mathcal{C})$, so $B \in \sigma(\mathcal{C})$; i.e. $\mathcal{O} \subset \sigma(\mathcal{C})$
- So $\sigma(\mathcal{O}) \subset \sigma(\mathcal{C})$
- For any $A \in \mathcal{C}$, we have $A^c \in \mathcal{O}$, so $A^c \in \sigma(\mathcal{O})$, so $A \in \sigma(\mathcal{O})$; i.e. $\mathcal{C} \subset \sigma(\mathcal{O})$,
- So $\sigma(\mathcal{C}) \subset \sigma(\mathcal{O})$
- That is, $\sigma(\mathcal{C}) = \sigma(\mathcal{O})$

Measure

Definition

A (positive) measure ν on a measurable space $(\Omega, \mathcal{F})$ is a non-negative function $\nu : \mathcal{F} \to \mathcal{R}$ such that

1. (non-negativity) $0 \leq \nu(A) \leq \infty$ for all $A \in \mathcal{F}$,
2. (empty is zero) $\nu(\emptyset) = 0$, and
3. (σ-additivity): $\sum_{i=1}^{\infty} \nu(A_i) = \nu\left(\bigcup_{i=1}^{\infty} A_i\right)$ if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ and $A_1, A_2, \ldots$ are disjoint.

- The triplet $(\Omega, \mathcal{F}, \nu)$ is called a *measure space*.
- If $\nu(\Omega) = 1$, it is also called a *probability space*, and the number $\nu(A)$ is interpreted as the probability of the event A to happen. And ν is often denoted by P or $\Pr$.

Examples

- Counting measure: $\nu(A) =$ the number of elements in A, $\forall A \subset \Omega$. (Note: $\nu(A)$ can be ∞)
- Lebesgue measure: There exists a unique measure m on $(\mathcal{R}, \mathcal{B})$ that satisfies $m([a, b]) = b - a$ for any $a < b$.
  When $\mathcal{R}$ is mentioned, it is by default endowed with the Borel σ-field and Lebesgue measure unless explicitly mentioned

σ-finite

- A measure ν on $(\Omega, \mathcal{F})$ is said to be *σ-finite* if there exists a sequence of measurable sets $A_1, A_2, \ldots$ such that $\bigcup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i.
- The Lebesgue measure is σ-finite: $\mathcal{R} = \bigcup_{i=1}^{\infty} A_i$ with $A_i = [-i, i]$ and $m(A_i) = 2i < \infty$.
- All finite measures are σ-finite.

σ-finite is required in some important theorems (Radon-Nikodym, Fubin's), and we focus on σ-finite measures in this course.

Product measure

How to introduce a measure on a *product space* $\Omega_1 \times \cdots \times \Omega_d$, like, $\mathcal{R}^d = \mathcal{R} \times \cdots \times \mathcal{R}$?

- For a product space $\Omega_1 \times \cdots \times \Omega_d$, where each Ω_i is endowed with a σ-field $\mathcal{F}_i$, the σ-field generated by $\prod_{i=1}^d \mathcal{F}_i = \{A_1 \times \cdots A_d : A_i \in \mathcal{F}_i\}$ is called the *product σ-field*.
- One can show: for $\mathcal{R}^d = \mathcal{R} \times \cdots \times \mathcal{R}$, the product σ-field is the same as $\mathcal{B}^d$.

Proposition

Suppose $(\Omega_i, \mathcal{F}_i, \nu_i)$, $i = 1, 2, \ldots, d$, are measure spaces and $\nu_1, \ldots, \nu_d$ are all σ-finite. There exists a unique σ-finite measure on the product σ-field, denoted by $\nu_1 \times \cdots \times \nu_d$, such that

$$\nu_1 \times \cdots \times \nu_d(A_1 \times \cdots \times A_d) = \prod_{i=1}^d \nu_i(A_i) \tag{1}$$

for all $A_i \in \mathcal{F}_i$.

Example (Lebesgue measure on $\mathcal{R}^d$)

- For $\mathcal{R}^d$, we use the Lebesgue measure m on $(\mathcal{R}, \mathcal{B})$ to define a unique product measure $m \times \cdots \times m$.
- It is called the Lebesgue measure on $(\mathcal{R}^d, \mathcal{B}^d)$.
- It is the standard/canonical measure on $\mathcal{R}^d$.
  Again, without otherwise explicitly mentioned, $\mathcal{R}^d$ is endowed with $\mathcal{B}^d$) and Lebesgue measure.

Measurable functions

Recall that $f : \Omega \to \Lambda$ is a *continuous function* between two topological spaces Ω and Λ, if for every open subset A in Λ, $f^{-1}(A) = \{x \in \Omega : f(x) \in A\}$ is an open set in Ω.

Definition

Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces and $f : \Omega \to \Lambda$ a function. The function f is called a *measurable function* from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ if $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{G}$.

- If f is measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{R}, \mathcal{B})$, then f is said to be *Borel measurable* or is called a *Borel function* on $(\Omega, \mathcal{F})$.
- If $f : \mathcal{R} \mapsto \mathcal{R}$ is continuous, then it is Borel measurable (left for homework)
- If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G}) = \{f^{-1}(A) : A \in \mathcal{G}\}$ is a sub-σ-field of $\mathcal{F}$ (left for exercise) Usually denoted by $\sigma(f)$ and called the σ-field generated by f.

Examples

- The indicator function I_A for a measurable set A is a Borel function. Here,

$$I_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A. \end{cases} \tag{2}$$

- A *simple function* of the form

$$f(\omega) = \sum_{i=1}^{k} c_i I_{A_i}(\omega), \tag{3}$$

is a Borel function for any real numbers $c_1, \ldots, c_k$ and measurable sets $A_1, \ldots, A_k$.

Some results

- **Richness:** Proposition 1.4 in Shao 2003 shows that many operations (addition, multiplication, division, composition, sup, limit) of measurable functions preserve the measurability
- **Approximation by simple functions:** For any non-negative Borel function f, there exists a sequence of simple functions φ_n's such that $0 \le \varphi_1 \le \varphi_2 \le \cdots \le f$ and $\lim_n \varphi_n(x) = f(x)$ for every $x \in \Omega$
  Let
  $$\varphi_n = \sum_{i=0}^{n2^n-1} \cdot \frac{i}{2^n} I_{\frac{i}{2^n} \le f < \frac{i+1}{2^n}} + n \cdot I_{f \ge n}, \forall n \in \mathcal{N}$$
  Check
  1. φ_n's are simple functions
  2. $0 \le \varphi_1(x) \le \varphi_2(x) \le \cdots \le f(x), \forall x \in \Omega$
  3. If $f(x) = \infty$, then $\varphi_n(x) = n \to \infty$
  4. If $n > f(x)$ then $f(x) - \varphi_n(x) \le \frac{1}{2^n}$, which implies $\lim_n \varphi_n(x) = f(x)$

Notations for probability

In probability theory, a measurable function is also called a *random element*, and often denoted by capital letters X, Y, Z, $\ldots$.
If X is real-valued, then it is called a *random variable* (r.v.); if it is vector-valued, then it is called a *random vector*.

Integration

We introduce the concept of Lebesgue integral by three steps:

- step 1: define integral of "simple" functions – easy case
- step 2: define integral of non-negative Borel functions by approximation of simple functions
- step 3: define integral of all Borel functions

Integral of non-negative simple functions

- Suppose $f : \Omega \to \mathcal{R}$ is a simple non-negative function: $f(x) = \sum_{i=1}^{k} c_i I_{A_i}(x)$ for $A_i \in \mathcal{F}$ and $c_i \geq 0$.
- Define the integral of f as

$$\int f \, \mathrm{d}\nu = \sum_{i=1}^{k} c_i \nu(A_i)$$

- Well defined even when $\nu(A_i) = \infty$ for some A_i: $c \cdot \infty = \infty$ when $c > 0$ and $c \cdot \infty = 0$ when $c = 0$.
- Note that $\int f \, \mathrm{d}\nu = \infty$ is possible and allowed.

Integral of a non-negative Borel function

- Any Borel function f can be approximated by a sequence of simple functions
- Use the integrals of these simple functions as proxy
- Let $\mathcal{S}_f$ be the collection of all non-negative simple functions such that $g \leq f$ if $g \in \mathcal{S}_f$.
- Define the integral of f as

$$\int f \, \mathrm{d}\nu = \sup\{\int g \, \mathrm{d}\nu : g \in \mathcal{S}_f\}. \tag{4}$$

Exercise: Show that if a sequence of simple functions φ_n's satisfies $0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$ and $\lim_n \varphi_n = f$, then

$$\int f \, \mathrm{d}\nu = \lim_n \int \varphi_n \, \mathrm{d}\nu$$

Integral of arbitrary Borel functions

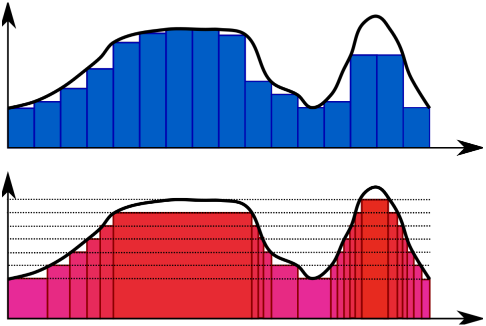
- Divide f into two parts: $f = f_+ - f_-$
  - Positive part: $f_+(x) = \max\{f(x), 0\}$
  - Negative part: $f_-(x) = -\min\{f(x), 0\} = \max\{-f(x), 0\}$. Note that the negative part is also a nonnegative function
- Define $\int f \, \mathrm{d}\nu$ as

$$\int f \, \mathrm{d}\nu = \int f_+ \, \mathrm{d}\nu - \int f_- \, \mathrm{d}\nu \tag{5}$$

if at least one of $\int f_+ \, \mathrm{d}\nu$ and $\int f_- \, \mathrm{d}\nu$ is finite.
  - if yes, we say the integral of f *exists*
  - if not, then we can the integral of f *does not exist*
- When both $\int f_+ \, \mathrm{d}\nu$ and $\int f_- \, \mathrm{d}\nu$ are finite, we say f is *integrable*

Comparison with Riemann Integral

- Both are defined as the limit of a sequence of sums.
- Riemann integral: dividing the X-domain
- Lebesgue integral: dividing the Y-range.



- If a function is Riemann integrable on a closed interval, then its Riemann integral will coincide with its Lebesgue integral.
- A bounded function on a closed interval is Riemann integrable if and only if the set of its discontinuities has Lebesgue measure 0.

Some notations

Integral over a subset $A \in \mathcal{F}$

- I_A is measurable, and so is the product $I_A f$.
- If the integral of $I_A f$ exists, then we can define

$$\int_A f \, \mathrm{d}\nu = \int I_A f \, \mathrm{d}\nu. \tag{6}$$

- Notation: $\int f \, \mathrm{d}\nu = \int_\Omega f \, \mathrm{d}\nu = \int f(x) \, \mathrm{d}\nu(x) = \int f(x) \nu(\, \mathrm{d}x)$

For a probability measure P and a r.v. X, the expectation of X is $\mathbb{E}X = \mathbb{E}(X) = \int X \, \mathrm{d}P$

Exercises

1. Let $(\Omega, \mathcal{F})$ be a measurable space and $C \in \mathcal{F}$. Show that $\mathcal{F}_C = \{C \cap A : A \in \mathcal{F}\}$ is a σ-field on C.
2. If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G}) = \{f^{-1}(A) : A \in \mathcal{G}\}$ is a sub-σ-field of $\mathcal{F}$
3. Show that if a sequence of simple functions φ_n's satisfies $0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$ and $\lim_n \varphi_n = f$, then

$$\int f \, \mathrm{d}\nu = \lim_n \int \varphi_n \, \mathrm{d}\nu$$

Exercises 1

Let $(\Omega, \mathcal{F})$ be a measurable space and $C \in \mathcal{F}$. Show that $\mathcal{F}_C = \{C \cap A : A \in \mathcal{F}\}$ is a σ-field on C.
Proof:

- Since $\emptyset \in \mathcal{F}$, so $\emptyset = C \cap \emptyset \in \mathcal{F}$
- If $C \cap A \in \mathcal{F}_C$, where $A \in \mathcal{F}$.
  Then $A^c \in \mathcal{F}_C$.
  Note that $(C \cap A)^c$ w.r.t. C is $C \setminus (C \cap A) = C \cap A^c \in \mathcal{F}_C$.
- Suppose $C \cap A_i \in \mathcal{F}_C$, where $A_i \in \mathcal{F}$, $\forall i \in \mathcal{N}$.
  $\cup_i A_i \in \mathcal{F}$
  $\cup_i (C \cap A_i) = C \cap (\cup_i A_i) \in \mathcal{F}_C$

Exercises 2

If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G}) = \{f^{-1}(A) : A \in \mathcal{G}\}$ is a sub-σ-field of $\mathcal{F}$

Proof:

- $\emptyset = f^{-1}(\emptyset) \in f^{-1}(\mathcal{G})$
- If $f^{-1}(A) \in f^{-1}(\mathcal{G})$, where $A \in \mathcal{G}$,
  then $A^c \in \mathcal{G}$.
  Note that $(f^{-1}(A))^c = f^{-1}(A^c) \in f^{-1}(\mathcal{G})$.
- If $f^{-1}(A_i) \in f^{-1}(\mathcal{G})$, where $A_i \in \mathcal{G}$
  then $\cup_i A_i \in \mathcal{G}$
  Note that $\cup_i f^{-1}(A_i) = f^{-1}(\cup_i A_i) \in \mathcal{G}$.

Exercises 3

Show that if a sequence of simple functions φ_n's satisfies
$0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$ and $\lim_n \varphi_n = f$, then

$$\int f \, \mathrm{d}\nu = \lim_n \int \varphi_n \, \mathrm{d}\nu$$

Proof: "$\geq$" : by definition of $\int f \, \mathrm{d}\nu$
"$\leq$" : If we can prove this result:

Lemma

For any simple function $g \leq f$, we have

$$\int g \, \mathrm{d}\nu \leq \lim_n \int \varphi_n \, \mathrm{d}\nu.$$

then $\sup_{g \in S_f} \int g \, \mathrm{d}\nu \leq \lim_n \int \varphi_n \, \mathrm{d}\nu$.

For any simple function $g \leq f$, we have

$$\int g \, \mathrm{d}\nu \leq \lim_n \int \varphi_n \, \mathrm{d}\nu.$$

Proof: Suppose $g = \sum_{j=1}^m b_j I_{B_j}$

- For any $c \in (0,1)$, let $A_n(c) = \{\varphi_n \geq c \cdot g\}$
- We have $\varphi_n I_{A_n(c)} \geq c \cdot g I_{A_n(c)}$.
- Hence

$$\int \varphi_n \, \mathrm{d}\nu \geq \int \varphi_n I_{A_n(c)} \, \mathrm{d}\nu \geq c \cdot \int g I_{A_n(c)} \, \mathrm{d}\nu$$
$$= c \sum_j b_j \nu(B_j \cap A_n(c)).$$

- Since $\varphi_n \uparrow f$, we have $A_n(c) \uparrow \Omega$ (see next slide)
- So $\nu(B_j \cap A_n(c)) \uparrow \nu(B_j)$ (by Proposition 1.1 in JS)
- Hence

$$\lim_n \int \varphi_n \, \mathrm{d}\nu \geq c \sum_j b_j \nu(B_j).$$

Proving $A_n(c) \uparrow \Omega$:
Fix any $x \in \Omega$.
If $f(x) = 0$, then $\phi_n(x) = g(x) = 0$, so $x \in A_n(c)$ for any n and c.
If $f(x) > 0$, then $f(x) = \lim \phi_n(x) > c \cdot g(x)$. So for all n large enough, $\phi_n(x) > c \cdot g(x)$, and $x \in A_n(c)$.
This shows $\lim_n A_n(c) = \cup_n A_n(c) = \Omega$.