Chapter 1: Probability Theory
Lecture 1: Measure space, measurable function, and integration

Random experiment: uncertainty in outcomes
Ω: sample space: a set containing all possible outcomes

Definition 1.1

A collection of subsets of Ω, $\mathscr{F}$, is a σ-field (or σ-algebra) if
(i) The empty set $\emptyset \in \mathscr{F}$;
(ii) If $A \in \mathscr{F}$, then the complement $A^c \in \mathscr{F}$;
(iii) If $A_i \in \mathscr{F}$, $i = 1, 2, \ldots$, then their union $\cup A_i \in \mathscr{F}$.
$(\Omega, \mathscr{F})$ is a measurable space if $\mathscr{F}$ is a σ-field on Ω

Two trivial examples: $\mathscr{F} = \{\emptyset, \Omega\}$ and $\mathscr{F}$ = all subsets of Ω (power set)
A nontrivial example: $\mathscr{F} = \{\emptyset, A, A^c, \Omega\}$, where $A \subset \Omega$
$\mathscr{C}$ = a collection of subsets of interest (may not be a σ-field)
$\sigma(\mathscr{C})$: the smallest σ-field containing $\mathscr{C}$ (the σ-field generated by $\mathscr{C}$)
$\sigma(\mathscr{C}) = \mathscr{C}$ if $\mathscr{C}$ itself is a σ-field
$\sigma(\{A\}) = \sigma(\{A, A^c\}) = \sigma(\{A, \Omega\}) = \sigma(\{A, \emptyset\}) = \{\emptyset, A, A^c, \Omega\}$

Borel σ-field

$\mathscr{R}^k$: the k-dimensional Euclidean space ($\mathscr{R}^1 = \mathscr{R}$ is the real line)
$\mathscr{O}$ = all open sets, $\mathscr{C}$ = all closed sets
$\mathscr{B}^k = \sigma(\mathscr{O}) = \sigma(\mathscr{C})$: the Borel σ-field on $\mathscr{R}^k$
$C \in \mathscr{B}^k$, $\mathscr{B}_C = \{C \cap B : B \in \mathscr{B}^k\}$ is the Borel σ-field on C

Definition 1.2.

Let $(\Omega, \mathscr{F})$ be a measurable space.
A set function ν defined on $\mathscr{F}$ is a *measure* if
(i) $0 \le \nu(A) \le \infty$ for any $A \in \mathscr{F}$;
(ii) $\nu(\emptyset) = 0$;
(iii) If $A_i \in \mathscr{F}$, $i = 1, 2, ...$, and A_i's are disjoint, i.e., $A_i \cap A_j = \emptyset$ for any $i \neq j$, then

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i).$$

$(\Omega, \mathscr{F}, \nu)$ is a measure space if ν is a measure on $\mathscr{F}$ in $(\Omega, \mathscr{F})$.

If $\nu(\Omega) = 1$, then ν is a probability measure.
We usually use P instead of ν; i.e., $(\Omega, \mathscr{F}, P)$ is a probability space.

Conventions

- For any $x \in \mathscr{R}$, $\infty + x = \infty$, $x\infty = \infty$ if $x > 0$, $x\infty = -\infty$ if $x < 0$
- $0\infty = 0$, $\infty + \infty = \infty$, $\infty^a = \infty$ for any $a > 0$;
- $\infty - \infty$ or ∞/∞ is not defined

Important examples of measures

- Let $x \in \Omega$ be a fixed point and
$$\delta_x(A) = \begin{cases} c & x \in A \\ 0 & x \notin A. \end{cases}$$
This is called a point mass at x
- Let $\mathscr{F}$ = all subsets of Ω and $\nu(A)$ = the number of elements in $A \in \mathscr{F}$ ($\nu(A) = \infty$ if A contains infinitely many elements).
Then ν is a measure on $\mathscr{F}$ and is called the counting measure.
- There is a unique measure m on $(\mathscr{R}, \mathscr{B})$ that satisfies $m([a,b]) = b - a$ for every finite interval $[a,b]$, $-\infty < a \leq b < \infty$.
This is called the Lebesgue measure.
If we restrict m to the measurable space $([0,1], \mathscr{B}_{[0,1]})$, then m is a probability measure (uniform distribution).

Proposition 1.1 (Properties of measures)

Let $(\Omega, \mathscr{F}, \nu)$ be a measure space.

1. (Monotonicity). If $A \subset B$, then $\nu(A) \le \nu(B)$.
2. (Subadditivity). For any sequence $A_1, A_2, \ldots$,
$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \nu(A_i).$$
3. (Continuity). If $A_1 \subset A_2 \subset A_3 \subset \cdots$ (or $A_1 \supset A_2 \supset A_3 \supset \cdots$ and $\nu(A_1) < \infty$), then
$$\nu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \nu(A_n),$$
where
$$\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i \quad \left(\text{or } = \bigcap_{i=1}^{\infty} A_i\right).$$

Let P be a probability measure on $(\mathscr{R}, \mathscr{B})$.
The cumulative distribution function (c.d.f.) of P is defined to be
$$F(x) = P((-\infty, x]), \quad x \in \mathscr{R}$$

Proposition 1.2 (Properties of c.d.f.'s)

(i) Let F be a c.d.f. on $\mathscr{R}$.

- (a) $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$;
- (b) $F(\infty) = \lim_{x\to\infty} F(x) = 1$;
- (c) F is nondecreasing, i.e., $F(x) \le F(y)$ if $x \le y$;
- (d) F is right continuous, i.e., $\lim_{y\to x, y>x} F(y) = F(x)$.

(ii) Suppose a real-valued function F on $\mathscr{R}$ satisfies (a)-(d) in part (i). Then F is the c.d.f. of a unique probability measure on $(\mathscr{R}, \mathscr{B})$.

Product space

$\mathscr{I} = \{1, ..., k\}$, k is finite or ∞

Γ_i, $i \in \mathscr{I}$, are some sets

$\prod_{i\in\mathscr{I}} \Gamma_i = \Gamma_1 \times \cdots \times \Gamma_k = \{(a_1, ..., a_k) : a_i \in \Gamma_i, i \in \mathscr{I}\}$

$\mathscr{R} \times \mathscr{R} = \mathscr{R}^2$, $\mathscr{R} \times \mathscr{R} \times \mathscr{R} = \mathscr{R}^3$

Let $(\Omega_i, \mathscr{F}_i)$, $i \in \mathscr{I}$, be measurable spaces

$\prod_{i\in\mathscr{I}} \mathscr{F}_i$ is not necessarily a σ-field

$\sigma(\prod_{i\in\mathscr{I}} \mathscr{F}_i)$ is called the *product σ-field* on the *product space* $\prod_{i\in\mathscr{I}} \Omega_i$

$(\prod_{i\in\mathscr{I}} \Omega_i, \sigma(\prod_{i\in\mathscr{I}} \mathscr{F}_i))$ is denoted by $\prod_{i\in\mathscr{I}} (\Omega_i, \mathscr{F}_i)$

Example: $\prod_{i=1,...,k} (\mathscr{R}, \mathscr{B}) = (\mathscr{R}^k, \mathscr{B}^k)$

Product measure

Consider a rectangle $[a_1, b_1] \times [a_2, b_2] \subset \mathscr{R}^2$.
The usual area of $[a_1, b_1] \times [a_2, b_2]$ is

$$(b_1 - a_1)(b_2 - a_2) = m([a_1, b_1])m([a_2, b_2])$$

Is $m([a_1, b_1])m([a_2, b_2])$ the same as the value of a measure defined on the product σ-field?

σ-finite

A measure ν on $(\Omega, \mathscr{F})$ is said to be σ-*finite* iff there exists a sequence $\{A_1, A_2, ...\}$ such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i

Any finite measure (such as a probability measure) is clearly σ-finite

The Lebesgue measure on $\mathscr{R}$ is σ-finite, since $\mathscr{R} = \cup A_n$ with $A_n = (-n, n)$, $n = 1, 2, ...$

The counting measure in is σ-finite if and only if Ω is countable

The measure $\nu(A) = \infty$ unless $A = \emptyset$ is not σ-finite

Proposition 1.3 (Product measure theorem)

Let $(\Omega_i, \mathscr{F}_i, \nu_i)$, $i = 1, ..., k$, be measure spaces with σ-finite measures. There exists a unique σ-finite measure on σ-field $\sigma(\mathscr{F}_1 \times \cdots \times \mathscr{F}_k)$, called the *product measure* and denoted by $\nu_1 \times \cdots \times \nu_k$, such that

$$\nu_1 \times \cdots \times \nu_k(A_1 \times \cdots \times A_k) = \nu_1(A_1) \cdots \nu_k(A_k)$$

for all $A_i \in \mathscr{F}_i$, $i = 1, ..., k$.

Jiont and marginal c.d.f.'s

The joint c.d.f. of a probability measure on $(\mathscr{R}^k, \mathscr{B}^k)$ is defined by

$$F(x_1, ..., x_k) = P((-\infty, x_1] \times \cdots \times (-\infty, x_k]), \quad x_i \in \mathscr{R}$$

and the ith *marginal* c.d.f. is defined by

$$F_i(x) = \lim_{x_j \to \infty, j=1,...,i-1,i+1,...,k} F(x_1, ..., x_{i-1}, x, x_{i+1}, ..., x_k)$$

There is a 1-1 correspondence between probability and c.d.f. on $\mathscr{R}^k$. The product measure corresponds to

$$F(x_1, ..., x_k) = F_1(x_1) \cdots F_k(x_k), \quad (x_1, ..., x_k) \in \mathscr{R}^k,$$

where F_i is the c.d.f. of the ith probability measure.

Measurable function

f: a function from Ω to Λ (often $\Lambda = \mathscr{R}^k$)
Inverse image of $B \subset \Lambda$ under f:

$$f^{-1}(B) = \{f \in B\} = \{\omega \in \Omega : f(\omega) \in B\}.$$

The inverse function f^{-1} need not exist for $f^{-1}(B)$ to be defined.

$$f^{-1}(B^c) = (f^{-1}(B))^c \quad \text{for any } B \subset \Lambda$$

$$f^{-1}(\cup B_i) = \cup f^{-1}(B_i) \quad \text{for any } B_i \subset \Lambda,\ i = 1, 2, ...$$

Let $\mathscr{C}$ be a collection of subsets of Λ.
Define $f^{-1}(\mathscr{C}) = \{f^{-1}(C) : C \in \mathscr{C}\}$

Definition 1.3

Let $(\Omega, \mathscr{F})$ and $(\Lambda, \mathscr{G})$ be measurable spaces.
Let f be a function from Ω to Λ.
f is called a *measurable function* from $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G})$ iff $f^{-1}(\mathscr{G}) \subset \mathscr{F}$.

Remarks

- f is measurable from $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G})$ iff for any $B \in \mathscr{G}$, $f^{-1}(B) = \{\omega : f(\omega) \in B\} \in \mathscr{F}$; we don't care about whether $\{f(\omega) : \omega \in A\}$ is in $\mathscr{G}$ or not, $A \in \mathscr{F}$.
- If $\mathscr{F}$ = all subsets of Ω, then any function f is measurable.
- If f is measurable from $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G})$, then $f^{-1}(\mathscr{G})$ is a sub-σ-field of $\mathscr{F}$ and is called the σ-field generated by f and denoted by $\sigma(f)$.
- $\sigma(f)$ may be much simpler than $\mathscr{F}$
- A measurable f from $(\Omega, \mathscr{F})$ to $(\mathscr{R}, \mathscr{B})$ is called a Borel function.
- f is Borel if and only if $f^{-1}(a, \infty) \in \mathscr{F}$ for all $a \in \mathscr{R}$.
- There are a lots of Borel functions (Proposition 1.4)
- In probability and statistics, a Borel function is also called a random variable. (A random variable = a variable that is random?)
- A random vector $(X_1, ..., X_n)$ is a function measurable from $(\Omega, \mathscr{F})$ to $(\mathscr{R}^n, \mathscr{B}^n)$.
- $(X_1, ..., X_n)$ is a random vector iff each X_i is a random variable.

Indicator and simple functions

The indicator function for $A \subset \Omega$ is:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

For any $B \subset \mathscr{R}$,

$$I_A^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ A & 0 \notin B, 1 \in B \\ A^c & 0 \in B, 1 \notin B \\ \Omega & 0 \in B, 1 \in B. \end{cases}$$

Then, $\sigma(I_A) = \{\emptyset, A, A^c, \Omega\}$ and I_A is Borel iff $A \in \mathscr{F}$
Note that $\sigma(I_A)$ is much simpler than $\mathscr{F}$.
Let $A_1, ..., A_k$ be measurable sets on Ω and $a_1, ..., a_k$ be real numbers.
A simple function is

$$\varphi(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega)$$

A simple function is nonnegative iff $a_i \geq 0$ for all i.
Any nonnegative Borel function can be the limit of a sequence of nonnegative simple functions.

Let $A_1,...,A_k$ be a partition of Ω, i.e., A_i's are disjoint and $A_1 \cup \cdots \cup A_k = \Omega$.
Then the simple function φ with distinct a_i's exactly characterizes this partition and $\sigma(\varphi) = \sigma(\{A_1,...,A_k\})$.

Distribution (law)

Let $(\Omega, \mathscr{F}, \nu)$ be a measure space and f be a measurable function from $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G})$.
The *induced measure* by f, denoted by $\nu \circ f^{-1}$, is a measure on $\mathscr{G}$ defined as

$$\nu \circ f^{-1}(B) = \nu(f \in B) = \nu\left(f^{-1}(B)\right), \quad B \in \mathscr{G}$$

If $\nu = P$ is a probability measure and X is a random variable or a random vector, then $P \circ X^{-1}$ is called the *law* or the *distribution* of X and is denoted by P_X.
The c.d.f. of P_X is also called the c.d.f. or joint c.d.f. of X and is denoted by F_X.

Integration

Integration is a type of “average".

Definition 1.4

(a) The integral of a nonnegative simple function φ w.r.t. ν is defined as

$$\int \varphi d\nu = \sum_{i=1}^{k} a_i \nu(A_i).$$

(b) Let f be a nonnegative Borel function and let $\mathscr{S}_f$ be the collection of all nonnegative simple functions satisfying $\varphi(\omega) \le f(\omega)$ for any $\omega \in \Omega$. The integral of f w.r.t. ν is defined as

$$\int f d\nu = \sup \left\{ \int \varphi d\nu : \ \varphi \in \mathscr{S}_f \right\}.$$

(Hence, for any Borel function $f \ge 0$, there exists a sequence of simple functions $\varphi_1, \varphi_2, \ldots$ such that $0 \le \varphi_i \le f$ for all i and $\lim_{n\to\infty} \int \varphi_n d\nu = \int f d\nu$.)

(c) Let f be a Borel function,

$$f_+(\omega) = \max\{f(\omega), 0\}$$

be the positive part of f, and

$$f_-(\omega) = \max\{-f(\omega), 0\}$$

be the negative part of f. (Note that f_+ and f_- are nonnegative Borel functions, $f(\omega) = f_+(\omega) - f_-(\omega)$, and $|f(\omega)| = f_+(\omega) + f_-(\omega)$.) We say that $\int f d\nu$ exists if and only if at least one of $\int f_+ d\nu$ and $\int f_- d\nu$ is finite, in which case

$$\int f d\nu = \int f_+ d\nu - \int f_- d\nu.$$

(d) When both $\int f_+ d\nu$ and $\int f_- d\nu$ are finite, we say that f is integrable. Let A be a measurable set and I_A be its indicator function. The integral of f over A is defined as

$$\int_A f d\nu = \int I_A f d\nu.$$

Note: A Borel function f is integrable if and only if $|f|$ is integrable.

Notation for integrals

- $\int f d\nu = \int_\Omega f d\nu = \int f(\omega) d\nu = \int f(\omega) d\nu(\omega) = \int f(\omega)\nu(d\omega)$.
- In probability and statistics, $\int X dP = EX = E(X)$ and is called the *expectation* or *expected value* of X.
- If F is the c.d.f. of P on $(\mathscr{R}^k, \mathscr{B}^k)$, $\int f(x) dP = \int f(x) dF(x) = \int f dF$.

Extended set

For convenience, we define the integral of a measurable f from $(\Omega, \mathscr{F}, \nu)$ to $(\bar{\mathscr{R}}, \bar{\mathscr{B}})$, where $\bar{\mathscr{R}} = \mathscr{R} \cup \{-\infty, \infty\}$, $\bar{\mathscr{B}} = \sigma(\mathscr{B} \cup \{\{\infty\}, \{-\infty\}\})$.
Let $A_+ = \{f = \infty\}$ and $A_- = \{f = -\infty\}$.
If $\nu(A_+) = 0$, we define $\int f_+ d\nu$ to be $\int I_{A_+^c} f_+ d\nu$; otherwise $\int f_+ d\nu = \infty$.
$\int f_- d\nu$ is similarly defined.
If at least one of $\int f_+ d\nu$ and $\int f_- d\nu$ is finite, then
$\int f d\nu = \int f_+ d\nu - \int f_- d\nu$ is well defined.

Example 1.5

For a countable Ω, $\mathscr{F}$ = all subsets of Ω, ν = the counting measure, and a Borel f,

$$\int f d\nu = \sum_{\omega \in \Omega} f(\omega).$$

Example 1.6.

If $\Omega = \mathscr{R}$ and ν is the Lebesgue measure, then the Lebesgue integral of f over an interval $[a,b]$ is written as

$$\int_{[a,b]} f(x)dx = \int_a^b f(x)dx,$$

which agrees with the Riemann integral in calculus when the latter is well defined.

However, there are functions for which the Lebesgue integrals are defined but not the Riemann integrals.

Proposition 1.5

Let $(\Omega, \mathscr{F}, \nu)$ be a measure space and f and g be Borel functions.

(i) If $\int f d\nu$ exists and $a \in \mathscr{R}$, then $\int (af) d\nu$ exists and is equal to $a \int f d\nu$.

(ii) If both $\int f d\nu$ and $\int g d\nu$ exist and $\int f d\nu + \int g d\nu$ is well defined, then $\int (f+g) d\nu$ exists and is equal to $\int f d\nu + \int g d\nu$.

It is often a good idea to break the proof into several steps: simple functions, nonnegative functions, and then general functions.

Proof of Proposition 1.5(i) (the proof of (ii) is an exercise)

If $a=0$, then $\int (af)d\nu = \int 0 d\nu = 0 = a\int f d\nu$.
Suppose that $a>0$.
If f is simple and ≥ 0, then af is also simple and ≥ 0 and
$a\int f d\nu = \int (af) d\nu$ follows from the definition of integration.
For $a>0$ and a general $f \geq 0$,

$$\begin{aligned}\int (af)d\nu &= \sup\left\{\int \varphi d\nu :\ \varphi \in \mathscr{S}_{af}\right\} = \sup\left\{\int a\phi d\nu : \phi = \varphi/a \in \mathscr{S}_f\right\} \\ &= \sup\left\{a\int \phi d\nu : \phi \in \mathscr{S}_f\right\} = a\sup\left\{\int \phi d\nu : \phi \in \mathscr{S}_f\right\} = a\int f d\nu\end{aligned}$$

For $a>0$ and general f, since $\int f d\nu$ exists,

$$\begin{aligned}a\int f d\nu &= a\left(\int f_+ d\nu - \int f_- d\nu\right) = a\int f_+ d\nu - a\int f_- d\nu \\ &= \int af_+ d\nu - \int af_- d\nu = \int (af)_+ d\nu - \int (af)_- d\nu = \int (af) d\nu\end{aligned}$$

For $a<0$, $af = |a|(-f)$