CHAPTER 2: BASIC MEASURE THEORY

Set Theory and Topology in Real Space

- Basic concepts in set theory
  - element, set, whole space (Ω)
  - power set: 2^{Ω}; empty set: $\emptyset$
  - set relationship: $A \subseteq B$, $\cap_\alpha A_\alpha$, $\cup_\alpha A_\alpha$, A^c, $A - B$

$$A - B = A \cap B^c$$

- Set operations
  - properties: for any B, $\{A_\alpha\}$,

$$B \cap \{\cup_\alpha A_\alpha\} = \cup_\alpha \{B \cap A_\alpha\}, \quad B \cup \{\cap_\alpha A_\alpha\} = \cap_\alpha \{B \cup A_\alpha\},$$

$$\{\cup_\alpha A_\alpha\}^c = \cap_\alpha A_\alpha^c, \quad \{\cap_\alpha A_\alpha\}^c = \cup_\alpha A_\alpha^c. \quad (\text{ de Morgan law})$$

  - partition of a set

$$A_1 \cup A_2 \cup A_3 \cup ... = A_1 \cup (A_2 - A_1) \cup (A_3 - A_1 \cup A_2) \cup ...$$

  - $\limsup_n A_n = \cap_{n=1}^\infty \{\cup_{m=n}^\infty A_m\}$
    $\liminf_n A_n = \cup_{n=1}^\infty \{\cap_{m=n}^\infty A_m\}.$

- Topology in the Euclidean space
  - *open set*, *closed set*, *compact set*
  - properties: the union of any number of open sets is open; A is closed if and only if for any sequence $\{x_n\}$ in A such that $x_n \to x$, x must belong to A
  - only $\emptyset$ and the whole real line are both open set and closed
  - any open-set covering of a compact set has finite number of open sets covering the compact set

Measure Space

- Motivating example: counting measure
  - $\Omega = \{x_1, x_2, ...\}$
  - a set function $\mu^{\#}(A)$ is the number of points in A.
  - (a) $\mu^{\#}(\emptyset) = 0$;
    (b) if $A_1, A_2, ...$ are disjoint sets of Ω, then

$$\mu^{\#}(\cup_n A_n) = \sum_n \mu^{\#}(A_n).$$

- Motivating example: Lebesgue measure
  - $\Omega = (-\infty, \infty)$
  - how to measure the sizes of possibly any subsets in R? a set function λ?
  - (a) $\lambda(\emptyset) = 0$;
    (b) for any disjoint sets $A_1, A_2, \ldots,$

$$\lambda(\cup_n A_n) = \sum_n \lambda(A_n)$$

  - assign the length to any est of $\mathcal{B}_0$

$$\cup_{i=1}^n (a_i, b_i] \cup (-\infty, b] \cup (a, \infty), \quad \text{disjoint intervals}$$

  - What about non-intervals? how about in R^k?

- Three components in defining a measure space
  - the whole space, Ω
  - a collection of subsets whose sizes are measurable, $\mathcal{A}$,
  - a set function μ assigns negative values (sizes) to each set of $\mathcal{A}$ and satisfies properties (a) and (b)

Field, σ-filed

- Some intuition
  - $\mathcal{A}$ contains the sets whose sizes are measurable
  - $\mathcal{A}$ should be closed under complement or union

Definition 2.1 (fields, σ-fields) A non-void class $\mathcal{A}$ of subsets of Ω is called a:

(i) *field* or *algebra* if $A, B \in \mathcal{A}$ implies that $A \cup B \in \mathcal{A}$ and $A^c \in \mathcal{A}$; equivalently, $\mathcal{A}$ is closed under complements and finite unions.

(ii) σ-*field* or σ-*algebra* if $\mathcal{A}$ is a field and $A_1, A_2, ... \in \mathcal{A}$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$; equivalently, $\mathcal{A}$ is closed under complements and countable unions. †

- Properties of σ-field

Proposition 2.1. (i) For a field $\mathcal{A}$, $\emptyset, \Omega \in \mathcal{A}$ and if $A_1, ..., A_n \in \mathcal{A}$, $\cap_{i=1}^{n} A_i \in \mathcal{A}$.
(ii) For a σ-field $\mathcal{A}$, if $A_1, A_2, ... \in \mathcal{A}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Proof

(i) For any $A \in \mathcal{A}$, $\Omega = A \cup A^c \in \mathcal{A} \Rightarrow \emptyset = \Omega^c \in \mathcal{A}$.

$A_1, ..., A_n \in \mathcal{A}$
$\Rightarrow \cap_{i=1}^n A_i = (\cup_{i=1}^n A_i^c)^c \in \mathcal{A}$.

(ii) $(\cap_{i=1}^\infty A_i)^c = \cup_{i=1}^\infty A_i^c$.

- Examples of σ-field
  - $\mathcal{A} = \{\emptyset, \Omega\}$ and $2^\Omega = \{A : A \subset \Omega\}$
  - $\mathcal{B}_0$ is a field but not a σ-field

$$(a, b) = \cup_{n=1}^{\infty}(a, b - 1/n] \notin \mathcal{B}_0$$

  - $\mathcal{A} = \{A : A \text{ is in } R \text{ and } A^c \text{ is countable}\}$
    $\mathcal{A}$ is closed under countable union but not complement

- Measure defined on a σ-field

Definition 2.2 (measure, probability measure)
(i) A *measure* μ is a function from a σ-field $\mathcal{A}$ to $[0, \infty)$ satisfying: $\mu(\emptyset) = 0$; $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for any countable (finite) disjoint sets $A_1, A_2, ... \in \mathcal{A}$. The latter is called the *countable additivity*.
(ii) Additionally, if $\mu(\Omega) = 1$, μ is a *probability measure* and we usually use P instead of μ to indicate a probability measure.

- Properties of measure

Proposition 2.2

(i) If $\{A_n\} \subset \mathcal{A}$ and $A_n \subset A_{n+1}$ for all n, then $\mu(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

(ii) If $\{A_n\} \subset \mathcal{A}$, $\mu(A_1) < \infty$ and $A_n \supset A_{n+1}$ for all n, then $\mu(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.

(iii) For any $\{A_n\} \subset \mathcal{A}$, $\mu(\cup_n A_n) \leq \sum_n \mu(A_n)$ (*countable sub-additivity*).

(iv) $\mu(\liminf_n A_n) = \lim_n \mu(\cap_{k=n}^{\infty} A_n) \leq \liminf_n \mu(A_n)$

Proof

(i) $\mu(\cup_{n=1}^{\infty} A_n) = \mu(A_1 \cup (A_2 - A_1) \cup ...) = \mu(A_1) + \mu(A_2 - A_1) + ...$
$= \lim_n \{\mu(A_1) + \mu(A_2 - A_1) + ... + \mu(A_n - A_{n-1})\} = \lim_n \mu(A_n)$.

(ii)
$\mu(\cap_{n=1}^{\infty} A_n) = \mu(A_1) - \mu(A_1 - \cap_{n=1}^{\infty} A_n) = \mu(A_1) - \mu(\cup_{n=1}^{\infty}(A_1 \cap A_n^c))$.
$A_1 \cap A_n^c$ is increasing
$\Rightarrow$ the second term equals $\lim_n \mu(A_1 \cap A_n^c) = \mu(A_1) - \lim_n \mu(A_n)$.

(iii)
$\mu(\cup_n A_n) = \lim_n \mu(A_1 \cup ... \cup A_n) = \lim_n \{\sum_{i=1}^n \mu(A_i - \cup_{j<i} A_j)\}$
$\leq \lim_n \sum_{i=1}^n \mu(A_i) = \sum_n \mu(A_n)$.

(iv) $\mu(\liminf_n A_n) = \lim_n \mu(\cap_{k=n}^{\infty} A_n) \leq \liminf_n \mu(A_n)$.

- Measures space

a triplet $(\Omega, \mathcal{A}, \mu)$

- set in $\mathcal{A}$ is called a *measurable set*
- If $\mu = P$ is a probability measure, $(\Omega, \mathcal{A}, P)$ is a *probability measure space*: *probability sample* and *probability event*
- a measure μ is called *σ-finite* if there exists a countable sets $\{F_n\} \subset \mathcal{A}$ such that $\Omega = \cup_n F_n$ and for each F_n, $\mu(F_n) < \infty$.

- Examples of measure space
  - discrete measure:

$$\mu(A) = \sum_{\omega_i \in A} m_i, \quad A \in \mathcal{A}.$$

  - counting measure $\mu^{\#}$ in any space, say R: it is not σ-finite.

Measure Space Construction

- Two basic questions
  - Can we find a σ-field containing all the sets of $\mathcal{C}$?
  - Can we obtain a set function defined for any set of this σ-field such that the set function agrees with μ in $\mathcal{C}$?

- Answer to the first question

Proposition 2.3 (i) Arbitrary intersections of fields (σ-fields) are fields (σ-fields).
(ii) For any class $\mathcal{C}$ of subsets of Ω, there exists a minimal σ-field containing $\mathcal{C}$ and we denote it as $\sigma(\mathcal{C})$.

Proof

(i) is obvious.

For (ii),

$$\sigma(\mathcal{C}) = \cap_{\mathcal{C}\subset\mathcal{A},\mathcal{A}\text{ is }\sigma\text{-field}}\mathcal{A},$$

i.e., the intersection of all the σ-fields containing $\mathcal{C}$.

- Answer to the second question

Theorem 2.1 (Caratheodory Extension Theorem) A measure μ on a field $\mathcal{C}$ can be extended to a measure on the minimal σ-field $\sigma(\mathcal{C})$. If μ is σ-finite on $\mathcal{C}$, then the extension is unique and also σ-finite.

Construction

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

- Application to measure construction
  - generate a σ-field containing $\mathcal{B}_0$: *Borel σ-field $\mathcal{B}$*
  - extend λ to $\mathcal{B}$: the Lebesgue measure
  - $(R, \mathcal{B}, \lambda)$ is named the *Borel measure space*
  - in R^k, we obtain $(R^k, \mathcal{B}^k, \lambda^k)$

- Other measure construction on $\mathcal{B}$
  - F is non-decreasing and right-continuous
  - a set function in $\mathcal{B}_0$: $\lambda_F((a, b]) = F(b) - F(a)$
  - measure extension λ_F in $\mathcal{B}$: *Lebesgue-Stieltjes measure generated by F*
  - the Lebesuge measure is a special case with $F(x) = x$
  - if F is a distribution function, this measure is a probability measure in R

- Completion after measure construction
  - motivation: any subsets of a zero-measure set should be given measure zero but may not be in $\mathcal{A}$
  - Completion: add these nuisance sets to $\mathcal{A}$

- Details of completion
  - obtain another measure space $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$

$$\bar{\mathcal{A}} = \{A \cup N : A \in \mathcal{A},$$

$$N \subset B \text{ for some } B \in \mathcal{A} \text{ such that } \mu(B) = 0\}$$

    and $\bar{\mu}(A \cup N) = \mu(A)$.
  - the completion of the Borel measure space is the *Lebesgue measure space* and the completed Borel σ-field is the σ-field of *Lebesgue sets*
  - we always assume that a measure space is completed

Measurable Function

- Definition

Definition 2.3 (measurable function) Let $X : \Omega \mapsto R$ be a function defined on Ω. X is *measurable* if for $x \in R$, the set $\{\omega \in \Omega : X(\omega) \leq x\}$ is measurable, equivalently, belongs to $\mathcal{A}$. Especially, if the measure space is a probability measure space, X is called a *random variable*.

- Property of measurable function

Proposition 2.4 If X is measurable, then for any $B \in \mathcal{B}$, $X^{-1}(B) = \{\omega : X(\omega) \in B\}$ is measurable.

Proof

$$\mathcal{B}^* = \{B : B \subset R, X^{-1}(B) \text{ is measurable in } \mathcal{A}\}$$

$(-\infty, x] \in \mathcal{B}^*$.

$B \in \mathcal{B}^* \Rightarrow X^{-1}(B) \in \mathcal{A} \Rightarrow X^{-1}(B^c) = \Omega - X^{-1}(B) \in \mathcal{A}$
then $B^c \in \mathcal{B}^*$.

$B_1, B_2, ... \in \mathcal{B}^* \Rightarrow X^{-1}(B_1), X^{-1}(B_2), ... \in \mathcal{A} \Rightarrow$
$X^{-1}(B_1 \cup B_2 \cup ...) = X^{-1}(B_1) \cup X^{-1}(B_2) \cup ... \in \mathcal{A}$.
$\Rightarrow B_1 \cup B_2 \cup ... \in \mathcal{B}^*$.

$\Rightarrow \mathcal{B}^*$ is a σ-field containing all intervals of the type $(-\infty, x] \Rightarrow$
$\mathcal{B} \subset \mathcal{B}^*$.
For any Borel set B, $X^{-1}(B)$ is measurable in $\mathcal{A}$.

- Construction of measurable function
  - *simple function*: $\sum_{i=1}^{n} x_i I_{A_i}(\omega), \quad A_i \in \mathcal{A}$
  - the finite summation and the maximum of simple functions are still simple functions
  - any elementary functions of measurable functions are measurable

Proposition 2.5 Suppose that $\{X_n\}$ are measurable. Then so are $X_1 + X_2, X_1X_2, X_1^2$ and $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$ and $\liminf_n X_n$.

Proof

$\{X_1 + X_2 \leq x\} = \Omega - \{X_1 + X_2 > x\} =$
$\Omega - \cup_{r \in Q} \{X_1 > r\} \cap \{X_2 > x - r\}$, Q={ all rational numbers}.

$$\{X_1^2 \leq x\} = \{X_1 \leq \sqrt{x}\} - \{X_1 < -\sqrt{x}\}.$$

$$X_1 X_2 = \left\{(X_1 + X_2)^2 - X_1^2 - X_2^2\right\}/2$$

$\{\sup_n X_n \leq x\} = \cap_n \{X_n \leq x\}$.
$\{\inf_n X_n \leq x\} = \{\sup_n(-X_n) \geq -x\}$.

$\{\limsup_n X_n \leq x\} = \cap_{r \in Q, r > 0} \cup_{n=1}^{\infty} \cap_{k \geq n} \{X_k < x + r\}$.
$\liminf_n X_n = -\limsup_n(-X_n)$.

• Approximating measurable function with simple functions

Proposition 2.6 For any measurable function $X \geq 0$, there exists an increasing sequence of simple functions $\{X_n\}$ such that $X_n(\omega)$ increases to $X(\omega)$ as n goes to infinity.

Proof

$$X_n(\omega) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} I\{\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\} + nI\{X(\omega) \geq n\}$$

$\Rightarrow$ X_n is increasing over n.

$\Rightarrow$ if $X(\omega) < n$, then $|X_n(\omega) - X(\omega)| < \frac{1}{2^n}$.

$\Rightarrow$ $X_n(\omega)$ converges to $X(\omega)$.

If X is bounded, $\sup_\omega |X_n(\omega) - X(\omega)| < \frac{1}{2^n}$

Integration

Definition 2.4 (i) For any simple function $X(\omega) = \sum_{i=1}^n x_i I_{A_i}(\omega)$, we define $\sum_{i=1}^n x_i \mu(A_i)$ as the *integral* of X with respect to measure μ, denoted as $\int X d\mu$.
(ii) For any $X \geq 0$, we define $\int X d\mu$ as

$$\int X d\mu = \sup_{Y \text{ is simple function, } 0 \leq Y \leq X} \int Y d\mu.$$

(iii) For general X, let $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Then $X = X^+ - X-$. If one of $\int X^+ d\mu$, $\int X^- d\mu$ is finite, we define

$$\int X d\mu = \int X^+ d\mu - \int X^- d\mu.$$

- Some notes
  - X is *integrable* if $\int |X| d\mu = \int X^+ d\mu + \int X^- d\mu$ is finite
  - the definition (ii) is consistent with (i) when X itself is a simple function
  - for a probability measure space and X is a random variable, $\int X d\mu \equiv E[X]$

- Fundamental properties of integration

Proposition 2.7 (i) For two measurable functions $X_1 \geq 0$ and $X_2 \geq 0$, if $X_1 \leq X_2$, then $\int X_1 d\mu \leq \int X_2 d\mu$.
(ii) For $X \geq 0$ and any sequence of simple functions Y_n increasing to X, $\int Y_n d\mu \to \int X d\mu$.

Proof

(i) For any simple function $0 \le Y \le X_1$, $Y \le X_2$.

$\Rightarrow \int Y d\mu \le \int X_2 d\mu$.

Take the supreme over all the simple functions less than X_1

$\Rightarrow \int X_1 d\mu \le \int X_2 d\mu$.

(ii) From (i), $\int Y_n d\mu$ is increasing and bounded by $\int X d\mu$.

It suffices to show that for any simple function $Z = \sum_{i=1}^{m} x_i I_{A_i}(\omega)$, where $\{A_i, 1 \le i \le m\}$ are disjoint measurable sets and $x_i > 0$, such that $0 \le Z \le X$,

$$\lim_n \int Y_n d\mu \ge \sum_{i=1}^{m} x_i \mu(A_i).$$

We consider two cases.
Case 1.$\int Zd\mu = \sum_{i=1}^m x_i\mu(A_i)$ is finite thus both x_i and $\mu(A_i)$ are finite.

Fix an $\epsilon > 0$, let $A_{in} = A_i \cap \{\omega : Y_n(\omega) > x_i - \epsilon\}$. $\Rightarrow$ A_{in} increases to A_i $\Rightarrow$ $\mu(A_{in})$ increases to $\mu(A_i)$.

When n is large,

$$\int Y_n d\mu \geq \sum_{i=1}^m (x_i - \epsilon)\mu(A_i).$$

$\Rightarrow$ $\lim_n \int Y_n d\mu \geq \int Zd\mu - \epsilon \sum_{i=1}^m \mu(A_i)$.
$\Rightarrow$ $\lim_n \int Y_n d\mu \geq \int Zd\mu$ by letting ϵ approach 0.

Case 2 suppose $\int Z d\mu = \infty$ then there exists some i from $\{1, ..., m\}$, say 1, so that $\mu(A_1) = \infty$ or $x_1 = \infty$.

Choose any $0 < x < x_1$ and $0 < y < \mu(A_1)$.

$A_{1n} = A_1 \cap \{\omega : Y_n(\omega) > x\}$ increases to A_1. n large enough,
$\mu(A_{1n}) > y$
$\Rightarrow \lim_n \int Y_n d\mu \geq xy$.

$\Rightarrow$ Letting $x \to x_1$ and $y \to \mu(A_1)$, conclude $\lim_n \int Y_n d\mu = \infty$.

$\Rightarrow \lim_n \int Y_n d\mu \geq \int Z d\mu$.

- Elementary properties

Proposition 2.8 Suppose $\int X d\mu, \int Y d\mu$ and $\int X d\mu + \int Y d\mu$ exit. Then
(i) $\int (X+Y) d\mu = \int X d\mu + \int Y d\mu, \quad \int cX d\mu = c \int X d\mu$;
(ii) $X \geq 0$ implies $\int X d\mu \geq 0$; $X \geq Y$ implies $\int X d\mu \geq \int Y d\mu$; and $X = Y$ a.e. implies that $\int X d\mu = \int Y d\mu$;
(iii) $|X| \leq Y$ with Y integrable implies that X is integrable; X and Y are integrable implies that $X + Y$ is integrable.

- Calculation of integration by definition

$$\int X d\mu = \lim_n \left\{ \sum_{k=1}^{n2^n-1} \frac{k}{2^n}\mu(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) + n\mu(X \geq n) \right\}.$$

- Integration w.r.t counting measure or Lebesgue measure
  - $\int g d\mu^{\#} = \sum_i g(x_i)$.
  - continuous function $g(x)$, $\int g d\lambda$ is equal to the usual Riemann integral $\int g(x)dx$
  - $(\Omega, \mathcal{B}, \lambda_F)$, where F is differentiable except discontinuous points $\{x_1, x_2, ...\}$,

    $$\int g d\lambda_F = \sum_i g(x_i)\left\{F(x_i) - F(x_i-)\right\} + \int g(x)f(x)dx,$$

    where $f(x)$ is the derivative of $F(x)$.

Convergence Theorems

- Monotone convergence theorem (MCT)

Theorem 2.2 If $X_n \geq 0$ and X_n increases to X, then $\int X_n d\mu \to \int X d\mu$.

Proof

Choose nonnegative simple function X_{km} increasing to X_k as $m \to \infty$. Define $Y_n = \max_{k \le n} X_{kn}$.

$\Rightarrow$ $\{Y_n\}$ is an increasing series of simple functions

$$X_{kn} \le Y_n \le X_n, \quad \text{so} \int X_{kn} d\mu \le \int Y_n d\mu \le \int X_n d\mu.$$

$\Rightarrow$ $n \to \infty$ $X_k \le \lim_n Y_n \le X$ and

$$\int X_k d\mu \le \int \lim_n Y_n d\mu = \lim_n \int Y_n d\mu \le \lim_n \int X_n d\mu$$

$\Rightarrow$ $k \to \infty$, $X \le \lim_n Y_n \le X$ and

$$\lim_k \int X_k d\mu \le \int \lim_n Y_n d\mu \le \lim_n \int X_n d\mu.$$

The result holds.

- Counter example

$X_n(x) = -I(x > n)/n$ in the Lebesgue measure space.
X_n increases to zero but $\int X_n d\lambda = -\infty$

- Fatou's Lemma

Theorem 2.3 If $X_n \geq 0$ then

$$\int \liminf_n X_n d\mu \leq \liminf_n \int X_n d\mu.$$

Proof

$$\liminf_n X_n = \sup_{n=1}^{\infty} \inf_{m \geq n} X_m.$$

$\Rightarrow$ $\{\inf_{m\geq n} X_m\}$ increases to $\liminf_n X_n$.

By the MCT,

$$\int \liminf_n X_n d\mu = \lim_n \int \inf_{m\geq n} X_m d\mu \leq \int X_n d\mu.$$

- Two definitions in convergence

Definition 2.4 A sequence X_n *converges almost everywhere* (a.e.) to X, denoted $X_n \rightarrow_{a.e.} X$, if $X_n(\omega) \rightarrow X(\omega)$ for all $\omega \in \Omega - N$ where $\mu(N) = 0$. If μ is a probability, we write a.e. as a.s. (almost surely). A sequence X_n *converges in measure* to a measurable function X, denoted $X_n \rightarrow_\mu X$, if $\mu(|X_n - X| \geq \epsilon) \rightarrow 0$ for all $\epsilon > 0$. If μ is a probability measure, we say X_n *converges in probability* to X.

- Properties of convergence

Proposition 2.9 Let $\{X_n\}$, X be finite measurable functions. Then $X_n \rightarrow_{a.e.} X$ if and only if for any $\epsilon > 0$,

$$\mu(\cap_{n=1}^{\infty} \cup_{m \geq n} \{|X_m - X| \geq \epsilon\}) = 0.$$

If $\mu(\Omega) < \infty$, then $X_n \rightarrow_{a.e.} X$ if and only if for any $\epsilon > 0$,

$$\mu(\cup_{m \geq n} \{|X_m - X| \geq \epsilon\}) \rightarrow 0.$$

Proof

$$\{\omega : X_n(\Omega) \to X(\omega)\}^c = \cup_{k=1}^{\infty} \cap_{n=1}^{\infty} \cup_{m \geq n} \left\{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\right\}.$$

$X_n \to_{a.e} X \Rightarrow$ the measure of the left-hand side is zero.
$\Rightarrow \cap_{n=1}^{\infty} \cup_{m \geq n} \{|X_m - X| \geq \epsilon\}$ has measure zero.

For the other direction, choose $\epsilon = 1/k$ for any k, then by countable sub-additivity,

$$\mu(\cup_{k=1}^{\infty} \cap_{n=1}^{\infty} \cup_{m \geq n} \left\{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\right\})$$

$$\leq \sum_k \mu(\cap_{n=1}^{\infty} \cup_{m \geq n} \left\{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\right\}) = 0.$$

$\Rightarrow X_n \to_{a.e.} X$.
When $\mu(\Omega) < \infty$, the latter holds by Proposition 2.2.

- Relationship between two convergence modes

Proposition 2.10 Let X_n be finite a.e.
(i) If $X_n \to_\mu X$, then there exists a subsequence $X_{n_k} \to_{a.e} X$.
(ii) If $\mu(\Omega) < \infty$ and $X_n \to_{a.e.} X$, then $X_n \to_\mu X$.

Proof

(i) Find n_k

$$P(|X_{n_k} - X| \geq 2^{-k}) < 2^{-k}.$$

$\Rightarrow \mu(\cup_{m \geq k} \{|X_{n_m} - X| \geq \epsilon\}) \leq \mu(\cup_{m \geq k} \{|X_{n_m} - X| \geq 2^{-k}\})$
$\leq \sum_{m \geq k} 2^{-m} \to 0.$

$\Rightarrow X_{n_k} \to_{a.e} X.$

(ii) is direct from the second part of Proposition 2.9.

- Examples of convergence
  - Let $X_{2^n+k} = I(x \in [k/2^n, (k+1)/2^n)), 0 \le k < 2^n$ in the Lebesgue measure space. Then $X_n \to_\lambda 0$ but does not converge to zero almost everywhere.
  - $X_n = nI(|x| > n) \to_{a.e.} 0$ but $\lambda(|X_n| > \epsilon) \to \infty$.

- Dominated Convergence Theorem (DCT)

Theorem 2.4 If $|X_n| \leq Y$ a.e. with Y integrable, and if $X_n \rightarrow_{\mu} X$ (or $X_n \rightarrow_{a.e.} X$), then $\int |X_n - X| d\mu \rightarrow 0$ and $\lim \int X_n d\mu = \int X d\mu$.

Proof

Assume $X_n \to_{a.e} X$. Define $Z_n = 2Y - |X_n - X|$. $Z_n \geq 0$ and $Z_n \to 2Y$.

$\Rightarrow$ From the Fatou's lemma,

$$\int 2Y d\mu \leq \liminf_n \int (2Y - |X_n - X|) d\mu.$$

$\Rightarrow$ $\limsup_n \int |X_n - X| d\mu \leq 0$.

If $X_n \to_\mu X$ and the result does not hold for some subsequence of X_n, by Proposition 2.10, there exits a further sub-sequence converging to X almost surely. However, the result holds for this further subsequence. Contradiction!

- Interchange of integral and limit or derivative

Theorem 2.5 Suppose that $X(\omega, t)$ is measurable for each $t \in (a, b)$.
(i) If $X(\omega, t)$ is a.e. continuous in t at t_0 and $|X(\omega, t)| \leq Y(\omega), a.e.$ for $|t - t_0| < \delta$ with Y integrable, then

$$\lim_{t \to t_0} \int X(\omega, t) d\mu = \int X(\omega, t_0) d\mu.$$

(ii) Suppose $\frac{\partial}{\partial t}X(\omega, t)$ exists for a.e. ω, all $t \in (a, b)$ and $|\frac{\partial}{\partial t}X(\omega, t)| \leq Y(\omega), a.e.$ for all $t \in (a, b)$ with Y integrable. Then

$$\frac{\partial}{\partial t}\int X(\omega, t)d\mu = \int \frac{\partial}{\partial t}X(\omega, t)d\mu.$$

Proof

(i) follows from the DCT and the subsequence argument.

(ii)

$$\frac{\partial}{\partial t}\int X(\omega,t)d\mu = \lim_{h\to 0}\int \frac{X(\omega,t+h)-X(\omega,t)}{h}d\mu.$$

Then from the conditions and (i), such a limit can be taken within the integration.

Product of Measures

- Definition
  - $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$
  - $\mathcal{A}_1 \times \mathcal{A}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\})$
  - $(\mu_1 \times \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. with its extension to all sets in the $\mathcal{A}_1 \times \mathcal{A}_2$

- Examples
  - $(R^k = R \times ... \times R, \mathcal{B} \times ... \times \mathcal{B}, \lambda \times ... \times \lambda)$

$$\lambda \times ... \times \lambda \equiv \lambda^k$$

  - $\Omega = \{1, 2, 3...\}$

$$(R \times \Omega, \mathcal{B} \times 2^{\Omega}, \lambda \times \mu^{\#})$$

- Integration on the product measure space
  - In calculus,
    $\int_{R^2} f(x,y)dxdy = \int_x \int_y f(x,y)dydx = \int_y \int_x f(x,y)dxdy$
  - Do we have the same equality in the product measure space?

Theorem 2.6 (Fubini-Tonelli Theorem) Suppose that $X : \Omega_1 \times \Omega_2 \to R$ is $\mathcal{A}_1 \times \mathcal{A}_2$ measurable and $X \geq 0$. Then

$$\int_{\Omega_1} X(\omega_1, \omega_2) d\mu_1 \text{ is } \mathcal{A}_2 \text{ measurable,}$$

$$\int_{\Omega_2} X(\omega_1, \omega_2) d\mu_2 \text{ is } \mathcal{A}_1 \text{ measurable,}$$

$$\int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2) d(\mu_1 \times \mu_2) = \int_{\Omega_1} \left\{ \int_{\Omega_2} X(\omega_1, \omega_2) d\mu_2 \right\} d\mu_1$$

$$= \int_{\Omega_2} \left\{ \int_{\Omega_1} X(\omega_1, \omega_2) d\mu_1 \right\} d\mu_2.$$

- Conclusion from Theorem 2.6
  - in general, $X = X^+ - X^-$. Then the above results hold for X^+ and X^-. Thus, if $\int_{\Omega_1 \times \Omega_2} |X(\omega_1, \omega_2)| d(\mu_1 \times \mu_2)$ is finite, then the above results hold.

- One example
  - let $(\Omega, 2^{\Omega}, \mu^{\#})$ be a counting measure space where $\Omega = \{1, 2, 3, ...\}$ and $(R, \mathcal{B}, \lambda)$ be the Lebesgue measure space
  - define $f(x, y) = I(0 \leq x \leq y) \exp\{-y\}$; then

$$\int_{\Omega \times R} f(x, y) d\{\mu^{\#} \times \lambda\} = \int_{\Omega} \{\int_R f(x, y) d\lambda(y)\} d\mu^{\#}(x)$$

$$= \int_{\Omega} \exp\{-x\} d\mu^{\#}(x) = \sum_{n=1}^{\infty} \exp\{-n\} = 1/(e-1).$$

Derivative of Measures

- Motivation
  - let $(\Omega, \mathcal{A}, \mu)$ be a measurable space and let X be a non-negative measurable function on Ω
  - a set function ν as $\nu(A) = \int_A X d\mu = \int I_A X d\mu$ for each $A \in \mathcal{A}$.
  - ν is a measure on $(\Omega, \mathcal{A})$
  - observe $X = d\nu/d\mu$

- Absolute continuity

Definition 2.5 If for any $A \in \mathcal{A}$, $\mu(A) = 0$ implies that $\nu(A) = 0$, then ν is said to be *absolutely continuous* with respect to μ, and we write $\nu \prec\prec \mu$. Sometimes it is also said that ν is *dominated* by μ.

- Equivalent conditions

Proposition 2.11 Suppose $\nu(\Omega) < \infty$. Then $\nu \prec\prec \mu$ if and only if for any $\epsilon > 0$, there exists a δ such that $\nu(A) < \epsilon$ whenever $\mu(A) < \delta$.

Proof

"$\Leftarrow$" is clear.

To prove "$\Rightarrow$", suppose there exists ϵ and a set A_n such that $\nu(A_n) > \epsilon$ and $\mu(A_n) < n^{-2}$.

Since $\sum_n \mu(A_n) < \infty$,

$$\mu(\limsup_n A_n) \le \sum_{m\ge n} \mu(A_n) \to 0.$$

$\Rightarrow$ $\mu(\limsup_n A_n) = 0$.

However, $\nu(\limsup_n A_n) = \lim_n \nu(\cup_{m\ge n} A_m) \ge \limsup_n \nu(A_n) \ge \epsilon$. Contradiction!

- Existence and uniqueness of the derivative

Theorem 2.7 (Radon-Nikodym theorem) Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space, and let ν be a measurable on $(\Omega, \mathcal{A})$ with $\nu \prec\prec \mu$. Then there exists a measurable function $X \geq 0$ such that $\nu(A) = \int_A X d\mu$ for all $A \in \mathcal{A}$. X is unique in the sense that if another measurable function Y also satisfies the equation, then $X = Y$, a.e.

- Transformation of integration using derivative

Proposition 2.13 Suppose ν and μ are σ-finite measure defined on a measure space $(\Omega, \mathcal{A})$ with $\nu \prec\prec \mu$, and suppose Z is a measurable function such that $\int Zd\nu$ is well defined. Then for any $A \in \mathcal{A}$,

$$\int_A Zd\nu = \int_A Z\frac{d\nu}{d\mu}d\mu.$$

Proof

(i) If $Z = I_B$ where $B \in \mathcal{A}$, then

$$\int_A Z d\nu = \nu(A \cap B) = \int_{A \cap B} \frac{d\nu}{d\mu} d\mu = \int_A I_B \frac{d\nu}{d\mu} d\mu.$$

(ii) If $Z \geq 0$, find a sequence of simple function Z_n increasing to Z. For Z_n, $\int_A Z_n d\nu = \int_A Z_n \frac{d\nu}{d\mu} d\mu$. Take limits on both sides and apply the MCT.

(iii) For any Z, write $Z = Z^+ - Z^-$.
$\int Z d\nu = \int Z^+ d\nu - \int Z^- d\nu = \int Z^+ \frac{d\nu}{d\mu} d\mu - \int Z^- \frac{d\nu}{d\mu} d\mu = \int Z \frac{d\nu}{d\mu} d\mu.$

Induced Measure

- Definition
  - let X be a measurable function defined on $(\Omega, \mathcal{A}, \mu)$.
  - for any $B \in \mathcal{B}$, define $\mu_X(B) = \mu(X^{-1}(B))$
  - μ_X is called a *measure induced by* X: $(R, \mathcal{B}, \mu_X)$.

- Density function of X
  - $(R, \mathcal{B}, \nu)$ is another measure space (often the counting measure or the Lebesgue measure)
  - suppose μ_X is dominated by ν with the derivative
  - $f \equiv d\mu_X/d\nu$ is called the *density of X with respect to the dominating measure ν*

- Comparison with usual density function
  - $(\Omega, \mathcal{A}, \mu) = (\Omega, \mathcal{A}, P)$ is a probability space
  - X is a random variable
  - if ν is the counting measure, $f(x)$ is in fact the probability mass function of X
  - if ν is the Lebesgue measure, $f(x)$ is the probability density function of X

- Integration using density
  - $\int_\Omega g(X(\omega))d\mu(\omega) = \int_R g(x)d\mu_X(x) = \int_R g(x)f(x)d\nu(x)$
  - the integration of $g(X)$ on the original measure space Ω can be transformed as the integration of $g(x)$ on R with respect to the induced-measure μ_X and can be further transformed as the integration of $g(x)f(x)$ with respect to the dominating measure ν

- Interpretation in probability space
  - in probability space, $E[g(X)] = \int_R g(x)f(x)d\nu(x)$
  - any expectations regarding random variable X can be computed via its probability mass function (ν is counting measure) or density function (ν is Lebesgue measure)
  - in statistical calculation, we do NOT need to specify whatever probability measure space X is defined on, while solely rely on $f(x)$ and ν.

Probability Measure

- A few important reminders
  - a probability measure space $(\Omega, \mathcal{A}, P)$ is a measure space with $P(\Omega) = 1$;
  - random variable (or random vector in multi-dimensional real space) X is any measurable function;
  - integration of X is equivalent to the expectation;

- the density or the mass function of X is the Radon-Nikydom derivative of the X-induced measure with respect to the Lebesgue measure or the counting measure in real space;
- using the mass function or density function, statisticians unconsciously ignore the underlying probability measure space $(\Omega, \mathcal{A}, P)$.

- Cumulative distribution function revisited
  - $F(x)$ is a nondecreasing function with $F(-\infty) = 0$ and $F(\infty) = 1$;
  - $F(x)$ is right-continuous;
  - λ_F, the Lebesgue-Stieljes measure generated by F is exactly the same measure as the one induced by X, i.e., P_X.

Conditional Probability

- A simple motivation
  - the conditional probability of an event A given another event B has two possibilities:
    $P(A|B) = P(A \cap B)/P(B)$
    $P(A|B^c) = P(A \cap B^c)/P(B^c)$;
  - equivalently, A given the event B is a measurable function assigned to the σ-field $\{\emptyset, B, B^c, \Omega\}$,

$$P(A|B)I_B(\omega) + P(A|B^c)I_{B^c}(\omega).$$

- Definition of conditional probability

An event A given a sub-σ-field $\aleph$, $P(A|\aleph)$

– it is a measurable and integrable function on $(\Omega, \aleph)$;

– for any $G \in \aleph$,

$$\int_G P(A|\aleph)dP = P(A \cap G).$$

- Existence and Uniqueness of Conditional Probability Function

Theorem 2.8 The measurable function $P(A|\aleph)$ exists and is unique in the sense that any two functions satisfying the definition are the same almost surely.

Proof

In $(\Omega, \aleph, P)$, define a set function ν on $\aleph$ such that $\nu(G) = P(A \cap G)$ for any $G \in \aleph$.

$\Rightarrow$ ν is a measure and $P(G) = 0$ implies that $\nu(G) = 0$ $\Rightarrow \nu \prec\prec P$.

$\Rightarrow$ By the Radon-Nikodym theorem, there exits a $\aleph$-measurable function X such that $\nu(G) = \int_G X dP$.
$\Rightarrow$ X satisfies the properties (i) and (ii).

Suppose X and Y both are measurable in $\aleph$ and $\int_G X dP = \int_G Y dP$ for any $G \in \aleph$. Choose choose $G = \{X - Y \geq 0\}$ and $G = \{X - Y < 0\}$ $\Rightarrow$ $\int |X - Y| dP = 0$ $\Rightarrow$ $X = Y$, a.s.

- Properties of conditional probability

Theorem 2.9 $P(\emptyset|\aleph) = 0, P(\Omega|\aleph) = 1$ a.e. and

$$0 \leq P(A|\aleph) \leq 1$$

for each $A \in \mathcal{A}$. if $A_1, A_2, ...$ is finite or countable sequence of disjoint sets in $\mathcal{A}$, then

$$P(\cup_n A_n|\aleph) = \sum_n P(A_n|\aleph).$$

Conditional Expectation

- Definition

X given $\aleph$, denoted $E[X|\aleph]$

- $E[X|\aleph]$ is measurable in $\aleph$ and integrable;
- for any $G \in \aleph$, $\int_G E[X|\aleph]dP = \int_G XdP$, equivalently;
  $E[E[X|\aleph]I_G] = E[XI_G], a.e.$
- The existence and the uniqueness of $E[X|\aleph]$ can be shown similar to Theorem 2.8.

- Properties of conditional expectation

Theorem 2.10 Suppose X, Y, X_n are integrable.
(i) If $X = a$ a.s., then $E[X|\aleph] = a$.
(ii) $E[aX + bY|\aleph] = aE[X|\aleph] + b[Y|\aleph]$.
(iii) If $X \leq Y$ a.s., then $E[X|\aleph] \leq E[Y|\aleph]$.
(iv) $|E[X|\aleph]| \leq E[|X||\aleph]$.
(v) If $\lim_n X_n = X$ a.s., $|X_n| \leq Y$ and Y is integrable, then $\lim_n E[X_n|\aleph] = E[X|\aleph]$.
(vi) If X is measurable in $\aleph$, $E[XY|\aleph] = XE[Y|\aleph]$. (vii) For two sub-$\sigma$ fields $\aleph_1$ and $\aleph_2$ such that $\aleph_1 \subset \aleph_2$,

$$E\left[E[X|\aleph_2]|\aleph_1\right] = E[X|\aleph_1].$$

(viii) $P(A|\aleph) = E[I_A|\aleph]$.

Proof

(i)-(iv) be shown directly using the definition.

To prove (v), consider $Z_n = \sup_{m \geq n} |X_m - X|$. Z_n decreases to 0. $\Rightarrow |E[X_n|\aleph] - E[X|\aleph]| \leq E[Z_n|\aleph]$. $E[Z_n|\aleph]$ decreases to a limit $Z \geq 0$.
Remains to show $Z = 0$ a.s. Note $E[Z_n|\aleph] \leq E[2Y|\aleph] \Rightarrow$ by the DCT, $E[Z] = \int E[Z|\aleph]dP \leq \int E[Z_n|\aleph]dP \to 0$. $\Rightarrow Z = 0$ a.s.

For (vii), for any $G \in \aleph_1 \subset \aleph_2$,

$$\int_G E[X|\aleph_2]dP = \int_G XdP = \int_G E[X|\aleph_1]dP.$$

(viii) is clear from the definition of the conditional probability.

To see (vi) holds, consider simple function first, $X = \sum_i x_i I_{B_i}$ where B_i are disjoint set in $\aleph$. For any $G \in \aleph$,

$$\int_G E[XY|\aleph]dP = \int_G XYdP = \sum_i x_i \int_{G\cap B_i} YdP$$

$$= \sum_i x_i \int_{G\cap B_i} E[Y|\aleph]dP = \int_G XE[Y|\aleph]d.$$

$\Rightarrow E[XY|\aleph] = XE[Y|\aleph]$.

For any X, a sequence of simple functions X_n converges to X and $|X_n| \leq |X|$. Then

$$\int_G X_n YdP = \int_G X_n E[Y|\aleph]dP.$$

Note that $|X_n E[Y|\aleph]| = |E[X_n Y|\aleph]| \leq E[|XY||\aleph]$. From the DCT, $\int_G XYdP = \int_G XE[Y|\aleph]dP$.

- Relation to classical conditional density
  - $\aleph = \sigma(Y)$: the σ-field generated by the class $\{\{Y \le y\} : y \in R\}$ $\Rightarrow P(X \in B|\aleph) = g(B, Y)$
  - $\int_{Y \le y_0} P(X \in B|\aleph) dP = \int I(y \le y_0) g(B, y) f_Y(y) dy = P(X \in B, Y \le y_0)$
$$= \int I(y \le y_0) \int_B f(x, y) dx dy.$$
  - $g(B, y) f_Y(y) = \int_B f(x, y) dx \Rightarrow P(X \in B|\aleph) = \int_B f(x|y) dx.$
  - the conditional density of $X|Y = y$ is the density function of the conditional probability measure $P(X \in \cdot|\aleph)$ with respect to the Lebesgue measure.

- Relation to classical conditional expectation
  - $E[X|\aleph] = g(Y)$ for some measurable function $g(\cdot)$
  - $\int I(Y \leq y_0)E[X|\aleph]dP = \int I(y \leq y_0)g(y)f_Y(y)dy$

$$= E[XI(Y \leq y_0)] = \int I(y \leq y_0)xf(x,y)dxdy$$

  - $g(y) = \int xf(x,y)dx/f_Y(y)$
  - $E[X|\aleph]$ is the same as the classical conditional expectation of X given $Y = y$

READING MATERIALS: Lehmann and Casella, Sections 1.2 and 1.3, Lehmann *Testing Statistical Hypotheses*, Chapter 2 (Optional)