Advanced Probability Theory

CONTENTS

I. Chapter 1. Set Theory

A. Sets

Definition: Let $A, A_1, A_2, \cdots \subset \Omega$,

(1) Infinitely often (i.o.)

$$\limsup_n A_n \triangleq \overline{\lim_n} A_n = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n = \{\omega : \forall k \geq 1, \exists n \geq k, \omega \in A_n\}.$$

(2) Ultimately (ult.)

$$\liminf_n A_n \triangleq \underline{\lim_n} A_n = \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n = \{\omega : \exists k \geq 1, \forall n \geq k, \omega \in A_n\}.$$

(3) The sequence $\{A_n\}$ converges to A, written as $A = \lim_{n\to\infty} A_n$ or simply $A_n \to A$ iff

$$\overline{\lim_n} A_n = \underline{\lim_n} A_n = A.$$

Theorem 1.3.1 $\overline{\lim}_n A_n \subset \underline{\lim}_n A_n$.

Theorem 1.3.2 (Monotone sequence of sets converges).

(1) $A_1 \subset A_2 \subset A_3 \subset \cdots \Rightarrow A_n \to A = \cup_{k=1}^{\infty} A_k$, written as $A_n \uparrow A$.

(2) $A_1 \supset A_2 \supset A_3 \supset \cdots \Rightarrow A_n \to A = \cap_{k=1}^{\infty} A_k$, written as $A_n \downarrow A$.

Theorem 1.3.3 $\liminf_n A_n = (\limsup_n A_n^c)^c$.

Definition: The **indicator function** of a set $A \subset \Omega$ is defined as

$$I_A(\omega) = I\{\omega \in A\} = 1 \text{ for } \omega \in A$$
$$= 0 \text{ for } \omega \in A^c.$$

Properties of indicator functions: $\forall A, B \subset \Omega, A_i \subset \Omega, i = 1, 2, \cdots$

1. $A = B \Leftrightarrow I_A \equiv I_B, A \subset B \Leftrightarrow I_A \leq I_B$.
2. $I_{A\cap B} = \min\{I_A, I_B\} = I_A I_B, I_{A\cup B} = \max\{I_A, I_B\} = I_A + I_B - I_A I_B \leq I_A + I_B$.
3. $I_{A^c} = 1 - I_A, I_{A-B} = I_{A\cap B^c} = I_A(1 - I_B), I_{A\Delta B} = |I_A - I_B|$.
4. $I_{\liminf_n A_n} = \liminf_n I_{A_n}, I_{\limsup_n A_n} = \limsup_n I_{A_n}, I_{\cap_1^\infty A_n} \leq \sum_1^\infty I_{A_n}$.

B. Semi-algebras, Algebras, and σ-algebras

Definition 1.5.1 A nonempty class $\mathcal{S}$ of subsets of Ω is an **semi-algebra** on Ω if

(1) $\forall A, B \in \mathcal{S}, A \cap B \in \mathcal{S}$ (*i.e.*, closed under intersection)

(2) $\forall A \in \mathcal{S}, \exists A_i \in \mathcal{S}, A_i \cap A_j = \emptyset$ $(i \neq j), A^c = \sum_{i=1}^n A_i$.

Definition 1.5.2 A nonempty class $\mathcal{S}$ of subsets of Ω is an **algebra** on Ω if

(1) $\forall A_1, A_2 \in \mathcal{S}, A_1 \cup A_2 \in \mathcal{S}$; (2) $\forall A \in \mathcal{S}, A^c \in \mathcal{S}$.

Definition 1.5.3 A nonempty class $\mathcal{S}$ of subsets of Ω is an σ**-algebra** on Ω if

(1) $\forall A_n \in \mathcal{S}(n \geq 1), \bigcup_{i=1}^{\infty} A_n \in \mathcal{S}$; (2) $\forall A \in \mathcal{S}, A^c \in \mathcal{S}$.

The pair $(\Omega, \mathcal{A})$ is called a **measurable space**. The sets of $\mathcal{A}$ are called **measurable sets**.

Remark 1.5.1 If $\mathcal{A}$ is an algebra (or a σ-algebra), then $\emptyset, \Omega \in \mathcal{A}$. However, the same may not hold for semi-algebras.

Remark 1.5.2 $\mathcal{A}$ is an algebra $\Leftrightarrow$ (1) $\Omega \in \mathcal{A}$. (2) $\forall A, B \in \mathcal{A}, A - B \in \mathcal{A}$.

Theorem 1.5.1 If $\mathcal{S}$ is a semi-algebra, then its generated algebra $\overline{\mathcal{S}}$ is {finite disjoint unions of sets in $\mathcal{S}$}.

C. Generated classes

Lemma 1.6.1 Let $\{\mathcal{A}_\lambda : \lambda \in \Lambda\}$ be a collection of σ-algebras. Then $\mathcal{A} = \cap_{\lambda \in \Gamma} \mathcal{A}_\gamma$ is also a σ-algebra (closed under arbitrary intersection).

Theorem 1.6.1 For any class $\mathcal{A}$, there exists a unique minimal σ-algebra containing $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, called the σ-algebra generated by $\mathcal{A}$. In other words, $\mathcal{A} \subset \sigma(\mathcal{A}), \forall \sigma$-algebra $\mathcal{B} \supset \mathcal{A}, \mathcal{B} \supset \sigma(\mathcal{A})$ and $\sigma(\mathcal{A})$ is unique.

Theorem 1.6.2 $\mathcal{S}$ is a semi-algebra, and $\overline{\mathcal{S}}$ is an algebra generated by $\mathcal{S}$. Then $\sigma(\overline{\mathcal{S}}) = \sigma(\mathcal{S})$.

Definition: The smallest σ-algebra generated by the collection of all finite open intervals on $\mathbb{R} = (-\infty, \infty)$ is called the **Borel σ-algebra**, denoted by $\mathcal{B}$. The elements of $\mathcal{B}$ are called **Borel sets**. The pair $(\mathbb{R}, \mathcal{B})$ is called the (1-dimensional) **Borel measurable space**.

Lemma 1.6.2 For $A \in \mathcal{B}$, let $\mathcal{B}_A = \{B \cap A : B \in \mathcal{B}\} = \mathcal{B} \cap A$. Then $(A, \mathcal{B}_A)$ is a measurable space, and $\mathcal{B}_A$ is called the Borel σ-algebra on A.

D. Monotone class, π-class, and λ-class

Definitions: Let $\mathcal{A}$ be a nonempty class of subsets of Ω.

(1) $\mathcal{A}$ is said to be a **monotone class (m-class)** on Ω if $\lim_{n\to\infty} A_n \in \mathcal{A}$ for every monotone sequence $A_n \in \mathcal{A}, n \geq 1$.

(2) $\mathcal{A}$ is a π**-class** on Ω if $A \cap B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$.

(3) $\mathcal{A}$ is a λ**-class** on Ω if

  (i) $\Omega \in \mathcal{A}$,

  (ii) $A - B \in \mathcal{A}$ for $A, B \in \mathcal{A}, B \subset A$.

  (iii) $\lim A_n \in \mathcal{A}$ for every increasing sequence $A_n \in \mathcal{A}, n \geq 1$.

Theorem 1.7.1 If $\mathcal{A}$ is a λ-class, it is an m-class.

Theorem 1.7.2 Suppose $\mathcal{A}$ is an algebra on Ω. Then $\mathcal{A}$ is an m-class $\Leftrightarrow$ $\mathcal{A}$ is a σ-algebra.

Theorem 1.7.3 $\mathcal{A}$ is a σ-algebra iff it is both a λ-class and π-class.

Lemma 1.7.1 The power set is an m-class (or λ-class, or π-class).

Lemma 1.7.2 Let $\{\mathcal{A}_\gamma : \gamma \in \Gamma\}$ be m-classes (or λ-classes, or π-classes). Then $\mathcal{A} = \cap_{\gamma\in\Gamma}\mathcal{A}_\gamma$ is also an m-class (or λ-class, or π-class).

Theorem 1.7.4 For any class $\mathcal{A}$, there exists a unique minimal m-class (or λ-class, or π-class) containing $\mathcal{A}$, denoted by $m(\mathcal{A})$ (or $\lambda(\mathcal{A})$, or $\pi(\mathcal{A})$), called the m-class (or λ-class, or π-class) generated by $\mathcal{A}$.

E. The monotone class theorem

Theorem 1.8.1 Let $\mathcal{A}$ be an algebra. Then

(1) $m(\mathcal{A}) = \sigma(\mathcal{A})$;

(2) If $\mathcal{B}$ is an m-class and $\mathcal{A} \subset \mathcal{B}$, then $\sigma(\mathcal{A}) \subset \mathcal{B}$.

Theorem 1.8.2 Let $\mathcal{A}$ be a π-class. Then

(1) $\lambda(\mathcal{A}) = \sigma(\mathcal{A})$;

(2) If $\mathcal{B}$ is an λ-class and $\mathcal{A} \subset \mathcal{B}$, then $\sigma(\mathcal{A}) \subset \mathcal{B}$.

Theorem 1.8.3 (Monotone Class Theorem) Let $\mathcal{A} \subset \mathcal{B}$ be two classes on Ω.

(1) If $\mathcal{A}$ is a π-class, and $\mathcal{B}$ is a λ-class, then $\sigma(\mathcal{A}) \subset \mathcal{B}$.

(2) If $\mathcal{A}$ is an algebra, and $\mathcal{B}$ is an m-class, then $\sigma(\mathcal{A}) \subset \mathcal{B}$.

F. Product spaces

Definitions: For any measurable spaces $(\Omega_i, \mathcal{A}_i), i = 1, 2, \cdots, n \geq 2$:

1. n-dimensional **rectangles** of the product space of $\prod_{i=1}^n \Omega_i$:
$$\prod_{i=1}^n A_i := A_1 \times \cdots \times A_n = \{(\omega_1, \cdots, \omega_n) : \omega_i \in A_i \subset \Omega_i, 1 \leq i \leq n\}.$$
Moreover, if $A_i \in \mathcal{A}_i, 1 \leq i \leq n$, they are called **measurable rectangles**.
2. n-dimensional **product σ-algebra**:
$$\prod_{i=1}^n \mathcal{A}_i = \sigma\left(\left\{\prod_{i=1}^n A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq n\right\}\right).$$
3. n-dimensional **product measurable space**:
$$\prod_{i=1}^n (\Omega_i, \mathcal{A}_i) = \left(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{A}_i\right).$$

II. Chapter 2. Measure Theory

A. Definitions

Let Ω be a space, $\mathcal{A}$ a class and $\mu : \mathcal{A} \to \mathbb{R} = [-\infty, \infty]$ a **set function**.

1. μ is **finite** on $\mathcal{A}$ if $|\mu(A)| < \infty, \forall A \in \mathcal{A}$.

2. μ is **σ-finite** on $\mathcal{A}$ if $\exists A_n \subset A$, such that for each n, $\bigcup_{i=1}^{\infty} A_i = \Omega$ and $|\mu(A_i)| < \infty$.

3. Assume that $A_n, \sum_1^n A_i, \sum_1^\infty A_i \in \mathcal{A}$ and A_i are disjoint.

μ is **additive** $\Leftrightarrow \mu(\sum_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. (If $\emptyset \in \mathcal{A}$, obviously $\mu(\emptyset)$=0.)

μ is **σ-additive** $\Leftrightarrow \mu(\sum_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$. (If $\emptyset \in \mathcal{A}$, obviously σ-additive $\Rightarrow$ additive.)

4. μ is a **measure** on $\mathcal{A}$ if it is both non-negative and σ-additive.

Definition of measure space: If μ is a measure on a σ-algebra $\mathcal{A}$ of subsets of Ω, the triplet $(\Omega, \mathcal{A}, \mu)$ is a measure space. The sets of $\mathcal{A}$ are called **measurable sets**, or $\mathcal{A}$-measurable. A measure space $(\Omega, \mathcal{A}, P)$ is a **probability space** if $P(\Omega) = 1$.

B. Properties of measure

Theorem 2.2.1 Let $\mathcal{A}$ be a **semi-algebra** including $\emptyset$ and Ω, μ a **nonnegative additive set function** on $\mathcal{A}$. Let $A, B, A_n, B_n (n = 1, 2, \cdots) \in \mathcal{A}$:

(1) Monotonicity: $A \subset B \Rightarrow \mu(A) \leq \mu(B)$

(2) σ-subadditivity: $\sum_{n=1}^\infty A_n \subset A \Rightarrow \sum_{n=1}^\infty \mu(A_n) \leq \mu(A)$,

if μ is σ-additive (measure): $B \subset \sum_{n=1}^\infty B_n \Rightarrow \mu(B) \leq \sum_{n=1}^\infty \mu(B_n)$.

All the properties for semi-algebras also hold for algebras. In addition:

Theorem 2.2.2 (σ-subadditivity) Let μ be a measure on an **algebra** $\mathcal{A}$.

$$A, A_n (n = 1, 2, \cdots) \in \mathcal{A}, A \subset \bigcup_{n=1}^\infty A_n \Rightarrow \mu(A) \leq \sum_{n=1}^\infty \mu(A_n).$$

All the properties for semi-algebras and algebras also hold for σ-algebras. In addition:

Theorem 2.2.3 Let μ be a measure on a σ-algebra $\mathcal{A}$, and $\{A_n\} \in \mathcal{A}$.

(1) (**Monotonicity**) $A_1 \subset A_2 \Rightarrow \mu(A_1) \leq \mu(A_2)$.

(2) (**Countable Sub-Additivity**) $\mu(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu(A_i)$.

(3) (**Continuity from below**) $A_n \nearrow A \Rightarrow \mu(A_n) \to \mu(A)$.

(4) (**Continuity from above**) $A_n \searrow A$ and $\mu(A_m) < \infty$ for some $m \geq 1$ except for $A_m = \emptyset \Rightarrow \mu(A_n) \to \mu(A)$.

(5) (**Continuity**) If μ is a finite measure and $A_n \to A$, then $\mu(A_n) \to \mu(A)$.

C. Probability measure

Definition: For a measure space $(\Omega, \mathcal{A}, \mu)$, if $\mu(\Omega) = 1$, then μ is a **probability measure**, usually written as P, then $(\Omega, \mathcal{A}, P)$ is called a **probability space**.

Properties:

(1) $\forall A \in \mathcal{A}, 0 \leq P(A) \leq 1$;

(2) $P(\sum_{n=1}^{N} A_n) = \sum_{n=1}^{N} P(A_n), P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$;

(3) $P(B - A) = P(B) - P(A)$ if $A \subset B$, $P(A^c) = 1 - P(A)$;

(4) $P(A) \leq P(B)$ if $A \subset B$;

(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B), P(\cup_{k=1}^{n} A_k) = \sum_k P(A_k) - \sum_{i<j} P(A_i \cap A_j) + \cdots$

(6) If $A_n \nearrow A$ or $A_n \searrow A$ or $A_n \to A$, then $P(A_n) \to P(A)$.

D. Extension of set functions from semi-algebras to algebras

Definition: Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of subsets of Ω with $\mathcal{A} \subset \mathcal{B}$. If μ and ν are two set functions (or measures) defined on $\mathcal{A}$ and $\mathcal{B}$, respectively such that $\nu(A) = \mu(A), \forall A \in \mathcal{A}$, ν is said to be an **extension of** μ **from** $\mathcal{A}$ **to** $\mathcal{B}$, and μ the restriction from $\mathcal{B}$ to $\mathcal{A}$.

Theorem 2.5.1 Extend a set function from a semi-algebra $\mathcal{S}$ to its generated algebra $\overline{\mathcal{S}}$

(1) Let μ be a **non-negative additive set function** (or measure) on a semi-algebra $\mathcal{S}$ and $\emptyset \in \mathcal{S}$, then μ has a unique extension $\overline{\mu}$ to $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$, such that $\overline{\mu}$ is additive.

(2) Moreover, if μ is σ-additive on $\mathcal{S}$ (a measure), then so is μ on $\overline{\mathcal{S}}$.

E. Outer measure

Definition of outer measure. Let μ be a measure on a semi-algebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$. For any $A \subset \Omega$ (*i.e.*, defined on the power set), define

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n); A \subset \bigcup_{n=1}^{\infty} A_n, A_n \in \mathcal{S} \right\}$$

to be the outer measure of A. μ^* is called the outer measure induced by the measure μ.

Properties of outer measure:

(1) $\forall A \in \mathcal{S}, \mu^*(A) = \mu(A), \mu^*(\emptyset) = \mu(\emptyset) = 0$.

(2) (**Monotonicity**) $\mu^*(A) \leq \mu^*(B)$ for $A \subset B \subset \Omega$

(3) (**σ-subadditivity**) $\mu^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$, for $\{A_n\} \subset \Omega$.

Definition: A set $A \subset \Omega$ is said to be **measurable w.r.t. an outer measure** μ^* if for any $D \subset \Omega$, one has $\mu^*(D) = \mu^*(A \cap D) + \mu^*(A^c \cap D)$.

Theorem 2.6.2 A set $A \subset \Omega$ is measurable w.r.t. an outer measure μ^* *iff* for any $D \subset \Omega$, one has $\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$.

Theorem 2.6.3 Let $\mathcal{A}^*$ be the class of all μ^*-measurable sets.

(1) $\mathcal{A}^*$ is a σ-algebra.

(2) If $A = \sum_{n=1}^{\infty} A_n$ with $\{A_n\} \in \mathcal{A}^*$, then for any $B \subset \Omega$:

$$\mu^*(A \cap B) = \sum_{n=1}^{\infty} \mu^*(A_n \cap B)$$

(3) $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ is a measure space. Furthermore, $\mu^*|_{\mathcal{A}^*}$ is an extension of μ from $\mathcal{S}$ to $\mathcal{A}^*$ (also a restriction of μ^* from the power set $\mathcal{P}(\mathcal{A})$ to $\mathcal{S}$).

F. Extension of measures from semi-algebras to σ-algebras

Theorem 2.7.1 A semi-algebra $\mathcal{S} \subset \mathcal{A}^* \Rightarrow \sigma(\mathcal{S}) \subset \mathcal{A}^*$.

Theorem 2.7.2 (Caratheodory Extension Theorem) Let μ be a measure on a semi-algebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$,

(1) μ has an extension to $\sigma(\mathcal{S})$, denoted by $\mu|_{\sigma(\mathcal{S})}$, so $(\Omega, \sigma(\mathcal{S}), \mu|_{\sigma(\mathcal{S})})$ is a measure space. Furthermore, $\mu|_{\sigma(\mathcal{S})} = \mu^*|_{\sigma(\mathcal{S})}$, i.e., this extension can be simply taken to be the restriction of measure $\mu^*|_{\mathcal{A}^*}$ to $\sigma(\mathcal{S})$.

(2) If μ is σ-finite, then the extension in (1) is unique.

If P is a probability defined on a semi-algebra $\mathcal{S}$ on Ω, then there exists a unique probability space $(\Omega, \sigma(\mathcal{S}), P^*)$ such that $P^*(A) = P(A), \forall A \in \mathcal{S}$.

G. Completion of a measure

Definition: Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and $N \subset \Omega$.

(1) N is a μ**-null set** iff $\exists B \in \mathcal{A}, \mu(B) = 0$ such that $N \subset B$.

(2) $(\Omega, \mathcal{A}, \mu)$ is a complete measure space if every μ-null set $N \in \mathcal{A}$.

Clearly, a μ-null set $N \subset \Omega$ may not be $\mathcal{A}$-measurable unless $(\Omega, \mathcal{A}, \mu)$ is complete. However, the next theorem shows that any measurable space can always be completed.

Theorem 2.8.1 Given a measure space $(\Omega, \mathcal{A}, \mu)$, there exists a complete space $(\Omega, \mathcal{A}, \mu)$ such that $\mathcal{A} \subset \overline{\mathcal{A}}$ and $\overline{\mu} = \mu$ on $\mathcal{A}$, which can be

$$\overline{\mathcal{A}} = \{A \cup N : A \in \mathcal{A}, N \text{ is a } \mu\text{-null set}\} = \{A \Delta N : A \in \mathcal{A}, N \text{ is a } \mu\text{-null set}\},$$

satisfying $\overline{\mu}(A \Delta N) = \overline{\mu}(A \cup N) = \mu(A)$.

Lemma 2.8.1 Let μ be a measure on a semi-algebra $\mathcal{S}$, and μ^* the outer measure induced by μ. If $A \subset \Omega$, and $\mu^*(A) < \infty$, then $\exists B \in \sigma(\mathcal{S})$ such that

(1) $A \subset B$,

(2) $\mu^*(A) = \mu^*(B)$,

(3) $\forall C \subset B - A$ and $C \in \sigma(\mathcal{S})$, we have $\mu^*(C) = 0$.

(Here, we call B to be a measurable cover of A.)

Theorem 2.8.2 Let μ be a σ-finite measure on a semi-algebra $\mathcal{S}$, μ^* be the outer measure induced by μ, and $\mathcal{A}^*$ the σ-algebra consists of all the μ^*-measurable sets. Then $(\Omega, \mathcal{A}^*, \mu|_{\mathcal{A}^*})$ is the completion of $(\Omega, \sigma(\mathcal{S}), \mu|_{\sigma(\mathcal{S})})$.

H. Construction of measures on a σ-algebra

Theorem 2.9.1 Let μ be a nonnegative set function on a semi-algebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$ and

(1) μ is additive on $\mathcal{S}$, *i.e.*, $\mu(A) = \sum_{i=1}^{n}(A_i)$ whenever $A_n \in \mathcal{S}$ and $A = \sum_{i=1}^{n} A_i \in \mathcal{S}$;

(2) μ is σ-subadditive on $\mathcal{S}$, *i.e.*, $\mu(A) \le \sum_{i=1}^{\infty}(A_i)$ whenever $A, A_n \in \mathcal{S}$ and $A \subset \sum_{n=1}^{\infty} A_n$ (or $A = \sum_{n=1}^{\infty} A_n$ or $A \subset \bigcup_{n=1}^{\infty} A_n$),

Then μ is a measure on $\mathcal{S}$.

Theorem 2.9.2 (Lebesgue-Stieltjes measure) Suppose that F is finite on $(-\infty, \infty)$ (*i.e.*, $\forall |t| < \infty, |F(t)| < \infty$), and F is non-decreasing and right continuous, then there is a unique measure (namely, **L-S measure**) μ on $(\mathbb{R}, \mathcal{B})$ with $\mu((a, b]) = F(b) - F(a), (-\infty \le a < b \le +\infty)$. (When $a = b = \infty$, the right hand is understood to be 0.)

Corollary 2.9.1 (Lebesgue measure) There is a unique measure μ on $(\mathbb{R}, \mathcal{B})$ with $\mu((a, b]) = b - a, (-\infty \le a < b \le \infty)$.

Remarks:

(1) A non-decreasing and right continuous function F is called a L-S measure function.

(2) The (completed) measure μ is called the L-S measure. The (incomplete) measure μ is called the B-L-S measure. (B stands for “Borel”).

(3) If $F(x) = x$, then (the complete) μ is called the Lebesgue measure. (note: the incomplete μ is called Borel measure). Lebesgue measure is not finite since $\mu(R) = \infty$, but it is σ-finite.

(4) F uniquely determines μ, but not visa versa, since we can write $\mu((a, b]) = F(b) - F(a) = (F(b) + c) - (F(a) + c)$. So there is no 1-1 correspondence between the class of all L-S measure functions and the class of all L-S measures.

(5) If we further restrict μ to the measurable space $([0,1], \mathcal{B} \cap [0,1])$, then μ is a probability measure (a uniform probability measure).

(6) When Ω is uncountable (e.g. $\Omega = \mathbb{R}$ or $[0,1]$), it is not possible to find a measure on all subsets of R and still satisfy $\mu((a,b]) = b - a$. This is why it is necessary to introduce σ-fields that are smaller than the power set, but large enough for all practical purposes.

Definition: A real-valued function F on $\mathbb{R}$ is distribution function (d.f.) if

(1) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0, F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

(2) F is non-decreasing, *i.e.*, $F(x) \le F(y)$ if $x \le y$.

(3) F is right continuous, *i.e.*, $F(y) \searrow F(x)$ if $y \searrow x$.

Theorem 2.9.3 (Correspondence theorem) The relation $F(x) = P((-\infty, x]), x \in \mathbb{R}$ establishes a 1-1 correspondence between all d.f.s and all probability measures on $(\mathbb{R}, \mathcal{B})$.

Remarks:

(1) This definition does not involve any random variables. Some authors define such functions involving r.v. to be d.f.s.: Let X be a random variable on $(\Omega, \mathcal{A}, P)$, and let $F_X(x) = P(X \le x)$. Then F is a right continuous, non-decreasing function with $F(-\infty) = 0$ and $F(\infty) = 1$.

(2) These are equivalent: $F(x) = P((-\infty, x])$; $P((a,b]) = F(b) - F(a)$; $P([a,b]) = F(b) - F(a-)$; $P([a,b)) = F(b-) - F(a-)$; $P((a,b)) = F(b-) - F(a)$.

(3) Other properties: $P((-\infty, b)) = F(b-)$, $P(\{a\}) = F(a) - F(a-)$.

(4) The set $S(F) = \{x : F(x+\epsilon) - F(x-\epsilon) > 0, \forall \epsilon > 0\}$ is called the **support** of F, and any $x \in S(F)$ is called a **point of increase**. (a) Each jump point of F belongs to the support and that each isolated point of the support is a jump point; (b) $S(F)$ is a closed set; (c) a discrete d.f. can have support $(-\infty, \infty)$ (*e.g.*, the discrete d.f. with positive jump size at each rational number).

Definitions of different types of distributions:

(1) δ_t is called a **degenerate** d.f. at t if $\delta_t(x) = I\{x \ge t\}$.

(2) F is called **discrete** if it can be represented in the form of $F(x) = \sum_{n=1}^{\infty} p_n \delta_{a_n}(x)$, where $\{a_n, n \ge 1\}$ is a countable set of real numbers, $p_j > 0$ for all $j \ge 1$ and $\sum_{n=1}^{\infty} p_n = 1$.

(3) A d.f F is called **continuous** if it is continuous everywhere.

Lemma 2.9.1 The set of discontinuities of a non-decreasing function is countable. Let $\{a_j\}$ be the countable set of points of jump of a d.f. F and $p_j = F(a_j) - F(a_j-) > 0$ the size at jump at a_j. Consider $F_d(x) = \sum_{j=1}^{\infty} p_j \delta_{a_j}(x)$, it is non-decreasing and right continuous with $F_d(-\infty) = 0, F_d(\infty) = \sum p_j$.

Theorem 2.9.4 Let $F_c(x) = F(x) - F_d(x)$, then it is non-negative, non-decreasing and continuous. $F(x)$ has a unique decomposition to a continuous and a discrete non-negative, non-decreasing functions, *i.e.*, $F(x) = F_c(x) + F_d(x)$. Furthermore, $F(x)$ can be uniquely written as the convex combination of a discrete and a continuous d.f.s, *i.e.*,

$$F(x) = \alpha F_1(x) + (1-\alpha)F_2(x), F_1(x) = \frac{F_d(x)}{F_d(\infty)}, F_2(x) = \frac{F_c(x)}{F_c(\infty)}, \alpha = F_d(\infty).$$

Definition: (1) A function F is called **absolutely continuous** [in $(-\infty, \infty)$ and w.r.t. the Lebesgue measure] iff there exists a function f in L^1 (*i.e.* $\int f(t)dt < \infty$ is defined and finite) such that for every $x < y$,

$$F(y) - F(x) = \int_x^y f(t)dt,$$

Here $f(t)$ is called the density of F. It can be shown that $F'(t) = f(t)$ a.e.

Alternative definition: A d.f. F is called **absolutely continuous** iff there exists a function $f \geq 0$ such that for every x,

$$F(x) = \int_{-\infty}^x f(t)dt,$$

Here $f(t)$ is called the probability density function (p.d.f.).

(2) A function F is called **singular** iff it is continuous, not identically zero, F' exists a.e., and $F'(t) = 0$ a.e.

Theorem 2.9.7 Every d.f. can be written as the convex combination of a discrete, a singular continuous, and an absolutely continuous d.f. Such a decomposition is unique.

Definition of Hausdorff measure:

1. For a subset $U \subset \mathbb{R}^n$, its **diagram** is defined as $|U| = \sup_{x,y \in U}\{\|x-y\|\}$.
2. For any subset $F \subset \mathbb{R}^n$, a countable class $\{U_i : 0 < |U_i| \leq \delta, U_i \subset \mathbb{R}^n\}$ is called a δ-cover of F if $F \subset \cup_{i=1}^{\infty} U_i$.
3. Denote $H_\delta^S(F) = \inf \sum_{i=1}^{\infty} |U_i|^S$ where any $\{U_i\}$ is a δ-cover of F ($\delta > 0$), then $H^S(F) = \lim_{\delta \to 0} H_\delta^S(F)$ is called the S**-dimensional Hausdorff outer measure**. It can uniquely determine a measure, *i.e.*, S**-dimensional Hausdorff measure**.
4. **Theorem**: For any subset $A \subset \mathbb{R}^n$, $\exists d_H (0 \leq d_H \leq n)$ such that $H^S(A) = 0$ when $S > d_H$ and $H^S(A) = +\infty$ when $S < d_H$, such d_H is called the **Hausdorff dimension** of A.

I. Radon-Nikodym Theorem

Definition 2.11.1 Let μ, ν be two measures on the measurable space $(\Omega, \mathcal{F})$, we say ν is **absolutely continuous w.r.t.** μ, written as $\nu << \mu$ if $\mu(A) \Rightarrow \nu(A) = 0$.

Theorem 2.11.1 (Radon-Nikodym Theorem) Given a measurable space (X, Σ), if a measure ν on (X, Σ) is absolutely continuous w.r.t. a σ**-finite** measure μ on (X, Σ), then there is a measurable function f on X and taking values in $[0, \infty)$, such that for any measurable set A,

$$\nu(A) = \int_A f d\mu.$$

The function f is uniquely defined up to a μ-null set, that is, if g is another function which satisfies the same property, then $f = g$, μ-almost everywhere. It is commonly written $d\nu/d\mu$ (**Radon-Nikodym derivative**).

Properties: Let ν, μ, λ be σ-finite measures on the same measure space.

(1) If $\nu << \lambda$ and $\mu << \lambda$ (ν and μ are absolutely continuous w.r.t. λ), then

$$\frac{d(\nu + \mu)}{d\lambda} = \frac{d(\mu)}{d\lambda} + \frac{d(\nu)}{d\lambda}$$

μ-almost everywhere.

(2) If $\nu << \mu << \lambda$,

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$

ν-almost everywhere.

(3) If $\mu << \lambda$ and g is a μ-integrable function, then

$$\int_X g d\mu = \int_X g \frac{d\mu}{d\lambda} d\lambda.$$

(4) If $\mu << \nu$ and $\nu << \mu$, then

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

(5) If ν is a finite signed or complex measure, then

$$\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|.$$

III. Chapter 3. Random Variables

A. *Mappings*

Definition: Let $X : \Omega_1 \to \Omega_2$ be a mapping.

(1) For every subset $B \subset \Omega_2$, the inverse image of B is $X^{-1}(B) = \{\omega : \omega \in \Omega_1, X(\omega) \in B\}$.

(2) For every class $\mathcal{G} \subset \mathcal{P}(\Omega_2)$, the inverse image of $\mathcal{G}$ is $X^{-1}(\mathcal{G}) = \{X^{-1}(B) : B \in \mathcal{G}\}$.

Properties of the inverse image:

1. $X^{-1}(\Omega_2) = \Omega_1, X^{-1}(\emptyset) = \emptyset$;
2. $X^{-1}(B^c) = [X^{-1}(B)]^c, X^{-1}(B_1 - B_2) = X^{-1}(B_1) - X^{-1}(B_2)$ for $B, B_1, B_2 \subset \Omega_2$;
3. $X^{-1}(\cup_{\gamma\in\Gamma} B_\gamma) = \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma), X^{-1}(\cap_{\gamma\in\Gamma} B_\gamma) = \cap_{\gamma\in\Gamma} X^{-1}(B_\gamma)$ for $B_\gamma \subset \Omega_2, \gamma \in \Gamma$;
4. $B_1 \subset B_2 \subset \Omega_2 \Rightarrow X^{-1}(B_1) \subset X^{-1}(B_2)$;
5. If $\mathcal{F}$ is a σ-algebra in Ω_2, then $X^{-1}(\mathcal{F})$ is a σ-algebra in Ω_1.
6. Let $\mathcal{C}$ be a nonempty class in Ω_2, then $X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C}))$.

B. *Measurable mapping*

Definitions: $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces. $X : \Omega_1 \to \Omega_2$ is a **measurable mapping** if $X^{-1}(A) \in \mathcal{A}_1, \forall A \in \mathcal{A}_2$. X is a **measurable function** if $(\Omega_2, \mathcal{A}_2) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. X is a **Borel (measurable) function** if $(\Omega_1, \mathcal{A}_1) = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ and $(\Omega_2, \mathcal{A}_2) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Theorem 3.2.1 $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ is a measurable mapping if $\mathcal{A}_2 = \sigma(\mathcal{C})$ and $X^{-1}(C) \in \mathcal{A}_1, \forall C \in \mathcal{C}$.

Theorem 3.2.2 If $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ and $f : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ are measurable mappings, then $f(X) = f \circ X : (\Omega_1, \mathcal{A}_1) \to (\Omega_3, \mathcal{A}_3)$ is also measurable.

C. *Random variables (vectors)*

Definition: A **random variable (r.v.)** X is a measurable function from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$. We then say that X is $\mathcal{A}$-measurable, or simply write it as $X \in \mathcal{A}$.

Another definition: A random variable (r.v.) X is a measurable mapping from a probability space $(\Omega, \mathcal{A}, P)$ to $(\mathbb{R}, \mathcal{B})$ such that $P(|X| = \infty) = P(\omega : |X(\omega)| = \infty) = 0$.

Theorem 3.3.1 X is a r.v. from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$, (*i.e.*, $X \in \mathcal{A}$)

$\Leftrightarrow \{X \le x\} = X^{-1}([-\infty, x]) \in \mathcal{A}, \forall x \in \mathbb{R}$

$\Leftrightarrow \{X \le x\} = X^{-1}([-\infty, x]) \in \mathcal{A}, \forall x \in \mathcal{D}$ which is a dense subset of $\mathbb{R}$.

Also, $\{X \le x\}$ in the theorem can be replaced by any of the following: $\{X \le x\}, \{X \ge x\}, \{X < x\}, \{X > x\}, \{x < X < y\}$.

Definition: $X = (X_1, \cdots, X_n)$ is a **random vector** if X_k is a r.v. on $(\Omega, \mathcal{A})$ for $1 \le k \le n$.

Theorem 3.3.2 If $X = (X_1, \cdots, X_n)$ is a random vector, then X is a measurable function from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

D. Construction of random variables

Unless stated, all r.v.'s are measurable functions from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ from here on.

Theorem 3.4.1 If X, Y are r.v.'s (*i.e.*, $X, Y \in \mathcal{A}$), so are $aX+bY, X \vee Y = \max\{X, Y\}, X \wedge Y = \min\{X, Y\}, X^2, XY, X/Y (Y \neq 0)$, and $X^+ = \max\{X, 0\}, X^- = \min\{X, 0\}, |X| = X^+ - X^-$.

Theorem 3.4.2 $X_1, X_2, \cdots$ are r.v. on $(\Omega, \mathcal{A})$, (i.e., $X_i \in \mathcal{A}$),

(1) $\sup_n X_n, \inf_n X_n, \overline{\lim}_n X_n$, and $\underline{\lim}_n X_n$ are r.v.'s (*i.e.*, they are all $\in \mathcal{A}$).

(2) If $X(\omega) = \lim_n X_n(\omega)$ for every ω, then X is a r.v. (*i.e.*, $X \in \mathcal{A}$).

(3) If $S(\omega) = \sum_{n=1}^{\infty} X_n(\omega)$ exists for every ω, then S is a r.v. (*i.e.*, $S \in \mathcal{A}$).

Definition of a.s.: Let $X_1, X_2, \cdots$ be a sequence of r.v.'s on a probability space $(\Omega, \mathcal{A}, P)$. Define $\Omega_0 = \{\omega : \lim_n X_n(\omega) \text{ exists}\} = \{\omega : \overline{\lim}_n X_n(\omega) = \underline{\lim}_n X_n(\omega)\}$. Clearly, Ω_0 is measurable. If $P(\Omega_0) = 1$, we say that X_n **converges almost surely (a.s.)** and write $X_n \xrightarrow{a.s.} X$.

Theorem 3.4.3 $X = (X_1, \cdots, X_n)$ is a random n-vector, f is a Borel function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then $f(X)$ is a random m-vector.

E. Approximations of r.v. by simple r.v.s

Theorem 3.5.1 Given a measurable space $(\Omega, \mathcal{A})$,

(1) (Indicator r.v.) If $A \in \mathcal{A}$, the indicator function I_A is a r.v. (Recall: $I_A(\omega) = I\{\omega \in A\}$ indicates whether A occurs or not.)

(2) (Simple r.v.) If $\Omega = \sum_{i=1}^{n} A_i$, where $A_i \in \mathcal{A}$, then $X = \sum_{i=1}^{n} a_i I_{A_i}$ is a r.v. (For simplicity, we assume that $\{a_1, \cdots, a_n\}$ are distinct.)

Theorem 3.5.2 Given a r.v. $X \ge 0$ on $(\Omega, \mathcal{A})$, there exists simple r.v.'s $0 \le X_1 \le X_2 \le \cdots$ with $X_n(\omega) \nearrow X(\omega)$ for every $\omega \in \Omega$. A construction is:

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} < X(\omega) \le \frac{k}{2^n}\right\} + nI\{X(\omega) > n\}.$$

F. σ-algebra generated by random variables

Definition: Let $\{X_\lambda, \lambda \in \Lambda\}$ be a nonempty family of r.v.'s on $(\Omega, \mathcal{A})$ (Λ may not be countable). Define $\sigma(X_\lambda, \lambda \in \Lambda) := \sigma(X_\lambda \in B, B \in \mathcal{B}, \lambda \in \Lambda) = \sigma(X_\lambda^{-1}(\mathcal{B}), \lambda \in \Lambda) = \sigma(\cup_{\lambda \in \Lambda} X_\lambda^{-1}(\mathcal{B}))$, which is called the σ-algebra generated by $\{X_\lambda, \lambda \in \Lambda\}$.

(1) For $\Lambda = \{1, 2, \cdots, n\}$ (n may be ∞), we have

$$\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{B})) = X_i^{-1}(\mathcal{B}) = \{X_i \in \mathcal{B}\},$$

$$\sigma(X_1, \cdots, X_n) = \sigma(\cup_{i=1}^n X_i^{-1}(\mathcal{B})) = \sigma(\cup_{i=1}^n \sigma(X_i)).$$

(2) For $\Lambda = \{1, 2, \cdots\}$, it is easy to check that

$$\sigma(X_1) \subset \sigma(X_1, X_2) \subset \cdots \subset \sigma(X_1, \cdots, X_n),$$

$$\sigma(X_1, X_2, \cdots) \supset \sigma(X_2, X_3, \cdots) \supset \cdots \supset \sigma(X_n, X_{n+1}, \cdots).$$

(3) The σ-algebra $\cap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \cdots)$ is referred to as the **tail σ-algebra** of $X_1, X_2, \cdots$.

An example: The σ-algebra generated by a discrete r.v. Consider a discrete r.v. X taking distinct values $\{x_i, 1 \leq i \leq n\}$ (where n could take ∞) and define $A_i = \{\omega : X(\omega) = x_i\}$. We have the following results.

(1) $A_i = \{\omega : X(\omega) = x_i\}$ constitute a disjoint partition of Ω (mutually exclusive and exhaustive). When this is not satisfied in a general case of generating a σ-algebra, we can use **disjointization** techniques to form a partition first and then apply this theorem.

(2) $\sigma(A_1, A_2, \cdots, A_n) = \sigma(A_0 = \emptyset, A_1, A_2, \cdots, A_n) = \{\cup_{i \in I} A_i, I \subset \{0, 1, 2, \cdots, n\}\}$, if n is finite, $|\sigma(A_1, A_2, \cdots, A_n)| = C_n^0 + C_n^1 + \cdots + C_n^n = 2^n$. When $A_1, A_2, \cdots, A_n$ are not exclusive to each other for all pairs in a general case of generating a σ-algebra, disjointization would generate 2^n mutually exclusive sets so we conjecture that $|\sigma(A_1, A_2, \cdots, A_n)| = 2^{2^n}$.

Theorem 3.6.1 Let $X_1, \cdots, X_n$ be r.v.'s on $(\Omega, \mathcal{A})$. A real function Y on Ω is $\sigma(X_1, \cdots, X_n)$-measurable (or a r.v. on the σ-algebra) iff Y has the form $f(X_1, \cdots, X_n)$, where f is a Borel function on $\mathbb{R}_n$.

G. Distributions and induced distribution functions

Theorem 3.7.1 A r.v. X on $(\Omega, \mathcal{A}, P)$ induces another probability space $(\mathbb{R}, \mathcal{B}, P_X)$ through

$$\forall B \in \mathcal{B}, P_X(B) = P(X^{-1}(B)) = P(X \in B).$$

We say that $P_X(\cdot)$ is the **distribution** of X, and $F_X(x) = P_X((-\infty, x]) = P(X \leq x)$ the **distribution function** of X.

Definition: X and Y are two r.v.s.

(1) X and Y are **identically distributed (i.d.)** if $F_X = F_Y$, denoted by $X \stackrel{d}{=} Y$.

(2) X and Y are **equal almost surely (a.s.)** if $P(X = Y) = 1$, denoted by $X \overset{a.s.}{=} Y$ (which is a much stronger concept than $X \overset{d}{=} Y$).

(3) X is a discrete r.v. if $\exists$ a countable subset C of $\mathbb{R}$ s.t. $P(X \in C) = 1$.

Theorem 3.7.2 X is discrete $\Leftrightarrow F_X$ is discrete.

Definition: $X = (X_1, \cdots, X_n)$ is a random vector. The **distribution** of X is $P_X(B) = P(X^{-1}(B)) = P(X \in B), \forall B \in \mathcal{B}$, and the **distribution function** of X is $F_X(x) = P(X_1 \le x_1, \cdots, X_n \le x_n)$.

Theorem 3.7.3 $X = (X_1, \cdots, X_n)$ is a random vector. Then for any subset $I = \{i_1, \cdots, i_m\}$ of $\{1, \cdots, n\}, m \le n$, we have

$$F_{X_{i_1}, \cdots, X_{i_m}}(x_{i_1}, \cdots, x_{i_m}) = \lim_{x_j \to \infty, j \notin I} F_{X_1, \cdots, X_n}(x_1, \cdots, x_n).$$

Definition: A random vector X is discrete if $\exists$ a countable subset C of $\mathbb{R}^n$ s.t. $P(X \in C) = 1$.

Theorem: A random vector $X = (X_1, \cdots, X_n)$ is discrete iff each $X_k (1 \le k \le n)$ is discrete.

H. Generating random variables with prescribed distributions

Definition: The inverse of a d.f. F, or **quantile function** associated with F, is defined by

$$F^{-1}(u) = \inf\{t : F(t) \ge u\}, \forall u \in (0, 1).$$

Theorem 3.8.1 Let $F^{-1}(u) = \inf\{t : F(t) \ge u\}, \forall u \in (0, 1)$, then

(1) $F^{-1}(u)$ is non-decreasing and left-continuous.

(2) $F^{-1}(F(x)) \le x, \forall x \in \mathbb{R}$.

(3) $F(F^{-1}(u)) \ge u, \forall u \in (0, 1)$.

(4) $F^{-1}(u) \le t \Leftrightarrow u \le F(t)$.

(5) If F is continuous, $F(F^{-1}(u)) = u, \forall u \in (0, 1)$.

Theorem 3.8.2 (Quantile transformation) F is a d.f. on $\mathbb{R}$, $U \sim$ Uniform$(0, 1)$, then the random variable $X := F^{-1}(U) \sim F$.

Based on Theorem 3.8.2, we can generate r.v.'s from uniform distributions such as:

(1) $X = -\ln(1 - U)$ and $X = -\ln U$ follow exponential distributions.

(2) $X_1 = \sqrt{-2\pi \ln U_1} \cos(2\pi U_2)$ and $X_2 = \sqrt{-2\pi \ln U_1} \sin(2\pi U_2)$ are independent r.v.'s following $\mathcal{N}(0, 1)$ where U_1, U_2 follow independent uniform distributions $U(0, 1)$.

Theorem 3.8.3 If a r.v. X has a continuous d.f. F, then $F(X) \sim U(0, 1)$.

IV. Chapter 4. Expectation and Integration

The expectation of a r.v. X on $(\Omega, \mathcal{A}, P)$ is defined from simple r.v.s to general ones.

A. Expectation

Definition: The expectation of a simple r.v. $X = \sum_{i=1}^n a_i I_{A_i}$ with $X = \sum_{i=1}^n A_i = \Omega, A_i \in \mathcal{A}$ is

$$E(X) = \sum_{i=1}^n a_i P(A_i).$$

Lemma 4.1.1 $E(X)$ is well defined in the sense: if $\sum_{i=1}^n a_i I_{A_i} = \sum_{i=1}^m b_i I_{B_i}$ with $\Omega = \sum_{i=1}^n A_i = \sum_{j=1}^m B_j$, then $\sum_{i=1}^n a_i P(A_i) = \sum_{i=1}^m b_i P(B_i)$.

Theorem 4.1.1 (Properties of the expectation of simple r.v.'s) X, Y are simple r.v.'s.

(1) $E(C) = C$ when C is a constant r.v.

(2) **(linearity)** $aX + bY$ is simple and $E(aX + bY) = aE(X) + bE(Y)$.

(3) **(non-negativity)** $X \geq 0 \Rightarrow E(X) \geq 0$.

(4) **(monotonicity)** $X \geq Y \Rightarrow E(X) \geq E(Y)$.

Definition: $X \geq 0$ is a r.v. on $(\Omega, \mathcal{A}, P)$ (according to Theorem 3.5.2).

(a) $E(X) := \lim_{n\to\infty} E(X_n) \leq \infty$, where X_n's are simple, non-negative, and $X_n \nearrow X$.

(b) The expectation of X over the event $A \in \mathcal{A}$ is $E_A(X) := E(XI_A)$.

(c) If $Y \leq 0$ is a r.v. on $(\Omega, \mathcal{A}, P)$, define $E(Y) := -E(-Y)$.

(d) The following notation is often used (for $X \geq 0$ or otherwise):

$$E(X) = \int_\Omega X(\omega)P(d\omega) = \int_\Omega XdP = \int XdP.$$

$$E_A(X) = \int_A X(\omega)P(d\omega) = \int_A XdP.$$

Theorem 4.1.4 (Properties of the expectation of non-negative r.v.'s)

(1) **(linearity)** $X, Y \geq 0, ab \geq 0 \Rightarrow E(aX + bY) = aE(X) + bE(Y)$.

(2) **(non-negativity)** $X \geq 0 \Rightarrow E(X) \geq 0$.

(3) **(monotonicity)** $X \geq Y \geq 0 \Rightarrow E(X) \geq E(Y)$.

Theorem 4.1.5 $X \geq 0$, then $E(X) = 0 \Leftrightarrow X \overset{a.s.}{=} 0$.

Theorem 4.1.6 If $X > 0$ a.s., then $E(X) > 0$. If $E(X) > 0$, then $P(X \geq 0) > 0$.

Theorem 4.1.7 (Fatou's lemma) (Equivalent to monotone convergence theorem)

1. Suppose that $X_n \geq Y$ a.s. for some Y with $E|Y| < \infty$, then $E(\underline{\lim}_n X_n) \leq \underline{\lim}_n E(X_n)$.
2. Suppose that $X_n \leq Y$ a.s. for some Y with $E|Y| < \infty$, then $E(\overline{\lim}_n X_n) \geq \overline{\lim}_n E(X_n)$.

Theorem 4.1.8 (Monotone convergence theorem) Let $X, X_1, X_2, \cdots$ be non-negative r.v.'s,

(1) $X_n \nearrow X \Rightarrow E(X_n) \nearrow E(X)$,

(2) $X_n \searrow X$ and for some $m \geq 1, E(X_m) < \infty \Rightarrow E(X_n) \searrow E(X)$.

Definition: X is a r.v. on $(\Omega, \mathcal{A}, P)$, based on $X = X^+ - X^-$ we can define:

(a) For general r.v. X, if either $E(X^+) < \infty$ or $E(X^-) > -\infty$, then $E(X) := E(X^+) - E(X^-)$. In this case, the expectation of X is said to exist and $E(X) \in [-\infty, \infty]$.

(b) If $E(X^+) = E(X^-) = \infty$ then $E(X)$ is not defined.

(c) X is **integrable** if $E(X) < \infty$.

(d) If X is integrable and $A \in \mathcal{A}$, the expectation of X over A is $E_A(X) = E(XI_A)$.

(e) Denote $L^1 = \{X : E|X| < \infty\}$.

Theorem 4.1.10 (Properties of the expectation of general r.v.'s)

(1) (**Absolute integrability**) $E(X)$ is finite if and only if $E|X|$ is finite.

(2) (**Linearity**) $X, Y \in L^1, a, b \in \mathbb{R} \Rightarrow aX + bY \in L^1, E(aX + bY) = aE(X) + bE(Y)$.

(3) (**Monotonicity**) $X, Y \in L^1, X \leq Y$ (maybe a.s.) $\Rightarrow E(X) \leq E(Y)$.

(4) (**Modulus inequality**) $\forall X \in L^1, |E(X)| \leq E|X|$.

(5) (**σ-additivity over sets**) $A = \sum_{i=1}^{\infty} A_i \Rightarrow E_A(X) = \sum_{i=1}^{\infty} E_{A_i}(X)$.

(6) (**Mean value theorem**) If $a \leq X \leq b$ a.s. on $A \in \mathcal{A}$, then $aP(A) \leq E_A(X) \leq bP(A)$.

(7) (**Integration term by term**) If $\sum_{i=1}^{\infty} E|X_n| < \infty$, then $\sum_{i=1}^{\infty} |X_n| < \infty$ so that $\sum_{i=1}^{\infty} X_n$ converges a.s. and $E(\sum_{i=1}^{\infty} X_n) = \sum_{i=1}^{\infty} E(X_n)$.

(8) $X \overset{a.s.}{=} Y \Rightarrow E(X) = E(Y)$.

Theorem 4.1.12 (Dominated Convergence Theorem) If $X_n \xrightarrow{a.s.} X, |X_n| < Y, E(Y) < \infty$,

$$\lim_{n\to\infty} E(X_n) = E(\lim_{n\to\infty} X_n) = E(X).$$

B. Integration

Definition: Let f be Borel measurable on $(\Omega, \mathcal{A}, \mu)$. The **integral** of f w.r.t. μ is denoted by

$$\int f(\omega)\mu(d\omega) = \int f d\mu = \int f.$$

(1) If $f = \sum_1^n a_{i=1} I_{A_i}$ with $a_i \geq 0$,

$$\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

(2) If $f \geq 0$, define

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu,$$

where $f_n \geq 0$ are simple functions and $f_n \nearrow f$.

(3) For a general function $f = f^+ - f^-$, define

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu$$

if either $\int f^+ d\mu < \infty$ or $\int f^- d\mu < \infty$; otherwise, the integral $\int f d\mu$ is not defined.

(4) f is said to be **integrable** w.r.t. μ if $\int |f| d\mu < \infty$ (or equivalently, $\int f^+ d\mu < \infty$ and $\int f^- d\mu < \infty$). We shall use L^1 to denote all integrable functions.

(5) If either $f \geq 0$ or $f \in L^1$, and $A \in \mathcal{A}$, then the integral of f w.r.t. μ over A is defined by

$$\int_A f d\mu = \int f I_A d\mu = \int f(\omega) I_A(\omega) \mu(d\omega).$$

Properties of integrals

- **(Absolute integrability)** $\int f d\mu$ is finite iff $\int |f| d\mu$ is finite.
- **(Linearity)** If $f, g \geq 0, a, b \geq 0$ or $f, g \in L^1, a, b \in \mathbb{R}$, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.
- **(σ-additivity over sets)** If $A = \sum_{i=1}^{\infty} A_i$, $\int_A f d\mu = \sum_{i=1}^{\infty} \int_{A_i} f d\mu$.
- **(Positivity)** If $f \geq 0$ a.e., then $\int f d\mu \geq 0$.
- **(Monotonicity)** If $f_1 \leq f \leq f_2$ a.e., $\int f_1 \leq \int f \leq \int f_2$.
- **(Mean value theorem)** If $a \leq f \leq b$ a.e. on $A \in \mathcal{A}$, then $a\mu(A) \leq \int_A f d\mu \leq b\mu(A)$.
- **(Modulus inequality)** $|\int f| \leq \int |f|$.
- **(Fatou's Lemma)** If $f_n \geq 0$ a.e., then $\int \liminf_n f_n \leq \liminf_n \int f_n$
- **(Monotone Convergence Theorem)** If $0 \leq f_n \nearrow f$, then $\lim_n \int f_n = \int f = \int \lim_n f_n$.
- **(Dominated Convergence Theorem)** If $f_n \to f$ a.s., $|f_n| \leq g$ a.e. for all n, $\int g < \infty$,

$$\lim_n \int f_n = \int f = \int \lim_n f_n.$$

- **(Integration term by term)** If $\sum_{i=1}^{\infty} \int |f_n| < \infty$, then $\sum_{i=1}^{\infty} |f_n| < \infty$ a.e., so that $\sum_{i=1}^{\infty} f_n$ converges a.e. and $\int \sum_{i=1}^{\infty} f_n = \sum_{i=1}^{\infty} \int f_n$.

Definition: f is a Borel measurable function on $(\Omega, \mathcal{A}, \mu)$.

(a) In the case of $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}, \mu)$, if we write $x = \omega \in \mathbb{R}$, then

$$\int f(\omega) \mu(d\omega) = \int f(x) \mu(dx)$$

is the **Lebesgue-Stieltjes integral of f w.r.t.** μ.

(b) In the case of $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}, \lambda)$, where λ is the Lebesgue measure, then

$$\int f(x) \lambda(dx) = \int f(x) dx$$

is the **Lebesgue integral of** f **w.r.t.** λ. (Note: $\lambda(dx) = dx$.)

(c) Let F be a non-decreasing and right-continuous function on $\mathbb{R}$ (i.e. L-S measure function). It is known that there exists a unique measure μ on the measurable space $(\mathbb{R}, \mathcal{B})$ such that $\mu((a,b]) = F(b) - F(a)$. Then we define the **Lebesgue-Stieljes integral of** f **w.r.t.** F as

$$\int f dF := \int f(x) dF(x) := \int f(x)\mu(dx) = \int f d\mu.$$

(d) In the special case $F(x) = x$, the unique measure μ reduces to the Lebesgue measure λ. As a consequence, the integral in (c) reduces to the **Lebesgue integral of** f:

$$\int f(x)dx = \int f(x)\lambda(dx) = \int f d\mu.$$

Some special cases of L-S integral of the form $\int_B f dG$ where B is a Borel set in $\mathbb{R}$

1. G **is a discrete function** with at most countably many jumps $\{x_1, x_2, \cdots\}$ where $\Delta G(x_n) = G(x_n) - G(x_n-) > 0$. The measure μ will be discrete with positive measure at each jump, so

$$\int_B f dG = \sum_{n:x_n \in B} f(x_n)\Delta G(x_n).$$

2. G **is an absolutely continuous function** with derivative g, then $\mu((s,t]) = \int_s^t g(x)dx$, thus

$$\int_B f dG = \int_B f d\mu = \int_B f(x)g(x)dx.$$

3. G **is a mixture of discrete and absolute continuous functions**

Suppose that $G : [a, \infty) \to \mathbb{R}$, is right-continuous on $[a, \infty)$, and is differentiable on $\mathbb{R}$ except at points in a countably infinite set $\{x_1, x_2, \cdots\}$, where each $x_i > a$. Hence

$$G(t) = G(a) + \int_{(a,t]} g(x)dx + \sum_{n:x_n \le t} \Delta G(x_n),$$

$$\int_{[a,t)} f(t)dG(t) = \int_{(a,t]} f(x)g(x)dx + \sum_{n:a<x_n \le t} f(x_n)\Delta G(x_n).$$

4. G **is a right-continuous function of bounded variation**, then we have $G = G_1 - G_2$, where both G_1 and G_2 are non-decreasing and right-continuous functions. In this case,

$$\int_B f dG = \int_B f dG_1 - \int_B f dG_2.$$

5. Integration by parts formula

If F and G are differentiable functions with respective derivatives f and g, then from calculus we have the following integration by parts formula:

$$F(t)G(t) - F(s)G(s) = \int_{(s,t]} F(x)g(x)dx + \int_{(s,t]} G(x)f(x)dx = \int_{(s,t]} F(x)dG(x) + \int_{(s,t]} G(x)dF(x).$$

Theorem 4.2.1 Let F and G be right-continuous functions of bounded variation. Then:

$$\begin{aligned} F(t)G(t) - F(s)G(s) &= \int_{(s,t]} F(x-)gG(x) + \int_{(s,t]} G(x)dF(x) \\ &= \int_{(s,t]} F(x)dG(x) + \int_{(s,t]} G(x-)dF(x), \end{aligned}$$

and

$$F(t)G(t) - F(s)G(s) = \int_{(s,t]} F(x-)dG(x) + \int_{(s,t]} G(x-)dF(x) + \sum_{n:s<x_n\leq t} \Delta F(x_n)\Delta G(x_n)$$

Tonelli Theorem: Let $f(x,y)$ be a non-negative measurable function on $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q$, then

(1) For almost every $x \in \mathbb{R}^p$, $f(x,y)$ is a measurable function w.r.t. y on $\mathbb{R}^q$.

(2) $F_f(x) = \int_{\mathbb{R}^q} f(x,y)dy$ is a non-negative measurable function on $\mathbb{R}^p$.

(3) $\int_{\mathbb{R}^p} F_f(x)dx = \int_{\mathbb{R}^p} dx \int_{\mathbb{R}^q} f(x,y)dy = \int_{\mathbb{R}^n} f(x,y)dxdy$.

Fubini Theorem: Let $f \in L(\mathbb{R}^n)$, $f(x,y) \in \mathbb{R}^p \times \mathbb{R}^q$, then

(1) For almost every $x \in \mathbb{R}^p$, $f(x,y)$ is a integrable function w.r.t. y on $\mathbb{R}^q$.

(2) $F_f(x) = \int_{\mathbb{R}^q} f(x,y)dy$ is a integrable function on $\mathbb{R}^p$.

(3) $\int_{\mathbb{R}^n} f(x,y)dxdy = \int_{\mathbb{R}^p} dx \int_{\mathbb{R}^q} f(x,y)dy = \int_{\mathbb{R}^q} dy \int_{\mathbb{R}^p} f(x,y)dx$.

C. How to compute expectation

Theorem 4.4.1 (Change of variable formula)

Let X be measurable from $(\Omega, \mathcal{A}, P)$ to $(\Omega_0, \mathcal{A}_0, P_X)$ where $P_X = P \circ X^{-1}$ is the induced probability on X. g is Borel on $(\Omega_0, \mathcal{A}_0)$. Either $g \geq 0$ or $E|g(X)| < \infty$. Then

$$E[g(X)] = \int_{\Omega_0} g(y)P_X(dy).$$

Lemma 4.4.1 Let X be an absolutely continuous r.v. with density function f, i.e., $F_X(x) = \int_{-\infty}^{x} f(t)dt$. Let X be the unique probability measure corresponding to F_X, then

$$P_X(B) = \int_B f d\lambda = \int_B f(x)dx, \forall B \in \mathcal{B},$$

where λ is the Lebesgue measure.

Theorem 4.4.2 Let X be an absolutely continuous r.v. with density function f, i.e., $F_X(x) = \int_{-\infty}^{x} f(t)dt$. Assume g is Borel, then

$$E[g(X)] = \int_{\mathcal{R}} g(x)f(x)dx,$$

provided that $\int_{\mathcal{R}} |g(x)|f(x)dx < \infty$.

Theorem 4.4.3 Let X be a discrete r.v. taking values $x_1, x_2, \cdots$ with probability $P(X = x_k) = p_k$ for $k \geq 1$, and g be a Borel, then

$$E[g(X)] = \sum_{k=1}^{\infty} g(x_k) P(X = x_k) = \sum_{k=1}^{\infty} g(x_k) p_k$$

provided that $\sum_{k=1}^{\infty} |g(x_k)| p_k < \infty$.

D. Relation between expectation and tail probability

Theorem 4.5.1 We have

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

So $E|X| < \infty$ if and only if $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$. (Hints of proof: one can show that $\sum_{n=1}^{\infty} nP(n \leq |X| \leq n+1) = \sum_{n=1}^{\infty} P(|X| \geq n)$).

Corollary 4.5.1 If X takes only integer values then $E|X| = \sum_{n=1}^{\infty} P(|X| \geq n)$.

Theorem 4.5.2 If $Y \geq 0$ then $E(Y) = \int_0^{\infty} P(Y \geq y) dy = \int_0^{\infty} P(Y > y) dy = \int_0^{\infty} [1 - F_Y(y)] dy$.

Corollary 4.5.2 If $Y \geq 0$ and $r > 0$, then

$$E(Y^r) = r \int_0^{\infty} x^{r-1} P(Y \geq x) dx = r \int_0^{\infty} x^{r-1} P(Y > x) dx.$$

Corollary 4.5.3 If Y is integrable, then

$$E(Y) = E(Y^+) - E(Y^-) = \int_0^{\infty} P(Y > x) dx - \int_0^{\infty} P(Y \leq -x) dx.$$

E. Moments and moment inequalities

Definition: Let X be a r.v. and $p > 0$,

- p-th moment: $E(X^p)$
- p-th absolute moment: $E(|X|^p)$
- p-th central moment: $E((X - E(X))^p)$
- p-th absolute central moment: $E|(X - E(X))^p|$
- L^p space is $\{X : E|X|^p < \infty\}$.

Young's inequality: Let h be continuous and strictly increasing function with $h(0) = 0$ and $h(\infty) = \infty$. Let $g = h^{-1}$ (the inverse). Then, for any $a > 0$ and $b > 0$, we have

$$ab \leq \int_0^a h(t) dt + \int_0^b g(t) dt.$$

Lemma 4.6.1 Let $a, b > 0, p, q \geq 1, \frac{1}{p} + \frac{1}{q} = 1$, then $\frac{1}{p} a^p + \frac{1}{q} b^q \geq ab$ with equality iff $a^p = b^q$.

Holder's inequality: Suppose that $p, q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$ and that $E|X|^p < \infty, E|X|^q < \infty$, then $E|XY| < \infty$ and

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}}(E|X|^q)^{\frac{1}{q}}.$$

Cauchy-Schwarz inequality:

$$E|XY| \leq \sqrt{(E|X|^2)(E|Y|^2)}.$$

Lyapunov's inequality:

(1) For $p \geq 1, E(|X|) \leq E(|X|^p)^{1/p}$.

(2) For $0 < r \leq s < \infty, (E|Z|^r)^{1/r} \leq (E|Z|^s)^{1/s}$.

Minkowski's inequality: $p \geq 1$, then $(E|X_1 + \cdots + X_n|^p)^{1/p} \leq (E|X_1|^p)^{1/p} + \cdots + (E|X_n|^p)^{1/p}$.

Jensen's inequality: Let ψ be convex, *i.e.*, $\forall \lambda \in (0, 1)$ and $x, y \in \mathbb{R}$, one has

$$\lambda\psi(x) + (1 - \lambda)\psi(y) \geq \psi(\lambda x + (1 - \lambda)y).$$

Suppose that $E|X| < \infty$ and $E|\psi(X)| < \infty$, then $\psi(E(X)) \leq E(\psi(X))$.

Chebyshev (Markov) inequality: If g is strictly increasing and positive on $(0, \infty)$, $g(x) = g(-x)$, and X is a r.v. such that $E[g(X)] < \infty$, then for each $a > 0$:

$$P(|X| \geq a) \leq \frac{E[g(X)]}{g(a)}.$$

A widely used case is

$$P(|X| \geq a) \leq \frac{E[|X|^p]}{a^p}.$$

Lemma 6.6.1 (**C_r-inequality**) $|x + y|^r \leq C_r(|x|^r + |y|^r)$, where $r > 0$ and $C_r = 1$ if $0 < r < 1$; $C_r = 2^{r-1}$ if $r > 1$.

V. Chapter 5. Independence

A. Definition

Definition: Let $(\Omega, \mathcal{A}, P)$ be a probability space.

- Events $A_1, \cdots, A_n \in \mathcal{A}$ are said to be independent iff for every subset J of $\{1, 2, \cdots, n\}$,
$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$
- Classes (including σ-algebras) $\mathcal{A}_1, \cdots, \mathcal{A}_n \subset \mathcal{A}$ are said to be independent iff for every subset J of $\{1, 2, \cdots, n\}$, and $A_i \in \mathcal{A}_i$,
$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$
- The r.v.'s $X_1, \cdots, X_n$ are said to be independent iff the events $\{X_i \in B_i\}$ are independent for any Borel sets B_i.
- The r.v.'s of an infinite (not necessarily countable) family are said to be independent iff those in every finite subfamily are.
- The r.v.'s of a family are said to be pairwise independent iff every two of them are independent.
- The r.v.'s that are independent and have the same d.f. are called independent and identically distributed (i.i.d.).

B. How to check independence?

Theorem 5.2.1 The r.v.'s $X_1, \cdots, X_n$ are independent iff $F_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = \prod_{i=1}^n F_{X_i}(t_i)$.

Theorem 5.2.2 If $\mathcal{G}$ and $\mathcal{D}$ are independent classes of events, and $\mathcal{D}$ is a π-class, then $\mathcal{G}$ and $\sigma(\mathcal{D})$ are independent.

Theorem 5.2.3 Suppose that $\mathcal{A}_1, \cdots, \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-class. Then $\sigma(\mathcal{A}_1), \cdots, \sigma(\mathcal{A}_n)$ are independent.

Theorem 5.2.4 Discrete r.v.'s $X_1, \cdots, X_n$, taking values in countable set C, are independent iff
$$P(X_1 = a_1, \cdots, X_n = a_n) = \prod_{i=1}^n P(X_i = a_i),\ \forall a_1, \cdots, a_n \in C.$$

Theorem 5.2.5 Let $X = (X_1, \cdots, X_n)$ be an absolutely continuous random vector. Then $X_1, \cdots, X_n$ are independent if the density function satisfies $f_X(t_1, \cdots, t_n) = \prod_{i=1}^n f_{X_i}(t_i)$.

C. Functions of independent r.v.'s

Theorem 5.3.1 If $X_1, \cdots, X_n$ are independent r.v.'s and $g_1, \cdots, g_n$ are Borel measurable functions, then $g_1(X_1), \cdots, g_n(X_n)$ are independent r.v.'s.

Theorem 5.3.2 Let $1 = n_0 \leq n_1 < n_2 < \cdots < n_k = n$, g_j be a Borel measurable function of $n_j - n_{j-1}$ variables. If $X_1, \cdots, X_n$ are independent r.v.'s, then

$$g_1(X_1, \cdots, X_{n_1}), g_2(X_{n_1+1}, \cdots, X_{n_2}), \cdots, g_n(X_{n_{k-1}+1}, \cdots, X_{n_k})$$

are independent.

Theorem 5.3.3 (Convolution) Let X, Y be independent and absolutely continuous. Then $X+Y$ is absolutely continuous and

$$f_{X+Y}(t) = \int_{-\infty}^{+\infty} f_X(t-s) f_Y(s) ds, t \in \mathcal{R}.$$

e.g., for independent $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Theorem 5.3.4 Let X, Y be non-negative and integer-valued. Then for each $n \geq 0$,

$$P(X + Y = n) = \sum_{k=0}^{n} P(X = k) P(Y = n - k).$$

e.g., for independent Poisson distributions $X \sim \mathcal{P}(\lambda_1)$ and $Y \sim \mathcal{P}(\lambda_2)$, $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

Definition The covariance of two random variables X and Y is defined to be $Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$.

Theorem 5.3.5 If X, Y are independent and integrable r.v.'s, then $Cov(X, Y) = 0$.

Theorem 5.3.6 If $X_1, \cdots, X_n$ are independent and all have finite expectations, then

$$E(X_1, \cdots, X_n) = \prod_{i=1}^{n} E(X_i).$$

Theorem 5.3.7 Let $u(x)$ and $v(x)$ be both non-decreasing or both non-increasing functions on $I = (a, b)$ (finite or infinite interval on $\mathbb{R}$), and $P(X \in I) = 1$. Then, $E[u(X)]E[v(X)] \leq E[u(X)v(X)]$, or $Cov(u(X), v(X)) \geq 0$, provided these means exist.

Specially, $(E|X|^r)(E|X|^s) \leq E|X|^{r+s}, r, s \geq 0$.

D. Borel-Cantelli Lemma and Kolmogorov 0-1 Law

Theorem 5.4.1 (Borel-Cantelli Lemma)

(1) $P(\limsup_n A_n) = 0 \Leftarrow \sum_{n \to \infty} P(A_n) < \infty$.

(2) $P(\limsup_n A_n) = 1 \Leftarrow \sum_{n \to \infty} P(A_n) = \infty$ and $A_1, A_2, \cdots$ are (pairwise) independent.

Corollary 5.4.1 (Borel 0-1 Law) Let $\{A_n\}$ be (pairwise) independent, then

$$\begin{aligned} P(\limsup_n A_n) &= 0, \text{ iff } \lim_{n\to\infty}\sum_{i=1}^{n} P(A_i) < \infty \\ &= 1, \text{ iff } \lim_{n\to\infty}\sum_{i=1}^{n} P(A_i) = \infty. \end{aligned}$$

Corollary 5.4.3 If A_n are (pairwise) independent and $A_n \to A$, then $P(A) = 0$ or 1.

Corollary 5.4.4 Let X_n be (pairwise) independent. Then

$$X_n \xrightarrow{a.s.} 0 \Leftrightarrow \sum_{n}^{\infty} P(|X_n| \geq \epsilon) < \infty, \forall \epsilon > 0.$$

(*i.e.*, convergence in probability fast enough implies convergence almost sure.)

Corollary 5.4.5 Let $\{X, X_n, n \geq 1\}$ be (pairwise) i.i.d., then

(1) $E|X| < \infty \Leftrightarrow X_n = o(n)$ a.s.

(2) $E|X|^r < \infty (r > 0) \Leftrightarrow X_n = o(n^{1/r})$ a.s.

Definition: The **tail σ-algebra** (or **remote future)** of a sequence $\{X_n, n \geq 1\}$ of r.v.'s on $(\Omega, \mathcal{A}, P)$ is $\bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \cdots)$. The sets of the tail σ-algebra are called **tail events**, and functions Y measurable relative to the tail σ-algebra are dubbed **tail functions** ($Y^{-1}(\mathcal{B})$ is the tail σ-algebra).

Theorem 5.4.3 (Kolmogorov 0-1 Law) Tail events of a sequence $\{X_n, n \geq 1\}$ of independent r.v.'s have probabilities 0 or 1.

Corollary 5.4.6 Tail functions of a sequence of independent r.v.'s are degenerate (constants a.s.).

Corollary 5.4.7 If $\{X_n, n \geq 1\}$ is a sequence of independent r.v.'s, then $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are degenerate a.s.

VI. Chapter 6. Convergence Concepts

A. *Modes of convergence*

Definitions: Let $X, X_1, X_2, \cdots$ be random variables on $(\Omega, \mathcal{A}, P)$.

(a) $X_n \to X$ almost surely, written as $X_n \xrightarrow{a.s.} X$ if

$$P(\lim_{n\to\infty} X_n = X) = P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

(b) $X_n \to X$ in r-th mean, or in L^r space, where $r > 0$, if $\lim_{n\to\infty} E|X_n - X|^r = 0$.

(c) $X_n \to X$ in probability, written as $X_n \xrightarrow{P} X$, if $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0, \forall \epsilon > 0$.

(d) $X_n \to X$ in distribution, written as $X_n \xrightarrow{D} X$ or $F_{X_n} \Rightarrow F_X$ if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ for all continuity points of } F_X(x).$$

Theorem 6.1.1 The following statements are equivalent:

(a) $X_n \xrightarrow{a.s.} X$;

(b) $\forall \epsilon > 0, \lim_{n\to\infty} P(\cap_{m=n}^{\infty}\{|X_m - X| < \epsilon\}) = 1$;

(c) $\forall \epsilon > 0, \lim_{n\to\infty} P(\cup_{m=n}^{\infty}\{|X_m - X| \geq \epsilon\}) = 0$;

(d) $\forall \epsilon > 0, \lim_{n\to\infty} P(\sup_{m=n}^{\infty}\{|X_m - X| \geq \epsilon\}) = 0$, i.e., $\sup_{m=n}^{\infty} |X_m - X| \xrightarrow{P} 0$;

(e) $\forall \epsilon > 0, P(\overline{\lim}_{n\to\infty}\{|X_n - X| \geq \epsilon\}) = 0$.

B. *Cauchy Criterion*

Theorem 6.2.1 (Cauchy Criterion of a.s.) X_n converges a.s.

$$\Leftrightarrow \forall \epsilon > 0, \lim_{n\to\infty} P(|X_m - X_{m'}| \leq \epsilon, \text{ all } m > m' \geq n) = 1.$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{n\to\infty} P(|X_m - X_{m'}| > \epsilon, \text{ some } m > m' \geq n) = 0.$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{M\to\infty} P\left(\sup_{m,n\geq M} |X_m - X_n| > \epsilon\right) = 0.$$

$$\Leftrightarrow \sup_{m,n\geq M} |X_m - X_n| \xrightarrow{P} 0.$$

C. *Relationships between modes of convergence*

Theorem 6.3.1 Relationships between modes of convergence:

(1) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X, X_n \xrightarrow{L^r} X \Rightarrow X_n \xrightarrow{P} X, X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$.

(2) If $r > s > 0$, then $X_n \xrightarrow{L^r} X \Rightarrow X_n \xrightarrow{L^s} X$.

(3) No other implications hold in general.

Theorem 6.4.1 $X_n \xrightarrow{D} C \Leftrightarrow X_n \xrightarrow{P} C$, where C is a constant.

Lemma 6.4.1 If $X_n \xrightarrow{P} X, |X_n| \leq Y$ a.s. (i.e. $P(|X_n| \leq Y) = 1$) for all n, then $|X| \leq Y$ a.s.

Lemma 6.4.2 If $E|Y| < \infty$, and $\lim_{n\to\infty} P(A_n) = 0$, then $E_{A_n}|Y| \to 0$.

Theorem 6.4.2 (Lebesgue Dominated Convergence Theorem) If $X_n \xrightarrow{P} X, |X_n| \leq Y$ a.s. for all n, and $E(Y^r) < \infty$ for $r > 0$, then $X_n \xrightarrow{L^r} X$, which in turn implies that $E(X_n^r) \to E(X^r)$.

Theorem 6.4.4 (Another Fatou's Lemma) If $X_n \geq 0$ and $X_n \xrightarrow{P} X$, then $E(X) \leq \underline{\lim}_n E(X_n)$.

Corollary 6.4.1 (Bounded convergence in probability implies mean convergence) If $P(|X_n| \leq C) = 1$ for all n and some constant C, then $X_n \xrightarrow{P} X \Leftrightarrow X_n \xrightarrow{L^r} X$ for all $r > 0$.

Theorem 6.4.5 (Dominated convergence a.s. implies mean convergence) If $X_n \xrightarrow{a.s.} X, P(|X_n| \leq Y) = 1$ for all n, and $E(Y^r) < \infty$ for $r > 0$, then $X_n \xrightarrow{L^r} X$.

Theorem 6.4.6 (Convergence in probability sufficiently fast implies a.s. convergence) If $\sum_{n=1}^\infty P(|X_n - X| > \epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$. $\{X_n, n \geq 1\}$ is said to **converge completely** to X if $\sum_{n=1}^\infty P(|X_n - X| > \epsilon) < \infty$ for all $\epsilon > 0$.

Theorem 6.4.7 (Convergence in mean sufficiently fast implies a.s. convergence)

If $\sum_{n=1}^\infty E|X_n - X|^r < \infty$ for some $r > 0$, then $X_n \xrightarrow{a.s.} X$.

Theorem 6.4.8 (Convergence sequences in probability contains a.s. subsequences)

If $X_n \xrightarrow{P} X$, then there exists a non-random integers $n_1 < n_2 < \cdots$ such that $X_{n_i} \xrightarrow{a.s.} X$.

Theorem 6.4.11 (Skorokhod's representation theorem) Suppose that $X_n \xrightarrow{P} X$. Then there exist r.v.'s Y and $\{Y_n, n \geq 1\}$ on $((0,1), \mathcal{B}_{(0,1)}, P_\lambda = \lambda_{(0,1)})$ ($\mathcal{B}_{(0,1)}$ denotes the Borel sets in $(0,1)$, $\lambda_{(0,1)}$ is the Lebesgue measure restricted to $(0,1)$) s.t. $Y \overset{d}{=} X, Y_n \overset{d}{=} X_n, Y_n \xrightarrow{a.s.} Y$.

D. Uniform integrability and convergence of moments

Definition: A sequence of r.v.'s $\{Y_n, n \geq 1\}$ on $(\Omega, \mathcal{A}, P)$ is **uniformly integrable u.i.** if and only if $\lim_{C\to\infty} \sup_{n\geq 1} E[|Y_n| I\{|Y_n| \geq C\}] = 0$.

Theorem 6.5.1 (An equivalent definition of u.i.): (1) $\sup_{n\geq 1} E|Y_n| < \infty$. (2) $\forall \epsilon > 0, \exists \delta > 0$ such that for any $A \in \mathcal{A}$ s.t. $P(A) < \delta$, $\sup_n E_A|Y_n| \equiv \sup_n E[|Y_n| I_A] < \epsilon$.

Theorem 6.5.2 (Properties of u.i.)

- $\{X_n\}$ is u.i. iff $\{|X_n|\}$ is u.i.
- If $|X_n| \leq |Y_n|$, and $\{Y_n\}$ is u.i., then $\{X_n\}$ is u.i.
- $\{X_n\}$ is u.i. iff $\{X_n^+\}$ and $\{X_n^-\}$ are both u.i.
- If $\{X_n\}$ and $\{Y_n\}$ are each u.i., so is $\{X_n + Y_n\}$.
- If $\{X_n\}$ is u.i., so is any subsequence of $\{X_n\}$.

- If $|X_n| \le Y \in L^1$, then $\{X_n\}$ is u.i. (i.e., the Lesbegue DCT.)
- If $E(\sup_n |X_n|) < \infty$, then $\{X_n\}$ is u.i.
- Let $\psi > 0$ satisfy $\lim_{x\to\infty} \frac{\psi(x)}{x} = \infty$. If $\sup_n E[\psi(|X_n|)] < \infty$, then $\{X_n\}$ is u.i.

Theorem 6.5.3 (Vitali's Theorem) Suppose that $X_n \xrightarrow{P} X$, and $E|X_n|^r < \infty$ for all n (i.e. $X_n \in L^r$). Then the following three statements are equivalent:
(1) $\{X_n^r\}$ is u.i.; (2) $X_n \xrightarrow{L^r} X, E|X|^r < \infty$; (3) $\lim_n E|X_n|^r = E|X|^r < \infty$.

Theorem 6.5.4 Suppose that $X_n \xrightarrow{L^r} X(r > 0)$, and $E|X|^r < \infty$. Then

$$\lim_{n\to\infty} E|X_n|^r = E|X|^r, \lim_{n\to\infty} E(X_n^r) = E(X^r).$$

Theorem 6.5.5 (Converge in dist. + u.i. $\Rightarrow$ converge in L^r) Suppose that $X_n \xrightarrow{D} X$, and $\{X_n^r\}(r > 0)$ is u.i. Then $E|X|^r < \infty, \lim_{n\to\infty} E|X_n|^r = E|X|^r, \lim_{n\to\infty} E(X_n^r) = E(X^r)$.

E. Some closed operations of convergence

Theorem 6.6.1 (Closed under addition)

- $X_n \xrightarrow{a.s.} X, Y_n \xrightarrow{a.s.} Y \Rightarrow X_n \pm Y_n \xrightarrow{a.s.} X \pm Y$.
- $X_n \xrightarrow{L^r} X, Y_n \xrightarrow{L^r} Y \Rightarrow X_n \pm Y_n \xrightarrow{L^r} X \pm Y$.
- $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$.
- However, if $X_n \xrightarrow{D} X, Y_n \xrightarrow{D} Y$, then it is not true in general that $X_n \pm Y_n \xrightarrow{D} X \pm Y$.

Theorem 6.6.2 (Continuous mapping theorem) Let $X_1, X_2, \cdots$ and X be k-dim random vectors, $g : \mathbb{R}^k \to \mathbb{R}$ be continuous. Then (1) $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$, (2) $X_n \xrightarrow{D} X \Rightarrow g(X_n) \xrightarrow{D} g(X)$, (3) $X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$. (Also holds when g is continuous a.s.)

Theorem 6.6.4 (Slutsky's Theorem) Let $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{P} C$ (constant). Then
(1) $X_n + Y_n \xrightarrow{D} X + C$; (2) $X_n Y_n \xrightarrow{D} CX$; (3) $X_n/Y_n \xrightarrow{D} X/C$ if $C \neq 0$.

F. Simple limit theorems

Theorem 6.7.1 Suppose that X_i's are uncorrelated and $\sup_{k\ge1} E(X_k^2) \le M < \infty$. Denote $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \mu_i = E(X_i)$ and $\overline{\mu} = \frac{1}{n}\sum_{i=1}^n \mu_i$. Then $\overline{X} - \overline{\mu} \xrightarrow{L^2} 0, \overline{X} - \overline{\mu} \xrightarrow{P} 0, \overline{X} - \overline{\mu} \xrightarrow{a.s.} 0$.

G. General Fatou's Lemma

Let $g(\cdot) \ge 0$ be continuous. If $X_n \to X$ in **any mode** (i.e., in probability, or distribution, or L^r, or a.s.), then $E[g(X)] \le \liminf_{n\to\infty} E[g(X_n)], E[g(X)] \ge \limsup_{n\to\infty} E[g(X_n)]$.

VII. Chapter 7. Weak Law of Large Numbers

A. *Equivalent sequences*

Definition: Two sequences of r.v.'s $\{X_n\}$ and $\{Y_n\}$ on $(\Omega, \mathcal{A}, P)$ are said to be **equivalent** if

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty.$$

Theorem 7.1.1 Suppose that $\{X_n\}$ and $\{Y_n\}$ are equivalent.

(1) $\sum_{n=1}^{\infty}(X_n - Y_n)$ converges a.s. (2) If $a_n \uparrow \infty$ then $\frac{1}{a_n}\sum_{k=1}^{n}(X_k - Y_k) \xrightarrow{a.s.} 0$.

Corollary 7.1.1 Suppose that $\{X_n\}$ and $\{Y_n\}$ are equivalent, and $a_n \uparrow \infty$. Then with probability one (a.s.) (1) $\sum_{k=1}^{n} X_k$ or $\frac{1}{n}\sum_{k=1}^{n} X_k$ converges, diverges to $+\infty$ or $-\infty$ or fluctuates in the same way as $\sum_{k=1}^{n} Y_k$ or $\frac{1}{n}\sum_{k=1}^{n} Y_k$. (2) In particular, if $\frac{1}{n}\sum_{k=1}^{n} X_k$ converges in probability, so does $\frac{1}{n}\sum_{k=1}^{n} Y_k$.

B. *Weak Law of Large Numbers (WLLN)*

Theorem 7.2.1 Let $\{X_i\}$ be **pairwise** independent and identically distributed r.v.'s with finite mean $\mu = E(X_1)$. Then $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$.

Theorem 7.2.2 Let $\{X_i\}$ be **pairwise** independent and identically distributed r.v.'s such that $E[X_1 I(|X_1| \leq n)] \to 0, nP(|X_1| > n) \to 0$. Then $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$.

Theorem 7.2.3 (Khinchine's WLLN) Let $\{X_i\}$ be i.i.d. r.v.'s with finite mean $\mu = E(X_1)$. Then $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$.

Theorem 7.2.4 (Chebyshev's WLLN) Let $\{X_i\}$ be **pairwise** independent r.v.'s such that $\forall i \in \mathbb{N}_+, E(X_i) < \infty, Var(X_i) \leq M < \infty$, then

$$\forall \epsilon > 0, \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1,$$

i.e., $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \lim_{n\to\infty}\sum_{k=1}^{n} E(X_k)$.

Theorem 7.2.5 (Markov's WLLN) Let $\{X_i\}$ be r.v.'s with finite $E(X_i)$ and

$$\lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = 0,$$

then

$$\forall \epsilon > 0, \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1,$$

i.e., $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \lim_{n\to\infty}\sum_{k=1}^{n} E(X_k)$.

VIII. Chapter 8. Strong Convergence

A. *Some maximal inequalities*

Theorem 8.1.1 (Hajek-Renyi maximal inequality) Let $X_1, X_2, \cdots$ be independent with $E(X_k) = 0$ and $\sigma_k^2 = Var(X_k) < \infty$. Write $S_k = \sum_{i=1}^k X_i$. Let $\{c_k\}$ be a positive and non-increasing sequence (i.e. $c_k > 0$ and $c_k \downarrow$). Then $\forall \epsilon > 0$, and $m < n$, we have

$$P\left(\max_{m \le k \le n} c_k |S_k| \ge \epsilon\right) \le \frac{1}{\epsilon^2}\left[c_m^2 \sum_{k=1}^m \sigma_k^2 + \sum_{k=m+1}^n c_k^2 \sigma_k^2\right].$$

Theorem 8.1.2 (Kolmogorov maximal inequality) Let $X_1, X_2, \cdots$ be independent with $E(X_k) = 0$ and $\sigma_k^2 = Var(X_k) < \infty$. Write $S_k = \sum_{i=1}^k X_i$. Let $\epsilon > 0$,

(a) (Upper bound)

$$P\left(\max_{1 \le k \le n} |S_k| \ge \epsilon\right) \le \frac{Var(S_n)}{\epsilon^2}.$$

(b) (Lower bound) If $|X_k| \le C \le \infty$, then $\forall k \ge 1$,

$$P\left(\max_{1 \le k \le n} |S_k| \ge \epsilon\right) \ge 1 - \frac{(\epsilon + C)^2}{Var(S_n)}.$$

B. *The a.s. convergence of series and three-series theorem*

Definitions of Cauchy convergence of r.v.'s:

1. The sequence $\{X_n, n \ge 1\}$ is almost sure (a.s.) Cauchy convergent

$$\Leftrightarrow \lim_{m,n\to\infty} P(|X_m - X_n| = 0) = 1;$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{M\to\infty} P(\sup_{m,n \ge M} |X_m - X_n| \le \epsilon) = 1;$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{M\to\infty} P(\sup_{m,n \ge M} |X_m - X_n| > \epsilon) = 0;$$

$$\Leftrightarrow \sup_{m,n \ge M} |X_m - X_n| \xrightarrow{P} 0 \text{ as } M \to \infty;$$

$$\Leftrightarrow \sup_{m > n} |X_m - X_n| \xrightarrow{P} 0 \text{ or } = o_p(1) \text{ as } n \to \infty.$$

2. The sequence $\{X_n, n \ge 1\}$ is Cauchy convergent in probability

$$\Leftrightarrow \forall \epsilon > 0, \lim_{m,n\to\infty} P(|X_m - X_n| \le \epsilon) = 1;$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{m,n\to\infty} P(|X_m - X_n| > \epsilon) = 0;$$

$$\Leftrightarrow \forall \epsilon > 0, \lim_{n\to\infty} \sup_{m > n} P(|X_m - X_n| > \epsilon) = 0;$$

$$\Leftrightarrow \forall \epsilon > 0, \sup_{m > n} P(|X_m - X_n| > \epsilon) = o(1) \text{ as } n \to \infty.$$

3. The sequence $\{X_n, n \geq 1\}$ is mean square Cauchy convergent iff $\lim_{m,n\to\infty} E|X_m - X_n|^2 = 0$.

Theorem 8.2.1 $X_n \xrightarrow{a.s.} X \Leftrightarrow \{X_n\}$ is a.s. Cauchy convergent.

Theorem 8.2.2 $X_n \xrightarrow{P} X \Leftrightarrow \{X_n\}$ is Cauchy convergent in probability.

Theorem 8.2.3 $X_n \xrightarrow{L^2} X \Leftrightarrow \{X_n\}$ is mean square Cauchy convergent.

Theorem 8.2.4 (Variance criterion for series) Let $X_1, X_2, \cdots$ be independent with variances $\sigma_k^2 = Var(X_k) < \infty$. If $\sum_{k=1}^{\infty} Var(X_k) < \infty$, then $\sum_{i=1}^{\infty} X_i$ converges a.s.

Corollary 8.2.1 (Kolmogorov SLLN) Let $X_1, X_2, \cdots$ be independent, a non-decreasing sequence $b_n \nearrow \infty$ and $\sum_{n=1}^{\infty} \frac{Var(X_n)}{b_n} < \infty$, then $\frac{1}{b_n} \sum_{i=1}^{n} (X_i - E(X_i)) \xrightarrow{a.s.} 0$.

Theorem 8.2.5 (Kolmogorov three series theorem) Let $X_1, X_2, \cdots$ be independent. Let $Y_n = X_n I(|X_n| < A)$, then $\sum_{k=1}^{\infty} X_k$ converge a.s. $\Leftrightarrow$ for some $A > 0$:

(1) $\sum_{n=1}^{\infty} P(|X_n| > A) = \sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$;

(2) $\sum_{n=1}^{\infty} E(Y_n)$ converges;

(3) $\sum_{n=1}^{\infty} Var(Y_n) < \infty$. (Variance criterion for truncated r.v.)

Theorem 8.2.6 (Kolmogorov two series theorem) Let $X_1, X_2, \cdots$ be independent. Then $\sum_{k=1}^{\infty} |X_k|$ converge a.s. $\Leftrightarrow$ for some $C > 0$,

(1) $\sum_{n=1}^{\infty} P(|X_n| \geq C) < \infty$;

(2) $\sum_{n=1}^{\infty} E|X_n| I(|X_n| < C) < \infty$.

Theorem 8.2.7 (mean convergence implies a.s. convergence) $\{X_n\}$ are independent r.v.'s. Then $\sum_{n=1}^{\infty} E|X_n|^r < \infty, (0 < r \leq 1) \Rightarrow \sum_{n=1}^{\infty} E|X_n|$ converges a.s.

Theorem 8.2.8 If $\{X_n\}$ is a sequence of non-negative, integrable r.v.'s, and $\sum_{n=1}^{\infty} E(X_n) < \infty$, then $S_n = \sum_{k=1}^{\infty} X_k$ converges a.s.

C. Strong Laws of Large Numbers (SLLN)

Lemma 8.3.1 (Cesaro's Lemma) Given two sequences $\{b_n\}, \{x_n\}$, assume that (1) $b_n \geq 0, a_n = \sum_{k=1}^{n} b_k \nearrow \infty$, (2) $\lim_{n\to\infty} x_n = x, |x| < \infty$. Then

$$\frac{1}{a_n} \sum_{k=1}^{n} b_k x_k \equiv \frac{\sum_{k=1}^{n} b_k x_k}{\sum_{k=1}^{n} b_k} \to x.$$

Lemma 8.3.2 (Abel's method of summation, "integration by parts") $\{a_n\}, \{x_n\}$ are two sequences with $a_0 = 0, S_k = \sum_{j=1}^{k} x_k, S_0 = 0$. Then

$$\sum_{k=1}^{n} a_k x_k = a_n S_n - \sum_{k=1}^{n} (a_k - a_{k-1}) S_{k-1}.$$

Lemma 8.3.3 If $a_n \nearrow \infty$ and $\sum_{n=1}^{\infty} x_n$ converges, then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} a_k x_k = 0.$$

Corollary 8.3.2 (Kronecker Lemma) If $a_n \nearrow \infty, \sum_{n=1}^{\infty} \frac{y_n}{a_n}$ converges, then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} y_k = 0.$$

Theorem 8.3.1 Let $\{X_n\}$ be independent r.v.'s. Assume that

(1) $\{g_n(x)\}$ are even functions, positive and non-decreasing for $x > 0$. Assume for all n, at least one of the following holds:

(i) $\frac{x}{g_n(x)} \nearrow$ for $x > 0$.

(ii) $\frac{x}{g_n(x)} \searrow$ and $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0; E(X_n) = 0$.

(iii) $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0$; X_n has a symmetric density function about 0.

(2) $\{a_n\}$ is a positive sequence, and $\sum_{n=1}^{\infty} \frac{E[g_n(X_n)]}{g_n(a_n)} < \infty$.

Then we have $\sum_{n=1}^{\infty} \frac{X_n}{a_n}$ converges a.s. If we further assume that $0 < a_n \nearrow \infty$, then $\frac{1}{a_n} \sum_{j=1}^{n} X_j \xrightarrow{a.s.} 0$.

Corollary 8.3.3 Let $\{X_n\}$ be independent r.v.'s. and $0 < a_n \nearrow \infty$. Assume that

$$\sum_{n=1}^{\infty} E\left|\frac{X_n}{a_n}\right|^r = \sum_{n=1}^{\infty} \frac{E\left|X_n\right|^r}{a_n^r} < \infty, 0 < r \leq 2.$$

Then we have

$$\frac{1}{a_n} \sum_{j=1}^{n} X_j \xrightarrow{a.s.} 0 \text{ if } 0 < r \leq 1;$$

$$\frac{1}{a_n} \sum_{j=1}^{n} (X_j - E(X_j)) \xrightarrow{a.s.} 0 \text{ if } 1 \leq r \leq 2.$$

Theorem 8.3.4 (Kolmogorov SLLN for iid r.v.'s) Let $X_1, X_2, \cdots$ be i.i.d. r.v.'s, then

$$E|X_1| < \infty \Leftrightarrow \frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{a.s.} E(X_1).$$

$$E|X_1| = \infty \Rightarrow \limsup_{n} \frac{1}{n} \left|\sum_{k=1}^{n} X_k\right| \stackrel{a.s.}{=} \infty.$$

Theorem 8.3.5 (Marcinkiewicz SLLN for i.i.d. r.v.'s) Let $\{X_n\}$ be i.i.d. r.v.'s and $0 < r < 2$. Then

$$\frac{1}{n^{1/r}} \sum_{k=1}^{n} (X_k - a) \xrightarrow{a.s.} 0$$

if and only if $E|X_1|^r < \infty$, where $a = E(X_1)$ if $1 \leq r < 2$, a can be arbitrary if $0 < r < 1$.

IX. Chapter 9. Weak Convergence

A. Definition of weak convergence

Definitions:

(a) A sequence of d.f.s $\{F_n, n \geq 1\}$ is said to **converge weakly** to a d.f. F, written as $F_n \Rightarrow F$, if $F_n(x) \Rightarrow F(x)$, for all $x \in C(F)$ (continuous points).

(b) A sequence of random variables (r.v.s) X_n is said to **converge weakly** or **in distribution** or **in law** to a limit X, written as $X_n \Rightarrow X$ or $X_n \xrightarrow{D} X$, if their d.f.s $F_n(x) = P(X_n \leq x)$ converge weakly to $F(x) = P(X \leq x)$.

Theorem 9.2.1 (Portmanteau Theorem) The following statements are equivalent.

(1) $X_n \Rightarrow X$, i.e., $X_n \xrightarrow{D} X$.

(2) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ for all open sets G.

(3) $\limsup_{n\to\infty} P(X_n \in K) \leq P(X \in K)$ for all closed sets K.

(4) $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$ for all sets A with $P(X \in \partial A) = 0$.

(5) $E[g(X_n)] \to E[g(X)]$ for all bounded continuous function g.

(6) $E[g(X_n)] \to E[g(X)]$ for all functions g of the form $g(x) = h(x)I_{[a,b]}(x)$ where $h(x)$ is continuous on $[a, b]$ and $a, b \in C(F)$.

(7) $\lim_{n\to\infty} \psi_n(t) = \psi(t)$ where $\psi_n(t)$ and $\psi(t)$ are the c.f.s of X_n and X, respectively.

B. Helly's selection theorem and tightness

Theorem 9.3.1 (Helly's Selection Theorem) For every sequence of d.f.'s F_n, there exists a subsequence F_{n_k} and a right continuous function F so that $\forall x \in C(F), \lim_{k\to\infty} F_{n_k}(x) = F(x)$.

Definition: A sequence of d.f.'s $\{F_n, n \geq 1\}$ is said to be **tight** if, for all $\epsilon > 0$, there is an $M > 0$ (free of n) so that $\limsup_{n\to\infty}[1-F_n(M)+F_n(-M)] = \limsup_{n\to\infty} P(|X_n| > M) \leq \epsilon$.

Theorem 9.3.2 Every subsequential limit is a d.f. iff the sequence $\{F_n, n \geq 1\}$ is tight.

Theorem (Integration to the limit) Suppose g, h are continuous with $g(x) > 0$, and $|h(x)|/g(x) \to 0$ as $|x| \to \infty$. If $F_n \Rightarrow F$ and $\int g(x)dF_n(x) \leq C < \infty$ then

$$\lim_{n\to\infty} \int h(x)dF_n(x) = \int h(x)dF(x).$$

Theorem 9.4.1 (Polya Theorem) If $F_n \Rightarrow F$, and F is continuous, then the point-wise weak convergence holds uniformly:

$$\lim_{n\to\infty} \sup_t |F_n(t) - F(t)| = 0.$$

X. Chapter 10. Charateristic Functions

A. *Definitions and examples*

Definition: The characteristic function (c.f.) for a random variable (r.v.) X in $\mathbb{R}$ with distribution function (d.f.) F is defined to be

$$\psi(t) = \psi_X(t) = E(e^{itX}) = \int_{\mathbb{R}} e^{itx} dF(x) = E(\cos(tX)) + iE(\sin(tX)).$$

1. Standard normal (Gaussian) distribution:
$$\text{p.d.f. } f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \Rightarrow \text{ c.f. } \psi(t) = e^{-t^2/2}.$$
2. Uniform distribution $[0, a]$:
$$\text{p.d.f. } f(x) = \frac{1}{a} I(0 \le x \le a) \Rightarrow \text{ c.f. } \psi(t) = \frac{e^{iat} - 1}{iat}.$$
3. Uniform distribution $[-a, a]$:
$$\text{p.d.f. } f(x) = \frac{1}{2a} I(-a \le x \le a) \Rightarrow \text{ c.f. } \psi(t) = \frac{\sin(at)}{at}.$$
4. Triangular distribution on $[-a, a]$:
$$\text{p.d.f. } f(x) = \frac{1}{a}\left(1 - \frac{|x|}{a}\right) I\{|x| < a\} \Rightarrow \text{ c.f. } \psi(t) = \frac{2(1 - \cos(at))}{a^2 t^2}.$$
5. Inverse triangular distribution on $[-a, a]$:
$$\text{p.d.f. } f(x) = \frac{1 - \cos(ax)}{\pi a x^2} \Rightarrow \text{ c.f. } \psi(t) = \left(1 - \frac{|t|}{a}\right) I\{|t| < a\}.$$
6. Exponential distribution:
$$\text{p.d.f. } f(x) = e^{-x} I(x \ge 0) \Rightarrow \text{ c.f. } \psi(t) = \frac{1}{1 - it}.$$
7. Gamma distribution:
$$\text{p.d.f. } f(x) = \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x} I(x \ge 0) \Rightarrow \text{ c.f. } \psi(t) = \frac{1}{(1 - it/\lambda)^c}.$$
8. Double exponential distribution:
$$\text{p.d.f. } f(x) = \frac{1}{2} e^{-|x|} \Rightarrow \text{ c.f. } \psi(t) = \frac{1}{1 + t^2}.$$
9. Binomial distribution: i.i.d. r.v's $X, X_1, \cdots, X_n \sim B(1, p)$ (Bernoulli distribution), i.e., $P(X = 1) = p, P(X = 0) = 1 - p =: q$, then $S_n = \sum_{i=1}^n X_n \sim B(n, p)$,
$$\psi_X(t) = pe^{it} + q, \psi_{S_n}(t) = (pe^{it} + q)^n.$$
10. Poisson (λ) distribution:
$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}(k \ge 0) \Rightarrow \text{ c.f. } \psi(t) = e^{\lambda(e^{it} - 1)}.$$

B. Properties of characteristic functions

1. $\psi(0) = 1, |\psi(t)| = |E(e^{itX})| \leq E|e^{itX}| = 1$ for all t.
2. $\psi_{aX+b}(t) = E[e^{it(aX+b)}] = e^{itb}E[e^{itaX}] = e^{itb}\psi(at)$.
3. $\psi_{-X}(t) = \psi_X(-t) = \overline{\psi_X(t)}$ (conjugate).
4. $\psi_X(t)$ is real iff X is symmetric about 0.
5. If X, Y are independent r.v.'s, then $\psi_{X+Y}(t) = E[e^{it(X+Y)}] = E(e^{itX})E(e^{itY}) = \psi_X(t)\psi_Y(t)$.
6. Let $F_1, \cdots, F_n$ are d.f.'s with c.f. $\psi_1, \cdots, \psi_n$. If $\lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1$, then $\sum_{i=1}^n \lambda_i F_i$ is a d.f. with c.f. given by $\sum_{i=1}^n \lambda_i \psi_i$.
7. If $\psi(t)$ is a c.f., so are $|\psi(t)|^2$ and $\mathrm{Re}(\psi(t))$, since i.i.d r.v.'s X, Y with c.f. $\psi(t)$ lead to $\psi_{X-Y}(t) = \psi_X(t)\psi_Y(-t) = |\psi(t)|^2$, and the d.f. $(F_X(x) + F_{-X}(x))/2$ has c.f. $\mathrm{Re}(\psi(t))$. However, $|\psi(t)|$ may not be a c.f.
8. If $|\psi_X(t)| \equiv 1$ for all t, then $\psi_X(t) = e^{ibt}$, *i.e.*, X is degenerate at b.
9. $\psi(t)$ is uniformly continuous in $t \in (-\infty, \infty)$, but could be nowhere differentiable.

Theorem 10.3.1

$$\mathrm{Re}(1 - \psi(t)) \geq \frac{1}{4}\mathrm{Re}(1 - \psi(2t)) \geq \cdots \geq \frac{1}{4^n}\mathrm{Re}(1 - \psi(2^n t)).$$

In particular, we have

$$1 - |\psi(t)|^2 \geq \frac{1}{4}(1 - |\psi(2t)|^2) \geq \cdots \geq \frac{1}{4^n}(1 - |\psi(2^n t)|^2),$$
$$1 - |\psi(t)| \geq \frac{1}{8}(1 - |\psi(2t)|) \geq \cdots \geq \frac{1}{8^n}(1 - |\psi(2^n t)|).$$

Corollary 10.3.1 Suppose that $|\psi(t)| \leq a < 1$ for $|t| \geq b > 0$. Then $|\psi(t)| \leq 1 - ct^2 \leq e^{-ct^2}$ for $|t| < b$ where $c = \frac{1-a^2}{8b^2}$.

Corollary 10.3.2 $\limsup_{|t|\to\infty} |\psi(t)| < 1 \Rightarrow \forall\delta > 0, \exists d \in (0, 1)$ such that $|\psi(t)| \leq d$ for $|t| > \delta$.

Theorem 10.3.2 Let X be a nondegenerate r.v. with c.f. ψ. There exist $\delta > 0$ and $\epsilon > 0$ such that $|\psi(t)| \leq 1 - \epsilon t^2$ for $|t| \leq \delta$.

Theorem 10.3.3 For any $t, h \in \mathbb{R}$, $|\psi(t + h) - \psi(t)|^2 \leq 2(1 - \mathrm{Re}[\psi(h)]) = 2E(1 - \cos(hX))$.

Corollary 10.3.3 If c.f.s $\psi_n(t) \to g(t)$ for all t, and g is continuous at 0, then g is continuous everywhere on $\mathbb{R}$.

Corollary 10.3.4 If there exists some $\delta > 0$ such that c.f.s $|\psi_n(t)| \to 1$ for $|t| < \delta$, then $|\psi_n(t)| \to 1$ for all $t \in \mathbb{R}$ (hence, $X_n \Rightarrow 0$.)

C. Inversion formula

Lemma 10.4.1

$$\lim_{T\to\infty}\int_0^T \frac{\sin at}{t}dt = \frac{\pi}{2}sgn(a).$$

Theorem 10.4.1 (The inversion formula) Let $\psi(t) = \int e^{itx}P(dx)$ where P is a probability measure, then for $a < b$ we have

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\psi(t)dt = P((a,b)) + \frac{1}{2}P(\{a,b\})$$

provided that the limit on the left hand side exists.

Theorem 10.4.2

$$P(\{a\}) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\psi(t)dt.$$

Theorem 10.4.3 (Uniqueness) Characteristic functions uniquely determines distribution functions. That is, there is a one-one correspondence between c.f.s and d.f.s.

Theorem 10.4.4 If $\int_{-\infty}^{\infty}|\psi(t)|dt < \infty$, then the corresponding probability measure P satisfies $P(\{a\}) = 0$ for all a and has bounded continuous density

$$f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\psi(t)dt.$$

Theorem 10.4.5 If $P(X \in b + h\mathbb{Z}) = 1$, then for $x \in b + h\mathbb{Z}$, we have

$$P(X = x) = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\psi(t)dt.$$

D. Levy Continuity Theorem

Lemma 10.5.1 (The tail probability behavior of a r.v. is actually determined by the behavior of its c.f. around the origin.)

$$\forall a > 0, P\left(|X| > \frac{2}{a}\right) \le \frac{1}{a}\int_{-a}^{a}(1-\psi(t))dt.$$

Lemma 10.5.2 Let F_n be a sequence of d.f.s with c.f.s ψ_n. If $\psi_n(t) \to g(t)$, and $g(t)$ that is continuous at 0, then F_n is tight.

Theorem 10.5.1 (Levy continuity theorem) Assume that X_n has d.f. F_n and c.f. ψ_n.

(1) If $X_n \xrightarrow{d} X_\infty (F_n \Rightarrow F_\infty)$, then $\psi_n(t) \to \psi_\infty(t)$ for all t.

(2) If $\psi_n(t) \to \psi(t)$ and $\psi(t)$ is continuous at 0, then there exists a r.v. X with d.f. F such that $X_n \xrightarrow{d} X(F_n \Rightarrow F)$ and ψ is the c.f. of X.

E. Moments of r.v.s and derivatives of their c.f.s

Theorem 10.6.1 If $E|X|^n < \infty$, then $\psi^{(n)}(t)$ exists and is a uniformly continuous function:

$$\psi^{(k)}(t) = i^k E(X^k e^{itX}) = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF(x), \forall k = 0, 1, 2, \cdots, n.$$

In particular, $\psi^{(k)}(0) = i^k E(X^k), \forall k = 0, 1, 2, \cdots, n.$

Theorem 10.6.2 If $\psi^{(n)}(0)$ exists and is finite for some $n \in \mathbb{N}_+$, then $E|X|^n < \infty$ if n is even.

Theorem 10.6.3 (1) If $E|X|^{n+\delta} < \infty$ for some non-negative integer n and some $\delta \in [0, 1]$, then the c.f. has Taylor expansion

$$\psi(t) = \sum_{k=0}^{n} (EX^k) \frac{(it)^k}{k!} + \theta \frac{2E|X|^{n+\delta}|t|^{n+\delta}}{n!}, |\theta| \leq 1.$$

(2) Conversely, suppose that the c.f. of a r.v. X can be written as

$$\psi(t) = \sum_{k=0}^{n} a_k \frac{(it)^k}{k!} + o(t^n) \text{ when } t \to 0,$$

then $E|X|^n < \infty$ if n is even. Furthermore, $a_k = E(X^k)$ whenever $E|X|^k < \infty$.

Theorem 10.6.4 Let $X, X_1, X_2, \cdots$ be i.i.d. r.v.'s with d.f. F. These statements are equivalent:

(1) $\psi'(0) = i\nu$.

(2) $tP(|X| > t) \to 0, E(XI\{|X| \leq t\}) \to \nu$.

(3) The WLLN holds: $(X_1 + X_2 + \cdots + X_n)/n \xrightarrow{P} \nu$.

(Also, we can easily prove WLLN for i.i.d. r.v.'s $\{X_i\}$ with $E(X_i) = \mu$ such that $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \nu$ through the Taylor expansion of c.f.s and Helly's selection theorem, Levy continuity theorem.)

F. Esseen's Smoothing Lemma

Lemma 10.8.1 Let F be a d.f. and G a function such that $G(-\infty) = 0, G(\infty) = 1$, and $\sup_x |G'(x)| \leq \lambda < \infty$. Put $\Delta(x) = F(x) - G(x)$ and let V_T be an inverse triangular d.f. with a p.d.f. $v_T(x) = \frac{1-\cos(Tx)}{\pi T x^2}$, denote $\Delta^T(t) = \Delta * V_T(x) = \int_{-\infty}^{\infty} \Delta(t-x) v_T(x) dx$, then

$$\sup_x |\Delta(x)| \leq 2 \sup_x |\Delta^T(x)| + \frac{24\lambda}{\pi T}.$$

Lemma 10.8.2 (Esseen's Smoothing Lemma) Let F be a d.f. with vanishing expectation and c.f. $\psi_F(t)$. Suppose that $F - G$ vanishes at $\pm\infty$ and that G has a derivative G' such that $\sup_x |G'(x)| \leq \lambda$. Finally, suppose that G has a continuously differentiable Fourier transform ψ_G such that $\psi_G(0) = 1$ and $\psi_G'(0) = 0$. Then for any $T > 0$,

$$\sup_x |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-\infty}^{\infty} \left| \frac{\psi_F(t) - \psi_G(t)}{t} \right| dt + \frac{24\lambda}{\pi T}.$$

XI. Chapter 11. Central Limit Theorems

Theorem 11.1.1 (Levy theorem) Let $X_1, \cdots, X_n$ be i.i.d. r.v's with $E(X_1) = 0, \sigma^2 = E(X_1^2) < \infty$. Let $F_n(x) = P(\sqrt{n}\overline{X}/\sigma \leq x)$ and $\Phi(x)$ be the standard Gaussian distribution function, then $\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \to 0$.

Theorem 11.1.2 (Lindeberg-Feller CLT) For each n, let $X_{n,k}(1 \leq k \leq n)$ be independent r.v.s with $E(X_{n,k}) = 0$ and $\sum_{k=1}^n \sigma_{n,k}^2 := \sum_{k=1}^n E(X_{n,k}^2) = 1$. Denote $F_n(x) = P(\sum_{k=1}^n X_{n,k} \leq x)$. Then the following two statements are equivalent.

(1) The Lindeberg condition holds: $\forall \epsilon > 0, \lim_{n\to\infty} \sum_{k=1}^n E(X_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\}) = 0$.

(2) $\lim_{n\to\infty} \max_{1\leq k\leq n} \sigma_{n,k}^2 = 0$ and $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| = 0$.

Theorem 11.1.3 (Lyapunov CLT) For each n, let $X_{n,k}(1 \leq k \leq n)$ be independent r.v.s with $E(X_{n,k}) = 0$ and $\sum_{k=1}^n \sigma_{n,k}^2 := \sum_{k=1}^n E(X_{n,k}^2) = 1$. Denote $F_n(x) = P(\sum_{k=1}^n X_{n,k} \leq x)$.

$$\exists \delta > 0, \lim_{n\to\infty} \sum_{k=1}^n E|X_{n,k}|^{2+\delta} = 0 \Rightarrow \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| = 0.$$

Corollary 11.1.1 (Lindeberg-Feller CLT) Let $X_1, \cdots, X_n$ be independent non-degenerate r.v.'s such that $E(X_j) = 0, Var(X_j) = \sigma_j^2 < \infty, j = 1, \cdots, n$. Let $S_n = \sum_{k=1}^n X_k, B_n = \sum_{k=1}^n \sigma_k^2$, and $F_n(x) = P(S_n/B_n \leq x)$. The following two statements are equivalent.

(1) The Lindeberg condition holds: $\forall \epsilon > 0, \lim_{n\to\infty} B_n^{-2} \sum_{k=1}^n E(X_k^2 I\{|X_k| \geq \epsilon B_n\}) = 0$.

(2) $\lim_{n\to\infty} \max_{1\leq k\leq n}(\sigma_k^2/B_k^2) = 0$ and $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| = 0$.

Corollary 11.1.2 (Lyapunov CLT)

$$\exists \delta > 0, \lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E|X_k|^{2+\delta} = 0 \Rightarrow \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| = 0.$$

Theorem 11.2.1 (Berry-Esseen bounds for independent r.v.'s) Let $X_1, \cdots, X_n$ be independent r.v.'s such that $E(X_k) = 0$ and $E|X_k|^{2+\delta} < \infty(k = 1, \cdots, n)$ for some $0 < \delta \leq 1$. Denote

$$L_{n,\delta} = B_n^{-(2+\delta)} \sum_{k=1}^n E|X_k|^{2+\delta}.$$

Then for all n, there exists some constant A,

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq AL_{n,\delta}.$$

Corollary 11.2.3 (Berry-Esseen bounds for i.i.d. r.v.'s) Let $X_1, \cdots, X_n$ be i.i.d. r.v.'s. Let $\delta \in (0, 1]$, and $E(X_1) = 0, E(X_1^2) = \sigma^2 > 0, E|X_1|^{2+\delta} < \infty$. Denote $\rho_\delta = E|X_1|^{2+\delta}/\sigma^{2+\delta}$, then for all n, there exists some constant A,

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{A\rho_\delta}{n^{\delta/2}}.$$