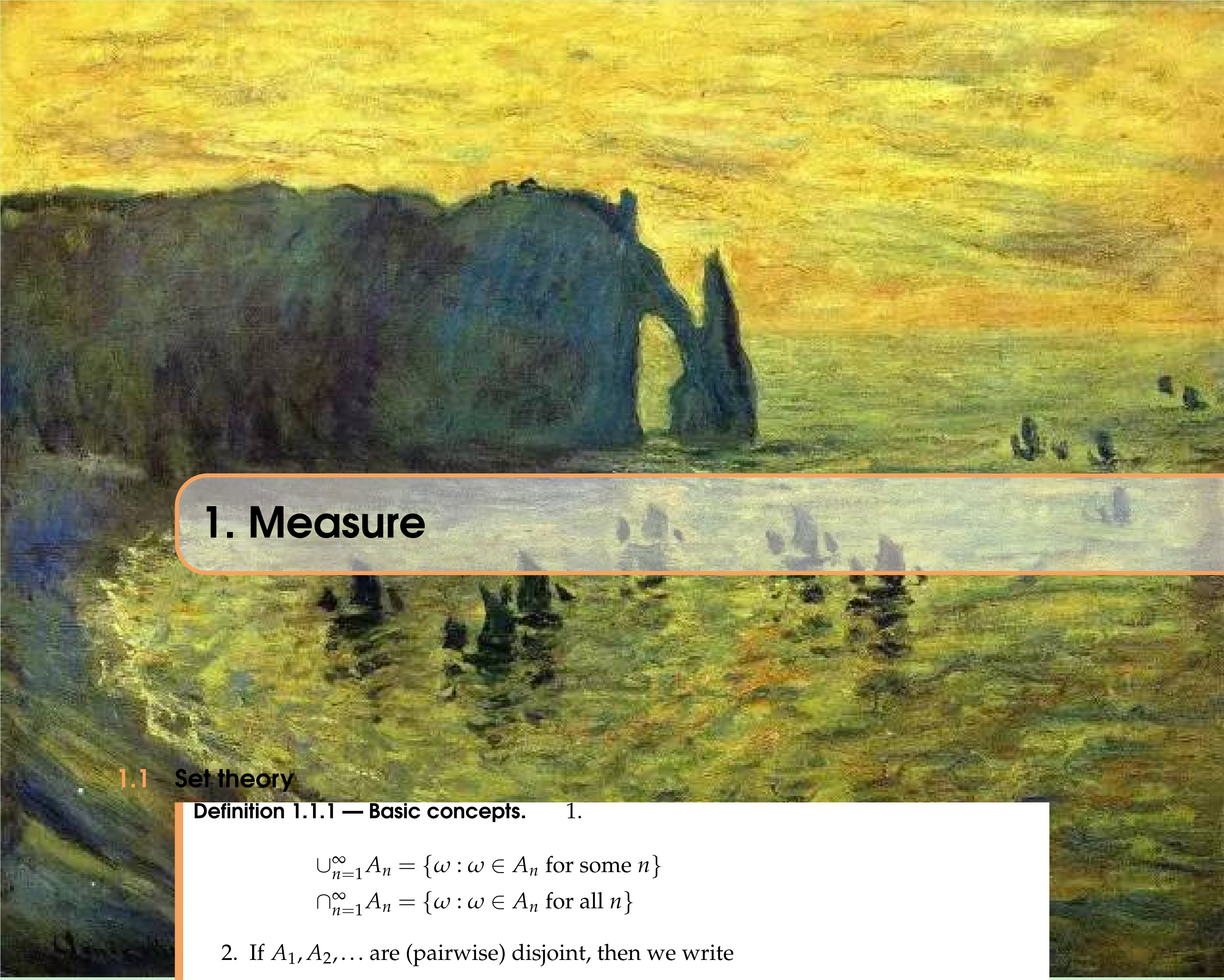
Advanced Probability

Xin Wen

Contents

Part One



1. Measure

1.1 Set theory

Definition 1.1.1 — Basic concepts. 1.

$$\cup_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for some } n\}$$
$$\cap_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for all } n\}$$

2. If $A_1, A_2, \dots$ are (pairwise) disjoint, then we write

$$\cup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n$$

3. (a) Infinitely often (i.o.)

$$\begin{aligned}\limsup_n A_n \equiv \overline{\lim_n} A_n &= \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n \\ &= \{\omega : \forall k \geq 1, \exists n \geq k, \text{ s.t. } \omega \in A_n\} \\ &= \{\omega : \omega \in A_n \text{ for infinitely many values of } n\} \\ &= \{A_n, \text{ i.o. }\}\end{aligned}$$

(b) Ultimately (ult.)

$$\begin{aligned}\liminf_n A_n \equiv \underline{\lim}_n A_n &= \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n \\ &= \{\omega : \exists k \geq 1, \forall n \geq k, \text{s.t. } \omega \in A_n\} \\ &= \{\omega : \omega \in A_n \text{ for all but finitely many values of } n\} \\ &= \{A_n, \text{ ult. }\}\end{aligned}$$

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

$$\liminf_n A_n = \limsup_n A_n = A$$

4.

$$\liminf_n A_n \subset \limsup_n A_n$$

(from definition)

R $\forall == \cap, \exists == \cup$. The more you intersect, the less the value

■ **Example 1.1 — $\liminf_n A_n$ and $\limsup_n A_n$ may not be equal.** Specify $\liminf_n A_n$ and $\limsup_n A_n$ when $A_{2j} = B$, and $A_{2j-1} = C, j = 1, 2, \ldots$. Clearly,

$$\begin{aligned}\lim_n \inf A_n &= \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n = \cup_{k=1}^{\infty} (B \cap C) = B \cap C \\ \limsup_n A_n &= \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n = \cap_{k=1}^{\infty} (B \cup C) = B \cup C\end{aligned}$$

Theorem 1.1.1 — Monotone sequence of sets converges.
1. If $A_1 \subset A_2 \subset A_3 \ldots$, then $A_n \to A = \cup_{k=1}^{\infty} A_k$, written as $A_n \uparrow A$
2. If $A_1 \supset A_2 \supset A_3 \ldots$, then $A_n \to A = \cap_{k=1}^{\infty} A_k$, written as $A_n \downarrow A$

Proof. We shall only prove a). Let $A = \cup_{k=1}^{\infty} A_k$. Clearly,

$$\liminf_n A_n = \cup_{k=1}^{\infty} (\cap_{n=k}^{\infty} A_n) = \cup_{k=1}^{\infty} A_k = A$$

$$\limsup_n A_n = \cap_{k=1}^{\infty} (\cup_{n=k}^{\infty} A_n) = \cap_{k=1}^{\infty} (\cup_{n=1}^{\infty} A_n) = \cap_{k=1}^{\infty} A = A$$

■

Corollary 1.1.2 Let $A_1, A_2, \ldots$ be subsets of Ω.

$$\{A_n, \text{i.o.}\} = \limsup_n A_n = \lim_{k\to\infty} \cup_{n=k}^{\infty} A_n$$

$$\{A_n, \text{ult.}\} = \liminf_n A_n = \lim_{k\to\infty} \cap_{n=k}^{\infty} A_n$$

Proof. We shall only prove the first one. Let $B_k = \cup_{n=k}^{\infty} A_n$. Clearly, $B_1 \supset B_2 \supset B_3 \ldots$ From the last theorem,

$$\lim_{k\to\infty} B_k = \cap_{k=1}^{\infty} B_k = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n = \limsup_n A_n$$

■

Theorem 1.1.3 — Using indicator function to prove some set relationships. $\forall A, B \subset \Omega$, we have

1. $A = B \Longleftrightarrow I_A = I_B \quad (\text{meaning } I_A(\omega) = I_B(\omega) \text{ for all } \omega \in \Omega)$
2. $A \subset B \Longleftrightarrow I_A \le I_B$
3. $A = \emptyset \text{ or } \Omega, \text{resp.} \Longleftrightarrow I_A = 0 \quad \text{or } 1, \text{resp.}$
4. $I_{A\cap B} = \min\{I_A, I_B\} = I_A I_B$
5. $I_{A\cup B} = \max\{I_A, I_B\} = I_A + I_B - I_{A\cap B} = I_A + I_B - I_A I_B$
6. $I_{A^c} = 1 - I_A$
7. $I_{A-B} = I_{A\cap B^c} = I_A(1 - I_B)$
8. $I_{A\Delta B} = |I_A - I_B|$
9. $I_{\liminf_n A_n} = \liminf_n I_{A_n}$
10. $I_{\limsup_n A_n} = \limsup_n I_{A_n}$
11. $I_{A\cup B} \le I_A + I_B$
12. $I_{\cup_1^{\infty} A_n} \le \sum_1^{\infty} I_{A_n}$

Theorem 1.1.4

$$I_{\cup_1^n A_j} = \sum_1^n I_{A_j} - \sum_{1\le j_1<j_2\le n} I_{A_{j_1}\cap A_{j_2}} + \sum_{1\le j_1<j_2<j_3\le n} I_{A_{j_1}\cap A_{j_2}\cap A_{j_3}} + \ldots\ldots\ldots + (-1)^{n-1} I_{A_1\cap A_2\cap\ldots\cap A_n}$$

Proof. Set $s_1 = \sum_1^n I_{A_j}, s_2 = \sum_{1\le j_1<j_2\le n} I_{A_{j_1}\cap A_{j_2}}, \ldots\ldots s_n = I_{A_1\cap A_2\cap\ldots\cap A_n}$. Then we need to show

$$I_{\cup_1^n A_j} = s_1 - s_2 + s_3 - \ldots + (-1)^{n-1} s_n \tag{1.1}$$

In proof of this, if for some $\omega \in \Omega, I_{\cup_1^n A_j}(\omega) = 0$, clearly $s_k(\omega) = 0, 1 \le k \le n$, whence (1.1) holds. On the other hand, if $I_{\cup_1^n A_j}(\omega) = 1$, then $\omega \in A_j$ for at least one $j, 1 \le j \le n$. Suppose that ω belongs toexactly m of the sets $A_1, \ldots, A_n$. Then $s_1(\omega) = m, s_2(\omega) =$

$\binom{m}{2}, \ldots, s_m(\omega) = 1, s_{m+1}(\omega) = \ldots = s_n(\omega) = 0$ whence

$$s_1 - s_2 + s_3 - \ldots + (-1)^{n-1}s_n = \binom{m}{1} - \binom{m}{2} + \ldots + (-1)^{m-1}\binom{m}{m}$$
$$= \binom{m}{0} - (1-1)^m = 1 = I_{\cup_1^n A_j}$$

This completes the proof. ■

Corollary 1.1.5 — inclusion-exclusion formula. By taking expectation, we get the "inclusion-exclusion formula":

$$P(\cup_1^n A_j) = \sum_{j=1}^{n} P(A_j) - \sum_{1 \le j_1 < j_2 \le n} P(A_{j_1} \cap A_{j_2}) + \ldots\ldots\ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Definition 1.1.2
1. Let $\mathbf{R}^d$ be the set of vectors $(x_1, \ldots x_d)$ of real numbers
2. $\mathcal{R}^d$ be the Borel sets, the smallest σ -field containing the open sets.

Definition 1.1.3 — semialgebra. A collection $\mathcal{S}$ of sets is said to be a **semialgebra** if
1. it is closed under intersection, i.e., $S, T \in \mathcal{S}$ implies $S \cap T \in \mathcal{S}$, and
2. if $S \in \mathcal{S}$ then S^c is a finite disjoint union of sets in $\mathcal{S}$.

■ **Example 1.2** $\{(a,b]\}$ is a semialgbra, the disadvantage of semialgebra is it doesnot include $(\infty, a]$, (b, ∞).

Definition 1.1.4 — Algebra. A collection $\mathcal{A}$ of subsets of Ω is called an **algebra** (or field) if $A, B \in \mathcal{A}$ implies A^c and $A \cup B$ are in $\mathcal{A}$. Since $A \cap B = (A^c \cup B^c)^c$, it follows that $A \cap B \in \mathcal{A}$.

R algebra does not closed under limit, e.g. $\lim \mu((1, 2 + \frac{1}{n}]) = 1$, but 2+1/$n$ − 1>1

Proposition 1.1.6
1. If $\mathcal{A}$ is an algebra (or a σ -algebra), then $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$. However, the same may not hold for semi-algebras. For the second part, consider $\mathcal{S} = \{\emptyset, A, A^c\}$. It is clearly a semi-algebra, but does not contain Ω.
2. $\mathcal{A}$ is an algebra $\Longleftrightarrow$ (1).$\Omega \in \mathcal{A}$ (2). $A, B \in \mathcal{A}$ implies $A - B \in \mathcal{A}$
3. a σ -algebra is an algebra.
4. Let $\{\mathcal{A}_\gamma : \gamma \in \Gamma\}$ be a collection (possibly uncountable) of σ -algebras. Then $\mathcal{A} = \cap_{\gamma \in \Gamma} \mathcal{A}_\gamma$ is also a σ -algebra.
5. 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,
   (a) $\mathcal{A} \subset \sigma(\mathcal{A})$

(b) For any σ-algebra $\mathcal{B}$ with $\mathcal{A} \subset \mathcal{B}, \sigma(\mathcal{A}) \subset \mathcal{B}$, and $\sigma(\mathcal{A})$ is unique.

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

$$\sigma(\mathcal{S}) = \sigma(\overline{\mathcal{S}})$$

Lemma 1.1 — **Algebra generated by semialgebra.** If $\mathcal{S}$ is a semialgebra then

$$\overline{\mathcal{S}} = \{\text{finite disjoint unions of sets in } \mathcal{S}\}$$

is an algebra, called the algebra generated by $\mathcal{S}$ ■

Proof. To prove $\overline{S}$ is closed under intersection and complement. ■

Definition 1.1.5 — σ -algebra. A σ **-algebra** on X is a family of subsets of X that contains X and is closed under complementation, under the formation of countable unions, and under the formation of countable intersections. Condition (a) is equal to $\mathscr{A}$ be nonempty.

If $\mathscr{A}$ is a σ -algebra on the set X, it is sometimes convenient to call a **subset** of X $\mathscr{A}$ **-measurable** if it belongs to $\mathscr{A}$.

Proposition 1.1.7 Let X be a set. Then the intersection of an arbitrary nonempty collection of σ-algebras on X is a σ-algebra on X.

Analysis 1.1 Let $\mathscr{C}$ be a nonempty collection of σ -algebras on X, and let $\mathscr{A}$ be the intersection of the σ -algebras that belong to $\mathscr{C}$. ■

Proposition 1.1.8 The union of a family of σ -algebras can fail to be a σ -algebra.

Corollary 1.1.9 Let X be a set, and let $\mathscr{F}$ be a family of subsets of X. Then there is a smallest σ -algebra on X that includes $\mathscr{F}$. The smallest σ -algebra is called the σ -algebra generated by $\mathscr{F}$ and is often denoted by $\sigma(\mathscr{F})$.

Theorem 1.1.10 — **Measure on semialgebra extended to algebra.** Let $\mathcal{S}$ be a semialgebra and let μ defined on $\mathcal{S}$ have $\mu(\emptyset) = 0$. Suppose

1. if $S \in \mathcal{S}$ is a finite disjoint union of sets $S_i \in \mathcal{S}$ then $\mu(S) = \sum_i \mu(S_i)$, and
2. if $S_i, S \in \mathcal{S}$ with $S = +_{i \geq 1} S_i$ then $\mu(S) \leq \sum_{i \geq 1} \mu(S_i)$.

Then μ has a unique extension $\bar{\mu}$ that is a measure on $\overline{\mathcal{S}}$ the algebra generated by $\mathcal{S}$. If $\bar{\mu}$ is sigma-finite then there is a unique extension ν that is a measure on $\sigma(\mathcal{S})$

Lemma 1.2 Suppose if $S \in \mathcal{S}$ is a finite disjoint union of sets $S_i \in \mathcal{S}$ then $\mu(S) = \sum_i \mu(S_i)$,

1. If $A, B_i \in \overline{\mathcal{S}}$ with $A = +_{i=1}^n B_i$ then $\bar{\mu}(A) = \sum_i \bar{\mu}(B_i)$
2. If $A, B_i \in \overline{\mathcal{S}}$ with $A \subset \cup_{i=1}^n B_i$ then $\bar{\mu}(A) \leq \sum_i \bar{\mu}(B_i)$

Definition 1.1.6 — Borel σ -algebra. The Borel σ -algebra on $\mathbb{R}^d$ is the σ -algebra on $\mathbb{R}^d$ generated by the collection of **open** subsets of $\mathbb{R}^d$; it is denoted by $\mathscr{B}\left(\mathbb{R}^d\right)$.

Proposition 1.1.11 The σ-algebra $\mathscr{B}(\mathbb{R})$ of Borel subsets of $\mathbb{R}$ is generated by each of the following collections of sets:

1. the collection of all closed subsets of $\mathbb{R}$
2. the collection of all subintervals of $\mathbb{R}$ of the form $(-\infty, b]$
3. the collection of all subintervals of $\mathbb{R}$ of the form $(a, b]$.

R Every "reasonable" subset of R is a Borel set. However, $\sigma(R) \neq \mathcal{P}(R)$.

Definition 1.1.7 — $\mathscr{F}_\sigma$,$\mathscr{G}_\delta$. Let $\mathscr{G}$ be the family of all open subsets of $\mathbb{R}^d$, and let $\mathscr{F}$ be the family of all closed subsets of $\mathbb{R}^d$. Let $\mathscr{G}_\delta$ be the collection of all intersections of sequences of sets in $\mathscr{G}$, and let $\mathscr{F}_\sigma$ be the collection of all unions of sequences of sets in $\mathscr{F}$. Sets in $\mathscr{G}_\delta$ are often called G_δ 's, and sets in $\mathscr{F}_\sigma$ are often called F_σ 's.

Proposition 1.1.12 Each closed subset of $\mathbb{R}^d$ is a G_δ(open intersect), and each open subset of $\mathbb{R}^d$ is an F_σ(close union)

Proof. Suppose that F is a closed subset of $\mathbb{R}^d$. We need to construct a sequence $\{U_n\}$ of open subsets of $\mathbb{R}^d$ such that $F = \cap_n U_n$. For this define U_n by

$$U_n = \left\{x \in \mathbb{R}^d : \|x - y\| < 1/n \text{ for some } y \text{ in } F\right\}$$

(Note that U_n is empty if F is empty.) It is clear that each U_n is open and that $F \subseteq \cap_n U_n$. The reverse inclusion follows from the fact that F is closed (note that each point in $\cap_n U_n$ is the limit of a sequence of points in F). Hence each closed subset of $\mathbb{R}^d$ is a G_δ

If U is open, then U^c is closed and so is a G_δ. Thus there is a sequence $\{U_n\}$ of open sets such that $U^c = \cap_n U_n$. The sets U_n^c are then closed, and $U = \cup_n U_n^c$; hence U is an F_σ ■

Proposition 1.1.13 — Condition of algebra to be sigma-alg. Let X be a set, and let $\mathscr{A}$ be an algebra on X. Then $\mathscr{A}$ is *a* σ -algebra if either

1. $\mathscr{A}$ is closed under the formation of unions of increasing sequences of sets, or
2. $\mathscr{A}$ is closed under the formation of intersections of decreasing sequences of sets.

Proof. First suppose that condition (a) holds. Since $\mathscr{A}$ is an algebra, we can check that it is a σ -algebra by verifying that it is closed under the formation of countable unions. Suppose that $\{A_i\}$ is a sequence of sets that belong to $\mathscr{A}$. For each n let $B_n = \cup_{i=1}^{n} A_i$. The sequence $\{B_n\}$ is increasing, and, since $\mathscr{A}$ is an algebra, each B_n belongs to $\mathscr{A}$; thus assumption (a) implies that $\cup_n B_n$ belongs to $\mathscr{A}$. However, $\cup_i A_i$ is equal to $\cup_n B_n$ and so belongs to $\mathscr{A}$. Thus $\mathscr{A}$ is closed under the formation of countable unions and so is a σ -algebra.

Now suppose that condition (b) holds. It is enough to check that condition holds. If $\{A_i\}$ is an increasing sequence of sets that belong to $\mathscr{A}$, then $\{A_i^c\}$ is a decreasing sequence of sets that belong to $\mathscr{A}$, and so condition (b) implies that $\cap_i A_i^c$ belongs to $\mathscr{A}$. since $\cup_i A_i = \left(\cap_i A_i^c\right)^c$, it follows that $\cup_i A_i$ belongs to $\mathscr{A}$. Thus condition (a) follows from condition (b), and the proof is complete. ∎

Definition 1.1.8 Let $\mathcal{A}$ be a nonempty class of subsets of Ω.

1. $\mathcal{A}$ is said to be a **monotone class** (m -class) on (M-class—under mono-limit) Ω if $\lim A_n \in \mathcal{A}$ for every monotone sequence $A_n \in \mathcal{A}, n \geq 1$. That is,
   (a) If $A_i \in \mathcal{A}$ and $A_i \uparrow$, then $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i \in \mathcal{A}$
   (b) If $A_i \in \mathcal{A}$ and $A_i \downarrow$, then $\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i \in \mathcal{A}$
   (i.e., m -class is closed under monotone operations). (The existence of $\lim A_n$ was established in Theorem 1.1.1)
2. $\mathcal{A}$ is a π -class on Ω if
   $$A \cap B \in \mathcal{A}, \quad \text{whenever } A, B \in \mathcal{A}$$
   (i.e. a π -class is closed under finite intersection.)
3. $\mathcal{A}$ is a λ -class on Ω if
   (a) $\Omega \in \mathcal{A}$
   (b) $A - B \in \mathcal{A}$ for $A, B \in \mathcal{A}$ and B is a proper subset of A, (i.e. $B \subset A$.)
   (c) $\lim A_n \in \mathcal{A}$ for every increasing sequence $A_n \in \mathcal{A}, n \geq 1$
   (i.e. a λ -class contains Ω and is closed under proper difference and countable increasing union. Note $\lambda \approx$ "increasing" limit). ($\lambda = \text{lam}\,da = \text{lam} + da \approx \text{lim} + \text{diff}$. Also $\lambda =$ "Large")

Proposition 1.1.14 1. If $\mathcal{A}$ is a λ -class, it is an m -class.

Proof. Part (iii) in the definition of λ -class implies that $\mathcal{A}$ is closed under the limit of increasing sequence of sets in $\mathcal{A}$. By part (i) and (ii) in the definition of λ -class, $\mathcal{A}$ is closed under the limit of decreasing sequence of sets in $\mathcal{A}$. Thus, it is an m -class. ∎

2. Suppose $\mathcal{A}$ is an algebra on Ω. Then $\mathcal{A}$ is an m -class $\Longleftrightarrow$ $\mathcal{A}$ is a σ -algebra.

Proof. " $\Longleftarrow$ ". Suppose that $\mathcal{A}$ is a σ -algebra, and we shall show that $\mathcal{A}$ is an m -class. Let $A_n \in \mathcal{A}$ and $A_n \uparrow$ (or $\downarrow$). From Theorem 1.1.1, $\lim_n A_n = \cup_{n=1}^{\infty} A_n$ (or $\cap_{n=1}^{\infty} A_n$), which belongs to $\mathcal{A}$(as $\mathcal{A}$ is a σ -algebra). Thus $\mathcal{A}$ is an m -class.

" $\Longrightarrow$ ". Suppose that $\mathcal{A}$ is an m -class, and we shall show that it is a σ -algebra. since $\mathcal{A}$ is an algebra, it suffices to show that $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$ for $A_n \in \mathcal{A}$. To show this, let $B_n = \cup_{k=1}^{n} A_k$. Then B_n is monotone (increasing) and $B_n \in \mathcal{A}$ as $\mathcal{A}$ is an algebra. So $\lim_n B_n = \lim_{n\to\infty} \cup_1^n A_n = \cup_{n=1}^{\infty} A_n \in \mathcal{A}$ as $\mathcal{A}$ is an m -class. ■

3. A σ -algebra must be an m -class, but an m -class may not be an algebra, let alone a σ -algebra. $\mathcal{A} = \{(-\infty, a), (-\infty, a], (-\infty, \infty), \emptyset : a \in R\}$. $\mathcal{A}$ is an m -class
4. $\mathcal{A}$ is a σ -algebra iff it is both a λ -class and π -class.

Proof. " $\Longrightarrow$ ". Suppose that $\mathcal{A}$ is a σ -algebra. Clearly, it is a π -class. Now we show that it is also a λ -class.

(a) It is clear that $\Omega \in \mathcal{A}$.
(b) $A - B = A \cap B^c \in \mathcal{A}$
(c) For every increasing sequence $A_n \in \mathcal{A}, n \geq 1$, by Theorem 1.1.1, $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$ which belongs to $\mathcal{A}$ (as $\mathcal{A}$ is a σ -algebra).

So it is a λ -class.

" $\Longrightarrow$ ". Now suppose that $\mathcal{A}$ is both a λ -class and π -class. First, $\forall A \in \mathcal{A}$ $A^c = \Omega - A \in \mathcal{A}$. (as $\mathcal{A}$ is a λ -class) So $\mathcal{A}$ is closed under complements. Now let us show that $\mathcal{A}$ is closed under countable union. For any sequence $A_n \in \mathcal{A}, n \geq 1$, using (7.2), we have

$$B_n = \cup_{j=1}^{n} A_j = \left(\cap_{j=1}^{n} A_j^c\right)^c \in \mathcal{A}, \quad (\text{ as } \mathcal{A} \text{ is a } \pi \text{ -class (intersect) and } \lambda \text{complement})$$

It follows that $\mathcal{A}$ is closed under finite union. Since B_n is increasing, by the property of λ -class, we have

$$\cup_{n=1}^{\infty} A_n = \lim_{n\to\infty} B_n \in \mathcal{A}$$

Hence $\mathcal{A}$ is closed under countable union. This completes the proof. ■

5. The power set $\mathcal{P}(\mathcal{A})$ is an m -class (or λ -class, or π -class).
6. 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).
7. 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 A. In other words,
   (a) $\mathcal{A} \subset m(\mathcal{A})$, (or λ -class, or π -class).

(b) For any m-class (or λ-class, or π-class) $\mathcal{B}$ with $\mathcal{A} \subset \mathcal{B}$, we have $m(\mathcal{A}) \subset \mathcal{B}$, (or $\lambda(\mathcal{A}) \subset \mathcal{B}$, or $\pi(\mathcal{A}) \subset \mathcal{B}$)

and $m(\mathcal{A})$ (or $\lambda(\mathcal{A})$, or $\pi(\mathcal{A})$) is unique.

1.1.1 The Monotone Class Theorem (MCT)

Theorem 1.1.15 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 $\sigma(\mathcal{A}) \subset \mathcal{B}$

Proof. 1. First note that $\mathcal{A} \subset m(\mathcal{A}) \subset \sigma(\mathcal{A})$ (since $\sigma(\mathcal{A})$ is an m -class containing $\mathcal{A}$ and $m(\mathcal{A})$ is the smallest m -class containing $\mathcal{A}$).

It remains to show that $m(\mathcal{A}) \supset \sigma(\mathcal{A})$. It suffices to show that $m(\mathcal{A})$ is a σ -algebra. In fact, by Theorem 1.1.14, we only need to show $m(\mathcal{A})$ is an algebra.

Proof. Define

$$\mathcal{C}_1 = \{A : A \in m(\mathcal{A}), A \cap B \in m(\mathcal{A}) \text{ for all } B \in \mathcal{A}\}$$
$$\mathcal{C}_2 = \{B : B \in m(\mathcal{A}), A \cap B \in m(\mathcal{A}) \text{ for all } A \in m(\mathcal{A})\}$$
$$\mathcal{C}_3 = \{A : A \in m(\mathcal{A}), A^c \in m(\mathcal{A})\}$$

We now show that $\mathcal{C}_j, j = 1,2,3$ are m -classes.

First we look at $\mathcal{C}_1$. Suppose $A_j \in \mathcal{C}_1, j \geq 1$, and $A_1 \subset A_2 \subset A_3 \ldots$ (increasing). By the definition of $\mathcal{C}_1$, we have $A_j \cap B \in m(\mathcal{A}), j \geq 1$, for all $B \in \mathcal{A}(\subset m(\mathcal{A}))$. Also we have $(A_1 \cap B) \subset (A_2 \cap B) \subset (A_3 \cap B) \subset \ldots$. That is, $A_j \cap B$'s form an increasing sequence in $m(\mathcal{A})$. In view of $m(\mathcal{A})$ being a m -class, it follows that $A_j \cap B \to \cup_{j=1}^{\infty} (A_j \cap B) \in m(\mathcal{A})$. Using the DeMorgan's law, we get

$$B \cap \left(\cup_{j=1}^{\infty} A_j\right) = \cup_{j=1}^{\infty} (B \cap A_j) \in m(\mathcal{A}), \quad \text{for all } B \in \mathcal{A}$$

Therefore, $\lim_{j\to\infty} A_j = \cup_{j=1}^{\infty} A_j \in \mathcal{C}_1$. Similarly, for a decreasing sequence of sets $D_1 \supset D_2 \supset D_3 \supset \ldots$ in $\mathcal{C}_1$, we can show that $B \cap D_j$'s form a decreasing sequence in $m(\mathcal{A})$. Thus,

$$B \cap (\cap_{i=1}^{\infty} D_j) = \cap_{i=1}^{\infty} (B \cap D_j) \in m(\mathcal{A})$$

Therefore, $\lim_j D_j = \cap_{j=1}^{\infty} D_j \in \mathcal{C}_1$. Thus, we have shown that $\mathcal{C}_1$ is a m -class. The proof for $\mathcal{C}_2$ is similar and hence omitted. Now we look at $\mathcal{C}_3$. By the identities $\left(\cup_{j=1}^{\infty} A_j\right)^c = \cap_{j=1}^{\infty} A_j^c$, $\left(\cap_{j=1}^{\infty} A_j\right)^c = \cup_{j=1}^{\infty} A_j^c$, we can easily show that $\mathcal{C}_3$ is a m -class. ■

Secondly, we shall show that $m(\mathcal{A}) = \mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}_3$. Clearly, $\mathcal{C}_i \subset m(\mathcal{A})$ by definition, $i = 1,2,3$. So we only need to show below that $\mathcal{C}_i \supset m(\mathcal{A}), i = 1,2,3$. (Just need to prove $\mathcal{A} \in C_i$ + $m(\mathcal{A})$ is smallest one)

Proof. (a) Since $\mathcal{A}$ is an algebra and $\mathcal{A} \subset m(\mathcal{A})$, it is clear that $\mathcal{A} \subset \mathcal{C}_1$. Hence, $m(\mathcal{A}) \subset \mathcal{C}_1$ as $m(\mathcal{A})$ is the smallest m -class. Noting $\mathcal{C}_1 \subset m(\mathcal{A})$, we get

$$m(\mathcal{A}) = \mathcal{C}_1$$

(b) For any $B \in \mathcal{A}$ and $A \in m(\mathcal{A}) = \mathcal{C}_1$, we have $A \cap B \in m(\mathcal{A})$ by the definition of $\mathcal{C}_1$, which in turn implies that $\mathcal{A} \subset \mathcal{C}_2$ (as $m(\mathcal{A})$ is the smallest m -class). Noting $\mathcal{C}_2 \subset m(\mathcal{A})$, we get

$$m(\mathcal{A}) = \mathcal{C}_2$$

which implies that $m(\mathcal{A})$ is closed under intersection.

(c) It is easy to see that $\mathcal{A} \subset \mathcal{C}_3$. (To see this, if $A \in \mathcal{A} \subset m(\mathcal{A})$, then $A^c \in \mathcal{A} \subset m(\mathcal{A})$ as $\mathcal{A}$ ia an algebra.) Thus, $\mathcal{C}_3$ is an m -class containing $\mathcal{A}$. Thus, $m(\mathcal{A}) \subset \mathcal{C}_3$. But it is clear that $\mathcal{C}_3 \subset m(\mathcal{A})$ by definition. Then

$$m(\mathcal{A}) = \mathcal{C}_3$$

which means that $m(\mathcal{A})$ is closed under complement.

■

Finally, we shall show $(*)$: $m(\mathcal{A})$ is an algebra, which follows from (ii) and (iii) above.

2. $\mathcal{A} \subset \mathcal{B}$ implies that $m(\mathcal{A}) \subset m(\mathcal{B}) = \mathcal{B}$ (as $\mathcal{B}$ is an m-class). The proof then follows easily from(1): $m(\mathcal{A}) = \sigma(\mathcal{A})$

■

Theorem 1.1.16 Let $\mathcal{A}$ be a π -class.

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

Proof. Clearly, $\lambda(\mathcal{A}) \subset \sigma(\mathcal{A})$ as $\sigma(\mathcal{A})$ is a λ -class and $\lambda(\mathcal{A})$ is the smallest λ -class. Now let us show that $\lambda(\mathcal{A}) \supset \sigma(\mathcal{A})$. From Theorem 1.1.14(4), it suffices to show that $\lambda(\mathcal{A})$ is a π -class (which implies that $\lambda(\mathcal{A})$ is a σ -algebra containing $\mathcal{A}$, thus $\lambda(\mathcal{A}) \supset \sigma(\mathcal{A})$). To this end, define

$$\mathcal{C}_1 = \{A : A \subset \Omega, A \cap B \in \lambda(\mathcal{A}) \text{ for all } B \in \mathcal{A}\}$$

Clearly, we can show that

1. $\mathcal{A} \subset \mathcal{C}_1$
2. $\mathcal{C}_1$ is a λ -class.

Proof. Proof of (a) and (b). Let $A \in \mathcal{A}$. Since $\mathcal{A}$ is a π -class, for any $B \in \mathcal{A}$, we have $A \cap B \subset \mathcal{A} \in \lambda(\mathcal{A})$. Thus, $A \in \mathcal{C}_1$. This proves (a). The proof of (b) is left as an exercise. ■

Hence, $\mathcal{C}_1$ is a λ -class containing $\mathcal{A}$. So $\mathcal{C}_1 \supset \lambda(\mathcal{A})$. Thus $A \in \lambda(\mathcal{A})$ implies that $A \in \lambda(\mathcal{A}) \in \mathcal{C}_1$. So for any $B \in \mathcal{A}$, by the definition of $\mathcal{C}_1$, we have $A \cap B \in \lambda(\mathcal{A})$. Define

$$\mathcal{C}_2 = \{B : B \subset \Omega, A \cap B \in \lambda(\mathcal{A}) \text{ for all } A \in \lambda(\mathcal{A})\}$$

Then, $\mathcal{A} \subset \mathcal{C}_2$. We can also show that $\mathcal{C}_2$ is a λ -class (Why?). Therefore, $\mathcal{C}_2$ is a λ -class containing $\mathcal{A}$. Consequently, $\mathcal{C}_2 \supset \lambda(\mathcal{A})$. So if $A, B \in \lambda(\mathcal{A})$, then $B \in \lambda(\mathcal{A}) \subset \mathcal{C}_2$, and by the definition of $\mathcal{C}_2$, we get $A \cap B \in \lambda(\mathcal{A})$. Thus, $\lambda(\mathcal{A})$ is a π -class. ■

Theorem 1.1.17 — 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 $\sigma(\mathcal{A}) \subset \mathcal{B}$

Proof. Applying the last two theorems, we get

1. $\mathcal{B}$ is a λ -class containing $\mathcal{A}$, so $\mathcal{B} \supset \lambda(\mathcal{A}) = \sigma(\mathcal{A})$.
2. $\mathcal{B}$ is a m -class containing $\mathcal{A}$, so $\mathcal{B} \supset m(\mathcal{A}) = \sigma(\mathcal{A})$

■

Theorem 1.1.18 — The Monotone Class Theorem is used in the following way. If $\mathcal{A}$ has some property $\mathcal{P}$, in order to show that $\sigma(\mathcal{A})$ has the same property $\mathcal{P}$, we can proceed as follows:

1. Define $\mathcal{B} = \{B : B \text{ has property } \mathcal{P}\}$, so that $\mathcal{A} \subset \mathcal{B}$.
2. Show that
   (a) $\mathcal{A}$ is a π -class, and $\mathcal{B}$ is a λ -class, or
   (b) $\mathcal{A}$ is an algebra, and $\mathcal{B}$ is an m -class,
3. From the Monotone Class Theorem, we get $\sigma(\mathcal{A}) \subset \mathcal{B}$.

Therefore, $\sigma(\mathcal{A})$ has property $\mathcal{P}$ as well.

Definition 1.1.9 — Product space. For any measurable spaces $(\Omega_i, \mathcal{A}_i)\,, i = 1, \ldots$ define for $n \geq 2$

1. n -dim rectangles of the product space of $\prod_{i=1}^n \Omega_i$:

$$\prod_{i=1}^n A_i := A_1 \times \ldots \times A_n = \{(\omega_1, \ldots, \omega_n) : \omega_i \in A_i \subset \Omega_i, 1 \leq i \leq n\}$$

2. n -dim 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 -dim product measurable space:

$$\prod_{i=1}^{n}(\Omega_i,\mathcal{A}_i)=\left(\prod_{i=1}^{n}\Omega_i,\prod_{i=1}^{n}\mathcal{A}_i\right)$$

Theorem 1.1.19 — Product measure theorem. Let $(\Omega_i,\mathcal{F}_i,\nu_i), i=1,\ldots,k$ be measure spaces with σ -finite measures, where $k\geq 2$ is an integer. Then there exists a unique σ -finite measure on the product σ -field $\sigma(\mathcal{F}_1\times\cdots\times\mathcal{F}_k)$ called the product measure and denoted by $\nu_1\times\cdots\times\nu_k$, such that

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

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

Theorem 1.1.20 — Extended to cases involving initely many measure spaces. If $(\mathcal{R}^k,\mathcal{B}^k,P_i), i=1,2,\ldots,$ are probability spaces, then there is a product probability measure P on $\prod_{i=1}^{\infty}(\mathcal{R}^k,\mathcal{B}^k)$ such that for any positive integer l and $B_i\in\mathcal{B}^k, i=1,\ldots,l$

$$P\left(B_1\times\cdots\times B_l\times\mathcal{R}^k\times\mathcal{R}^k\times\cdots\right)=P_1(B_1)\cdots P_l(B_l)$$

Lemma 1.3 The intersection (but not the union) of any two measurable rectangles of a given product space is a measurable rectangle in that space. In other words, the class of measurable rectangles of $\prod_{i=1}^{n}\Omega_i$ is a π -class. ▪

Theorem 1.1.21 Let $(\Omega_i,\mathcal{A}_i), \leq i\leq n$ be measurable spaces, and

$$\mathcal{A}=\left\{\text{finite union of disjoint rectangles}\prod_{i=1}^{n}A_i \text{ with } A_i\in\mathcal{A}_i, 1\leq i\leq n.\right\}$$

Show that $\mathcal{A}$ is the algebra generated by the class of all measurable rectangles of the product space $\prod_{i=1}^{n}\Omega_i$

Proof. Let $\mathcal{G}$ denote the class of all measurable rectangles of $\prod_{i=1}^{n}\Omega_i$, which clearly forms a π -class. Then we need to show that

$$\mathcal{A}=\mathcal{A}(\mathcal{G})$$

We shall show this in several steps.

1. First we show that $\mathcal{A}$ is also a π -class. If $A_i=\cup_{j=1}^{n_i}E_{ij}\in\mathcal{A}, i=1,2$, with $E_{ij}\in\mathcal{G}$, then $E_{1j}\cap E_{2k}\in\mathcal{G}$ from the above lemma, hence

$$A_1\cap A_2=\cup_{j=1}^{n_1}\cup_{k=1}^{n_k}\left(E_{1j}\cap E_{2k}\right)\in\mathcal{A}$$

Moreover, if $E = E_1 \times \ldots \times E_n \in \mathcal{G}$, from

$$\begin{aligned}
\Omega_1 \times \ldots \times \Omega_n &= (E_1 \times \Omega_2 \ldots \times \Omega_n) + (E_1^c \times \Omega_2 \ldots \times \Omega_n) \\
&= (E_1 \times E_2 \times \Omega_3 \times \ldots \times \Omega_n) + (E_1 \times E_2^c \times \Omega_3 \ldots \times \Omega_n) \\
&= + (E_1^c \times \Omega_2 \ldots \times \Omega_n) \\
&= \ldots\ldots\ldots \\
&= (E_1 \times \ldots \times E_n) + (E_1 \times \ldots \times E_{n-1} \times E_n^c) \\
&\quad + (E_1 \times \ldots \times E_{n-2} \times E_{n-1} \times \Omega_n) + \cdots\cdots \\
&\quad \cdots + (E_1^c \times \Omega_2 \ldots \times \Omega_n)
\end{aligned}$$

Therefore,

$$\begin{aligned}
E^c =& \Omega_1 \times \ldots \times \Omega_n - (E_1 \times E_2 \times \ldots \times E_{n-1} \times E_n) \\
&= (E_1 \times E_2 \times \ldots \times E_{n-1} \times E_n^c) + (E_1 \times \ldots \times E_{n-2} \times E_{n-1} \times \Omega_n) \\
&\quad + \ldots\ldots\ldots + (E_1^c \times \Omega_2 \ldots \times \Omega_n) \\
&= \cup_{i=1}^n D_i(say) \in \mathcal{A}
\end{aligned}$$

2. Next we show that $\mathcal{A}$ is an algebra. From (1), $\mathcal{A}$ is closed under intersection. It suffices to show that it is also closed under complements. Let $A = \cup_{j=1}^r E_j \in \mathcal{A}$ with $E_j \in \mathcal{G}$, then from (1), we have

$$A^c = \cap_{j=1}^r E_j^c = \bigcap_{j=1}^r \cup_{i=1}^n D_i^{(j)} = \cup_{i=1}^n \left(\cap_{j=1}^r D_i^{(j)} \right) \in \mathcal{A}$$

So $\mathcal{A}$ is indeed closed under complements. Hence, $\mathcal{A}$ is an algebra.

3. Clearly, $\mathcal{G} \subset \mathcal{A}$, so $\mathcal{A}(\mathcal{G}) \subset \mathcal{A}$ from (2) . On the other hand, every finite union of disjoint rectangles with measurable sides $\in \mathcal{A}(\mathcal{G})$, and so $\mathcal{A} \subset \mathcal{A}(\mathcal{G})$. Thus we showed that $\mathcal{A} = \mathcal{A}(\mathcal{G})$, as required.

■

Corollary 1.1.22 The σ -algebra $\prod_{i=1}^n \mathcal{A}_i$, generated by the rectangles with measurable sides, is also the σ-algebra generated by the algebra $\mathcal{A}$ in the last theorem.

In the special case where $(\Omega_i, \mathcal{A}_i) = (\Omega, \mathcal{A})$ for all i, then we can write

$$\Omega^n = \prod_{i=1}^n \Omega_i, \quad \mathcal{A}^n = \prod_{i=1}^n \mathcal{A}_i, \quad \text{etc.}$$

1.2 Measure theory

Definition 1.2.1 — Probability Space. A **probability space** is a measure space with total measure one. The standard notation is $(\Omega, \mathcal{F}, \mathbb{P})$ where:

1. Ω is a set (sometimes called a sample space). Elements of Ω are denoted ω and are sometimes called outcomes

2. $\mathcal{F}$ is a σ -algebra (or σ -field) of subsets of Ω. Sets in $\mathcal{F}$ are called events.
3. $\mathbb{P}$ is a function from $\mathcal{F}$ to [0,1] with $\mathbb{P}(\Omega) = 1$ and such that if $E_1, E_2, \ldots \in \mathcal{F}$ are disjoint

$$\mathbb{P}\left[\bigcup_{j=1}^{\infty} E_j\right] = \sum_{j=1}^{\infty} \mathbb{P}\left[E_j\right]$$

Definition 1.2.2 — Measure. A **measure** is a nonnegative countably additive set function; that is, a function $\mu : \mathcal{F} \to \mathbf{R}$ with

1. $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$, and
2. (countable additive)if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu\left(\cup_i A_i\right) = \sum_i \mu\left(A_i\right)$$

If $\mu(\Omega) = 1$, we call μ a probability measure, which denoted by $\mathcal{P}$.

R If Ω is countable in the sense that there is a one-to-one correspondence between Ω and the set of all integers, then one can usually consider the trivial σ -field that contains all subsets of Ω and a measure that assigns a value to every subset of Ω. When Ω is uncountable (e.g. ,$\Omega = \mathcal{R}$ or $[0,1]$) it is not possible to define a reasonable measure for every subset of Ω; for example, it is not possible to find a measure on all subsets of $\mathcal{R}$ and still satisfy property $m([a,b]) = b - a$. This is why it is necessary to introduce σ -fields that are smaller than the power set.

Definition 1.2.3 — finitely additive measure. A **finitely additive measure** on the algebra $\mathscr{A}$ is a function $\mu : \mathscr{A} \to [0, +\infty]$ that satisfies $\mu(\varnothing) = 0$ and is finitely additive.

Theorem 1.2.1 — Properties of Measure. Let μ be a measure on $(\Omega, \mathcal{F})$

1. **monotonicity**. If $A \subset B$ then $\mu(A) \leq \mu(B)$
2. **subadditivity**. If $A \subset \cup_{m=1}^{\infty} A_m$ then $\mu(A) \leq \sum_{m=1}^{\infty} \mu\left(A_m\right)$
3. **continuity from below**. If $A_i \uparrow A$ (i.e., $A_1 \subset A_2 \subset \ldots$ and $\cup_i A_i = A$) then $\mu\left(A_i\right) \uparrow \mu(A)$
4. **continuity from above**. If $A_i \downarrow A$ (i.e., $A_1 \supset A_2 \supset \ldots$ and $\cap_i A_i = A$), with $\mu\left(A_1\right) < \infty$, then $\mu\left(A_i\right) \downarrow \mu(A)$
5. **Continuity at** A. If $A_n \to A$ implies $\mu\left(A_n\right) \to \mu(A), \mu$ is said to be continuous at A

Analysis 1.2 Construct the disjoint sets according to condtions and use the the countable additive and subadditive. Be familiar with the language. ■

Proof. Prove the third and fourth properties

1. First suppose that $\{A_k\}$ is an increasing sequence of sets that belong to $\mathscr{A}$, and define a sequence $\{B_i\}$ of sets by letting $B_1 = A_1$ and letting $B_i = A_i - A_{i-1}$ if $i > 1$. The sets just constructed are disjoint, belong to $\mathscr{A}$, and satisfy $A_k = \cup_{i=1}^k B_i$ for each k. It follows that $\cup_k A_k = \cup_i B_i$ and hence that

$$\mu\left(\cup_k A_k\right) = \sum_i \mu\left(B_i\right) = \lim_k \sum_{i=1}^k \mu\left(B_i\right) = \lim_k \mu\left(\cup_{i=1}^k B_i\right) = \lim_k \mu\left(A_k\right)$$

This completes the proof of (a).

2. However, measure μ may NOT be continuous from above if without the condition of $\mu\left(A_m\right) < \infty$ for some finite m . For example, define a measure μ on the Borel measurable space $(R, \mathcal{B})$ by

$$\begin{aligned} \mu(A) &= 0, \quad \text{if } A = \varnothing \\ &= \infty, \quad \text{if } A \neq \varnothing \end{aligned}$$

It is easy to check that μ is σ -additive, thus a measure. Take $A_n = (0, 1/n)$, then $A_n \to \varnothing$ forms a decreasing sequence of sets in $\mathcal{B}$, but $\mu\left(A_n\right) = \infty + \mu(\varnothing) = 0$

3. Now suppose that $\{A_k\}$ is a decreasing sequence of sets that belong to $\mathscr{A}$ and that $\mu\left(A_n\right) < +\infty$ holds for some n. We can assume that $n = 1$. For each k let $C_k = A_1 - A_k$. Then $\{C_k\}$ is an increasing sequence of sets that belong to $\mathscr{A}$ and satisfy

$$\cup_k C_k = A_1 - \left(\cap_k A_k\right) = A_1 \cap \left(\cap_k A_k\right)^C = A_1 \cap \left(\cup_k A_k^C\right)$$

It follows from part (a) that $\mu\left(\cup_k C_k\right) = \lim_k \mu\left(C_k\right)$ and hence that

$$\mu\left(A_1 - \left(\cap_k A_k\right)\right) = \mu\left(\cup_k C_k\right) = \lim_k \mu\left(C_k\right) = \lim_k \mu\left(A_1 - A_k\right)$$

In view of Proposition 1.2.2 and the assumption that $\mu\left(A_1\right) < +\infty$ (promising the cancel law), this implies that

$$\mu\left(\cap_k A_k\right) = \lim_k \mu\left(A_k\right)$$

■

R Using $+$ to denote disjoint union. And using $B_n = A'_n - \cup_{m=1}^{n-1} A'_m$ and $B_n = A_n - A_{n-1}$ to construct disjoint subset.

■ **Example 1.3 — Lebesgue measure.** Consider the set $\mathbb{R}$ of all real numbers and the σ -algebra $\mathscr{B}(\mathbb{R})$ of Borel subsets of $\mathbb{R}$. A measure on $\mathscr{B}(\mathbb{R})$ that assigns to each subinterval of $\mathbb{R}$ **its length**; this measure is known as **Lebesgue measure** and will be denoted by λ in this book.

Proposition 1.2.2 — Monotone of measure. Let$(X,\mathscr{A},\mu)$ be a measure space, and let A and B be subsets of X that belong to $\mathscr{A}$ and satisfy $A \subseteq B$. Then $\mu(A) \leq \mu(B)$. If in addition A satisfies $\mu(A) < +\infty$, then $\mu(B-A) = \mu(B) - \mu(A)$

Proof. Proved by constructing disjoint set and using countable additive. ■

Definition 1.2.4 — Finite/σ-finite.
1. Let μ be a measure on a measurable space $(X,\mathscr{A})$. Then μ is a **finite measure** if $\mu(X) < +\infty$ and is a σ **-finite** measure if X is the union of a sequence $A_1, A_2, \ldots$ of sets that belong to $\mathscr{A}$ and satisfy $\mu(A_i) < +\infty$ for each i.
2. More generally, a set in $\mathscr{A}$ is σ **-finite** under μ if it is the union of a sequence of sets that belong to $\mathscr{A}$ and have finite measure under μ.
3. If the measure space $(X,\mathscr{A},\mu)$ is σ -finite, then X is the union of a sequence $\{B_i\}$ of disjoint sets that belong to $\mathscr{A}$ and have finite measure under μ; such a sequence $\{B_i\}$ can be formed by choosing a sequence $\{A_i\}$ as in the definition of σ -finiteness, and then letting $B_1 = A_1$ and $B_i = A_i - \left(\cup_{j=1}^{i-1} A_j\right)$ if $i > 1$

Theorem 1.2.3 — Countable additivity of μ implies the countable subadditivity of μ. Let $(X,\mathscr{A},\mu)$ be a measure space. If $\{A_k\}$ is an arbitrary sequence of sets that belong to $\mathscr{A}$, then

$$\mu\left(\cup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu\left(A_k\right)$$

Proof. Define a sequence $\{B_k\}$ of subsets of X by letting $B_1 = A_1$ and letting

$$B_k = A_k - \left(\cup_{i=1}^{k-1} A_i\right) \quad \text{if } k > 1$$

Then each B_k belongs to $\mathscr{A}$ and is a subset of the corresponding A_k, and so satisfies $\mu(B_k) \leq \mu(A_k)$. Since in addition the sets B_k are disjoint and satisfy $\cup_k B_k = \cup_k A_k$, it follows that

$$\mu\left(\cup_k A_k\right) = \mu\left(\cup_k B_k\right) = \sum_k \mu\left(B_k\right) \leq \sum_k \mu\left(A_k\right)$$

■

Proposition 1.2.4 — A finitely additive measure under some situations is in fact countably additive. Let $(X,\mathscr{A})$ be a measurable space, and let μ be a finitely additive measure on $(X,\varnothing)$. Then μ is a measure if either

1. $\lim_k \mu(A_k) = \mu(\cup_k A_k)$ holds for each increasing sequence $\{A_k\}$ of sets that belong to $\mathscr{A}$, or
2. $\lim_k \mu(A_k) = 0$ holds for each decreasing sequence $\{A_k\}$ of sets that belong to $\mathscr{A}$ and satisfy $\cap_k A_k = \varnothing$

Proof. We need to verify the countable additivity of μ. Let $\{B_j\}$ be a sequence of disjoint sets that belong to $\mathscr{A}$; we will prove that $\mu(\cup_j B_j) = \sum_j \mu(B_j)$. First assume that condition (a) holds, and for each k, let $A_k = \cup_{j=1}^k B_j$. Then the finite additivity of μ implies that $\mu(A_k) = \sum_{j=1}^k \mu(B_j)$, while condition (a) implies that $\mu(\cup_{k=1}^\infty A_k) = \lim_k \mu(A_k)$; since $\cup_{j=1}^\infty B_j = \cup_{k=1}^\infty A_k$, it follows that

$$\mu\left(\cup_{j=1}^\infty B_j\right) = \mu(\cup_{k=1}^\infty A_k) = \lim_k \mu(A_k) = \sum_{j=1}^\infty \mu(B_j)$$

Now assume that condition (b) holds, and for each k let $A_k = \cup_{j=k}^\infty B_j$. Then the finite additivity of μ implies that

$$\mu\left(\cup_{j=1}^\infty B_j\right) = \sum_{j=1}^k \mu(B_j) + \mu(A_{k+1})$$

while condition (b) implies that $\lim_k \mu(A_{k+1}) = 0$; hence $\mu\left(\cup_{j=1}^\infty B_j\right) = \sum_{j=1}^\infty \mu(B_j)$ ∎

Definition 1.2.5 — Borel measure.
1. A measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ is often called a **Borel measure** on $\mathbb{R}^d$.
2. More generally, if X is a Borel subset of $\mathbb{R}^d$ and if $\mathscr{A}$ is the σ -algebra consisting of those Borel subsets of $\mathbb{R}^d$ that are included in X, then a measure on $(X, \mathscr{A})$ is called a Borel measure on X.

Definition 1.2.6 — Continue or discrete measure. Suppose that $(X, \mathscr{A})$ is a measurable space such that for each x in X the set $\{x\}$ belongs to $\mathscr{A}$. A finite or σ -finite measure μ on $(X, \mathscr{A})$ is **continuous** if $\mu(\{x\}) = 0$ holds for each x in X and is **discrete** if there is a countable subset D of X such that $\mu(D^c) = 0$.

Definition 1.2.7 — Measure Space/Measurable Space.
1. If X is a set, if $\mathscr{A}$ is a σ -algebra on X, and if μ is a measure on $\mathscr{A}$, then the triplet $(X, \mathscr{A}, \mu)$ is often called a **measure space**.
2. If X is a set and if $\mathscr{A}$ is a σ algebra on X, then the pair $(X, \mathscr{A})$ is often called a measurable space.

1.2.1 Properties of measure

Case I: semialgebras

Let $\mathcal{A}$ be a semialgebra. For simplicity, we assume that

$$\{\emptyset, \Omega\} \subset \mathcal{A}$$

Theorem 1.2.5 Let μ be a nonnegative additive set function on a semialgebra $\mathcal{A}$. Let $A, B \in \mathcal{A}$ and $\{A_n, B_n, n \geq 1\} \in \mathcal{A}$

1. (Monotonicity):

$$A \subset B \Longrightarrow \mu(A) \leq \mu(B)$$

2. ($\sigma-$ Subadditivity) :
   (a) $(\sum_1^\infty A_n \subset A, \quad \Longrightarrow \quad \sum_1^\infty \mu(A_n) \leq \mu(A)$
   (b) Further assume that μ is σ -additive (hence a measure). Then, (disjoint)

$$B \subset \sum_{n=1}^{\infty} B_n, \quad \Longrightarrow \quad \mu(B) \leq \sum_{n=1}^{\infty} \mu(B_n)$$

Proof. 1. Note $B = A + (B \cap A^c) = A + B \cap (\sum_{i=1}^n A_i) = A + \sum_1^n (A_i \cap B)$

$$\mu(B) = \mu(A) + \sum_1^n \mu(A_i \cap B) \geq \mu(A)$$

2. The proof will be given at the end of the section "Extensions of set functions" later.

■

1.2.2 Case II: algebras

All the properties for semialgebras also hold for algebras. In addition, we have

Theorem 1.2.6 — σ -subadditivity. Let μ be a measure on an algebra $\mathcal{A}$. Then,

$$A \subset \cup_1^\infty A_n, \text{ where } A \in \mathcal{A}, \{A_n, n \geq 1\} \in \mathcal{A}, \quad \Longrightarrow \quad \mu(A) \leq \sum_1^\infty \mu(A_n)$$

(Note the difference from the last theorem. Here we have $\cup_1^\infty$, not $\sum_1^\infty$)

Proof. Let $B_n = A \cap A_n \in \mathcal{A}$, and $C_n = B_n - \cup_{i=1}^{n-1} B_i \in \mathcal{A}$ (as it only involves a finite number of operations). So

$$\begin{aligned}
A &= A \cap (\cup_1^\infty A_n) = \cup_1^\infty (A \cap A_n) = \cup_1^\infty B_n \\
&= B_1 + (B_2 - B_1) + (B_3 - [B_1 \cup B_2]) + \ldots\ldots \\
&= \sum_{n=1}^{\infty} C_n
\end{aligned}$$

By the σ -additivity and monotonicity of the measure μ on $\mathcal{A}$, we get

$$\mu(A) = \mu\left(\sum_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} \mu(C_n) \leq \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

as $C_n \subset B_n \subset A_n$

1.2.3 Case III: σ -algebras.

Theorem 1.2.7 Let μ be a measure on a σ -algebra $\mathcal{A}$, and $\{A_n\} \in \mathcal{A}$.

1. (Monotonicity) $A_1 \subset A_2, \Longrightarrow \mu(A_1) \leq \mu(A_2)$
2. (Boole's inequality or Countable Sub-Additivity)
$$\mu\left(\cup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$
(this property does not require disjoint sets.)
3. (Continuity from below)
$$A_n \nearrow A, \Longrightarrow \mu(A_n) \to \mu(A)$$
4. (Continuity from above)
$$A_n \searrow A \text{ and } \mu(A_m) < \infty \text{ for some } m \geq 1 \text{ except for } A_m = \varnothing \Longrightarrow \mu(A_n) \to \mu(A)$$
5. (Continuity at A) If μ is a finite measure, and $A_n \to A$, then $\mu(A_n) \to \mu(A)$. [The claim (5) may not be true if μ is not a finite measure; see the example given in Remark 2.1.1.

Proposition 1.2.8 — Properties of probability measure. If $\mu(\Omega) = 1$, then μ is a probability measure, usually written as P (thus finite). The probability space is $(\Omega, \mathcal{A}, P)$. Then by definition,

1. For any $A \in \mathcal{A}$, we have $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $P\left(\sum_1^{\infty} A_n\right) = \sum_1^{\infty} P(A_n)$
4. $P\left(\sum_1^{n} A_i\right) = \sum_1^{n} P(A_i)$
5. $P(B - A) = P(B) - P(A)$ if $A \subset B$. In particular, $P(A^c) = 1 - P(A)$
6. $P(A) \leq P(B)$ if $A \subset B$
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
8. $P\left(\cup_{k=1}^{n} A_k\right) = \sum_k P(A_k) - \sum_{i<j} P\left(A_i \cap A_j\right) + \ldots\ldots\ldots$
9. If $A_n \nearrow A$ or $A_n \searrow A$, then $P(A_n) \to P(A)$
10. If $A_n \to A$, then $P(A_n) \to P(A)$

■ **Example 1.4**

$$\begin{aligned} \mu(A) &= 0 \quad A = \varnothing \\ &= \infty \quad A \in \mathcal{A}, \quad A \neq \varnothing \end{aligned}$$

This is neither a finite nor a σ -finite measure.

1.2.4 Extension of set functions (or measures) from semialgebras to algebras

Definition 1.2.8 — Extension of measure. 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

$$\mu(A) = \nu(A), \quad \text{for all } A \in \mathcal{A}$$

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 1.2.9 — set function on semialg extend to algebra.
1. Let μ be a non-negative additive set function (or measure) on a semialgebra $\mathcal{S}$ (containing $\emptyset$), then μ has a unique extension $\bar{\mu}$ to $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$, such that $\bar{\mu}$ is additive.
2. Moreover, if μ is σ -additive on $\mathcal{S}$ (which implies that μ is a measure on $\mathcal{S}$), then so is $\bar{\mu}$ on $\overline{\mathcal{S}}$

Proof. 1. (1). For every $A \in \overline{\mathcal{S}}$, we can write $A = \sum_1^m A_i$ with $A_i \in \mathcal{S}$. Define

$$\bar{\mu}(A) = \sum_1^m \mu(A_i)$$

Then $\bar{\mu}$ is well-defined on $\mathcal{S}$, since if A has a distinct partition $A = \sum_{j=1}^n B_j, B_j \in \mathcal{S}$, then

$$\bar{\mu}(A) = \sum_{i=1}^m \mu(A_i) = \sum_{i=1}^m \mu(A_i \cap A) = \sum_{i=1}^m \mu\left(\sum_{j=1}^n A_i \cap B_j\right)$$

$$= \sum_{i=1}^m \sum_{j=1}^n \mu(A_i \cap B_j) = \sum_{j=1}^n \mu\left(\sum_{i=1}^m A_i \cap B_j\right)$$
$$= \sum_{j=1}^n \mu(A \cap B_j) = \sum_{j=1}^n \mu(B_j)$$

That is, $\bar{\mu}(A)$ is the same irrespective of the different partitions of A. Clearly, $\bar{\mu}(A)$ is nonnegative and $\bar{\mu}(\emptyset) = \mu(\emptyset) = 0$. It's easy to see that $\bar{\mu}$ is additive. To prove uniqueness, let $\hat{\mu}$ be another extension of μ to $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$ such that $\hat{\mu}$ is additive. For $A = \sum_{i=1}^l A_i \in \mathcal{A}(\mathcal{S})$ with $A_i \in \mathcal{S}$, we have

$$\hat{\mu}(A) = \sum_1^l \hat{\mu}(A_i) = \sum_1^l \mu(A_i) = \bar{\mu}(A)$$

(2). We will show that $\bar{\mu}$ is σ -additive below. Suppose that $B_n \in \overline{\mathcal{S}}$ and disjoint and $A := \sum_{n=1}^\infty B_n \in \overline{\mathcal{S}}$, we wish to show

$$\bar{\mu}(A) = \sum_{n=1}^\infty \bar{\mu}(B_n)$$

Since $A, B_n \in \overline{\mathcal{S}}$, we have

$$A = \sum_{i=1}^l A_i \quad \text{with } A_i \in \mathcal{S}; \quad \text{and} \quad B_n = \sum_{j=1}^{J_n} B_{nj} \quad \text{with } B_{nj} \in \mathcal{S}$$

Note that $A_i \cap B_{nj} \in \mathcal{S}$ (because of semialgebra) and they are disjoint. So we have

$$
\begin{aligned}
\bar{\mu}(A) &= \sum_{i=1}^{l} \mu(A_i) = \sum_{i=1}^{l} \mu(A_i \cap A) \quad \text{ as } A_i \subset A \\
&= \sum_{i=1}^{l} \mu\left(\sum_{n=1}^{\infty} \sum_{j=1}^{J_n} A_i \cap B_{nj} \right) \quad \text{ as } A = \sum_{n=1}^{\infty} B_n = \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} B_{nj} \\
&= \sum_{i=1}^{l} \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} \mu(A_i \cap B_{nj}) \quad \text{ as } \mu \text{ is } \sigma\text{-additive on } \mathcal{S}
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\sum_{n=1}^{\infty} \bar{\mu}(B_n) = \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} \mu(B_{nj}) &= \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} \mu(B_{nj} \cap A) \quad \text{ as } B_{nj} \in A \\
&= \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} \mu\left(\sum_{i=1}^{l} B_{nj} \cap A_i \right) \\
&= \sum_{n=1}^{\infty} \sum_{j=1}^{J_n} \sum_{i=1}^{l} \mu(B_{nj} \cap A_i)
\end{aligned}
$$

as μ is σ -additive on $\mathcal{S}$.

■

Theorem 1.2.10 (σ -Subadditivity) Let μ be a nonnegative additive set function on a semialgebra $\mathcal{S}$. Let $A, A_n \in \mathcal{S}$. Then

1. $\sum_1^\infty A_n \subset A, \implies \mu(A) \geq \sum_1^\infty \mu(A_n)$
2. Further assume that μ is σ -additive (hence a measure). Then,

$$
A \subset \sum_{n=1}^{\infty} A_n, \implies \mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)
$$

Proof. Let μ be the unique extension $\bar{\mu}$ on $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$ (from the last theorem).

1. Since $\sum_1^n A_i \subset A$ and $A, \sum_1^n A_i \in \overline{\mathcal{S}}$, by the monotonicity of μ, we get

$$
\mu(A) = \bar{\mu}(A) \geq \bar{\mu}\left(\sum_1^n A_i \right) = \sum_1^n \bar{\mu}(A_i) = \sum_1^n \mu(A_i)
$$

   Letting $n \to \infty$, we get the desired result.
2. Since $\bar{\mu}$ is a measure on the algebra $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$, applying Theorem 1.2.6 , we have $\bar{\mu}(A) \leq \sum_1^\infty \bar{\mu}(A_i)$, implying $\mu(A) \leq \sum_1^\infty \mu(A_i)$

■

1.2.5 Outer measure

Definition 1.2.9 — Outer measure induced by a measure. Let μ be a measure on a semialgebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$. For any $A \subset \Omega$, define

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

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

R

1. $a^* := \inf_n a_n$ is the largest lower bound of the series $\{a_n, n \geq 1\}$, i.e.
   (a) $a^* := \inf_n a_n$ is a lower bound, that is, $a_n \geq a$ for all $n \geq 1$
   (b) $\forall \epsilon > 0, a^* + \epsilon$ can not be a lower bound, that is, there exists an a_m s.t.
   $$a_m \leq a^* + \epsilon$$
2. $\mu^*(A)$ is defined for all $A \in \Omega$ since $A \subset \Omega$. So the domain of $\mu^*(A)$ is the power set $\mathcal{P}(\Omega)$. The range of μ^* is $[0, \infty]$
3. $\mu^*(A)$ may not be a measure itself.
4. One can think of A as any set on R, and $\cup_1^{\infty} A_n$ as a countable covering of A.

Theorem 1.2.11 — Properties of outer measure. Let μ be a measure on a semialgebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$, and μ^* be the outer measure induced by μ

1. $\mu^*(A) = \mu(A)$ for $A \in \mathcal{S}$. In particular, $\mu^*(\emptyset) = \mu(\emptyset) = 0$
2. (Monotonicity)
   $$\mu^*(A) \leq \mu^*(B) \quad \text{for } A \subset B \subset \Omega$$
3. ($\sigma-$ subadditivity)
   $$\mu^*\left(\cup_1^{\infty} A_n\right) \leq \sum_1^{\infty} \mu^*(A_n), \quad \text{for } \{A_n\} \subset \Omega$$

Proof. 1. Suppose $A \in \mathcal{S}$. Since $A \subset A$, by definition of outer measure, $\mu^*(A) \leq \mu(A)$. On the other hand, if $A \subset \cup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{S}$, then $\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$ by the σ -subadditivity of measure μ on $\mathcal{S}$. Taking inf on both sides, we get $\mu(A) \leq \mu^*(A)$. So (1) is true.

2. Suppose $A \subset B$. If $B \subset \cup_{n=1}^{\infty} B_n$, then $A \subset \cup_{n=1}^{\infty} B_n$. (i.e., if $\{B_n\}$ is a countable covering of B in $\mathcal{S}$, it is also a countable covering of A in $\mathcal{S}$.) So by definition,
   $$\mu^*(A) \leq \sum_{n=1}^{\infty} \mu(B_n)$$
   Taking inf on both sides, we get $\mu^*(A) \leq \mu^*(B)$.

3. If $\sum_1^\infty \mu^*(A_n) = \infty$, (3) is true. Now assume $\sum_1^\infty \mu^*(A_n) < \infty$. Let ϵ be an arbitrary positive number. For each $A_n \subset \Omega$, by the definition of $\mu^*(A_n)$, there exists $\{A_{nk}\} \in \mathcal{S}, k \geq 1$, such that $A_n \subset \cup_{k=1}^\infty A_{nk}$ (i.e., a countable covering of A_n in $\mathcal{S}$) and

$$\mu^*(A_n) \leq \sum_{k=1}^\infty \mu(A_{nk}) \leq \mu^*(A_n) + \frac{\epsilon}{2^n}$$

It is clear that $\cup_1^\infty A_n \subset \cup_{n=1}^\infty \cup_{k=1}^\infty A_{nk}$ with $A_{nk} \in \mathcal{S}$. (i.e., the latter is a countable covering of the former in $\mathcal{S}$.) So, in view of the last inequality, we have

$$\mu^*(\cup_1^\infty A_n) \leq \sum_{n=1}^\infty \sum_{k=1}^\infty \mu(A_{nk}) \leq \sum_{n=1}^\infty \mu^*(A_n) + \epsilon$$

Since ϵ can be arbitrarily small, the proof is complete.

■

Definition 1.2.10 — A set is measurable. 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 1.2.12 A set $A \subset \Omega$ is measurable w.r.t. an outer measure μ^* iff for any $D \subset \Omega$, one has

$$\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$$

Proof. It follows from Theorem 1.2.11 that μ^* is σ -subadditve, i.e.,

$$\mu^*(D) \leq \mu^*(A \cap D) + \mu^*(A^c \cap D)$$

Therefore, $\mu^*(D) = \mu^*(A \cap D) + \mu^*(A^c \cap D) \iff \mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$ ■

Theorem 1.2.13 Let $\mathcal{A}^*$ be the class of all μ^* -measurable sets.

1. The class $\mathcal{A}^*$ is a σ -algebra.
2. If $A = \sum_1^\infty A_n$ with $\{A_n\} \in \mathcal{A}^*$, then for any $B \subset \Omega$

$$\mu^*(A \cap B) = \sum_1^\infty \mu^*(A_n \cap B)$$

3. $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ is a measure space. Furthermore, $\mu^*|_{\mathcal{A}^*}$ is an extension of μ from $\mathcal{S}$ to $\mathcal{A}^*$ (i.e., Although μ^* is defined on $\mathcal{P}(\Omega)$, but the restriction of μ^* to $\mathcal{A}^*$ is a measure.) (We' ll still use μ^* to denote $\mu^*|_{\mathcal{A}}{}^*$ if no confusion occurs.)

Proof. 1. We first show that $\mathcal{A}^*$ is closed under complement. If $A \in \mathcal{A}^*$, then for any $D \subset \Omega$, one has

$$\mu^*(D) = \mu^*(A \cap D) + \mu^*(A^c \cap D) = \mu^*(A^c \cap D) + \mu^*\left((A^c)^c \cap D\right)$$

which implies that $A^c \in \mathcal{A}^*$. So $\mathcal{A}^*$ is closed under complement. It remains to show that $\mathcal{A}^*$ is closed under countable union, or countable disjoint union as given in the next lemma.

Lemma 1.4 $\mathcal{F}$ is an algebra. $\mathcal{F}$ is a σ-algebra $\Longleftrightarrow \sum_1^\infty A_n \in \mathcal{F}$ whenever $\{A_n\} \in \mathcal{F}$ and disjoint. (i.e. closed under countable disjoint union.) ■

To use the lemma, take $\{A_n\} \in \mathcal{A}^*$ and they are disjoint. We need to show that $\sum_1^\infty A_n \in \mathcal{A}^*$. For any $D \in \Omega$

$$\begin{aligned}
\mu^*(D) =& \mu^*(A_1 \cap D) + \mu^*(A_1^c \cap D) \\
=& \mu^*(A_1 \cap D) + \mu^*(A_1^c \cap A_2 \cap D) + \mu^*(A_1^c \cap A_2^c \cap D) \\
=& \mu^*(A_1 \cap D) + \mu^*(A_2 \cap A_1^c \cap D) + \mu^*(A_3 \cap A_2^c \cap A_1^c \cap D) \\
=& \dots\dots\dots\dots \\
=& \mu^*(A_1 \cap D) + \mu^*(A_2 \cap A_1^c \cap D) + \mu^*(A_3 \cap A_2^c \cap A_1^c \cap D) \\
& + \dots\dots\dots\dots \\
& + \mu^*(A_n \cap A_{n-1}^c \cap \dots\dots \cap A_2^c \cap A_1^c \cap D) \\
& + \mu^*(A_n^c \cap A_{n-1}^c \cap \dots\dots \cap A_2^c \cap A_1^c \cap D)
\end{aligned}$$

Since $\{A_n\}$ are disjoint, we have

$$\begin{aligned}
& A_k \cap A_{k-1}^c \cap \dots\dots \cap A_1^c = A_k \\
& A_n^c \cap A_{n-1}^c \cap \dots\dots \cap A_1^c = \left(\textstyle\sum_{i=1}^n A_i\right)^c \supset \left(\textstyle\sum_{i=1}^\infty A_i\right)^c
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu^*(D) =& \mu^*(A_1 \cap D) + \mu^*(A_2 \cap D) + \ldots + \mu^*(A_n \cap D) \\
& + \mu^*\left(\left(\sum_{i=1}^n A_i\right)^c \cap D\right) \\
\geq & \sum_{i=1}^n \mu^*(A_i \cap D) + \mu^*\left(\left(\sum_{i=1}^\infty A_i\right)^c \cap D\right)
\end{aligned}$$

Letting $n \to \infty$, we get

$$\begin{aligned}
\mu^*(D) &\geq \sum_{i=1}^\infty \mu^*(A_i \cap D) + \mu^*\left(\left(\sum_{i=1}^\infty A_i\right)^c \cap D\right) \\
&\geq \mu^*\left(\left(\sum_{i=1}^\infty A_i\right) \cap D\right) + \mu^*\left(\left(\sum_{i=1}^\infty A_i\right)^c \cap D\right)
\end{aligned} \tag{1.2}$$

This shows that $\sum_{i=1}^\infty A_i \in \mathcal{A}^*$. Thus, $\mathcal{A}^*$ is a σ -algebra.

2. Continuing from above, since $\sum_{i=1}^\infty A_i \in \mathcal{A}^*$, and so all the inequalities in (1.2) can

be replaced by equalities. That is,

$$\begin{aligned}\mu^*(D) &= \sum_{i=1}^{\infty}\mu^*(A_i\cap D)+\mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)^c\cap D\right)\\ &= \mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)\cap D\right)+\mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)^c\cap D\right)\end{aligned} \tag{1.3}$$

In particular, we let $D=(\sum_{i=1}^{\infty}A_i)\cap B$, we get

$$\begin{aligned}\mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)\cap B\right) &= \sum_{i=1}^{\infty}\mu^*(A_i\cap B)+\mu^*(\emptyset)\\ &= \mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)\cap B\right)+\mu^*(\emptyset)\end{aligned}$$

Since $\mu^*(\emptyset)=0$, from Theorem 1.2.11, we get

$$\sum_{i=1}^{\infty}\mu^*(A_i\cap B)=\mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)\cap B\right)=\mu^*(A\cap B)$$

which proves (2). (Remark: Note that, from 3.25 , we can not directly get

$$\sum_{i=1}^{\infty}\mu^*(A_i\cap D)=\mu^*\left(\left(\sum_{i=1}^{\infty}A_i\right)\cap D\right)=\mu^*(A\cap D)$$

Since $A_1+B=A_2+B$ does not necessarily imply $A_1=A_2$ unless $|B|<\infty$.)

3. Clearly, μ^* is nonnegative on $\mathcal{A}^*$ and σ -additive from (2) by taking $B=A$ or Ω. So μ^* is a measure on $\mathcal{A}^*$. The fact that $\mu^*|_{\mathcal{A}^*}$ is an extension of μ from $\mathcal{S}$ to $\mathcal{A}^*$ follows from Theorem 1.2.11(1)

■

1.2.6 Extension of measures from semialgebras to σ -algebras

Theorem 1.2.14 We have $\mathcal{S}\subset\mathcal{A}^*$, hence $\sigma(\mathcal{S})\subset\mathcal{A}^*$

Proof. For any $A\in\mathcal{S}$, we need to show $A\in\mathcal{A}^*$, namely

$$\mu^*(D)\geq\mu^*(A\cap D)+\mu^*(A^c\cap D)\qquad\text{for any } D\subset\Omega$$

By the definition of $\mu^*(D)$, for any arbitrary $\epsilon>0$, there exists $\{A_n\}\in\mathcal{S},\cup_1^{\infty}A_n\supset D$ (i.e. there exists a countable covering of D in $\mathcal{S}$), such that

$$\mu^*(D)+\epsilon\geq\sum_{1}^{\infty}\mu(A_n)\geq\mu^*(D) \tag{1.4}$$

As $A\in\mathcal{S}$, we have $A^c=\sum_{k=1}^{m}B_k$ with $B_k\in\mathcal{S}$. So

$$A_n=A\cap A_n+A^c\cap A_n=A\cap A_n+\sum_{k=1}^{m}(B_k\cap A_n)$$

So $\mu(A_n) = \mu(A \cap A_n) + \sum_{k=1}^{m} \mu(B_k \cap A_n)$. From this and 1.4, we have

$$\begin{aligned}\mu^*(D) + \epsilon &\geq \sum_1^\infty \mu(A_n) \\ &= \sum_1^\infty \mu(A \cap A_n) + \sum_{n=1}^\infty \sum_{k=1}^m \mu(B_k \cap A_n) \\ &\geq \mu^*(A \cap D) + \mu^*(A^c \cap D)\end{aligned}$$

where the last inequality holds by the definition of μ^* and the fact that $\cup_1^\infty (A \cap A_n)$ and $\cup_{n=1}^\infty \cup_{k=1}^m (B_k \cap A_n)$ are countable coverings of $A \cap D$ and $A^c \cap D$ in $\mathcal{S}$, respectively, since

$$\begin{aligned}&\cup_1^\infty (A \cap A_n) = A \cap (\cup_1^\infty A_n) \supset A \cap D \\ &\cup_{n=1}^\infty \cup_{k=1}^m (B_k \cap A_n) = \cup_{n=1}^\infty (A^c \cap A_n) = A^c \cap (\cup_{n=1}^\infty A_n) \supset A^c \cap D\end{aligned}$$

Since ϵ can be arbitrarily small, we have shown that $A \in \mathcal{A}^*$. Therefore, we showed that $\mathcal{S} \subset \mathcal{A}^*$. To prove the second part, we take σ on both sides of $\mathcal{S} \subset \mathcal{A}^*$ to get $\sigma(\mathcal{S}) \subset \sigma(\mathcal{A}^*) = \mathcal{A}^*$ as $\mathcal{A}^*$ is a σ-algebra. ■

R We have this relationship: $\mathcal{S} \subset \overline{\mathcal{S}} \subset \sigma(\mathcal{S}) \subset \mathcal{A}^* \subset \mathcal{P}(\Omega)$

R We have seen from Theorem 1.2.13 that the measure μ can be extended from $\mathcal{S}$ to $\mathcal{A}^*$. We 've also shown that $\sigma(\mathcal{S}) \subset \mathcal{A}^*$. In the next theorem, however, we shall extend the measure μ directly from $\mathcal{S}$ to $\sigma(\mathcal{S})$. (It is possible to extend measure from a semialgebra $\mathcal{S}$ to $\mathcal{A}(\mathcal{S})$, and then extend it again to σ-algebra.

Theorem 1.2.15 — Caratheodory Extension Theorem. Let μ be a measure on a semialgebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$

1. μ has an extension to $\sigma(\mathcal{S})$, denoted by $\mu|_{\sigma(\mathcal{S})}$, so $\left(\Omega, \sigma(\mathcal{S}), \mu|_{\sigma(\mathcal{S})}\right)$ 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 (i) is unique. (i.e., if μ_1 and μ_2 are both extensions of μ to $\sigma(\mathcal{S})$, then $\mu_1 = \mu_2$)

(Remark: The extension from $\mathcal{S}$ to $\sigma(\mathcal{S})$ may not be unique if μ is not σ-finite.)

Proof.
1. From Theorem 1.2.14, we know $\sigma(\mathcal{S}) \subset \mathcal{A}^*$. Then, by Theorem 1.2.13 , the restriction of μ^* on $\sigma(\mathcal{S})$ is an extension of μ from $\mathcal{S}$ to $\sigma(\mathcal{S})$
2. μ is σ-finite $\Longrightarrow$ there exists disjoint $\{D_n\} \subset \mathcal{S}$, such that $\sum_{i=1}^\infty D_i = \Omega$ and $\mu(D_n) < \infty$ for each n. Suppose μ_1 and μ_2 are both extensions of measure μ from $\mathcal{S}$ to $\sigma(\mathcal{S})$. We shall first show:

$$\mu_1(A \cap D_n) = \mu_2(A \cap D_n), \quad \forall A \in \sigma(\mathcal{S}), \text{ and } \forall n \geq 1 \tag{1.5}$$

Proof. $\forall n \geq 1$, define

$$\mathcal{M} = \{A : A \in \sigma(\mathcal{S}), \mu_1(A \cap D_n) = \mu_2(A \cap D_n)\}$$

[This is, we collect all sets satisfying (1.5), and need to show that $\mathcal{M} = \sigma(\mathcal{S})$. Since $\mathcal{M} \subset \sigma(\mathcal{S})$, we only need to show that $\mathcal{M} \supset \sigma(\mathcal{S})$.]
Obviously, $\mathcal{S} \subset \mathcal{M}$. (To see this, if $A \in \mathcal{S}$, then clearly, $A \in \sigma(\mathcal{S})$ and also $\mu_1(A \cap D_n) = \mu_2(A \cap D_n)$ since $A \cap D_n \in \mathcal{S}$ and μ_1 and μ_2 are both extensions of μ from $\mathcal{S}$ to $\sigma(\mathcal{S})$. Hence, $A \in \mathcal{M}$.) So μ_1 and μ_2 are both extensions of measure μ from $\mathcal{S}$ to $\sigma(\mathcal{S})$. From Theorem 1.2.9, we know μ has a unique extension to $\mathcal{A}(\mathcal{S})$, i.e.,

$$\mu_1(A) = \mu_2(A), \quad \forall A \in \overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$$

Therefore, $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S}) \subset \mathcal{M}$. (To see this, if $A \in \overline{\mathcal{S}}$, then clearly, $A \in \sigma(\mathcal{S})$ and also

$$\mu_1(A \cap D_n) = \mu_2(A \cap D_n)$$

Since $A \cap D_n \in \overline{\mathcal{S}}$ and μ_1 and μ_2 are both unique extensions of μ from $\mathcal{S}$ to $\overline{\mathcal{S}}$. Hence, $A \in \mathcal{M}$.) From the Monotone Class Theorem of the last chapter, [i.e., if $\mathcal{M}$ is an m -class containing $\mathcal{S}$, then $\sigma(\mathcal{S}) \subset \mathcal{M}$], it suffices to show that $\mathcal{M}$ is an m -class, which implies that $\mathcal{M} \supset \sigma(\overline{\mathcal{S}}) = \sigma(\mathcal{S})$. By definition, $\mathcal{M} \subset \sigma(\mathcal{S})$. This shows $\mathcal{M} = \sigma(\mathcal{S})$ which in turn proves (1.5) ■

It remains to show that $\mathcal{M}$ is an m -class.

Proof. Let $\{A_k\} \in \mathcal{M}$ and $A_k \nearrow A = \cup_1^\infty A_k$. Note $\{A_k\} \in \mathcal{M} \Longrightarrow \mu_1(A_k \cap D_n) = \mu_2(A_k \cap D_n), n \geq 1, k \geq 1$. Also $\{A_n\} \in \mathcal{M} \subset \sigma(\mathcal{S})$, by the property of continuity from below for a measure on a σ -algebra, we get

$$\mu_1(A \cap D_n) = \lim_{k \to \infty} \mu_1(A_k \cap D_n) = \lim_{k \to \infty} \mu_2(A_k \cap D_n) = \mu_2(A \cap D_n)$$

Therefore, $A = \cup_1^\infty A_k \in \mathcal{M}$. Now Let $\{A_k\} \in \mathcal{M}$ and $A_k \searrow A = \cap_1^\infty A_k$. Since μ is σ -finite, we have $\mu_i(A_k \cap D_n) \leq \mu_i(D_n) = \mu(D_n) < \infty$. By the property of continuity from above for a measure on a σ algebra, we get

$$\mu_1(A \cap D_n) = \lim_{k \to \infty} \mu_1(A_k \cap D_n) = \lim_{k \to \infty} \mu_2(A_k \cap D_n) = \mu_2(A \cap D_n)$$

Therefore, $A = \cap_1^\infty A_k \in \mathcal{M}$. Thus, $\mathcal{M}$ is an m -class. ■

Finally, we show the uniqueness.

Proof. $\forall A \in \sigma(\mathcal{S})$, in view of (1.5) and σ -additivity of a measure, we have

$$\mu_1(A) = \mu_1\left(A \cap \sum_1^\infty D_n\right) = \sum_1^\infty \mu_1(A \cap D_n) = \sum_1^\infty \mu_2(A \cap D_n) = \mu_2(A)$$

Thus, μ_1 and μ_2 are the same on $\sigma(\mathcal{S})$ The measure extension theorem has a useful application in probability. ■

■

Corollary 1.2.16 If P is a probability defined on a semialgebra $\mathcal{S}$ on Ω, then there exists a unique probability space $(\Omega, \sigma(\mathcal{S}), P^*)$ such that

$$P^*(A) = P(A), \quad \forall A \in \mathcal{S}$$

1.2.7 Completion of a measure

Definition 1.2.11 — μ-null/complete. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and $N \subset \Omega$

1. N is a μ **-null** set iff $\exists B \in \mathcal{A}$ with $\mu(B) = 0$ 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 1.2.17 — Any measurable space can always be completed. Given a measure space $(\Omega, \mathcal{A}, \mu)$, there exists a complete space $(\Omega, \overline{\mathcal{A}}, \bar{\mu})$ such that $\mathcal{A} \subset \overline{\mathcal{A}}$ and $\bar{\mu} = \mu$ on $\mathcal{A}$

Proof. We shall prove the theorem in several steps.

1. Take

$$\begin{aligned} \overline{\mathcal{A}} &= \{A \cup N : A \in \mathcal{A}, \quad N \text{ is a } \mu \text{ -null set } \} \\ \overline{\mathcal{B}} &= \{A \Delta N : A \in \mathcal{A}, \quad N \text{ is a } \mu \text{ -null set } \} \end{aligned} \tag{1.6}$$

We shall show that $\overline{\mathcal{A}} = \overline{\mathcal{B}}$

Proof. Let $A \in \mathcal{A}$ and N be a μ -null set, i.e., $\exists B \in \mathcal{A}$ with $\mu(B) = 0$ such that $N \subset B$. It is easy to show (via diagrams, e.g.) that

$$\begin{aligned} A \cup N &= (A - B) + (B \cap (A \cup N)) = (A - B)\Delta(B \cap (A \cup N)) \\ A \Delta N &= (A - B) + (B \cap (A \Delta N)) = (A - B) \cup (B \cap (A \Delta N)) \end{aligned} \tag{1.7}$$

(By using B to construct two new sets to connect two operators.) Note that $A - B \in \mathcal{A}$. Also $(B \cap (A \cup N)) \subset B$ and $(B \cap (A \Delta N)) \subset B$ are both μ -null sets.
To finish the proof, let $\bar{A} := A \cup N \in \overline{\mathcal{A}}$. From the above relation, it is clear that $\bar{A} := A \cup N \in \overline{\mathcal{B}}$. Therefore, $\overline{\mathcal{A}} \subset \overline{\mathcal{B}}$. Similarly, we can show that $\overline{\mathcal{B}} \subset \overline{\mathcal{A}}$ ■

2. Secondly, we'll show that $\overline{\mathcal{A}}$ is a σ -algebra.

Proof. If $E_i = A_i \cup N_i$ where $A_i \in \mathcal{A}, N_i \subset B_i$ where $\mu(B_i) = 0$, then $\cup_i A_i \in \mathcal{A}$, and subadditivity implies $\mu(\cup_i B_i) \leq \sum_i \mu(B_i) = 0$, so $\cup_i E_i \in \overline{\mathcal{A}}$. As for complements, if $E = A \cup N$ and $N \subset B$, then $N^c \supset B^c$, so

$$E^c = A^c \cap N^c = (A^c \cap B^c) \cup (A^c \cap N^c \cap B)$$

where $A^c \cap B^c \in \mathcal{A}$ and $A^c \cap N^c \cap B \subset B$ (i.e., a μ -null set). So $E^c \in \overline{\mathcal{A}}$. Thus, $\overline{\mathcal{A}}$ is a σ -algebra.

■

3. Define a set function on $\overline{\mathcal{A}}$ by

$$\bar{\mu}(A \cup N) = \mu(A), \quad \text{for } A \cup N \in \overline{\mathcal{A}} \tag{1.8}$$

Then, we can show that

(a) $\bar{\mu}$ is well defined;

(b) $\bar{\mu}(A \Delta N) = \mu(A)$, for $A \Delta N \in \overline{\mathcal{A}}$

Proof. (a) We need to show that if $E = A_1 \Delta N_1 = A_2 \Delta N_2$ are two decompositions then $\mu(A_1) = \mu(A_2)$. Since$(A \Delta B) \Delta C = A \Delta (B \Delta C)$, we get

$$\begin{aligned} &(A_1 \Delta A_2) \Delta (N_1 \Delta N_2) = [(A_1 \Delta A_2) \Delta N_1] \Delta N_2 = [(A_1 \Delta N_1) \Delta A_2] \Delta N_2 \\ =&(A_1 \Delta N_1) \Delta (A_2 \Delta N_2) = (A_1 \Delta N_1) \Delta (A_1 \Delta N_1) = \emptyset \end{aligned}$$

(where the last equality follows since $I_{A \Delta A} = |I_A - I_A| = 0$). Therefore,

$$(A_1 \Delta A_2) = (N_1 \Delta N_2)$$

Since $N_1 \Delta N_2 \subset N_1 \cup N_2$, thus $N_1 \Delta N_2$ and $A_1 \Delta A_2$ are μ -null sets. So $\mu(A_1 \Delta A_2) = 0$. Consequently, $\mu(A_1 - A_2) = \mu(A_2 - A_1) = 0$. Therefore,

$$\begin{aligned} \mu(A_1) &= \mu(A_1 \cap A_2) + \mu(A_1 - A_2) = \mu(A_1 \cap A_2) \\ \mu(A_2) &= \mu(A_1 \cap A_2) + \mu(A_2 - A_1) = \mu(A_1 \cap A_2) \end{aligned}$$

(b) $\mu(B) = 0 \Longrightarrow \mu(A) = \mu(A - B) + \mu(A \cap B) = \mu(A - B)$. Then, from (1.7) and 1.8 , we get

$$\bar{\mu}(A \Delta N) = \bar{\mu}((A - B) \cup (B \cap (A \Delta N))) = \mu(A - B) = \mu(A)$$

∎

4. Next we show that $(\Omega, \overline{\mathcal{A}}, \bar{\mu})$ is a measure space. From (b) ,$\overline{\mathcal{A}}$ is a σ -algebra. From (1.6) and (1.8),$\bar{\mu}$ is σ -additive on $\overline{\mathcal{A}}$.

5. Finally, we show that $(\Omega, \overline{\mathcal{A}}, \bar{\mu})$ is the completion of $(\Omega, \mathcal{A}, \mu)$. That is, we need to show any $\bar{\mu}$ -null set is $\overline{\mathcal{A}}$ -measurable.

Proof. Let $\overline{N}$ be a $\bar{\mu}$ -null set, i.e.,

$$\exists \overline{B} \in \overline{\mathcal{A}} \text{ such that } \overline{N} \subset \overline{B} \text{ and } \bar{\mu}(\overline{B}) = 0$$

Note that

$$\begin{aligned} &\text{“}\exists\, \overline{B} \in \overline{\mathcal{A}}\text{”} &&\Longrightarrow \exists A \in \mathcal{A} \text{ and a } \mu\text{-null set } N \text{s.t. } \overline{B} = A \Delta N \\ &\text{“}\bar{\mu}(\overline{B}) = 0\text{”} &&\Longrightarrow \bar{\mu}(\overline{B}) = \bar{\mu}(A \Delta N) = \mu(A) = 0 \\ &\text{“}N \text{ a } \mu \text{ -null set ”} &&\Longrightarrow \exists B \in \mathcal{A} \text{ and } N \subset B \text{ s.t. } \mu(B) = 0 \end{aligned}$$

Thus,

$$\overline{N} \subset \overline{B} = A\Delta N \subset A \cup N \subset A \cup B \in \mathcal{A}$$

$$\mu(A \cup B) \leq \mu(A) + \mu(B) = 0$$

That is, $\overline{N}$ is μ -null set. Therefore,

$$\overline{N} = \emptyset \Delta \overline{N} \in \overline{\mathcal{A}}$$

■

■

R

1. We shall call $(\Omega, \overline{\mathcal{A}}, \bar{\mu})$ to be the completion of $(\Omega, \mathcal{A}, \mu)$.)
2. From the proof below, we shall see that one can take

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

$$\bar{\mu}(A\Delta N) = \bar{\mu}(A \cup N) = \mu(A), \text{ for } A\Delta N \in \overline{\mathcal{A}} \text{ and } A \cup N \in \overline{\mathcal{A}}$$

Lemma 1.5 — Measurable cover on algebra. Let μ be a measure on a semialgebra $\mathcal{S}$, 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.)

Proof. $\forall n \geq 1, \exists \{F_{nk}, k \geq 1\}$ such that $A \subset \cup_{k=1}^{\infty} F_{nk} := B_n$ with $F_{nk} \subset \mathcal{S}$ and

$$\mu^*(A) \leq \sum_{k=1}^{\infty} \mu(F_{nk}) \leq \mu^*(A) + \frac{1}{n}$$

(definition of outer measure). Since μ^* is an extension of the measure μ from $\mathcal{S}$ to $\sigma(\mathcal{S})$, we get

$$\mu^*(A) \leq \mu^*(B_n) \leq \sum_{k=1}^{\infty} \mu(F_{nk}) \leq \mu^*(A) + \frac{1}{n}$$

(the first leq is for monotonicity $A \subset B_n$, the second leq is for subadditive) Let $B = \cap_{n=1}^{\infty} B_n$. Then $B \in \sigma(\mathcal{S})$ and $B \supset A$, which proves (i). Hence,

$$\mu^*(A) \leq \mu^*(B) \leq \mu^*(B_n) \leq \mu^*(A) + \frac{1}{n}$$

Letting $n \to \infty$ results in $\mu^*(A) = \mu^*(B)$, which proves (ii). Note that μ^* is also a measure on $(\Omega, \mathcal{A}^*)$ where $\mathcal{A}^* \supset \sigma(\mathcal{S})$. So if $C \subset B - A$ and $C \in \sigma(\mathcal{S})$, then $\mu^*(C) \leq \mu^*(B) - \mu^*(A) = 0$, which proves (iii). ■

Theorem 1.2.18 Let μ be a σ -finite measure on a semialgebra $\mathcal{S}$, 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 $\left(\Omega, \sigma(\mathcal{S}), \mu^*|_{\sigma(\mathcal{S})}\right)$.

Proof. For simplicity, we shall write

$$\mu^* := \mu^*|_{\mathcal{A}^*}, \quad \mu_\sigma := \mu^*|_{\sigma(\mathcal{S})} = \mu|_{\sigma(\mathcal{S})}$$

Let $\left(\Omega, \overline{\sigma(\mathcal{S})}, \bar{\mu}_\sigma\right)$ be the completion of $(\Omega, \sigma(\mathcal{S}), \mu_\sigma)$ as in the last theorem. So

$$\overline{\sigma(\mathcal{S})} = \{A \cup N : A \in \sigma(\mathcal{S}), \quad N \text{ is a } \mu_\sigma\text{-null set}\},$$
$$\bar{\mu}_\sigma(A \cup N) = \mu_\sigma(A), \text{ for } A \cup N \in \overline{\sigma(\mathcal{S})}$$

It suffices to show that

$$\overline{\sigma(\mathcal{S})} = \mathcal{A}^* \tag{1.9}$$

the proof of which will be given a little later. To see why (1.9) is enough, let $E \in \mathcal{A}^*, \Longrightarrow E \in \overline{\sigma(\mathcal{S})}. \Longrightarrow \exists A \in \sigma(\mathcal{S})$ and a μ_σ-null set N such that $E = A \cup N$. Since μ^* is a measure on $(\Omega, \mathcal{A}^*)$, and N is μ^* -measurable with $\mu^*(N) = 0$, then we have

$$\mu^*(A) \leq \mu^*(E) \leq \mu^*(A \cup N) \leq \mu^*(A) + \mu^*(N) = \mu^*(A) = \mu_\sigma(A)$$

Hence, we have

$$\mu^*(A \cup N) = \mu^*(E) = \mu_\sigma(A)$$

which implies that $(\Omega, \mathcal{A}^*, \mu^*)$ is indeed the completion of $(\Omega, \sigma(\mathcal{S}), \mu_\sigma)$.

It remains to prove (1.9) . Proof of (1.9)

1. We first show that $\overline{\sigma(\mathcal{S})} \subset \mathcal{A}^*$. If $E := A \cup N \in \overline{\sigma(\mathcal{S})}$, where $A \in \sigma(\mathcal{S}) \subset \mathcal{A}^*$ and N is μ_σ -null set. It is easy to show that $N \in \mathcal{A}^*$

   Proof. N is μ_σ -null set, $\Longrightarrow N \subset B \in \sigma(\mathcal{S}) \subset \mathcal{A}^*$ such that $\mu_\sigma(B) = 0. \Rightarrow 0 \leq \mu^*(N) \leq \mu^*(B) = \mu_\sigma(B) = 0$, i.e., $\mu^*(N) = 0, \Rightarrow \forall D \subset \Omega, \mu^*(D) \geq \mu^*(D \cap N^c) + 0 = \mu^*(D \cap N^c) + \mu^*(N) \geq \mu^*(D \cap N^c) + \mu^*(D \cap N). \Rightarrow N$ is μ^* -measurable, i.e., $N \in \mathcal{A}^*$. ■

   Since $A, N \in \mathcal{A}^*$ and $\mathcal{A}^*$ is a σ -algebra, we have $E := A \cup N \in \mathcal{A}^*$. Thus, $\overline{\sigma(\mathcal{S})} \subset \mathcal{A}^*$
2. Now we show that $\mathcal{A}^* \subset \overline{\sigma(\mathcal{S})}$. Let $A \in \mathcal{A}^*$

(a) Case I: $\mu^*(A) < \infty$. From Lemma 1.5, $\exists B \in \sigma(\mathcal{S})$ such that $A \subset B$ and $\mu^*(A) = \mu^*(B)$[i.e. B is a measurable cover of A in $\sigma(\mathcal{S})$]. Since μ^* is a measure on $(\Omega, \mathcal{A}^*)$, we get

$$\mu^*(B - A) = \mu^*(B) - \mu^*(A) = 0$$

Let C be a measurable cover of $B - A$ in $\sigma(\mathcal{S})$, i.e.,

$$C \supset B - A, \quad C \in \sigma(\mathcal{S}), \quad \mu^*(C) = \mu^*(B - A) = 0$$

Clearly, $A = (B - C) + (A \cap C)$ (draw a diagram to illustrate), and

$$B - C \in \sigma(\mathcal{S}), \quad A \cap C \subset C, \quad C \in \sigma(\mathcal{S}), \quad \mu^*(C) = 0 = \mu_\sigma(C)$$

Thus, $A \cap C$ is a μ_σ -null set. This implies that $A \in \overline{\sigma(\mathcal{S})}$

(b) Case II: $\mu^*(A) = \infty$. (Using the σ-finite of μ) Since μ_σ is σ -finite, therefore its extension μ^* is also σ -finite. Thus, $\forall A \in \mathcal{A}^*$, we have $A = \sum_{i=1}^{\infty} A_n$ with $A_n \in \mathcal{A}^*$ and $\mu^*(A_n) < \infty$

Proof. μ^* is σ -finite, $\Longrightarrow \Omega = \sum_1^\infty E_n$ with $E_n \in \mathcal{A}^*$ and $\mu^*(E_n) < \infty$. $\Rightarrow A = A \cap \Omega = \sum_1^\infty A \cap E_n := \sum_1^\infty A_n$ with $A_n := A \cap E_n \in \mathcal{A}^*$ and $\mu^*(A_n) \le \mu^*(E_n) < \infty$ ■

It follows from the last paragraph that $A_n \in \overline{\sigma(\mathcal{S})}$, which implies that $A = \sum_{i=1}^{\infty} A_n \in \overline{\sigma(\mathcal{S})}$ as $\overline{\sigma(\mathcal{S})}$ is a σ -algebra. This proves that $\mathcal{A}^* \subset \overline{\sigma(\mathcal{S})}$

Combining (i) and (ii), we have shown $\mathcal{A}^* = \overline{\sigma(\mathcal{S})}$ ■

R From the proof of the last theorem, we have seen that $\mathcal{A}^* = \overline{\sigma(\mathcal{S})}$. Therefore, we can write

$$\mathcal{A}^* = \sigma(\mathcal{S}) + \{\text{all } \mu_\sigma - \text{null sets}\}$$

In other words, the gap between $\mathcal{A}^*$ and $\sigma(\mathcal{S})$ is filled with all all μ_σ -null sets.

1.2.8 Construction of measures on a σ -algebra $\mathcal{A}$

Theorem 1.2.19 — General procedures. Here is one useful way of constructing measures on a σ -algebra $\mathcal{A}$.

1. Identify a semialgebra $\mathcal{S}$ so that $\mathcal{A} = \sigma(\mathcal{S})$.
2. Define a map $\mu : \mathcal{S} \to R$ so that μ is a measure on $\mathcal{S}$. (The next theorem is useful in proving μ is a measure on $\mathcal{S}$.)
3. Extend the measure from $\mathcal{S}$ to $\mathcal{A} = \sigma(\mathcal{S})$ by the measure extension theorem.

$$\begin{array}{ccccccccc} \mathcal{S} & \Longrightarrow & \overline{\mathcal{S}} & \Longrightarrow & \mathcal{A}^* & \Longrightarrow & \sigma(\mathcal{S}) & = & \mathcal{A} \\ \mu & \Longrightarrow & \bar{\mu}_\sigma & \Longrightarrow & \mu^*|_{\mathcal{A}^*} & \Longrightarrow & \mu^*|_{\sigma(\mathcal{S})} & = & \mu|_{\sigma(\mathcal{S})} \end{array}$$

Note that the first two relations are extensions while the last one is a restriction. Usually, μ^* can not be extended further to the power set $\mathcal{P}(\Omega)$.

Theorem 1.2.20 — Useful in finding a measure on semialgebra $\mathcal{S}$. Let μ be a nonnegative set function on a semi-algebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$. Assume that

1. μ is finite additive on $\mathcal{S}$.

$$\left(\text{i.e. } \mu(A) = \sum_1^n \mu(A_i) \text{ whenever } A_n \in \mathcal{S} \text{ and } A = \sum_1^n A_i \in \mathcal{S}.\right)$$

2. μ is σ -subadditive on $\mathcal{S}$. (i.e. $\mu(A) \leq \sum_1^\infty \mu(A_i)$ whenever $A, A_n \in \mathcal{S}$ and $A \subset \sum_1^\infty A_n$ ([or $A = \sum_1^\infty A_n$, or $A \subset \cup_1^\infty A_n$])

Then μ is a measure on $\mathcal{S}$.

Proof. Since μ is nonnegative and finitely additive, it can be easily shown that $\mu(\emptyset) = 0$. It remains to check if μ is σ -additive. That is, if $A = \sum_1^\infty A_n$, where $A, A_n \in \mathcal{S}$, we'd like to show

$$\mu(A) = \sum_1^\infty \mu(A_n)$$

From Theorem 1.2.10, we have

$$\mu(A) \geq \sum_1^\infty \mu(A_n)$$

Combining this with assumption (ii), we get $\mu(A) = \sum_1^\infty \mu(A_n)$ ∎

Lebesgue and Lebesgue-Stieltjes measures

Theorem 1.2.21 — L-S measure induced by L-S function. Suppose that F is finite on $(-\infty, \infty)$ (i.e. $|F(t)| < \infty$ for $|t| < \infty$) and

1. F is nondecreasing;
2. F is right continuous.

Then there is a unique measure μ on $(R, \mathcal{B})$ with

$$\mu((a,b]) = F(b) - F(a), \quad -\infty \leq a \leq b \leq \infty$$

(When $a = b = \infty$ or $-\infty$, the right hand is understood to be 0.)

Corollary 1.2.22 — Lebesgue measure. There is a unique measure λ on $(R, \mathcal{B})$ with

$$\mu((a,b]) = b - a, \quad -\infty \leq a < b \leq \infty$$

$(F(x) = x)$

R

1. A function F which is nondecreasing and right continuous is called a **Lebesgue-Stieltjes (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. Clearly, 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 function 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. $\Omega = R$ 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.

Proof of Theorem 1.2.21

Proof. Let $\mathcal{S} = \{(a,b] : -\infty \leq a < b \leq \infty\} \cup \{R, \emptyset, \{-\infty\}\}$. It can be shown that $\mathcal{S}$ is a semi-algebra and $\mathcal{B} = \sigma(\mathcal{S})$. Define

$$\mu((a,b]) = F(b) - F(a), \quad \text{if } -\infty \leq a < b \leq \infty$$
$$\mu(\{-\infty\}) = 0$$
$$\mu(R) = F(\infty) - F(-\infty)$$

where

$$F(\infty) = \lim_{x \nearrow \infty} F(x), \quad F(-\infty) = \lim_{x \searrow -\infty} F(x)$$

R Several remarks are in order before we move on:

1. Recall $\Omega = R = [-\infty, \infty]$. Since we assume that $\Omega = R \in \mathcal{S}$, we need to add R to $\mathcal{S}$.*As* a consequence, we also need to add $\{-\infty\} = [-\infty,\infty] - (-\infty,\infty] = (-\infty,\infty]^c$ to $\mathcal{S}$ as well.
2. Note that both $F(-\infty)$ and $F(\infty)$ exist (but may take $\pm\infty$ value) as F is monotone.
3. The definition of $F(\infty)$ implies that F is left continuous at ∞.
4. If we consider $\Omega = (-\infty, \infty]$ instead of $[-\infty, \infty]$, then $\mathcal{S} = \{(a,b] : -\infty \leq a \leq b \leq \infty\}$ will form a semialgebra. The construction would be simpler in that case. (One can then extend the measure again to $[-\infty, \infty]$.)

Let us come back to the proof of the theorem.

1. First we check that μ is well-defined, i.e. we should not have $\infty - \infty$ or $(-\infty) - (-\infty)$. Clearly, there is no problem when both a and b are finite, or one of them is finite and the other one is infinite. Finally, when $a = -\infty$ and $b = \infty, \mu((a,b])$ is also well defined since $F(\infty) > -\infty$ and $F(-\infty) < \infty$, otherwise, we would have either $F(t) = \infty$ for all $t \in R$ or $F(t) = -\infty$ for all $t \in R$, which is excluded from our consideration.
2. Secondly, we show that μ defined above is a measure on the semialgebra $\mathcal{S}$ by way of Theorem 1.2.20.(finite additive +σ-subadd) Clearly, μ is nonnegative.
   (a) Checking finite additivity: Let $S = \sum_{i=1}^{n} S_i$, where $S, S_i \in \mathcal{S}$
      i. If $S = \{\emptyset\}$, or $= \{-\infty\}$, finite additivity is trivial since $S = S$. If $S = R$ then finite additivity holds for $R = \sum_{i=1}^{n} S_i$ iff it holds for $R - \{-\infty\} = \sum_{i=1}^{n} (S_i - \{-\infty\})$. (Note that one and only one of S_i 's contains $\{-\infty\}$.) But it is easy to see that $R - \{-\infty\} = (-\infty, \infty]$ and $S_i - \{-\infty\}$'s are all half intervals of the form $(a,b]$, where $-\infty \le a \le b \le \infty$. So the case is reduced to the following case (ib).
      ii. If $S = (a,b]$, then we must have $S_i = (a_i, b_i]$. That is, $(a,b] = \sum_1^n (a_i, b_i]$. Then after some relabelling the intervals, we must have $a_1 = a, b_n = b$, and $a_i = b_{i-1}$ for $2 \le i \le n$. So by definition, we have $\mu((a,b]) = \sum_1^n \mu((a_i, b_i]) = F(b) - F(a)$, i.e., μ is finitely additive.
   (b) Checking σ -subadditivity. Let $S \subset \cup_{i=1}^{\infty} S_i$, where $S, S_i \in \mathcal{S}$. If $S \in \{R, \emptyset, \{-\infty\}\}, \sigma$ -subadditivity is easy to check, similar to (i) above. It suffices to consider $S = (a,b]$ and $S_i = (a_i, b_i]$. That is, let $(a,b] \subset \cup_1^{\infty} (a_i, b_i]$.
      i. Case $I : -\infty < a < b < \infty$. W.L.O.G., assume that $-\infty < a_i < b_i < \infty$ for all i. By right continuity of F, we can pick $\delta > 0$ and $\eta_i > 0$ so that

$$
\begin{aligned}
&F(a+\delta) < F(a) + \epsilon \\
&F(b_i + \eta_i) < F(b_i) + \epsilon/2^i
\end{aligned}
$$

Note that $[a+\delta, b] \subset (a,b] \subset \cup_1^{\infty} (a_i, b_i] \subset \cup_1^{\infty} (a_i, b_i + \eta_i)$, i.e., the closed interval has a countable cover by the open intervals $(a_i, b_i + \eta_i)$. So there is a finite subcover $(\alpha_i, \beta_i), 1 \le j \le J$(Compact). By monotonicity of F, we get

$$\begin{aligned}F(b)-F(a) &= [F(b)-F(a+\delta)]+[F(a+\delta)-F(a)]\\ &\le [F(b)-F(a+\delta)]+\epsilon\\ &\le \sum_{j=1}^{J}\left[F\left(\beta_j\right)-F\left(\alpha_j\right)\right]+\epsilon\\ &\le \sum_{i=1}^{\infty}\left[F\left(b_i+\eta_i\right)-F\left(a_i\right)\right]+\epsilon\\ &\le \sum_{i=1}^{\infty}\left[F\left(b_i\right)-F\left(a_i\right)\right]+2\epsilon\end{aligned}$$

Letting $\epsilon \to 0$, we get

$$\mu((a,b]) \le \sum_{i=1}^{\infty}\mu\left(\left(a_i,b_i\right]\right)$$

ii. Case II: $-\infty = a < b < \infty$, i.e., $(-\infty,b] \subset \cup_1^{\infty}\left(a_i,b_i\right]$. Let $-\infty < M < b < \infty$. Then $[M,b] \subset \cup_1^{\infty}\left(a_i,b_i\right]$. From Case I, we get

$$F(b)-F(M) \le \sum_{i=1}^{\infty}\left[F\left(b_i\right)-F\left(a_i\right)\right]$$

Letting $M \to -\infty$, we get $F(b)-F(-\infty) \le \sum_{i=1}^{\infty}\left[F\left(b_i\right)-F\left(a_i\right)\right]$, i.e.,

$$\mu((a,b]) \le \sum_{i=1}^{\infty}\mu\left(\left(a_i,b_i\right]\right)$$

iii. Case III: $-\infty < a < b = \infty$. Proof is similar to (b).

iv. Case IV : $-\infty = a < b = \infty$. Proof is similar to (b)

Combining (i) and (ii), we have shown that μ is indeed a measure on the semialgebra $\mathcal{S}$ by Theorem 1.2.20

Also we can see that μ is σ -finite as

$$\Omega = \mathcal{R} = [-\infty,\infty] = \{-\infty\} + \sum_{n=-\infty}^{\infty}(n,n+1] + \{\infty\}$$

and each component on the right hand side has finite measure.

Finally, we can apply Caratheodory's Extension Theorem to get the desired result. ∎

Relationship between probability measures and distribution functions

R We mentioned that there is no $1-1$ correspondence between all L-S measures on $(\mathcal{R},\mathcal{B})$ and all L-S measure functions. However, there is a $1-1$ correspondence between all probability measures on $(\mathcal{R},\mathcal{B})$ and distribution functions.

Definition 1.2.12 A real-valued function F on $\mathcal{R}$ is **distribution function** (d.f.) if

1. $F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \quad F(\infty) = \lim_{x\to\infty} F(x) = 1$
2. F is nondecreasing, i.e., $F(x) \le F(y)$ if $x \le y$
3. F is right continuous, i.e., $F(y) \searrow F(x)$ if $y \searrow x$

The difference between L-S function and d.f. is the first condition.

Theorem 1.2.23 — Correspondence theorem. The relation

$$F(x) = P((-\infty, x]), \quad x \in \mathcal{R} \tag{1.10}$$

establishes a $1-1$ correspondence between all d.f.'s and all probability measures on $(\mathcal{R}, \mathcal{B})$.

Proof. 1. Given a probability measure P on $(\mathcal{R}, \mathcal{B})$, we first show that F determined by (1.10) is a d.f. (i.e. $(a)-(c)$ above hold)

(a) Given $x_n \searrow -\infty$, it follows that $(-\infty, x_n] \searrow 0$. Since P is continuous from above, we have

$$F(x_n) = P((-\infty, x_n]) \searrow P(\emptyset) = 0$$

Thus, $\lim_{x\searrow-\infty} F(x) = 0$. Similarly, $\lim_{x\nearrow\infty} F(x) = 1$

(b) If $x \le y \Longrightarrow (-\infty, x] \subset (-\infty, y]$. By the monotonicity of P, we get

$$F(x) = P((-\infty, x]) \le P((-\infty, y]) = F(y)$$

(c) Given $x_n \searrow x$, it follows that $(-\infty, x_n] \searrow (-\infty, x]$. Since P is continuous from above, we have

$$F(x_n) = P((-\infty, x_n]) \searrow P((-\infty, x]) = F(x)$$

Thus, $\lim_{y\searrow x} F(x) = F(y)$, i.e. F is right continuous.

2. Given a d.f. F, it must be a L-S measure function. Applying Theorem 1.2.21, there exists a unique probability measure P (induced by L-S function) on $(\mathcal{R}, \mathcal{B})$ satisfying $F(x) - F(y) = P((y, x])$ for $x \le y$. Letting $y \searrow -\infty$ we have

$$F(x) = \lim_{y\searrow-\infty} [F(x) - F(y)] = \lim_{y\searrow-\infty} P((y, x]) = P((-\infty, x])$$

■

Proposition 1.2.24 1. From the above theorem, we can define a distribution function (d.f.) by

$$F(x) = P((-\infty, x]), \quad x \in \mathcal{R}$$

This definition does not involve any random variables (which we have not discussed yet).

2. F is a d.f. and P is a probability measure. Then the following are equivalent
   (a) $P((-\infty,b]) = F(b)$
   (b) $P((a,b]) = F(b) - F(a)$
   (c) $P([a,b]) = F(b) - F(a-)$
   (d) $P([a,b)) = F(b-) - F(a-)$
   (e) $P((a,b)) = F(b-) - F(a)$
   (f) $P((-\infty,b)) = F(b-)$
   (g) $P(\{a\}) = F(a) - F(a-)$
3. Let X be a random variable (to be introduced later) on $(\Omega, \mathcal{A}, P)$, and let $F_X(x) = P(X \leq x)$. Then F is a right continuous, nondecreasing function with $F(-\infty) = 0$ and $F(\infty) = 1$.
4. For a d.f. F, the set

$$S(F) = \{x : F(x+\epsilon) - F(x-\epsilon) > 0, \text{ for all } \epsilon > 0\}$$

is called the **support** of F. Furthermore, any point $x \in S(F)$ 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 $(-\infty, \infty)$.

Proof. 1. x is a jump point if $F(x) - F(x-) > 0$, clearly $F(x+\epsilon) - F(x-\epsilon) > F(x) - F(x-) > 0$ for all $\epsilon > 0$. Thus, $x \in S(F)$
2. Let $\{x_n, n \geq 1\} \in S(F)$ and $x_n \to x$. We need to show that $x \in S(F)$. Given $\epsilon > 0$, $\exists N_0 > 0$ such that for all $n \geq N_0$, we have $|x_n - x| \leq \epsilon/2$, or $x_n - \epsilon/2 < x < x_n + \epsilon/2$ Therefore, choosing any $n_0 \geq N_0$, we have

$$\begin{aligned} F(x+\epsilon) - F(x-\epsilon) &\geq F\left(x_{n_0} - \epsilon/2 + \epsilon\right) - F\left(x_{n_0} + \epsilon/2 - \epsilon\right) \\ &\geq F\left(x_{n_0} + \epsilon/2\right) - F\left(x_{n_0} - \epsilon/2\right) > 0 \end{aligned}$$

3. The discrete d.f. with positive jump size at each rational number is such an example.

■

Definition 1.2.13 — Different types of distributions. 1. A d.f F is called **discrete** if it can be represented in the form

$$F(x) = \sum_{1}^{\infty} p_n \delta_{a_n}(x)$$

(check that it is indeed a d.f.) where $\{a_n, n \geq 1\}$ is a **countable** set of real numbers, $p_j > 0$ for all $j \geq 1$ and $\sum p_j = 1$

2. A d.f F is called **continuous** if it is continuous everywhere.

Definition 1.2.14 1. Degenerate distribution functions. The unique L-S probability measure on $(\mathcal{R}, \mathcal{B})$ for a degenerate d.f. $\delta_t(x) = I\{x \geq t\}$ is

$$\begin{aligned} P((a,b]) = \delta_t(b) - \delta_t(a) &= 1 \quad t \in (a,b] \\ &= 0 \quad t \notin (a,b] \end{aligned}$$

(In particular, we have $P(\{t\}) = \delta_t(t) - \delta_t(t-) = 1 - 0 = 1$.)
Another probability measure for $\delta_t(x) = I\{x \geq t\}$ defined on a different measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\Omega = \{t\}$ and so $\mathcal{P}(\Omega) = \{\emptyset, \Omega\} = \{\emptyset, \{t\}\}$, is

$$P_1(\{t\}) = 1, \quad P_1(\emptyset) = 0$$

2. Discrete distribution functions. The unique L-S probability measure on $(\mathcal{R}, \mathcal{B})$ for a discrete d.f. $F(x) = \sum_{n=1}^{\infty} p_n \delta_{a_n}(x)$ is

$$\begin{aligned} P((a,b]) = F(b) - F(a) &= \sum_{n=1}^{\infty} p_n \delta_{a_n}(b) - \sum_{n=1}^{\infty} p_n \delta_{a_n}(a) \\ &= \sum_{n=1}^{\infty} p_n \left[\delta_{a_n}(b) - \delta_{a_n}(a)\right] = \sum_{\{n: a < a_n \leq b\}} p_n \end{aligned}$$

Since $\delta_{a_n}(b) - \delta_{a_n}(a) = 1$ iff $a < a_n \leq b$.
Another probability measure for $F(x)$ defined on a different measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\Omega = \{a_1, a_2, \ldots, a_n, \ldots\}$ (either finite or countably infinite), is

$$P_2(A) = \sum_{k \geq 1} p_{i_k}, \quad \text{for } A = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}, \ldots\}$$

3. The set of jump points for a discrete d.f. can be dense. Let $\{a_n, n \geq 1\}$ be any given enumeration of the set of all rational numbers, and let $\{p_n, n \geq 1\}$ be a set of positive numbers such that $\sum p_n = 1$, e.g., $p_n = 2^{-n}$. Define

$$F(x) = \sum_{1}^{\infty} p_n \delta_{a_n}(x)$$

Since $0 \leq \delta_t(x) \leq 1$, the sequence on the right hand side of F is absolutely and uniformly convergent(M test). Also since $\delta_t(x)$ is increasing, we have for $x < y$

$$F(y) - F(x) = \sum_{1}^{\infty} p_n \left(\delta_{a_n}(y) - \delta_{a_n}(x)\right) \geq 0$$

Hence, F is nondecreasing. Due to the uniform convergence, we may deduce that

$$\begin{aligned} F(x) - F(x-) &= \sum_{1}^{\infty} p_n \left(\delta_{a_n}(x) - \delta_{a_n}(x-)\right) \\ &= 0 \quad x \notin \{a_n, n \geq 1\} \\ &= p_n \quad x = a_n, \quad n \geq 1 \end{aligned}$$

Therefore, $F(x)$ has jumps at all the rational points and nowhere else.

R The example shows that the set of points of jump of an increasing function may be countably infinite and everywhere dense.

Decomposition of distribution functions

Lemma 1.6 The set of discontinuities of a non-decreasing function is countable.

Proof. Every discontinuous point of an increasing function must be a jump and every jump contains a rational number. ∎

Definition 1.2.15 — F_d. 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

$$F_d(x) = \sum_{j=1}^{\infty} p_j \delta_{a_j}(x) = \text{sum of all the jumps of } F \text{ in } (-\infty, x] \tag{1.11}$$

It is clearly nondecreasing, right continuous with

$$F_d(-\infty) = 0, \quad F_d(\infty) = \sum_j p_j$$

Theorem 1.2.25 Let $F_c(x) = F(x) - F_d(x)$. Then F_c is nonnegative, nondecreasing, and continuous.

Proof. Let $x < y$, then

$$\begin{aligned} F_d(y) - F_d(x) = & \text{ sum of jumps in } (-\infty, y] - \text{sum of jumps in } (-\infty, x] \\ &= \sum_{x < a_j \leq y} p_j = \sum_{x < a_j \leq y} \left(F(a_j) - F(a_j-)\right) \\ &\leq F(y) - F(x) \end{aligned} \tag{1.12}$$

1. From (1.12), we get

$$F_c(y) - F_c(x) = [F(y) - F(x)] - [F_d(y) - F_d(x)] \geq 0$$

Thus, F_c is nondecreasing (so are F_d and F).

2. Letting $x \to -\infty$ in (1.12), we get $F_d(y) \le F(y)$. Thus $F_c(y) = F(y) - F_d(y) \ge 0$. So F_c is nonnegative.
3. Finally, by definition of jump point, we have

$$\begin{aligned} F(x) - F(x-) =& p_j \quad \text{if } x = a_j, j \ge 1 \\ =& 0 \quad \text{otherwise} \end{aligned}$$

Similarly, by definition of F_d in (1.11), we have

$$\begin{aligned} F_d(x) - F_d(x-) =& p_j \quad \text{if } x = a_j, j \ge 1 \\ =& 0 \quad \text{otherwise.} \end{aligned}$$

Therefore,

$$F_c(x) - F_c(x-) = [F(x) - F(x-)] - [F_d(x) - F_d(x-)] = 0$$

That is, F_c is continuous.

■

Theorem 1.2.26 [The decomposition is unique] If $F(x) = G_c(x) + G_d(x)$, where G_c is continuous and G_d is discrete, i.e. $G_d(x) = \sum_i p'_i \delta_{a'_i}(x)$. Then

$$G_c(x) = F_c(x), \quad \text{and} \quad G_d(x) = F_d(x)$$

Proof. If $F_d \neq G_d$, then either

1. $\{a_n, n \ge 1\} \neq \{a'_n, n \ge 1\}$, or
2. $a_i = a'_i$ for all $i \ge 1$ (after some reordering), but $p_i \neq p'_i$ for some $i \ge 1$
   In either case, there exists at least one $\tilde{a} = a_i$ or a'_i for some $i \ge 1$, such that

$$F_d(\tilde{a}) - F_d(\tilde{a}-) \neq G_d(\tilde{a}) - G_d(\tilde{a}-)$$

In view of $F(x) = F_c(x) + F_d(x) = G_c(x) + G_d(x)$, we have

$$\begin{aligned} 0 &= F_c(\tilde{a}) - F_c(\tilde{a}-) \\ &= [F(\tilde{a}) - F_d(\tilde{a})] - [F(\tilde{a}-) - F_d(\tilde{a}-)] \\ &= [F(\tilde{a}) - F(\tilde{a}-)] - [F_d(\tilde{a}) - F_d(\tilde{a}-)] \\ &=\neq [F(\tilde{a}) - F(\tilde{a}-)] - [G_d(\tilde{a}) - G_d(\tilde{a}-)] \\ &= [F(\tilde{a}) - G_d(\tilde{a})] - [F(\tilde{a}-) - G_d(\tilde{a}-)] \\ &= G_c(\tilde{a}) - G_c(\tilde{a}-) \\ &= 0 \end{aligned}$$

This is a contradiction. Therefore, we must have $F_d = G_d$, and consequently $F_c = G_c$

■

Theorem 1.2.27 Every d.f. can be written as the convex combination of a discrete and a continuous one:

$$F(x) = \alpha F_1(x) + (1-\alpha)F_2(x). \quad F_1 : \text{discrete and} F_2 : \text{continuous}.$$

Such a decomposition is unique.

Proof. Given a d.f. $F(x)$, from the last theorem we can write $F(x) = F_d(x) + F_c(x)$ uniquely. Let $\alpha = F_d(\infty)$, then $1-\alpha = F_c(\infty)$

1. If $\alpha = 0$, then $0 \leq F_d(x) \leq F_d(\infty) = \alpha = 0$ for all x, i.e., $F_d(x) = 0$ for all x. Hence, $F_c(x)$ is a proper d.f., and we can write $F(x) = 0 \times F_d(x) + 1 \times F_c(x)$.
2. If $\alpha = 1$, then $0 \leq F_c(x) \leq F_c(\infty) = 1-\alpha = 0$ for all x, i.e., $F_c(x) = 0$ for all x. Hence, $F_d(x)$ is a proper d.f., and we can write $F(x) = 1 \times F_d(x) + 0 \times F_c(x)$
3. If $0 < \alpha < 1$, then F can be written as

$$\begin{aligned} F(x) = F_d(x) + F_c(x) &= \alpha \frac{F_d(x)}{F_d(\infty)} + (1-\alpha)\frac{F_c(x)}{F_c(\infty)} \\ &= \alpha F_1(x) + (1-\alpha)F_2(x) \end{aligned}$$

where it is easy to see that F_1 and F_2 are both d.f.'s.

■

Corollary 1.2.28 Every d.f. F is either a discrete d.f. or a continuous d.f., or a combination (or a mixture) of both.

Further decomposition of a continuous d.f. F

We can further decompose a continuous d.f. as

A continuous d.f. = an absolutely continuous d.f. + a singular d.f.

Definition 1.2.16
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.
2. Alternative defintion: A d.f F is called absolutely continuous iff there exists a function $f \geq 0$ such that

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

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

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

Theorem 1.2.29 Every d.f. 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.

■ **Example 1.5 — a singular distribution–Uniform distribution on the Cantor set..** The Cantor set C is defined by removing $(1/3,2/3)$ from $[0,1]$ and then removing the middle third of each interval that remains. We define an associated d.f. by setting

$$F(x) = 0 \text{ for } x \leq 0$$
$$F(x) = 1 \text{ for } x \geq 1$$
$$F(x) = 1/2 \text{ for } x \in [1/3,2/3]$$
$$F(x) = 1/4 = 1/2^2 \text{ for } x \in [1/9,2/9] = \left[1/3^2,2/3^2\right]$$
$$F(x) = 3/4 = 1-1/2^2 \text{ for } x \in [7/9,8/9] = \left[\left(3^2-2\right)/3^2,\left(3^2-1\right)/3^2\right]$$

It can be shown that the resulting function F is defined for all all $x \in R$. Further it is nondecreasing, continuous, and $F(-\infty) = 0$ and $F(\infty) = 1$, i.e., it is a d.f. (Check carefully about the continuity of F). Also, it is easy to see that $F'(x) = 0$ for all $x \in R$ except perhaps for those on the Cantor set (i.e., those points $m/3^n, m,n \in \mathcal{N}$)(and the measure of Cantor set is zero). By definition, F is clearly a singular d.f., which is called "Lebesgue's singular function".

1.2.9 Radon-Nikodym theorem

Let μ and ν be two measures on the measurable space $(\Omega,\mathcal{F})$.

Definition 1.2.17 — Absolutely continuous. We say that ν is **absolutely continuous** w.r.t. μ, written as $\nu << \mu$, if

$$\mu(A) = 0 \text{ implies } \nu(A) = 0$$

Theorem 1.2.30 — Radon-Nikodym theorem. Given a measurable space (X,Σ), if a measure ν on (X,Σ) is absolutely continuous with respect to 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

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

for any measurable set A. The function f satisfying the above equality is uniquely defined up to a μ -null set, that is, if g is another function which satisfies the same property, then $f = g, \mu$ -almost everywhere. It is commonly written $d\nu/d\mu$, and is called the "**Radon-Nikodym derivative**".

Proposition 1.2.31 — **Properties of Radon-Nikodym derivative.** Let ν, μ, and λ be σ -finite measures on the same measure space.

1. If $\nu << \lambda$ and $\mu << \lambda (\nu$ and μ are absolutely continuous in respect to $\lambda)$, then
$$\frac{d(\nu+\mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda}$$
λ -almost everywhere.
2. If $\nu << \mu << \lambda$, then
$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$
λ -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
$$\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|$$

The assumption of sigma-finiteness

The Radon-Nikodym theorem assumes that the measure μ is σ -finite. Here is an example when μ is not σ -finite and the Radon-Nikodym theorem fails to hold.

■ **Example 1.6** Let $\mathcal{B}(R)$ be the Borel σ -algebra on the real line, (i.e. the minimal σ -algebra containing all open intervals). For $A \in \mathcal{B}(R)$, define

$$\mu(A) = |A|$$

the cardinality of A. That is, $\mu(A)$ is the number of elements of A if A is finite, and infinity otherwise. Let ν be the usual Lebesgue measure on this Borel algebra. Then we can show the following.

1. μ is indeed a measure, but it is not sigma-finite;

   Proof. The first part is easy to check. For the second part, note that not every Borel set (e.g. (0,1)) is at most a countable union of finite sets. ■

2. ν is absolutely continuous w.r.t. μ

   Proof. For a set A, one has $\mu(A) = 0$ if and only if A is the empty set, and then this implies that $\nu(A) = 0$ ■

3. The Radon-Nikodym theorem does not hold.

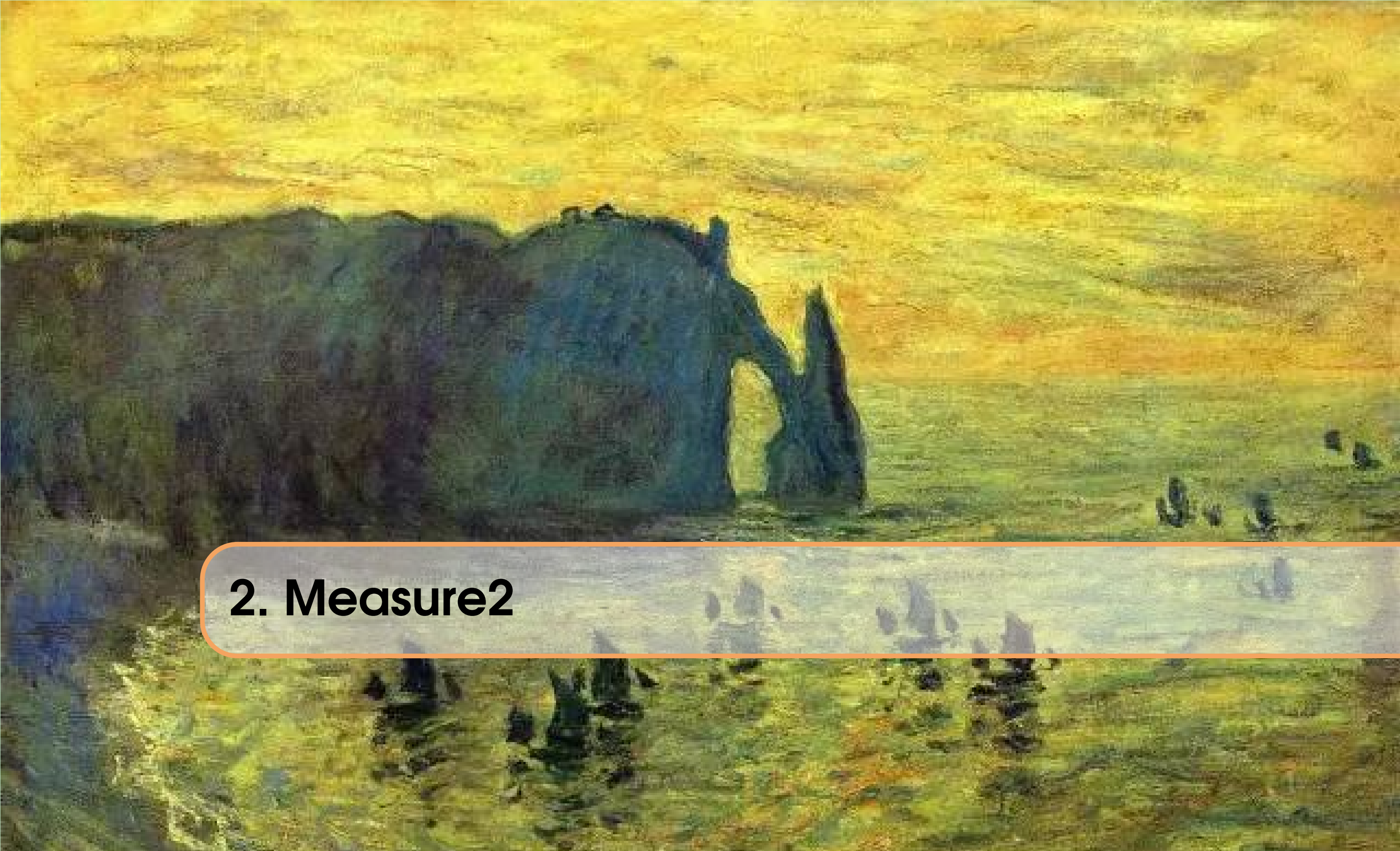
   Proof. Assume that the Radon-Nikodym theorem holds, that is, for some measurable function f one has

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

   for all Borel sets. Taking A to be a singleton set, $A = \{a\}$, and using the above equality, one finds

   $$0 = f(a)$$

   for $a \in R$. This implies that $f \equiv 0$, which further implies that the Lebesgue measure $\nu \equiv 0$, which is a contradiction. ■



2. Measure2

Definition 2.0.1 — Outer measure. Let X be a set, and let $\mathscr{P}(X)$ (which is a special σ-algebra) be the collection of all subsets of X. An outer measure on X is a function $\mu^* : \mathscr{P}(X) \to [0,+\infty]$ such that

1. $\mu^*(\varnothing) = 0$
2. if $A \subseteq B \subseteq X$, then $\mu^*(A) \leq \mu^*(B)$, and
3. if $\{A_n\}$ is an infinite sequence of subsets of X, then $\mu^*(\cup_n A_n) \leq \sum_n \mu^*(A_n)$

Thus an outer measure on X is a monotone and countably subadditive function from $\mathscr{P}(X)$ to $[0,+\infty]$ whose value at $\varnothing$ is 0.

R

1. A measure on X is an outer measure if and only if its domain $\mathcal{A}$ is $\mathscr{P}(X)$
2. An outer measure generally fails to be countably additive

Proposition 2.0.1 Countable union of countable sets is still countable set. (construct the one-to-one map to N.)

Definition 2.0.2 — Lebesgue outer measure. Lebesgue outer measure on $\mathbb{R}$, which we will denote by λ^*, is defined as follows. For each subset A of $\mathbb{R}$, let $\mathscr{C}_A$ be the set of all infinite sequences $\{(a_i,b_i)\}$ of bounded open intervals such that $A \subseteq \cup_i (a_i,b_i)$. Then $\lambda^* : \mathscr{P}(\mathbb{R}) \to [0,+\infty]$ is defined by

$$\lambda^*(A) = \inf\left\{\sum_i (b_i - a_i) : \{(a_i,b_i)\} \in \mathscr{C}_A\right\}$$

Theorem 2.0.2 Lebesgue outer measure on $\mathbb{R}$ is an outer measure, and it assigns to each subinterval of $\mathbb{R}$ its length.

Proof. We begin by verifying that λ^* is an outer measure. The the monotonicity is easliy to be verified. Now consider the countable subadditivity of λ^*. Let $\{A_n\}_{n=1}^{\infty}$ be an arbitrary sequence of subsets of $\mathbb{R}$. If $\sum_n \lambda^*(A_n) = +\infty$, then $\lambda^*(\cup_n A_n) \leq \sum_n \lambda^*(A_n)$ certainly holds. So suppose that $\sum_n \lambda^*(A_n) < +\infty$, and let ε be an arbitrary positive number. For each n choose a sequence $\{(a_{n,i}, b_{n,i})\}_{i=1}^{\infty}$ that covers A_n and satisfies

$$\sum_{i=1}^{\infty}(b_{n,i} - a_{n,i}) < \lambda^*(A_n) + \varepsilon/2^n$$

If we combine these sequences into one sequence $\{(a_j, b_j)\}$, then the combined sequence satisfies (countable union is still countable)

$$\cup_n A_n \subseteq \cup_j (a_j, b_j)$$

and

$$\sum_j (b_j - a_j) < \sum_n (\lambda^*(A_n) + \varepsilon/2^n) = \sum_n \lambda^*(A_n) + \varepsilon$$

These relations, together with the fact that ε is arbitrary, imply that $\lambda^*(\cup_n A_n) \leq \sum_n \lambda^*(A_n)$. Thus λ^* is an outer measure.

Now we compute the outer measure of the subintervals of $\mathbb{R}$. First consider a closed bounded interval $[a,b]$. It is easy to see that $\lambda^*([a,b]) \leq b - a$(cover $[a,b]$ with sequences of open intervals in which the first interval is barely larger than $[a,b]$, and the sum of the lengths of the other intervals is very small)(just this example could cover this interval, let alone the infimum). We turn to the reverse inequality. Let $\{(a_i, b_i)\}$ be a sequence of bounded open intervals whose union includes $[a,b]$. Since $[a,b]$ is compact, there is a positive integer n such that $[a,b] \subseteq \cup_{i=1}^{n}(a_i, b_i)$. It is easy to check that $b - a \leq \sum_{i=1}^{n}(b_i - a_i)$ (use induction on n) and hence that $b - a \leq \sum_{i=1}^{\infty}(b_i - a_i)$. Since $\{(a_i, b_i)\}$ was an arbitrary sequence whose union includes $[a,b]$, it follows that $b - a \leq \lambda^*([a,b])$. Thus $\lambda^*([a,b]) = b - a$. ■

R The outer measure of an arbitrary bounded interval is its length.

Definition 2.0.3 — Lebesgue outer measure on $\mathbb{R}^d$. Lebesgue outer measure on $\mathbb{R}^d$, which we will denote by λ^* (or, if necessary in order to avoid ambiguity, by λ_d^*) is defined as follows. A d -dimensional interval is a subset of $\mathbb{R}^d$ of the form $I_1 \times \cdots \times I_d$, where $I_1, \ldots, I_d$ are subintervals of $\mathbb{R}$ and $I_1 \times \cdots \times I_d$ is given by

$$I_1 \times \cdots \times I_d = \{(x_i, \ldots, x_d) : x_i \in I_i \text{ for } i = 1, \ldots, d\}$$

Note that the intervals $I_1, \ldots, I_d$, and hence the d -dimensional interval $I_1 \times \cdots \times I_d$ can be open, closed, or neither open nor closed. The volume of the d -dimensional interval $I_1 \times \cdots \times I_d$ is the product of the lengths of the intervals $I_1, \ldots, I_d$, and will be denoted by $\operatorname{vol}(I_1 \times \cdots \times I_d)$. For each subset A of $\mathbb{R}^d$ let $\mathscr{C}_A$ be the set of all sequences $\{R_i\}$ of bounded and open d -dimensional intervals for which $A \subseteq \cup_{i=1}^{\infty} R_i$ Then $\lambda^*(A)$, the outer measure of A, is the infimum of the set

$$\inf \left\{ \sum_{i=1}^{\infty} \operatorname{vol}(R_i) : \{R_i\} \in \mathscr{C}_A \right\}$$

Theorem 2.0.3 Lebesgue outer measure on $\mathbb{R}^d$ is an outer measure, and it assigns to each d -dimensional interval its volume.

Definition 2.0.4 Let X be a set, and let μ^* be an outer measure on X. A subset B of X is μ^*- measurable (or measurable with respect to μ^*) if

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$$

holds for every subset A of X. Thus a μ^* -measurable subset of X is one that divides each subset of X in such a way that the sizes (as measured by μ^*) of the pieces add properly. An equal formula to valid measurable set

$$\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c)$$

Thus the μ^* -measurability of B can be verified by checking previous formula holds for each A that satisfies $\mu^*(A) < +\infty$.

Proposition 2.0.4 Let X be a set, and let μ^* be an outer measure on X. Then each subset B of X that satisfies $\mu^*(B) = 0$ or that satisfies $\mu^*(B^c) = 0$ is μ^* -measurable.

Analysis 2.1 Just to check $\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c)$ ■

Corollary 2.0.5 The sets $\varnothing$ and X are measurable for every outer measure on X.

Theorem 2.0.6 Let X be a set, let μ^* be an outer measure on X, and let $\mathscr{M}_{\mu^*}$ be the collection of all μ^* -measurable subsets of X. Then

1. $\mathscr{M}_{\mu^*}$ is a σ -algebra, and
2. the restriction of μ^* to $\mathscr{M}_{\mu^*}$ is a measure on $\mathscr{M}_{\mu^*}$

Proof. $\mathscr{M}_{\mu^*}$ closed under complementation is obviously. Now suppose that B_1 and B_2 are μ^* -measurable subsets of X; we will show that $B_1 \cup B_2$ is μ^* measurable. For this, let A

be an arbitrary subset of X. The μ^* -measurability of B_1 implies

$$\begin{aligned}\mu^*\left(A\cap\left(B_1\cup B_2\right)\right)&=\mu^*\left(A\cap\left(B_1\cup B_2\right)\cap B_1\right)+\mu^*\left(A\cap\left(B_1\cup B_2\right)\cap B_1^c\right)\\&=\mu^*\left(A\cap B_1\right)+\mu^*\left(A\cap B_1^c\cap B_2\right)\end{aligned}$$

(Using the distribution law) we find

$$\begin{aligned}&\mu^*(A\cap\left(B_1\cup B_2\right))+\mu^*\left(A\cap\left(B_1\cup B_2\right)^c\right)\\&\quad=\mu^*\left(A\cap B_1\right)+\mu^*\left(A\cap B_1^c\cap B_2\right)+\mu^*\left(A\cap B_1^c\cap B_2^c\right)\\&\quad=\mu^*\left(A\cap B_1\right)+\mu^*\left(A\cap B_1^c\right)\\&\quad=\mu^*(A)\end{aligned}$$

Since A was an arbitrary subset of X, the set $B_1\cup B_2$ must be measurable. Thus $\mathscr{M}_{\mu^*}$ is an algebra.

Next suppose that $\{B_i\}$ is an infinite sequence of disjoint μ^* -measurable sets; we will show by induction that

$$\mu^*(A)=\sum_{i=1}^n\mu^*\left(A\cap B_i\right)+\mu^*\left(A\cap\left(\cap_{i=1}^nB_i^c\right)\right)$$

As to the induction step, note that the μ^* -measurability of B_{n+1} and the disjointness of the sequence $\{B_i\}$ imply that

$$\begin{aligned}&\mu^*\left(A\cap\left(\cap_{i=1}^nB_i^c\right)\right)\\&\quad=\mu^*\left(A\cap\left(\cap_{i=1}^nB_i^c\right)\cap B_{n+1}\right)+\mu^*\left(A\cap\left(\cap_{i=1}^nB_i^c\right)\cap B_{n+1}^c\right)\\&\quad=\mu^*\left(A\cap B_{n+1}\right)+\mu^*\left(A\cap\left(\cap_{i=1}^{n+1}B_i^c\right)\right)\end{aligned}$$

(The first equality is for B_{n+1} is disjoint with B_i, so B_{n+1} is a subset of $\left(\cap_{i=1}^nB_i^c\right)$) Replace $\mu^*\left(A\cap\left(\cap_{i=1}^nB_i^c\right)\right)$ with $\mu^*\left(A\cap\left(\cap_{i=1}^\infty B_i^c\right)\right)$, and thus with $\mu^*\left(A\cap\left(\cup_{i=1}^\infty B_i\right)^c\right)$; by letting the n in the sum in the resulting inequality approach infinity, we find

$$\mu^*(A)\geq\sum_{i=1}^\infty\mu^*\left(A\cap B_i\right)+\mu^*\left(A\cap\left(\cup_{i=1}^\infty B_i\right)^c\right)\tag{2.1}$$

This and the countable subadditivity of μ^* imply that

$$\begin{aligned}\mu^*(A)&\geq\sum_{i=1}^\infty\mu^*\left(A\cap B_i\right)+\mu^*\left(A\cap\left(\cup_{i=1}^\infty B_i\right)^c\right)\\&\geq\mu^*\left(A\cap\left(\cup_{i=1}^\infty B_i\right)\right)+\mu^*\left(A\cap\left(\cup_{i=1}^\infty B_i\right)^c\right)\\&\geq\mu^*(A)\end{aligned}$$

it follows that each inequality in the preceding calculation must in fact be an equality and hence that $\cup_{i=1}^\infty B_i$ is μ^* -measurable. Thus $\mathscr{M}_\mu*$ is closed under the formation of unions of disjoint sequences of sets. Since the union of an arbitrary sequence $\{B_i\}$ of sets in $\mathscr{M}_{\mu^*}$

is the union of a disjoint sequence of sets in $\mathscr{M}_{\mu^*}$, namely of the sequence (Skill: just need to prove under the disjoint case.)

$$B_1, B_1^c \cap B_2, \ldots, B_1^c \cap B_2^c \cap \cdots \cap B_{n-1}^c \cap B_n, \ldots$$

the algebra $\mathscr{M}_{\mu^*}$ is closed under the formation of countable unions. With this we have proved that $\mathscr{M}_{\mu^*}$ is a σ -algebra.

To show that the restriction of μ^* to $\mathscr{M}_{\mu^*}$ is a measure, we need to verify its countable additivity. If $\{B_i\}$ is a sequence of disjoint sets in $\mathscr{M}_{\mu^*}$, then replacing A with $\cup_{i=1}^{\infty} B_i$ in inequality (2.6) yields

$$\mu^*\left(\cup_{i=1}^{\infty} B_i\right) \geq \sum_{i=1}^{\infty} \mu^*\left(B_i\right) + 0$$

Since the reverse inequality is automatic, the countable additivity of the restriction of μ^* to $\mathscr{M}_{\mu^*}$ follows. ■

Theorem 2.0.7 Every Borel subset of $\mathbb{R}$ is Lebesgue measurable.

Proof. We begin by checking that every interval of the form $(-\infty, b]$ is Lebesgue measurable. Let B be such an interval. We need only check that

$$\lambda^*(A) \geq \lambda^*(A \cap B) + \lambda^*\left(A \cap B^c\right) \tag{2.2}$$

holds for each subset A of $\mathbb{R}$ for which $\lambda^*(A) < +\infty$. Let A be such a set, let ε be an arbitrary positive number, and let $\{(a_n, b_n)\}$ be a sequence of open intervals that covers A and satisfies $\sum_{n=1}^{\infty}(b_n - a_n) < \lambda^*(A) + \varepsilon$ (Called highly accurate open cover). Then for each n the sets $(a_n, b_n) \cap B$ and $(a_n, b_n) \cap B^c$ are disjoint intervals (one of which may instead be the empty set) whose union is (a_n, b_n), and so

$$b_n - a_n = \lambda^*\left((a_n, b_n)\right) = \lambda^*\left((a_n, b_n) \cap B\right) + \lambda^*\left((a_n, b_n) \cap B^c\right)$$

(see Proposition 2.0.2 Lebesgue outer measure sends borel set to its length) . Since the sequence $\{(a_n, b_n) \cap B\}$ covers $A \cap B$ and the sequence $\{(a_n, b_n) \cap B^c\}$ covers $A \cap B^c$, we have the countable subadditivity of λ^* that

$$\begin{aligned}
\lambda^*(A \cap B) + \lambda^*\left(A \cap B^c\right) &\leq \sum_n \lambda^*\left((a_n, b_n) \cap B\right) + \sum_n \lambda^*\left((a_n, b_n) \cap B^c\right) \\
&= \sum_n (b_n - a_n) < \lambda^*(A) + \varepsilon
\end{aligned}$$

However, ε was arbitrary, and so inequality (2.7) and the Lebesgue measurability of B follow.

Thus the collection $\mathscr{M}_{\lambda^*}$ of Lebesgue measurable sets is a σ -algebra on $\mathbb{R}$ (Theorem 2.0.6) that contains each interval of the form $(-\infty, b]$. However $\mathscr{B}(\mathbb{R})$ is the smallest σ -algebra on $\mathbb{R}$ that contains all these intervals and so $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{M}_{\lambda^*}$ ■

Proposition 2.0.8 Every Borel subset of $\mathbb{R}^d$ is Lebesgue measurable.

Definition 2.0.5 — Lebesgue measure. The restriction of Lebesgue outer measure on $\mathbb{R}$ (or on $\mathbb{R}^d$) to the collection $\mathscr{M}_\lambda$ * of Lebesgue measurable subsets of $\mathbb{R}$ (or of $\mathbb{R}^d$) is called **Lebesgue measure** and will be denoted by λ or by λ_d.

Theorem 2.0.9 Let μ be a finite measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$, and let $F_\mu : \mathbb{R} \to \mathbb{R}$ be defined by $F_\mu(x) = \mu((-\infty,x])$. Then F_μ is bounded, nondecreasing, and rightcontinuous, and satisfies $\lim_{x\to-\infty} F_\mu(x) = 0$

Proof. It follows from monotonic of measure that $0 \le \mu((-\infty,x]) \le \mu(\mathbb{R})$ holds for all x in $\mathbb{R}$ and that $\mu((-\infty,x]) \le \mu((-\infty,y])$ holds for all x and y in $\mathbb{R}$ such that $x \le y$; hence F_μ is bounded(finite measure) and nondecreasing.

Next suppose that $x \in \mathbb{R}$ and that $\{x_n\}$ is the sequence defined by $x_n = x + 1/n$. Then $(-\infty,x] = \cap_{n=1}^{\infty}(-\infty,x_n]$, and so Proposition 1.2 .5 (limit measure of decreasing sets) implies that $F_\mu(x) = \lim_n F_\mu(x_n)$. The right continuity of F_μ follows (note that if $x < y < x_n$, then, since F_μ is nondecreasing, $|F_\mu(y) - F_\mu(x)| \le |F_\mu(x_n) - F_\mu(x)|$). A similar argument shows that $\lim_{x\to-\infty} F_\mu(x) = 0$. ■

Proposition 2.0.10 — **measure μ is continuous if and only if the function F_μ is continuous .** Let μ and F_μ be as defined in condition. The interval $(a,b]$ is the difference of the intervals $(-\infty,b]$ and $(-\infty,a]$, and so Proposition 1.2 .2(the difference of two sets) implies that

$$\mu((a,b]) = F_\mu(b) - F_\mu(a) \tag{2.3}$$

Since F_μ is bounded and nondecreasing, the limit of $F_\mu(t)$ as t approaches x from the left exists for each x in $\mathbb{R}$; this limit is equal to $\sup\{F_\mu(t) : t < x\}$ and will be denoted by $F_\mu(x-)$. Now let $\{a_n\}$ be a sequence that increases to the real number b; if we apply Eq. (3.37) to each interval $(a_n,b]$ and then use Proposition 1.2.5 (monotonic of measure), we find that

$$\mu(\{b\}) = F_\mu(b) - F_\mu(b-)$$

Consequently F_μ is continuous at b if $\mu(\{b\}) = 0$, and is discontinuous there, with a jump of size $\mu(\{b\})$ in its graph, if $\mu(\{b\}) \ne 0$. Thus the measure μ is continuous (according to the definition of measure continuity) if and only if the function F_μ is continuous.

Previous equations allow one to use F_μ to recover the measure under μ of certain subsets of $\mathbb{R}$.

Theorem 2.0.11 — **Measure is determined by F_μ.** For each bounded, nondecreasing, and right-continuous function $F : \mathbb{R} \to \mathbb{R}$ that satisfies $\lim_{x\to-\infty} F(x) = 0$, there is a unique finite measure μ on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ such that $F(x) = \mu((-\infty,x])$ holds at each x in

$\mathbb{R}$

Proof. Let F be as in the statement of the proposition. We begin by constructing the required measure μ. Define a function $\mu^* : \mathscr{P}(\mathbb{R}) \to [0, +\infty]$ by letting $\mu^*(A)$ be the infimum of the set of sums $\sum_{n=1}^{\infty} (F(b_n) - F(a_n))$, where $\{(a_n, b_n]\}$ ranges over the set of sequences of half-open intervals that cover A, in the sense that $A \subseteq \cup_{n=1}^{\infty} (a_n, b_n]$. Then μ^* is an outer measure on $\mathbb{R}$; (check this by modifying some of the arguments used in the proof of Proposition 2.0.2)

Next we verify that $\mu^*((-\infty, x]) = F(x)$ holds for each x in $\mathbb{R}$. The inequality

$$\mu^*((-\infty, x]) \leq F(x)$$

holds, since $(-\infty, x]$ can be covered by the intervals in the sequence $\{(x-n, x-n+1]\}_{n=1}^{\infty}$, for which we have $\sum_{n=1}^{\infty}(F(x-n+1) - F(x-n)) = F(x)$. We turn to the reverse inequality. Let $\{(a_n, b_n]\}$ be a sequence that covers $(-\infty, x]$, and let ε be a positive number. Use the fact that $\lim_{t \to -\infty} F(t) = 0$ to choose a number t such that $t < x$ and $F(t) < \varepsilon$, and for each n use the right continuity of F to choose a positive number δ_n such that $F(b_n + \delta_n) < F(b_n) + \varepsilon/2^n$. Then the interval $[t, x]$ is compact, each interval $(a_n, b_n + \delta_n)$ is open, $[t, x] \subseteq \cup_{n=1}^{\infty} (a_n, b_n + \ \delta_n)$, and $\sum_n (F(b_n + \delta_n) - F(a_n)) \leq \sum_n (F(b_n) - F(a_n)) + \varepsilon$. The compactness of $[t, x]$ implies that there is a positive integer N such that $[t, x] \subseteq \cup_{n=1}^{N} (a_n, b_n + \delta_n)$. It follows that $(t, x]$ is the union of a finite collection of disjoint intervals $(c_j, d_j]$, each of which is included in some $(a_n, b_n + \delta_n]$. Consequently

$$F(x) - F(t) = \sum_j (F(d_j) - F(c_j)) \leq \sum_{n=1}^{\infty} (F(b_n + \delta_n) - F(a_n))$$

and so

$$F(x) - \varepsilon \leq \sum_{n=1}^{\infty} (F(b_n) - F(a_n)) + \varepsilon$$

Since ε and the sequence $\{(a_n, b_n]\}$ are arbitrary, the inequality $F(x) \leq \mu^*((-\infty, x])$ follows ($\{(a_n, b_n]\}$ be a sequence that covers $(-\infty, x]$). With this we have shown that $F(x) = \mu^*((-\infty, x])$.

Easily check that the proof of Proposition 2.0.7 can be modified so as to show that each interval $(-\infty, b]$ is μ^* -measurable and then that each Borel subset of $\mathbb{R}$ is μ^* -measurable.

Let μ be the restriction of μ^* to $\mathscr{B}(\mathbb{R})$. The preceding steps of our proof, together with Theorem 2.0.6, show that μ is a measure and that it satisfies $\mu((-\infty, x]) = F(x)$ at each x in $\mathbb{R}$. Since F is bounded, while $\mu(\mathbb{R}) = \lim_{n\to\infty} \mu((-\infty, n]) = \lim_{n\to\infty} F(n))$ (measure of monotonic sets), the measure μ is finite.

Finally we check the uniqueness of μ. Let μ be as constructed above, and let v be a possibly different measure such that $v((-\infty, x]) = F(x)$ holds for each x in $\mathbb{R}$. We first show that

$$v(A) \leq \mu(A) \tag{2.4}$$

is true for each Borel subset A of $\mathbb{R}$. To see this, note that if A is a Borel set and if $\{(a_n, b_n]\}$ is a sequence such that $A \subseteq \cup_n (a_n, b_n]$, then (according to (3.37) and subadditivity, applied to v)

$$v(A) \leq \sum_n v((a_n, b_n]) = \sum_n (F(b_n) - F(a_n)) \tag{2.5}$$

Since $\mu^*(A)$ was defined to be the infimum of the set of values that can occur as sums on the right side of (2.5), inequality (3.39) follows. If we apply inequality (3.39) to A and to A^c, we find

$$v(\mathbb{R}) = v(A) + v(A^c) \leq \mu(A) + \mu(A^c) = \mu(\mathbb{R})$$

Since $v(\mathbb{R}) = \mu(\mathbb{R}) < +\infty$, it follows that $v(A)$ and $v(A^c)$ are equal to $\mu(A)$ and $\mu(A^c)$, respectively. With this the proof that $v = \mu$ is complete. ■

2.1 Lebesgue measure

Theorem 2.1.1 — Lebesgue measure is regular. Let A be a Lebesgue measurable subset of $\mathbb{R}^d$. Then

1. $\lambda(A) = \inf\{\lambda(U) : U \text{ is open and } A \subseteq U\}$, and
2. $\lambda(A) = \sup\{\lambda(K) : K \text{ is compact and } K \subseteq A\}$

Proof. Note that the monotonicity of λ implies that $\lambda(A) \leq \inf\{\lambda(U) : U \text{ is open and } A \subseteq U\}$ and $\lambda(A) \geq \sup\{\lambda(K) : K \text{ is compact and } K \subseteq A\}$. Hence we need only prove the reverse inequalities. We begin with part (a).

Since the required equality clearly holds if $\lambda(A) = +\infty$, we can assume that $\lambda(A) < +\infty$. Let ε be an arbitrary positive number. Then according to the definition of Lebesgue measure, there is a sequence $\{R_i\}$ of open d -dimensional intervals such that $A \subseteq \cup_i R_i$ and

$$\sum_i \operatorname{vol}(R_i) < \lambda(A) + \varepsilon$$

Let U be the union of these intervals. Then U is open, $A \subseteq U$, and (subadditivity and definition of d-dim lebgesue measure)

$$\lambda(U) \leq \sum_i \lambda(R_i) = \sum_i \operatorname{vol}(R_i) < \lambda(A) + \varepsilon$$

Since ε is arbitrary, part (a) is proved.

We turn to part (b) and deal first with the case where A is bounded. Let C be a closed and bounded set that includes A, and let ε be an arbitrary positive number. Use part (a) to choose an open set U that includes $C-A$ and satisfies

$$\lambda(U)<\lambda(C-A)+\varepsilon=\lambda C-\lambda A-\varepsilon$$

Let $K=C-U$. Then K is a closed and bounded (and hence compact) subset of A; furthermore, $C\subseteq K\cup U$ and so

$$\lambda(C)\leq\lambda(K)+\lambda(U)$$

Last two inequalities imply that $\lambda(A)-\varepsilon<\lambda(K)$. Since ε was arbitrary, part (b) is proved in the case where A is bounded.

Finally, consider the case where A is not bounded. Suppose that b is a real number less than $\lambda(A)$; we will produce a compact subset K of A such that $b<\lambda(K)$. Let $\{A_j\}$ be an increasing sequence of bounded measurable subsets of A such that $A=\cup_j A_j$ (for example, we might let A_j be the intersection of A with the closed ball of radius j about the origin). Measure of monotonic sets implies that $\lambda(A)=\lim_j\lambda\left(A_j\right)$, and so we can choose j_0 such that $\lambda\left(A_{j_0}\right)>b$. Now apply to A_{j_0} (which is bounded) the weakened form of part (b) that was proved in the preceding paragraph; this gives a compact subset K of A_{j_0} (and hence of A) such that $\lambda(K)>b$. Since b was an arbitrary number less than $\lambda(A)$, the proof is complete. ■

Theorem 2.1.2 Each open subset of $\mathbb{R}^d$ is the union of a countable disjoint collection of half-open cubes, each of which is of the form given in expression

$$\left\{(x_1,\ldots,x_d):j_i2^{-k}\leq x_i<(j_i+1)2^{-k}\text{ for }i=1,\ldots,d\right\}\tag{2.6}$$

for some integers $j_1,\ldots,j_d$ and some positive integer k.

Proof. For each positive integer k, let $\mathscr{C}_k$ be the collection of all cubes of the form

$$\left\{(x_1,\ldots,x_d):j_i2^{-k}\leq x_i<(j_i+1)2^{-k}\text{ for }i=1,\ldots,d\right\}$$

where $j_1,\ldots,j_d$ are arbitrary integers. It is easy to see that

1. each $\mathscr{C}_k$ is a countable partition of $\mathbb{R}^d$, and
2. if $k_1<k_2$, then each cube in $\mathscr{C}_{k_2}$ is included in some cube in $\mathscr{C}_{k_1}$

Suppose that U is an open subset of $\mathbb{R}^d$. We construct a collection $\mathscr{D}$ of cubes inductively by letting $\mathscr{D}$ be empty at the start, and then at step $k($ for $k=1,2,\ldots)$ adding to $\mathscr{D}$ those cubes in $\mathscr{C}_k$ that are included in U but are disjoint from all the cubes put into $\mathscr{D}$ at earlier steps.(Image the seven-piece puzzle.) It is clear that $\mathscr{D}$ is a countable disjoint collection of cubes whose union is included in U. It remains only to check that its union includes U.

Let x be a member of U. Since U is open, the cube in $\mathscr{C}_k$ that contains x is included in U if k is sufficiently large. Let k_0 be the smallest such k. Then the cube in $\mathscr{C}_{k_0}$ that contains x belongs to $\mathscr{D}$, and so x belongs to the union of the cubes in $\mathscr{D}$. ■

Theorem 2.1.3 Lebesgue measure is the only measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ that assigns to each d -dimensional interval, or even to each half-open cube of the form given in expression (2.6), its volume.

Proof. We need only assume that μ is a measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ that assigns to each cube of the form given in expression (2.6) its volume and prove that $\mu = \lambda$. First suppose that U is an open subset of $\mathbb{R}^d$. Then according to Lemma 2.1.2, there is a disjoint sequence $\{C_j\}$ of half-open cubes that have the form given in expression (2.6) and whose union is U, and so

$$\mu(U) = \sum_j \mu\left(C_j\right) = \sum_j \lambda\left(C_j\right) = \lambda(U)$$

hence μ and λ agree on the open subsets of $\mathbb{R}^d$. Next suppose that A is an arbitrary Borel subset of $\mathbb{R}^d$. If U is an open subset of $\mathbb{R}^d$ that includes A, then $\mu(A) \leq \mu(U) = \lambda(U)$; it follows that $\mu(A) \leq \inf\{\lambda(U) : U \text{ is open and } A \subseteq U\}$. The regularity of λ (Proposition 1.4.1) now implies that

$$\mu(A) \leq \lambda(A) \tag{2.7}$$

We need to show that this inequality can be replaced with an equality. First suppose that A is a bounded Borel subset of $\mathbb{R}^d$ and that V is a bounded open set that includes A. Then inequality (2.7), applied to the sets A and $V - A$, implies that

$$\mu(V) = \mu(A) + \mu(V - A) \leq \lambda(A) + \lambda(V - A) = \lambda(V)$$

Since the extreme members of this inequality are equal(two measures are agree on open sets), and since $\mu(A)$ and $\mu(V - A)$ are no larger than $\lambda(A)$ and $\lambda(V - A)$, respectively, it follows that $\mu(A)$ and $\lambda(A)$ are equal. Finally, an arbitrary Borel subset A of $\mathbb{R}^d$ is the union of a sequence of disjoint bounded Borel sets and so must satisfy $\mu(A) = \lambda(A)$ ■

Definition 2.1.1 — Translate of set. For each element x and subset A of $\mathbb{R}^d$ we will denote by $A + x$ the subset of $\mathbb{R}^d$ defined by

$$A + x = \left\{y \in \mathbb{R}^d : y = a + x \text{ for some } a \text{ in } A\right\}$$

the set $A + x$ is called the translate of A by x.

Theorem 2.1.4 Lebesgue outer measure on $\mathbb{R}^d$ is translation invariant, in the sense that if $x \in \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$, then $\lambda^*(A) = \lambda^*(A+x)$. Furthermore, a subset B of $\mathbb{R}^d$ is Lebesgue measurable if and only if $B+x$ is Lebesgue measurable.

Proof. The equality of $\lambda^*(A)$ and $\lambda^*(A+x)$ follows from the definition of λ^* and the fact that the volume of a d -dimensional interval is invariant under translation. The second assertion follows from the first, together with the definition of a Lebesgue measurable set-note that a set B satisfies

$$\lambda^*(A-x) = \lambda^*((A-x)\cap B) + \lambda^*\left((A-x)\cap B^c\right)$$

for all sets $A-x$ if and only if $B+x$ satisfies

$$\lambda^*(A) = \lambda^*(A\cap(B+x)) + \lambda^*\left(A\cap(B+x)^c\right)$$

for all sets A. ∎

Lebesgue measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ is characterized up to constant multiples by the following result; see Chap. 9 for analogous results that hold in more general situations.

Theorem 2.1.5 Let μ be a nonzero measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ that is finite on the bounded Borel subsets of $\mathbb{R}^d$ and is translation invariant, in the sense that $\mu(A) = \mu(A+x)$ holds for each A in $\mathscr{B}\left(\mathbb{R}^d\right)$ and each x in $\mathbb{R}^d$. Then there is a positive number c such that $\mu(A) = c\lambda(A)$ holds for each A in $\mathscr{B}\left(\mathbb{R}^d\right)$.

Note that for the concept of translation invariance for measures on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ to make sense, the Borel σ -algebra on $\mathbb{R}^d$ must be translation invariant, in the sense that if $A \in \mathscr{B}\left(\mathbb{R}^d\right)$ and $x \in \mathbb{R}^d$, then $A+x \in \mathscr{B}\left(\mathbb{R}^d\right)$. To check this translation invariance of $\mathscr{B}\left(\mathbb{R}^d\right)$, note that $\left\{A \subseteq \mathbb{R}^d : A+x \in \mathscr{B}\left(\mathbb{R}^d\right)\right\}$ is a σ -algebra that contains the open sets and hence includes $\mathscr{B}\left(\mathbb{R}^d\right)$.

Proof. Let $C = \{(x_1, \ldots, x_d) : 0 \le x_i < 1 \text{ for each } i\}$, and let $c = \mu(C)$. Then c is finite (since μ is finite on the bounded Borel sets) and positive (if it were 0, then $\mathbb{R}^d$, as the union of a sequence of translates of C, would have measure zero underμ, $\mu(\mathbb{R}) = 0)$. Define a measure v on $\mathscr{B}\left(\mathbb{R}^d\right)$ by letting $v(A) = (1/c)\mu(A)$ hold for each A in $\mathscr{B}\left(\mathbb{R}^d\right)$. Then v is translation invariant, and it assigns to the set C defined above its Lebesgue measure, namely 1. If D is a half-open cube that has the form given in expression (2.6) and whose edges have length 2^{-k}, then C is the union of 2^{dk} translates of D, and so

$$1/2^{-dk}v(D) = v(C) = \lambda(C) = 2^{dk}\lambda(D)$$

thus v and λ agree on all such cubes. Proposition 2.1.3 now implies that $v = \lambda$ and hence that $\mu = c\lambda$. ∎

Example 1.4.6 (The Cantor Set). We should note a few facts about the Cantor set, a set which turns out to be a useful source of examples. Recall that it is defined as follows. Let K_0 be the interval $[0,1]$. Form K_1 by removing from K_0 the interval $(1/3,2/3)$. Thus $K_1 = [0,1/3] \cup [2/3,1]$. Continue this procedure, forming K_n by removing from K_{n-1} the open middle third of each of the intervals making up K_{n-1}. Thus K_n is the union of 2^n disjoint closed intervals, each of length $(1/3)^n$ The Cantor set (which we will temporarily denote by K) is the set of points that remain; thus $K = \cap_n K_n$

Of course K is closed and bounded. Furthermore, K has no interior points, since an open interval included in K would for each n be included in one of the intervals making up K_n and so would have length at most $(1/3)^n$. The cardinality of K is that of the continuum: it is easy to check that the map that assigns to a sequence $\{z_n\}$ of 0 's and 1 's the number $\sum_{n=1}^{\infty} 2z_n/3^n$ is a bijection of the set of all such sequences onto K; hence the cardinality of K is that of the set of all sequences of 0 's and 1 's and so that of the continuum (see Appendix A).

Proposition 1.4.7. The Cantor set is a compact set that has the cardinality of the continuum but has Lebesgue measure zero.

Proof. We have already noted that the Cantor set (again call it K) is compact and has the cardinality of the continuum. To compute the measure of K, note that for each n it is included in the set K_n constructed above and that $\lambda(K_n) = (2/3)^n$. Thus $\lambda(K) \leq (2/3)^n$ holds for each n, and so $\lambda(K)$ must be zero. (For an alternative proof, check that the sum of the measures of the intervals removed from [0,1] during the construction of K is the sum of the geometric series

$$\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} + \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} + \left(\frac{2}{3}\right)^3 \cdot \frac{1}{3} + \ldots$$

and so is 1.) Example 1.4 .8 (A Nonmeasurable Set). We now return to one of the promises made in Sect. 1.3 and prove that there is a subset of $\mathbb{R}$ that is not Lebesgue measurable. Note that our proof of this uses the axiom of choice. [4] Whether the use of this axiom is essential was an open question until the mid- 1960 s, when R.M. Solovay showed that if a certain consistency assumption holds, then the existence of a subset of $\mathbb{R}$ that is not Lebesgue measurable cannot be proved from the axioms of Zermelo-Frankel set theory without the use of the axiom of choice.

Theorem 1.4.9. There is a subset of $\mathbb{R}$, and in fact of the interval $(0,1)$, that is not Lebesgue measurable. Proof. Define a relation $\sim$ on $\mathbb{R}$ by letting $x \sim y$ hold if and only if $x - y$ is rational. It is easy to check that $\sim$ is an equivalence relation: it is reflexive ($x \sim x$ holds for each x), symmetric ($x \sim y$ implies $y \sim x$), and transitive ($x \sim y$ and $y \sim z$ imply $x \sim z$). Note that each equivalence class under $\sim$ has the form $\mathbb{Q} + x$ for some x and so is dense in $\mathbb{R}$. since these equivalence classes are disjoint, and since each intersects the interval $(0,1)$, we can use the axiom of choice to form a subset E of (0,1) that contains exactly one element from each equivalence class. We will prove that the set E is not Lebesgue measurable.

Let $\{r_n\}$ be an enumeration of the rational numbers in the interval $(-1,1)$, and for each n let $E_n = E + r_n$. We will check that (a) the sets E_n are disjoint, (b) $\cup_n E_n$ is included in the interval $(-1,2)$, and (c) the interval (0,1) is included in $\cup_n E_n$ To check (a), note that if $E_m \cap E_n \neq \varnothing$, then there are elements e and e' of E such that $e + r_m = e' + r_n$; it follows that $e \sim e'$ and hence that $e = e'$ and $m = n$. Thus (a) is proved. Assertion (b) follows from the inclusion $E \subseteq (0,1)$ and the fact that each term of the sequence $\{r_n\}$ belongs to $(-1,1)$. Now consider assertion (c). Let x bean arbitrary member of $(0,1)$, and let e be the member of E that satisfies $x \sim e$. Then $x - e$ is rational and belongs to (-1,1) (recall that both x and e belong to (0,1)) and so has the form r_n for some n. Hence $x \in E_n$, and assertion (c) is proved.

Suppose that the set E is Lebesgue measurable. Then for each n the set E_n is measurable (Proposition 1.4.4), and so property (a) above implies that

$$\lambda\left(\cup_n E_n\right) = \sum_n \lambda\left(E_n\right)$$

furthermore, the translation invariance of λ implies that $\lambda\left(E_n\right) = \lambda(E)$ holds for each n. Hence if $\lambda(E) = 0$, then $\lambda\left(\cup_n E_n\right) = 0$, contradicting assertion (c) above, while if $\lambda(E) \neq 0$, then $\lambda\left(\cup_n E_n\right) = +\infty$, contradicting assertion (b). Thus the assumption that E is measurable leads to a contradiction, and the proof is complete.

Definition 2.1.2 Let A be a subset of $\mathbb{R}$. Then $\operatorname{diff}(A)$ is the subset of $\mathbb{R}$ defined by

$$\operatorname{diff}(A) = \{x - y : x \in A \text{ and } y \in A\}$$

The following fact about such sets is occasionally useful.

Proposition 2.1.6 Let A be a Lebesgue measurable subset of $\mathbb{R}$ such that $\lambda(A) > 0$. Then $\operatorname{diff}(A)$ includes an open interval that contains 0.

Proof. According to Proposition 1.4.1, there is a compact subset K of A such that $\lambda(K) > 0$. since $\operatorname{diff}(K)$ is then included in $\operatorname{diff}(A)$, it is enough to prove that $\operatorname{diff}(K)$ includes an open interval that contains 0. Note that a real number x belongs to $\operatorname{diff}(K)$ if and only if K intersects $x + K$; thus it suffices to prove that if $|x|$ is sufficiently small, then K intersects $x + K$

Use Proposition 1.4 .1 to choose an open set U such that $K \subseteq U$ and $\lambda(U) < 2\lambda(K)$. The distances between the points in K and the points outside U are bounded away from 0 (since the distance from a point x of U to the complement of U is a continuous strictly positive function of x and so has a positive minimum on the compact set K; see D.27 and D.18). Thus there is a positive number ε such that if $|x| < \varepsilon$, then $x + K$ is included in U. Suppose that $|x| < \varepsilon$. If $x + K$ were disjoint from K, then it would follow from the translation invariance of λ and the relation $x+K \subseteq U$ that

$$2\lambda(K)=\lambda(K)+\lambda(x+K)=\lambda(K\cup(x+K))\leq\lambda(U)$$

However this contradicts the inequality $\lambda(U)<2\lambda(K)$, and so K and $x+K$ cannot be disjoint. Therefore, $x\in\operatorname{diff}(K)$. Consequently the interval $(-\varepsilon,\varepsilon)$ is included in $\operatorname{diff}(K)$, and thus in $\operatorname{diff}(A)$ We can use Proposition 1.4.10, plus a modification of the proof of Theorem 1.4.9, to prove the following rather strong result (see the remark at the end of this section and the one following the proof of Proposition 1.5 .4).

Proposition 2.1.7 There is a subset A of $\mathbb{R}$ such that each Lebesgue measurable set that is included in A or in A^c has Lebesgue measure zero.

Proof. Define subsets G, G_0, and G_1 of $\mathbb{R}$ by

$$G=\{x: x=r+n\sqrt{2} \text{ for some } r \text{ in } \mathbb{Q} \text{ and } n \text{ in } \mathbb{Z}\}$$
$$G_0=\{x: x=r+2n\sqrt{2} \text{ for some } r \text{ in } \mathbb{Q} \text{ and } n \text{ in } \mathbb{Z}\}, \text{ and}$$
$$G_1=\{x: x=r+(2n+1)\sqrt{2} \text{ for some } r \text{ in } \mathbb{Q} \text{ and } n \text{ in } \mathbb{Z}\}$$

It is easy to see that G and G_0 are subgroups of $\mathbb{R}$ (under addition), that G_0 and G_1 are disjoint, that $G_1=G_0+\sqrt{2}$, and that $G=G_0\cup G_1$. Define a relation $\sim$ on $\mathbb{R}$ by letting $x\sim y$ hold when $x-y\in G$; the relation $\sim$ is then an equivalence relation on $\mathbb{R}$. Use the axiom of choice to form a subset E of $\mathbb{R}$ that contains exactly one representative of each equivalence class of $\sim$. Let $A=E+G_0$ (that is, let A consist of the points that have the form $e+g_0$ for some e in E and some g_0 in G_0).

We now show that there does not exist a Lebesgue measurable subset B of A such that $\lambda(B)>0$. For this let us assume that such a set exists; we will derive a contradiction. Proposition 1.4 .10 implies that there is an interval $(-\varepsilon,\varepsilon)$ that is included in $\operatorname{diff}(B)$ and hence in $\operatorname{diff}(A)$. since G_1 is dense in $\mathbb{R}$, it meets the interval $(-\varepsilon,\varepsilon)$ and hence meets diff (A). This, however, is impossible, since each element of $\operatorname{diff}(A)$ is of the form $e_1-e_2+g_0$ (where e_1 and e_2 belong to E and g_0 belongs to G_0) and so cannot belong to G_1 (the relation $e_1-e_2+g_0=g_1$ would imply that $e_1=e_2$ and $g_0=g_1$, contradicting the disjointness of G_0 and G_1). This completes our proof that every Lebesgue measurable subset of A must have Lebesgue measure zero.

It is easy to check that $A^c=E+G_1$ and hence that $A^c=A+\sqrt{2}$. It follows that each Lebesgue measurable subset of A^c is of the form $B+\sqrt{2}$ for some Lebesgue measurable subset B of A. since A has no Lebesgue measurable subsets of positive measure, it follows that A^c also has no such subsets, and with this the proof is complete.

Note that the set A of Proposition 1.4 .11 is not Lebesgue measurable: if it were, then both A and A^c would include (in fact, would be) Lebesgue measurable sets of positive Lebesgue measure. Thus we could have presented Theorem 1.4 .9 as a corollary of Proposition 1.4.11. (Of course, the proof of Theorem 1.4 .9 presented earlier is simpler than the proofs of Propositions 1.4 .10 and 1.4 .11 taken together and is in fact a classical and well-known argument; hence it was included.)

2.1.1 Completeness and Regularity

Let $(X, \mathscr{A}, \mu)$ be a measure space. The measure μ (or the measure space $(X, \mathscr{A}, \mu)$) is complete if the relations $A \in \mathscr{A}, \mu(A)=0$, and $B \subseteq A$ together imply that $B \in \mathscr{A}$. It is sometimes convenient to call a subset B of $X \mu$ -negligible (or μ -null) if there is a subset A of X such that $A \in \mathscr{A}, B \subseteq A$, and $\mu(A)=0$. Thus the measure μ is complete if and only if every μ -negligible subset of X belongs to $\mathscr{A}$.

It follows from Proposition 1.3 .5 that if μ^* is an outer measure on the set X and if $\mathscr{M}_{\mu^*}$ is the σ -algebra of all μ^* -measurable subsets of X, then the restriction of μ^* to $\mathscr{M}_{\mu^*}$ is complete. In particular, Lebesgue measure on the σ -algebra of Lebesgue measurable subsets of $\mathbb{R}^d$ is complete. On the other hand, as we will soon see, the restriction of Lebesgue measure to the σ -algebra of Borel subsets of $\mathbb{R}$ is not complete.

It is sometimes convenient to be able to deal with arbitrary subsets of sets of measure zero, and at such times complete measures are desirable. In many such situations the following construction proves useful.

Let $(X, \mathscr{A})$ be a measurable space, and let μ be a measure on $\mathscr{A}$. The completion of $\mathscr{A}$ under μ is the collection $\mathscr{A}_\mu$ of subsets A of X for which there are sets E and F in $\mathscr{A}$ such that

$$E \subseteq A \subseteq F \tag{2.8}$$

and

$$\mu(F-E)=0 \tag{2.9}$$

A set that belongs to $\mathscr{A}_\mu$ is sometimes said to be μ -measurable.

Suppose that A, E, and F are as in the preceding paragraph. It follows immediately that $\mu(E)=\mu(F)$. Furthermore, if B is a subset of A that belongs to $\mathscr{A}$, then

$$\mu(B) \leq \mu(F)=\mu(E)$$

Hence

$$\mu(E)=\sup \{\mu(B): B \in \mathscr{A} \text{ and } B \subseteq A\}$$

and so the common value of $\mu(E)$ and $\mu(F)$ depends only on the set A (and the measure μ), and not on the choice of sets E and F satisfying (1) and (2). Thus we can define a function $\bar{\mu}: \mathscr{A}_\mu \rightarrow[0,+\infty]$ by letting $\bar{\mu}(A)$ be the common value of $\mu(E)$ and $\mu(F)$, where E and F belong to $\mathscr{A}$ and satisfy (1) and (2). This function $\bar{\mu}$ is called the completion of μ

Proposition 2.1.8 Let $(X, \mathscr{A})$ be a measurable space, and let μ be a measure on A. Then $\mathscr{A}_\mu$ is a σ -algebra on X that includes $\mathscr{A}$, and $\bar{\mu}$ is a measure on $\mathscr{A}_\mu$ that is complete and whose restriction to $\mathscr{A}$ is μ.

Proof. Proof. It is clear that $\mathscr{A}_\mu$ includes $\mathscr{A}$ (for A in $\mathscr{A}$ let the sets E and F in (2.8) and (2.9) equal A), and in particular that $X \in \mathscr{A}_\mu$. Note that the relations $E \subseteq A \subseteq F$ and $\mu(F-E) = 0$ imply the relations $F^c \subseteq A^c \subseteq E^c$ and $\mu(E^c - F^c) = 0$; thus $\mathscr{A}_\mu$ is closed under complementation. Next suppose that $\{A_n\}$ is a sequence of sets in $\mathscr{A}\mu$. For each n choose sets E_n and F_n in $\mathscr{A}$ such that $E_n \subseteq A_n \subseteq F_n$ and $\mu(F_n - E_n) = 0$ Then $\cup_n E_n$ and $\cup_n F_n$ belong to $\mathscr{A}$ and satisfy $\cup_n E_n \subseteq \cup_n A_n \subseteq \cup_n F_n$ and

$$\mu(\cup_n F_n - \cup_n E_n) \leq \mu(\cup_n (F_n - E_n)) \leq \sum_n \mu(F_n - E_n) = 0$$

thus $\cup A_n$ belongs to $\mathscr{A}_\mu$. This completes the proof that $\mathscr{A}_\mu$ is a σ -algebra on X that includes $\mathscr{A}$.

Now consider the function $\bar{\mu}$. It is an extension of μ, since for A in $\mathscr{A}$ we can again let E and F equal A. It is clear that $\bar{\mu}$ has nonnegative values and satisfies $\bar{\mu}(\varnothing) = 0$, and so we need only check its countable additivity. Let $\{A_n\}$ be a sequence of disjoint sets in $\mathscr{A}_\mu$, and for each n again choose sets E_n and F_n in $\mathscr{A}$ that satisfy $E_n \subseteq A_n \subseteq F_n$ and $\mu(F_n - E_n) = 0$. The disjointness of the sets A_n implies the disjointness of the sets E_n, and so we can conclude that

$$\bar{\mu}(\cup_n A_n) = \mu(\cup_n E_n) = \sum_n \mu(E_n) = \sum_n \bar{\mu}(A_n)$$

Thus $\bar{\mu}$ is a measure. It is easy to check that $\bar{\mu}$ is complete. ■

Proposition 2.1.9 Lebesgue measure on $(\mathbb{R}^d, \mathscr{M}_{\lambda^*})$ is the completion of Lebesgue measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$

Lemma 2.1 Let A be a Lebesgue measurable subset of $\mathbb{R}^d$. Then there exist Borel subsets E and F of $\mathbb{R}^d$ such that $E \subseteq A \subseteq F$ and $\lambda(F-E) = 0$

Proof. First suppose that A is a Lebesgue measurable subset of $\mathbb{R}^d$ such that $\lambda(A) < +\infty$. For each positive integer n, use Proposition 2.1.1 to choose a compact set K_n such that $K_n \subseteq A$ and $\lambda(A) - 1/n < \lambda(K_n)$ and an open set U_n such that $A \subseteq U_n$ and $\lambda(U_n) < \lambda(A) + 1/n$. Let $E = \cup_n K_n$ and $F = \cap_n U_n$. Then E and F belong to $\mathscr{B}(\mathbb{R}^d)$ and satisfy $E \subseteq A \subseteq F$. The relation

$$\lambda(F-E) \leq \lambda(U_n - K_n) = \lambda(U_n - A) + \lambda(A - K_n) < 2/n$$

holds for each n, and so $\lambda(F-E) = 0$. Thus the lemma is proved in the case where $\lambda(A) < +\infty$ If A is an arbitrary Lebesgue measurable subset of $\mathbb{R}^d$, then A is the union of a sequence $\{A_n\}$ of Lebesgue measurable sets of finite Lebesgue measure. For each n we can choose Borel sets E_n and F_n such that $E_n \subseteq A_n \subseteq F_n$ and $\lambda(F_n - E_n) = 0$ The sets E and F defined by $E = \cup_n E_n$ and $F = \cup_n F_n$ then satisfy $E \subseteq A \subseteq F$ and $\lambda(F - E) = 0$ (note that $F - E \subseteq \cup_n (F_n - E_n)$)

■

Proof. Proof of Proposition 2.1.9. Let λ be Lebesgue measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, let $\bar{\lambda}$ be the completion of λ, and let λ_m be Lebesgue measure on $(\mathbb{R}^d, \mathscr{M}_{\lambda^*})$. Lemma 1.5 .3 implies that $\mathscr{M}_{\lambda^*}$ is included in the completion of $\mathscr{B}(\mathbb{R}^d)$ under λ and that λ_m is the restriction of $\bar{\lambda}$ to $\mathscr{M}_{\lambda^*}$. Thus we need only check that each set A that belongs to the completion of $\mathscr{B}(\mathbb{R}^d)$ under λ is Lebesgue measurable. For such a set A there exist Borel sets E and F such that $E \subseteq A \subseteq F$ and $\lambda(F - E) = 0$. since $A - E \subseteq F - E$ and $\lambda_m(F - E) = \lambda(F - E) = 0$, the completeness of Lebesgue measure on $\mathscr{M}_{\lambda^*}$ implies that $A - E \in \mathscr{M}_{\lambda^*}$. Thus A, since it is the union of $A - E$ and E, must belong to $\mathscr{M}_{\lambda^*}$ ■

Question 2.1
1. there are Lebesgue measurable subsets of $\mathbb{R}$ that are not Borel sets, and
2. the restriction of Lebesgue measure to $\mathscr{B}(\mathbb{R})$ is not complete.

It should be noted that although replacing a measure space $(X, \mathscr{A}, \mu)$ with its completion $(X, \mathscr{A}_\mu, \bar{\mu})$ enables one to avoid some difficulties, it introduces others. Some difficulties arise because the completed σ -algebra $\mathscr{A}_\mu$ is often more complicated than the original σ -algebra $\mathscr{A}$. Others are caused by the fact that for measures μ and v defined on a common σ -algebra $\mathscr{A}$, the completions $\mathscr{A}_\mu$ and $\mathscr{A}_v$ of $\mathscr{A}$ under μ and v may not be equal (see Exercise 3). Because of these complications it seems wise whenever possible to avoid arguments that depend on completeness; it turns out that in the basic parts of measure theory this can almost always be done.

Let $(X, \mathscr{A})$ be a measurable space, let μ be a measure on $\mathscr{A}$, and let A be an arbitrary subset of X. Then $\mu^*(A)$, the outer measure of A, is defined by

$$\mu^*(A) = \inf\{\mu(B) : A \subseteq B \text{ and } B \in \mathscr{A}\}$$

and $\mu_*(A)$, the inner measure of A, is defined by

$$\mu_*(A) = \sup\{\mu(B) : B \subseteq A \text{ and } B \in \mathscr{A}\}$$

It is easy to check that $\mu_*(A) \leq \mu^*(A)$ holds for each subset A of X.

Proposition 2.1.10 Let $(X, \mathscr{A})$ be a measurable space, and let μ be a measure on $(X, \mathscr{A})$. Then the function $\mu^* : \mathscr{P}(X) \to [0, +\infty]$ defined by Eq.(3) is an outer measure (as defined in Sect. 1.3) on X.

Proof. Certainly μ^* satisfies $\mu^*(\varnothing)=0$ and is monotone. We turn to its subadditivity. Let $\{A_n\}$ be a sequence of subsets of X. The inequality $\mu^*(\cup_n A_n) \leq \sum_n \mu^*(A_n)$ is clear if $\sum_n \mu^*(A_n)=+\infty$. So suppose that $\sum_n \mu^*(A_n)<+\infty$. Let ε be an arbitrary positive number, and for each n choose a set B_n that belongs to $\mathscr{A}$, includes A_n, and satisfies $\mu(B_n) \leq \mu^*(A_n)+\varepsilon/2^n$. Then the set B defined by $B=\cup_n B_n$ belongs to $\mathscr{A}$, includes $\cup_n A_n$, and satisfies $\mu(B) \leq \sum_n \mu^*(A_n)+\varepsilon$ (see Proposition 1.2.4); thus $\mu^*(\cup_n A_n) \leq \sum_n \mu^*(A_n)+$ ε. since ε is arbitrary, the proof is complete. ■

Note that Proposition 1.4 .11 can now be rephrased: there is a subset A of $\mathbb{R}$ such that $\lambda_*(A)=0$ and $\lambda_*(A^c)=0$ Proposition 1.5.5. Let $(X,\mathscr{A})$ be a measurable space, let μ be a measure on $\mathscr{A}$, and let A be a subset of X such that $\mu^*(A)<+\infty$. Then A belongs to $\mathscr{A}_\mu$ if and only if $\mu_*(A)=\mu^*(A)$ Proof. If A belongs to $\mathscr{A}_\mu$, then there are sets E and F that belong to $\mathscr{A}$ and satisfy $E \subseteq A \subseteq F$ and $\mu(F-E)=0$. Then

$$\mu(E) \leq \mu_*(A) \leq \mu^*(A) \leq \mu(F)$$

and since $\mu(E)=\mu(F)$, the relation $\mu_*(A)=\mu^*(A)$ follows. One can obtain a proof that the relation $\mu_*(A)=\mu^*(A)<+\infty$ implies that A belongs to $\mathscr{A}_\mu$ by modifying the first paragraph of the proof of Lemma 1.5.3; the details are left to the reader (replace appeals to Proposition 1.4.1 with appeals to the definitions of μ_* and μ^*).

In this section we have been dealing with one way of approximating sets from above and from below by measurable sets. We turn to another such approximation. Let $\mathscr{A}$ be a σ -algebra on $\mathbb{R}^d$ that includes the σ -algebra $\mathscr{B}(\mathbb{R}^d)$ of Borel sets. A measure μ on $(\mathbb{R}^d,\mathscr{A})$ is regular if

1. each compact subset K of $\mathbb{R}^d$ satisfies $\mu(K)<+\infty$
2. each set A in $\mathscr{A}$ satisfies

$$\mu(A)=\inf\{\mu(U): U \text{is open and} A \subseteq U\}, \text{and}$$

3. each open subset U of $\mathbb{R}^d$ satisfies

$$\mu(U)=\sup\{\mu(K): K \text{is compact and} K \subseteq U\}$$

Proposition 1.4 .1 implies that Lebesgue measure, whether on $(\mathbb{R}^d,\mathscr{M}_{\lambda^*})$ or on $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$, is regular. Part (b) of that proposition appears to be stronger than condition (c) in the definition of regularity; however, we will see in Chap. 7 that every regular measure on $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$ satisfies the analogue of part (b) of Proposition 1.4.1. In Chap. 7 we will also see that on more general spaces, the analogue of condition (c) above, rather than of part (b) of Proposition 1.4.1, is the condition that should be used in the definition of regularity.

Proposition 1.5.6. Let μ be a finite measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$. Then μ is regular. Moreover, each Borel subset A of $\mathbb{R}^d$ satisfies

$$\mu(A) = \sup\{\mu(K) : K \subseteq A \text{ and } K \text{ is compact }\}$$

Let us first prove the following weakened form of Proposition 1.5 .6 Lemma 1.5.7. Let μ be a finite measure on $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$. Then each Borel subset A of $\mathbb{R}^d$ satisfies

$$\begin{aligned}
&\mu(A) = \inf\{\mu(U) : A \subseteq U \text{ and } U \text{ is open }\} \text{ and} \\
&\mu(A) = \sup\{\mu(C) : C \subseteq A \text{ and } C \text{ is closed }\}
\end{aligned}$$

Proof. Let $\mathscr{R}$ be the collection of those Borel subsets A of $\mathbb{R}^d$ that satisfy (5) and (6)

We begin by showing that $\mathscr{R}$ contains the open subsets of $\mathbb{R}^d$. Let V be an open subset of $\mathbb{R}^d$. Of course V satisfies

$$\mu(V) = \inf\{\mu(U) : V \subseteq U \text{ and } U \text{ is open }\}$$

According to Proposition 1.1.6, there is a sequence $\{C_n\}$ of closed subsets of $\mathbb{R}^d$ such that $V = \cup_n C_n$. We can assume that the sequence $\{C_n\}$ is increasing (replace C_n with $\cup_{i=1}^n C_i$ if necessary). Proposition 1.2 .5 implies that $\mu(V) = \lim_n \mu\left(C_n\right)$, and so V satisfies $\mu(V) = \sup\{\mu(C) : C \subseteq V$ and C is closed $\}$. With this we have proved that $\mathscr{R}$ contains all the open subsets of $\mathbb{R}^d$ It is easy to check (do so) that $\mathscr{R}$ consists of the Borel sets A that satisfy for each positive ε there exist an open set U and a closed set C such that $C \subseteq A \subseteq U$ and $\mu(U - C) < \varepsilon$.

We now show that $\mathscr{R}$ is a σ -algebra. If contains $\mathbb{R}^d$, since $\mathbb{R}^d$ is open. If $A \in \mathscr{R}$, if ε is a positive number, and if C and U are, respectively, closed and open and satisfy $C \subseteq A \subseteq U$ and $\mu(U - C) < \varepsilon$, then U^c and C^c are respectively closed and open and satisfy $U^c \subseteq A^c \subseteq C^c$ and $\mu\left(C^c - U^c\right) < \varepsilon$; thus it follows (from (7)) that $\mathscr{R}$ is closed under complementation. Now let $\{A_k\}$ be a sequence of sets in $\mathscr{R}$ and let ε be a positive number. For each k choose a closed set C_k and an open set U_k such that $C_k \subseteq A_k \subseteq U_k$ and $\mu\left(U_k - C_k\right) < \varepsilon/2^k$. Let $U = \cup_k U_k$ and $C = \cup_k C_k$. Then U and C satisfy the relations $C \subseteq \cup_k A_k \subseteq U$ and

$$\mu(U - C) \leq \mu\left(\cup_k\left(U_k - C_k\right)\right) \leq \sum_k \left(U_k - C_k\right) < \varepsilon$$

The set U is open, but the set C can fail to be closed. However, for each n the set $\cup_{k=1}^n C_k$ is closed, and it follows from (8), together with the fact that $\mu(U - C) = \lim_n \mu\left(U - \cup_{k=1}^n C_k\right)$ that there is a positive integer n such that $\mu\left(U - \cup_{k=1}^n C_k\right) < \varepsilon$ Then U and $\cup_{k=1}^n C_k$ are the sets required in (7), and $\mathscr{R}$ is closed under the formation of countable unions.

We have now shown that $\mathscr{R}$ is a σ -algebra on $\mathbb{R}^d$ that contains the open sets. since $\mathscr{B}\left(\mathbb{R}^d\right)$ is the smallest σ -algebra on $\mathbb{R}^d$ that contains the open sets, it follows that $\mathscr{B}\left(\mathbb{R}^d\right) \subseteq \mathscr{R}$. With this Lemma 1.5 .7 is proved.

Proof. Proof of Proposition 1.5.6. Condition (a) in the definition of regularity follows from the finiteness of μ, while condition (b) follows from Lemma 1.5.7. We turn to condition (c) and Eq. (4). Let A be a Borel subset of $\mathbb{R}^d$ and let ε be a positive number. Then according to Lemma 1.5 .7 there is a closed subset C of A such that $\mu(C) > \mu(A) - \varepsilon$. Choose an increasing sequence $\{C_n\}$ of closed and bounded (hence compact) sets whose union is C (these sets can, for example, be constructed by letting $C_n = C \cap \{x \in \mathbb{R}^d : \|x\| \leq n\}$). Proposition 1.2 .5 implies that $\mu(C) = \lim_n \mu(C_n)$, and so if n is large enough, then C_n is a compact subset of A such that $\mu(C_n) > \mu(A) - \varepsilon$. Equation (4) and condition (c) follow. ■

2.1.2 Dynkin Classes

This section is devoted to a technique that is often useful for verifying the equality of measures and the measurability of functions (measurable functions will be defined in Sect. 2.1). We begin with a basic definition.

Let X be a set. A collection $\mathscr{D}$ of subsets of X is a d -system (or a Dynkin class) on X if

1. $X \in \mathscr{D}$
2. $A - B \in \mathscr{D}$ whenever $A, B \in \mathscr{D}$ and $A \supseteq B$, and
3. $\cup_n A_n \in \mathscr{D}$ whenever $\{A_n\}$ is an increasing sequence of sets in $\mathscr{D}$.

A collection of subsets of X is a π -system on X if it is closed under the formation of finite intersections.

■ **Example 2.1** Suppose that X is a set and that $\mathscr{A}$ is a σ -algebra on X. Then $\mathscr{A}$ is certainly a d -system. Furthermore, if μ and v are finite measures on $\mathscr{A}$ such that $\mu(X) = v(X)$, then the collection $\mathscr{S}$ of all sets A that belong to $\mathscr{A}$ and satisfy $\mu(A) = v(A)$ is a d -system; it is easy to show by example that $\mathscr{S}$ is not necessarily a σ -algebra (see Exercise 3). The fact that such families $\mathscr{S}$ are d -systems forms the basis for many of the applications of d -systems.

Note that the intersection of a nonempty family of d -systems on a set X is a d system on X and that an arbitrary collection of subsets of X is included in some d -system on X, namely the collection of all subsets of X. Hence if $\mathscr{C}$ is an arbitrary collection of subsets of X, then the intersection of all the d -systems on X that include $\mathscr{C}$ is a d -system on X that includes $\mathscr{C}$; this intersection is the smallest such d -system and is called the d -system generated by $\mathscr{C}$. We will sometimes denote this d -system by $d(\mathscr{C})$

Theorem 2.1.11 Let X be a set, and let $\mathscr{C}$ be a π -system on X. Then the σ -algebra generated by $\mathscr{C}$ coincides with the d -system generated by $\mathscr{C}$.

Proof. Let $\mathscr{D}$ be the d -system generated by $\mathscr{C}$, and, as usual, let $\sigma(\mathscr{C})$ be the σ algebra generated by $\mathscr{C}$. since every σ -algebra is a d -system, the σ -algebra $\sigma(\mathscr{C})$ is a d -system that includes $\mathscr{C}$; hence $\mathscr{D} \subseteq \sigma(\mathscr{C})$. We can prove the reverse inclusion by showing that $\mathscr{D}$ is a σ -algebra, for then $\mathscr{D}$, as a σ -algebra that includes $\mathscr{C}$, must include the σ -algebra generated by $\mathscr{C}$, namely $\sigma(\mathscr{C})$

We begin the proof that $\mathscr{D}$ is a σ -algebra by showing that $\mathscr{D}$ is closed under the formation of finite intersections. Define a family $\mathscr{D}_1$ of subsets of X by letting

$$\mathscr{D}_1 = \{A \in \mathscr{D} : A \cap C \in \mathscr{D} \text{ for each } C \text{ in } \mathscr{C}\}$$

The fact that $\mathscr{C} \subseteq \mathscr{D}$ implies that $X \in \mathscr{D}_1$; furthermore, the identities

$$(A - B) \cap C = (A \cap C) - (B \cap C)$$

and

$$(\cup_n A_n) \cap C = \cup_n (A_n \cap C)$$

together with the fact that $\mathscr{D}$ is a d -system, imply that $\mathscr{D}_1$ is closed under the formation of proper differences and under the formation of unions of increasing sequences of sets. Thus $\mathscr{D}_1$ is a d -system. since $\mathscr{C}$ is closed under the formation of finite intersections and is included in $\mathscr{D}$, it is included in $\mathscr{D}_1$. Thus $\mathscr{D}_1$ is a d -system that includes $\mathscr{C}$; hence it must include $\mathscr{D}$. With this we have proved that we get a set in $\mathscr{D}$ whenever we take the intersection of a set in $\mathscr{D}$ and a set in $\mathscr{C}$. Next define $\mathscr{D}_2$ by letting

$$\mathscr{D}_2 = \{B \in \mathscr{D} : A \cap B \in \mathscr{D} \text{ for each } A \text{ in } \mathscr{D}\}$$

The previous step of this proof shows that $\mathscr{C} \subseteq \mathscr{D}_2$, and a straightforward modification of the argument in the previous step shows that $\mathscr{D}_2$ is a d -system. It follows that $\mathscr{D} \subseteq \mathscr{D}_2 -$ in other words, that $\mathscr{D}$ is closed under the formation of finite intersections. It is now easy to complete the proof. Parts (a) and (b) of the definition of a d system imply that $X \in \mathscr{D}$ and that $\mathscr{D}$ is closed under complementation. As we have just seen, $\mathscr{D}$ is also closed under the formation of finite intersections, and so it is an algebra. Finally $\mathscr{D}$, as a d -system, is closed under the formation of unions of increasing sequences of sets, and so by Proposition 1.1 .7 it must be a σ -algebra; with that the proof is complete. ■

Corollary 2.1.12 Let $(X, \mathscr{A})$ be a measurable space, and let $\mathscr{C}$ be a π -system on X such that $\mathscr{A} = \sigma(\mathscr{C})$. If μ and v are finite measures on $\mathscr{A}$ that satisfy $\mu(X) = v(X)$ and that satisfy $\mu(C) = v(C)$ for each C in $\mathscr{C}$, then $\mu = v$

Proof. Let $\mathscr{D} = \{A \in \mathscr{A} : \mu(A) = v(A)\}$. As we noted above, $\mathscr{D}$ is a d -system. since $\mathscr{C}$ is a π -system and is included in $\mathscr{D}$, it follows from Theorem 1.6 .2 that $\mathscr{D} \supseteq \sigma(\mathscr{C}) = \mathscr{A}$. Thus $\mu(A) = v(A)$ holds for each A in $\mathscr{A}$, and the proof is complete. ■

Now suppose that μ and v are finite Borel measures on $\mathbb{R}$ such that $\mu(I)=v(I)$ holds for each interval I of the form $(-\infty, b]$. Note that $\mathbb{R}$ is the union of an increasing sequence of intervals of the form $(-\infty, b]$ and hence that $\mu(\mathbb{R})=v(\mathbb{R})$ since the collection of all intervals of the form $(-\infty, b]$ is a π -system that generates $\mathscr{B}(\mathbb{R})$ (see Proposition 1.1 .4), it follows from Corollary 1.6 .3 that $\mu=v$. With this we have another proof of the uniqueness assertion in Proposition 1.3 .10 .

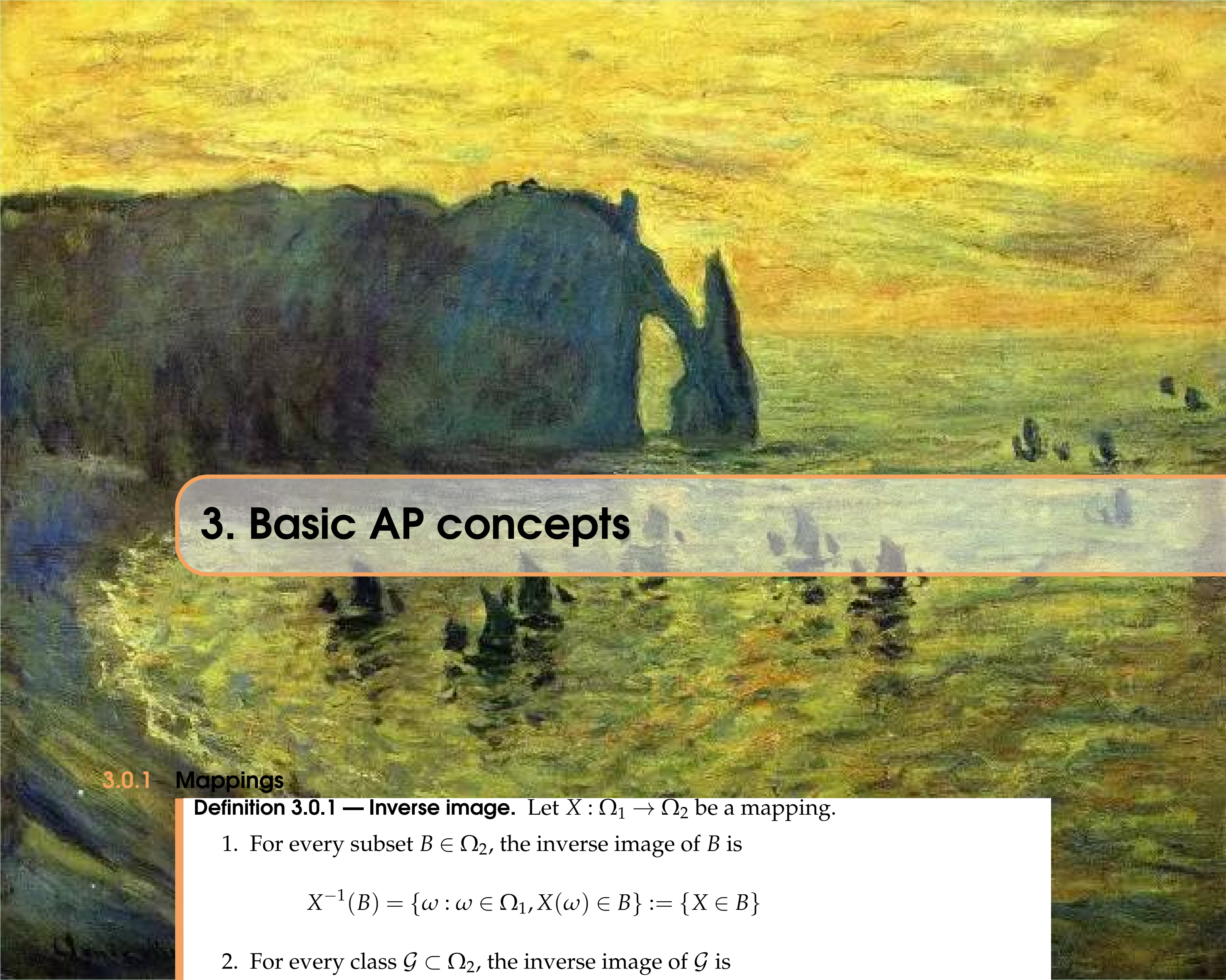
The following result is essentially an extension of Corollary 1.6 .3 to the case of σ -finite measures. Note that it implies that Lebesgue measure is the only measure on $\mathscr{B}\left(\mathbb{R}^d\right)$ that assigns to each d -dimensional interval its volume, and so it provides a second proof of part of Proposition 1.4.3.

Corollary 2.1.13 Let $(X, \mathscr{A})$ be a measurable space, and let $\mathscr{C}$ be a π -system on X such that $\mathscr{A}=\sigma(\mathscr{C})$. If μ and v are measures on $(X, \mathscr{A})$ that agree on $\mathscr{C}$, and if there is an increasing sequence $\{C_n\}$ of sets that belong to $\mathscr{C}$, have finite measure under μ and v, and satisfy $\cup_n C_n=X$, then $\mu=v$

Proof. Choose an increasing sequence $\{C_n\}$ of sets that belong to $\mathscr{C}$, have finite measure under μ and v, and satisfy $\cup_n C_n=X$. For each positive integer n define measures μ_n and v_n on $\mathscr{A}$ by $\mu_n(A)=\mu\left(A \cap C_n\right)$ and $v_n(A)=v\left(A \cap C_n\right)$ Corollary 1.6 .3 implies that for each n we have $\mu_n=v_n$. since

$$\mu(A)=\lim_n \mu_n(A)=\lim_n v_n(A)=v(A)$$

holds for each A in $\mathscr{A}$, the measures μ and v must be equal. ■



3. Basic AP concepts

3.0.1 Mappings

Definition 3.0.1 — Inverse image. Let $X:\Omega_1 \to \Omega_2$ be a mapping.

1. For every subset $B \in \Omega_2$, the inverse image of B is
$$X^{-1}(B) = \{\omega : \omega \in \Omega_1, X(\omega) \in B\} := \{X \in B\}$$
2. For every class $\mathcal{G} \subset \Omega_2$, the inverse image of $\mathcal{G}$ is
$$X^{-1}(\mathcal{G}) = \left\{X^{-1}(B) : B \in \mathcal{G}\right\}$$
So $\left(A \in X^{-1}(\mathcal{G}) \text{ means that } \exists B \in \mathcal{G} \text{ s.t. } A = X^{-1}(B).\right)$

Theorem 3.0.1 — Properties of the inverse image.

1. X is a mapping from Ω_1 to Ω_2. Then
   (a) $X^{-1}(\Omega_2) = \Omega_1, X^{-1}(\emptyset) = \emptyset$
   (b) $X^{-1}(B^c) = \left[X^{-1}(B)\right]^c$
   (c) $X^{-1}(\cup_{\gamma\in\Gamma} B_\gamma) = \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma)$ for $B_\gamma \subset \Omega_2, \gamma \in \Gamma$
   (d) $X^{-1}(\cap_{\gamma\in\Gamma} B_\gamma) = \cap_{\gamma\in\Gamma} X^{-1}(B_\gamma)$ for $B_\gamma \subset \Omega_2, \gamma \in \Gamma$ where Γ is an index set, not necessarily countable.
   (e) $X^{-1}(B_1 - B_2) = X^{-1}(B_1) - X^{-1}(B_2)$ for $B_1, B_2 \subset \Omega_2$
   (f) monotoncity: $B_1 \subset B_2 \subset \Omega_2$ implies that $X^{-1}(B_1) \subset X^{-1}(B_2)$
2. If $\mathcal{B}$ is a σ-algebra in Ω_2, then $X^{-1}(\mathcal{B})$ is a σ -algebra in Ω_1.

3. Let $\mathcal{C}$ be a nonempty class in Ω_2, then

$$X^{-1}(\sigma(\mathcal{C})) = \sigma\left(X^{-1}(\mathcal{C})\right)$$

Proof. 1. (a) Note that

$$X^{-1}(\Omega_2) = \{\omega \in \Omega_1 : X(\omega) \in \Omega_2\} = \Omega_1$$
$$X^{-1}(\varnothing) = \{\omega \in \Omega_1 : X(\omega) \in \varnothing\} = \varnothing$$

(b) $X^{-1}(B^c) = \{\omega : X(\omega) \in B^c\} = \{\omega : X(\omega) \notin B\} = \{\omega : \omega \notin X^{-1}(B)\} = \left\{\omega : \omega \in \left[X^{-1}(B)\right]^c\right\} = \left[X^{-1}(B)\right]^c$

(c) $\omega \in X^{-1}(\cup_{\gamma\in\Gamma} B_\gamma) \iff X(\omega) \in \cup_{\gamma\in\Gamma} B_\gamma \iff X(\omega) \in B_\gamma$ for some $\gamma \in \Gamma \iff \omega \in X^{-1}(B_\gamma)$ for some $\gamma \in \Gamma \iff \omega \in \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma)$

(d) similar to (iii).

(e) $X^{-1}(B_1 - B_2) = X^{-1}(B_1 \cap B_2^c) = X^{-1}(B_1) \cap X^{-1}(B_2^c) = X^{-1}(B_1) \cap \left[X^{-1}(B_2)\right]^c = X^{-1}(B_1) - X^{-1}(B_2)$

(f) $X^{-1}(B_1) = \{\omega : X(\omega) \in B_1\} \subset \{\omega : X(\omega) \in B_2\} = X^{-1}(B_2)$

2. First $X^{-1}(\mathcal{B})$ is nonempty as $\Omega_1 = X^{-1}(\Omega_2) \in X^{-1}(\mathcal{B})$ ($\mathcal{B}$ is a σ-algebra). Next, let $A \in X^{-1}(\mathcal{B})$. Then $\exists B \in \mathcal{B}$ s.t. $A = X^{-1}(B)$. So $A^c = X^{-1}(B^c) \in X^{-1}(\mathcal{B})$. So $X^{-1}(\mathcal{B})$ is closed under complement. Similarly, it can be shown that it is closed under countable union. Thus it is a σ -algebra.
3. Clearly, $X^{-1}(\mathcal{C}) \subset X^{-1}(\sigma(\mathcal{C}))$, which is a σ -algebra from (2). Thus $\sigma\left(X^{-1}(\mathcal{C})\right) \subset X^{-1}(\sigma(\mathcal{C}))$. It remains to show

$$X^{-1}(\sigma(\mathcal{C})) \subset \sigma\left(X^{-1}(\mathcal{C})\right)$$

Define

$$\mathcal{G} = \left\{G : X^{-1}(G) \in \sigma\left(X^{-1}(\mathcal{C})\right)\right\}$$

It suffices to show that $\sigma(\mathcal{C}) \subset \mathcal{G}$. since $\mathcal{C} \subset \mathcal{G}$, we only need to show that $\mathcal{G}$ is a σ -algebra. (Proof of $\mathcal{C} \subset \mathcal{G}$: take $G \in \mathcal{C}$, then $X^{-1}(G) \in X^{-1}(\mathcal{C}) \in \sigma\left(X^{-1}(\mathcal{C})\right)$, hence $G \in \mathcal{G}$. It remains to show that $\mathcal{G}$ is a σ -algebra. Let $G \in \mathcal{G} \Longrightarrow X^{-1}(G) \in \sigma\left(X^{-1}(\mathcal{C})\right), \Longrightarrow X^{-1}(G^c) = \left[X^{-1}(G)\right]^c \in \sigma\left(X^{-1}(\mathcal{C})\right) \Longrightarrow G^c \in \mathcal{G} \Longrightarrow \mathcal{G}$ is closed under complement. Similarly, we can show that it is closed under countable union. The proof is complete.

■

R Clearly, $X^{-1}(\cdot)$ on Ω_2 preserves **all the set operations** on Ω_1

Definition 3.0.2 — Measurable mapping. 1. $(\Omega_1, \mathcal{A})$ and $(\Omega_2, \mathcal{B})$ are measurable spaces. $X : \Omega_1 \to \Omega_2$ is a measurable mapping if

$$X^{-1}(B) \equiv \{X \in B\} \in \mathcal{A}, \quad \forall B \in \mathcal{B}$$

2. X is a measurable function if $(\Omega_2, \mathcal{B}) = (\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$ in (1).
3. X is a Borel (measurable) function if $(\Omega_1, \mathcal{A}) = (\mathcal{R}^m, \mathcal{B}(\mathcal{R}^m))$ and $(\Omega_2, \mathcal{B}) = (\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$ in(1)

Definition 3.0.3 If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G})$ is a sub- σ -field of $\mathcal{F}$(verify). It is called the σ -field generated by f and is denoted by $\sigma(f)$

The next theorem is useful in checking if X is measurable or not.

Theorem 3.0.2 — X is measurable or not, just need to check the generator. $X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ is a measurable mapping if $\mathcal{B} = \sigma(\mathcal{C})$ and $X^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$

Proof. From Theorem 3.0.1 (iii), we have

$$X^{-1}(\mathcal{B}) = X^{-1}(\sigma(\mathcal{C})) = \sigma\left(X^{-1}(\mathcal{C})\right) \subset \sigma(\mathcal{A}) = \mathcal{A}$$

■

Theorem 3.0.3 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 $f(X) = f \cdot X : (\Omega_1, \mathcal{A}_1) \to (\Omega_3, \mathcal{A}_3)$ is also measurable.

Proof. $\forall A_3 \in \mathcal{A}_3, \{f \cdot X \in A_3\} = \left\{X \in f^{-1}(A_3) \in \mathcal{A}_2\right\} \in \mathcal{A}_1$

A more detailed proof is: $\forall A_3 \in \mathcal{A}_3$,

$$\begin{aligned}(f \cdot X)^{-1}(A_3) &= \{\omega : f \cdot X(\omega) \in A_3\} = \left\{\omega : X(\omega) \in f^{-1}(A_3) \in \mathcal{A}_2\right\} \\ &= \left\{\omega : \omega \in X^{-1}\left(f^{-1}(A_3)\right)\right\} \in \mathcal{A}_1\end{aligned}$$

■

3.1 Random Variables

Definition 3.1.1 — Random variables–general. A **random variable** (r.v.) X is a **measurable function** from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$. $\Longleftrightarrow \{X \in B\} = X^{-1}(B) \in \mathcal{A}$ for all Borel set $B \in \mathcal{B}$. We then say that X is $\mathcal{A}-$ measurable, or simply write it as

$$X \in \mathcal{A}$$

(R) This definition does not involve any probability measures.

Definition 3.1.2 — Another definition of random variable. A random variable (r.v.) X is a measurable mapping from $(\Omega, \mathcal{A}, P)$ to $(\mathcal{R}, \mathcal{B})$ such that $P(|X| = \infty) = P(\{\omega : |X(\omega)| = \infty\}) = 0$

Definition 3.1.3 — Random Variable–adapt to probability space. A random variable X is a measurable function from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the reals. i.e., it is a function

$$X : \Omega \longrightarrow (-\infty, \infty)$$

such that for every Borel set B

$$X^{-1}(B) = \{X \in B\} \in \mathcal{F}$$

Here we use the shorthand notation

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$$

If X is a random variable, then for every Borel subset B of $\mathbb{R}, X^{-1}(B) \in \mathcal{F}$. We can define a function on Borel sets by

$$\mu_X(B) = \mathbb{P}\{X \in B\} = \mathbb{P}\left[X^{-1}(B)\right]$$

This function is in fact a **measure**, and $(\mathbb{R}, \mathcal{B}, \mu_X)$ is a probability space.

Definition 3.1.4 — Random vectors. $X = (X_1, \ldots, X_n)$ is a random vector if X_k is a r.v. on $(\Omega, \mathcal{A})$ for $1 \leq k \leq n$

Theorem 3.1.1 $X = (X_1, \ldots, X_n)$ is a random vector $\Longrightarrow X$ is a measurable function from $(\Omega, \mathcal{A})$ to $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$

Proof. Let $I_k = (a_k, b_k], -\infty \leq a_k \leq b_k \leq \infty, k \geq 1$. Since $\{X_i \in I_k\} \in \mathcal{A}$

$$\{X = (X_1, \ldots, X_n) \in I_1 \times \ldots \times I_n\} = \bigcap_{k=1}^{n} \{X_i \in I_k\} \in \mathcal{A}$$

The proof follows from this and Theorem 3.0.2 as $\mathcal{B}(\mathcal{R}^n) = \sigma(\{\prod_{k=1}^n I_k\})$. ∎

3.1.1 Construction of random variables

Algebraic operations $(+, -, \times, \div)$

Theorem 3.1.2 If X, Y are $r.v.'s$ (i.e., $X, Y \in \mathcal{A}$), so are

$$aX + bY, \quad X \vee Y = \max\{X, Y\}, \quad X \wedge Y = \min\{X, Y\}$$
$$X^2, \quad XY, \quad X/Y (Y(\omega) \neq 0)$$

Proof.

$$
\begin{aligned}
\{aX \le t\} =& \{X \le t/a\} \quad \text{if } a > 0 \\
& \{X \ge t/a\} \quad \text{if } a < 0 \\
& \in \mathcal{A} \\
\{X+Y < t\} =& \bigcup_{r \in \mathcal{Q}} (\{X < r\} \cap \{Y < t - r\}), \quad (\mathcal{Q} = \text{ all rational numbers }) \\
& \in \mathcal{A}
\end{aligned}
$$

(the proof of this is given at the end)

$$
\begin{aligned}
\{X^2 \le t\} =& \{|X| \le \sqrt{t}\} = \{X \le \sqrt{t}\} - \{X < -\sqrt{t}\} \quad \text{if } t \ge 0 \\
& \emptyset, \quad \text{if } t < 0 \\
& \in \mathcal{A} \\
XY =& \frac{1}{2}\left((X+Y)^2 - X^2 - Y^2\right) \in \mathcal{A}
\end{aligned}
$$

$$
\begin{aligned}
\{1/Y \le t\} &= \{1/Y \le t\} \cap (\{Y < 0\} \cup \{Y > 0\}) \;\; (\text{which is equal to} \Omega) \\
&= (\{1/Y \le t\} \cap \{Y < 0\}) \cup (\{1/Y \le t\} \cap \{Y > 0\}) \\
&= \{Y \ge 1/t\} \cup \{Y < 0\} \quad \text{if } t \ge 0 \\
&= \{Y \ge 1/t\} \cap \{Y < 0\} \quad \text{if } t < 0 \\
&\in \mathcal{A} \\
X/Y &= X(1/Y) \in \mathcal{A} \\
\{X \wedge Y \ge t\} &= \{X \ge t\} \cap \{Y \ge t\} \in \mathcal{A} \\
\{X \vee Y \le t\} &= \{X \le t\} \cap \{Y \le t\} \in \mathcal{A}
\end{aligned}
$$

Proof of (4.1).

. Clearly, RHS $\subset$ LHS. If LHS holds, i.e. $X+Y < t, \Longrightarrow$ Clearly, for any point (r,s) in the open triangle with vertexes $(X,Y),(X,t-X)$, and $(t-Y,Y)$, denoted by Δ, we have $r+s < t$. $\Rightarrow$ In particular, we can choose a rational point $(r,s) \in \Delta$. Then we have $X < r$ and $Y < s < t-r$. ■

Definition 3.1.5 The positive and negative parts of a function $X : \Omega \to \mathcal{R}$ are

$$
X^+ = \max\{X, 0\}, \quad X^- = -\min\{X, 0\}
$$

It is clear that $X = X^+ - X^-$ and $|X| = X^+ + X^-$

Corollary 3.1.3 If X is a r.v. (i.e., $X \in \mathcal{A}$), so are X^+, X^- and $|X|$.

Limiting operations

Theorem 3.1.4 $X_1, X_2, \ldots$ are $r.v.$ on $(\Omega, \mathcal{A}), (i.e., X_i \in \mathcal{A})$

1. $\sup_n X_n, \inf_n X_n, \limsup_n X_n$, and $\liminf_n X_n$ 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., $S \in \mathcal{A}$)

Proof. 1.

$$\left\{\sup_n X_n \leq t\right\} = \bigcap_{n=1}^{\infty} \{X_n \leq t\} \quad \left\{\inf_n X_n \geq t\right\} = \bigcap_{n=1}^{\infty} \{X_n \geq t\}$$

$$\limsup_n X_n = \inf_k \sup_{m \geq k} X_m \quad \liminf_n X_n = \sup_k \inf_{m \geq k} X_m$$

2. $X(\omega) = \lim_n X_n(\omega) = \limsup_n X_n(\omega)$ is a r.v.
3. This follows from (2) .

■

Definition 3.1.6 — Converges almost surely. Let $X_1, X_2, \ldots$ be a sequence of r. v.'s on $(\Omega, \mathcal{A}, P)$. Define $\Omega_0 \equiv \{\omega : \lim_n X_n(\omega) \text{ exists }\} = \{\omega : \limsup_n X_n(\omega) - \liminf_n X_n(\omega) = 0\}$. Clearly, Ω_0 is measurable from the last theorem.

If $P(\Omega_0) = 1$, we say that X_n **converges almost surely** (a.s.) and write $X_n \to X$ a.s.

Transformations

Theorem 3.1.5 $X = (X_1, \ldots, X_n)$ is a random n-vector, f is a Borel function from $\mathcal{R}^n$ to $\mathcal{R}^m$. Then $f(X)$ is a random m -vector.

Proof. The proof follows directly from Theorems 3.3 .2 and 3.2.2. ■

R

1. Note that continuous functions are Borel measurable. So if $X, X_1, \ldots, X_n$ are r.v.'s, so are

$$X_1 + \ldots + X_n, \quad \sin(X), \quad e^X, \quad \text{Polynomial }(X), \quad \text{etc.}$$

2. We illustrate how to show that the function $f(x_1, x_2) = x_1 + x_2$ is measurable. We need to show that $\{(x_1, x_2) : x_1 + x_2 < a\}$ is an open set in $\mathcal{R}^2$, which is true since $\{(x_1, x_2) : x_1 + x_2 < a\} = \bigcup_{r \in \mathcal{Q} \cap A} \text{Rectangle }(r, \epsilon_r) \in \mathcal{B}^2$

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

First we introduce a few simple r.v.'s, which forms the basis of all other r.v.'s.

Theorem 3.1.6
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. The class of **simple functions** is obtained by taking linear combinations of indicators of measurable sets, i.e.,

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

where $A_1, \ldots, A_k$ are measurable sets on Ω and $a_1, \ldots, a_k$ are real numbers.
3. (simple r.v.). If $\Omega = \sum_1^n A_i$, where $A_i \in \mathcal{A}$, then $X = \sum_1^n a_i I_{A_i}$ is a r.v. (For simplicity, we assume that $\{a_1, \ldots, a_n\}$ are distinct.)

Proof.
1. Method I: by using the definition.
   (a) $\forall B \in \mathcal{B}$, note that

$$\begin{aligned} \{I_A \in B\} = \emptyset &\text{ if } 0 \notin B, 1 \notin B \\ A^c &\text{ if } 0 \in B, 1 \notin B \\ A &\text{ if } 0 \notin B, 1 \in B \\ \Omega &\text{ if } 0 \in B, 1 \in B \end{aligned}$$

   Since $A \in \mathcal{A}$, we see that $\{I_A \in B\} \in \mathcal{A}$. Thus I_A is a r.v.
   (b) $\forall B \in \mathcal{B}$, note that $\{X \in B\} = \bigcup_{\{i: a_i \in B\}} A_i \in \mathcal{A}$. The proof is complete.
2. Method II: by using Theorem 3.1.8
   (a) Clearly,

$$\begin{aligned} \{I_A \le t\} = \emptyset &\text{ if } t < 0 \\ A^c &\text{ if } 0 \le t < 1 \\ \Omega &\text{ if } t \ge 1 \end{aligned}$$

   Thus $\{I_A \in B\} \in \mathcal{A}$, i.e., I_A is a r.v.
   (b) It follows from (1) that I_{A_i} 's are r.v.'s. Then the proof follows from Theorem 3.1.2.

■

R Any random variable can be approximated by simple ones (this is crucial to the definition of expectation later on). From now on, when no confusion arises, we often write

$$I_A \equiv I\{A\}$$

Theorem 3.1.7 — Non-negetive random variables can bu approximated by simple ones. Given a r.v. $X \geq 0$ on $(\Omega, \mathcal{A})$, there exists simple r.v. 's $0 \leq X_1 \leq X_2 \leq \ldots \leq \ldots$ with $X_n(\omega) \nearrow X(\omega)$ for every $\omega \in \Omega$

Proof. Proof. $\forall n \geq 1$, let

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

Now let us give a rigorous proof of the theorem. Clearly, $X_n(\omega) \geq 0$ for all n. Next we show that $X_n(\omega) \nearrow$ for all $\omega \in \Omega$. For any $n \geq 1$ and $\omega \in \Omega$, we have

$$\begin{aligned}
X_{n+1}(\omega) &= \sum_{k=1}^{(n+1)2^{n+1}} \frac{k-1}{2^{n+1}} I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \leq \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\
&= \left(\sum_{k=1}^{n2^{n+1}} + \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}}\right) \frac{k-1}{2^{n+1}} I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \leq \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\
&:= A + B + C
\end{aligned}$$

Now

$$\begin{aligned}
A &= \sum_{k=1}^{n2^{n+1}} \frac{k-1}{2^{n+1}} I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \leq \frac{k}{2^{n+1}}\right\} \\
&= 0 \times I\left\{0 < X(\omega) \leq \frac{1}{2^{n+1}}\right\} + \frac{1}{2^{n+1}} I\left\{\frac{1}{2^{n+1}} < X(\omega) \leq \frac{2}{2^{n+1}}\right\} \\
&+ \frac{2}{2^{n+1}} I\left\{\frac{2}{2^{n+1}} < X(\omega) \leq \frac{3}{2^{n+1}}\right\} + \frac{3}{2^{n+1}} I\left\{\frac{3}{2^{n+1}} < X(\omega) \leq \frac{4}{2^{n+1}}\right\} \\
&+ \cdots\cdots\cdots\cdots \\
&+ \cdots\cdots\cdots\cdots \\
&+ \frac{n2^{n+1}-2}{2^{n+1}} I\left\{\frac{n2^{n+1}-2}{2^{n+1}} < X(\omega) \leq \frac{n2^{n+1}-1}{2^{n+1}}\right\} + \frac{n2^{n+1}-1}{2^{n+1}} I\left\{\frac{n2^{n+1}-1}{2^{n+1}} < X(\omega) \leq \frac{n2^{n+1}}{2^{n+1}}\right\} \\
&\geq 0 \times I\left\{0 < X(\omega) \leq \frac{1}{2^{n+1}}\right\} + 0 \times I\left\{\frac{1}{2^{n+1}} < X(\omega) \leq \frac{1}{2^n}\right\} \\
&+ \frac{1}{2^n} I\left\{\frac{1}{2^n} < X(\omega) \leq \frac{3}{2^{n+1}}\right\} + \frac{1}{2^n} I\left\{\frac{3}{2^{n+1}} < X(\omega) \leq \frac{2}{2^n}\right\} \\
&+ \cdots\cdots\cdots\cdots \\
&+ \cdots\cdots\cdots\cdots \\
&+ \frac{n2^n-1}{2^n} I\left\{\frac{n2^n-1}{2^n} < X(\omega) \leq \frac{n2^{n+1}-1}{2^{n+1}}\right\} + \frac{n2^n-1}{2^n} I\left\{\frac{n2^{n+1}-1}{2^{n+1}} < X(\omega) \leq \frac{n2^{n+1}}{2^{n+1}}\right\} \\
&= \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} < X(\omega) \leq \frac{k}{2^n}\right\}
\end{aligned}$$

$$\begin{aligned}
B+C &= \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}} \frac{k-1}{2^{n+1}} I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\
&\ge \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}} \frac{n2^{n+1}}{2^{n+1}} I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + nI\{X(\omega) > n+1\} \\
&\ge nI\{n < X(\omega) \le n+1\} + nI\{X(\omega) > n+1\} \\
&= nI\{X(\omega) > n\}
\end{aligned}$$

Therefore,

$$X_{n+1}(\omega) = A+B+C \ge \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} < X(\omega) \le \frac{k}{2^n}\right\} + nI\{X(\omega) > n\} = X_n(\omega) \ge 0$$

Thus, $X_n(\omega) \nearrow$ for every $\omega \in \Omega$. So $\lim_{n\to\infty} X_n(\omega)$ exists (maybe ∞). It remains to show that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. First, if $X(\omega) = \infty$, then by definition, we have $X_n(\omega) = n \to \infty = X(\omega)$. If $X(\omega) < \infty$, then for n large enough, we have

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

We observe that there exists $(1 \le k \le n2^n)$ such that $\left\{\frac{k-1}{2^n} < X(\omega) \le \frac{k}{2^n}\right\}$ in which case $X_n(\omega) = \frac{k-1}{2^n}$ the left end point of the set. Therefore,

$$0 \le X(\omega) - \frac{k-1}{2^n} = X(\omega) - X_n(\omega) \le \frac{1}{2^n} \to 0$$

■

Definition 3.1.7 For any $x \in \mathcal{R}$, let

$$[x] \text{be the largest integer } \le x; \quad \lfloor x \rfloor \text{be the largest integer } < x$$

It is easy to see that

$$\lfloor x \rfloor = [x] \qquad \text{if } x \text{ is a non-integer}$$
$$[x] - 1 \qquad \text{if } x \text{ is an integer}$$

Then, it can easily shown that the expression $X_n(\omega)$ given above can also be written down simply as

$$X_n(\omega) = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n$$

How to check a random variable?

To verify X is a r.v., we don't need to check that $\{X \in B\} \in \mathcal{A}$ for all Borel sets B. One only needs to check this for all intervals.

Theorem 3.1.8 — Check a random variable. X is a r.v. from $(\Omega,\mathcal{A})$ to $(\mathcal{R},\mathcal{B})$, (i.e., $X\in\mathcal{A}$)

$$\begin{aligned}&\Longleftrightarrow \{X\le x\}=X^{-1}([-\infty,x])\in\mathcal{A} \text{ for all } x\in\mathcal{R}\\&\Longleftrightarrow \{X\le x\}=X^{-1}([-\infty,x])\in\mathcal{A}, x\in\mathcal{D} \text{ which is a dense subset of } \mathcal{R}\end{aligned}$$

Proof. Take $\mathcal{C}=\{[-\infty,b]:b\in\mathcal{R}\}$ or $\mathcal{C}=\{[-\infty,b]:b\in\mathcal{D}\}$ in Theorem 3.0.2 ■

R

1. We can take $\mathcal{D}$ to be all rational numbers (which is dense).
2. $\{X\le x\}$ in the theorem can be replaced by any of the following:

$$\{X\le x\},\quad \{X\ge x\},\quad \{X<x\},\quad \{X>x\},\quad \{x<X<y\}, \text{ etc.}$$

σ-algebra generated by random variables.

Definition 3.1.8 Let $\{X_\lambda,\lambda\in\Lambda\}$ be a nonempty family of r. v.'s on $(\Omega,\mathcal{A})(\Lambda$ may not be countable). Define

$$\sigma(X_\lambda,\lambda\in\Lambda):=\sigma(X_\lambda\in B, B\in\mathcal{B},\lambda\in\Lambda)=\sigma\left(X_\lambda^{-1}(\mathcal{B}),\lambda\in\Lambda\right)=\sigma\left(\cup_{\lambda\in\Lambda}X_\lambda^{-1}(\mathcal{B})\right)$$

which is called the σ -algebra generated by $X_\lambda,\lambda\in\Lambda$.

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

$$\begin{aligned}\sigma(X_i)&=\sigma\left(X_i^{-1}(\mathcal{B})\right)=X_i^{-1}(\mathcal{B})=\{X_i\in\mathcal{B}\}\\\sigma(X_1,\ldots,X_n)&=\sigma\left(\cup_{i=1}^n X_i^{-1}(\mathcal{B})\right)=\sigma\left(\cup_{i=1}^n\sigma(X_i)\right)\end{aligned}$$

2. For $\Lambda=\{1,2,\ldots,\}$, it is easy to check that

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

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

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

Examples

■ **Example 3.1 — The σ -field generated by a discrete r.v..** Consider a discrete r.v. X taking distinct values $\{x_i, 1\le i\le n\}$ (where n could take ∞) and define $A_i=\{\omega:X(\omega)=x_i\}$. We have the following results.

1. $\{A_i, i\ge 1\}$ constitute a disjoint partition of Ω.
2. Choose $\mathcal{C}=\{A_1,A_2,\ldots,A_n\}$, then

$$\sigma(\mathcal{C})=\sigma(A_1,A_2,\ldots,A_n)=\sigma(A_0,A_1,A_2,\ldots,A_n)=\{\cup_{i\in I}A_i: I\subset\{0,1,2,\ldots,n\}\}$$

where $A_0 = \emptyset$. (Hint: show that the RHS forms a σ -algebra.)

R One key assumption in the above example is that the sets A_i 's form a partition of Ω (i.e., they are mutually exclusive and exhaustive.) When this is not satisfied, we can use disjointization techniques to form a **partition** first and then apply this theorem. See the next example.

R Note that when n is finite, the total number of elements in $\sigma(A_1, A_2, \ldots, A_n)$ is

$$\binom{n}{0} + \binom{n}{1} + \ldots\ldots + \binom{n}{n} = (1+1)^n = 2^n$$

■ **Example 3.2** Let Ω be a sample space, and $A, B, C \subset \Omega$. (For simplicity, we assume that all three sets are *NOT* mutually exclusive to each other and *NOT* equal to $\emptyset$ or Ω.) Let $|A|$ denote the total number of elements in A. Find

1. $\sigma(\emptyset)$ and $|\sigma(\emptyset)|$
2. $\sigma(A)$ and $|\sigma(A)|$
3. $\sigma(A,B)$ and $|\sigma(A,B)|$
4. $\sigma(A,B,C)$ and $|\sigma(A,B,C)|$

Proof. We provide two different methods to do this.

1. Method 1
   (a) $\sigma(\emptyset) = \{\emptyset, \Omega\}$ and $|\sigma(\emptyset)| = 2$
   (b) $\sigma(A) = \{\emptyset, A, A^c, \Omega\}, \quad |\sigma(A)| = 4 = 2^2$
   (c) The answer to this is more complicated:

$$\sigma(A,B) = \{\emptyset, A, B, A^c, B^c, \quad A\cup B, A\cup B^c, A^c\cup B, A^c\cup B^c, \quad A^c\cap B^c, A^c\cap B, A\cap B^c, A\cap B, \\ (A\cap B)\cup(A^c\cap B^c), ((A\cap B)\cup(A^c\cap B^c))^c, \quad \Omega\}$$

   and hence $|\sigma(A,B)| = 16 = 2^4 = 2^{2^2}$
   (d) can be seen that it will get messier as the number of subsets increases. Just try to find $\sigma(A,B,C)$ this way yourself. You will find that $|\sigma(A,B,C)| = 2^{2^3} = 256$, certainly too many elements to write down by hand.
2. Method 2. One alternative method is given below, which provides a general method to do this. The key is disjointization. If

$$\Omega = A_1 + A_2 + \ldots + A_n$$

   then, $\sigma(A_1, \ldots, A_n) = \{\cup_{i\in J} A_i, J \text{ is a subset of } \{0,1,\ldots,n\}\}$ where we denote $A_0 = \emptyset$.
   (a) $\sigma(\emptyset) = \{\emptyset, \Omega\}$ and $|\sigma(\emptyset)| = 2$

(b) First we will find $\sigma(A)$. From the Venn diagram, we can disjointize Ω as

$$\Omega = A + A^c =: A_1 + A_2$$

Following Remark 3.1.2, $|\sigma(A)| = 2^2 = 4$. From the last example, we have

$$\begin{aligned}\sigma(A) &= \sigma(A, A^c) = \sigma(A_1, A_2)\\ &= \{\cup_{i\in J} A_i, J \text{ is a subset of } \{1, \ldots, n\}\}\\ &= \{\emptyset, A_1, A_2, A_1 \cup A_2\}\\ &= \{\emptyset, A, A^c, \Omega\}\end{aligned}$$

(c) 3. Next, we will find $\sigma(A, B)$. From the Venn diagram, we can disjointize Ω as

$$\Omega = (A - B) + (B - A) + (A \cap B) + (A \cup B)^c =: A_1 + \ldots + A_4$$

Following Remark 3.6.3, $|\sigma(A, B)| = 2^4 = 16 = 4^2$. From the last example, we have

$$\begin{aligned}\sigma(A, B) =&\sigma(A_1, \ldots, A_4)\\ =&\{\cup_{i\in J} A_i, J \text{ is a subset of } \{1, \ldots, n\}\}\\ =&\{\emptyset, A_1, A_2, A_3, A_4, A_1 + A_2, A_1 + A_3, A_1 + A_4, A_2 + A_3, A_2 + A_4, A_3 + A_4\\ &A_1 + A_2 + A_3, A_1 + A_2 + A_4, \ldots, A_2 + A_3 + A_4, A_1 + A_2 + A_3 + A_4\}\\ =&\{\emptyset, \Omega, A, B, A^c, B^c\\ &A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c\\ &A^c \cap B^c, A^c \cap B, A \cap B^c, A \cap B\\ &(A \cap B) \cup (A^c \cap B^c), ((A \cap B) \cup (A^c \cap B^c))^c\}\end{aligned}$$

(d) Finally, we can find $\sigma(A, B, C)$, just as above. Similarly to the above, we have $|\sigma(A, B, C)| = 2^{2^3} = 256$. We leave the rest as an exercise.

■

R If $A_1, \ldots, A_n$ are not mutually exclusive to each other for all pairs, then from the last ω, it may or may not belong to $A_i, i = 1, \ldots, n$. Thus the total number of mutually exclusive sets is $2 \times 2 \times \ldots . \times 2 = 2^n$. Following Remark , we get $|\sigma(A_1, \ldots, A_n)| = 2^{2^n}$

Now we take $A = \{X_1 = 1\}, B = \{X_2 = 1\}$, and $C = \{X_3 = 1\}$. Also let

$$Y = X_1 + X_2 + X_3$$

to be the total number of heads in 3 tosses. Define

$$\mathcal{F}_1 = \sigma(A), \quad \mathcal{F}_2 = \sigma(A, B), \quad \mathcal{F}_3 = \sigma(A, B, C)$$

Find

$$E(Y \mid \mathcal{F}_1) = ???? \quad E(Y \mid \mathcal{F}_2) = ???? \quad E(Y \mid \mathcal{F}_3) = ????$$

■ **Example 3.3** Toss a coin three times. Then we can construct a probability space $(\Omega, \mathcal{F}, P)$, where

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} =: \{A_1, A_2, \ldots.., A_7, A_8\}$$
$$\mathcal{F} = \text{all possible subsets of } \Omega, \quad |\mathcal{F}| = 2^8 = 256$$

Let X_i =be the number of H 's in the i-th toss, $i = 1,2,3$. Namely, $X_i = I\{$ the ith toss is a Head$\}$ for $i = 1,2,3$. So X_i 's are r.v.s taking two possible values 0, and 1. Now

$$\begin{aligned}
\sigma(X_1) &= \sigma(\{X_1 = 0\}, \{X_1 = 1\}) = \sigma\left(\{X_1 = 1\}^c, \{X_1 = 1\}\right) = \sigma(\{X_1 = 1\}) \\
&= \sigma(\{HHH, HHT, HTH, HTT\}) = \sigma(\{A_1, A_2, A_3, A_4\}) \\
&= \{\emptyset, \Omega, \{A_1, A_2, A_3, A_4\}, \{A_5, A_6, A_7, A_8\}\}. \quad 2^2 = 4 \text{ elements}
\end{aligned}$$

similarly,

$$\begin{aligned}
\sigma(X_2) &= \sigma(\{X_2 = 1\}) \\
&= \{\emptyset, \Omega, \{A_1, A_2, A_5, A_6\}, \{A_3, A_4, A_7, A_8\}\}. \\
\sigma(X_3) &= \sigma(\{X_3 = 1\}) \\
&= \{\emptyset, \Omega, \{A_1, A_3, A_5, A_7\}, \{A_2, A_4, A_6, A_8\}\} \\
\sigma(X_1, X_2) &= \sigma(\sigma(X_1), \sigma(X_2)) = \sigma(\{X_1 = 1\}, \{X_2 = 1\}) \\
&= \sigma(\{A_1, A_2, A_3, A_2\}, \{A_1, A_2, A_2\} \\
&= 2^{2^2} = 16 \text{ elements} \\
\sigma(X_1, X_3) &= 2^{2^2} = 16 \text{ elements} \\
\sigma(X_2, X_3) &= 2^{2^2} = 16 \text{ elements} \\
\sigma(X_1, X_2, X_3) &= \sigma(\{X_1 = 1\}, \{X_2 = 1\}, \{X_3 = 1\}) \\
&= 2^{2^3} = 256 \text{ elements.}
\end{aligned}$$

In fact, we can show that $\mathcal{F} = \sigma(X_1, X_2, X_3)$

The $\sigma-$ field generated by a continuous r.v.

For discrete r.v. Y, we have seen that $\sigma(Y)$ can be generated from $Y = y_i$ for all y_i 's. However, if Y is a continuous r.v., the σ -algebra generated by the sets $\{\omega : Y(\omega) = y\}, y \in R$, is, from a mathematical point of view, not rich enough. For instance, sets of the form $\{\omega : a < Y(\omega) \le b\}$ do not belong to such a σ -algebra. However, it turns out that it is enough to have a σ -algebra generated by all (open, half open, or closed) intervals.

Theorem 3.1.9 Let $X_1, \ldots, X_n$ be r.v. 's on $(\Omega, \mathcal{A})$. A real function Y on Ω is $\sigma(X_1, \ldots, X_n)$ -measurable (or a r.v. on the σ -algebra) iff Y has the form $f(X_1, \ldots, X_n)$, where f is a Borel function on $\mathcal{R}^n$.

Proof. See Chow and Teicher, page 17 . ■

3.1.3 Distributions and induced distribution functions

Case I: Random variables

Associated with every r.v. is a probability measure on $\mathcal{R}$.

Theorem 3.1.10 Let $(\Omega,\mathcal{F},\nu)$ be a measure space and f be a measurable function from $(\Omega,\mathcal{F})$ to $(\Lambda,\mathcal{G})$. **The induced measure by** f, denoted by $\nu\circ f^{-1}$, is a measure on $\mathcal{G}$ defined as

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

It is usually easier to deal with $\nu\circ f^{-1}$ than to deal with ν since $(\Lambda,\mathcal{G})$ is usually simpler than $(\Omega,\mathcal{F})$. Furthermore, subsets not in $\sigma(f)$ are not involved in the definition of $\nu\circ f^{-1}$. As we discussed earlier, in some cases we are only interested in subsets in $\sigma(f)$

If $\nu=P$ is a probability measure and X is a random variable or a random vector, then $P\circ X^{-1}$ is called the law or the distribution of X and is denoted by P_X. For any c.d.f. or joint c.d.f. F, there exists at least one random variable or vector (usually there are many) defined on some probability space for which $F_X=F$.

Theorem 3.1.11 A r.v. X on $(\Omega,\mathcal{A},P)$ induces another probability space $(\mathcal{R},\mathcal{B},P_X)$ through

$$\forall B\in\mathcal{B}:\quad P_X(B)=P\left(X^{-1}(B)\right)=P(X\in B)$$

Proof. Clearly, $P_X(B)$ is nonnegative and $P_X(\mathcal{R})=P(\Omega)=1$. It is also σ -additive as

$$P_X\left(\sum_i B_i\right)=P\left(X^{-1}\left(\sum_i B_i\right)\right)=\sum_i P\left(X^{-1}(B_i)\right)=\sum_i P_X(B_i)$$

■

Definition 3.1.9 X is a r.v.

1. The distribution of X :

$$P_X(B)=P\left(X^{-1}(B)\right)=P(X\in B),\quad B\in\mathcal{B}$$

2. The distribution function of X :

$$F_X(x)=P_X((-\infty,x])=P(X\leq x)$$

Definition 3.1.10 — Identically Distributed,equal almost surely.

1. Given two r.v.'s X and Y on $(\Omega_1,\mathcal{A}_1,P_1)$ and $(\Omega_2,\mathcal{A}_2,P_2)$ respectively, X and Y are **identically distributed** (i.d.) if $F_X=F_Y$, denoted by $X=_d Y$.
2. X and Y on $(\Omega,\mathcal{A},P)$ are **equal almost surely** (a.s.) if $P(X=Y)=1$, denoted by

$X =_{a.s.} Y$

R

1. X and Y in Definition (i) does not have to be in the same probability space while in Definition (ii) they must be
2. $X =_d Y$ is a much weaker concept than $X =$ a.s. Y. The former may not have much bearing on the latter. One could have $X =_d Y$ even if $P(X \neq Y) = 1$. For example, $X \sim N(0,1)$ and $Y = -X$. Clearly $P(X = Y) = P(X = 0) = 0$, but $X =_d Y$

Definition 3.1.11 — Discrete random variable. A r.v. X on $(\Omega, \mathcal{A}, P)$ is discrete if $\exists$ a countable subset C of $\mathcal{R}$ s.t. $P(X \in C) = 1$

Theorem 3.1.12 X is discrete $\Longleftrightarrow F_X$ is discrete.

Proof. " $\Longleftarrow$ If F_X is discrete, then $F_X(x) = \sum_{i=1}^{\infty} p_i \delta_{a_i}(x)$, where $\sum_{i=1}^{\infty} p_i = 1$. Let $C = \{a_i, i \geq 1\}$, then we have

$$P_X(C) = P_X\left(\cup_{i=1}^{\infty} \{a_i\}\right) = \sum_{i=1}^{\infty} P_X(\{a_i\}) = \sum_{i=1}^{\infty} [F_X(a_i) - F_X(a_i -)] = \sum_{i=1}^{\infty} p_i = 1$$

That is, X is a discrete r.v.

$\Longrightarrow$ " If X is a discrete r.v., then $P_X(C) = 1$, where $C = \{a_i, i \geq 1\}$. Then

$$\begin{aligned} F_X(x) &= P(X \in [-\infty, x]) = P(X \in [-\infty, x] \cap C) = \textstyle\sum_{a_i \in [-\infty, x]} P_X(\{a_i\}) \\ &= \textstyle\sum_{i=1}^{\infty} P_X(\{a_i\}) I\{a_i \leq x\} = \sum_{i=1}^{\infty} P_X(\{a_i\}) \delta_{a_i}(x) := \sum_{i=1}^{\infty} p_i \delta_{a_i}(x) \end{aligned}$$

That is, F_X is discrete. ■

Case II: Random vectors

Definition 3.1.12 $X = (X_1, \ldots, X_n)$ is a random vector.

1. The distribution of X :
$$P_X(B) = P\left(X^{-1}(B)\right) = P(X \in B), \quad B \in \mathcal{B}^n$$
2. The (joint) distribution function of X :
$$F_X(x) = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$$

Marginal d.f. can be uncovered from the joint d.f. while the reverse is not true (need independence).

Theorem 3.1.13 — Marginal d.f.. $X = (X_1, \ldots, X_n)$ is a random vector. Then for any subset $I = \{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$ $m \le n$, we have

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

Proof. As $x_j \nearrow \infty, j \notin I$, we have

$$\{X_1 \le x_1, \cdots, X_n \le x_n\} = \bigcap_{i=1}^{n} \{X_i \le x_i\} \nearrow_{x_j \nearrow \infty} \bigcap_{j=1}^{m} \left\{X_{i_j} \le x_{i_j}\right\}$$

■

Definition 3.1.13 A random vector X on $(\Omega, \mathcal{A}, P)$ is discrete if $\exists$ a countable subset C of $\mathcal{R}^n$ s.t. $P(X \in C) = 1$

Theorem 3.1.14 A random vector $X = (X_1, \ldots, X_n)$ is discrete iff X_k is discrete for each $1 \le k \le n$

Proof. Let C be a countable subset of $\mathcal{R}^n$, define $C_i = \{x_i : (x_1, \ldots, x_n) \in C\}$, which is clearly countable. Then we have $C = C_1 \times \ldots \times C_n$. Then

$$\begin{aligned} P(X \in C) = 1, \Longleftrightarrow \quad & P(\cap_1^n \{X_i \in C_i\}) = 1 \\ \Longleftrightarrow \quad & P(\cup_1^n \{X_i \notin C_i\}) = 0 \Longleftrightarrow P(\{X_i \notin C_i\}) = 0, 1 \le i \le n \\ \Longleftrightarrow \quad & P(\{X_i \in C_i\}) = 1, 1 \le i \le n, \quad \Longleftrightarrow \quad X_i \text{ is discrete for all } i \end{aligned}$$

■

Definition 3.1.14 — Distribution. The measure μ_X is called the **distribution** of the random variable.

$$\mu_X(B) = \mathbb{P}\{X \in B\} = \mathbb{P}\left[X^{-1}(B)\right]$$

1. If μ_X gives measure one to a countable set of reals, then X is called a **discrete random variable**.
2. If μ_X gives zero measure to every singleton set, and hence to every countable set, X is called a **continuous random variable**.
3. Every random variable can be written as a sum of a discrete random variable and a continuous random variable.
4. All random variables defined on a discrete probability space are discrete.

Definition 3.1.15 — Distribution Function. The distribution μ_X is often given in terms of the distribution function defined by

$$F_X(x) = \mathbb{P}\{X \le x\} = \mu_X(-\infty, x]$$

Note that $F = F_X$ is a Stieltjes measure function and satisfies the following:

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

Theorem 3.1.15 If f is continuous at t, then the fundamental theorem of calculus implies that

$$f(x) = F'(x)$$

A density f satisfies

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Conversely, any nonnegative function that integrates to one is the density of a random variable.

■ **Example 3.4 — Every probability measure on $\mathbb{R}$ is the distribution of a random variable.** Let μ be any probability measure on $(\mathbb{R}, \mathcal{B})$. Consider the trivial random variable

$$X = x$$

defined on the probability space $(\mathbb{R}, \mathcal{B}, \mu)$. Then X is a random variable and $\mu_X = \mu$. Hence every probability measure on $\mathbb{R}$ is the distribution of a random variable.

Theorem 3.1.16 — Composition. If X is a random variable and

$$g : (\mathbb{R}, \mathcal{B}) \to \mathbb{R}$$

is a Borel measurable function, then $Y = g(X)$ is also a random variable.

■ **Example 3.5 — Continuous random variable with no density.** Recall that the Cantor function is a continuous function $F : [0,1] \to [0,1]$ with $F(0) = 0, F(1) = 1$ and such that $F'(x) = 0$ for all $x \in [0,1] \setminus A$ where A denotes the "middle thirds" Cantor set. Extend F to $\mathbb{R}$ by setting $F(x) = 0$ for $x \leq 0$ and $F(x) = 1$ for $x \geq 1$. Then F is a distribution function. A random variable with this distribution function is continuous, since F is continuous. However, such a random variable has no density.

3.1.4 Generating random variables with prescribed distributions

Given a r.v. X on a probability space $(\Omega, \mathcal{A}, P)$, we can induce a d.f. of X by $F_X(x) = P(X \leq x)$. Now given given a d.f. F, can we find a r.v. Y such that Y has d.f. F? The answer is affirmative.

Definition 3.1.16 — Quantile function. The inverse of a d.f. F, or **quantile function**

associated with F, is defined by

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

Theorem 3.1.17 — Properties of quantile function. Let $F^{-1}(u)=\inf\{t:F(t)\geq u\},\forall u\in(0,1)$. Then

1. $F^{-1}(u)$ is non-decreasing and left continuous.
2. $F^{-1}(F(x))\leq x,\forall x\in R$
3. $F\left(F^{-1}(u)\right)\geq u,\forall u\in(0,1)$
4. $F^{-1}(u)\leq t\Longleftrightarrow u\leq F(t)$
5. If F is continuous, then $F\left(F^{-1}(u)\right)=u$ for $u\in(0,1)$

Proof. 1. Monotonicity. Let $u_1<u_2\Longrightarrow$ if $F(t)\geq u_2$, then $F(t)\geq u_2>u_1$

$$\begin{aligned}&\Longrightarrow\{t:F(t)\geq u_1\}\supset\{t:F(t)\geq u_2\}\\&\Longrightarrow F^{-1}(u_1)=\inf\{t:F(t)\geq u_1\}\leq\inf\{t:F(t)\geq u_2\}=F^{-1}(u_2)\end{aligned}$$

Left-continuity. Let $u_n\nearrow u\Longrightarrow F^{-1}(u_n)\leq F^{-1}(u)$ and $F^{-1}(u_n)\nearrow b$, say. (Monotonicity shown above) $\Longrightarrow F^{-1}(u_n)\leq b\leq F^{-1}(u)$ for all $n\Longrightarrow u_n\leq F(b)$ from (4) (proof given below). $\Longrightarrow F(b)\geq\lim_n u_n=u\Longrightarrow b\in\{t:F(t)\geq u\}$, hence, $b\geq F^{-1}(u)$ $\Longrightarrow\lim_n F^{-1}(u_n)=b=F^{-1}(u)$

2. $F^{-1}(F(x))=\inf\{t:F(t)\geq F(x)\}\leq x$ as $x\in\{t:F(t)\geq F(x)\}$
3. First we claim that the set $\{t:F(t)\geq u\}$ must be a half interval in the form of

$$\{t:F(t)\geq u\}=(a,\infty),\quad\text{or}\quad[a,\infty)\tag{3.1}$$

(In fact, we will see that it CAN not be of the first type (a,∞).) To see this, suppose that $r\in\{t:F(t)\geq u\}$ (implying that $F(r)\geq u$), then for any $r'>r$ we must have $F(r')\geq F(r)\geq u$, implying that $r'\in\{t:F(t)\geq u\}$. From (3.1), we obtain that $F^{-1}(u)=\inf\{t:F(t)\geq u\}=a$. Now since $a+n^{-1}\in\{t:F(t)\geq u\}$, we have $F\left(a+n^{-1}\right)\geq u$. Letting $n\to\infty$ and using the right continuity of F, we have

$$F\left(F^{-1}(u)\right)=F(a)=\lim_{n\to\infty}F\left(a+n^{-1}\right)\geq u$$

R The last line in the above proof of (3) states that $a=F^{-1}(u)\in\{t:F(t)\geq u\}$. So we must have

$$\{t:F(t)\geq u\}=[a,\infty)=\left[F^{-1}(u),\infty\right)$$

Therefore, we have

$$F^{-1}(u)=\inf\{t:F(t)\geq u\}=\min\{t:F(t)\geq u\}$$

4. " $\Longleftarrow$ ". If $F(t) \geq u \Longrightarrow t \in \{t : F(t) \geq u\} \Longrightarrow t \geq \inf\{t : F(t) \geq u\} = F^{-1}(u)$
   "$\Longrightarrow$ ". If $F^{-1}(u) \leq t$, then since F is non-decreasing, we have $F(t) \geq F\left(F^{-1}(u)\right) \geq u$ from (3)
5. From (3), $F\left(F^{-1}(u)\right) \geq u$. We now show that the equality must hold. If not, denoting $a = F^{-1}(u)$, we would have $F(a) > u$. By continuity and monotonicity of F, we can find $\delta > 0$ such that $F(a-\delta) \geq u$, implying that $a - \delta \in \{t : F(t) \geq u\}$. We then would have $a - \delta \geq \inf\{t : F(t) \geq u\} = a$, implying that $\delta \leq 0$. Contradiction.

■

Theorem 3.1.18 — Quantile transformation.. F is a d.f. on $R, U \sim$ Uniform (0,1) . Then

$$X := F^{-1}(U) \sim F$$

Proof. Since $F^{-1}(u)$ is non-decreasing (monotone), it is Borel measurable. Thus, $X := F^{-1}(U)$ is a r.v. Furthermore, from the last theorem (4), we have

$$P(X \leq x) = P\left(F^{-1}(U) \leq x\right) = P(U \leq F(x)) = F(x)$$

■

R Theorem 3.1.18 is the basis for many simulation procedures. First one needs to generate "random number" U from the uniform d.f. Uniform$(0,1)$, and then apply Theorem 3.1.18. Feasibility of this technique, of course, depends on either having F^{-1} available in closed form, or being able to approximate it numerically, or using some other techniques.

■ **Example 3.6** Think about how to generate r.v.'s from

1. Exponential(1),
2. $N(0,1)$
3. Cauchy,
4. the empirical d.f. of observations $x_1, \ldots, x_n : F_n(x) = n^{-1}\sum_{i=1}^{n} I\{X_i \leq x\}$.

Proof. Let $U, V \sim$Uniform$(0,1)$, and they are independent.

1. Take $X = -\ln U$. (Using Theorem 3.1.18, we get $X = -\ln(1-U)$.)
2. Let $\theta = 2\pi U, R = \sqrt{-2\ln V}$. Then $X = R\cos\theta$ and $Y = R\sin\theta$ are both $N(0,1)$ and independent. (Note $F(t) = \int_{-\infty}^{t} e^{-x^2/2}dx/\sqrt{2\pi}$. So $F^{-1}(u)$ has no close form.)
3. Let $X, Y \sim N(0,1)$ and they are independent, then $Z = X/Y$ is Cauchy. (There is no close form for $F^{-1}(u)$.)

4. We note that $F_n(x)$ is a step function. Let $x_{(1)},\dots,x_{(n)}$ are ordered observations of $x_1,\dots,x_n$ in ascending order. For any $U \sim$ Uniform $(0,1)$, we can easily see that

$$\begin{aligned} F^{-1}(U) &= x_{(1)}, \quad 0 < U \le \frac{1}{n} \\ &= x_{(2)}, \quad \frac{1}{n} < U \le \frac{2}{n} \\ &= \dots\dots \\ &= x_{(n)}, \quad \frac{n-1}{n} < U \le 1 \end{aligned}$$

Clearly, the range of $F^{-1}(U)$ is $\left\{x_{(1)},\dots,x_{(n)}\right\}$ and

$$P\left(F^{-1}(U) = x_{(k)}\right) = P\left(\frac{k-1}{n} < U \le \frac{k}{n}\right) = \frac{1}{n}, \quad k = 1,2,\dots,n$$

That is, to draw a random number from $F_n(x)$, we can simply draw from the observations with equal probability.

■

Theorem 3.1.19 If a r.v. X has a continuous d.f. F, then $F(X) \sim$ Uniform(0,1)

.

Proof. Proof. For $x \in (0,1)$

$$\begin{aligned} P(F(X) \le x) =& 1 - P(F(X) > x) \\ =& 1 - P(F(X) \ge x) \quad \text{(from Lemma 3.1 below)} \\ =& 1 - P\left(X \ge F^{-1}(x)\right) \quad (\text{ from } (4) \text{ of Theorem 3.1.17}) \\ =& 1 - P\left(X > F^{-1}(x)\right) \quad (X \text{ is a continuous r.v. }) \\ =& P\left(X \le F^{-1}(x)\right) \\ =& F\left(F^{-1}(x)\right) = x \quad \left(\text{from } (5) \text{ of Theorem 3.1.17}\right) \end{aligned}$$

■

Lemma 3.1 If a r.v. X has a continuous d.f. F, then $F(X)$ is also a continuous r.v.

Proof. The range of $F(X)$ is $[0,1]$. We need to show that $P(F(X) = t) = 0$ for any $t \in (0,1)$. For any $t \in (0,1)$, since F is continuous, from the proof of (3) and the ensuing remark in the proof of Theorem 3.1.17, we see that

$$\{x : F(x) \ge t\} = [\inf\{x : F(x) \ge t\}, \infty) = [\min\{x : F(x) \ge t\}, \infty) = [\inf\{x : F(x) = t\}, \infty)$$

Similarly, one can show that

$$\{x : F(x) \le t\} = (-\infty, \sup\{x : F(x) \le t\}] = (-\infty, \max\{x : F(x) \le t\}] = (-\infty, \sup\{x : F(x) = t\}]$$

Therefore,

$$\{x : F(x) = t\} = \{x : F(x) \ge t\} \bigcap \{x : F(x) \le t\} = [\inf\{x : F(x) = t\}, \sup\{x : F(x) = t\}] := [a, b]$$

It then follows that $F(a) = F(b) = t$. Therefore,

$$\begin{aligned} P(F(X) = t) &= P(X \in \{x : F(x) = t\}) = P(X \in [a, b]) \\ &= P(a \le X \le b) = P(a < X \le b) = F(b) - F(a) = t - t0 \end{aligned}$$

■

R Lemma 3.1 states that if X is continuous with d.f. F, so is $F(X)$. What about $G(X)$ for any continuous function G ?

3.2 Expectation

3.2.1 Expectation for simple r.v.'s

Definition 3.2.1 The expectation of a simple r.v. $X = \sum_1^n a_i I_{A_i}$ with $\sum_1^n A_i = \Omega, A_i \in \mathcal{A}$ is

$$EX = \sum_1^n a_i P(A_i)$$

In other words, given a (measurable) partition $\{A_i\}$ of Ω, we assign value a_i to A_i, then the EX is simply the weighted average of a_i with weights being the probability of A_i

R

1. Since probability measure is a finite measure, we won't encounter the situation $\infty - \infty$. For general integrations w.r.t. some measure (not necessarily finite), we can define integrals for nonnegative measurable function, in order to avoid the possibility of $\infty - \infty$.
2. We have mentioned in the last chapter that if X is a r.v. on $(\Omega, \mathcal{A}, P)$, then we require $P(|X| = \infty) = 0$. For a simple r.v. $X = \sum_1^n a_i I_{A_i}$, this implies that $|a_i| < \infty$ if $P(A_i) > 0$ Clearly, this implies that $E|X| < \infty$.

Lemma 3.2 EX is well defined in the sense: if $\sum_{i=1}^n a_i I_{A_i} = \sum_{j=1}^m b_j I_{B_j}$ with $\Omega = \sum_{i=1}^n A_i = \sum_{j=1}^m B_j$, then

$$\sum_{i=1}^n a_i P(A_i) = \sum_{j=1}^m b_j P(B_j)$$

Proof. If $\sum_{i=1}^{n} a_i I_{A_i} = \sum_{j=1}^{m} b_j I_{B_j}$ with $\Omega = \sum_{i=1}^{n} A_i = \sum_{j=1}^{m} B_j$, then we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_i I_{A_i \cap B_j} = \sum_{i=1}^{n}\sum_{j=1}^{m} b_j I_{A_i \cap B_j}$$

which follows since

$$LHS = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i I_{A_i \cap B_j} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i I_{A_i} I_{B_j} = \left(\sum_{i=1}^{n} a_i I_{A_i}\right)\left(\sum_{j=1}^{m} I_{B_j}\right) = \sum_{i=1}^{n} a_i I_{A_i}$$
$$RHS = \sum_{i=1}^{n}\sum_{j=1}^{m} b_j I_{A_i \cap B_j} = \sum_{j=1}^{m}\sum_{i=1}^{n} b_j I_{A_i} I_{B_j} = \left(\sum_{j=1}^{m} b_j I_{B_j}\right)\left(\sum_{i=1}^{n} I_{A_i}\right) = \sum_{j=1}^{m} b_j I_{B_j}$$

So $a_i = b_j$ if $A_j \cap B_j \neq \emptyset$. Therefore,

$$\sum_{i=1}^{n} a_i P(A_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i P(A_i \cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_j P(A_i \cap B_j) = \sum_{j=1}^{m} b_j P(B_j)$$

■

Theorem 3.2.1 — Properties of simple r.v.. X, Y are simple $r.v.$'s.

1. $EC = C$
2. $EI_A = P(A)$
3. $aX + bY$ is simple and $E(aX + bY) = aEX + bEY$. (Linearity)
4. $X \geq 0 \Longrightarrow EX \geq 0$. (Nonnegativity)
5. $X \geq Y \Longrightarrow EX \geq EY$. (Monotonicity)

Proof.
1. $EC = E(CI_\Omega) = CP(\Omega) = C$
2. $EI_A = E(1 \times I_A + 0 \times I_{A^c}) = P(A)$
3. X, Y are simple, so $X = \sum_{i=1}^{n} a_i I_{A_i}$ and $Y = \sum_{j=1}^{m} b_j I_{B_j}$. So

$$aX + bY = \sum_{i=1}^{n}\sum_{j=1}^{m} (aa_i + bb_j) I_{A_i \cap B_j} \tag{3.2}$$

is a simple r.v. and therefore,

$$E(aX + bY) = \sum_{i=1}^{n}\sum_{j=1}^{m} (aa_i + bb_j) P(A_i \cap B_j)$$
$$= a\sum_{i=1}^{n} a_i P(A_i) + b\sum_{j=1}^{m} b_j P(B_j) = aEX + bEY$$

4. $X = \sum_{i=1}^{n} a_i I_{A_i} \geq 0, \Longrightarrow a_i \geq 0$ for all i with $P(A_i) > 0$,

$$\Longrightarrow EX = \sum_{i=1}^{n} a_i P(A_i) \geq 0$$

5. Since $X \geq Y$, from (3) and (4), we have $0 \leq E(X-Y) = EX - EY$

■

R

1. Probability is continuous. However, **expectation is NOT continuous**. e.g. Take $\Omega = [0,1], \mathcal{A} = \mathcal{B} \cap [0,1]$ and $P =$ Lebesgue measure (i.e. a Uniform distribution on [0,1]). Define

$$X_n = nI_{(0,1/n)}$$

Then $\forall \omega \in \Omega : X_n(\omega) \to 0 \equiv X$, but for each n, we have $EX_n = nP((0,1/n)) = 1 \neq 0 = EX$

2. Expectation is continuous for the following two special cases:
   (a) Monotone convergent sequence (see the Monotone Convergent Theorem later)
   (b) Dominated convergence sequence (see the Dominated Convergence Theorem later

   We shall see that the example in (1) does not belong to these cases. To illustrate that (i) is not satisfied, we see that even though $X_n \to 0$, the convergence is NOT monotone.

Expectation for nonnegative r.v.'s

Definition 3.2.2 Any nonnegative r.v. X can be approximated by an increasing sequence of simple r.v.'s whose expectations are already defined.

Theorem 3.2.2 Given a nonnegative r.v. X, there exists simple r.v.'s $0 \leq X_1 \leq X_2 \leq \dots$ such that $X_n(\omega) \nearrow X(\omega)$ for every ω

Proof. This has been shown in the last chapter. We just give a brief summary here. For each n, let

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right\} + nI_{\{X \geq n\}} = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n$$

Since X is a r.v. $\Longrightarrow I_{\{a \leq X < b\}}$ are indicator r.v.'s $\Longrightarrow X_n$ are simple r.v.'s. On the measurable set (event) $\{(k-1)/2^n \leq X < k/2^n\}, X_n$ equals the left endpoint $(k-1)/2^n$, which ensures

1. $X_n(\omega) \leq X_{n+1}(\omega) \quad \forall \omega \in \Omega$ (Monotonicity)
2. $0 \leq X(\omega) - X_n(\omega) \leq 1/2^n \quad \forall n \geq 1$

   Then $X_n(\omega)$ converges (from (a)) to $X(\omega)$ (from (b)).

■

Definition 3.2.3 $X \geq 0$ is a r.v. on $(\Omega, \mathcal{A}, P)$.

1. The expectation of X is $EX = \lim_{n\to\infty} EX_n \leq \infty$, where X_n 's are simple, nonnegative, and $X_n \nearrow X$.
2. The expectation of X over the event $A \in \mathcal{A}$ is $E_A X := E(XI_A)$.
3. If $Y \leq 0$ is a r.v. on $(\Omega, \mathcal{A}, P)$, define $EY := -E(-Y)$.

R

1. The definition forces monotone continuity from below for simple $r.v.$'s $\{X_n\}$: $X_n \nearrow X$ implies $EX_n \nearrow EX$
2. The following notation is often used (for $X \geq 0$ or otherwise):

$$EX = \int_\Omega X(\omega)P(d\omega) = \int_\Omega XdP = \int XdP$$
$$E_A X = \int_A X(\omega)P(d\omega) = \int_A XdP$$

Theorem 3.2.3 — Uniqueness of simple r.v.. $X_n \geq 0, Y_n \geq 0$ are simple r.v.'s and $X_n \nearrow X, Y_n \nearrow X$. Then

1. $\lim_{n\to\infty} EX_n$ and $\lim_{n\to\infty} EY_n$ exist.
2. $\lim_{n\to\infty} EX_n = \lim_{n\to\infty} EY_n$

Proof. 1. $X_n \nearrow X \Longrightarrow EX_n \leq EX_{n+1} \Longrightarrow \lim_{n\to\infty} EX_n$ exists (maybe ∞). Similarly, $\lim_{n\to\infty} EY_n$ exists.

2. Given $k \geq 1, Y_k$ is simple and $Y_k \leq X$. If we can show that

$$\forall k \geq 1: \quad EY_k \leq \lim_n EX_n \tag{3.3}$$

which implies that $\lim_k EY_k \leq \lim_n EX_n$. By symmetry, $\lim_k EY_k \geq \lim_n EX_n$. This proves the theorem.

Proof of (3.3). Fix $\epsilon > 0$ and $k \geq 1$. For each n, define $A_n = \{X_n > Y_k - \epsilon\}$, which is measurable. Note that

(a) $A_n \nearrow \Omega$ since $X_n \nearrow X$

(b) $X_n \geq (Y_k - \epsilon) I_{A_n}$ (note that both sides are still simple $r.v.$'s.)

Therefore,

$$\begin{aligned} EX_n &\geq E\left[(Y_k - \epsilon) I_{A_n}\right] = E\left(Y_k\left[1 - I_{A_n^c}\right]\right) - \epsilon P(A_n) \\ &\geq EY_k - \left(\max_{\omega\in\Omega} Y_k(\omega)\right) P(A_n^c) - \epsilon P(A_n) \end{aligned}$$

(Y_k is simple, so "max" exists and is finite) (the second term is taking Y_k out of expectation and taking I_{A^c} into expectation) ($A_n \nearrow \Omega$ and $A_n^C \searrow \emptyset$)

$$\to \quad EY_k - \epsilon, \qquad \text{as } n \to \infty$$

Since ϵ is arbitrary, we have $\lim_n EX_n \geq EY_k$. The proof is finished.

■

ALGEBRAIC OPERATIONS

Theorem 3.2.4 $X, Y \geq 0$ are r.v.'s.

1. $X, Y \geq 0$, and $ab \geq 0$, then $E(aX + bY) = aEX + bEY$. (Linearity)("$ab \geq 0$" $\Longleftrightarrow$ "a and b have the same signs", so to avoid the possibility $\infty - \infty$.)
2. $X \geq 0, \Longrightarrow EX \geq 0$. (Nonnegativity)
3. $X \geq Y \geq 0 \Longrightarrow EX \geq EY \geq 0$. (Monotonicity)

Proof. Let $X_n \geq 0, Y_n \geq 0$ be simple r.v.'s and $X_n \nearrow X, Y_n \nearrow X$

1. Clearly, $aX_n + bY_n$ are also simple from (3.2) . Thus

$$
\begin{aligned}
E(aX + bY) &= \lim_n E\left(aX_n + bY_n\right) = \lim_n \left(aEX_n + bEY_n\right) \\
&= a \lim_n EX_n + b \lim_n EY_n = aEX + bEY
\end{aligned}
$$

2. $EX = \lim_n EX_n \geq 0$
3. $X - Y \geq 0$ and $(1) \Longrightarrow EX = E(Y + [X - Y]) = EY + E(X - Y)$. $\Longrightarrow EX - EY = E(X - Y) \geq 0$ from (2)

■

Theorem 3.2.5 $X \geq 0$. Then $EX = 0 \Longleftrightarrow X = a.s.0$

Proof. 1. "$\Longrightarrow$". Suppose that $X =_{a.s.} 0$ is not true, i.e., $P(X = 0) < 1 \Longrightarrow F_X(0) = P(X \leq 0) = P(X = 0) < 1$, $(X \geq 0) \Longrightarrow \exists \epsilon > 0$ such that $F_X(\epsilon) < 1$. Since F_X is right continuous, Therefore,

$$
\begin{aligned}
EX &= E\left(XI_{\{X > \epsilon\}}\right) + E\left(XI_{\{X \leq \epsilon\}}\right) \\
&\geq \epsilon E\left(I_{\{X > \epsilon\}}\right) = \epsilon P(X > \epsilon) = \epsilon(1 - F_X(\epsilon)) > 0
\end{aligned}
$$

Contradiction. This proves that $X = a.s.0$

2. " $\Longleftarrow$ Noting $X \geq 0$, we have

$$
\begin{aligned}
0 \leq EX &= E(XI\{X = 0\}) + E\left(XI_{\{X \neq 0\}}\right) = E(XI\{X > 0\}) \\
&\leq \left(\sup_{\omega \in \Omega} X(\omega)\right) EI_{\{X > 0\}} = \left(\sup_{\omega \in \Omega} X(\omega)\right) P(X > 0) \leq \infty \times 0 =^{?} 0
\end{aligned}
$$

■

Theorem 3.2.6 1. If $X > 0$ a.s., then $EX > 0$. (i.e., strict inequality is preserved by expectation)
2. If $EX \geq 0$, then $P(X \geq 0) > 0$

Proof. 1. $X > 0$ implies that $EX \geq 0$. We show that EX can not be equal to 0. Otherwise, if $EX = 0$, coupled with the assumption $X > 0 \geq 0$, we get $X = 0$ a.s. from the last theorem. This contradicts with $X > 0$ a.s. Thus we have shown that $EX > 0$.

2. If the conclusion is not true, then $P(X \geq 0) = 0$, i.e., $P(X < 0) = 1$, or $X < 0$ a.s. Similarly to the proof of (i), we can show that $EX < 0$. This contradicts with the assumption $EX \geq 0$.

■

LIMITING OPERATIONS

Theorem 3.2.7 — Fatou's lemma. 1. Suppose that $X_n \geq Y$ a.s. for all n and some Y with $E|Y| < \infty$.

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} EX_n. \quad \Longleftrightarrow \quad E\left(\underline{\lim}_n X_n\right) \leq \underline{\lim}_n EX_n$$

(Typically, one can choose $Y = 0$ in practice.)

2. Suppose that $X_n \leq Y$ a.s. for all n and some Y with $E|Y| < \infty$.

$$E\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} EX_n. \quad \Longleftrightarrow \quad E(\overline{\lim}_n X_n) \geq \overline{\lim}_n EX_n$$

Proof. 1. Without loss of generality, we can choose $Y = 0$, otherwise, we could consider $X_n - Y$ instead. For each m, let $Z_m := \inf_{k\geq m} X_k$, so that $Z_m \nearrow \underline{\lim}_n X_n$. Recall ($\underline{\lim}_n = \liminf_{n\to\infty} X_n = \sup_{m\geq 1} \inf_{k\geq m} X_k \equiv \sup_{m\geq 1} Z_m = \lim_m Z_m$)
Note $X_n \geq 0 \Longrightarrow \underline{\lim}\, X_n \geq 0 \Longrightarrow$ there exists a sequence of simple r.v.'s $\{Y_k, k \geq 1\}$ such that $0 \leq Y_k \nearrow \underline{\lim}\, X_n$.
Fix $k \geq 1$ and $\epsilon > 0$, and for each m, define $A_m = \{Z_m > Y_k - \epsilon\}$, which is measurable. Since $X_m \geq Z_m \nearrow \underline{\lim}_n X_n \geq Y_k > Y_k - \epsilon$, then $A_m \nearrow \Omega$. Thus,

$$\forall m: \quad X_m \geq (Y_k - \epsilon)\, I_{A_m}$$

Therefore,

$$\begin{aligned} EX_m &\geq E[(Y_k - \epsilon)\, I_{A_m}] = E\left(Y_k\left[1 - I_{A_m^c}\right]\right) - \epsilon P(A_m) \\ &\geq EY_k - \left(\max_{\omega\in\Omega} Y_k(\omega)\right) P(A_m^c) - \epsilon P(A_m) \end{aligned}$$

(Y_k is simple, so "max" exists and is finite)

$$\to \quad EY_k - \epsilon, \quad \text{as } m \to \infty$$

Since ϵ is arbitrary, we have $\underline{\lim}_m EX_m \geq EY_k$. Letting $k \to \infty$, we get $\underline{\lim}_m EX_m \geq \lim_{k\to\infty} EY_k = E(\underline{\lim}_n X_n)$

2. The proof is similar to that for lim inf and hence omitted.

■

Proposition 3.2.8 Fatou's Lemma $\Longleftrightarrow$ the Monotone Convergence Theorem (MCT). (The MCT will be introduced later.)

1. Fatou's Lemma implies the Monotone Convergence Theorem (MCT). This can be seen from the proof of MCT later.
2. The Monotone Convergence Theorem implies Fatou's Lemma.

Proof. Assume that $X_n \geq 0$. First we have

$$\inf_{k \geq n} X_k \leq X_n$$

Taking expectation on both sides, we get

$$E \inf_{k \geq n} X_k \leq EX_k$$

Note that $\inf_{k \geq n} X_k \nearrow \liminf_n X_n$, and hence the LHS is non-decreasing. Take $\liminf_n$ on both sides and then applying the MCT, we get

$$\liminf_n E \inf_{k \geq n} X_k = \lim_n E \inf_{k \geq n} X_k = E \lim_n \inf_{k \geq n} X_k \leq E \liminf_n X_n$$

1. The inequality in Fatou's lemma can be strict. For example, in the last example, we have

   $$\lim_n X_n(\omega) = 0 \text{ and } EX_n = 1. \text{ So } E \lim X_n = 0 \neq \underline{\lim} \lim EX_n = 1$$

2. Suppose that the sequence of functions is a sequence of random variables, $X_1, X_2, \ldots$, with $X_n Y$ (almost surely) for some Y such that $E(|Y|) < \infty$. Then by Fatou's lemma

   $$E\left(\liminf_{n\to\infty} X_n Y\right) \leq \liminf_{n\to\infty} EX_n Y$$

   It is often useful to assume that Y is a constant. For example, taking $Y = 0$ it becomes clear that Fatou's lemma can be applied to any sequence of non-negative random variables.

■

Theorem 3.2.9 — Monotone convergence theorem (MCT)–monotone continuity. Let $X, X_1, X_2, \ldots$ be nonnegative r.v.'s. Then,

1. $X_n(\omega) \nearrow X(\omega) \Longrightarrow EX_n \nearrow EX$
2. $X_n(\omega) \searrow X(\omega)$ and $EX_m < \infty$ for some $m \geq 1, \Rightarrow EX_n \searrow EX$

Proof.

1. $X_n \leq X \Longrightarrow EX_n \leq EX \Longrightarrow \limsup_n EX_n \leq EX$. So

   $$EX = E \lim_n X_n = E \liminf X_n \leq \liminf_n EX_n \leq \limsup_n EX_n \leq EX$$

   That is, $\liminf_n EX_n = \limsup_n EX_n = \lim_n EX_n = EX$

2. WLOG, assume that $EX_1 < \infty$. Then, $X_n(\omega) \searrow X(\omega) \Longrightarrow X_1(\omega) - X_n(\omega) \nearrow X_1(\omega) -$

   $$X(\omega) \Longrightarrow EX_1 - EX_n \nearrow EX_1 - EX \Longrightarrow EX_n \searrow EX$$

■

Corollary 3.2.10 If $Y_k \geq 0$ and $\sum_{k=1}^{\infty} Y_k(\omega) < \infty$, then

$$E\left(\sum_{k=1}^{\infty} Y_k\right) = \sum_{k=1}^{\infty} EY_k$$

Proof. Letting $X_n = \sum_1^n Y_k$ and $X = \sum_1^{\infty} Y_k$. Clearly, $X_n(\omega) \nearrow X(\omega)$ for every ω from the assumptions. Applying the Monotone Convergence Theorem: $RHS = \lim_n EX_n = EX = LHS$ ■

Theorem 3.2.11 EXPECTATIONS EXTEND PROBABILITIES

1. $A_n \nearrow A \Longrightarrow I_{A_n} \nearrow I_A \Longrightarrow P(A_n) = EI_{A_n} \nearrow EI_A = P(A)$. Similarly, $A_n \searrow A \Longrightarrow P(A_n) \searrow P(A)$. So the Monotone Convergence Theorem implies monotone continuity of P.
2. Take $Y_k = I_{A_k}$ where A_k are disjoint. Then

$$P\left(\sum_1^{\infty} A_k\right) = EI_{\sum_1^{\infty} A_k} = \sum_1^{\infty} EI_{A_k} = \sum_1^{\infty} P(A_k)$$

   the σ -additivity of P.
3. Take $X_n = I_{A_n}$, then since $\liminf_n I_{A_n} = I_{\liminf_n A_n}$, Fatou's lemma implies that

$$P\left(\liminf_n A_n\right) = EI_{\liminf_n A_n} = E\liminf_n I_{A_n} \leq \liminf_n P(A_n)$$

   which was used to prove continuity of P.
4. Measures are always continuous from below, and conditionally continuous from above. Expectations are always continuous from below, and conditionally continuous from above.
5. Finite measures (including probability measure P) is always continuous, but E is not.

To understand why this is so, recall a question in Homework 2, stated below for easy reference:

Theorem 3.2.12 If $(\Omega, \mathcal{A}, \mu)$ is a measure space, and $A_n \in \mathcal{A}$

1. Prove that $\mu(\underline{\lim}\, A_n) \leq \underline{\lim}\, \mu(A_n)$. Analogously, if $\mu(\cup_{i=n}^{\infty} A_i) < \infty$ for some $n \geq 1$, then $\mu\left(\overline{\lim} A_n\right) \geq \overline{\lim}\ \mu(A_n)$.
2. If μ is a finite measure, and $\underline{\lim}\, A_n = \overline{\lim}\, A_n = A$, (i.e. $\lim A_n = A$), then $\lim \mu(A_n) = \mu(A)$

Proof. 1. $\mu(\underline{\lim}\, A_n) = \mu(\lim_n \cap_{i=n}^{\infty} A_i) = \lim_n \mu(\cap_{i=n}^{\infty} A_i) \leq \lim_n \inf_{i \geq n} \mu(A_i) = \underline{\lim}_n\, \mu(A_n)$ and $\mu\left(\overline{\lim} A_n\right) = \mu(\lim_n \cup_{i=n}^{\infty} A_i) = \lim_n \mu(\cup_{i=n}^{\infty} A_i) \geq \overline{\lim}_n\, \mu(A_n)$ where in the very last inequality, we used the assumption $\mu(\cup_{i=n}^{\infty} A_i) < \infty$ for some $n \geq 1$.

2. Since μ is finite, we apply (i) to get

$$\overline{\lim_n}\,\mu(A_n) \le \mu\left(\overline{\lim} A_n\right) = \mu(A) = \mu\left(\underline{\lim}_n A_n\right) \le \underline{\lim}_n\,\mu(A_n)$$

■

It is clear that the critical steps for continuity of finite measure μ is the two inequalities:

$$\mu\left(\underline{\lim}_n A_n\right) \le \underline{\lim}_n\,\mu(A_n), \quad \mu\left(\overline{\lim} A_n\right) \ge \overline{\lim_n}\,\mu(A_n)$$

If we can show that

$$E\left(\underline{\lim}_n A_n\right) \le \underline{\lim}_n\,E(X_n), \quad E\left(\overline{\lim} X_n\right) \ge \overline{\lim_n}\,E(X_n)$$

then E would also be continuous. However, the second inequality is not always true, e.g., $X_n = nI_{(0,1/n)} \to 0$, but $EX_n = 1 \ne 0$

Expectation for general r.v.'s

Definition 3.2.4 Recall $X^+ = \max\{X,0\} = XI_{\{X\ge 0\}}$, $\quad X^- = \max\{-X,0\} = -XI_{\{X\le 0\}}$, and

$$\begin{aligned} X =& XI_{\{X\ge 0\}} + XI_{\{X\le 0\}} = XI_{\{X\ge 0\}} - \left(-XI_{\{X\le 0\}}\right) = X^+ - X^- \\ |X| =& |X|I_{\{X\ge 0\}} + |X|I_{\{X\le 0\}} = XI_{\{X\ge 0\}} + \left(-XI_{\{X\le 0\}}\right) = X^+ + X^- \end{aligned}$$

So EX can be defined by EX^+ and EX^-. But we need to be careful to avoid $EX^+ - EX^- = \infty - \infty$

Definition 3.2.5 Let X be a r.v. on $(\Omega, \mathcal{A}, P)$

1. For general r.v. X, if either $EX^+ < \infty$ or $EX^- < \infty$ (but not both), then the expectation of X is

$$EX = EX^+ - EX^-$$

In this case, the expectation of X is said to exist and $EX \in [-\infty, \infty]$

2. If $EX^+ = EX^- = \infty$, then EX is not defined.
3. X is **integrable** if $E|X| < \infty$ (L^1).
4. If X is integrable and $A \in \mathcal{A}$, the expectation of X over A is

$$E_A X = E(XI_A)$$

R

1. Let $L^1 = \{X : E|X| < \infty\}$. This defines the class of all integrable r.v.'s on $(\Omega, \mathcal{A}, P)$
2. X is integrable, i.e. $X \in L^1 \iff E|X| < \infty \iff EX^+ < \infty$ and $EX^- < \infty$. So EX in (a) is well defined since we don't have $\infty - \infty$.)
3. First XI_A is a r.v. Secondly, $E|X| < \infty \implies E|XI_A| < \infty$. That is, $E(XI_A)$ in (d) is well defined.)

ALGEBRAIC OPERATIONS

Theorem 3.2.13
1. $X, Y \in L^1$, and $a, b \in \mathcal{R}$, then $aX + bY \in L^1$, and
$$E(aX + bY) = aEX + bEY. \quad \text{(Linearity)}$$
2. For $X \in L^1, |EX| \leq E|X|$
3. $X, Y \in L^1$, and $X \leq Y$, then $EX \leq EY$. (Monotonicity)

Proof. 1. By the triangle inequality: $|aX + bY| \leq |a||X| + |b||Y|$. Note that each term is nonnegative, so their expectations are well defined and monotonicity of expectations for nonnegative r.v.'s implies

$$E|aX + bY| \leq |a|E|X| + |b|E|Y| < \infty$$

Thus, $aX + bY \in L^1$. To complete the proof, it remains to show that
(a) $E(X + Y) = EX + EY$
(b) $E(aX) = aEX$
Proof of (a).

Proof. We next show that if $Z_1, Z_2 \geq 0$, then $E(Z_1 - Z_2) = EZ_1 - EZ_2$. If suffices to show

$$\begin{aligned} E(Z_1 - Z_2) &\equiv E(Z_1 - Z_2)^+ - E(Z_1 - Z_2)^- \\ &\equiv E(Z_1 - Z_2) I_{\{Z_1 \geq Z_2\}} - E(Z_2 - Z_1) I_{\{Z_1 \leq Z_2\}} \\ &= EZ_1 - EZ_2 \end{aligned}$$

or equivalently we need to show that

$$EZ_1 + E(Z_2 - Z_1) I_{\{Z_1 \leq Z_2\}} = EZ_2 + E(Z_1 - Z_2) I_{\{Z_1 \geq Z_2\}}$$

which follows easily from the identity

$$Z_1 + (Z_2 - Z_1) I_{\{Z_1 \leq Z_2\}} = Z_2 + (Z_1 - Z_2) I_{\{Z_1 \geq Z_2\}}$$

(Note LHS $=$ RHS under 3 cases: $Z_1 < Z_2, Z_1 = Z_2, Z_1 > Z_2$).
It then follows that

$$\begin{aligned} E(X + Y) &= E\left(X^+ - X^- + Y^+ - Y^-\right) \\ &= E\left(\left(X^+ + Y^+\right) - \left(X^- + Y^-\right)\right) \\ &= E\left(X^+ + Y^+\right) - E\left(X^- + Y^-\right) \\ &= \left(EX^+ + EY^+\right) - \left(EX^- + EY^-\right) \\ &= EX + EY \end{aligned}$$

■

Proof of (b).

Proof. If $a \geq 0$, then

$$
\begin{aligned}
E(aX) &= E\left[(aX)^+\right] - E\left[(aX)^-)\right] = E\left[aX^+\right] - E\left[aX^-\right] \\
&= a\left(EX^+ - EX^-\right) = aEX
\end{aligned}
$$

If $a < 0$, then

$$(aX)^+ = aXI_{\{aX\geq 0\}} = (-a)(-X)I_{\{X\leq 0\}} = (-a)X^-$$

Similarly, $(aX)^- = (-a)X^+$. Therefore,

$$
\begin{aligned}
E(aX) &= E\left[(aX)^+\right] - E\left[(aX)^-)\right] = E\left[(-a)X^-\right] - E\left[(-a)X^+\right] \\
&= -a\left(EX^- - EX^+\right) = aEX
\end{aligned}
$$

■

2. $|EX| = |EX^+ - EX^-| \leq EX^+ + EX^- = E|X|$
3. $X - Y \in L^1$ and from (1), we get $EY - EX = E(Y - X) = E(Y - X)^+ \geq 0$

■

Theorem 3.2.14 $X =$ a.s. $Y \Longrightarrow EX = EY$.

Proof. Let $Z = X - Y$, then

$$1 = P(X = Y) = P(X - Y = 0) = P(Z = 0) = P(|Z| = 0) = P\left(Z^+ = 0, Z^- = 0\right)$$

Therefore, $P\left(Z^+ = 0\right) = 1$ and $P\left(Z^- = 0\right) = 1$ (here we used the fact that $P(A \cap B) = 1$ implies $P(A) = P(B) = 1$.) Thus, $EZ^+ = EZ^- = 0$. It follows that

$$E(X - Y) = EZ^+ - EZ^- = 0$$

Finally

$$EX = E[Y + (X - Y)] = EY + E(X - Y) = EY$$

■

LIMITING OPERATIONS

Below, we shall write $X_n \to X$ to denote $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$

Theorem 3.2.15 — Dominated Convergence Theorem. If $X_n \to X$ a.s., $|X_n| < Y$ for all n, and $EY < \infty$, then

$$\lim_n EX_n = EX\left(= E\lim_n X_n\right)$$

Proof. First, from the assumptions, we have that $|X| \leq |X_n - X| + |X_n| \leq C + Y$ a.s. Thus, $E|X| \leq C + EY < \infty$. Hence, $X \in L^1$.

Secondly, $|X_n| \leq Y \iff -Y \leq X_n \leq Y$ for each n.

Now $Y - X_n \geq 0 \Longrightarrow$

$$\begin{aligned} EY - EX = E(Y-X) &= E\left[\liminf_n (Y - X_n)\right] \leq \liminf_n E\left(Y - X_n)\right] \\ &= EY + \liminf_n [-EX_n] = EY - \limsup_n EX_n \end{aligned}$$

Thus, $\limsup_n EX_n \leq EX$.

On the other hand, $Y + X_n \geq 0 \Longrightarrow$

$$\begin{aligned} EY + EX = E(Y+X) &= E\left[\liminf_n (Y + X_n)\right] \leq \liminf_n E\left(Y + X_n)\right] \\ &= EY + \liminf_n EX_n \end{aligned}$$

Thus, $\liminf_n EX_n \geq EX$. Combining the two, we get $\limsup_n EX_n \leq EX \leq \liminf_n EX_n$. This implies that $\lim_n EX_n = EX$. ■

R

1. Going back to our earlier example, if Y dominates $X_n = nI_{(0,1/n)}$ for each n, then we must have $Y \geq \sum_1^m nI_{(1/(n+1),1/n)}$ for all $m \geq 1$, thus letting $m \to \infty$, we get $Y \geq \sum_1^\infty nI_{(1/(n+1),1/n)}$. This implies that

$$\begin{aligned} EY \geq E\sum_1^\infty nI_{(1/(n+1),1/n)} &= E\lim_{m\to\infty}\sum_1^m nI_{(1/(n+1),1/n)} \\ &= \lim_{m\to\infty} E\sum_1^m nI_{(1/(n+1),1/n)} \quad \text{(by Monotone convergence theorem)} \\ &= \sum_1^\infty nEI_{(1/(n+1),1/n)} = \sum_1^\infty nP((1/(n+1),1/n)) \\ &= \sum_1^\infty n(1/n - 1/(n+1)) = \sum_1^\infty 1/(n+1) \\ &= \infty \end{aligned}$$

In other words, if there exists a r.v. $Y \geq 0$ which dominates all $X_n (n \geq 1)$, then $EY = \infty$. Thus, the condition in Dominated Convergence Theorem is violated.

2. If $A_n \to A$, then $I_{A_n} \to I_A$. Taking $X_n = I_{A_n}$, and $Y = 1$, we get

$$P(A_n) = EI_{A_n} \to EI_A = P(A)$$

the full-fledged continuity of P.

Theorem 3.2.16 — Summary. Assume that $X, Y, X_1, \ldots, X_n$ below are all r.v.'s on $(\Omega, \mathcal{A}, \mu)$

1. (Absolute integrability). EX is finite if and only if $E|X|$ is finite.
2. (Linearity). If the RHS below is meaningful, namely not $+\infty - \infty$ or $-\infty - \infty$, (e.g. if $X, Y \geq 0$ and $a, b \geq 0$, or if $X, Y \in L^1$ and $a, b \in \mathcal{R}$), then
$$E(aX + bY) = aEX + bEY$$
3. ($\sigma-$ additivity over sets). If $A = \sum_{i=1}^{\infty} A_i$
$$E_A X = \sum_{i=1}^{\infty} E_{A_i} X$$
4. (Positivity). If $X \geq 0$ a.s., then
$$EX \geq 0$$
5. (Monotonicity). If $X_1 \leq X \leq X_2$ a.s., then
$$EX_1 \leq EX \leq EX_2$$
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. (Modulus inequality).
$$|EX| \leq E|X|$$
8. (Fatou's Lemma). If $X_n \geq 0$ a.s., then
$$E\left(\liminf_n X_n\right) \leq \liminf_n EX_n$$
9. (Monotone Convergence Theorem). If $0 \leq X_n \nearrow X$, then
$$\lim_n EX_n = EX = E \lim_n X_n$$
10. (Dominated Convergence Theorem). If $X_n \to X$ a.s., $|X_n| < Y$ a.s. for all n, and $EY < \infty$, then
$$\lim_n EX_n = EX = E \lim_n X_n$$

11. (Integration term by term). If $\sum_{i=1}^{\infty} E|X_n| < \infty$, then

$$\sum_{i=1}^{\infty} |X_n| < \infty, a.s$$

so that $\sum_{i=1}^{\infty} X_n$ converges a.s., and

$$E\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} EX_n$$

R

1. Note that when EX is well defined, i.e., $EX = EX^+ - EX^-$ is not of the form $\infty - \infty$ or $-\infty + \infty$, then EX has only three possibilities:

$$\infty, \quad -\infty, \quad \text{finite.}$$

2. If a relation involving r.v.'s is true a.s. , then we can simply pretend it is true everywhere when we calculate integration or expectations. This is justified by the following theorem:

Theorem 3.2.17 If $P(A) = 1$, then $E_\Omega X = E_A X$

Proof. $E_\Omega X = E_A X + E_{A^c} X$, but $0 \le |E_{A^c} X| \le E_{A^c}|X| \le \infty \times P(A^c) = \infty \times 0 = 0$ ■

3. If the mean of a nonnegative r.v. is bounded, the r.v. is bounded a.s.

Theorem 3.2.18 If $E|X| < \infty$, then $|X| < \infty$ a.s.

$$(\text{Namely, if } P(|X| = \infty) > 0, \text{ then } |EX| = \infty)$$

Proof. If the theorem is not true, then we have $P(|X| = \infty) > 0$. As a result,

$$E|X| = E|X|I_{|X|<\infty} + E|X|I_{|X|=\infty} \ge E|X|I_{|X|=\infty} = \infty \times P(|X| = \infty) = \infty$$

This implies that $|EX| = \infty$, which contradicts with our assumption. ■

Proof. PROOFS OF THE SUMMARY

1. (Absolute integrability).

$$EX < \infty \iff EX^+ < \infty \text{ and } EX^- < \infty \iff E|X| < \infty$$

2. (Linearity). There are four possible cases:
   (a) $X, Y \in L^1$. This has been shown.
   (b) $X \in L^1$ and $Y \notin L^1$. Similar to (3) below.
   (c) $X \notin L^1$ and $Y \in L^1$. WLOG, assume that $EX^+ = \infty$, thus $EX^- < \infty$ $EY^+ < \infty$, and $EY^- < \infty$. Also assume that $a, b \ge 0$. So RHS $= \infty$ and

$$LHS = E\left[\left(aX^+ + bY^+\right) - \left(aX^- + bY^-\right)\right]$$

$$\text{But } E\left[\left(aX^+ + bY^+\right) \ge E\left(aX^+\right) = aEX^+ = \infty, E\left(aX^- + bY^-\right) < \infty\right.$$

therefore, LHS $= \infty =$ RHS

(d) $X \notin L^1$ and $Y \notin L^1$. WLOG, assume that $EX^+ = \infty$, and $EY^+ = \infty$ Thus $EX^- < \infty$, and $EY^- < \infty$. In order for the RHS to be meaningful, we must have $ab \geq 0$. Assume that $a, b > 0$. Then the rest of the proof is similar to that in (3) above.

3. (σ -additivity over sets). If $X \geq 0$, then applying the Monotone Convergence Theorem, we get

$$E_A X = EXI_A = E\left(\sum_{i=1}^{\infty} I_{A_i} X\right) = \sum_{i=1}^{\infty} E\left(XI_{A_i}\right) = \sum_{i=1}^{\infty} E_{A_i} X$$

For general r.v. $X = X^+ - X^-$.

4. (Positivity). If $X \geq 0$ a.s., then $EX \geq 0$.
5. (Monotonicity). It suffices to show that $X \leq Y$ a.s. implies $EX \leq EY$. If both EX and EY are ∞(or $-\infty$), then the inequality clearly holds. Otherwise, since $Y - X \geq 0$ a.s., by the linearity and positivity of expectations, the proof follows from

$$EY - EX = E(Y - X) \geq 0$$

6. (Mean value theorem). $a \leq X \leq b$ a.s. implies $aI_A \leq XI_A \leq bI_A$ a.s. The proof is done by taking expectation on both sides.
7. (Integration term by term). By the Monotone Convergence Theorem, we get

$$E\left(\sum_{i=1}^{\infty} |X_n|\right) = \sum_{i=1}^{\infty} E|X_n| < \infty$$

By Theorem 3.2.18, we get $\sum_{i=1}^{\infty} |X_n| < \infty$ a.s., which in turn implies that $\sum_{i=1}^{\infty} X_n$ converges a.s. Finally, apply the Dominated Convergence Theorem, we have $E\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} EX_n$

■

3.2.2 Integration

Just need to be familiar with the language of integral.

Definition 3.2.6 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 $f = \sum_1^n a_i I_{A_i}$ with $a_i \geq 0$

$$\int f d\mu = \sum_1^n a_i \mu(A_i)$$

2. If $f \geq 0$, define
$$\int f d\mu = \lim_n \int f_n d\mu$$
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 $\int f^- d\mu < \infty$. If $\int f^+ d\mu = \infty$ and $\int f^- d\mu = \infty$, then $\int f d\mu$ 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
$$\int_A f d\mu = \int f I_A d\mu = \int f(\omega) I_A(\omega) \mu(d\omega)$$

R An equivalent definition is (see Shao, 1998, or Durrett)
$$\int f d\mu := \sup \left\{ \int \psi d\mu : \psi \in S_f \right\}$$
where $S_f =$ the collection of all nonnegative simple functions ψ such that $\psi(\omega) \leq f(\omega)$ for any $\omega \in \Omega$

R By taking $f = I_A$, then it follows from (a) and (e) that
$$\mu(A) = \int I_A d\mu = \int_A d\mu$$

Proposition 3.2.19 — Some properties of integrals. Assume that all functions below are Borel measurable on $(\Omega, \mathcal{A}, \mu)$.

1. (Absolute integrability). $\int f$ is finite if and only if $\int |f|$ is finite.
2. (Leariity). If the RHS below is meaningful, namely not $+\infty - \infty$ or $-\infty - \infty$, (e.g. if $f, g \geq 0$ and $a, b \geq 0$, or if $f, g \in L^1$ and $a, b \in \mathcal{R}$) , then
$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$$
3. (σ -additivity over sets). If $A = \sum_{i=1}^\infty A_i$
$$\int_A f d\mu = \sum_{i=1}^\infty \int_{A_i} f d\mu$$

4. (Positivity). If $f \geq 0$ a.e., then
$$\int f d\mu \geq 0$$
5. (Monotonicity). If $f_1 \leq f \leq f_2$ a.e., then
$$\int f_1 \leq \int f \leq \int f_2$$
6. (Mean value theorem). If $a \leq f \leq b$ a.e. on $A \in \mathcal{A}$, then
$$a\mu(A) \leq \int_A f d\mu \leq b\mu(A)$$
7. (Modulus inequality).
$$\left|\int f\right| \leq \int |f|$$
8. (Fatou's Lemma). If $f_n \geq 0$ a.e., then
$$\int \liminf_n f_n \leq \liminf_n \int f_n$$
9. (Monotone Convergence Theorem). If $0 \leq f_n \nearrow f$, then
$$\lim_n \int f_n = \int f = \int \lim_n f_n$$
10. (Dominated Convergence Theorem). If $f_n \to f$ a.e., $|f_n| < g$ a.e. for all n, and $\int g < \infty$, then
$$\lim_n \int f_n = \int f = \int \lim_n f_n$$
11. (Integration term by term). If $\sum_{i=1}^{\infty} \int |f_n| < \infty$, then
$$\sum_{i=1}^{\infty} |f_n| < \infty, \text{ a.e.}$$
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 3.2.7 — Lebesgue-Stieltjes integral and Lebesgue integral. f is a Borel measurable function on $(\Omega, \mathcal{A}, \mu)$.

1. In the case of $(\Omega, \mathcal{A}, \mu) = (\mathcal{R}, \mathcal{B}, \mu)$, if we write $x = \omega \in \mathcal{R}$, then
$$\int f(\omega)\mu(d\omega) = \int f(x)\mu(dx)$$
is just the ordinary **Lebesgue-Stieltjes integral** of f w.r.t. μ
2. In the case of $(\Omega, \mathcal{A}, \mu) = (\mathcal{R}, \mathcal{B}, \lambda)$, where λ is the Lebesgue measure, then
$$\int f(x)\lambda(dx) = \int f(x)dx$$

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

Definition 3.2.8 Let F be a nondecreasing and right-continuous function on $\mathcal{R}$ (i.e. L-S measure function). It is known that there exists a unique measure μ on the measurable space $(\mathcal{R}, \mathcal{B})$ such that

$$\mu((a,b]) = F(b) - F(a) \tag{3.4}$$

1. Then we can define

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

to be the L-S integral of f w.r.t. F.
2. In the special case $F(x) = x$, the unique measure μ determined from (3.4) reduces to the Lebesgue measure λ. As a consequence, the integral in (a) reduces to

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

which is the Lebesgue integral of f.(Note: $\lambda(dx) = dx$.)

R X is a r.v. on $(\Omega, \mathcal{A}, P)$.
1. Expectations are special cases of integrals:

$$EX = \int X(\omega)P(d\omega) = \int X(t)dF_X(t), \quad \text{where } F_X(t) = P(X \le t)$$

2. For L-S integral, $\int_{(a,b]} f d\mu$ may not be the same as $\int_{(a,b)} f d\mu$ ect. So we don't write $\int_a^b f d\mu$. For L-integral, it is OK to do this.
3. We know (or do we really?) the relationship between Riemann integral and Lebesgue integral. Now there exist some similar relationship between Riemann-Stieltjes (R -S) integral and Lebesgue-Stieltjes (L-S) integral. Now let us point out one more difference between them. Well, roughly speaking, the R − S integral $\int f(x)dg(x)$ is defined as the limit of R − S sums as the mesh goes to 0. We must require that f and g do not have discontinuities at the same point in order for the limit to exist. On the other hand, no such requirement is necessary for the L-S integral.

Some special cases

Consider the L-S integral of the form

$$\int_B f dG$$

where B is a Borel set in R.

1. G **is a discrete (i.e., a step) function.** When G is a step function, it will have at most countably many jumps $\{x_1, x_2, \ldots\}$, where $\Delta G(x_n) = G(x_n) - G(x_n-) > 0$.

The measure μ will be discrete with positive measure at each of the points $x_1, x_2, \ldots$, so

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

In particular, defining $\int_s^t := \int_{(s,t]}$, we have

$$\int_s^t f dG := \int_{(s,t]} f dG = \sum_{n: s < x_n \le t} f(x_n) \Delta G(x_n)$$

In this case, the L-S integral is a short and convenient notation for a sum with a finite or countably infinite number of terms.

2. G **is an absolutely continuous function.** When 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 $G : [a, \infty) \to R$, is right-continuous on $[a, \infty)$, and is differentiable on R except at points in a countably infinite set $\{x_1, x_2, \ldots\}$, where each $x_i > a$. In most applications, one can take $a = -\infty$ or $a = 0$. Then, G can be written

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

In this case,

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

4. G **is a right-continuous function of bounded variation.** When 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 nondecreasing and right-continuous functions. In this case,

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

Using this decomposition, it is easy to show that all of the standard results from Lebesgue integration hold for $\int_B f dG$

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:

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

If either F or G have discontinuities, then a bit more care must be taken with the $L-S$ integral.

Theorem 3.2.20 Let F,G be right-continuous functions of bounded variation. Then

$$\begin{aligned} F(t)G(t)-F(s)G(s) &= \int_{(s,t]} F(x-)dG(x) + \int_{(s,t]} G(x)dF(x) \\ &= \int_{(s,t]} F(x)dG(x) + \int_{(s,t]} G(x-)dF(x) \end{aligned} \tag{3.5}$$

and

$$F(t)G(t)-F(s)G(s) = \int_{(s,t]} F(x-)dG(x) + \int_{(s,t]} G(x-)dF(x) + \sum_{n:s<x_n\le t} \Delta F(x)\Delta G(x) \tag{3.6}$$

Proof. WLOG, we will only prove the case for $s=0$. Note that, whenever $s\le t$, by definition,

$$\begin{aligned} F(s)-F(0) = \mu((0,s]) = \textstyle\int_{(0,s]} d\mu = \int_{(0,s]} dF(x) = \int_{(0,t]} I_{\{0<x\le s\}} dF(x) \\ F(s-)-F(0) = \mu((0,s)) = \textstyle\int_{(0,s)} d\mu = \int_{(0,s)} dF(x) = \int_{(0,t]} I_{\{0<x<s\}} dF(x) \end{aligned}$$

By Fubini's Theorem, (although we did not discuss this theorem in this course yet!)

$$\begin{aligned} & \{F(t)-F(0)\}\{G(t)-G(0)\} \\ =& F(t)G(t)-F(0)G(t)-G(0)F(t)-F(0)G(0) \\ =& \int_{(0,t]} dF(x) \times \int_{(0,t]} dG(x) \\ =& \int_{(0,t]}\int_{(0,t]} dF(x)dG(y) \\ =& \int_{(0,t]}\int_{(0,t]} I_{\{0<x<y\}} dF(x)dG(y) + \int_{(0,t]}\int_{(0,t]} I_{\{0<y\le x\}} dF(x)dG(y) \\ =& \int_{(0,t]} \left(\int_{(0,t]} I_{\{0<x<y\}} dF(x)\right) dG(y) + \int_{(0,t]} \left(\int_{(0,t]} I_{\{0<y\le x\}} dG(y)\right) dF(x) \\ =& \int_{(0,t]} \{F(y-)-F(0)\} dG(y) + \int_{(0,t]} \{G(x)-G(0)\} dF(x) \\ =& \int_{(0,t]} F(y-)dG(y) - F(0)\int_{(0,t]} dG(y) + \int_{(0,t]} G(x)dF(x) - G(0)\int_{(0,t]} dF(x) \\ =& \int_{(0,t]} F(y-)dG(y) + \int_{(0,t]} G(x)dF(x) - F(0)G(t) - G(0)F(t) \end{aligned}$$

Algebraic simplification leads to (3.5). To prove (3.6), in view of (3.5), it suffices to show that

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

or equivalently $(\Delta G(x) = G(x) - G(x-) > 0.)$,

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

Let $F(x) = F_c(x) + F_d(x) = F_c(x) + \sum_{s<x} \Delta F(s)$, where F_c and F_d are the continuous and discrete parts of F, respectively. Then,

$$\begin{aligned}\int_{(s,t]} \Delta G(x) dF(x) &= \int_{(s,t]} \Delta G(x) dF_c(x) + \int_{(s,t]} \Delta G(x) dF_d(x) \\ &= 0 + \sum_{0<x\le t} \Delta G(x) \Delta F(x)\end{aligned}$$

where we used

$$\int_{(s,t]} \Delta G(x) dF_c(x) = 0 \tag{3.7}$$

which is true since F_c is continuous (and hence assigns measure 0 to any single point) and $\Delta G = 0$ at all but at most a countable number of points in $(0,t]$. See the remark below for a more detailed proof. ■

R Here we provide a more detailed justification of (3.7). Denote D_G and C_G to be the set of all jump points and continuity points of G, respectively. Then,

$$\begin{aligned}\int_B \Delta G(x) dF_c(x) &= \int_{B\cap C_G} \Delta G(x) dF_c(x) + \int_{B\cap D_G} \Delta G(x) dF_c(x) \\ &= \int_{B\cap C_G} 0 dF_c(x) + \int_{B\cap D_G} \Delta G(x) dF_c(x) \\ &= \int_{B\cap C_G} 0 dF_c(x) + \int_{B\cap D_G} \Delta G(x) d\mu_c(x) \\ &\qquad (\mu_c \text{ is the corresponding measure to } F_c) \\ &= 0\end{aligned}$$

Since $B \cap C_G$ is a countable set, and

$$\mu_c(B\cap C_G) = \sum_{x_n \in B\cap D_G} \mu_c(\{x_n\}) = \sum_{x_n\in B\cap D_G} [F_c(x_n) - F_c(x_n-)] = 0$$

R In the integration by parts formula,

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

We notice that $F(x-)$ and $G(x-)$ are continuous functions. Therefore, there is no common jump points between $F(x-)$ and $G(x)$, and also between $F(x)$ and $G(x-)$. All common jump points have been absorbed by the last term.

Theorem 3.2.21 — Measure-theoretic and probabilistic languages.

Measure	Probability
Integral	Expectation
Measurable set	Event
Measurable function	Random variable
Almost everywhere (a.e.)	Almost surely (a.s.)

3.2.3 How to compute expectation

Theorem 3.2.22 — Change of variable formula. Assume the following holds:

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

Then

$$Eg(X) = \int_{\Omega_0} g(y) P_X(dy)$$

R

1. To explain why this is called "Change of variable formula", we note

$$Eg(X) = \int_{\Omega} g(X(\omega)) dP = \int_{\Omega_0} g(y) dP_X$$

It is as if we changed our variable from ω to $y = X(\omega)$
2. In most cases, we choose

$$(\Omega_0, \mathcal{A}_0, P_X) = (\mathcal{R}^n, \mathcal{B}^n, P_X)$$

where $P_X(B) = P(X \in B)$ for $B \in \mathcal{B}^n$ is the distribution induced by X. In this case, X will be an n -dim random vector, and $g(X)$ will be a random variable as g is Borel function.
3. One practical implication of the above theorem is that we can compute expected values of functions of random variables by performing L-S integrals on the real line $\mathcal{R}$. Below, we shall first change it into L-integrals which equal R-integrals when the latter exists.

Proof. 1. Case I: Indicator functions. If $g = I_B$ with $B \in \mathcal{A}_0$, then the relevant definitions show

$$\begin{aligned} Eg(X) = EI_B(X) = P(X \in B) = P_X(B) &= \int I_B dP_X \\ = \int_{\Omega_0} I_B(y) P_X(dy) &= \int_{\Omega_0} g(y) P_X(dy) \end{aligned}$$

2. Case II: Simple functions. Let $g = \sum_{i=1}^{n} b_i I_{B_i}$ with $B_i \in \mathcal{A}_0$. The linearity of expected value, the result of Case I, and the linearity of integration imply

$$Eg(X) = E\left(\sum_{i=1}^{n} b_i I_{B_i}(X)\right) = \sum_{i=1}^{n} b_i E I_{B_i}(X) = \sum_{i=1}^{n} b_i \int_{\Omega_0} I_{B_i}(y) P_X(dy)$$
$$= \int_{\Omega_0} \left(\sum_{i=1}^{n} b_i I_{B_i}(y)\right) P_X(dy) = \int_{\Omega_0} g(y) P_X(dy)$$

3. Case III: Nonnegative functions. Now if $g \geq 0$, then there exists a sequence of simple functions $\{g_n, n \geq 1\}$ such that $0 \leq g_n \nearrow g$. For instance, we could choose

$$g_n(x) = ([2^n g(x)]/2^n) \wedge n$$

where $[x]$ is the integer part of x. From Case II and the Monotone Convergence Theorem, we get

$$Eg(X) = \lim_n E g_n(X) = \lim_n \int_{\Omega_0} g_n(y) P_X(dy) = \int_{\Omega_0} g(y) P_X(dy)$$

4. Case IV: Integrable functions. For the general case, we can write $g(x) = g(x)^+ - g(x)^-$. The condition that g is integrable guarantees that $Eg(X)^+ < \infty$ and $Eg(X)^- < \infty$. So from Case (III) for nonnegative functions and linearity of expected value and integration

$$Eg(X) = Eg(X)^+ - Eg(X)^-$$
$$= \int_{\Omega_0} g(y)^+ P_X(dy) - \int_{\Omega_0} g(y)^- P_X(dy)$$
$$= \int_{\Omega_0} g(y) P_X(dy)$$

■

Expected values of absolutely continuous r.v.'s

Lemma 3.3 Let X be an absolutely continuous r.v. with density function f, i.e., $F_X(x) = \int_{-\infty}^{x} f(t)dt$. Let P_X be the unique probability measure corresponding to F_X. Then

$$P_X(B) = \int_B f d\lambda = \int_B f(x)dx, \quad \forall B \in \mathcal{B} \tag{3.8}$$

where λ is the L -measure.

Proof. We shall give two different proofs.

1. Method 1. Denote $\mu(B) = \int_B f(x)dx$. It is easy to show that both RHS $\mu(\cdot)$ and LHS $P_X(\cdot)$ of (3.8) are probability measures on $(\mathcal{R}, \mathcal{B})$. It follows from (3.9) below that $P_X|_s = \mu|_s$ on the semialgebra $\mathcal{S} = \{(a,b] : -\infty \leq a \leq b \leq \infty\}$. By the uniqueness of the extensions of measures from a semialgebra $\mathcal{S}$ to the σ -algebra $\mathcal{B} = \sigma(\mathcal{S})$, we prove (3.8)

2. Method 2. Let $\mathcal{A} = \{A \in \mathcal{B} : P_X(A) = \int_A f(x)dx\}$. It is easy to show that $\mathcal{A}$ is a σ -algebra, and $\mathcal{A} \supset \mathcal{S} := \{(-\infty, x], x \in \mathcal{R}\}$. Therefore, $\mathcal{A} \supset \sigma(\mathcal{S}) = \mathcal{B}$. The proof is done. (In fact, we have $\mathcal{A} = \mathcal{B}$.)

■

Theorem 3.2.23 Let X be an absolutely continuous r.v. with density function f, i.e., $F_X(x) = \int_{-\infty}^{x} f(t)dt$. Assume further that g is Borel. Then

$$Eg(X) = \int_{\mathcal{R}} g(x)f(x)dx$$

provided that $\int_{\mathcal{R}} |g(x)|f(x)dx < \infty$ (Thus, L-S integral is changed into L-integral, which equals R -integral if the later exists.)

Proof. Let P_X be the unique probability measure corresponding to F_X such that

$$P_X((a,b]) = F_X(b) - F_X(a) = \int_{(a,b]} f(t)dt \tag{3.9}$$

From the last lemma, we have

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

From Theorem 3.2.22, we have $Eg(X) = \int_{\mathcal{R}} g(x)P_X(dx)$. To complete our proof, we only need to show that

$$\int_{\mathcal{R}} g(x)P_X(dx) = \int_{\mathcal{R}} g(x)f(x)dx \tag{3.10}$$

Proof of (3.10). We shall employ the same method used in the last theorem.

1. Case I: Indicator functions. If $g = I_B$ with $B \in \mathcal{B}$, then

$$LHS = \int I_B(x)P_X(dx) = P(X \in B) = P_X(B) = \int_{\mathcal{R}} I_B(y)f(y)dy = RHS$$

   where the second last equality comes from (3.8) .

2. Case II: Simple functions. If $g = \sum_{i=1}^{n} b_i I_{B_i}$ with $B_i \in \mathcal{B}$. The linearity of expected value, the result of Case I, and the linearity of integration imply

$$\begin{aligned} LHS &= \int \left(\sum_{i=1}^{n} b_i I_{B_i}(x) \right) P_X(dx) = \sum_{i=1}^{n} b_i \int I_{B_i}(x)P_X(dx) \\ &= \sum_{i=1}^{n} b_i \int I_{B_i}(y)f(y)dy = \int_{\mathcal{R}} \sum_{i=1}^{n} b_i I_{B_i}(y)f(y)dy = \int g(y)f(y)dy \\ &= RHS \end{aligned}$$

3. Case III: Nonnegative functions. Now if $g \geq 0$, then there exists a sequence of simple functions $\{g_n, n \geq 1\}$ such that $0 \leq g_n \nearrow g$. From Case II and the Monotone Convergence Theorem, we get

$$LHS = \lim_n \int g_n(y)P_X(dy) = \lim_n \int g_n(y)f(y)dy = \int g(y)f(y)dy = RHS$$

4. Case IV: Integrable functions. For the general case, we can write $g(x) = g(x)^+ - g(x)^-$. The condition implies that g is integrable, i.e., $\mathrm{E}g(X)^+ < \infty$ and $Eg(X)^- < \infty$. So from Case (III) for nonnegative functions and linearity of expected value and integration

$$\begin{aligned} LHS &= \int g(X)^+ P_X(dy) - \int g(X)^- P_X(dy) \\ &= \int g(y)^+ f(y)dy - \int g(y)^- f(y)dy = RHS \end{aligned}$$

This proves (3.10), and hence the theorem.

■

R

1. For an absolutely continuous r.v. X, we have several equivalent expressions:

$$Eg(X) = \int_{\mathcal{R}} g(x)P_X(dx) = \int_{\mathcal{R}} g(x)dF_X(x) = \int_{\mathcal{R}} g(x)f(x)dx$$

2. The last integral in (i) is L-integral, which equals R-integral when the latter exists. This will greatly facilitate our calculations.

Theorem 3.2.24 — Expected values of discrete r.v.'s. Let X be a discrete r.v. taking values $x_1, x_2, \ldots$, with probability $P(X = x_k) = p_k$ for $k \geq 1$, and g be Borel. Then

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

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

Proof. Clearly, $g(X)$ is a r.v. taking values $g(x_1), g(x_2), \ldots$, and we can write

$$g(X) = \sum_{k=1}^{\infty} g(x_k) I_{\{X=x_k\}}$$

1. Case I: If $g(X) \geq 0$, then $g(x_k) \geq 0$ for all $k \geq 1$. Define

$$Z_n = \sum_{k=1}^{n} g(x_k) I_{\{X=x_k\}}$$

a form of truncated r.v. Clearly, given X, we have $0 \leq Z_n \nearrow Z_\infty \equiv g(X)$, and Z_n are simple r.v.'s. Either by the definition of expectation for nonnegative r.v., or simply applying the Monotone Convergence Theorem, we get

$$Eg(X) = \lim_n EZ_n = \lim_n \sum_{k=1}^{n} g(x_k) P(X = x_k) = \sum_{k=1}^{\infty} p_k g(x_k)$$

2. Case II: Let us consider general g. It follows from Case I and the assumption that $E|g(X)| = \sum_{k=1}^{\infty} p_k |g(x_k)| < \infty$. Therefore,

$$\begin{aligned} Eg(X) = Eg(X)^+ - Eg(X)^- &= \sum_{k=1}^{\infty} p_k g(x_k)^+ - \sum_{k=1}^{\infty} p_k g(x_k)^- \\ &= \sum_{k=1}^{\infty} p_k \left(g(x_k)^+ - g(x_k)^- \right) = \sum_{k=1}^{\infty} p_k g(x_k) \end{aligned}$$

■

3.2.4 Relation between expectation and tail probability

Theorem 3.2.25 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 $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$

Proof. By the σ -additivity of expectation over sets, if $A_n = \{n \leq |X| < n+1\}$,

$$E|X| = E_{\sum_{n=0}^{\infty} A_n} |X| = \sum_{n=0}^{\infty} E_{A_n} |X|$$

By the Mean Value Theorem, $nP(A_n) \leq E_{A_n}|X| \leq (n+1)P(A_n)$, thus

$$\sum_{n=1}^{\infty} nP(A_n) = \sum_{n=0}^{\infty} nP(A_n) \leq E|X| \leq \sum_{n=0}^{\infty} (n+1)P(A_n) = 1 + \sum_{n=1}^{\infty} nP(A_n) \tag{3.11}$$

It remains to show that

$$\sum_{n=1}^{\infty} nP(A_n) = \sum_{n=1}^{\infty} P(|X| \geq n) \tag{3.12}$$

Proof. Proof of (3.12). Note that

$$\begin{aligned} \sum_{n=1}^{m} nP(A_n) = \sum_{n=1}^{m} nP(n \leq |X| < n+1) &= \sum_{n=1}^{m} n[P(|X| \geq n) - P(|X| \geq n+1)] \\ &= P(|X| \geq 1) - P(|X| \geq 2) \\ &+ 2P(|X| \geq 2) - 2P(|X| \geq 3) \\ &+ 3P(|X| \geq 3) - 3P(|X| \geq 4) \\ &+ \cdots\cdots\cdots\cdots \\ &+ mP(|X| \geq m) - mP(|X| \geq m+1) \\ &= \sum_{n=1}^{m} P(|X| \geq n) - mP(|X| \geq m+1) \end{aligned}$$

That is, $\sum_{n=1}^{m} P(|X| \geq n) = \sum_{n=1}^{m} nP(A_n) + mP(|X| \geq m+1)$. So

$$\sum_{n=1}^{m} nP(A_n) \leq \sum_{n=1}^{m} P(|X| \geq n) \leq \sum_{n=1}^{m} nP(A_n) + mP(|X| \geq m+1) \tag{3.13}$$

where the last term satisfies

$$mP(|X| \geq m+1) = EmI_{(|X| \geq m+1)} \leq E|X|I_{(|X| \geq m+1)}$$

1. Case I : $E|X| < \infty$. Here, $E|X|I_{(|X| \geq m+1)} = E|X| - E|X|I_{(|X| < m+1)} \to 0$ as $m \to \infty$ by the Monotone Convergence Theorem. Thus, (3.12) is true.
2. Case II: $E|X| = \infty$. Here, from (3.11), we have $\sum_{n=1}^{\infty} nP(A_n) = \infty$. And then from (3.13), we get $\sum_{n=1}^{\infty} P(|X| \geq n) = \infty$. Thus, (3.12) is true as well. Hence, we proved (3.12), and therefore the theorem.

■

R If we assume that $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$, then the proof becomes much simpler, as given below.

Proof. Recall that $[a]$ denotes the integer part of a. Note

$$\begin{aligned}
[|X|] &= \{ \text{ total number of positive integers } \leq |X| \} \\
&= \sum_{k=1}^{\infty} I\{k \leq |X|\} = \sum_{k=1}^{\infty} I\{|X| \geq k\}
\end{aligned}$$

Therefore, $\sum_{k=1}^{\infty} I\{|X| \geq k\} \leq [|X|] \leq |X| \leq [|X|] + 1 \leq \sum_{k=1}^{\infty} I\{|X| \geq k\} + 1$. The theorem follows by taking expectation on both sides, and then apply the rule of integration by parts. For integer-valued r.v.s, we have the following identity (not inequality). ■

Corollary 3.2.26 — Integer-valued r.v.. If X takes only integer values, then

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

Proof. $E|X| := \sum_{n=0}^{\infty} nP(|X| = n) = \sum_{n=1}^{\infty} nP(A_n) = \sum_{n=1}^{\infty} P(|X| \geq n)$ from (3.12) ■

For any non-negative r.v. (discrete or continuous), we have the following integral expression.

Theorem 3.2.27 — For any non-negative r.v.. If $Y \geq 0$, then

$$EY = \int_0^{\infty} P(Y \geq y)dy = \int_0^{\infty} P(Y > y)dy = \int_0^{\infty} [1 - F_Y(y)]\, dy$$

Proof. 1. We'd like to use Corollary 3.2.26, so we first need to turn Y into integer-valued r.v. Let $Y_n = [2^n Y]/2^n$, and $X_n = 2^n Y_n = [2^n Y]$. Then $0 \leq Y_n \nearrow Y$, and by the Monotone Convergence Theorem,

$$EY = \lim_{n \to \infty} EY_n = \lim_{n \to \infty} 2^{-n} EX_n \tag{3.14}$$

Now that X_n is a nonnegative and integer-valued r.v., from the last corollary (3.2.26), we get

$$EX_n = \sum_{j=1}^{\infty} P(X_n \geq j) = \sum_{j=1}^{\infty} P([2^n Y] \geq j) = \sum_{j=1}^{\infty} P(2^n Y \geq j)$$

But

$$\int_0^\infty P(Y \geq y) dy = \sum_{j=0}^{\infty} \int_{j/2^n}^{(j+1)/2^n} P(Y \geq y) dy$$

The above two relations give us

$$\int_0^\infty P(Y \geq y) dy \leq \sum_{j=0}^{\infty} \frac{1}{2^n} P\left(Y \geq \frac{j}{2^n}\right) = \frac{1}{2^n} \sum_{j=0}^{\infty} P(2^n Y \geq j)$$
$$= \frac{1}{2^n} \left(\sum_{j=1}^{\infty} P(2^n Y \geq j) + P(2^n Y \geq 0) \right) = \frac{1}{2^n} (EX_n + 1)$$
$$\int_0^\infty P(Y \geq y) dy \geq \sum_{j=0}^{\infty} \frac{1}{2^n} P\left(Y \geq \frac{j+1}{2^n}\right) = \frac{1}{2^n} \sum_{j=1}^{\infty} P(2^n Y \geq j) = \frac{1}{2^n} EX_n$$

Therefore,

$$EY_n = \frac{1}{2^n} EX_n \leq \int_0^\infty P(Y \geq y) dy \leq \frac{1}{2^n} (EX_n + 1) = EY_n + \frac{1}{2^n}$$

Letting $n \to \infty$ and using (3.14), we prove the theorem.

2. The second inequality follows from the fact $\int_0^\infty P(Y = x) dx = 0$, which will be left as an exercise.

■

For moments of any positive order, the following identity holds.

Corollary 3.2.28 — For non-negative r.v.. If $Y \geq 0$ and $r > 0$, then

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

Proof. Applying the last theorem, we get

$$EY^r = \int_0^\infty P(Y^r \geq x) dx = \int_0^\infty P\left(Y \geq x^{1/r}\right) dx = r \int_0^\infty z^{r-1} P(Y \geq z) dz, \text{ where } z = x^{1/r}$$

The second equality holds due to the property of L-integrals. ■

Corollary 3.2.29 — For general r.v.. If Y is integrable, then

$$EY = EY^+ - EY^- = \int_0^\infty P(Y > x) dx - \int_0^\infty P(Y \leq -x) dx$$

Proof. 1. First, $EY^+ = \int_0^\infty P(Y I\{Y \geq 0\} > x) dx = \int_0^\infty P(Y > x) dx$.

2. Secondly , $EY^- = \int_0^\infty P(-YI\{-Y \geq 0\} > x)dx = \int_0^\infty P(-Y > x)dx = \int_0^\infty P(Y \leq -x)dx$ ■

R The thinner the tails, the higher moments the r.v. will have.

3.2.5 Moments and Moment inequalities

Definition 3.2.9 Let X be a r.v. and $r > 0$,

1. Define
   (a) r th Moment: EX^r
   (b) r th Absolute Moment: $E|X|^r$
   (c) r th Central Moment: $E(X-EX)^r$
   (d) r th Absolute Central Moment: $E|X-EX|^r$
2. L^r Spaces $= \{X : E|X|^r < \infty\}$

Theorem 3.2.30 — 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 of h). Then, for any $a > 0$ and $b > 0$, we have

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

Proof. Try to give a direct proof. However, one picture is worth a thousand words. ■

Theorem 3.2.31 — Holder's inequality. Suppose that $p, q > 1$ satisfy $1/p + 1/q = 1$, and that $E|X|^p < \infty, E|Y|^q < \infty$. Then, $E|XY| < \infty$ and

$$E|XY| \leq [E|X|^p]^{1/p} [E|Y|^q]^{1/q}$$

Proof. Take $h(t) = t^{p-1}$ in Young's inequality, then

$$g(s) = s^{1/(p-1)} = s^{(1/p)/(1-1/p)} = s^{(1-1/q)/(1/q)} = s^{q-1}$$
$$\int_0^a h(t)dt = \frac{a^p}{p}, \quad \int_0^b g(s)ds = \frac{b^q}{q}$$

Therefore,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Setting $a = |X| / [E|X|^p]^{1/p}, b = |Y| / [E|Y|^q]^{1/q}$, we get

$$\frac{|XY|}{[E|X|^p]^{1/p} [E|Y|^q]^{1/q}} \leq \frac{|X|^p}{pE|X|^p} + \frac{|Y|^q}{qE|Y|^q}$$

The result follows by taking expectations on both sides. ■

R (One could try to prove this directly without resort to Young's inequality and use this as a starting point; see the Lemma below).

Lemma 3.4 Let $a,b>0$, and $p,q\geq 1$ such that $1/p+1/q=1$, then

$$\frac{1}{p}a^p+\frac{1}{q}b^q\geq ab$$

with equality if and only if $a^p=b^q$

Proof. Fix b, and consider the function

$$g(a)=\frac{1}{p}a^p+\frac{1}{q}b^q-ab$$

To minimize $g(a)$, differentiate and set equal to 0:

$$\frac{d}{da}g(a)=0,\quad \Longrightarrow a^{p-1}-b=0,\quad \Longrightarrow b=a^{p-1}$$

A check on the second derivative will establish that this is indeed a minimum. The value of the function at the minimum is

$$\frac{1}{p}a^p+\frac{1}{q}\left(a^{p-1}\right)^q-a\left(a^{p-1}\right)=\frac{1}{p}a^p+\frac{1}{q}a^q-a^p=0$$

Hence the minimum is 0 and the inequality is proved. since the minimum is unique, equality holds only if $b=a^{p-1}$, which is equivalent to $a^p=b^q$. ∎

Theorem 3.2.32 — Cauchy-Schwarz inequality.

$$E|XY|\leq\sqrt{[E|X|^2]\,[E|Y|^2]}$$

Proof. Take $p=q=2$ in Holder's inequality. ∎

Theorem 3.2.33 — Lyapunov's inequality.
1. $E(|X|)\leq E\left(|X|^p\right)^{1/p}$ for $p\geq 1$
2. $[E|Z|^r]^{1/r}\leq[E|Z|^s]^{1/s}$, for $0<r\leq s<\infty$

Proof.
1. Take $Y=1$ in Holder's inequality.
2. Take $X=Z^r$ in (1), and also let $rp=s$, we get

$$E|Z|^r\leq[E|Z|^{rp}]^{1/p}=[E|Z|^{rp}]^{r/(rp)}=\left\{[E|Z|^s]^{1/s}\right\}^r$$

(Alternative proof can be given by Jensen's inequality given below).

∎

Theorem 3.2.34 — Minkowski's inequality. Suppose $p \geq 1$, then

$$[E|X+Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}$$

or more generally,

$$\left[E\,|X_1 + \ldots + X_n|^p\right]^{1/p} \leq \left[E\,|X_1|^p\right]^{1/p} + \ldots + \left[E\,|X_n|^p\right]^{1/p}$$

Proof. The proof for $p = 1$ is trivial. Now assume $p > 1$. In order to have $p^{-1} + q^{-1} = 1$, we need $q = \left(1 - p^{-1}\right)^{-1} = p/(p-1)$. From Holder's inequality, we have

$$\begin{aligned}
E|X+Y|^p &= E\left\{|X+Y|^{p-1}|X+Y|\right\} \\
&\leq E\left\{|X+Y|^{p-1}|X|\right\} + E\left\{|X+Y|^{p-1}|Y|\right\} \\
&\leq \left(E|X+Y|^{q(p-1)}\right)^{1/q}(E|X|^p)^{1/p} + \left(E|X+Y|^{q(p-1)}\right)^{1/q}(E|Y|^p)^{1/p} \\
&\leq (E|X+Y|^p)^{1/q}(E|X|^p)^{1/p} + (E|X+Y|^p)^{1/q}(E|Y|^p)^{1/p} \\
&\leq (E|X+Y|^p)^{1/q}\left[(E|X|^p)^{1/p} + (E|Y|^p)^{1/p}\right]
\end{aligned}$$

So $\{E|X+Y|^p\}^{1-1/q} = [E|X+Y|^p]^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$ ■

Jensen's inequality

Theorem 3.2.35 Let ψ be convex, that is, for all $\lambda \in (0,1)$ and $x, y \in \mathcal{R}$, one has

$$\lambda\psi(x) + (1-\lambda)\psi(y) \geq \psi(\lambda x + (1-\lambda)y) \tag{3.15}$$

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

$$\psi(EX) \leq E[\psi(X)]$$

Proof. Denote $\mu = EX$. Since ψ is convex, then there exists $\lambda \in R$ (in nice cases, $\lambda = \psi'(\mu)$) such that for all x

$$\psi(x) \geq \psi(\mu) + \lambda(x - \mu) \tag{3.16}$$

Put $x = X$ and take expectation, we get $E\psi(X) \geq \psi(\mu) + \lambda E(X-\mu) = \psi(\mu) = \psi(EX)$ ■

R Inequality (3.16) means that there always exists a line (i.e. tangent line on the *LHS* of (3.16)) which lies below the convex curve $\psi(x)$. The tangent line and the convex curve touches at $x = \mu$ and λ is the slope of the tangent line, which is the gradient of $\psi(x)$ at μ if it exists.

Proof. proof of (3.16). We shall prove the inequality in several steps. Here are some of the implications.

1. For $x \leq t \leq y$

$$\frac{\psi(t)-\psi(x)}{t-x} \leq \frac{\psi(y)-\psi(x)}{y-x} \leq \frac{\psi(y)-\psi(t)}{y-t}$$

(The geometric meaning of the inequalities is very clear, showing the relationships of the three slopes.) We only show the first part, which is equivalent to

$$\psi(t)=\psi\left(\frac{y-t}{y-x}x+\frac{t-x}{y-x}y\right) \leq \psi(x)+\frac{t-x}{y-x}[\psi(y)-\psi(x)]=\frac{y-t}{y-x}\psi(x)+\frac{t-x}{y-x}\psi(y)$$

This is certainly true by taking $\lambda=\frac{y-t}{y-x}$ in (3.15)

2. Since $t-h_1 \leq t \leq t+h_2$ for $h_1, h_2>0$, from (1) we get

$$\frac{\psi(t)-\psi(t-h_1)}{h_1} \leq \frac{\psi(t+h_2)-\psi(t)}{h_2}$$

LHS (or RHS) is an increasing (or decreasing) function of h_1 (or h_2) which is bounded from above (or below) by the RHS (or LHS). Letting $h_1, h_2 \searrow 0$ results in

$$\psi'(t-) \leq \psi'(t+)$$

3. Let $t \searrow x$ and $t \nearrow y$ in (1), we get

$$\frac{\psi(y)-\psi(x)}{y-x} \geq \psi'(x+), \quad \frac{\psi(y)-\psi(x)}{y-x} \leq \psi'(y-)$$

Changing x into t in the first and y into t in the second, we get

$$\frac{\psi(y)-\psi(t)}{y-t} \geq \psi'(t+), \quad \frac{\psi(t)-\psi(x)}{t-x} \leq \psi'(t-)$$

Using (2), we have

$$\frac{\psi(t)-\psi(x)}{t-x} \leq \psi'(t-) \leq \psi'(t+) \leq \frac{\psi(y)-\psi(t)}{y-t}$$

4. For any fixed t, choose a so that $\psi'(t-) \leq a \leq \phi'(t+)$, and let

$$l(z)=a(z-t)+\psi(t)$$

then $l(t)=\psi(t)$, and $\psi(z) \geq l(z)$

Proof. $l(l)=\psi(l)$ is trivial. The second part is equivalent to $l(z)=a(z-l)+\psi(l) \leq \psi(z)$ Or

$$\psi(z)-\psi(t) \geq a(z-t)$$

This follows from (3). (If $z=t$, trivial. If $z>t$, from (2),$[\psi(z)-\psi(t)]/(z-t) \geq \phi'(t+) \geq a$. If $z<t$, from (2), $[\psi(z)-\psi(t)]/(z-t) \leq \phi'(t-) \leq a)$ ■

■

Theorem 3.2.36 — 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 $Eg(X) < \infty$ then for each $a > 0$

$$P(|X| \geq a) \leq \frac{Eg(X)}{g(a)}$$

Proof. $Eg(X) \geq Eg(X)I_{\{g(X)\geq g(a)\}} \geq g(a)EI_{\{g(X)\geq g(a)\}} = g(a)EI_{\{|X|\geq a\}} = g(a)P(|X| \geq a)$
Some special cases:

$$X \in L^1 \Longrightarrow P(|X| \geq a) \leq \frac{E|X|}{a}$$
$$X \in L^p \Longrightarrow P(|X| \geq a) \leq \frac{E|X|^p}{a^p}$$
$$X \in L^2 \Longrightarrow P(|X - EX| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$
$$Ee^{t|X|} < \infty, t \geq 0 \Longrightarrow P(|X| \geq a) \leq \frac{Ee^{t|X|}}{e^{ta}}$$

■

Definition 3.2.10 — Expectation.
1. If X is a **nonnegative** random variable, the **expectation** of X, denoted $\mathbb{E}(X)$, is
$$\mathbb{E}(X) = \int X d\mathbb{P}$$
where the integral is the Lebesgue integral.
2. If X is a random variable with $\mathbb{E}(|X|) < \infty$, then we also define the expectation by
$$\mathbb{E}(X) = \int X d\mathbb{P}$$
3. For the discrete random variable, $\mathbb{E}(X) = \sum_{j=1}^{\infty} a_j \mathbb{P}\{X = a_j\}$

If X takes on positive and negative values, and $\mathbb{E}(|X|) = \infty$, the expectation is not defined.

Lemma 3.5 Suppose X is a random variable with distribution μ_X. Then

$$\mathbb{E}(X) = \int_{\mathbb{R}} x d\mu_X$$

(Either side exists if and only if the other side exists.)

Analysis 3.1 Using the sample variables to approach.

Proof. First assume that $X \geq 0$. If n, k are positive integers let

$$A_{k,n} = \left\{\omega : \frac{k-1}{n} \leq X(\omega) < \frac{k}{n}\right\}$$

For every $n < \infty$, consider the discrete random variable X_n taking values in $\{k/n : k \in \mathbb{Z}\}$

$$X_n = \sum_{k=1}^{\infty} \frac{k}{n} 1_{A_{k,n}}$$

Then,

$$X_n - \frac{1}{n} \leq X \leq X_n$$

Hence

$$\mathbb{E}\left[X_n\right] - \frac{1}{n} \leq \mathbb{E}[X] \leq \mathbb{E}\left[X_n\right]$$

But,

$$\mathbb{E}\left[X_n - \frac{1}{n}\right] = \sum_{k=1}^{\infty} \frac{k-1}{n} \mathbb{P}\left(A_{k,n}\right)$$

$$\mathbb{E}\left[X_n\right] = \sum_{k=1}^{\infty} \frac{k}{n} \mathbb{P}\left(A_{k,n}\right)$$

and

$$\frac{k-1}{n} \mu_X \left[\frac{k-1}{n}, \frac{k}{n}\right) \leq \int_{\left[\frac{k-1}{n}, \frac{k}{n}\right)} x d\mu_X \leq \frac{k}{n} \mu_X \left[\frac{k-1}{n}, \frac{k}{n}\right)$$

By summing we get

$$\mathbb{E}\left[X_n - \frac{1}{n}\right] \leq \int_{[0,\infty)} x d\mu_X \leq \mathbb{E}\left[X_n\right]$$

By letting n go to infinity we get the result. The general case can be done by writing $X = X^+ - X^-$. ■

In particular, the expectation of a random variable depends **only on its distribution** and not on the probability space on which it is defined. If X has a density f, then the measure μ_X is the same as $f(x)dx$ so we can write (as in elementary courses)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

where again the expectation exists if and only if

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

Theorem 3.2.37 If $X_1, X_2, \ldots$ are nonnegative we can even take countable sums,

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[X_n\right]$$

Proof. Using the rule of finite sums and the monotone converge theorem. ∎

Theorem 3.2.38 Suppose X is a random variable with distribution μ_X, and g is a Borel measurable function. Then,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mu_X$$

(Either side exists if and only if the other side exists.)

Proof. Still use the simple variables and MCT. Assume first that g is a nonnegative function. Then there exists an increasing sequence of nonnegative simple functions with g_n approaching g. Note that $g_n(X)$ is then a sequence of nonnegative simple random variables approaching $g(X)$. ∎

Theorem 3.2.39 — Two important inequality. Let X be a random variable.

1. (Markov's Inequality) If $a > 0$,

$$\mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}[|X|]}{a}$$

2. (Chebyshev's Inequality) If $a > 0$

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\} \leq \frac{\operatorname{Var}(X)}{a^2}$$

Proof. Let $X_a = a1_{\{|X| \geq a\}}$. Then $X_a \leq |X|$ and hence

$$a\mathbb{P}\{|X| \geq a\} = \mathbb{E}[X_a] \leq \mathbb{E}[X]$$

This gives Markov's inequality. For Chebyshev's inequality, we apply Markov's inequality to the random variable $[X - \mathbb{E}(X)]^2$ to get

$$\mathbb{P}\left\{[X - \mathbb{E}(X)]^2 \geq a^2\right\} \leq \frac{\mathbb{E}\left([X - \mathbb{E}(X)]^2\right)}{a^2}$$

∎

Corollary 3.2.40 — Generalized Chebyshev Inequality. Let $f : [0, \infty) \to [0, \infty)$ be a nondecreasing Borel function and let X be a nonnegative random variable. Then for all $a > 0$,

$$\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[f(X)]}{f(a)}$$

3.3 Independenct

Definition 3.3.1 — Independent. Let $(\Omega, \mathcal{A}, P)$ be a probability space.

1. **Events** $A_1, \ldots, A_n \in \mathcal{A}$ are said to be **independent** iff
$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$
for every subset J of $\{1, 2, \ldots, n\}$
2. **Classes** $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are said to be independent iff
$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$
for every subset J of $\{1, 2, \ldots, n\}$, and $A_i \in \mathcal{A}_i$. In particular, σ **-algebras** $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are said to be independent iff
$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) \quad \text{for any } A_i \in \mathcal{A}_i$$
(Note we can choose some $A_i = \Omega \in \mathcal{A}_i$.)
3. The **r.v.'s** $X_1, \ldots, X_n$ are said to be independent iff the events $\{X_i \in B_i\}$ are independent, i.e.,
$$P\left(\bigcap_{i \in J} \{X_i \in B_i\}\right) = \prod_{i \in J} P(\{X_i \in B_i\})$$
for every subset J of $\{1, 2, \ldots, n\}$. This is clearly equivalent to
$$P\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} P(\{X_i \in B_i\}) \tag{3.17}$$
for any Borel sets $B_i \in \mathcal{B}$ (as one can take some $B_i = \mathcal{R}_i$).
4. The r.v.'s of an infinite (not necessarily countable) family are said to be independent iff those in every finite subfamily are.
5. A collection of events $\{A_\alpha\}$ is called **pairwise independent** if for each distinct $A_{\alpha_1}, A_{\alpha_2}$,
$$\mathbb{P}(A_{\alpha_1} \cap A_{\alpha_2}) = \mathbb{P}(A_{\alpha_1})\mathbb{P}(A_{\alpha_2})$$
6. The r.v.'s that are independent and have the same d.f. are called **independent and identically distributed** (i.i.d.).

■ **Example 3.7 — Pairwise indep but not totally indep.** Let X_1, X_2, X_3 are independent r.v.'s with $P(X_i = 0) = P(X_i = 1) = 1/2$. Let $A_1 = \{X_2 = X_3\}, A_2 = \{X_3 = X_1\}, A_3 = \{X_1 = X_2\}$. Then A_i 's are pairwise independent but not (totally) independent.

Proof. Note for $i \neq j$,

$$\begin{aligned}P(A_1) &= P(X_2 = X_3) = P(X_2 = X_3 = 1) + P(X_2 = X_3 = 0)\\ &= P(X_2 = 1, X_3 = 1) + P(X_2 = 0, X_3 = 0)\\ &= P(X_2 = 1)P(X_3 = 1) + P(X_2 = 0)P(X_3 = 0) = 1/4 + 1/4 = 1/2\\ P(A_i \cap A_j) &= P(X_1 = X_2 = X_3)\\ &= P(X_1 = 0, X_2 = 0, X_3 = 0) + P(X_1 = 1, X_2 = 1, X_3 = 1)\\ &= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) + P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)\\ &= 0.5^3 + 0.5^3 = 1/4\end{aligned}$$

Similarly, $P(A_2) = P(A_3) = 1/2$. Thus, $P(A_i \cap A_j) = P(A_i)P(A_j)$. Thus, A_i 's are pairwise independent. But they are not independent since

$$P(A_1 \cap A_2 \cap A_3) = P(X_1 = X_2 = X_3) = 1/4 \neq 1/8 = P(A_1)P(A_2)P(A_3)$$

■

3.3.1 How to check independence

In order to check if $X_1, \ldots, X_n$ are independent, one only needs to verify (3.17) for $B_i = (-\infty, t_i]$.

Theorem 3.3.1 The r.v.'s $X_1, \ldots, X_n$ are independent iff

$$F_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = F_{X_1}(t_1) \ldots\ldots F_{X_n}(t_n) \tag{3.18}$$

for all $t_1, \ldots, t_n \in \mathcal{R}$

Proof. "$\Longrightarrow$". If $X_1, \ldots, X_n$ are independent, then taking $B_i = (-\infty, t_i]$ in (3.17) results in (3.18) . "$\Longleftarrow$". Assume now that (3.18) holds, we shall show that $X_1, \ldots, X_n$ are independent. As the first step, we shall show that

$$P\left(X_1 \in B, \cap_{j=2}^{n} \{X_j \leq t_j\}\right) = P(X_1 \in B) \prod_{j=2}^{n} P(X_j \leq t_j), \quad \forall B \in \mathcal{B} \tag{3.19}$$

To show this, define

$$\mathcal{B}_1 = \left\{B \in \mathcal{B} : P\left(X_1 \in B, \cap_{j=2}^{n} \{X_j \leq t_j\}\right) = P(X_1 \in B) \prod_{j=2}^{n} P(X_j \leq t_j), t_j \in \mathcal{R}, 2 \leq j \leq n\right\}$$

It can be shown:

1. the π -class: $\mathcal{A} = \{(-\infty, t] : t \in \mathcal{R}\}$ is contained in $\mathcal{B}_1$.
2. $\mathcal{B}_1$ is a λ -class.

Then by the Monotone Class Theorem (MCT), we get that $\mathcal{B}_1 \supset \sigma(\mathcal{A}) = \mathcal{B}$. In fact, $\mathcal{B}_1 = \mathcal{B}$ as $\mathcal{B}_1 \subset \mathcal{B}$. This proves (3.19). It remains to show (ii): $\mathcal{B}_1$ is a λ -class.

1. $\mathcal{R} \in \mathcal{B}_1$ by (3.18)
2. If $A \subset B \in \mathcal{B}_1$, then for all $t_2, \ldots, t_n \in \mathcal{R}$

$$\begin{aligned}
&P\left(X_1 \in B - A, \cap_{j=2}^n \{X_j \le t_j\}\right) \\
=&P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) - P\left(X_1 \in A, \cap_{j=2}^n \{X_j \le t_j\}\right) \\
=&P(X_1 \in B) \prod_{j=2}^n P(X_j \le t_j) - P(X_1 \in A) \prod_{j=2}^n P(X_j \le t_j) \\
=&P(X_1 \in B - A) \prod_{j=2}^n P(X_j \le t_j)
\end{aligned}$$

which confirms that $B - A \in \mathcal{B}_1$.
3. If $B_k \in \mathcal{B}_1$ and $B_k \nearrow B$, then for all $t_2, \ldots, t_n \in \mathcal{R}$

$$\begin{aligned}
P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) &= \lim_{k\to\infty} P\left(X_1 \in B_k, \cap_{j=2}^n \{X_j \le t_j\}\right) \\
&= \lim_{k\to\infty} P(X_1 \in B_k) \prod_{j=2}^n P(X_j \le t_j) \\
&= P(X_1 \in B) \prod_{j=2}^n P(X_j \le t_j)
\end{aligned}$$

which implies that $B \in \mathcal{B}_1$

Combining (i)-(iii), we prove that $\mathcal{B}_1$ is a λ -class.

We continue this procedure iteratively until the same extension has been accomplished for $X_2, \ldots, X_n$. For instance, the second step is as follows. Define

$$\begin{aligned}
\mathcal{B}_2 = \Big\{B' \in \mathcal{B} : P&\left(X_1 \in B, X_2 \in B', \cap_{j=3}^n \{X_j \le t_n\}\right) \\
&= P(X_1 \in B) P(X_2 \in B') \prod_{j=3}^n P(X_j \le t_j), B \in \mathcal{B}, t_j \in \mathcal{R}, 3 \le j \le n\Big\}
\end{aligned}$$

Using the similar arguments to the above, we can show that $\mathcal{B}_2 = \sigma(\mathcal{A}) = \mathcal{B}$ ■

Theorem 3.3.2 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.

Proof. For any $B \in \mathcal{G}$, define

$$\mathcal{D}^* = \{A : A \in \sigma(\mathcal{D}) \text{ and } P(A \cap B) = P(A)P(B)\}$$

Then it is easy to show that $\mathcal{D}^*$ is a λ -class containing a π -class $\mathcal{D}$. By the Monotone Class Theorem, $\mathcal{D}^* \supset \sigma(\mathcal{D})$. This completes our proof. ■

Applying the above lemma, we get (Durrett, page 25.)

Theorem 3.3.3 Suppose that $\mathcal{A}_1,\ldots,\mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π -class. Then $\sigma(\mathcal{A}_1),\ldots,\sigma(\mathcal{A}_n)$ are independent.

Question 3.1 Question: Is the condition: "each $\mathcal{A}_i$ is a π -class" really necessary in the last theorem?

Theorem 3.3.4 — Discrete r.v.'s. Discrete r.v.'s $X_1,\ldots,X_n$, taking values in countable set C, are independent iff

$$P(X_1=a_1,\ldots,X_n=a_n)=\prod_{i=1}^{n}P(X_i=a_i) \tag{3.20}$$

for all $a_1,\ldots,a_n\in C$

Proof. If $X_1,\ldots,X_n$ are independent, then (3.20) is obviously true. On the other hand, if (3.20) is true, then

$$\begin{aligned}
F_{X_1,\ldots,X_n}(t_1,\ldots,t_n)&=P(X_1\le t_1,\ldots,X_n\le t_n)\\
&=\sum_{\{a_1\in C:a_1\le t_1\}}\cdots\sum_{\{a_n\in C:a_n\le t_n\}}P(X_1=a_1,\ldots,X_n=a_n)\\
&=\sum_{\{a_1\in C:a_1\le t_1\}}\cdots\sum_{\{a_n\in C:a_n\le t_n\}}P(X_1=a_1)\cdots P(X_n=a_n)\\
&=\left(\sum_{\{a_1\in C:a_1\le t_1\}}P(X_1=a_1)\right)\cdots\left(\sum_{\{a_n\in C:a_n\le t_n\}}P(X_n=a_n)\right)\\
&=P(X_1\le t_1)\cdots P(X_n\le t_n)\\
&=F_{X_1}(t_1)\ldots\ldots F_{X_n}(t_n)
\end{aligned}$$

Therefore, $X_1,\ldots,X_n$ are independent. ■

Theorem 3.3.5 — Absolutely continuous r.v.'s. Let $X=(X_1,\ldots,X_n)$ be an absolutely continuous random vector. Then $X_1,\ldots,X_n$ are independent iff

$$f_X(y_1,\ldots,y_n)=\prod_{i=1}^{n}f_{X_i}(y_i) \tag{3.21}$$

for all $y_1,\ldots,y_n\in\mathcal{R}$

Proof. If $X_1,\ldots,X_n$ are independent, then

$$\begin{aligned}
\prod_{i=1}^{n}\int_{-\infty}^{t_i}f_{X_i}(y_i)\,dy_i&=\prod_{i=1}^{n}P(X_i\le t_i)=P(X_1\le t_1,\ldots,X_n\le t_n)\\
&=\int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_n}f_X(y_1,\ldots,y_n)\,dy_1\ldots dy_n
\end{aligned}$$

Hence,

$$\int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} \left(f_X(y_1, \ldots, y_n) - \prod_{i=1}^{n} f_{X_i}(y_i) \right) dy_1 \ldots dy_n = 0$$

Differentiating w.r.t. $t_1, \ldots, t_n$ results in (3.21). On the other hand, if (3.21) is true, then

$$\begin{aligned} P(X_1 \le t_1, \ldots, X_n \le t_n) &= \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f_X(y_1, \ldots, y_n)\, dy_1 \ldots dy_n \\ &= \prod_{i=1}^{n} \int_{-\infty}^{t_i} f_{X_i}(y_i)\, dy_i = \prod_{i=1}^{n} P(X_i \le t_i) \end{aligned}$$

Therefore, $X_1, \ldots, X_n$ are independent. ∎

3.3.2 Functions of independent r.v.'s

Transformation properties

Theorem 3.3.6 If $X_1, \ldots, X_n$ are independent r.v.'s and $g_1, \ldots, g_n$ are Borel measurable functions, then $g_1(X_1), \ldots, g_n(X_n)$ are independent r.v.'s.

Proof. For $B_i \in \mathcal{B}$, we have $g_i^{-1}(B_i) \in \mathcal{B}$. So

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} \{g_i(X_i) \in B_i\} \right) &= P\left(\bigcap_{i=1}^{n} \left\{ X_i \in g_i^{-1}(B_i) \right\} \right) = \prod_{i=1}^{n} P\left(X_i \in g_i^{-1}(B_i) \right) \\ &= \prod_{i=1}^{n} P(g_i(X_i) \in B_i) \end{aligned}$$

∎

Theorem 3.3.7 Let $1 = n_0 \le n_1 < n_2 < \ldots < n_k = n; g_j$ be a Borel measurable function of $n_j - n_{j-1}$ variables. If $X_1, \ldots, X_n$ are independent r.v.'s, then

$$g_1(X_1, \ldots, X_{n_1}), \quad g_2(X_{n_1+1}, \ldots, X_{n_2}), \ldots\ldots\ldots, \quad g_k(X_{n_{k-1}}, \ldots, X_{n_k})$$

are independent.

Proof. For simplicity, we shall only prove it for $k = 2$. Denote $Z_1 = (X_1, \ldots, X_m)$ and $Z_2 = (X_{m+1}, \ldots, X_n)$. Then $Y_1 \equiv g_1(Z_1)$ and $Y_2 \equiv g_2(Z_2)$ are independent iff for all $B_1, B_2 \in \mathcal{B}$, we have

$$P\left(Z_1 \in g_1^{-1}(B_1), Z_2 \in g_2^{-1}(B_2) \right) = P\left(Z_1 \in g_1^{-1}(B_1) \right) P\left(Z_2 \in g_2^{-1}(B_2) \right)$$

which is implied by the stronger condition

$$P(Z_1 \in A_1, Z_2 \in A_2) = P(Z_1 \in A_1) P(Z_2 \in A_2) \tag{3.22}$$

for all $A_1 \in \mathcal{B}^m$ and $A_2 \in \mathcal{B}^{n-m}$. To show this, define

$$\begin{aligned}\mathcal{B}_1 = \{A \in \mathcal{B}^m : P(Z_1 \in A, Z_2 \in B_1 \times \ldots \times B_{n-m}) = P(Z_1 \in A)\, P(Z_2 \in B_1 \times \ldots \times B_{n-m}),\\ \text{for any } \; B_i \in \mathcal{B}, \quad 1 \leq i \leq n-m\}\end{aligned}$$

Similar to the proof of Theorem 3.3.1, it can be shown:

1. the π -class: $\mathcal{A} = \{A_1 \times \ldots \times A_m : A_i \in \mathcal{B}, 1 \leq i \leq m\}$ is contained in $\mathcal{B}_1$
2. $\mathcal{B}_1$ is a λ -class.

Then by the Monotone Class Theorem, we get that $\mathcal{B}_1 \supset \sigma(\mathcal{A}) = \mathcal{B}^m$. In fact, $\mathcal{B}_1 = \mathcal{B}^m$ as $\mathcal{B}_1 \subset \mathcal{B}^m$. In a similar fashion, we can complete our proof (3.22) . ■

Theorem 3.3.8 — Convolutions. Let X, Y be independent and absolutely continuous. Then $X+Y$ is absolutely continuous *and*

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

Proof.

$$\begin{aligned}P(X+Y \leq t) &= P((X,Y) \in \{(x,y) : x+y \leq t\})\\ &= \iint_{x+y \leq t} f_{X,Y}(x,y) dxdy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{t-y} f_X(x) f_Y(y) dxdy\\ &= \int_{y=-\infty}^{\infty} \left[\int_{x=-\infty}^{t-y} f_X(x) dx \right] f_Y(y) dy\\ &= \int_{y=-\infty}^{\infty} \left[\int_{z=-\infty}^{t} f_X(z-y) dz \right] f_Y(y) dy \quad (z = x+y)\\ &= \int_{z=-\infty}^{t} \left[\int_{y=-\infty}^{\infty} f_X(z-y) f_Y(y) dy \right] dz\end{aligned}$$

■

■ **Example 3.8** If $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, and X, Y are independent, then $X+Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Theorem 3.3.9 Let X, Y be nonnegative and integer-valued. Then for each $n \geq 0$

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

Proof.

$$\begin{aligned}P(X+Y=n) &= P\left(\sum_{k=0}^{n} \{X=k, Y=n-k\}\right) = \sum_{k=0}^{n} P(X=k, Y=n-k)\\ &= \sum_{k=0}^{n} P(X=k) P(Y=n-k)\end{aligned}$$

■

■ **Example 3.9** If $X \sim$ Poisson $(\lambda_1), Y \sim$ Poisson (λ_2), and X, Y are independent, then $X+Y \sim$ Poisson $(\lambda_1 + \lambda_2)$

■ **Example 3.10**
1. Show that if X and Y are independent, then $\sigma(X)$ and $\sigma(Y)$ are too. Therefore, X and Y are independent iff $\sigma(X)$ and $\sigma(Y)$ are independent.
2. Show that events A and B are independent, then so are A^c and B, A and B^c, A^c and B^c.

   Proof. If A and B are independent, then

$$P(A \cap B^c) = P(A \cap (\Omega - B)) = P(A - (A \cap B))$$
$$= P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(B^c)$$

■

3. Events $A_1, \ldots, A_n$ are independent iff $I_{A_1}, \ldots, I_{A_n}$ are independent.

   Proof. We shall do this for $n = 2$. If A_1, A_2 are independent,

$$\begin{aligned}
P(I_{A_1}=1, I_{A_2}=1) &= P(A_1 \cap A_2) = P(A_1)P(A_2) = P(I_{A_1}=1)P(I_{A_2}=1)\\
P(I_{A_1}=1, I_{A_2}=0) &= P(A_1 \cap A_2^c) = P(A_1)P(A_2^c) = P(I_{A_1}=1)P(I_{A_2}=0)\\
P(I_{A_1}=0, I_{A_2}=1) &= P(A_1^c \cap A_2) = P(A_1^c)P(A_2) = P(I_{A_1}=0)P(I_{A_2}=1)\\
P(I_{A_1}=0, I_{A_2}=0) &= P(A_1^c \cap A_2^c) = P(A_1^c)P(A_2^c) = P(I_{A_1}=0)P(I_{A_2}=0)
\end{aligned}$$

   Thus, I_{A_1}, I_{A_2} are independent.
   On the other hand, if I_{A_1}, I_{A_2} are independent, then

$$P(A_1 \cap A_2) = P(I_{A_1}=1, I_{A_2}=1) = P(I_{A_1}=1)P(I_{A_2}=1) = P(A_1)P(A_2)$$

   which implies that A_1, A_2 are independent. ■

Definition 3.3.2 — Correlation. The covariance of two r.v.'s X and Y is defined to be

$$\text{Cov}(X,Y) := E(X-EX)(Y-EY) = E(XY) - EXEY$$

X, Y are said to be positively correlated, uncorrelated, or negatively correlated iff $\text{Cov}(X,Y) > 0, = 0$ or < 0, respectively.

Theorem 3.3.10 If X, Y are independent and integrable r.v.'s, then

$$\text{Cov}(X,Y) = 0$$

That is, independence implies uncorrelatedness.

(Remark: The theorem has been proved for discrete and absolute continuous r.v.'s in an elementary probability course.)

Proof. We shall divide the proof in several steps.

1. Step 1. If X, Y are simple r.v.'s., i.e., $X = \sum_{i=1}^{n} a_i I_{\{X=a_i\}}$ and $Y = \sum_{j=1}^{m} b_j I_{\{Y=b_j\}}$. For simplicity, we assume that a_i's and b_j's are different. Then, $EX = \sum_{i=1}^{n} a_i P(\{X = a_i\})$ and $EY = \sum_{j=1}^{m} b_j P(\{Y = b_j\})$. Note $XY = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j I_{\{X=a_i\}\cap\{Y=b_j\}}$ is also a discrete r.v. and

$$
\begin{aligned}
E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P\left(\{X = a_i\} \cap \{Y = b_j\}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j P\left(\{X = a_i\}\right) P\left(\{Y = b_j\}\right) = EXEY
\end{aligned}
$$

2. If $X \geq 0, Y \geq 0$, then take $X_n = \min\left([2^n X]/2^n, n\right)$ and $Y_n = \min\left([2^n Y]/2^n, n\right)$. So
   (a) X_n and Y_n are both simple r.v.'s.
   (b) $0 \leq X_n \nearrow X$ and $0 \leq Y_n \nearrow Y$, which in turn implies that
      i. $EX = \lim_n EX_n$, and $EY = \lim_n EY_n$, by the Monotone Convergence Theorem.
      ii. $0 \leq X_n Y_n \nearrow XY$. (Proof. The first part $0 \leq X_n Y_n \nearrow$ is obvious. To show $X_n Y_n \nearrow XY$, we note $0 \leq XY - X_n Y_n = X(Y - Y_n) + Y_n(X - X_n) \to 0$ a.s., using the fact: $\tilde{X}_n \to 0$ and $\tilde{Y}_n \to \tilde{Y}$ implies $\tilde{X}_n \tilde{Y}_n \to 0$ a.s.)
   (c) X_n and Y_n are independent as Borel functions of X and Y, respectively.
   Now applying the Monotone Convergence Theorem, we get

$$
E(XY) = \lim_n E(X_n Y_n) = \lim_n E(X_n) E(Y_n) = (EX)(EY)
$$

3. Step 3. For general integrable r.v.'s, note that independence of X and Y implies that of X^+ and Y^+; X^+ and Y^-; and so on. Therefore,

$$
\begin{aligned}
EXY &= E\left(X^+ - X^-\right)\left(Y^+ - Y^-\right) \\
&= E\left(X^+Y^+\right) - E\left(X^+Y^-\right) - \left(EX^-Y^+\right) + E\left(X^-Y^-\right) \\
&= EX^+EY^+ - EX^+EY^- - EX^-EY^+ + EX^-EY^- \\
&= \left(EX^+ - EX^-\right)\left(EY^+ - EY^-\right) \\
&= EXEY
\end{aligned}
$$

■

Theorem 3.3.11 If $X_1, \ldots, X_n$ are independent and all have finite expectations, then

$$
E(X_1 \cdots X_n) = (EX_1) \cdots (EX_n)
$$

The following provides a covariance inequality.

Theorem 3.3.12 Let $u(x)$ and $v(x)$ be both non-decreasing or both non-increasing functions on $I = (a,b)$ (finite or infinite interval on R), and $P(X \in I) = 1$. Then,

$$Eu(X)Ev(X) \leq E[u(X)v(X)], \quad \text{or} \quad \mathrm{Cov}(u(X),v(X)) \geq 0$$

provided these means exist.

Proof. Let X and Y be independent and identically distributed r.v.'s. Clearly, we have $[u(X) - u(Y)][v(X) - v(Y)] \geq 0$. Hence

$$\begin{aligned}
0 &\leq E[u(X) - u(Y)][v(X) - v(Y)] \\
&= Eu(X)v(X) + Eu(Y)v(Y) - Eu(X)v(Y) - Eu(Y)v(X) \\
&= 2Eu(X)v(X) - 2Eu(X)v(Y) \\
&= 2Eu(X)v(X) - 2Eu(X)Ev(Y)
\end{aligned}$$

from which we have $Eu(X)v(X) \geq Eu(X)Ev(Y) = Eu(X)Ev(X)$ ■

R The theorem implies that $\mathrm{Cov}(u(X),v(X)) \geq 0$, i.e., $u(X)$ and $v(X)$ are positively correlated (in the broad sense). This makes sense since both $u(X)$ and $v(X)$ tend to increase or decrease together as X increases or decreases. Some simple application of this inequality yields

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

For instance, we have

$$(E|X|)^2 \leq EX^2, \quad (E|X|)\left(EX^2\right) \leq E|X|^3, \quad \text{etc.}$$

Definition 3.3.3 Let $\{A_n\}$'s be a sequence of events on $(\Omega, \mathcal{A}, P)$. Recall

$$\limsup_n A_n = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m = \lim_{n\to\infty} \cup_{m=n}^{\infty} A_m = \{A_n \text{ i.o.}\}$$

$$\liminf_n A_n = \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m = \lim_{n\to\infty} \cap_{m=n}^{\infty} A_m = \{A_n \text{ ult. }\}$$

$$\liminf_n A_n = \left(\limsup_n A_n^c\right)^c$$

Theorem 3.3.13 — Borel-Cantelli Lemma. 1. $P(A_n, i.o) = 0$ if $\sum_{n\to\infty} P(A_n) < \infty$
2. $P(A_n, i.o) = 1$ if $\sum_{n\to\infty} P(A_n) = \infty$ and $A_1, A_2, \ldots,$ are independent.

Proof. 1. $P(A_n, i.o) = P(\lim_n \cup_{m=n}^{\infty} A_m) = \lim_n P(\cup_{m=n}^{\infty} A_m) \leq \lim_n \sum_{m=n}^{\infty} P(A_m) \to 0$

2. Noting $1-x \le e^{-x}$ for all $x \in \mathcal{R}$ and independence of A_n, we have

$$
\begin{aligned}
0 \le & 1-P\left(A_n, i.o.\right)=P\left(A_n^c, ult.\right)=P\left(\lim_n \cap_{m=n}^{\infty} A_m^c\right) \\
& =\lim_{n \rightarrow \infty} P\left(\lim_{r \rightarrow \infty} \cap_{m=n}^r A_m^c\right)=\lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} P\left(\cap_{m=n}^r A_m^c\right) \\
& =\lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \prod_{m=n}^r\left[1-P\left(A_m\right)\right] \quad \text{(independence)} \\
& =\lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \prod_{m=n}^r e^{-P\left(A_m\right)}\left(\text{ as } 1-x \le e^{-x}\right) \\
& =\lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} e^{-\sum_{m=n}^r P\left(A_m\right)}=\lim_{n \rightarrow \infty} e^{-\sum_{m=n}^{\infty} P\left(A_m\right)} \\
& =\lim_{n \rightarrow \infty} 0=0
\end{aligned}
$$

■

Proposition 3.3.14 The inequality $e^x \ge 1+x$, all $x \in R$, can be shown by different methods.

1. Method 1. $f(x):=e^x-(1+x), f^{\prime}\left(x_0\right)=0 \Longrightarrow x_0=0, f^{\prime \prime}\left(x_0\right)=e^{x_0}>0$, convex at x_0. So

$$
f(x) \ge f\left(x_0\right)=0
$$

2. Method 2. It can be seen easily by comparing the plots of e^x v.s. $1+x$, and by noting e^x is convex.
3. Method 3. If $x \ge 0, e^x \ge e^0=1, \Longrightarrow e^x-1=\int_0^x e^x dx \ge \int_0^x dx=x$.
On the other hand, if $x \le 0, e^x \le e^0=1, \Longrightarrow 1-e^x=\int_x^0 e^x dx \le \int_x^0 dx=-x$. In fact, we can continue doing this to get $e^x \ge 1+x^2+\ldots+x^m / m!$ (for $x \ge 0$).

R Part (a) holds irrespectively of A_n 's being independent or not. However, part (b) may not hold if A_n 's are (strongly) dependent. For instance, take $A_n=A$ with $0<P(A)<1$, then $\sum_n P\left(A_n\right)=\infty$ but $P\left(A_n, i.o\right)=P(A) \in(0,1)$

Theorem 3.3.15 — Independence can be reduced to pairwise independence. $P\left(A_n, i.o\right)=1$ if $\sum_{n=1}^{\infty} P\left(A_n\right)=\infty$ and $\left\{A_n, n \ge 1\right\}$ are (pairwise) independent.

Corollary 3.3.16 — Borel 0-1 Law. Let $\left\{A_n\right\}$'s be (pairwise) independent. Then

$$
\begin{aligned}
P\left(A_n, i.o\right)=0 & \quad \text{if } \sum_n P\left(A_n\right)<\infty \\
=1 & \quad \text{if } \sum_n P\left(A_n\right)=\infty
\end{aligned}
$$

i.e., when A_n are independent, $P\left(\lim\sup A_n\right)=0$, or 1 according to $\sum P\left(A_n\right)<$ or $=\infty$.

Corollary 3.3.17 — Alternative form of Borel 0-1 Law. Let $\{A_n\}$'s be (pairwise) independent. Then

1. $P(A_n, \text{ i.o.}) = 0 \iff \sum_n P(A_n) < \infty$
2. $P(A_n, \text{ i.o.}) = 1 \iff \sum_n P(A_n) = \infty$

R

1. We have seen that, if A_n 's are not independent, $P(A_n, i.o)$ is either 0 or 1. Now, if A_n 's are not independent, is it possible $0 < P(A_n, i.o) < 1$?

Corollary 3.3.18 If A_n are (pairwise) independent and $A_n \to A$, then $P(A) = 0$, or 1.

Proof. $P(A) = P(\lim_n A_n) = P(\limsup_n A_n) = P(A_n, i.o.)$, apply Borel $0-1$ Law. ■

Corollary 3.3.19 Let X_n be (pairwise) independent. Then

$$X_n \to 0 \text{ a.s.} \iff \sum_n P(|X_n| \geq \epsilon) < \infty, \forall \epsilon > 0$$

(That is, convergence in probability fast enough implies convergence almost sure.)

Proof.

$$\begin{aligned}
& X_n \to 0 \quad a.s. \\
\iff & P\left(\lim_n X_n = 0\right) = 1 \\
\iff & P\left(\cap_{m=1}^{\infty} \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < 1/m\}\right) = 1 \\
\iff & \forall m \geq 1 : P\left(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < 1/m\}\right) = 1 \\
& \qquad (\text{as } P\left(\cap_{m=1}^{\infty} B_m\right) = 1 \implies \forall m \geq 1 : P(B_m) = 1, \text{ see the proof given below. }) \\
\iff & \forall \epsilon > 0 : P\left(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < \epsilon\}\right) = 1 \\
\iff & \forall \epsilon > 0 : P\left(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} \{|X_k| \geq \epsilon\}\right) = 0 \\
\iff & P\left(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k\right) = 0, \text{ where } A_k = \{|X_k| \geq \epsilon\} \\
\iff & P(A_n, \text{i.o.}) = 0 \\
\iff & \sum_n P(A_n) = \sum_n P(|X_n| \geq \epsilon) < \infty \\
& \qquad \text{(By Borel 0-1 law for (pairwise) independent events).}
\end{aligned}$$

■

R In the above proof, we have used the fact: $P\left(\cap_{m=1}^{\infty} B_m\right) = 1 \implies \forall j \geq 1 : P(B_j) = 1$.

Proof. $\forall j \geq 1$, we have $0 = P\left(\cup_{m=1}^{\infty} B_m^c\right) \geq P\left(B_j^c\right) = 1 - P\left(B_j\right)$. So $P\left(B_j\right) = 1$ ■

Corollary 3.3.20 Let $\{X, X_n, n \geq 1\}$ be (pairwise) i.i.d., then

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

Proof.
1. $E|X| < \infty \iff \sum_n P\left(\left|X_n\right|/n \geq \epsilon\right) = \sum_n P(|X| \geq \epsilon n) < \infty, \forall \epsilon > 0. \iff \left|X_n\right|/n \to$ 0 a.s (from Corollary 5.4.4)
2. It follows from (i).

■

Kolmogorov 0-1 laws

Definition 3.3.4 — tail σ -algebra. 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

$$\bigcap_{n=1}^{\infty} \sigma\left(X_j, j \geq n\right) \equiv \bigcap_{n=1}^{\infty} \sigma\left(X_n, X_{n+1}, \ldots\ldots\right)$$

The sets of the tail σ -algebra are called tail events, and functions measurable relative to the tail σ -algebra are dubbed tail functions.

R

1. Recall $\sigma\left(X_j, j \geq n\right) \equiv \sigma\left(X_n, X_{n+1}, \ldots\ldots\right) =$ the future after time $n =$ the smallest σ algebra w.r.t. which all $X_m, m \geq n$ are measurable.
2. Intuitively, A is a tail event if and only if changing a finite number of values does not affect the occurrence of the event.
3. A is a tail event if the event depends entirely on the "tail series".

■ **Example 3.11 — Examples of tail events..**

1. If $B_n \in \mathcal{B}$, then $\{X_n \in B_n \text{ i.o.}\}$ is a tail event.

   Proof.

   $$\{X_n \in B_n \text{ i.o.}\} = \limsup_n \{X_n \in B_n\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{X_m \in B_m\} \in \bigcap_{n=1}^{\infty} \sigma\left(X_n, X_{n+1}, \ldots\ldots\right)$$

   ■

2. $\{A_n \text{ i.o. }\}$ is a tail event.

   Proof. Taking $X_n = I_{A_n}, B_n = \{1\}$ in (i), we get

   $$\{A_n \text{ i.o.}\} = \{I_{A_n} = 1 \text{ i.o.}\} = \{X_n \in B_n \text{ i.o.}\}$$

   Then apply (i). ■

3. Let $S_n = X_1 + \cdots + X_n$. It is easy to check that (Durrett, p62.)
   (a) $\{\lim_n S_n \text{ exists}\}$ is a tail event.
   (b) $\{\limsup_n S_n > x\}$ is NOT a tail event.
   (c) $\{\limsup_n S_n / C_n > x\}$ is a tail event if $C_n \to \infty$

Proof. 1.

$$\left\{\lim_n S_n \text{ exists}\right\} = \left\{\sum_{n=1}^{\infty} X_n \text{ converges}\right\} = \bigcap_{n=1}^{\infty} \left\{\sum_{m=n}^{\infty} X_m \text{ converges}\right\}$$
$$\in \bigcap_{n=1}^{\infty} \sigma\left(X_n, X_{n+1}, \ldots\ldots\right)$$

2. $\{\limsup_n S_n > x\}$ may depend on the initial values of X_i 's. For instance, assume that $\limsup_n S_n = \limsup_n S_n = 2 > 1$ (i.e. take $x = 1$). $\forall m \geq 1$, if we change X_m to $X_m - 2$, then $\lim_n S_n = 0 < 1$
3. Since $C_n \to \infty, \forall m \geq 1$, we have

$$\limsup_n S_n / C_n = \lim_n S_m / C_n + \limsup_n \left(S_n - S_m\right) / C_n = \limsup_n \left(S_n - S_m\right) / C_n$$

Hence,

$$\left\{\limsup_n S_n / C_n > x\right\} = \bigcap_{m=1}^{\infty} \left\{\limsup_n \left[\left(X_{m+1} + \ldots + X_n\right)\right] / C_n > x\right\}$$
$$\in \bigcap_{m=1}^{\infty} \sigma\left(X_m, X_{m+1}, \ldots..\right)$$

■

Theorem 3.3.21 — Kolmogorov 0-1 Law. Tail events of a sequence $\{X_n, n \geq 1\}$ of independent r.v.'s have probabilities zero or one.

Proof. (The proof is from Chow and Teicher, p64.) Since $\{X_n, n \geq 1\}$ are independent,

$\Longrightarrow \forall n \geq 1 : \sigma\left(X_i, 1 \leq i \leq n\right)$ and $\sigma\left(X_j, j > n\right)$ are independent.

$\Longrightarrow \forall n \geq 1 : \sigma\left(X_i, 1 \leq i \leq n\right)$ and $\cap_{m=0}^{\infty} \sigma\left(X_j, j > m\right) =: \mathcal{D}$ are independent. (as $\mathcal{D} \subset \sigma\left(X_j, j > n\right)$).

$\Longrightarrow \mathcal{A} =: \cup_{n=1}^{\infty} \sigma\left(X_i, 1 \leq i \leq n\right)$ and $\mathcal{D}$ are independent.

[as for any $A \in \mathcal{A}$, is an algebra (but not necessarily a σ -algebra), hence a π -class.]

$\Longrightarrow$ By Theorem 3.3.2,$\sigma(\mathcal{A})$ and $\mathcal{D}$ are independent. [as $\mathcal{A}$ is an algebra (but not necessarily a σ -algebra), hence a π-class].

$\Longrightarrow \mathcal{D}$ and $\mathcal{D}$ are independent since $\mathcal{D} \subset \sigma\left(X_n, n \geq 1\right) = \sigma(\mathcal{A})$. (Why?)

$\Longrightarrow \forall B \in \mathcal{D}$, we have $P(B) = P(B \cap B) = P^2(B)$, implying $P(B) = 1$ or 0 ■

R When $\mathcal{D}$ is independent of itself, then we must have $\mathcal{D} = \{\emptyset, \Omega\}$.

Corollary 3.3.22 Tail functions of a sequence of independent r.v.'s are degenerate, i.e., constants a.s.

Proof. Let Y be a tail function, by the $0-1$ law, $P(Y \leq c)=0$ or 1 for any $c \in R$.

1. If $P(Y \leq c)=0$ for all $c \in R$, then $P(Y=\infty)=1$
2. If $P(Y \leq c)=1$ for all $c \in R$, then $P(Y=-\infty)=1$
3. Otherwise, $c_0=\inf\{c: P(Y \leq c)=1\}$ is finite, hence $P\left(Y=c_0\right)=1$ by definition.

■

Corollary 3.3.23 If $\left\{X_n, n \geq 1\right\}$ is a sequence of independent r.v.'s, then $\limsup_{n \rightarrow \infty} X_n$ and $\liminf_{n \rightarrow \infty} X_n$ are degenerate a.s.

Proof. For each $n \geq k \geq 1, X_n$ is $\sigma\left(X_j, j \geq k\right)$ -measurable, and $Y_k=: \sup_{n \geq k} X_n$ is $\sigma\left(X_j, j \geq k\right)$ measurable. Hence, Y_n is $\sigma\left(X_j, j \geq n\right)$ -measurable and hence $\sigma\left(X_j, j \geq k\right)$ -measurable for every $n \geq k \geq 1$. Since $\left[\sigma\left(X_j, j \geq n\right) \subset \sigma\left(X_j, j \geq k\right)\right]$. This implies that $\limsup_{n \rightarrow \infty} X_n= \lim_{n \rightarrow \infty} \sup_{j \geq n} X_j=\lim_{n \rightarrow \infty} Y_n$ is $\sigma\left(X_j, j \geq k\right)$ -measurable for every $k \geq 1$. Thus, $\limsup_{n \rightarrow \infty} X_n$ is $\cap_{k=1}^{\infty} \sigma\left(X_j, j \geq k\right)-$ measurable, i.e., a tail function, and similarly for $\liminf_{n \rightarrow \infty} X_n$. The proof follows from the last corollary. ■

R Some comparisons about Kolmogorov 0-1 law and Borel-Cantelli Lemma.

1. The Kolmogorov 0-1 law confines the value of $P\left(A_n\right.$, i.o. $)$ to 0 or 1 (since $\left\{A_n\right.$, i.o. $\}$ is a tail event), but it does not specify which value it takes. On the other hand, the Borel-Cantelli Lemma enables us to specify exactly which value it takes.
2. Borel-Cantelli Lemma is applicable to independent and dependent events, while the Kolmogorov 0-1 law applies to independent r.v.'s.

Definition 3.3.5 — Random variable equal to measurable function. If X is a random variable and $\mathcal{F}$ is a σ -algebra, we say that X is $\mathcal{F}$ -measurable if $X^{-1}(B) \in \mathcal{F}$ for every Borel set B, i.e., if X is a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define $\mathcal{F}(X)$ to be the smallest σ -algebra $\mathcal{F}$ for which X is $\mathcal{F}$ -measurable.$\mathcal{F}(X)= \left\{X^{-1}(B): B \text{ Borel }\right\}$

Lemma 3.6 Suppose $\mathcal{F}^0$ and $\mathcal{G}^0$ are two algebras of events that are independent, i.e., $\mathbb{P}(A \cap B)=\mathbb{P}(A) \mathbb{P}(B)$ for $A \in \mathcal{F}^0, B \in \mathcal{G}^0$. Then $\mathcal{F}:=\sigma\left(\mathcal{F}^0\right)$ and $\mathcal{G}:=\sigma\left(\mathcal{G}^0\right)$ are independent σ -algebras. ■

Proof. Let $B \in \mathcal{G}^0$ and define the measure

$$\mu_B(A)=\mathbb{P}(A \cap B), \quad \tilde{\mu}_B(A)=\mathbb{P}(A) \mathbb{P}(B)$$

Note that $\mu_B, \tilde{\mu}_B$ are finite measures and they agree on $\mathcal{F}^0$. Hence (using Cartheodory extension) they must agree on $\mathcal{F}$. Hence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for $A \in \mathcal{F}, B \in \mathcal{G}^0$. Now take $A \in F$ and consider the measures

$$\nu_A(B) = \mathbb{P}(A \cap B), \quad \tilde{\nu}_A(B) = \mathbb{P}(A)\mathbb{P}(B)$$

■

Definition 3.3.6 — Independence of Random variables. Random variables $X_1, X_2, \ldots, X_n$ are said to be independent if any of these (equivalent) conditions hold:

1. Event: For all Borel sets $B_1, \ldots, B_n$

$$\mathbb{P}\{X_1 \in B_1, \ldots, X_n \in B_n\} = \mathbb{P}\{X_1 \in B_1\}\mathbb{P}\{X_2 \in B_2\} \cdots \mathbb{P}\{X_n \in B_n\}$$

2. σ -algebras: The σ -algebras $\mathcal{F}(X_1), \ldots, \mathcal{F}(X_n)$ are independent.
3. distribution: $\mu = \mu_{X_1} \times \mu_{X_2} \times \cdots \times \mu_{X_n}$
4. Distribution function: For all real $t_1, \ldots, t_n$

$$F(t_1, \ldots, t_n) = F_{X_1}(t_1) F_{X_2}(t_2) \cdots F_{X_n}(t_n)$$

5. Density: Random variables with a joint density are independent if and only if e

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

where $f_{X_j}(x_j)$ denotes the density of X_j. An infinite collection of random variables is called independent if each finite subcollection is independent.

Proposition 3.3.24 If $X_1, \ldots, X_n$ are independent random variables and $g_1, \ldots, g_n$ are Borel measurable functions, then $g_1(X_1), \ldots, g_n(X_n)$ are independent.

Proposition 3.3.25 If X, Y are independent random variables with $\mathbb{E}|X| < \infty, \mathbb{E}|Y| < \infty$, then $\mathbb{E}[|XY|] < \infty$ and

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Proved by simple variables and MCT.

The multiplication rule for expectation can also be viewed as an application of Fubini's Theorem. Let μ, ν denote the distributions of X and Y. If they are independent, then the distribution of (X, Y) is $\mu \times \nu$ Hence

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy d(\mu \times \nu) = \int \left[\int xy d\mu(x) \right] d\nu(y) = \int \mathbb{E}[X] y d\nu(y) = \mathbb{E}[X]\mathbb{E}[Y]$$

Definition 3.3.7 — orthogonal. Random variables X, Y that satisfy $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ are called **orthogonal**.

Proposition 3.3.26 — a generalization of the Pythagorean Theorem.. Suppose $X_1, \ldots, X_n$ are random variables with finite variance. If $X_1, \ldots, X_n$ are pairwise orthogonal, then

$$\operatorname{Var}[X_1 + \cdots + X_n] = \operatorname{Var}[X_1] + \cdots + \operatorname{Var}[X_n]$$

Just to prove the situation when all random variables have zero mean.

R Recall $\langle X, Y \rangle = \mathbb{E}[XY]$.

3.4 Convergence

Definition 3.4.1 — Modes of convergence. Let $X, X_1, X_2, \ldots$ be r.v.'s on $(\Omega, \mathcal{A}, P)$. We say that

1. **Almost surely:** $X_n \to X$ a.s. (i.e., with probability 1) if

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

2. r **th mean:** $X_n \to X$ in r th mean, or in L_r space, where $r > 0$, if

$$\lim_{n\to\infty} E|X_n - X|^r = 0$$

3. **In prob:** $X_n \to X$ in prob, written as $X_n \to_p X$, if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

4. **In distribution:** $X_n \to X$ in distribution, written as $X_n \to_d X$ or $F_{X_n} \Longrightarrow F_X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

   for all continuity points of $F_X(x)$ (All discontinuity points of F_X has Lesbegue measure 0, hence convergence at continuity points will be enough to identify the limit.)

Proposition 3.4.1

1. Note that

$$\left\{\omega \in \Omega : \lim_n X_n(\omega) \text{ converges }\right\} = \left\{\omega \in \Omega : \limsup_n X_n(\omega) = \liminf_n X_n(\omega)\right\}$$

   is clearly $\mathcal{A}$ -measurable, thus an event.
2. A **metric space** $\{X, \rho\}$ is a nonempty set X of elements together with a real-valued function ρ defined on $X \times X$ such that for all x, y, and $z \in X$:
   (a) $\rho(x,y) \geq 0$, and $\rho(x,y) = 0$ iff $x = y$; (nonnegativity)

(b) $\rho(x,y) = \rho(y,x)$; (symmetry)

(c) $\rho(x,y) \le \rho(x,z) + \rho(z,y)$; (triangle inequality).

The function ρ is called a **metric**.

3. $\rho(Y,Z) = \|Y - Z\|_r$ (actually a norm) is a metric in $L_r = \{X : E|X|^r < \infty\}$, where

$$\|X\|_r = \begin{array}{ll} E|X|^r, & 0 < r < 1 \\ (E|X|^r)^{1/r}, & r \ge 1 \end{array}$$

Proof. Nonnegativity and symmetry are easy to check. It remains to show the triangle inequality.

If $r \ge 1$, Minkowski's inequality implies $\|X+Y\|_r \le \|X\|_r + \|Y\|_r$.

If $0 < r < 1$, then for any $0 \le \lambda \le 1$, we have $\lambda^r + (1-\lambda)^r \ge \lambda + (1-\lambda) = 1$. The proof follows by taking $\lambda = |x|/(|x|+|y|)$ ■

So convergence in r th mean may be interpreted as convergence in the L_r metric. Thus, $X_n \to X$ in r th mean iff $\|X_n - X\|_r \to 0$

4. For convergence in probability, we can define a measure of "pseudo-distance" between X and Y by

$$\rho_\epsilon(X,Y) = P(|X - Y| > \epsilon) = \int_E dP, \quad \text{where } E = \{\omega \in \Omega : |X(\omega) - Y(\omega)| > \epsilon\}$$

Note that ρ_e is NOT a metric since for fixed $\epsilon > 0, \rho_\epsilon(X,Y) = 0 \ne X = Y$ a.s.

5. **Levy metric**: For two d.f.'s F, G, let

$$\rho(F,G) = \inf\{\delta > 0 : F(x-\delta) - \delta \le G(x) \le F(x+\delta) + \delta, \text{ for all } x \in R\}$$

Then it can be shown that

(a) ρ is a metric on the space of d.f.'s.

(b) Convergence in distribution is equivalent to convergence w.r.t. Levy metric.

6. The convergence $\to_p, \to$ a.s and $\to_{L_r}$ each represent a sense in which, for n sufficiently large, $X_n(\omega)$ and $X(\omega)$ approximate each other as functions of $\omega \in \Omega$. This means that the d.f.'s of X_n and X can not be too dissimilar.

7. On the other hand, $\to_d$ depends only the d.f.'s involved and does not necessitate that the relevant r.v.'s approximate each other as functions of ω. In fact, X_n and X may not be defined on the same probability space.

Equivalent definition of a.s. convergence

Theorem 3.4.2 The following statements are equivalent:

1. $X_n \to X$ a.s.

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

2. $\forall \epsilon > 0 : \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} \{|X_m - X| < \epsilon\}\right) = 1$

3. $\forall \epsilon > 0 : \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) = 0$
4. $\forall \epsilon > 0 : \lim_{n\to\infty} P\left(\{\sup_{m=n}^{\infty} |X_m - X|\} \geq \epsilon\right) = 0$, i.e. $\sup_{m=n} |X_m - X| \longrightarrow_p 0$
5. $\forall \epsilon > 0 : P\left(|X_n - X| \geq \epsilon, i.o.\right) = 0$

Furthermore, $\geq \epsilon$ " and " $< \epsilon$ " above may be replaced by " $> \epsilon$ " and " $\leq \epsilon$ ", respectively.

Proof. 1. " $(a) \iff (b)$". First note that

$$\begin{aligned}
&\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} \\
&= \bigcap_{\epsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{m=n} \{\omega : |X_m(\omega) - X(\omega)| < \epsilon\} \\
&\quad (\text{ i.e. },\forall \epsilon > 0, \exists n \geq 1, \text{ s.t. } |X_m(\omega) - X(\omega)| < \epsilon \quad \forall m \geq n) \\
&= \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\right\} \\
&= \bigcap_{k=1}^{\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\right\} \\
&= \lim_{k\to\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\right\} \\
&= \lim_{k\to\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{|X_m - X| < \frac{1}{k}\right\}
\end{aligned} \tag{3.23}$$

(Recall If $A_n \uparrow A, B_n \downarrow B$, then $A = \lim_n A_n = \cup_{n=1}^{\infty} A_n, B = \lim_n B_n = \cap_{n=1}^{\infty} B_n$) If (a) holds, then it follows from (3.23) that

$$1 = P\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) \leq \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} \left\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\right\}\right) \leq 1$$

which implies that (b) holds. If (b) holds, since probability is continuous, we have

$$P\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) \to X(\omega)\right\}\right) = \lim_{k\to\infty} \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} \left\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\right\}\right) = 1$$

which implies that (a) holds.

2. "$(b) \iff (c)$". Trivial.
3. "$(c) \iff (e)$". Trivial since

$$\begin{aligned}
P\left(|X_n - X| \geq \epsilon, i.o.\right) &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) \\
&= P\left(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) \\
&= \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right)
\end{aligned}$$

4. $''(c) \iff (d)''.\forall \epsilon > 0$, let

$$A_{n,\epsilon} = \left\{\omega \in \Omega : \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right\}, \quad B_{n,\epsilon} = \left\{\omega \in \Omega : \sup_{m \geq n} |X_m - X| \geq \epsilon\right\}$$

Then the equivalence of (c) and (d) follows from the fact: $A_{n,\epsilon} \subset B_{n,\epsilon} \subset A_{n,\epsilon/2}$. (Why?)

■

Cauchy Criterion

Theorem 3.4.3 — (Cauchy Criterion of a.s.–When no limit is specified). X_n converges a.s.

$$\iff \forall \epsilon > 0 : \lim_{n \to \infty} P\left(|X_m - X_{m'}| \leq \epsilon, \text{all } m > m' \geq n\right) = 0$$

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

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

$$\iff \sup_{m,n \geq M} |X_m - X_n| \longrightarrow_p 0$$

Relationships between modes of convergence

Theorem 3.4.4 1. For $r \geq 1$

$$\begin{array}{l} X_n \to X \text{ a.s.} \\ \qquad\qquad \Longrightarrow X_n \to_p X \Longrightarrow X_n \to_d X \\ X_n \to X \text{ in } L_r \end{array}$$

2. If $r > s > 0$, then $X_n \to X$ in $L_r \Longrightarrow X_n \to X$ in L_s
3. No other implications hold in general.

Proof. 1. (a) If $X_n \to X$ a.s., then $X_n \to_p X$.

Proof. Note that $0 \leq P\left(|X_n - X| \geq \epsilon\right) \leq P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) \to 0$ as $n \to \infty$ by Theorem 3.4.2 (c). Hence, $X_n \to_p X$ ■

The converse may not hold: Let

$$P\left(X_n = 0\right) = 1 - n^{-1}, \quad P\left(X_n = 1\right) = n^{-1}$$

and X_n 's are independent. Then $X_n \to_p 0$ since $P\left(|X_n - 0| > \epsilon\right) \leq n^{-1} \to 0$.

However, $X_n \not\to 0$ a.s. since for any $0 < \epsilon < 1$, we have

$$P\left(\bigcap_{m\geq n}\{|X_m - 0| \leq \epsilon\}\right) = P\left(\lim_{r\to\infty}\bigcap_{m=n}^{r}\{|X_m| \leq \epsilon\}\right) = \lim_{r\to\infty} P\left(\bigcap_{m=n}^{r}\{|X_m| \leq \epsilon\}\right)$$
$$= \lim_{r\to\infty}\prod_{m=n}^{r} P(|X_m| \leq \epsilon) = \lim_{r\to\infty}\prod_{m=n}^{r}\left(1 - m^{-1}\right)$$
$$= \lim_{r\to\infty}\frac{n-1}{n}\frac{n}{n+1}\cdots\frac{r-1}{r} = \lim_{r\to\infty}\frac{n-1}{r} = 0$$

By the equivalent definition of a.s., we see that $X_n \not\to 0$ a.s.

(b) If $X_n \to X$ in L_r, then $X_n \to_p X$.

Proof. By Markov inequality, $0 \leq P(|X_n - X| \geq \epsilon) \leq E|X_n - X|^r / \epsilon^r \to 0$. Thus, $X_n \to_p X$. ■

The converse may not hold: Let $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = n) = n^{-1}$. Then $X_n \to_p 0$, but $EX_n = 1 \to 0$

(c) If $X_n \to_p X$, then $X_n \to_d X$

Proof. Denote $F_n(x) = P(X_n \leq x)$ and $F(x) = P(X \leq x)$. First we have

$$\begin{aligned}F_n(x) &= P(X_n \leq x, |X_n - X| \leq \epsilon) + P(X_n \leq x, |X_n - X| > \epsilon)\\ &\leq P(X \leq x - (X_n - X), |X_n - X| \leq \epsilon) + P(|X_n - X| > \epsilon)\\ &\leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon)\\ &= F(x + \epsilon) + P(|X_n - X| > \epsilon)\end{aligned}$$

On the other hand,

$$\begin{aligned}F_n(x) &= 1 - P(X_n > x)\\ &= 1 - P(X_n > x, |X_n - X| \leq \epsilon) - P(X_n > x, |X_n - X| > \epsilon)\\ &\geq 1 - P(X > x - (X_n - X), |X_n - X| \leq \epsilon) - P(|X_n - X| > \epsilon)\\ &\geq 1 - P(X > x - \epsilon) - P(|X_n - X| > \epsilon)\\ &= F(x - \epsilon) - P(|X_n - X| > \epsilon)\end{aligned}$$

Combining the two, we have

$$F(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) + P(|X_n - X| > \epsilon)$$

Letting $n \to \infty$, we obtain

$$F(x - \epsilon) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x + \epsilon)$$

If $F(x)$ is continuous at x, then as $\epsilon \downarrow 0$, we have $F(x - \epsilon) \uparrow F(x)$ and $F(x + \epsilon) \downarrow F(x)$, the result is proved.

The converse may not hold: Let $X \sim N(0,1)$ and $X_n = -X \sim N(0,1)$. Then $X_n =_d X$ but $X_n \not\to_p X$ as $P(|X_n - X| \geq \epsilon) = P(2|X| \geq \epsilon) \neq 0$ ■

2. Trivial since $0 \leq \left(E|X_n - X|^s\right)^{1/s} \leq \left(E|X_n - X|^r\right)^{1/r} \to 0$ as $n \to \infty$.

   The converse may not hold: Let $P(X_n = 0) = 1 - n^{-2}$ and $P(X_n = n) = n^{-2}$. Then $X_n \to 0$ in L^1 as $E|X_n - 0| = n \cdot \frac{1}{n^2} = 1/n \to 0$, but $X_n \neq 0$ in L^2 as $E|X_n - 0|^2 = n^2 \cdot \frac{1}{n^2} = 1 \neq 0$

3. We now show that "a.s. convergence" and "mean convergence" do not imply each other.

   (a) Example (a): Let $P(X_n = 0) = 1 - n^{-2}$ and $P\left(X_n = n^3\right) = n^{-2}$. Then $X_n \to 0$ a.s. but $X_n \not\to 0$ in L^1

   Proof. $X_n \to 0$ a.s. Since $\sum_{n=1}^{\infty} P(|X_n - 0| \geq \epsilon) = \sum_{n=1}^{\infty} n^{-2} < \infty$, by p series test. However, $X_n \neq 0$ in L^1 as $E|X_n - 0| = n \to \infty$. (The good idea to make a example to prove that r.v. does not converge in L^1 is let $P(X_n \neq 0) = n^{-2}$.) ■

   (b) Example (b): Let $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = 1) = n^{-1}$, and they are independent. Then $X_n \to 0$ in L^1, but $X_n \not\to 0$ a.s.

   Proof. Then $X_n \to 0$ in L^1 as $E|X_n - 0| = 1/n \to 0$ However, it was shown earlier that $X_n \neq 0$ a.s. ■

■

3.4.1 Partial converses

We now provide some partial converses to results in the last section, i.e., converse results with some additional assumptions.

Theorem 3.4.5 — Convergence in probability and distribution to constants are equivalent. $X_n \to_d C \iff X_n \to_p C$, where C is a constant.

Proof. We only need to show the part "$\Longrightarrow$". Given $\epsilon > 0$, we have

$$\begin{aligned}
P(|X_n - C| > \epsilon) &= P(X_n > C + \epsilon) + P(X_n < C - \epsilon) \\
&\leq P(X_n > C + \epsilon) + P(X_n \leq C - \epsilon) \\
&= [1 - F_{X_n}(C + \epsilon)] + F_{X_n}(C - \epsilon) \\
&\to [1 - F_C(C + \epsilon)] + F_C(C - \epsilon) = 1 - 1 + 0 \\
&= 0
\end{aligned}$$

where we used that $C \pm \epsilon$ are continuity points of $F_C(x)$, which is degenerate at C. ■

Dominated convergence in probability implies convergence in mean

Lemma 3.7 — Similar to uniformly convergence. If $X_n \to_p X, |X_n| \leq Y$ a.s. (i.e. $P(|X_n| \leq Y) = 1$) for all n, then $|X| \leq Y$ a.s. ■

Proof. Given $\delta > 0$, as $n \to \infty$, we have

$$\begin{aligned} P(|X| > Y + \delta) &= P(|X| > Y + \delta, |X_n| \le Y) + P(|X| > Y + \delta, |X_n| > Y) \\ &\le P(|X| > |X_n| + \delta, |X_n| \le Y) + P(|X_n| > Y) \\ &\le P(|X_n - X| > \delta) + 0 \end{aligned}$$

Letting $n \to \infty$, we get $P(|X| > Y + \delta) = 0$, i.e., $|X| \le Y + \delta$ a.s. for any $\delta > 0$. Letting $\delta = 0$, we get the desired result. ■

Lemma 3.8 If $E|Y| < \infty$, and $P(A_n) \to 0, n \to \infty$, then $E_{A_n}|Y| \to 0$. (In particular, $E[|Y|I_{\{|Y|>n\}}] \to 0$ by the Monotone Convergence Theorem. The result is also trivial by using u.i. of X to be introduced later.)

Proof. Since $E|Y| < \infty, \forall \epsilon > 0, \exists A_\epsilon > 0$ s.t. $E[|Y|I_{\{|Y|>A_\epsilon\}}] \le \epsilon$. We thus have

$$\begin{aligned} E_{A_n}|Y| &= E\left[|Y|I_{\{|Y|>A_\epsilon\}}I_{A_n}\right] + E\left[|Y|I_{\{|Y|\le A_\epsilon\}}I_{A_n}\right] \\ &\le E\left[|Y|I_{\{|Y|>A_\epsilon\}}\right] + A_\epsilon E I_{A_n} \\ &\le \epsilon + A_\epsilon P(A_n) \end{aligned}$$

Since $P(A_n) \to 0, n \to \infty$, the RHS $\le 2\epsilon$ for sufficiently large n. ■

Theorem 3.4.6 — Lebesgue Dominated Convergence Theorem. If $X_n \to_p X, |X_n| \le Y$ a.s. for all n, and $EY^r < \infty$ for $r > 0$, then $X_n \to X$ in L_r, which in turn implies that $EX_n^r \to EX^r$

Proof. We shall give three proofs.

1. Method 1: via u.i.. It is special case of Theorem 3.4.24.
2. Method 2: direct approach. From Lemma 3.7, we have $|X_n - X| \le 2Y$ a.s. Now choose and fix $\epsilon > 0$. Since $EY^r < \infty$ there exists a finite constant $A_\epsilon > \epsilon > 0$ s.t. $E\left[Y^r I_{\{2Y>A_\epsilon\}}\right] \le \epsilon$. We thus have

$$\begin{aligned} E|X_n - X|^r =& E|X_n - X|^r I_{\{|X_n-X|>A_\epsilon\}} + E|X_n - X|^r I_{\{|X_n-X|\le\epsilon\}} \\ &+ E|X_n - X|^r I_{\{\epsilon<|X_n-X|\le A_\epsilon\}} \\ \le& E(2Y)^r I_{\{2Y>A_\epsilon\}} + \epsilon^r + A_\epsilon^r P(|X_n - X| > \epsilon) \\ \le& 2^r E(Y)^r I_{\{2Y>A_\epsilon\}} + \epsilon^r + A_\epsilon^r P(|X_n - X| > \epsilon) \\ \le& 2^r\epsilon + \epsilon^r + A_\epsilon^r P(|X_n - X| > \epsilon) \end{aligned}$$

Since $P(|X_n - X| > \epsilon) \to 0, n \to \infty$, the RHS $\le 2^r\epsilon + 2\epsilon^r$ for sufficiently large n

3. Method 3. From Lemma 3.7 , we have $|X_n - X| \leq 2Y$ a.s. So

$$\begin{aligned} E|X_n - X|^r &= E|X_n - X|^r I_{\{|X_n - X| > \epsilon\}} + E|X_n - X|^r I_{\{|X_n - X| \leq \epsilon\}} \\ &\leq 2^r EY^r I_{\{|X_n - X| > \epsilon\}} + \epsilon^r \end{aligned}$$

From Lemma 3.8, the first term on RHS, $EY^r I_{\{|X_n - X| > \epsilon\}} \to 0$ for sufficiently large n

■

Proposition 3.4.7 1. The main theorem weakens the a.s. convergence condition in the Dominated Convergence Theorem to just convergence in probability. Recall that the Dominated Convergence Theorem states:

> **Theorem 3.4.8 — Dominated Convergence Theorem.** If $X_n \to X$ a.s. and $P(|X_n| \leq Y) = 1$ for all n, and $EY < \infty$, then $EX_n \to EX$, namely, $X_n \to X$ in L^1

2. Similar to Remark (1), the a.s. convergence in the well-known Fatou's Lemma can also be weakened to just convergence in probability, or even in distribution.

> **Theorem 3.4.9 — Another Fatou's Lemma.** If $X_n \geq 0$ and $X_n \to_p X$, then
>
> $$EX = E \liminf_n X_n \leq \liminf_n EX_n$$

3. The theorem states: convergence in probability and plus dominance (i.e. $|X_n| < Y$ with $EY < \infty$) imply convergence in mean. However, the dominance condition is often too strong, and can be replaced by a weaker condition uniformly integrability condition, to be studied later.

> **Corollary 3.4.10 — Bounded convergence in probability implies mean convergence–See Y as constant.** If $X_n \to p\ X$ and $P(|X_n| \leq C) = 1$ for all n and some C, then $X_n \to X$ in L_r for all $r > 0$

> **Corollary 3.4.11** Suppose that $|X_n| \leq C$ a.s. for all n and some C. Then for $r > 0$,
>
> $$X_n \to X \text{ in } L_r \quad \Longleftrightarrow \quad X_n \to_p X$$

> **Corollary 3.4.12** $X_n \to_p 0$ iff $E(|X_n| / (1 + |X_n|)) \to 0$

Proof. Note that $|X_n| / (1 + |X_n|)$ is bounded by 1. Hence

$$E\left(\frac{|X_n|}{1 + |X_n|}\right) \to 0, \quad \Longleftrightarrow \quad \frac{|X_n|}{1 + |X_n|} \to_p 0 \quad \Longleftrightarrow \quad X_n \to_p 0$$

The last equivalence follows since

$$P(|X_n| \geq \epsilon) = P\left(\frac{|X_n|}{1 + |X_n|} \geq \frac{\epsilon}{1 + \epsilon}\right)$$

■

Dominated convergence a.s. implies convergence in mean

Theorem 3.4.13 If $X_n \to X$ a.s., $P(|X_n| \le Y) = 1$ for all n, and $EY^r < \infty$ for $r \ge 0$, then $X_n \to X$ in L_r

Convergence in probability sufficiently fast implies a.s. convergence

Theorem 3.4.14 If $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \to X$ a.s.

Proof. Given $\epsilon > 0$, we have $P(\bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\}) \le \sum_{m=n}^{\infty} P(|X_m - X| > \epsilon) \to 0$ ■

Definition 3.4.2 — Converge Completely. $\{X_n, n \ge 1\}$ is said to **converge completely** to X if $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$. Thus, complete convergence implies a.s. convergence

Convergence in mean sufficiently fast implies a.s. convergence

Theorem 3.4.15 If $\sum_{n=1}^{\infty} E|X_n - X|^r < \infty$ for some $r > 0$, then $X_n \to X$ a.s.

Proof. By Markov's inequality, $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) \le \sum_{n=1}^{\infty} E|X_n - X|^r / \epsilon^r < \infty$, then use Theorem 3.4.14. ■

Theorem 3.4.16 — Convergence sequences in probability contains a.s. subsequences. If $X_n \to_p X$, then there exists a non-random integers $n_1 < n_2 < \ldots$ such that $X_{n_i} \to X$ a.s.

Proof. $X_n \to_p X$ implies, for any $\epsilon > 0, \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$. Thus, for every $k \ge 1$, we can find some large positive integer n_k such that $P(|X_{n_k} - X| > 1/k) \le 1/k^2$. Clearly, $\{n_k, k \ge 1\}$ can be easily made to be increasing. Thus,

$$\sum_{k>1/\epsilon} P(|X_{n_k} - X| > \epsilon) \le \sum_{k>1/\epsilon} P(|X_{n_k} - X| > 1/k) \le \sum_{k>1/\epsilon} 1/k^2 < \infty$$

$(\epsilon > \frac{1}{k})$ The proof then follows from Theorem 3.4.14. ■

Theorem 3.4.17 — Convergence in distribution plus uniform integrability implies convergence in moments. Suppose that $X_n \to_d X$, and the sequence $\{X_n^r\}$ is uniformly integrable $(u.i.)$, where $r > 0$. Then $E|X|^r < \infty$, and

$$\lim_n EX_n^r = EX^r, \quad \lim_n E|X_n|^r = E|X|^r$$

Convergence in distribution implies a.s. convergence in another probability space.

Let $\mathcal{B}_{[0,1]}$ denote the Borel sets in [0,1] and $\lambda_{[0,1]}$ the Lebesgue measure restricted to [0,1]

Theorem 3.4.18 — Skorokhod's representation theorem. Suppose that $X_n \to_d X$. Then there exist r.v.'s Y and $\{Y_n, n \ge 1\}$ on $\left((0,1), \mathcal{B}_{(0,1)}, P_\lambda = \lambda_{(0,1)}\right)$ s.t.

1. Y_n and Y have the same d.f.'s as X_n and X. That is, $X_n =_d Y_n, X =_d Y$.

2. $Y_n \to Y$ a.s. as $n \to \infty$

Proof. 1. Let $F_n(x)$ and $F(x)$ denote the d.f.'s of X_n and X, respectively. For $t \in (0,1)$, define

$$Y_n(t) = F_n^{-1}(t) = \inf\{x : F_n(x) \geq t\}, \quad Y(t) = F^{-1}(t) = \inf\{x : F(x) \geq t\}$$

From Theorem 3.1.17, we have

$$t \leq F_n(x) \Longleftrightarrow Y_n(t) \leq x. \quad \text{and}$$

$$t \leq F(x) \Longleftrightarrow Y(t) \leq x$$

Therefore,

$$P_\lambda(Y \leq x) = P_\lambda(\{t : Y(t) \leq x\}) = P_\lambda(\{t : t \leq F(x)\}) = F(x)$$

Similarly, we can show that $P_\lambda(Y_n \leq x) = F_n(x)$.

2. First fix $\epsilon > 0$ and $t \in (0,1)$. Pick a point x of continuity of F such that

$$Y(t) - \epsilon < x < Y(t)$$

[Such an x must exist in $(Y(t) - \epsilon, Y(t))$ since all discontinuity points are countable.] By (a) above, we get $F(x) < t$ (for $x < Y(t)$). Since $F_n \to F$, then $F_n(x) < t$ for all large n, implying that $Y_n(t) > x$ for all large n. Then we have

$$Y(t) - \epsilon < x < Y_n(t), \quad \text{for all large } n$$

Letting $n \to \infty$, we get

$$\liminf_n Y_n(t) \geq Y(t), \quad \text{for all } t \in (0,1)$$

Next, if $0 < t < t' < 1$, pick a point x of continuity of F such that

$$Y(t') < x < Y(t') + \epsilon$$

By (a) above again, $t < t' \leq F(x)$ and so $t < F_n(x)$ for all large n. Since $F_n \to F$, giving that

$$Y_n(t) \leq x < Y(t') + \epsilon, \quad \text{for all } t \in (0,1)$$

Letting $n \to \infty$ and $\epsilon \to 0$, we obtain

$$\limsup_n Y_n(t) \leq Y(t') \quad \text{whenever } 0 < t < t' < 1$$

Thus,

$$Y(t) \leq \liminf_n Y_n(t) \leq \limsup_n Y_n(t) \leq Y(t') \quad \text{whenever } 0 < t < t' < 1$$

Now if t is a point of continuity of $Y(t)$, then letting $t' \downarrow t$, we obtain

$$\lim_n Y_n(t) = Y(t)$$

However, since $Y(t)$ is nondecreasing, and so the set D of discontinuities of Y is countable, thus,

$$0 \le P_\lambda\left(\left\{t : \lim_n Y_n(t) \neq Y(t)\right\}\right) \le P_\lambda(D) = 0$$

■

R

1. The theorem is "constructive", not existential, as is demonstrated by the proof.
2. The theorem is useful in proving theorems on convergence of moments. Some applications can be found in Theorem 3.4.26 (c) or Helly's Theorem later.

3.4.2 Convergence of moments; uniform integrability

We know that convergence in probability does not imply convergence in mean (e.g. Let $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = n) = n^{-1}$). We also learnt that with an extra condition $|X_n| < Y$ with $EY < \infty$, then $EX_n \to EX$ (i.e. Dominated Convergence Theorem.) However, the condition $|X_n| < Y$ with $EY < \infty$ is often too strong in cases of interest. A weaker condition is uniform integrability, which is, in fact, a necessary and sufficient condition for a convergent sequence in probability to converge in mean.

First definition of u.i.

Proposition 3.4.19 For a single r.v. X, it can be easily shown from the DCT that X is integrable, i.e., $X \in L^1 = L^1(\Omega, \mathcal{F}, P)$

$\Longleftrightarrow E|X|I\{|X| > K\} \to 0$ as $K \to \infty$

$$\left(\text{as } P(|X|I\{|X| > K\} > \epsilon) = P(|X| > K) \le E|X|/K \to 0 \text{ and } |X|I\{|X| > K\} \le |X| \in L^1.\right)$$

$\Longleftrightarrow \forall \varepsilon > 0, \exists K > 0$ such that $E|X|I\{|X| > K\} \le \varepsilon$ as $K \to \infty$.

This motivates the notion of uniform integrability for a collection of r.v.s $Y_n, n \ge 1$ by requiring $E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \to 0$ as $C \to \infty$ uniformly in n :

Definition 3.4.3 — Definition of uniform integrability (u.i.). A sequence of r.v.'s $\{Y_n, n \ge 1\}$ on $(\Omega, \mathcal{A}, P)$ is **u.i.** if and only if

$$\lim_{C \to \infty} \sup_{n \ge 1} E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} = 0$$

R

1. As can be seen from Theorem 3.4.20 below, Definition 3.4.3 implies

$$\sup_n E|Y_n| \le M < \infty$$

which means that Y_n are all "integrable together". But this does not mean "uniform integrable"; see Theorem 3.4.20 below.

2. A more general definition than Definition 3.4.3 can be given as follows: A collection of r.v.s $\{X_i, i \in I\}$ is said to be u.i. if $\lim_{C\to\infty}\sup_{i\in I} E\left\{|X_i| I_{\{|X_i|>C\}}\right\} = 0$

Second definition of u.i.

Lemma 3.9 — An "absolute continuity" property. If X is integrable, i.e., $X \in L^1$, then, $Q(A) := E_A|X|$ is absolutely continuous, i.e., $\forall \varepsilon > 0, \exists \delta > 0$ such that for any $A \in \mathcal{A}$,

$$Q(A) := E_A|X| \equiv E|X|I_A < \epsilon, \quad \text{whenever} \quad P(A) < \delta$$

Proof. If the conclusion is false, then for some $\varepsilon_0 > 0$, we can find a sequence (A_n) of r. v.s such that

$$P(A_n) < 2^{-n}, \quad \text{and} \quad E|X|I_{A_n} \ge \epsilon_0$$

Let $H := \limsup A_n$. Since $\sum_n P(A_n) < \infty$, then by the Borel-Contelli lemma, we have

$$P(H) = P(\limsup A_n) = P(A_n, \text{ i.o }) = 0$$

which implies that

$$E_H|X| = E|X|I_H = 0$$

On the other hand, by the reverse Fatou's lemma, we have

$$E_H|X| = E|X|I_H = E|X|I_{\limsup A_n} = E\limsup |X|I_{A_n} \ge \limsup E|X|I_{A_n} \ge \epsilon_0$$

A contradiction. ∎

Combining the above "integrable together" property and "absolute continuity" property, we will have another definition of u.i.

Theorem 3.4.20 — An equivalent definition of u.i.. A sequence of r.v.'s $\{Y_n\}$ on $(\Omega, \mathcal{A}, P)$ is u.i. if and only if

1. $\sup_n E|Y_n| < \infty$, and

2. $\forall \epsilon > 0, \exists \delta > 0$ such that for any $A \in \mathcal{A}$

$$\sup_n E_A |Y_n| \equiv \sup_n E |Y_n| I_A < \epsilon, \quad \text{whenever} \quad P(A) < \delta$$

Proof. 1. Sufficiency. Suppose $\{Y_n\}$ is u.i., let us show that (a), (b) hold. Note that

$$\begin{aligned}\sup_n E|Y_n| &\leq \sup_n E\left\{|Y_n| I_{\{|Y_n| \leq C\}}\right\} + \sup_n E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \\ &\leq C + \sup_n E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \\ \sup_n E_A|Y_n| &\leq \sup_n E_A\left\{|Y_n| I_{\{|Y_n| \leq C\}}\right\} + \sup_n E_A\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \\ &\leq CP(A) + \sup_n E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\}\end{aligned}$$

For any $\epsilon > 0$, choose and fix a large enough C so that $\sup_n E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \leq \epsilon$. Also choose $\delta \leq \epsilon / C$. Then $\sup{}_n E|Y_n| \leq C + 1 < \infty$ and $E_A|Y_n| \leq \epsilon$

2. Necessity. Suppose (a) and (b) hold, we now show $\{Y_n\}$ is u.i. First, from (a), $M =: \sup_n E|Y_n| < \infty$. For any $\epsilon > 0$, choosing $\delta > 0$ as in (b), the Chebyshev's inequality ensures that, for every $n \geq 1$

$$P(|Y_n| > C) \leq \frac{E|Y_n|}{C} \leq \frac{\sup_n E|Y_n|}{C} \leq \frac{M}{C} < \delta$$

If we choose $C > M/\delta$. Consequently, from (b), with the choice of $A =: \{|Y_n| > C\}$, we have

$$\sup_n E_A|Y_n| = \sup_n E|Y_n| I_{\{|Y_n| > C\}} < \epsilon$$

That is, $\lim_{C \to \infty} \sup_n E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} = 0$

■

R Recall that Lemma 3.8 states: If $E|Y| < \infty$, and $P(A_n) \to 0, n \to \infty$, then $E[|Y| I_{A_n}] \to 0$. Then it seems that (a) could imply (b) in the above theorem. However, this is not true. We must have uniform integrability condition.

Theorem 3.4.21 1. $\{X_n\}$ is u. i. iff $\{|X_n|\}$ is u.i.

2. If $|X_n| \leq |Y_n|$, and $\{Y_n\}$ is u.i., then $\{X_n\}$ is u.i.

   Proof. $\sup_{n \geq 1} E\left\{|X_n| I_{\{|X_n| > C\}}\right\} \leq \sup_{n \geq 1} E\left\{|Y_n| I_{\{|Y_n| > C\}}\right\} \to 0$ ■

3. $\{X_n\}$ is u.i. iff $\{X_n^+\}$ and $\{X_n^-\}$ are both u.i.
4. If $\{X_n\}$ and $\{Y_n\}$ are each u.i., so is $\{X_n + Y_n\}$.
5. If $\{X_n\}$ is u.i., so is any subsequence of $\{X_n\}$.
6. If $|X_n| \leq Y \in L^1$, then $\{X_n\}$ is u.i. (i.e., the Lesbegue DCT.)

Proof. $\sup_{n\geq 1} E\left\{|X_n| I_{\{|X_n|>C\}}\right\} \leq E\left\{|Y| I_{\{|Y|>C\}}\right\} \to 0$ ■

7. If $E\left(\sup_n |X_n|\right) < \infty$, then $\{X_n\}$ is u.i. (Note that $E\left(\sup_n |X_n|\right) \geq \sup_n E|X_n|$.)

   Proof. Take $Y = \sup_n |X_n|$ in the *DCT* above. ■

8. 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*

   Proof. $\forall M = \epsilon^{-1} > 0, \exists x_0 > 0$ such that $\frac{|\psi(x)|}{|x|} \geq M$ for large $|x| \geq x_0$

$$\sup_n E\left\{|Y_n| I_{\{|Y_n|>C\}}\right\} \leq M^{-1} \sup_n E\left\{\psi(|Y_n|) I_{\{|Y_n|>C\}}\right\} \leq M^{-1} \sup_n E\psi(|Y_n|) =: \epsilon A$$

   Letting $C \to \infty$ and then letting $\epsilon \to 0$. ■

R We need x to be dominated by $\psi(x)$, e.g. $\psi(x) = x\log^+ x$). In practice, we may choose $\psi(x) = x^p$ with $p > 1$, and u.i. is checked by using moments. For instance, when $p = 2$, square integrability implies u.i.

Theorem 3.4.22 — Vitali's Theorem–Convergence in prob. + u.i. $\Rightarrow$ convergence in mean. Suppose that $X_n \to_p X$, and $E|X_n|^r < \infty$ all n (i.e. $X_n \in L_r$). Then the following three statements are equivalent.

1. $\{X_n^r\}$ is u.i.
2. $X_n \to X$ in L_r; and $E|X|^r < \infty$
3. $E|X_n|^r \to E|X|^r < \infty$

Proof. 1. "$(i) \implies (ii)$". Suppose (i) holds. We show (ii) in 3 steps.

(a) We first show that $E|X|^r < \infty$.

Proof. Since $X_n \to_p X, \exists$ a subsequence $\{n_k\}$ such that $\lim_{k\to\infty} X_{n_k} = X$ a.s., thus $\lim_{k\to\infty} |X_{n_k}|^r = |X|^r$ a.s. (See Theorem 3.4.26.) By Fatou's Lemma,

$$E|X|^r = E\lim_{k\to\infty} |X_{n_k}|^r = E\liminf_k |X_{n_k}|^r \leq \liminf_k E|X_{n_k}|^r \leq \sup_n E|X_n|^r < \infty$$

where the last inequality follows from assumption (i): $\{X_n^r\}$ is u.i. ■

(b) Secondly, we show that $\left\{|X_n - X|^r\right\}$ is u.i.

Proof. We will use C_r inequality $|X_n - X|^r \leq C_r\left(|X_n|^r + |X|^r\right)$. See Lemma 3.10 later on. Then apply Theorem 3.4.21, part 2 . ■

(c) Finally, we show that $X_n \to X$ in L_r

Proof. Fix $\epsilon_0 > 0$, we have $E|X_n - X|^r I_{\{|X_n - X|>\epsilon_0\}} \to 0$ as $P(|X_n - X| > \epsilon_0) \to 0$.

Hence,

$$\begin{aligned} E\left|X_n - X\right|^r &= E\left|X_n - X\right|^r I_{\{|X_n - X| \le \epsilon_0\}} + E\left|X_n - X\right|^r I_{\{|X_n - X| > \epsilon_0\}} \\ &\le \epsilon_0^r + E\left|X_n - X\right|^r I_{\{|X_n - X| > \epsilon_0\}} \\ &\to \epsilon_0^r \end{aligned}$$

Thus, $X_n \to X$ in L_r ■

2. "$(ii) \implies (iii)$". See Theorem 3.4.23 below.
3. "$(iii) \implies (i)$"

Proof. Suppose (iii) is true. To prove (i), let $A > 0$ and construct a nonnegative and continuous function by

$$\begin{aligned} f_A(x) &= |x|^r & |x|^r \le A \\ &\le |x|^r & A < |x|^r \le A+1 \\ &= 0 & |x|^r > A+1 \end{aligned}$$

Since $f_A(x)$ is bounded and continuous, by Helly's theorem, $\lim_n Ef_A(X_n) = Ef_A(X)$. Also note that $0 \le |x|^r I_{\{|x|^r \le A\}} \le f_A(x) \le |x|^r I_{\{|x|^r \le A+1\}}$, so

$$\begin{aligned} \liminf_n E\left|X_n\right|^r I_{\left\{|X_n|^r \le A+1\right\}} \ge \liminf_n Ef_A(X_n) = \lim_n Ef_A(X_n) = Ef_A(X) \\ \ge E|X|^r I_{\{|X|^r \le A\}} \end{aligned}$$

Subtracting this from the assumption (iii): $\lim_n E\left|X_n\right|^r = E|X|^r < \infty$, we get

$$\limsup_n E\left|X_n\right|^r I_{\{|X_n|^r > A+1\}} \le E|X|^r I_{\{|X|^r > A\}}$$

The last integral does not depend on n and converges to zero as $A \to \infty$. This means: $\forall \epsilon > 0\ \exists A_0 = A_0(\epsilon)$, and $n_0 = n_0\left(A_0(\epsilon)\right)$ such that we have

$$\sup_{n \ge n_0} E\left|X_n\right|^r I_{\left\{|X_n|^r > A+1\right\}} \le \epsilon, \quad \text{if } A \ge A_0$$

However, since each X_n^r is integrable, $\exists A_1 = A_1(\epsilon)$ such that

$$\sup_{n \ge 1} E\left|X_n\right|^r I_{\left\{|X_n|^r > A+1\right\}} \le \epsilon, \quad \text{if } A \ge \max\left\{A_0, A_1\right\}$$

That proves (i). ■

■

Theorem 3.4.23 — Relationship between L_r convergence and convergence of moments. Suppose that $X_n \to X$ in $L_r (r > 0)$, and $E|X|^r < \infty$. Then

1. $\lim_n E\left|X_n\right|^r = E|X|^r$
2. $\lim_n EX_n^r = EX^r$

Proof. 1. For $0 < r \leq 1$, apply the C_r inequality $|x+y|^r \leq |x|^r + |y|^r$ (see Lemma 3.10 later) to write $|x|^r - |y|^r \leq |x-y|^r$ and thus

$$|E|X_n|^r - E|X|^r| \leq E||X_n|^r - |X|^r| \leq E|X_n - X|^r$$

From the assumptions, we get $\lim_n E|X_n|^r = E|X|^r$. For $r > 1$, apply Minkowski's inequality to obtain

$$\left|\left(E|X_n|^r\right)^{1/r} - \left(E|X|^r\right)^{1/r}\right| \leq \left(E|X_n - X|^r\right)^{1/r}$$

From the assumptions, we get $\lim_n \left(E|X_n|^r\right)^{1/r} = \left(E|X|^r\right)^{1/r}$, i.e. $\lim_n E|X_n|^r = E|X|^r$

2. It follows from (i) and Vitali Theorem that $\{X_n^r\}$ is u.i. Also from the assumptions, we have $X_n \to_d X$. Then (ii) follows from Theorem 3.4.24.

■

R

1. The mean convergence $X_n \to X$ in $L_r (r > 0)$ does not imply $X_n \in L_r$ or $X \in L_r$. For example, take $X_n \equiv X = Cauchy$, then $X_n \to X$ in L^1, but $X_n = X \notin L_1$
2. If $X_n \to X$ in L_r, then the mean convergence $X_n \to X$ in L_r implies convergence in moments (from the last theorem). The converse may not be true. Take $r = 2, X_{2n+1} = X$, and $X_{2n} = -X$ with $0 < EX^2 < \infty$. Clearly, $X_n \to X$ in L_r, but $X_n \not\to X$ in L_r

Convergence in dist. + u.i. ⇒ convergence in mean

Theorem 3.4.24 Suppose that $X_n \to_d X$, and $\{X_n^r\}$ is u.i. $(r > 0)$. Then

1. $E|X|^r < \infty$
2. $\lim_n EX_n^r = EX^r$
3. $\lim_n E|X_n|^r = E|X|^r$

Proof. 1. Method 1. Use Skorokhod's representation theorem3.4.18, we find $\{Y_n\}$ and Y such that $Y_n \to Y$ a.s.
   (a) Y_n and Y have the same d.f.'s as X_n and X. That is, $X_n =_d Y_n, X = {}_dY$.
   (b) $Y_n \to Y$ a.s. as $n \to \infty$

   Then the proof follows from the results on Section 3.4.22 to Y_n : Convergence in prob. + u.i. $\Longrightarrow$ convergence in mean.

2. Method 2: Direct approach.
   (a) Let $F(x) = P(X \leq x)$. Fix $\epsilon > 0$. Choose C s.t. $\pm C$ are continuity points of F (This can be done a.s. since the set of all discontinuities are countable). By the u.i. assumption, we have, as C is large enough,

$$\sup_n E|X_n^r| I_{\{|X_n| \geq C\}} \leq \epsilon$$

For any $D > C$ s.t. $\pm D$ are also continuity points of F, we obtain from the Helly's theorems that

$$\lim_n E|X_n^r| I_{\{C \le |X_n| \le D\}} = E|X^r| I_{\{C \le |X| \le D\}}$$

It follows that $E|X^r| I_{\{C \le |X| \le D\}} < \epsilon$ for all such choices of D. Letting $D \to \infty$, we get $E|X^r| I_{\{|X| \ge C\}} < \epsilon$. So we prove (i).

(b) For the same C as above, write

$$|EX_n^r - EX^r| \le \left|EX_n^r I_{\{|X_n| \le C\}} - EX^r I_{\{|X| \le C\}}\right| + E|X_n|^r I_{\{|X_n| > C\}} + E|X|^r I_{\{|X| > C\}}$$

By the Helly's theorems again, the first term on RHS tends to 0 as $n \to \infty$ while the other two terms are less than ϵ. Thus, (ii) is proved.

(c) Proof is similar to that of (ii).

■

3.4.3 Some closed operations of convergence

Algebraic operations

Theorem 3.4.25 — In prob, a.s. and r th is closed in addition, while dist is not.

1. If $X_n \to_p X$ and $Y_n \to_p Y$, then $X_n \pm Y_n \to_p X \pm Y$
2. If $X_n \to X$ a.s. and $Y_n \to Y$ a.s., then $X_n \pm Y_n \to X \pm Y$ a.s.
3. If $X_n \to_r X$ and $Y_n \to_r Y$, then $X_n \pm Y_n \to_r X \pm Y$
4. However, if $X_n \to_d X$ and $Y_n \to_d Y$, then it is not true in general that $X_n \pm Y_n \to_d X \pm Y$.

Proof. It is clear that $Y_n \to Y$ in prob, a.s. or L_r implies $-Y_n \to -Y$ in prob, a.s. or L_r. So we only need to show $+$ relations in the theorem.

1. Given $\epsilon > 0$

$$\begin{aligned}
0 &\le P(|X_n + Y_n - X - Y| > 2\epsilon) \\
&\le P\left(\{|X_n - X| > \epsilon\} \bigcup \{|Y_n - Y| > \epsilon\}\right) \\
&\le P(|X_n - X| > \epsilon) + P(|Y_n - Y| > \epsilon) \to 0
\end{aligned}$$

2. Given $\epsilon > 0$

$$
\begin{aligned}
0 \leq & P\left(\bigcup_{m \geq n} \{|X_m + Y_m - X - Y| > 2\epsilon\}\right) \\
\leq & P\left(\bigcup_{m \geq n} \left[\{|X_m - X| > \epsilon\} \bigcup \{|Y_m - Y| > \epsilon\}\right]\right) \\
\leq & P\left(\left[\bigcup_{m \geq n} \{|X_m - X| > \epsilon\}\right] \bigcup \left[\bigcup_{m \geq n} \{|Y_m - Y| > \epsilon\}\right]\right) \\
\leq & P\left(\bigcup_{m \geq n} \{|X_m - X| > \epsilon\}\right) + P\left(\bigcup_{m \geq n} \{|Y_m - Y| > \epsilon\}\right) \\
\to & 0
\end{aligned}
$$

3. First we need a lemma:

Lemma 3.10 — C_r-inequality. $|x + y|^r \leq C_r\left(|x|^r + |y|^r\right)$, where $r > 0$ and

$$
\begin{aligned}
C_r = & 1 \qquad \text{if } 0 < r < 1 \\
& 2^{r-1} \qquad \text{if } r \geq 1
\end{aligned}
$$

(The lemma clearly implies that $|x + y|^r \leq 2^r\left(|x|^r + |y|^r\right)$.)

Proof. When $r \geq 1$, then the proof follows from $|(x+y)/2|^r \leq \left(|x|^r + |y|^r\right)/2$. When $0 < r < 1$, then $\lambda^r + (1-\lambda)^r \geq \lambda + (1-\lambda) = 1$ for $\lambda = |x|/(|x| + |y|)$. ■

Proof of (c)

$$
\begin{aligned}
0 \leq & E|X_n + Y_n - X - Y|^r = E|(X_n - X) + (Y_n - Y)|^r \\
\leq & C_r\left(E|X_n - X|^r + E|Y_n - Y|^r\right) \to 0
\end{aligned}
$$

Note that when $r \geq 1$, one could also use the Minkovski's inequality.

4. $X_n \to_d X, Y_n \to_d Y$, but we may not have $X_n + Y_n \to_d X + Y$

■ **Example 3.12** Take $X_n \sim N(0,1), Y_n = -X_n \sim N(0,1)$. Then $X_n + Y_n = 0$

Now take independent $X, Y \sim N(0,1)$, then $X + Y \sim N(0,2)$

■

Exercise 3.1 (See Q3, Chung, K.L. p70.) If $X_n \to_p X$ and $Y_n \to_p Y$, then $X_n Y_n \to_p XY$. What about a.s. and L_r convergence?

R We have just seen that (d) may not hold in general. However, it does hold if $Y_n \to_d Y$ is replaced by $Y_n \to_d C$, which is also equivalent to $Y_n \to_p C$. This is the well-known Slutsky's Theorem (see the next chapter for details.)

Transformations (Continuous Mapping)

Theorem 3.4.26 — Continuous mapping theorem. Let $X_1, X_2, \ldots$ and X be k -dim random vectors, $g : \mathcal{R}_k \to \mathcal{R}$ be continuous. Then

1. $X_n \to X$ a.s. $\Longrightarrow g(X_n) \to g(X)$ a.s.
2. $X_n \to_p X \Longrightarrow g(X_n) \to_p g(X)$
3. $X_n \to_d X \Longrightarrow g(X_n) \to_d g(X)$

Proof. We shall only treat the case $k = 1$. The extension to general k is trivial.

1. By the continuity of g

$$A =: \{\omega : X_n(\omega) \to X(\omega)\} \subset \{\omega : g(X_n(\omega)) \to g(X(\omega))\} =: B$$

Thus $1 = P(A) \leq P(B) \leq 1$, i.e., $P(B) = 1$

2. Let $\epsilon > 0$. We may pick some large M such that $P(|X| \geq M) \leq \epsilon$. The continuous function g is uniformly continuous on the bounded interval $|x| \leq M + \epsilon$. There exists $\delta > 0$ such that

$$|g(x) - g(y)| \leq \epsilon, \quad \text{if} \quad |x - y| \leq \delta, \text{ and } |x| \leq M \tag{3.24}$$

First we notice that

$$|g(X_n) - g(X)| > \epsilon, \quad \text{and} \quad |X_n - X| \leq \delta \Longrightarrow |X| \geq M$$

Proof. To see this, if this were not true, we would have $|X| < M$, which, in combination of $|X_n - X| \leq \delta$, implies that $|X_n - X| \leq M + \delta \leq M + \epsilon$ (we can always choose $\delta \leq \epsilon$). Then from (3.24), we would have $|g(x) - g(y)| \leq \epsilon$, which contradicts with our assumption. ∎

Therefore,

$$\begin{aligned}
&P(|g(X_n) - g(X)| > \epsilon) \\
=&P(|g(X_n) - g(X)| > \epsilon, |X_n - X| > \delta) + P(|g(X_n) - g(X)| > \epsilon, |X_n - X| \leq \delta) \\
=&P(|X_n - X| > \delta) + P(|g(X_n) - g(X)| > \epsilon, |X_n - X| \leq \delta) \\
\leq&P(|X_n - X| > \delta) + P(|X| \geq M) \\
\to&P(|X| \geq M) \leq \epsilon
\end{aligned}$$

in the limit as $n \to \infty$. It follows that $g(X_n) \to_p g(X)$.

3. As $X_n \to_d X$, by Skorokhod construction, $\exists Y, Y_1, Y_2, \ldots$ on another probability space $(\Omega', \mathcal{A}', P')$ such that $Y_n \to Y$ a.s. and $X_n =_d Y_n, X =_d Y$. By $(a), g(Y_n) \to g(Y)$ a.s., which in turn implies that $g(Y_n) \to_d g(Y)$. But this is the same as $g(X_n) \to_d g(X)$

∎

1. What is wrong with the following "proof' of part (b) in the continuous mapping theorem?

   Proof. Wrong proof. By the continuity of $g, \forall \epsilon > 0, \exists \delta > 0$ such that

   $$|X_n - X| < \delta \quad \Longrightarrow \quad |g(X_n) - g(X)| < \epsilon$$

   Thus $P(|X_n - X| < \delta) \leq P(|g(X_n) - g(X)| < \epsilon)$. Take $\limsup_{n\to\infty}$ on both sides, we get $1 = \lim_{n\to\infty} P(|X_n - X| < \delta) \leq \liminf_{n\to\infty} P(|g(X_n) - g(X)| < \epsilon) \leq \limsup_{n\to\infty} P(|g(X_n) - g(X$
   Thus, we have

   $$\limsup_{n\to\infty} P(|g(X_n) - g(X)| < \epsilon) = 1$$

   ■

2. A different proof of (b) is given in Chow and Teicher, Corollary 3. page 68 .

Theorem 3.4.27 — The continuity assumption of g everywhere can be weakened. Let $X_1, X_2, \ldots$ and X be k -dim random vectors. Assume that g is continuous a.s. w.r.t. the probability measure P_X, i.e. $P_X(X : g(\cdot)$ is discontinous at $X) = P(\omega : g(\cdot)$ is discontinous at $X(\omega)) = 0$. Then, Theorem 3.4.26 still holds.

Proof. We shall prove the case $k = 1$. Using the same notation as in Theorem 3.4.26, we have

$$\begin{aligned} A =: &\{\omega : X_n(\omega) \to X(\omega)\} \\ &\subset \{\omega : g(X_n(\omega)) \to g(X(\omega))\} + \{\omega : g(\cdot) \text{ is discontinues at } X(\omega)\} \\ &=: B + D \end{aligned}$$

Thus $1 = P(A) = P(B + D) \leq P(B) + P(D) \leq P(B) \leq 1$, i.e., $P(B) = 1$ ■

1. In Theorem 3.4.27, if $X = C$, then $g(x)$ only needs to be continuous at C.
2. We shall now give an example, if $P_X(X : g(\cdot)$ is discontinous at $X) \neq 0$, then $X_n \to_p X \not\Rightarrow g(X_n) \to_p g(X)$ e.g. Let $P(X_n = -1/n) = 1$, and $P(X = 0) = 1$. Define

   $$\begin{aligned} g(t) = &\quad t - 1, t < 0 \\ = &\quad t + 1, t \geq 0 \end{aligned}$$

   Clearly, $P_X(X : g(\cdot)$ is discontinous at $X) = P_X(X = 0) = 1 \neq 0$. However, we note $X_n \to_p X = 0$ whereas $\lim_n P(|g(X_n) + 1| \geq \epsilon) = \lim_n P(|-1/n - 1 + 1| \geq \epsilon) = 0 \quad \forall \epsilon > 0$. That is,

   $$g(X_n) \to_p -1 \neq 1 = g(0)$$

■ **Example 3.13**
1. If $X_n \to_d N(0,1)$, then $X_n^2 \to_d N^2(0,1) = \chi_1^2$
2. If $(X_n, Y_n) \to_d N(\mathbf{0}, \mathbf{I})$, then $X_n/Y_n \to_d$ Cauchy.

Proof. Let $(X,Y) =_d N(\mathbf{0}, \mathbf{I})$, then X, Y are independent, $X/Y =_d$ Cauchy, and $P(g(X,Y) =: X/Y$ is discontinous$) = P(Y=0) = 0$. Apply the above remark.) ■

3.4.4 Slutsky's Theorem

We have seen from Theorem 3.4.25 that: $X_n \to X$ and $Y_n \to Y$ implies that $X_n \pm Y_n \to X \pm Y$ a.s., in probability, or in r -th mean, but not true in distribution in general. However, it does hold if $Y = C$ (in which case $Y_n \to_d C \iff Y_n \to_p C$)

Theorem 3.4.28 — Slutsky's Theorem. Let $X_n \to_d X, Y_n \to_p C$ (i.e. $Y_n \to_d C$) . Then
1. $X_n + Y_n \to_d X + C$
2. $X_n Y_n \to_d CX$
3. $X_n/Y_n \to_d X/C$ if $C \neq 0$

Proof. We shall only prove (a). The proofs of (b) and (c) are similar. First note that

$$
\begin{aligned}
F_{X_n+Y_n}(x) &= P(X_n + Y_n \le x) \\
&= P(X_n + Y_n \le x, |Y_n - C| \le \epsilon) + P(X_n + Y_n \le x, |Y_n - C| > \epsilon) \\
&\le P(X_n + C \le x - (Y_n - C), |Y_n - C| \le \epsilon) + P(|Y_n - C| > \epsilon) \\
&\le P(X_n + C \le x + \epsilon) + P(|Y_n - C| > \epsilon) \\
&= F_{X_n+C}(x+\epsilon) + P(|Y_n - C| > \epsilon)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
F_{X_n+Y_n}(x) &= 1 - P(X_n + Y_n > x) \\
&= 1 - P(X_n + Y_n > x, |Y_n - C| \le \epsilon) - P(X_n + Y_n > x, |Y_n - C| > \epsilon) \\
&\ge 1 - P(X_n + C > x - (Y_n - C), |Y_n - C| \le \epsilon) - P(|Y_n - C| > \epsilon) \\
&\ge 1 - P(X_n + C > x - \epsilon) - P(|Y_n - C| > \epsilon) \\
&= F_{X_n+C}(x-\epsilon) - P(|Y_n - C| > \epsilon)
\end{aligned}
$$

Combining the two, we have

$$
F_{X_n+C}(x-\epsilon) - P(|Y_n - C| > \epsilon) \le F_{X_n+Y_n}(x) \le F_{X_n+C}(x+\epsilon) + P(|Y_n - C| > \epsilon)
$$

Letting $n \to \infty$, we obtain

$$
F_{X+C}(x-\epsilon) \le \liminf_n F_{X_n+Y_n}(x) \le \limsup_n F_{X_n+Y_n}(x) \le F_{X+C}(x+\epsilon)
$$

If $F_{X+C}(x)$ is continuous at x, then as $\epsilon \downarrow 0$, we have $F_{X+C}(x-\epsilon) \uparrow F_{X+C}(x)$ and $F_{X+C}(x+\epsilon) \downarrow F_{X+C}(x)$ the result is proved. ■

R

1. The proof of the theorem is almost the same as that of Theorem **??**
2. Slutsky's theorems are very useful in proving limiting distributions in so-called δ **-method**.

Theorem 3.4.29 Suppose that X_i 's are uncorrelated and $\sup_{k\geq 1} EX_k^2 \leq M < \infty$. Denote $\overline{X} = n^{-1}\sum_{i=1}^n X_i$, $\mu_i = EX_i$ and $\bar{\mu} = n^{-1}\sum_{i=1}^n \mu_i$. Then

1. $\overline{X} - \bar{\mu} \to_{L_2} 0$
2. $\overline{X} - \bar{\mu} \to_p 0$
3. $\overline{X} - \bar{\mu} \to_{\text{a.s.}} 0$

Proof. Note that $\sigma_i^2 =: \operatorname{Var}(X_i) = EX_i^2 - (EX_i)^2 \leq EX_i^2 \leq M$, and μ_i exists. Write $S_n = \sum_{i=1}^n X_i$ then

$$\overline{X} - \bar{\mu} = \frac{S_n - ES_n}{n}$$

$E(\overline{X} - \bar{\mu})^2 = \frac{1}{n^2}\sum_i\sum_j E(X_i - \mu_i)(X_j - \mu_j) = \frac{1}{n^2}\sum_i E(X_i - \mu_i)^2 = \frac{1}{n^2}\sum \sigma_i^2 \leq \frac{M}{n} \to 0$

This is implied by (i) below.

Then $\operatorname{Var}(S_n) = \sum_{i=1}^n \sigma_i^2$. By Chebyshev's inequality,

$$P(|S_n - ES_n| > n\epsilon) \leq \frac{\operatorname{Var}(S_n)}{n^2\epsilon^2} \leq \frac{M}{n\epsilon^2}$$

Summing over n on RHS, the resulting series diverges. However, if we confine our attention to the subsequence $\{n^2; n \geq 1\}$, then

$$\sum_n P\left(|S_{n^2} - ES_{n^2}| > n^2\epsilon\right) \leq \sum_n \frac{M}{n^2\epsilon^2} < \infty \tag{3.25}$$

Hence by Borel-Cantelli Lemma, we have

$$P\left(|S_{n^2} - ES_{n^2}|/n^2 > \epsilon, i.o.\right) = 0$$

Consequently, by Theorem 3.4.2, we have

$$\frac{S_{n^2} - ES_{n^2}}{n^2} \to 0, a.s \tag{3.26}$$

(In fact, (3.26) is implied directly by (3.25) since convergence in prob. fast enough implies a.s. convergence.)

Hence we have proved (iii) only for a subsequence $n^2 : n \geq 1$. We need to fill in the gaps in the limit process. W.L.O.G., assume that $\mu_i = 0$ all i. Let

$$D_n = \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$$

Then we have

$$
\begin{aligned}
ED_n^2 \leq & E \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|^2 \\
\leq & \sum_{k=n^2}^{(n+1)^2-1} E|S_k - S_{n^2}|^2 \\
= & \sum_{k=n^2}^{(n+1)^2-1} \sum_{j=n^2+1}^{k} \sigma^2(X_j) \\
\leq & \sum_{k=n^2}^{(n+1)^2-1} \sum_{j=n^2+1}^{(n+1)^2-1} M \\
\leq & 4n^2 M
\end{aligned}
$$

and consequently by Chebyshev's inequality

$$
P\left(D_n > n^2\epsilon\right) \leq \frac{4M}{\epsilon^2 n^2}
$$

It follows as before that

$$
\frac{D_n}{n^2} \to 0, a.s \tag{3.27}
$$

Now it is clear from (3.26) and (3.27) that, for $n^2 \leq k < (n+1)^2$

$$
0 \leq \frac{|S_k|}{k} = \frac{|S_{n^2} + (S_k - S_{n^2})|}{k} \leq \frac{|S_{n^2}| + |S_k - S_{n^2}|}{k} \leq \frac{|S_{n^2}| + |D_n|}{n^2} \to 0 \quad \text{a.s.}
$$

■

R

1. The result in the last theorem involves only the first moment, but we have operated with the second. In the next section, we shall remove the extra second moment condition.
2. The method used in the proof is called "subsequence method", very useful in other contexts as well. It first proves the result for a subsequence and then fill in the gap.

3.4.5 Fatou's Lemma Revisited

We shall derive a most general Fatou's lemma, (see Durrett, Chapter 2, p48.)

Theorem 3.4.30 — General Fatou's Lemma. Let $g(\cdot) \geq 0$ be continuous. If $X_n \to X_\infty$ in any mode (i.e., in prob, or L^r, or distribution, or a.s.), then

$$
\liminf_n Eg(X_n) \geq Eg(X_\infty)
$$

and

$$\limsup_n Eg(X_n) \le Eg(X_\infty)$$

Proof. We only prove the first inequality here. Applying the Skorokhod's representation theorem, there exist Y_n and Y such that $Y_n \to_d Y_\infty$ such that $X_n =_d Y_n$ and $X_\infty = dY_\infty$

$$\begin{aligned} Eg(X_\infty) = Eg(Y_\infty) = Eg(\lim_n Y_n) = E\lim_n Eg(Y_n) = E\liminf_n g(Y_n) \\ \le \liminf_n Eg(Y_n) = \liminf_n Eg(X_n) \end{aligned}$$

∎

Corollary 3.4.31 $\{X_n\}$ is a sequence of r.v.'s.

1. If $X_n \ge Y$ in any mode (i.e., in prob, or L^r, or distribution, or a.s.), then

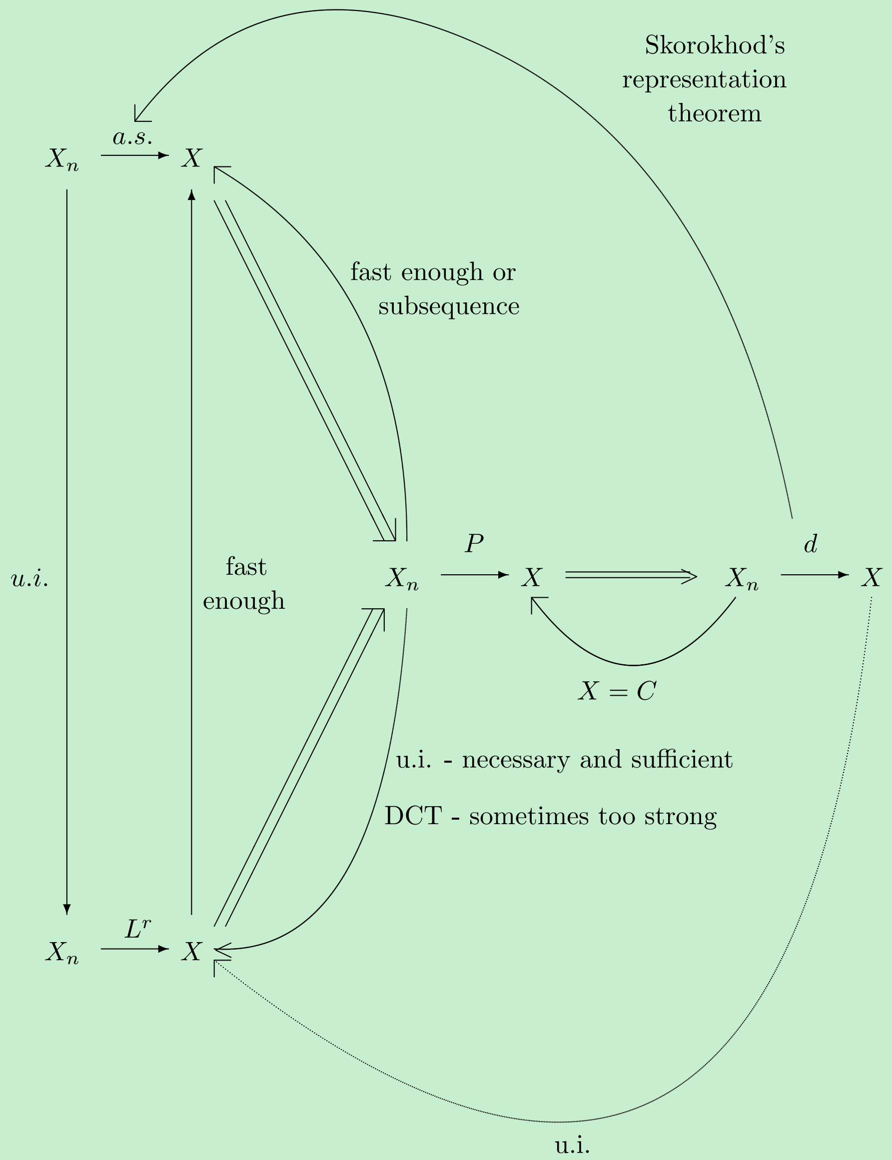
$$\liminf_n EX_n \ge E\left(\liminf_n X_n\right)$$

2. If $X_n \le Y$ in any mode (i.e., in prob, or L^r, or distribution, or a.s.), then

$$\limsup_n EX_n \le E\left(\limsup_n X_n\right)$$

The proof follows from the last theorem. The case $Y = 0$ is most often used.

Summary: relationships amongst four modes of converges.



Definition 3.4.4 — Convergence.

1. Converge in probability: A sequence of random variables Y_n **converges** to Y **in probability** if for every $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\{|Y_n - Y| > \epsilon\} = 0$$

2. Converge almost surely or with probability one: A sequence of random variables Y_n converges to Y almost surely (a.s.) or with probability one (w.p.1) if there is an event E with $P(E) = 1$ and such that for $\omega \in E, Y_n(\omega) \to Y(\omega)$.

R "in probability" is the same as "in measure" and "almost surely" is the analogue of "almost everywhere"

Lemma 3.11 Suppose $Y_1, Y_2, \ldots$ is a sequence of random variables with $\mathbb{E}[Y_n] \to \mu$ and $\mathrm{Var}[Y_n] \to 0$. Show that $Y_n \to \mu$ in probability. Give an example to show that it is not necessarily true that $Y_n \to \mu$ w.p.1. ■

Theorem 3.4.32 — Weak Law of Large Numbers. If $X_1, X_2, \ldots$ are independent random variables such that $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}[X_n] \le \sigma^2$ for each n, then

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow \mu$$

in probability.

The Strong Law of Large Numbers $(SLLN)$ is a statement about almost sure convergence of the weighted sums. We will prove a version of the strong law here. We start with an easy, but very useful lemma. Recall that the limsup of events is defined by

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Probabilists often write $\{A_n \text{ i.o. }\}$ for $\limsup A_n$; here "i.o." stands for "infinitely often".

Lemma 3.12 — Borel-Cantelli Lemma. Suppose $A_1, A_2, \ldots$ is a sequence of events.

1. If $\sum \mathbb{P}(A_n) < \infty$, then

$$\mathbb{P}\{A_n \text{ i.o. }\} = \mathbb{P}(\limsup A_n) = 0$$

2. If $\sum \mathbb{P}(A_n) = \infty$, and the $A_1, A_2, \ldots$ are independent, then

$$\mathbb{P}\{A_n \text{ i.o. }\} = \mathbb{P}(\limsup A_n) = 1$$

■

Proof. 1. If $\sum \mathbb{P}(A_n) < \infty$

$$\mathbb{P}(\limsup A_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_n\right) \le \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0$$

The second inequality is for subaddition, and the last equality is for the general term of converge series approaches to zero.

2. Assume $\sum \mathbb{P}(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent. We will show that

$$\mathbb{P}\left[(\limsup A_n)^c\right] = \mathbb{P}\left[\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c\right] = 0$$

To show this it suffices to show that for each n

$$\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$$

and for this we need to show for each n

$$\lim_{M\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{M} A_m^c\right) = 0$$

But by independence,

$$\mathbb{P}\left(\bigcap_{m=n}^{M} A_m^c\right) = \prod_{m=n}^{M}\left[1-\mathbb{P}\left(A_m\right)\right] \le \exp\left\{-\sum_{n=m}^{M}\mathbb{P}\left(A_m\right)\right\} \to 0$$

The last equality is for $e^x \ge 1+x$, then $e^{-\mathbb{P}A_m} \ge 1-\mathbb{P}A_m$

■

3.5 Law Of Large Numbers

3.5.1 Weak Law Of Large Numbers

Theorem 3.5.1 In probability theory, the following limit theorems are of major concern:

1. Weak law of large numbers (WLLN): What are the necessary and sufficient conditions for $\frac{S_n}{B_n} - A_n \longrightarrow_p 0$, where $\{A_n\}$ and $\{B_n\}$ are two non-random sequences, $B_n > 0$ and $B_n \nearrow \infty$?
2. Strong law of large numbers (SLLN): What are the necessary and sufficient conditions for $\frac{S_n}{B_n} - A_n \longrightarrow$ a.s. 0, where $\{A_n\}$ and $\{B_n\}$ are as in (i)?
3. Law of iterated logorithms (LIL): When do we have $\limsup_{n\to\infty} \frac{S_n}{B_n} = 1, a.s$
4. Central limit theorem (CLT): What are the necessary and sufficient conditions for $\frac{S_n}{B_n} - A_n \longrightarrow_d N(0,1)$?
5. Refinements of CLT: Berry-Esseen bounds, Edgeworth Expansions, Large deviations, saddlepoint approximations.

When $\{X_n\}$ are independent, the above issues are well understood. Extensions are still continuing in different directions:

1. dependent sequences (mixing conditions, martingales);
2. limit theorems for other statistics than S_n
3. functional limit theorems (stochastic processes)

3.5.2 Equivalent sequences; truncation

Definition 3.5.1 — Equivalent. Two sequences of r.v.'s $\{X_n\}$ and $\{Y_n\}$ on $(\Omega, \mathcal{A}, P)$ are said to be **equivalent** iff

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

Theorem 3.5.2 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 $\frac{1}{a_n}\sum_{j=1}^{n}(X_j - Y_j) \to 0$ a.s.

Proof. By the Borel-Cantelli lemma, the assumption of equivalence implies

$$P(\{\omega : X_n(\omega) \neq Y_n(\omega)\}, \text{i.o.}) = P(X_n \neq Y_n, \text{i.o.}) = 0$$

$(P(A_n, i.o.) = P(\{A_n^c, ult.\}^c))$ Hence,

$$P(\{\omega : X_n(\omega) = Y_n(\omega)\}, \text{ult.}) = 1 - P(\{X_n = Y_n\}^c, \text{i.o.}) = 1 - P(X_n \neq Y_n, \text{i.o.}) = 1$$

Thus, $\exists$ a P -null set N with the property: if $\omega \in \Omega - N, \exists n_0(\omega)$ such that

$$n \geq n_0(\omega) \quad \Longrightarrow \quad X_n(\omega) = Y_n(\omega)$$

For such an ω, the two numerical sequences $\{X_n(\omega)\}$ and $\{Y_n(\omega)\}$ differ only in a finite number of terms (how many depending on ω). In other words, the series $\sum_{n=1}^{\infty}(X_n(\omega) - Y_n(\omega))$ consists of zeros from a certain point on. Both (a) and (b) of the theorem follow from this fact. ■

Corollary 3.5.3 Suppose that $\{X_n\}$ and $\{Y_n\}$ are equivalent, and $a_n \uparrow \infty$. Then with probability one (a.s.)

1. $\sum_{j=1}^{n} X_j$ or $\frac{1}{a_n}\sum_{j=1}^{n} X_j$ converges, diverges to $+\infty$ or $-\infty$, or fluctuates in the same way as $\sum_{j=1}^{n} Y_j$ or $\frac{1}{a_n}\sum_{j=1}^{n} Y_j$
2. In particular, if $a_n^{-1}\sum_{j=1}^{n} X_j$ converges in probability, so does $a_n^{-1}\sum_{j=1}^{n} Y_j$

Proof. (a) follows from the proof of the last theorem. To show (b), if $a_n^{-1}\sum_{j=1}^{n} X_j \to_p X$, then

$$\frac{1}{a_n}\sum_{j=1}^{n} Y_j = \frac{1}{a_n}\sum_{j=1}^{n} X_j + \frac{1}{a_n}\sum_{j=1}^{n}(Y_j - X_j) \to_p X$$

■

3.5.3 Weak Law of Large Numbers

Theorem 3.5.4 Let $\{X_i\}$ be pairwise independent and identically distributed r.v.'s with finite mean $\mu = EX_1$. Then $\overline{X} \to_p \mu$

Proof. $|EX_1| < \infty, \quad \Longleftrightarrow \quad E|X_1| < \infty \quad \Longleftrightarrow$

$$\sum_{n=1}^{\infty} P(|X_1| > n) = \sum_{n=1}^{\infty} P(|X_n| > n) < \infty \quad \text{(identical distribution)} \tag{3.28}$$

Define the "truncated" r.v.'s $\{Y_n\}$ by

$$Y_n = X_n I_{\{|X_n| \le n\}}$$

Then $\{X_n\}$ and $\{Y_n\}$ are equivalent since by (3.28)

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

Let $T_n = \sum_{i=1}^n Y_i$ and $\overline{Y} = n^{-1} \sum_{j=1}^n Y_j$. By Theorem 3.5.2, $\overline{X}$ and $\overline{Y}$ converges or diverges at the same time. Therefore, it suffices to show, as $n \to \infty$

1. $E\overline{Y} \to \mu$
2. $\mathrm{Var}(\overline{Y}) \to 0$

To see why, applying Markov's inequality,

$$0 \le P(|\overline{Y} - \mu| > \epsilon) \le E|\overline{Y} - \mu|^2/\epsilon^2 = \epsilon^{-2}\left[\mathrm{Var}(\overline{Y}) + \mathrm{bias}^2(\overline{Y})\right] \to 0$$

($\mathrm{bias}^2(\overline{Y}) \to 0 \Longleftrightarrow E\overline{Y} \to \mu$)

1. Proof of (i).

$$\begin{aligned} 0 \le |E\overline{Y} - \mu| &= \frac{1}{n}\left|\sum_{j=1}^n (EY_j - EX_j)\right| = \frac{1}{n}\left|\sum_{j=1}^n \left(EX_j I_{\{|X_j| \le j\}} - EX_i\right)\right| \\ &= \frac{1}{n}\left|-\sum_{j=1}^n EX_j I_{\{|X_j| > j\}}\right| \le \frac{1}{n}\sum_{j=1}^n E|X_j| I_{\{|X_j| > j\}} \longrightarrow 0 \end{aligned}$$

   where we have used the fact: "$a_n \to a$" $\Longrightarrow$ "$\bar{a}_n \to a$", where $a_n = E|X_n| I_{\{|X_n| > n\}} \to 0$ (Stolz)

2. Proof of (ii). It is equivalent to show $\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) = o\left(n^2\right)$. Note that Y_n are independent (as functions of X_n) and bounded. Thus

$$\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{k=1}^n \mathrm{Var}(Y_k) \le \sum_{k=1}^n EY_k^2 = \sum_{k=1}^n EX_k^2 I_{\{|X_k| \le k\}} = \sum_{k=1}^n EX_1^2 I_{\{|X_1| \le k\}}$$

   The crudest estimate of this is

$$\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) \le \sum_{k=1}^n kE|X_1| I_{\{|X_1| \le k\}} \le \sum_{k=1}^n kE|X_1| = \frac{1}{2}n(n+1)E|X_1| = O\left(n^2\right)$$

   which is not good enough. (Note that in the above we have used the bound $EX_1^2 I_{\{|X_1| \le k\}} \le kE|X_1| I_{\{|X_1| \le k\}}$ for all $k = 1, \ldots, n$. But when k is small, this bound may be too rough. This suggests that we should perhaps consider k to be small and large separately.) To improve upon it, we shall use another level of truncation. Let

$\{a_n\}$ be a sequence of integers such that $0 < a_n < n, a_n \uparrow \infty$, but $a_n = o(n)$. We have

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n} Y_i\right) &\leq \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \leq k\}} \\
&= \sum_{k=1}^{a_n} EX_1^2 I_{\{|X_1| \leq k\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{|X_1| \leq k\}} \\
&= \sum_{k=1}^{a_n} EX_1^2 I_{\{|X_1| \leq k\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{|X_1| \leq a_n\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{a_n < |X_1| \leq k \leq n\}} \\
&\leq \sum_{k=1}^{a_n} kE|X_1| I_{\{|X_1| \leq k\}} + a_n \sum_{k=a_n}^{n} E|X_1| I_{\{|X_1| \leq a_n\}} + n \sum_{k=a_n}^{n} E|X_1| I_{\{a_n < |X_1| \leq n\}} \\
&\leq a_n \sum_{k=1}^{a_n} E|X_1| + a_n \sum_{k=a_n}^{n} E|X_1| + n \sum_{k=a_n}^{n} E|X_1| I_{\{|X_1| > a_n\}} \\
&\leq n a_n E|X_1| + n^2 E|X_1| I_{\{|X_1| > a_n\}}
\end{aligned}
$$

which implies that

$$
0 \leq \operatorname{Var}(\overline{Y}) = \frac{1}{n^2} \operatorname{Var}\left(\sum_{i=1}^{n} Y_i\right) \leq \frac{a_n}{n} E|X_1| + E|X_1| I_{\{|X_1| > a_n\}} \to 0
$$

This completes our proof.

■

R A finer truncation could be used to bound $\operatorname{Var}(\overline{Y})$, which goes as follows.

$$
\operatorname{Var}(\overline{Y}) \leq \frac{1}{n^2} \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \leq k\}}
$$

But

$$
\begin{aligned}
EX_1^2 I_{\{|X_1| \le n\}} &= \sum_{j=1}^{n} EX_1^2 P\{j-1 \le |X_1| < j\} \\
&\le \sum_{j=1}^{n} j^2 P\{j-1 \le |X_1| < j\} \\
&\le \sum_{j=1}^{n} j(j+1) P\{j-1 \le |X_1| < j\} \\
&\le 2 \sum_{j=1}^{n} \sum_{i=1}^{j} i P\{j-1 \le |X_1| < j\} \\
&\le 2 \sum_{i=1}^{n} \sum_{j=i}^{n} i P\{j-1 \le |X_1| < j\} \\
&= 2 \sum_{i=1}^{n} i \sum_{j=i}^{n} P\{j-1 \le |X_1| < j\} \\
&= 2 \sum_{i=1}^{n} i P\{i-1 \le |X_1| < n\} \\
&\le 2 \sum_{i=1}^{n} i P\{|X_1| \ge i-1\} \\
&\le_{???} 2 \left(1 + \sum_{i=1}^{\infty} i P\{|X_1| \ge i\}\right) \\
&\le_{???} 2(2 + E|X_1|) \\
&< \infty
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathrm{Var}(\overline{Y}) &\le \frac{1}{n^2} \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \le k\}} \\
&\le \frac{1}{n^2} \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \le n\}} \\
&\le \frac{1}{n} 2(2 + E|X_1|) \to 0
\end{aligned}
$$

The above theorem may be slightly generalized as follows; compare with Corollary 3.5.9 later.

Theorem 3.5.5 Let $\{X_i\}$ be pairwise independent and identically distributed r.v.'s such that

$$EX_1 I\{|X_1| \le n\} \longrightarrow 0, \quad \text{and} \quad nP(|X_1| > n) \to 0$$

Then $\overline{X} \to_p 0$

R If X_i 's are independent, then we can use a totally different approach, i.e., the characteristic function approach (to be introduced later) to prove the above result.

Theorem 3.5.6 Let $\{X_i\}$ be i.i.d. r.v.'s with finite mean $\mu = EX_1$. Then

$$\overline{X} \to_p \mu$$

Proof. Let $\psi_X(t) = Ee^{itX}$. Since $|EX_1| < \infty$, we have

$$\psi_{X_1}(t) = \psi_{X_1}(0) + \psi'_{X_1}(0)t + o(t) = 1 + i\mu t + o(t), \quad |t| < \delta$$

Then

$$\psi_{\bar{X}}(t) = (\psi_{X_1}(t/n))^n = \left(1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right)^n \longrightarrow e^{it\mu}$$

This implies that $\overline{X} \to_d \mu$, or $\overline{X} \to_p \mu$.

If $\mu = EX_1$ exists, we could prove a stronger result: $\overline{X} \to_{a.s.} \mu$. (see the next chapter.) ■

Theorem 3.5.7 — Classical forms of the WILLN–Kolmogorov (n) - Feller (a_n). Let $\{X_n\}$ be independent r.v.'s with $F_n(x) = P(X_n \le x)$. Let $a_n > 0$ and $a_n \uparrow \infty$. Then

$$\frac{1}{a_n}\sum_{k=1}^{n} X_k \to_p 0$$

if and only if, as $n \to \infty$

1. $\sum_{k=1}^{n} P(|X_1| \ge a_n) \longrightarrow 0$
2. $E\left(\sum_{k=1}^{n} \frac{X_1}{a_n} I_{\{|X_1| < a_n\}}\right) \longrightarrow 0$
3. $\mathrm{Var}\left(\sum_{k=1}^{n} \frac{X_1}{a_n} I_{\{|X_1| < a_n\}}\right) \longrightarrow 0$

if and only if, by writing $Y_{nk} = \frac{X_1}{a_n} I\left\{\frac{X_1}{a_n} < 1\right\}$, as $n \to \infty$

1. $\sum_{k=1}^{n} P(|X_1| \ge a_n) \longrightarrow 0$
2. $E(\sum_{k=1}^{n} Y_{nk}) \to 0$
3. $\mathrm{Var}(\sum_{k=1}^{n} Y_{nk}) \longrightarrow 0$

R

1. Compare this with Kolmogorov's three series theorem in the next chapter.
2. For illustration, take $a_n = n$. Let $Y_1 = X_1 I\{|X_1| < n\}$, then (ii)-(iii) become

$$E(\overline{Y}) \to 0, \quad \mathrm{var}(\overline{Y}) \to 0$$

which ensures that $\overline{Y} \to_p 0$. On the other hand, (i) implies that $\{X_n\}$ and $\{Y_n\}$ are equivalent, hence, $\overline{X} \to_p 0$

When $\{X_n\}$ are i.i.d. r.v.'s, we have the following theorem.

Theorem 3.5.8 Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.'s with common d.f. F. Then the following statements are equivalent:

1. $\overline{X} := n^{-1} \sum_{k=1}^{n} X_1 \to_p C$ for some constant C
2. $nP(|X_1| \geq n) \to 0$ and $E[X_1 I\{|X_1| < n\}] \to C$ as $n \to \infty$
3. The characteristic function (c.f.) $\psi(t)$ of the X_1 is differentiable at $t = 0$ and $\psi'(0) = iC$.

Corollary 3.5.9 Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.'s with common d.f. F. The existence of a sequence of real numbers $\{a_n\}$ for which

$$\frac{1}{n} \sum_{i=1}^{n} X_i - a_n \longrightarrow_p 0$$

if and only if

$$nP(|X_1| \geq n) \to 0$$

in which case we may take $a_n = E[X_1 I\{|X_1| < n\}]$

R We make some remarks about Condition (ii) in Theorem 3.5.8

1. Recall that $nP(|X| \geq n) = o(1) \Rightarrow E|X|^{1-\epsilon} < \infty$, but $\neq E|X| < \infty$. Hence, $nP(|X| \geq n) = o(1)$ is weaker than the requirement $E|X| < \infty$ (the latter in fact implies a stronger conclusion $\bar{X} \to_{a.s.} \mu$; see the next chapter)
2. The constant C in condition (ii) may not be the mean EX_1, which may not even exist. Note that $C = 0$ whenever X_1 is a symmetric r.v. around 0.
3. A sequence of i.i.d. r.v.'s satisfies the WLLN but the SLLN whenever condition (ii) holds but $E|X_1| = \infty$. For instance, if X_1 is symmetric around 0, but its d.f. satisfies

$$F(x) \approx 1 - \frac{1}{x \log x}, \quad x \to \infty$$

3.6 Strong Convergence

The central theme of this chapter is to characterize the a.s. convergence of

1. series $\sum_{k=1}^{\infty} X_k$ (when X_i s are independent)
2. averages $n^{-a} \sum_{k=1}^{n-1} X_k$, where $\alpha > 0$ (when X_i s are i.i.d.)

The two problems are related by the elementary Kronecker lemma, and the main results are the

1. Kolmogorov three series criterion (for series,)
2. the strong law of large numbers (for averages)

In a nutsbell, the chapter simply studies the relationship between moment conditions and a.s. convergence for random sequences, series, and averages.

There are many special properties unique to independent Sequences, which are not true for general equivalent.

3.6.1 Some maximal inequalities

Strong Laws deal with a.s. convergence. One criterion is given in Theorem 6.1 .1 (c) or (e). That is, $X_n \to X$ a.s. iff $\forall \epsilon > 0$, as $n \to \infty$

$$P\left(\left\{\sup_{m:m\geq n} |X_m - X|\right\} \geq \epsilon\right) = \lim_{r\to\infty} P\left(\left\{\max_{n\leq m\leq r} |X_m - X|\right\} \geq \epsilon\right) \to 0$$

Therefore, it is necessary to estimate the "maximal" probability on the right hand side.

Theorem 3.6.1 — Hajek-Renyi maximal inequality. Let $X_1, X_2, \ldots$ be independent with $EX_k = 0$ and $\sigma_k^2 = \operatorname{Var}(X_i) < \infty$. Write $S_k = \sum_{i=1}^k X_i$. Let $\{c_k\}$ be a positive and nonincreasing sequence (i.e. $c_k > 0$ and $c_k \downarrow$). Then $\forall \epsilon > 0$, and $m < n$, we have

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

Proof. Let

$$E_m = \{c_m |S_m| \geq \epsilon\}$$

$$E_j = \left\{\max_{m\leq k<j} \{c_k |S_k|\} < \epsilon, c_j |S_j| \geq \epsilon\right\}, \quad m+1 \leq j \leq n \text{construct disjoint sets}$$

$$A = \sum_{j=m}^n E_j = \left\{\max_{m\leq k\leq n} \{c_k |S_k|\} \geq \epsilon\right\}$$

$$Y = c_m^2 S_m^2 + \sum_{k=m+1}^n c_k^2 \left(S_k^2 - S_{k-1}^2\right)$$

Note that $ES_k^2 = \sum_{j=1}^k \sigma_j^2, E_j$'s are mutually exclusive. It suffices to show that

$$P(A) = P\left(\sum_{j=m}^n E_j\right) = \sum_{j=m}^n P\left(E_j\right) \leq \frac{1}{\epsilon^2} EY$$

Now we can rewrite

$$\begin{aligned} Y &= c_m^2 S_m^2 + \sum_{k=m+1}^n c_k^2 S_k^2 - \sum_{k=m+1}^n c_k^2 S_{k-1}^2 \\ &= \sum_{k=m}^n c_k^2 S_k^2 - \sum_{k=m+1}^n c_k^2 S_{k-1}^2 \\ &= \sum_{k=m}^n c_k^2 S_k^2 - \sum_{k=m}^{n-1} c_{k+1}^2 S_k^2 \\ &= \sum_{k=m}^{n-1} \left(c_k^2 - c_{k+1}^2\right) S_k^2 + c_n^2 S_n^2 \\ &\geq 0, \quad \text{as} \quad c_k \searrow \end{aligned}$$

So

$$EY \geq E(YI_A) = E\left(YI_{\sum_{j=m}^n E_j}\right) = \sum_{j=m}^{n} EYI_{E_j}$$

For $m \leq j \leq n$, we have

$$\begin{aligned}EYI_{E_j} &= \sum_{k=m}^{n-1} \left(c_k^2 - c_{k+1}^2\right) ES_k^2 I_{E_j} + c_n^2 ES_n^2 I_{E_j} \\ &\geq \sum_{k=j}^{n-1} \left(c_k^2 - c_{k+1}^2\right) ES_k^2 I_{E_j} + c_n^2 ES_n^2 I_{E_j}\end{aligned}$$

But for $j \leq k \leq n$, we have

$$\begin{aligned}ES_k^2 I_{E_j} =& E\left[\left(S_j + (S_k - S_j)\right]^2 I_{E_j} \\ =& ES_j^2 I_{E_j} + E\left(S_k - S_j\right)^2 I_{E_j} + 2ES_j\left(S_k - S_j\right) I_{E_j} \\ =& ES_j^2 I_{E_j} + E\left(S_k - S_j\right)^2 I_{E_j} \\ & \Big(\text{as } E\left[S_j I_{E_j}\left(S_k - S_j\right)\right] = E\left[S_j I_{E_j}\right] E\left[(S_k - S_j)\right] = 0 \\ & \quad \text{Since } S_j I_{E_j} \text{ and } S_k - S_j \text{ are independent}\Big) \\ & \geq ES_j^2 I_{E_j} \\ & \geq \epsilon^2 P\left(E_j\right)/c_j^2, \text{ (as } \left|c_j S_j\right| \geq \epsilon \text{ on } E_j)\end{aligned}$$

Thus,

$$EYI_{E_j} \geq \left[\sum_{k=j}^{n-1}\left(c_k^2 - c_{k+1}^2\right) + c_n^2\right] \frac{\epsilon^2 P\left(E_j\right)}{c_j^2} = \epsilon^2 P\left(E_j\right)$$

Finally, we get

$$EY \geq EYI_A = \sum_{j=m}^{n} \epsilon^2 P\left(E_j\right) = \epsilon^2 P\left(\sum_{j=m}^{n} E_j\right) = \epsilon^2 P(A)$$

■

Theorem 3.6.2 — Kolmogorov maximal inequality. Let $X_1, X_2, \ldots$ be independent with $EX_k = 0$ and $\sigma_k^2 = \operatorname{Var}(X_i) < \infty$. Write $S_k = \sum_{i=1}^k X_i$. Let $\epsilon > 0$

1. (Upper bound)

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon\right) \leq \frac{\operatorname{Var}(S_n)}{\epsilon^2}$$

2. (Lower bound). If $|X_k| \leq C \leq \infty$, then $\forall k \geq 1$,

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

(Note that the $RHS = -\infty$ when $C = \infty$.)

Proof. 1. Take $m = 1$ and $c_k = 1, k \geq 1$ in Hajek-Renyi maximal inequality. Direct proof. Let

$$
\begin{aligned}
E_1 &= \{|S_1| \geq \epsilon\} \\
E_j &= \left\{ \left\{ \max_{1 \leq k < j} |S_k| \right\} < \epsilon, |S_j| \geq \epsilon \right\}, \quad \text{for } 2 \leq j \leq n \\
A &= \sum_{j=1}^{n} E_j = \left\{ \left\{ \max_{1 \leq k \leq n} |S_k| \right\} \geq \epsilon \right\}
\end{aligned}
$$

So

$$
\begin{aligned}
\operatorname{Var}(S_n) = ES_n^2 \geq ES_n^2 I_A = \sum_{j=1}^{n} ES_n^2 I_{E_j} &= \sum_{j=1}^{n} E\left[\left(S_j + (S_n - S_j)\right]^2 I_{E_j} \right. \\
&= \sum_{j=1}^{n} ES_j^2 I_{E_j} + \sum_{j=1}^{n} E\left(S_n - S_j\right)^2 I_{E_j} + 2\sum_{j=1}^{n} ES_j\left(S_n - S_j\right) I_{E_j} \\
&\geq \sum_{j=1}^{n} ES_j^2 I_{E_j} \\
&\Big(\text{as } E\left[S_j I_{E_j}\left(S_n - S_j\right)\right] = E\left[S_j I_{E_j}\right] E\left[\left(S_n - S_j\right)\right] = 0 \\
&\quad \text{since } S_j I_{E_j} \text{ and } S_k - S_j \text{ are independent} \Big) \\
&\geq \sum_{j=1}^{n} \epsilon^2 P\left(E_j\right) \quad \left(\text{ as } \left|c_j S_j\right| > \epsilon \text{ on } E_j\right) \\
&= \epsilon^2 P(A) \\
&= \epsilon^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon \right)
\end{aligned}
$$

2. From (a), we have

$$
\begin{aligned}
ES_n^2 I_A &= \sum_{j=1}^{n} ES_j^2 I_{E_j} + \sum_{j=1}^{n} E\left(S_n - S_j\right)^2 I_{E_j} \\
&\leq \sum_{j=1}^{n} E\left(\left|S_{j-1}\right| + C\right)^2 I_{E_j} + \sum_{j=1}^{n} \sum_{k=j+1}^{n} \left(EX_k^2\right) P\left(E_j\right) \\
&\quad (\text{since } \left(S_n - S_j\right) \text{ and } I_{E_j} \text{ are independent.}) \\
&\leq \sum_{j=1}^{n} E(\epsilon + C)^2 I_{E_j} + \left(ES_n^2\right) \sum_{j=1}^{n} P\left(E_j\right) \\
&= \left((\epsilon + C)^2 + ES_n^2\right) P(A)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
ES_n^2 I_A &= ES_n^2 - ES_n^2 I_{A^c} \\
&\geq ES_n^2 - \epsilon^2 P\left(A^c\right) \quad \text{as } A^c = \left\{ \left\{ \max_{1 \leq k \leq n} |S_k| \right\} < \epsilon \right\} \\
&= ES_n^2 - \epsilon^2 + \epsilon^2 P(A)
\end{aligned}
$$

Combining the above, we get

$$ES_n^2 - \epsilon^2 + \epsilon^2 P(A) \le ES_n^2 I_A \le \left((\epsilon + C)^2 + ES_n^2\right) P(A)$$

Hence,

$$P(A) \ge \frac{ES_n^2 - \epsilon^2}{ES_n^2 - \epsilon^2 + (\epsilon + C)^2} = 1 - \frac{(\epsilon + C)^2}{ES_n^2 - \epsilon^2 + (\epsilon + C)^2} \ge 1 - \frac{(\epsilon + C)^2}{ES_n^2}$$

■

Corollary 3.6.3 Let $X_1, X_2, \ldots$ be independent with $EX_k = 0$ and $\sigma_k^2 = \text{Var}(X_i) < \infty$. Write $S_k = \sum_{i=1}^k X_i$. If $|X_k| \le C \le \infty$, then $\forall k \ge 1$ and $\epsilon > 0$

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

R Chebyshev inequality is a special case of Kolmogorov maximal inequality by taking $n = 1$

$$P(|X - \mu| \ge \epsilon) \le \frac{E(X - \mu)^2}{\epsilon^2}$$

3.6.2 The a.s. convergence of series; three-series theorem

Definition 3.6.1 — Series converge. $\sum_{n=1}^{\infty} a_n$ is said to **converge** (in whatever sense) iff $\lim_{n \to \infty} \sum_{k=1}^{n} a_k$ exists.

To show that $\sum_{n=1}^{\infty} a_n$ converges, one of the most useful tools is **Cauchy criterion** since no limit is specified here.

Definition 3.6.2 — Review: Cauchy convergent a.s. or in probability.
1. The sequence $\{X_n, n \ge 1\}$ is **almost sure (a.s.) Cauchy convergent**

$$\begin{aligned}
&\Longleftrightarrow P\left(\lim_{m,n\to\infty} |X_m - X_n| = 0\right) = 1 \\
&\Longleftrightarrow \forall \epsilon > 0 : \lim_{M\to\infty} P\left(\sup_{m,n\ge M} |X_m - X_n| \le \epsilon\right) = 1 \\
&\Longleftrightarrow \forall \epsilon > 0 : \lim_{M\to\infty} P\left(\sup_{m,n\ge M} |X_m - X_n| > \epsilon\right) = 0 \\
&\Longleftrightarrow \sup_{m,n\ge M} |X_m - X_n| \longrightarrow_p 0 \text{ as } M \to \infty \\
&\Longleftrightarrow \sup_{m>n} |X_m - X_n| \longrightarrow_p 0 \text{ as } n \to \infty \\
&\Longleftrightarrow \sup_{m>n} |X_m - X_n| = o_p(1) \text{ as } n \to \infty
\end{aligned}$$

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

$$\Longleftrightarrow \forall \epsilon > 0: \lim_{m,n \to \infty} P(|X_m - X_n| \leq \epsilon) = 1$$

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

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

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

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

Theorem 3.6.4 $X_n \to X$ a.s. $\Longleftrightarrow X_n$ is a.s. Cauchy convergent.

Proof. "$\Longrightarrow$". $\exists$N : a $P-$ null set such that $\forall \omega \in N^c, \lim_n X_n(\omega) = X(\omega)$. Therefore,

$$0 \leq |X_n(\omega) - X_m(\omega)| \leq |X_n(\omega) - X(\omega)| + |X_n(\omega) - X(\omega)| \to 0$$

i.e. X_n is Cauchy convergent on N^c.

$'' \Longleftarrow ''$.$\exists N_0$: a $P-$ null set such that $\forall \omega \in N_0^c, \lim_{m,n \to \infty} |X_n(\omega) - X_m(\omega)| = X(\omega)$. Since $X_n(\omega)$ is a real sequence, then $\lim_n X_n(\omega) = X(\omega), \forall \omega \in N_0^c$, where X_n is a r.v. ■

Theorem 3.6.5 $X_n \to X$ in probability $\Longleftrightarrow \{X_n\}$ is Cauchy convergent in probability.

Proof. "$\Longrightarrow$". $\forall \epsilon > 0: \lim_n P(|X_n - X| > \epsilon) = 0$. Therefore,

$$0 \leq P(|X_n - X_m| > 2\epsilon) \leq P(|X_n - X| > \epsilon) + P(|X_m - X| > \epsilon) \to 0$$

X_n is Cauchy convergent in probability. $'' \Longleftarrow ''$. $\{X_n\}$ is Cauchy convergent in probability. Then, $\forall \epsilon > 0$, we have

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

Then for any integer $k \geq 1, \exists$ an integer m_k such that

$$P\left(|X_m - X_n| > 2^{-k}\right) \leq 2^{-k}, \quad \text{for all } m > n \geq m_k$$

W.L.O.G, we can assume that m_k is strictly increasing sequence. Then setting,

$$Y_k := X_{n_k}, \quad A_k := \left\{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\right\} = \left\{|Y_{k+1} - Y_k| > 2^{-k}\right\}$$

we have

$$P(A_k) := P\left(|Y_{k+1} - Y_k| > 2^{-k}\right) \leq 2^{-k}, \quad \text{for all } m > n \geq m_k$$

Since $\sum_{k=1}^{\infty} P(A_k) < \infty$, by the Borel-Cantelli Lemma, $P(A_k, i.o.) = 0$ or $P(A_k^c, ult.) = 1$. That is, apart from an ω -set N of measure zero, we have

$$|Y_{k+1}(\omega) - Y_k(\omega)| \leq 2^{-k}$$

provided $k \geq$ some integer $k_0(\omega)$. Hence, for $\omega \in N^c$, as $n \to \infty$, we have

$$\sup_{m>n} |Y_m(\omega) - Y_n(\omega)| \leq \sum_{k=n}^{\infty} |Y_{k+1}(\omega) - Y_k(\omega)| \leq \sum_{k=n}^{\infty} 2^{-k} = 2^{-(n-1)} \to 0$$

Then,

$$P\left(\limsup_{n\to\infty} Y_k = \liminf_{n\to\infty} Y_k = \lim_{n\to\infty} Y_k := X\right) = 1$$

That is, $Y_k = X_{n_{k+1}} \to X$ a.s., hence in probability as well. Then, as $k \to \infty$

$$P(|X_k - X| \geq 2\epsilon) \leq P\left(\left|X_k - X_{n_{k+1}}\right| \geq \epsilon\right) + P\left(\left|X_{n_{k+1}} - X\right| \geq \epsilon\right) = o(1)$$

So $X_k \to_p X$ ■

Theorem 3.6.6 $X_n \to X$ in $L_2 \Longleftrightarrow \{X_n\}$ is mean square Cauchy convergent.

Proof. "$\Longrightarrow$". Suppose that $X_n \to X$ in L_2, i.e., $E|X_n - X|^2 \to 0$. By Minkowski's inequality,

$$\left(E|X_m - X_n|^2\right)^{1/2} \leq \left(E|X_m - X|^2\right)^{1/2} + \left(E|X_n - X|^2\right)^{1/2} \to 0$$

as $m, n \to \infty$. So $\{X_n\}$ is mean square Cauchy convergent.

"$\Longleftarrow$" (Subsequence approach.) Suppose that $\{X_n\}$ is mean square Cauchy convergent, i.e., $E|X_m - X_n|^2 \to 0$ as $m, n \to \infty$. By Chebyshev inequality, it is easy to see that $\{X_n\}$ is Cauchy convergent in probability, and therefore converges in probability to some limit X_∞. It follows that there exists a subsequence $\{X_{n_k}, k \geq 1\}$ which converges to X_∞ almost surely. Now as $n \to \infty$

$$\begin{aligned}
E|X_n - X_\infty|^2 &= E \lim_{k\to\infty} |X_n - X_{n_k}|^2 = E \liminf_{k\to\infty} |X_n - X_{n_k}|^2 \\
&\leq \liminf_{k\to\infty} E|X_n - X_{n_k}|^2 \qquad \text{(Fatou's lemma)} \\
&\to 0 \quad \text{(Cauchy convergent in } L_2\text{)}
\end{aligned}$$

Therefore, $X_n \to X_\infty$ in L_2. ■

R Summarizing some of these theorems, we note that

$$\begin{aligned}
X_n \to X \text{ a.s.} \quad &\Longleftrightarrow \quad \sup_{m>n} |X_m - X_n| = o_p(1), \text{ as } n \to \infty \\
X_n \to_p X \quad &\Longleftrightarrow \quad \sup_{m>n} P(|X_m - X_n| \geq \epsilon) = o(1), \text{ as } n \to \infty
\end{aligned}$$

Variance criterion for random series

Theorem 3.6.7 — (Variance criterion for series, due to Khinchin and Kolmogorov). Let $X_1, X_2, \ldots$ be independent with $EX_k = 0$ and $\sigma_k^2 = \mathrm{Var}(X_k) < \infty$. If $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty$, then $\sum_{k=1}^{\infty} X_k$ converges a.s. (i.e., L_2 convergence implies a.s. convergence for independent series)

Proof. 1. We shall give two different proofs, both of which use Kolmogorov inequality. Method 1: Direct approach. Write $S_n = \sum_{k=1}^{n} X_k$. By Kolmogorov inequality,

$$P\left(\left\{\max_{\{m:M<m\leq n\}} |S_m - S_M|\right\} > \epsilon\right) \leq \frac{\mathrm{Var}(S_m - S_M)}{\epsilon^2} = \frac{\sum_{k=M+1}^{n} \mathrm{Var}(X_k)}{\epsilon^2}$$

Let $n \to \infty$

$$P\left(\left\{\sup_{\{m:m>M\}} |S_m - S_M|\right\} > \epsilon\right) \leq \frac{\sum_{k=M+1}^{\infty} \mathrm{Var}(X_k)}{\epsilon^2}$$

Therefore,

$$\lim_{M\to\infty} P\left(\left\{\sup_{\{m:m>M\}} |S_m - S_M|\right\} > \epsilon\right) = 0$$

Note that $\sup_{m,n>M} |S_m - S_n| \searrow$ as $M \nearrow$, and

$$P\left(\left\{\sup_{m,n>M} |S_m - S_n|\right\} > 2\epsilon\right) \leq P\left(\sup_{m>M} |S_m - S_M| > \epsilon\right) + P\left(\sup_{n>M} |S_n - S_M| > \epsilon\right)$$
$$= 2P\left(\sup_{m>M} |S_m - S_M| > \epsilon\right) \to 0, \quad \text{as } M \to \infty$$

Therefore, S_n is a Cauchy sequence a.s., hence $\lim_{n\to\infty} S_n$ exists.

2. Method 2: Subsequence method. Write $S_n = \sum_{k=1}^{n} X_k$. By Chebyshve's inequality, for $m < n$

$$P(|S_n - S_m| > \epsilon) \leq \frac{\sum_{k=m+1}^{n} \mathrm{Var}(X_k)}{\epsilon^2} \to 0, \quad \text{as } m \to \infty$$

Therefore, S_n is a Cauchy sequence in probability, i.e., $S_n \to_p S_\infty$, say. Hence, $\exists$ a subsequence $\{n_k\} \nearrow \infty$ such that $S_{n_k} \to a.s. S_\infty$ as $k \to \infty$. Now $\forall n \geq 1, \exists k \geq 1$, such that $n_k < n \leq n_{k+1}$, and

$$\begin{aligned} 0 &\leq |S_n - S_\infty| \leq \left|S_{n_{k+1}} - S_\infty\right| + \left|S_{n_{k+1}} - S_n\right| \\ &\leq \left|S_{n_{k+1}} - S_\infty\right| + \max_{n_k<j\leq n_{k+1}} \left|S_{n_{k+1}} - S_j\right| \\ &=: A_k + B_k \end{aligned}$$

We have shown that $A_k \to 0$ a.s. To show $B_k \to 0$ a.s., we apply Kolmogorov inequality,

$$\sum_{k=1}^{\infty} P(|B_k| \geq \epsilon) \leq \sum_{k=1}^{\infty} \frac{1}{\epsilon^2} E\left(S_{n_{k+1}} - S_{n_k}\right)^2 = \sum_{k=1}^{\infty} \frac{1}{\epsilon^2} \sum_{j=n_k+1}^{n_{k+1}} EX_j^2 \leq \frac{\sum_{i=1}^{\infty} EX_i^2}{\epsilon^2} < \infty$$

which implies that $B_k \to 0$ a.s. Hence, $S_n \to S_\infty$ a.s.

∎

Corollary 3.6.8 — Kolmogorov SLLN. Let $X_1, X_2, \dots$ be independent with $\mu_k = EX_k$ and $\sigma_k^2 = EX_k^2 < \infty$. Let $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ and $\bar{\mu} = n^{-1}\sum_{i=1}^n \mu_i$. If $\sum_{k=1}^\infty EX_k^2/k^2 < \infty$, then $\overline{X} - \bar{\mu} \to 0$ a.s.

Proof. 1. Method 1. Use Hajek-Renyi's inequality and Kronecker Lemma.

2. Method 2. Let $Y_k = (X_k - EX_k)/k$. Since $\sum_{k=1}^\infty \text{var}(Y_k) < \infty$, from the last theorem, we get $\sum_{k=1}^\infty (X_k - EX_k)/k$ converges a.s. By the Kronecker Lemma,

$$n^{-1}\sum_{k=1}^{n}(X_k - EX_k) \to 0 \text{ a.s.}$$

■

R Let $X, X_1, X_2, \dots$ be i.i.d. r.v.'s with $\mu = EX$.

1. If $EX^4 < \infty$, by Chebyshev's inequality, we can show $\overline{X} \to \mu$ a.s.
2. If $EX^2 < \infty$, by Kolmogorov's SLLN (Theorem 3.6.4) or more directly Hajek-Renyi's inequality, we can show that $\overline{X} \to \mu$ a.s.
3. If $E|X| < \infty$, we can apply Kolmogorov's three series theorem (see the next section) or equivalently the truncation method to show that $\overline{X} \to \mu$ a.s.

Kolmogorov three series theorem for random series

Variance criterion are useful only when variance of each term X_n has finite second moment. However, for general r.v., where this may be violated, we may still use the variance criterion after proper truncation. This is the well-known Kolmogorov three series theorem.

Theorem 3.6.9 — Kolmogorov three series theorem. Let $X_1, X_2, \dots$ be independent r.v.'s. Let

$$Y_n = X_n I_{\{|X_n| \le A\}} = \begin{cases} X_n, & |X_n| \le A \\ 0, & |X_n| > A \end{cases}$$

Then $\sum_{k=1}^\infty X_k$ a.s. $\Longleftrightarrow$ for some $A > 0$, the following three series converge:

1. $\sum_{n=1}^\infty P(|X_n| > A) = \sum_{n=1}^\infty P(X_n \ne Y_n) < \infty$
2. $\sum_{n=1}^\infty EY_n$ converges;
3. $\sum_{n=1}^\infty \text{Var}(Y_n) < \infty$ (i.e. Variance criterion for truncated r.v.)

Proof. " $\Longleftarrow$ " Assume that the three series converge.

$$\Longrightarrow \sum_{n=1}^{\infty} E\left(Y_n - EY_n\right)^2 < \infty [\text{ from (iii) }]$$

$$\Longrightarrow \sum_{n=1}^{\infty}\left(Y_n - EY_n\right) \text{converges a.s. (by Theorem 3.6.7, the variance criterion)}$$

$$\Longrightarrow \sum_{n=1}^{\infty} Y_n \text{ converges a.s. [from (ii)]}$$

$$\Longrightarrow \sum_{n=1}^{\infty} X_n \text{ converges a.s.} [\text{ as } \{X_n\} \text{ and } \{Y_n\} \text{ are equivalent from } (i)]$$

"$\Longrightarrow$ " Assume that $\sum_{n=1}^{\infty} X_n$ converges a.s.

1. Proof of (i). Clearly, the assumption $\Longrightarrow X_n = S_n - S_{n-1} \to 0$ a.s. $\Longrightarrow P\left(|X_n| > A, i.o.\right) = 0, \forall A > 0 \Longleftrightarrow \sum_{n=1}^{\infty} P\left(|X_n| > A\right) < \infty$, from Borel-Cantelli Lemma for independent r.v.'s $\{X_n\}$. Thus, (i) holds.
2. Proof of (iii). From (i), $\{X_n\}$ and $\{Y_n\}$ are equivalent. $\Longrightarrow \sum_{n=1}^{\infty} Y_n$ converges a.s. But since $|Y_n - EY_n| \leq 2A$, the applying Kolmogorov maximal inequality3.6.2 of lower bound, we get

$$P\left(\max_{n\leq j\leq r}\left|\sum_{j=n}^{j} Y_j\right| \leq \epsilon\right) \leq \frac{(\epsilon + 2A)^2}{\operatorname{Var}\left(\sum_{j=n}^{r} Y_j\right)} = \frac{(\epsilon + 2A)^2}{\sum_{j=n}^{r} \operatorname{Var}\left(Y_j\right)}, \quad \forall \epsilon > 0$$

Were the series (iii) to diverge, then $\sum_{j=n}^{r} \operatorname{Var}\left(Y_j\right) \to \infty$ as $r \to \infty$ for each fixed n. Thus, letting $r \to \infty$, we get

$$0 \leq P\left(\sup_n \left|\sum_{j=n}^{\infty} Y_j\right| \leq \epsilon\right) \leq \frac{(\epsilon + 2A)^2}{\sum_{j=n}^{\infty} \operatorname{Var}\left(Y_j\right)} = 0, \quad \forall \epsilon > 0$$

or equivalently,

$$P\left(\sup_n \left|\sum_{j=n}^{\infty} Y_j\right| \geq \epsilon\right) = 1$$

That is, the tail of $\sum_{j=1}^{\infty} Y_j$ almost surely would not be bounded by any fixed constant ϵ, so that the series could not converge almost surely. This contradiction proves that (iii) must converge a.s.
3. Proof of (ii). We just proved (i) and (iii), which imply, respectively,
   (a) $\sum_{n=1}^{\infty} Y_n$, converges a.s. as $\{X_n\}$ and $\{Y_n\}$ are equivalent from (i).
   (b) $\sum_{n=1}^{\infty} \left(Y_n - EY_n\right)$ converges a.s., which follows from (iii).

   Thus,

$$\sum_{n=1}^{\infty} EY_n = \sum_{n=1}^{\infty} Y_n - \sum_{n=1}^{\infty}\left(Y_n - EY_n\right) \text{ converges a.s.}$$

   This proves (ii).

■

R Note that the Kolmogorov three series theorem gives a necessary and sufficient condition for a series to converge a.s.: namely, the moments series (up to the order 2) of the truncated r.v.'s converge.

R Note that the proof only used the second half of the Kolmogorov's inequality.

Review on Borel 0-1 Criterion: Suppose that X_n 's are independent. Then

$$\begin{aligned} X_n \to 0, a.s. &\iff P(|X_n| \geq \epsilon, i.o.) = 0, \quad \forall \epsilon > 0 \\ &\iff \sum_{n=1}^{\infty} P(|X_n| \geq \epsilon) < \infty, \quad \forall \epsilon > 0 \end{aligned}$$

Kolmogorov two series theorem for absolute random series

To prove a.s absolute convergence, we have Kolmogorov two series theorem.

Theorem 3.6.10 — Kolmogorov two series theorem for a.s absolute convergence. $\{X_n\}$ are independent. Then $\sum_{n=1}^{\infty} |X_n|$ converges a.s. iff

$$\sum_{n=1}^{\infty} P(|X_n| \geq C) < \infty; \quad \sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} < \infty \tag{3.29}$$

Proof. By Kolmogorov three series theorem, we only need to show that the conditions in (3.29) imply that $\sum_{n=1}^{\infty} EX_n^2 I_{\{|X_n|<C\}} < \infty$, which is true since

$$\sum_{n=1}^{\infty} EX_n^2 I_{\{|X_n|<C\}} \leq C \sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} < \infty$$

$(X_n^2 = |X_n| \cdot |X_n|)$ ■

Theorem 3.6.11 — mean convergence implies a.s. convergence. $\{X_n\}$ are independent. Then $\sum_{n=1}^{\infty} E|X_n|^r < \infty (0 < r \leq 1)$ implies that $\sum_{n=1}^{\infty} |X_n|$ converges a.s.

Proof. Apply the last result,

1. $\sum_{n=1}^{\infty} P(|X_n| \geq C) \leq \sum_{n=1}^{\infty} E|X_n|^r / C^r < \infty$
2. $\sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} \leq C^{1-r} \sum_{n=1}^{\infty} E|X_n|^r I_{\{|X_n|<C\}} \leq C^{1-r} \sum_{n=1}^{\infty} E|X_n|^r < \infty$

■

R Note that $\sum_{n=1}^{\infty} E|X_n|^r < \infty (0 < r < 1)$ implies that $\sum_{n=1}^{\infty} E|X_n| < \infty$ as $E|X_n| \to 0$ and hence $E|X_n| \leq E|X_n|^r$ for large n. Therefore, we can replace the condition $\sum_{n=1}^{\infty} E|X_n|^r < \infty$ $(0 < r \leq 1)$ by $\sum_{n=1}^{\infty} E|X_n| < \infty$. Does the theorem still hold for $r > 1$?

R In fact, we can even remove the independence condition here, which of course will require different proof.

Theorem 3.6.12 If $\{X_n, n \geq 1\}$ is a sequence of **nonnegative**(non-i.i.d.), integrable r.v.'s with $S_n = X_1 + \ldots + X_n$ and $\sum_{n=1}^{\infty} EX_n < \infty$, then S_n converges a.s.

Proof. We shall use Cauchy criterion and subsequence method. By Chebyshev's inequality, and the assumption that $\sum_{n=1}^{\infty} EX_n < \infty$, we have, for $m \geq n$

$$P\left(|S_m - S_n| > \epsilon\right) \leq \frac{1}{\epsilon} E\left|S_m - S_n\right| = \frac{1}{\epsilon} \sum_{k=n+1}^{m} EX_k \to 0$$

i.e., S_n is a Cauchy sequence in probability. Therefore, $S_n \to S_\infty$ in probability, say. Hence, there exists a subsequence n_k such that $S_{n_k} \to S_\infty$ a.s. But since S_n is a monotone sequence, for any n, there exists k such that

$$S_{n_k} \leq S_n \leq S_{n_{k+1}}$$

Since $S_{n_k} \to S_\infty$ a.s. and $S_{n_{k+1}} \to S_\infty$ a.s., we get $S_n \to S_\infty$ a.s. ■

3.6.3 Strong Laws of Large Numbers (SILLN)

Definition 3.6.3 A sequence of r.v.'s $X_1, X_2, \ldots$ with partial sum $S_n = \sum_{k=1}^{n} X_k$ is said to obey the strong (weak) law of large numbers iff S_n / a_n converges to a constant a.s. (in probability). The important Kronecker lemma enables us to convert convergence results for random series into convergence of averages, i.e., into laws of large numbers.

Lemma 3.13 — Cesaro's Lemma. Given two sequences $\{b_n\}$ and $\{x_n\}$, assume that

1. $b_n \geq 0$ and $a_n = \sum_{k=1}^{n} b_k \nearrow \infty$
2. $x_n \to x, |x| < \infty$

Then

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

(i.e., weighted average of a convergent sequence converges to the same value.)

Proof. $\forall \epsilon > 0, \exists n_0$ such that $|x_n - x| < \epsilon$ for $n \geq n_0$. Therefore,

$$\left| \frac{1}{a_n} \sum_{k=1}^{n} b_k x_k - x \right| = \left| \frac{\sum_{k=1}^{n} b_k \left(x_k - x\right)}{\sum_{k=1}^{n} b_k} \right| \leq \frac{\sum_{k=1}^{n_0} b_k \left|x_k - x\right|}{\sum_{k=1}^{n} b_k} + \frac{\sum_{k=n_0}^{n} b_k \left|x_k - x\right|}{\sum_{k=1}^{n} b_k}$$
$$\leq \frac{\sum_{k=1}^{n_0} b_k \left|x_k - x\right|}{\sum_{k=1}^{n} b_k} + \frac{\sum_{k=n_0}^{n} b_k \epsilon}{\sum_{k=1}^{n} b_k} \leq \frac{C\left(n_0\right)}{a_n} + \epsilon$$

Letting $n \to \infty$, and noting $a_n \to \infty$, we obtain the theorem. ■

Corollary 3.6.13 If $x_n \to x$ (finite), then $\bar{x} = n^{-1}\sum_{k=1}^{n} x_k \to x$

Theorem 3.6.14 — Abel's method of summation, "integration by parts". $\{a_n\}$ and $\{x_n\}$ are two sequences with $a_0 = 0, S_k = \sum_{j=1}^{k} x_k$, and $S_0 = 0$. Then

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

or more vividly, by denoting $\Delta S_k = S_k - S_{k-1}$ etc.,

$$\sum_{k=1}^{n} a_k \Delta S_k = a_n S_n - \sum_{k=1}^{n} S_{k-1} \Delta a_k$$

Compare with $\int_0^n f(s)ds = sf(s)|_0^n - \int_0^n \mathrm{sdf}(s))$

Proof.

$$\begin{aligned}
\sum_{k=1}^{n} a_k x_k &= \sum_{k=1}^{n} a_k (S_k - S_{k-1}) = \sum_{k=1}^{n} a_k S_k - \sum_{k=1}^{n} a_k S_{k-1} \\
&= \sum_{k=0}^{n} a_k S_k - \sum_{k=0}^{n-1} a_{k+1} S_k = \left(a_n S_n + \sum_{k=0}^{n-1} a_k S_k \right) - \sum_{k=0}^{n-1} a_{k+1} S_k \\
&= a_n S_n - \sum_{k=0}^{n-1} (a_{k+1} - a_k) S_k = a_n S_n - \sum_{k=1}^{n} (a_k - a_{k-1}) S_{k-1}
\end{aligned}$$

■

Lemma 3.14 — Sums into weighted sums, Kronecker. If $a_n \nearrow \infty$ and $\sum_{n=1}^{\infty} x_n$ converges, then

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

Proof. Let $S_k = \sum_{j=1}^{k} x_k$ and $S_0 = 0$. So $S_n \to S_\infty$, say. Applying Abel's method and Cesaro's Lemma, we get

$$\begin{aligned}
\frac{1}{a_n} \sum_{k=1}^{n} a_k x_k &= \frac{1}{a_n} \left(a_n S_n - \sum_{k=1}^{n} (a_k - a_{k-1}) S_{k-1} \right) \\
&= S_n - \frac{\sum_{k=1}^{n} (a_k - a_{k-1}) S_{k-1}}{\sum_{k=1}^{n} (a_k - a_{k-1})} \\
&\to S_\infty - S_\infty = 0
\end{aligned}$$

■

Corollary 3.6.15 — Kronecker lemma. If $a_n \nearrow \infty$ and $\sum_{n=1}^{\infty} \frac{y_n}{a_n}$ converges, then

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

SLLN for independent $r.v.'s$

All theorems in this section point to the fact: For independent r.v.'s, moment convergence implies a.s. convergence.

Theorem 3.6.16 Let $\{X_n\}$ be independent r.v.'s. Assume that

1. $\{g_n(x)\}$ are even functions, positive and nondecreasing for $x > 0$. Assume for all n, at least one of the following holds:
   (a) $\frac{x}{g_n(x)} \nearrow$ for $x > 0$
   (b) $\frac{x}{g_n(x)} \searrow$ and $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0$; $\quad EX_n = 0$
   (c) $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0$; $\quad X_n$ has a symmetric d.f. about 0
2. $\{a_n\}$ is a positive sequence, and

$$\sum_{n=1}^{\infty} \frac{Eg_n(X_n)}{g_n(a_n)} < \infty \tag{3.30}$$

Then we have

$$\sum_{n=1}^{\infty} \frac{X_n}{a_n} \quad \text{converges a.s.} \tag{3.31}$$

If we further assume that $0 < a_n \nearrow \infty$, then

$$\frac{1}{a_n} \sum_{j=1}^{n} X_j \to 0 \tag{3.32}$$

Proof. The proof of (3.32) follows from (3.31) and Kronecker's Lemma. Hence, we shall prove (3.31) next. Let

$$F_n(x) = P(X_n \le x), \text{ and } Y_n = X_n I_{\{|X_n| < a_n\}} \Longleftrightarrow Y_n/a_n = (X_n/a_n) I_{\{|X_n|/a_n < 1\}}$$

By the three series theorem, it suffices to show the convergence of the random variables $\{X_n/a_n\}$ and $C = 1$, i.e. convergence of series

1. $\sum_{n=1}^{\infty} P\left(\frac{|X_n|}{a_n} \ge 1\right)$
2. $\sum_{n=1}^{\infty} E\left(\frac{Y_n}{a_n}\right)$
3. $\sum_{n=1}^{\infty} E\left(\frac{Y_n^2}{a_n^2}\right)$

(Note that (c) implies that $\sum_{n=1}^{\infty} \mathrm{Var}(Y_n/a_n) < \infty$.)

1. Proof of (a). If $|X_n| \ge a_n$, then $g_n(X_n) \ge g_n(a_n)$ (as $g_n(x) \nearrow$ for $x \ge 0$ and is even).

Thus, using (3.30), we get

$$\sum_{n=1}^{\infty} P(|X_n| \geq a_n) \leq \sum_{n=1}^{\infty} P(g(X_n) \geq g(a_n)) \leq_{\text{Markov ineq}} \sum_{n=1}^{\infty} \frac{Eg(X_n)}{g(a_n)} < \infty$$

2. Proof of (c). If $|X_n| < a_n$, then clearly $g(X_n) \leq g(a_n)$; also we can show that

$$\frac{X_n^2}{a_n^2} \leq \frac{g_n(X_n)}{g_n(a_n)} \tag{3.33}$$

Proof. Proof of (3.33). We shall look at it under assumptions (i)-(iii) separately. If (i) holds, then $|X_n| < a_n$ implies

$$\frac{|X_n|}{g_n(X_n)} \leq \frac{a_n}{g_n(a_n)}, \Longrightarrow \frac{X_n^2}{a_n^2} \leq \frac{g_n^2(X_n)}{g_n^2(a_n)} \leq \frac{g_n(X_n)}{g_n(a_n)}, \left(\text{as } \frac{g_n(X_n)}{g_n(a_n)} \leq 1\right) \tag{3.34}$$

If (ii) or (iii) holds, then $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0$. Thus, $|X_n| < a_n$ implies

$$\frac{X_n^2}{g_n(X_n)} \leq \frac{a_n^2}{g_n(a_n)}, \quad \Longrightarrow \quad \frac{X_n^2}{a_n^2} \leq \frac{g_n(X_n)}{g_n(a_n)}$$

This proves (3.33) ∎

Now it follows easily from (3.33) that

$$\sum_{n=1}^{\infty} \frac{EY_n^2}{a_n^2} \leq \sum_{n=1}^{\infty} E\left(\frac{X_n^2}{a_n^2} I_{\{|X_n|<a_n\}}\right) \leq \sum_{n=1}^{\infty} E\left(\frac{g(X_n)}{g(a_n)} I_{\{|X_n|<a_n\}}\right) \leq \sum_{n=1}^{\infty} E\left(\frac{g(X_n)}{g(a_n)}\right) < \infty$$

3. Proof of (b). We shall look at it under assumptions (i)-(iii) separately. Assume first (i) holds. Noting $|Y_n| < a_n$, it follows from (3.34) that

$$\left|\sum_{n=1}^{\infty} \frac{EY_n}{a_n}\right| \leq \sum_{n=1}^{\infty} \frac{Eg_n(Y_n)}{g_n(a_n)}$$

Assume now that (ii) holds. If $|X_n| \geq a_n$, then

$$\frac{|X_n|}{g_n(X_n)} \leq \frac{a_n}{g_n(a_n)}, \quad \Longrightarrow \quad \frac{|X_n|}{a_n} \leq \frac{g_n(X_n)}{g_n(a_n)}$$

But from $EX_n = 0$

$$\left|\sum_{n=1}^{\infty} \frac{EY_n}{a_n}\right| = \left|\sum_{n=1}^{\infty} E\frac{X_n}{a_n} I_{\{|X_n|<a_n\}}\right| = \left|-\sum_{n=1}^{\infty} E\frac{X_n}{a_n} I_{\{|X_n|\geq a_n\}}\right| \leq \sum_{n=1}^{\infty} \frac{Eg_n(Y_n)}{g_n(a_n)} < \infty$$

Finally assume that (iii) holds, then $EY_n = 0$. Naturally, $|\sum_{n=1}^{\infty} EY_n/a_n| < \infty$

∎

Take $g_n(x) \equiv g(x) \equiv |x|^r$ for all n and $r > 0$ in the above corollary, we get

Corollary 3.6.17 $\{X_n\}$ are independent r.v. s, and $0 < a_n \nearrow \infty$. Assume that

$$\sum_{n=1}^{\infty} E\left|\frac{X_n}{a_n}\right|^r = \sum_{n=1}^{\infty} \frac{E|X_n|^r}{a_n^r} < \infty, \quad 0 < r \leq 2 \tag{3.35}$$

Then, we have

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

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

Proof. Take $g_n(x) \equiv |x|^r$

1. Case I: If $0 < r \leq 1$, then (3.35) is equivalent to (3.30) . From the last theorem, we get $\sum_{n=1}^{\infty} X_n / a_n$ converges a.s. and then apply Kronecker's Lemma.
2. Case II: If $1 \leq r \leq 2$, then (3.35) implies that

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{Eg_n\left(X_n - EX_n\right)}{g_n\left(a_n\right)} &= \sum_{n=1}^{\infty} \frac{E\left|X_n - EX_n\right|^r}{a_n^r} \\
&\leq C_r \sum_{n=1}^{\infty} \frac{E\left|X_n\right|^r}{a_n^r} + C_r \sum_{n=1}^{\infty} \frac{\left|EX_n\right|^r}{a_n^r} \quad (\text{by } C_r \text{ -inequality }) \\
&\leq 2C_r \sum_{n=1}^{\infty} \frac{E\left|X_n\right|^r}{a_n^r} \quad \left(\text{ as } \left|EX_n\right|^r \leq E\left|X_n\right|^r, r \geq 1\right) \\
&< \infty
\end{aligned}$$

From the last theorem, we get $\sum_{n=1}^{\infty}\left(X_n - EX_n\right)/a_n$ converges a.s. and then apply Kronecker's Lemma.

■

R

1. When $r = 1$, we can either add or drop the term EX_j in the above theorem. To see why, note that we now have $\sum_{n=1}^{\infty} \frac{E|X_n|}{a_n} < \infty$. It then follows from (1.3) that $\sum_{n=1}^{\infty} \frac{EX_n}{a_n} < \infty$. Using Kronecker's Lemma, we have $\frac{1}{a_n}\sum_{k=1}^{n} EX_k \to 0$
2. When $r > 2$, the above corollary may not hold. Some variants of the the corollary exists; see the exercises for example.

SLLN for independent r.v.'s: necessary and sufficient moment conditions

We provided some necessary and sufficient conditions for the WLLN for independent r.v.'s in the last chapter. Here we do the same for the SLLN.

Theorem 3.6.18 Let $\{X_n\}$ be a sequence of independent r.v.'s, and let $\{a_n\}$ be a sequence of positive numbers such that $a_n \nearrow \infty$. Put $Y_{nk} = \frac{X_k}{a_n} I\{|X_k| < a_n\}$ for $1 \leq k \leq n$.

Assume that

$$\sum_{n=1}^{\infty} EY_{nn}^2 < \infty$$

Then the relation $\frac{1}{a_n}\sum_{j=1}^n X_j \to 0$ a.s. if and only if

$$\sum_{n=1}^{\infty} P(|X_n| \geq a_n) < \infty, \quad \text{and} \quad \sum_{k=1}^{n} EY_{nk} \longrightarrow 0$$

Applying Theorem 3.6.18 to i.i.d. r.v.'s, we get

Theorem 3.6.19 Let $\{X_n\}$ be a sequence of i.i.d. r.v.'s, and let $\{a_n\}$ be a sequence of positive numbers such that $a_n \nearrow \infty$ and

$$\sum_{k=n}^{\infty} 1/a_k^2 = O\left(n/a_n^2\right)$$

Then the relation $\frac{1}{a_n}\sum_{j=1}^n X_j \to 0$ a.s. if and only if

$$\sum_{n=1}^{\infty} P(|X_n| \geq a_n) < \infty, \quad \text{and} \quad nE\left(\frac{X_1}{a_n}\right) I\{|X_1| < a_n\} \longrightarrow 0$$

SLLN for i.i.d. r.v.'s: necessary and sufficient moment conditions

Theorem 3.6.20 — Kolmogorov SLLN for iid r.v.'s. Let $X_1, X_2, \ldots$ be i.i.d. r.v. 's, and $S_n = \sum_{k=1}^n X_k$ Then

1. $E|X_1| < \infty \quad \Longrightarrow \quad \frac{S_n}{n} \to EX_1$ a.s.
2. $E|X_1| = \infty \quad \Longrightarrow \quad \limsup_n \frac{|S_n|}{n} = \infty \quad a.s$

Proof. 1. Assume $E|X_1| < \infty$. Write $Y_n = X_n I_{\{|X_n| \leq n\}}$. Clearly,

$$\sum_{n=1}^{\infty} P(|X_n| \neq Y_n) = \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \leq E|X_1| < \infty$$

Therefore, $\{X_n\}$ and $\{Y_n\}$ are equivalent sequences. So it suffices to show that

(a) $\frac{1}{n}\sum_{j=1}^n EY_j \to EX_1$

(b) $\frac{1}{n}\sum_{j=1}^n (Y_j - EY_j) \longrightarrow 0$

Since these would imply

$$\frac{1}{n}\sum_{j=1}^n Y_j = \frac{1}{n}\sum_{j=1}^n EY_j + \frac{1}{n}\sum_{j=1}^n (Y_j - EY_j) \longrightarrow EX_1 \quad \text{a.s.}$$

which in turn implies that $n^{-1}\sum_{j=1}^n X_j \to EX_1$

(a) Proof of (a). Now

$$\begin{aligned} EY_n &= EX_n I_{\{|X_n| \leq n\}} = EX_1 I_{\{|X_1| \leq n\}} = E\left(X_1^+ - X_1^-\right) I_{\{|X_1| \leq n\}} \\ &= EX_1^+ I_{\{|X_1| \leq n\}} - EX_1^- I_{\{|X_1| \leq n\}} \\ &\longrightarrow EX_1^+ - EX_1^- = EX_1 \end{aligned}$$

where in the last line we have used the Monotone Convergence Theorem or the Dominated Convergence Theorem. Thus, $n^{-1}\sum_{j=1}^{n} EY_j \to EX_1$

(b) Proof of (b). Applying Corollary 3.6.17 with $a_n = n$ to $\{Y_n\}$, we get

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{EY_n^2}{n^2} &= \sum_{n=1}^{\infty} \frac{1}{n^2} EX_n^2 I_{\{|X_n| \le n\}} \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{n^2} E|X_1|^2 I_{\{k-1<|X_1|\le k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^2} EX_1^2 I_{\{k-1<|X_1|\le k\}} \\
&= \sum_{k=1}^{\infty} \left[E\left(X_1^2 I_{\{k-1<|X_1|\le k\}}\right) \left(\sum_{n=k}^{\infty} \frac{1}{n^2} \right) \right] \qquad (3.36) \\
&\le \sum_{k=1}^{\infty} \left[kE\left(|X_1| I_{\{k-1<|X_1|\le k\}}\right) \left(\frac{C}{k}\right) \right] \\
&= C \sum_{k=1}^{\infty} E\left(|X_1| I_{\{k-1<|X_1|\le k\}}\right) \\
&\le CE|X_1| < \infty
\end{aligned}
$$

alternatively, we could also carry on from (3.36) as follows

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{EY_n^2}{n^2} &\le \sum_{k=1}^{\infty} \left[k^2 P(k-1<|X_1|\le k) \left(\frac{C}{k}\right) \right] \\
&\le C \sum_{k=1}^{\infty} kP(k-1<|X_1|\le k) \\
&\le C(1+E|X_1|) < \infty
\end{aligned}
$$

where we used the elementary estimate $\sum_{n=k}^{\infty} \frac{1}{n^2} \le C/k$ for some $C>0$ and all $k \ge 1$. (For instance, if $k \ge 2$, then $\sum_{n=k}^{\infty} \frac{1}{n^2} \le \sum_{n=k}^{\infty} \frac{1}{(n-1)n} \le 1/(k-1) \le 2/k$.) Then it follows from Corollary 3.6.17 that (ii) holds.

(ii). Suppose that $E|X_1| = \infty$. Then $\forall A > 0, E(|X_1|/A) = \infty$. From this and also $\{X_n\}$ are iid, we have

$$
\sum_{n=1}^{\infty} P(|X_1| > An) = \sum_{n=1}^{\infty} P(|X_n| > An) = \infty
$$

Therefore,

$$
\begin{aligned}
1 =& P(|X_n| > An, \text{ i.o.}), \qquad \text{(By Borel 0-1 law for independent r.v.'s)} \\
=& P(|S_n - S_{n-1}| > An, \text{ i.o.}) \\
=& P(|S_n| > An/2, \text{ or } |S_{n-1}| > A(n-1)/2, \text{ i.o.}) \\
& \quad (\text{ as } |S_n - S_{n-1}| > An = |S_n| > An/2, \text{ or } |S_{n-1}| > A(n-1)/2) \\
=& P(|S_n| > An/2, \text{ i.o. })
\end{aligned}
$$

Thus, for each $A > 0$, there exists a null set $N(A)$ such that if $\omega \in \Omega - N(A)$, we have

$$\limsup_n \frac{|S_n|}{n} \geq \frac{A}{2}$$

Let $N = \bigcup_{m=1}^{\infty} N(m)$, which is also a null set. Then if $\omega \in N^c$, $\limsup_n \frac{S_n}{n} \geq \frac{A}{2}$ is still true for all A, and therefore the upper limit is ∞.

■

(R) In the proof, we need to estimate the second moment Y_n in terms of the first moment, since this is the only one assumed in the hypothesis. The standard technique is to spit the interval of integration and then invert the repeated summation.

Corollary 3.6.21 Let $X_1, X_2, \dots$ be i.i.d. r.v.'s, and $S_n = \sum_{k=1}^{n} X_k$. Then

$$\overline{X} =: \frac{S_n}{n} \to C, \quad a.s.$$

if and only if EX_1 exists and $EX_1 = C$.

Proof. If $n^{-1}S_n \to C$ a.s., then

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n-1}\frac{n-1}{n} \to 0, \quad a.s.$$

Thus, $P(|X_n| \geq n, \text{ i.o.}) = 0$. From the Borel 0 - 1 law for independent sequences, and identical distribution assumption, we get

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty$$

which implies that $E|X_1| < \infty$. ■

Now if $E|X_1| < \infty$, it then follows from the last theorem that $S_n/n \to EX_1$ a.s.

Theorem 3.6.22 — Marcinkiewicz SLLN for iid r.v.'s. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s, and $0 < r < 2$. Then

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

if and only if $E|X_1|^r < \infty$, where

$$\begin{aligned} a &= EX_1 \quad \text{if } 1 \leq r < 2 \\ &= \text{arbitrary} \quad \text{if } 0 < r < 1. \text{ (Usually, simply chooses } a = 0) \end{aligned}$$

Proof. Denote $S_n = \sum_{k=1}^{n} (X_k - a)$. If $n^{-1/r} \sum_{k=1}^{n} (X_k - a) \longrightarrow 0$ a.s., then

$$\frac{X_n}{n^{1/r}} = \frac{a}{n^{1/r}} + \frac{S_n}{n^{1/r}} - \frac{S_{n-1}}{n^{1/r}} = \frac{a}{n^{1/r}} + \frac{S_n}{n^{1/r}} - \frac{S_{n-1}}{(n-1)^{1/r}} \frac{(n-1)^{1/r}}{n^{1/r}} \to 0, \quad \text{a.s.}$$

Thus, $P\left(|X_n| \geq n^{1/r}, \text{ i.o.}\right) = 0$. From the Borel $0-1$ law for independent sequences, and identical distribution assumption, we get

$$\sum_{n=1}^{\infty} P\left(|X_n| \geq n^{1/r}\right) = \sum_{n=1}^{\infty} P\left(|X_1|^r \geq n\right) < \infty$$

which implies that $E|X_1|^r < \infty$.

Now we shall show that if $E|X_1|^r < \infty$, we shall show that $\frac{1}{n^{1/r}} \sum_{k=1}^{n} (X_k - a) \to 0$ a.s. The proof is very similar to Theorem 3.6.20. However, we shall write it down below for completeness.

Write $Y_n = X_n I_{\{|X_n| \leq n^{1/r}\}}$. Clearly,

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} P\left(|X_n| > n^{1/r}\right) = \sum_{n=1}^{\infty} P\left(|X_1|^r > n\right) \leq E|X_1|^r < \infty$$

Therefore, $\{X_n\}$ and $\{Y_n\}$ are equivalent sequences. Therefore, it suffices to show that

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

1. Case I: $r = 1$. This is Theorem 3.6.20.
2. Case II: $0 < r < 1$. Applying Corollary 3.6.17 with $a_n = n^{1/r}$ to $\{Y_n\}$, we get

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{E|Y_n|}{n^{1/r}} &= \sum_{n=1}^{\infty} \frac{1}{n^{1/r}} E|X_n| I_{\{|X_n| \leq n^{1/r}\}} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \leq k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \leq k\}} \\
&= \sum_{k=1}^{\infty} \left[E\left(|X_1| I_{\{k-1 < |X_1|^r \leq k\}}\right) \left(\sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \right) \right] \\
&\qquad \text{(Note that the series } \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \text{converges when} 0 < r < 1.) \\
&\leq \sum_{k=1}^{\infty} \left[k^{1/r} P\left(I_{\{k-1 < |X_1|^r \leq k\}}\right) \left(\frac{C}{k^{1/r-1}} \right) \right] \\
&= C \sum_{k=1}^{\infty} k P\left(I_{\{k-1 < |X_1|^r \leq k\}}\right) \\
&\leq C\left(1 + E|X_1|^r\right) < \infty
\end{aligned}$$

where we used the elementary estimate $\sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \leq C/k^{1/r-1}$ for some $C > 0$ and all $k \geq 1$ when $0 < r < 1$. It follows from Corollary 3.6.17 that

$$\frac{1}{n^{1/r}} \sum_{k=1}^{n} Y_k \to 0$$

which in turn implies that

$$\frac{1}{n^{1/r}}\sum_{k=1}^{n}(Y_k-a)=\frac{1}{n^{1/r}}\sum_{k=1}^{n}Y_k-\frac{a}{n^{1/r-1}}\to 0$$

3. Case III: $1<r<2$ with $a=EX_1$ Applying Corollary 3.6.17 with $a_n=n^{1/r}$ to $\{Y_n\}$,

$$\begin{aligned}
\sum_{n=1}^{\infty}\frac{EY_n^2}{n^{2/r}} &= \sum_{n=1}^{\infty}\frac{1}{n^{2/r}}EX_n^2 I_{\{|X_n|\le n^{1/r}\}}\\
&= \sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{1}{n^{2/r}}EX_1^2 I_{\{k-1<|X_1|^r\le k\}}\\
&= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty}\frac{1}{n^{2/r}}EX_1^2 I_{\{k-1<|X_1|^r\le k\}}\\
&= \sum_{k=1}^{\infty}\left[E\left(X_1^2 I_{\{k-1<|X_1|r\le k\}}\right)\left(\sum_{n=k}^{\infty}\frac{1}{n^{2/r}}\right)\right]\\
&\qquad \text{(Note that the series }\sum_{n=k}^{\infty}\frac{1}{n^{2/r}}\text{converges when}1<r<2.)\\
&\le \sum_{k=1}^{\infty}\left[k^{2/r}P\left(k-1<|X_1|^r\le k\right)\left(\frac{C}{k^{2/r-1}}\right)\right]\\
&= C\sum_{k=1}^{\infty}kP\left(k-1<|X_1|^r\le k\right)\\
&\le C\left(1+E|X_1|^r\right)<\infty
\end{aligned}$$

where we used the elementary estimate $\sum_{n=k}^{\infty}\frac{1}{n^{2/r}}\le C/k^{2/r-1}$ for some $C>0$ and all $k\ge 1$ when $1<r<2$. Thus, (i) follows from Corollary 3.6.17.

Proof of (ii) and (iii). Note that

$$\begin{aligned}
\sum_{n=1}^{\infty}\frac{EX_n-EY_n}{n^{1/r}}\,| &\le \sum_{n=1}^{\infty}\frac{E|X_n-Y_n|}{n^{1/r}}\\
&= \left|\sum_{n=1}^{\infty}\frac{1}{n^{1/r}}EX_n I_{\{|X_n|>n^{1/r}\}}\right| \le \sum_{n=1}^{\infty}\frac{1}{n^{1/r}}E|X_1|\,I_{\{|X_1|r>n\}}\\
&\le \sum_{n=1}^{\infty}\sum_{k=n}^{\infty}\frac{1}{n^{1/r}}E|X_1|\,I_{\{k-1<|X_1|^r\le k\}}\\
&= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty}\frac{1}{n^{1/r}}E|X_1|\,I_{\{k-1<|X_1|^r\le k\}}\\
&\le \sum_{k=1}^{\infty}k^{1/r}P\left(k-1<|X_1|^r\le k\right)\left(\sum_{n=k}^{\infty}\frac{1}{n^{1/r}}\right)\\
&\le \sum_{k=1}^{\infty}\left[k^{2/r}P\left(k-1<|X_1|^r\le k\right)\left(\frac{C}{k^{2/r-1}}\right)\right]\\
&= C\sum_{k=1}^{\infty}kP\left(k-1<|X_1|^r\le k\right)\\
&\le C\left(1+E|X_1|^r\right)<\infty
\end{aligned}$$

From Kronecker Lemma and Corollary 3.6.17, we obtain (ii) and (iii). Note that (iii) also follows from the equivalence sequence theorem in the last chapter on WLLN. Combining (i)-(iii), we get $\frac{1}{n^{1/r}}\sum_{k=1}^{n}(X_k - EX_1) = \frac{1}{n^{1/r}}\sum_{k=1}^{n}(X_k - Y_k) + \frac{1}{n^{1/r}}\sum_{k=1}^{n}(Y_k - EY_k) + \frac{1}{n^{1/r}}\sum_{k=1}^{n}(EY_k - EX_k) \to 0 \quad \text{a.s.}$

■

Theorem 3.6.23 — Strong Law of Large Numbers. Let $X_1, X_2, \ldots$ be independent random variables each with . mean μ. Suppose there exists an $M < \infty$ such that $\mathbb{E}\left[X_n^4\right] \leq M$ for each n (this holds, for example, when the sequence is i.i.d. and $\mathbb{E}\left[X_1^4\right] < \infty$). Then w.p.1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu$$

Proof. Without loss of generality we may assume that $\mu = 0$ for otherwise we can consider the random variables $X_1 - \mu, X_2 - \mu, \ldots$ Our first goal is to show that

$$\mathbb{E}\left[(X_1 + X_2 + \cdots + X_n)^4\right] \leq 3Mn^2 \tag{3.37}$$

To see this, we first note that if $i \notin \{j, k, l\}$,

$$\mathbb{E}\left[X_i X_j X_k X_l\right] = \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j X_k X_l\right] = 0 \tag{3.38}$$

(for the indepence.) When we expand $(X_1 + X_2 + \cdots + X_n)^4$ and then use the sum rule for expectation we get n^4 terms. We get n terms of the form $\mathbb{E}\left[X_i^4\right]$ and $3n(n-1)$ terms of the form $\mathbb{E}\left[X_i^2 X_j^2\right]$ with $i \neq j$. There are also terms of the form $\mathbb{E}\left[X_i X_j^3\right], \mathbb{E}\left[X_i X_j X_k^2\right]$ and $\mathbb{E}\left[X_i X_j X_k X_l\right]$ for distinct i, j, k, l; all of these expectations are zero by 3.38 so we will not bother to count the exact number. We know that $\mathbb{E}\left[X_i^4\right] \leq M$. Also from the Holder inequality, if $i \neq j$

$$\mathbb{E}\left[X_i^2 X_j^2\right] \leq \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[X_j^2\right] \leq \left(\mathbb{E}\left[X_i^4\right]\mathbb{E}\left[X_j^4\right]\right)^{1/2} \leq M$$

Since the sum consists of at most $3n^2$ terms each bounded by M, we get 3.37. Let A_n be the event

$$A_n = \left\{\left|\frac{X_1 + \cdots + X_n}{n}\right| \geq \frac{1}{n^{1/8}}\right\} = \left\{|X_1 + X_2 + \cdots + X_n| \geq n^{7/8}\right\}$$

By the generalized Chebyshev inequality (Exercise 2.5),

$$\mathbb{P}(A_n) \leq \frac{\mathbb{E}\left[(X_1 + \cdots + X_n)^4\right]}{\left(n^{7/8}\right)^4} \leq \frac{3M}{n^{3/2}}$$

In particular, $\sum \mathbb{P}(A_n) < \infty$. Hence by the Borel-Cantelli Lemma,

$$\mathbb{P}(\limsup A_n) = 0$$

Suppose for some $\omega, (X_1(\omega)+\cdots+X_n(\omega))/n$ does not converge to zero. Then there is an $\epsilon > 0$ such that

$$|X_1(\omega)+X_2(\omega)+\cdots+X_n(\omega)| \geq \epsilon n$$

for infinitely many values of n. This implies that $\omega \in \limsup A_n$. Hence the set of ω at which we do not have convergence has probability zero. ■

Corollary 3.6.24 The strong law of large numbers holds in greater generality than under the conditions given here. In fact, if $X_1, X_2, \ldots$ is any i.i.d. sequence of random variables with $\mathbb{E}(X_1) = \mu$, then the strong law holds

$$\frac{X_1+\cdots+X_n}{n} \to \mu, \quad \text{w.p.1.}$$

Question 3.2 Let $X_1, X_2, \ldots$ be independent random variables. Suppose that

$$\mathbb{P}\{X_n = 2^n\} = 1 - \mathbb{P}\{X_n = 0\} = 2^{-n}$$

Show that that $\mathbb{E}[X_n] = 1$ for each n and that w.p.1,

$$\frac{X_1+X_2+\cdots+X_n}{n} \longrightarrow 0$$

Verify that the hypotheses of Theorem 3.6.23 do not hold in this case.

Question 3.3

1. Let X be a random variable. Show that $\mathbb{E}[|X|] < \infty$ if and only if
$$\sum_{n=1}^{\infty} \mathbb{P}\{|X| \geq n\} < \infty$$
2. Let $X_1, X_2, \ldots$ be i.i.d. random variables. Show that
$$\mathbb{P}\{|X_n| \geq n \text{ i.o.}\} = 0$$
if and only if $\mathbb{E}[|X_1|] < \infty$.

Following to introduce the Kolmogorov Zero-One Law.

Assume $X_1, X_2, \ldots$ are independent. Let $\mathcal{F}_n$ be the σ-algebra generated by $X_1, \ldots, X_n$, and let $\mathcal{G}_n$ be the σ-algebra generated by $X_{n+1}, X_{n+2}, \ldots$ Note that $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent σ-algebras, and

$$F_1 \subset \mathcal{F}_2 \subset \cdots, \quad \mathcal{G}_1 \supset \mathcal{G}_2 \supset \cdots$$

Let $\mathcal{F}$ be the σ-algebra generated by $X_1, X_2, \ldots$, i.e., the smallest σ-algebra containing the algebra

$$\mathcal{F}^0 := \bigcup_{n=1}^{\infty} \mathcal{F}_n$$

The tail σ -algebra $\mathcal{T}$ is defined by

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{G}_n$$

(The intersection of σ -algebras is a σ -algebra, so $\mathcal{T}$ is a σ -algebra.) An example of an event in $\mathcal{T}$ is

$$\left\{ \lim_{m\to\infty} \frac{X_1 + \cdots + X_m}{m} = 0 \right\}$$

It is easy to see that this event is in $\mathcal{G}_n$ for each n since

$$\left\{ \lim_{m\to\infty} \frac{X_1 + \cdots + X_m}{m} = 0 \right\} = \left\{ \lim_{m\to\infty} \frac{X_{n+1} + \cdots + X_m}{m} = 0 \right\}$$

Lemma 3.15 — $\mathcal{F}_0$ is "dense" in $\mathcal{F}$. Suppose $\mathcal{F}^0$ is an algebra of events and $\mathcal{F} = \sigma\left(\mathcal{F}^0\right)$. If $A \in \mathcal{F}$, then for every $\epsilon > 0$ there is a n $A_0 \in \mathcal{F}^0$ with

$$\mathbb{P}\left(A \triangle A_0\right) < \epsilon$$

Proof. Let $\mathcal{B}$ be the collection of all events A such that for every $\epsilon > 0$ there is an $A_0 \in \mathcal{F}^0$ such that $\mathbb{P}\left(A \triangle A_0\right) < \epsilon$. (This is the common skill to prove a conclusion when dealing with sigma-algbra). Trivially, $\mathcal{F}^0 \subset \mathcal{B}$. We will show that $\mathcal{B}$ is a σ -algebra and this will imply that $\mathcal{F} \subset \mathcal{B}$. Obviously, $\Omega \in \mathcal{B}$. Suppose $A \in \mathcal{B}$ and let$\epsilon > 0$. Find $A_0 \in \mathcal{F}^0$ such that $\mathbb{P}\left(A \triangle A_0\right) < \epsilon$. Then $A_0^c \in \mathcal{F}^0$ and

$$\mathbb{P}\left(A^c \triangle A_0^c\right) = \mathbb{P}\left(A \triangle A_0\right) < \epsilon$$

Hence $A^c \in \mathcal{B}$. Suppose $A_1, A_2, \ldots \in \mathcal{B}$ and let $A = \cup A_j$. Let $\epsilon > 0$, and find n such that

$$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] \geq \mathbb{P}[A] - \frac{\epsilon}{2} \quad \text{or} \quad \mathbb{P}[A] - \mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] \leq \frac{\epsilon}{2}$$

For $j = 1, \ldots, n$, let $A_{j,0} \in \mathcal{F}^0$ such that

$$\mathbb{P}\left[A_j \triangle A_{j,0}\right] \leq \epsilon 2^{-j-1}$$

Let $A_0 = A_{1,0} \cup \cdots \cup A_{n,0}$ and note that

$$A \triangle A_0 = \cup_{j=1}^{\infty} A_j \triangle \cup_{j=1}^{n} A_{j,0} \subset \left[\bigcup_{j=1}^{n} A_j \Delta A_{j,0}\right] \cup \left[A \backslash \bigcup_{j=1}^{n} A_j\right]$$

and hence $\mathbb{P}\left(A \triangle A_0\right) < \epsilon$ and $A \in \mathcal{B}$. ■

Theorem 3.6.25 — Kolmogorov Zero-One Law. If $A \in \mathcal{T}$ then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proof. Let $A \in \mathcal{T}$ and let $\epsilon > 0$. Find a set $A_0 \in \mathcal{F}^0$ with $\mathbb{P}(A \triangle A_0) < \epsilon$. Then $\mathcal{A}_0 \in \mathcal{F}_n$ for some n. Since $\mathcal{T} \subset \mathcal{G}_n$ and $\mathcal{F}_n$ is independent of $\mathcal{G}_n$, A and A_0 are independent. Therefore $\mathbb{P}(A \cap A_0) = \mathbb{P}(A)\mathbb{P}(A_0)$. Note that $\mathbb{P}(A \cap A_0) \geq \mathbb{P}(A) - \mathbb{P}(A \triangle A_0) \geq \mathbb{P}(A) - \epsilon$. Letting $\epsilon \to 0$, we get $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(A)$. This implies $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ ■

As an example, suppose $X_1, X_2, \ldots$ are independent random variables. Then the probability of the event

$$\left\{\omega : \lim_{n\to\infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = 0\right\}$$

is zero or one.

3.7 Central Limit Theorem

Definition 3.7.1 — Fourier transform. If $g : \mathbb{R} \to \mathbb{R}$, the **Fourier transform** of g is defined by

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-ixy} g(x)dx$$

Definition 3.7.2 g a Schwartz function it is C^∞ and all of its derivatives decay at $\pm\infty$ faster than every polynomial. If g is Schwartz, then $\hat{g}$ is Schwartz.

(Roughly speaking, the Fourier transform sends smooth functions to bounded functions and bounded functions to smooth functions. Very smooth and very bounded functions are sent to very smooth and very bounded functions.) The inversion formula

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy} \hat{g}(y)dy \tag{3.39}$$

holds for Schwartz functions. A straightforward computation shows that

$$\widehat{e^{-x^2/2}} = \sqrt{2\pi} e^{-y^2/2}$$

We also need

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx &= \int_{-\infty}^{\infty} f(x) \left[\int_{-\infty}^{\infty} g(y)e^{-ixy}dy\right] \\
&= \int_{-\infty}^{\infty} g(y) \left[\int_{-\infty}^{\infty} f(x)e^{-ixy}dx\right] dy \\
&= \int_{-\infty}^{\infty} \hat{f}(y)g(y)dy
\end{aligned}$$

R The characteristic function is a version of the Fourier transform.

Definition 3.7.3 The characteristic function of a a random variable X is the function $\phi = \phi_X : \mathbb{R} \longrightarrow \mathbb{C}$ defined by

$$\phi(t) = \mathbb{E}\left[e^{iXt}\right]$$

Proposition 3.7.1 — Properties of Characteristic Functions.

1. $\phi(0) = 1$ and for all t

$$|\phi(t)| = \left|\mathbb{E}\left[e^{iXt}\right]\right| \leq \mathbb{E}\left[\left|e^{iXt}\right|\right] = 1$$

2. ϕ is a continuous function of t. To see this note that the collection of random variables $Y_s = e^{isX}$ are dominated by the random variable 1 which has finite expectation. Hence by the dominated convergence theorem,

$$\lim_{s \to t} \phi(s) = \lim_{s \to t} \mathbb{E}\left[e^{isX}\right] = \mathbb{E}\left[\lim_{s \to t} e^{isX}\right] = \phi(t)$$

In fact, it is uniformly continuous.

3. If $Y = aX + b$ where a, b are constants, then

$$\phi_Y(t) = \mathbb{E}\left[e^{i(aX+b)t}\right] = e^{ibt}\mathbb{E}\left[e^{iX(at)}\right] = e^{ibt}\phi_X(at)$$

4. The function $M_X(t) = e^{tX}$ is often called the **moment generating function** of X. Unlike the characteristic function, the moment generating function does not always exist for all values of t. When it does exist, however, we can use the formal relation $\phi(t) = M(it)$

5. If a random variable X has a density f then

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx = \hat{f}(-t)$$

6. If two random variables have the same characteristic function, then they have the same distribution. This fact is not obvious $-$ this will follow from Proposition 5.6 which gives an inversion formula for the characteristic function similar to 3.39 .

7. If $X_1, \ldots, X_n$ are independent random variables, then

$$\begin{aligned}
\phi_{X_1+\cdots+X_n}(t) &= \mathbb{E}\left[e^{it(X_1+\cdots+X_n)}\right] \\
&= \mathbb{E}\left[e^{itX_1}\right]\mathbb{E}\left[e^{itX_2}\right]\cdots\mathbb{E}\left[e^{-tX_n}\right] \\
&= \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t)
\end{aligned}$$

8. If X is a normal random variable, mean zero, variance 1, then by completing the square in the exponential one can compute

$$\phi(t) = \int_{-\infty}^{\infty} e^{ixt}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = e^{-t^2/2}$$

If Y is normal mean μ, variance σ^2, then Y has the same distribution as $\sigma X + \mu$, hence

$$\phi_Y(t) = \phi_{\sigma X+\mu}(t) = e^{i\mu t}\phi_X(\sigma t) = e^{i\mu t}e^{-\sigma^2 t^2/2}$$

Question 3.4 Show that ϕ is a uniformly continuous function of t.

Theorem 3.7.2 Suppose X is a random variable with characteristic function ϕ. Suppose $\mathbb{E}[|X|] < \infty$ ($\mathbb{E}X$ exists). Then ϕ is continuously differentiable and

$$\phi'(0) = i\mathbb{E}(X)$$

Analysis 3.2 Mainly by the definition of derivation

Proof. Note that for real θ

$$\left|e^{i\theta} - 1\right| \leq |\theta|$$

This can be checked geometrically or by noting that

$$\left|e^{i\theta} - 1\right| = \left|\int_0^\theta ie^{is} ds\right| \leq \int_0^\theta \left|ie^{is}\right| ds = \theta$$

We write the difference quotient

$$\frac{\phi(t+\delta) - \phi(t)}{\delta} = \int_{-\infty}^{\infty} \frac{e^{i(t+\delta)x} - e^{itx}}{\delta} d\mu(x)$$

Note that

$$\left|\frac{e^{i(t+\delta)x} - e^{itx}}{\delta}\right| = \left|\frac{e^{itx}(e^{i\delta x} - 1)}{\delta}\right| \leq \frac{|\delta x|}{\delta} \leq |x|$$

Saying $\mathbb{E}[|X|] < \infty$ is equivalent to saying that $|x|$ is integrable with respect to the measure μ. By the dominated convergence theorem,

$$\begin{aligned}\lim_{\delta \to 0} \int_{-\infty}^{\infty} \frac{e^{i(t+\delta)x} - e^{itx}}{\delta} d\mu(x) &= \int_{-\infty}^{\infty} \left[\lim_{\delta \to 0} \frac{e^{i(t+\delta)x} - e^{itx}}{\delta}\right] d\mu(x) \\ &= \int_{-\infty}^{\infty} ixe^{itx} d\mu(x)\end{aligned}$$

Hence,

$$\phi'(t) = \int_{-\infty}^{\infty} ixe^{itx} d\mu(x)$$

and, in particular,

$$\phi'(0) = i\mathbb{E}[X]$$

∎

Corollary 3.7.3 Suppose X is a random variable with characteristic function ϕ. Suppose $\mathbb{E}\left[|X|^k\right] < \infty$ for some positive integer k. Then ϕ has k continuous derivatives and

$$\phi^{(j)}(t) = i^j \mathbb{E}\left[X^j\right], \quad j = 0, 1 \ldots, k$$

Proof.

$$\mathbb{E}\left[e^{iXt}\right] = \mathbb{E}\left[\sum_{j=0}^{\infty} \frac{(iX)^j t^j}{j!}\right] = \sum_{j=0}^{\infty} \frac{i^j \mathbb{E}\left(X^j\right)}{j!} t^j$$

differentiating both sides, and evaluating at $t = 0$. ■

R The existence of $\mathbb{E}\left[|X|^k\right]$ allows one to justify this rigorously up to the k th derivative.

Theorem 3.7.4 Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with mean μ and variance σ^2 and characteristic function ϕ. Let

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{\sigma^2 n}}$$

and let ϕ_n be the characteristic function of Z_n. Then for each t,

$$\lim_{n\to\infty} \phi_n(t) = e^{-t^2/2}$$

Proof. Without loss of generality we will assume $\mu = 0, \sigma^2 = 1$ for otherwise we can consider $Y_j = \left(X_j - \mu\right)/\sigma$. By Corollary 3.7.3, we have

$$\phi(t) = 1 + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + \epsilon_t t^2 = 1 - \frac{t^2}{2} + \epsilon_t t^2$$

where $\epsilon_t \to 0$ as $t \to 0$. Note that

$$\phi_n(t) = \left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

and hence for fixed t

$$\lim_{n\to\infty} \phi_n(t) = \lim_{n\to\infty}\left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 - \frac{t^2}{2n} + \frac{\delta_n}{n}\right]^n$$

where $\delta_n \to 0$ as $n \to \infty$. Once n is sufficiently large that $\left|\delta_n - \left(t^2/2\right)\right| < n$ we can take logarithms and expand in a Taylor series,

$$\log\left[1 - \frac{t^2}{2n} + \frac{\delta_n}{n}\right] = -\frac{t^2}{2n} + \frac{\rho_n}{n}, \quad n \to \infty$$

where $\rho_n \to 0$. (Since ϕ can take on complex values, we need to take a complex logarithm with $\log 1 = 0$,, but there is no problem provided that $\left|\delta_n - (t^2/2)\right| < n$.) Therefore

$$\lim_{n\to\infty} \log \phi_n(t) = -\frac{t^2}{2}$$

■

Theorem 3.7.5 — Central limit Theorem. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with mean μ and finite variance. If $-\infty < a < b < \infty$, then

$$\lim_{n\to\infty} \mathbb{P}\left\{a \le \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Proof. Let ϕ_n be the characteristic function of $n^{-1/2}\sigma^{-1}(X_1 + \cdots + X_n - n\mu)$. We have already shown that for each $t, \phi_n(t) \to e^{-t^2/2}$. It suffices to show that if μ_n is any sequence of distributions such that their characteristic functions ϕ_n converge pointwise to $e^{-t^2/2}$ then for every $a < b$,

$$\lim_{n\to\infty} \mu_n[a,b] = \Phi(b) - \Phi(a)$$

where

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Fix a, b and let $\epsilon > 0$. We can find a C^∞ function $g = g_{a,b,\epsilon}$ such that

$$\begin{aligned} 0 \le\, & g(x) \le 1, \quad -\infty < x < \infty \\ & g(x) = 1, \quad a \le x \le b \\ & g(x) = 0, \quad x \notin [a-\epsilon, b+\epsilon] \end{aligned}$$

We will show that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) d\mu_n(x) = \int_{-\infty}^{\infty} g(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \tag{3.40}$$

since for any distribution μ

$$\mu[a,b] \le \int_{-\infty}^{\infty} g(x) d\mu(x) \le \mu[a-\epsilon, b+\epsilon]$$

we then get the theorem by letting $\epsilon \to 0$. Since g is a C^∞ function with compact support, it is a Schwartz function. Let

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-ixy} g(x) dx$$

be the Fourier transform. Then, using the inversion formula 3.39 ,

$$\begin{aligned}\int_{-\infty}^{\infty} g(x)d\mu_n(x) &= \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixy}\hat{g}(y)dy\right]d\mu_n(x)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} e^{ixy}d\mu_n(x)\right]\hat{g}(y)dy\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_n(y)\hat{g}(y)dy\end{aligned}$$

since $|\phi_n\hat{g}| \leq |\hat{g}|$ and $\hat{g}$ is L^1, we can use the DCT (dominated convergence theorem) to conclude

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} g(x)d\mu_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-y^2/2}\hat{g}(y)dy$$

Recalling that $e^{-x^2/2} = \sqrt{2\pi}e^{-y^2/2}$, and that

$$\int \hat{f}g = \int f\hat{g}$$

we get 3.40 . ■

Lemma 3.16

$$\int_{-T}^{T}\frac{e^{icx}}{ix}dx = 2\int_{0}^{|c|T}\frac{\sin x}{x}dx$$

Proof.

$$\begin{aligned}\int_{-T}^{T}\frac{\cos cx + i\sin cx}{ix}dx &= \int_{-T}^{T}\frac{\sin cx}{x}dx\\ &=2\int_{0}^{T}\frac{\sin cx}{x}dx\end{aligned}$$

Let $cx = t$, and $x = t/c$

$$= 2\int_{0}^{ct}\frac{\sin t}{t/c}\frac{1}{c}dx$$

■

Lemma 3.17

$$\lim_{T\to\infty}\int_{0}^{T}\frac{\sin x}{x}dx = \frac{\pi}{2}$$

Theorem 3.7.6 Let X be a random variable with distribution μ, distribution function F, and characteristic function ϕ. Then for every $a < b$ such that F is continuous at a and b,

$$\mu[a,b] = F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy} \phi(y) dy \tag{3.41}$$

Proof. Firstly to check the existing of this integral. Fix $T < \infty$. Note that

$$\left| \frac{e^{-iya} - e^{-iyb}}{iy} \right| |\phi(y)| \le (b-a) \tag{3.42}$$

and hence the integral on the right-hand side of (3.41) is well defined. The bound (3.42) allows us to use Fubini's Theorem:

$$\begin{aligned}\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy} \phi(y) dy &= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy} \left[\int_{-\infty}^{\infty} e^{iyx} d\mu(x) \right] dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy \right] d\mu(x)\end{aligned}$$

It is easy to check that for any real c,

$$\int_{-T}^{T} \frac{e^{icx}}{ix} dx = 2 \int_{0}^{|c|T} \frac{\sin x}{x} dx$$

We will need the fact that

$$\lim_{T\to\infty} \int_0^T \frac{\sin x}{x} dx = \frac{\pi}{2}$$

There are two relevant cases.

1. If $x-a$ and $x-b$ have the same sign (i.e., $x < a$ or $x > b$),
$$\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy = \lim_{T\to\infty} \int_0^{T\cdot|x-a|} \frac{\sin x}{x} dx = \int_0^{T\cdot|x-b|} \frac{\sin x}{x} dx = 0$$
2. If $x-a > 0$ and $x-b < 0$, $(x \in (a,b))$
$$\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy = 4 \lim_{T\to\infty} \int_0^T \frac{\sin x}{x} dx = 2\pi$$

We do not need to worry about what happens at $x = a$ or $x = b$ since μ gives zero measure to these points. Since the function

$$g_T(a,b,x) = \int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy$$

is uniformly bounded in T, a, b, x, we can use dominated convergence theorem to conclude that

$$\lim_{T\to\infty} \int_{-\infty}^{\infty} \left[\int_{-T}^{T} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy \right] d\mu(x) = \int 2\pi 1_{(a,b)} d\mu(x) = 2\pi\mu[a,b]$$

(exchange limit with integral) Therefore,

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iya} - e^{-iyb}}{iy} \phi(y) dy = \mu[a,b]$$

∎

3.8 Conditional expection

Definition 3.8.1 — Conditional expection. Let us first consider the case when $\mathcal{G}$ is a finite σ -algebra. In this case there exist disjoint, nonempty events $A_1, A_2, \ldots, A_k$ that generate $\mathcal{G}$ in the sense that $\mathcal{G}$ consists precisely of those sets obtained from unions of the events $A_1, \ldots, A_k$. If $\mathbb{P}\left(A_j\right) > 0$, we define $\mathbb{E}[X \mid \mathcal{G}]$ on A_j to be the average of X on A_j

$$\mathbb{E}[X \mid \mathcal{G}] = \frac{\int_{A_j} X d\mathbb{P}}{\mathbb{P}\left(A_j\right)} = \frac{\mathbb{E}\left[X 1_{A_j}\right]}{\mathbb{P}\left(A_j\right)}, \quad \text{on } A_j \tag{3.43}$$

Note that $\mathbb{E}[X \mid \mathcal{G}]$ is a $\mathcal{G}$ -measurable random variable and that its definition is unique except perhaps on a set of total probability zero.

Theorem 3.8.1 If $\mathcal{G}$ is a finite σ -algebra and $\mathbb{E}[X \mid \mathcal{G}]$ is defined as in (3.43), then for every event $A \in \mathcal{G}$

$$\int_A \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} = \int_A X d\mathbb{P}$$

Theorem 3.8.2 — for infinite $\mathcal{G}$. Suppose X is an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ is a σ -algebra. Then there exists a $\mathcal{G}$ -measurable random variable Y such that if $A \in \mathcal{G}$,

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} \tag{3.44}$$

Moreover, if $\tilde{Y}$ is another $\mathcal{G}$ -measurable random variable satisfying (3.44), then $\tilde{Y} = Y$ a.s.

Proof. The uniqueness is immediate since if $Y, \tilde{Y}$ are $\mathcal{G}$ -measurable random variables satisfying (15), then

$$\int_A (Y - \tilde{Y}) d\mathbb{P} = 0$$

for all $A \in \mathcal{G}$ and hence $Y - \tilde{Y} = 0$ almost surely. To show existence consider the probability space $(\Omega, \mathcal{G}, \mathbb{P})$. Consider the (signed) measure on $(\Omega, \mathcal{G})$,

$$\nu(A) = \int_A X d\mathbb{P}$$

Note that $\nu << P$. Hence by the Radon-Nikodym theorem there is a $\mathcal{G}$ -measurable random variable Y with

$$\nu(A) = \int_A Y d\mathbb{P}$$

■

Corollary 3.8.3 Using the proposition, we can define the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ to be the random variable Y in the proposition. It is characterized by the properties

1. $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ -measurable.
2. For every $A \in \mathcal{G}$

$$\int_A \mathbb{E}[X \mid \mathcal{G}] dP = \int_A X dP$$

The random variable $\mathbb{E}[X \mid \mathcal{G}]$ is only defined up to a set of probability zero.

Proposition 3.8.4 — Properties of Conditional Expectation.

1. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$
2. If $a, b \in \mathbb{R}, \mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$.
3. If X is $\mathcal{G}$ -measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$.
4. If $\mathcal{G}$ is independent of X, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.($\mathcal{G}$ is independent of X if $\mathcal{G}$ is independent of the σ -algebra generated by X. Equivalently, they are independent if for every $A \in \mathcal{G}, 1_A$ is independent of X.)

   Proof. The constant random variable $\mathbb{E}[X]$ is certainly $\mathcal{G}$ -measurable. Also, if $A \in \mathcal{G}$,

$$\int_A \mathbb{E}[X] d\mathbb{P} = \mathbb{P}(A)\mathbb{E}[X] = \mathbb{E}\left[1_A\right]\mathbb{E}[X] = \mathbb{E}\left[1_A X\right] = \int_A X d\mathbb{P}$$

■

5. If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ are σ -algebras, then

$$\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}]$$

   Proof. Clearly, $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}]$ is $\mathcal{H}$ -measurable. If $A \in \mathcal{H}$, then $A \in \mathcal{G}$, and hence

$$\int_A \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] d\mathbb{P} = \int_A \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} = \int_A X d\mathbb{P}$$

■

6. If Y is $\mathcal{G}$ -measurable, then

$$\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$$

   Proof. Clearly $Y\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ -measurable. We need to show that for every $A \in \mathcal{G}$

$$\int_A Y\mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} = \int_A YX d\mathbb{P} \tag{3.45}$$

   We will first consider the case where $X, Y \geq 0$. Note that this implies that $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s., since $\int_A \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} = \int_A X d\mathbb{P} \geq 0$ for every $A \in \mathcal{G}$. Find simple $\mathcal{G}$ -measurable random variables $0 \leq Y_1 \leq Y_2 \leq \cdots$ such that $Y_n \to Y$. Then $Y_n X$ converges monotonically to YX and $Y_n \mathbb{E}[X \mid \mathcal{G}]$ converges monotonically to $Y\mathbb{E}[X \mid \mathcal{G}]$. If $A \in \mathcal{G}$ and

$$Z = \sum_{j=1}^{n} c_j 1_{B_j}, \quad B_j \in \mathcal{G}$$

is a $\mathcal{G}$ -measurable simple random variable,

$$\begin{aligned}
\int_A Z\mathbb{E}[X \mid \mathcal{G}]d\mathbb{P} &= \sum_{j=1}^n c_j \int_A 1_{B_j}\mathbb{E}[X \mid \mathcal{G}]d\mathbb{P} \\
&= \sum_{j=1}^n c_j \int_{A\cap B_j} \mathbb{E}[X \mid \mathcal{G}]d\mathbb{P} \\
&= \sum_{j=1}^n c_j \int_{A\cap B_j} Xd\mathbb{P} \\
&= \int_A \left[\sum_{j=1}^n c_j 1_{B_j}\right] Xd\mathbb{P} \\
&= \int_A ZXd\mathbb{P}
\end{aligned}$$

Hence (3.45) holds for simple nonnegative Y and nonnegative X, and hence by the monotone convergence theorem it holds for all nonnegative Y and nonnegative X. For general Y, X, we write $Y = Y^+ - Y^-, X = X^+ - X^-$ and use linearity of expectation and conditional expectation. ■

3.9 Weak convergence

Definition 3.9.1 — Converge Weakly.
1. A sequence of d.f.s $\{F_n, n \geq 1\}$ is said to **converge weakly** to a d.f. F, written as $F_n \Longrightarrow F$, if $F_n(x) \longrightarrow F(x)$, for all $x \in C(F)$(the continuous condition is to make the limit function F to be a candidate of distribution function.)
2. 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 $X_n \Longrightarrow X$ or $X_n \longrightarrow_d X$, if their d.f.s $F_n(x) = P(X_n \leq x)$ converge weakly $F(x) = P(X \leq x)$

R The r.v.'s Y_n 's are not required to be on the same probability space.

■ **Example 3.14** Let $X \sim F$ and $X_n = X + n^{-1}$. Then,

$$F_n(x) = P\left(X + n^{-1} \leq x\right) = F\left(x - n^{-1}\right) \to F(x-)$$

Thus, we observe the following.
1. The limit of d.f.s may not be a d.f. In fact, in the current example, $F(x-)$ is left continuous. If we turn $F(x-)$ into a right-continuous function (in this case $F(s)$), then F is a proper d.f. and hence we have $F_n \Longrightarrow F$.
2. $\lim_{n\to\infty} F_n(x) = F(x-)$ which equals to $F(x)$ iff $x \in C(F)$. This is why we restrict attention to continuity points in the definition of weak convergence.

3.9.1 Equivalent definitions of weak convergence

Theorem 3.9.1 — Portmanteau Theorem. The following statements are equivalent.

(a) $X_n \Longrightarrow X$ or $X_n \rightarrow_d X$

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

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

(d) $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$ for all sets A with $P(X \in \partial A) = 0$ where ∂A is the boundary of A.

(e) $\mathrm{E}g(X_n) \to Eg(X)$ for all bounded continuous function g

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

(g) $\lim_n \psi_n(t) = \psi(t)$, where $\psi_n(t)$ and $\psi(t)$ are the c.f.s of X_n and X, respectively.

R

1. Note that (d) uses convergence in distribution to define weak convergence, a concept which is closest to the original definition. Also, (b) and (c) are similar to (d).
2. To help remember (b) and (c), think about what can happen when $P(X_n = x_n) = 1, x_n \to x$, and x lies on the boundary of the set. If G is open, we can have $x_n \in G$ for all n, but $x \notin G$ (so RHS of (b) could be 0). If K is closed, we can have $x_n \notin K$ for all n, but $x \in K$ (so RHS of (c) could be 1)

Proof. Proof of Theorem 3.9.1

1. "(a) $\Longrightarrow$ (b)": Let Y, Y_n have the same d.f.'s as X, X_n and $Y_n \to Y$ a.s. Since G is open, we can show

$$\liminf_n I_G(Y_n) \geq I_G(Y) \quad \text{a.s.}$$

This can be seen as follows.

(a) If $I_G(Y) = 0$, the proof is trivial;

(b) If $I_G(Y) = 1, \Longleftrightarrow Y \in G$. Since G is open (the open set on R would not only contain one point Y), for large enough n, we have $Y_n \in G, \Longleftrightarrow I_G(Y_n) = 1$. Hence, $LHS = 1 = I_G(Y) = RHS$

Now applying Fatou's Lemma, we get

$$P(Y \in G) = EI_G(Y) \leq E\liminf_n I_G(Y_n) \leq_F \liminf_n EI_G(Y_n) = \liminf_n P(Y_n \in G)$$

2. "(b) $\Longleftrightarrow$ (c) ". A is open $\Longleftrightarrow A^c$ is closed, and $P(A) = 1 - P(A^c)$

3. "(b) + (c) $\Longrightarrow$ (d)." Let $\bar{A}$ and A^0 be the closure and interior of A, respectively. Then, $\partial A = \bar{A} - A^0$ and $0 = P(X \in \partial A) = P(\bar{A}) - P(A^0)$, so

$$P(X \in \bar{A}) = P(X \in A) = P(X \in A^0)$$

Using (b) and (c) now,

$$P(X \in A) = P(X \in \bar{A}) \geq \limsup_{n\to\infty} P(X_n \in \bar{A}) \geq \limsup_{n\to\infty} P(X_n \in A)$$
$$P(X \in A) = P\left(X \in A^0\right) \leq \liminf_{n\to\infty} P\left(X_n \in A^0\right) \leq \liminf_{n\to\infty} P(X_n \in A)$$

from which the proof follows.

4. "(d) $\Longrightarrow$ (a)." Let $x \in C(F)$ and $A = (-\infty, x]$, so $P(X \in \partial A) = P(X = x) = 0$. From (d), we have $P(X_n \leq x) = P(X_n \in A) \to P(X \in A) = P(X \leq x)$
5. "(a) $\Longrightarrow$ (e)." By Skorokhod representation theorem, let $Y_n =_d X_n$ and $Y =_d X$ with $Y_n \to Y$ almost surely. Since g is continuous $g(Y_n) \to g(Y)$ almost surely and the bounded convergence theorem implies

$$Eg(X_n) = Eg(Y_n) \to Eg(Y) = Eg(X)$$

For convenience, we give Skorokhod's representation theorem below.

Theorem 3.9.2 — Skorokhod's representation theorem. Suppose that $F_n \Longrightarrow F$. Let $\mathcal{B}_{[0,1]}$ denote the Borel sets in [0,1] and $\lambda_{[0,1]}$ the Lebesgue measure restricted to $[0,1]$. Then there exist r.v.'s Y and

$$\{Y_n, n \geq 1\} \text{ on } \left((0,1), \mathcal{B}_{(0,1)}, P_\lambda = \lambda_{(0,1)}\right) \text{ s.t.}$$

(a) $Y_n \sim F_n, Y \sim F$
(b) $Y_n \to Y$ a.s. as $n \to \infty$

6. "(e) $\Longrightarrow$ (a)" (the connection between expectation and r.v. is inductor function.)Let

$$g_{x,\epsilon}(y) = \begin{cases} 1 & y \leq x \\ 0 & y \geq x + \epsilon \\ \text{linear} & x \leq y \leq x + \epsilon \end{cases}$$

Since $g_{x,\epsilon}(y) = 1$ for $y \leq x, g_{x,\epsilon}$ is continuous, and $g_{x,\epsilon}(y) = 0$ for $y > x + \epsilon$

$$\begin{aligned} \limsup_{n\to\infty} P(X_n \leq x) &= \limsup_{n\to\infty} EI\{X_n \leq x\} \leq_{\text{by figure}} \limsup_{n\to\infty} Eg_{x,\epsilon}(X_n) =_e Eg_{x,\epsilon}(X) \\ &\leq EI\{X \leq x + \epsilon\} = P(X \leq x + \epsilon) \end{aligned}$$

Letting $\epsilon \to 0$ gives $\limsup_{n\to\infty} P(X_n \leq x) \leq P(X \leq x)$. The last conclusion is valid for any x. To get the other direction, we observe

$$\begin{aligned} \liminf_{n\to\infty} P(X_n \leq x) &= \liminf_{n\to\infty} EI\{X_n \leq x\} \geq \liminf_{n\to\infty} Eg_{x-\epsilon,\epsilon}(X_n) = Eg_{x-\epsilon,\epsilon}(X) \\ &\geq EI\{X \leq x - \epsilon\} = P(X \leq x - \epsilon) \end{aligned}$$

Letting $\epsilon \to 0$ gives $\liminf_{n\to\infty} P(X_n \leq x) \geq P(X < x) = P(X \leq x)$ if x is continuity point. The results for lim sup and lim inf combine to give the desired result.

7. "$(e) \Longrightarrow (f)$." We will prove (f) for $g(x) = h(x)I_{(-\infty,b]}(x)$, where h is bounded and continuous, and $b \in C(F)$; the general result follows similarly. To apply (e), we need to approximate g by a continuous function g_1 defined by

$$\begin{aligned} g_1(x) =& g(x) && \text{if } x \notin (b, b+\delta) \\ =& g(b)[1+(b-x)/\delta] && \text{if } x \in (b, b+\delta) \quad \text{which is a linear function} \end{aligned}$$

Also define

$$\begin{aligned} g_2(x) =& \quad g(b)[1+(x-b)/\delta] && \text{if } x \in (b-\delta, b) \\ =& \quad g(b)[1+(b-x)/\delta] && \text{if } x \in [b, b+\delta) \\ =& \quad 0 && \text{otherwise.} \end{aligned}$$

Now $|E[g(X_n) - g_1(X_n)]| \le |Eg_2(X_n)|$ and $|E[g(X) - g_1(X)]| \le |Eg_2(X)|$, so

$$\begin{aligned} |E[g(X_n) - g(X)]| &\le |E[g(X_n) - g_1(X_n)]| + |E[g_1(X_n) - g_1(X)]| + |E[g_1(X) - g(X)]| \\ &\le |Eg_2(X_n)| + |E[g_1(X_n) - g_1(X)]| + |Eg_2(X)| \\ &\to_e 2|Eg_2(X)| \quad \text{as } n \to \infty \end{aligned}$$

since by assumption (e), $E[g_1(X_n) - g_1(X)] \to 0$ and $Eg_2(X_n) \to Eg_2(X)$. But observe that $|g_2(x)| \le |g(b)|I_{\{(b-\delta,b+\delta)\}}$, we have

$$|Eg_2(X)| \le |g(b)|P(b-\delta < X < b+\delta) \to 0 \quad \text{as } \delta \downarrow 0$$

by the assumption that $b \in C$, i.e., $P(X = b) = 0$. Hence (f) holds.

8. "$(f) \Longrightarrow (a)$." Suppose that (c) holds now. Let $b \in C$. Take $h(x) \equiv 1$ for all x, we have that, if $c \in C$

$$\begin{aligned} P(X_n \le b) &\ge P(a \le X_n \le b) \\ &\longrightarrow_f P(a \le X \le b) \text{ as } n \to \infty \\ &\longrightarrow P(X \le b) \text{ as } a \to -\infty \text{ through } C \end{aligned}$$

A similar argument, but taking the limit in the other direction, yields for $b' \in C$

$$\begin{aligned} P(X_n \ge b') &\ge P(c \ge X_n \ge b') \text{ if } c \ge b' \\ &\longrightarrow P(c \ge X \ge b') \text{ as } n \to \infty \\ &\longrightarrow P(X \ge b') \quad \text{as } c \to \infty \text{ through } C \end{aligned}$$

It follows that, if $b, b' \in C$ and $b < b'$, then for any $\epsilon > 0$, there exists N such that

$$P(X \le b) - \epsilon \le P(X_n \le b) \le P(X_n \le b') \le P(X \le b') + \epsilon$$

for all $n \ge N$. Now letting $n \to \infty, \epsilon \downarrow 0$, and $b' \downarrow b$ through C, we obtain that $F_{X_n}(b) \equiv P(X_n \le b) \to P(X \le b) \equiv F_X(b)$. This proves (a)

9. "$(g) \Longleftrightarrow (a)$." This will be proved somewhere else.

∎

3.9.2 Helly's selection theorem and tightness

Theorem 3.9.3 — 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

$$\lim_{k\to\infty} F_{n_k}(x) = F(x), \quad \text{for all } x \in C(F)$$

Lemma 3.18 Suppose that G is a bounded nondecreasing function of D, which is a dense subset of $(-\infty,\infty)$. Define

$$F(x) = \lim_{y\in D, y\to x-} G(y), \quad H(x) = \lim_{y\in D, y\to x+} G(y)$$

Then,

1. $F(x)$ (or $H(x)$) is a left (or right) continuous, nondecreasing function on $(-\infty,\infty)$ with $C(F) \supset C(G)$
2. $F(x) = G(x)$ (or $H(x) = G(x)$) for $x \in C(G)$

Proof. Let $F(x) = a$. For any $\varepsilon > 0$, there exists $x' \in D$ with $x' < x$ and $a \geq G(x') > a-\varepsilon$. Hence, for $y \in D \cap (x',x)$, we have $F(y) = \lim_{z\in D, z\to y-} G(z) \geq G(x') > a-\varepsilon$, implying $F(x-) \geq a-\varepsilon$, and thus $F(x-) \geq a$ by letting $\varepsilon \searrow 0$. Since F inherits the monotonicity of G, necessarily $F(x-) \leq F(x) = a$, whence $F(x-) = a = F(x)$. Now let $x \in C(G)$. Choose $y_n \nearrow x$ where $y_n \in D$; also choose $x_n \searrow x \in C(G)$ where $x_n \in D$. It follows that

$$G(x) \longleftarrow G(y_n) \leq F(y_{n+1}) \leq F(x) \leq F(x_n) \leq G(x_n) \longrightarrow G(x)$$

yielding the final statement of the lemma. The second part can be shown similarly. ∎

Proof. Proof of Theorem 3.9.3: The proof uses a so-called "diagonalisation method". Let $r_1, r_2, \ldots$ be an enumeration of all the rational numbers, denoted by D. The number sequence $\{F_n(r_1), n \geq 1\}$ is bounded, hence by Bolzano-Weierstrass theorem, there exists a subsequence $\{F_{1k}, k \geq 1\}$ of the given sequence such that the limit

$$\lim_{k\to\infty} F_{1k}(r_1) = l_1 \in [0,1]$$

exists. Next, the sequence $\{F_{1k}(r_2), k \geq 1\}$ is bounded, hence, there exists a subsequence $\{F_{2k}, k \geq 1\}$ of $\{F_{1k}, k \geq 1\}$ such that the limit

$$\lim_{k\to\infty} F_{2k}(r_2) = l_2 \in [0,1]$$

since $\{F_{2k}\}$ is a subsequence of $\{F_{1k}\}$, it converges also at r_1 to l_1. Continuing, we obtain

$$\begin{array}{ll} F_{11}, F_{12}, \ldots, F_{1k}, \ldots & \text{converging at } r_1 \\ F_{21}, F_{22}, \ldots, F_{2k}, \ldots & \text{converging at } r_1, r_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ F_{j1}, F_{j2}, \ldots, F_{jk}, \ldots & \text{converging at } r_1, \ldots, r_j \end{array}$$

Now consider the diagonal sequence $\{F_{kk}, k \geq 1\}$. We assert that it converges at every $r_j, j \geq 1$. To see this, let r_j be given. Apart from the first $j-1$ terms, the sequence $\{F_{kk}, k \geq 1\}$ is a subsequence of $\{F_{jk}, k \geq 1\}$, which converges at r_j and hence $\lim_k F_{kk}\left(r_j\right) = l_j$, as desired.

We have thus proved the existence of an infinite subsequence $\{n_k\}$ and a function G defined and increasing on D such that

$$\forall r \in D : \lim_k F_{n_k}(r) = G(r)$$

Note that G is bounded and nondecreasing, but it is not necessarily right continuous. To fix that, we define a function F from G on R as follows:

$$F(x) = \lim_{y \in D, y \to x+} G(y)$$

From the above lemma, $F(x)$ is a right continuous, nondecreasing function on $(-\infty, \infty)$.

To complete the proof, let $x \in C(F)$. For any $\epsilon > 0$, there exists $\delta > 0$ such that, when $|y - x| \leq \delta$, we have $|F(y) - F(x)| < \epsilon$. In particular, we can choose two rationals $r_1 < x$ and $s > x$ with $r_1, s \in (x - \delta, x + \delta)$, such that $F(x) - F(r_1) < \varepsilon$ and $F(s) - F(x) < \varepsilon$. Furthermore, we can choose another rational r_2 with $r_1 < r_2 < x < s$ so that

$$F(x) - \varepsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(s) < F(x) + \varepsilon$$

Since $F_{n_k}(r_2) \to G(r_2) \geq F(r_1)$, and $F_{n_k}(s) \to G(s) \leq F(s)$, it follows that if k is large, we have

$$F(x) - \varepsilon < F_{n_k}(r_2) \leq F_{n_k}(x) \leq F_{n_k}(s) < F(x) + \varepsilon$$

namely,

$$|F_{n_k}(x) - F(x)| < \varepsilon$$

■

R

1. (i) We should note that the limit of a sequence of d.f.s may not be a d.f. (Consequently, the limit in Helly Selection Theorem may not be a d.f. either.) For example, if a $+b+c=1$, and

   $$F_n(x) = aI_{\{x\geq n\}} + bI_{\{x\geq -n\}} + cG(x)$$

   where G is a d.f., then $\lim_n F_n(x) = F(x) := b + cG(x)$. But, F is *NOT* a d.f. as $F(-\infty) = b$ and $F(\infty) = b + c = 1 - a$. In another words, an amount of mass a escapes to ∞, and mass b escapes to $-\infty$
   This type of convergence is sometimes called "**vague convergence**", which is weaker than the weak convergence since it allows mass to escape. For convenience, we write $F_n \Longrightarrow_v F$ if $F_n(x) \longrightarrow F(x)$ for all $x \in C(F)$
2. The last example raises a question: how do we make sure that the limit of d.f.s is still a d.f., or althernatively, when can we conclude that no mass is lost after taking the limit? To answer this question, we need a new concept: tight, as given below.

Definition 3.9.2 — Tight. A sequence of d.f.'s $\{F_n, n \geq 1\}$ is said to be **tight** if, for all $\epsilon > 0$, there is an $M = M_\epsilon$ (free of n) so that

$$\limsup_{n\to\infty} [1 - F_n(M) + F_n(-M)] \leq \epsilon, \quad \text{or} \quad \limsup_{n\to\infty} P(|X_n| > M) \leq \epsilon$$

That is, all of the probability measures give most of their mass to the same finite interval; mass does not "escape to infinity".

Theorem 3.9.4 Every subsequential limit is the d.f. of a probability measure iff the sequence F_n is tight.

Proof. Suppose that the sequence is tight and $F_{n_k} \Longrightarrow_v F$. Let $r < -M_\varepsilon$ and $s > M_\varepsilon$ be continuity points of F. Since $F_{n_k}(r) \to F(r)$ and $F_{n_k}(s) \to F(s)$, we have

$$\begin{aligned}
1 - F(s) + F(r) &= \lim_{k\to\infty} (1 - F_{n_k}(s) + F_{n_k}(r)) \\
&\leq \limsup_{k\to\infty} (1 - F_{n_k}(M_\varepsilon) + F_{n_k}(-M_\varepsilon)) \\
&\leq \lim_{n\to\infty} (1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)) \\
&\leq \varepsilon
\end{aligned}$$

which in turn implies that

$$\lim_{x\to\infty} [1 - F(x) + F(-x)] \leq \varepsilon$$

(LHS does have a limit since it is a non-increasing function of x and also has lower bound 0.) Since ε is arbitrary, it follows $1 - F(x) + F(-x) \to 0$ as $x \to \infty$. Hence, $1 - F(x) \to 0$ and $F(-x) \to 0$ as $x \to \infty$ Therefore, F is the d.f. of a probability measure.

To prove the converse, now suppose that F_n is not tight. In this case, there is an $\varepsilon_0 > 0$ and a subsequence $n_k \to \infty$ so that

$$1 - F_{n_k}(k) + F_{n_k}(-k) \geq \varepsilon_0$$

for all k. By passing to a further subsequence $F_{n_{k_j}}$, we can suppose that $F_{n_{k_j}} \Longrightarrow vF$. Let $r < 0 < s$ be continuity points of F. Since $F_{n_{k_j}}(r) \to F(r)$ and $F_{n_{k_j}}(s) \to F(s)$, we have

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{j\to\infty}\left(1 - F_{n_{k_j}}(s) + F_{n_{k_j}}(r)\right) \\ &\geq \liminf_{j\to\infty}\left(1 - F_{n_{k_j}}(k_j) + F_{n_{k_j}}(-k_j)\right) \\ &\geq \varepsilon_0 \end{aligned}$$

which in turn implies that

$$\lim_{x\to\infty}[1 - F(x) + F(-x)] \geq \varepsilon_0$$

Hence, we can NOT have $1 - F(x) \to 0$ and $F(-x) \to 0$ to hold true at the same time as $x \to \infty$ Therefore, F is NOT the d.f. of a probability measure. ■

Corollary 3.9.5 A sequence of d.f.'s $\{F_n, n \geq 1\}$ converges vaguely to $F(x)$, denoted by $F_n \Longrightarrow_v F$. Then F is a d.f. $\Longleftrightarrow$ F_n is tight.

The following sufficient condition for tightness is often useful.

Theorem 3.9.6 If there is a $\psi \geq 0$ so that $\psi(x) \to \infty$ as $|x| \to \infty$ and

$$C := \sup_n \int \psi(x) dF_n(x) = \sup_n E\psi(X_n) < \infty$$

then F_n is tight. (Here we assume that $X_n \sim F_n$.)

Proof. From

$$C = \sup_n \int \psi(x) dF_n(x) \geq \int_{[-M,M]} \psi(x) dF_n(x) \geq \left(\inf_{|x|\geq M} \psi(x)\right) \int_{[-M,M]} dF_n(x)$$

we get

$$1 - F_n(M) + F_n(-M) \leq \frac{C}{\inf_{|x|\geq M} \psi(x)} \longrightarrow 0$$

■

Theorem 3.9.7 — Polya Theorem–Pointwise weak convergence of F_n to F holds uniformly

if F is continuous.. If $F_n \Longrightarrow F$, and F is continuous, then

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

Proof. Note that

$$\lim_{t\to\infty} [1 - F_n(t)] = \lim_{t\to\infty} [1 - F(t)] = 0, \quad \lim_{t\to-\infty} F_n(t) = \lim_{t\to-\infty} F(t) = 0$$

For any $\epsilon > 0$, we can choose sufficiently large M such that

$$\sup_{t\in(-\infty,-M]} |F_n(t) - F(t)| \le \epsilon, \quad \sup_{t\in[M,\infty]} |F_n(t) - F(t)| \le \epsilon$$

Since F is continuous, it is uniformly continuous on $[-M, M]$. So choose n sufficiently large to get

$$\sup_{t\in[-M,M]} |F_n(t) - F(t)| \le \epsilon$$

Combining the above results, for n sufficiently large, we get

$$\sup_{t\in[-\infty,\infty]} |F_n(t) - F(t)| \le 3\epsilon$$

This proves our theorem. ■

3.9.3 Additional topic: Stable convergence and mixing"

Renyi introduced and developed the ideas of limit theorems which are mixing or stable. These concepts are a strengthening of the idea of weak convergence of r.v.s. In this expository note we point out some equivalent definitions of mixing and stability and discuss the use of these concepts in several contexts. Further, we show how a central limit theorem for martingales can be obtained directly using stability. Recall that, if $\{Y_n\}$ is a sequence of r.v.'s with d.f. F_n, then Y_n is said to converge in distribution to Y, a r.v. with d.f. F, if

$$\lim_{n\to\infty} F_n(x) = F(x), \quad x \in C(F)$$

where $C(F)$ is the continuity points of F. We shall write this as

$$Y_n \longrightarrow_d Y, \quad \text{or} \quad F_n \Longrightarrow F$$

A strengthening of convergence in distribution is stable convergence in distribution, which is a property of the sequence of rv's $\{Y_n\}$ on the same probability space rather than of the corresponding sequence of d.f.s.

Definition 3.9.3 Suppose that $Y_n \longrightarrow_d Y$, where all the Y_n are on the same probability space $(\Omega, \mathcal{F}, P)$, we say that the convergence is stable if

1. for all continuity points of Y and all events $E \in \mathcal{F}, \lim_{n\to\infty} P(\{Y_n \leq y\} \cap E) = Q_y(E)$ exists, and
2. $Q_y(E) \to P(E)$ as $y \to \infty$

We write this as

$$Y_n \longrightarrow_d Y(\text{stably}), \quad \text{or} \quad F_n \Longrightarrow F(\text{stably})$$

In other words, the first part of the definition is equivalent to saying: for all events E such that $P(E) > 0$, the distribution of Y, conditional on B, converges in law to some distribution which may depend on B and which must, as the $\{Y_n\}$ are tight, be proper.

We now give an example of convergence in distribution but not stably.

Example 3.15 Let X and X' be i.i.d. non-degenerate r.v.'s. Let

$$Z_n = \begin{array}{ll} X & \text{for } n \text{ odd} \\ X' & \text{for } n \text{ even} \end{array}$$

Then we have $Z_n \longrightarrow_d X$, but we don't have $Z_n \longrightarrow_d X(\text{stably})$

Proof. Now $Z_n \longrightarrow_d X$ holds since

$$\begin{aligned} P(Z_n \leq x) = P(X \leq x) & \qquad \text{for } n \text{ odd} \\ & \qquad P(X' \leq x) \text{ for } n \text{ even} \\ = P(X \leq x) = F(x) & \end{aligned}$$

To see why we don't have $Z_n \longrightarrow_d X$ (stably), take $E = \{X \leq y\}$. Then,

$$\begin{aligned} P(Z_n \leq x, E) =& P(X \leq x, X \leq y) \quad \text{for } n \text{ odd} \\ & P(X' \leq x, X \leq y) \quad \text{for } n \text{ even} \\ =& P(X \leq x \wedge y) \quad \text{for } n \text{ odd} \\ & P(X' \leq x) P(X \leq y) \quad \text{for } n \text{ even} \\ =& F(x \wedge y) \quad \text{for } n \text{ odd} \\ & F(x)F(y) \quad \text{for } n \text{ even.} \end{aligned}$$

since $F(x \wedge y) > F(x)F(y)$ whenever $0 < F(x) \vee F(y) < 1$, the limit $P(Z_n \leq x, E)$ does not exist, which proves our claim. ■

Despite this dependence on the sequence $\{Y_n\}$, the requirement that a limit theorem be stable is quite weak. Most known limit theorems are in fact stable and if a limit theorem is not stable one can choose a subsequence along which it will be stable.

Definition 3.9.4 A sequence $\{Z_n\}$ of L_1 r.v.s is said to converge weakly in L_1, to Z if

$$\lim_{n\to\infty} E(Z_n\eta) = E(Z\eta)$$

for all bounded $\mathcal{F}$ -measurable r.v.'s η or equivalently

$$\lim_{n\to\infty} E(Z_n I_E) = E(ZI_E)$$

for all $\mathcal{F}$ -measurable events $E, P(E) > 0$. We denote this by

$$Z_n \longrightarrow Z\left(\text{ weakly in } L^1\right)$$

As an example, if $\exp(itY_n) \longrightarrow \exp(itY)$ (weakly in L_1) for each real t, then clearly, $Y_n \to_d Y$. Therefore, weak convergence in L^1 is a useful tool in proving convergence in distribution; see an example on martingale CLT given later.

The condition of weak convergence in L^1 is stronger than uniform integrability but weaker than L^1 -convergence. That is,

$$''L^1 \text{ -convergence" } \Longrightarrow \text{ "weak convergence in } L^{1''} \Longrightarrow \text{ "uniform integrability".}$$

The first relation follows from

$$|E(Z_n I_E) - E(ZI_E)| = |E[(Z_n - Z) I_E)]| \leq E|Z_n - Z| \longrightarrow 0$$

In fact, Neveu (1965, Propositions II.5.3 and IV.2.2, "Mathematical Foundations of the Calculus of Probability") showed even stronger results: suppose that Z and Z_n 's are r.v.'s,

1. $Z_n \longrightarrow Z$ in $L^1 \Longleftrightarrow E(Z_n I_E) \longrightarrow E(ZI_E)$ uniformly in $E \in \mathcal{F}$
2. if $Z_n \longrightarrow Z$ (weakly in L^1), then $\{Z_n\}$ is uniformly integrable (u.i.).

Despite this dependence on the sequence $\{Y_n\}$, the requirement that a limit theorem be stable is quite weak. Most known limit theorems are in fact stable and if a limit theorem is not stable one can choose a subsequence along which it will be stable.

The following proposition gives a number of equivalent definitions of stability. See (Aldous and Eagleson, 1978).

Proposition 3.9.8 Suppose that $F_n \Longrightarrow F$, The following conditions are equivalent:

1. $Y_n \Longrightarrow F_Y$ (stably);
2. For all fixed $\mathcal{F}$ -measurable rv's σ, the sequence of random vectors (Y_n, σ) converges jointly in distribution;
3. For each fixed real t, the sequence of (complex-valued) rv's $\exp\{itY_n\}$ converges weakly in L^1
4. For all fixed k and $B \in \sigma(Y_1, \ldots, Y_k), P(B) > 0, \lim_{n\to\infty} P(Y_n \leq x \mid B)$ exists for a countable dense set of points x

Theorem 3.9.9 — Continuous mapping theorem. Let g be a measurable function and $D_g = \{x : g \text{ is discontinuous at } x\}$. If $X_n \Longrightarrow X_\infty$ and $P(X_n \in D_g) = 0$ then $g(X_n) \Longrightarrow g(X)$. If in addition g is bounded then $Eg(X_n) \to Eg(X_\infty)$

R D_g is always a Borel set.

Proof. We wish to apply Theorem ??. Let f be any bounded continuous function. By Skorokhod representation theorem, let $Y_n =_d X_n$ with $Y_n \to Y_\infty$ almost surely. since f is continuous, then $D_{f\circ g} \subset D_g$ so $P(Y_\infty \in D_{f\circ g}) = 0$. Then, for $\omega \in \{\omega : Y_\infty(\omega) \notin D_{f\circ g}\} \cap \{\omega : Y_n(\omega) \to Y_\infty(\omega)\} := A_1 \cap A_2$, we have $f(g(Y_n(\omega))) \to f(g(Y_\infty(\omega)))$. since $P(A_1 \cap A_2) = 1 - P(A_1^c \cup A_2^c) \geq 1 - P(A_1^c) - P(A_2^c) = 1$, we have

$$f(g(Y_n)) \to f(g(Y_\infty))$$

since f is also bounded then the bounded convergence theorem implies $Ef(g(Y_n) \to Ef(g(Y_\infty)))$. Then we apply Theorem ?? to get the desired result.

The second conclusion is easier. since $P(Y_\infty \in D_g) = 0$, $f(g(Y_n)) \to f(g(Y_\infty))$ almost surely, and desired result follows from the bounded convergence theorem. ■

3.10 Characteristic Function

Definition 3.10.1 The characteristic function (c.f.) for a random variable (r.v.) X in $\mathcal{R}$ with distribution function (d.f.) F is defined to be

$$\psi(t) = \psi_X(t) = Ee^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x) = E\cos(tX) + i\sin(tX)$$

The subscript X in $\psi_X(t)$ can be omitted if there is no confusion.

3.10.1 Some examples of characteristic functions

Definition 3.10.2 1. Standard normal:

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \\ \text{c.f.} & \psi(t) = e^{-t^2/2} \end{array}$$

Proof.

$$\psi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx$$

Therefore,

$$\begin{aligned}\psi'(t) &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\sin(tx)e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sin(tx)de^{-x^2/2}\\ &= \frac{1}{\sqrt{2}}\left(\sin(tx)\cdot e^{-x^2/2}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -t\cos(tx)e^{-x^2/2}dx\right)\\ &= \frac{t}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(tx)e^{-x^2/2}dx = -t\psi(t)\end{aligned}$$

Then, $\psi'(t)/\psi(t) = \frac{d}{dt}\ln\psi(t) = -t$, resulting in $\ln\psi(t) = -t^2/2 + C$. Setting $t = 0$, we get $C = 0$ Finally, we get $\psi(t) = e^{-t^2/2}$ ■

2. Uniform $[0,a]$

$$\text{p.d.f.}\quad f(x) = \frac{1}{a}I\{0 \le x \le a\}$$

$$\text{c.f.}\quad \psi(t) = \int_0^a e^{itx}\frac{1}{a}dx = \frac{1}{a}\frac{e^{itx}}{it}\Big|_{x=0}^{a} = \frac{e^{iat}-1}{iat}$$

Uniform $[-a,a]$

$$\text{p.d.f.}\quad f(x) = \frac{1}{2a}I\{-a \le x \le a\}$$

$$\text{c.f.}\quad \psi(t) = \int_{-a}^a e^{itx}\frac{1}{2a}dx = \frac{1}{2a}\frac{e^{itx}}{it}\Big|_{x=-a}^{a} = \frac{\sin(at)}{at}$$

3. Triangular $[-a,a]$:

$$\text{p.d.f.}\quad f(x) = \frac{1}{a}\left(1-\frac{|x|}{a}\right)I\{|x|<a\} = \frac{1}{a}\left(1-\frac{|x|}{a}\right)^{+}$$

$$\text{c.f.}\quad \psi(t) = \frac{2(1-\cos(at))}{a^2t^2}$$

Proof. Let X,Y be i.i.d. r.v.'s from Uniform $(-b,b)$ with $b = a/2$. Then,

$$\psi_X(t) = \psi_Y(t) = \frac{\sin(bt)}{bt}$$

It is easy to show that the convolution of X and Y (or the d.f. of $X+Y$) is Triangular $[-a,a]$, whose c.f. is

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) = \frac{\sin^2(bt)}{b^2t^2} = 2\frac{2\sin^2(at/2)}{a^2t^2} = \frac{2(1-\cos(at))}{a^2t^2}$$

■

Proof. Direct proof. By definition,

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} e^{itx} f(x)dx \\
&= \frac{2}{a}\int_0^a \left(1-\frac{x}{a}\right)\cos(tx)dx \\
&= \frac{2}{at}\int_{x=0}^a \left(1-\frac{x}{a}\right)d\sin(tx) \\
&= \frac{2}{2t}\left[\left(1-\frac{x}{a}\right)\sin(tx)\right]\Big|_{x=0}^a + \frac{2}{a^2t}\int_{x=0}^a \sin(tx)dx \\
&= \frac{1-\cos(ax)}{a^2t^2}
\end{aligned}$$

■

4. Inverse Triangular (or Polya distribution):

$$\text{p.d.f.} \quad f(x) = \frac{1}{\pi}\frac{1-\cos(ax)}{ax^2}$$

$$\text{c.f.} \quad \psi(t) = \left(1-\frac{|t|}{a}\right)I\{|t|<a\} = \left(1-\frac{|t|}{a}\right)^+$$

Note: Here $\psi(t)$ has a bounded support, which proves useful later on. See the section on Esseen's smooth lemma.

Proof. Apply Theorem 3.10.10 below (i.e. $f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\psi(t)dt$) to the triangular distribution above to get

$$\frac{1}{a}\left(1-\frac{|x|}{a}\right)^+ = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{2(1-\cos(at))}{a^2t^2}dt = \frac{1}{a}\int_{-\infty}^{\infty} e^{is(-x)}\left(\frac{1}{\pi}\frac{(1-\cos(as))}{as^2}\right)ds$$

Let $s=y$, and $x=-t$, we have

$$\left(1-\frac{|t|}{a}\right)^+ = \int_{-\infty}^{\infty} e^{ity}\left(\frac{1}{\pi}\frac{(1-\cos(ay))}{ay^2}\right)dy = \int_{-\infty}^{\infty} e^{ity} f(y)dy$$

■

5. Exponential distribution

$$\text{p.d.f.} \quad f(x) = e^{-x}, x \geq 0$$

$$\text{c.f.} \quad \psi(t) = \frac{1}{1-it}$$

Proof. Integrating gives

$$\psi(t) = \int_0^\infty e^{itx} e^{-x} dx = \int_0^\infty e^{x(it-1)} dx = \left. \frac{e^{x(it-1)}}{(it-1)} \right|_{x=0}^{\infty} = 0 - \frac{1}{it-1} = \frac{1}{1-it}$$

■

6. Gamma distribution

$$\text{p.d.f.} \quad f(x) = \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x}, x \geq 0$$

$$\text{c.f.} \quad \psi(t) = \frac{1}{(1 - it/\lambda)^c}$$

7. Double exponential distribution

$$\text{p.d.f.} \quad f(x) = \frac{1}{2} e^{-|x|}, x \geq 0$$

$$\text{c.f.} \quad \psi(t) = \frac{1}{1+t^2}$$

Proof. Integrating gives

$$\begin{aligned}
\psi(t) &= \frac{1}{2} \int_{-\infty}^{\infty} e^{itx} e^{-|x|} dx = \frac{1}{2} \int_{-\infty}^{0} e^{itx} e^{x} dx + \frac{1}{2} \int_0^\infty e^{itx} e^{-x} dx \\
&= \frac{1}{2} \int_{-\infty}^{0} e^{x(it+1)} dx + \frac{1}{2} \int_0^\infty e^{x(it-1)} dx = \left. \frac{e^{x(it+1)}}{2(it+1)} \right|_{x=-\infty}^{0} + \left. \frac{e^{x(it-1)}}{2(it-1)} \right|_{x=0}^{\infty} \\
&= \frac{1}{2} \left(\frac{1}{1-it} + \frac{1}{1+it} \right) = \frac{1}{2} \left(\frac{2}{1-(it)^2} \right) = \frac{1}{1+t^2}
\end{aligned}$$

■

8. The Cauchy distribution:

$$\text{p.d.f.} \quad f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

$$\text{c.f.} \quad \psi(t) = \exp(-|t|)$$

Proof. Apply Theorem 3.10.10 below to the double exponential distribution above to get

$$\frac{1}{2} e^{-|x|} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1+t^2} dt$$

That is,

$$e^{-|x|} = \int_{-\infty}^{\infty} e^{-itx} \frac{1}{\pi} \frac{1}{1+t^2} dt = \int_{-\infty}^{\infty} e^{it(-x)} f(t) dt$$

■

9. Binomial distribution. $X, X_1, \ldots, X_n \sim_{\text{iid}}$ Bernoulli(p)(coin flips), i.e., $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$, and $S_n = X_1 + \ldots + X_n \sim \text{Bin}(n, p)$. Then

$$\psi_X(t) = Ee^{itX} = pe^{it} + q$$

and

$$\psi_{S_n}(t) = \psi^n(t) = \left(pe^{it} + q\right)^n$$

10. Poisson (λ) distribution.

p.m.f.	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \geq 0$
c.f.	$\psi(t) = \exp\left\{\lambda\left(e^{it} - 1\right)\right\}$

Proof.

$$\psi(t) = Ee^{itX} = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda\left[e^{it} - 1\right]}$$

■

11. Centered Poisson distribution. $X_t \sim \text{Poisson}(\lambda)$, and $\tilde{X}_t = X_t - \lambda t$, whose c.f. is

$$\psi(t) = Ee^{itX} e^{-i\lambda t} = e^{\lambda\left[e^{it} - 1 - it\right]}$$

3.10.2 Definition and some properties of c.f.s

Proposition 3.10.1 Some elementary properties of c.f.s

1. $\psi(0) = 1$ and $|\psi(t)| = \left|Ee^{itX}\right| \leq E\left|e^{itX}\right| = 1$ for all t
2. $\psi(t)$ is uniformly continuous in $t \in (-\infty, \infty)$

   Proof. For any real t and $h \to 0$,

   $$|\psi(t+h) - \psi(t)| = \left|Ee^{i(t+h)X} - Ee^{itX}\right| = \left|E\left[e^{itX}\left(e^{ihX} - 1\right)\right]\right| \leq E\left|e^{ihX} - 1\right| \to 0$$

   ■

3. $\psi_{aX+b}(t) = Ee^{it(aX+b)} = e^{itb} Ee^{i\,\mathrm{ta}\,X} = e^{itb}\psi_X(at)$
4. $\psi_{-X}(t) = \psi_X(-t) = \overline{\psi_X(t)}$, where $\bar{z}$ denotes the complex conjugate of z
5. $\psi_X(t)$ is real iff X is symmetric about zero.

   Proof. If X is symmetric about zero, then $X =_d -X$, which implies that $\psi_X(t) = \psi_{-X}(t) = \psi_X(-t) = \overline{\psi_X(t)}$. Then $\psi_X(t)$ is real. The above argument can be reversed, but in one of the steps we need the following fact: $\psi_X(t) = \psi_Y(t)$ implies $X =_d Y$, which will be proved later. ■

6. If X and Y are independent r.v.'s, then

   $$\psi_{X+Y}(t) = Ee^{it(X+Y)} = Ee^{itX} Ee^{itY} = \psi_X(t)\psi_Y(t)$$

In particular, if $\psi(t)$ is a c.f., so is $\psi^m(t)$, where m is a positive integer.

7. Let $F_1, \ldots, F_n$ are d.f.'s with c.f. $\psi_1, \ldots, \psi_n$. 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 $\sum_{i=1}^n \lambda_i \psi_i$
8. If $\psi(t)$ is a c.f., so are $|\psi(t)|^2$ and $\operatorname{Re}\psi$. In particular, $|\psi(t)|^2$ is the c.f. of a symmetric r.v. $X-Y$, where X, Y are i.i.d. with c.f. ψ;$\operatorname{Re}\psi$ is the c.f. of the d.f. $(F_X(x) + F_{-X}(x))/2$

Proof. (a) First, let X, Y be i.i.d. with c.f. $\psi(t)$, then

$$\psi_{X-Y}(t) = \psi_X(t)\psi_Y(-t) = \psi(t)\bar{\psi}(t) = |\psi(t)|^2$$

(b) Secondly, $-X$ has c.f. $\overline{\psi(t)}$. So the d.f. $(F_X(x) + F_{-X})/2$ has c.f.

$$\frac{1}{2}\left(\psi_X(t) + \overline{\psi_X(t)}\right) = \operatorname{Re}\psi_X(t)$$

■

9. If $\psi(t)$ is a c.f., however, $|\psi|$ may not necessarily be a c.f.

Proof. Let X_k be i.i.d. $\text{Bin}(1,1/3)$, i.e. $P(X_k = 1) = 1/3, P(X_k = 0) = 2/3$ so that its c.f. is $\phi(t) = \frac{2}{3} + \frac{1}{3}e^{it}$. Suppose Y_j are i.i.d. r.v.s with c.f. $\psi(t) = |\phi(t)|$. Then $\psi^2(t) = |\phi(t)|^2$, which means that $Y_1 + Y_2 =_d X_1 - X_2$. Since $X_k \in \{0,1\}$, so $X_1 - X_2 \in \{-1,0,1\}$, and therefore, $Y_j \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}$. Write $\alpha = P\left(Y_j = 1/2\right)$. Then,

$$\alpha^2 = P\left(Y_1 = \frac{1}{2}\right)P\left(Y_2 = \frac{1}{2}\right) = P(Y_1 + Y_2 = 1) = P(X_1 - X_2 = 1) = \frac{2}{9}$$
$$(1-\alpha)^2 = P\left(Y_1 = -\frac{1}{2}\right)P\left(Y_2 = -\frac{1}{2}\right) = P(Y_1 + Y_2 = -1) = P(X_1 - X_2 = -1) = \frac{2}{9}$$

implying $\alpha^2 = (1-\alpha)^2$ so that $\alpha = 1/2$, contradicting with the fact that $\alpha^2 = 2/9$. Thus, no such r.v. Y exists. ■

R We often need to study the properties of $|\psi(t)|$ near the origin. However, as has been just seen, $|\psi(t)|$ may not be a c.f., hence we may not be able to use many nice properties of c.f.s. To overcome this problem, one often studies $|\psi(t)|^2$ first, which is indeed a c.f.

10. If $|\psi(t)| \equiv 1$ for all t, then $\psi(t) = e^{ibt}$, that is, X is degenerate at b

Proof. Let X, Y be i.i.d. r.v.s with c.f. $\psi(t)$, and denote $Z = X - Y$. From the properties of c.f.s, we have $\psi_Z(t) = |\psi(t)|^2 = 1^2 = 1$ for all t. Since 1 is the c.f. of a degenerate r.v. at 0, by the one-to-one correspondence between c.f. and d.f. (to be studied later), we have $P(Z = 0) = 1$ for some constant c. So $0 = \operatorname{Var}(Z) = 2\operatorname{Var}(X)$. Then $P(X = b) = 1$ for some constant b. Hence, $\psi_X(t) = e^{ibt}$ ■

> **R** When trying to show that $Z=0$ a.s., instead of using the one-to-one correspondence theorem as done above, one could also use Theorem 3.10.8 to calculate it directly as follows:
>
> $$\begin{aligned} P(\{a\}) &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\psi(t)dt = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}dt \\ &= \lim_{T\to\infty} 1 = 1, \quad \text{if } a=0 \\ &\lim_{T\to\infty}\frac{\sin Ta}{Ta} = 0, \quad \text{if } a\neq 0 \end{aligned}$$

11. A c.f could be nowhere differentiable. For instance, take $P\left(X=5^k\right)=1/2^{k+1}, k=0,1,\ldots$, then $\psi(t)=\sum_{k=0}^{\infty}\exp\left\{it5^k\right\}/2^{k+1}$ is nowhere differentiable.

Relationship between local and global properties of c.f.

The following theorem is useful in extending local properties of c.f. around the origin to global ones, and visa versa.

Theorem 3.10.2

$$\operatorname{Re}(1-\psi(t)) \geq \frac{1}{4}\operatorname{Re}(1-\psi(2t)) \geq \ldots \geq \frac{1}{4^n}\operatorname{Re}\left(1-\psi\left(2^n t\right)\right) \tag{3.46}$$

In particular, we have

1. $1-|\psi(t)|^2 \geq \frac{1}{4}\left(1-|\psi(2t)|^2\right) \geq \ldots \geq \frac{1}{4^n}\left(1-\left|\psi\left(2^n t\right)\right|^2\right)$
2. $(1-|\psi(t)|) \geq \frac{1}{8}(1-|\psi(2t)|) \geq \frac{1}{8^n}\left(1-\left|\psi\left(2^n t\right)\right|\right)$

Proof. We prove (3.46) first, which follows by taking expectation to

$$1-\cos 2tX = 2(\sin(tX))^2 = 2(2\sin(tX/2)\cos(tX/2))^2 \leq 8\sin^2(tX/2) = 4(1-\cos tX)$$

1. Applying the above to c.f. $|\psi(t)|^2$.
2. Noting $0 \leq |\psi| \leq 1$, we get

$$\begin{aligned} 1-|\psi(2t)| &\leq (1-|\psi(2t)|)(1+|\psi(2t)|) \leq 1-|\psi(2t)|^2 \\ &\leq 4\left(1-|\psi(t)|^2\right) \leq 4(1-|\psi(t)|)(1+|\psi(t)|) \leq 8(1-|\psi(t)|) \end{aligned}$$

■

Corollary 3.10.3 Suppose that $|\psi(t)| \leq a < 1$ for $|t| \geq b > 0$. Then

$$|\psi(t)| \leq 1-ct^2 \leq e^{-ct^2}, \quad \text{for } |t|<b, \text{ where } c=\frac{1-a^2}{8b^2}$$

Proof. Here we know the behavior of $|\psi(t)|$ away from 0. We would like to find the behavior near 0. We need to show $1-|\psi(t)| \geq ct^2$. (The second inequality follows from

$1+x \le e^x$ for all x.) Since $|t| \ge b$, there exists an m such that $b/2^m \le |t| < 2b/2^m$, i.e., $b \le |2^m t| < 2b$. Thus,

$$1-|\psi(t)|^2 \ge \frac{1}{4^m}\left(1-|\psi(2^m t)|^2\right) \ge \left(\frac{1}{2^m}\right)^2 (1-a^2) \ge \left(\frac{t}{2b}\right)^2 (1-a^2) = 2ct^2$$

In view of the inequality $(1-x)^{1/2} \le 1-x/2$ for $|x| \le 1$, we have

$$|\psi(t)| = \left(|\psi(t)|^2\right)^{1/2} \le \left(1-2ct^2\right)^{1/2} \le 1-ct^2$$

■

Corollary 3.10.4 Under Cramer's condition: $\limsup_{|t|\to\infty} |\psi(t)| < 1$, then for any $\delta > 0$, there exists $d \in (0,1)$ such that

$$|\psi(t)| \le d, \quad \text{for } |t| \ge \delta$$

Proof. Cramer's condition implies that there exists some $a < 1$ and $b > 0$ such that $|\psi(t)| \le a < 1$ for $|t| \ge b > 0$. Then from the last corollary, we have $|\psi(t)| \le 1-ct^2$ for $|t| < b$

Now for $b > |t| \ge \delta > 0$ we have $|\psi(t)| \le 1-ct^2 \le 1-c\delta^2$. The proof follows by choosing $c = \max\left\{1-c\delta^2, a\right\}$ ■

Theorem 3.10.5 Let X be a nondegenerate r.v. with c.f ψ. There exist $\delta > 0$ and $\epsilon > 0$ such that

$$|\psi(t)| \le 1-\epsilon t^2 \quad \text{for } |t| \le \delta$$

Proof. Let Y be an independent copy of X, and $Z = X - Y$. Then,

$$\begin{aligned} 1-|\psi(t)| &\ge (1-|\psi(t)|)\frac{1+|\psi(t)|}{2} = \frac{1}{2}(1-|\psi(t)|^2) \\ &= \frac{1}{2}(1-\psi_Z(t)) = \frac{1}{2}E(1-\cos tZ) \end{aligned}$$

By Taylor expansion, for $|t| < 1$,

$$1-\cos t = 1-\left(1-\frac{t^2}{2!}+\frac{t^4}{4!}-\dots\right) = \frac{t^2}{2!}-\frac{t^4}{4!}+\dots \ge \frac{t^2}{2!}-\frac{t^4}{4!} = \frac{t^2}{2}\left(1-\frac{t^2}{12}\right)$$

Thus,

$$\begin{aligned} E[1-\cos tZ] &\ge \frac{t^2}{2}E\left(Z^2\left(1-\frac{t^2Z^2}{12}\right)I\{|tZ|<1\}\right) \\ &\ge \frac{t^2}{2}\left(1-\frac{1}{12}\right)E\left(Z^2 I\{|Z|<1/|t|\}\right) \end{aligned}$$

Since X is nondegenerate, so is Z. Therefore, when $|t|$ is small enough, say, $|t| \le \delta$, we have

$$E\left(Z^2 I\{|Z|<1/|t|\}\right) \ge E\left(Z^2 I\{|Z|<1/\delta\}\right) > 0$$

Combining all the above, we have, as $|t| \le \delta$,

$$1 - |\psi(t)| \ge \frac{1}{2}E(1-\cos tZ) \ge \left(\frac{11}{48}E\left(Z^2 I\{|Z| < 1/\delta\}\right)\right)t^2 = \epsilon t^2$$

■

Theorem 3.10.6 For any $t, h \in R$, we have

$$|\psi(t+h) - \psi(t)|^2 \le 2(1 - Re\psi(h)) = 2E[1-\cos(hX)]$$

Proof.

$$\begin{aligned}
|\psi(t+h) - \psi(t)|^2 &= \left|Ee^{i(t+h)X} - Ee^{itX}\right|^2 = \left|E\left[e^{itX}\left(e^{ihX}-1\right)\right]\right|^2 \\
&\le E\left|e^{ihX}-1\right|^2 = E\left[\left(e^{ihX}-1\right)\left(\overline{e^{ihX}-1}\right)\right] \\
&= E\left[\left(e^{ihX}-1\right)\left(e^{-ihX}-1\right)\right] \\
&= E\left(e^{ihX}e^{-ihX} - e^{ihX} - e^{-ihX} + 1\right) \\
&= 2E[1-\cos(hX)]
\end{aligned}$$

■

As an application, we give the following example. Compare this with Levy Continuity Theorem.

■ **Example 3.16** 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 R

Proof. We have

$$\begin{aligned}
|g(t+h) - g(t)|^2 &= \lim_{n\to\infty} |\psi_n(t+h) - \psi_n(t)|^2 \\
&\le 2\lim_{n\to\infty} [1 - \mathrm{Re}\,\psi_n(h)] \quad \text{(by Theorem 3.10.6)} \\
&= 2[1 - \mathrm{Re}\,g(h)] \\
&\to 2[1 - \mathrm{Re}\,g(0)] = 0, \quad \text{as } h \to 0
\end{aligned}$$

■

■ **Example 3.17** 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 $t \in R$ (hence, $X_n \Longrightarrow 0$.)

Proof. By Theorem 3.10.6

$$||\psi_n(2t)| - |\psi_n(t)||^2 \le |\psi_n(2t) - \psi_n(t)|^2 \le 2(1 - \mathrm{Re}\,\psi_n(t))$$

Letting $n \to \infty$ and in view of $\lim_{n\to\infty} |\psi_n(t)| = 1$ for $|t| \le \delta$, we have $\lim_{n\to\infty} |\psi_n(2t)| = 1$ for $|t| \le \delta$. That is, $|\psi_n(t)| \to 1$ for $|t| \le 2\delta$. Continuing on, we see that $|\psi_n(t)| \to 1$ for all $t \in R$. ■

Proof. Alternative proof. One can prove this by using Theorem 3.46, which will extend the local properties to global ones. We will leave the details here. In the following, we shall instead use Theorem 3.10.6 to prove the statement. ∎

3.10.3 Inversion formula

We will prove a very important fact: the c.f. uniquely determines the distribution.

Theorem 3.10.7 — The inversion formula. Let $\psi(t) = \int e^{itx}\mu(dx)$, where μ is a probability measure. If $a < b$, then

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

provided that the limit on the left hand side exists.

Proof. Let

$$\begin{aligned}
I(T) =: &\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \psi(t) dt \\
&= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \left(\int_{-\infty}^{\infty} e^{itx}\mu(dx) \right) dt \\
&= \frac{1}{2\pi} \int_{-T}^{T} \left(\int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\mu(dx) \right) dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt \right) \mu(dx) \\
&\left(\text{by Fubini's theorem since} \left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| = \left| \int_a^b e^{-itx} dx \right| \left| e^{itx} \right| \le b - a \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \right) \mu(dx) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{1}{it} (\cos[t(x-a)] - \cos[t(x-b)]) dt \right) \mu(dx) \\
&+ \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{i}{it} (\sin[t(x-a)] - \sin[t(x-b)]) dt \right) \mu(dx) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} \frac{1}{t} (\sin[t(x-a)] - \sin[t(x-b)]) dt \right) \mu(dx) \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\int_{0}^{T} \frac{1}{t} (\sin[t(x-a)] - \sin[t(x-b)]) dt \right) \mu(dx) \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} [I(x-a,T) - I(x-b,T)] \mu(dx)
\end{aligned}$$

where $I(a,T)$ is defined in Lemma 3.19 below. Again from Lemma 3.19 we get

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{\pi}I(x-a,T)-I(x-b,T) &= \left(-\frac{1}{2}\right)-\left(-\frac{1}{2}\right)=0, \quad x<a\\
&=0-\left(-\frac{1}{2}\right)=\frac{1}{2}, \quad x=a\\
&=\frac{1}{2}-\left(-\frac{1}{2}\right)=1, \quad a<x<b\\
&=\frac{1}{2}-0=\frac{1}{2}, \quad x=b\\
&=\left(\frac{1}{2}\right)-\left(\frac{1}{2}\right)=0, \quad x>b
\end{aligned}$$

Note that from (3.47), for every θ, we have

$$|I(\theta,T)|\leq\int_0^{\theta T}\frac{\sin v}{v}dv\leq\int_0^{\infty}\frac{\sin v}{v}dv=I(1,\infty)=\frac{\pi}{2}$$

Using the dominated (or bounded) convergence theorem, we have

$$\begin{aligned}
\lim_{T\to\infty}I(T) &= \int_{-\infty}^{\infty}\frac{1}{\pi}\lim_{T\to\infty}[I(x-a,T)-I(x-b,T)]\mu(dx)\\
&=\int_{(-\infty,a)\cup(b,\infty)}0\mu(dx)+\int_{(a,b)}1\mu(dx)+\int_a\frac{1}{2}\mu(dx)+\int_b\frac{1}{2}\mu(dx)\\
&=\mu(a,b)+\frac{1}{2}\mu(\{a,b\})
\end{aligned}$$

■

Lemma 3.19 We have

$$\lim_{T\to\infty}\int_0^T\frac{\sin au}{u}du=\frac{\pi}{2}\operatorname{sgn}\{a\}$$

Proof. Write

$$I(a,T)=:\int_0^T\frac{\sin au}{u}du=\int_0^{aT}\frac{\sin v}{v}dv \tag{3.47}$$

Then, we have

$$\lim_{T\to\infty}I(a,T)=\operatorname{sgn}\{a\}\lim_{T\to\infty}I(1,T)$$

This can be seen as follows:

1. If $a>0$, we have $\lim_{T\to\infty}I(a,T)=\lim_{T\to\infty}\int_0^{aT}\frac{\sin v}{v}dv=\lim_{T\to\infty}I(1,T)$
2. If $a<0$, we have

$$\lim_{T\to\infty}I(a,T)=\lim_{T\to\infty}(-1)\int_0^{aT}\frac{\sin(-v)}{-v}d(-v)=-\lim_{T\to\infty}\int_0^{-aT}\frac{\sin(w)}{w}dw=-\lim_{T\to\infty}I(1,T)$$

3. If $a = 0$, we have $\lim_{T\to\infty} I(a,T) = 0 = \text{sgn}\{0\} \lim_{T\to\infty} I(1,T)$

   Then it suffices to show that

$$\lim_{T\to\infty} I(1,T) = \lim_{T\to\infty} \int_0^T \frac{\sin u}{u} du = \frac{\pi}{2}$$

To show this, note that

$$\begin{aligned}
I(1,T) &= \int_0^T \frac{\sin u}{u} du = \int_0^T \sin u \left(\int_0^\infty e^{-uv} dv \right) du \\
&= \int_0^\infty \left(\int_0^T \sin(u) e^{-uv} dv \right) dv \\
&= \int_0^\infty \left(\frac{1}{1+v^2} - \frac{v \sin T + \cos T}{1+v^2} e^{-vT} \right) dv \\
&\quad \text{(this can be checked by differentiation)} \\
&= \frac{\pi}{2} - \int_0^\infty \frac{s \sin T + T \cos T}{T^2 + s^2} e^{-s} ds
\end{aligned}$$

The integrand in the last integral is dominated by e^{-s}, and so by Lebesgue Dominated Convergence Theorem, the integral tends to 0 as $T \to \infty$. ■

Theorem 3.10.8

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

Proof. Here are the main steps involved.

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T e^{-ita} \psi(t) dt &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T \int_\infty^\infty e^{it(y-a)} d\mu(y) dt \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^\infty \int_{-T}^T e^{it(y-a)} dt d\mu(y) \\
&= \lim_{T\to\infty} \int_{-\infty}^\infty \frac{\sin[(y-a)T]}{(y-a)T} d\mu(y) \\
&= \lim_{T\to\infty} \int_{(-\infty,a)} + \int_{\{a\}} + \int_{(a,\infty)} \frac{\sin[(y-a)T]}{(y-a)T} d\mu(y) \\
&= 0 + \mu(\{a\}) + 0 \\
&= \mu(\{a\})
\end{aligned}$$

■

Theorem 3.10.9 — One-to-one correspondence between d.f. and c.f.–Uniqueness. Characteristic functions uniquely determines distribution functions. That is, there is a one-one correspondence between c.f.s and d.f.s.

Just using definition. Suppose that two d.f.s F_1 and F_2 have the same c.f. $\psi(t)$, we need to show that $F_1 \equiv F_2$. Let $a, b \in C(F_1) \cap C(F_2)$ with $a < b$. From the inversion formula,

$$F_1(b) - F_1(a) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \psi(t) dt = F_2(b) - F_2(a)$$

Let $a \to -\infty$ along $C(F_1) \cap C(F_2)$, we get $F_1(b) = F_2(b)$ for all $b \in C(F_1) \cap C(F_2)$. By the right continuity of d.f., we have $F_1(b) = F_2(b)$ for all $b \in R$. ∎

The case when ψ is integrable

It follows from the last theorem that if $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$, we have $\mu(\{a\}) = 0$ for all a. That is, μ is a continuous measure. In fact, we can get a stronger result than this.

Theorem 3.10.10 — Density induced by c.f.. If $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$, then μ has bounded continuous density

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

Proof. Since $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$, it follows from Theorem 3.10.8 that $\mu(\{a, b\}) = 0$ for all $a < b$. Take $a = y$ and $b = y + h$ in Theorem 3.10.7 with $h > 0$, we have

$$\frac{\mu(y, y+h)}{h} = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t) dt$$

as

$$\int_{-\infty}^{\infty} \left| \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t) \right| dt = \int_{-\infty}^{\infty} \left| \frac{1 - e^{-ith}}{ith} \psi(t) \right| dt \le \int_{-\infty}^{\infty} |\psi(t)| dt < \infty$$

Then

$$\begin{aligned}
\lim_{h \searrow 0} \frac{\mu(y, y+h)}{h} &= \lim_{h \searrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t) dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h \searrow 0} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t) dt \\
&\quad \text{(dominated convergence theorem)} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h \searrow 0} \frac{ite^{-it(y+h)}}{it} \psi(t) dt \quad \text{(L'Hospital rule)} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \psi(t) dt
\end{aligned}$$

∎

R When $\int_{-\infty}^{\infty}|\psi(t)|dt < \infty$, we have the following interesting relationships:

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

Therefore, if we know one relation, we know the other one straightaway. One can use these relationships to find some c.f.s easily.

R The condition $\int_{-\infty}^{\infty}|\psi(t)|dt < \infty$ is only a sufficient condition for X to have a p.d.f. There are examples where $\int_{-\infty}^{\infty}|\psi(t)|dt = \infty$,but a p.d.f. still exists.

The case when ψ is not integrable

Theorem 3.10.11 If $P(X \in b + h\mathbf{Z}) = 1$, where $\mathbf{Z} = \{0, \pm 1, \pm 2, \ldots.\}$, then for $x \in b + h\mathbf{Z}$, we have

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

Proof. Assume that $x = b + jh$ for some $j \in \mathbf{Z}$. Denote $p_k = P(X = b + kh)$ for $k \in \mathbf{Z}$. Then,

$$\begin{aligned}
\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\psi(t)dt &= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-it(b+jh)}\sum_{k=-\infty}^{\infty} e^{it(b+kh)}p_k dt \\
&= \frac{h}{2\pi}\sum_{k=-\infty}^{\infty} p_k \int_{-\pi/h}^{\pi/h} e^{it(k-j)h}dt \\
&= \frac{h}{2\pi}\left(\sum_{k\neq j} + \sum_{k=j}\right) p_k \int_{-\pi/h}^{\pi/h} e^{it(k-j)h}dt \\
&= \frac{h}{2\pi}\left(0 + p_j\int_{-\pi/h}^{\pi/h} 1dt\right) \quad (e^{i\pi} = -1) \\
&= p_j = P(X = b + jh) \\
&= P(X = x)
\end{aligned}$$

■

3.10.4 Levy Continuity Theorem

Instead of studying d.f.s directly, we could study their corresponding c.f.s. This can be done due to Levy continuity theorem.

Lemma 3.20 For any a > 0, we have

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

Proof.

$$\begin{aligned}
\int_{-a}^{a}(1-\psi(t))dt &= 2a - \int_{-a}^{a} Ee^{itX}dt \\
&= 2a - E\left(\int_{-a}^{a} e^{itX}dt\right) \quad (|e^{itX}| < 1 \quad \text{by Fubin theorem }) \\
&= 2a - E\left(\int_{-a}^{a} \cos(tX)dt\right) \quad \text{by symmetry} \\
&= 2a - E\left(\frac{2\sin(aX)}{X}\right) \\
&= 2aE\left(1 - \frac{\sin(aX)}{aX}\right) \\
&\ge 2aE\left(1 - \frac{\sin(aX)}{aX}\right) I\{|aX| > 2\} \\
&\ge 2aE\left(1 - \frac{1}{2}\right) I\{|aX| > 2\} \\
&\ge aP(|aX| > 2)
\end{aligned}$$

■

R When a is chosen to be very small, Lemma 3.20 shows that the tail probability behavior of a r.v. X is actually determined by the behavior of its c.f. around the origin.

Lemma 3.21 — Test tight by d.f.. 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.

Proof. First note that $g(0) = \lim_n \psi_n(0) = 1$ and $g(t)$ is continuous at 0. Therefore, $\forall \varepsilon > 0, \exists a_0 > 0$ such that $|1 - g(t)| = |g(t) - g(0)| < \varepsilon/2$ whenever $|t| < a_0$. Therefore, as $n \to \infty$, we have

$$\begin{aligned}
P\left(|X_n| > \frac{2}{a_0}\right) &\le \left|\frac{1}{a_0}\int_{-a_0}^{a_0}(1-\psi_n(t))\,dt\right| \quad \text{Lemma3.20} \\
&\longrightarrow \left|\frac{1}{a_0}\int_{-a_0}^{a_0}(1-g(t))dt\right| \quad \text{(dominated convergence theorem)} \\
&\le \frac{1}{a_0}\int_{-a_0}^{a_0}|1-g(t)|dt \le \frac{1}{a_0}\int_{-a_0}^{a_0}\varepsilon/2dt \le \varepsilon
\end{aligned}$$

Thus, $\exists N_0 > 0$ such that, for all $n > N_0$, one has $P\left(|X_n| > \frac{2}{a_0}\right) \le \varepsilon$ ■

Theorem 3.10.12 — Levy continuity theorem. Assume that X_n has d.f. F_n and c.f. ψ_n for $1 \le n \le \infty$

1. If $X_n \to_d X_\infty$, (i.e., $F_n \Longrightarrow F_\infty$), 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 $X_n \to_d X$ (i.e., $F_n \Longrightarrow F$), and ψ is the c.f. of X.

Proof. 1. The proof follows from the bounded convergence theorem.

2. Now suppose that $\psi_n(t) \to \psi(t)$, and that $\psi(t)$ is continuous at 0. From Lemma 3.21, F_n is tight. Now suppose that $F_{n_k} \Longrightarrow_v \tilde{F}$ for some subsequence n_k and some limit $\tilde{F}$. Since F_n is tight, we have $F_{n_k} \Longrightarrow \tilde{F}$, i.e., the limit $\tilde{F}$ is a d.f. From part (i) of the current theorem, we see that $\psi_{n_k}(t) \to \psi_{\tilde{F}}(t)$ for all t. On the other hand, from the assumption, we have $\psi_{n_k}(t) \to \psi(t)$ for all t. Therefore,

$$\psi_{\tilde{F}}(t) = \psi(t)$$

Clearly, $\psi(t)$ is a c.f. Suppose that its corresponding d.f. is F. By the uniqueness theorem, Theorem 3.10.9, we get $\tilde{F} = F$. This shows that F is the only possible weak limit of the F_n Therefore,

$$F_n \Longrightarrow F$$

■

The following corollary is immediate.

Corollary 3.10.13 — Levy continuity theorem. $X_n \to_d X$ if and only if $\psi_{X_n}(t) \to \psi_X(t)$ for all t

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

The smoothness of the c.f. $\psi(t)$ at $t = 0$ is closely related to how many moments that X possesses, and hence to the tail behavior of the d.f. of X.

Theorem 3.10.14 If $E|X|^n < \infty$, then $\psi^{(n)}(t)$ exists and is a uniformly continuous function given by

$$\psi^{(k)}(t) = i^k E\left(X^k e^{itX}\right) = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF(x), \quad k = 0, 1, 2, \ldots, n$$

In particular,

$$\psi^{(k)}(0) = i^k E X^k, \quad k = 0, 1, \ldots, n$$

Proof. Note that

$$\frac{\psi(t+h) - \psi(t)}{h} = \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF(x)$$

Using Lemma 3.22, the integrand is dominated by $|x|$.(using Taylor to second term) So the first derivative of $\psi(t)$ exists by the dominated convergence theorem, and is given by

$$\begin{aligned}\psi'(t) &= \lim_{h\to 0}\frac{\psi(t+h)-\psi(t)}{h} = \lim_{h\to 0}\int_{-\infty}^{\infty} e^{itx}\frac{e^{ihx}-1}{h}dF(x)\\ &= \int_{-\infty}^{\infty}\lim_{h\to 0} ix\frac{e^{ihx}-e^0}{ihx}dF(x) = i\int_{-\infty}^{\infty} xe^{itx}dF(x)\end{aligned}$$

The uniform continuity of $\psi'(t)$ follows from

$$|\psi'(t+\delta)-\psi'(t)| = \left|\int_{-\infty}^{\infty} xe^{itx}\left(e^{it\delta}-1\right)dF(x)\right| \to 0$$

by the dominated convergence theorem. Therefore, the assertion is true for $n=1$. The general case follows by induction. ∎

A partial converse is given by the following theorem.

Theorem 3.10.15 If $\psi^{(n)}(0)$ exists and is finite for some $n = 1,2,\ldots$, then $E|X|^n < \infty$ if n is even. (Consequently, $E|X|^{n-1} < \infty$ if n is odd.)

Proof. We proceed by induction. Note that

$$\begin{aligned}-\psi^{(2)}(0) &= -\lim_{h\to 0}\frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2}\\ &= -\lim_{h\to 0} E\frac{\cos(hX)-i\sin(hx)+\cos(-hx)+i\sin(hx)-2}{h^2}\\ &= 2\lim_{h\to 0} E\frac{1-\cos(hX)}{h^2}\\ &\geq 2E\lim_{h\to 0}\frac{1-\cos(hX)}{h^2} \qquad \text{(by Fatou's lemma)}\\ &= \quad EX^2\end{aligned} \tag{3.48}$$

Where the first equality is for

$$\lim_{h\to 0}\frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2} = \lim_{h\to 0} = \frac{\psi'(h)-\psi'(-h)}{2h} = \lim_{h\to 0}\frac{\psi''(h)+\psi''(-h)}{2}$$

Therefore, the finiteness of $\psi^{(2)}(0)$ implies $EX^2 < \infty$. Now assuming that $\psi^{(2j)}(0) < \infty$ implies $EX^{2j} < \infty (j \geq 1)$. From Theorem 3.10.14

$$\psi^{(2j)}(t) = i^{2j}\int_{-\infty}^{\infty} x^{2j}e^{itx}dF(x) = (-1)^j E\left(X^{2j}e^{itX}\right)$$

So if $\psi^{(2(j+1))}(0) < \infty$, then

$$\begin{aligned}
(-1)^{j+1}\psi^{(2(j+1))}(0) &= (-1)^{j+1}\lim_{h\to 0}\frac{\psi^{(2j)}(h)+\psi^{(2j)}(-h)-2\psi^{(2j)}(0)}{h^2} \\
&= 2(-1)^{2j+2}\lim_{h\to 0}\int_{-\infty}^{\infty}\frac{x^{2j}[1-\cos(hx)]}{h^2}dF(x) \\
&= 2\lim_{h\to 0}E\frac{X^{2j}[1-\cos(hX)]}{h^2} \\
&\geq 2E\lim_{h\to 0}\frac{X^{2j}[1-\cos(hX)]}{h^2} \qquad \text{(by Fatou's lemma)} \\
&= EX^{2(j+1)}
\end{aligned}$$

Therefore, $\psi^{(2(j+1))}(0) < \infty$ implies $EX^{2(j+1)} < \infty$ ■

R If $g^{(2)}(0)$ exists, then applying L 'Hospital Rule, we can see that $\lim_{h\to 0}\frac{g(h)+g(-h)-2g(0)}{h^2}$ exists and equals $g^{(2)}(0)$ In fact,

$$\lim_{h\to 0}\frac{g(h)+g(-h)-2g(0)}{h^2} = \lim_{h\to 0}\frac{g'(h)-g'(-h)}{2h} = g^{(2)}(0)$$

However, the reverse is not true in general. For instance, take

$$g(t) = a_0 + a_1 t + \frac{1}{2}a_2 t^2 + D(t)t^3, \quad \text{as } t \to 0$$

where $D(t)$ takes values 0 or 1, depending on whether t is rational or irrational. Clearly, we do have

$$\lim_{h\to 0}\frac{g(h)+g(-h)-2g(0)}{h^2} = \lim_{h\to 0}\frac{a_2h^2 + D(h)h^3 - D(-h)h^3}{h^2} = a_2$$

However, for any fixed $t = t_0, g(t)$ is not continuous at t_0, let alone differentiable at t_0

R However, if $g(t) = \psi(t)$ is a c.f., then

$$\psi^{(2)}(0) \text{ exists } \iff \lim_{h\to 0}\frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2} \text{ exists,}$$

and furthermore they have the same limit. We have seen that the former implies the later. To see that the later implies the former, we see from (3.48) in the proof of Theorem **??** above that $EX^2 < \infty$. Applying Theorem 3.10.14 again to obtain $\psi^{(2)}(0)$ exists.

R From Theorem 3.10.15, $\psi^{(n)}(0) < \infty$ implies $E|X|^n < \infty$ when n is even. However, this may not be true when n is odd, as illustrated by the next example.

■ **Example 3.18** Define $P(X = \pm n) = C/\left(2n^2\log n\right), n = 2,3,\ldots$. Show that $\psi'(0)$ exists, but $E|X| = \infty$

Proof. The c.f. of X is

$$\psi(t) = Ee^{itX} = \sum_{n\geq 2} P(X = \pm n)e^{i(\pm n)t} = C\sum_{n=2}^{\infty} \frac{\cos(nt)}{n^2 \log n}$$

The series is uniformly convergent, and $\psi'(t)$ exists and

$$\psi'(t) = -C\sum_{n=2}^{\infty} \frac{\sin(nt)}{n \log n}$$

and hence $\psi'(0) = 0$. However,

$$E|X| = C\sum_{n=2}^{\infty} \frac{1}{n \log n} = \infty$$

■

Moments of a r.v. and Taylor expansion of its c.f.

The following theorem establishes the link between the existence of moments of a r.v. and Taylor expansion of its c.f. around the origin. The first part of the theorem is particularly useful in studying the limit theorems by the c.f. approach.

Theorem 3.10.16 1. If $E|X|^{n+\delta} < \infty$ for some nonnegative integer n and some $\delta \in [0,1]$, then the c.f. has Taylor expansion

$$\psi(t) = \sum_{k=0}^{n} \left(EX^k\right) \frac{(it)^k}{k!} + \theta \frac{2E|X|^{n+\delta}|t|^{n+\delta}}{n!} \text{ with } |\theta| \leq 1$$
$$\psi(t) = \sum_{k=0}^{n} a_k \frac{(it)^k}{k!} + o\left(t^n\right), \quad \text{as } t \to 0$$

where $a_k = EX^k$ for $k = 0,1,2,\ldots,n$

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

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

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

R Suppose that

$$g(t) = \sum_{k=0}^{2} a_k \frac{(it)^k}{k!} + o\left(t^2\right) = a_0 + a_1 t + \frac{1}{2}a_2 t^2 + o\left(t^2\right), \quad \text{as } t \to 0 \tag{3.49}$$

we can easily show that

$$g(0) = \lim_{t\to 0} g(t) = a_0$$
$$g'(0) = \lim_{t\to 0} \frac{g(t) - g(0)}{t} = \lim_{t\to 0} \frac{a_1 t + \frac{1}{2}a_2 t^2 + o\left(t^2\right)}{t}$$
$$= \lim_{t\to 0} \left(a_1 + \frac{1}{2}a_2 t + o(t)\right) = a_1$$

However, in order to find $g''(0)$, we need to establish that $g'(t)$ exist in a neighborhood of 0. But this can not be derived just from the expansion (3.49), as can be easily seen from the same example given in Remark 3.10.5.

However, if we let $g(t) = \psi(t)$ be a c.f., then we could show that $\psi''(0)$ exist iff $\lim_{h\to 0} \frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2}$ exists. The latter condition is often easier to check in practice. We will discuss it in detail in the proof of the theorem.

Proof. Proof of Theorem 3.10.16

1. The first equation in (1) follows directly from Lemma 3.23. Let us show the second one. By Taylor expansion and Theorem 3.10.14,

$$\psi(t) = \sum_{k=0}^{n-1} \frac{\psi^{(k)}(0)}{k!} t^k + \frac{\psi^{(n)}(\theta t)}{n!} t^n = \sum_{k=0}^{n} \left(EX^k\right) \frac{(it)^k}{k!} + R_n(t)$$

   where $R_n(t) = t^n \left[\psi^{(n)}(\theta t) - \psi^{(n)}(0)\right] / n!$ with $\theta \in [0,1]$. Note that

$$\frac{R_n(t)}{t^n} = \int_{-\infty}^{\infty} \frac{(ix)^n}{n!} \left(e^{i\theta tx} - 1\right) dF(x)$$

   The integrand is dominated by $2|x|^n/n!$. By the dominated convergence theorem, we have $R_n(t)/t^n \to 0$ as $t \to 0$

2. Assume that $n = 2m$. For simplicity, we shall only prove this for the special case $n = 2$ and leave the more general case as an exercise. Since

$$\psi(t) = a_0 + a_1(it) + \frac{a_2}{2}(it)^2 + o\left(t^2\right)$$

   we have

$$\psi(t) + \psi(-t) = 2\left(a_0 + \frac{a_2}{2}(it)^2 + o\left(t^2\right)\right)$$

   Thus,

$$-\lim_{h\to 0} \frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2} = \lim_{h\to 0}(a_2 + o(1)) = a_2$$

■

Application to Weak Law of Large Numbers

■ **Example 3.19 — Weak Law of Large Numbers.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with $EX_1 = 0$ and common c.f. $\psi(t)$. From Theorem 3.10.16

$$\psi(t) = 1 + i(EX)t + o(t) = 1 + o(t)$$

Therefore, the c.f. of $\overline{X}$ is

$$\psi_{\bar{X}}(t) = \psi^n(t/n) = \left(1 + o\left(n^{-1}\right)\right)^n \to 1$$

as $t \to 0$ by taking logarithms. Thus, $\overline{X} \to_d 0$, and so $\overline{X} \to_p 0$. This is the Weak Law of Large Numbers (WLLN).

We actually assumed the existence of EX above, under which we can in fact obtain the strong law of large numbers (SLLN) : $\overline{X} \to 0$ a.s. If we are content with WLLN, we can do without assuming the existence of the first moment; see Theorem 3.10.17 below.

In Theorem 3.10.14, we have seen that if $E|X| < \infty, \psi'(t)$ exists and $\psi'(0) = iEX$. The converse, however, is false. That is, that $\psi'(t)$ exists and $\psi'(0) = ia$ does not necessarily imply $E|X| < \infty$; see an earlier example given in this section. In fact, the differentiability of ψ is closely connected with the weak law of large numbers (WLLN) for i.i.d. r.v.'s.

Theorem 3.10.17 Let $X, X_1, X_2, \ldots$ be i.i.d. r.v.'s with d.f. F. The following three statements are equivalent:

1. $\psi'(0) = i\nu$
2. $t[1 - F(t) - F(-t)] \longrightarrow 0$ and $\int_{-t}^{t} xF\{dx\} \to \nu$ as $t \to \infty$ (or equivalently, $tP(|X| > t) \to 0$ and $E\left[XI_{\{|X| \le t\}}\right] \to \nu$.)
3. The $WLLN$ holds: $(X_1 + \ldots X_n)/n \to \nu$ in probability.

R It should be noted that ν is not necessarily the mean EX, which may not even exist here. For instance, from (ii) of Theorem 3.10.17, we can take X to be symmetric around 0 and

$$P(X > t) = \frac{1}{t \log t}, \quad \text{as } t \to \infty$$

The mean does not exist, but $\nu = 0$ and the WLLN holds here.

Lemma 3.22 For $n = 0, 1, 2, \ldots$ and any real t

$$\left| e^{it} - 1 - it - \frac{(it)^2}{2!} - \ldots - \frac{(it)^n}{n!} \right| \le \min\left\{ \frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right\}$$

Proof. By integration by parts, we have, for any $m \ge 0$,

$$\begin{aligned} \int_0^t (t-s)^m e^{is} ds &= \frac{-1}{m+1} \int_0^t e^{is} d(t-s)^{m+1} \\ &= \frac{t^{m+1}}{m+1} + \frac{i}{m+1} \int_0^t (t-s)^{m+1} e^{is} ds \end{aligned}$$

Therefore, by iteration we get

$$
\begin{aligned}
e^{it} &= 1+\left(e^{it}-1\right)=1+i\int_0^t e^{is}ds \\
&= 1+it+i^2\int_0^t (t-s)e^{is}ds \quad (t=1, s=0, m=0) \\
&= \cdots\cdots \\
&= 1+it+\frac{(it)^2}{2!}+\ldots+\frac{(it)^n}{n!}+\frac{i^{n+1}}{n!}\int_0^t (t-s)^n e^{is}ds
\end{aligned}
$$

Note that

$$
\left|\int_0^t (t-s)^n e^{is}ds\right| \le \int_0^{|t|} |t-s|^n ds \le \frac{|t|^{n+1}}{n+1}
$$

By integration by parts,

$$
\begin{aligned}
\int_0^t (t-s)^n e^{is}ds = (-i)\int_0^t (t-s)^n de^{is} \\
&= -it^n + in\int_0^t (t-s)^{n-1}e^{is}ds \\
&= in\int_0^t (t-s)^{n-1}\left[e^{is}-1\right]ds
\end{aligned}
$$

and hence

$$
\left|\int_0^t (t-s)^n e^{is}ds\right| \quad \le 2n\int_0^{|t|} |t-s|^{n-1}ds = 2|t|^n
$$

Using these relationships, we get

$$
\left|e^{it}-1-it-\frac{(it)^2}{2!}-\ldots-\frac{(it)^n}{n!}\right| \le \min\left\{\frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!}\right\}
$$

The proof is complete. ■

Lemma 3.23 For $n=0,1,2,\ldots$ and any real t

$$
\left|e^{it}-1-it-\frac{(it)^2}{2!}-\ldots-\frac{(it)^n}{n!}\right| \le \frac{2|t|^{n+\delta}}{n!}, \qquad \text{for every } \delta\in[0,1]
$$

Proof. If $|t|\le 1$, then $|t|^{n+1}/(n+1)! \le |t|^{n+\delta}/n!$. If $|t|\ge 1$, then $2|t|^n/n! \le 2|t|^{n+\delta}/n!$. The lemma follows from these and Lemma 3.22. ■

3.10.6 When is a function a c.f.?

Example 3.20 1. A non-constant function $\psi(t)$ such that $\psi''(0)=0$ can not be a c.f., since the corresponding d.f. would have $EX^2=-\psi''(0)=0$, implying that $P(X=b)=1$ for some constant b and thus $\psi(t)=e^{itb}$

2. $\psi(t) = e^{-|t|^\alpha}$ is not a c.f. for $\alpha > 2$ since $\psi'(t) = e^{-|t|^\alpha}\alpha|t|^\alpha$ and
$$\psi''(t) = e^{-|t|^\alpha}\left[\left(\alpha|t|^{\alpha-1}\right)^2 + \alpha(\alpha-1)|t|^{\alpha-2}\right)\right]$$
which gives $\psi''(0) = 0$. Therefore, we get $\psi(t) = e^{itb} \neq e^{-|t|^\alpha}$
3. On the other hand, for $\alpha \leq 2, \psi(t) = e^{-|t|^\alpha}$ is a c.f. of a stable d.f. For a proof, see Feller $(1971$, Vol. $2, p509)$, or Durrett $(1991, p87)$. In particular, $\alpha = 1$ corresponds to Cauchy d.f. and $\alpha = 2$ corresponds to Normal d.f.

3.10.7 Esseen's Smoothing Lemma

Often we are interested in difference of two functions. For instance, if a r.v. T_n has an asymptotic normal distribution, as in the central limit theorem, then
$$\sup_x |P(T_n \leq x) - \Phi(x)| \to 0 \quad \text{as } n \to \infty$$
The natural question is then how fast this limit goes to zero. In other words, we are interested in the rates of convergence to normality. One fundamental tool in studying the difference in two functions is the "smoothing lemma".

The word "smoothing" is derived from the fact that: any r.v. X perturbed by an independent continuous r.v. Y, will also be a continuous r.v.. That is, if X and Y are independent and Y is a continuous r.v., then $X+Y$ is a continuous r.v. for all X. Furthermore, the degree of smoothness for $X+Y$ also depends on the degree of smoothness for Y. This follows from the following identity: (making discrete r.v. be a smooth r.v. by convolution)
$$F_{X+Y}(t) = \int_{-\infty}^{\infty} F_Y(t-y)dF_X(y)$$
Let V_T be the d.f. with a p.d.f. (i.e. inverse triangular d.f.)(This r.v. is like to normal but it is bounded–Making $(-\infty,\infty)$ to be $(-T,T)$)
$$v_T(x) = \frac{1-\cos(Tx)}{\pi T x^2}$$
which is the p.d.f of the sum of two independent $U[-1/(2T), 1/(2T)]$ (try to plot it!) The corresponding c.f. is given by
$$\omega_T(t) = \left(1 - \frac{t}{T}\right) I\{|t| \leq T\}$$
The explicit form of $\omega_T(t)$ is of no importance. What matters is that $\omega_T(t)$ vanishes for $|t| \geq T$, since this eliminates all questions of convergence. For any function $\Delta(x)$, we denote its convolution with $V_T(x)$ by
$$\Delta^T(t) \equiv \Delta \star V_T(t) := \int_{-\infty}^{\infty} \Delta(t-x)v_T(x)dx$$
Our objective is to estimate the maximum of $|\Delta|$ in terms of the maximum of $|\Delta^T|$

Lemma 3.24 Let F be a d.f. and G a function such that $G(-\infty)=0, G(\infty)=1$, and $\sup_x |G'(x)| \leq \lambda < \infty . Put$

$$\Delta(x)=F(x)-G(x)$$

Then

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

Proof. Denote

$$\eta=\sup_x |\Delta(x)|, \quad \eta_T=\sup_x \left|\Delta^T(x)\right|, \quad h=\frac{\eta}{2\lambda}$$

Then it suffices to show that

$$\eta \leq 2\eta_T+\frac{24\lambda}{\pi T}$$

Since $\Delta(x) \rightarrow 0$ as $x \rightarrow \pm\infty$, and $\Delta(x)$ is right continuous, then it is clear that at some point x_0, either $|\Delta(x_0+)|=|\Delta(x_0)|=\eta$ or $|\Delta(x_0-)|=\eta$. Without loss of generality, we may assume that $\Delta(x_0)=\eta$ since $\Delta(x)$ attains its maximum at $x=x_0$, then we can imagine that $\Delta^T(x)$ should attain its maximum around $x=x_0$. Therefore,

$$\begin{aligned}
\eta_T =: \sup_x \left|\Delta^T(x)\right| &\geq \Delta^T(x_0+h)=\int_{-\infty}^{\infty} \Delta(x_0+h-x)\, v_T(x) dx \\
&=\left(\int_{|x| \leq h}+\int_{|x|>h}\right) \Delta(x_0+h-x)\, v_T(x) dx \\
&=\int_{|x| \leq h} \Delta(x_0+h-x)\, v_T(x) dx+\int_{|x|>h} \Delta(x_0+h-x)\, v_T(x) dx
\end{aligned} \tag{3.50}$$

We will get a lower bound for $\Delta(x_0+h-x)$ in the first and second integral.

1. In the first integral, we have $|x| \leq h$, which implies that $h-x \geq 0$. Thus,

$$\begin{aligned}
\Delta(x_0+h-x) &=\Delta(x_0)+[\Delta(x_0+h-x)-\Delta(x_0)] \\
&=\eta+[F(x_0+h-x)-F(x_0)]-[G(x_0+h-x)-G(x_0)] \\
&\geq \eta-[G(x_0+h-x)-G(x_0)] \quad (\text{ as } h-x \geq 0) \\
&=\eta-G'(x_0+\theta(h-x))(h-x) \quad (\text{by MVT where } 0 \leq \theta \leq 1) \\
&\geq \eta-\lambda(h-x) \\
&=\frac{\eta}{2}+\lambda x \quad (\text{ as } \lambda h=\eta/2)
\end{aligned} \tag{3.51}$$

In the second integral, we have $|x|>h$. We will use the following trivial bound

$$\Delta(x_0+h-x) \geq -\Delta(x_0)=-\eta \tag{3.52}$$

Puting (3.51) and (3.52) into (3.50), we get

$$\begin{aligned}\eta_T &\geq \int_{|x|\leq h}\left(\frac{\eta}{2}+\lambda x\right)v_T(x)dx+\int_{|x|>h}(-\eta)v_T(x)dx\\ &=\frac{\eta}{2}P(|V_T|\leq h)-\eta P(|V_T|>h) \quad \text{by symmetry}\\ &=\frac{\eta}{2}[1-P(|V_T|>h)]-\eta P(|V_T|>h)\\ &=\frac{\eta}{2}-\frac{3\eta}{2}P(|V_T|>h)\end{aligned}$$

Now

$$\begin{aligned}P(|V_T|>h) &= \int_{|x|>h}\frac{1-\cos(Tx)}{\pi Tx^2}dx=\int_{|x|>h}\frac{2\sin^2(Tx/2)}{\pi Tx^2}dx\\ &=\int_{|x|>h}\frac{\sin^2(Tx/2)}{\pi(Tx/2)^2}d(Tx/2)=\int_{|y|>Th/2}\frac{\sin^2 y}{\pi y^2}dy\\ &=\frac{2}{\pi}\int_{y=Th/2}^{\infty}\frac{\sin^2 y}{y^2}dy\\ &\leq\frac{2}{\pi}\int_{y=Th/2}^{\infty}\frac{1}{y^2}dy\\ &=\frac{4}{\pi Th}\end{aligned}$$

we finally have

$$\eta_T \geq \frac{\eta}{2}-\frac{3\eta}{2}\times\frac{4}{\pi Th}=\frac{\eta}{2}-\frac{6\eta}{\pi Th}=\frac{\eta}{2}-\frac{12\lambda}{\pi T}$$

Namely

$$\eta \leq 2\eta_T+\frac{24\lambda}{\pi T}$$

■

Lemma 3.25 — 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 $T>0$

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

Proof. Define $F^T(x)=V_T\star F(x)$ and $G^T(x)=V_T\star G(x)$ to be the convolution of F and G with V_T, respectively. Also, let $f^T(x)$ and $g^T(x)$ be the respective p.d.f.'s of $F^T(x)$

and $G^T(x)$. Clearly, the convolutions F^T and G^T have Fourier transforms $\psi_F(t)\omega_T(t)$ and $\psi_G(t)\omega_T(t)$, respectively. Then by the inversion formula,

$$f^T(y) - g^T(y) = \frac{1}{2\pi}\int_{-T}^{T} e^{-ity}\left(\psi_F(t) - \psi_G(t)\right) w_T(t)dt$$

Integrating w.r.t. x, and using Fubini's theorem, we obtain

$$\begin{aligned}
\Delta^T(x) &:= V_T \star [F(x) - G(x)] \\
&= V_T \star F(x) - V_T \star G(x) \\
&= \int_{-\infty}^{x} \left(f^T(y) - g^T(y)\right) dy \\
&= \int_{-\infty}^{x} \left(\frac{1}{2\pi}\int_{-T}^{T} e^{-ity}\left(\psi_F(t) - \psi_G(t)\right) w_T(t)dt\right) dy \qquad (3.53) \\
&= \frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{x} e^{-ity}\left(\psi_F(t) - \psi_G(t)\right) w_T(t)dydt \\
&\quad \text{(by Fubini theorem, with justifications give below)} \\
&= \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\frac{\psi_F(t) - \psi_G(t)}{-it} w_T(t)dt + C
\end{aligned}$$

Use of Fubini's theorem above can be justified as follows. Since F is a d.f. with mean 0, we have $\psi_F(0) = 1$ and $\psi_F'(0) = 0$. Also by assumption $\psi_G(0) = 1$ and $\psi_G'(0) = 0$. Therefore, the integrand in (3.53) is a continuous function vanishing at the origin. So no problem of convergence arises.

Let $|x| \to \infty$, then $\Delta^T(x) \to 0$ as $F(x) - G(x) \to 0$. On the other hand, the integral in (3.53) also goes to 0 as $|x| \to \infty$ by the Riemann-Lebesgue lemma. Hence we must have $C = 0$. Then from Lemma 3.25 and (3.53), we have

$$\begin{aligned}
\sup_x |F(x) - G(x)| &\le \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{\psi_F(t) - \psi_G(t)}{t}\right| w_T(t)dt + \frac{24\lambda}{\pi T} \\
&\le \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\psi_F(t) - \psi_G(t)}{t}\right| dt + \frac{24\lambda}{\pi T}
\end{aligned}$$

■

Lemma 3.26 — Riemann–Lebesgue lemma. If f is L^1 integrable on R^d, that is to say, if the Lebesgue integral of $|f|$ is finite, then the Fourier transform of f satisfies

$$\hat{f}(z) \equiv \int_{\mathbb{R}^d} f(x)\exp(-iz\cdot x)dx \to 0 \text{ as } |z| \to \infty$$

Proof. First suppose that $f(x) = \chi_{(a,b)}(x)$, the indicator function of an open interval. Then:

$$\int f(x)e^{i\xi x}dx = \int_a^b e^{i\xi x}dx = \frac{e^{i\xi b} - e^{i\xi a}}{i\xi} \to 0 \text{ as } |\xi| \to \infty$$

By additivity of limits, the same holds for an arbitrary step function. That is, for any function f of the form:

$$f = \sum_{i=1}^{N} c_i \chi_{(a_i, b_i)}, \quad c_i \in \mathbb{R}, \quad a_i \leq b_i \in \mathbb{R}$$

We have that:

$$\lim_{|\xi| \to \infty} \int f(x) e^{i\xi x} dx = 0$$

Finally, let $f \in L^1$ be arbitrary. Let $\varepsilon \in \mathbb{R} > 0$ be fixed. since the simple functions are dense in L^1, there exists a simple function g such that:

$$\int |f(x) - g(x)| dx < \varepsilon$$

By our previous argument and the definition of a limit of a complex function, there exists $N \in \mathbb{N}$ such that for all $|\xi| > N$:

$$\left| \int g(x) e^{i\xi x} dx \right| < \varepsilon$$

By additivity of integrals:

$$\int f(x) e^{i\xi x} dx = \int (f(x) - g(x)) e^{i\xi x} dx + \int g(x) e^{i\xi x} dx$$

By the triangle inequality for complex numbers, the [triangle inequality] for integrals, multiplicativity of the absolute value, and Euler's Formula:

$$\left| \int f(x) e^{i\xi x} dx \right| \leq \int |f(x) - g(x)| dx + \left| \int g(x) e^{i\xi x} dx \right|$$

For all $|\xi| > N$, the right side is bounded by 2ε by our previous arguments. since ε was arbitrary, this establishes:

$$\lim_{|\xi| \to \infty} \int f(x) e^{i\xi x} dx = 0$$

for all $f \in L^1$ ■

3.10.8 Characteristic functions and smoothness condition

Definition 3.10.3 — Lattice d.f. and nonlattice d.f.. If all points of increase of F are among $b, b \pm h, b \pm 2h, \ldots,$, then we say that F is *a* lattice d.f. with span h

Theorem 3.10.18 — Characterization of lattice distribution. If $\lambda \neq 0$, the following three statements are equivalent:

1. $\psi(\lambda) = 1$
2. $\psi(t)$ has period λ, i.e., $\psi(t + n\lambda) = \psi(t)$ for all t and n
3. All points of increase of F are among $0, \pm h, \pm 2h, \ldots$, where $h = 2\pi/\lambda$.

Proof. We shall show that $(c) \to (b) \to (a) \to (c)$

1. If (c) is true and F attributes weight p_k to $kh, k = 0, \pm 1, \pm 2, \ldots$, then $\psi(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikht}$ which has period $2\pi/h = \lambda$. So (c) implies (b).
2. If (b) is true, by taking $n = 1$ and $t = 0$, we get $\psi(\lambda) = \psi(0) = 1$, which proves (a)
3. If (a) is true, $\psi(\lambda) = E\cos(\lambda X) + iE\sin(\lambda X) = 1$, then $\int_{-\infty}^{\infty}[1 - \cos(\lambda x)]dF(x) = 0$. Note that the integrand is nonnegative. So at every point x of increase for F, we must have $1 - \cos(\lambda x) = 0$. Thus F is concentrated on the multiples of $2\pi/\lambda$, and hence (c) is true.

■

Corollary 3.10.19 If $\lambda \neq 0$, the following three statements are equivalent:

1. $\psi(\lambda) = e^{ib\lambda}$
2. $\psi(t)$ satisfies $\psi(t + n\lambda) = \psi(t)e^{in\lambda b}$ for all t and n
3. All points of increase of F are among $b, b \pm h, b \pm 2h, \ldots$, where $h = 2\pi/\lambda$.

Theorem 3.10.20 There exist only the following three possibilities:

1. $|\psi(t)| \equiv 1$ for all t. In this case, $\psi(t) = e^{ibt}$ (degenerate at b).
2. $|\psi(\lambda)| = 1$ and $|\psi(t)| < 1$ for $0 < t < \lambda$ (lattice with span $h = 2\pi/\lambda$.)
3. $|\psi(t)| < 1$ for all $t \neq 0$ (non-lattice distribution).

Strongly nonlattice d.f.

We often need a condition slightly stronger than nonlattice, called strongly nonlattice.

Definition 3.10.4 — Strongly nonlattice. Suppose X is a r.v. with d.f. $F(x)$ and c.f. $\psi(t)$.F is said to be **strongly non-lattice** if Cramer's condition holds, i.e.,

$$\limsup_{|t|\to\infty} |\psi(t)| < 1$$

The next theorem shows that strong non-latticeness is stronger than latticeness.

Theorem 3.10.21 If a d.f. F is strongly non-lattice, it is non-lattice.

R The converse is not true. For example, let $P(X = 0) = p_0, P(X = 1) = p_1, P(X = a) = p_2$, where a is irrational, and $p_0 + p_1 + p_2 = 1$ Show that X is non-lattice, but not strongly non-lattice.

Proof. That X is non-lattice is obvious. It is easy to see that

$$\psi(t) = Ee^{itX} = p_0 + p_1 e^{it} + p_2 e^{ita}$$

Now since a is irrational, by Hurwitz's theorem (see "An Introduction to the Theory of Numbers", 5th edition, I. Niven, H.S. Zuckerman and H.L. Montgomery, 1991, p 342), there exists infinitely many rational numbers h/k such that

$$\left|a - \frac{h}{k}\right| < \frac{1}{\sqrt{5}k^2}$$

Arrange all such rational numbers h/k so that the denumerator k are in asending order and denote the corresponding numerators and denumerators by $\{m_k, k \geq 1\}$ and $\{n_k, k \geq 1\}$. Then n_k 's are strictly increasing to ∞ and

$$\left|a - \frac{m_k}{n_k}\right| < \frac{1}{\sqrt{5}n_k^2}, \quad k = 1, 2, \ldots$$

Now choosing $t_k = 2\pi n_k$, we get

$$\begin{aligned}
\psi(t_k) &= p_0 + p_1 e^{i2\pi n_k} + p_2 e^{i2\pi n_k a} \\
&= p_0 + p_1 + p_2 e^{i2\pi n_k \left(\frac{m_k}{n_k} + \left[a - \frac{m_k}{n_k}\right]\right)} \\
&= 1 - p_2 + p_2 e^{i2\pi m_k} e^{i2\pi n_k \left[a - \frac{m_k}{n_k}\right]} \\
&= 1 - p_2 + p_2 e^{i2\pi (n_k a - m_k)} \\
&= 1 + p_2 \left(e^{i2\pi (n_k a - m_k)} - 1\right)
\end{aligned}$$

Therefore,

$$|\psi(t_k) - 1| \leq 2\pi p_2 |n_k a - m_k| < \frac{2\pi p_2}{\sqrt{5}n_k} \to 0$$

■

Absolutely continuity implies strong non-latticeness

Theorem 3.10.22 If the d.f. F has a non-zero absolutely continuous component($c_1 \neq 0$), it is strongly non-lattice (hence non-lattice).

Proof. By the Lebesgue decomposition theorem, the d.f. F has a unique decomposition

$$F(x) = c_1 F_1(x) + c_2 F_2(x) + c_3 F_3(x)$$

where $c_k \geq 0 (k = 1, 2, 3), c_1 + c_2 + c_3 = 1$, and $F_1(x), F_2(x), F_3(x)$ are absolutely continuous, discrete, and singular d.f.s respectively. Letting $\psi_i(t) = \int e^{itx} dF_i(x)$ for $i = 1, 2, 3$, then,

$$\psi(t) = c_1 \psi_1(t) + c_2 \psi_2(t) + c_3 \psi_3(t)$$

From the assumption, $0 < c_1 \leq 1$ and $F_1(x)$ has density $f_1(x)$. By Riemann-Lebesgue lemma,

$$\lim_{|t| \to \infty} \psi_1(t) = \lim_{|t| \to \infty} \int e^{itx} f_1(x) dx = 0$$

Therefore,

$$\limsup_{|t|\to\infty} |\psi(t)| \le c_1 \limsup_{|t|\to\infty} |\psi_1(t)| + c_2 + c_3 = c_2 + c_3 < 1$$

■

Continuous d.f. must be nonlattice, but is not necessarily strongly nonlattice

The d.f. in the above example is of discrete type, but also non-lattice. The next example shows that if the d.f. is singular (which is continuous but not absolute continuous), Cramer's condition may also fail.

■ **Example 3.21 — Continuity (but not absolute continuity) $\not\Longrightarrow$ strongly nonlattice.** Any distribution having jumps only on the Cantor set does not satisfy Cramer's condition.

Proof. Denote the Cantor set by $\{x_{kj} = j/3^k, k = 1,2,\ldots, j = 1,2,\ldots,\} \cap [0,1]$ which has mass p_{kj} at x_{kj}. It is clear that the d.f. is singular (which must be nonlattice). We now show that it is NOT strongly nonlattice. Now

$$\psi(t) = Ee^{itX} = \sum_{k,j\ge 1} \exp\{itx_{kj}\} p_{kj} = \sum_{k,j\ge 1} \exp\left\{itj/3^k\right\} p_{kj}$$

Choosing $t = t_m = 2\pi 3^m \to \infty$ as $m \to \infty$, we get

$$\begin{aligned}
\psi(t_m) &= \sum_{k,j\ge 1} \exp\left\{i(2\pi 3^m)j/3^k\right\} p_{kj} \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \exp\left\{i\left(2\pi j 3^{m-k}\right)\right\} p_{kj} \\
&= \left(\sum_{k=1}^{m} + \sum_{k=m+1}^{\infty}\right) \sum_{j=1}^{\infty} \exp\left\{i\left(2\pi j 3^{m-k}\right)\right\} p_{kj} \\
&= \sum_{k=1}^{m}\sum_{j=1}^{\infty} p_{kj} + \sum_{k=m+1}^{\infty}\sum_{j=1}^{\infty} \exp\left\{i\left(2\pi j 3^{m-k}\right)\right\} p_{kj} \quad (e^{i\pi} = 1) \\
&\to 1
\end{aligned}$$

Therefore, $\limsup_{|t|\to\infty} |\psi(t)| = 1$. That is, Cramer's condition does not hold. ■

A quick summary

We can classify d.f.s as follows. (A tree diagram or Venn diagram is better).

1. Discrete
   (a) Lattice (e.g. Binomial, Poisson, Negative Binomial, Geometric)
   (b) Nonlattice
      i. Strongly nonlattice (e.g. an empirical distribution function (edf) with pdf f, such as a bootstrap sample from a absolute continuous dist.)
      ii. Weakly nonlattice (i.e., non-lattice, but not strongly nonlattice) (e.g. $P(X = 0) = p_0, P(X = 1) = p_1, P(X = \sqrt{2}) = 1 - p_0 - p_1$

2. Absolutely continuous (which must be nonlattice)
3. Singular (which is weakly nonlattice)
4. Mixtures of $I - -III$. (e.g. any d.f. with a non-zero absolutely continuous component is strongly nonlattice.)

3.11 Central limit Theorems and Related Expansions

limit theorems for sums of independent r.v.'s occupy a special place in probability and mathematical statistics. First of all, they are the simplest case to study. Second, the theory for sums of independent r.v.'s are the most complete. Third, studies on sum of independent r.v.'s can shed light on other classes of statistics such as the function of the sum of independent r.v.'s, U -statistics, L -statistics and symmetric statistics, etc.

In this chapter, we shall discuss some important issues concerning the sum of independent r.v.'s. In particular, we shall discuss the asymptotic normality, Berry-Esseen bounds (uniform or nonuniform), Edgeworth expansions (uniform or nonuniform), large deviations and saddlepoint approximations, and others. This will pave the way for other classes of statistics.

3.11.1 Central Limit Theorems (ClLT)

CLT for i.i.d. r.v.s

When $X_1, \ldots, X_n$ are i.i.d., we get the following simple CLT.

Theorem 3.11.1 — (Levy theorem). Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with $EX_1 = 0$, and $\sigma^2 = EX_1^2 < \infty$. Let $F_n(x) = P(\sqrt{n}\bar{X}/\sigma \leq x)$. Then

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \to 0$$

Proof. We provide two methods.

1. Method 1. Since $EX_1^2 < \infty, \psi(t)$ is twice differentiable and has the following Taylor expansion,

$$\psi(t) = \psi(0) + \psi'(0)t + \frac{1}{2}\psi''(0)t^2 + o\left(t^2\right) = 1 - \frac{\sigma^2 t^2}{2} + o\left(t^2\right)$$

Then the c.f. of F_n is

$$\psi_{\sqrt{n}\bar{X}/\sigma}(t) = \psi^n\left(\frac{t}{\sqrt{n}\sigma}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-t^2/2}$$

2. Method 2. The Lindeberg condition in Corollary 3.11.5 holds since

$$\frac{1}{B_n^2}\sum_{k=1}^{n} EX_k^2 I\left\{|X_k| \geq \epsilon B_n\right\} = \frac{EX_1^2 I\left\{|X_1| \geq \epsilon\sqrt{nEX_1^2}\right\}}{EX_1^2} \to 0$$

■

CLT for triangular arrays with finite variances

The following theorem states that a sum of a large number of small independent effects has approximately a normal distribution.

Theorem 3.11.2 — Lindeberg-Feller CLT. For each n, let $X_{n,k}, 1 \leq k \leq n$, be independent r.v.s with $EX_{n,k} = 0$ and $\sum_{k=1}^{n} \sigma_{n,k}^2 := \sum_{k=1}^{n} EX_{n,k}^2 = 1$. Denote $F_n(x) = P\left(\sum_{m=1}^{n} X_{n,k} \leq x\right)$. Then the following two statements are equivalent.

1. The Lindeberg condition holds:

$$\forall \epsilon > 0: \quad \lim_{n \to \infty} \sum_{m=1}^{n} EX_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\} = 0$$

2. (a) $\max_{1 \leq m \leq n} \sigma_{n,k}^2 \to 0$ and
   (b) $\sup_{x \in R} |F_n(x) - \Phi(x)| \to 0$

R Clearly, the Lindeberg condition implies that

$$\forall \epsilon > 0: \quad \lim_{n \to \infty} \sum_{m=1}^{n} P\left(|X_{n,k}| \geq \epsilon\right) = 0$$

which further implies that

$$\forall \epsilon > 0: \quad \sup_n P\left(|X_{n,k}| \geq \epsilon\right) \longrightarrow 0$$

That is, all the individual terms $X_{n,k}$ are uniformly small.

Proof. 1. "(i) $\Longrightarrow$ (ii)". We first prove (a) of (ii). For $1 \leq k \leq n$, we have

$$\begin{aligned}
\sigma_{n,k}^2 &= EX_{n,k}^2 I\{|X_{n,k}| < \epsilon\} + EX_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\} \\
&\leq \epsilon^2 + EX_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\}
\end{aligned}$$

Thus $\max_{1 \leq k \leq n} \sigma_{n,k}^2 \leq \epsilon^2 + \sum_{k=1}^{n} EX_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\}$. Letting $n \to \infty$, the Lindeberg condition implies that $\max_{1 \leq k \leq n} \sigma_{n,k}^2 \leq 2\epsilon^2$. Since ϵ can be chosen to be arbitrarily small, we have

$$\max_{1 \leq m \leq n} \sigma_{n,k}^2 \to 0$$

We now prove (b) of (ii). Write $\psi_{n,k}(t) = Ee^{itX_{n,k}}$. It suffices to show that, $\forall t \in R$

$$\prod_{k=1}^{n} \psi_{n,k}(t) \longrightarrow e^{-t^2/2} \Longleftrightarrow \sum_{k=1}^{n} \ln \psi_{n,k}(t) + t^2/2 \longrightarrow 0 \tag{3.54}$$

It suffices to show that, as $n \to \infty, \forall t \in R$

$$\sum_{k=1}^{n} \ln \psi_{n,k}(t) - \sum_{k=1}^{n} (\psi_{n,k}(t) - 1) \to 0 \tag{3.55}$$

$$\sum_{k=1}^{n} (\psi_{n,k}(t) - 1) + \frac{t^2}{2} \to 0 \tag{3.56}$$

Let us prove (3.55) first. From the inequality $\left|e^{it} - 1 - it\right| \le t^2/2$ for any real t, we have

$$|\psi_{n,k}(t) - 1| = \left|Ee^{itX_{n,k}} - 1 - itEX_{n,k}\right| \le E\left|e^{itX_{n,k}} - 1 - itX_{n,k}\right| \le \frac{1}{2}t^2 EX_{n,k}^2 = \frac{1}{2}t^2\sigma_{n,k}^2$$

($EX_{n,k} = 0$) Thus, as $n \to \infty$

$$\max_{1\le k\le n} |\psi_{n,k}(t) - 1| \le \frac{1}{2}t^2 \max_{1\le k\le n} \sigma_{n,k}^2 = o(1), \quad \text{and} \quad \sum_{k=1}^{n} |\psi_{n,k}(t) - 1| \le \frac{t^2}{2}$$

Hence, from Theorem 11.8.1 : $|\ln(1+z) - z| \le |z|^2$ for $|z| \le 1/2$, (3.55) follows from

$$\sum_{k=1}^{n} |\ln \psi_{n,k}(t) - (\psi_{n,k}(t) - 1)| \le \sum_{k=1}^{n} |\psi_{n,k}(t) - 1|^2 \le o(1) \sum_{k=1}^{n} |\psi_{n,k}(t) - 1| = o(1)$$

($\max_{1\le k\le n} |\psi_{n,k}(t) - 1| \le o(1)$) Next let us prove (3.56). By using the inequality $\left|e^{it} - 1 - it - \frac{1}{2}(it)^2\right| \le \min\left\{t^2, \frac{1}{6}|t|^3\right\}$ for any real t, we have

$$\begin{aligned}
\left|\sum_{k=1}^{n} (\psi_{n,k}(t) - 1) + \frac{t^2}{2}\right| &= \left|\sum_{k=1}^{n} E\left(e^{itX_{n,k}} - 1 - itX_{n,k} - \frac{1}{2}(itX_k)^2\right)\right| \\
&\le \sum_{k=1}^{n} E\min\left\{t^2X_{n,k}^2, \frac{1}{6}|tX_{n,k}|^3\right\} \\
&\le t^2 \sum_{k=1}^{n} EX_{n,k}^2 I\{|X_k| \ge \epsilon\} + \frac{|t|^3\epsilon}{6} \sum_{k=1}^{n} E|X_{n,k}|^2 I\{|X_{n,k}| < \epsilon\} \\
&\le t^2 \sum_{k=1}^{n} EX_{n,k}^2 I\{|X_k| \ge \epsilon\} + \frac{1}{6}|t|^3\epsilon
\end{aligned}$$

($\sum_{k=1}^{n} EX_{n,k}^2 = 1$) Then (3.56) follows from this, the lindeberg condition and by choosing ϵ arbitrarily small.

2. "(ii) $\Longrightarrow$ (i)". Assume that (ii) holds. First, part (b) of (ii) implies (3.54). Secondly, from the proceeding proof, we can see that that (3.55) is implied from part (a) of (ii). Putting these two together, we see that (3.56) still holds. In particular, the real part

of the left hand side in (3.56) should tend to 0, i.e.,

$$\begin{aligned}
0 \longleftarrow \operatorname{Re}\left(\sum_{k=1}^{n}\left(\psi_{n,k}(t)-1\right)+\frac{t^2}{2}\right) \\
&= \sum_{k=1}^{n} E\left(\cos\left(tX_{n,k}\right)-1+\frac{1}{2}t^2X_{n,k}^2\right) \\
&\geq \sum_{k=1}^{n} E\left(\cos\left(tX_{n,k}\right)-1+\frac{1}{2}t^2X_{n,k}^2\right) I\left\{\left|X_{n,k}\right|\geq\epsilon\right\} \\
&\qquad \left(\text{as } \cos(y)-1+\frac{1}{2}y^2\geq 0\right) \\
&\geq \sum_{k=1}^{n} E\left(\frac{1}{2}t^2X_{n,k}^2-2\right) I\left\{\left|X_{n,k}\right|\geq\epsilon\right\} \quad (\text{as} \cos(y)\geq -1) \\
&= \sum_{k=1}^{n} EX_{n,k}^2\left(\frac{1}{2}t^2-\frac{2}{X_{n,k}^2}\right) I\left\{\left|X_{n,k}\right|\geq\epsilon\right\} \\
&\geq \left(\frac{t^2}{2}-\frac{2}{\epsilon^2}\right)\sum_{k=1}^{n} EX_{n,k}^2 I\left\{\left|X_{n,k}\right|\geq\epsilon\right\}
\end{aligned}$$

So long as t is chosen so that $t^2/2-2/\epsilon^2>0$, i.e., $t^2>4/\epsilon^2$. Thus the right side tends to zero. Hence the lindeberg condition holds.

■

Theorem 3.11.3 — Lyapunov CLT. For each n, let $X_{n,k}, 1\leq k\leq n$, be independent r.v.s with $EX_{n,k}=0$ and $\sum_{k=1}^{n}\sigma_{n,k}^2 := \sum_{k=1}^{n} EX_{n,k}^2 = 1$. Denote $F_n(x)=P\left(\sum_{m=1}^{n}X_{n,k}\leq x\right)$. If

$$\sum_{k=1}^{n} E\left|X_{n,k}\right|^{2+\delta} \to 0 \tag{3.57}$$

then

$$\sup_{x\in R}\left|F_n(x)-\Phi(x)\right|\to 0$$

Proof. The Lindeberg condition holds since

$$\begin{aligned}
\sum_{k=1}^{n} EX_{n,k}^2 I\left\{\left|X_{n,k}\right|\geq\epsilon\right\} &\leq \frac{1}{\epsilon^\delta}\sum_{k=1}^{n} E\left|X_{n,k}\right|^{2+\delta} I\left\{\left|X_{n,k}\right|\geq\epsilon\right\} \\
&\leq \frac{1}{\epsilon^\delta}\sum_{k=1}^{n} E\left|X_{n,k}\right|^{2+\delta} \to 0
\end{aligned}$$

■

An application: CLT for independent r.v.s

Let $X_1,\cdots,X_n$ be a sequence of independent non-degenerate r.v.'s such that $EX_j=0$ and $\operatorname{var}\left(X_j\right)=\sigma_j^2<\infty, j=1,\ldots,n$. Let

$$S_n=\sum_{j=1}^{n}X_j, \quad B_n^2=\sum_{j=1}^{n}\sigma_j^2$$

and

$$F_n(x) = P(S_n/B_n \le x)$$

Corollary 3.11.4 — Lindeberg-Feller CLT. The following two statements are equivalent.

1. The Lindeberg condition holds: for every fixed $\epsilon > 0$,

$$B_n^{-2}\sum_{k=1}^{n} EX_k^2 I\{|X_k| \ge \epsilon B_n\} \to 0, \quad \text{as } n \to \infty$$

2. (a) $\max_{1\le k\le n}\{\sigma_k^2/B_n^2\} \to 0$ and
   (b) $\sup_{x\in R}|F_n(x) - \Phi(x)| \to 0$

Proof. The theorem following by taking $X_{n,k} = X_k/B_n$ in Theorem 3.11.2 ■

Corollary 3.11.5 — Lyapunov CLT. If

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n} E|X_k|^{2+\delta} \to 0 \tag{3.58}$$

then

$$\sup_{x\in R}|F_n(x) - \Phi(x)| \to 0$$

Proof. The theorem following by taking $X_{n,k} = X_k/B_n$ in Corollary 3.11.2 ■

R

1. Both the Lindeberg condition in Theorem 3.11.2 and the Lyapunov condition (3.58) in Theorem 3.11.3 are sufficient but not necessary for the CLT to hold. This can be seen from the example where $X_1 \sim N(0,1)$, and $X_k = 0$ if $k \ge 2$. In other words, these conditions are stronger than needed to establish the asymptotic normality. In fact, one can give a more precise description about the rate of convergence to normality under these conditions. For instance, under the Lindeberg condition, one can show that (see Petrov, 1995)

$$\sup_x |F_n(x) - \Phi(x)| \le A(\Lambda_n(\epsilon) + \epsilon) \quad \text{for any } \epsilon > 0$$

where

$$\Lambda_n(\epsilon) = B_n^{-2}\sum_{k=1}^{n} EX_k^2 I\{|X_k| \ge \epsilon B_n\}$$

On the other hand, under the Lyapunov condition, one can show that (see Section 11.2 for more detail

$$\sup_x |F_n(x) - \Phi(x)| \le AL_n$$

where $L_{n,\delta} = B_n^{-(2+\delta)}\sum_{k=1}^{n} E|X_k|^{2+\delta}$ and A is an absolute positive constant. Note that both inequalities clearly imply the CLT.

2. The Lindeberg condition in Corollary 3.11.4 is equivalent to

$$\Lambda_n'(\epsilon) := B_n^{-2} \sum_{k=1}^{n} EX_k^2 I\{|X_k| \geq \epsilon B_k\} \to 0$$

where $B_k^2 = \sum_{i=1}^k EX_k^2$. To see this, first assume that $\Lambda_n'(\epsilon) \to 0$, then monotonicity of B_k^2 yields $\Lambda_n(\epsilon) \to 0$. On the other hand, if $\Lambda_n(\epsilon) \to 0$, then the reverse implication follows by noting that for all $\epsilon > 0$ and arbitrarily small $\delta > 0$,

$$\begin{aligned}
\Lambda_n'(\epsilon) =& B_n^{-2} \left(\sum_{\{k: B_k \leq \delta B_n\}} + \sum_{\{k: B_k > \delta B_n\}} \right) EX_k^2 I\{|X_k| \geq \epsilon B_k\} \\
\leq & B_n^{-2} \sum_{\{k: B_k \leq \delta B_n\}} EX_k^2 + \sum_{k=1}^{n} EX_k^2 I\{|X_k| \geq \epsilon\delta B_n\} \\
\leq & B_n^{-2} B_m^2 + \Lambda_n(\epsilon) \\
& \left(\text{where } m = \sup\{k : B_k \leq \delta B_n\}. \text{ Note: } B_1^2 \leq B_2^2 \leq \ldots \leq B_n^2 \right) \\
=& \delta^2 + \Lambda_n(\epsilon) \\
\to & \delta^2
\end{aligned}$$

Central Limit Theorems with infinite variances

So far, we have discussed the CLT under the second moment condition. In fact, CLT can hold under slightly weaker condition.

Theorem 3.11.6 If $X, \{X_n, n \geq 1\}$ are i.i.d. r.v.s with non-degenerate d.f. F, then

$$\lim_{n\to\infty} P\left(\frac{1}{B_n} \sum_{i=1}^{n} X_i - A_n \leq x \right) = \Phi(x)$$

for some $B_n > 0$ and A_n iff

$$\lim_{C\to\infty} \frac{P(|X| > C)}{C^{-2} EX^2 I\{|X| \leq C\}} = 0$$

Moreover, A_n, B_n may be chosen as

$$B_n = \sup\left\{ C : C^{-2} EX^2 I\{|X| \leq C\} \geq \frac{1}{n} \right\}$$

$$A_n = \frac{n}{B_n} EXI\{|X| < B_n\}$$

3.11.2 Uniform Berry-Esseen Bounds

The Central limit Theorems state that the d.f. of the sum of independent r.v.'s converges to normal d.f. under appropriate conditions. However, they give no information how fast this convergence is. It is therefore of interest to investigage the rates of convergence to the normal law. These are provided by the Berry-Esseen type inequalities. A key tool we shall use is the Smoothing Lemma.

Throughout the section, let $X_1, \ldots, X_n$ be independent r.v.'s such that $EX_j = 0$ and $E\left|X_j\right|^{2+\delta} < \infty$ for some $0 < \delta \le 1 (j = 1, \ldots, n)$. Put

$$EX_j^2 = \sigma_j^2, \quad B_n^2 = \sum_{j=1}^n \sigma_j^2, \quad L_{n,\delta} = \frac{\sum_{j=1}^n E\left|X_j\right|^{2+\delta}}{B_n^{2+\delta}}$$

In particular, in the i.i.d. case, where $EX_j^2 = \sigma^2$, we have

$$B_n^2 = n\sigma^2, \quad L_{n,\delta} = \frac{nE\left|X_1\right|^{2+\delta}}{(n\sigma^2)^{(2+\delta)/2}} = \frac{C}{n^{\delta/2}}, \quad L_{n,1} = \frac{C}{n^{1/2}}$$

Some useful lemmas

In this section, we write $X_{n,k} = X_k / B_n, \psi_{n,k}(t) = Ee^{itX_{n,k}}$ and $\psi_n(t) = Ee^{it\sum_{k=1}^n X_{n,k}}$

Lemma 3.27

$$\left|\psi_n(t) - e^{-t^2/2}\right| \le 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/2} \quad \text{for } |t| < \frac{1}{2}L_{n,\delta}^{-1/(2+\delta)} \tag{3.59}$$

R In the i.i.d. case, $L_{n,\delta}^{-1/(2+\delta)} = Cn^{\delta/[2(2+\delta)]}$

Proof. Note that

$$\begin{aligned}\left|\psi_n(t) - e^{-t^2/2}\right| &= \left|\prod_{k=1}^n \psi_{n,k}(t) - e^{-t^2/2}\right| \\ &= \left|\exp\left\{\sum_{k=1}^n \ln\psi_{n,k}(t)\right\} - e^{-t^2/2}\right| \\ &= e^{-t^2/2}\left|\exp\left\{\sum_{k=1}^n \ln\psi_{n,k}(t) + \frac{t^2}{2}\right\} - 1\right|\end{aligned} \tag{3.60}$$

Using Theorem 3.10.16, we have

$$\begin{aligned}\psi_{n,k}(t) &= 1 + itEX_{n,k} + \frac{1}{2}(it)^2EX_{n,k}^2 + \theta|t|^{2+\delta}E\left|X_{n,k}\right|^{2+\delta} \\ &= 1 - \frac{1}{2}t^2\sigma_{n,k}^2 + \theta|t|^{2+\delta}E\left|X_{n,k}\right|^{2+\delta}\end{aligned}$$

where θ is a complex number with $|\theta| \le 1$. Noting that

$$|t|\sigma_{n,k} \le |t|\left(E\left|X_{n,k}\right|^{2+\delta}\right)^{1/(2+\delta)} \le |t|L_{n,\delta}^{1/(2+\delta)} < \frac{1}{2}$$

Thus,

$$
\begin{aligned}
|\psi_{n,k}(t)-1| &\le \frac{1}{2}\times\frac{1}{4}+\frac{1}{2^{2+\delta}}<\frac{3}{8}\\
|\psi_{n,k}(t)-1|^2 &\le 2\left[\left(\frac{1}{2}t^2\sigma_{n,k}^2\right)^2+\left(|t|^{2+\delta}E\,|X_{n,k}|^{2+\delta}\right)^2\right]\\
&=2\left(\frac{1}{4}\,|t\sigma_{n,k}|^{2+\delta}\,|t\sigma_{n,k}|^{2-\delta}+\left(|t|^{2+\delta}E\,|X_{n,k}|^{2+\delta}\right)\left(|t|^{2+\delta}E\,|X_{n,k}|^{2+\delta}\right)\right)\\
&\le 2\left(\frac{1}{8}+\frac{1}{2^{2+\delta}}\right)|t|^{2+\delta}E\,|X_{n,k}|^{2+\delta}\\
&=\frac{3}{4}|t|^{2+\delta}E\,|X_k|^{2+\delta}
\end{aligned}
$$

By Taylor expansion of $\log(1+z)$, we can easily find that (see Appendix)

$$\log(1+z)=z+\frac{4}{5}\theta|z|^2,\quad \text{where } |\theta|\le 1 \text{ and } |z|<\frac{3}{8}$$

Hence, taking $z=\psi_{n,k}(t)-1$, we get

$$
\begin{aligned}
\ln\psi_{n,k}(t)&=\ln\left(1+[\psi_{n,k}(t)-1]\right)=(\psi_{n,k}(t)-1)+\frac{4}{5}\theta_k\,|\psi_{n,k}(t)-1|^2\\
&=-\frac{1}{2}t^2\sigma_{n,k}^2+\frac{8}{5}\theta_k|t|^{2+\delta}E\,|X_{n,k}|^{2+\delta},\quad |\theta_k|\le 1\quad \left(1+\frac{4}{5}\cdot\frac{3}{4}=\frac{8}{5}\right)
\end{aligned}
$$

and thus

$$s=:\sum_{k=1}^{n}\ln\psi_{n,k}(t)+\frac{t^2}{2}=\sum_{k=1}^{n}\ln\left(1+r_k\right)+\frac{t^2}{2}=\frac{8}{5}\theta|t|^{2+\delta}L_{n,\delta},\quad |\theta|\le\max_{1\le k\le n}|\theta_k|\le 1$$

Clearly, $|s|\le 2/5$, $(\frac{8}{5}\cdot\frac{1}{2^{2+\delta}}=\frac{2}{5})$. So from (3.60) and the inequality $|e^z-1|\le|z|e^{|z|}$ for every complex z, we get

$$
\begin{aligned}
\left|\psi_n(t)-e^{-t^2/2}\right| &\le e^{-t^2/2}|s|e^{|s|}\\
&\le e^{-t^2/2}\frac{8}{5}|t|^{2+\delta}L_{n,\delta}e^{2/5}\\
&\le 3|t|^{2+\delta}L_{n,\delta}e^{-t^2/2}
\end{aligned}
$$

■

Lemma 3.28

$$\left|\psi_n(t)-e^{-t^2/2}\right|\le 16L_{n,\delta}|t|^{2+\delta}e^{-t^2/3}\quad \text{for } |t|\le\left(\frac{1}{36L_{n,\delta}}\right)^{1/\delta}$$

Proof. Consider the symmetrized r.v. $X_{n,j}^s=X_{n,j}-Y_{n,j}$, where $X_{n,j}$ and $Y_{n,j}$ are independent with the same distribution. Clearly, $X_{n,j}^s$ has the c.f. $\left|\psi_{n,j}(t)\right|^2$ and the variance $2\sigma_{n,j}^2$. Furthermore, by using the inequality $|a+b|^r\le 2^{r-1}\left(|a|^r+|b|^r\right)$, we have

$$E\left|X_{n,j}^s\right|^{2+\delta}\le 2^{1+\delta}\left(E\left|X_{n,j}\right|^{2+\delta}+E\left|Y_{n,j}\right|^{2+\delta}\right)\le 8E\left|X_{n,j}\right|^{2+\delta}$$

Using Theorem **??** and the inequality $e^x \geq 1+x$ for all real x, we have

$$\begin{aligned}
\left|\psi_{n,j}(t)\right|^2 &= 1+(it)EX_{n,j}^s - \frac{1}{2}t^2 E\left(X_{n,j}^s\right)^2 + \theta|t|^{2+\delta}E\left|X_{n,j}^s\right|^{2+\delta} \\
&\leq 1-\sigma_{n,j}^2 t^2 + 8|t|^{2+\delta}E\left|X_{n,j}\right|^{2+\delta} \\
&\leq \exp\left\{-\sigma_{n,j}^2 t^2 + 8|t|^{2+\delta}E\left|X_{n,j}\right|^{2+\delta}\right\}
\end{aligned}$$

From the assumption that $|t|^\delta \leq 1/\left(36L_{n,\delta}\right)$, we have

$$\begin{aligned}
|\psi_n(t)|^2 &= \prod_{j=1}^n \left|\psi_{n,j}(t)\right|^2 \leq \exp\left\{-t^2+8|t|^{2+\delta}L_{n,\delta}\right\} \leq \exp\left\{-t^2+8t^2/36\right\} \\
&= \exp\left\{-t^2+2t^2/9\right\} = \exp\left\{-7t^2/9\right\} \leq \exp\left\{-2t^2/3\right\}
\end{aligned} \tag{3.61}$$

1. If $|t| < \left(2L_{n,\delta}^{1/(2+\delta)}\right)^{-1}$, then from Lemma 3.27

$$\left|\psi_n(t)-e^{-t^2/2}\right| \leq 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/2} \leq 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/3}$$

2. If $\left(2L_{n,\delta}^{1/(2+\delta)}\right)^{-1} \leq |t| \leq \left(36L_{n,\delta}\right)^{-1/\delta}$, then $\left(2^{2+\delta}\right)^{-1} \leq |t|^{2+\delta}$, i.e., $|t|^{2+\delta}L_{n,\delta} \geq 2^{-(2+\delta)} \geq 2^{-3} = 1/8$, i.e. $8|t|^{2+\delta}L_{n,\delta} \geq 1s$. Thus, then from (3.61)

$$\left|\psi_n(t)-e^{-t^2/2}\right| \leq |\psi_n(t)|+e^{-t^2/2} \leq 2\exp\left\{-t^2/3\right\} \leq 16L_{n,\delta}|t|^{2+\delta}e^{-t^2/3}$$

■

If we take $\delta = 1$ above, we have

Corollary 3.11.7 Let $X_1,\ldots,X_n$ be independent r.v.'s such that $EX_j = 0$ and $E\left|X_j\right|^3 < \infty (j=1,\ldots,n).Put$

$$EX_j^2 = \sigma_j^2, \quad B_n^2 = \sum_{j=1}^n \sigma_j^2, \quad L_{n,1} = B_n^{-3/2}\sum_{j=1}^n E\left|X_j\right|^3$$

Let $\psi_n(t)$ be the c.f. of the random variable $B_n^{-1/2}\sum_{j=1}^n X_j$. Then

1. $\left|\psi_n(t)-e^{-t^2/2}\right| \leq 3|t|^3 L_{n,1}e^{-t^2/2}$ for $|t| < \frac{1}{2L_{n,1}^{1/3}}$
2. $\left|\psi_n(t)-e^{-t^2/2}\right| \leq 16|t|^3 L_{n,1}e^{-t^2/3}$ for $|t| \leq \frac{1}{36L_{n,1}}$

The corollary becomes even more apparent in the i.i.d. case.

Corollary 3.11.8 Let $X_1,\ldots,X_n$ be i.i.d. r.v.'s. Let

$$EX_1 = 0, \quad EX_1^2 = \sigma^2 > 0, \quad E\left|X_1\right|^3 < \infty, \quad \rho = E\left|X_1\right|^3/\sigma^3$$

Let $\psi_n(t)$ be the c.f. of the random variable $\sigma^{-1}n^{-1/2}\sum_{j=1}^n X_j$. Then

1. $\left|\psi_n(t)-e^{-t^2/2}\right| \leq 3|t|^3\rho e^{-t^2/2}n^{-1/2}$ for $|t| < \left(2\rho^{1/3}\right)^{-1}n^{1/6}$

2. $\left|\psi_n(t) - e^{-t^2/2}\right| \le 16|t|^3 \rho e^{-t^2/3} n^{-1/2}$ for $|t| \le (36\rho)^{-1} n^{1/2}$

Berry-Esseen bounds

Theorem 3.11.9 — Berry-Esseen bounds for independent r.v.'s. Let $X_1, \ldots, X_n$ be independent r.v.'s such that $EX_j = 0$ and $E\left|X_j\right|^{2+\delta} < \infty (j = 1, \ldots, n)$ for some $0 < \delta \le 1$. Put

$$L_{n,\delta} = B_n^{-(2+\delta)} \sum_{j=1}^{n} E\left|X_j\right|^{2+\delta}$$

Then for all n

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \le A L_{n,\delta}$$

Proof. In the Smoothing Lemma we set $T = (36 L_{n,\delta})^{-1/\delta}$. Note that $\Phi'(x) = \phi(x) = (2\pi)^{-1/2} e^{-x^2/2} \le (2\pi)^{-1/2}$. Then we have

$$\begin{aligned} \sup_x |F_n(x) - \Phi(x)| &\le \frac{2}{\pi} \int_0^T |t|^{-1} \left|\psi_n(t) - e^{-t^2/2}\right| dt + \frac{24\lambda}{\pi T} \\ &\le \pi^{-1} \int_0^T 16 L_{n,\delta} t^{1+\delta} e^{-t^2/3} dt + \frac{24\lambda}{\pi} (36 L_{n,\delta})^{1/\delta} \\ &\le C_\delta L_{n,\delta} + C_\delta L_{n,\delta}^{1/\delta} \end{aligned}$$

If $L_{n,\delta}^{1/\delta} \le 1$, then $\sup_x |F_n(x) - \Phi(x)| \le 2 C_\delta L_{n,\delta}$ with $A = 2C_\delta$. On the other hand, if $L_{n,\delta}^{1/\delta} > 1$, then we can simply take $A = 1$ ■

In the i.i.d. case, Theorem 3.11.9 reduces to the following corollary.

Corollary 3.11.10 — Berry-Esseen bounds for i.i.d. r.v.'s. Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s. Let $\delta \in (0,1]$, and

$$EX_1 = 0, \quad EX_1^2 = \sigma^2 > 0, \quad E|X_1|^{2+\delta} < \infty, \quad \rho_\delta = E|X_1|^{2+\delta} / \sigma^{2+\delta}$$

Then for all n

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

In particular, the case $\delta = 1$ gives the most familiar Berry-Esseen bound.

Corollary 3.11.11 Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s. Let

$$EX_1 = 0, \quad EX_1^2 = \sigma^2 > 0, \quad E|X_1|^3 < \infty, \quad \rho = E|X_1|^3 / \sigma^3$$

Then for all n

$$\sup_{x\in R}|F_n(x)-\Phi(x)|\leq\frac{A\rho}{\sqrt{n}}$$

R

1. The constant A in Theorem **??** can be taken to be 3, see Feller (1971). However, the best known constant is $A=0.80$ given by Berk (1972)
2. Berry-Esseen bounds hold for all n and x, so in that sense, they are not asymptotic results.
3. The order of estimates $n^{-1/2}$ in Cor 3.11.11 can not be improved without additional conditions on the distributions of our r.v.'s. In other words, the third moment condition is minimal.

   ■ **Example 3.22** Let $P(X_i=\pm 1)=1/2, i=1,\ldots,n$. Denote $S_n=\sum_{i=1}^n X_i$. Suppose that $n=2m$ is even. From Sterling's formula: $n!=\sqrt{2\pi n}(n/e)^n(1+o(1))$, we have

$$\begin{aligned}
P(S_n=0)&=P\left(\sum_{i=1}^n I\{X_i=-1\}=m\right)=\left(\begin{array}{c}n\\m\end{array}\right)\frac{1}{2^n}=\frac{1}{2^n}\frac{n!}{m!^2}\\
&=\frac{\sqrt{2\pi n}(n/e)^n(1+o(1))}{\left(\sqrt{2\pi m}(m/e)^m(1+o(1))\right)^2}\cdot\frac{1}{2^n}\\
&=\frac{\sqrt{2}}{\sqrt{\pi n}}(1+o(1))
\end{aligned}$$

   That is, the d.f. of S_n has a jump of size $O\left(n^{-1/2}\right)$ at 0, and can not be approximated by a continuous function at 0 with error size of less than $O\left(n^{-1/2}\right)$. More precisely, we have

$$\begin{aligned}
F_n(0)&=P\left((S_n-ES_n)/\sqrt{\operatorname{var}(S_n)}\leq 0\right)=P(S_n\leq 0)\\
&=\frac{1}{2}\left(P(S_n\leq 0)+P(S_n\geq 0)\right)\quad(\textit{by}\text{ symmetry})\\
&=\frac{1}{2}\left(P(S_n\leq 0)+P(S_n>0)+P(S_n=0)\right)=\frac{1}{2}\left(1+P(S_n=0)\right)\\
&=\Phi(0)+\frac{1}{\sqrt{2\pi n}}(1+o(1))
\end{aligned}$$

   Hence

$$\sup_x|F_n(x)-\Phi(x)|\geq|F_n(0)-\Phi(0)|=\frac{1}{\sqrt{2\pi n}}(1+o(1))$$

   In other words, the discrete d.f. $F_n(x)$ can not be approximated by a continuous d.f. Φ to an accuracy smaller than $(2\pi n)^{-1/2}(1+o(1))$

4. The above example also shows that even if the r.v. has moments of all orders, the order of error in the normal approximation is still $O\left(n^{-1/2}\right)$. This is because we are approximating a distribution with jump size $O\left(n^{-1/2}\right)$ with a continuous distribution. Better approximations are possible in this case by using Edgeworth type expansions, which will be discussed later.

5. It also follows from the example that the absolute constant A in Theorem 3.11.11 is not less than $(2\pi n)^{-1/2} \approx 0.4$. This is also true for other Berry-Esseen bounds.

A generalization of Berry-Esseen bounds

Here we shall give Berry-Esseen bounds assuming only second-order moments.

Theorem 3.11.12 Let $X_1,\dots,X_n$ be independent r.v.'s such that $EX_j = 0$ and $E\left|X_j\right|^{2+\delta} < \infty (j = 1,\dots,n).Put$

$$\Lambda_n(\epsilon) = B_n^{-2} \textstyle\sum_{j=1}^n E\left|X_j\right|^2 I\left\{\left|X_j\right| > \epsilon B_n\right\}$$
$$\lambda_n(\epsilon) = B_n^{-3} \textstyle\sum_{j=1}^n E\left|X_j\right|^3 I\left\{\left|X_j\right| \le \epsilon B_n\right\}$$

Then for all n and $\epsilon > 0$

$$\sup_{x\in R} |F_n(x) - \Phi(x)| \le A\left(\Lambda_n(\epsilon) + \lambda_n(\epsilon)\right)$$

Proof. 1. From the definition, we have, for all $\epsilon > 0$,

$$\lambda_n(\epsilon) \le \epsilon B_n^{-2} \sum_{j=1}^n E\left|X_j\right|^2 I\left\{\left|X_j\right| \le \epsilon B_n\right\} \le \epsilon$$

Therefore, we have the inequality,

$$\sup_{x\in R} |F_n(x) - \Phi(x)| \le A\left(\Lambda_n(\epsilon) + \epsilon\right)$$

which, in turn, implies the Lindeberg CLT since $\Lambda_n(\epsilon) \to 0$ as $\epsilon \to 0$

2. On the other hand, we note that, for every $\delta \in (0,1]$,

$$\Lambda_n(\epsilon) + \lambda_n(\epsilon) \le \frac{\sum_{j=1}^n E\left|X_j\right|^{2+\delta}}{B_n^{2+\delta}}$$

Then we derive the Berry-Esseen bounds

$$\sup_{x\in R} |F_n(x) - \Phi(x)| \le \frac{A\sum_{j=1}^n E\left|X_j\right|^{2+\delta}}{B_n^{2+\delta}}$$

which, in turn, implies the Lynapnov CLT.

■

3.11.3 Non-Uniform Berry-Esseen Bounds

We have seen in Section 11.2 that the minimal condition for Berry-Esseen bounds of order $O\left(n^{-1/2}\right)$ is the existence of the third moment. Also we have seen that higher moment conditions can not improve the error order $O\left(n^{-1/2}\right)$ uniformly for all x. Although

Berry-Esseen bounds are true for all n and x they are only useful in the center of the distribution. They do not accurately describe the tail behavior.

On the other hand, by using Chebyshev's inequality, if $E|X|^r < \infty$ for some $r \geq 2$, then we have, for $x > 0$

$$\begin{aligned}
|F_n(x) - \Phi(x)| &= \max\{\Phi(x) - F_n(x), F_n(x) - \Phi(x)\} \\
&\leq \max\{1 - F_n(x), 1 - \Phi(x)\} \\
&= \max\{P(\sqrt{n}\bar{X} > x), P(N(0,1) > x)\} \\
&\leq \frac{1}{x^r} \max\{E|\sqrt{n}\bar{X}|^r, E|N(0,1)|^r\} \\
&\leq \frac{C}{x^r}
\end{aligned}$$

where the last line follows from Rosenthal inequality (e.g., Petrov, 1995, p59) :

$$\begin{aligned}
E|\sqrt{n}\bar{X}|^r &= n^{-r/2} E\left|\sum_1^n X_k\right|^r \\
&\leq n^{-r/2} C_r \left(\sum_1^n E|X_k|^r + \left(\sum_1^n EX_k^2\right)^{r/2}\right) \\
&\leq n^{-r/2} C\left(n + n^{r/2}\right) \leq C
\end{aligned}$$

Similarly, it is also true for $x < 0$. Therefore, for all x, we have

$$|F_n(x) - \Phi(x)| \leq \frac{C}{(1 + |x|)^r}$$

Although this bound is good at the tails, it is not good at the center since it does not involve n as in Berry-Esseen bounds.

So a more informative way to describe rates of convergence to normality is via the Non-uniform Berry-Esseen bounds, which involve both n and x.

Theorem 3.11.13 Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s. Let

$$EX_1 = 0, \quad EX_1^2 = \sigma^2 > 0, \quad E|X_1|^3 < \infty, \quad \rho = E|X_1|^3/\sigma^3$$

and

$$F_n(x) = P\left(\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j \leq x\right)$$

Then for all x and n

$$|F_n(x) - \Phi(x)| \leq \frac{A\rho}{\sqrt{n}} \frac{1}{(1 + |x|^3)}$$

For the independent case, we have

Theorem 3.11.14 Let $X_1,\ldots,X_n$ be independent r.v.'s such that $EX_j=0$ and $E\left|X_j\right|^3<\infty (j=1,\ldots,n)$ *Put*

$$EX_j^2=\sigma_j^2,\quad B_n^2=\sum_{j=1}^{n}\sigma_j^2,\quad L_n=B_n^{-3/2}\sum_{j=1}^{n}E\left|X_j\right|^3$$

$$F_n(x)=P\left(B_n^{-1}\sum_{j=1}^{n}X_j\leq x\right)$$

Then for all n and x

$$|F_n(x)-\Phi(x)|\leq\frac{AL_n}{1+|x|^3}$$

Theorem 3.11.15 Let $X_1,\ldots,X_n$ be i.i.d. with $E\left|X_1\right|^r<\infty, r\geq 3$. Then for all x and n

$$|F_n(x)-\Phi(x)|\leq C_r\left(\frac{E\left|X_1\right|^3/\sigma^3}{\sqrt{n}}+\frac{E\left|X_1\right|^r/\sigma^r}{n^{(r-2)/2}}\right)\frac{1}{(1+|x|)^r}$$

Theorem 3.11.16 Let $X_1,\ldots,X_n$ be i.i.d. with $E\left|X_1\right|^{2+\delta}<\infty, \delta\in(0,1]$. Then for all x and n,

$$|F_n(x)-\Phi(x)|\leq\frac{C_\delta E\left|X_1\right|^{2+\delta}}{\sigma^{2+\delta}n^{\delta/2}}\frac{1}{1+|x|^{2+\delta}}$$

3.11.4 Figeworth Expansions

The CLT can be established under second moment condition. Under higher moment condition, BerryEsseen bounds give convergence rates to normality. For instance, when $E\left|X_1\right|^3<\infty$, the error in the normal approximation is of size $O\left(n^{-1/2}\right)$. This suggests that one may include more correction terms in the CLT in order to get better approximations than normality. This is so-called Edgeworth Expansions.

Heuristic argument for informal Edgeworth expansions

For simplicity, assume that $X, X_1, X_2, \ldots$ are i.i.d. r.v.s with $EX=0, EX^2=1$, and $\rho=EX^3$. Let

$$F_n(x)=P(\sqrt{n}(\bar{X}-0)/1\leq x)=P(\sqrt{n}\,\bar{X}\leq x)$$

It is known that $F_n(x)\Longrightarrow\Phi(x)$. We will use the c.f. approach to derive a more accurate approximation. Note that $\psi_X(t)=1+itEX+(it)^2EX^2+\frac{1}{6}(it)^3EX^3+o\left(t^3\right)=1-t^2+\frac{1}{6}(it)^3\rho+o\left(t^3\right)$. Hence,

$$\psi_{\sqrt{n}\bar{X}}(t)=\psi_X^n(t/\sqrt{n})=\left(1-\frac{1}{2n}t^2+\frac{1}{6n^{3/2}}(it)^3\rho+n^{-3/2}o\left(t^3\right)\right)^n$$

Hence, using $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots$, we have

$$\begin{aligned}
\ln \psi_{\sqrt{n}\bar{X}}(t) &= n \ln\left(1 - \frac{1}{2n}t^2 + \frac{1}{6n^{3/2}}(it)^3\rho + n^{-3/2}o\left(t^3\right)\right) \\
&= n\left(-\frac{1}{2n}t^2 + \frac{1}{6n^{3/2}}(it)^3\rho + n^{-3/2}o\left(t^3\right) + \dots\right) \\
&= -\frac{t^2}{2} + \frac{1}{6n^{1/2}}(it)^3\rho + n^{-1/2}o\left(t^3\right) + \dots
\end{aligned}$$

Thus,

$$\begin{aligned}
\psi_{\sqrt{n}\bar{X}}(t) = \psi_X^n(t/\sqrt{n}) &= e^{-t^2/2}\exp\left\{\frac{1}{6n^{1/2}}(it)^3\rho + n^{-1/2}o\left(t^3\right)\right\} \\
&= e^{-t^2/2}\left(1 + \frac{1}{6n^{1/2}}(it)^3\rho + \dots\right) \\
&= e^{-t^2/2} + \frac{\rho}{6n^{1/2}}(it)^3 e^{-t^2/2} + \dots
\end{aligned}$$

Therefore, the "formal density" of $\sqrt{n}\bar{X}$ is

$$\begin{aligned}
f_{\sqrt{n}\bar{X}}(x) &= \frac{1}{2\pi}\int e^{-itx}\psi_{\sqrt{n}\bar{X}}(t)dt \\
&= \frac{1}{2\pi}\int e^{-itx}e^{-t^2/2}dt + \frac{\rho}{6n^{1/2}}\frac{1}{2\pi}\int e^{-itx}(it)^3 e^{-t^2/2}dt + \dots \\
&= \phi(x) + \frac{\rho}{6n^{1/2}}H_3(x)\phi(x) + \dots
\end{aligned}$$

Finally, we integrate the "density" to get the d.f.

$$\begin{aligned}
P(\sqrt{n}X \le x) &= \int_{-\infty}^{x}\left(\phi(x) + \frac{\rho}{6n^{1/2}}H_3(x)\phi(x) + \dots\right)dx \\
&= \Phi(x) - \frac{\rho}{6n^{1/2}}H_2(x)\phi(x) + \dots
\end{aligned}$$

where the last line follows since

$$\begin{aligned}
\tfrac{d}{dx}H_2(x)\phi(x) = H_2'(x)\phi(x) + H_2(x)(-x)\phi(x) &= 2x\phi(x) + \left(-x^3 + x\right)\phi(x) \\
&= -\left(x^3 - 3x\right)\phi(x) = -H_3(x)\phi(x)
\end{aligned}$$

Hermite polynomials

It is known that the standard normal r.v. has the c.f. $e^{-t^2/2}$, i.e.,

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-t^2/2}e^{-itx}dt$$

By repeated differentiation w.r.t. x, we get the identity

$$\phi^{(k)}(x) = (-1)^k H_k(x)\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(-it)^k e^{-t^2/2}e^{-itx}dt, \quad k \ge 0 \tag{3.62}$$

where H_k is a polynomial of degree k, and it is even or odd whenever k is even or odd. The H_k are called **Hermite polynomials**. It is easy to check that the first few Hermite polynomials are

$$
\begin{aligned}
H_1(x) &= x \\
H_2(x) &= x^2 - 1 \\
H_3(x) &= x^3 - 3x \\
H_4(x) &= x^4 - 6x^2 + 3 \\
H_5(x) &= x^5 - 10x^3 + 15x \\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15
\end{aligned}
$$

Finally from (3.62), we have, for $k \geq 0$

$$
\begin{aligned}
&H_k(x)\phi(x) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty}(it)^k e^{-t^2/2}e^{-itx}dt \\
&(it)^k e^{-t^2/2} = \int_{-\infty}^{\infty} H_k(x)\phi(x)e^{itx}dt
\end{aligned}
\tag{3.63}
$$

That is, the Fourier transform of $H_k(x)\phi(x)$ is $(it)^k e^{-t^2/2}$

Relationship between moments and cumulants

Recall the c.f. of X is defined as $\psi(t) = Ee^{itX}$. The relationship between the moments and the derivatives of the c.f. (when both exist) is given by

$$
\psi^{(k)}(0) = i^k EX^k, \quad k = 0, 1, 2, \ldots
$$

Define $\kappa(t) = \ln\psi(t) = \ln Ee^{itX}$ to be the cumulant generating function, and define the cumulants κ_k 's by

$$
\kappa^{(k)}(0) = i^k \kappa_k \quad k = 0, 1, 2, \ldots
$$

The moments and cumulants are related. For instance,

$$
\begin{aligned}
\mu_1 &= EX^1 = \kappa_1 \\
\mu_2 &= EX^2 = \kappa_2 + \kappa_1^2 \\
\mu_3 &= EX^3 = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 \\
\mu_4 &= EX^4 = \kappa_4 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4
\end{aligned}
$$

and

$$
\begin{aligned}
\kappa_1 &= \mu_1 = EX_1 \\
\kappa_2 &= \mu_2 - \mu_1^2 = E(X - \mu_1)^2 \\
\kappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 = E(X - \mu_1)^3 \\
\kappa_4 &= \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 \\
&= E(X - \mu_1)^4 - 3\left(E(X - \mu_1)^2\right)^2
\end{aligned}
$$

Edgeworth expansions

Berry-Esseen bounds state that the error incurred by approximating the d.f. of a normalized sum by a standard normal d.f. is of size $O\left(n^{-1/2}\right)$. This suggests that it is possible to incorporate the error term and obtain a more accurate approximation under some restrictions. For simplicity, we shall only concentrate on the i.i.d. case here.

Theorem 3.11.17 — Edgeworth expansion for i.i.d. r.v.'s. Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s from a d.f. F with $EX_1 = 0, EX_1^2 = \sigma^2 > 0$, and finite third moment $\mu_3 = EX_1^3$. Suppose that F is a nonlattice d.f.,then

$$\sup_{x \in R} \left| F_n(x) - \Phi(x) + \frac{\mu_3}{6\sigma^3\sqrt{n}} H_2(x)\phi(x) \right| = o\left(n^{-1/2}\right)$$

Proof. In the Smoothing Lemma, we put

$$F(x) = F_n(x), \quad \text{and} \quad G(x) = \Phi(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}} H_2(x)\phi(x)$$

The derivative of $G(x)$ is

$$G'(x) = \phi(x) + \frac{\mu_3}{6\sigma^3\sqrt{n}} H_3(x)\phi(x)$$

whose Fourier transform is

$$\psi_G(t) = \int_{-\infty}^{\infty} \left(\phi(x) + \frac{\mu_3}{6\sigma^3\sqrt{n}} H_3(x)\phi(x) \right) e^{itx} dx = e^{-t^2/2} \left(1 + \frac{\mu_3}{6\sigma^3\sqrt{n}} (it)^3 \right)$$

First fix $\epsilon > 0$. We then choose $T = a\sqrt{n}$ in the Smoothing lemma, where the constant a is chosen so large that $24 \sup_x |G'(x)| < \epsilon a$, then we have

$$\Delta := \sup_x |F_n(x) - G(x)| \leq \frac{1}{\pi} \int_{-a\sqrt{n}}^{a\sqrt{n}} |t|^{-1} \left| \psi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - \psi_G(t) \right| dt + \frac{\epsilon}{\sqrt{n}} \tag{3.64}$$

We partition the interval of integration into two parts:

1. $\delta \leq \frac{|t|}{\sqrt{n}\sigma} \leq a/\sigma$
2. $\frac{|t|}{\sqrt{n}\sigma} \leq \delta$

where $\delta > 0$ is arbitrary, but fixed.

(i) We first consider $\delta \leq |t|/(\sqrt{n}\sigma) \leq a/\sigma$. Since F is nonlattice, the maximum of $|\psi(t_n)| < q < 1$ for $\delta \leq |t_n| \leq a/\sigma$. Therefore, the contribution of the intervals $\delta \leq |t_n| \leq a/\sigma$ to the integral in (3.64) is

$$\leq \int_{\delta\sigma\sqrt{n}}^{a\sqrt{n}} t^{-1} \left[q^n + e^{-t^2/2} \left(1 + \left| \frac{\mu_3 t^3}{\sigma^3} \right| \right) \right] dt$$

and this tends to zero more rapidly than any power of $1/n$.

(ii) Secondly, let us consider $|t| \leq \delta\sigma\sqrt{n}$. Note that we can rewrite Δ as

$$\Delta \leq \int_{-a\sqrt{n}}^{a\sqrt{n}} |t|^{-1} e^{-t^2/2} \left| \exp\left\{ n\xi\left(\frac{t}{\sigma\sqrt{n}}\right) \right\} - \left[1 + \frac{\mu_3}{6\sigma^3\sqrt{n}} (it)^3 \right] \right| dt + \frac{\epsilon}{\sqrt{n}} \tag{3.65}$$

where

$$\xi(t) = \ln \psi(t) + \frac{\sigma^2 t^2}{2}$$

The integrand will be estimated using the following inequality: for any complex α and β,

$$\begin{aligned} |e^\alpha - 1 - \beta| &\le \left|e^\alpha - e^\beta\right| + \left|e^\beta - 1 - \beta\right| \\ &\le \max\left\{e^{|\alpha|}, e^{|\beta|}\right\}\left(|\alpha - \beta| + \frac{1}{2}\beta^2\right) \end{aligned} \tag{3.66}$$

To see this, letting $\gamma = \max\{|\alpha|, |\beta|\}$, then

$$\begin{aligned} \left|e^\alpha - e^\beta\right| &= \left|\left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots\right) - \left(1 + \beta + \frac{\beta^2}{2!} + \frac{\beta^3}{3!} + \dots\right)\right| \\ &= \left|(\alpha - \beta) + \frac{1}{2!}\left(\alpha^2 - \beta^2\right) + \frac{1}{3!}\left(\alpha^3 - \beta^3\right) + \dots\right| \\ &= |\alpha - \beta|\left|1 + \frac{1}{2!}(\alpha + \beta) + \frac{1}{3!}\left(\alpha^2 + \alpha\beta + \beta^2\right) + \dots\right| \\ &\le |\alpha - \beta|\left(1 + \frac{1}{2!}(2\gamma) + \frac{1}{3!}\left(3\gamma^2\right) + \dots\right) \\ &= |\alpha - \beta|\left(1 + \gamma + \frac{1}{2!}\gamma^2 + \dots\right) \\ &= |\alpha - \beta| e^\gamma \end{aligned}$$

and

$$\left|e^\beta - 1 - \beta\right| = \left|\frac{\beta^2}{2!} + \frac{\beta^3}{3!} + \dots\right| \le \frac{\beta^2}{2}\left(1 + |\beta| + \frac{|\beta|^2}{2!} + \frac{|\beta|^3}{3!} + \dots\right) \le \frac{\beta^2}{2} e^{|\beta|}$$

Since $E|X_1|^3 < \infty$, the function $\xi(t)$ is three times differentiable and $\xi(0) = \xi'(0) = \xi''(0) = 0$ while $\xi'''(0) = i^3\mu_3$, and $\xi'''(t)$ is continuous. Thus by the three-term Taylor expansion,

$$\xi(t) = \xi(0) + \xi'(0)t + \frac{1}{2}\xi''(0)t^2 + \frac{1}{6}\xi'''(0)t^3 + o\left(t^3\right) = \frac{1}{6}(it)^3\mu_3 + o\left(t^3\right)$$

We conclude that, there exists some small $\delta > 0$ such that

$$\left|\xi(t) - \frac{1}{6}\mu_3(it)^3\right| < \epsilon\sigma^3|t|^3 \quad \text{for } |t| < \delta$$

Here we can choose δ small enough so that

$$|\xi(t)| < \frac{1}{4}\sigma^2 t^2, \quad \left|\frac{1}{6}\mu_3(it)^3\right| < \frac{1}{4}\sigma^2 t^2$$

for $|t| < \delta$

$\alpha = n\xi\left(\frac{t}{\sigma\sqrt{n}}\right), \beta = \frac{\mu_3}{6\sigma^3\sqrt{n}}(it)^3, \xi(t) = \frac{1}{6}(it)^3\mu_3,$

$$\begin{aligned}
|\alpha - \beta| &= \left| n\frac{1}{6}i^3\left(\frac{t}{\sqrt{n}\sigma}\right)^3 \mu_3 - \frac{\mu_3}{6\sigma^3\sqrt{n}}(it)^3 \right| \\
&= \frac{1}{\sigma^3\sqrt{n}} \cdot \left| \xi(t) - \frac{1}{6}\mu_3(it)^3 \right| \\
&< \frac{1}{\sigma^3\sqrt{n}} \cdot \epsilon\sigma^3|t|^3
\end{aligned}$$

Hence, using (3.66), the integrand in (3.65) is

$$\leq |t|^{-1}e^{-t^2/4}\left(\frac{\epsilon|t|^3}{\sqrt{n}} + \frac{\mu_3^2}{72\sigma^6 n}t^6\right)$$

Therefore, the contribution of the intervals $|t| \leq \delta\sigma\sqrt{n}$ to the integral in (3.65) is $o\left(n^{-1/2}\right)$ since ϵ is arbitrary. Therefore, we have shown that $\Delta = o\left(n^{-1/2}\right)$. ■

The argument in the above proof breaks down for lattice distributions because their c.f. are periodic and so the contribution of $|t| > \delta\sigma\sqrt{n}$ does not tend to zero. In fact, for a lattice d.f., it can be shown from the inversion formula that the largest jump of F_n is of the order of magnitude $n^{-1/2}$, and hence Theorem 3.11.17 can not be true of any lattice distribution. For a special example, see the remarks in Section 11.2

Theorem 3.11.18 — Higher-order Edgeworth expansion for i.i.d. r.v.'s. Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s from a d.f. F with $EX_1 = 0, EX_1^2 = \sigma^2 > 0$, and finite r -th moment $\mu_r = EX_1^r$ with $r \geq 3$. Furthermore, assume that

$$\limsup_{|t|\to\infty} |\psi(t)| < 1$$

(Cramer's condition) then we have

$$\sup_{x\in R}\left| F_n(x) - \Phi(x) - \sum_{k=3}^{r} n^{-(k-2)/2}R_k(x)\phi(x) \right| = o\left(n^{-(r-2)/2}\right)$$

where $R_k(x)$ is a polynomial depending only on population moments of order $\leq r$, but not on n and r.

Proof. The proof of this theorem is very similar to that of Theorem 3.11.17. In the Smoothing Lemma, we put $F(x) = F_n(x)$ and

$$G(x) = \Phi(x) + \sum_{k=3}^{r} n^{-(k-2)/2}R_k(x)\phi(x)$$

whose derivative has Fourier transform

$$\psi_G(t) = e^{-t^2/2}\left[1 + \sum_{j=3}^{r} \frac{\kappa_j(it)^j}{\sigma^j j!} n^{-(j-2)/2}\right]$$

where κ_j are the cumulants of order j. Choosing $T = an^{(r-2)/2}$ in the Smoothing Lemma, where the constant a is chosen so large that $24\sup_x |G'(x)| < \epsilon a$, then we have

$$\Delta := \sup_x |F_n(x) - G(x)| \leq \frac{1}{\pi}\int_{-T}^{T} \left| \frac{\psi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - \psi_G(t)}{t} \right| dt + \frac{\epsilon}{n^{(r-2)/2}} \tag{3.67}$$

We partition the interval of integration into two parts: $|t| \leq \delta\sigma\sqrt{n}$ and $\delta\sigma\sqrt{n} \leq |t| \leq T$, where $\delta > 0$ is arbitrary, but fixed. First, consider $\delta\sigma\sqrt{n} \leq |t| \leq T$. From the Cramer's condition, the maximum of $|\psi(t)| < q < 1$ for $|t| \geq \delta$. Therefore, the contribution of the intervals $\delta\sigma\sqrt{n} \leq |t| \leq T$ to the integral in (3.67) is

$$\leq \int_{\delta\sigma\sqrt{n}}^{T} t^{-1}\left[q^n + e^{-t^2/2}\left(1 + \left|\sum_{j=3}^{r} \frac{\kappa_j t^j}{\sigma^j j!} n^{-(j-2)/2}\right|\right)\right] dt$$

and this tends to zero more ripidly than any power of $1/n$. Second, consider $|t| \leq \delta\sigma\sqrt{n}$. Let

$$\xi(t) = \ln\psi(t) + \frac{\sigma^2 t^2}{2}, \quad v^*(t) = 1 + \sum_{j=3}^{r} \frac{\kappa_j (it)^j}{\sigma^j j!}$$

Then we can rewrite Δ as

$$\Delta \leq \int_{-T}^{T} |t|^{-1} e^{-t^2/2} \left| \exp\left\{ n\xi\left(\frac{t}{\sigma\sqrt{n}}\right)\right\} - \left[1 + nv^*\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\right| dt + \frac{\epsilon}{n^{(r-2)/2}} \tag{3.68}$$

The integrand will be estimated using the following inequality: for any complex α and β,

$$\begin{aligned} \left| e^\alpha - 1 - \beta - \ldots - \frac{\beta^m}{m!} \right| &\leq \left| e^\alpha - e^\beta \right| + \left| e^\beta - 1 - \beta - \ldots - \frac{\beta^m}{m!} \right| \\ &\leq \max\left\{ e^{|\alpha|}, e^{|\beta|} \right\} \left(|\alpha - \beta| + \frac{\beta^{m+1}}{(m+1)!} \right) \end{aligned} \tag{3.69}$$

Since $E|X_1|^r < \infty$, the function $\xi(t) - v^*(t)$ has j continuous derivatives at 0 and it is easy to check that all of them are 0. Thus by the r -term Taylor expansion, we conclude that

$$|\xi(t) - v^*(t)| < \epsilon\sigma^r |t|^r \quad \text{for } |t| < \delta$$

We can further choose δ small enough so that

$$|\xi(t)| < \frac{1}{4}\sigma^2 t^2, \quad |v^*(t)| < C\sigma^3 |t|^3 \quad \text{for } |t| < \delta$$

Thus, if $|t| \leq \delta\sigma\sqrt{n}$, we have

$$\left| n\xi\left(\frac{t}{\sqrt{n}\sigma}\right) - nv^*\left(\frac{t}{\sqrt{n}\sigma}\right) \right| < \epsilon |t|^r n^{-(r-2)/2}$$

and

$$\left|n\xi\left(\frac{t}{\sqrt{n}\sigma}\right)\right| \le \frac{t^2}{4}, \quad \left|nv^*\left(\frac{t}{\sqrt{n}\sigma}\right)\right| \le C\frac{|t|^3}{\sqrt{n}}$$

Hence, using (3.69), the integrand in (3.68) is

$$\le |t|^{-1}e^{-t^2/4}\left(\frac{\epsilon|t|^3}{\sqrt{n}} + \frac{\mu_3^2}{72\sigma^6 n}t^6\right)$$

Therefore, the contribution of the intervals $|t| \le \delta\sigma\sqrt{n}$ to the integral in (3.68) is $o\left(n^{-1/2}\right)$ since ϵ is arbitrary. Therefore, we have shown that $\Delta = o\left(n^{-1/2}\right)$.

∎

Theorem 3.11.19 — One-term Edgeworth expansion for independent r.v.'s. Let $X_1, \ldots, X_n$ be independent r.v.'s with c.f. ψ_j and $EX_j = 0, EX_j^2 = \sigma_j^2 > 0, (1 \le j \le n)$. Let $B_n^2 = \sum_{j=1}^n \sigma_j^2$ and $\mu_{3n} = \sum_{j=1}^n EX_j^3$. Further assume that

1. $n^{-1}\sum_{j=1}^n E\left|X_j\right|^3 < C_0$ where $0 < C_0 < \infty$
2. $cn < B_n^2 < Cn$, where $0 < c < C < \infty$
3. $\left|\prod_{k=1}^n \psi_k(t)\right| = o\left(n^{-1/2}\right)$ uniformly in $|t| > \delta > 0$

Then

$$\sup_{x\in R}\left|F_n(x) - \Phi(x) + \frac{\mu_3 n}{6B_n^3}H_2(x)\phi(x)\right| = o\left(n^{-1/2}\right)$$

Proof. In the Smoothing lemma, we put $F(x) = F_n(x)$ and

$$G(x) = \Phi(x) - \frac{\mu_{3n}}{6B_n^3}H_2(x)\phi(x)$$

The derivative of $G(x)$ is

$$G'(x) = \phi(x) + \frac{\mu_{3n}}{6B_n^3}H_2(x)\phi(x)$$

whose Fourier transform is

$$\psi_G(t) = e^{-t^2/2}\left(1 + \frac{\mu_{3n}}{6B_n^3}(it)^3\right)$$

Choosing $T = a\sqrt{n}$ in the Smoothing lemma, where the constant a is chosen so large that $24\sup_x |G'(x)| < \epsilon a$, then we have

$$\Delta := \sup_x |F_n(x) - G(x)|$$
$$\le \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\prod_{k=1}^n \psi_k\left(\frac{t}{B_n}\right) - \psi_G(t)}{t}\right|dt + \frac{\epsilon}{\sqrt{n}} \tag{3.70}$$

We partition the interval of integration into two parts: $|t| \le \delta B_n$ and $\delta B_n \le |t| \le T$, where $\delta > 0$ is arbitrary, but fixed. First, consider $\delta B_n \le |t| \le T$. From the assumption, we see that the contribution of the intervals $\delta B_n \le |t| \le T$ to the integral in (3.70) is

$$\begin{aligned}
&\le \textstyle\int_{\delta B_n}^{T} t^{-1} \left|\prod_{k=1}^{n} \psi_k(t)\right| dt + \int_{\delta B_n}^{T} t^{-1} e^{-t^2/2} \left(1 + \left|\frac{\mu_{3n}}{6B_n^3}\right|\right) dt \\
&\le o\left(n^{-1/2}\right) \textstyle\int_{\delta B_n}^{T} t^{-1} dt + o\left(n^{-1/2}\right) \\
&\le o\left(n^{-1/2}\right) \ln\left(T/\delta B_n\right) + o\left(n^{-1/2}\right) \\
&= o\left(n^{-1/2}\right)
\end{aligned}$$

Second, consider $|t| \le \delta B_n$. Note that we can rewrite Δ as

$$\Delta \le \int_{-T}^{T} |t|^{-1} e^{-t^2/2} \left| \exp\left\{ \xi\left(\frac{t}{B_n}\right) \right\} - \left[1 + \frac{\mu_{3n}}{6}\left(\frac{it}{B_n}\right)^3\right] \right| dt + \frac{\epsilon}{\sqrt{n}} \tag{3.71}$$

where

$$\xi(t) = \sum_{k=1}^{n} \ln \psi_k(t) + \frac{1}{2} B_n^2 t^2$$

The integrand will be estimated using the following inequality: for any complex α and β,

$$\begin{aligned}
\left|e^{\alpha} - 1 - \beta\right| &\le \left|e^{\alpha} - e^{\beta}\right| + \left|e^{\beta} - 1 - \beta\right| \\
&\le \max\left\{e^{|\alpha|}, e^{|\beta|}\right\} \left(|\alpha - \beta| + \frac{1}{2}\beta^2\right)
\end{aligned} \tag{3.72}$$

since $E\left|X_k\right|^3 < \infty$, the function $\xi(t)$ is three times differentiable and $\xi(0) = \xi'(0) = \xi''(0) = 0$ while $\xi'''(0) = i^3 \mu_{3n}/n$, and $\xi'''(t)$ is continuous. Thus by the three-term Taylor expansion, we conclude that

$$\left|\xi(t) - \frac{\mu_{3n}}{6n}(it)^3\right| < \frac{\epsilon B_n^3 |t|^3}{n^{3/2}} \quad \text{for } |t| < \delta$$

Here we can choose δ small enough so that

$$|\xi(t)| < \frac{B_n^2 t^2}{4n}, \quad \left|\frac{\mu_3 n}{6n}(it)^3\right| < \frac{B_n^2 t^2}{4n} \quad \text{for } |t| < \delta$$

Hence, using (3.72), the integrand in (3.71) is

$$\le |t|^{-1} e^{-t^2/4} \left(\frac{\epsilon |t|^3}{\sqrt{n}} + \frac{\mu_3^2}{72\sigma^6 n} t^6\right)$$

Therefore, the contribution of the intervals $|t| \le \delta\sigma\sqrt{n}$ to the integral in (3.71) is $o\left(n^{-1/2}\right)$ since ϵ is arbitrary. Therefore, we have shown that $\Delta = o\left(n^{-1/2}\right)$. ■

Theorem 3.11.20 — Higher-order Edgeworth expansion for independent r.v.'s. Let $X_1, \ldots, X_n$ be independent r.v.'s with c.f. ψ_j. Assume

1. (moment conditions)

$$C_1 < n^{-1} \sum_{j=1}^{n} EX_j^2 < C_2, \quad n^{-1} \sum_{j=1}^{n} E\left|X_j\right|^r < C_1, \quad (r \ge 3)$$

2. (smoothness condition)

$$\left|\prod_{k=1}^{n} \psi_k(t)\right| = o\left(n^{-(r-2)/2}\right) \qquad \text{uniformly in ??????} > |t| > \delta > 0$$

Then,

$$\sup_{x \in R}\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| = o\left(n^{-(r-2)/2}\right)$$

In the i.i.d. case, it reduces to the following theorem.

Theorem 3.11.21 — Edgeworth expansion for i.i.d. r.v.'s. Let $X, X_1, \ldots, X_n$ be i.i.d. r.v.'s with c.f. ψ. Assume

1. (moment conditions:) $E\left|X_j\right|^r < C, \quad (r \geq 3)$
2. (smoothness condition)
   (a) The d.f. of X is nonlattice if $r = 3$
   (b) $\limsup_{|t| \to \infty} |\psi(t)| < 1 \quad$ if $r \geq 3$

Then,

$$\sup_{x \in R}\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_n(x)\phi(x)\right| = o\left(n^{-(r-2)/2}\right)$$

3.11.5 Non-uniform Edgeworth expansions

Theorem 3.11.22 — Non-uniform Edgeworth expansion for i.i.d. r.v.'s. Let $X, X_1, \ldots, X_n$ be i.i.d. r.v.'s with mean 0, variance $\sigma^2 > 0$, and c.f. $\psi(t)$. If $E|X|^r < \infty$ for some integer $k \geq 3$, then for all n and x

$$\begin{aligned}
&\leq \left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| \\
&\leq C_r \frac{E|X|^r I\{|X| \geq \sigma\sqrt{n}(1+|x|)\}}{\sigma^r(1+|x|)^r n^{(r-2)/2}} \\
&\quad + C_r \frac{E|X|^{r+1} I\{|X| < \sigma\sqrt{n}(1+|x|)\}}{\sigma^{r+1}(1+|x|)^{r+1} n^{(r-1)/2}} \\
&\quad + C_r \left(\sup_{|t| \geq \delta} |\psi(t)| + \frac{1}{2n}\right)^n \frac{n^{r(r+1)/2}}{(1+|x|)^{r+1}}
\end{aligned}$$

where $\delta = \frac{\sigma^2}{12E|X|^3}$ and $C_r > 0$ is a constant depending only on r.

If $\limsup_{|t| \to \infty} |\psi(t)| < 1$, then for any $\delta > 0, \sup_{|t| \geq \delta} |\psi(t)| < 1$, so that the factor $\left(\sup_{|t| \geq \delta} |\psi(t)| + \frac{1}{2n}\right)^n$ decreases to 0 faster than n^{-a} for any $a > 0$. Some consequences of the last theorem are given below.

Theorem 3.11.23 — Non-uniform Edgeworth expansion for i.i.d. r.v.'s. Suppose that $\limsup_{|t|\to\infty}|\psi(t)| < 1$ and $E|X|^r < \infty$ for some $k \geq 3$, then there exists a positive function $\epsilon(u)$ such that $\lim_{u\to\infty} = 0$ and

$$\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| \leq \frac{\epsilon(\sqrt{n}(1+|x|))}{n^{(r-2)/2}(1+|x|)^r}$$

Theorem 3.11.24 — Non-uniform Edgeworth expansion for i.i.d. r.v.'s. Suppose that $\limsup_{|t|\to\infty}|\psi(t)| < 1$ and $E|X|^r < \infty$ for some $k \geq 3$, then there exists a positive function $\epsilon(u)$ such that $\lim_{u\to\infty} = 0$ and

$$(1+|x|)^r \left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| = o\left(\frac{1}{n^{(r-2)/2}}\right)$$

3.11.6 Large Deviations

All the limit theorems and expansions derived so far deal with absolute errors. Although they are useful for moderate values of x, they are less meaningful for large x. For instance, CLT states that

$$\sup_{x\in R} |F_n(x) - \Phi(x)| \to 0$$

However, for large $|x|$, both $F_n(x)$ and $\Phi(x)$ are either close to 1 or 0, therefore, the statement of CLT becomes empty. In this section, we will look at the tail probability $1 - F_n(x) = P(\sqrt{n}\bar{X} > x)$ as $x =: x_n \to \infty$. For simplicity, we shall assume that Cramer condition (to be given later) holds.

Cramer conditioin

Let $X, X_1, \ldots, X_n$ be i.i.d. r.v.'s. Let the following Cramer's condition hold:

$$Ee^{tX} < \infty, \quad \text{in} |t| < H \text{ for some constant } H > 0.$$

Cramer's condition simply means that the moment generating function exists near the origin, and implies that moments of all orders exist. Several equivalent forms are given below.

Lemma 3.29 The following assertions are equivalent:

1. $Ee^{tX} < \infty \quad$ in $|t| < H$ for some constant $H > 0$
2. $Ee^{a|X|} < \infty$ for some constant $a > 0$

3. $P(|X| \geq x) \leq be^{-cx}$ for some constants $b, c > 0$ and all $x > 0$

Proof. 1. "(1) $\to$ (2)". This follows from

$$Ee^{a|X|} = \int_{-\infty}^{0} e^{-ax} dF(x) + \int_{0}^{\infty} e^{ax} dF(x) \leq Ee^{aX} + Ee^{-aX} < \infty$$

2. "(2) $\to$ (1)". This follows from $e^{tX} \leq e^{t|X|}$
3. "(2) $\to$ (3)". This follows from

$$P(|X| \geq x) = P(a|X| \geq ax) \leq e^{-ax} Ee^{a|X|}, \quad x, a > 0$$

4. "(3) $\to$ (2)". Note that

$$\begin{aligned} Ee^{a|X|} &= Ee^{-aX} I\{X < 0\} + Ee^{aX} I\{X \geq 0\} \\ &= \int_{-\infty}^{0} e^{-ax} dF(x) - \int_{0}^{\infty} e^{ax} d(1 - F(x)) \\ &= I_1 + I_2 \end{aligned}$$

Now by integration by parts, for $0 \leq a < c$

$$\begin{aligned} I_2 &= -\left[e^{ax}(1 - F(x))\right]_0^{\infty} + \int_0^{\infty} (1 - F(x)) de^{ax} \\ &\leq [1 - F(0)] - \lim_{t \to \infty} be^{(a-c)t} + a \int_0^{\infty} be^{(a-c)x} dx \\ &< \infty \end{aligned}$$

Similarly, we can show that $I_2 < \infty$.

■

Part (3) of the above lemma states that the tail probability of X decreases to zero exponentially fast under the Cramer's condition.

Conjugate (or associated) distributions

One of the key tools in large deviation theory is the so-called conjugate (or associated) distribution. Let $X, X_1, \ldots, X_n$ be i.i.d. r.v.'s with a common d.f. F. Denote the moment generating function (m.g.f.) and cumulant generating function (c.g.f.) of X by

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tX} dF(x), \quad K_X(t) = \ln M_X(t)$$

Definition 3.11.1 Given a d.f. F, define its conjugate (or associated) distribution by

$$G(y) = \frac{\int_{-\infty}^{y} e^{sx} dF(x)}{\int_{-\infty}^{\infty} e^{sx} dF(x)} = \frac{\int_{-\infty}^{y} e^{sx} dF(x)}{M_X(t)} = \int_{-\infty}^{y} e^{sy - K_X(s)} dF(x)$$

or equivalently,

$$dG(x) = \frac{e^{sx}dF(x)}{\int_{-\infty}^{\infty} e^{sx}dF(x)} = \frac{e^{sx}dF(x)}{M_X(s)} = e^{sx-K_X(s)}dF(x)$$

Lemma 3.30 It is easy to see that m.g.f.'s and c.g.f.'s of X and Y are related by

$$M_Y(t) = \frac{M_X(t+s)}{M_X(s)}, \quad K_Y(t) = K_X(t+s) - K_X(s)$$

In particular,

$$EY = K_Y'(t)\big|_{t=0} = K_X'(s), \quad \mathrm{Var}(Y) = K_Y''(t)\big|_{t=0} = K_X''(s)$$

R After a change of measure from F to G, the mean of X is changed from $E_F X = 0$ under F to $E_G X = K_X'(s)$ under G. By choosing different s, we can freely change the mean to any (allowable) location. In particular, in estimating $P(\bar{X} > x)$ for some large x, we can choose appropriate s so that the point x becomes the center of d.f. rather than the tail. The reason for doing this is that one can apply many nice results such as CLT, Edgewoth expansions, etc, which behave very nicely near the center of the d.f.

Lemma 3.31 Let $Y, Y_1, \ldots, Y_n$ be i.i.d. r.v.'s with a common d.f. G. Define

$$F_n(x) = P\left(\sum_{i=1}^n X_i \le x\right), \quad G_n(y) = P\left(\sum_{i=1}^n Y_i \le y\right)$$

Then

$$G_n(y) = \int_{-\infty}^{y} e^{sx-nK_X(s)}dF_n(x) = \frac{\int_{-\infty}^{y} e^{sx}dF_n(x)}{\left(e^{K_X(s)}\right)^n} = \frac{\int_{-\infty}^{y} e^{sx}dF_n(x)}{M_X^n(s)} = \frac{\int_{-\infty}^{y} e^{sx}dF_n(x)}{\int_{-\infty}^{\infty} e^{sx}dF_n(x)}$$

or equivalently,

$$dG_n(x) = e^{sx-nK_X(s)}dF_n(x) = \frac{e^{sx}dF_n(x)}{\left(e^{K_X(s)}\right)^n} = \frac{e^{sx}dF_n(x)}{M_X^n(s)} = \frac{e^{sx}dF_n(x)}{\int_{-\infty}^{\infty} e^{sx}dF_n(x)}$$

Proof. The LHS has m.g.f.

$$M_{\sum_{i=1}^n Y_i}(t) = E\exp\left(t\sum_{i=1}^n Y_i\right) = M_Y^n(t) = \left(\frac{M_X(t+s)}{M_X(s)}\right)^n = \frac{M_{\sum_{i=1}^n X_i}(t+s)}{M_{\sum_{i=1}^n X_i}(s)}$$

The RHS has m.g.f.

$$\int_{-\infty}^{\infty} e^{tx}\frac{e^{sx}dF_n(x)}{\int_{-\infty}^{\infty} e^{sx}dF_n(x)} = \frac{\int_{-\infty}^{\infty} e^{(t+s)x}dF_n(x)}{\int_{-\infty}^{\infty} e^{sx}dF_n(x)} = \frac{M\sum_{i=1}^n x_i(t+s)}{M\sum_{i=1}^n X_i(s)}$$

By the uniqueness theorem, we hence proved the lemma.

■

The following lemma is critical in deriving large deviation results.

Lemma 3.32 If we choose τ such that $K'_X(\tau) = z$, then we have

$$
\begin{aligned}
P(\overline{X} > z) &= e^{n[K_X(\tau)-\tau K'_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}dP(T_n \le y)} \\
&= e^{-n[z\tau - K_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}dP(T_n \le y)} \\
&= e^{-n\sup_{t>0}[zt-K_X(t)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}dP(T_n \le y)}
\end{aligned}
$$

where

$$
T_n = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nVar(Y)}} = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nK''_X(s)}}
$$

Proof.

$$
\begin{aligned}
P(\overline{X} > z) &= P\left(\sum_{i=1}^n X_i > nz\right) \\
&= \int_{nz}^\infty dP\left(\sum_{i=1}^n X_i \le x\right) \\
&= \int_{nz}^\infty e^{-sx+nK_X(s)} dP\left(\sum_{i=1}^n Y_i \le x\right) \\
&= e^{nK_X(s)} \int_{nz}^\infty e^{-sx} dP\left(\frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nVar(Y)}} \le \frac{x - nK'_X(s)}{\sqrt{nK''_X(s)}}\right) \\
&= e^{nK_X(s)} \int_{\frac{\sqrt{n}[z-K'_X(s)]}{\sqrt{K''_X(s)}}}^\infty e^{-s\left(nK'_X(s)+y\sqrt{nK''_X(s)}\right)} dP\left(T_n \le y\right) \\
&= e^{n[K_X(s)-sK'_X(s)]} \int_{\frac{\sqrt{n}[z-K'_X(s)]}{\sqrt{K''_X(s)}}}^\infty e^{-sy\sqrt{nK''_X(s)}} dP\left(T_n \le y\right)
\end{aligned}
$$

where

$$
T_n = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{n\mathrm{Var}(Y)}}, \quad y = \frac{x - nK'_X(s)}{\sqrt{nK''_X(s)}}
$$

If we choose τ such that $K'_X(\tau) = z$, then we have

$$
P(\overline{X} > z) = e^{n[K_X(\tau)-sK'_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right)
$$

■

Large Deviations

From Lemma 3.29, under Cramer's condition, $P(S_n > na)$ converges to 0 exponentially fast. More careful analysis can give a precise convergence rate.

Theorem 3.11.25 Let $X_1, X_2, \ldots$ be i.i.d. with mean 0 and m.g.f. $M(t) = Ee^{tX} < \infty$ for $|t| < \delta$, and $K(t) = \log M(t)$. Then

$$\lim_n P(S_n > na)^{1/n} = \lim_n P(\overline{X} > a)^{1/n} = e^{-\sup_{t>0}[at-K(t)]} =: e^{-\eta(a)}$$

or equivalently,

$$-\lim_n \frac{1}{n} \ln P(S_n > na) = \sup_{t>0}[at - K(t)] =: \eta(a)$$

where $\eta(a) = \sup_{t>0}[at - K(t)]$.

R From the theorem, the bound is interesting if $\eta(a) > 0$. It suffices to show that $at - K(t) > 0$ for some $t > 0$. To prove this, we note

$$\begin{aligned} at - K(t) &= at - \left[K(0) + K'(0)t + \frac{1}{2}K''(0)t^2 + \ldots\right] \\ &= at - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right) = at\left(1 - \frac{1}{2}\sigma^2 t + o(t)\right) > 0 \end{aligned}$$

when $t > 0$ is chosen to be small enough.

Proof. We first give an upper bound. For any $t > 0$, by Chebyshev inequality, we have

$$P(S_n > na)^{1/n} = P\left(e^{tS_n} > e^{nat}\right)^{1/n} \le \left(\frac{Ee^{tS_n}}{e^{nat}}\right)^{1/n} = \left(\frac{\left(Ee^{tX_1}\right)^n}{e^{nat}}\right)^{1/n} = \frac{M(t)}{e^{at}} = e^{K(t)-at}$$

Take $\inf_{t>0}$ on both sides, we get

$$P(S_n > na)^{1/n} \le \inf_{t>0} e^{K(t)-at} = e^{\inf_{t>0}[K(t)-at]} = e^{-\sup_{t>0}[at-K(t)]}$$

Next we will give a lower bound. From the last lemma, we have

$$\begin{aligned} P(\overline{X} > a)^{1/n} &= e^{[K_X(\tau) - sK'_X(\tau)]}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP(T_n \le y)\right)^{1/n} \\ &= e^{-\psi(a)}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP(T_n \le y)\right)^{1/n} \\ &=: e^{-\psi(a)} A_n^{1/n} \end{aligned}$$

It suffices to show that $A_n^{1/n} \geq 1$ as $n \to \infty$. To do this, note that

$$
\begin{aligned}
A_n =& \int_0^\infty e^{-\sqrt{n}\tau_0 y} dP\left(T_n \leq y\right) \quad \left(\text{where } \tau_0 = \tau\sqrt{K_X''(\tau)}\right) \\
=:& \frac{\int_0^\infty e^{-\sqrt{n}\tau_0 y} dP\left(T_n \leq y\right)}{\int_0^\infty dP\left(T_n \leq y\right)} \int_0^\infty dP\left(T_n \leq y\right) \\
=& E\left(e^{-\sqrt{n}\tau_0 T_n} \mid T_n > 0\right) P\left(T_n > 0\right) \\
\geq& \exp\left(\left\{-\sqrt{n}\tau_0 E\left[T_n \mid T_n > 0\right]\right\}\right) P\left(T_n > 0\right), \quad \text{(by Jensen's inequality)}
\end{aligned}
$$

Now

$$
E\left[T_n \mid T_n > 0\right] = \frac{E\left[T_n I\left(T_n > 0\right)\right]}{P\left(T_n > 0\right)} \leq \frac{E\left|T_n\right|}{P\left(T_n > 0\right)} \leq \frac{\left(ET_n^2\right)^{1/2}}{P\left(T_n > 0\right)} \leq \frac{1}{P\left(T_n > 0\right)} \leq \frac{1}{0.5 - \epsilon}
$$

and therefore,

$$
A_n^{1/n} \geq \exp\left(\frac{-\sqrt{n}\tau_0}{nP\left(T_n > 0\right)}\right) P^{1/n}\left(T_n > 0\right) \longrightarrow 1
$$

since $P\left(T_n > 0\right) \to 1/2$ and $P^{1/n}\left(T_n > 0\right) \to 1$ ■

Cramer-type large deviations

We indicated at the beginning of this section that normal approximation is of limited use when we look into the far tail of the d.f. of standardized sums. The CLT only works well not too far away from the center of the distribution. The question is then how far can we actually safely use the CLT as we move away from the center of the distribution.

Since $1 - F_n(x)$ and $1 - \Phi(x)$ are both close to 0 as $x \to \infty$, using $|F_n(x) - \Phi(x)|$ as a measure of closeness of the two d.f.s may not be very helpful to us. A more useful measure in this case is to estimate the relative error in approximating $1 - F_n(x)$ by $1 - \Phi(x)$ when $x \to \infty$, or approximating $F_n(x)$ by $\Phi(x)$ when $x \to -\infty$. In other words, we would like to have

$$
\frac{1 - F_n(x)}{1 - \Phi(x)} \to 1, \quad \frac{F_n(-x)}{\Phi(-x)} \to 1 \tag{3.73}
$$

when both x and n tend to ∞. However, the statement (3.73) can not be true in general. For instance, if $X_1, \ldots, X_n$ are i.i.d. with $P\left(X_1 = 0\right) = P\left(X_1 = 1\right) = 1/2$, then $P\left(S_n > n\right) = 1 - F_n(\sqrt{n}) = 0$. Thus, for $|x| > \sqrt{n}$, the ratios in (3.73) are 0. But as we shall see next, the statement is true if x varies with n such that $x = x_n \to \infty$ at a certain rate.

Theorem 3.11.26

$$
\frac{1 - F_n(x)}{1 - \Phi(x)} = 1 + O\left(\frac{x^3}{\sqrt{n}}\right)
$$

Proof. Note that

$$
1 - F_n(x) = P(\sqrt{n}(\overline{X} - \mu)/\sigma > x) = P\left(\overline{X} > \mu + \frac{x\sigma}{\sqrt{n}}\right)
$$

Then from Lemma (3.32), we have

$$1 - F_n(x) = e^{n[K_X(\tau) - \tau K'_X(\tau)]} \int_0^\infty e^{-\tau y \sqrt{nK''_X(\tau)}} dP(T \le y) \tag{3.74}$$

where τ satisfies

$$K'_X(\tau) = \mu + \frac{x\sigma}{\sqrt{n}}$$

Let

$$\Delta(y) = P(T \le y) - \Phi(y)$$

We can approximate $P(T \le y)$ by the standard normal d.f., resulting in

$$\begin{aligned} 1 - F_n(x) &= e^{n[K_X(\tau) - \tau K'_X(\tau)]} \int_0^\infty e^{-\tau y \sqrt{nK''_X(\tau)}} d[\Phi(y) + \Delta(y)] \\ &= A_n + B_n \end{aligned}$$

First from

$$K_X(\tau) = \mu\tau + \frac{1}{2}\sigma^2\tau^2 + \frac{1}{6}EX^3\tau^3 + \ldots\ldots$$

we get

$$K'_X(\tau) = \mu + \sigma^2\tau + \frac{1}{2}EX^3\tau^2 + \ldots\ldots = \mu + \frac{x\sigma}{\sqrt{n}}$$

Thus

$$\tau \sim \frac{x}{\sigma\sqrt{n}} \quad \text{if } \frac{x}{\sqrt{n}} \to 0$$

Now

$$\begin{aligned} A_n &= e^{n[K_X(\tau) - \tau K'_X(\tau)]} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\tau y \sqrt{nK''_X(\tau)} - \frac{1}{2}y^2} dy \\ &= e^{n\left[K_X(\tau) - \tau K'_X(\tau) + \frac{1}{2}\tau^2 K''_X(\tau)\right]} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(y + \tau\sqrt{nK''_X(\tau)})^2} dy \\ &= e^{n\left[K_X(\tau) - \tau K'_X(\tau) + \frac{1}{2}\tau^2 K''_X(\tau)\right]} \left(1 - \Phi(\tau\sqrt{nK''_X(\tau)})\right) \end{aligned}$$

Denote the exponent in A_n by $u(\tau) = K_X(\tau) - \tau K'_X(\tau) + \frac{1}{2}\tau^2 K''_X(\tau)$. It is easy to check that $u(0) = u'(0) = u''(0) = 0$ and $u'''(0) = EX^3$. So its power series expansion starts with cubic terms. Thus

$$A_n = e^{nO(\tau^3)} \left(1 - \Phi(\tau\sqrt{nK''_X(\tau)})\right) = \left(1 - \Phi(\tau\sqrt{nK''_X(\tau)})\right) \left(1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right)$$

Next we put $x_\tau = \tau\sqrt{nK''_X(\tau)}$

$$\begin{aligned} \frac{x_\tau - x}{\sqrt{n}} &= \tau\sqrt{K''_X(\tau)} - \frac{K'_X(\tau)}{\sigma} \\ &= \sigma\tau\left(1 + \frac{(EX^3)}{2\sigma^2}\tau + O\left(\tau^2\right)\right) - \sigma\tau\left(1 + \frac{(EX^3)}{2\sigma^2}\tau + O\left(\tau^2\right)\right) \\ &= O\left(\tau^3\right) \end{aligned}$$

Thus,

$$x_\tau - x = O\left(\sqrt{n}\tau^3\right) = O\left(\frac{x^3}{n}\right)$$

Now

$$\begin{aligned}
\log\left(\frac{1-\Phi(x_\tau)}{1-\Phi(x)}\right) &= \log(1-\Phi(x_\tau)) - \log(1-\Phi(x)) \\
&= \frac{-\phi(x_0)}{1-\Phi(x_0)}(x_\tau - x) \quad \text{where } x_0 \text{ is between } x \text{ and } x_\tau \\
&\sim -x_0(x_\tau - x) = O\left(\frac{x^4}{n}\right)
\end{aligned}$$

Hence,

$$\frac{1-\Phi(x_\tau)}{1-\Phi(x)} = 1 + O\left(\frac{x^4}{n}\right)$$

Therefore, for $x = o(\sqrt{n})$

$$A_n = (1-\Phi(x))\left[1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right]$$

For B_n, by Berry-Esseen bounds,

$$\begin{aligned}
|B_n| &\le e^{n[K_X(\tau)-\tau K_X'(\tau)]}\int_0^\infty e^{-\tau y\sqrt{nK_X''(\tau)}}d|\Delta(y)| \\
&\le 2e^{n[K_X(\tau)-\tau K_X'(\tau)]}\sup_y|\Delta(y)| \\
&\le 4e^{-\frac{1}{2}n\sigma^2\tau^2+nO(\tau^3)}\frac{E|X|^3}{(EX^2)^{3/2}\sqrt{n}} \\
&\le C\phi(x)\left[1+O\left(\frac{x^3}{\sqrt{n}}\right)\right]\frac{E|X|^3}{(EX^2)^{3/2}\sqrt{n}} \\
&\sim C(1-\Phi(x))\left[1+O\left(\frac{x^3}{\sqrt{n}}\right)\right]\frac{x}{\sqrt{n}} \\
&\sim A_nO\left(\frac{x}{\sqrt{n}}\right)
\end{aligned}$$

Finally, we get

$$1 - F_n(x) = A_nO\left(1 + \frac{x}{\sqrt{n}}\right) = (1-\Phi(x))\left[1+O\left(\frac{x^3}{\sqrt{n}}\right)\right]$$

■

3.11.7 Saddlepoint Approximations

Saddlepoint Approximations can provide extremely accurate approximations for the tail probabilities where most of the alternatives fail. The approximations can work even for very small sample size n.

From our earlier calculations, we have

$$\begin{aligned} P(\overline{X} > x) &= e^{n[K_X(\tau) - \tau K_X'(\tau)]} \int_0^\infty e^{-\tau y \sqrt{nK_X''(\tau)}} dP(T \le y) \\ &= e^{-n[\tau x - K_X(\tau)]} \int_0^\infty e^{-\tau y \sqrt{nK_X''(\tau)}} dP(T \le y) \\ &= e^{-\hat{w}^2/2} \int_0^\infty e^{-\hat{z} y} dP(T \le y) \end{aligned}$$

where $K_X'(\tau) = x$ and

$$\hat{w} = \sqrt{2n\left[\tau x - K_X(\tau)\right]}\, \mathrm{sgn}\{\tau\}, \quad \hat{z} = \tau \sqrt{nK_X''(\tau)}$$

First order approximations.

By the CLT, we can use the normal distribution to $\Phi(y)$ to approximate $P(T \le y)$, resulting in

$$P(\overline{X} > x) = e^{-\hat{w}^2/2} \int_0^\infty e^{-\hat{z} y} d[\Phi(y) + \Delta(y)] =: A_n + B_n$$

where $\Delta(y) = P(T \le y) - \Phi(y)$. First

$$\begin{aligned} A_n &= e^{-\hat{w}^2/2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\hat{z} y - \frac{1}{2} y^2} dy \\ &= e^{-\hat{w}^2/2} e^{\frac{1}{2}\hat{z}^2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(y+\hat{z})^2} dy \\ &= e^{-\hat{w}^2/2} e^{\frac{1}{2}\hat{z}^2} (1 - \Phi(\hat{z})) \\ &= \phi(\hat{w}) \frac{1 - \Phi(\hat{z})}{\phi(\hat{z})} \\ &= \phi(\hat{w}) M(\hat{z}) \end{aligned}$$

where

$$M(t) = \frac{1 - \Phi(t)}{\phi(t)}$$

is the Miller's ratio. For B_n, by integration by parts and Berry-Esseen bounds, we get

$$\begin{aligned} |B_n| &\le e^{-\hat{w}^2/2} \int_0^\infty e^{-\hat{z} y} d\Delta(y) \\ &\le e^{-\hat{w}^2/2} \left[\Delta(0) - \int_0^\infty \Delta(y) de^{-\hat{z} y} \right] \\ &\le 2e^{-\hat{w}^2/2} \sup_y |\Delta(y)| \\ &\le e^{-\hat{w}^2/2} \frac{E|Y_1|^3}{(EY_1^2)^{3/2} \sqrt{n}} \\ &\le Ce^{-\hat{w}^2/2} \frac{E|X_1|^3}{(EX_1^2)^{3/2} \sqrt{n}} \\ &\le C\phi(\hat{w}) / \sqrt{n} \end{aligned}$$

Therefore,

$$P(\overline{X} > x) = \phi(\hat{w})\left(M(\hat{z}) + O\left(n^{-1/2}\right)\right)$$

R

1. The Miller's ratio has the following asymptotic expansions:

$$1 - \Phi(t) = \phi(t)\left(\frac{1}{t} - \frac{1}{t^3} + \frac{1\cdot 3}{t^5} - \frac{1\cdot 3\cdot 5}{t^7} + \ldots\ldots\right), \quad \text{as } t \to \infty$$

   Proof of this can be seen by integration by parts iteratively ($\phi(x)' = \phi(x)\cdot -x$),

$$1 - \Phi(t) = \int_t^\infty \phi(x)dx = -\int_t^\infty \frac{d\phi(x)}{x} = \frac{\phi(t)}{t} + \int_t^\infty \frac{\phi(x)}{x^2}dx = \ldots\ldots$$

2. The following inequalities are immediate:

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\phi(t) < 1 - \Phi(t) < \frac{\phi(t)}{t} \quad \text{for } t > 1$$

$$\frac{\phi(t)}{2t} < 1 - \Phi(t) < \frac{\phi(t)}{t} \quad \text{for } t > \sqrt{2}$$

3. From (2), we can rewrite

$$\begin{aligned} P(\overline{X} > x) &= \phi(\hat{w})M(\hat{z})\left(1 + \hat{z}O\left(n^{-1/2}\right)\right) \\ &= [1 - \Phi(\hat{w})]\frac{M(\hat{z})}{M(\hat{w})}\left(1 + \hat{z}O\left(n^{-1/2}\right)\right) \end{aligned}$$

   In particular,
   (a) if $|x-\mu| = O\left(n^{-1/2}\right)$, then $\tau = O\left(n^{-1/2}\right)$ and so $\hat{z} = O(1)$. Therefore, the above saddlepoint approximation has relative error of size $O\left(n^{-1/2}\right)$.
   (b) if $|x-\mu| = o(1)$, then $\tau = o(1)$ and so $\hat{z} = O(\sqrt{n})$. Therefore, the above saddlepoint approximation has relative error of size $o(1)$.

Second-order approximations.

We can use the one-term Edgeworth expansion to approximate $P(T \le y)$:

$$E_1(y) = \Phi(y) - \frac{\kappa_3(\tau)}{6\sqrt{n}}H_3(y)\phi(y), \quad \Delta_1(y) = P(T \le y) - E_1(y)$$

Then we have

$$\begin{aligned} P(\overline{X} > x) &= e^{-n\hat{w}^2/2}\int_0^\infty e^{-\hat{z}y}d\left[E_1(y) + \Delta_1(y)\right] \\ &= A_{n1} + B_{n1} \end{aligned}$$

where $\Delta(y) = P(T \le y) - \Phi(y)$. First

$$\begin{aligned}
A_n &= e^{-\hat{w}^2/2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\hat{z}y - \frac{1}{2}y^2} dy \\
&= e^{-\hat{w}^2/2} e^{\frac{1}{2}\hat{z}^2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(y+\hat{z})^2} dy \\
&= e^{-\hat{w}^2/2} e^{\frac{1}{2}\hat{z}^2} (1 - \Phi(\hat{z})) \\
&= \phi(\hat{w}) \frac{1 - \Phi(\hat{z})}{\phi(\hat{z})} \\
&= \phi(\hat{w}) M(\hat{z})
\end{aligned}$$

where $M(t) = \frac{1-\Phi(t)}{\phi(t)}$ is the Miller's ratio. For B_n, by integration by parts and Berry-Esseen bounds, we get

$$\begin{aligned}
|B_n| &\le e^{-\hat{w}^2/2} \int_0^\infty e^{-\hat{z}y} d\Delta(y) \\
&\le e^{-\hat{w}^2/2} \left[\Delta(0) - \int_0^\infty \Delta(y) de^{-\hat{z}y} \right] \\
&\le 2e^{-\hat{w}^2/2} \sup_y |\Delta(y)| \\
&\le e^{-\hat{w}^2/2} \frac{E|Y_1|^3}{(EY_1^2)^{3/2}\sqrt{n}} \\
&\le Ce^{-\hat{w}^2/2} \frac{E|X_1|^3}{(EX_1^2)^{3/2}\sqrt{n}} \\
&\le C\phi(\hat{w})/\sqrt{n}
\end{aligned}$$

Therefore,

$$P(\bar{X} > x) = \phi(\hat{w}) \left(M(\hat{z}) + O\left(n^{-1/2}\right) \right)$$

R

1. (1). The Miller's ratio has the following asymptotic expansions:

$$1 - \Phi(t) = \phi(t) \left(\frac{1}{t} - \frac{1}{t^3} + \frac{1 \cdot 3}{t^5} - \frac{1 \cdot 3 \cdot 5}{t^7} + \ldots\ldots \right), \quad \text{as } t \to \infty$$

Proof of this can be seen by integration by parts iteratively,

$$1 - \Phi(t) = \int_t^\infty \phi(x) dx = -\int_t^\infty \frac{d\phi(x)}{x} = \frac{\phi(t)}{t} + \int_t^\infty \frac{\phi(x)}{x^2} dx = \ldots\ldots$$

(2). The following inequalities are immediate:

$$\begin{aligned}
\left(\frac{1}{t} - \frac{1}{t^3} \right) \phi(t) < 1 - \Phi(t) < \frac{\phi(t)}{t} \quad & \text{for } t > 1 \\
\frac{\phi(t)}{2t} < 1 - \Phi(t) < \frac{\phi(t)}{t} \quad & \text{for } t > \sqrt{2}
\end{aligned}$$

(3). From (2), we can rewrite

$$\begin{aligned} P(\bar{X} > x) &= \phi(\hat{w})M(\hat{z})\left(1 + \hat{z}O\left(n^{-1/2}\right)\right) \\ &= [1 - \Phi(\hat{w})]\frac{M(\hat{z})}{M(\hat{w})}\left(1 + \hat{z}O\left(n^{-1/2}\right)\right) \end{aligned}$$

In particular,

1. if $|x-\mu| = O\left(n^{-1/2}\right)$, then $\tau = O\left(n^{-1/2}\right)$ and so $\hat{z} = O(1)$. Therefore, the above saddlepoint approximation has relative error of size $O\left(n^{-1/2}\right)$.
2. if $|x-\mu| = o(1)$, then $\tau = o(1)$ and so $\hat{z} = O(\sqrt{n})$. Therefore, the above saddlepoint approximation has relative error of size $o(1)$.

Theorem 3.11.27 For a complex z,

1. $|e^z - 1| \le |z|e^{|z|}$
2. $|e^z - 1| \le e^{|z|} - 1$
3. $|\ln(1+z) - z| \le \frac{|z|^2}{2(1-|z|)}$, for $|z| \le 1$. For instance,
   (a) $|\ln(1+z) - z| \le |z|^2$, for $|z| \le 1/2$
   (b) $|\ln(1+z) - z| \le \frac{4}{5}|z|^2$, for $|z| \le 3/8$

Proof.

$$\begin{aligned} |e^z - 1| &= \left|z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots\right| \\ &\le |z|\left|1 + |z| + \frac{|z|^2}{2!} + \frac{|z|^3}{3!} + \frac{|z|^4}{4!} + \cdots\right| \\ &\le |z|e^{|z|} \end{aligned}$$

$$\begin{aligned} |e^z - 1| &\le \left|z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots\right| \\ &= \left||z| + \frac{|z|^2}{2!} + \frac{|z|^3}{3!} + \frac{|z|^4}{4!} + \cdots\right| \\ &\le e^{|z|} - 1 \end{aligned}$$

$$\begin{aligned} |\ln(1+z) - z| &= \left|\frac{z^2}{2} - \frac{z^3}{3} + \frac{z^4}{4} + \cdots\right| \\ &\le \frac{|z|^2}{2}|1 + |z| + |z|^2 + |z|^3 + \cdots| \\ &\le \frac{|z|^2}{2(1-|z|)} \end{aligned}$$

$\blacksquare$

3.12 Innitely divisible distributions

The family of infinitely divisible distributions is an important one in probability theory for at least two reasons:

1. it is closely associated with Levy Process;
2. it characterizes the family of limiting distributions for the sum of independent r.v.s.

Definition 3.12.1 A d.f. F and its corresponding c.f. $\psi(t)$ are called **infinitely divisible (i.d.)** if for every positive integer n

1. there exists a c.f. $\psi_n(t)$ such that
$$\psi(t) = (\psi_n(t))^n$$
2. or equivalently, there exists a d.f. F_n such that
$$F = F_n^{*n}, \quad \text{the } n-\text{fold convolution of the function } F_n$$
3. or equivalently, there exist a r.v. $X \sim F$ with c.f. ψ and i.i.d. r.v.s $X_{n1}, \ldots, X_{nn}$ with c.f. ψ_n such that
$$X =_d X_{n1} + \ldots + X_{nn}$$

■ **Example 3.23**

1. Degenerate d.f. (i.e. $P(X = C) = 1$)
$$\psi(t) = e^{itC} = \left(e^{it(C/n)}\right)^n = (\psi_n(t))^n$$
2. Normal d.f.
$$\psi(t) = \exp\left\{i\mu t - \sigma^2 t^2/2\right\} = \left(\exp\left\{i(\mu/n)t - \left(\sigma^2/n\right)t^2/2\right\}\right)^n = (\psi_n(t))^n$$
3. Poisson d.f.
$$\psi(t) = \exp\left\{\lambda\left(e^{it} - 1\right)\right\} = \left(\exp\left\{(\lambda/n)\left(e^{it} - 1\right)\right\}\right)^n = (\psi_n(t))^n$$
4. Compound Poisson d.f.: $S_N = X_1 + \ldots + X_N$, where $X \sim F$ and $N \sim$ Poisson (λ). So
$$\psi(t) = e^{\lambda(\varphi_x(t)-1)} = \left(e^{(\lambda/n)(\varphi_x(t)-1)}\right)^n = (\psi_n(t))^n$$
Remark: if $P(X = 1) = 1$, this reduces to Poisson d.f.
5. Cauchy d.f. with p.d.f. $f(x) = \frac{a}{\pi}\frac{1}{a^2+x^2}$
$$\psi(t) = \exp\{-a|t|\} = (\exp\{-(a/n)|t|\})^n = (\psi_n(t))^n$$
6. $\alpha-$ stable d.f. (including Cauchy d.f.)
7. Gamma d.f. with p.d.f. $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$:
$$\psi(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha} = \left(\left(1 - \frac{it}{\beta}\right)^{-\alpha/n}\right)^n = (\psi_n(t))^n$$
Two special cases:

(a) The χ^2 -distribution is i.d. since it is a special case of Gamma d.f.

(b) The t -distribution is i.d. since it is a special case of the χ^2 -distribution (with degree of freedom 1)

8. log-normal d.f.
9. Pareto d.f.
10. Student d.f.

Theorem 3.12.1 1. Let $\psi(t)$ be an i.d. c.f. Then $\psi(t) \neq 0$ for every t.

2. If $\psi_i(t), i = 1, 2, \ldots, m$, are i.d.c.f.'s, so is $\prod_{i=1}^m \psi_i(t)$
3. If $\psi(t)$ is i.d. c.f., so is $|\psi(t)|$. (Remark: If $\psi(t)$ is a c.f., $|\psi(t)|$ may not be a c.f. in general.)
4. Let $\left\{\psi^{(m)}(t); m = 1, 2, \ldots\right\}$ be a sequence of i.d.c.f.s converging to some c.f. $\psi(t)$. Then, $\psi(t)$ must be i.d.c.f.

Proof. 1. For every c.f. $\psi(t)$, it is continuous and $\psi(0) = 1$. Thus, given $\epsilon \in (0, 1)$, we have $|\psi(t)| = |\psi(0) - [\psi(0) - \psi(t)]| \geq |\psi(0)| - |\psi(0) - \psi(t)| > 1 - \epsilon$ for $|t| \leq \delta$. Since $\psi(t) = (\psi_n(t))^n$ for every $n \geq 1$, we have

$$|\psi_n(t)| = |\psi(t)|^{1/n} > (1-\epsilon)^{1/n} \to 1, \quad \text{as } n \to \infty$$

which implies that, for $n \geq N_0$, we have $|\psi_n(t)| > 1 - \epsilon/8 > 0$ for $|t| \leq \delta$ and $n \geq N_0$ Next we will show that $|\psi_n(t)| > 0$ for $|t| \leq 2\delta$ and $n \geq N_0$. To do that, we use Lemma ??????? to get

$$1 - |\psi_n(2t)| \leq 8\left(1 - |\psi_n(t)|\right) < \epsilon$$

so $|\psi_n(2t)| \geq 1 - \epsilon > 0$ for $|t| \leq \delta$ and $n \geq N_0$. That is, $|\psi_n(t)| > 0$ for $|t| \leq 2\delta$ and $n \geq N_0$. Continuing like this, we get $|\psi_n(t)| > 0$ for all $t \in R$ and $n \geq N_0$. Therefore,

$$|\psi(t)| = \left|(\psi_{N_0}(t))^{N_0}\right| = |\psi_{N_0}(t)|^{N_0} > 0$$

2. Take $m = 2$ for instance. Since $\psi_i(t) = (\psi_{in}(t))^n, i = 1, 2$, we have $\psi_1(t)\psi_2(t) = (\psi_{1n}(t)\psi_{2n}(t))^n$
3. If $\psi(t)$ is i.d., $\psi(t) = (\psi_{2n}(t))^{2n}$ for all $n \geq 1$. Hence,

$$|\psi(t)|^2 = \left|(\psi_{2n}(t))^{2n}\right|^2 = \left(|\psi_{2n}(t)|^2\right)^{2n}$$

therefore,

$$|\psi(t)| = \left(|\psi_{2n}(t)|^2\right)^n =: (\tilde{\psi}_n(t))^n$$

where $\tilde{\psi}_n(t) = |\psi_{2n}(t)|^2$ is a c.f.

4. We have $\psi^{(m)}(t) = \left(\psi_n^{(m)}(t)\right)^n$ for all m and n. From the assumption, we have

$$\lim_{m\to\infty} \psi^{(m)}(t) = \lim_{m\to\infty} \left(\psi_n^{(m)}(t)\right)^n = \left(\lim_{m\to\infty} \psi_n^{(m)}(t)\right)^n = \psi(t)$$

for each fixed n, namely,

$$\lim_{m\to\infty} \psi_n^{(m)}(t) = (\psi(t))^{1/n}$$

Since $\left\{\psi_n^{(m)}(t), m \geq 1\right\}$ are c.f.s, and $(\psi(t))^{1/n}$ is continuous at 0, then by Levy continuity theorem, $(\psi(t))^{1/n}$ is a c.f. as well. Therefore, $\psi(t) = \left(\psi^{1/n}(t)\right)^n$ is i.d.

■

■ **Example 3.24** We can use Theorem 3.12.1 (part 1) to judge whether a c.f. is NOT i.d.

1. Let $X \sim$ Uniform [-1,1] with c.f. $f(t) = (\sin t)/t$. Since $f(k\pi) = 0$, then, $f(t)$ is NOT i.d.
2. Let $P(X = \pm 1)$ with c.f. $f(t) = \cos t$. Similarly, it is NOT i.d.

3.12.1 Levy-Khintchine representation of infinitely divisible c.f.s

Theorem 3.12.2 — "Levy-Khintchine" representation. A function $\psi(t)$ is an i.d.c.f. if and only if it admits the representation

$$\psi(t) = \exp\left\{it\gamma + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x)\right\} =: e^{\eta(t)} \tag{3.75}$$

where γ is a real constant, and G is a bounded non-decreasing function. ($W.L.O.G.$, we assume that G is left-continuous and $G(-\infty) = 0$; see the remark below.)

R

1. Denote the function under the integral sign by

$$g(t,x) = \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}$$

It is easy to see that $\lim_{x\to 0} g(t,x) = -\frac{t^2}{2}$, so we can define $g(t,0) = -\frac{t^2}{2}$

2. First, we note that the values of $G(x)$ at points of discontinuity do not influence the value of the integral on the RHS of (3.75) since $g(t,x)$ is continuous in x. Secondly, adding any constant, C, to $G(x)$ does not influence the value of the integral on the RHS of (3.75) either. For the purpose of definiteness, we may assume that it is left-continuous and $G(-\infty) = 0$ from now on.

Proof of necessity

Theorem 3.12.3 If $\psi(t)$ is i.d.c.f., then $\psi(t)$ has the representation (3.75)

Proof. We start with $\psi(t) = (\psi_n(t))^n$. Let F_n denote the d.f. corresponding to the c.f. ψ_n. Now as $0 < |\psi(t)| \le 1, \ln\psi(t)$ exists and is finite. Furthermore,

$$\psi^{1/n}(t) = \exp\left(\frac{1}{n}\ln\psi(t)\right) = 1 + \frac{1}{n}\ln\psi(t) + O\left(\frac{1}{n^2}\right) =: 1 + \frac{1}{n}\eta(t) + O\left(\frac{1}{n^2}\right)$$

from which we get

$$\begin{aligned}
\eta(t) &= \lim n\left(\psi^{1/n}(t) - 1 - O\left(\frac{1}{n^2}\right)\right) \\
&= \lim_{n\to\infty} n\left(\psi^{1/n}(t) - 1\right) \\
&= \lim_{n\to\infty} n\left(\psi_n(t) - 1\right) = \lim_{n\to\infty}\int_{-\infty}^{\infty} n\left(e^{itx} - 1\right) dF_n(x) \\
&= \lim_{n\to\infty}\int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2} + \frac{itx}{1+x^2}\right) dF_n(x) \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) dF_n(x)\right\} \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{x^2}{1+x^2}dF_n(x)\right\} \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}d\int_{-\infty}^{x}\frac{ny^2}{1+y^2}dF_n(y)\right\} \\
&=: \lim_{n\to\infty}\left\{it\gamma_n + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG_n(x)\right\}
\end{aligned}$$

where

$$\eta_n(t) =: it\gamma_n + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG_n(x) \tag{3.76}$$

and

$$\gamma_n = \int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x), \quad G_n(x) = \int_{-\infty}^{x}\frac{ny^2}{1+y^2}dF_n(y)$$

We have shown that

$$\lim_{n\to\infty}\eta_n(t) = \eta(t)$$

where $\eta(t)$ is continuous at 0. From Lemma 3.35 below, we have that $\gamma_n \to \gamma$ and $G_n \Longrightarrow G$ for some constant γ and some "nice" function G as described in Theorem 3.12.2. ∎

Appendix: Several useful lemmas

Let

$$A(y) =: \left(1 - \frac{\sin y}{y}\right)\frac{1+y^2}{y^2}$$

Note that for $y = o(1)$, we have

$$A(y) = \left(1 - \frac{1}{y}\left(y - \frac{y^3}{3!} + \frac{y^5}{5!} - \dots\right)\right)\frac{1+y^2}{y^2} = \left(\frac{1}{3!} - \frac{y^2}{5!} - \dots\right)(1+y^2)$$

so if we define $A(y)$ to be $1/3!$ at $y=0$, so that $A(y)$ is a nonnegative bounded continuous function. Furthermore, it can be shown easily that

$$0 < c_1 \le A(y) \le c_2 < \infty, \quad \text{for all } y$$

Now define

$$\Lambda(x) = \int_{-\infty}^{x} A(y)dG(y), \quad \text{and} \quad \Lambda_n(x) = \int_{-\infty}^{x} A(y)dG_n(y)$$

Therefore, $\Lambda(x)$ and $\Lambda_n(x)$ are bounded and non-decreasing, $\Lambda(-\infty)=0$ and $\Lambda(\infty) < \infty$ as $G(x)$ is bounded and non-decreasing. Furthermore, we can easily work out their Fourier transforms.

Lemma 3.33 We have

$$\int_{-\infty}^{\infty} e^{itx} d\Lambda(x) = \eta(t) - \frac{1}{2}\int_0^1 [\eta(t+h)+\eta(t-h)]dh =: \lambda(t)$$
$$\int_{-\infty}^{\infty} e^{itx} d\Lambda_n(x) = \eta_n(t) - \frac{1}{2}\int_0^1 [\eta_n(t+h)+\eta_n(t-h)]\,dh =: \lambda_n(t)$$

Proof.

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{itx} d\Lambda(x) &= \int_{-\infty}^{\infty} e^{itx}\left(1-\frac{\sin x}{x}\right)\frac{1+x^2}{x^2}dG(x) \\
&= \int_{-\infty}^{\infty}\int_0^1 e^{itx}(1-\cos hx)\frac{1+x^2}{x^2}dhdG(x) \\
&\quad \text{(by Fubini's theorem since the integrand is bounded} \\
&\qquad \text{continuous in}[0,1]\times(-\infty,\infty)) \\
&= \int_0^1\int_{-\infty}^{\infty} e^{itx}(1-\cos hx)\frac{1+x^2}{x^2}dG(x)dh \\
&= \int_0^1 \left(\eta(t) - \frac{1}{2}[\eta(t+h)+\eta(t-h)]\right) dh \\
&= \eta(t) - \frac{1}{2}\int_0^1 [\eta(t+h)+\eta(t-h)]dh \\
&= \lambda(t)
\end{aligned}$$

■

Lemma 3.34 There is a 1 -to-1 correspondence between the functions $\eta(t)$ given by (3.75) and the pair (γ, G), where γ is a real constant and G is a non-decreasing bounded function with $G(-\infty) = 0$. So we can use the notation

$$\eta = (\gamma, G)$$

Proof. Given the pair $(\gamma, G), \eta$ is uniquely determined. On the other hand, given the function $\eta(t)$, it uniquely defines $\lambda(t)$ (which is a c.f. up to a constant multiplier). From the last lemma, $\lambda(t)$ uniquely defines $\Lambda(t)$. In turn, $\Lambda(t)$ uniquely defines the function

$$G(x) = \int_{-\infty}^{x} \frac{d\Lambda(y)}{A(y)}$$

Finally, ψ and G together uniquely defines γ. ∎

Lemma 3.35
1. If $\gamma_n \to \gamma$ and $G_n \Longrightarrow G$, then $\eta_n(t) \to \eta(t) = (\gamma, G)$
2. If $\eta_n(t) \to \eta(t)$ and $\eta(t)$ is continuous at 0, then there exist some real constant γ and bounded non-decreasing left-continuous function $G(x)$ such that $\gamma_n \to \gamma$ and $G_n \Longrightarrow G$, and $\eta(t) = (\gamma, G)$

Proof.
1. This part is trivial.
2. Since $\psi_n(t)$'s are i.d. c.f.s (verify this as an exercise!) and $\psi_n(t) = e^{\eta_n(t)} \to e^{\eta(t)}$ and $e^{\eta(t)}$ is continuous at 0, so by Lemma ??, we know that $e^{\eta(t)}$ is also i.d.c.f. Also, from Theorem 3.12.1, we have $e^{\eta(t)} \neq 0$. Therefore, $|\eta(t)|$ is finite, and $\eta_n(t) \to \eta(t)$ uniformly in any interval $t \in [a, b]$. Therefore, we have

$$\lambda_n(t) \to \lambda(t)$$

since $\lambda(t)$ is continuous at $t = 0$, by Levy continuity theorem, we have

$$\Lambda_n(t) \to \Lambda(t)$$

Noting that $\Lambda_n(-\infty) = \Lambda(-\infty) = 0$, and $\lambda_n(0) \to \lambda(0)$, i.e.

$$\lambda_n(0) = \int_{-\infty}^{\infty} d\Lambda_n(x) = \Lambda_n(\infty) - 0 \to \lambda(0) = \int_{-\infty}^{\infty} d\Lambda(x) = \Lambda(\infty) - 0$$

Therefore,

$$\Lambda_n \Longrightarrow \Lambda$$

From (4.3), we have $G_n(x) \to G(x)$. Now

$$\begin{aligned} it\gamma_n &= \eta_n(t) - \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG_n(x) \\ &\to \eta(t) - \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x) \end{aligned}$$

Thus, there exists γ such that $\lim_n \gamma_n = \gamma$; and from part (i), we know $\psi = (\gamma, G)$. ∎

Proof of sufficiency

Theorem 3.12.4 If $\psi(t)$ has the representation (4.1), then $\psi(t)$ is i.d.c.f.

Proof. Denote the integral in (3.75) by $I(t)$, which can be written as

$$\begin{aligned} I(t) &= \int_{-\infty}^{\infty} g(t,x) dG(x) \\ &= \int_{\{x>0\}} + \int_{\{x<0\}} + \int_{\{x=0\}} g(t,x) dG(x) \\ &= I_+(t) + I_-(t) + g(t,0)[G(0+) - G(0-)] \\ &= I_+(t) + I_-(t) - \frac{t^2}{2}[G(0+) - G(0-)] \end{aligned}$$

Then, we have

$$\psi(t) = e^{it\gamma} \exp\{I(t)\} = e^{it\gamma} \exp\{I_+(t)\} \exp\{I_-(t)\} \exp\left\{-\frac{t^2}{2}[G(0+) - G(0-)]\right\}$$

Therefore, it suffices to show that $\exp\{I_+(t)\}$ and $\exp\{I_-(t)\}$ are i.d.c.f. We will look at the first one only since the second can be done similarly. Note that

$$\exp\{I_+(t)\} = \lim_m \exp\left\{I_+^{1/m}(t)\right\}$$

where

$$\begin{aligned} I_+^{\epsilon}(t) &= \int_{\epsilon}^{1/\epsilon} g(t,x) dG(x) \\ &= \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1 - \frac{it\xi_k}{1+\xi_k^2}\right) \frac{1+\xi_k^2}{\xi_k^2} \left[G(x_{k+1}) - G(x_{k+1})\right] \\ &\qquad (\epsilon = x_0 < x_1 < \ldots < x_n = 1/\epsilon, \quad x_k \le \xi_k < x_{k+1}) \\ &= \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1 - \frac{it\xi_k}{1+\xi_k^2}\right) \lambda_{nk} \\ &=: \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1\right) \lambda_{nk} - it \frac{\xi_k \lambda_{nk}}{1+\xi_k^2} \\ &=: \lim_n \sum_{k=0}^{n-1} \left(ita_{nk} + \left(e^{it\xi_k} - 1\right) \lambda_{nk}\right) \\ &=: \lim_n \sum_{k=0}^{n-1} T_{nk} \end{aligned}$$

where

$$\lambda_{nk} = \frac{1+\xi_k^2}{\xi_k^2} \left[G(x_{k+1}) - G(x_{k+1})\right], \quad a_{nk} = -\frac{\xi_k \lambda_{nk}}{1+\xi_k^2}$$

Note that $\exp(T_{nk}) = \exp\left(ita_{nk} + \left(e^{it\xi_k} - 1\right) \lambda_{nk}\right)$ is the c.f. of a Poisson d.f. Therefore, $\prod_{k=0}^{n-1} \exp\{T_{nk}\}$ is also a c.f. and furthermore, we have

$$\lim_n \prod_{k=0}^{n-1} \exp\{T_{nk}\} = \lim_n \exp\left\{\sum_{k=0}^{n-1} T_{nk}\right\} = \exp\{I_+^{\epsilon}(t)\}$$

Since $\exp\{I_+^\epsilon(t)\}$ is continuous at 0, by Levy continuity theorem, $\exp\{I_+^\epsilon(t)\}$ is the c.f. of some d.f. It then follows from Theorem ?? that $\exp\{I_+^\epsilon(t)\}$ is i.d.c.f. ■

Several examples

Given an i.d. c.f $\psi(t) = e^{\eta(t)}$, we sometimes can find out the Levy-Khintchine representation directly

$$\eta(t) = it\gamma + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x) = it\gamma + \int_{-\infty}^{\infty} g(t,x)dG(x)$$

Recall $g(t,0) = -t^2/2$. Alternatively, we use $\eta = (\gamma, G)$, where

$$\gamma_n = \int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) \to \gamma, \quad G_n(x) = \int_{-\infty}^{x}\frac{ny^2}{1+y^2}dF_n(y) \to G(x)$$

■ **Example 3.25** Give the Levy-Khintchine representation for the following r.v.'s

1. $N(\mu, \sigma^2)$
2. Poisson(λ),
3. Degenerate r.v.: $P(X = C) = 1$
4. Cauchy r.v.: $f(x) = \frac{a}{\pi}\frac{1}{a^2+x^2}$ and $\psi(t) = e^{-a|t|}$

Proof. Let $\delta_h(x)$ be the (left-continuous) Dirac measure with mass concentrated at h, namely,

$$\delta_h(x) = I\{x > h\}$$

1. For Normal d.f., we have

$$\eta(t) = it\mu - \frac{1}{2}\sigma^2 t^2$$

So we can take $dG(0) = \sigma^2$, and $dG(x) = 0$ for $x \neq 0$. That is,

$$\gamma = \mu, \quad G(x) = \sigma^2\delta_0(x) = \sigma^2 I\{x > 0\}$$

2. For Poisson (λ) d.f., we have

$$\begin{aligned}
\eta(t) &= \lambda\left(e^{it} - 1\right) = \lambda\left(e^{itx} - 1\right)\Big|_{x=1} \\
&= \left(e^{itx} - 1 - \frac{itx}{1+x^2} + \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}\Big|_{x=1} \\
&= it\frac{\lambda}{2} + \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}\Big|_{x=1} \\
&= it\frac{\lambda}{2} + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}d\delta_1(x)
\end{aligned}$$

So

$$\gamma = \frac{\lambda}{2}, \quad G(x) = \frac{\lambda}{2}\delta_1(x) = \frac{\lambda}{2}I\{x > 1\}$$

3. For degenerate d.f. $X = C$ a.s., we have

$$\eta(t) = itC$$

So

$$\gamma = C, \quad G(x) \equiv 0$$

4. since $\psi(t) = \left(e^{-(a/n)|t|}\right)^n$, so

$$f_n(x) = \frac{dF_n(x)}{dx} = \frac{a/n}{\pi} \frac{1}{a^2/n^2 + x^2}$$

Now

$$\begin{aligned}
G_n(x) &= n \int_{-\infty}^{x} \frac{y^2}{1+y^2} dF_n(y) = n \int_{-\infty}^{x} \frac{y^2}{1+y^2} \frac{a}{n\pi} \frac{1}{a^2/n^2 + y^2} dy \\
&= \frac{a}{\pi} \int_{-\infty}^{x} \frac{y^2}{1+y^2} \frac{1}{a^2/n^2 + y^2} dy \\
&\to \frac{a}{\pi} \int_{-\infty}^{x} \frac{1}{1+y^2} dy =: G(x) \\
\gamma_n &= n \int_{-\infty}^{\infty} \frac{x}{1+x^2} dF_n(x) = n \frac{a/n}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+x^2} \frac{1}{a^2/n^2 + x^2} dx = 0 =: \gamma
\end{aligned}$$

That is,

$$\gamma = 0, \quad G(x) = \frac{a}{\pi} \int_{-\infty}^{x} \frac{1}{1+y^2} dy$$

■

3.12.2 Levy formula

We now give another very useful representation: Levy formula. Let us rewrite the Levy-Khintchine representation as

$$\begin{aligned}
\psi(t) &= \exp\left\{ it\gamma + \left(\int_{\{x=0\}} + \int_{\{|x|>0\}} \right) \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x) \right\} \\
&= \exp\left\{ it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{\{|x|>0\}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) dL(x) \right\}
\end{aligned}$$

where

$$\begin{aligned}
\sigma^2 &= G(0+) - G(0-) \\
dL(x) &= \frac{1+x^2}{x^2} dG(x), \quad x \neq 0
\end{aligned}$$

Several remarks about $L(x)$ are given below.

1. $L(x)$ is defined on $\mathcal{R}-\{0\}$. It is easy to see that
$$\begin{aligned} L(x) = & C_1 + \int_{-\infty}^{x} \frac{1+y^2}{y^2} dG(y), \quad \text{if } x<0 \\ & C_2 - \int_{x}^{\infty} \frac{1+y^2}{y^2} dG(y), \quad \text{if } x>0 \end{aligned}$$
for any constants C_1 and C_2. (Verify that the integrals are well-defined!) Furthermore, it is easy to see that $L(x)$ is non-decreasing on $(-\infty,0)$ and $(0,\infty)$, respectively, and satisfies
$$\lim_{x\to-\infty} L(x) = C_1, \quad \lim_{x\to\infty} L(x) = C_2 \tag{3.77}$$
2. Note that, for every finite $\delta>0$, we have
$$\int_{0<|x|<\delta} x^2 dL(x) = \int_{0<|x|<\delta} (1+x^2)\, dG(x) \le (1+\delta^2) \int_{0<|x|<\delta} dG(x) < \infty$$
3. In view of (3.77), the following are equivalent:
$$\int_{0<|x|<\delta} x^2 dL(x) < \infty, \quad \Longleftrightarrow \quad \int_{|x|>0} (x^2 \wedge 1)\, dL(x) < \infty, \quad \Longleftrightarrow \quad \int_{|x|>0} \frac{x^2}{1+x^2} dL(x) < \infty$$
4. $L(x)$ is finite for $x \ne 0$. But at $x=0$, they might not be well-defined. Namely, as $x \nearrow 0$ or $x \searrow 0$, we might have $|L(x)| \to \infty$ and / or $|L'(x)| = \infty$ (if L' exists). On the other hand, it is easy to see that, for every finite $\epsilon>0$, we have
$$L((-\epsilon,\epsilon)^c) = \int_{|x|>\epsilon} dL(x) = \int_{|x|>\epsilon} \frac{1+x^2}{x^2} dG(x) \le (1+\epsilon^{-2}) \int_{|x|>\epsilon} dG(x) < \infty$$
5. $L(x)$ is often called "Levy measure", a very important concept in the studies of Levy processes.

We summarize everything in the next theorem.

Theorem 3.12.5 — Levy formula. *A function* $\psi(t)$ is an i.d.c.f. if and only if it admits the following (unique) representation
$$\psi(t) = \exp\left\{ it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{|x|>0} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) dL(x) \right\}$$
where γ is a real constant, σ^2 is a non-negative constant, and the function L is non-decreasing on the intervals $(-\infty,0)$ and $(0,\infty)$, and satisfies
$$\int_{0<|x|<\delta} x^2 dL(x) < \infty$$
for every finite $\delta>0$

R An alternative Levy formula takes the following form:

$$\psi(t) = \exp\left\{it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{|x|>0}\left(e^{itx} - 1 - itxI\{|x|<1\}\right)dL(x)\right\}$$
$$= \exp\left\{it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{0<|x|<1}\left(e^{itx} - 1 - itx\right)dL(x) + \int_{|x|\geq 1}\left(e^{itx} - 1\right)dL(x)\right\}$$

This corresponds to the decomposition of a Levy process, which can be written as the sum of a Brownian motion, small jump component (a martingale), and a compound Poisson process.

■ **Example 3.26** Redo Example 3.25.

Proof. 1. For Normal d.f., we have $\eta(t) = it\mu - \frac{1}{2}\sigma^2 t^2 + 0$ So

$$\gamma = \mu, \quad \sigma^2 = \sigma^2, \quad L(x) = 0$$

The other two cases are left as exercises.

■

3.12.3 Kolmogorov formula

In the special case where the r.v. X has finite second moment EX^2, we know that $\psi(t)$ is twice differentiable. In this case, we have the following simpler representation.

Theorem 3.12.6 — Komogorov formula. *A* function $\psi(t)$ is an i.d.c.f. with a finite variance if and only if it admits the following (unique) representation

$$\psi(t) = \exp\left\{it\gamma + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - itx\right)\frac{1}{x^2}dK(x)\right\}$$

where γ is a real constant, the function K is a bounded non-decreasing function.

3.12.4 Relationship between the sum of independent r.v. and i.d.

What do the limiting d.f. of sums of independent r.v.s look like? We start with some simple examples:

1. If $E|X_1| < \infty$, then $\bar{X} \to_d \mu$, whose c.f is $e^{it\mu}$
2. If $E|X_1| < \infty$, then $\sqrt{n}(\bar{X} - \mu)/\sigma \to_d N(0,1)$, whose c.f is $e^{-t^2/2}$
3. If X_{nk} are i.i.d. $\text{Bin}(n, p = \lambda/n)$, then $\sum X_{nk} \to_d$ Poisson (λ), whose c.f. is $e^{\lambda(e^{it}-1)}$
4. If X_k are i.i.d. $\text{Cauchy}(0,1)$, then $\bar{X} \to_d X_1$, whose c.f. $e^{-|t|}$.

Note that the c.f.s of the limiting distributions are all of the form $e^{\eta(t)}$. In fact, all the above limiting distributions are i.d., this is no accident. See the next theorem.

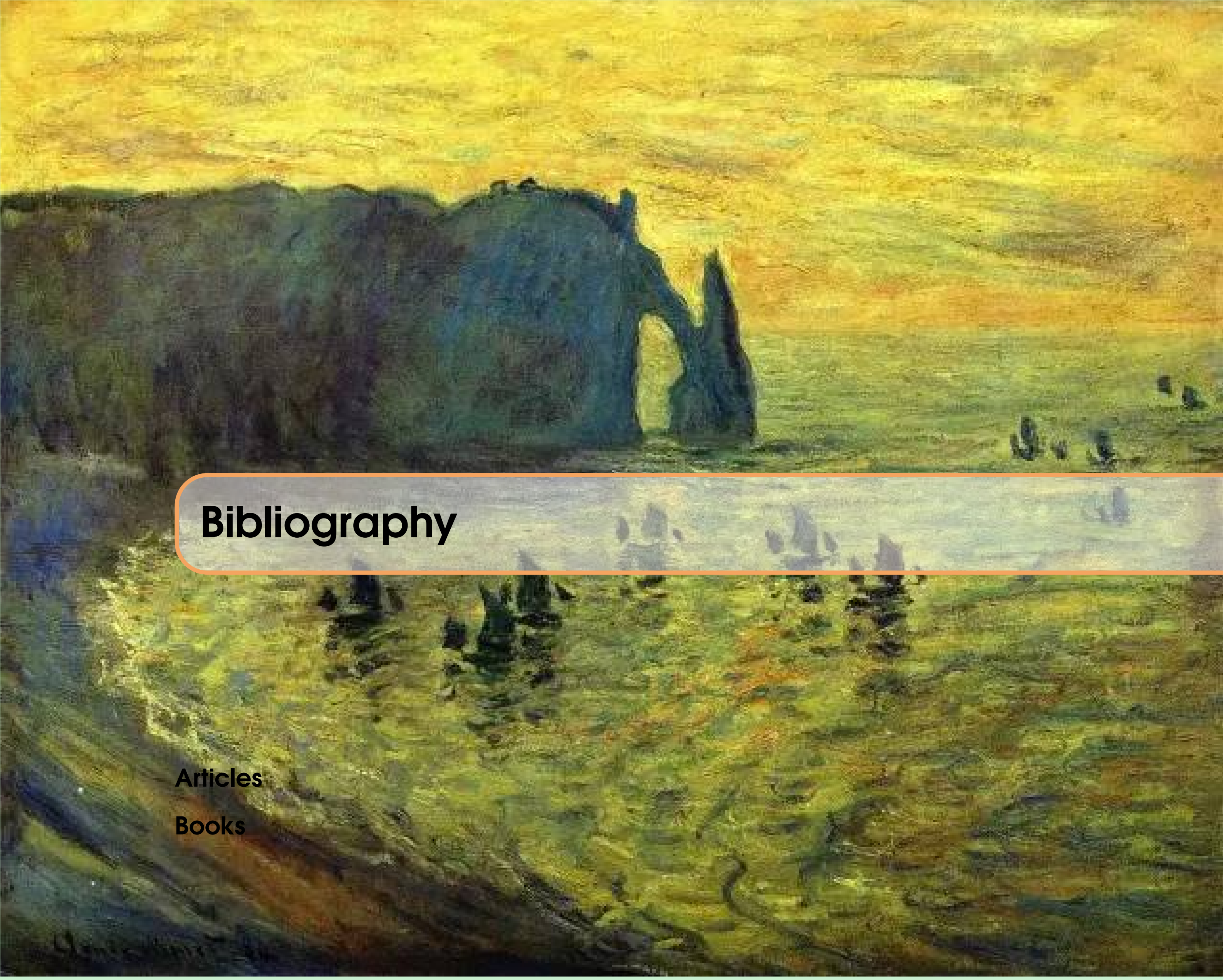
Theorem 3.12.7 Let $\sum X_{nk}$ be independent r.v.s satisfying the following infinitesimal condition:

$$\max_{1 \le k \le n} P\left(|X_{nk}| > \epsilon\right) \to 0, \quad \text{as } n \to \infty$$

Then,

$$\left\{ \text{ all limiting d.f.s of } \sum X_{nk} \right\} = \{\text{all i.d. d.f.s}\}$$

Part Four



Bibliography

Articles

Books