# Advanced Probability Theory

**Bing-Yi JING**

Department of Mathematics,
Hong Kong University of Science and Technology

Email: majing@ust.hk

July, 2012

# Contents


**1 Set Theory** **1**

1.1 Sets . . . 1

1.2 Basic set operations . . . 1

1.3 Operations of sequence of sets . . . 2

1.4 Indicator functions . . . 3

1.5 Semi-algebras, Algebras, and $\sigma$-algebras . . . 5

1.5.1 Definitions . . . 5

1.5.2 Relationships . . . 6

1.5.3 Some special $\sigma$-algebras. . . . 6

1.5.4 How to generate algebras from semi-algebras . . . 7

1.6 Generated classes (Minimal classes) . . . 7

1.7 Monotone class ($m$-class), $\pi$-class, and $\lambda$-class . . . 9

1.7.1 Definitions . . . 9

1.7.2 Relationships with $\sigma$-algebras . . . 10

1.7.3 Minimal $m$-class, $\lambda$-class and $\pi$-class . . . 10

1.7.4 Graphical illustration of different classes . . . 11

1.8 The Monotone Class Theorem (MCT) . . . 12

1.9 Product Spaces . . . 15

1.10 Exercise . . . 17

**2 Measure Theory** **18**

2.1 Definitions . . . 18

2.2 Properties of measure . . . 20

2.2.1 Case I: semialgebras. . . . 20

2.2.2 Case II: algebras. . . . 20

2.2.3 Case III: $\sigma$-algebras. . . . 21

2.3 Arithmetics with $\infty$ . . . 22

2.4 Probability measure . . . 22

2.5 Some examples of measure . . . 22

2.6 Extension of set functions (or measures) from semialgebras to algebras . . . 23

2.7 Outer measure . . . 25

2.8 Extension of measures from semialgebras to $\sigma$-algebras . . . 28

2.9 Completion of a measure . . . 31

2.10 Construction of measures on a $\sigma$-algebra $\mathcal{A}$ . . . 35

2.10.1 General procedures . . . 35

2.10.2 Lebesgue and Lebesgue-Stieltjes measures . . . 36

2.10.3 Relationship between probability measures and distribution functions . . . . . . . 38
2.10.4 Different types of distributions . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
2.10.5 Decomposition of distribution functions . . . . . . . . . . . . . . . . . . . . . . . 41
2.10.6 Further decomposition of a continuous d.f. $F$ . . . . . . . . . . . . . . . . . . . . 43
2.10.7 Some examples of distributions . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
2.11 Radon-Nikodym theorem . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
2.11.1 Applications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
2.11.2 Properties . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
2.11.3 The assumption of sigma-finiteness . . . . . . . . . . . . . . . . . . . . . . . . . 46
2.12 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48

**3 Random Variables** **49**
3.1 Mappings . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
3.2 Measurable mapping . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
3.3 Random Variables (Vectors) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
3.3.1 Random variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
3.3.2 How to check a random variable? . . . . . . . . . . . . . . . . . . . . . . . . . . 52
3.3.3 Random vectors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
3.4 Construction of random variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
3.4.1 Algebraic operations $(+, -, \times, \div)$ . . . . . . . . . . . . . . . . . . . . . . . . . 53
3.4.2 Limiting operations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
3.4.3 Transformations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
3.5 Approximations of r.v. by simple r.v.s . . . . . . . . . . . . . . . . . . . . . . . . . 54
3.6 $\sigma$-algebra generated by random variables. . . . . . . . . . . . . . . . . . . . . . . . 57
3.6.1 Definition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
3.6.2 Some examples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
3.6.3 The $\sigma-$field generated by a continuous r.v. . . . . . . . . . . . . . . . . . . . . . 60
3.6.4 A useful theorem . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.7 Distributions and induced distribution functions . . . . . . . . . . . . . . . . . . . . 60
3.7.1 Case I: Random variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.7.2 Case II: Random vectors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.8 Generating random variables with prescribed distributions . . . . . . . . . . . . . . . 62
3.9 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

**4 Expectation and Integration** **67**
4.1 Expectation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.1.1 Expectation for simple r.v.'s . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.1.2 Expectation for nonnegative r.v.'s . . . . . . . . . . . . . . . . . . . . . . . . . . 69
4.1.3 Expectation for general r.v.'s . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
4.1.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
4.2 Integration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
4.2.1 Definition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
4.2.2 Some properties of integrals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
4.2.3 Special cases: Lebesgue-Stieltjes integral and Lebesgue integral . . . . . . . . . . . 81
4.3 Measure-theoretic and probabilistic languages . . . . . . . . . . . . . . . . . . . . . . 84
4.4 How to compute expectation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

4.4.1 Expected values of absolutely continuous r.v.'s . . . 86
4.4.2 Expected values of discrete r.v.'s . . . 87
4.5 Relation between expectation and tail probability . . . 88
4.6 Moments and Moment inequalities . . . 90
4.6.1 Young's inequality . . . 90
4.6.2 Holder's inequality . . . 91
4.6.3 Cauchy-Schwarz inequality . . . 91
4.6.4 Lyapunov's inequality . . . 92
4.6.5 Minkowski's inequality . . . 92
4.6.6 Jensen's inequality . . . 92
4.6.7 Chebyshev (Markov) inequality . . . 93
4.7 Exercises . . . 94

**5 Independence** **96**
5.1 Definition . . . 96
5.2 How to check independence . . . 97
5.2.1 Discrete r.v.'s . . . 99
5.2.2 Absolutely continuous r.v.'s . . . 99
5.3 Functions of independent r.v.'s . . . 100
5.3.1 Transformation properties . . . 100
5.3.2 Convolutions . . . 100
5.3.3 An example . . . 101
5.3.4 Correlation . . . 102
5.4 Borel-Cantelli Lemma and Kolmogorov $0-1$ Law . . . 104
5.4.1 Borel-Cantelli Lemma . . . 104
5.4.2 Kolmogorov 0-1 laws . . . 106
5.5 Exercises . . . 109

**6 Convergence Concepts** **111**
6.1 Modes of convergence . . . 111
6.1.1 Equivalent definition of a.s. convergence . . . 112
6.2 Cauchy Criterion . . . 113
6.3 Relationships between modes of convergence . . . 114
6.4 Partial converses . . . 115
6.4.1 Convergence in probability and distribution to constants are equivalent . . . 115
6.4.2 Dominated convergence in probability implies convergence in mean . . . 116
6.4.3 Dominated convergence a.s. implies convergence in mean . . . 117
6.4.4 Convergence in probability sufficiently fast implies a.s. convergence . . . 118
6.4.5 Convergence in mean sufficiently fast implies a.s. convergence . . . 118
6.4.6 Convergence sequences in probability contains a.s. subsequences . . . 118
6.4.7 Convergence in distribution plus uniform integrability implies convergence in moments 118
6.4.8 Convergence in prob plus uniform integrability implies convergence in mean. . . . 118
6.4.9 Convergence in distribution implies a.s. convergence in another probability space. 119
6.5 Convergence of moments; uniform integrability . . . 120
6.5.1 Definition of uniform integrability (u.i.) . . . 120
6.5.2 Convergence in prob. + u.i. $\Longrightarrow$ convergence in mean . . . 122

6.5.3 Relationship between $L_r$ convergence and convergence of moments . . . 123
6.5.4 Convergence in dist. + u.i. $\Longrightarrow$ convergence in mean . . . 124
6.6 Some closed operations of convergence . . . 125
6.6.1 Algebraic operations . . . 125
6.6.2 Transformations (Continuous Mapping) . . . 126
6.6.3 Slutsky's Theorem . . . 127
6.7 Simple limit theorems . . . 128
6.8 Fatou's Lemma Revisited . . . 130
6.9 Summary: relationships amongst four modes of converges . . . 130
6.10 Exercises . . . 132

**7 Weak Law of Large Numbers** **134**
7.1 Equivalent sequences; truncation . . . 134
7.2 Weak Law of Large Numbers . . . 135
7.3 Classical forms of the WLLN . . . 138
7.4 Exercises . . . 140

**8 Strong Convergence** **141**
8.1 Some maximal inequalities . . . 141
8.2 The a.s. convergence of series; three-series theorem . . . 144
8.2.1 Review: Cauchy convergent a.s. or in probability . . . 144
8.2.2 Variance criterion for random series . . . 146
8.2.3 Kolmogorov three series theorem for random series . . . 147
8.2.4 Kolmogorov two series theorem for absolute random series . . . 149
8.3 Strong Laws of Large Numbers (SLLN) . . . 150
8.3.1 Several useful lemmas . . . 150
8.3.2 SLLN for independent r.v.'s . . . 151
8.3.3 SLLN for i.i.d. r.v.'s: necessary and sufficient moment conditions . . . 154
8.4 Some applications of the WLLN and SLLN . . . 158
8.5 Some results on LIL . . . 159
8.6 Exercises . . . 160

**9 Weak convergence** **162**
9.1 Definition . . . 162
9.2 Equivalent definitions of weak convergence . . . 162
9.3 Helly's selection theorem and tightness . . . 165
9.4 Polya Theorem . . . 168
9.5 Additional topic: Stable convergence and mixing* . . . 169
9.6 Exercise . . . 172
9.7 Some useful theorems . . . 172

**10 Characteristic Functions** **173**
10.1 Some examples of characteristic functions . . . 173
10.2 Definition and some properties of c.f.s . . . 177
10.3 Inversion formula . . . 180
10.4 Levy Continuity Theorem . . . 184
10.5 Moments of r.v.s and derivatives of their c.f.s . . . 185

10.6 When is a function a c.f.? . . . 190
10.7 Esseen's Smoothing Lemma . . . 191
10.8 Characteristic functions and smoothness condition . . . 194
10.9 Exercises . . . 197

**11 Central Limit Theorems and Related Expansions** **199**
11.1 Central Limit Theorems (CLT) . . . 199
11.1.1 CLT for i.i.d. r.v.s . . . 199
11.1.2 CLT for triangular arrays with finite variances . . . 200
11.1.3 Central Limit Theorems with infinite variances . . . 203
11.2 Uniform Berry-Esseen Bounds . . . 203
11.2.1 Some useful lemmas . . . 204
11.2.2 Berry-Esseen bounds . . . 206
11.2.3 A generalization of Berry-Esseen bounds . . . 208
11.3 Non-Uniform Berry-Esseen Bounds . . . 208
11.4 Edgeworth Expansions . . . 209
11.4.1 Heuristic argument for informal Edgeworth expansions . . . 210
11.4.2 Hermite polynomials . . . 210
11.4.3 Relationship between moments and cumulants . . . 211
11.4.4 Edgeworth expansions . . . 211
11.5 Non-uniform Edgeworth expansions . . . 216
11.6 Large Deviations . . . 218
11.6.1 Cramer conditioin . . . 218
11.6.2 Conjugate (or associated) distributions . . . 219
11.6.3 Large Deviations . . . 220
11.6.4 Cramer-type large deviations . . . 221
11.7 Saddlepoint Approximations . . . 223
11.8 Appendix: Some useful elementary inequalities . . . 226
11.9 Exercises . . . 228

**12 Infinitely divisible distributions** **229**
12.1 Definitions . . . 229
12.2 Some examples . . . 229
12.3 Some properties . . . 230
12.4 Levy-Khintchine representation of infinitely divisible c.f.s . . . 231
12.4.1 Proof of necessity . . . 232
12.4.2 Proof of sufficiency . . . 234
12.4.3 Several examples . . . 235
12.5 Levy formula . . . 236
12.6 Kolmogorov formula . . . 238
12.7 Relationship between the sum of independent r.v. and i.d. . . . 238
12.8 Exercises . . . 238

## Notation and abbreviations

| | |
|---|---|
| a.s. | almost surely |
| d.f. | distribution function |
| i.i.d. | independent and identically distributed |
| i.o. | infinitely often |
| w.r.t. | with respect to |
| W.L.O.G. | without loss of generality |
| $\mathcal{R}$ | the set of real numbers |
| $I_A$ or $I\{A\}$ | indicator function of the set $A$ |
| $\blacksquare$ | end of a proof |

Unless otherwise specified, all limits are taken as $n \to \infty$.

$$a_n = O(b_n) \iff |a_n/b_n| \le C \iff \limsup |a_n/b_n| < \infty$$

$$a_n = o(b_n) \iff a_n/b_n \to 0$$

$$a_n \sim b_n \iff a_n/b_n \to 1$$

# Chapter 1

# Set Theory

## 1.1 Sets

**Definition**:

$\Omega$: a space (a nonempty reference set). $\emptyset$: an empty set.

**A (sub)set** is a collection of objects (or elements) in $\Omega$, denoted by $A, B, C$, ...

**A class (or family)** is a collection of subsets of $\Omega$, denoted by

$\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$, ...

$\omega \in A$ : $\omega$ is an element of $A$.

$A \subset B$ : the set $A$ is contained in the set $B$.

Convention: $\emptyset \subset A$ for any set $A$.

## 1.2 Basic set operations

Let $A$ and $B$ be subsets of $\Omega$, define

$$\begin{aligned} A^c &= \{\omega : \omega \notin A\} \qquad \text{(complement)} \\ A \cup B &= \{\omega : \omega \in A, \text{ or } \omega \in B\} \qquad \text{(union)} \\ A \cap B &= \{\omega : \omega \in A, \text{ and } \omega \in B\} \qquad \text{(intersection)} \\ A - B &= \{\omega : \omega \in A, \text{ and } \omega \notin B\} \qquad \text{(difference)} \\ A \Delta B &= (A - B) \cup (B - A) \qquad \text{(symmetric difference).} \end{aligned}$$

(Note: sometimes we write $A \cap B = AB$, the product of $A$ and $B$.)

**Venn diagrams** are often helpful.

## 1.3 Operations of sequence of sets

Let $A_1, A_2, ...$ be subsets of $\Omega$, denote their union and intersection by

$$\begin{aligned}\cup_{n=1}^{\infty} A_n &= \{\omega : \omega \in A_n \text{ for some } n\}\\ \cap_{n=1}^{\infty} A_n &= \{\omega : \omega \in A_n \text{ for all } n\}.\end{aligned}$$

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

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

**Definition**: Let $A, A_1, A_2, ...$ be subsets of $\Omega$. Define

(a) Infinitely often (i.o.)

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

(b) Ultimately (ult.)

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

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

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

Theorem **1.3.1** *We have*

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

*Proof.* By definition,

$$\begin{aligned}\liminf_n A_n &= \{\omega : \omega \in A_n \text{ for all but finitely many values of } n\}\\ &\subset \{\omega : \omega \in A_n \text{ for infinitely many values of } n\}\\ &= \limsup_n A_n. \quad \blacksquare\end{aligned}$$

**Example**. ($\liminf_n A_n$ and $\limsup_n A_n$ may not be equal). Specify $\liminf_n A_n$ and $\limsup_n A_n$ when $A_{2j} = B$, and $A_{2j-1} = C$, $j = 1, 2, ....$

**Solution**. Clearly,

$$\begin{aligned}&\liminf_n A_n = \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n = \cup_{k=1}^{\infty}(B \cap C) = B \cap C,\\ &\limsup_n A_n = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n = \cap_{k=1}^{\infty}(B \cup C) = B \cup C. \quad \blacksquare\end{aligned}$$

THEOREM **1.3.2 (Monotone sequence of sets converges)** .

*a) If $A_1 \subset A_2 \subset A_3 ...$, then $A_n \to A = \cup_{k=1}^{\infty} A_k$, written as $A_n \uparrow A$.*

*b) If $A_1 \supset A_2 \supset A_3 ...$, 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,

$$\begin{aligned} \liminf_n A_n &= \cup_{k=1}^{\infty} \left(\cap_{n=k}^{\infty} A_n\right) = \cup_{k=1}^{\infty} A_k = A \\ \limsup_n A_n &= \cap_{k=1}^{\infty} \left(\cup_{n=k}^{\infty} A_n\right) = \cap_{k=1}^{\infty} \left(\cup_{n=1}^{\infty} A_n\right) = \cap_{k=1}^{\infty} A = A. \quad \blacksquare \end{aligned}$$

COROLLARY **1.3.1** *Let $A_1, A_2, ...$ be subsets of $\Omega$.*

$$\begin{aligned} \{A_n, i.o.\} &= \limsup_n A_n = \lim_{k\to\infty} \cup_{n=k}^{\infty} A_n \\ \{A_n, ult.\} &= \liminf_n A_n = \lim_{k\to\infty} \cap_{n=k}^{\infty} A_n \end{aligned}$$

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

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

## 1.4 Indicator functions

Let $A \subset \Omega$, define

$$\begin{aligned} I_A(\omega) = I\{\omega \in A\} &= 1 \qquad \text{for } \omega \in A \\ & \quad\; 0 \qquad \text{for } \omega \in A^c. \end{aligned}$$

Thus, $I_A$ "indicates" whether $A$ occurs or not, depending on whether it is 1 or 0. Therefore, the next theorem is obvious.

The indicator function transforms operations of sets into algebraic operations of 1's and 0's, which are often easier to deal with.

THEOREM **1.4.1** *$\forall A, B \subset \Omega$, we have*

$$\begin{aligned} A = B \quad &\Longleftrightarrow \quad I_A = I_B, \\ & \qquad (\textit{meaning } I_A(\omega) = I_B(\omega) \textit{ for all } \omega \in \Omega) \\ A \subset B \quad &\Longleftrightarrow \quad I_A \le I_B, \\ A = \emptyset \quad \textit{or} \quad \Omega, \quad \textit{resp.} \quad &\Longleftrightarrow \quad I_A = 0 \quad \textit{or} \quad 1, \quad \textit{resp.} \end{aligned}$$

*Proof.* Let us prove the first one only. Clearly, $A = B$ implies that $I_A(\omega) = I_B(\omega)$ for all $\omega \in \Omega$. On the other hand, if $A \neq B$, then $\exists \omega \in A$ but $\omega \notin B$, or $\exists \omega \in B$ but $\omega \notin A$. In either case, for that $\omega$, we have $I_A(\omega) \neq I_B(\omega)$. The proof is completed. $\blacksquare$

THEOREM **1.4.2**

$$\begin{aligned} I_{A\cap B} &= \min\{I_A, I_B\} = I_A I_B \\ I_{A\cup B} &= \max\{I_A, I_B\} = I_A + I_B - I_{A\cap B} \\ &= I_A + I_B - I_A I_B \\ I_{A^c} &= 1 - I_A \\ I_{A-B} &= I_{A\cap B^c} = I_A\left(1 - I_B\right) \end{aligned}$$

$$\begin{aligned}
I_{A\Delta B} &= |I_A - I_B| \\
I_{\liminf_n A_n} &= \liminf_n I_{A_n} \\
I_{\limsup_n A_n} &= \limsup_n I_{A_n} \\
I_{A\cup B} &\leq I_A + I_B \\
I_{\cup_1^\infty A_n} &\leq \sum_1^\infty I_{A_n}
\end{aligned}$$

*Proof.* We shall only prove a couple of these relationships.

(1) **First we show that** $I_{A\cup B} = \max\{I_A, I_B\}$. $\forall \omega \in \Omega$, if $I_{A\cup B}(\omega) = 1$, then $\omega \in A \cup B$. So either $\omega \in A$ or $\omega \in B$. So either $I_A(\omega) = 1$, or $I_B(\omega) = 1$. Thus $I_{A\cup B}(\omega) = \max\{I_A(\omega), I_B(\omega)\}$. Similarly, we can show that this is true if $I_{A\cup B}(\omega) = 0$. Thus, we proved $I_{A\cup B} = \max\{I_A, I_B\}$.

(2) **Next we show that** $I_{\liminf_n A_n} = \liminf_n I_{A_n}$. $\forall \omega \in \Omega$, if $I_{\liminf_n A_n}(\omega) = 1$, then $\omega \in \liminf_n A_n = \{A_n, ult.\}$. That is, $\omega$ is contained in all but finite many values of $n$. Thus, $I_{A_n}(\omega) = 1$ except for a finite many values of $n$. It implies that $\liminf_n I_{A_n}(\omega) = \lim_n I_{A_n}(\omega) = 1$. On the other hand, if $I_{\liminf_n A_n}(\omega) = 0$, then $\omega \in (\liminf_n A_n)^c = \limsup_n A_n^c$ (by using the DeMorgan Law) $= \{A_n^c, i.o.\}$. Therefore, $I_{A_n^c}(\omega) = 1$ i.o. and so $\limsup_n I_{A_n^c}(\omega) = 1$. Hence, $\limsup_n [1 - I_{A_n}(\omega)] = 1 - \liminf_n I_{A_n}(\omega) = 1$. So $\liminf_n I_{A_n}(\omega) = 0 = I_{\liminf_n A_n}(\omega)$. The proof is complete.

(3) **Finally, we shall prove** $I_{A\Delta B}(\omega) = |I_A(\omega) - I_B(\omega)|$. Consider 4 cases: $\forall \omega \in \Omega$,

(a) $I_A(\omega) = 1$, $I_B(\omega) = 1$, then RHS = 0. Also we note that $\omega \in A$ and $\omega \in B$, thus $\omega \in A \cap B$. That is, $\omega \notin A\Delta B$, implying LHS=0. So LHS = RHS =0.

Similarly, we can show the following:

(b) $I_A(\omega) = 1$, $I_B(\omega) = 0$, then LHS=1=RHS.

(c) $I_A(\omega) = 0$, $I_B(\omega) = 1$, then LHS=1=RHS.

(d) $I_A(\omega) = 0$, $I_B(\omega) = 0$, then LHS=0=RHS.

This completes the proof. ∎

Indicator functions can sometimes facilitate set operations, as illustrated below.

**Example**. Show $(A\Delta B)\Delta C = A\Delta(B\Delta C)$.

*Proof.* Note

$$I_{LHS} = |I_{A\Delta B} - I_C| = ||I_A - I_B| - I_C|,$$
$$I_{RHS} = |I_A - I_{B\Delta C}| = |I_A - |I_B - I_C||.$$

If $I_B = 0$, then $I_{LHS} = |I_A - I_C| = I_{RHS}$.

If $I_B = 1$, then $I_{RHS} = ||I_B - I_C| - I_A| = |1 - I_C - I_A| = I_{LHS}$. ∎

THEOREM **1.4.3**

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

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

$$I_{\cup_1^n A_j} = s_1 - s_2 + s_3 - \ldots + (-1)^{n-1} s_n. \tag{4.1}$$

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 (4.1) 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 $\omega$ belongs to exactly $m$ of the sets $A_1, ..., A_n$. Then $s_1(\omega) = m$, $s_2(\omega) = \binom{m}{2}$, ..., $s_m(\omega) = 1$, $s_{m+1}(\omega) = ... = s_n(\omega) = 0$, whence

$$\begin{aligned} s_1 - s_2 + s_3 - ... + (-1)^{n-1} s_n &= \binom{m}{1} - \binom{m}{2} + ... + (-1)^{m-1}\binom{m}{m} \\ &= \binom{m}{0} - (1-1)^m = 1 = I_{\cup_1^n A_j}. \end{aligned}$$

This completes the proof. ∎

REMARK **1.4.1** *By taking expectation (to be studied later), we get the "inclusion-exclusion formula":*

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

## 1.5 Semi-algebras, Algebras, and $\sigma$-algebras

### 1.5.1 Definitions

Let $\Omega$ be a space.

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

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

(ii). if $A \in \mathcal{S}$, then $A^c$ is a finite disjoint union of sets in $\mathcal{S}$, i.e.,

$$A^c = \sum_{i=1}^n A_i, \quad \text{where} \quad A_i \in \mathcal{S}, \quad A_i \cap A_j = \emptyset, \ i \ne j.$$

**Definition**: A nonempty class $\mathcal{A}$ of subsets of $\Omega$ is an **algebra** on $\Omega$ if

(i). $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$,

(ii). $A_1 \cup A_2 \in \mathcal{A}$ whenever $A_j \in \mathcal{A}$, $j = 1, 2$.

(i.e. $\mathcal{A}$ is closed under complement and *finite* union.)

**Definition**: A nonempty class $\mathcal{A}$ of subsets of $\Omega$ is a **$\sigma$-algebra** on $\Omega$ if

(i). $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$,

(ii). $\cup_{n=1}^\infty A_n \in \mathcal{A}$ whenever $A_n \in \mathcal{A}$, $n \ge 1$.

(i.e. $\mathcal{A}$ is closed under complement and *countable* union.) The pair $(\Omega, \mathcal{A})$ is called a **measurable space**. The sets of $\mathcal{A}$ are called **measurable sets**.

**[Note: $\sigma$ or $\Sigma$ = "(countable) sum" = "(countable) union"].**

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

**Proof.** Assume $\mathcal{A}$ is a nonempty algebra or $\sigma$-algebra. $\exists A \in \mathcal{A}$, so $A^c \in \mathcal{A}$. Thus, $\Omega = A \cup A^c \in \mathcal{A}$ and $\emptyset = \Omega^c \in \mathcal{A}$.

For the second part, consider $\mathcal{S} = \{\emptyset, A, A^c\}$. It is clearly a semi-algebra, but does not contain $\Omega$.

REMARK **1.5.2** *$\mathcal{A}$ is an algebra $\iff$ (1). $\Omega \in \mathcal{A}$. (2). $A, B \in \mathcal{A}$ implies $A - B \in \mathcal{A}$.*

**Proof.** "$\Rightarrow$" part is easy. We now show "$\Leftarrow$" part. Let $A, B \in \mathcal{A}$. Since $\Omega \in \mathcal{A}$, thus $A^c = \Omega - A \in \mathcal{A}$. Also

$$A \cap B = A - (A - B) \in \mathcal{A}, \qquad \text{as } (A - B) \in \mathcal{A}.$$

So $\mathcal{A}$ is closed under complements and finite intersection thus union. Then $\mathcal{A}$ is an algebra.
∎

### 1.5.2 Relationships

An algebra is a semi-algebra, and a $\sigma$-algebra is an algebra. The reverse may not be true. See the next two examples.

**Example 1: A semi-algebra may not be an algebra.**

Let $\Omega = (-\infty, \infty]$. Let $\mathcal{S}$ be a collection of half open intervals in $\Omega$, i.e.,

$$\mathcal{S} = \{(a, b] : -\infty \le a \le b \le \infty\}.$$

Then $\mathcal{S}$ is a semi-algebra, but NOT an algebra. (It is easy to define Lebesgue measure (i.e. length in this case) on $\mathcal{S}$, and then generalize to the Borel $\sigma$-algebra.)

*Proof.* For the first part, note that $(a, a] = \emptyset \in \mathcal{S}$. It is easy to check that the intersection of two half open intervals are still half-intervals (including $\emptyset$). So it is closed under intersection. On the other hand, if $A = (a, b] \in \mathcal{S}$, then $A^c = (-\infty, a] + (b, \infty]$, the union of two disjoint half open intervals. Thus, $\mathcal{S}$ is a semialgebra.

For the second part, note that $(n, n+1] \in \mathcal{S}$, but $(0, 1] \cup (2, 3] \notin \mathcal{S}$. So it is not closed under finite union, thus not an algebra.

**Example 2: An algebra may not be a $\sigma$-algebra.**

Again, let $\Omega = (-\infty, \infty]$. Define

$$\begin{aligned} \overline{\mathcal{S}} &= \{\cup_{i=1}^{n} (a_i, b_i] : -\infty \le a_i \le b_i \le \infty\} \\ &= \{\cup_{i=1}^{m} (c_i, d_i] : -\infty \le c_1 \le d_1 \le c_2 \le d_2 ... \le d_m \le \infty\}. \\ &\quad \text{(After ordering and merging of intervals.)} \end{aligned}$$

Then $\overline{\mathcal{S}}$ is an algebra, but NOT a $\sigma$-algebra.

*Proof.* The first part is easy. For the second part, note that

$$\cup_{n=1}^{\infty} (0, 1 - 1/n] = (0, 1) \notin \overline{\mathcal{S}},$$

or as another example

$$\cup_{n=1}^{\infty} (n - 1/4, n + 1/4] \notin \overline{\mathcal{S}}.$$

So it is not closed under countable union, thus not a $\sigma$-algebra.

### 1.5.3 Some special $\sigma$-algebras.

(1). **Trivial $\sigma$-algebra**: $\{\emptyset, \Omega\}$. This is the smallest $\sigma$-algebra.

(2). **Power set:** = all subsets of $\Omega$, denoted by $\mathcal{P}(\Omega)$. This is the largest $\sigma$-algebra, often too big to define probability.

(3). **The smallest $\sigma$-algebra containing** $A \in \Omega$: $\{\emptyset, A, A^c, \Omega\}$.

### 1.5.4 How to generate algebras from semi-algebras

THEOREM **1.5.1** *If $\mathcal{S}$ is a semi-algebra, then*

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

*is an algebra, called the* **algebra generated by** $\mathcal{S}$. *We sometimes also denote* $\overline{\mathcal{S}}$ *by* $\mathcal{A}(\mathcal{S})$.

**Proof.** Let $A, B \in \overline{\mathcal{S}}$, then $A = \sum_{i=1}^{m} A_i$ and $B = \sum_{j=1}^{n} B_j$ with $A_i, B_j \in \mathcal{S}$. Clearly, $A_i \cap B_j \in \mathcal{S}$ by the definition of $\mathcal{S}$. Thus,

$$A \cap B = \sum_{i=1}^{m} \sum_{j=1}^{n} A_i \cap B_j \in \overline{\mathcal{S}}.$$

So $\overline{\mathcal{S}}$ is closed under (finite) intersection.

As for complements, note that $A^c = (\sum_{i=1}^{m} A_i)^c = \cap_{i=1}^{m} A_i^c$. The definition of $\mathcal{S}$ implies $A_i^c \in \overline{\mathcal{S}}$. But we have just shown that $\overline{\mathcal{S}}$ is closed under (finite) intersection, so $A^c \in \overline{\mathcal{S}}$. So $\overline{\mathcal{S}}$ is also closed under complement. This completes our proof. ∎

REMARK **1.5.3** *It is easy to define (Lesbegue) measure on the semialgebra $\mathcal{S}$, and then easily extend it to the algebra $\overline{\mathcal{S}}$. Can we extend it further to some $\sigma$-algebra. We shall see later that the answer to this question is affirmative. Since there are many $\sigma$-algebras, we shall first consider the smallest one containing $\mathcal{S}$. This is our next topic: the* **$\sigma$-algebra generated by** $\mathcal{S}$, *or more generally by any class.*

## 1.6 Generated classes (Minimal classes)

(Probability) Measure is defined on $\sigma$-algebras. Given a class $\mathcal{A}$ of subsets of $\Omega$, there always exists a *unique minimal* $\sigma$-algebra containing $\mathcal{A}$. Minimality is important since the power set $\mathcal{P}(\Omega)$ is too large unless $\Omega$ is finite or countable.

LEMMA **1.6.1** *Let $\{\mathcal{A}_\gamma : \gamma \in \Gamma\}$ be a collection of $\sigma$-algebras. Then $\mathcal{A} = \cap_{\gamma\in\Gamma}\mathcal{A}_\gamma$ is also a $\sigma$-algebra.*

*[i.e., $\sigma$-algebras are closed under (possibly uncountable) intersection.]*

*Proof.* (1). $A \in \mathcal{A}$ implies $A \in \mathcal{A}_\gamma$ for all $\gamma \in \Gamma$, which in turn implies $A^c \in \mathcal{A}_\gamma$ for all $\gamma \in \Gamma$. Therefore, $A^c \in \mathcal{A}$. That is, $\mathcal{A}$ is closed under complements.

(2). If $A_i \in \mathcal{A}$, $i \geq 1$, then $A_i \in \mathcal{A}_\gamma$ for all $i \geq 1$ and all $\gamma \in \Gamma$. Hence, $\cup_{i=1}^{\infty} A_i \in \mathcal{A}_\gamma$ for all $\gamma \in \Gamma$. So $\cup_{i=1}^{\infty} A_i \in \cap_{\gamma\in\Gamma}\mathcal{A}_\gamma = \mathcal{A}$. So $\mathcal{A}$ is closed under countably infinite union. This completes the proof. ∎

THEOREM **1.6.1** *For any class $\mathcal{A}$, there exists a unique minimal $\sigma$-algebra containing $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, called the $\sigma$-algebra generated by $\mathcal{A}$. In other words,*

*a)* $\mathcal{A} \subset \sigma(\mathcal{A})$,

*b) For any $\sigma$-algebra $\mathcal{B}$ with $\mathcal{A} \subset \mathcal{B}$, $\sigma(\mathcal{A}) \subset \mathcal{B}$,*

*and $\sigma(\mathcal{A})$ is unique.*

**Proof. Existence.** Let $Q_{\mathcal{A}} = \{\mathcal{B} : \mathcal{B} \supset \mathcal{A},\ \mathcal{B}$ is a $\sigma$-algebra on $\Omega\}$. Clearly, $Q_{\mathcal{A}}$ is not empty since it contains the power set $\mathcal{P}(\mathcal{A})$. Define

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{B}\in Q_{\mathcal{A}}} \mathcal{B}.$$

From the above lemma, $\sigma(\mathcal{A})$ is a $\sigma$-algebra.

**Uniqueness.** Let $\sigma_1(\mathcal{A})$ be another $\sigma$-algebra satisfying (a) and (b). By definition, $\sigma(\mathcal{A}) \subset \sigma_1(\mathcal{A})$. By symmetry, $\sigma_1(\mathcal{A}) \subset \sigma(\mathcal{A})$. Thus, $\sigma(\mathcal{A}) = \sigma_1(\mathcal{A})$. ∎

Remark **1.6.1** *Similar theorem holds for algebras, monotone classes, $\lambda$-class, and $\pi$-class, see later sections.*

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

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

**Proof.** By definition, $\mathcal{S} \subset \overline{\mathcal{S}} \subset \sigma(\overline{\mathcal{S}})$. So $\sigma(\overline{\mathcal{S}})$ is a $\sigma$-algebra containing $\mathcal{S}$, thus $\sigma(\mathcal{S}) \subset \sigma(\overline{\mathcal{S}})$.

We now show that $\sigma(\overline{\mathcal{S}}) \subset \sigma(\mathcal{S})$. For any given $A \in \overline{\mathcal{S}}$, by definition, we have $A = \sum_{i=1}^{n} A_i$, where $A_i \in \mathcal{S} \subset \sigma(S)$. Therefore, $A \in \sigma(S)$. This shows that $\overline{\mathcal{S}} \subset \sigma(\mathcal{S})$. So $\sigma(\mathcal{S})$ is a $\sigma$-algebra containing $\overline{\mathcal{S}}$, thus $\sigma(\overline{\mathcal{S}}) \subset \sigma(\mathcal{S})$. This completes our proof.

We now introduce the most important $\sigma$-algebra in measure theory.

**Definition:** The smallest $\sigma$-algebra generated by the collection of all finite open intervals on the real line $\mathcal{R} = (-\infty, \infty)$ (or $\mathcal{R} = (-\infty, \infty]$ or $\mathcal{R} = [-\infty, \infty]$ as in Chow and Teicher, page 11.) is called the **Borel $\sigma$-algebra**, denoted by $\mathcal{B}$. The elements of $\mathcal{B}$ are called **Borel sets**. The pair $(\mathcal{R}, \mathcal{B})$ is called the (1-dimensional) **Borel measurable space**.

Remark **1.6.2** .

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

(2). *$k$-dimensional Borel $\sigma$-algebra, Borel sets and measurable space can be similarly defined.*

(3). *Borel $\sigma$-algebra can be generated by many other ways. For instance, it can be generated by the class of all finite closed intervals, or the class of all open sets, or the class of all closed sets.*

(4). *For $A \in \mathcal{B}$, let*

$$\mathcal{B}_A = \{B \cap A : B \in \mathcal{B}\} = \mathcal{B} \cap A.$$

*Then $(A, \mathcal{B}_A)$ is a measurable space, and $\mathcal{B}_A$ is called the Borel $\sigma$-algebra on $A$.*

**Example: Borel $\sigma$-algebras generated by semialgebras**.

(1). Let $R_0 = (-\infty, \infty)$ (the real line), and define

$$\mathcal{S}_0 = \{A : A = R_0 \text{ or } [a, b) \text{ or } (-\infty, b) \text{ or } [a, \infty), \quad a, b \in R\}.$$

Then $\mathcal{S}_0$ is a semi-algebra on $R_0$. The $\sigma(\mathcal{S}_0)$ is called the **Borel $\sigma$-algebras on the real line** $R$.

(2). Let $R_1 = (-\infty, \infty]$ (half extended real line), and define

$$\mathcal{S}_1 = \{(a, b], \quad a, b \in R_1\}.$$

Then $\mathcal{S}_1$ is a semi-algebra on $R_1$. The $\sigma(\mathcal{S}_1)$ is called the **Borel $\sigma$-algebras on the half extended real line** $R_1$.

(3). Let $R_2 = [-\infty, \infty]$ (the extended real line), and define

$$\mathcal{S}_2 = \{(a, b], \quad a, b \in R_2\} \cup \{R_2, \{\infty\}, \{-\infty\}\}.$$

Then $\mathcal{S}_2$ is a semi-algebra on $R_2$. The $\sigma(\mathcal{S}_2)$ is called the **Borel $\sigma$-algebras on the extended real line** $R_2$.

Note that the introduction of $\infty$ and $-\infty$ simplifies the structures of semialgebras.

## 1.7 Monotone class ($m$-class), $\pi$-class, and $\lambda$-class

Checking $\mathcal{A}$ to be a $\sigma$-algebra directly may not be easy in practice. This can be made easier by introducing new and more primitive classes such as monotone classes, $\pi$-classes, and $\lambda$-classes.

### 1.7.1 Definitions

**Definition**: Let $\mathcal{A}$ be a nonempty class of subsets of $\Omega$.

(1). $\mathcal{A}$ is said to be a *monotone class* ($m$-class) on $\Omega$ 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 1.3.2).

(2). $\mathcal{A}$ is a $\pi$*-class on* $\Omega$ if

$$A \cap B \in \mathcal{A}, \quad \text{whenever } A, B \in \mathcal{A}.$$

(i.e. a $\pi$-class is closed under *finite intersection*. Note $\pi \approx \prod$.)

(3). $\mathcal{A}$ is a $\lambda$*-class on* $\Omega$ if

(i). $\Omega \in \mathcal{A}$,
(ii). $A - B \in \mathcal{A}$ for $A, B \in \mathcal{A}$ and $B$ is a proper subset of $A$, (i.e. $B \subset A$.)
(iii). $\lim A_n \in \mathcal{A}$ for every increasing sequence $A_n \in \mathcal{A}$, $n \geq 1$.
(i.e. a $\lambda$-class contains $\Omega$ and is closed under *proper difference* and *countable increasing union*. Note $\lambda \approx$ "increasing" limit).
($\lambda = lamda = lam + da \approx lim + diff$.  Also $\lambda =$ "Large", see Theorem 1.8.3)

**Remark.** The notion of a $\pi$-class and $\lambda$-class comes from Dynkin (1961), *Theory of Markov Processes*, Prentice-Hall.

THEOREM **1.7.1** *If $\mathcal{A}$ is a $\lambda$-class, it is an $m$-class.*

**Proof.** Part (iii) in the definition of $\lambda$-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 $\lambda$-class, $\mathcal{A}$ is closed under the limit of decreasing sequence of sets in $\mathcal{A}$. Thus, it is an $m$-class. ∎

### 1.7.2 Relationships with $\sigma$-algebras

THEOREM **1.7.2** *Suppose $\mathcal{A}$ is an algebra on $\Omega$. Then*

$$\mathcal{A} \text{ is an } m\text{-class} \iff \mathcal{A} \text{ is a } \sigma\text{-algebra.}$$

**Proof.** "$\Longleftarrow$". Suppose that $\mathcal{A}$ is a $\sigma$-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 1.3.2, $\lim_n A_n = \cup_{n=1}^{\infty} A_n$ (or $\cap_{n=1}^{\infty} A_n$), which belongs to $\mathcal{A}$ (as $\mathcal{A}$ is a $\sigma$-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 $\sigma$-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. ∎

A $\sigma$-algebra must be an $m$-class, but an $m$-class may not be an algebra, let alone a $\sigma$-algebra (see the next example). But if $\mathcal{A}$ is already an algebra, then an $m$-class is a $\sigma$-algebra. *To check an algebra $\mathcal{A}$ is a $\sigma$-algebra, it suffices to show that it is an $m$-class, which may be easier in practice.*

**Example**. Let $\Omega = R$, $\mathcal{A} = \{(-\infty, a), (-\infty, a], (-\infty, \infty), \emptyset : a \in R\}$. So $\mathcal{A}$ is an $m$-class, but is not an algebra (thus not a $\sigma$-algebra) since it is not closed under differences. e.g. $A = (-\infty, a]$, $B = (-\infty, b)$, $b > a$, then $B - A = (a, b) \notin \mathcal{A}$.

THEOREM **1.7.3** *$\mathcal{A}$ is a $\sigma$-algebra iff it is both a $\lambda$-class and $\pi$-class.*

**Proof.** "$\Longrightarrow$". Suppose that $\mathcal{A}$ is a $\sigma$-algebra. Clearly, it is a $\pi$-class. Now we show that it is also a $\lambda$-class.

(i). It is clear that $\Omega \in \mathcal{A}$.

(ii). $A - B = A \cap B^c \in \mathcal{A}$.

(iii). For every increasing sequence $A_n \in \mathcal{A}$, $n \geq 1$, by Theorem 1.3.2, $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$, which belongs to $\mathcal{A}$ (as $\mathcal{A}$ is a $\sigma$-algebra).

So it is a $\lambda$-class.

"$\Longleftarrow$". Now suppose that $\mathcal{A}$ is both a $\lambda$-class and $\pi$-class. First, $\forall\, A \in \mathcal{A}$,

$$A^c = \Omega - A \in \mathcal{A}. \quad \text{(as } \mathcal{A} \text{ is a } \lambda\text{-class)} \tag{7.2}$$

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

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

It follows that $\mathcal{A}$ is closed under *finite union.* Since $B_n$ is increasing, by the property of $\lambda$-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. ∎

### 1.7.3 Minimal $m$-class, $\lambda$-class and $\pi$-class

For any class $\mathcal{A}$, we have seen that there exists a unique minimal $\sigma$-algebra containing $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, (the $\sigma$-algebra generated by $\mathcal{A}$). The same holds true for $m$-class, $\lambda$-class and $\pi$-class.

LEMMA **1.7.1** *The power set $\mathcal{P}(\mathcal{A})$ is an $m$-class (or $\lambda$-class, or $\pi$-class).*

The proof is trivial.

LEMMA **1.7.2** *Let $\{\mathcal{A}_\gamma : \gamma \in \Gamma\}$ be $m$-classes (or $\lambda$-classes or $\pi$-classes). Then $\mathcal{A} = \cap_{\gamma\in\Gamma}\mathcal{A}_\gamma$ is also an $m$-class (or $\lambda$-class or $\pi$-class).*

The proof is similar to that of $\sigma$-algebras.

THEOREM **1.7.4** *For any class $\mathcal{A}$, there exists a unique minimal $m$-class (or $\lambda$-class, or $\pi$-class) containing $\mathcal{A}$, denoted by $m(\mathcal{A})$ (or $\lambda(\mathcal{A})$, or $\pi(\mathcal{A})$), called the $m$-class (or $\lambda$-class, or $\pi$-class) generated by $\mathcal{A}$. In other words,*

(a) $\mathcal{A} \subset m(\mathcal{A})$, *(or $\lambda$-class, or $\pi$-class).*

(b) *For any $m$-class (or $\lambda$-class, or $\pi$-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.*

The proof is left as an exercise.

### 1.7.4 Graphical illustration of different classes

$$
\begin{array}{ccccc}
\sigma-algebra & \Longleftrightarrow & \lambda-class & + & \pi-class, \\
\Downarrow & & \Downarrow & & \\
algebra & & m-class & & \\
\Downarrow & & & & \\
semi-algebra & & & & \\
\Downarrow & & & & \\
\pi-class & & & &
\end{array}
$$

## 1.8 The Monotone Class Theorem (MCT)

THEOREM **1.8.1** *Let $\mathcal{A}$ be an algebra. Then,*

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

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

**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 $\sigma$-algebra. In fact, by Theorem 1.7.2, we only need to show

$$m(\mathcal{A}) \text{ is an algebra.} \qquad (*)$$

**Proof of (*):** Define

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

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

*Proof.* First we look at $\mathcal{C}_1$. Suppose $A_j \in \mathcal{C}_1$, $j \geq 1$, and $A_1 \subset A_2 \subset A_3 ...$ (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 $(A_1 \cap B) \subset (A_2 \cap B) \subset (A_3 \cap B) \subset ....$ 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

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

Therefore, $\lim_{j\to\infty} A_j = \cup_{j=1}^{\infty} A_j \in \mathcal{C}_1$. Similarly, for a decreasing sequence of sets $D_1 \supset D_2 \supset D_3 \supset ...$ 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 $(\cup_{j=1}^{\infty} A_j)^c = \cap_{j=1}^{\infty} A_j^c$, $\quad (\cap_{j=1}^{\infty} A_j)^c = \cup_{j=1}^{\infty} A_j^c$, 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 $\mathcal{C}_i \supset m(\mathcal{A})$, $i = 1, 2, 3$ .

*Proof.* (i). 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

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

(ii). 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.

(iii). 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 **1.8.2** *Let $\mathcal{A}$ be a $\pi$-class.*

(1) *Then $\lambda(\mathcal{A}) = \sigma(\mathcal{A})$.*

(2) *If $\mathcal{B}$ is an $\lambda$-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 $\lambda$-class and $\lambda(\mathcal{A})$ is the smallest $\lambda$-class.

Now let us show that $\lambda(\mathcal{A}) \supset \sigma(\mathcal{A})$. From Theorem 1.7.3, it suffices to show that $\lambda(\mathcal{A})$ is a $\pi$-class (which implies that $\lambda(\mathcal{A})$ is a $\sigma$-algebra containing $\mathcal{A}$, thus $\lambda(\mathcal{A}) \supset \sigma(\mathcal{A})$). To this end, define

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

Clearly, we can show that

$$(a).\ \mathcal{A} \subset \mathcal{C}_1, \qquad\qquad (b).\ \mathcal{C}_1 \text{ is a } \lambda\text{-class.}$$

**Proof of (a) and (b)**. Let $A \in \mathcal{A}$. Since $\mathcal{A}$ is a $\pi$-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 $\lambda$-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

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

Then, $\mathcal{A} \subset \mathcal{C}_2$. We can also show that $\mathcal{C}_2$ is a $\lambda$-class (Why?). Therefore, $\mathcal{C}_2$ is a $\lambda$-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 $\pi$-class. ∎

The following theorem is the direct consequence of the last two theorems.

THEOREM **1.8.3 (Monotone Class Theorem)** .

*Let $\mathcal{A} \subset \mathcal{B}$ be two classes on $\Omega$.*

(1). *If $\mathcal{A}$ is a $\pi$-class, and $\mathcal{B}$ is a $\lambda$-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 $\lambda$-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 $\mathcal{B} \supset m(\mathcal{A}) = \sigma(\mathcal{A})$. ∎

**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:

(i) Define $\mathcal{B} = \{B : B \text{ has property } \mathcal{P}\}$, so that $\mathcal{A} \subset \mathcal{B}$.

(ii) Show that

(1). $\mathcal{A}$ is a $\pi$-class, and $\mathcal{B}$ is a $\lambda$-class, or

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

(iii) From the Monotone Class Theorem, we get $\sigma(\mathcal{A}) \subset \mathcal{B}$.

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

## 1.9 Product Spaces

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

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

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

Moreover, if $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$, they are dubbed **measurable rectangles** or **rectangles with measurable sides**.

(2). $n$-dim **product $\sigma$-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**:

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

The following lemma is easy to check.

LEMMA **1.9.1** *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 $\pi$-class.*

THEOREM **1.9.1** *Let $(\Omega_i, \mathcal{A}_i), \leq i \leq n$ be measurable spaces, and*

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

*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 $\pi$-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 $\pi$-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

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

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

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

Therefore,

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

(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

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

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 **1.9.1** *The $\sigma$-algebra $\prod_{i=1}^n \mathcal{A}_i$, generated by the rectangles with measurable sides, is also the $\sigma$-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

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

## 1.10 Exercise

1. Show that

   (i) if $\mathcal{A}_1 \subset \mathcal{A}_2 \subset ......$ are $\sigma$-algebras, then $\cup_{n=1}^{\infty} \mathcal{A}_n$ is an algebra.
   (ii) [Optional] Give an example to show that $\cup_{n=1}^{\infty} \mathcal{A}_n$ may not be a $\sigma$-algebra.

2. If $\mathcal{A}$ is an algebra, and $\sum_1^{\infty} A_n \in \mathcal{A}$ for every disjoint $\{A_n, n \geq 1\}$ in $\mathcal{A}$, then $\mathcal{A}$ is a $\sigma$-algebra.

3. Let $g$ be a minimal operation, e.g. $g = \sigma, m, \lambda, \pi$. Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of subsets of $\Omega$. Show that (only need to show one of them)

   (1). $g(g(\mathcal{A})) = g(\mathcal{A})$.
   (2). $A \subset g(\mathcal{B}), \implies g(\mathcal{A}) \subset g(\mathcal{B})$.
   (3). $g(\mathcal{A}) \subset \mathcal{B}, \implies g(\mathcal{A}) \subset g(\mathcal{B})$
   (4). $\mathcal{A} \subset \mathcal{B}, \implies g(\mathcal{A}) \subset g(\mathcal{B})$.

4. Show that the power set $\mathcal{P}(\Omega)$ is a $\sigma$-algebra (hence it is also a semialgebra, an algebra, a $\lambda$-class, an $m$-class, and a $\pi$-class).

5. Define

$$\begin{aligned} \mathcal{C}_1 &= \{A : A \subset \Omega,\ A \cap B \in \lambda(\mathcal{A}) \text{ for all } B \in \mathcal{A}\}, \\ \mathcal{C}_2 &= \{B : B \subset \Omega,\ A \cap B \in \lambda(\mathcal{A}) \text{ for all } A \in \lambda(\mathcal{A})\}. \end{aligned}$$

   Show that $\mathcal{C}_1$ and $\mathcal{C}_2$ are $\lambda$-classes

6. Define $\sup_{\gamma \in \Gamma} A_\gamma$ and $\inf_{\gamma \in \Gamma} A_\gamma$ to be the smallest upper bound and largest lower bound of $\{A_\gamma : \gamma \in \Gamma\}$, respectively. Then show

$$\sup_{\gamma \in \Gamma} A_\gamma = \cup_{\gamma \in \Gamma} A_\gamma, \qquad \inf_{\gamma \in \Gamma} A_\gamma = \cap_{\gamma \in \Gamma} A_\gamma.$$

   (Compare these with the definitions of $\sup_{\gamma \in \Gamma} a_\gamma$ and $\inf_{\gamma \in \Gamma} a_\gamma$).

7. Show that

   (1) $(A \Delta B) \Delta (B \Delta C) = A \Delta C$,
   (2) $(A \Delta B) \Delta (C \Delta D) = (A \Delta C) \Delta (B \Delta D)$,
   (3) $A \Delta B = C$ iff $A = B \Delta C$,
   (4) $A \Delta B = C \Delta D$ iff $A \Delta C = B \Delta D$.

8. (Optional) Given a sequence of sets $\{A_n, n \geq 1\}$, let $B_1 = A_1$, $B_{n+1} = B_n \Delta A_{n+1}$, $n \geq 1$. Show that $\lim_{n\to\infty} B_n$ exists iff $\lim_{n\to\infty} A_n = \emptyset$.

# Chapter 2

# Measure Theory

## 2.1 Definitions

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

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

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

**Remarks:**

(a) Convention: $\exists A \in \mathcal{A}$ such that $|\mu(A)| < \infty$, otherwise, there is not much point if all values are infinite.

(b) $\cup_{i=1}^{\infty} A_i = \Omega$ can be replaced by $\sum_{i=1}^{\infty} A_i = \Omega$. See one of the exercises.

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

(i) $\mu$ is additive $\Longleftrightarrow \mu(\sum_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$.

(ii) $\mu$ is $\sigma$-additive $\Longleftrightarrow \mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

**Definition**: $\mu$ is a **measure** on $\mathcal{A}$ if

(i) it is nonnegative, i.e., $\forall A \in \mathcal{A}: \ \mu(A) \geq 0$;

(ii) it is $\sigma$-additive.

LEMMA **2.1.1** *Assume that $\emptyset \in \mathcal{A}$ and $\mu$ is additive. Then*
*(i) $\mu(\emptyset) = 0$;* *(ii) $\sigma$-additivity $\Longrightarrow$ additivity.*

*Proof.* Since $\exists A \in \mathcal{A}$ s.t. $\mu(A) < \infty$, so $\mu(A) = \mu(A) + \mu(\emptyset)$. Hence $\mu(\emptyset) = 0$, which proves (i). Proof of (ii) follows from (i). ∎

**Definition**:

(i). If $\mu$ is a measure on a $\sigma$-algebra $\mathcal{A}$ of subsets of $\Omega$, the triplet $(\Omega, \mathcal{A}, \mu)$ is a **measure space**. The sets of $\mathcal{A}$ are called **measurable sets**, or **$\mathcal{A}$-measurable**.

(ii). A measure space $(\Omega, \mathcal{A}, P)$ is a **probability space** if $P(\Omega) = 1$.

[Note: $(\Omega, \mathcal{A})$ = **measurable space** $\neq$ **measure space** = $(\Omega, \mathcal{A}, \mu)$.]

**Definition**: Assume that $\{A_n, n \geq 1\} \in \mathcal{A}$ and $A \in \mathcal{A}$, and $\mu$ is a measure.

(1) (**Continuity from above**)

If $A_n \searrow A$ implies $\mu(A_n) \to \mu(A)$, $\mu$ is said to be continuous at $A$ from above.

(2) (**Continuity from below**)

If $A_n \nearrow A$ implies $\mu(A_n) \to \mu(A)$, $\mu$ is said to be continuous at $A$ from below.

(3) (**Continuity at** $A$)

If $A_n \to A$ implies $\mu(A_n) \to \mu(A)$, $\mu$ is said to be continuous at $A$.

REMARK **2.1.1** .

(*i*) Additivity is a rather natural requirement while that of $\sigma$-additivity is not.

(*ii*) **Measure $\mu$ is continuous from below.**

To see this, let $B_n \nearrow \lim_{n\to\infty} B_n = \cup_{n=1}^{\infty} B_n$. Then we can write $B_n = B_1 + (B_2 - B_1) + ... + (B_n - B_{n-1}) := \sum_{i=1}^{n} A_i$. Note that $\sigma$-additivity implies that

$$\mu\left(\lim_{n\to\infty} B_n\right) = \mu\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) = \lim_{n\to\infty} \mu(B_n).$$

That is, $\mu$ **is continuous from below**.

(*iii*) **However, measure $\mu$ may NOT be continuous from above**.

For example, define a measure $\mu$ on the Borel measurable space $(R, \mathcal{B})$ by

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

It is easy to check that $\mu$ is $\sigma$-additive, thus a measure. Take $A_n = (0, 1/n)$, then $A_n \to \emptyset$ forms a decreasing sequence of sets in $\mathcal{B}$, but $\mu(A_n) = \infty \not\to \mu(\emptyset) = 0$.

(*iv*) **Measure $\mu$ will be continuous from above conditionally on the assumption that $\mu(A_m) < \infty$ for some finite** $m$.

For example, let us show that the measure $\mu$ given above is indeed continuous from above conditionally at $A = \emptyset$. In fact, given a decreasing sequence $A_n \searrow \emptyset$, the assumption $\mu(A_m) < \infty$ for some finite $m$ implies that $A_k = \emptyset$ for all $k \geq m$, thus $\mu(A_k) = 0 = \mu(\emptyset)$ for all $k \geq m$.

(*v*) **Finite measures (such as probability measure) are always continuous.** See the next section (i.e. Case III).

## 2.2 Properties of measure

### 2.2.1 Case I: semialgebras.

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

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

Hence, Lemma 2.1.1 applies.

THEOREM **2.2.1** *Let $\mu$ 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**): $\quad A \subset B \implies \mu(A) \leq \mu(B).$

(2). ($\sigma$-**Subadditivity**):

(a). $\sum_1^\infty A_n \subset A, \implies \sum_1^\infty \mu(A_n) \leq \mu(A).$

(b). *Further assume that $\mu$ is $\sigma$-additive (hence a measure). Then,*

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

**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. ∎

### 2.2.2 Case II: algebras.

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

THEOREM **2.2.2** ($\sigma$-**subadditivity**) *Let $\mu$ be a* **measure** *on an algebra $\mathcal{A}$. Then,*

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

(**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 $\sigma$-additivity and monotonicity of the measure $\mu$ 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 $C_n \subset B_n \subset A_n$. ∎

### 2.2.3 Case III: $\sigma$-algebras.

All the properties for semialgebras and algebras also hold for $\sigma$-algebras. In addition, we have the following.

THEOREM **2.2.3** *Let $\mu$ be a measure on a $\sigma$-algebra $\mathcal{A}$, and $\{A_n\} \in \mathcal{A}$.*

(1) (**Monotonicity**) $A_1 \subset A_2, \Longrightarrow \mu(A_1) \le \mu(A_2)$.

(2) (**Boole's inequality** *or* **Countable Sub-Additivity**)

$$\mu(\cup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} \mu(A_i).$$

(3) (**Continuity from below**)

$$A_n \nearrow A, \Longrightarrow \mu(A_n) \to \mu(A).$$

(4) (**Continuity from above**)

$$A_n \searrow A \text{ and } \mu(A_m) < \infty \text{ for some } m \ge 1 \text{ except for } A_m = \emptyset \implies \mu(A_n) \to \mu(A).$$

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

[The claim (5) may not be true if $\mu$ is not a finite measure; see the example given in Remark 2.1.1.]

**Proof.** Even though the proofs of (1) and (2) are implied by those for semi-algebras and algebras, we will still give very simple proofs for $\sigma$-algebras.

(1) $\mu(A_2) = \mu(A_1) + \mu(A_2 - A_1) \ge \mu(A_1)$.

(2) $A := \cup_{i=1}^{\infty} A_i = A_1 + (A_2 - A_1) + (A_3 - (A_1 \cup A_2)) + ...$, so

$$\begin{aligned} \mu(A) &= \mu(A_1) + \mu(A_2 - A_1) + \mu\left(A_3 - (A_1 \cup A_2)\right) + ... \\ &\le \mu(A_1) + \mu(A_2) + \mu(A_3) + ... \end{aligned}$$

(3) First assume that $\mu(A_m) = \infty$ for some $m$. Note $A_n \supset A_m$ for $n \ge m$ and $A \equiv \cup_1^{\infty} A_n \supset A_m$. Then, monotonicity implies that $\mu(A_n) \ge \mu(A_m) = \infty$ for $n \ge m$ and $\mu(A) \ge \infty$. Therefore,

$$\lim_n \mu(A_n) = \infty = \mu(A).$$

Now assume that for all $m$, $\mu(A_m) < \infty$. Then $A = \cup_{i=1}^{\infty} A_i = A_1 + (A_2 - A_1) + (A_3 - A_2) + ...$, so

$$\begin{aligned} \mu(A) &= \mu(A_1) + \mu(A_2 - A_1) + \mu(A_3 - A_2) + ... \\ &= \mu(A_1) + \sum_{i=2}^{\infty} [\mu(A_i) - \mu(A_{i-1})] \\ &= \mu(A_1) + \lim_{n\to\infty} \sum_{i=2}^{n} [\mu(A_i) - \mu(A_{i-1})] \\ &= \lim_{n\to\infty} \mu(A_n). \end{aligned}$$

(4) Assume that for some $m$, $\mu(A_m) < \infty$. (There is no point of discussion if $\mu(A_m) = \infty$ for all $m \ge 1$.) Then $A_m - A_n$, $n \ge 1$ forms an increasing sequence and $A_m - A_n \to A_m - A = A_m - \cap_{i=1}^{\infty} A_i$. Then from (3), for $n \ge m$,

$$\mu(A_m - A_n) = \mu(A_m) - \mu(A_n) \to \mu\left(A_m\right) - \mu\left(\cap_{i=1}^{\infty} A_i\right).$$

Therefore,

$$\lim_n \mu(A_n) = \mu\left(\cap_{i=1}^{\infty} A_i\right) = \mu(A).$$

(5) See one of the exercises at the end of the Chapter. ∎

## 2.3 Arithmetics with $\infty$

Since a measure can be $\infty$, we need to know arithmetics with it. For any $x \in (-\infty, \infty)$, define

1. $\infty + x = \infty, \qquad x\infty = \infty \ (x > 0), \qquad x\infty = -\infty \ (x < 0)$,
2. $0\infty = 0, \quad \infty + \infty = \infty, \quad \infty\infty = \infty$
3. $\frac{x}{\infty} = \frac{x}{-\infty} = 0$.
4. $\infty - \infty$ and $\infty/\infty$ are not defined.

## 2.4 Probability measure

If $\mu(\Omega) = 1$, then $\mu$ is a probability measure, usually written as $P$. The probability space is $(\Omega, \mathcal{A}, P)$. Then by definition,

1. For any $A \in \mathcal{A}$, we have $0 \le P(A) \le 1$.
2. $P(\Omega) = 1$.
3. $P(\sum_1^\infty A_n) = \sum_1^\infty P(A_n)$.

Here are some more properties:

4. $P(\sum_1^n A_i) = \sum_1^n P(A_i)$
5. $P(B - A) = P(B) - P(A)$ if $A \subset B$. In particular, $P(A^c) = 1 - P(A)$.
6. $P(A) \le P(B)$ if $A \subset B$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
8. $P(\cup_{k=1}^n A_k) = \sum_k P(A_k) - \sum_{i<j} P(A_i \cap A_j) + \ldots\ldots..$
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 $P(A_n) \to P(A)$.

(So probability measure is continuous, as well as continuous from above and below.)

*Proof.* We shall only prove a selected few.

(5). $P(B) = P(A + [B - A]) = P(A) + P(B - A)$.

(7). $P(A \cup B) = P(A + [B - A]) = P(A) + P(B - A)$ and

$$P(B - A) = P(B - A \cap B) = P(B) - P(A \cap B).$$

(8). By induction. The details are omitted. ∎

## 2.5 Some examples of measure

**Example**. $(\Omega, \mathcal{A})$ is a measurable space. The following set function $\mu$ on $\mathcal{A}$ is a measure:

$$\begin{aligned} \mu(A) &= 0 & & A = \emptyset, \\ &= \infty & & A \in \mathcal{A}, \quad A \ne \emptyset. \end{aligned}$$

This is neither a finite nor a $\sigma$-finite measure.

This example shows that measures can be rather abstract, even though they are extensions of length, area, volume.

**Example**. (Counting measure). $\Omega$ is a space. For any $A \subset \Omega$, define

$$\mu(A) = |A| := \text{number of elements in } A, \ (\infty \text{ if it contains infinitely many elements})$$

Then $\mu$ is the **counting measure** on the measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\mathcal{P}(\Omega)$ is the power set. Note that

(a). $\mu$ is a finite measure if $|\Omega| < \infty$;

(b). $\mu$ is a $\sigma$-finite measure if $\Omega$ is countable;

(c). $\mu$ is not a $\sigma$-finite measure if $\Omega$ is not countable.

**Proof of (c)**. Otherwise, if $\mu$ is a $\sigma$-finite measure, then $\exists \cup_1^\infty A_n = \Omega$ such that $\mu(A_n) = |A_n| < \infty$, which makes $\Omega$ countable. Contradiction. ∎

**Example**. (Discrete Probability Spaces). Let $\Omega$ be either finite or countably infinite. For any $A \in \mathcal{P}(\Omega)$, define

$$P(A) = \sum_{\omega \in A} p(\omega), \qquad \text{where } p(\omega) \geq 0 \text{ and } \sum_{\omega \in \Omega} p(\omega) = 1.$$

Then $(\Omega, \mathcal{P}(\Omega), P)$ is a (**discrete**) probability space.

**Example**. (Geometric Probability). Continuing from the last example, when $\Omega$ is finite, we may take $p(w) = 1/|\Omega|$ so that $P(A) = |A|/|\Omega|$. Note that we have assumed certain kind of symmetry in our definition. Examples include

(1). flipping a fair coin: $\Omega = \{Head, Tail\}$,

(2). rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

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

**Definition:** Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of subsets of $\Omega$ with $\mathcal{A} \subset \mathcal{B}$. If $\mu$ and $\nu$ are two set functions (or measures) defined on $\mathcal{A}$ and $\mathcal{B}$, respectively such that

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

$\nu$ is said to be an **extension of** $\mu$ **from** $\mathcal{A}$ **to** $\mathcal{B}$, and $\mu$ the **restriction** from $\mathcal{B}$ to $\mathcal{A}$.

First we shall extend a set function (or a measure) from a semialgebra $\mathcal{S}$ to $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$.

THEOREM **2.6.1** .

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

(2). *Moreover, if $\mu$ is $\sigma$-additive on $\mathcal{S}$ (which implies that $\mu$ is a measure on $\mathcal{S}$ ), then so is $\overline{\mu}$ on $\overline{\mathcal{S}}$.*

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

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

Then $\overline{\mu}$ 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

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

$$
\begin{aligned}
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\mu(A_i\cap B_j) = \sum_{j=1}^{n}\mu(\sum_{i=1}^{m}A_i\cap B_j)\\
&= \sum_{j=1}^{n}\mu(A\cap B_j) = \sum_{j=1}^{n}\mu(B_j).
\end{aligned}
$$

That is, $\overline{\mu}(A)$ is the same irrespective of the different partitions of $A$.

Clearly, $\overline{\mu}(A)$ is nonnegative and $\overline{\mu}(\emptyset) = \mu(\emptyset) = 0$. It's easy to see that $\overline{\mu}$ is additive.

To prove uniqueness, let $\hat{\mu}$ be another extension of $\mu$ to $\bar{\mathcal{S}} = \mathcal{A}(\mathcal{S})$ 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

$$
\hat{\mu}(A) = \sum_{1}^{l}\hat{\mu}(A_i) = \sum_{1}^{l}\mu(A_i) = \overline{\mu}(A).
$$

(2). We will show that $\overline{\mu}$ is $\sigma$-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

$$
\overline{\mu}(A) = \sum_{n=1}^{\infty}\overline{\mu}(B_n). \tag{6.1}
$$

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

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

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

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

On the other hand, we have

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

This proves (6.1). ∎

We are now ready to prove part 2 of Theorem 2.2.1, stated earlier but not proved.

THEOREM **2.6.2 ($\sigma$-Subadditivity)** *Let $\mu$ be a nonnegative additive set function on a semialgebra $\mathcal{S}$. Let $A, A_n \in \mathcal{S}$. Then*

(1). $\sum_1^\infty A_n \subset A, \quad \Longrightarrow \quad \mu(A) \geq \sum_1^\infty \mu(A_n)$.

(2). *Further assume that $\mu$ is $\sigma$-additive (hence a measure). Then,*

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

**Proof.** Let $\mu$ be the unique extension $\overline{\mu}$ 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 $\mu$, we get

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

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 2.2.2, we have $\bar{\mu}(A) \leq \sum_1^\infty \bar{\mu}(A_i)$, implying $\mu(A) \leq \sum_1^\infty \mu(A_i)$. ∎

## 2.7 Outer measure

**Definition**: Let $\mu$ 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 $A$. $\mu^*$ is called the outer measure induced by the measure $\mu$.

**Remarks:**

(0) Recall: $a^* := \inf_n a_n$ is the largest lower bound of the series $\{a_n, n \geq 1\}$, i.e.

*(i)* $a^* := \inf_n a_n$ is a lower bound, that is, $a_n \geq a$ for all $n \geq 1$.

*(ii)* $\forall \epsilon > 0$, $a^* + \epsilon$ can not be a lower bound, that is, there exists an $a_m$ s.t. $a_m \leq a^* + \epsilon$.

(1). $\mu^*(A)$ 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 $\mu^*$ is $[0, \infty]$.

(2). $\mu^*(A)$ may not be a measure itself.

(3). One can think of $A$ as any set on $R$, and $\cup_1^\infty A_n$ as a countable covering of $A$.

Theorem **2.7.1** *Let $\mu$ be a measure on a semialgebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$, and $\mu^*$ be the outer measure induced by $\mu$.*

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

(2) (**Monotonicity**)

$$\mu^*(A) \leq \mu^*(B) \quad \textit{for } A \subset B \subset \Omega.$$

(3) ($\sigma$**-subadditivity**)

$$\mu^*\left(\cup_1^\infty A_n\right) \leq \sum_1^\infty \mu^*(A_n), \qquad \textit{for } \{A_n\} \subset \Omega.$$

**Proof.** (1). Suppose $A \in \mathcal{S}$. Since $A \subset A$, by definition, $\mu^*(A) \le \mu(A)$.

On the other hand, if $A \subset \cup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{S}$, then $\mu(A) \le \sum_{n=1}^{\infty} \mu(A_n)$ by the $\sigma$-subadditivity of $\mu$ on $\mathcal{S}$, see Theorem 2.6.2. Taking inf on both sides, we get $\mu(A) \le \mu^*(A)$. 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,

$$\mu^*(A) \le \sum_{n=1}^{\infty} \mu(B_n).$$

Taking inf on both sides, we get $\mu^*(A) \le \mu^*(B)$.

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

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

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

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

Since $\epsilon$ can be arbitrarily small, the proof is complete. ∎

We've shown that $\mu^*$ is $\sigma$-subadditive, but it may not be $\sigma$-additive, and therefore may not be used as a measure (since a measure must be $\sigma$-additive). Although it is possible to introduce an inner measure $\mu_*$ and then a measure induced by $\mu$ (if inner measure = outer measure), it is more convenient to introduce the following definition, due to Caratheodory.

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

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

Although this definition is not very intuitive, it works well.

THEOREM **2.7.2** *A set $A \subset \Omega$ is measurable w.r.t. an outer measure $\mu^*$ iff for any $D \subset \Omega$, one has*

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

*Proof.* It follows from Theorem 2.7.1 that $\mu^*$ is $\sigma$-subadditve, i.e.,

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

Therefore, $\mu^*(D) = \mu^*(A \cap D) + \mu^*(A^c \cap D) \Longleftrightarrow \mu^*(D) \ge \mu^*(A \cap D) + \mu^*(A^c \cap D)$. ∎

THEOREM **2.7.3** . *Let $\mathcal{A}^*$ be the class of all $\mu^*$-measurable sets.*

(1). *The class $\mathcal{A}^*$ is a $\sigma$-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 $\mu$ from $\mathcal{S}$ to $\mathcal{A}^*$.*

(*i.e., Although* $\mu^*$ *is defined on* $\mathcal{P}(\Omega)$*, but the restriction of* $\mu^*$ *to* $\mathcal{A}^*$ *is a measure.*)
(*We'll still use* $\mu^*$ *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

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

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 **2.7.1** *$\mathcal{F}$ is an algebra. $\mathcal{F}$ is a $\sigma$-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.*)

(The proof of this is left as homework, see homework 1).

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 $D \in \Omega$,

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

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

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

Therefore,

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

Letting $n \to \infty$, we get

$$\begin{aligned}
\mu^*(D) &\geq \sum_{i=1}^\infty \mu^*(A_i \cap D) + \mu^*\left(\left(\sum_{i=1}^\infty A_i\right)^c \cap D\right) \\
&\geq \mu^*\left(\left(\sum_{i=1}^\infty A_i\right) \cap D\right) + \mu^*\left(\left(\sum_{i=1}^\infty A_i\right)^c \cap D\right). \qquad (7.2)
\end{aligned}$$

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

(2). Continuing from above, since $\sum_{i=1}^{\infty} A_i \in \mathcal{A}^*$, and so all the inequalities in (7.2) can be replaced by equalities. That is,

$$\begin{aligned}
\mu^*(D) &= \sum_{i=1}^{\infty} \mu^*(A_i \cap D) + \mu^*\left(\left(\sum_{i=1}^{\infty} A_i\right)^c \cap D\right) &(7.3)\\
&= \mu^*\left(\left(\sum_{i=1}^{\infty} A_i\right) \cap D\right) + \mu^*\left(\left(\sum_{i=1}^{\infty} A_i\right)^c \cap D\right). &(7.4)
\end{aligned}$$

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

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

Since $\mu^*(\emptyset) = 0$ from Theorem 2.7.1, 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 (7.3) and (7.4), 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, $\mu^*$ is nonnegative on $\mathcal{A}^*$ and $\sigma$-additive from (2) by taking $B = A$ or $\Omega$. So $\mu^*$ is a measure on $\mathcal{A}^*$.

The fact that $\mu^*|_{\mathcal{A}^*}$ is an extension of $\mu$ from $\mathcal{S}$ to $\mathcal{A}^*$ follows from Theorem 2.7.1 (1). ∎

## 2.8 Extension of measures from semialgebras to $\sigma$-algebras

What is the relationship between the two $\sigma$-algebras $\mathcal{A}^*$ and $\sigma(\mathcal{S})$?

Theorem **2.8.1** *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

$$\mu^*(D) + \epsilon \geq \sum_{1}^{\infty} \mu(A_n) \geq \mu^*(D). \tag{8.5}$$

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 (8.5), 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 $\mu^*$ 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 $\epsilon$ 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 $\sigma$ 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 $\sigma$-algebra. ∎

REMARK **2.8.1** *From the last theorem and earlier results, we see that*

$$\mathcal{S} \quad\subset\quad \overline{\mathcal{S}} \quad\subset\quad \sigma(\mathcal{S}) \quad\subset\quad \mathcal{A}^* \quad\subset\quad \mathcal{P}(\Omega).$$

REMARK **2.8.2** *We have seen from Theorem 2.7.3 that the measure $\mu$ 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 $\mu$ 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 $\sigma$-algebra; see Durrett (1996) for example.)*

THEOREM **2.8.2 (Caratheodory Extension Theorem)** .

*Let $\mu$ be a measure on a semialgebra $\mathcal{S}$ with $\emptyset,\Omega\in\mathcal{S}$.*

*(i). $\mu$ 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 $\sigma(\mathcal{S})$.*

*(ii). If $\mu$ is* **$\sigma$-finite**, *then the extension in (i) is* **unique**. (i.e., if $\mu_1$ and $\mu_2$ are both extensions of $\mu$ to $\sigma(\mathcal{S})$, then $\mu_1=\mu_2$.)

(**Remark**: The extension from $\mathcal{S}$ to $\sigma(\mathcal{S})$ may not be unique if $\mu$ is not $\sigma$-finite; see Exercise 2, page 63, Yan S.J. *et al.* )

**Proof.** (i). From Theorem 2.8.1, we know $\sigma(\mathcal{S})\subset\mathcal{A}^*$. Then, by Theorem 2.7.3, the restriction of $\mu^*$ on $\sigma(\mathcal{S})$ is an extension of $\mu$ from $\mathcal{S}$ to $\sigma(\mathcal{S})$.

(2). $\mu$ is $\sigma$-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 $\mu_1$ and $\mu_2$ are both extensions of measure $\mu$ from $\mathcal{S}$ to $\sigma(\mathcal{S})$.

We shall first show:

$$\mu_1(A\cap D_n)=\mu_2(A\cap D_n),\qquad \forall A\in\sigma(\mathcal{S}),\ \text{and}\ \forall n\geq 1. \tag{8.6}$$

*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 (8.6), 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 $\mu_1$ and $\mu_2$ are both extensions of $\mu$ from $\mathcal{S}$ to $\sigma(\mathcal{S})$. Hence, $A \in \mathcal{M}$.*) So $\mu_1$ and $\mu_2$ are both extensions of measure $\mu$ from $\mathcal{S}$ to $\sigma(\mathcal{S})$. From Theorem 2.6.1, we know $\mu$ has a unique extension to $\mathcal{A}(\mathcal{S})$, i.e.,

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

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 $\mu_1$ and $\mu_2$ are both unique extensions of $\mu$ 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 (8.6).

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 $\sigma$-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 $\mu$ is $\sigma$-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 $\sigma$-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 (8.6) and $\sigma$-additivity of a measure, we have

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

Thus, $\mu_1$ and $\mu_2$ are the same on $\sigma(\mathcal{S})$. ∎

The measure extension theorem has a useful application in probability.

COROLLARY **2.8.1** *If $P$ is a probability defined on a semialgebra $\mathcal{S}$ on $\Omega$, then there exists a unique probability space $(\Omega, \sigma(\mathcal{S}), P^*)$ such that*

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

## 2.9 Completion of a measure

We have seen that $\mu$ also has an extension $\mu^*$ to $\mathcal{A}^*$, which in turn has a restriction $\mu^*|_{\mathcal{A}^*}$ to the $\sigma$-algebra $\sigma(\mathcal{S})$. But what is the relationship between these two?

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

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

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

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

THEOREM **2.9.1** *Given a measure space $(\Omega, \mathcal{A}, \mu)$, there exists a complete space $\left(\Omega, \overline{\mathcal{A}}, \overline{\mu}\right)$ such that $\mathcal{A} \subset \overline{\mathcal{A}}$ and $\overline{\mu} = \mu$ on $\mathcal{A}$.*

*Proof.* We shall prove the theorem in several steps.

(a). Take

$$\begin{aligned} \overline{\mathcal{A}} &= \{A \cup N : A \in \mathcal{A}, \quad N \text{ is a } \mu\text{-null set}\}. && (9.7) \\ \overline{\mathcal{B}} &= \{A \Delta N : A \in \mathcal{A}, \quad N \text{ is a } \mu\text{-null set}\}. && (9.8) \end{aligned}$$

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

**Proof.** Let $A \in \mathcal{A}$ and $N$ be a $\mu$-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

$$\begin{aligned} A \cup N &= (A - B) + (B \cap (A \cup N)) = (A - B)\Delta(B \cap (A \cup N)), && (9.9) \\ A \Delta N &= (A - B) + (B \cap (A \Delta N)) = (A - B) \cup (B \cap (A \Delta N)). && (9.10) \end{aligned}$$

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 $\mu$-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 $\overline{\mathcal{B}} \subset \overline{\mathcal{A}}$. ∎

(b). Secondly, we'll show that $\overline{\mathcal{A}}$ is a $\sigma$-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 $\mu$-null set). So $E^c \in \overline{\mathcal{A}}$. Thus, $\overline{\mathcal{A}}$ is a $\sigma$-algebra.

(c). Define a set function on $\overline{\mathcal{A}}$ by

$$\overline{\mu}(A \cup N) = \mu(A), \qquad \text{for } A \cup N \in \overline{\mathcal{A}}, \qquad (9.11)$$

Then, we can show that

(c1) $\overline{\mu}$ is well defined; (c2) $\overline{\mu}(A \Delta N) = \mu(A)$, for $A \Delta N \in \overline{\mathcal{A}}$.

**Proof.** (c1) 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 $(A\Delta B)\Delta C = A\Delta(B\Delta C)$ (see the example in the last chapter), we get

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

(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 $\mu$-null sets. So $\mu(A_1\Delta A_2) = 0$, consequently, $\mu(A_1 - A_2) = \mu(A_2 - A_1) = 0$. Therefore,

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

(c2) $\mu(B) = 0 \Longrightarrow \mu(A) = \mu(A - B) + \mu(A \cap B) = \mu(A - B)$. Then, from (9.10) and (9.11), we get

$$\overline{\mu}(A\Delta N) = \mu(A - B) = \mu(A). \quad ∎$$

(d). Next we show that $(\Omega, \overline{\mathcal{A}}, \overline{\mu})$ is a measure space. From (b), $\overline{\mathcal{A}}$ is a $\sigma$-algebra. From (9.7) and (9.11), $\overline{\mu}$ is $\sigma$-additive on $\overline{\mathcal{A}}$.

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

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

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

Note that

"$\exists \overline{B} \in \overline{\mathcal{A}}$" $\Longrightarrow \exists A \in \mathcal{A}$ and a $\mu$-null set $N$ s.t. $\overline{B} = A\Delta N$.

"$\overline{\mu}(\overline{B}) = 0$" $\Longrightarrow \overline{\mu}(\overline{B}) = \overline{\mu}(A\Delta N) = \mu(A) = 0$.

"$N$ a $\mu$-null set" $\Longrightarrow \exists B \in \mathcal{A}$ and $N \subset B$ s.t. $\mu(B) = 0$.

Thus,

$$\begin{aligned}
\overline{N} \subset \overline{B} = A\Delta N \subset A \cup N \subset A \cup B \in \mathcal{A}. \\
\mu(A \cup B) \le \mu(A) + \mu(B) = 0.
\end{aligned}$$

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

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

**Remarks:**

(1). We shall call $\left(\Omega, \overline{\mathcal{A}}, \overline{\mu}\right)$ to be the completion of $(\Omega, \mathcal{A}, \mu)$.)

(2). From the proof below, we shall see that one can take

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

We now introduce a lemma useful in the next theorem.

Lemma **2.9.1** *Let $\mu$ be a measure on a semialgebra $\mathcal{S}$, and $\mu^*$ the outer measure induced by $\mu$. If $A \subset \Omega$, and $\mu^*(A) < \infty$, then $\exists\ B \in \sigma(\mathcal{S})$ such that* *(i)* $A \subset B$,

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

*(iii)*. $\forall C \subset B - A$ *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

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

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

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

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 $\mu^*$ 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). ∎

We are now ready to state the main theorem of this section.

Theorem **2.9.2** *Let $\mu$ be a $\sigma$-finite measure on a semialgebra $\mathcal{S}$, $\mu^*$ be the outer measure induced by $\mu$, and $\mathcal{A}^*$ the $\sigma$-algebra consists of all the $\mu^*$-measurable sets. Then $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ is the completion of $(\Omega, \sigma(\mathcal{S}), \mu^*|_{\sigma(\mathcal{S})})$.*

**Proof.** For simplicity, we shall write

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

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

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

It suffices to show that

$$\overline{\sigma(\mathcal{S})} = \mathcal{A}^*, \tag{9.12}$$

the proof of which will be given a little later.

To see why (9.12) is enough, let $E \in \mathcal{A}^*$, $\Longrightarrow E \in \overline{\sigma(\mathcal{S})}$. $\Longrightarrow \exists A \in \sigma(\mathcal{S})$ and a $\mu_\sigma$-null set $N$ such that $E = A \cup N$. Since $\mu^*$ is a measure on $(\Omega, \mathcal{A}^*)$, and $N$ is $\mu^*$-measurable with $\mu^*(N) = 0$, then we have

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

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 (9.12).

**Proof of (9.12)**:

(i). 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 $\mu_\sigma$-null set. It is easy to show that $N \in \mathcal{A}^*$.

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

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

(ii). Now we show that $\mathcal{A}^* \subset \overline{\sigma(\mathcal{S})}$. Let $A \in \mathcal{A}^*$.

**Case I:** $\mu^*(A) < \infty$.

From Lemma 2.9.1, $\exists\, B \in \sigma(\mathcal{S})$ such that $A \subset B$ and $\mu^*(A) = \mu^*(B)$ [i.e. $B$ is a measurable cover of $A$ in $\sigma(\mathcal{S})$]. Since $\mu^*$ 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.,

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

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

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

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

**Case II:** $\mu^*(A) = \infty$**.**

Since $\mu_\sigma$ is $\sigma$-finite, therefore its extension $\mu^*$ is also $\sigma$-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.* $\mu^*$ is $\sigma$-finite, $\Longrightarrow$ $\Omega = \sum_1^\infty E_n$ with $E_n \in \mathcal{A}^*$ and $\mu^*(E_n) < \infty$. $\Longrightarrow$ $A = A \cap \Omega = \sum_1^\infty A \cap E_n := \sum_1^\infty A_n$ with $A_n := A \cap E_n \in \mathcal{A}^*$ and $\mu^*(A_n) \leq \mu^*(E_n) < \infty$.

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 $\sigma$-algebra. This proves that $\mathcal{A}^* \subset \overline{\sigma(\mathcal{S})}$.

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

Remark **2.9.1** *From the proof of the last theorem, we have seen that* $\mathcal{A}^* = \overline{\sigma(\mathcal{S})}$. *Therefore, we can write*

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

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

## 2.10 Construction of measures on a $\sigma$-algebra $\mathcal{A}$

### 2.10.1 General procedures

Here is one useful way of constructing measures on a $\sigma$-algebra $\mathcal{A}$.

(i). Identify a semialgebra $\mathcal{S}$ so that $\mathcal{A} = \sigma(\mathcal{S})$.

(ii). Define a map $\mu : \mathcal{S} \to R$ so that $\mu$ is a measure on $\mathcal{S}$.

(The next theorem is useful in proving $\mu$ is a measure on $\mathcal{S}$.)

(iii). Extend the measure from $\mathcal{S}$ to $\mathcal{A} = \sigma(\mathcal{S})$ by the measure extension theorem.

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

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

The following theorem is useful in finding a measure on $\mathcal{S}$.

THEOREM **2.10.1** *Let $\mu$ be a nonnegative set function on a semi-algebra $\mathcal{S}$ with $\emptyset, \Omega \in \mathcal{S}$. Assume that*

*(i) $\mu$ is additive on $\mathcal{S}$.*
*(i.e. $\mu(A) = \sum_1^n \mu(A_i)$ whenever $A_n \in \mathcal{S}$ and $A = \sum_1^n A_i \in \mathcal{S}$.)*

*(ii) $\mu$ is $\sigma$-subadditive on $\mathcal{S}$.*
*(i.e. $\mu(A) \le \sum_1^\infty \mu(A_i)$ 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 $\mu$ is a measure on $\mathcal{S}$.*

**Proof.** Since $\mu$ is nonnegative and finitely additive, it can be easily shown that $\mu(\emptyset) = 0$. It remains to check if $\mu$ is $\sigma$-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 2.6.2, we have

$$\mu(A) \ge \sum_1^\infty \mu(A_n).$$

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

### 2.10.2 Lebesgue and Lebesgue-Stieltjes measures

We are now ready to construct Lebesgue and Lebesgue-Stieltjes measures on 1-dim Borel space $(R, \mathcal{B})$, where $R = [-\infty, \infty]$ and $\mathcal{B} =$ Borel sets (i.e., the $\sigma$-algebra generated by all open intervals).

THEOREM **2.10.2 (L-S measure)** *Suppose that $F$ is finite on $(-\infty, \infty)$ (i.e. $|F(t)| < \infty$ for $|t| < \infty$), and*

*(i) $F$ is nondecreasing;* *(ii) $F$ is right continuous.*

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

$$\mu((a, b]) = F(b) - F(a), \qquad -\infty \le a \le b \le \infty,$$

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

COROLLARY **2.10.1 (Lebesgue measure)** *There is a unique measure $\lambda$ on $(R, \mathcal{B})$ with*

$$\mu((a, b]) = b - a, \qquad -\infty \le a < b \le \infty,$$

Remarks:

(1) A function $F$ which is nondecreasing and right continuous is called a **Lebesgue-Stieltjes (L-S) measure function**.

(2) The (completed) measure $\mu$ is called the **L-S measure**. The (incomplete) measure $\mu$ is called the **B-L-S measure**. (B stands for "Borel").

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

(4) Clearly, $F$ uniquely determines $\mu$, 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 $\mu$ to the measurable space $([0,1], \mathcal{B} \cap [0,1])$, then $\mu$ is a probability measure, (a uniform probability measure).

(6) When $\Omega$ 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 $\sigma$-fields that are smaller than the power set, but large enough for all practical purposes.

**Proof of Theorem 2.10.2.** Let $\mathcal{S} = \{(a,b] : -\infty \le a < b \le \infty\} \cup \{R,\ \emptyset,\ \{-\infty\}\}$. It can be shown that $\mathcal{S}$ is a semi-algebra and $\mathcal{B} = \sigma(\mathcal{S})$. Define

$$\begin{aligned} \mu((a,b]) &= F(b) - F(a), \qquad \text{if } -\infty \le a < b \le \infty, \\ \mu(\{-\infty\}) &= 0, \\ \mu(R) &= F(\infty) - F(-\infty), \end{aligned}$$

where

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

*Several remarks are in order before we move on:*

*(a) 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.*

*(b) Note that both $F(-\infty)$ and $F(\infty)$ exist (but may take $\pm\infty$ value) as $F$ is monotone.*

*(c) The definition of $F(\infty)$ implies that $F$ is left continuous at $\infty$.*

*(d) If we consider $\Omega = (-\infty, \infty]$ instead of $[-\infty, \infty]$, then $\mathcal{S} = \{(a,b] : -\infty \le a \le b \le \infty\}$ 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.

(I). First we check that $\mu$ 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.

(II). Secondly, we show that $\mu$ defined above is a measure on the semialgebra $\mathcal{S}$ by way of Theorem 2.10.1. Clearly, $\mu$ is nonnegative.

(i). Checking finite additivity (i.e., checking condition (i) of Theorem 2.10.1): Let $S = \sum_{i=1}^n S_i$, where $S, S_i \in \mathcal{S}$.

(ia). 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).

(ib). 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., $\mu$ is finitely additive.

(ii). Checking $\sigma$-subadditivity (i.e., checking condition (ii) of Theorem 2.10.1). 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 $(a,b] \subset \cup_1^\infty (a_i, b_i]$.

(a). 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

$$F(a+\delta) < F(a) + \epsilon,$$
$$F(b_i + \eta_i) < F(b_i) + \epsilon/2^i.$$

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 $(\alpha_i, \beta_i)$, $1 \le j \le J$. By monotonicity of $F$, we get

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

Letting $\epsilon \to 0$, we get

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

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

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

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

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

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

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

Combining (i) and (ii), we have shown that $\mu$ is indeed a measure on the semialgebra $\mathcal{S}$ by Theorem 2.10.1.

Also we can see that $\mu$ is $\sigma$-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. ∎

### 2.10.3 Relationship between probability measures and distribution functions

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:** A real-valued function $F$ on $\mathcal{R}$ is distribution function (d.f.) if (a). $F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \qquad F(\infty) = \lim_{x\to\infty} F(x) = 1.$

(b). $F$ is nondecreasing, i.e., $F(x) \le F(y)$ if $x \le y$

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

THEOREM **2.10.3 (Correspondence theorem)** *The relation*

$$F(x) = P((-\infty, x]), \qquad x \in \mathcal{R} \tag{10.13}$$

*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 (10.13) is a d.f. (i.e. (a)-(c) above hold).

(a) Given $x_n \searrow -\infty$, it follows that $(-\infty, x_n] \searrow \emptyset$. 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, $\lim_{x\searrow\infty} F(x) = 1$.

(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 2.10.2, there exists a unique probability measure $P$ on $(\mathcal{R}, \mathcal{B})$ satisfying $F(x) - F(y) = P((y, x])$ for $x \le y$. Letting $y \searrow -\infty$, we have

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

**Remarks:**

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

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

This definition does not involve any random variables (which we have not discussed yet). Different versions of d.f.'s involving r.v.'s can be found in Durrett, for instance.

(2) $F$ is a d.f. and $P$ is a probability measure. Then the following are equivalent

$$\begin{array}{ll} (1) & P((-\infty, b]) = F(b), \\ (2) & P((a, b]) = F(b) - F(a), \\ (3) & P([a, b]) = F(b) - F(a-), \\ (4) & P([a, b)) = F(b-) - F(a-), \\ (5) & P((a, b)) = F(b-) - F(a). \end{array}$$

(3) Other useful relations are

$$\begin{array}{ll} (6) & P((-\infty, b)) = F(b-) \\ (7) & P(\{a\}) = F(a) - F(a-). \end{array}$$

(4) 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 $F(\infty) = 1$. Some authors define such functions to be d.f.s.

(5). 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**. It can be shown that

(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.* (a) $x$ 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, $x \in S(F)$.

(b). 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

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

(c) The discrete d.f. with positive jump size at each rational number is such an example. ∎

### 2.10.4 Different types of distributions

**Definition:**

(i). A d.f. $\delta_t$ is called **degenerate** at $t$ if

$$\begin{aligned} \delta_t(x) &= 0 \qquad x < t \\ &= 1 \qquad x \geq t. \end{aligned}$$

(check that it is indeed a d.f.)

(ii). 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 $\sum p_j = 1$.

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

**Remarks:**

(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

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

(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

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

(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

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

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 $\Omega = \{a_1, a_2, ..., a_n, ...\}$ (either finite or countably infinite), is

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

(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. 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

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

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

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

### 2.10.5 Decomposition of distribution functions

LEMMA **2.10.1** *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.

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

$$F_d(x) = \sum_{j=1}^{\infty} p_j \delta_{a_j}(x) = \text{sum of all the jumps of } F \text{ in } (-\infty, x]. \tag{10.14}$$

It is clearly nondecreasing, right continuous with

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

THEOREM **2.10.4** *Let $F_c(x) = F(x) - F_d(x)$. Then $F_c$ is nonnegative, nondecreasing, and continuous.*

**Proof.** Let $x < y$, then

$$\begin{aligned} F_d(y) - F_d(x) &= \text{sum of jumps in } (-\infty, y] - \text{sum of jumps in } (-\infty, x] \\ &= \sum_{x < a_j \le y} p_j = \sum_{x < a_j \le y} (F(a_j) - F(a_j-)) \\ &\le F(y) - F(x). \end{aligned} \tag{10.15}$$

(i). From (10.15), we get

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

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

(ii). Letting $x \to -\infty$ in (10.15), we get $F_d(y) \le F(y)$. Thus $F_c(y) = F(y) - F_d(y) \ge 0$. So $F_c$ is nonnegative.

(iii). 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 (10.14), 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 **2.10.5** *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*

$$G_c(x) = F_c(x), \qquad \textit{and} \qquad G_d(x) = F_d(x)$$

*(i.e., the decomposition is unique).*

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

(i) $\{a_n, n \ge 1\} \neq \{a'_n, n \ge 1\}$, or

(ii) $a_i = a'_i$ 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 $i \geq 1$, such that

$$F_d(\tilde{a}) - F_d(\tilde{a}-) \neq G_d(\tilde{a}) - G_d(\tilde{a}-). \tag{10.16}$$

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

$$\begin{aligned}
0 &= F_c(\tilde{a}) - F_c(\tilde{a}-) \\
&= [F(\tilde{a}) - F_d(\tilde{a})] - [F(\tilde{a}-) - F_d(\tilde{a}-)] \\
&= [F(\tilde{a}) - F(\tilde{a}-)] - [F_d(\tilde{a}) - F_d(\tilde{a}-)] \\
&\neq [F(\tilde{a}) - F(\tilde{a}-)] - [G_d(\tilde{a}) - G_d(\tilde{a}-)] \qquad \text{from (10.16)} \\
&= [F(\tilde{a}) - G_d(\tilde{a})] - [F(\tilde{a}-) - G_d(\tilde{a}-)] \\
&= G_c(\tilde{a}) - G_c(\tilde{a}-) \\
&= 0.
\end{aligned}$$

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

THEOREM **2.10.6** *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 $1 - \alpha = F_c(\infty)$.

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 $F(x) = 0 \times F_d(x) + 1 \times F_c(x)$.

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 $F(x) = 1 \times F_d(x) + 0 \times F_c(x)$.

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 **2.10.2** *Every d.f. $F$ is either a discrete d.f. or a continuous d.f., or a combination (or a mixture) of both.*

### 2.10.6 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.**

But the above further analysis of d.f.'s requires the theory of Lebesgue measure, and integration, which we have not done so far. We only list some definitions and conclusions.

**Definition:**

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

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

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

**Alternative 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.).

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

THEOREM **2.10.7** *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.*

Let us give an example of a singular distribution.

**Exmaple. (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

$F(x) = 0$ for $x \leq 0$,

$F(x) = 1$ for $x \geq 1$,

$F(x) = 1/2$ for $x \in [1/3, 2/3]$,

$F(x) = 1/4 = 1/2^2$ for $x \in [1/9, 2/9] = [1/3^2, 2/3^2]$,

$F(x) = 3/4 = 1 - 1/2^2$ for $x \in [7/9, 8/9] = [(3^2 - 2)/3^2, (3^2 - 1)/3^2]$,

..............................

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 $m/3^n$, $m, n \in \mathcal{N}$). By definition, $F$ is clearly a singular d.f., which is called "Lebesgue's singular function".

REMARK **2.10.1** *For Cantor ternary set, one can see Question 10, page 25, Chow and Teicher.*

### 2.10.7 Some examples of distributions

**Absolutely continuous d.f. $F$'s with p.d.f. $f$'s**

(i) **Uniform distribution**: $f(x) = \frac{1}{b-a} I\{a \le x \le b\}$.

(ii) **Normal distribution**: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

(iii) **Cauchy distribution** $f(x) = \frac{C}{(1+x^2)}$.

(iv) **Exponential distribution** $f(x) = e^{-x} I\{x \ge 0\}$.

**Discrete distribution**

(i) **Binomial d.f.**: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},\ 0 < p < 1,\ x = 0, ..., n.$

(ii) **Possion d.f.**: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, ......$

(iii) **Empirical distribution function (e.d.f.) of observations** $\{x_1, ..., x_n\}$:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{x_i \le x\}.$$

(i.e., $F_n(x)$ is the proportion of data $\le x$.)

## 2.11 Radon-Nikodym theorem

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

DEFINITION **2.11.1** *We say that $\nu$ is* **absolutely continuous w.r.t.** *$\mu$, written as $\nu << \mu$, if*

$$\mu(A) = 0 \textit{ implies } \mu(A) = 0. \quad \blacksquare$$

THEOREM **2.11.1 (Radon-Nikodym theorem)** *Given a measurable space $(X, \Sigma)$, if a measure $\nu$ on $(X, \Sigma)$ is absolutely continuous with respect to a sigma-finite measure $\mu$ on $(X, \Sigma)$, 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 $\mu$-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".*

The theorem is named after Johann Radon, who proved the theorem for the special case where the underlying space is $R^n$ in 1913, and for Otton Nikodym who proved the general case in 1930.

For a proof of the theorem, see Wikipedia.

### 2.11.1 Applications

The theorem is very important in extending the ideas of probability theory from probability masses and probability densities defined over real numbers to probability measures defined over arbitrary sets. It tells if and how it is possible to change from one probability measure to another.

For example, it is necessary when proving the existence of conditional expectation for probability measures. The latter itself is a key concept in probability theory, as conditional probability is just a special case of it.

Amongst other fields, financial mathematics uses the theorem extensively. Such changes of probability measure are the cornerstone of the rational pricing of derivative securities.

### 2.11.2 Properties

Let $\nu$, $\mu$, and $\lambda$ be $\sigma$-finite measures on the same measure space.

- If $\nu << \lambda$ and $\mu << \lambda$ ($\nu$ and $\mu$ are absolutely continuous in respect to $\lambda$), then
$$\frac{d(\nu+\mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda}$$
$\lambda$-almost everywhere.
- If $\nu << \mu << \lambda$, then
$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$
$\lambda$-almost everywhere.
- If $\mu << \lambda$ and $g$ is a $\mu$-integrable function, then
$$\int_X g\, d\mu = \int_X g\frac{d\mu}{d\lambda}\, d\lambda.$$
- If $\mu << \nu$ and $\nu << \mu$, then
$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$
- If $\nu$ is a finite signed or complex measure, then
$$\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|.$$

### 2.11.3 The assumption of sigma-finiteness

The Radon-Nikodym theorem assumes that the measure $\mu$ is $\sigma$-finite. Here is an example when $\mu$ is not $\sigma$-finite and the Radon-Nikodym theorem fails to hold.

Example **2.11.1** *Let $\mathcal{B}(R)$ be the Borel $\sigma$-algebra on the real line, (i.e. the minimal $\sigma$-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 $\nu$ be the usual Lebesgue measure on this Borel algebra. Then we can show the following.*

- *$\mu$ 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.*
- *$\nu$ is absolutely continuous w.r.t. $\mu$;*

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

- *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.* ∎

## 2.12 Exercises

1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $A_i \in \mathcal{A}$, $i = 1, ..., n$ $(n \geq 2)$. Show that

$$P\left(\cup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j).$$

2. If $(\Omega, \mathcal{A}, P)$ is a probability space, and $\{A_n, n \geq 1\} \in \mathcal{A}$ (i.e., they are a sequence of events).

   (i) Verify that $P(\underline{\lim}\ A_n) = \lim P\left(\cap_{i=n}^{\infty} A_i\right)$.
   (ii) Show that $P\left(\cap_{i=1}^{\infty} A_i\right) = 1$ if $P\left(A_i\right) = 1$, $n \geq 1$.

3. If $(\Omega, \mathcal{A}, \mu)$ is a measure space, and $A_n \in \mathcal{A}$.

   (i) 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 $\mu(\overline{\lim}\ A_n) \geq \overline{\lim}\ \mu(A_n)$.
   (ii) If $\mu$ is a finite measure, and $\underline{\lim}\ A_n = \overline{\lim}\ A_n = A$, (i.e. $\lim A_n = A$), then $\lim \mu(A_n) = \mu(A)$.
   (iii) If $\sum_1^{\infty} \mu(A_n) < \infty$, then $\mu(\overline{\lim}\ A_n) = 0$. (This is half Borel-Cantelli Lemma when $\mu$ is a probability measure.)

4. $A_1, A_2, ....$ are a sequence of events on the probability space $(\Omega, \mathcal{A}, P)$

   (1) If $P(A_n) \to 0$ and $\sum_{n=1}^{\infty} P(A_n A_{n+1}^c) < \infty$, then $P(A_n, i.o) = 0$.
   (2) If $P(A_n) \to 0$ and $\sum_{n=1}^{\infty} P(A_n^c A_{n+1}) < \infty$, then $P(A_n, i.o) = 0$.
   (3) If $A_n$'s are independent, provide a simple proof for (1).
   (4) (**Optional.**) Give examples to show that (1) may not be true if only one of two conditions holds.

5. Let $\mu$ be a nonnegative additive set function on an algebra $\mathcal{A}$ and $\{A_n, n \geq 0\} \subset \mathcal{A}$. Show

   (1) $\mu$ is $\sigma$-additive (hence a measure on $\mathcal{A}$) iff for $A_n \nearrow A$, all $\in \mathcal{A}$, we have $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.

   (2) if for every $A_n \searrow \emptyset$, we have $\lim_n \mu(A_n) = 0$, then $\mu$ is $\sigma$-additive (hence a measure).

6. (**Alternative definition of $\sigma$-finite measure**). Let $\mu$ be a $\sigma$-finite measure on a semialgebra $\mathcal{S}$. Then there exists <u>disjoint</u> $\{B_n, n \geq 1\} \subset \mathcal{S}$, such that $\sum_{i=1}^{\infty} B_i = \Omega$ and $\mu(B_n) < \infty$ for each $n$.

# Chapter 3

# Random Variables

## 3.1 Mappings

Since $\Omega_1$ is arbitrary, it is often convenient to consider a function (mapping) $X$ from $\Omega_1$ to a simpler space $\Omega_2$ (e.g. $\Omega_2 = \mathcal{R}$).

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

(i). For every subset $B \in \Omega_2$, the **inverse image** of $B$ is

$$X^{-1}(B) = \{\omega : \omega \in \Omega_1, X(\omega) \in B\} := \{X \in B\}.$$

(ii). For every class $\mathcal{G} \subset \Omega_2$, the **inverse image** of $\mathcal{G}$ is

$$X^{-1}(\mathcal{G}) = \{X^{-1}(B) : B \in \mathcal{G}\}.$$

(So $A \in X^{-1}(\mathcal{G})$ means that $\exists B \in \mathcal{G}$ s.t. $A = X^{-1}(B)$.)

Theorem **3.1.1 (Properties of the inverse image)** .

(1). *$X$ is a mapping from $\Omega_1$ to $\Omega_2$. Then*

*(i)* $X^{-1}(\Omega_2) = \Omega_1$, $X^{-1}(\emptyset) = \emptyset$.
*(ii)* $X^{-1}(B^c) = \left[X^{-1}(B)\right]^c$.
*(iii)* $X^{-1}\left(\cup_{\gamma\in\Gamma} B_\gamma\right) = \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma)$ *for* $B_\gamma \subset \Omega_2$, $\gamma \in \Gamma$.
*(vi)* $X^{-1}\left(\cap_{\gamma\in\Gamma} B_\gamma\right) = \cap_{\gamma\in\Gamma} X^{-1}(B_\gamma)$ *for* $B_\gamma \subset \Omega_2$, $\gamma \in \Gamma$,
*where $\Gamma$ is an index set, not necessarily countable.*
*(v)* $X^{-1}(B_1 - B_2) = X^{-1}(B_1) - X^{-1}(B_2)$ *for* $B_1, B_2 \subset \Omega_2$.
*(vi)* $B_1 \subset B_2 \subset \Omega_2$ *implies that* $X^{-1}(B_1) \subset X^{-1}(B_2)$.

(2) *If $\mathcal{B}$ is a $\sigma$-algebra in $\Omega_2$, then $X^{-1}(\mathcal{B})$ is a $\sigma$-algebra in $\Omega_1$.*

(3) *Let $\mathcal{C}$ be a nonempty class in $\Omega_2$, then*

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

**Proof.** (1).
(i) Note that

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

(ii) $X^{-1}(B^c) = \{\omega : X(\omega) \in B^c\} = \{\omega : X(\omega) \notin B\} = \{\omega : \omega \notin X^{-1}(B)\} = \{\omega : \omega \in [X^{-1}(B)]^c\} = [X^{-1}(B)]^c$.

(iii) $\omega \in X^{-1}(\cup_{\gamma\in\Gamma} B_\gamma) \Longleftrightarrow X(\omega) \in \cup_{\gamma\in\Gamma} B_\gamma \Longleftrightarrow X(\omega) \in B_\gamma$ for some $\gamma \in \Gamma \Longleftrightarrow \omega \in X^{-1}(B_\gamma)$ for some $\gamma \in \Gamma \Longleftrightarrow \omega \in \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma)$.

(iv) Similar to (iii).

(v) $X^{-1}(B_1 - B_2) = X^{-1}(B_1 \cap B_2^c) = X^{-1}(B_1) \cap X^{-1}(B_2^c) = X^{-1}(B_1) \cap [X^{-1}(B_2)]^c = X^{-1}(B_1) - X^{-1}(B_2)$.

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

(2). First $X^{-1}(\mathcal{B})$ is nonempty as $\Omega_1 = X^{-1}(\Omega_2) \in X^{-1}(\mathcal{B})$. 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 $\sigma$-algebra.

(3). Clearly, $X^{-1}(\mathcal{C}) \subset X^{-1}(\sigma(\mathcal{C}))$, which is a $\sigma$-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

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

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 $\sigma$-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 $\sigma$-algebra. Let $G \in \mathcal{G} \Longrightarrow X^{-1}(G) \in \sigma\left(X^{-1}(\mathcal{C})\right)$, $\Longrightarrow X^{-1}(G^c) = [X^{-1}(G)]^c \in \sigma\left(X^{-1}(\mathcal{C})\right) \Longrightarrow G^c \in \mathcal{G} \Longrightarrow \mathcal{G}$ is closed under complement. Similarly, we can show that it is closed under countable union. The proof is complete.

**Remark:** Clearly, $X^{-1}(\cdot)$ on $\Omega_2$ preserves all the set operations on $\Omega_1$.

## 3.2 Measurable mapping

**Definition:**

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

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

(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).

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

Theorem **3.2.1** *$X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ 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 3.1.1 (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 **3.2.2** *If $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ and $f : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ are measurable mappings, then $f(X) = f \cdot X : (\Omega_1, \mathcal{A}_1) \to (\Omega_3, \mathcal{A}_3)$ is also measurable.*

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

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

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

(A picture will make this very clear.)

## 3.3 Random Variables (Vectors)

### 3.3.1 Random variables

**Definition:** 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

$$X \text{ is } \mathcal{A}\text{−measurable},$$

or simply write it as

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

**Remark:** This definition does not involve any probability measures. Some authors define r.v.'s on a probability space with extra condition.

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

We shall use either definition whichever is more convenient for us.

### 3.3.2 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. This is justified by the next theorem, which is a simple consequence of Theorem 3.2.1.

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

$\Longleftrightarrow \{X \le x\} = X^{-1}([-\infty, x]) \in \mathcal{A}$ *for all* $x \in \mathcal{R}$.

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

**Proof.** Take $\mathcal{C} = \{[-\infty, b] : b \in \mathcal{R}\}$ or $\mathcal{C} = \{[-\infty, b] : b \in \mathcal{D}\}$ in Theorem 3.2.1.

**Remark:**

(i). We can take $\mathcal{D}$ to be all rational numbers.

(ii). $\{X \le x\}$ in the theorem can be replaced by any of the following:

$$\{X \le x\}, \quad \{X \ge x\}, \quad \{X < x\}, \quad \{X > x\}, \quad \{x < X < y\},\ etc.$$

### 3.3.3 Random vectors

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

Theorem **3.3.2** *$X = (X_1, ..., X_n)$ 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 $I_k = (a_k, b_k], -\infty \le a_k \le b_k \le \infty$, $k \ge 1$. Since $\{X_i \in I_k\} \in \mathcal{A}$,

$$\{X = (X_1, ..., X_n) \in I_1 \times ... \times I_n\} = \bigcap_{k=1}^{n} \{X_i \in I_k\} \in \mathcal{A}.$$

The proof follows from this and Theorem 3.2.1 as $\mathcal{B}(\mathcal{R}^n) = \sigma\left(\{\prod_{k=1}^n I_k\}\right)$.

## 3.4 Construction of random variables

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

### 3.4.1 Algebraic operations $(+, -, \times, \div)$

THEOREM **3.4.1** *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.**

$$\begin{aligned}
\{aX \le t\} &= \{X \le t/a\} \quad if\ a > 0 \\
&\phantom{=} \{X \ge t/a\} \quad if\ a < 0, \\
&\in \mathcal{A}
\end{aligned}$$

$$\begin{aligned}
\{X + Y < t\} &= \bigcup_{r \in \mathcal{Q}} \left(\{X < r\} \cap \{Y < t - r\}\right), \quad (Q = \text{all rational numbers}), \qquad (4.1)\\
&\phantom{=} \text{(the proof of this is given at the end)} \\
&\in \mathcal{A}
\end{aligned}$$

$$\begin{aligned}
\{X^2 \le t\} &= \{|X| \le \sqrt{t}\} = \{X \le \sqrt{t}\} - \{X < -\sqrt{t}\} \qquad \text{if } t \ge 0 \\
&\phantom{=} \emptyset, \qquad \text{if } t < 0 \\
&\in \mathcal{A}
\end{aligned}$$

$$XY = \frac{1}{2}\left((X+Y)^2 - X^2 - Y^2\right) \in \mathcal{A}$$
$$,$$

$$\begin{aligned}
\{1/Y \le t\} &= \{1/Y \le t\} \cap (\{Y < 0\} \cup \{Y > 0\}) \\
&= (\{1/Y \le t\} \cap \{Y < 0\}) \cup (\{1/Y \le t\} \cap \{Y > 0\}) \\
&= \{Y \ge 1/t\} \cup \{Y < 0\} \qquad if\ t \ge 0 \\
&\phantom{=} \{Y \ge 1/t\} \cap \{Y < 0\} \qquad if\ t < 0, \\
&\in \mathcal{A}
\end{aligned}$$

$$X/Y = X\,(1/Y) \in \mathcal{A}$$

$$\{X \wedge Y \ge t\} = \{X \ge t\} \cap \{Y \ge t\} \in \mathcal{A},$$

$$\{X \vee Y \le t\} = \{X \le t\} \cap \{Y \le t\} \in \mathcal{A}.$$

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 $\Delta$, we have $r+s < t$. $\Longrightarrow$ In particular, we can choose a rational point $(r,s) \in \Delta$. Then we have $X < r$ and $Y < s < t - r$.

**Definition:** The **positive** and **negative** parts of a function $X : \Omega \to \mathcal{R}$ are

$$X^+ = \max\{X, 0\}, \qquad X^- = -\min\{X, 0\}.$$

It is clear that $X = X^+ - X^-$ and $|X| = X^+ + X^-$.

Corollary **3.4.1** *If $X$ is a r.v. (i.e., $X \in \mathcal{A}$), so are $X^+$, $X^-$ and $|X|$.*

### 3.4.2 Limiting operations

Theorem **3.4.2** *$X_1, X_2, \ldots$ 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 $\omega$, 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 $\omega$, then $S$ is a r.v., (i.e., $S \in \mathcal{A}$).*

**Proof.** (1)

$$\begin{aligned} \{\sup_n X_n \le t\} &= \bigcap_{n=1}^{\infty} \{X_n \le t\} \qquad \{\inf_n X_n \ge t\} = \bigcap_{n=1}^{\infty} \{X_n \ge t\} \\ \limsup_n X_n &= \inf_k \sup_{m \ge k} X_m \qquad \liminf_n X_n = \sup_k \inf_{m \ge k} X_m. \end{aligned}$$

(2) $X(\omega) = \lim_n X_n(\omega) = \limsup_n X_n(\omega)$ is a r.v.

(3) This follows from (2).

**Definition:**

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, $\Omega_0$ 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.

### 3.4.3 Transformations

Theorem **3.4.3** *$X = (X_1, \ldots, X_n)$ 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.

**Remark**:

1. Note that continuous functions are Borel measurable. So if $X, X_1, \ldots, X_n$ are r.v.'s, so are

$$X_1 + \ldots + X_n, \qquad \sin(X), \qquad e^X, \qquad Polynomial(X), \quad etc.$$

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 $\{(x_1, x_2) : x_1 + x_2 < a\} = \bigcup_{r \in \mathcal{Q} \cap A} Rectangle(r, \epsilon_r) \in \mathcal{B}^2$.

## 3.5 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 **3.5.1** .

(1). (*Indicator r.v.*) *If $A \in \mathcal{A}$, the indicator function $I_A$ is a r.v.*

(*Recall: $I_A(\omega) = I\{\omega \in A\}$ indicates whether $A$ occurs or not.*)

(2). (*Simple r.v.*). *If $\Omega = \sum_1^n A_i$, 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. Method I: by using the definition.**

(1). $\forall B \in \mathcal{B}$, note that

$$\begin{aligned}\{I_A \in B\} &= \emptyset \quad if\ 0 \notin B,\ 1 \notin B \\ & \quad A^c \quad if\ 0 \in B,\ 1 \notin B \\ & \quad A \quad if\ 0 \notin B,\ 1 \in B \\ & \quad \Omega \quad if\ 0 \in B,\ 1 \in B.\end{aligned}$$

Since $A \in \mathcal{A}$, we see that $\{I_A \in B\} \in \mathcal{A}$. Thus $I_A$ is a r.v.

(2). $\forall B \in \mathcal{B}$, note that $\{X \in B\} = \bigcup_{\{i: a_i \in B\}} A_i \in \mathcal{A}$. The proof is complete.

**Method II: by using Theorem 3.3.1.**

(1). Clearly,

$$\begin{aligned}\{I_A \le t\} &= \emptyset \quad if\ t < 0 \\ & \quad A^c \quad if\ 0 \le t < 1 \\ & \quad \Omega \quad if\ t \ge 1.\end{aligned}$$

Thus $\{I_A \in B\} \in \mathcal{A}$, i.e., $I_A$ is a r.v.

(2). It follows from (1) that $I_{A_i}$'s are r.v.'s. Then the proof follows from Theorem 3.4.1.

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

**Proof.** $\forall n \ge 1$, let

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

One can draw a picture to see that the claim in the theorem is true.

Now let us give a rigorous proof of the theorem. Clearly, $X_n(\omega) \ge 0$ for all $n$. Next we show that $X_n(\omega) \nearrow$ for all $\omega \in \Omega$. For any $n \ge 1$ and $\omega \in \Omega$, we have

$$\begin{aligned}X_{n+1}(\omega) &= \sum_{k=1}^{(n+1)2^{n+1}} \frac{k-1}{2^{n+1}}\, I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\ &= \left(\sum_{k=1}^{n2^{n+1}} + \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}}\right) \frac{k-1}{2^{n+1}}\, I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\ &:= A + B + C.\end{aligned}$$

Now

$$
\begin{aligned}
A &= \sum_{k=1}^{n2^{n+1}} \frac{k-1}{2^{n+1}}\, I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} \\
&= 0 \times I\left\{0 < X(\omega) \le \frac{1}{2^{n+1}}\right\} + \frac{1}{2^{n+1}}\, I\left\{\frac{1}{2^{n+1}} < X(\omega) \le \frac{2}{2^{n+1}}\right\} \\
&\quad + \frac{2}{2^{n+1}}\, I\left\{\frac{2}{2^{n+1}} < X(\omega) \le \frac{3}{2^{n+1}}\right\} + \frac{3}{2^{n+1}}\, I\left\{\frac{3}{2^{n+1}} < X(\omega) \le \frac{4}{2^{n+1}}\right\} \\
&\quad + \cdots\cdots \\
&\quad + \cdots\cdots \\
&\quad + \frac{n2^{n+1}-2}{2^{n+1}}\, I\left\{\frac{n2^{n+1}-2}{2^{n+1}} < X(\omega) \le \frac{n2^{n+1}-1}{2^{n+1}}\right\} + \frac{n2^{n+1}-1}{2^{n+1}}\, I\left\{\frac{n2^{n+1}-1}{2^{n+1}} < X(\omega) \le \frac{n2^{n+1}}{2^{n+1}}\right\} \\
&\ge 0 \times I\left\{0 < X(\omega) \le \frac{1}{2^{n+1}}\right\} + 0 \times\ I\left\{\frac{1}{2^{n+1}} < X(\omega) \le \frac{1}{2^{n}}\right\} \\
&\quad + \frac{1}{2^n}\, I\left\{\frac{1}{2^n} < X(\omega) \le \frac{3}{2^{n+1}}\right\} + \frac{1}{2^n}\, I\left\{\frac{3}{2^{n+1}} < X(\omega) \le \frac{2}{2^n}\right\} \\
&\quad + \cdots\cdots \\
&\quad + \cdots\cdots \\
&\quad + \frac{n2^n-1}{2^n}\, I\left\{\frac{n2^n-1}{2^n} < X(\omega) \le \frac{n2^{n+1}-1}{2^{n+1}}\right\} + \frac{n2^n-1}{2^n}\, I\left\{\frac{n2^{n+1}-1}{2^{n+1}} < X(\omega) \le \frac{n2^{n+1}}{2^{n+1}}\right\} \\
&= \sum_{k=1}^{n2^n} \frac{k-1}{2^n}\, I\left\{\frac{k-1}{2^n} < X(\omega) \le \frac{k}{2^n}\right\}.
\end{aligned}
$$

and

$$
\begin{aligned}
B + C &= \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}} \frac{k-1}{2^{n+1}}\, I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + (n+1)I\{X(\omega) > n+1\} \\
&\ge \sum_{k=n2^{n+1}+1}^{(n+1)2^{n+1}} \frac{n2^{n+1}}{2^{n+1}}\, I\left\{\frac{k-1}{2^{n+1}} < X(\omega) \le \frac{k}{2^{n+1}}\right\} + nI\{X(\omega) > n+1\} \\
&\ge nI\left\{n < X(\omega) \le n+1\right\} + nI\{X(\omega) > n+1\} \\
&= nI\{X(\omega) > n\}.
\end{aligned}
$$

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 $\infty$).

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,

$$
0 \le X(\omega) - \frac{k-1}{2^n} = X(\omega) - X_n(\omega) \le \frac{1}{2^n} \to 0. \quad \blacksquare
$$

Remark **3.5.1** *For any $x \in \mathcal{R}$, let*

$[x]$ *be the largest integer* $\le x$; $\lfloor x \rfloor$ *be the largest integer* $< x$.

*It is easy to see that*

$$\begin{aligned} \lfloor x \rfloor &= [x] && \textit{if } x \textit{ is a non-integer} \\ & \quad [x]-1 && \textit{if } x \textit{ is an integer} \end{aligned}$$

*Then, it can easily shown that the expression $X_n(\omega)$ given above can also be written down simply as*

$$X_n(\omega) = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n.$$

*Proof.* Left as an exercise. ∎

## 3.6 $\sigma$-algebra generated by random variables.

### 3.6.1 Definition

**Definition.** 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 $\sigma$-algebra generated by $X_\lambda, \lambda \in \Lambda$.

(i) For $\Lambda = \{1, 2, ..., n\}$ ($n$ may be $\infty$), we have

$$\begin{aligned} \sigma(X_i) &= \sigma\left(X_i^{-1}(\mathcal{B})\right) = X_i^{-1}(\mathcal{B}) = \{X_i \in \mathcal{B}\}, \\ \sigma(X_1, ..., X_n) &= \sigma\left(\cup_{i=1}^n X_i^{-1}(\mathcal{B})\right) = \sigma\left(\cup_{i=1}^n \sigma(X_i)\right). \end{aligned}$$

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

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

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

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

### 3.6.2 Some examples

EXAMPLE **3.6.1** *(The $\sigma-$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 $\infty$) and define $A_i = \{\omega : X(\omega) = x_i\}$. We have the following results.*

*(i) $\{A_i, i \ge 1\}$ constitute a disjoint partition of $\Omega$.*

*(ii) Choose $\mathcal{C} = \{A_1, A_2, ..., A_n\}$, then*

$$\sigma(\mathcal{C}) = \sigma(A_1, A_2, ..., A_n) = \sigma(A_0, A_1, A_2, ..., A_n) = \{\cup_{i \in I} A_i : \ I \subset \{0, 1, 2, ..., n\}\},$$

*where $A_0 = \emptyset$. (Hint: show that the RHS forms a $\sigma$-algebra.)*

*Proof.* Left as an exercise. ∎

REMARK **3.6.1** *One key assumption in the above example is that the sets $A_i$'s form a partition of $\Omega$ (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.*

Remark **3.6.2** *Note that when $n$ is finite, the total number of elements in $\sigma(A_1, A_2, ..., A_n)$ is*

$$\binom{n}{0} + \binom{n}{1} + ...... + \binom{n}{n} = (1+1)^n = 2^n.$$

Example **3.6.2** *Let $\Omega$ 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 $\Omega$.) 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 $|\sigma(A, B, C)|$.*

**Solution**. We provide two different methods to do this.

**Method 1.**

1. $\sigma(\emptyset) = \{\emptyset, \Omega\}$ and $|\sigma(\emptyset)| = 2$
2. $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$, $\qquad |\sigma(A)| = 4 = 2^2$.
3. 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 $|\sigma(A, B)| = 16 = 2^4 = 2^{2^2}$.
4. It 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.

**Method 2.** One alternative method is given below, which provides a general method to do this. The key is disjointization. If

$$\Omega = A_1 + A_2 + ... + A_n,$$

then,

$$\sigma(A_1, ..., A_n) = \{\cup_{i \in J} A_i, J \text{ is a subset of } \{0, 1, ..., n\}\},$$

where we denote $A_0 = \emptyset$.

1. $\sigma(\emptyset) = \{\emptyset, \Omega\}$ and $|\sigma(\emptyset)| = 2$
2. First we will find $\sigma(A)$. From the Venn diagram, we can disjointize $\Omega$ as

$$\Omega = A + A^c =: A_1 + A_2.$$

   Following Remark 3.6.3, $|\sigma(A)| = 2^2 = 4$. From the last example, we have

$$\begin{aligned} \sigma(A) &= \sigma(A, A^c) = \sigma(A_1, A_2) \\ &= \{\cup_{i \in J} A_i, J \text{ is a subset of } \{1, ..., n\}\} \\ &= \{\emptyset, A_1, A_2, A_1 \cup A_2\} \\ &= \{\emptyset, A, A^c, \Omega\} \end{aligned}$$

3. Next, we will find $\sigma(A,B)$. From the Venn diagram, we can disjointize $\Omega$ as
$$\Omega = (A-B)+(B-A)+(A\cap B)+(A\cup B)^c =: A_1+...+A_4.$$
Following Remark 3.6.3, $|\sigma(A,B)| = 2^4 = 16 = 4^2$. From the last example, we have
$$\begin{aligned}\sigma(A,B) &= \sigma(A_1,...,A_4)\\ &= \{\cup_{i\in J}A_i, J \text{ is a subset of } \{1,...,n\}\}\\ &= \{\emptyset, A_1, A_2, A_3, A_4, A_1+A_2, A_1+A_3, A_1+A_4, A_2+A_3, A_2+A_4, A_3+A_4,\\ &\quad A_1+A_2+A_3, A_1+A_2+A_4, ...., A_2+A_3+A_4, A_1+A_2+A_3+A_4\}\\ &= \{\emptyset, \Omega,\ A, B, A^c, B^c,\\ &\quad A\cup B, A\cup B^c, A^c\cup B, A^c\cup B^c,\\ &\quad A^c\cap B^c, A^c\cap B, A\cap B^c, A\cap B,\\ &\quad (A\cap B)\cup(A^c\cap B^c), ((A\cap B)\cup(A^c\cap B^c))^c\}.\end{aligned}$$

4. 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. ∎

Remark **3.6.3** *If $A_1,...,A_n$ are not mutually exclusive to each other for all pairs, then from the last example, we conjecture that $|\sigma(A_1,...,A_n)| = 2^{2^n}$. In fact, this can be argued as follows. Pick any $\omega$, it may or may not belong to $A_i$, $i=1,...,n$. Thus the total number of mutually exclusive sets is $2\times 2\times ....\times 2 = 2^n$. Following Remark , we get $|\sigma(A_1,...,A_n)| = 2^{2^n}$.*

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
$$\mathcal{F}_1=\sigma(A), \qquad \mathcal{F}_2=\sigma(A,B), \qquad \mathcal{F}_3=\sigma(A,B,C).$$
Find
$$E(Y|\mathcal{F}_1) = ???? \qquad E(Y|\mathcal{F}_2) = ???? \qquad E(Y|\mathcal{F}_3) = ????$$

Example **3.6.3** *Toss a coin three times. Then we can construct a probability space $(\Omega,\mathcal{F},P)$, where*
$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} =: \{A_1, A_2, ......, A_7, A_8\}.$$
$$\mathcal{F} = \textit{all possible subsets of } \Omega, \qquad |\mathcal{F}| = 2^8 = 256,$$
*Let $X_i$ = be the number of H's in the i-th toss, i=1,2,3. Namely, $X_i = I\{\textit{the ith toss is a Head}\}$ for $i=1,2,3$. So $X_i$'s are r.v.s taking two possible values 0, and 1. Now*
$$\begin{aligned}\sigma(X_1) &= \sigma(\{X_1=0\},\{X_1=1\}) = \sigma(\{X_1=1\}^c,\{X_1=1\}) = \sigma(\{X_1=1\})\\ &= \sigma(\{HHH,HHT,HTH,HTT\}) = \sigma(\{A_1,A_2,A_3,A_4\})\\ &= \{\emptyset,\Omega,\{A_1,A_2,A_3,A_4\},\{A_5,A_6,A_7,A_8\}\}. \qquad 2^2 = 4 \textit{ elements}\end{aligned}$$
*Similarly,*
$$\begin{aligned}\sigma(X_2) &= \sigma(\{X_2=1\})\\ &= \{\emptyset,\Omega,\{A_1,A_2,A_5,A_6\},\{A_3,A_4,A_7,A_8\}\}. \qquad 2^2=4 \textit{ elements}\\ \sigma(X_3) &= \sigma(\{X_3=1\})\\ &= \{\emptyset,\Omega,\{A_1,A_3,A_5,A_7\},\{A_2,A_4,A_6,A_8\}\}. \qquad 2^2=4 \textit{ elements}\\ \sigma(X_1,X_2) &= \sigma(\sigma(X_1),\sigma(X_2)) = \sigma(\{X_1=1\},\{X_2=1\})\\ &= \sigma(\{A_1,A_2,A_3,A_4\},\{A_1,A_2,A_5,A_6\})\\ &= 2^{2^2} = 16 \textit{ elements}\\ \sigma(X_1,X_3) &= 2^{2^2} = 16 \textit{ elements}\\ \sigma(X_2,X_3) &= 2^{2^2} = 16 \textit{ elements}\\ \sigma(X_1,X_2,X_3) &= \sigma(\{X_1=1\},\{X_2=1\},\{X_3=1\})\\ &= 2^{2^3} = 256 \textit{ elements.}\end{aligned}$$

*In fact, we can show that $\mathcal{F} = \sigma(X_1, X_2, X_3)$.* ∎

### 3.6.3 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 $\sigma$-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 $\{\omega : a < Y(\omega) \leq b\}$ do not belong to such a $\sigma$-algebra. However, it turns out that it is enough to have a $\sigma$-algebra generated by all (open, half open, or closed) intervals.

### 3.6.4 A useful theorem

THEOREM **3.6.1** *Let $X_1, ..., X_n$ be r.v.'s on $(\Omega, \mathcal{A})$. A real function $Y$ on $\Omega$ is $\sigma(X_1, ..., X_n)$-measurable (or a r.v. on the $\sigma$-algebra) iff $Y$ has the form $f(X_1, ..., X_n)$, where $f$ is a Borel function on $\mathcal{R}^n$.*

**Proof.** See Chow and Teicher, page 17.

## 3.7 Distributions and induced distribution functions

### 3.7.1 Case I: Random variables

Associated with every r.v. is a probability measure on $\mathcal{R}$.

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

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

**Proof.** Clearly, $P_X(B)$ is nonnegative and $P_X(\mathcal{R}) = P(\Omega) = 1$. It is also $\sigma$-additive as

$$P_X(\sum_i B_i) = P\left(X^{-1}(\sum_i B_i)\right) = \sum_i P\left(X^{-1}(B_i)\right) = \sum_i P_X(B_i).$$

**Definition:** $X$ is a r.v.

a) The **distribution** of $X$:

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

b) The **distribution function** of $X$:

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x).$$

**Definition:**

(i) 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 $X =_d Y$.

(ii) $X$ and $Y$ on $(\Omega, \mathcal{A}, P)$ are **equal almost surely (a.s.)** if $P(X = Y) = 1$, denoted by $X =_{a.s.} Y$.

**Remarks:**

(i) $X$ and $Y$ in Definition (i) does not have to be in the same probability space while in Definition (ii) they must be .

(ii) $X =_d Y$ 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:** 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 **3.7.2** *$X$ 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

$$P_X(C) = P_X\left(\cup_{i=1}^\infty \{a_i\}\right) = \sum_{i=1}^\infty P_X\left(\{a_i\}\right) = \sum_{i=1}^\infty [F_X(a_i) - F_X(a_i-)] = \sum_{i=1}^\infty p_i = 1.$$

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

$$F_X(x) = P(X \in [-\infty, x]) = P(X \in [-\infty, x] \cap C) = \sum_{a_i \in [-\infty, x]} P_X(\{a_i\})$$

$$= \sum_{i=1}^\infty P_X(\{a_i\}) I\{a_i \leq x\} = \sum_{i=1}^\infty P_X(\{a_i\}) \delta_{a_i}(x) := \sum_{i=1}^\infty p_i \delta_{a_i}(x).$$

That is, $F_X$ is discrete.

### 3.7.2 Case II: Random vectors

**Definition:** $X = (X_1, ..., X_n)$ is a random vector.

a) The **distribution** of $X$:

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

b) 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.

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

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

**Proof.** As $x_j \nearrow \infty$, $j \notin I$, we have

$$\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = \bigcap_{i=1}^n \{X_i \leq x_i\} \nearrow \bigcap_{j=1}^m \{X_{i_j} \leq x_{i_j}\}$$

**Definition:** 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 **3.7.4** *A random vector $X = (X_1, ..., X_n)$ 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 $C_i = \{x_i : (x_1, ..., x_n) \in C\}$, which is clearly countable. Then we have $C = C_1 \times ... \times C_n$. Then

$$\begin{aligned} P(X \in C) = 1, \quad &\Longleftrightarrow \quad P\left(\cap_1^n \{X_i \in C_i\}\right) = 1 \\ \Longleftrightarrow \quad P\left(\cup_1^n \{X_i \notin C_i\}\right) = 0 \quad &\Longleftrightarrow \quad P\left(\{X_i \notin C_i\}\right) = 0, \ 1 \le i \le n \\ \Longleftrightarrow \quad P\left(\{X_i \in C_i\}\right) = 1, 1 \le i \le n, \quad &\Longleftrightarrow \quad X_i \text{ is discrete for all } i. \quad \blacksquare \end{aligned}$$

## 3.8 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 $F_X(x) = P(X \le x)$. 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:** The inverse of a d.f. $F$, or quantile function associated with $F$, is defined by

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

An example of the plot of $F^{-1}(u)$ is given below.

(An example of the plot of $F^{-1}(u)$).

**Remarks.** From the above plot, we make the following observations, which turn out to be true in general.

(1). For $u \in (0, 1)$, the set $\{t : F(t) \ge u\} \ne \emptyset$ since $F(t) \to 1$ as $t \to \infty$. So $F^{-1}(u)$ is well defined.

(2). $F^{-1}(u)$ is **left continuous** (see the next theorem).

(3). $F^{-1}$ jumps when $F$ is flat. $F^{-1}$ is flat when $F$ jumps.

(4). In fact, $F^{-1}(u)$ is a mirror image of $F(t)$ along the line $u = t$.

Some of the properties of $F^{-1}(u)$ are listed in the next theorem.

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

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

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

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

(4). *$F^{-1}(u) \le t \Longleftrightarrow \quad u \le F(t)$.*

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

**Proof.**

(1). **Monotonicity.** Let $u_1 < u_2$

$\Longrightarrow$ if $F(t) \geq u_2$, then $F(t) \geq u_2 > u_1$,

$\Longrightarrow \{t : F(t) \geq u_1\} \supset \{t : F(t) \geq u_2\}$

$\Longrightarrow F^{-1}(u_1) = \inf\{t : F(t) \geq u_1\} \leq \inf\{t : F(t) \geq u_2\} = F^{-1}(u_2)$.

**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

$$\{t : F(t) \geq u\} = (a, \infty), \quad or \quad [a, \infty). \tag{8.2}$$

(In fact, we will see that it CAN not be of the first type $(a, \infty)$.)

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 $F(r') \geq F(r) \geq r$, implying that $r' \in \{t : F(t) \geq u\}$.

From (8.2), 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 $F(a + n^{-1}) \geq u$. Letting $n \to \infty$ and using the right continuity of $F$, we have $F(F^{-1}(u)) = F(a) = \lim_{n\to\infty} F(a + n^{-1}) \geq u$. ∎

**Remark.** 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

$$\{t : F(t) \geq u\} = [a, \infty) = [F^{-1}(u), \infty).$$

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 $F(t) \geq F(F^{-1}(u)) \geq u$ from (3).

(5). From (3), $F(F^{-1}(u)) \geq u$. 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 **3.8.2 (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

$$P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x). \quad \blacksquare$$

REMARK **3.8.1** *Theorem 3.8.2 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 3.8.2. 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**. Think about how to generate r.v.'s from

(i) Exponential(1), (ii) $N(0,1)$, (iii) Cauchy,
(iv) the empirical d.f. of observations $x_1, ..., x_n$: $F_n(x) = n^{-1}\sum_{i=1}^n I\{X_i \le x\}$.

**Solution.** Let $U, V \sim Uniform(0,1)$, and they are independent.

(i) Take $X = -\ln U$. (Using Theorem 3.8.2, we get $X = -\ln(1-U)$.)

(ii) 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.)

(iii) Let $X, Y \sim N(0,1)$ and they are independent, then $Z = X/Y$ is Cauchy.
(There is no close form for $F^{-1}(u)$.)

(iv). We note that $F_n(x)$ is a step function. Let $x_{(1)}, ..., x_{(n)}$ are ordered observations of $x_1, ..., x_n$ in ascending order. For any $U \sim$ Uniform(0,1), we can easily see that

$$\begin{aligned} F^{-1}(U) &= x_{(1)}, & 0 < U \le \frac{1}{n} \\ &= x_{(2)}, & \frac{1}{n} < U \le \frac{2}{n} \\ &= ...... \\ &= x_{(n)}, & \frac{n-1}{n} < U \le 1. \end{aligned}$$

Clearly, the range of $F^{-1}(U)$ is $\{x_{(1)}, ..., x_{(n)}\}$ and

$$P\left(F^{-1}(U) = x_{(k)}\right) = P\left(\frac{k-1}{n} < U \le \frac{k}{n}\right) = \frac{1}{n}, \qquad k = 1, 2, ..., n.$$

That is, to draw a random number from $F_n(x)$, we can simply draw from the observations with equal probability. $\blacksquare$

When $F$ is continuous (not necessarily absolutely continuous), a converse to Theorem 3.8.2 holds as well.

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

**Proof.** For $x \in (0,1)$,

$$\begin{aligned}
P(F(X) \le x) &= 1 - P(F(X) > x) \\
&= 1 - P(F(X) \ge x) \qquad \text{(from Lemma 3.8.1 below)} \\
&= 1 - P\left(X \ge F^{-1}(x)\right) \qquad \text{(from (4) of Theorem 3.8.1)} \\
&= 1 - P\left(X > F^{-1}(x)\right) \qquad (X \text{ is a continuous r.v.}) \\
&= P\left(X \le F^{-1}(x)\right) \\
&= F\left(F^{-1}(x)\right) = x \qquad \text{(from (5) of Theorem 3.8.1).} \quad \blacksquare
\end{aligned}$$

The following lemma is used in the above proof.

LEMMA **3.8.1** *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 3.8.1, 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,

$$\begin{aligned}
P(F(X) = t) &= P(X \in \{x : F(x) = t\}) = P(X \in [a, b]) \\
&= P(a \le X \le b) = P(a < X \le b) = F(b) - F(a) = 0. \quad \blacksquare
\end{aligned}$$

REMARK **3.8.2** *Lemma 3.8.1 states that if $X$ is continuous with d.f. $F$, so is $F(X)$. What about $G(X)$ for any continuous function $G$?*

## 3.9 Exercises

1. Let $\Omega = \{1,2,3,4,5,6\}$, and let $\mathcal{A} = \sigma(\{1,2,3,4\},\{3,4,5,6\})$.

   (a) List all sets in $\mathcal{A}$.
   (b) Is the function

$$\begin{aligned} X(\omega) &= 2 \qquad \omega = 1,2,3,4 \\ &= 7 \qquad \omega = 5,6 \end{aligned}$$

   a r.v. over $(\Omega, \mathcal{A})$?

2. Let $X, Y$ be r.v.'s on $(\Omega, \mathcal{A})$, and $A \in \mathcal{A}$.

   a) Show that $\{X \le Y\}$, $\{X < Y\}$, and $\{X = Y\}$ are events, i.e., they are all $\mathcal{A}-$measurable.
   b) Show that

$$\begin{aligned} Z(\omega) &= X(\omega) \qquad \omega \in A \\ &= Y(\omega) \qquad \omega \in A^c \end{aligned}$$

   is a r.v.

3. Show that $X^- = (-X)^+$, $(X+Y)^+ \le X^+ + Y^+$, $(X+Y)^- \le X^- + Y^-$, and $X^+ \le (X+Y)^+ + Y^-$.

# Chapter 4

# Expectation and Integration

## 4.1 Expectation

Let $X$ be a r.v. on $(\Omega, \mathcal{A}, P)$.

### 4.1.1 Expectation for simple r.v.'s

**Definition:** 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

$$EX = \sum_1^n a_i P(A_i). \quad \blacksquare$$

In other words, given a (measurable) partition $\{A_i\}$ of $\Omega$, 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 $A_i$.

**Remarks:**

(i). 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 $\infty - \infty$.

(ii). 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 **4.1.1** *$EX$ 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

$$\begin{aligned} LHS &= \sum_{i=1}^n \sum_{j=1}^m a_i I_{A_i \cap B_j} = \sum_{i=1}^n \sum_{j=1}^m a_i I_{A_i} I_{B_j} = \left( \sum_{i=1}^n a_i I_{A_i} \right) \left( \sum_{j=1}^m I_{B_j} \right) = \sum_{i=1}^n a_i I_{A_i} \\ RHS &= \sum_{i=1}^n \sum_{j=1}^m b_j I_{A_i \cap B_j} = \sum_{j=1}^m \sum_{i=1}^n b_j I_{A_i} I_{B_j} = \left( \sum_{j=1}^m b_j I_{B_j} \right) \left( \sum_{i=1}^n I_{A_i} \right) = \sum_{j=1}^m b_j I_{B_j}. \end{aligned}$$

So $a_i = b_j$ if $A_j \cap B_j \neq \emptyset$. Therefore,

$$\sum_{i=1}^{n} a_i P(A_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i P(A_i \cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_j P(A_i \cap B_j) = \sum_{j=1}^{m} b_j P(B_j). \quad \blacksquare$$

Some properties are listed below.

THEOREM **4.1.1** *$X, Y$ 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

$$aX + bY = \sum_{i=1}^{n}\sum_{j=1}^{m}(aa_i + bb_j) I_{A_i \cap B_j} \tag{1.1}$$

is a simple r.v. and therefore,

$$\begin{aligned} E(aX + bY) &= \sum_{i=1}^{n}\sum_{j=1}^{m}(aa_i + bb_j) P(A_i \cap B_j) \\ &= a\sum_{i=1}^{n} a_i P(A_i) + b\sum_{j=1}^{m} b_j P(B_j) = aEX + bEY. \end{aligned}$$

(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 $0 \leq E(X - Y) = EX - EY$.

**Remarks:**

(1). Probability is continuous. However, expectation is NOT.

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 $EX_n = nP((0, 1/n)) = 1 \not\to 0 = EX$.

(2). Expectation is continuous for the following two special cases:

(i) Monotone convergent sequence (see the Monotone Convergent Theorem later)

(ii) 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.

### 4.1.2 Expectation for nonnegative r.v.'s

**(I). DEFINITION**

Any nonnegative r.v. $X$ can be approximated by an increasing sequence of simple r.v.'s whose expectations are already defined.

THEOREM **4.1.2** *Given a nonnegative r.v. $X$, there exists simple r.v.'s $0 \le X_1 \le X_2 \le ...$ such that $X_n(\omega) \nearrow X(\omega)$ for every $\omega$.*

**Proof.** This has been shown in the last chapter. We just give a brief summary here. For each $n$, let

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} \le X < \frac{k}{2^n}\right\} + nI_{\{X \ge n\}} = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n.$$

Since $X$ is a r.v. $\Longrightarrow I_{\{a \le X < b\}}$ are indicator r.v.'s $\Longrightarrow X_n$ are simple r.v.'s.
On the measurable set (event) $\{(k-1)/2^n \le X < k/2^n\}$, $X_n$ equals the left endpoint $(k-1)/2^n$, which ensures

(a) $X_n(\omega) \le X_{n+1}(\omega) \quad \forall \omega \in \Omega.$ (**Monotonicity**)
(b) $0 \le X(\omega) - X_n(\omega) \le 1/2^n \quad \forall n \ge 1.$

Then $X_n(\omega)$ converges (from (a)) to $X(\omega)$ (from (b)).
**Definition:** $X \ge 0$ is a r.v. on $(\Omega, \mathcal{A}, P)$.

(a) The expectation of $X$ is $EX = \lim_{n\to\infty} EX_n \le \infty$, where $X_n$'s are simple, nonnegative, and $X_n \nearrow X$.

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

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

REMARK **4.1.1** (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 $X \ge 0$ or otherwise):*

$$\begin{aligned} EX &= \int_\Omega X(\omega)P(d\omega) = \int_\Omega XdP = \int XdP, \\ E_A X &= \int_A X(\omega)P(d\omega) = \int_A XdP. \end{aligned}$$

$EX$ is well defined in the sense:

THEOREM **4.1.3** *$X_n \ge 0$, $Y_n \ge 0$ 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) $X_n \nearrow X \Longrightarrow EX_n \le EX_{n+1} \Longrightarrow \lim_{n\to\infty} EX_n$ exists (maybe $\infty$). Similarly, $\lim_{n\to\infty} EY_n$ exists.

(2) Given $k \ge 1$, $Y_k$ is simple and $Y_k \le X$. If we can show that

$$\forall k \ge 1: \quad EY_k \le \lim_n EX_n, \tag{1.2}$$

which implies that $\lim_k EY_k \le \lim_n EX_n$. By symmetry, $\lim_k EY_k \ge \lim_n EX_n$. This proves the theorem.

**Proof of (1.2).** Fix $\epsilon > 0$ and $k \ge 1$. For each $n$, define $A_n = \{X_n > Y_k - \epsilon\}$, which is measurable. Note that

(i) $A_n \nearrow \Omega$ since $X_n \nearrow X$.
(ii) $X_n \geq (Y_k - \epsilon) I_{A_n}$ (*note that both sides are still simple r.v.'s.*)

Therefore,

$$
\begin{aligned}
EX_n &\geq E[(Y_k - \epsilon) I_{A_n}] = E\left(Y_k[1 - I_{A_n^c}]\right) - \epsilon P(A_n) \\
&\geq EY_k - \left(\max_{\omega \in \Omega} Y_k(\omega)\right) P(A_n^c) - \epsilon P(A_n) \\
&\quad (Y_k \text{ is simple, so "max" exists and is finite}) \\
&\to EY_k - \epsilon, \qquad \text{as } n \to \infty.
\end{aligned}
$$

Since $\epsilon$ is arbitrary, we have $\lim_n EX_n \geq EY_k$. The proof is finished.

### (II). ALGEBRAIC OPERATIONS

THEOREM **4.1.4** *$X, Y \geq 0$ 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 (1.1). Thus

$$
\begin{aligned}
E(aX + bY) &= \lim_n E(aX_n + bY_n) = \lim_n (aEX_n + bEY_n) \\
&= a \lim_n EX_n + b \lim_n EY_n = aEX + bEY.
\end{aligned}
$$

(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 **4.1.5** *$X \geq 0$. Then $EX = 0 \Longleftrightarrow X =_{a.s.} 0$.*

**Proof.** "$\Longrightarrow$". Suppose that $X =_{a.s.} 0$ is not true, i.e., $P(X = 0) < 1 \Longrightarrow F_X(0) = P(X \leq 0) = P(X = 0) < 1$, $\Longrightarrow \exists \epsilon > 0$ such that $F_X(\epsilon) < 1$ since $F_X$ is right continuous, Therefore,

$$
\begin{aligned}
EX &= E(XI_{\{X > \epsilon\}}) + E(XI_{\{X \leq \epsilon\}}) \\
&\geq \epsilon E(I_{\{X > \epsilon\}}) = \epsilon P(X > \epsilon) > 0.
\end{aligned}
$$

Contradiction. This proves that $X =_{a.s.} 0$
"$\Longleftarrow$". Noting $X \geq 0$, we have

$$
\begin{aligned}
0 &\leq EX = E(XI_{\{X=0\}}) + E(XI_{\{X \neq 0\}}) = E(XI_{\{X>0\}}) \\
&\leq \left(\sup_{\omega \in \Omega} X(\omega)\right) EI_{\{X>0\}} \leq \left(\sup_{\omega \in \Omega} X(\omega)\right) P(X > 0) \leq \infty \times 0 = 0.
\end{aligned}
$$

THEOREM **4.1.6** .

(*i*) *If $X > 0$ a.s., then $EX > 0$.* (i.e., strict inequality is preserved by expectation)
(*ii*) *If $EX \geq 0$, then $P(X \geq 0) > 0$.*

**Proof.** (i) We first show that if $X > 0$ a.s., then $EX > 0$. Clearly, $X > 0$ 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 $EX > 0$.

(ii) 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 $EX \geq 0$.

## (III). LIMITING OPERATIONS

The next theorem is critical in proving monotone continuity of expectations. It is also important in its own right.

THEOREM **4.1.7 (Fatou's lemma)**

1. *Suppose that $X_n \geq Y$ a.s. for all $n$ and some $Y$ with $E|Y| < \infty$.*

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} EX_n. \qquad \Longleftrightarrow \qquad E\left(\underline{\lim}_n X_n\right) \leq \underline{\lim}_n EX_n.$$

*(Typically, one can choose $Y = 0$ in practice.)*

2. *Suppose that $X_n \leq Y$ a.s. for all $n$ and some $Y$ with $E|Y| < \infty$.*

$$E\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} EX_n. \qquad \Longleftrightarrow \qquad E\left(\overline{\lim}_n X_n\right) \geq \overline{\lim}_n EX_n.$$

*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 $Z_m \nearrow \underline{\lim} X_n$. (Recall $\underline{\lim} X_n = \liminf_{n\to\infty} X_n = \sup_{m\geq 1}\inf_{k\geq m} X_k \equiv \sup_{m\geq 1} Z_m = \lim_m Z_m$)

   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,

$$\begin{aligned} EX_m &\geq E[(Y_k - \epsilon)I_{A_m}] = E\left(Y_k[1 - I_{A_m^c}]\right) - \epsilon P(A_m) \\ &\geq EY_k - \left(\max_{\omega\in\Omega} Y_k(\omega)\right) P(A_m^c) - \epsilon P(A_m) \\ &\quad (Y_k \text{ is simple, so "max" exists and is finite}) \\ &\to EY_k - \epsilon, \qquad \text{as } m \to \infty. \end{aligned}$$

   Since $\epsilon$ is arbitrary, we have $\underline{\lim}_m EX_m \geq EY_k$. Letting $k \to \infty$, we get $\underline{\lim}_m EX_m \geq \lim_{k\to\infty} EY_k = E\left(\underline{\lim}_n X_n\right)$.

2. The proof is similar to that for lim inf and hence omitted. ∎

REMARK **4.1.2** *We have the following simple relationships.*

$$\textit{Fatou's Lemma} \qquad \Longleftrightarrow \qquad \textit{the Monotone Convergence Theorem (MCT).}$$

*(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*

$$\liminf_n E \inf_{k\geq n} X_k = \lim_n E \inf_{k\geq n} X_k = E \lim_n \inf_{k\geq n} X_k \leq E \liminf_n X_n. \quad ∎$$

REMARK **4.1.3**

*(a) The inequality in Fatou's lemma can be strict. For example, in the last example, we have* $\lim_n X_n(\omega) = 0$ *and* $EX_n = 1$*. So* $E\underline{\lim}X_n = 0 \neq \underline{\lim}EX_n = 1$*.*

*(b) Fatou's lemma is particularly useful in establishing results about the convergence of the expectations of the elements of a sequence of random variables.*

*(c) 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*

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} EX_n.$$

*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.*

Now we shall introduce the first form of continuity for expectations: monotone continuity.

THEOREM **4.1.8 (Monotone convergence theorem)** *Let* $X, X_1, X_2, \ldots$ *be nonnegative r.v.'s. Then,*

(1) $X_n(\omega) \nearrow X(\omega) \implies EX_n \nearrow EX$.

(2) $X_n(\omega) \searrow X(\omega)$ *and* $EX_m < \infty$ *for some* $m \geq 1$, $\implies EX_n \searrow EX$.

**Proof.**

(1) $X_n \leq X \implies EX_n \leq EX \implies \limsup_n EX_n \leq EX$. 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, $X_n(\omega) \searrow X(\omega) \implies X_1(\omega) - X_n(\omega) \nearrow X_1(\omega) - X(\omega) \implies EX_1 - EX_n \nearrow EX_1 - EX \implies EX_n \searrow EX$.

THEOREM **4.1.9** *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 $\omega$ from the assumptions. Applying the Monotone Convergence Theorem: $RHS = \lim_n EX_n = EX = LHS$.

**IV. EXPECTATIONS EXTEND PROBABILITIES**

(1). $A_n \nearrow A \implies I_{A_n} \nearrow I_A \implies P(A_n) = EI_{A_n} \nearrow EI_A = P(A)$. Similarly, $A_n \searrow A \implies P(A_n) \searrow P(A)$. So the Monotone Convergence Theorem implies monotone continuity of $P$.

(2). Take $Y_k = I_{A_k}$ where $A_k$ are disjoint. Then

$$P(\sum_1^\infty A_k) = EI_{\sum_1^\infty A_k} = \sum_1^\infty EI_{A_k} = \sum_1^\infty P(A_k),$$

the $\sigma$-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

$$P(\liminf_n A_n) = EI_{\liminf_n A_n} = E\liminf_n I_{A_n} \leq \liminf_n P(A_n),$$

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:

If $(\Omega, \mathcal{A}, \mu)$ is a measure space, and $A_n \in \mathcal{A}$.

$(i)$ 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 $\mu(\overline{\lim}\ A_n) \geq \overline{\lim}\ \mu(A_n)$.

$(ii)$ If $\mu$ 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.**

$(i)$ $\mu(\underline{\lim} A_n) = \mu\left(\lim_n \cap_{i=n}^{\infty} A_i\right) = \lim_n \mu\left(\cap_{i=n}^{\infty} A_i\right) \leq \underline{\lim}_n \mu(A_n)$,

and

$\mu(\overline{\lim} A_n) = \mu\left(\lim_n \cup_{i=n}^{\infty} A_i\right) = \lim_n \mu\left(\cup_{i=n}^{\infty} A_i\right) \geq \overline{\lim}_n \mu(A_n)$,

where in the very last inequality, we used the assumption $\mu(\cup_{i=n}^{\infty} A_i) < \infty$ for some $n \geq 1$.

$(ii)$ Since $\mu$ is finite, we apply (i) to get

$$\overline{\lim_n} \mu(A_n) \leq \mu(\overline{\lim} A_n) = \mu(A) = \mu(\underline{\lim} A_n) \leq \underline{\lim}_n \mu(A_n).$$

It is clear that the critical steps for continuity of **finite measure** $\mu$ is the two inequalities:

$$\mu(\underline{\lim} A_n) \leq \underline{\lim}_n \mu(A_n), \qquad \mu(\overline{\lim} A_n) \geq \overline{\lim}_n \mu(A_n).$$

If we can show that

$$E(\underline{\lim} X_n) \leq \underline{\lim}_n E(X_n), \qquad E(\overline{\lim} X_n) \geq \overline{\lim}_n E(X_n),$$

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 $EX_n = 1 \not\to 0$.

### 4.1.3 Expectation for general r.v.'s

**(I). DEFINITION**

Recall $X^+ = \max\{X, 0\} = XI_{\{X \geq 0\}}$, $X^- = \max\{-X, 0\} = -XI_{\{X \leq 0\}}$, and

$$\begin{aligned} X &= XI_{\{X\geq 0\}} + XI_{\{X\leq 0\}} = XI_{\{X\geq 0\}} - (-XI_{\{X\leq 0\}}) = X^+ - X^-, \\ |X| &= |X|I_{\{X\geq 0\}} + |X|I_{\{X\leq 0\}} = XI_{\{X\geq 0\}} + (-XI_{\{X\leq 0\}}) = X^+ + X^-. \end{aligned}$$

So $EX$ can be defined by $EX^+$ and $EX^-$. But we need to be careful to avoid $EX^+ - EX^- = \infty - \infty$.

**Definition:** Let $X$ be a r.v. on $(\Omega, \mathcal{A}, P)$ (not necessarily nonnegative).

(a) 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 $EX \in [-\infty, \infty]$.

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

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

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

$$E_A X = E(XI_A)$$

**Remark:**

(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. $X \in L^1 \Longleftrightarrow E|X| < \infty \Longleftrightarrow EX^+ < \infty$ 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, $E|X| < \infty \Longrightarrow E|XI_A| < \infty$. That is, $E(XI_A)$ in (d) is well defined.)

**(II). ALGEBRAIC OPERATIONS**

THEOREM **4.1.10** .

(1). *$X, Y \in L^1$, and $a, b \in \mathcal{R}$, then $aX + bY \in L^1$, and*

$$E(aX + bY) = aEX + bEY. \qquad \textbf{(Linearity)}$$

(2). *For $X \in L^1$, $|EX| \le E|X|$.*

(3). *$X, Y \in L^1$, and $X \le Y$, then $EX \le EY$. (**Monotonicity**)*

**Proof.**

(1) By the triangle inequality: $|aX+bY| \le |a||X|+|b||Y|$. Note that each term is nonnegative, so their expectations are well defined and monotonicity of expectations for nonnegative r.v.'s implies

$$E|aX + bY| \le |a|E|X| + |b|E|Y| < \infty.$$

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)**. We next show that if $Z_1, Z_2 \ge 0$, then $E(Z_1 - Z_2) = EZ_1 - EZ_2$. If suffices to show

$$\begin{aligned} E(Z_1 - Z_2) &\equiv E(Z_1 - Z_2)^+ - E(Z_1 - Z_2)^- \\ &\equiv E(Z_1 - Z_2)I_{\{Z_1 \ge Z_2\}} - E(Z_2 - Z_1)I_{\{Z_1 \le Z_2\}} \\ &= EZ_1 - EZ_2, \end{aligned}$$

or equivalently we need to show that

$$EZ_1 + E(Z_2 - Z_1)I_{\{Z_1 \le Z_2\}} = EZ_2 + E(Z_1 - Z_2)I_{\{Z_1 \ge Z_2\}},$$

which follows easily from the identity

$$Z_1 + (Z_2 - Z_1)I_{\{Z_1 \le Z_2\}} = Z_2 + (Z_1 - Z_2)I_{\{Z_1 \ge Z_2\}}.$$

(Note LHS = RHS under 3 cases: $Z_1 < Z_2$, $Z_1 = Z_2$, $Z_1 > Z_2$.)

It then follows that

$$\begin{aligned} E(X + Y) &= E(X^+ - X^- + Y^+ - Y^-) \\ &= E\left((X^+ + Y^+) - (X^- + Y^-)\right) \\ &= E(X^+ + Y^+) - E(X^- + Y^-) \\ &= (EX^+ + EY^+) - (EX^- + EY^-) \\ &= EX + EY \end{aligned}$$

**Proof of (b)**. If $a \ge 0$, then

$$\begin{aligned} E(aX) &= E[(aX)^+] - E[(aX)^-)] = E[aX^+] - E[aX^-] \\ &= a(EX^+ - EX^-) = aEX. \end{aligned}$$

If $a < 0$, then

$$(aX)^+ = aXI_{\{aX \ge 0\}} = (-a)(-X)I_{\{X \le 0\}} = (-a)X^-$$

Similarly, $(aX)^- = (-a)X^+$. Therefore,

$$\begin{aligned} E(aX) &= E[(aX)^+] - E[(aX)^-)] = E[(-a)X^-] - E[(-a)X^+] \\ &= -a(EX^- - EX^+) = aEX. \end{aligned}$$

(2) $|EX| = |EX^+ - EX^-| \le EX^+ + EX^- = E|X|$.

(3) $X - Y \in L^1$ and from (1), we get $EY - EX = E(Y - X) = E(Y - X)^+ \ge 0$. ∎

Theorem **4.1.11** $X =_{a.s.} Y \Longrightarrow EX = EY$.

**Proof.** Let $Z = X - Y$, then

$$1 = P(X = Y) = P(X - Y = 0) = P(Z = 0) = P(|Z| = 0) = P(Z^+ = 0, Z^- = 0).$$

Therefore, $P(Z^+ = 0) = 1$ and $P(Z^- = 0) = 1$ (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,

$$EX = E[Y + (X - Y)] = EY + E(X - Y) = EY. \quad \blacksquare$$

## (III). LIMITING OPERATIONS

We are now ready to introduce the second form of continuity for expectations: the dominated continuity. Below, we shall write $X_n \to X$ to denote $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$.

Theorem **4.1.12 (Dominated Convergence Theorem)** *If $X_n \to X$ a.s., $|X_n| < Y$ for all $n$, and $EY < \infty$, then*

$$\lim_n EX_n = EX \ (= E \lim_n X_n).$$

**Proof.** First, from the assumptions, we have that $|X| \le |X_n - X| + |X_n| \le C + Y$ a.s. Thus, $E|X| \le C + EY < \infty$. Hence, $X \in L^1$.

Secondly, $|X_n| \le Y \Longleftrightarrow -Y \le X_n \le Y$ for each $n$.

Now $Y - X_n \ge 0 \Longrightarrow$

$$\begin{aligned} EY - EX &= E(Y - X) = E[\liminf_n (Y - X_n)] \le \liminf_n E(Y - X_n)] \\ &= EY + \liminf_n [-EX_n] = EY - \limsup_n EX_n. \end{aligned}$$

Thus, $\limsup_n EX_n \le EX$.

On the other hand, $Y + X_n \ge 0 \Longrightarrow$

$$\begin{aligned} EY + EX &= E(Y + X) = E[\liminf_n (Y + X_n)] \le \liminf_n E(Y + X_n)] \\ &= EY + \liminf_n EX_n. \end{aligned}$$

Thus, $\liminf_n EX_n \ge EX$.

Combining the two, we get $\limsup_n EX_n \le EX \le \liminf_n EX_n$. This implies that $\lim_n EX_n = EX$. $\blacksquare$

**Remark:**

(1) Going back to our earlier example, if $Y$ dominates $X_n = nI_{(0,1/n)}$ for each $n$, then we must have $Y \ge \sum_1^m nI_{(1/(n+1),1/n)}$ for all $m \ge 1$, thus letting $m \to \infty$, we get $Y \ge \sum_1^\infty nI_{(1/(n+1),1/n)}$. This implies that

$$\begin{aligned} EY &\ge E\sum_1^\infty nI_{(1/(n+1),1/n)} = E\lim_{m\to\infty}\sum_1^m nI_{(1/(n+1),1/n)} \\ &= \lim_{m\to\infty} E\sum_1^m nI_{(1/(n+1),1/n)} \qquad \text{(by Monotone convergence theorem)} \\ &= \sum_1^\infty nEI_{(1/(n+1),1/n)} = \sum_1^\infty nP\left((1/(n+1), 1/n)\right) \\ &= \sum_1^\infty n\left(1/n - 1/(n+1)\right) = \sum_1^\infty 1/(n+1) \\ &= \infty. \end{aligned}$$

In other words, if there exists a r.v. $Y \ge 0$ which dominates all $X_n$ $(n \ge 1)$, 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 $P$.

### 4.1.4 Summary

Assume that $X, Y, X_1, ..., X_n$ below are all r.v.'s on $(\Omega, \mathcal{A}, \mu)$.

**(Absolute integrability).** $EX$ is finite if and only if

$$E|X| \text{ is finite.}$$

**(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

$$E(aX + bY) = aEX + bEY.$$

**($\sigma$-additivity over sets).** If $A = \sum_{i=1}^{\infty} A_i$

$$E_A X = \sum_{i=1}^{\infty} E_{A_i} X.$$

**(Positivity).** If $X \geq 0$ a.s., then

$$EX \geq 0.$$

**(Monotonicity).** If $X_1 \leq X \leq X_2$ a.s., then

$$EX_1 \leq EX \leq EX_2.$$

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

$$aP(A) \leq E_A X \leq bP(A).$$

**(Modulus inequality).**

$$|EX| \leq E|X|.$$

**(Fatou's Lemma).** If $X_n \geq 0$ a.s., then

$$E\left(\liminf_n X_n\right) \leq \liminf_n EX_n.$$

**(Monotone Convergence Theorem).** If $0 \leq X_n \nearrow X$, then

$$\lim_n EX_n = EX = E \lim_n X_n.$$

**(Dominated Convergence Theorem).** If $X_n \to X$ a.s., $|X_n| < Y$ a.s. for all $n$, and $EY < \infty$, then

$$\lim_n EX_n = EX = E \lim_n X_n.$$

**(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

$$E\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} EX_n.$$

**Remarks.**

(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:

$$\infty, \qquad -\infty, \qquad finite.$$

(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 **4.1.13** *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 **4.1.14** *If $E|X| < \infty$, then $|X| < \infty$ a.s.*
*(Namely, if $P(|X| = \infty) > 0$, 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.

**PROOFS OF THE SUMMARY.**

**(Absolute integrability).** $EX$ is finite $\Longleftrightarrow$ $E|X|$ is finite.

*Proof.* $EX < \infty \Longleftrightarrow EX^+ < \infty$ and $EX^- < \infty \Longleftrightarrow E|X| < \infty$.

**(Linearity).** If the RHS below is meaningful, namely not $+\infty - \infty$ or $-\infty - \infty$, (e.g. if $X, Y \ge 0$ and $a, b \ge 0$, or if $X, Y \in L^1$ and $a, b \in \mathcal{R}$), then

$$E(aX + bY) = aEX + bEY.$$

*Proof.* There are four possible cases:
(1) $X, Y \in L^1$. This has been shown.
(2) $X \in L^1$ and $Y \notin L^1$. Similar to (3) below.
(3) $X \notin L^1$ 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

$$LHS = E[(aX^+ + bY^+) - (aX^- + bY^-)].$$

But $E[(aX^+ + bY^+) \ge E(aX^+) = aEX^+ = \infty$, $E(aX^- + bY^-) < \infty$, therefore, LHS$= \infty =$ RHS.
(4) $X \notin L^1$ 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 $ab \ge 0$. Assume that $a, b > 0$. Then the rest of the proof is similar to that in (3) above.

**($\sigma$-additivity over sets).** If $A = \sum_{i=1}^\infty A_i$

$$E_A X = \sum_{i=1}^\infty E_{A_i} X.$$

*Proof.* If $X \ge 0$, 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. $X = X^+ - X^-$, the proof is left as an exercise.

**(Positivity).** If $X \geq 0$ a.s., then $EX \geq 0$.

**(Monotonicity).** If $X_1 \leq X \leq X_2$ a.s., then $EX_1 \leq EX \leq EX_2$.

*Proof.* It suffices to show that $X \leq Y$ a.s. implies $EX \leq EY$.
If both $EX$ and $EY$ are $\infty$ (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

$$EY - EX = E(Y - X) \geq 0.$$

**(Mean value theorem).** If $a \leq X \leq b$ a.s. on $A$, then

$$aP(A) \leq E_A X \leq bP(A).$$

*Proof.* $a \leq X \leq b$ a.s. implies $aI_A \leq XI_A \leq bI_A$ a.s. The proof is done by taking expectation on both sides.

**(Modulus inequality).** $|EX| \leq E|X|$.

**(Fatou's Lemma).** If $X_n \geq 0$ a.s., then

$$E\left(\liminf_n X_n\right) \leq \liminf_n EX_n.$$

**(Monotone Convergence Theorem).** If $0 \leq X_n \nearrow X$, then

$$\lim_n EX_n = EX = E\lim_n X_n.$$

**(Dominated Convergence Theorem).** If $X_n \to X$ a.s., $|X_n| < Y$ a.s. for all $n$, and $EY < \infty$, then

$$\lim_n EX_n = EX = E\lim_n X_n.$$

**(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

$$E\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} EX_n.$$

*Proof.* 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 4.1.14, 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 $E\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} EX_n$.

## 4.2 Integration

### 4.2.1 Definition

Integration over a measure space $(\Omega, \mathcal{A}, \mu)$ is conceptually the same as expectation. The integral is defined via the same procedure, and the properties are *nearly* identical.

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

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

a) If $f = \sum_1^n a_i I_{A_i}$ with $a_i \geq 0$,

$$\int f d\mu = \sum_1^n a_i \mu(A_i).$$

b) 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 $f_n \nearrow f$.

**Remark**: 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 $\psi$ such that $\psi(\omega) \leq f(\omega)$ for any $\omega \in \Omega$.

c) 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.

d) $f$ is said to be **integrable** w.r.t. $\mu$ 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.

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

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

**Remark:** By taking $f = I_A$, then it follows from (a) and (e) that

$$\mu(A) = \int I_A d\mu = \int_A d\mu$$

### 4.2.2 Some properties of integrals

Here are some properties. Proofs are similar to expectations and hence ommitted.

Assume that all functions below are Borel measurable on $(\Omega, \mathcal{A}, \mu)$.

**(Absolute integrability).** $\int f$ is finite if and only if $\int |f|$ is finite.

**(Reariity).** 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

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

**($\sigma$-additivity over sets).** If $A = \sum_{i=1}^{\infty} A_i$

$$\int_A f d\mu = \sum_{i=1}^{\infty} \int_{A_i} f d\mu.$$

**(Positivity).** If $f \geq 0$ a.e., then

$$\int f d\mu \geq 0.$$

**(Monotonicity).** If $f_1 \leq f \leq f_2$ a.e., then

$$\int f_1 \leq \int f \leq \int f_2.$$

**(Mean value theorem).** If $a \leq f \leq b$ a.e. on $A \in \mathcal{A}$, then

$$a\mu(A) \leq \int_A f d\mu \leq b\mu(A).$$

**(Modulus inequality).**

$$|\int f| \leq \int |f|.$$

**(Fatou's Lemma).** If $f_n \geq 0$ a.e., then

$$\int \liminf_n f_n \leq \liminf_n \int f_n.$$

**(Monotone Convergence Theorem).** If $0 \leq f_n \nearrow f$, then

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

**(Dominated Convergence Theorem).** If $f_n \to f$ a.e., $|f_n| < g$ a.e. for all $n$, and $\int g < \infty$, then

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

**(Integration term by term).** If $\sum_{i=1}^{\infty} \int |f_n| < \infty$, then

$$\sum_{i=1}^{\infty} |f_n| < \infty, \ a.e.$$

so that $\sum_{i=1}^{\infty} f_n$ converges a.e., and

$$\int \sum_{i=1}^{\infty} f_n = \sum_{i=1}^{\infty} \int f_n.$$

### 4.2.3 Special cases: Lebesgue-Stieltjes integral and Lebesgue integral

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

(a). 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. $\mu$.**

(b). In the case of $(\Omega, \mathcal{A}, \mu) = (\mathcal{R}, \mathcal{B}, \lambda)$, where $\lambda$ 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. $\lambda$. (Note: $\lambda(dx) = dx$.)**

**Definition:** 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 $\mu$ on the measurable space $(\mathcal{R}, \mathcal{B})$ such that

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

(a). 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. $F$.**

(b). In the special case $F(x) = x$, the unique measure $\mu$ determined from (2.3) reduces to the Lebesgue measure $\lambda$. 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: $\lambda(dx) = dx$.)**

**Remark:** $X$ is a r.v. on $(\Omega, \mathcal{A}, P)$.

(1) Expectations are special cases of integrals:

$$EX = \int X(\omega)P(d\omega) = \int X(t)dF_X(t), \qquad \text{where } F_X(t) = P(X \le t).$$

(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 [actually we have not discussed the R-S integral in this course yet] 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 $R$. The abstract definition quickly translate into straightforward formulas when $G$ has some properties.

1. $G$ **is a discrete (i.e., a step) function.**

   When $G$ is a step function, it will have at most countably many jumps $\{x_1, x_2, ...\}$, where $\Delta G(x_n) = G(x_n) - G(x_n-) > 0$. The measure $\mu$ will be discrete with positive measure at each of the points $x_1, x_2, ...$, 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 $\{x_1, x_2, ...\}$, 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 **4.2.1** *Let $F, G$ be right-continuous functions of bounded variation. Then*

$$F(t)G(t) - F(s)G(s) = \int_{(s,t]} F(x-)dG(x) + \int_{(s,t]} G(x)dF(x) \tag{2.4}$$

$$= \int_{(s,t]} F(x)dG(x) + \int_{(s,t]} G(x-)dF(x), \tag{2.5}$$

*and*

$$F(t)G(t) - F(s)G(s) = \int_{(s,t]} F(x-)dG(x) + \int_{(s,t]} G(x-)dF(x) + \sum_{n:s<x_n\le t} \Delta F(x)\Delta G(x). \tag{2.6}$$

*Proof.* WLOG, we will only prove the case for $s = 0$. Note that, whenever $s \le t$, by definition,

$$F(s) - F(0) = \mu((0,s]) = \int_{(0,s]} d\mu = \int_{(0,s]} dF(x) = \int_{(0,t]} I_{\{0<x\le s\}} dF(x),$$

$$F(s-) - F(0) = \mu((0,s)) = \int_{(0,s)} d\mu = \int_{(0,s)} dF(x) = \int_{(0,t]} I_{\{0<x<s\}} dF(x).$$

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 (2.4). Proof of (2.5) is similar.
To prove (2.8), in view of (2.4), 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,

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

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

$$\int_{(s,t]} \Delta G(x) dF_c(x) = 0, \tag{2.7}$$

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. ∎

REMARK **4.2.1** *Here we provide a more detailed justification of (2.7). Denote $D_G$ and $C_G$ to be the set of all jump points and continuity points of $G$, respectively. Then,*

$$\begin{aligned}
\int_B \Delta G(x) dF_c(x) &= \int_{B\cap C_G} \Delta G(x) dF_c(x) + \int_{B\cap D_G} \Delta G(x) dF_c(x) \\
&= \int_{B\cap C_G} 0 dF_c(x) + \int_{B\cap D_G} \Delta G(x) dF_c(x) \\
&= \int_{B\cap C_G} 0 dF_c(x) + \int_{B\cap D_G} \Delta G(x) d\mu_c(x) \\
&\qquad \textit{($\mu_c$ is the corresponding measure to $F_c$)} \\
&= 0
\end{aligned}$$

*since $B \cap C_G$ is a countable set, and*

$$\mu_c(B \cap C_G) = \sum_{x_n \in B\cap D_G} \mu_c(\{x_n\}) = \sum_{x_n \in B\cap D_G} [F_c(x_n) - F_c(x_n-)] = 0. \quad ∎$$

REMARK **4.2.2** *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.*

## 4.3 Measure-theoretic and probabilistic languages

We now have two parallel languages, measure-theoretic and probabilistic:

| Measure | Probability |
|---|---|
| Integral | Expectation |
| Measurable set | Event |
| Measurable set | Random variable |
| Almost everywhere (a.e.) | Almost surely (a.s.) |

## 4.4 How to compute expectation

Integrating over $(\Omega, \mathcal{A}, P)$ is nice in theory, but to do computations we have to shift to a space on which we can do calculus. The following theorem will prove most useful.

THEOREM **4.4.1 (Change of variable formula)** .
*Assume the following holds:*

*(i) 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 $X$.*

*(ii) $g$ is Borel on $(\Omega_0, \mathcal{A}_0)$.*

*(iii) Either $g \geq 0$ or $E|g(X)| < \infty$.*

*Then*

$$Eg(X) = \int_{\Omega_0} g(y) P_X(dy).$$

**Remark:**

(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 $\omega$ 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.

(4) We will prove this result by verifying it in four increasingly more general special cases. The method employed here is a common one, and will be used several times later.

**Proof of the theorem:**

**Case I: Indicator functions**. If $g = I_B$ with $B \in \mathcal{A}_0$, then the relevant definitions show

$$\begin{aligned} Eg(X) &= EI_B(X) = P(X \in B) = P_X(B) = \int I_B dP_X \\ &= \int_{\Omega_0} I_B(y) P_X(dy) = \int_{\Omega_0} g(y) P_X(dy). \end{aligned}$$

**Case II: Simple functions**. Let If $g = \sum_{i=1}^n b_i I_{B_i}$ with $B_i \in \mathcal{A}_0$. The linearity of expected value, the result of Case I, and the linearity of integration imply

$$\begin{aligned} Eg(X) &= E\left(\sum_{i=1}^n b_i I_{B_i}(X)\right) = \sum_{i=1}^n b_i E I_{B_i}(X) = \sum_{i=1}^n b_i \int_{\Omega_0} I_{B_i}(y) P_X(dy) \\ &= \int_{\Omega_0} \left(\sum_{i=1}^n b_i I_{B_i}(y)\right) P_X(dy) = \int_{\Omega_0} g(y) P_X(dy). \end{aligned}$$

**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

$$Eg(X) = \lim_n Eg_n(X) = \lim_n \int_{\Omega_0} g_n(y) P_X(dy) = \int_{\Omega_0} g(y) P_X(dy).$$

**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

$$\begin{aligned} Eg(X) &= Eg(X)^+ - Eg(X)^- \\ &= \int_{\Omega_0} g(y)^+ P_X(dy) - \int_{\Omega_0} g(y)^- P_X(dy) \\ &= \int_{\Omega_0} g(y) P_X(dy). \quad \blacksquare \end{aligned}$$

### 4.4.1 Expected values of absolutely continuous r.v.'s

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

$$P_X(B) = \int_B f d\lambda = \int_B f(x)dx, \qquad \forall B \in \mathcal{B}, \tag{4.8}$$

*where $\lambda$ is the L-measure.*

*Proof.* We shall give two different proofs.

**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 (4.8) are probability measures on $(\mathcal{R}, \mathcal{B})$. It follows from (4.9) below that $P_X|_\mathcal{S} = \mu|_\mathcal{S}$ on the semialgebra $\mathcal{S} = \{(a, b] : -\infty \le a \le b \le \infty\}$. By the uniqueness of the extensions of measures from a semialgebra $\mathcal{S}$ to the $\sigma$-algebra $\mathcal{B} = \sigma(\mathcal{S})$, we prove (4.8).

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

$$P_X((a, b]) = F_X(b) - F_X(a) = \int_{(a,b]} f(t)dt. \tag{4.9}$$

From the last lemma, we have

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

From Theorem 4.4.1, we have $Eg(X) = \int_\mathcal{R} g(x)P_X(dx)$. To complete our proof, we only need to show that

$$\int_\mathcal{R} g(x)P_X(dx) = \int_\mathcal{R} g(x)f(x)dx. \tag{4.10}$$

**Proof of (4.10).** We shall employ the same method used in the last theorem.

**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 (4.8).

**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

$$\begin{aligned} LHS &= \int \left( \sum_{i=1}^n b_i I_{B_i}(x) \right) P_X(dx) = \sum_{i=1}^n b_i \int I_{B_i}(x) P_X(dx) \\ &= \sum_{i=1}^n b_i \int I_{B_i}(y) f(y) dy = \int_{\mathcal{R}} \sum_{i=1}^n b_i I_{B_i}(y) f(y) dy = \int g(y) f(y) dy \\ &= RHS. \end{aligned}$$

**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

$$LHS = \lim_n \int g_n(y) P_X(dy) = \lim_n \int g_n(y) f(y) dy = \int g(y) f(y) dy = RHS.$$

**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., $Eg(X)^+ < \infty$ 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 (4.10), and hence the theorem. ∎

**Remarks:**

(i) For an absolutely continuous r.v. $X$, we have several equivalent expressions:

$$Eg(X) = \int_{\mathcal{R}} g(x) P_X(dx) = \int_{\mathcal{R}} g(x) dF_X(x) = \int_{\mathcal{R}} g(x) f(x) dx.$$

(ii) The last integral in (i) is L-integral, which equals R-integral when the latter exists. This will greatly facilitate our calculations.

### 4.4.2 Expected values of discrete r.v.'s

The expectation for discrete r.v.'s is easier to calculate.

THEOREM **4.4.3** *Let $X$ be a discrete r.v. taking values $x_1, x_2, ...,$ 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 $g(x_1), g(x_2), ...,$ and we can write

$$g(X) = \sum_{k=1}^\infty g(x_k) I_{\{X = x_k\}}$$

**Case I:** If $g(X) \geq 0$, then $g(x_k) \geq 0$ for all $k \geq 1$. Define

$$Z_n = \sum_{k=1}^n g(x_k) I_{\{X = x_k\}}, \qquad \text{a form of truncated r.v.}$$

Clearly, given $X$, we have $0 \le Z_n \nearrow Z_\infty \equiv g(X)$, 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

$$Eg(X) = \lim_n EZ_n = \lim_n \sum_{k=1}^n g(x_k)P(X = x_k) = \sum_{k=1}^\infty p_k g(x_k).$$

**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,

$$\begin{aligned} Eg(X) &= Eg(X)^+ - Eg(X)^- = \sum_{k=1}^\infty p_k g(x_k)^+ - \sum_{k=1}^\infty p_k g(x_k)^- \\ &= \sum_{k=1}^\infty p_k \left(g(x_k)^+ - g(x_k)^-\right) = \sum_{k=1}^\infty p_k g(x_k). \quad \blacksquare \end{aligned}$$

## 4.5 Relation between expectation and tail probability

We shall introduce several powerful inequalities relating expectations to tail probabilities.

THEOREM **4.5.1** *We have*

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

*So* $E|X| < \infty$ *if and only if* $\sum_{n=1}^\infty P(|X| \ge n) < \infty$.

*Proof.* By the $\sigma$-additivity of expectation over sets, if $A_n = \{n \le |X| < n+1\}$,

$$E|X| = E_{\sum_{n=0}^\infty A_n}|X| = \sum_{n=0}^\infty E_{A_n}|X|.$$

By the Mean Value Theorem, $nP(A_n) \le E_{A_n}|X| \le (n+1)P(A_n)$, thus

$$\sum_{n=1}^\infty nP(A_n) = \sum_{n=0}^\infty nP(A_n) \le E|X| \quad \le \quad \sum_{n=0}^\infty (n+1)P(A_n) = 1 + \sum_{n=1}^\infty nP(A_n). \tag{5.11}$$

It remains to show that

$$\sum_{n=1}^\infty nP(A_n) = \sum_{n=1}^\infty P(|X| \ge n). \tag{5.12}$$

**Proof of (5.12).** Note that

$$\begin{aligned} \sum_{n=1}^m nP(A_n) &= \sum_{n=1}^m nP(n \le |X| < n+1) = \sum_{n=1}^m n\left[P(|X| \ge n) - P(|X| \ge n+1)\right] \\ &= P(|X| \ge 1) - P(|X| \ge 2) \\ &\quad +2P(|X| \ge 2) - 2P(|X| \ge 3) \\ &\quad +3P(|X| \ge 3) - 3P(|X| \ge 4) \\ &\quad +\ldots\ldots\ldots\ldots \\ &\quad +mP(|X| \ge m) - mP(|X| \ge m+1) \\ &= \sum_{n=1}^m P(|X| \ge n) - mP(|X| \ge m+1). \end{aligned}$$

That is, $\sum_{n=1}^m P(|X| \ge n) = \sum_{n=1}^m nP(A_n) + mP(|X| \ge m+1)$. So

$$\sum_{n=1}^m nP(A_n) \le \sum_{n=1}^m P(|X| \ge n) \le \sum_{n=1}^m nP(A_n) + mP(|X| \ge m+1), \tag{5.13}$$

where the last term satisfies

$$mP(|X| \geq m+1) = EmI_{(|X|\geq m+1)} \leq E|X|I_{(|X|\geq m+1)}$$

**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, (5.12) is true.

**Case II:** $E|X| = \infty$. Here, from (5.11), we have $\sum_{n=1}^{\infty} nP(A_n) = \infty$. And then from (5.13), we get $\sum_{n=1}^{m} P(|X| \geq n) = \infty$. Thus, (5.12) is true as well.

Hence, we proved (5.12), and therefore the theorem. ∎

**Remark.** 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 **4.5.1** *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 (5.12). ∎

For any non-negative r.v. (discrete or continuous), we have the following integral expression.

Theorem **4.5.2** *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.** We'd like to use Corollary 4.5.1, 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,

$$EY = \lim_{n\to\infty} EY_n = \lim_{n\to\infty} 2^{-n} EX_n. \tag{5.14}$$

Now that $X_n$ is a nonnegative and integer-valued r.v., from the last corollary, we get

$$EX_n = \sum_{j=1}^{\infty} P(X_n \geq j) = \sum_{j=1}^{\infty} P([2^n Y] \geq j) = \sum_{j=1}^{\infty} P\left(2^n Y \geq j\right).$$

But

$$\int_0^{\infty} P(Y \geq y)dy = \sum_{j=0}^{\infty} \int_{j/2^n}^{(j+1)/2^n} P\left(Y \geq y\right) dy$$

The above two relations give us

$$\begin{aligned} \int_0^{\infty} P(Y \geq y)dy &\leq \sum_{j=0}^{\infty} \frac{1}{2^n} P\left(Y \geq \frac{j}{2^n}\right) = \frac{1}{2^n} \sum_{j=0}^{\infty} P\left(2^n Y \geq j\right) \\ &= \frac{1}{2^n}\left(\sum_{j=1}^{\infty} P\left(2^n Y \geq j\right) + P(2^n Y \geq 0)\right) = \frac{1}{2^n}(EX_n + 1), \\ \int_0^{\infty} P(Y \geq y)dy &\geq \sum_{j=0}^{\infty} \frac{1}{2^n} P\left(Y \geq \frac{j+1}{2^n}\right) = \frac{1}{2^n} \sum_{j=1}^{\infty} P\left(2^n Y \geq j\right) = \frac{1}{2^n} EX_n \end{aligned}$$

Therefore,

$$EY_n = \frac{1}{2^n}EX_n \le \int_0^\infty P(Y \ge y)dy \le \frac{1}{2^n}(EX_n + 1) = EY_n + \frac{1}{2^n}.$$

Letting $n \to \infty$ and using (5.14), we prove the theorem.

The second inequality follows from the fact $\int_0^\infty P(Y = x)dx = 0$, which will be left as an exercise. See the homework. ∎

For moments of any positive order, the following identity holds.

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

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

**Proof.** Applying the last theorem, we get

$$EY^r = \int_0^\infty P(Y^r \ge x)dx = \int_0^\infty P(Y \ge x^{1/r})dx = r\int_0^\infty z^{r-1}P(Y \ge z)dz, \text{ where } z = x^{1/r}.$$

The second equality holds due to the property of L-integrals. ∎

For any integral r.v., we have

Corollary **4.5.3** *If $Y$ is integrable, then*

$$EY = EY^+ - EY^- = \int_0^\infty P(Y > x)dx - \int_0^\infty P(Y \le -x)dx.$$

Proof. First, $EY^+ = \int_0^\infty P(YI\{Y \ge 0\} > x)dx = \int_0^\infty P(Y > x)dx$. Secondly, $EY^- = \int_0^\infty P(-YI\{-Y \ge 0\} > x)dx = \int_0^\infty P(-Y > x)dx = \int_0^\infty P(Y \le -x)dx$. ∎

**Remark.** The above identities and inequalities show the close relationship between the existence of moments and the tail probabilities. The thinner the tails, the higher moments the r.v. will have.

## 4.6 Moments and Moment inequalities

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

(1) Define

$r$th Moment: $EX^r$

$r$th Absolute Moment: $E|X|^r$

$r$th Central Moment: $E(X - EX)^r$

$r$th Absolute Central Moment: $E|X - EX|^r$

(2) $L^r$ Spaces $= \{X : E|X|^r < \infty\}$.

Some very useful moment inequalities are given below.

### 4.6.1 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

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

**Proof**. **Try to give a direct proof**. However, one picture is worth a thousand words.

### 4.6.2 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

$$g(s) = s^{1/(p-1)} = s^{(1/p)/(1-1/p)} = s^{(1-1/q)/(1/q)} = s^{q-1},$$

$$\int_0^a h(t)dt = \frac{a^p}{p}, \qquad \int_0^b g(s)ds = \frac{b^q}{q}.$$

Therefore,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

(*One could try to prove this directly without resort to Young's inequality and use this as a starting point; see the Lemma below*).

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. ∎

**Remark.** Holder's inequality is one of the most important inequalities in analysis (including probability). For one thing, it can be used to derive a series of other important inequalities; see below. The inequality is very sharp and very flexible as one can choose $p$ (hence $q$) and $X$, $Y$ creatively to derive some powerful results.

Finally, let us prove the next inequality directly.

LEMMA **4.6.1** *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:

$$\frac{d}{da}g(a) = 0, \qquad \Longrightarrow a^{p-1} - b = 0, \qquad \Longrightarrow b = a^{p-1}.$$

A check on the second derivative will establish that this is indeed a minimum. The value of the function at the minimum is

p not q (here is the power)

$$\frac{1}{p}a^p + \frac{1}{q}(a^{p-1})^q - a(a^{p-1}) = \frac{1}{p}a^p + \frac{1}{q}a^q - a^p = 0.$$

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 $a^p = b^q$.

### 4.6.3 Cauchy-Schwarz inequality

$$E|XY| \leq \sqrt{[E|X|^2][E|Y|^2]}.$$

**Proof**. Take $p = q = 2$ in Holder's inequality. ∎

### 4.6.4 Lyapunov's inequality

(1). $E(|X|) \le E(|X|^p)^{1/p}$ for $p \ge 1$.

(2). $[E|Z|^r]^{1/r} \le [E|Z|^s]^{1/s}$, for $0 < r \le s < \infty$.

**Proof**. (1). Take $Y = 1$ in Holder's inequality.

(2). Take $X = Z^r$ in (1), and also let $rp = s$, we get

$$E|Z|^r \le [E|Z|^{rp}]^{1/p} = [E|Z|^{rp}]^{r/(rp)} = \{[E|Z|^s]^{1/s}\}^r.$$

(Alternative proof can be given by Jensen's inequality given below). ∎

### 4.6.5 Minkowski's inequality

Suppose $p \ge 1$, then

$$[E|X+Y|^p]^{1/p} \le [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}.$$

or more generally,

$$[E|X_1 + ... + X_n|^p]^{1/p} \le [E|X_1|^p]^{1/p} + ... + [E|X_n|^p]^{1/p}.$$

**Proof**. The proof for $p = 1$ is trivial. Now assume $p > 1$. In order to have $p^{-1} + q^{-1} = 1$, we need $q = (1 - p^{-1})^{-1} = p/(p-1)$. From Holder's inequality, we have

$$\begin{aligned} E|X+Y|^p &= E\{|X+Y|^{p-1}\ |X+Y|\} \\ &\le E\{|X+Y|^{p-1}\ |X|\} + E\{|X+Y|^{p-1}\ |Y|\} \\ &\le (E|X+Y|^{q(p-1)})^{1/q}(E|X|^p)^{1/p} + (E|X+Y|^{q(p-1)})^{1/q}(E|Y|^p)^{1/p} \\ &\le (E|X+Y|^p)^{1/q}(E|X|^p)^{1/p} + (E|X+Y|^p)^{1/q}(E|Y|^p)^{1/p} \\ &\le (E|X+Y|^p)^{1/q}\left[(E|X|^p)^{1/p} + (E|Y|^p)^{1/p}\right]. \end{aligned}$$

So $\{E|X+Y|^p\}^{1-1/q} = [E|X+Y|^p]^{1/p} \le (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$. ∎

### 4.6.6 Jensen's inequality

Let $\psi$ be convex, that is, for all $\lambda \in (0,1)$ and $x, y \in \mathcal{R}$, one has

$$\lambda\psi(x) + (1-\lambda)\psi(y) \ge \psi\left(\lambda x + (1-\lambda)y\right). \tag{6.15}$$

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

$$\psi(EX) \le E[\psi(X)].$$

**Proof**. Denote $\mu = EX$. Since $\psi$ is convex, then there exists $\lambda \in R$ (in nice cases, $\lambda = \psi'(\mu)$) such that for all $x$,

$$\psi(x) \ge \psi(\mu) + \lambda(x - \mu). \tag{6.16}$$

Put $x = X$ and take expectation, we get $E\psi(X) \ge \psi(\mu) + \lambda E(X - \mu) = \psi(\mu) = \psi(EX)$. ∎

REMARK **4.6.1** *Inequality (6.16) means that there always exists a line (i.e. tangent line on the LHS of (6.16)) which lies below the convex curve $\psi(x)$. The tangent line and the convex curve touches at $x = \mu$, and $\lambda$ is the slope of the tangent line, which is the gradient of $\psi(x)$ at $\mu$ if it exists.*

REMARK **4.6.2** *Convex functions have many interesting properties. For instance, they are always continuous functions, have both left and right derivatives. One can refer to many books devoted to this topic.*

**Appendix: proof of (6.16).** We shall prove the inequality in several steps. Here are some of the implications.

(1) For $x \le t \le y$,
$$\frac{\psi(t)-\psi(x)}{t-x} \le \frac{\psi(y)-\psi(x)}{y-x} \le \frac{\psi(y)-\psi(t)}{y-t}.$$
(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
$$\psi(t) = \psi\left(\frac{y-t}{y-x}x + \frac{t-x}{y-x}y\right) \le \psi(x) + \frac{t-x}{y-x}[\psi(y)-\psi(x)] = \frac{y-t}{y-x}\psi(x) + \frac{t-x}{y-x}\psi(y).$$
This is certainly true by taking $\lambda = \frac{y-t}{y-x}$ in (6.15).

(2) Since $t - h_1 \le t \le t + h_2$ for $h_1, h_2 > 0$, from (1) we get
$$\frac{\psi(t)-\psi(t-h_1)}{h_1} \le \frac{\psi(t+h_2)-\psi(t)}{h_2}.$$
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
$$\psi'(t-) \le \psi'(t+).$$

(3) Let $t \searrow x$ and $t \nearrow y$ in (1), we get
$$\frac{\psi(y)-\psi(x)}{y-x} \ge \psi'(x+), \qquad \frac{\psi(y)-\psi(x)}{y-x} \le \psi'(y-),$$
Changing $x$ into $t$ in the first and $y$ into $t$ in the second, we get
$$\frac{\psi(y)-\psi(t)}{y-t} \ge \psi'(t+), \qquad \frac{\psi(t)-\psi(x)}{t-x} \le \psi'(t-),$$
Using (2), we have
$$\frac{\psi(t)-\psi(x)}{t-x} \le \psi'(t-) \le \psi'(t+) \le \frac{\psi(y)-\psi(t)}{y-t}.$$

(4) For any fixed $t$, choose $a$ so that $\psi'(t-) \le a \le \phi'(t+)$, and let
$$l(z) = a(z-t) + \psi(t),$$
then $l(t) = \psi(t)$, and $\psi(z) \ge l(z)$.

*Proof.* $l(t) = \psi(t)$ is trivial. The second part is equivalent to $l(z) = a(z-t) + \psi(t) \le \psi(z)$ or
$$\psi(z) - \psi(t) \ge a(z-t).$$
This follows from (3). ( If $z = t$, trivial. If $z > t$, from (2), $[\psi(z)-\psi(t)]/(z-t) \ge \phi'(t+) \ge a$. If $z < t$, from (2), $[\psi(z)-\psi(t)]/(z-t) \le \phi'(t-) \le a$. ) ∎

### 4.6.7 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 $a > 0$:
$$P(|X| \ge a) \le \frac{Eg(X)}{g(a)}$$
**Proof**. $Eg(X) \ge Eg(X)I_{\{g(X)\ge g(a)\}} \ge g(a)EI_{\{g(X)\ge g(a)\}} = g(a)EI_{\{|X|\ge a\}} = g(a)P(|X| \ge a)$.

Some special cases:
$$\begin{aligned}
X \in L^1 &\implies P(|X| \ge a) \le \frac{E|X|}{a} \\
X \in L^p &\implies P(|X| \ge a) \le \frac{E|X|^p}{a^p} \\
X \in L^2 &\implies P(|X - EX| \ge a) \le \frac{Var(X)}{a^2} \\
Ee^{t|X|} < \infty, t \ge 0 &\implies P(|X| \ge a) \le \frac{Ee^{t|X|}}{e^{ta}}.
\end{aligned}$$

## 4.7 Exercises

1. Let $X \geq 0$ be a r.v. on $(\Omega, \mathcal{A}, P)$ and $0 < EX < \infty$. Then
$$\nu(A) = E_A X / EX$$
is a probability measure on $(\Omega, \mathcal{A})$.

2. Betteley (1977) provides an interesting addition law for expectations. Let $X, Y$ be two r.v.'s. Define
$$X \wedge Y = \min\{X, Y\}, \qquad and \qquad X \vee Y = \max\{X, Y\}.$$
Analogous to the probability law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, show that
$$E(X \vee Y) = EX + EY - E(X \wedge Y).$$

3. Let $C > 0$ be a constant. Then $E|X| < \infty$ iff $\sum_1^\infty P(|X| \geq Cn) < \infty$.

4. If $E|X| < \infty$, for any $\epsilon > 0$, there is a simple function $X_\epsilon$ such that $E|X - X_\epsilon| < \epsilon$.

5. For any $r > 0$, $E|X|^r < \infty$ iff $\sum_1^\infty n^{r-1} P(|X| \geq n) < \infty$.

6. (a). If $E|X|^r < \infty$ for $r > 0$, then $x^r P(|X| > x) = o(1)$ as $x \to \infty$,
(b). Give an example to show that the converse to (a) does not hold.
(c). A partial converse to (a) holds: if $x^r P(|X| > x) = o(1)$, then $E|X|^{r-\epsilon} < \infty$ for $0 < \epsilon < r$.
(d). (Optional.) Can we replace $x^r P(|X| > x) = o(1)$ in (c) by $x^r P(|X| > x) = O(1)$.

7. Assume that $EX^2 < \infty$ and $a$ is real. (i) Let
$$\begin{aligned} Y &= XI\{X \leq a\} + aI\{X > a\}, \\ Z &= XI\{X \geq b\} + bI\{X < b\}, \\ W &= XI\{b \leq X \leq a\} + aI\{X > a\} + bI\{X < b\}. \end{aligned}$$
Show that
$$Var(W) \leq \min\{Var(Y), Var(Z)\} \leq Var(X).$$
(Intuitively, truncation of a r.v. makes it less dispersive and hence has smaller variance.)

8. If $\mu = EX \geq 0$ and $0 \leq \lambda < 1$, then
$$P(X > \lambda\mu) \geq \frac{(1-\lambda)^2 \mu^2}{EX^2}.$$
Consequently, if $E|Y| = 1$, $P(|Y| > \lambda) \geq (1-\lambda)^2 / EY^2$. (This gives a lower bound complementing Chebyshev's inequality.)

9. Chebyshev's inequlaity states: $a^2 P(|X| \geq a) \leq EX^2$. Show that (i) it can be sharp for some fixed $a$ by giving an example; (ii) it can not be sharp for all $a$ since $\lim_{a\to\infty} a^2 P(|X| \geq a)/EX^2 \to 0$.

10. Let $f$ be measurable on $(\Omega, \mathcal{A}, \mu)$ and $A \in \mathcal{A}$. If $\mu(A) = 0$, then $\int_A f d\mu = 0$. (i.e., integration over a set $A$ of measure 0 is 0.) (Hint: use Mean Value Theorem.)

11. For any r.v. $Y$, show that $\int_{-\infty}^\infty P(Y = x)dx = 0$. (Hint: all jump points form a countable set which has L-measure 0.)

12. Let $X > 0$ a.s. Show:
$$(i) \qquad \lim_{y\to\infty} yE\left(\frac{1}{X} I\{X \geq y\}\right) \longrightarrow 0,$$
$$(ii) \qquad \lim_{y\to 0} yE\left(\frac{1}{X} I\{X \geq y\}\right) \longrightarrow 0.$$

13. Let $X, X_1, X_2, ......$ be i.i.d. r.v.'s with $EX^2 < \infty$. Show that, as $n \to \infty$, we have

(i) $nP(|X| > \epsilon\sqrt{n}) \to 0$;
(ii) $n^{-1/2}\max_{1\le k\le n}\{|X_k|\} \to 0$ in probability.

14. (i) If $X \ge 0$ and $E|X|^p < \infty$ for all $p > 0$, and

$$g(p) = \ln EX^p, \qquad 0 \le p < \infty,$$

then $g$ is convex on $[0, \infty)$.
(ii) Verify for $0 < a < b < d$ and any nonnegative r.v. $Y$ that

$$EY^b \le (EY^a)^{\frac{d-b}{d-a}} \left(EY^d\right)^{\frac{b-a}{d-a}}$$

(iii) Let

$$h(\alpha) := \sum_{n=1}^{\infty} c_n^{\alpha} EY_n^{\alpha}.$$

Utilize the above result to show that if $h(\alpha) < \infty$ for $\alpha = \alpha_1 > 0$ and $\alpha = \alpha_2 > 0$, where $\alpha_1 < \alpha_2$, then $h(\alpha) < \infty$ for all $\alpha \in [\alpha_1, \alpha_2]$.
(Hint: the following inequality might be useful: $a^{\lambda}b^{1-\lambda} \le (a+b)^{\lambda}(a+b)^{1-\lambda} = (a+b)$, where $a, b \ge 0$.)

15. If $X \ge 0$ and $Y \ge 0$, $p > 0$, then $E\{(X+Y)^p\} \le C_p\left(E\{X^p\} + E\{Y^p\}\right)$, where

$$\begin{aligned} C_p &= 2^{p-1}, \qquad \text{if } p > 1 \\ &= 1, \qquad \text{if } 0 \le p \le 1. \end{aligned}$$

**The following questions are optional.**

16. Show that an equivalent definition of integration for a nonnegative function $f \ge 0$ is

$$\int f d\mu := \sup\left\{\int \psi d\mu : \psi \in S_f\right\},$$

where $S_f$ = the collection of all nonnegative simple functions $\psi$ such that $\psi(\omega) \le f(\omega)$ for any $\omega \in \Omega$.

17. If $X_j \ge 0$, then

$$\begin{aligned} \left\{\left(\sum_{i=1}^n X_i\right)^p\right\} &\le \sum_{i=1}^n \{X_i^p\}, \qquad \text{if } p \le 1 \\ &\ge \sum_{i=1}^n \{X_i^p\}, \qquad \text{if } p \ge 1, \end{aligned}$$

Hence,

$$\begin{aligned} E\left\{\left(\sum_{i=1}^n X_i\right)^p\right\} &\le \sum_{i=1}^n E\{X_i^p\}, \qquad \text{if } p \le 1 \\ &\ge \sum_{i=1}^n E\{X_i^p\}, \qquad \text{if } p \ge 1. \end{aligned}$$

18. Suppose that $p \ge 1$. Show that

$$E\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i\right|^p\right\} \le \frac{1}{n}\sum_{i=1}^n E|X_i|^p, \tag{7.17}$$

$$E\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i\right|^p\right\} \le \left\{\frac{1}{n}\sum_{i=1}^n \{E|X_i|^p\}^{1/p}\right\}^p. \tag{7.18}$$

Compare these inequalities to see which one is better.

19. Establish that $g(t) = (\sin t)/t$ is Riemann but not Lebesgue integrable over $(-\infty, \infty)$ and find a function $h(t)$ wiich is Lebesgue integrable but Riemann integrable.

20. Discuss the relationships between the Riemann but not Lebesgue integrations.

# Chapter 5

# Independence

## 5.1 Definition

Independence is a fundamental new concept peculiar to the theory of probability.

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

(i) **Events** $A_1, ..., A_n \in \mathcal{A}$ are said to be **independent** iff

$$P\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} P\left(A_i\right).$$

for every subset $J$ of $\{1, 2, ..., n\}$.

(ii) **Classes** $\mathcal{A}_1, ..., \mathcal{A}_n$ are said to be **independent** iff

$$P\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} P\left(A_i\right).$$

for every subset $J$ of $\{1, 2, ..., n\}$, and $A_i \in \mathcal{A}_i$.

In particular, $\sigma$**-algebras** $\mathcal{A}_1, ..., \mathcal{A}_n$ are said to be **independent** iff

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i\right) \qquad \text{for any } A_i \in \mathcal{A}_i.$$

(Note we can choose some $A_i = \Omega \in \mathcal{A}_i$.)

(iii) The **r.v.'s** $X_1, ..., X_n$ are said to be **independent** iff the events $\{X_i \in B_i\}$ are independent, i.e.,

$$P\left(\bigcap_{i\in J} \{X_i \in B_i\}\right) = \prod_{i\in J} P\left(\{X_i \in B_i\}\right).$$

for every subset $J$ of $\{1, 2, ..., n\}$. This is clearly equivalent to

$$P\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} P\left(\{X_i \in B_i\}\right) \tag{1.1}$$

for any Borel sets $B_i \in \mathcal{B}$ (as one can take some $B_i = \mathcal{R}_i$).

(iv) The **r.v.'s** of an infinite (not necessarily countable) family are said to be **independent** iff those in every finite subfamily are.

(v) The r.v.'s of a family are said to be **pairwise independent** iff every two of them are independent.

(vi) The r.v.'s that are independent and have the same d.f. are called independent and identically distributed **(i.i.d.)**.

**Example.** 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

$$P(A_1 \cap A_2 \cap A_3) = P(X_1 = X_2 = X_3) = 1/4 \neq 1/8 = P(A_1)P(A_2)P(A_3).$$

## 5.2 How to check independence

In order to check if $X_1, ..., X_n$ are independent, one only needs to verify (1.1) for $B_i = (-\infty, t_i]$.

THEOREM **5.2.1** *The r.v.'s $X_1, ..., X_n$ are independent iff*

$$F_{X_1,...,X_n}(t_1, ..., t_n) = F_{X_1}(t_1)......F_{X_n}(t_n) \tag{2.2}$$

*for all $t_1, ..., t_n \in \mathcal{R}$.*

**Proof.** "$\Longrightarrow$". If $X_1, ..., X_n$ are independent, then taking $B_i = (-\infty, t_i]$ in (1.1) results in (2.2).

"$\Longleftarrow$". Assume now that (2.2) holds, we shall show that $X_1, ..., X_n$ are independent. As the first step, we shall show that

$$P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) = P\left(X_1 \in B\right) \prod_{j=2}^n P(X_j \le t_j), \qquad \forall B \in \mathcal{B} \tag{2.3}$$

To show this, define

$$\mathcal{B}_1 = \left\{ B \in \mathcal{B} : P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) = P\left(X_1 \in B\right) \prod_{j=2}^n P(X_j \le t_j),\ t_j \in \mathcal{R}, 2 \le j \le n \right\}.$$

It can be shown:

(i) the $\pi$-class: $\mathcal{A} = \{(-\infty, t] : \ t \in \mathcal{R}\}$ is contained in $\mathcal{B}_1$.

(ii) $\mathcal{B}_1$ is a $\lambda$-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 (2.3).

It remains to show (ii): $\mathcal{B}_1$ is a $\lambda$-class.

(i) $\mathcal{R} \in \mathcal{B}_1$ by (2.2).

(ii) If $A \subset B \in \mathcal{B}_1$, then for all $t_2, ..., t_n \in \mathcal{R}$,

$$\begin{aligned} & P\left(X_1 \in B - A, \cap_{j=2}^n \{X_j \le t_j\}\right) \\ = \; & P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) - P\left(X_1 \in A, \cap_{j=2}^n \{X_j \le t_j\}\right) \\ = \; & P\left(X_1 \in B\right) \prod_{j=2}^n P(X_j \le t_j) - P\left(X_1 \in A\right) \prod_{j=2}^n P(X_j \le t_j) \\ = \; & P\left(X_1 \in B - A\right) \prod_{j=2}^n P(X_j \le t_j), \end{aligned}$$

which confirms that $B - A \in \mathcal{B}_1$.

(iii) If $B_k \in \mathcal{B}_1$ and $B_k \nearrow B$, then for all $t_2, ..., t_n \in \mathcal{R}$,

$$\begin{aligned} P\left(X_1 \in B, \cap_{j=2}^n \{X_j \le t_j\}\right) &= \lim_{k \to \infty} P\left(X_1 \in B_k, \cap_{j=2}^n \{X_j \le t_j\}\right) \\ &= \lim_{k \to \infty} P\left(X_1 \in B_k\right) \prod_{j=2}^n P(X_j \le t_j) \\ &= P\left(X_1 \in B\right) \prod_{j=2}^n P(X_j \le t_j), \end{aligned}$$

which implies that $B \in \mathcal{B}_1$.

Combining (i)-(iii), we prove that $\mathcal{B}_1$ is a $\lambda$-class.

We continue this procedure iteratively until the same extension has been accomplished for $X_2, ..., X_n$. For instance, the second step is as follows. Define

$$\begin{aligned} \mathcal{B}_2 \;=\; & \Big\{B' \in \mathcal{B} : P\left(X_1 \in B, X_2 \in B', \cap_{j=3}^n \{X_j \le t_n\}\right) \\ & \left. = P\left(X_1 \in B\right) P\left(X_2 \in B'\right) \prod_{j=3}^n P(X_j \le t_j), \; B \in \mathcal{B}, \; t_j \in \mathcal{R}, 3 \le j \le n \right\}. \end{aligned}$$

Using the similar arguments to the above, we can show that $\mathcal{B}_2 = \sigma(\mathcal{A}) = \mathcal{B}$.

Using the same techniques in the last theorem, we can show the next theorem. (Chow, p62.)

THEOREM **5.2.2** *If $\mathcal{G}$ and $\mathcal{D}$ are independent classes of events, and $\mathcal{D}$ is a $\pi$-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 $\lambda$-class containing a $\pi$-class $\mathcal{D}$. By the Monotone Convergence Theorem, $\mathcal{D}^* \supset \sigma(\mathcal{D})$. This completes our proof.

Applying the above lemma, we get (Durrett, page 25.)

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

**Question:** Is the condition: "*each $\mathcal{A}_i$ is a $\pi$-class*" really necessary in the last theorem?

### 5.2.1 Discrete r.v.'s

THEOREM **5.2.4** *Discrete r.v.'s $X_1, ..., X_n$, taking values in countable set $C$, are independent iff*

$$P(X_1 = a_1, ..., X_n = a_n) = \prod_{i=1}^{n} P(X_i = a_i) \tag{2.4}$$

*for all $a_1, ..., a_n \in C$.*

*Proof.* If $X_1, ..., X_n$ are independent, then (2.4) is obviously true.

On the other hand, if (2.4) is true, then

$$\begin{aligned}
F_{X_1,...,X_n}(t_1, ..., t_n) &= P(X_1 \le t_1, ..., X_n \le t_n) \\
&= \sum_{\{a_1 \in C: a_1 \le t_1\}} ... \sum_{\{a_n \in C: a_n \le t_n\}} P(X_1 = a_1, ..., X_n = a_n) \\
&= \sum_{\{a_1 \in C: a_1 \le t_1\}} ... \sum_{\{a_n \in C: a_n \le t_n\}} P(X_1 = a_1) \cdots P(X_n = a_n) \\
&= \left( \sum_{\{a_1 \in C: a_1 \le t_1\}} P(X_1 = a_1) \right) \cdots \left( \sum_{\{a_n \in C: a_n \le t_n\}} P(X_n = a_n) \right) \\
&= P(X_1 \le t_1) \cdots P(X_n \le t_n) \\
&= F_{X_1}(t_1)......F_{X_n}(t_n).
\end{aligned}$$

Therefore, $X_1, ..., X_n$ are independent.

### 5.2.2 Absolutely continuous r.v.'s

THEOREM **5.2.5** *Let $X = (X_1, ..., X_n)$ be an absolutely continuous random vector. Then $X_1, ..., X_n$ are independent iff*

$$f_X(y_1, ..., y_n) = \prod_{i=1}^{n} f_{X_i}(y_i) \tag{2.5}$$

*for all $y_1, ..., y_n \in \mathcal{R}$.*

*Proof.* If $X_1, ..., X_n$ are independent, then

$$\begin{aligned}
\prod_{i=1}^{n} \int_{-\infty}^{t_i} f_{X_i}(y_i) dy_i &= \prod_{i=1}^{n} P(X_i \le t_i) = P(X_1 \le t_1, ..., X_n \le t_n) \\
&= \int_{-\infty}^{t_1} ... \int_{-\infty}^{t_n} f_X(y_1, ..., y_n) dy_1 ... dy_n.
\end{aligned}$$

Hence,

$$\int_{-\infty}^{t_1} ... \int_{-\infty}^{t_n} \left( f_X(y_1, ..., y_n) - \prod_{i=1}^{n} f_{X_i}(y_i) \right) dy_1 ... dy_n = 0.$$

Differentiating w.r.t. $t_1, ..., t_n$ results in (2.5).

On the other hand, if (2.5) is true, then

$$\begin{aligned}
P(X_1 \le t_1, ..., X_n \le t_n) &= \int_{-\infty}^{t_1} ... \int_{-\infty}^{t_n} f_X(y_1, ..., y_n) dy_1 ... dy_n \\
&= \prod_{i=1}^{n} \int_{-\infty}^{t_i} f_{X_i}(y_i) dy_i = \prod_{i=1}^{n} P(X_i \le t_i).
\end{aligned}$$

Therefore, $X_1, ..., X_n$ are independent.

## 5.3 Functions of independent r.v.'s

### 5.3.1 Transformation properties

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

*Proof.* For $B_i \in \mathcal{B}$, we have $g_i^{-1}(B_i) \in \mathcal{B}$. So

$$\begin{aligned} P\left(\bigcap_{i=1}^{n}\{g_i(X_i) \in B_i\}\right) &= P\left(\bigcap_{i=1}^{n}\{X_i \in g_i^{-1}(B_i)\}\right) = \prod_{i=1}^{n} P\left(X_i \in g_i^{-1}(B_i)\right) \\ &= \prod_{i=1}^{n} P\left(g_i(X_i) \in B_i\right). \end{aligned}$$

Similarly, we can show that

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

$$g_1(X_1, ..., X_{n_1}), \quad g_2(X_{n_1+1}, ..., X_{n_2}), \quad .........., \quad g_k(X_{n_{k-1}}, ..., X_{n_k})$$

*are independent.*

*Proof.* For simplicity, we shall only prove it for $k = 2$. Denote $Z_1 = (X_1, ..., X_m)$ and $Z_2 = (X_{m+1}, ..., X_n)$. 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

$$P\left(Z_1 \in A_1, Z_2 \in A_2\right) = P\left(Z_1 \in A_1\right) \; P\left(Z_2 \in A_2\right), \tag{3.6}$$

for all $A_1 \in \mathcal{B}^m$ and $A_2 \in \mathcal{B}^{n-m}$. To show this, define

$$\begin{aligned} \mathcal{B}_1 = \{ & A \in \mathcal{B}^m : P\left(Z_1 \in A, Z_2 \in B_1 \times ... \times B_{n-m}\right) = P\left(Z_1 \in A\right) P\left(Z_2 \in B_1 \times ... \times B_{n-m}\right), \\ & \text{for any } B_i \in \mathcal{B}, \; 1 \le i \le n-m\}. \end{aligned}$$

Similar to the proof of Theorem 5.2.1, it can be shown:

(i) the $\pi$-class: $\mathcal{A} = \{A_1 \times ... \times A_m : A_i \in \mathcal{B}, 1 \le i \le m\}$ is contained in $\mathcal{B}_1$.

(ii) $\mathcal{B}_1$ is a $\lambda$-class.

Then by the Monotone Convergence 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 (3.6).

### 5.3.2 Convolutions

THEOREM **5.3.3** *Let $X, Y$ be independent and absolutely continuous. Then $X+Y$ is absolutely continuous and*

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

*Proof.*

$$\begin{aligned} P(X+Y \le t) &= P\left((X,Y) \in \{(x,y) : x+y \le t\}\right) \\ &= \int\int_{x+y \le t} f_{X,Y}(x,y) dx dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{t-y} f_X(x) f_Y(y) dx dy \end{aligned}$$

$$
\begin{aligned}
&= \int_{y=-\infty}^{\infty}\left[\int_{x=-\infty}^{t-y} f_X(x)dx\right] f_Y(y)dy \\
&= \int_{y=-\infty}^{\infty}\left[\int_{z=-\infty}^{t} f_X(z-y)dz\right] f_Y(y)dy \qquad (z=x+y) \\
&= \int_{z=-\infty}^{t}\left[\int_{y=-\infty}^{\infty} f_X(z-y)f_Y(y)dy\right] dz.
\end{aligned}
$$

**Example.** 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 **5.3.4** *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.** If $X \sim Poisson(\lambda_1)$, $Y \sim Poisson(\lambda_2)$, and $X, Y$ are independent, then $X+Y \sim Poisson(\lambda_1 + \lambda_2)$.

### 5.3.3 An example

**Example**. (Durrett, page 23.)

(i) 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.

(ii) 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

$$
\begin{gathered}
P(A \cap B^c) = P(A \cap (\Omega - B)) = P(A - (A \cap B)) \\
= P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(B^c)
\end{gathered}
$$

(iii) Events $A_1, ..., A_n$ are independent iff $I_{A_1}, ..., I_{A_n}$ 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.

### 5.3.4 Correlation

**Definition.** The covariance of two r.v.'s $X$ and $Y$ is defined to be

$$Cov(X,Y) := E(X - EX)(Y - EY) = E(XY) - EXEY.$$

$X, Y$ are said to be positively correlated, uncorrelated, or negatively correlated iff $Cov(X,Y) > 0$, $= 0$ or $< 0$, respectively.

THEOREM **5.3.5** *If $X, Y$ are independent and integrable r.v.'s, then*

$$Cov(X,Y) = 0.$$

*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.
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

$$\begin{aligned} E(XY) &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j P(\{X = a_i\} \cap \{Y = b_j\}) \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j P(\{X = a_i\}) P(\{Y = b_j\}) = EXEY. \end{aligned}$$

Step 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

(i) $X_n$ and $Y_n$ are both simple r.v.'s.

(ii) $0 \leq X_n \nearrow X$ and $0 \leq Y_n \nearrow Y$, which in turn implies that

(a) $EX = \lim_n EX_n$, and $EY = \lim_n EY_n$, by the Monotone Convergence Theorem.
(b) $0 \leq X_n Y_n \nearrow XY$. (**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.)

(iii) $X_n$ and $Y_n$ are independent as Borel functions of $X$ and $Y$, respectively.

Now applying the Monotone Convergence Theorem, we get

$$E(XY) = \lim_n E(X_n Y_n) = \lim_n E(X_n) E(Y_n) = (EX)(EY).$$

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,

$$\begin{aligned} EXY &= E(X^+ - X^-)(Y^+ - Y^-) \\ &= E(X^+Y^+) - E(X^+Y^-) - (EX^-Y^+) + E(X^-Y^-) \\ &= EX^+EY^+ - E(X^+EY^- - EX^-EY^+ + EX^-EY^- \\ &= (EX^+ - EX^-)(EY^+ - EY^-) \\ &= EXEY. \quad \blacksquare \end{aligned}$$

In exactly the same manner, we can show that

THEOREM **5.3.6** *If $X_1, ..., X_n$ are independent and all have finite expectations, then*

$$E(X_1 \cdots X_n) = (EX_1) \cdots (EX_n). \quad \blacksquare$$

The following provides a covariance inequality.

THEOREM **5.3.7** *Let $u(x)$ and $v(x)$ be both non-decreasing or both non-increasing functions on $I = (a, b)$ (finite or infinite interval on $R$), and $P(X \in I) = 1$. Then,*

$$Eu(X)Ev(X) \le E[u(X)v(X)], \quad \text{or} \quad Cov\,(u(X), v(X)) \ge 0,$$

*provided these means exist.*

**Proof.** Let $X$ and $Y$ be independent and identically distributed r.v.'s. Clearly, we have $[u(X) - u(Y)][v(X) - v(Y)] \ge 0$. Hence,

$$\begin{aligned} 0 &\le E[u(X) - u(Y)][v(X) - v(Y)] \\ &= Eu(X)v(X) + Eu(Y)v(Y) - Eu(X)v(Y) - Eu(Y)v(X) \\ &= 2Eu(X)v(X) - 2Eu(X)v(Y) \\ &= 2Eu(X)v(X) - 2Eu(X)Ev(Y), \end{aligned}$$

from which we have $Eu(X)v(X) \ge Eu(X)Ev(Y) = Eu(X)Ev(X)$. $\blacksquare$

REMARK **5.3.1** *The theorem implies that $Cov(u(X), v(X)) \ge 0$, 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*

$$(E|X|^r)(E|X|^s) \le E|X|^{r+s}, \qquad r, s \ge 0.$$

For instance, we have

$$(E|X|)^2 \le EX^2, \qquad (E|X|)(EX^2) \le E|X|^3, \quad \text{etc.}$$

## 5.4 Borel-Cantelli Lemma and Kolmogorov $0-1$ Law

### 5.4.1 Borel-Cantelli Lemma

Let $\{A_n\}$'s be a sequence of events on $(\Omega, \mathcal{A}, P)$. Recall

$$\limsup_n A_n = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m = \lim_{n\to\infty} \cup_{m=n}^{\infty} A_m = \{A_n \ i.o.\}.$$

$$\liminf_n A_n = \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m = \lim_{n\to\infty} \cap_{m=n}^{\infty} A_m = \{A_n \ ult.\}.$$

$$\liminf_n A_n = \left(\limsup_n A_n^c\right)^c.$$

Theorem **5.4.1 (Borel-Cantelli Lemma)**

*(a). $P(A_n, i.o) = 0$ if $\sum_{n\to\infty} P(A_n) < \infty$.*
*(b). $P(A_n, i.o) = 1$ if $\sum_{n\to\infty} P(A_n) = \infty$ and $A_1, A_2, ...,$ are independent.*

*Proof.*
(a). $P(A_n, i.o) = P\left(\lim_n \cup_{m=n}^{\infty} A_m\right) = \lim_n P(\cup_{m=n}^{\infty} A_m) \le \lim_n \sum_{m=n}^{\infty} P(A_m) \to 0.$
(b). Noting $1 - x \le e^{-x}$ for all $x \in \mathcal{R}$ and independence of $A_n$, we have

$$\begin{aligned}
0 \ &\le \ 1 - P(A_n, i.o.) = P(A_n^c, ult.) = P(\lim_n \cap_{m=n}^{\infty} A_m^c) \\
&= \lim_{n\to\infty} P(\lim_{r\to\infty} \cap_{m=n}^{r} A_m^c) = \lim_{n\to\infty}\lim_{r\to\infty} P(\cap_{m=n}^{r} A_m^c) \\
&= \lim_{n\to\infty}\lim_{r\to\infty} \prod_{m=n}^{r} [1 - P(A_m)] \quad \text{(independence)} \\
&\le \lim_{n\to\infty}\lim_{r\to\infty} \prod_{m=n}^{r} e^{-P(A_m)} \quad (\text{as } 1 - x \le e^{-x}) \\
&= \lim_{n\to\infty}\lim_{r\to\infty} e^{-\sum_{m=n}^{r} P(A_m)} = \lim_{n\to\infty} e^{-\sum_{m=n}^{\infty} P(A_m)} \\
&= \lim_{n\to\infty} 0 = 0. \quad \blacksquare
\end{aligned}$$

Remark **5.4.1** *The inequality $e^x \ge 1 + x$, all $x \in R$, can be shown by different methods.*

**Method 1**. *$f(x) := e^x - (1+x)$, $f'(x_0) = 0 \Longrightarrow x_0 = 0$, $f''(x_0) = e^{x_0} > 0$, convex at $x_0$. So $f(x) \ge f(x_0) = 0$.*

**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.*

**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 $e^x \ge 1 + x^2 + ... + x^m/m!$ (for $x \ge 0$).* $\blacksquare$

Remark **5.4.2** *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 $\sum_n P(A_n) = \infty$ but $P(A_n, i.o) = P(A) \in (0, 1)$.*

Independence can be reduced to pairwise independence in part (b) of Borel-Cantelli Lemma.

Theorem **5.4.2** *$P(A_n, i.o) = 1$ if $\sum_{n=1}^{\infty} P(A_n) = \infty$ and $\{A_n, n \ge 1\}$ are (pairwise) independent.*

**Proof.** See Chung, K.L. page 76, Theorem 4.2.5. ∎

Corollary **5.4.1 (Borel 0-1 Law)** *Let $\{A_n\}$'s be (pairwise) independent. Then*

$$\begin{aligned} P(A_n, i.o) &= 0 \quad \text{if } \textstyle\sum_n P(A_n) < \infty \\ &= 1 \quad \text{if } \textstyle\sum_n P(A_n) = \infty. \end{aligned}$$

*i.e., when $A_n$ are independent, $P(\limsup A_n) = 0, or 1$ according to $\sum P(A_n) < or = \infty$.*

The above corollary quickly yields the next one:

Corollary **5.4.2 (Alternative form of Borel 0-1 Law)** *Let $\{A_n\}$'s be (pairwise) independent. Then*

$$\begin{aligned} (1).\quad P(A_n, i.o.) &= 0 \quad \Longleftrightarrow \quad \sum_n P(A_n) < \infty \\ (2).\quad P(A_n, i.o.) &= 1 \quad \Longleftrightarrow \quad \sum_n P(A_n) = \infty. \end{aligned}$$

**Remarks.**

(i) Borel 0-1 Law is a special case of the more general 0-1 laws for tail $\sigma$-algebras, see the next section.

(ii) 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 **5.4.3** *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 **5.4.4** *Let $X_n$ be (pairwise) independent. Then*

$$X_n \to 0 \ a.s. \iff \textstyle\sum_n P(|X_n| \geq \epsilon) < \infty, \ \forall \epsilon > 0.$$

**(That is, convergence in probability fast enough implies convergence almost sure.)**

**Proof.**

$$\begin{aligned}
& X_n \to 0 \qquad a.s. \\
\iff\quad & P(\lim_n X_n = 0) = 1. \\
\iff\quad & P(\cap_{m=1}^{\infty} \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < 1/m\}) = 1, \\
\iff\quad & \forall m \geq 1: \ P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < 1/m\}) = 1, \\
& \text{(as } P(\cap_{m=1}^{\infty} B_m) = 1 \Longrightarrow \forall m \geq 1\text{: } P(B_m) = 1\text{, see the proof given below.)} \\
\iff\quad & \forall \epsilon > 0: \ P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{|X_k| < \epsilon\}) = 1, \\
\iff\quad & \forall \epsilon > 0: \ P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} \{|X_k| \geq \epsilon\}) = 0, \\
\iff\quad & P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k) = 0, \ \text{where} \ \ A_k = \{|X_k| \geq \epsilon\}. \\
\iff\quad & P(A_n, i.o.) = 0. \\
\iff\quad & \sum_n P(A_n) = \sum_n P(|X_n| \geq \epsilon) < \infty. \\
& \text{(By Borel 0-1 law for (pairwise) independent events).} \ ∎
\end{aligned}$$

**Remark.** In the above proof, we have used the fact: $P(\cap_{m=1}^{\infty} B_m) = 1 \Longrightarrow \forall j \geq 1$: $P(B_j) = 1$.

**Proof.** $\forall j \geq 1$, we have $0 = P(\cup_{m=1}^{\infty} B_m^c) \geq P(B_j^c) = 1 - P(B_j)$. So $P(B_j) = 1$. ∎

An immediate consequence is

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

(*i*) $E|X| < \infty \iff X_n = o(n)$ *a.s.*

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

**Proof.**

(*i*) $E|X| < \infty \iff \sum_n P(|X_n|/n \geq \epsilon) = \sum_n P(|X| \geq \epsilon n) < \infty, \forall \epsilon > 0. \iff |X_n|/n \to 0$ a.s. (from Corollary 5.4.4)

(ii) It follows from (i). ∎

**Example.** Chow and Teicher, p61.

### 5.4.2 Kolmogorov 0-1 laws

It is a remarkable fact that probabilities of sets of certain class defined in terms of independent r.v.'s can only be 0 or 1.

**Definition:** The **tail $\sigma$-algebra** (or **remote future**) of a sequence $\{X_n, n \geq 1\}$ of r.v.'s on $(\Omega, \mathcal{A}, P)$ is

$$\bigcap_{n=1}^{\infty} \sigma(X_j, j \geq n) \equiv \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, ......),$$

The sets of the tail $\sigma$-algebra are called **tail events**, and functions measurable relative to the tail $\sigma$-algebra are dubbed **tail functions**.

**Remarks.**

(a). Recall $\sigma(X_j, j \geq n) \equiv \sigma(X_n, X_{n+1}, ......) =$ the future after time $n =$ the smallest $\sigma$-algebra w.r.t. which all $X_m, m \geq n$ are measurable.

(b). Intuitively, $A$ is a tail event if and only if changing a finite number of values does not affect the occurrence of the event.

(c). $A$ is a tail event if the event depends entirely on the "tail series".

**Examples of tail events.**

(i) If $B_n \in \mathcal{B}$, then $\{X_n \in B_n \ i.o.\}$ is a tail event.

**Proof.**

$$\{X_n \in B_n \ i.o.\} = \limsup_n \{X_n \in B_n\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{X_m \in B_m\} \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, ......).$$

(ii) $\{A_n \ i.o.\}$ is a tail event.

**Proof.** Taking $X_n = I_{A_n}$, $B_n = \{1\}$ in (i), we get

$$\{A_n \ i.o.\} = \{I_{A_n} = 1 \ i.o.\} = \{X_n \in B_n \ i.o.\}$$

Then apply (i).

(iii) Let $S_n = X_1 + \cdots + X_n$. It is easy to check that (Durrett, p62.)

(a) $\{\lim_n S_n \ exists\}$ 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.** (a)

$$\begin{aligned} \{\lim_n S_n \ exists\} &= \left\{ \sum_{n=1}^{\infty} X_n \ converges \right\} = \bigcap_{n=1}^{\infty} \left\{ \sum_{m=n}^{\infty} X_m \ converges \right\} \\ &\in \bigcap_{n=1}^{\infty} \sigma\left(X_n, X_{n+1}, ......\right). \end{aligned}$$

(b) $\{\limsup_n S_n > x\}$ 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 $\lim_n S_n = 0 < 1$.

(c) Since $C_n \to \infty$, $\forall m \geq 1$, we have

$$\limsup_n S_n/C_n = \lim_n S_m/C_n + \limsup_n (S_n - S_m)/C_n = \limsup_n (S_n - S_m)/C_n.$$

Hence,

$$\begin{aligned} \{\limsup_n S_n/C_n > x\} &= \bigcap_{m=1}^{\infty} \left\{ \limsup_n [(X_{m+1} + ... + X_n)]/C_n > x \right\} \\ &\in \bigcap_{m=1}^{\infty} \sigma\left(X_m, X_{m+1}, ......\right). \end{aligned}$$

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

**Proof.** (The proof is from Chow and Teicher, p64.) Since $\{X_n, n \geq 1\}$ are independent,

$\Longrightarrow$ $\forall n \geq 1$: $\sigma(X_i, 1 \leq i \leq n)$ and $\sigma(X_j, j > n)$ are independent.

$\Longrightarrow$ $\forall n \geq 1$: $\sigma(X_i, 1 \leq i \leq n)$ and $\cap_{m=0}^{\infty} \sigma(X_j, j > m) =: \mathcal{D}$ are independent,

(as $\mathcal{D} \subset \sigma(X_j, j > n)$.)

$\Longrightarrow$ $\mathcal{A} =: \cup_{n=1}^{\infty} \sigma(X_i, 1 \leq i \leq n)$ and $\mathcal{D}$ are independent.

[as for any $A \in \mathcal{A}$,

?????????????? is an algebra (but not necessarily a $\sigma$-algebra), hence a $\pi$-class.]

$\Longrightarrow$ By Theorem 5.2.2, $\sigma(\mathcal{A})$ and $\mathcal{D}$ are independent.

[as $\mathcal{A}$ is an algebra (but not necessarily a $\sigma$-algebra), hence a $\pi$-class.]

$\Longrightarrow$ $\mathcal{D}$ and $\mathcal{D}$ are independent since $\mathcal{D} \subset \sigma(X_n, n \geq 1) = \sigma(\mathcal{A})$. (**Why?**)

$\Longrightarrow$ $\forall B \in \mathcal{D}$, we have $P(B) = P(B \cap B) = P^2(B)$, implying $P(B) = 1$ or 0. ∎

**Remark.** When $\mathcal{D}$ is independent of itself, then we must have $\mathcal{D} = \{\emptyset, \Omega\}$.

COROLLARY **5.4.6** *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 $P(Y = c_0) = 1$ by definition.

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

**Proof.** For each $n \geq k \geq 1$, $X_n$ is $\sigma(X_j, j \geq k)$-measurable, and $Y_k =: \sup_{n \geq k} X_n$ is $\sigma(X_j, j \geq k)$-measurable. Hence, $Y_n$ is $\sigma(X_j, j \geq n)$-measurable and hence $\sigma(X_j, j \geq k)$-measurable for every $n \geq k \geq 1$ [since $\sigma(X_j, j \geq n) \subset \sigma(X_j, j \geq k)$]. This implies that $\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{j\geq n} X_j = \lim_{n\to\infty} Y_n$ is $\sigma(X_j, j \geq k)$-measurable for every $k \geq 1$. Thus, $\limsup_{n\to\infty} X_n$ is $\cap_{k=1}^{\infty}\sigma(X_j, j \geq k)$-measurable, i.e., a tail function, and similarly for $\liminf_{n\to\infty} X_n$. The proof follows from the last corollary.

**Remarks:** Some comparisons about Kolmogorov 0-1 law and Borel-Cantelli Lemma.

1. The Kolmogorov 0-1 law confines the value of $P(A_n, i.o.)$ to 0 or 1 (since $\{A_n, i.o.\}$ 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.

## 5.5 Exercises

1. (i) If $X \sim N(0,1)$, $Y \sim N(0,1)$, and $X, Y$ are independent, then $X + Y \sim N(0,2)$.
   (ii) If $X \sim Poisson(\lambda_1)$, $Y \sim Poisson(\lambda_2)$, and $X, Y$ are independent, then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$.

2. If $\{E_j, 1 \le j < \infty\}$ are independent events on $(\Omega, \mathcal{A}, P)$, then

$$P\left(\bigcap_{j=1}^{\infty} E_j\right) = \prod_{j=1}^{\infty} P(E_j),$$

   where the infinite product is defined to be the obvious limit; similarly,

$$P\left(\bigcup_{j=1}^{\infty} E_j\right) = 1 - \prod_{j=1}^{\infty} (1 - P(E_j)).$$

3. Let $\{X_j, 1 \le j \le n\}$ be independent with d.f.'s $\{F_j, 1 \le j \le n\}$. Find the d.f. of $\max_j X_j$ and $\min_j X_j$.

4. If $X$ and $Y$ are independent and $EX$ exists, then for any Borel set $B$, we have

$$\int_{Y \in B} X dP = (EX)P(Y \in B).$$

5. Let $\{X, X_n, n \ge 1\}$ be i.i.d. Show that

$$\overline{\lim}_{n\to\infty} |X_n|/n \le C \ a.s. \ (C > 0) \iff E|X| < \infty.$$

6. Let $\{X, X_n, n \ge 1\}$ be i.i.d. with d.f. given by

$$\begin{aligned} F(x) &= \frac{1}{2}\exp(-|x|^5/\pi), & x \le 0 \\ & \quad 1 - \frac{1}{2}\exp(-|x|^5/\pi), & x \ge 0 \end{aligned}$$

   Show that

$$\limsup_{n\to\infty} \frac{X_n}{(\pi \ln n)^{1/5}} = 1 \ a.s.$$

   (Hint: Show $P(X_n/(\pi \ln n)^{1/5} \le 1 + \epsilon, i.o.) = 1$, $P(X_n/(\pi \ln n)^{1/5} \ge 1 - \epsilon, i.o.) = 1$.)

7. Let $\{X, X_n, n \ge 1\}$ be i.i.d. with d.f. given by

$$F(x) = 1 - x^{-5}, \qquad x \ge 1.$$

   Show that

$$\limsup_{n\to\infty} \frac{\ln X_n}{\ln n} = c \qquad a.s.$$

   for some number $c$ and find $c$.

8. Let $A_n$ be a sequence of independent events with $P(A_n) < 1$ for all $n$. Show that $P(\cup_{n=1}^{\infty} A_n) = 1$ implies $P(A_n, i.o.) = 1$.

9. If $X_n$ is any sequence of r.v.'s, there are constants $c_n \to \infty$ so that $X_n/c_n \to 0$ a.s.

10. Let $\{X, X_n, n \ge 1\}$ be i.i.d. Show that $P\left(\overline{\lim}_{n\to\infty} X_n = \infty\right) = 1 \iff X$ is unbounded above, i.e., $P(X < C) < 1$, all $C < \infty$.

11. Let $\{X, X_n, n \ge 1\}$ be i.i.d. r.v.'s. Let $\alpha, \beta > 0$, $\alpha\beta > 1$. Assume that $E|X|^\alpha < \infty$. Then, $P(\max_{j=1}^{n} |X_j| > \epsilon n^\beta) \to 0$ as $n \to \infty$ for any $\epsilon > 0$.
    (i.e., $\max_{1 \le j \le n} |X_j| = o(n^\beta)$ in probability, whose definition will be given later in the course.)

12. (Optional.) Let $X_1$ and $X_2$ be two independent r.v.'s.

    (a). Show that: for all large $\lambda$,

$$P\left(|X_1| > \lambda\right) \leq 2P\left(|X_1| > \lambda, |X_2| < \lambda/2\right) \leq 2P\left(|X_1 + X_2| > \lambda\right).$$

    (b). Use (a) to show: if $X_1 + X_2 \in L_r$ for some $r \in (0, \infty)$, then $X_i \in L_r$, $i = 1, 2$.
    (Hint: $E|Y|^r = \int_0^\infty r t^{r-1} P(|Y| > t) dt$.)

13. (Optional.) If $X$ and $Y$ are independent, $E(|X|^p) < \infty$ for some $p \geq 1$ and $EY = 0$, then $E(|X|^p) \leq E(|X + Y|^p)$.

# Chapter 6

# Convergence Concepts

## 6.1 Modes of convergence

**Definition:** Let $X, X_1, X_2, ...$ be r.v.'s on $(\Omega, \mathcal{A}, P)$. We say that

(a) $X_n \to X$ a.s. (i.e., with probability 1) if

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

(b) $X_n \to X$ in $r$th mean, or in $L_r$ space, where $r > 0$, if

$$\lim_{n\to\infty} E|X_n - X|^r = 0.$$

(c) $X_n \to X$ in prob, written as $X_n \to_p X$, if

$$\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0, \qquad \text{for all } \epsilon > 0.$$

(d) $X_n \to X$ in distribution, written as $X_n \to_d X$ or $F_{X_n} \Longrightarrow F_X$, if

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

(All discontinuity points of $F_X$ has Lesbegue measure 0, hence convergence at continuity points will be enough to identify the limit.)

**Remarks:**

(a) 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.

(b) A **metric space** $\{X, \rho\}$ is a nonempty set $X$ of elements together with a real-valued function $\rho$ defined on $X \times X$ such that for all $x, y$, and $z \in X$:

(i) $\rho(x, y) \geq 0$, and $\rho(x, y) = 0$ iff $x = y$; (nonnegativity)
(ii) $\rho(x, y) = \rho(y, x)$; (symmetry)
(iii) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$; (triangle inequality).

The function $\rho$ is called a **metric**.

(b) $\rho(Y, Z) = \|Y - Z\|_r$ is a metric in $L_r = \{X : E|X|^r < \infty\}$, where

$$\begin{aligned} \|X\|_r &= E|X|^r, \qquad & 0 < r < 1, \\ & \quad (E|X|^r)^{1/r}, \qquad & r \geq 1. \end{aligned}$$

**Proof.** Nonnegativity and symmetry are easy to check. It remains to show the triangle inequality.

If $r \geq 1$, Minkowski's inequality implies $\|X+Y\|_r \leq \|X\|_r + \|Y\|_r$.

If $0 < r < 1$, then for any $0 \leq \lambda \leq 1$, we have $\lambda^r + (1-\lambda)^r \geq \lambda + (1-\lambda) = 1$. 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 $\|X_n - X\|_r \to 0$.

(c) For convergence in probability, we can define a measure of "pseudo-distance" between $X$ and $Y$ by

$$\rho_\epsilon(X,Y) = P\left(|X-Y| > \epsilon\right) = \int_E dP, \quad \text{where } E = \{\omega \in \Omega : |X(\omega) - Y(\omega)| > \epsilon\}.$$

Note that $\rho_\epsilon$ is NOT a metric since for fixed $\epsilon > 0$, $\rho_\epsilon(X,Y) = 0 \not\Longrightarrow X = Y$ a.s.

(d) **Levy metric.** For two d.f.'s $F, G$, let

$$\rho(F,G) = \inf\{\delta > 0 : F(x-\delta) - \delta \leq G(x) \leq F(x+\delta) + \delta, \text{ for all } x \in R\}.$$

Then it can be shown that (see Grimmit and Stirzaker, 1992)

(i) $\rho$ is a metric on the space of d.f.'s. (p.274)

(ii) Convergence in distribution is equivalent to convergence w.r.t. Levy metric. (p.285)

(e). 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. 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 $\omega$. In fact, $X_n$ and $X$ may not be defined on the same probability space.

### 6.1.1 Equivalent definition of a.s. convergence

We shall give several equivalent definitions of a.s. convergence which will prove useful in the next chapter on Laws of Large Numbers.

Theorem **6.1.1** *The following statements are equivalent:*

(a) $X_n \to X$ *a.s.;*

(b) $\forall \epsilon > 0: \ \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty}\{|X_m - X| < \epsilon\}\right) = 1$*;*

(c) $\forall \epsilon > 0: \ \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty}\{|X_m - X| \geq \epsilon\}\right) = 0$*;*

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

$$\sup_{m=n}^{\infty} |X_m - X| \longrightarrow_p 0.$$

(e) $\forall \epsilon > 0: \ P(|X_n - X| \geq \epsilon, i.o.) = 0$*;*

*Furthermore, "$\geq \epsilon$" and "$< \epsilon$" above may be replaced by "$> \epsilon$" and "$\leq \epsilon$", respectively.*

**Proof.** "$(a) \Longleftrightarrow (b)$". First note that

$$\begin{aligned} &\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} \\ = \ &\bigcap_{\epsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \epsilon\} \end{aligned}$$

$$\begin{aligned}
& \text{(i.e., } \forall \epsilon > 0, \exists n \geq 1, \text{ s.t. } |X_m(\omega) - X(\omega)| < \epsilon \ \forall m \geq n \text{ )} \\
= & \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{ \omega : |X_m(\omega) - X(\omega)| < \frac{1}{k} \right\} \\
= & \bigcap_{k=1}^{\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{ \omega : |X_m(\omega) - X(\omega)| < \frac{1}{k} \right\} \qquad (1.1) \\
= & \lim_{k\to\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{ \omega : |X_m(\omega) - X(\omega)| < \frac{1}{k} \right\} \\
= & \lim_{k\to\infty} \lim_{n\to\infty} \bigcap_{m=n}^{\infty} \left\{ |X_m - X| < \frac{1}{k} \right\}
\end{aligned}$$

(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 (1.1) 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.

"$(b) \Longleftrightarrow (c)$". Trivial.

"$(c) \Longleftrightarrow (e)$". Trivial since

$$\begin{aligned}
P(|X_n - X| \geq \epsilon, i.o.) &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) \\
&= P\left(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right) \\
&= \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\right).
\end{aligned}$$

"$(c) \Longleftrightarrow (d)$". $\forall \epsilon > 0$, let

$$A_{n,\epsilon} = \{\omega \in \Omega : \bigcup_{m=n}^{\infty} \{|X_m - X| \geq \epsilon\}\}, \qquad B_{n,\epsilon} = \{\omega \in \Omega : \sup_{m\geq n} |X_m - X| \geq \epsilon\}$$

Then the equivalence of (c) and (d) follows from the fact: $A_{n,\epsilon} \subset B_{n,\epsilon} \subset A_{n,\epsilon/2}$. (Why?) ∎

## 6.2 Cauchy Criterion

When no limit is specified, the next theorem is useful.

Theorem **6.2.1 (Cauchy Criterion of a.s.)**

$$\begin{aligned}
& X_n \textit{ converges a.s.} \\
\Longleftrightarrow \quad & \forall \epsilon > 0 : \lim_{n\to\infty} P(|X_m - X_{m'}| \leq \epsilon, \textit{all } m > m' \geq n) = 0.
\end{aligned}$$

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

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

$$\iff \quad \sup_{m,n\geq M} |X_m - X_n| \longrightarrow_p 0.$$

Proof. Left as an exercise. ∎

## 6.3 Relationships between modes of convergence

Theorem **6.3.1**

(1) *For $r \geq 1$,*

$$\begin{array}{c} X_n \to X \ a.s. \\ \\ X_n \to X \ in \ L_r \end{array} \implies X_n \to_p X \implies X_n \to_d X$$

(2) *If $r > s > 0$, then $X_n \to X$ in $L_r \implies X_n \to X$ 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 $0 \leq P(|X_n - X| \geq \epsilon) \leq P\left(\bigcup_{m=n}^{\infty}\{|X_m - X| \geq \epsilon\}\right) \to 0$ as $n \to \infty$ by Theorem 6.1.1(c). Hence, $X_n \to_p X$.

**The converse may not hold:** Let

$$P(X_n = 0) = 1 - n^{-1}, \qquad P(X_n = 1) = n^{-1},$$

and $X_n$'s are independent. Then $X_n \to_p 0$ since $P(|X_n - 0| > \epsilon) \leq n^{-1} \to 0$. However, $X_n \not\to 0$ a.s. since for any $0 < \epsilon < 1$, we have

$$\begin{aligned} P\left(\bigcap_{m\geq n}\{|X_m - 0| \leq \epsilon\}\right) &= P\left(\lim_{r\to\infty}\bigcap_{m=n}^{r}\{|X_m| \leq \epsilon\}\right) = \lim_{r\to\infty} P\left(\bigcap_{m=n}^{r}\{|X_m| \leq \epsilon\}\right) \\ &= \lim_{r\to\infty}\prod_{m=n}^{r} P(|X_m| \leq \epsilon) = \lim_{r\to\infty}\prod_{m=n}^{r}\left(1 - m^{-1}\right) \\ &= \lim_{r\to\infty}\frac{n-1}{n}\frac{n}{n+1}\ldots\frac{r-1}{r} = \lim_{r\to\infty}\frac{n-1}{r} = 0. \end{aligned}$$

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 $EX_n = 1 \not\to 0$.

(c) If $X_n \to_p X$, then $X_n \to_d X$.

*Proof.* Denote $F_n(x) = P(X_n \le x)$ and $F(x) = P(X \le x)$. First we have

$$\begin{aligned}
F_n(x) &= P(X_n \le x, |X_n - X| \le \epsilon) + P(X_n \le x, |X_n - X| > \epsilon) \\
&\le P(X \le x - (X_n - X), |X_n - X| \le \epsilon) + P(|X_n - X| > \epsilon) \\
&\le P(X \le x + \epsilon) + P(|X_n - X| > \epsilon) \\
&= F(x + \epsilon) + P(|X_n - X| > \epsilon).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
F_n(x) &= 1 - P(X_n > x) \\
&= 1 - P(X_n > x, |X_n - X| \le \epsilon) - P(X_n > x, |X_n - X| > \epsilon) \\
&\ge 1 - P(X > x - (X_n - X), |X_n - X| \le \epsilon) - P(|X_n - X| > \epsilon) \\
&\ge 1 - P(X > x - \epsilon) - P(|X_n - X| > \epsilon) \\
&= F(x - \epsilon) - P(|X_n - X| > \epsilon).
\end{aligned}$$

Combining the two, we have

$$F(x - \epsilon) - P(|X_n - X| > \epsilon) \le F_n(x) \le F(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Letting $n \to \infty$, we obtain

$$F(x - \epsilon) \le \liminf_n F_n(x) \le \limsup_n F_n(x) \le F(x + \epsilon).$$

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 $P(|X_n - X| \ge \epsilon) = P(2|X| \ge \epsilon) \not\to 0$.

(2). Trivial since $0 \le (E|X_n - X|^s)^{1/s} \le (E|X_n - X|^r)^{1/r} \to 0$ 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 $E|X_n - 0| = 1/n \to 0$, but $X_n \not\to 0$ in $L^2$ as $E|X_n - 0|^2 = 1 \not\to 0$.

(3). We now show that "a.s. convergence" and "mean convergence" do not imply each other.

**Example (a):** Let $P(X_n = 0) = 1 - n^{-2}$ and $P(X_n = n^3) = n^{-2}$. Then $X_n \to 0$ a.s. but $X_n \not\to 0$ in $L^1$.
**Proof.** $X_n \to 0$ a.s. since $\sum_{n=1}^{\infty} P(|X_n - 0| \ge \epsilon) = \sum_{n=1}^{\infty} n^{-2} < \infty$, by a criterion presented later.
However, $X_n \not\to 0$ in $L^1$ as $E|X_n - 0| = n \to \infty$.

**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 $X_n \not\to 0$ a.s.

## 6.4 Partial converses

We now provide some partial converses to results in the last section, i.e., converse results with some additional assumptions.

### 6.4.1 Convergence in probability and distribution to constants are equivalent

THEOREM **6.4.1** *$X_n \to_d C \iff X_n \to_p C$, where $C$ is a constant.*

*Proof.* We only need to show the part "$\Longrightarrow$". Given $\epsilon > 0$, we have

$$\begin{aligned} P(|X_n - C| > \epsilon) &= P(X_n > C + \epsilon) + P(X_n < C - \epsilon) \\ &\le P(X_n > C + \epsilon) + P(X_n \le C - \epsilon) \\ &= [1 - F_{X_n}(C + \epsilon)] + F_{X_n}(C - \epsilon) \\ &\to [1 - F_C(C + \epsilon)] + F_C(C - \epsilon) = 1 - 1 + 0 \\ &= 0. \end{aligned}$$

where we used that $C \pm \epsilon$ are continuity points of $F_C(x)$, which is degenerate at $C$. ∎

### 6.4.2 Dominated convergence in probability implies convergence in mean

We first prove two useful lemmas, which are also interesting in their own right.

LEMMA **6.4.1** *If $X_n \to_p X$, $|X_n| \le Y$ a.s. (i.e. $P(|X_n| \le Y) = 1$ ) for all $n$, then $|X| \le Y$ 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 **6.4.2** *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

$$\begin{aligned} E_{A_n}|Y| &= E[|Y|I_{\{|Y|>A_\epsilon\}}I_{A_n}] + E[|Y|I_{\{|Y|\le A_\epsilon\}}I_{A_n}] \\ &\le E[|Y|I_{\{|Y|>A_\epsilon\}}] + A_\epsilon E I_{A_n} \\ &\le \epsilon + A_\epsilon P(A_n). \end{aligned}$$

Since $P(A_n) \to 0$, $n \to \infty$, the RHS $\le 2\epsilon$ for sufficiently large $n$. ∎

We are ready to state the main result of this section.

THEOREM **6.4.2 (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.

**Method 1: via u.i.**. It is special case of Theorem 6.5.5. ∎

**Method 2: direct approach**.

From Lemma 6.4.1, 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. $E[Y^r I_{\{2Y>A_\epsilon\}}] \le \epsilon$. We thus have

$$\begin{aligned} E|X_n - X|^r &= E|X_n - X|^r I_{\{|X_n-X|>A_\epsilon\}} + E|X_n - X|^r I_{\{|X_n-X|\le\epsilon\}} \\ &\quad + E|X_n - X|^r I_{\{\epsilon<|X_n-X|\le A_\epsilon\}} \\ &\le E(2Y)^r I_{\{2Y>A_\epsilon\}} + \epsilon^r + A_\epsilon^r P(|X_n - X| > \epsilon) \\ &\le 2^r\epsilon + \epsilon^r + A_\epsilon^r P(|X_n - X| > \epsilon). \end{aligned}$$

Since $P(|X_n - X| > \epsilon) \to 0$, $n \to \infty$, the RHS $\le 2^r\epsilon + 2\epsilon^r$ for sufficiently large $n$. ∎

**Method 3**. From Lemma 6.4.1, we have $|X_n - X| \le 2Y$ a.s. So

$$\begin{aligned} E|X_n - X|^r &= E|X_n - X|^r I_{\{|X_n - X|>\epsilon\}} + E|X_n - X|^r I_{\{|X_n - X|\le\epsilon\}} \\ &\le 2^r EY^r I_{\{|X_n - X|>\epsilon\}} + \epsilon^r. \end{aligned}$$

From Lemma 6.4.2, the first term on RHS, $EY^r I_{\{|X_n - X|>\epsilon\}} \to 0$ for sufficiently large $n$. ∎

**Remarks:**

(i). 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 **6.4.3 (Dominated Convergence Theorem)** *If $X_n \to X$ a.s. and $P(|X_n| \le Y) = 1$ for all $n$, and $EY < \infty$, then $EX_n \to EX$, namely, $X_n \to X$ in $L^1$.*

(ii). Similar to Remark (i), the a.s. convergence in the well-known **Fatou's Lemma** can also be weakened to just convergence in probability, or even in distribution. See Section 6.8 or Durrett, Chapter 2, p48.

> THEOREM **6.4.4 (Another Fatou's Lemma)** *If $X_n \ge 0$ and $X_n \to_p X$, then $EX = E\liminf_n X_n \le \liminf_n EX_n$.*

(iii). The proof involves some interesting **truncation techniques**. There are several truncations working at different levels.

(iv). 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.

(iv). As indicated from the proof of Method 1, Dominated Moment Converge is a special case of uniform integrability.

COROLLARY **6.4.1 (Bounded convergence in probability implies mean convergence)** *If $X_n \to_p X$ and $P(|X_n| \le C) = 1$ for all $n$ and some $C$, then $X_n \to X$ in $L_r$ for all $r > 0$.* ∎

The following corollary is immediate.

COROLLARY **6.4.2** *Suppose that $|X_n| \le C$ a.s. for all $n$ and some $C$. Then for $r > 0$,*

$$X_n \to X \text{ in } L_r \quad \Longleftrightarrow \quad X_n \to_p X. \; ∎$$

One simple application is as follows:

COROLLARY **6.4.3** *$X_n \to_p 0$ 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

$$P(|X_n| \ge \epsilon) = P\left(\frac{|X_n|}{1+|X_n|} \ge \frac{\epsilon}{1+\epsilon}\right). \; ∎$$

### 6.4.3 Dominated convergence a.s. implies convergence in mean

THEOREM **6.4.5** *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 $L_r$.*

*Proof.* This follows from the last theorem. ∎

This is simply the **Dominated Convergence Theorem**; see Remark 1 in the last section.

### 6.4.4 Convergence in probability sufficiently fast implies a.s. convergence

THEOREM **6.4.6** *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 $P\left(\bigcup_{m=n}^{\infty}\{|X_m - X| > \epsilon\}\right) \le \sum_{m=n}^{\infty} P(|X_m - X| > \epsilon) \to 0$. ∎

DEFINITION **6.4.1** *$\{X_n, n \ge 1\}$ is said to converge* **completely** *to $X$ if $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$, see Hsu and Robbins (1947). Thus, complete convergence implies a.s. convergence.*

### 6.4.5 Convergence in mean sufficiently fast implies a.s. convergence

THEOREM **6.4.7** *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 6.4.6. ∎

### 6.4.6 Convergence sequences in probability contains a.s. subsequences

THEOREM **6.4.8** *If $X_n \to_p X$, then there exists a non-random integers $n_1 < n_2 < ...$ 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,

$$\sum_{k>1/\epsilon} P(|X_{n_k} - X| > \epsilon) \le \sum_{k>1/\epsilon} P(|X_{n_k} - X| > 1/k) \le \sum_{k>1/\epsilon} 1/k^2 < \infty.$$

The proof then follows from Theorem 6.4.6. ∎

### 6.4.7 Convergence in distribution plus uniform integrability implies convergence in moments

THEOREM **6.4.9** *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*

$$\lim_n EX_n^r = EX^r, \qquad \lim_n E|X_n|^r = E|X|^r.$$

*Proof.* See the section on uniform integrability later. ∎

### 6.4.8 Convergence in prob plus uniform integrability implies convergence in mean.

THEOREM **6.4.10** *Let $X_n \to_p X$, and $X_n \in L_r$ for $0 < r < \infty$. The the following three statements are equivalent.*

*(i)* $\{|X_n|^r\}$ *is u.i.*

*(ii)* $X_n \to X$ *in $L_r$. (i.e. $E|X_n - X|^r \to 0$.)*

*(iii)* $E|X_n|^r \to E|X|^r < \infty$.

*Proof.* See the section on uniform integrability later. ∎

### 6.4.9 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 **6.4.11 (Skorokhod's representation theorem)** *Suppose that $X_n \to_d X$. Then there exist r.v.'s $Y$ and $\{Y_n, n \geq 1\}$ on $((0,1), \mathcal{B}_{(0,1)}, P_\lambda = \lambda_{(0,1)})$ 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

$$Y_n(t) = F_n^{-1}(t) = \inf\{x : F_n(x) \geq t\}, \qquad Y(t) = F^{-1}(t) = \inf\{x : F(x) \geq t\}.$$

From Theorem 3.8.1, we have

$$t \leq F_n(x) \Longleftrightarrow Y_n(t) \leq x. \qquad \text{and,} \qquad t \leq F(x) \Longleftrightarrow Y(t) \leq x.$$

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 $F(x) < t$. 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

$$Y(t) - \epsilon < x < Y_n(t), \qquad \text{for all large } n.$$

Letting $n \to \infty$, we get

$$\liminf_n Y_n(t) \geq Y(t), \qquad \text{for all } t \in (0,1).$$

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

$$Y_n(t) \leq x < Y(t') + \epsilon, \qquad \text{for all } t \in (0,1).$$

Letting $n \to \infty$ and $\epsilon \to 0$, we obtain

$$\limsup_n Y_n(t) \leq Y(t') \qquad \text{whenever } 0 < t < t' < 1.$$

Thus,

$$Y(t) \leq \liminf_n Y_n(t) \leq \limsup_n Y_n(t) \leq Y(t') \qquad \text{whenever } 0 < t < t' < 1.$$

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,

$$0 \leq P_\lambda\left(\{t : \lim_n Y_n(t) \neq Y(t)\}\right) \leq P_\lambda(D) = 0. \quad \blacksquare$$

**Remarks:**

> (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 6.6.2 (c) or Helly's Theorem **??** later.

## 6.5 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; see Theorem 6.5.3 below.

### 6.5.1 Definition of uniform integrability (u.i.)

**First definition of u.i.**

For a single r.v. $X$, it can be easily shown from the DCT that

$$
\begin{aligned}
& X \text{ is integrable, i.e., } X \in L^1 = L^1(\Omega, \mathcal{F}, P) \\
\iff\ & E|X|I\{|X| > K\} \to 0 \text{ as } K \to \infty, \\
& (\text{as } P(|X|I\{|X| > K\} > \epsilon) = P(|X| > K) \le E|X|/K \to 0 \text{ and } |X|I\{|X| > K\} \le |X| \in L^1.) \\
\iff\ & \forall \varepsilon > 0, \exists K > 0 \text{ such that } E|X|I\{|X| > K\} \le \varepsilon \text{ as } K \to \infty,
\end{aligned}
$$

This motivates the notion of **uniform integrability** for a collection of r.v.s $Y_n$, $n \ge 1$ by **requiring** $E\{|Y_n|I_{\{|Y_n|>C\}}\} \to 0$ **as** $C \to \infty$ **uniformly in** $n$:

DEFINITION **6.5.1** *A sequence of r.v.'s $\{Y_n, n \ge 1\}$ on $(\Omega, \mathcal{A}, P)$ is* **u.i.** *if and only if*

$$\lim_{C\to\infty} \sup_{n\ge 1} E\{|Y_n|I_{\{|Y_n|>C\}}\} = 0. \quad \blacksquare$$

REMARK **6.5.1**

- *As can be seen from Theorem 6.5.1 below, Definition 6.5.1 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 6.5.1 below.*

- *A more general definition than Definition 6.5.1 can be given as follows: A collection of r.v.s $\{X_i, i \in I\}$ is said to be u.i. if $\lim_{C\to\infty} \sup_{i\in I} E\{|X_i|I_{\{|X_i|>C\}}\} = 0$.*

**Second definition of u.i.**

We will give an alternative definition of u.i. below, which is sometimes very useful. To motivate us, we give the following lemma first, which proves a stronger statement than $E|X|I\{|X| > K\} \to 0$ as $K \to \infty$, given earlier.

LEMMA **6.5.1 (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 $A \in \mathcal{A}$,*

$$Q(A) := E_A|X| \equiv E|X|I_A < \epsilon, \qquad \textit{whenever} \quad P(A) < \delta.$$

*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

$$P(A_n) < 2^{-n}, \qquad \text{and} \qquad E|X|I_{A_n} \ge \epsilon_0.$$

Let $H := \limsup A_n$. Since $\sum_n P(A_n) < \infty$, then by the Borel-Contelli lemma, we have

$$P(H) = P(\limsup A_n) = P(A_n,\ i.o) = 0,$$

which implies that

$$E_H|X| = E|X|I_H = 0.$$

On the other hand, by the reverse Fatou's lemma, we have

$$E_H|X| = E|X|I_H = E|X|I_{\limsup A_n} = E\limsup|X|I_{A_n} \geq \limsup E|X|I_{A_n} \geq \epsilon_0.$$

A contradiction. ∎.

Combining the above "integrable together" property and "absolute continuity" property, we will have another definition of u.i.

THEOREM **6.5.1 (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*

(a) $\sup_n E|Y_n| < \infty$, *and*

(b) $\forall \epsilon > 0$, $\exists \delta > 0$ *such that for any* $A \in \mathcal{A}$,

$$\sup_n E_A|Y_n| \equiv \sup_n E|Y_n|I_A < \epsilon, \qquad \textit{whenever} \quad P(A) < \delta.$$

*Proof.*

**Sufficiency.** Suppose $\{Y_n\}$ is **u.i.**, let us show that (a), (b) hold. Note that

$$\begin{aligned}
\sup_n E|Y_n| &\leq \sup_n E\{|Y_n|I_{\{|Y_n|\leq C\}}\} + \sup_n E\{|Y_n|I_{\{|Y_n|>C\}}\} \\
&\leq C + \sup_n E\{|Y_n|I_{\{|Y_n|>C\}}\}, \\
\sup_n E_A|Y_n| &\leq \sup_n E_A\{|Y_n|I_{\{|Y_n|\leq C\}}\} + \sup_n E_A\{|Y_n|I_{\{|Y_n|>C\}}\} \\
&\leq CP(A) + \sup_n E\{|Y_n|I_{\{|Y_n|>C\}}\}.
\end{aligned}$$

For any $\epsilon > 0$, choose and fix a large enough $C$ so that $\sup_n E\{|Y_n|I_{\{|Y_n|>C\}}\} \leq \epsilon$. Also choose $\delta \leq \epsilon/C$. Then $\sup_n E|Y_n| \leq C + 1 < \infty$ and $E_A|Y_n| \leq \epsilon$.

**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

$$\sup_n E_A|Y_n| = \sup_n E|Y_n|I_{\{|Y_n|>C\}} < \epsilon. \quad ∎$$

That is, $\lim_{C\to\infty}\sup_n E\{|Y_n|I_{\{|Y_n|>C\}}\} = 0$. ∎

REMARK **6.5.2** *Recall that Lemma 6.4.2 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.*

The following results are easy to prove, either by definition or by Theorem 6.5.1.

THEOREM **6.5.2**

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.* $\sup_{n\geq 1} E\{|X_n|I_{\{|X_n|>C\}}\} \leq \sup_{n\geq 1} E\{|Y_n|I_{\{|Y_n|>C\}}\} \to 0$. ∎
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.* $\sup_{n\geq 1} E\{|X_n| I_{\{|X_n|>C\}}\} \leq E\{|Y| I_{\{|Y|>C\}}\} \to 0.$ ∎

*7. If $E(\sup_n |X_n|) < \infty$, then $\{X_n\}$ is u.i.*

*(Note that $E(\sup_n |X_n|) \geq \sup_n E|X_n|$.)*

*Proof. Take $Y = \sup_n |X_n|$ in the DCT above.* ∎

*8. Let $\psi > 0$ satisfy $\lim_{x\to\infty} \dfrac{\psi(x)}{x} = \infty$. 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* $|x| \geq x_0$.

$$\sup_n E\{|Y_n| I_{\{|Y_n|>C\}}\} \leq M^{-1} \sup_n E\{\psi(|Y_n|) I_{\{|Y_n|>C\}}\} \leq M^{-1} \sup_n E\psi(|Y_n|) =: \epsilon A.$$

*Letting $C \to \infty$ and then letting $\epsilon \to 0$.* ∎

*(Remark: 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.)*

### 6.5.2 Convergence in prob. + u.i. $\Longrightarrow$ convergence in mean

The next theorem weakens the dominance condition in Theorem 6.4.2 to u.i. condition.

THEOREM **6.5.3 (Vitali's Theorem)** *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.*

*(i)* $\{X_n^r\}$ *is u.i.;*

*(ii)* $X_n \to X$ *in* $L_r$*; and* $E|X|^r < \infty$.

*(iii)* $E|X_n|^r \to E|X|^r < \infty$.

*Proof.*

"$(i) \Longrightarrow (ii)$". 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 6.6.2.) 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 $\{|X_n - X|^r\}$ is u.i.

**Proof.** We will use $C_r$ inequality $|X_n - X|^r \leq C_r(|X_n|^r + |X|^r)$. See Lemma 6.6.1 later on. Then apply Theorem 6.5.2, 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,

$$\begin{aligned} E|X_n - X|^r &= E|X_n - X|^r I_{\{|X_n - X|\leq\epsilon_0\}} + E|X_n - X|^r I_{\{|X_n - X|>\epsilon_0\}} \\ &\leq \epsilon_0^r + E|X_n - X|^r I_{\{|X_n - X|>\epsilon_0\}} \\ &\to \epsilon_0^r. \end{aligned}$$

Thus, $X_n \to X$ in $L_r$.

"$(ii) \Longrightarrow (iii)$". See Theorem 6.5.4 below.

"$(iii) \Longrightarrow (i)$".

Suppose (iii) is true. To prove (i), let $A > 0$ and construct a **nonnegative** and **continuous** function by

$$\begin{aligned} f_A(x) &= |x|^r & |x|^r \le A, \\ &\le |x|^r & A < |x|^r \le A+1, \\ &= 0 & |x|^r > A+1. \end{aligned}$$

(Draw a picture to illustrate)

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

$$\begin{aligned} \liminf_n E|X_n|^r I_{\{|X_n|^r \le A+1\}} &\ge \liminf_n Ef_A(X_n) = \lim_n Ef_A(X_n) = Ef_A(X) \\ &\ge E|X|^r I_{\{|X|^r \le A\}} \end{aligned}$$

Subtracting this from the assumption (iii): $\lim_n E|X_n|^r = E|X|^r < \infty$, we get

$$\limsup_n E|X_n|^r I_{\{|X_n|^r > A+1\}} \le E|X|^r I_{\{|X|^r > A\}}.$$

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 $n_0 = n_0(A_0(\epsilon))$ such that we have

$$\sup_{n \ge n_0} E|X_n|^r I_{\{|X_n|^r > A+1\}} \le \epsilon, \qquad \text{if } A \ge A_0.$$

However, since each $X_n^r$ is integrable, $\exists A_1 = A_1(\epsilon)$ such that

$$\sup_{n \ge 1} E|X_n|^r I_{\{|X_n|^r > A+1\}} \le \epsilon, \qquad \text{if } A \ge \max\{A_0, A_1\}.$$

That proves (i). ∎

### 6.5.3 Relationship between $L_r$ convergence and convergence of moments

THEOREM **6.5.4** *Suppose that $X_n \to X$ in $L_r$ ($r > 0$), and $E|X|^r < \infty$. Then*

$$(i).\ \lim_n E|X_n|^r = E|X|^r, \qquad (ii).\ \lim_n EX_n^r = EX^r.$$

*Proof.*

(i). For $0 < r \le 1$, apply the $C_r$ inequality $|x+y|^r \le |x|^r + |y|^r$ (see Lemma 6.6.1 later) to write $||x|^r - |y|^r| \le |x-y|^r$ and thus

$$|E|X_n|^r - E|X|^r| \le E\,||X_n|^r - |X|^r| \le E|X_n - X|^r$$

From the assumptions, we get $\lim_n E|X_n|^r = E|X|^r$.

For $r > 1$, apply Minkowski's inequality to obtain

$$\left|(E|X_n|^r)^{1/r} - (E|X|^r)^{1/r}\right| \le (E|X_n - X|^r)^{1/r}.$$

From the assumptions, we get $\lim_n (E|X_n|^r)^{1/r} = (E|X|^r)^{1/r}$, i.e. $\lim_n E|X_n|^r = E|X|^r$.

(ii). 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 6.5.5. ∎

**Remarks.**

(i). 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 $X_n = X \notin L_1$.

(ii). 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 $L_r$.

### 6.5.4 Convergence in dist. + u.i. $\Longrightarrow$ convergence in mean

We've seen in Section 6.5.2 that

**Convergence in prob. + u.i. $\Longrightarrow$ convergence in mean.**

In fact, convergence in prob. can be further weakened to convergence in dist.

THEOREM **6.5.5** *Suppose that $X_n \to_d X$, and $\{X_n^r\}$ is u.i. ($r > 0$). Then*

$$(i).\ E|X|^r < \infty, \qquad (ii).\ \lim_n EX_n^r = EX^r, \qquad (iii).\ \lim_n E|X_n|^r = E|X|^r.$$

*Proof.*

**Method 1.** Use Skorokhod's representation theorem, we find $\{Y_n\}$ and $Y$ such that $Y_n \to Y$ a.s.

(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 $n \to \infty$.

Then the proof follows from the results on Section 6.5.2 to $Y_n$:

**Convergence in prob. + u.i. $\Longrightarrow$ convergence in mean.**

**Method 2: Direct approach.**

(i). Let $F(x) = P(X \le x)$. 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,

$$\sup_n E|X_n^r| I_{\{|X_n| \ge C\}} \le \epsilon.$$

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).

(ii). 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 $\epsilon$. Thus, (ii) is proved.

(iii). Proof is similar to that of (ii). ∎

## 6.6 Some closed operations of convergence

### 6.6.1 Algebraic operations

THEOREM **6.6.1**

(a) *If* $X_n \to_p X$ *and* $Y_n \to_p Y$*, then* $X_n \pm Y_n \to_p X \pm Y$.

(b) *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.*

(c) *If* $X_n \to_r X$ *and* $Y_n \to_r Y$*, then* $X_n \pm Y_n \to_r X \pm Y$.

(d) *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.
(a) Given $\epsilon > 0$,

$$\begin{aligned} 0 &\le P(|X_n + Y_n - X - Y| > 2\epsilon) \\ &\le P\left(\{|X_n - X| > \epsilon\} \bigcup \{|Y_n - Y| > \epsilon\}\right) \\ &\le P(|X_n - X| > \epsilon) + P(|Y_n - Y| > \epsilon) \to 0. \end{aligned}$$

(b) Given $\epsilon > 0$,

$$\begin{aligned} 0 &\le P\left(\bigcup_{m \ge n} \{|X_m + Y_m - X - Y| > 2\epsilon\}\right) \\ &\le P\left(\bigcup_{m \ge n} \left[\{|X_m - X| > \epsilon\} \bigcup \{|Y_m - Y| > \epsilon\}\right]\right) \\ &\le P\left(\left[\bigcup_{m \ge n} \{|X_m - X| > \epsilon\}\right] \bigcup \left[\bigcup_{m \ge n} \{|Y_m - Y| > \epsilon\}\right]\right) \\ &\le P\left(\bigcup_{m \ge n} \{|X_m - X| > \epsilon\}\right) + P\left(\bigcup_{m \ge n} \{|Y_m - Y| > \epsilon\}\right) \\ &\to 0. \end{aligned}$$

(c) First we need a lemma:

LEMMA **6.6.1 ($C_r$-inequality.)** $|x + y|^r \le C_r(|x|^r + |y|^r)$*, where* $r > 0$ *and*

$$\begin{aligned} C_r &= 1 \qquad \text{if } 0 < r < 1 \\ &\quad\, 2^{r-1} \qquad \text{if } r \ge 1. \ . \end{aligned}$$

*(The lemma clearly implies that* $|x + y|^r \le 2^r(|x|^r + |y|^r)$.*)*

*Proof.* When $r \ge 1$, then the proof follows from $|(x + y)/2|^r \le (|x|^r + |y|^r)/2$.

When $0 < r < 1$, then $\lambda^r + (1 - \lambda)^r \ge \lambda + (1 - \lambda) = 1$ for $\lambda = |x|/(|x| + |y|)$.

**Proof of (c).**

$$\begin{aligned} 0 &\le E|X_n + Y_n - X - Y|^r = E|(X_n - X) + (Y_n - Y)|^r \\ &\le C_r(E|X_n - X|^r + E|Y_n - Y|^r) \to 0. \end{aligned}$$

Note that when $r \ge 1$, one could also use the Minkovski's inequality.

(d) $X_n \to_d X$, $Y_n \to_d Y$, but we may not have $X_n + Y_n \to_d X + Y$.

**Example**: 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. (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?

**Remark.** 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.)

### 6.6.2 Transformations (Continuous Mapping)

THEOREM **6.6.2 (Continuous mapping theorem)**

*Let $X_1, X_2, ...$ and $X$ be $k$-dim random vectors, $g : \mathcal{R}_k \to \mathcal{R}$ be continuous. Then*

$(a)$. $X_n \to X$ *a.s.* $\Longrightarrow g(X_n) \to g(X)$ *a.s.*

$(b)$. $X_n \to_p X \Longrightarrow g(X_n) \to_p g(X)$.

$(c)$. $X_n \to_d X \Longrightarrow g(X_n) \to_d g(X)$.

*Proof.* We shall only treat the case $k = 1$. The extension to general $k$ is trivial.

(a). 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 $1 = P(A) \le P(B) \le 1$, i.e., $P(B) = 1$.

(b). Let $\epsilon > 0$. We may pick some large $M$ such that $P(|X| \ge M) \le \epsilon$. The continuous function $g$ is uniformly continuous on the bounded interval $|x| \le M + \epsilon$. There exists $\delta > 0$ such that

$$|g(x) - g(y)| \le \epsilon, \qquad if \quad |x - y| \le \delta, \ and \ |x| \le M. \tag{6.2}$$

First we notice that

$$|g(X_n) - g(X)| > \epsilon, \quad and \quad |X_n - X| \le \delta \Longrightarrow |X| \ge M.$$

**Proof.** To see this, if this were not true, we would have $|X| < M$, which, in combination of $|X_n - X| \le \delta$, implies that $|X_n - X| \le M + \delta \le M + \epsilon$ (we can always choose $\delta \le \epsilon$). Then from (6.2), we would have $|g(x) - g(y)| \le \epsilon$, which contradicts with our assumption.

Therefore,

$$\begin{aligned}
& P\left(|g(X_n) - g(X)| > \epsilon\right) \\
= & P\left(|g(X_n) - g(X)| > \epsilon, |X_n - X| > \delta\right) + P\left(|g(X_n) - g(X)| > \epsilon, |X_n - X| \le \delta\right) \\
= & P\left(|X_n - X| > \delta\right) + P\left(|g(X_n) - g(X)| > \epsilon, |X_n - X| \le \delta\right) \\
\le & P\left(|X_n - X| > \delta\right) + P(|X| \ge M) \\
\to & P(|X| \ge M) \le \epsilon,
\end{aligned}$$

in the limit as $n \to \infty$. It follows that $g(X_n) \to_p g(X)$.

(c). As $X_n \to_d X$, by Skorokhod construction, $\exists Y, Y_1, Y_2, ...$ 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 $g(X_n) \to_d g(X)$. ∎

**Remarks.**

(i) What is wrong with the following "proof" of part (b) in the continuous mapping theorem?

**Wrong proof.** By the continuity of $g$, $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$|X_n - X| < \delta \qquad \Longrightarrow \qquad |g(X_n) - g(X)| < \epsilon.$$

Thus $P(|X_n - X| < \delta) \leq P(|g(X_n) - g(X)| < \epsilon)$. Take $\limsup_{n\to\infty}$ on both sides, we get $1 = \lim_{n\to\infty} P(|X_n - X| < \delta) \leq \liminf_{n\to\infty} P(|g(X_n) - g(X)| < \epsilon) \leq \limsup_{n\to\infty} P(|g(X_n)-g(X)| < \epsilon)$. Thus, we have $\limsup_{n\to\infty} P(|g(X_n)-g(X)| < \epsilon) = 1$. ∎

(ii) A different proof of (b) is given in Chow and Teicher, Corollary 3. page 68. ∎

The continuity assumption of $g$ everywhere can be weakened in the last theorem.

Theorem **6.6.3** *Let $X_1, X_2, ...$ and $X$ be $k$-dim random vectors. Assume that $g$ is continuous a.s. w.r.t. the probability measure $P_X$, i.e.*

$$P_X\left(X : g(\cdot) \textit{ is discontinous at } X\right) = P\left(\omega : g(\cdot) \textit{is discontinous at } X(\omega)\right) = 0.$$

*Then, Theorem 6.6.2 still holds.*

*Proof.* We shall prove the case $k = 1$. Using the same notation as in Theorem 6.6.2, 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., $P(B) = 1$.

Remark **6.6.1** *(i) In Theorem 6.6.3, if $X = C$, then $g(x)$ only needs to be continuous at $C$.*

*(ii) We shall now give an example, if $P_X\left(X : g(\cdot) \textit{is discontinous at } X\right) \neq 0$, then $X_n \to_p X$ $\not\Longrightarrow g(X_n) \to_p g(X)$.*

*e.g. Let $P(X_n = -1/n) = 1$, and $P(X = 0) = 1$. Define*

$$\begin{aligned} g(t) &= t - 1, \qquad t < 0 \\ &= t + 1, \qquad t \geq 0. \end{aligned}$$

*Clearly, $P_X\left(X : g(\cdot) \textit{is discontinous at } X\right) = P_X\left(X = 0\right) = 1 \neq 0$. However, we note $X_n \to_p X = 0$ whereas $\lim_n P\left(|g(X_n) + 1| \geq \epsilon\right) = \lim_n P\left(| - 1/n + 1| \geq \epsilon\right) = 1\ \forall \epsilon > 0$. That is, $g(X_n) \to_p -1 \neq 1 = g(0)$.*

**Examples:**

(i). If $X_n \to_d N(0, 1)$, then $X_n^2 \to_d N^2(0, 1) = \chi_1^2$.

(ii). 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 $P(g(X, Y) =: X/Y \text{is discontinous}) = P(Y = 0) = 0$. Apply the above remark.)

### 6.6.3 Slutsky's Theorem

We have seen from Theorem 6.6.1 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 $Y_n \to_d C \Longleftrightarrow Y_n \to_p C$).

Theorem **6.6.4 (Slutsky's Theorem)** *Let $X_n \to_d X$, $Y_n \to_p C$ (i.e. $Y_n \to_d C$). Then*

*(a). $X_n + Y_n \to_d X + C$.*

*(b). $X_n Y_n \to_d CX$.*

*(c). $X_n/Y_n \to_d X/C$ 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.

**Remark:**

(i) The proof of the theorem is almost the same as that of Theorem 6.3.1.

(ii) Slutsky's theorems are very useful in proving limiting distributions in so-called $\delta$-method. For an example, see Theorem A, page 118, Serfling (1980).

## 6.7 Simple limit theorems

We shall use what we learnt in this chapter to prove the following simple theorem.

THEOREM **6.7.1** *Suppose that $X_i$'s are uncorrelated and* $\sup_{k\ge1} EX_k^2 \le M < \infty$. *Denote* $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $\mu_i = EX_i$ *and* $\bar{\mu} = n^{-1}\sum_{i=1}^n \mu_i$. *Then*

*(i)* $\bar{X} - \bar{\mu} \to_{L_2} 0$;

*(ii)* $\bar{X} - \bar{\mu} \to_p 0$;

*(iii)* $\bar{X} - \bar{\mu} \to_{a.s.} 0$.

*Proof.* Note that $\sigma_i^2 =: Var(X_i) = EX_i^2 - (EX_i)^2 \le EX_i^2 \le M$, and $\mu_i$ exists. Write $S_n = \sum_{i=1}^n X_i$, then

$$\bar{X} - \bar{\mu} = \frac{S_n - ES_n}{n}.$$

(i) $E(\bar{X} - \bar{\mu})^2 = \frac{1}{n^2}\sum_i\sum_j E(X_i - \mu_i)(X_j - \mu_j) = \frac{1}{n^2}\sum_i E(X_i - \mu_i)^2 = \frac{1}{n^2}\sum \sigma_i^2 \le \frac{M}{n} \to 0$

(ii). This is implied by (i) below.

(iii) Then $Var(S_n) = \sum_{i=1}^n \sigma_i^2$. By Chebyshev's inequality,

$$P\left(|S_n - ES_n| > n\epsilon\right) \le \frac{Var(S_n)}{n^2\epsilon^2} \le \frac{M}{n\epsilon^2}.$$

Summing over $n$ on RHS, the resulting series diverges. However, if we confine our attention to the subsequence $\{n^2; n \geq 1\}$, then

$$\sum_n P\left(|S_{n^2} - ES_{n^2}| > n^2\epsilon\right) \leq \sum_n \frac{M}{n^2\epsilon^2} < \infty. \tag{7.3}$$

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 6.1.1, we have

$$\frac{S_{n^2} - ES_{n^2}}{n^2} \to 0, a.s. \tag{7.4}$$

(In fact, (7.4) is implied directly by (7.3) 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

$$P(D_n > n^2\epsilon) \leq \frac{4M}{\epsilon^2 n^2}.$$

It follows as before that

$$\frac{D_n}{n^2} \to 0, a.s. \tag{7.5}$$

Now it is clear from (7.4) and (7.5) that, for $n^2 \leq k < (n+1)^2$

$$0 \leq \frac{|S_k|}{k} = \frac{|S_{n^2} + (S_k - S_{n^2})|}{k} \leq \frac{|S_{n^2}| + |S_k - S_{n^2}|}{k} \leq \frac{|S_{n^2}| + |D_n|}{n^2} \to 0 \quad a.s. \quad \blacksquare$$

**Remark:**

(i). 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.

(ii). 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.

## 6.8 Fatou's Lemma Revisited

We shall derive a most general Fatou's lemma, (see Durrett, Chapter 2, p48.)

THEOREM **6.8.1 (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*

$$\limsup_n Eg(X_n) \leq Eg(X_\infty).$$

*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 $X_\infty =_d Y_\infty$.

$$Eg(X_\infty) = Eg(Y_\infty) = Eg(\lim_n Y_n) = E\lim_n Eg(Y_n) = E\liminf_n g(Y_n)$$

$$\leq \liminf_n Eg(Y_n) = \liminf_n Eg(X_n). \quad \blacksquare$$

COROLLARY **6.8.1** *$\{X_n\}$ is a sequence of r.v.'s.*

*(i) If $X_n \geq Y$ in any mode (i.e., in prob, or $L^r$, or distribution, or a.s.), then*

$$\liminf_n EX_n \geq E(\liminf_n X_n).$$

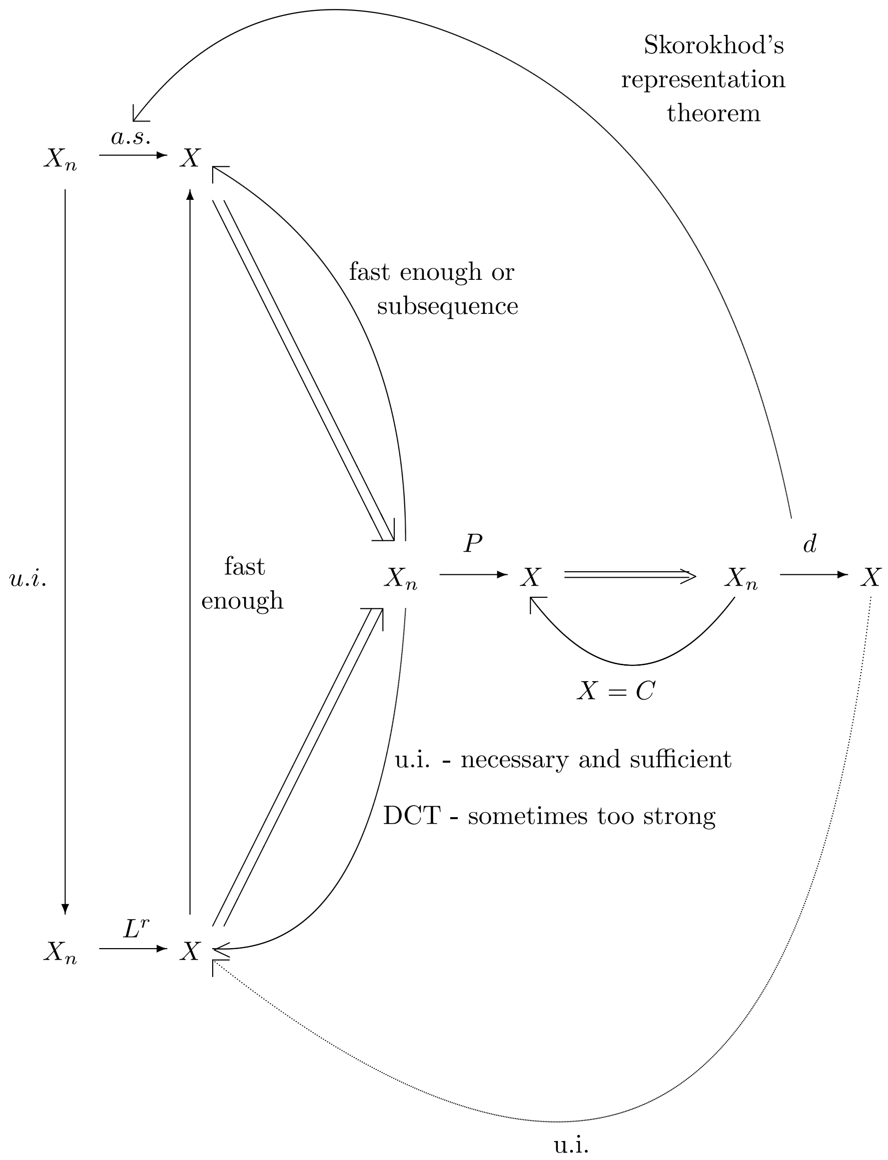
*(ii) If $X_n \leq Y$ in any mode (i.e., in prob, or $L^r$, or distribution, or a.s.), then*

$$\limsup_n EX_n \leq E(\limsup_n X_n). \quad \blacksquare$$

The proof follows from the last theorem. The case $Y = 0$ is most often used.

## 6.9 Summary: relationships amongst four modes of converges

The relationships amongst four modes of converges can be summarized in the following diagram.


Skorokhod's
representation
theorem
$X_n$
$a.s.$
$X$
fast enough or
subsequence
$u.i.$
fast
enough
$X_n$
$P$
$X$
$X_n$
$d$
$X$
$X = C$
u.i. - necessary and sufficient
DCT - sometimes too strong
$X_n$
$L^r$
$X$
u.i.

## 6.10 Exercises

1. $X_n \to_d X$ i.e. $\lim_n F_n(x) = F(x)$ for all $x \in C(F) =$ all continuity points of $x$. Give an example to show that $\lim_n F_n(x) = F(x)$ may not be true if $x \notin C(F)$.

   (Hint: consider $\delta_{n^{-1}}(x) \Longrightarrow \delta_0(x)$. But $\delta_n(0) \equiv 0 \neq 1 = \delta_0(0)$.)

2. If $|X| \leq 1$ a.s. then $P(|X| \geq \epsilon) \geq EX^2 - \epsilon^2$.

3. Show that

   (a) $X_n \to_p X \Longrightarrow X_n - X \to_p 0$.
   (b) If $X_n \to_p X$ and $X_n \to_p Y$, then $X = Y$ a.s.
   (c) $X_n \to_p 1 \Longrightarrow 1/X_n \to_p 1$.
   (d) $X_n \to_p 0$, $Y_n \to_p 0 \Longrightarrow X_n Y_n \to_p 0$.
   (e) $X_n \to_p a$, $Y_n \to_p b \Longrightarrow X_n Y_n \to_p ab$.
   (f) $X_n \to_p X$, and $Y$ is a r.v. $\Longrightarrow X_n Y \to_p XY$.
   (g) $X_n \to_p X$, $Y_n \to_p Y \Longrightarrow X_n Y_n \to_p XY$.

4. Check whether the following statements are true or false. Either prove your claim or give a counter-example. (Compare with (g) in the last question.)

   (i) If $X_n \to_{a.s.} X$ and $Y_n \to_{a.s.} Y$, $\Longrightarrow X_n Y_n \to_{a.s.} XY$.
   (ii) If $X_n \to X$ in $L_r$ and $Y_n \to Y$ in $L_r$ $(r > 0)$, $\Longrightarrow X_n Y_n \to XY$ in $L_r$.
   (iii) If $X_n \to_d X$ and $Y_n \to_d Y$, $\Longrightarrow X_n Y_n \to_d XY$.

5. $X_n \leq Y_n \leq Z_n$, $X_n \to_{a.s.} Y$, and $Z_n \to_{a.s.} Y \Longrightarrow Y_n \to_{a.s.} Y$.

6. $X_n \downarrow$, and $X_n \to_p 0$, $\Longrightarrow X_n \to_{a.s.} 0$.

7. $X_1, ..., X_n$ are i.i.d. r.v.'s with $\mu = EX_1$ and $\sigma^2 = Var(X_1) < \infty$. Show that

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} i X_i \to_p \mu.$$

8. $X_1, ..., X_n$ are i.i.d. r.v.'s with $\mu = EX_1$ and $EX_i^4 < \infty$. Let $S_n = \sum_1^n X_i$. Show that

$$\frac{S_n}{n} \to \mu, \qquad a.s.$$

9. $\{X_n\}$ is a sequence of r.v.'s (not necessarily independent), $S_n = \sum_1^n X_i$. Then for any $r \geq 1$,

   (a) $X_n \to 0$ a.s. $\Longrightarrow S_n/n \to 0$ a.s.
   (b) $X_n \to 0$ in $L^r \Longrightarrow S_n/n \to 0$ in $L^r$.
   (c) If $0 < r < 1$, then (b) may not hold.
   (d) $X_n \to_p 0 \not\Longrightarrow S_n/n \to_p 0$.

   (Hint: Let $\{X_n\}$ be independent and $P(X_n = 2^n) = n^{-1}$ and $P(X_n = 0) = 1 - n^{-1}$.)

   (e) $S_n/n \to_p 0 \Longrightarrow X_n/n \to_p 0$.
   (f) $S_n/a_n \to_p 0$ and $a_n/a_{n-1} \to 1 \Longrightarrow X_n/a_n \to_p 0$.

10. Let $\{X_n\}$ be i.i.d. r.v.'s. Then

    (a) $n^{-1} \max_{1 \leq k \leq n} |X_k| \to_p 0 \Longleftrightarrow nP(|X_1| > n) = o(1)$.
    (b) $n^{-1} \max_{1 \leq k \leq n} |X_k| \to 0$ a.s. $\Longleftrightarrow E|X_1| < \infty$.

11. Let $P(X_n = a_n > 0) = 1/n = 1 - P(X_n = 0)$, $n \geq 1$. Is $\{X_n, n \geq 1\}$ u.i. if

(i) $a_n = o(n)$,
(ii) $a_n = cn > 0$?

12. If $\{X_n : n \ge 1\}$ and $\{Y_n : n \ge 1\}$ are both u.i., then so is $\{X_n + Y_n : n \ge 1\}$.

13. If $\exists Y$ such that $E|Y| < \infty$ and $P(|Y_n| \ge y) \le P(|Y| \ge y)$ for all $n \ge 1$ and all $y > 0$, then $Y_n$ is u.i.

    (Remark: The condition can also be written as $F_n(y) \ge F(y)$, meaning that $Y_n \le_{stoch} Y$ (i.e. $Y_n$ is stochastically smaller than $Y$), which in turn implies that $E|Y_n| \le E|Y|$. So the condition is rather like a **dominated convergence condition**.)

14. Is the limit of distribution functions (d.f.) necessarily a d.f.? To be more precise, let $\lim_{n\to\infty} F_n(x) = F(x)$, where $F_n(x)$'s are d.f.'s for all $n \ge 1$. Is $F(x)$ necessarily a d.f.? Either prove your statement or give a counterexample.

15. Let $\{X_n, n \ge 1\}$ be a sequence of uniformly bounded r.v.'s (i.e. $|X_n| < C$ for all $n$ and some constant $C$). Write $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Show that $\bar{X} - E\bar{X} \to_p 0$ iff $Var(\bar{X}) \longrightarrow 0$.

16. **Pratt's Lemma**. Suppose that

    (i) $X_n \le Y_n \le Z_n$ a.s.
    (ii) $X_n \to_{a.s.} X, \qquad Y_n \to_{a.s.} Y, \qquad Z_n \to_{a.s.} Z.$
    (iii) $EX_n \to EX$ and $EZ_n \to EZ$, where $E|X| < \infty$ and $E|Z| < \infty$.

    Show that $EY_n \to EY$. (Hint: Apply Fatou's Lemma.)

17. The following proof is a WRONG proof of the last question. Where did it go wrong?

    **Wrong Proof.** For $n$ large enough, we have

$$\begin{aligned} |Y_n| &\le \max\{|X_n|, |Z_n|\} \quad a.s. \quad \text{[from (i)]} \\ &\le |X_n| + |Z_n| \quad a.s. \\ &\le |X| + |Z| + 2\epsilon, \quad a.s. \quad \text{[from (ii)]}. \end{aligned}$$

    Then apply the Dominated Convergence Theorem to $Y_n$ since $E(|X| + |Z| + 2\epsilon) < \infty$.

18. Give an example to illustrate that $X_n \longrightarrow X$ a.s. may not imply

$$X_n \le X + \epsilon, \qquad \text{for large enough } n.$$

19. Let $\{X_n\}$ be a sequence of independent r.v.'s which converges in probability to the limit $X$. Show that $P(X = C) = 1$ for some constant $C$.

20. Let $X, X_1, X_2, \ldots$ be r.v.'s such that $X_n \longrightarrow_{a.s.} X$. Show that $\sup_n |X_n| = O_p(1)$.

# Chapter 7

# Weak Law of Large Numbers

In the last chapter, we introduced various modes of convergence (a.s., in probability, in mean, in distribution) for a general sequence $X_1, X_2, \ldots\ldots$. In the next two chapters, we shall study these issues for some more "meaningful" random variables, such as $\bar{X} = (X_1 + \ldots + X_n)/n$ (called "sample mean" in statistics), or equivalently the partial sum $S_n = X_1 + \ldots + X_n$. In probability theory, the following **limit theorems** are of major concern:

(i) **Weak law of large numbers (WLLN)**:

What are the necessary and sufficient conditions for $\dfrac{S_n}{B_n} - A_n \longrightarrow_p 0$, where $\{A_n\}$ and $\{B_n\}$ are two non-random sequences, $B_n > 0$ and $B_n \nearrow \infty$?

(ii) **Strong law of large numbers (SLLN)**:

What are the necessary and sufficient conditions for $\dfrac{S_n}{B_n} - A_n \longrightarrow_{a.s.} 0$, where $\{A_n\}$ and $\{B_n\}$ are as in (i)?

(iii) **Law of iterated logorithms (LIL)**:

When do we have $\limsup_{n\to\infty} \dfrac{S_n}{B_n} = 1, a.s.$

(iv) **Central limit theorem (CLT)**:

What are the necessary and sufficient conditions for $\dfrac{S_n}{B_n} - A_n \longrightarrow_d N(0,1)$?

(v) **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:

(a) dependent sequences (mixing conditions, martingales);

(b) limit theorems for other statistics than $S_n$

(c) functional limit theorems (stochastic processes)

## 7.1 Equivalent sequences; truncation

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

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

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

(a) $\sum_{n=1}^{\infty}(X_n - Y_n)$ *converges a.s.;*

(b) *If* $a_n \uparrow \infty$, *then* $\dfrac{1}{a_n}\displaystyle\sum_{j=1}^{n}(X_j - Y_j) \to 0$ *a.s.*

*Proof.* By the Borel-Cantelli lemma, the assumption of equivalence implies

$$P(\{\omega : X_n(\omega) \neq Y_n(\omega)\}, i.o.) = P(X_n \neq Y_n, i.o.) = 0.$$

Hence,

$$P\left(\{\omega : X_n(\omega) = Y_n(\omega)\}, ult.\right) = 1 - P(\{X_n \neq Y_n\}^c, i.o.) = 1 - P(X_n \neq Y_n, i.o.) = 1.$$

Thus, $\exists$ a $P$-null set $N$ with the property: if $\omega \in \Omega - N$, $\exists n_0(\omega)$ such that

$$n \geq n_0(\omega) \qquad \Longrightarrow \qquad X_n(\omega) = Y_n(\omega).$$

For such an $\omega$, the two numerical sequences $\{X_n(\omega)\}$ and $\{Y_n(\omega)\}$ differ only in a finite number of terms (how many depending on $\omega$). 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. ∎

An easy and useful consequence of the last theorem is the following.

Corollary **7.1.1** *Suppose that* $\{X_n\}$ *and* $\{Y_n\}$ *are equivalent, and* $a_n \uparrow \infty$. *Then with probability one (a.s.)*

(a) $\sum_{j=1}^{n} X_j$ *or* $\dfrac{1}{a_n}\displaystyle\sum_{j=1}^{n} X_j$ *converges, diverges to* $+\infty$ *or* $-\infty$, *or fluctuates in the same way as* $\sum_{j=1}^{n} Y_j$ *or* $\dfrac{1}{a_n}\displaystyle\sum_{j=1}^{n} Y_j$.

(b) *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

$$\frac{1}{a_n}\sum_{j=1}^{n} Y_j = \frac{1}{a_n}\sum_{j=1}^{n} X_j + \frac{1}{a_n}\sum_{j=1}^{n}(Y_j - X_j) \to_p X. \quad ∎$$

## 7.2 Weak Law of Large Numbers

Theorem **7.2.1** *Let* $\{X_i\}$ *be* **pairwise** *independent and identically distributed r.v.'s with finite mean* $\mu = EX_1$. *Then*

$$\bar{X} \to_p \mu.$$

*Proof.* $|EX_1| < \infty, \qquad \Longleftrightarrow \qquad E|X_1| < \infty \qquad \Longleftrightarrow$

$$\sum_{n=1}^{\infty} P(|X_1| > n) = \sum_{n=1}^{\infty} P(|X_n| > n) < \infty, \qquad \text{(identical distribution)} \tag{2.1}$$

Define the "**truncated**" r.v.'s $\{Y_n\}$ by

$$Y_n = X_n I_{\{|X_n| \leq n\}}$$

Then $\{X_n\}$ and $\{Y_n\}$ are equivalent since by (2.1)

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

Let $T_n = \sum_{i=1}^{n} Y_i$ and $\bar{Y} = n^{-1}\sum_{j=1}^{n} Y_j$. By Theorem 7.1.1, $\bar{X}$ and $\bar{Y}$ converges or diverges at the same time. Therefore, it suffices to show, as $n \to \infty$,

$$(i).\ \ E\bar{Y} \to \mu, \qquad\qquad (ii).\ \ Var(\bar{Y}) \to 0.$$

To see why, applying Markov's inequality,

$$0 \le P(|\bar{Y} - \mu| > \epsilon) \le E|\bar{Y} - \mu|^2/\epsilon^2 = \epsilon^{-2}\left[Var(\bar{Y}) + bias^2(\bar{Y})\right] \to 0.$$

**Proof of (i).**

$$\begin{aligned}
0 \quad \le \quad |E\bar{Y} - \mu| &= \frac{1}{n}\left|\sum_{j=1}^{n}(EY_j - EX_j)\right| = \frac{1}{n}\left|\sum_{j=1}^{n}\left(EX_j I_{\{|X_j| \le j\}} - EX_i\right)\right| \\
= \quad \frac{1}{n}&\left|-\sum_{j=1}^{n} EX_j I_{\{|X_j| > j\}}\right| \le \frac{1}{n}\sum_{j=1}^{n} E|X_j| I_{\{|X_j| > j\}} \longrightarrow 0,
\end{aligned}$$

where we have used the fact: "$a_n \to a$" $\Longrightarrow$ "$\bar{a} \to a$", where $a_n = E|X_n|I_{\{|X_n|>n\}} \to 0$.

**Proof of (ii).** It is equivalent to show $Var\left(\sum_{i=1}^n Y_i\right) = o(n^2)$. Note that $Y_n$ are independent (as functions of $X_n$) and bounded. Thus

$$Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{k=1}^{n} Var(Y_k) \le \sum_{k=1}^{n} EY_k^2 = \sum_{k=1}^{n} EX_k^2 I_{\{|X_k| \le k\}} = \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \le k\}}$$

The crudest estimate of this is

$$Var\left(\sum_{i=1}^{n} Y_i\right) \le \sum_{k=1}^{n} kE|X_1| I_{\{|X_1| \le k\}} \le \sum_{k=1}^{n} kE|X_1| = \frac{1}{2}n(n+1)E|X_1| = O(n^2),$$

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** $k = 1, ..., n$. *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

$$\begin{aligned}
Var\left(\sum_{i=1}^{n} Y_i\right) &\le \sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \le k\}} \\
&= \sum_{k=1}^{a_n} EX_1^2 I_{\{|X_1| \le k\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{|X_1| \le k\}} \\
&= \sum_{k=1}^{a_n} EX_1^2 I_{\{|X_1| \le k\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{|X_1| \le a_n\}} + \sum_{k=a_n}^{n} EX_1^2 I_{\{a_n < |X_1| \le k \le n\}} \\
&\le \sum_{k=1}^{a_n} kE|X_1| I_{\{|X_1| \le k\}} + a_n \sum_{k=a_n}^{n} E|X_1| I_{\{|X_1| \le a_n\}} + n \sum_{k=a_n}^{n} E|X_1| I_{\{a_n < |X_1| \le n\}} \\
&\le a_n \sum_{k=1}^{a_n} E|X_1| + a_n \sum_{k=a_n}^{n} E|X_1| + n \sum_{k=a_n}^{n} E|X_1| I_{\{|X_1| > a_n\}} \\
&\le n a_n E|X_1| + n^2 E|X_1| I_{\{|X_1| > a_n\}},
\end{aligned}$$

which implies that

$$0 \le Var(\bar{Y}) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} Y_i\right) \le \frac{a_n}{n} E|X_1| + E|X_1| I_{\{|X_1| > a_n\}} \to 0.$$

This completes our proof. ∎

Remark **7.2.1** *A finer truncation could be used to bound* $Var\left(\bar{Y}\right)$*, which goes as follows.*

$$Var\left(\bar{Y}\right) \quad \le \quad \frac{1}{n^2}\sum_{k=1}^{n} EX_1^2 I_{\{|X_1| \le k\}}$$

*But*

$$
\begin{aligned}
EX_1^2 I_{\{|X_1|\le n\}} &= \sum_{j=1}^{n} EX_1^2 P\{j-1 \le |X_1| < j\} \\
&\le \sum_{j=1}^{n} j^2 P\{j-1 \le |X_1| < j\} \\
&\le \sum_{j=1}^{n} j(j+1) P\{j-1 \le |X_1| < j\} \\
&\le 2\sum_{j=1}^{n}\sum_{i=1}^{j} iP\{j-1 \le |X_1| < j\} \\
&\le 2\sum_{i=1}^{n}\sum_{j=i}^{n} iP\{j-1 \le |X_1| < j\} \\
&= 2\sum_{i=1}^{n} i\sum_{j=i}^{n} P\{j-1 \le |X_1| < j\} \\
&= 2\sum_{i=1}^{n} iP\{i-1 \le |X_1| < n\} \\
&\le 2\sum_{i=1}^{n} iP\{|X_1| \ge i-1\} \\
&\le 2\left(1+\sum_{i=1}^{\infty} iP\{|X_1| \ge i\}\right) \\
&\le 2\,(2+E|X_1|)??????????Check????????????? \\
&< \infty.
\end{aligned}
$$

*It follows that*

$$
\begin{aligned}
Var\left(\bar{Y}\right) &\le \frac{1}{n^2}\sum_{k=1}^{n} EX_1^2 I_{\{|X_1|\le k\}} \\
&\le \frac{1}{n^2}\sum_{k=1}^{n} EX_1^2 I_{\{|X_1|\le n\}} \\
&\le \frac{1}{n}2\,(2+E|X_1|) \to 0. \quad \blacksquare
\end{aligned}
$$

The above theorem may be slightly generalized as follows; compare with Corollary 7.3.1 later.

THEOREM **7.2.2** *Let* $\{X_i\}$ *be* **pairwise** *independent and identically distributed r.v.'s such that*

$$EX_1 I\{|X_1| \le n\} \longrightarrow 0, \qquad \text{and} \qquad nP(|X_1| > n) \to 0.$$

*Then* $\bar{X} \to_p 0$. $\blacksquare$

**Remark.** We should appreciate the power of **truncation** from the proof of the theorem.

**Remark.** 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 **7.2.3** *Let* $\{X_i\}$ *be i.i.d. r.v.'s with finite mean* $\mu = EX_1$. *Then*

$$\bar{X} \to_p \mu.$$

*Proof.* Let $\psi_X(t) = Ee^{itX}$. Since $|EX_1| < \infty$, we have

$$\psi_{X_1}(t) = \psi_{X_1}(0) + \psi'_{X_1}(0)t + o(t) = 1 + i\mu t + o(t), \qquad |t| < \delta.$$

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 $\bar{X} \to_d \mu$, or $\bar{X} \to_p \mu$. ∎

If $\mu = EX_1$ exists, we could prove a stronger result: $\bar{X} \to_{a.s.} \mu$. (see the next chapter.)

## 7.3 Classical forms of the WLLN

Theorem **7.3.1 (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 $n \to \infty$,*

$$\begin{aligned} (i) &\qquad \sum_{k=1}^n P\left(|X_1| \ge a_n\right) \longrightarrow 0 \\ (ii) &\qquad E\left(\sum_{k=1}^n \frac{X_1}{a_n} I_{\{|X_1|<a_n\}}\right) \longrightarrow 0 \\ (iii) &\qquad Var\left(\sum_{k=1}^n \frac{X_1}{a_n} I_{\{|X_1|<a_n\}}\right) \longrightarrow 0 \end{aligned}$$

*if and only if, by writing $Y_{nk} = \frac{X_1}{a_n} I\left\{\frac{X_1}{a_n} < 1\right\}$, as $n \to \infty$,*

$$\begin{aligned} (i) &\qquad \sum_{k=1}^n P\left(|X_1| \ge a_n\right) \longrightarrow 0 \\ (ii) &\qquad E\left(\sum_{k=1}^n Y_{nk}\right) \longrightarrow 0 \\ (iii) &\qquad Var\left(\sum_{k=1}^n Y_{nk}\right) \longrightarrow 0. \end{aligned}$$

*Proof.* Omitted. For details, please refer to Petrov (1995, p132), and Chung, who give two different treatments.

Remark **7.3.1** .

(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*

$$E(\bar{Y}) \to 0, \qquad var(\bar{Y}) \to 0,$$

*which ensures that $\bar{Y} \to_p 0$. On the other hand, (i) implies that $\{X_n\}$ and $\{Y_n\}$ are equivalent, hence, $\bar{X} \to_p 0$.*

When $\{X_n\}$ are i.i.d. r.v.'s, we have the following theorem.

THEOREM **7.3.2** *Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.'s with common d.f. $F$. Then the following statements are equivalent:*

- *(i)* $\bar{X} := n^{-1} \sum_{k=1}^{n} X_1 \to_p C$ *for some constant $C$,*
- *(ii)* $nP(|X_1| \geq n) \to 0$ *and* $E[X_1 I\{|X_1| < n\}] \to C$ *as* $n \to \infty$.
- *(iii) the characteristic function (c.f.) $\psi(t)$ of the $X_1$ is differentiable at $t = 0$ and $\psi'(0) = iC$.*

COROLLARY **7.3.1** *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 $a_n = E[X_1 I\{|X_1| < n\}]$.*

REMARK **7.3.2** *We make some remarks about Condition (ii) in Theorem 7.3.2.*

*(a). Recall that $nP(|X| \geq n) = o(1) \Longrightarrow E|X|^{1-\epsilon} < \infty$, but $\not\Longrightarrow E|X| < \infty$. 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).*

*(b). 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.*

*(c). 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*

$$F(x) \approx 1 - \frac{1}{x \log x}, \qquad x \to \infty.$$

## 7.4 Exercises

1. Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. r.v.'s, and $g$ a bounded continuous function. Write $\mu = EX_1$ and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Show that
$$\lim_{n\to\infty} Eg(\bar{X}) = g(\mu).$$
(Hint: Apply the WLLN.)

2. Let $f$ and $g$ be continuous functions on $[0, 1]$ satisfying $0 \leq f(x) \leq Cg(x)$ for all $x$ and some constant $C > 0$. Show that
$$\lim_{n\to\infty} \int_0^1 ... \int_0^1 \frac{\sum_{i=1}^n f(x_i)}{\sum_{i=1}^n g(x_i)} dx_1...dx_n = \frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx}, \qquad \text{where } \int_0^1 g(x)dx \neq 0.$$
In fact, a more general result is: for i.i.d. r.v. $X, X_1, X_2, ...$, we have
$$\lim_{n\to\infty} E\frac{\sum_{i=1}^n f(X_i)}{\sum_{i=1}^n g(X_i)} = \frac{Ef(X)}{Eg(X)}, \qquad \text{where } Eg(X) \neq 0.$$
(Hint: Apply the WLLN.)

# Chapter 8

# Strong Convergence

The central theme of this chapter is to characterize the a.s. convergence of

(i) series $\sum_{k=1}^{\infty} X_k$ (when $X_i$'s are independent),
(ii) averages $n^{-\alpha} \sum_{k=1}^{n} X_k$, 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

**(i) Kolmogorov three series criterion (for series,)**
**(ii) the strong law of large numbers (for averages)**.

In a nutshell, 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 random sequences. For instance, convergence in probability and a.s. convergence for series $\sum_{k=1}^{\infty} X_k$ are equivalent. (We shall however not cover this here, for details, see Loeve, e.g.).

## 8.1 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 **8.1.1 (Hajek-Renyi maximal inequality)** *Let $X_1, X_2, ...$ be independent with $EX_k = 0$ and $\sigma_k^2 = Var(X_i) < \infty$. 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

$$E_m = \{c_m |S_m| \geq \epsilon\},$$

$$\begin{aligned}
E_j &= \left\{ \max_{m \le k < j} \{c_k |S_k|\} < \epsilon, \ c_j |S_j| \ge \epsilon \right\}, \quad m+1 \le j \le n, \\
A &= \sum_{j=m}^{n} E_j = \left\{ \max_{m \le k \le n} \{c_k |S_k|\} \ge \epsilon \right\}, \\
Y &= c_m^2 S_m^2 + \sum_{k=m+1}^{n} c_k^2 \left(S_k^2 - S_{k-1}^2\right).
\end{aligned}$$

Note that $ES_k^2 = \sum_{j=1}^{k} \sigma_j^2$, $E_j$'s are mutually exclusive. It suffices to show that

$$P(A) = P\left(\sum_{j=m}^{n} E_j\right) = \sum_{j=m}^{n} P(E_j) \le \frac{1}{\epsilon^2} EY.$$

Now we can rewrite

$$\begin{aligned}
Y &= c_m^2 S_m^2 + \sum_{k=m+1}^{n} c_k^2 S_k^2 - \sum_{k=m+1}^{n} c_k^2 S_{k-1}^2 \\
&= \sum_{k=m}^{n} c_k^2 S_k^2 - \sum_{k=m+1}^{n} c_k^2 S_{k-1}^2 \\
&= \sum_{k=m}^{n} c_k^2 S_k^2 - \sum_{k=m}^{n-1} c_{k+1}^2 S_k^2 \\
&= \sum_{k=m}^{n-1} \left(c_k^2 - c_{k+1}^2\right) S_k^2 + c_n^2 S_n^2 \\
&\ge 0, \qquad \text{as } c_k \searrow.
\end{aligned}$$

So

$$EY \ge E(YI_A) = E(YI_{\sum_{j=m}^{n} E_j}) = \sum_{j=m}^{n} EYI_{E_j}$$

For $m \le j \le n$, we have

$$\begin{aligned}
EYI_{E_j} &= \sum_{k=m}^{n-1} \left(c_k^2 - c_{k+1}^2\right) ES_k^2 I_{E_j} + c_n^2 ES_n^2 I_{E_j} \\
&\ge \sum_{k=j}^{n-1} \left(c_k^2 - c_{k+1}^2\right) ES_k^2 I_{E_j} + c_n^2 ES_n^2 I_{E_j}.
\end{aligned}$$

But for $j \le k \le n$, we have

$$\begin{aligned}
ES_k^2 I_{E_j} &= E[(S_j + (S_k - S_j)]^2 I_{E_j} \\
&= ES_j^2 I_{E_j} + E(S_k - S_j)^2 I_{E_j} + 2ES_j(S_k - S_j) I_{E_j} \\
&= ES_j^2 I_{E_j} + E(S_k - S_j)^2 I_{E_j} \\
&\quad (\text{as } E[S_j I_{E_j}(S_k - S_j)] = E[S_j I_{E_j}] E[(S_k - S_j)] = 0, \\
&\quad \text{since } S_j I_{E_j} \text{ and } S_k - S_j \text{ are independent}) \\
&\ge ES_j^2 I_{E_j} \\
&\ge \epsilon^2 P(E_j)/c_j^2, \qquad (\text{as } |c_j S_j| \ge \epsilon \text{ on } E_j)
\end{aligned}$$

Thus,

$$EYI_{E_j} \quad \ge \quad \left[\sum_{k=j}^{n-1} \left(c_k^2 - c_{k+1}^2\right) + c_n^2\right] \frac{\epsilon^2 P(E_j)}{c_j^2} = \epsilon^2 P(E_j).$$

Finally, we get

$$EY \ge EYI_A = \sum_{j=m}^{n} \epsilon^2 P(E_j) = \epsilon^2 P\left(\sum_{j=m}^{n} E_j\right) = \epsilon^2 P(A).$$

This completes our proof.

One important special case is the following.

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

(*a*) (**Upper bound**)

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

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

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

(*Note that the RHS $= -\infty$ when $C = \infty$.*)

*Proof.* (a) Take $m = 1$ and $c_k = 1$, $k \ge 1$ in Hajek-Renyi maximal inequality.

**Direct proof.** Let

$$\begin{aligned}
E_1 &= \{|S_1| \ge \epsilon\}, \\
E_j &= \left\{\{\max_{1\le k<j} |S_k|\} < \epsilon,\ |S_j| \ge \epsilon\right\}, \qquad \text{for } 2 \le j \le n. \\
A &= \sum_{j=1}^n E_j = \left\{\{\max_{1\le k\le n} |S_k|\} \ge \epsilon\right\},
\end{aligned}$$

So

$$\begin{aligned}
Var(S_n) &= ES_n^2 \ge ES_n^2 I_A = \sum_{j=1}^n ES_n^2 I_{E_j} = \sum_{j=1}^n E[(S_j + (S_n - S_j)]^2 I_{E_j} \\
&= \sum_{j=1}^n ES_j^2 I_{E_j} + \sum_{j=1}^n E(S_n - S_j)^2 I_{E_j} + 2\sum_{j=1}^n ES_j(S_n - S_j) I_{E_j} \\
&\ge \sum_{j=1}^n ES_j^2 I_{E_j} \\
&\quad \text{(as } E[S_j I_{E_j}(S_n - S_j)] = E[S_j I_{E_j}]E[(S_n - S_j)] = 0, \\
&\quad \text{since } S_j I_{E_j} \text{ and } S_k - S_j \text{ are independent)} \\
&\ge \sum_{j=1}^n \epsilon^2 P(E_j) \qquad (\text{as } |c_j S_j| > \epsilon \text{ on } E_j) \\
&= \epsilon^2 P(A) \\
&= \epsilon^2 P\left(\max_{1\le k\le n} |S_k| \ge \epsilon\right).
\end{aligned}$$

(b). From (a), we have

$$\begin{aligned}
ES_n^2 I_A &= \sum_{j=1}^n ES_j^2 I_{E_j} + \sum_{j=1}^n E(S_n - S_j)^2 I_{E_j} \\
&\le \sum_{j=1}^n E(|S_{j-1}| + C)^2 I_{E_j} + \sum_{j=1}^n \sum_{k=j+1}^n (EX_k^2) P(E_j)
\end{aligned}$$

$$
\begin{aligned}
& \quad \text{(since } (S_n - S_j) \text{ and } I_{E_j} \text{ are independent.)} \\
&\le \sum_{j=1}^{n} E(\epsilon + C)^2 I_{E_j} + \left(ES_n^2\right) \sum_{j=1}^{n} P(E_j) \\
&= \left((\epsilon + C)^2 + ES_n^2\right) P(A).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
ES_n^2 I_A &= ES_n^2 - ES_n^2 I_{A^c} \\
&\ge ES_n^2 - \epsilon^2 P(A^c) \qquad \text{as } A^c = \{\{\max_{1\le k\le n} |S_k|\} < \epsilon\} \\
&= ES_n^2 - \epsilon^2 + \epsilon^2 P(A).
\end{aligned}
$$

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 **8.1.1** *Let $X_1, X_2, \ldots$ be independent with $EX_k = 0$ and $\sigma_k^2 = Var(X_i) < \infty$. Write $S_k = \sum_{i=1}^{k} X_i$. If $|X_k| \le C \le \infty$, then $\forall k \ge 1$ and $\epsilon > 0$,*

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

**Remark.** Chebyshev inequality is a special case of Kolmogorov maximal inequality by taking $n = 1$.

$$
P\left(|X - \mu| \ge \epsilon\right) \le \frac{E(X - \mu)^2}{\epsilon^2}
$$

## 8.2 The a.s. convergence of series; three-series theorem

**Definition:** $\sum_{n=1}^{\infty} a_n$ 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.

### 8.2.1 Review: Cauchy convergent a.s. or in probability

**Definition:**

1. The sequence $\{X_n, n \ge 1\}$ is almost sure (a.s.) Cauchy convergent
$\iff P(\lim_{m,n\to\infty} |X_m - X_n| = 0) = 1$;
$\iff \forall \epsilon > 0$: $\lim_{M\to\infty} P(\sup_{m,n\ge M} |X_m - X_n| \le \epsilon) = 1$;
$\iff \forall \epsilon > 0$: $\lim_{M\to\infty} P(\sup_{m,n\ge M} |X_m - X_n| > \epsilon) = 0$;
$\iff \sup_{m,n\ge M} |X_m - X_n| \longrightarrow_p 0$ as $M \to \infty$;
$\iff \sup_{m>n} |X_m - X_n| \longrightarrow_p 0$ as $n \to \infty$;
$\iff \sup_{m>n} |X_m - X_n| = o_p(1)$ as $n \to \infty$.

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

$\iff \forall \epsilon > 0: \ \lim_{m,n\to\infty} P(|X_m - X_n| \le \epsilon) = 1;$
$\iff \forall \epsilon > 0: \ \lim_{m,n\to\infty} P(|X_m - X_n| > \epsilon) = 0;$
$\iff \forall \epsilon > 0: \ \lim_{n\to\infty} \sup_{m>n} P(|X_m - X_n| > \epsilon) = 0;$
$\iff \forall \epsilon > 0: \ \sup_{m>n} P(|X_m - X_n| > \epsilon) = o(1)$ as $n \to \infty$.

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

Theorem **8.2.1** *$X_n \to X$ a.s. $\iff$ $X_n$ 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,

$$0 \le |X_n(\omega) - X_m(\omega)| \le |X_n(\omega) - X(\omega)| + |X_n(\omega) - X(\omega)| \to 0,$$

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 **8.2.2** *$X_n \to X$ in probability $\iff$ $\{X_n\}$ is Cauchy convergent in probability.*

*Proof.* "$\Longrightarrow$". $\forall \epsilon > 0 : \lim_n P(|X_n - X| > \epsilon) = 0$. Therefore,

$$0 \le P(|X_n - X_m| > 2\epsilon) \le P(|X_n - X| > \epsilon) + P(|X_m - X| > \epsilon) \to 0.$$

i.e. $X_n$ 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 $k \ge 1$, $\exists$ an integer $m_k$ such that

$$P(|X_m - X_n| > 2^{-k}) \le 2^{-k}, \qquad \text{for all } m > n \ge m_k.$$

W.L.O.G, we can assume that $m_k$ is strictly increasing sequence. Then setting,

$$Y_k := X_{n_k}, \qquad A_k := \{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\} = \{|Y_{k+1} - Y_k| > 2^{-k}\},$$

we have

$$P(A_k) := P(|Y_{k+1} - Y_k| > 2^{-k}) \le 2^{-k}, \qquad \text{for all } m > n \ge m_k.$$

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 $\omega$-set $N$ of measure zero, we have

$$|Y_{k+1}(\omega) - Y_k(\omega)| \le 2^{-k},$$

provided $k \ge$ some integer $k_0(\omega)$. Hence, for $\omega \in N^c$, as $n \to \infty$, we have

$$\sup_{m>n} |Y_m(\omega) - Y_n(\omega)| \le \sum_{k=n}^{\infty} |Y_{k+1}(\omega) - Y_k(\omega)| \le \sum_{k=n}^{\infty} 2^{-k} = 2^{-(n-1)} \to 0.$$

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 $k \to \infty$,

$$P\left(|X_k - X| \ge 2\epsilon\right) \le P\left(|X_k - X_{n_{k+1}}| \ge \epsilon\right) + P\left(|X_{n_{k+1}} - X| \ge \epsilon\right) = o(1).$$

So $X_k \to_p X$. ∎

THEOREM **8.2.3** *$X_n \to X$ 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,

$$\left(E|X_m - X_n|^2\right)^{1/2} \le \left(E|X_m - X|^2\right)^{1/2} + \left(E|X_n - X|^2\right)^{1/2} \to 0,$$

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 $X_\infty$. It follows that there exists a subsequence $\{X_{n_k}, k \ge 1\}$ which converges to $X_\infty$ almost surely. Now as $n \to \infty$,

$$\begin{aligned} E|X_n - X_\infty|^2 &= E \lim_{k\to\infty} |X_n - X_{n_k}|^2 = E \liminf_{k\to\infty} |X_n - X_{n_k}|^2 \\ &\le \liminf_{k\to\infty} E|X_n - X_{n_k}|^2 \qquad \text{(Fatou's lemma)} \\ &\to 0 \qquad \text{(Cauchy convergent in } L_2\text{)}. \end{aligned}$$

Therefore, $X_n \to X_\infty$ in $L_2$. ∎

**Remark.** Summarizing some of these theorems, we note that

$$X_n \to X \ a.s. \qquad \Longleftrightarrow \qquad \sup_{m>n} |X_m - X_n| = o_p(1), \ \text{as } n \to \infty$$

$$X_n \to_p X \qquad \Longleftrightarrow \qquad \sup_{m>n} P(|X_m - X_n| \ge \epsilon) = o(1), \ \text{as } n \to \infty.$$

### 8.2.2 Variance criterion for random series

THEOREM **8.2.4 (Variance criterion for series, due to Khinchin and Kolmogorov)** *Let $X_1, X_2, ...$ be independent with $EX_k = 0$ and $\sigma_k^2 = Var(X_k) < \infty$. If $\sum_{k=1}^{\infty} Var(X_k) < \infty$, then $\sum_{k=1}^{\infty} X_k$ converges a.s.*

(i.e., $L_2$ convergence implies a.s. convergence for independent series)

*Proof.* 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,

$$P\left(\left\{\max_{\{m: M<m\le n\}} |S_m - S_M|\right\} > \epsilon\right) \le \frac{Var(S_m - S_M)}{\epsilon^2} = \frac{\sum_{k=M+1}^{n} Var(X_k)}{\epsilon^2}$$

Let $n \to \infty$,

$$P\left(\left\{\sup_{\{m: m>M\}} |S_m - S_M|\right\} > \epsilon\right) \le \frac{\sum_{k=M+1}^{\infty} Var(X_k)}{\epsilon^2}.$$

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

$$\begin{aligned} P\left(\left\{\sup_{m,n>M} |S_m - S_n|\right\} > 2\epsilon\right) &\le P\left(\sup_{m>M} |S_m - S_M| > \epsilon\right) + P\left(\sup_{n>M} |S_n - S_M| > \epsilon\right) \\ &= 2P\left(\sup_{m>M} |S_m - S_M| > \epsilon\right) \to 0, \qquad \text{as } M \to \infty. \end{aligned}$$

Therefore, $S_n$ is a Cauchy sequence a.s., hence $\lim_{n\to\infty} S_n$ exists. ∎

**Method 2: Subsequence method.** Write $S_n = \sum_{k=1}^n X_k$. By Chebyshve's inequality, for $m < n$,

$$P\left(|S_n - S_m| > \epsilon\}\right) \le \frac{\sum_{k=m+1}^n Var(X_k)}{\epsilon^2} \to 0, \qquad \text{as } m \to \infty.$$

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 $\forall n \ge 1$, $\exists k \ge 1$, such that $n_k < n \le n_{k+1}$, and

$$\begin{aligned} 0 &\le |S_n - S_\infty| \le |S_{n_{k+1}} - S_\infty| + |S_{n_{k+1}} - S_n| \\ &\le |S_{n_{k+1}} - S_\infty| + \max_{n_k < j \le n_{k+1}} |S_{n_{k+1}} - S_j| \\ &=: A_k + B_k. \end{aligned}$$

We have shown that $A_k \to 0$ a.s. To show $B_k \to 0$ a.s., we apply Kolmogorov inequality,

$$\sum_{k=1}^\infty P\left(|B_k| \ge \epsilon\right) \le \sum_{k=1}^\infty \frac{1}{\epsilon^2} E\left(S_{n_{k+1}} - S_{n_k}\right)^2 = \sum_{k=1}^\infty \frac{1}{\epsilon^2} \sum_{j=n_k+1}^{n_{k+1}} EX_j^2 \le \frac{\sum_{i=1}^\infty EX_i^2}{\epsilon^2} < \infty.$$

which implies that $B_k \to 0$ a.s. Hence, $S_n \to S_\infty$ a.s. ∎

COROLLARY **8.2.1 (Kolmogorov SLLN)** *Let $X_1, X_2, ...$ be independent with $\mu_k = EX_k$ and $\sigma_k^2 = EX_k^2 < \infty$. Let $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $\bar{\mu} = n^{-1}\sum_{i=1}^n \mu_i$ If $\sum_{k=1}^\infty EX_k^2/k^2 < \infty$, then $\bar{X} - \bar{\mu} \to 0$ a.s.*

*Proof.* **Method 1.** Use Hajek-Renyi's inequality and Kronecker Lemma.

**Method 2.** Let $Y_k = (X_k - EX_k)/k$. Since $\sum_{k=1}^\infty var(Y_k) < \infty$, from the last theorem, we get $\sum_{k=1}^\infty (X_k - EX_k)/k$ converges a.s. By the Kronecker Lemma, $n^{-1}\sum_{k=1}^n (X_k - EX_k) \to 0$ a.s. ∎

**Remark:** More general theorems like the last one will be given later.

**Remark:** Let $X, X_1, X_2, ...$ be i.i.d. r.v.'s with $\mu = EX$.

(1) If $EX^4 < \infty$, by Chebyshev's inequality, we can show $\bar{X} \to \mu$ a.s.

(2) If $EX^2 < \infty$, by Kolmogorov's SLLN (Theorem 8.2.1) or more directly Hajek-Renyi's inequality, we can show that $\bar{X} \to \mu$ 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 $\bar{X} \to \mu$ a.s.

*It can be seen that the weaker the condition, the more sofisticated techniques will be required.*

### 8.2.3 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 **8.2.5 (Kolmogorov three series theorem)** *Let $X_1, X_2, ...$ be independent r.v.'s. Let*

$$\begin{aligned} Y_n = X_n I_{\{|X_n| \le A\}} &= X_n, \quad |X_n| \le A \\ & \quad\ 0, \quad\ \ |X_n| > A. \end{aligned}$$

*Then $\sum_{k=1}^\infty X_k$ a.s. $\iff$ for some $A > 0$, the following three series converge:*

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

(*ii*) $\sum_{n=1}^{\infty} EY_n$ *converges;*

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

*Proof.*

"$\Longleftarrow$" Assume that the three series converge.

$\Longrightarrow$ $\sum_{n=1}^{\infty} E(Y_n - EY_n)^2 < \infty$ [from (iii)].

$\Longrightarrow$ $\sum_{n=1}^{\infty}(Y_n - EY_n)$ converges a.s. (by Theorem 8.2.4, the variance criterion)

$\Longrightarrow$ $\sum_{n=1}^{\infty} Y_n$ converges a.s. [from (ii)].

$\Longrightarrow$ $\sum_{n=1}^{\infty} X_n$ converges a.s. [as $\{X_n\}$ and $\{Y_n\}$ are equivalent from (i)]

"$\Longrightarrow$" Assume that $\sum_{n=1}^{\infty} X_n$ converges a.s.

**Proof of (i)**. Clearly, the assumption $\Longrightarrow X_n = S_n - S_{n-1} \to 0$ a.s. $\Longrightarrow P(|X_n| > A, i.o.) = 0$, $\forall A > 0$. $\Longleftrightarrow \sum_{n=1}^{\infty} P(|X_n| > A) < \infty$, from Borel-Cantelli Lemma for independent r.v.'s $\{X_n\}$. Thus, (i) holds.

**Proof of (iii)**. From (i), $\{X_n\}$ and $\{Y_n\}$ are equivalent. $\Longrightarrow \sum_{n=1}^{\infty} Y_n$ converges a.s. But since $|Y_n - EY_n| \le 2A$, the applying Kolmogorov maximal inequality, we get

$$P\left(\max_{n\le j\le r}\left|\sum_{j=n}^{j} Y_j\right| \le \epsilon\right) \le \frac{(\epsilon+2A)^2}{Var\left(\sum_{j=n}^{r} Y_j\right)} = \frac{(\epsilon+2A)^2}{\sum_{j=n}^{r} Var(Y_j)}, \qquad \forall \epsilon > 0.$$

Were the series (iii) to diverge, then $\sum_{j=n}^{r} Var(Y_j) \to \infty$ as $r \to \infty$ for each fixed $n$. Thus, letting $r \to \infty$, we get

$$0 \le P\left(\sup_{n}\left|\sum_{j=n}^{\infty} Y_j\right| \le \epsilon\right) \le \frac{(\epsilon+2A)^2}{\sum_{j=n}^{\infty} Var(Y_j)} = 0, \qquad \forall \epsilon > 0.$$

or equivalently,

$$P\left(\sup_{n}\left|\sum_{j=n}^{\infty} Y_j\right| \ge \epsilon\right) = 1.$$

That is, the tail of $\sum_{j=1}^{\infty} Y_j$ almost surely would not be bounded by any fixed constant $\epsilon$, so that the series could not converge almost surely. This contradiction proves that (iii) must converge a.s.

**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). $\sum_{n=1}^{\infty}(Y_n - EY_n)$ converges a.s., which follows from (iii).

Thus,

$$\sum_{n=1}^{\infty} EY_n = \sum_{n=1}^{\infty} Y_n - \sum_{n=1}^{\infty}(Y_n - EY_n) \text{ converges a.s.}$$

This proves (ii).

REMARK **8.2.1** *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**.

Remark 8.2.2 *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

$$X_n \to 0, a.s. \quad \Longleftrightarrow \quad P(|X_n| \ge \epsilon, i.o.) = 0, \quad \forall \epsilon > 0,$$
$$\Longleftrightarrow \quad \sum_{n=1}^{\infty} P(|X_n| \ge \epsilon) < \infty, \quad \forall \epsilon > 0. \quad \blacksquare$$

### 8.2.4 Kolmogorov two series theorem for absolute random series

To prove a.s absolute convergence, we have Kolmogorov two series theorem.

Theorem 8.2.6 **(Kolmogorov two series theorem for a.s absolute convergence)** *$\{X_n\}$ are independent. Then $\sum_{n=1}^{\infty} |X_n|$ converges a.s. iff*

$$\sum_{n=1}^{\infty} P(|X_n| \ge C) < \infty; \qquad \sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} < \infty. \tag{2.1}$$

*Proof.* By Kolmogorov three series theorem, we only need to show that the conditions in (2.1) imply that $\sum_{n=1}^{\infty} EX_n^2 I_{\{|X_n|<C\}} < \infty$, which is true since

$$\sum_{n=1}^{\infty} EX_n^2 I_{\{|X_n|<C\}} \le C \sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} < \infty.$$

Theorem 8.2.7 **(mean convergence implies a.s. convergence)** *$\{X_n\}$ are independent. Then $\sum_{n=1}^{\infty} E|X_n|^r < \infty$ $(0 < r \le 1)$ implies that $\sum_{n=1}^{\infty} |X_n|$ converges a.s.*

*Proof.* Apply the last result,

(i) $\sum_{n=1}^{\infty} P(|X_n| \ge C) \le \sum_{n=1}^{\infty} E|X_n|^r / C^r < \infty$.

(ii) $\sum_{n=1}^{\infty} E|X_n| I_{\{|X_n|<C\}} \le C^{1-r} \sum_{n=1}^{\infty} E|X_n|^r I_{\{|X_n|<C\}} \le C^{1-r} \sum_{n=1}^{\infty} E|X_n|^r < \infty$

**Remark:** 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 $E|X_n| \le E|X_n|^r$ for large $n$. Therefore, we can replace the condition $\sum_{n=1}^{\infty} E|X_n|^r < \infty$ $(0 < r \le 1)$ by $\sum_{n=1}^{\infty} E|X_n| < \infty$. Does the theorem still hold for $r > 1$?

**Remark:** In fact, **we can even remove the independence condition here, which of course will require different proof.**

Theorem 8.2.8 *If $\{X_n, n \ge 1\}$ is a sequence of nonnegative, integrable r.v.'s with $S_n = X_1 + ... + X_n$, 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 $m \ge n$,

$$P\left(|S_m - S_n| > \epsilon\right) \le \frac{1}{\epsilon} E|S_m - S_n| = \frac{1}{\epsilon} \sum_{k=n+1}^{m} EX_k \to 0$$

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

$$S_{n_k} \le S_n \le S_{n_{k+1}}.$$

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.

## 8.3 Strong Laws of Large Numbers (SLLN)

**Definition:** A sequence of r.v.'s $X_1, X_2, ...$ 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.

### 8.3.1 Several useful lemmas

LEMMA **8.3.1 (Cesaro's Lemma)** . *Given two sequences $\{b_n\}$ and $\{x_n\}$, assume that*

*(i) $b_n \geq 0$ and $a_n = \sum_{k=1}^n b_k \nearrow \infty$,* *(ii) $x_n \to x$, $|x| < \infty$.*

*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,

$$\begin{aligned}
\left|\frac{1}{a_n}\sum_{k=1}^n b_k x_k - x\right| &= \left|\frac{\sum_{k=1}^n b_k(x_k - x)}{\sum_{k=1}^n b_k}\right| \leq \frac{\sum_{k=1}^{n_0} b_k|x_k - x|}{\sum_{k=1}^n b_k} + \frac{\sum_{k=n_0}^n b_k|x_k - x|}{\sum_{k=1}^n b_k} \\
&\leq \frac{\sum_{k=1}^{n_0} b_k|x_k - x|}{\sum_{k=1}^n b_k} + \frac{\sum_{k=n_0}^n b_k\epsilon}{\sum_{k=1}^n b_k} \leq \frac{C(n_0)}{a_n} + \epsilon.
\end{aligned}$$

Letting $n \to \infty$, and noting $a_n \to \infty$, we obtain the theorem. ∎

COROLLARY **8.3.1** *If $x_n \to x$ (finite), then $\bar{x} = n^{-1}\sum_{k=1}^n x_k \to x$.*

LEMMA **8.3.2 (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.,*

$$\sum_{k=1}^n a_k \Delta S_k = a_n S_n - \sum_{k=1}^n S_k \Delta a_k.$$

(*Compare with $\int_0^n f(s)ds = sf(s)|_0^n - \int_0^n sdf(s)$* )

*Proof.*

$$\begin{aligned}
\sum_{k=1}^n a_k x_k &= \sum_{k=1}^n a_k\left(S_k - S_{k-1}\right) = \sum_{k=1}^n a_k S_k - \sum_{k=1}^n a_k S_{k-1} \\
&= \sum_{k=0}^n a_k S_k - \sum_{k=0}^{n-1} a_{k+1} S_k = \left(a_n S_n + \sum_{k=0}^{n-1} a_k S_k\right) - \sum_{k=0}^{n-1} a_{k+1} S_k \\
&= a_n S_n - \sum_{k=0}^{n-1}\left(a_{k+1} - a_k\right) S_k = a_n S_n - \sum_{k=1}^n \left(a_k - a_{k-1}\right) S_{k-1}.
\end{aligned}$$ ∎

LEMMA **8.3.3 (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 **8.3.2 (Kronecker lemma)** *If $a_n \nearrow \infty$ and $\sum_{n=1}^{\infty}\frac{y_n}{a_n}$ converges, then*

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

### 8.3.2 SLLN for independent r.v.'s

**SLLN for independent r.v.'s: Sufficient moment conditions**

All theorems in this section point to the fact: **For independent r.v.'s, moment convergence implies a.s. convergence.**

There are other interesting facts associated with independent r.v.'s. For instance, convergence in probability is equivalent to convergence a.s. (See Loeve).

The following theorem is very useful in the following sections. Its proof is based on the three series theorem.

THEOREM **8.3.1** *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:*

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

*(ii)* $\dfrac{x}{g_n(x)} \searrow$ *and* $\frac{x^2}{g_n(x)} \nearrow$ *for $x > 0$;* $\quad EX_n = 0$.

*(iii)* $\dfrac{x^2}{g_n(x)} \nearrow$ *for $x > 0$; $X_n$ has a symmetric d.f. about 0.*

*(2). $\{a_n\}$ is a positive sequence, and*

$$\sum_{n=1}^{\infty}\frac{Eg_n(X_n)}{g_n(a_n)} < \infty. \tag{3.2}$$

*Then we have*

$$\sum_{n=1}^{\infty}\frac{X_n}{a_n} \quad \text{converges a.s.} \tag{3.3}$$

*If we further assume that* $0 < a_n \nearrow \infty$, *then*

$$\frac{1}{a_n}\sum_{j=1}^{n} X_j \to 0 \qquad a.s. \tag{3.4}$$

**Proof.** The proof of (3.4) follows from (3.3) and Kronecker's Lemma. Hence, we shall prove (3.3) next. Let $F_n(x) = P(X_n \le x)$, 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

$$(a).\ \sum_{n=1}^{\infty} P\left(\frac{|X_n|}{a_n} \ge 1\right), \qquad (b).\ \sum_{n=1}^{\infty} E\left(\frac{Y_n}{a_n}\right), \qquad (c).\ \sum_{n=1}^{\infty} E\left(\frac{Y_n^2}{a_n^2}\right).$$

(*Note that* (*c*) *implies that* $\sum_{n=1}^{\infty} Var\,(Y_n/a_n) < \infty$.)

**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 (3.2), we get

$$\sum_{n=1}^{\infty} P(|X_n| \ge a_n) \le \sum_{n=1}^{\infty} P\left(g(X_n) \ge g(a_n)\right) \le \sum_{n=1}^{\infty} \frac{Eg(X_n)}{g(a_n)} < \infty.$$

**Proof of (c)**. If $|X_n| < a_n$, then clearly $g(X_n) \le g(a_n)$; also we can show that

$$\frac{X_n^2}{a_n^2} \le \frac{g_n(X_n)}{g_n(a_n)}. \tag{3.5}$$

**Proof of (3.5)**. We shall look at it under assumptions (i)–(iii) separately.

If (i) holds, then $|X_n| < a_n$ implies

$$\frac{|X_n|}{g_n(X_n)} \le \frac{a_n}{g_n(a_n)}, \implies \frac{X_n^2}{a_n^2} \le \frac{g_n^2(X_n)}{g_n^2(a_n)} \le \frac{g_n(X_n)}{g_n(a_n)}, \text{ (as } \tfrac{g_n(X_n)}{g_n(a_n)} \le 1). \tag{3.6}$$

If (ii) or (iii) holds, then $\frac{x^2}{g_n(x)} \nearrow$ for $x > 0$. Thus, $|X_n| < a_n$ implies

$$\frac{X_n^2}{g_n(X_n)} \le \frac{a_n^2}{g_n(a_n)}, \qquad \implies \qquad \frac{X_n^2}{a_n^2} \le \frac{g_n(X_n)}{g_n(a_n)}.$$

This proves (3.5).

Now it follows easily from (3.5) that

$$\sum_{n=1}^{\infty} \frac{EY_n^2}{a_n^2} \le \sum_{n=1}^{\infty} E\left(\frac{X_n^2}{a_n^2} I_{\{|X_n|<a_n\}}\right) \le \sum_{n=1}^{\infty} E\left(\frac{g(X_n)}{g(a_n)} I_{\{|X_n|<a_n\}}\right) \le \sum_{n=1}^{\infty} E\left(\frac{g(X_n)}{g(a_n)}\right) < \infty.$$

**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 (3.6) that

$$\left|\sum_{n=1}^{\infty} \frac{EY_n}{a_n}\right| \le \sum_{n=1}^{\infty} \frac{Eg_n(Y_n)}{g_n(a_n)}.$$

Assume now that (ii) holds. If $|X_n| \ge a_n$, then

$$\frac{|X_n|}{g_n(X_n)} \le \frac{a_n}{g_n(a_n)}, \qquad \implies \qquad \frac{|X_n|}{a_n} \le \frac{g_n(X_n)}{g_n(a_n)}.$$

But from $EX_n = 0$,

$$\left|\sum_{n=1}^{\infty} \frac{EY_n}{a_n}\right| \;=\; \left|\sum_{n=1}^{\infty} E\frac{X_n}{a_n} I_{\{|X_n|<a_n\}}\right| = \left|-\sum_{n=1}^{\infty} E\frac{X_n}{a_n} I_{\{|X_n|\ge a_n\}}\right| \le \sum_{n=1}^{\infty} \frac{Eg_n(Y_n)}{g_n(a_n)} < \infty.$$

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 **8.3.3** *$\{X_n\}$ are independent r.v.'s, and $0 < a_n \nearrow \infty$. Assume that*

$$\sum_{n=1}^{\infty} E\left|\frac{X_n}{a_n}\right|^r = \sum_{n=1}^{\infty} \frac{E|X_n|^r}{a_n^r} < \infty, \qquad 0 < r \le 2. \tag{3.7}$$

*Then, we have*

$$\begin{aligned} \frac{1}{a_n}\sum_{j=1}^{n} X_j &\to 0 \qquad a.s. \quad if\ \ 0 < r \le 1; \\ \frac{1}{a_n}\sum_{j=1}^{n}(X_j - EX_j) &\to 0 \qquad a.s. \quad if\ \ 1 \le r \le 2. \end{aligned}$$

*Proof.* Take $g_n(x) \equiv |x|^r$.

**Case I**: If $0 < r \le 1$, then (3.7) is equivalent to (3.2). From the last theorem, we get $\sum_{n=1}^{\infty} X_n/a_n$ converges a.s. and then apply Kronecker's Lemma.

**Case II**: If $1 \le r \le 2$, then (3.7) implies that

$$\begin{aligned} &\sum_{n=1}^{\infty} \frac{Eg_n(X_n - EX_n)}{g_n(a_n)} = \sum_{n=1}^{\infty} \frac{E|X_n - EX_n|^r}{a_n^r} \\ \le\ & C_r\sum_{n=1}^{\infty}\frac{E|X_n|^r}{a_n^r} + C_r\sum_{n=1}^{\infty}\frac{|EX_n|^r}{a_n^r} \qquad \text{(by } C_r\text{-inequality)} \\ \le\ & 2C_r\sum_{n=1}^{\infty}\frac{E|X_n|^r}{a_n^r} \qquad\qquad \text{(as } |EX_n|^r \le E|X_n|^r,\ r \ge 1) \\ <\ & \infty. \end{aligned}$$

From the last theorem, we get $\sum_{n=1}^{\infty}(X_n - EX_n)/a_n$ converges a.s. and then apply Kronecker's Lemma. ∎

**Remark.** 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 $\frac{1}{a_n}\sum_{k=1}^{n} EX_k \to 0$.

**Remark.** 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 **8.3.2** *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 $1 \le k \le n$. 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*

$$\sum_{n=1}^{\infty} P(|X_n| \ge a_n) < \infty, \qquad \textit{and} \qquad \sum_{k=1}^{n} EY_{nk} \longrightarrow 0.$$

**Proof.** See Petrov, 1995, page 211, Theorem 6.8.

Applying Theorem 8.3.2 to i.i.d. r.v.'s, we get

THEOREM **8.3.3** *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*

$$\sum_{k=n}^{\infty} 1/a_k^2 = O(n/a_n^2).$$

*Then the relation* $\frac{1}{a_n}\sum_{j=1}^{n} X_j \to 0$ *a.s. if and only if*

$$\sum_{n=1}^{\infty} P(|X_n| \ge a_n) < \infty, \qquad \textit{and} \qquad nE\left(\frac{X_1}{a_n}\right) I\{|X_1| < a_n\} \longrightarrow 0.$$

**Proof.** See Petrov, 1995, page 212, Theorem 6.9.

### 8.3.3 SLLN for i.i.d. r.v.'s: necessary and sufficient moment conditions

**Kolmogorov SLLN for i.i.d. r.v.'s**

THEOREM **8.3.4 (Kolmogorov SLLN for iid r.v.'s)** *Let* $X_1, X_2, ...$ *be i.i.d. r.v.'s, and* $S_n = \sum_{k=1}^{n} X_k$*. Then*

$$(i) \quad E|X_1| < \infty \qquad \Longrightarrow \qquad \frac{S_n}{n} \to EX_1 \quad a.s. \tag{3.8}$$

$$(ii) \quad E|X_1| = \infty \qquad \Longrightarrow \qquad \limsup_n \frac{|S_n|}{n} = \infty \quad a.s. \tag{3.9}$$

**Proof.** (i). Assume $E|X_1| < \infty$. Write $Y_n = X_n I_{\{|X_n| \le n\}}$. Clearly,

$$\sum_{n=1}^{\infty} P\left(|X_n| \ne Y_n\right) = \sum_{n=1}^{\infty} P\left(|X_n| > n\right) = \sum_{n=1}^{\infty} P\left(|X_1| > n\right) \le E|X_1| < \infty.$$

Therefore, $\{X_n\}$ and $\{Y_n\}$ are equivalent sequences. So it suffices to show that

$$(a).\ \frac{1}{n}\sum_{j=1}^{n} EY_j \to EX_1, \qquad\qquad (b).\ \frac{1}{n}\sum_{j=1}^{n}(Y_j - EY_j) \longrightarrow 0 \quad a.s.$$

since these would imply

$$\frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{n}\sum_{j=1}^{n} EY_j + \frac{1}{n}\sum_{j=1}^{n}(Y_j - EY_j) \longrightarrow EX_1 \qquad a.s.$$

which in turn implies that $n^{-1}\sum_{j=1}^{n} X_j \to EX_1$.

**Proof of (a).** Now

$$\begin{aligned} EY_n &= EX_n I_{\{|X_n|\le n\}} = EX_1 I_{\{|X_1|\le n\}} = E(X_1^+ - X_1^-) I_{\{|X_1|\le n\}} \\ &= EX_1^+ I_{\{|X_1|\le n\}} - EX_1^+ I_{\{|X_1|\le n\}} \\ &\longrightarrow EX_1^+ - EX_1^- = EX_1, \end{aligned}$$

where in the last line we have used the Monotone Convergence Theorem or the Dominated Convergence Theorem. Thus, $n^{-1}\sum_{j=1}^n EY_j \to EX_1$.

**Proof of (b).** Applying Corollary 8.3.3 with $a_n = n$ to $\{Y_n\}$, we get

$$\begin{aligned} \sum_{n=1}^\infty \frac{EY_n^2}{n^2} &= \sum_{n=1}^\infty \frac{1}{n^2} EX_n^2 I_{\{|X_n|\le n\}} \\ &= \sum_{n=1}^\infty \sum_{k=1}^n \frac{1}{n^2} E|X_1|^2 I_{\{k-1<|X_1|\le k\}} \\ &= \sum_{k=1}^\infty \sum_{n=k}^\infty \frac{1}{n^2} EX_1^2 I_{\{k-1<|X_1|\le k\}} \\ &= \sum_{k=1}^\infty \left[ E\left(X_1^2 I_{\{k-1<|X_1|\le k\}}\right) \left(\sum_{n=k}^\infty \frac{1}{n^2}\right)\right] \qquad (3.10) \\ &\le \sum_{k=1}^\infty \left[ kE\left(|X_1| I_{\{k-1<|X_1|\le k\}}\right) \left(\frac{C}{k}\right)\right] \\ &= C\sum_{k=1}^\infty E\left(|X_1| I_{\{k-1<|X_1|\le k\}}\right) \\ &\le CE|X_1| < \infty, \end{aligned}$$

alternatively, we could also carry on from (3.10) as follows

$$\begin{aligned} \sum_{n=1}^\infty \frac{EY_n^2}{n^2} &\le \sum_{k=1}^\infty \left[k^2 P\left(k-1<|X_1|\le k\right)\left(\frac{C}{k}\right)\right] \\ &\le C\sum_{k=1}^\infty kP\left(k-1<|X_1|\le k\right) \\ &\le C(1+E|X_1|) < \infty, \end{aligned}$$

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 8.3.3 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,

$$\begin{aligned} 1 &= P(|X_n| > An,\ i.o.), \qquad \text{(By Borel 0-1 law for independent r.v.'s)} \\ &= P(|S_n - S_{n-1}| > An,\ i.o.) \\ &= P(|S_n| > An/2,\ or\ |S_{n-1}| > A(n-1)/2,\ i.o.) \\ &\qquad \text{(as } |S_n - S_{n-1}| > An \Longrightarrow |S_n| > An/2, \text{ or } |S_{n-1}| > A(n-1)/2). \\ &= P(|S_n| > An/2,\ i.o.). \end{aligned}$$

Thus, for each $A>0$, there exists a null set $N(A)$ such that if $\omega \in \Omega - N(A)$, we have

$$\limsup_n \frac{|S_n|}{n} \ge \frac{A}{2}$$

Let $N = \bigcup_{m=1}^{\infty} N(m)$, which is also a null set. Then if $\omega \in N$, $\limsup_n \frac{S_n}{n} \geq \frac{A}{2}$ is still true for all $A$, and therefore the upper limit is $\infty$.

**Remark.** 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 **8.3.4** *Let $X_1, X_2, ...$ be i.i.d. r.v.'s, and $S_n = \sum_{k=1}^{n} X_k$. Then*

$$\bar{X} =: \frac{S_n}{n} \to C, \qquad a.s.$$

*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, $P(|X_n| \geq n, i.o.) = 0$. From the Borel 0-1 law for independent sequences, and identical distribution assumption, we get

$$\sum_{n=1}^{\infty} P\left(|X_n| \geq n\right) = \sum_{n=1}^{\infty} P\left(|X_1| \geq n\right) < \infty,$$

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.

## Marcinkiewicz SLLN for iid r.v.'s

THEOREM **8.3.5 (Marcinkiewicz SLLN for iid r.v.'s)** *Let $X_1, X_2, ...$ be i.i.d. r.v.'s, and $0 < r < 2$. Then*

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

*if and only if $E|X_1|^r < \infty$, where*

$$\begin{aligned} a &= EX_1 && \text{if } 1 \leq r < 2 \\ &= \text{arbitrary} && \text{if } 0 < r < 1. \text{ (Usually, simply chooses } a = 0). \end{aligned}$$

*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

$$\frac{X_n}{n^{1/r}} = \frac{a}{n^{1/r}} + \frac{S_n}{n^{1/r}} - \frac{S_{n-1}}{n^{1/r}} = \frac{a}{n^{1/r}} + \frac{S_n}{n^{1/r}} - \frac{S_{n-1}}{(n-1)^{1/r}}\frac{(n-1)^{1/r}}{n^{1/r}} \to 0, \quad a.s.$$

Thus, $P(|X_n| \geq n^{1/r}, i.o.) = 0$. 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 8.3.4. However, we shall write it down below for completeness.

Write $Y_n = X_n I_{\{|X_n| \le n^{1/r}\}}$. Clearly,

$$\sum_{n=1}^{\infty} P(X_n \ne Y_n) = \sum_{n=1}^{\infty} P\left(|X_n| > n^{1/r}\right) = \sum_{n=1}^{\infty} P(|X_1|^r > n) \le E|X_1|^r < \infty.$$

Therefore, $\{X_n\}$ and $\{Y_n\}$ are equivalent sequences. Therefore, it suffices to show that

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

**Case I: $r = 1$.** This is Theorem 8.3.4.

**Case II: $0 < r < 1$.**
Applying Corollary 8.3.3 with $a_n = n^{1/r}$ to $\{Y_n\}$, we get

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{E|Y_n|}{n^{1/r}} &= \sum_{n=1}^{\infty} \frac{1}{n^{1/r}} E|X_n| I_{\{|X_n| \le n^{1/r}\}} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \le k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \le k\}} \\
&= \sum_{k=1}^{\infty} \left[ E\left(|X_1| I_{\{k-1 < |X_1|^r \le k\}}\right) \left( \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \right) \right] \\
&\quad \text{(Note that the series } \textstyle\sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \text{ converges when } 0 < r < 1.) \\
&\le \sum_{k=1}^{\infty} \left[ k^{1/r} P\left(I_{\{k-1 < |X_1|^r \le k\}}\right) \left( \frac{C}{k^{1/r-1}} \right) \right] \\
&= C \sum_{k=1}^{\infty} k P\left(I_{\{k-1 < |X_1|^r \le k\}}\right) \\
&\le C\left(1 + E|X_1|^r\right) < \infty,
\end{aligned}$$

where we used the elementary estimate $\sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \le C/k^{1/r-1}$ for some $C > 0$ and all $k \ge 1$ when $0 < r < 1$.
It follows from Corollary 8.3.3 that

$$\frac{1}{n^{1/r}} \sum_{k=1}^{n} Y_k \to 0 \qquad a.s.$$

which in turn implies that

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

**Case III: $1 < r < 2$ with $a = EX_1$.**
Applying Corollary 8.3.3 with $a_n = n^{1/r}$ to $\{Y_n\}$,

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{EY_n^2}{n^{2/r}} &= \sum_{n=1}^{\infty} \frac{1}{n^{2/r}} EX_n^2 I_{\{|X_n| \le n^{1/r}\}} \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{n^{2/r}} EX_1^2 I_{\{k-1 < |X_1|^r \le k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^{2/r}} EX_1^2 I_{\{k-1 < |X_1|^r \le k\}} \\
&= \sum_{k=1}^{\infty} \left[ E\left(X_1^2 I_{\{k-1 < |X_1|^r \le k\}}\right) \left( \sum_{n=k}^{\infty} \frac{1}{n^{2/r}} \right) \right]
\end{aligned}$$

(Note that the series $\sum_{n=k}^{\infty} \frac{1}{n^{2/r}}$ converges when $1 < r < 2$.)

$$
\begin{aligned}
&\le \sum_{k=1}^{\infty} \left[ k^{2/r} P\left(k-1 < |X_1|^r \le k\right) \left( \frac{C}{k^{2/r-1}} \right) \right] \\
&= C \sum_{k=1}^{\infty} k P\left(k-1 < |X_1|^r \le k\right) \\
&\le C\left(1 + E|X_1|^r\right) < \infty,
\end{aligned}
$$

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 8.3.3.

**Proof of (ii) and (iii)**. Note that

$$
\begin{aligned}
\left| \sum_{n=1}^{\infty} \frac{EX_n - EY_n}{n^{1/r}} \right| &\le \sum_{n=1}^{\infty} \frac{E|X_n - Y_n|}{n^{1/r}} \\
&= \left| \sum_{n=1}^{\infty} \frac{1}{n^{1/r}} EX_n I_{\{|X_n| > n^{1/r}\}} \right| \le \sum_{n=1}^{\infty} \frac{1}{n^{1/r}} E|X_1| I_{\{|X_1|^r > n\}} \\
&\le \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \le k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} E|X_1| I_{\{k-1 < |X_1|^r \le k\}} \\
&\le \sum_{k=1}^{\infty} k^{1/r} P\left(k-1 < |X_1|^r \le k\right) \left( \sum_{n=k}^{\infty} \frac{1}{n^{1/r}} \right) \\
&\le \sum_{k=1}^{\infty} \left[ k^{2/r} P\left(k-1 < |X_1|^r \le k\right) \left( \frac{C}{k^{2/r-1}} \right) \right] \\
&= C \sum_{k=1}^{\infty} k P\left(k-1 < |X_1|^r \le k\right) \\
&\le C\left(1 + E|X_1|^r\right) < \infty.
\end{aligned}
$$

From Kronecker Lemma and Corollary 8.3.3, 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

$$
\frac{1}{n^{1/r}} \sum_{k=1}^{n} (X_k - EX_1) = \frac{1}{n^{1/r}} \sum_{k=1}^{n} (X_k - Y_k) + \frac{1}{n^{1/r}} \sum_{k=1}^{n} (Y_k - EY_k) + \frac{1}{n^{1/r}} \sum_{k=1}^{n} (EY_k - EX_k) \to 0 \quad a.s. \qquad \blacksquare
$$

## 8.4 Some applications of the WLLN and SLLN

See Chung, page 138.

THEOREM **8.4.1 (Glivenko and Cantelli)** *??????????????*

THEOREM **8.4.2 (S. Bernstein's proof of Weierstrass' theorem)** *??????????????*

## 8.5 Some results on LIL

1. Let $X_1, \ldots, X_n$ be independent r.v.'s with $EX_i = 0$ and $EX_i^2 < \infty$. Let $S_k = \sum_{i=1}^k X_k$ and $\sigma_n^2 = Var(S_n)$. Prove that for $x > 0$,

$$P\left(\max_{1\le k\le n} S_k \ge x\right) \le 2P(S_n > x - \sqrt{2}\sigma_n)$$

   and

$$P\left(\max_{1\le k\le n} |S_k| \ge x\right) \le 2P(|S_n| > x - \sqrt{2}\sigma_n)$$

2. (a) Prove that

$$e^x + e^{-x} \le 2e^{x^2/2}, \qquad \forall x \in R.$$

   (b) Let $X$ be a r.v. with $P(X = 1) = P(X = -1) = 1/2$. Prove that

$$Ee^{tX} \le e^{x^2/2},$$

   (c) Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with $P(X_i = 1) = P(X_i = -1) = 1/2$, and let $S_n = \sum_{i=1}^n X_i$. Prove that

$$P\left(|S_n| \ge x\sqrt{n}\right) \le e^{x^2/2}, \qquad \forall x \in R.$$

   (Hint: apply the Chebyshev inequality to $e^{tS_n}$)

3. Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with $P(X_i = 1) = P(X_i = -1) = 1/2$, and let $S_n = \sum_{i=1}^n X_i$. Prove that

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\log\log n}} \le 1 \quad a.s.$$

   (Hint: use results in Problems 1 and 2(c), the subsequence method and the Borel-Cantelli lemma.)

4. Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with the common d.f. $F$, and let $M_n = \max_{k\le n} X_k$. Assume that $F(x) = 1 - 1/x$ for $x \ge 1$.

   (i) Find the d.f. of $M_n$.

   (ii) Prove that $P(M_n/n \le y) \to \exp(-1/y)$.

   (iii) Prove that

$$\liminf_{n\to\infty} \frac{M_n}{n/\log n} \ge 1 \qquad a.s.$$

   (Hint: using the subsequence method, one can prove $\liminf_{n\to\infty} \frac{M_n}{n/\log\log n} \ge 1$ $a.s$)

## 8.6 Exercises

1. Let Let $X, X_1, X_2, ...$ be i.i.d. r.v.'s. Check if the WLLN and the SLLN hold if the common d.f. is given by

$$(i). \qquad P(X=n)=P(X=-n)=\frac{C}{2n^2\ln^2 n}, \qquad n\geq 3.$$

$$(ii). \qquad P(X=n)=P(X=-n)=\frac{C}{2n^2\ln n}, \qquad n\geq 3.$$

2. Let $X_1, X_2, X_3, ...$ be independent r.v.'s such that $P(X_1=0)=P(X_2=0)=1$ and

$$P(X_n=n)=P(X_n=-n)=\frac{1}{2n\ln n}, \quad P(X_n=0)=1-\frac{1}{n\ln n}, \qquad n\geq 3.$$

Show that this sequence obeys the WLLN but not the SLLN. In other words, $\bar{X}\to_p 0$, but $\bar{X}\not\to_{a.s.} 0$.

3. $\{X_n\}$ 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.

4. Let $\{X_n\}$ be i.i.d. r.v.'s, and $\{C_n, n\geq 1\}$ is a bounded sequence. Assume that $EX_1=0$. Show that

$$\frac{1}{n}\sum_{j=1}^{n} C_j X_j \longrightarrow 0 \qquad a.s.$$

(Hint: Use truncation.)

5. Show that if $X_1, X_2, ...$ are independent with $EX_n=0$ and

$$\sum_{n=1}^{\infty} E\left(X_n^2 I\{|X_n|\leq 1\}+|X_n|I\{|X_n|>1\}\right)<\infty,$$

then $\sum_{n=1}^{\infty} X_n$ converges a.s.

6. $X_1, X_2, ...$ are independent r.v.'s. Suppose that

$$\sum_{n=1}^{\infty} E|X_n|^{p_n}<\infty,$$

where $0<p_n\leq 2$ for all $n$ and $EX_n=0$ when $p_n>1$. Show that $\sum_{n=1}^{\infty} X_n$ converges a.s.

**The following are optional questions.**

7. Let $\{X_n\}$ be independent r.v.'s with $EX_n=0$ for all $n$, and $\displaystyle\sum_{1}^{\infty}\frac{E|X_n|^{2r}}{n^{r+1}}<\infty$ for some $r>1$, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \qquad a.s.$$

8. Let $\{X_n\}$ be independent r.v.'s with

$$P(X_n=n^\alpha)=P(X_n=-n^\alpha)=1/2, \qquad n=1,2,......$$

Show that $\{X_n\}$ satisfies the SLLN if and only if $\alpha<1/2$.

9. Let $\{X_n\}$ be a sequence of non-negative r.v.'s such that $\sup_n EX_n<\infty$, $EX_mX_n\leq EX_mEX_n$ for $m\neq n$, and

$$\sum_{n=1}^{\infty}\frac{var(X_n)}{n^2}<\infty.$$

Then,

$$\frac{S_n-ES_n}{n}\longrightarrow 0 \qquad a.s.$$

10. Let $\{X_n\}$ be a sequence of pairwise independent r.v.'s satisfying

$$\sum_{n=1}^{\infty} \frac{var(X_n)}{n^2} < \infty, \qquad \text{and} \qquad \frac{1}{n}\sum_{k=1}^{n} E|X_k - EX_k| = O(1).$$

Then,

$$\frac{S_n - ES_n}{n} \longrightarrow 0 \qquad a.s.$$

One can not omit the second part of the conditions in this proposition.

# Chapter 9

# Weak convergence

Weak convergence or convergence in distribution is one of the most useful concepts in statistical inference.

## 9.1 Definition

Definition **9.1.1**

> *(a) A sequence of d.f.s $\{F_n, n \geq 1\}$ is said to* **converge weakly** *to a d.f. $F$, written as $F_n \Longrightarrow F$, if $F_n(x) \longrightarrow F(x)$, for all $x \in C(F)$.*
>
> *(b) A sequence of random variables (r.v.s) $X_n$ is said to* **converge weakly** *or* **in distribution** *or* **in law** *to a limit $X$, written as $X_n \Longrightarrow X$ or $X_n \longrightarrow_d X$, if their d.f.s $F_n(x) = P(X_n \leq x)$ converge weakly $F(x) = P(X \leq x)$.*

Note that convergence in distribution is a property of distribution functions; the r.v.'s $Y_n$'s are not required to be on the same probability space. Furthermore, some authors define weak convergence slightly differently, and call the above definition *complete convergece.*

Example **9.1.1** *Let $X \sim F$ and $X_n = X + n^{-1}$. Then,*

$$F_n(x) = P(X + n^{-1} \leq x) = F(x - n^{-1}) \to F(x-).$$

*Thus, we observe the following.*

- **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 $F_n \Longrightarrow F$.*

- $\lim_{n\to\infty} F_n(x) = F(x-)$ **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.* ∎

## 9.2 Equivalent definitions of weak convergence

There are many definitions of weak convergence of a sequence of measures or r.v.s, some of which are more general than others. The following equivalence result is sometimes known as the "**portmanteau theorem**".

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

(a) $X_n \Longrightarrow X$ *or* $X_n \to_d X$.

(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* $\partial A$ *is the boundary of* $A$.

(*e*) $Eg(X_n) \to Eg(X)$ *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* $a, b \in C(F)$.

(*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.*

Remark **9.2.1**

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 of Theorem 9.2.1**.

- "(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
$$\liminf_n I_G(Y_n) \geq I_G(Y) \quad a.s. \tag{2.1}$$
This can be seen as follows.
(i) If $I_G(Y) = 0$, the proof is trivial;
(ii) If $I_G(Y) = 1$, $\Longleftrightarrow Y \in G$. Since $G$ is open, for large enough $n$, we have $Y_n \in G$, $\Longleftrightarrow$ $I_{G_n}(Y) = 1$. Hence, $LHS = 1 = I_G(Y) = RHS$.
Now applying Fatou's Lemma, we get
$$P(Y \in G) = EI_G(Y) \leq E \liminf_n I_G(Y_n) \leq \liminf_n EI_G(Y_n) = \liminf_n P(Y_n \in G).$$
- "(b) $\Longleftrightarrow$ (c)". $A$ is open $\Longleftrightarrow$ $A^c$ is closed, and $P(A) = 1 - P(A^c)$.
- "(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,
$$P(X \in A) = P(X \in \bar{A}) \geq \limsup_{n\to\infty} P(X_n \in \bar{A}) \geq \limsup_{n\to\infty} P(X_n \in A),$$
$$P(X \in A) = P(X \in A^0) \leq \liminf_{n\to\infty} P(X_n \in A^0) \leq \liminf_{n\to\infty} P(X_n \in A),$$
from which the proof follows.
- "(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 $P(X_n \leq x) = P(X_n \in A) \to P(X \in A) = P(X \leq x)$.
- "(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
$$Eg(X_n) = Eg(Y_n) \to Eg(Y) = Eg(X), \quad \blacksquare$$

For convenience, we give Skorokhod's representation theorem below.

THEOREM 9.2.2 **(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 $\{Y_n, n \geq 1\}$ on $((0,1), \mathcal{B}_{(0,1)}, P_\lambda = \lambda_{(0,1)})$ s.t.*

(1) $Y_n \sim F_n$, $Y \sim F$; (2) $Y_n \to Y$ *a.s. as* $n \to \infty$.

- "(e) $\Longrightarrow$ (a)". 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 $y > x + \epsilon$,

$$\begin{aligned} \limsup_{n\to\infty} P(X_n \leq x) = \limsup_{n\to\infty} EI\{X_n \leq x\} &\leq \limsup_{n\to\infty} Eg_{x,\epsilon}(X_n) = Eg_{x,\epsilon}(X) \\ &\leq EI\{X \leq x + \epsilon\} = P(X \leq x + \epsilon) \end{aligned}$$

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. ∎

- "$(e) \Longrightarrow (f)$". 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

$$\begin{aligned} g_1(x) &= g(x) && \text{if } x \notin (b, b+\delta) \\ & \quad\; g(b)\,[1 + (b-x)/\delta] && \text{if } x \in (b, b+\delta). \end{aligned}$$

Also define

$$\begin{aligned} g_2(x) &= g(b)\,[1 + (x-b)/\delta] && \text{if } x \in (b-\delta, b) \\ &= g(b)\,[1 + (b-x)/\delta] && \text{if } x \in [b, b+\delta) \\ &= 0 && \text{otherwise.} \end{aligned}$$

(Draw a picture here)

Now $|E[g(X_n) - g_1(X_n)]| \leq |Eg_2(X_n)|$ and $|E[g(X) - g_1(X)]| \leq |Eg_2(X)|$, so

$$\begin{aligned} |E[g(X_n) - g(X)]| &\leq |E[g(X_n) - g_1(X_n)]| + |E[g_1(X_n) - g_1(X)]| + |E[g_1(X) - g(X)]| \\ &\leq |Eg_2(X_n)| + |E[g_1(X_n) - g_1(X)]| + |Eg_2(X)| \\ &\to 2|Eg_2(X)| \quad \text{as } n \to \infty, \end{aligned}$$

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 $|g_2(x)| \leq |g(b)| I_{\{(b-\delta,b+\delta)\}}$, we have

$$|Eg_2(X)| \leq |g(b)| P\,(b - \delta < X < b + \delta) \to 0 \quad \text{as } \delta \downarrow 0$$

by the assumption that $b \in C$, i.e., $P(X = b) = 0$. Hence (f) holds.

- "$(f) \Longrightarrow (a)$". Suppose that (c) holds now. Let $b \in C$. Take $h(x) \equiv 1$ for all $x$, we have that, if $c \in C$,

$$\begin{aligned} P(X_n \leq b) &\geq P(a \leq X_n \leq b) \\ &\longrightarrow P(a \leq X \leq b) \quad \text{as } n \to \infty \\ &\longrightarrow P(X \leq b) \qquad \text{as } a \to -\infty \text{ through } C. \end{aligned}$$

A similar argument, but taking the limit in the other direction, yields for $b' \in C$

$$\begin{aligned} P(X_n \geq b') &\geq P(c \geq X_n \geq b') \quad \text{if } c \geq b' \\ &\longrightarrow P(c \geq X \geq b') \quad \text{as } n \to \infty \\ &\longrightarrow P(X \geq b') \qquad \text{as } c \to \infty \text{ through } C \end{aligned}$$

It follows that, if $b, b' \in C$ and $b < b'$, then for any $\epsilon > 0$, there exists $N$ such that

$$P(X \leq b) - \epsilon \leq P(X_n \leq b) \leq P(X_n \leq b') \leq P(X \leq b') + \epsilon$$

for all $n \geq N$. Now letting $n \to \infty$, $\epsilon \downarrow 0$, and $b' \downarrow b$ through $C$, we obtain that $F_{X_n}(b) \equiv P(X_n \leq b) \to P(X \leq b) \equiv F_X(b)$. This proves (a).

- "$(g) \Longleftrightarrow (a)$". This will be proved somewhere else. ∎

## 9.3 Helly's selection theorem and tightness

The next result is useful in studying limits of sequence of d.f.'s.

Theorem **9.3.1 (Helly's Selection Theorem)** *For every sequence of d.f.'s $F_n$, there exists a subsequence $F_{n_k}$ and a right continuous function $F$ so that*

$$\lim_{k\to\infty} F_{n_k}(x) = F(x), \qquad \text{for all } x \in C(F). \quad ∎$$

Lemma **9.3.1** *Suppose that $G$ is a bounded nondecreasing function of $D$, which is a dense subset of $(-\infty, \infty)$. Define*

$$F(x) = \lim_{y\in D, y\to x-} G(y), \qquad H(x) = \lim_{y\in D, y\to x+} G(y).$$

*Then,*

- *$F(x)$ (or $H(x)$) is a left (or right) continuous, nondecreasing function on $(-\infty, \infty)$ with $C(F) \supset C(G)$;*
- *$F(x) = G(x)$ (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 $F(x-) \leq a$, 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 of Theorem 9.3.1**: The proof uses a so-called "**diagonalisation method**". Let $r_1, r_2, ...$ be an enumeration of all the rational numbers, denoted by $D$. The 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

$$F_{11}, F_{12}, ..., F_{1k}, ... \qquad \text{converging at } r_1;$$

$$F_{21}, F_{22}, ..., F_{2k}, ... \qquad \text{converging at } r_1, r_2;$$

$$..... \; ...... \; ......$$

$$F_{j1}, F_{j2}, ..., F_{jk}, ... \qquad \text{converging at } r_1, ..., r_j;$$

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 $\lim_k F_{kk}(r_j) = l_j$, 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 $F_{n_k}(r_2) \to G(r_2) \geq F(r_2)$, 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,

$$|F_{n_k}(x) - F(x)| < \varepsilon. \quad \blacksquare$$

Remark **9.3.1**

(*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* $\infty$, *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* $x \in C(F)$.

(*ii*) *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 **9.3.1** *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*

$$\limsup_{n \to \infty} [1 - F_n(M) + F_n(-M)] \leq \epsilon, \qquad or \qquad \limsup_{n \to \infty} P(|X_n| > M) \leq \epsilon. \quad \blacksquare$$

That is, all of the probability measures give most of their mass to the same finite interval; mass does not "escape to infinity".

Theorem **9.3.2** *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

$$\begin{aligned}
1 - F(s) + F(r) &= \lim_{k\to\infty} (1 - F_{n_k}(s) + F_{n_k}(r)) \\
&\leq \limsup_{k\to\infty} (1 - F_{n_k}(M_\varepsilon) + F_{n_k}(-M_\varepsilon)) \\
&\leq \limsup_{n\to\infty} (1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)) \\
&\leq \varepsilon,
\end{aligned}$$

which in turn implies that

$$\lim_{x\to\infty} [1 - F(x) + F(-x)] \quad \leq \quad \varepsilon.$$

(LHS does have a limit since it is a non-increasing function of $x$ and also has lower bound 0.) Since $\varepsilon$ 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 $F_{n_{k_j}} \Longrightarrow_v F$. 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

$$\lim_{x\to\infty} [1 - F(x) + F(-x)] \quad \geq \quad \varepsilon_0.$$

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. ∎

The next corollary is an easy consequence of the above theorem.

COROLLARY **9.3.1** *A sequence of d.f.'s $\{F_n, n \geq 1\}$ converges vaguely to $F(x)$, denoted by $F_n \Longrightarrow_v F$. Then*

$$F \text{ is a d.f.} \Longleftrightarrow F_n \text{ is } \textbf{tight}. \quad ∎$$

The following sufficient condition for tightness is often useful.

THEOREM **9.3.3** *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

$$1 - F_n(M) + F_n(-M) \leq \frac{C}{\inf_{|x|\geq M} \psi(x)} \longrightarrow 0. \quad ∎$$

## 9.4 Polya Theorem

Pointwise weak convergence of $F_n$ to $F$ holds uniformly if $F$ is continuous.

THEOREM **9.4.1 (Polya Theorem)** *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

$$\lim_{t\to\infty}[1 - F_n(t)] = \lim_{t\to\infty}[1 - F(t)] = 0, \qquad \lim_{t\to-\infty} F_n(t) = \lim_{t\to-\infty} F(t) = 0.$$

For any $\epsilon > 0$, we can choose sufficiently large $M$ such that

$$\sup_{t\in(-\infty,M]} |F_n(t) - F(t)| \le \epsilon, \qquad \sup_{t\in[M,\infty]} |F_n(t) - F(t)| \le \epsilon.$$

Since $F$ is continuous, it is uniformly continuous on $[-M, M]$. So choose $n$ sufficiently large to get

$$\sup_{t\in M-M,M]} |F_n(t) - F(t)| \le \epsilon.$$

Combining the above results, for $n$ sufficiently large, we get

$$\sup_{t\in M-M,M]} |F_n(t) - F(t)| \le 3\epsilon.$$

This proves our theorem.

**Remark:** For another proof, see Petrov (1995).

## 9.5 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

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad x \in C(F),$$

where $C(F)$ is the continuity points of $F$. We shall write this as

$$Y_n \longrightarrow_d Y, \qquad \textit{or} \qquad F_n \Longrightarrow F.$$

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 **9.5.1** *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*

*(a) for all continuity points of $Y$ and all events $E \in \mathcal{F}$, $\lim_{n\to\infty} P(\{Y_n \le y\} \cap E) = Q_y(E)$ exists, and*

*(b) $Q_y(E) \to P(E)$ as $y \to \infty$.*

*We write this as*

$$Y_n \longrightarrow_d Y \quad (\textit{stably}), \quad \textit{or} \qquad F_n \Longrightarrow F \quad (\textit{stably}). \quad ∎$$

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 **9.5.1** *Let $X$ and $X'$ be i.i.d. non-degenerate r.v.'s. Let*

$$\begin{aligned} Z_n &= X \qquad \textit{for } n \textit{ odd} \\ &\quad\; X' \qquad \textit{for } n \textit{ even.} \end{aligned}$$

*Then we have $Z_n \longrightarrow_d X$, but we don't have $Z_n \longrightarrow_d X$ (stably).*

**Proof.** Now $Z_n \longrightarrow_d X$ holds since

$$\begin{aligned} P(Z_n \le x) &= P(X \le x) \qquad \text{for } n \text{ odd} \\ &\quad\; P(X' \le x) \qquad \text{for } n \text{ even} \\ &= P(X \le x) = F(x). \end{aligned}$$

To see why we don't have $Z_n \longrightarrow_d X$ (stably), take $E = \{X \le y\}$. Then,

$$\begin{aligned} P(Z_n \le x, E) &= P(X \le x, X \le y) \qquad \text{for } n \text{ odd} \\ &\quad\; P(X' \le x, X \le y) \qquad \text{for } n \text{ even} \\ &= P(X \le x \wedge y) \qquad \text{for } n \text{ odd} \\ &\quad\; P(X' \le x) P(X \le y) \qquad \text{for } n \text{ even} \\ &= F(x \wedge y) \qquad \text{for } n \text{ odd} \\ &\quad\; F(x)F(y) \qquad \text{for } n \text{ even.} \end{aligned}$$

Since $F(x \wedge y) > F(x)F(y)$ whenever $0 < F(x) \vee F(y) < 1$, the limit $P(Z_n \le x, E)$ 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.** A sequence $\{Z_n\}$ of $L_1$ r.v.s is said to converge **weakly in** $L_1$, to $Z$ if

$$\lim_{n\to\infty} E(Z_n\eta) = E(Z\eta), \qquad \text{for all bounded } \mathcal{F}\text{-measurable r.v.'s } \eta$$

or equivalently

$$\lim_{n\to\infty} E(Z_n I_E) = E(Z I_E), \qquad \text{for all } \mathcal{F}\text{-measurable events } E,\ P(E) > 0.$$

We denote this by

$$Z_n \longrightarrow Z \text{ (weakly in } L^1). \quad \blacksquare$$

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,

$$\text{“}L^1\text{-convergence”} \Longrightarrow \text{“weak convergence in } L^1\text{”} \Longrightarrow \text{“uniform integrability”}.$$

The first relation follows from

$$|E(Z_n I_E) - E(Z I_E)| = |E[(Z_n - Z) I_E)]| \le E|Z_n - Z| \longrightarrow 0.$$

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,

(a) $Z_n \longrightarrow Z$ in $L^1 \Longleftrightarrow E(Z_n I_E) \longrightarrow E(Z I_E)$ **uniformly** in $E \in \mathcal{F}$;

(b) 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 1. Suppose that $F_n \Longrightarrow F$,. The following conditions are equivalent:

(A) $Y_n \Longrightarrow F_Y$ (stably);

(B) For all fixed $\mathcal{F}$-measurable rv's $\sigma$, the sequence of random vectors $(Y_n, \sigma)$ converges jointly in distribution;

(C) For each fixed real $t$, the sequence of (complex-valued) rv's $\exp\{itY_n\}$ converges weakly in $L^1$;

(D) For all fixed $k$ and $B \in \sigma(Y_1, ..., Y_k)$, $P(B) > 0$, $\lim_{n\to\infty} P(Y_n \leq x|B)$ exists for a countable dense set of points $x$.

**Proof.** Omitted. ∎

## 9.6 Exercise

1. There exist random variables which do not have moments of any order: for example, the density function given by $f(x) = \dfrac{1}{x(\ln x)^2}$, $x \geq \mathrm{e}$.

## 9.7 Some useful theorems

THEOREM **9.7.1 (Continuous mapping theorem)** *Let $g$ be a measurable function and $D_g = \{x : g$ 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 $Eg(X_n) \to Eg(X_\infty)$.*

***Remark.*** *$D_g$ 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

$$f(g(Y_n)) \to f(g(Y_\infty)) \quad a.s.$$

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. ∎

# Chapter 10

# Characteristic Functions

DEFINITION **10.0.1** *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.*

## 10.1 Some examples of characteristic functions

For ease of reference we give a table of the c.f.'s of some common densities and describe the method of deriving them. For more details, see Feller (1971, page 502).

1. **Standard normal:**

$$\begin{aligned} &\text{p.d.f.} \quad f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \\ &\text{c.f.} \quad \psi(t) = e^{-t^2/2}. \end{aligned}$$

*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,

$$\begin{aligned} \psi'(t) &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\sin(tx) e^{-x^2/2} dx = \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(tx) de^{-x^2/2} \\ &= \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx = -t\psi(t). \end{aligned}$$

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**$[0, a]$**.**

$$\begin{aligned} &\text{p.d.f.} \quad f(x) = \frac{1}{a} I\{0 \le x \le a\} \\ &\text{c.f.} \quad \psi(t) = \int_0^a e^{itx} \frac{1}{a} dx = \frac{1}{a} \left. \frac{e^{itx}}{it} \right|_{x=0}^{a} = \frac{e^{iat} - 1}{iat}. \end{aligned}$$

**Uniform**$[-a, a]$**.**

$$\begin{aligned} &\text{p.d.f.} \quad f(x) = \frac{1}{2a} I\{-a \le x \le a\} \\ &\text{c.f.} \quad \psi(t) = \int_{-a}^a e^{itx} \frac{1}{2a} dx = \frac{1}{2a} \left. \frac{e^{itx}}{it} \right|_{x=-a}^{a} = \frac{\sin(at)}{at}. \end{aligned}$$

3. **Triangular**$[-a,a]$**:**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \frac{1}{a}\left(1 - \frac{|x|}{a}\right) I\{|x| < a\} = \frac{1}{a}\left(1 - \frac{|x|}{a}\right)^{+} \\ \text{c.f.} & \psi(t) = \frac{2(1-\cos(at))}{a^2t^2}. \end{array}$$

**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

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) = \frac{\sin^2(bt)}{b^2t^2} = 2\frac{2\sin^2(at/2)}{a^2t^2} = \frac{2(1-\cos(at))}{a^2t^2}. \quad \blacksquare$$

**Direct proof.** By definition,

$$\begin{aligned} \psi(t) &= \int_{-\infty}^{\infty} e^{itx} f(x)dx \\ &= \frac{2}{a}\int_0^a \left(1 - \frac{x}{a}\right)\cos(tx)dx \\ &= \frac{2}{at}\int_{x=0}^a \left(1 - \frac{x}{a}\right) d\sin(tx) \\ &= \frac{2}{2t}\left[\left(1 - \frac{x}{a}\right)\sin(tx)\right]\Big|_{x=0}^a + \frac{2}{a^2t}\int_{x=0}^a \sin(tx)dx \\ &= \frac{1-\cos(ax)}{a^2t^2}. \quad \blacksquare \end{aligned}$$

4. **Inverse Triangular (or Polya distribution):**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \frac{1}{\pi}\frac{1-\cos(ax)}{ax^2} \\ \text{c.f.} & \psi(t) = \left(1 - \frac{|t|}{a}\right) I\{|t| < a\} = \left(1 - \frac{|t|}{a}\right)^{+}. \end{array}$$

**Note**: Here $\psi(t)$ has a bounded support, which proves useful later on. See the section on Esseen's smooth lemma.

**Proof.** Apply Theorem 10.3.4 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

$$\left(1 - \frac{|t|}{a}\right)^{+} = \int_{-\infty}^{\infty} e^{ity}\left(\frac{1}{\pi}\frac{(1-\cos(ay))}{ay^2}\right)dy = \int_{-\infty}^{\infty} e^{ity}f(y)dy. \quad \blacksquare$$

5. **Exponential distribution**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = e^{-x},\ x \geq 0 \\ \text{c.f.} & \psi(t) = \frac{1}{1-it}. \end{array}$$

*Proof.* Integrating gives

$$\psi(t) = \int_0^{\infty} e^{itx}e^{-x}dx = \int_0^{\infty} e^{x(it-1)}dx = \frac{e^{x(it-1)}}{(it-1)}\Big|_{x=0}^{\infty} = 0 - \frac{1}{it-1} = \frac{1}{1-it}. \quad \blacksquare$$

6. **Gamma distribution**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \dfrac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x},\ x \geq 0 \\ \text{c.f.} & \psi(t) = \dfrac{1}{(1 - it/\lambda)^c}. \end{array}$$

*Proof.* Exercise. ∎

7. **Double exponential distribution**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \dfrac{1}{2} e^{-|x|},\ x \geq 0 \\ \text{c.f.} & \psi(t) = \dfrac{1}{1 + t^2}. \end{array}$$

*Proof.* Integrating gives

$$\begin{aligned} \psi(t) &= \frac{1}{2}\int_{-\infty}^{\infty} e^{itx} e^{-|x|} dx = \frac{1}{2}\int_{-\infty}^{0} e^{itx} e^{x} dx + \frac{1}{2}\int_{0}^{\infty} e^{itx} e^{-x} dx \\ &= \frac{1}{2}\int_{-\infty}^{0} e^{x(it+1)} dx + \frac{1}{2}\int_{0}^{\infty} e^{x(it-1)} dx = \left.\frac{e^{x(it+1)}}{2(it+1)}\right|_{x=-\infty}^{0} + \left.\frac{e^{x(it-1)}}{2(it-1)}\right|_{x=0}^{\infty} \\ &= \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{2}\left(\frac{2}{1-(it)^2}\right) = \frac{1}{1+t^2}. \quad ∎ \end{aligned}$$

8. **The Cauchy distribution:**

$$\begin{array}{ll} \text{p.d.f.} & f(x) = \dfrac{1}{\pi}\dfrac{1}{1 + x^2} \\ \text{c.f.} & \psi(t) = \exp(-|t|). \end{array}$$

**Proof.** Apply Theorem 10.3.4 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**. $X, X_1, ..., X_n \sim_{iid} Bernoulli(p)$ (coin flips), i.e., $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$, and $S_n = X_1 + ... + X_n \sim Bin(n, p)$. Then

$$\psi_X(t) = Ee^{itX} = pe^{it} + q,$$

and

$$\psi_{S_n}(t) = \psi^n(t) = (pe^{it} + q)^n.$$

10. **Poisson($\lambda$) distribution**.

$$\begin{array}{lll} \text{p.m.f.} & P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}, & k \geq 0. \\ \text{c.f.} & \psi(t) = \exp\{\lambda(e^{it} - 1)\}. & \end{array}$$

**Proof.**

$$\psi(t) = Ee^{itX} = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda[e^{it} - 1]}. \quad ∎$$

11. **Centered Poisson distribution**. $X_t \sim$ Poisson($\lambda$), and $\widetilde{X}_t = X_t - \lambda t$, whose c.f. is

$$\psi(t) = Ee^{itX} e^{-i\lambda t} = e^{\lambda[e^{it} - 1 - it]}. \quad ∎$$

12. **Compound Poisson distribution**.

Here is a tally of some of popular d.f.s and their corresponding c.f.s.

See Feller, or Chow and Teicher, p294.

## 10.2 Definition and some properties of c.f.s

**Some elementary properties of c.f.s**

1. $\psi(0) = 1$ and $|\psi(t)| = |Ee^{itX}| \le E|e^{itX}| = 1$ for all $t$.
2. $\psi(t)$ is uniformly continuous in $t \in (-\infty, \infty)$.

   *Proof.* For any real $t$ and $h \to 0$,

   $$|\psi(t+h) - \psi(t)| = |Ee^{i(t+h)X} - Ee^{itX}| = |E[e^{itX}(e^{ihX} - 1)]| \le E|e^{ihX} - 1| \to 0. \quad \blacksquare$$

3. $\psi_{aX+b}(t) = Ee^{it(aX+b)} = e^{itb}Ee^{itaX} = e^{itb}\psi_X(at)$.
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. $\blacksquare$

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 $F_1, ..., F_n$ are d.f.'s with c.f. $\psi_1, ..., \psi_n$. If $\lambda_i \ge 0$, $\sum_{i=1}^n \lambda_i = 1$, 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 $\mathrm{Re}\psi$.

   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. $\psi$; $\mathrm{Re}\psi$ is the c.f. of the d.f. $(F_X(x) + F_{-X}(x))/2$.

   *Proof.* First, let $X, Y$ be i.i.d. with c.f. $\psi(t)$, then

   $$\psi_{X-Y}(t) = \psi_X(t)\psi_Y(-t) = \psi(t)\overline{\psi(t)} = |\psi(t)|^2.$$

   Secondly, $-X$ has c.f. $\overline{\psi(t)}$. So the d.f. $(F_X(x) + F_{-X})/2$ has c.f.

   $$\frac{1}{2}\left(\psi_X(t) + \overline{\psi_X(t)}\right) = \mathrm{Re}\psi_X(t). \quad \blacksquare$$

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. $Bin(1, 1/3)$, 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, $Y_j \in \{-\frac{1}{2}, \frac{1}{2}\}$. Write $\alpha = P(Y_j = 1/2)$. Then,

   $$\begin{aligned} \alpha^2 &= P(Y_1 = \frac{1}{2})P(Y_2 = \frac{1}{2}) = P(Y_1 + Y_2 = 1) = P(X_1 - X_2 = 1) = \frac{2}{9} \\ (1-\alpha)^2 &= P(Y_1 = -\frac{1}{2})P(Y_2 = -\frac{1}{2}) = P(Y_1 + Y_2 = -1) = P(X_1 - X_2 = -1) = \frac{2}{9} \end{aligned}$$

   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. $\blacksquare$

   **Remark.** 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 $0 = Var(Z) = 2Var(X)$. Then $P(X = b) = 1$ for some constant $b$. Hence, $\psi_X(t) = e^{ibt}$. ∎

**Remark.** 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 10.3.2 to calculate it directly as follows:

$$
\begin{aligned}
P(\{a\}) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\psi(t)dt = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}dt \\
&= \lim_{T\to\infty} 1 = 1, \qquad \text{if } a = 0 \\
&\quad \lim_{T\to\infty} \frac{\sin Ta}{Ta} = 0, \qquad \text{if } a \neq 0.
\end{aligned}
$$

11. A c.f could be nowhere differentiable.

For instance, take $P(X = 5^k) = 1/2^{k+1}$, $k = 0, 1, ...$, then $\psi(t) = \sum_{k=0}^{\infty} \exp\{it5^k\}/2^{k+1}$ 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 **10.2.1**

$$
Re\,(1-\psi(t)) \geq \frac{1}{4} Re\,(1-\psi(2t)) \geq ... \geq \frac{1}{4^n} Re\,(1-\psi(2^n t)). \tag{2.1}
$$

*In particular, we have*

*(a)* $1 - |\psi(t)|^2 \geq \frac{1}{4}\left(1 - |\psi(2t)|^2\right) \geq ... \geq \frac{1}{4^n}\left(1 - |\psi(2^n t)|^2\right).$
*(b)* $(1 - |\psi(t)|) \geq \frac{1}{8}(1 - |\psi(2t)|) \geq \frac{1}{8^n}(1 - |\psi(2^n t)|).$

*Proof.* We prove (2.1) first, which follows by taking expectation to

$$
1 - \cos 2tX = 2\left(\sin(tX)\right)^2 = 2\left(2\sin(tX/2)\cos(tX/2)\right)^2 \leq 8\sin^2(tX/2) = 4\left(1 - \cos tX\right).
$$

(a) Applying the above to c.f. $|\psi(t)|^2$.
(b) Noting $0 \leq |\psi| \leq 1$, we get

$$
\begin{aligned}
&1 - |\psi(2t)| \leq (1 - |\psi(2t)|)(1 + |\psi(2t)|) \leq 1 - |\psi(2t)|^2 \\
&\leq 4\left(1 - |\psi(t)|^2\right) \leq 4\left(1 - |\psi(t)|\right)\left(1 + |\psi(t)|\right) \leq 8\left(1 - |\psi(t)|\right). \quad \blacksquare
\end{aligned}
$$

Corollary **10.2.1** *Suppose that* $|\psi(t)| \leq a < 1$ *for* $|t| \geq b > 0$. *Then*

$$
|\psi(t)| \leq 1 - ct^2 \leq e^{-ct^2}, \qquad \textit{for } |t| < b, \textit{ where } c = \tfrac{1-a^2}{8b^2}.
$$

*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 $1 + x \leq e^x$ for all $x$.)

Since $|t| \leq b$, there exists an $m$ such that $b/2^m \leq |t| < 2b/2^m$, i.e., $b \leq |2^m t| < 2b$. Thus,

$$
1 - |\psi(t)|^2 \geq \frac{1}{4^m}\left(1 - |\psi(2^m t)|^2\right) \geq \left(\frac{1}{2^m}\right)^2 \left(1 - a^2\right) \geq \left(\frac{t}{2b}\right)^2 \left(1 - a^2\right) = 2ct^2.
$$

In view of the inequality $(1-x)^{1/2} \leq 1 - x/2$ for $|x| \leq 1$, we have

$$
|\psi(t)| = \left(|\psi(t)|^2\right)^{1/2} \leq \left(1 - 2ct^2\right)^{1/2} \leq 1 - ct^2. \quad \blacksquare
$$

COROLLARY **10.2.2** *Under Cramer's condition:* $\limsup_{|t|\to\infty} |\psi(t)| < 1$, *then for any* $\delta > 0$, *there exists* $d \in (0,1)$ *such that*

$$|\psi(t)| \le d, \qquad \textit{for } |t| \ge \delta.$$

*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 $c = \max\{1 - c\delta^2, a\}$. ∎

THEOREM **10.2.2** *Let* $X$ *be a nondegenerate r.v. with c.f* $\psi$. *There exist* $\delta > 0$ *and* $\epsilon > 0$ *such that*

$$|\psi(t)| \le 1 - \epsilon t^2 \qquad \textit{for } |t| \le \delta.$$

*Proof.* Let $Y$ be an independent copy of $X$, and $Z = X - Y$. Then,

$$\begin{aligned} 1 - |\psi(t)| &\ge (1 - |\psi(t)|)\frac{1 + |\psi(t)|}{2} = \frac{1}{2}\left(1 - |\psi(t)|^2\right) \\ &= \frac{1}{2}(1 - \psi_Z(t)) = \frac{1}{2}E(1 - \cos tZ) \end{aligned}$$

By Taylor expansion, for $|t| < 1$,

$$1 - \cos t = 1 - \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - ...\right) = \frac{t^2}{2!} - \frac{t^4}{4!} + ... \ge \frac{t^2}{2!} - \frac{t^4}{4!} = \frac{t^2}{2}\left(1 - \frac{t^2}{12}\right)$$

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 $|t| \le \delta$,

$$1 - |\psi(t)| \ge \frac{1}{2}E(1 - \cos tZ) \ge \left(\frac{11}{48}E\left(Z^2 I\{|Z| < 1/\delta\}\right)\right)t^2 = \epsilon t^2. \quad ∎$$

THEOREM **10.2.3** *For any* $t, h \in R$, *we have*

$$|\psi(t+h) - \psi(t)|^2 \le 2\left(1 - Re\ \psi(h)\right) = 2E[1 - \cos(hX)].$$

*Proof.*

$$\begin{aligned} |\psi(t+h) - \psi(t)|^2 &= |Ee^{i(t+h)X} - Ee^{itX}|^2 = |E[e^{itX}(e^{ihX} - 1)]|^2 \\ &\le E|e^{ihX} - 1|^2 = E[(e^{ihX} - 1)\overline{(e^{ihX} - 1)}] \\ &= E[(e^{ihX} - 1)(e^{-ihX} - 1)] \\ &= E(e^{ihX}e^{-ihX} - e^{ihX} - e^{-ihX} + 1) \\ &= 2E[1 - \cos(hX)]. \quad ∎ \end{aligned}$$

As an application, we give the following example. Compare this with Levy Continuity Theorem.

EXAMPLE **10.2.1** *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

$$
\begin{aligned}
|g(t+h)-g(t)|^2 &= \lim_{n\to\infty}|\psi_n(t+h)-\psi_n(t)|^2 \\
&\le 2\lim_{n\to\infty}[1-Re\ \psi_n(h)] \qquad \text{(by Theorem 10.2.3)} \\
&= 2[1-Re\ g(h)] \\
&\to 2[1-Re\ g(0)]=0, \qquad \text{as } h\to 0.
\end{aligned}
$$

EXAMPLE **10.2.2** *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 10.2.3,

$$||\psi_n(2t)|-|\psi_n(t)||^2 \le |\psi_n(2t)-\psi_n(t)|^2 \le 2\left(1-Re\ \psi_n(t)\right).$$

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 $t\in R$.

**Alternative proof.** One can prove this by using Theorem 10.2.1, which will extend the local properties to global ones. We will leave the details here. In the following, we shall instead use Theorem 10.2.3 to prove the statement.

## 10.3 Inversion formula

We will prove a very important fact: the c.f. uniquely determines the distribution. To do that, we need to prove the inversion formula.

### The inversion formula

THEOREM **10.3.1 (The inversion formula.)** *Let $\psi(t)=\int e^{itx}\mu(dx)$, where $\mu$ 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

$$
\begin{aligned}
I(T) &=: \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\psi(t)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\left(\int_{-\infty}^{\infty}e^{itx}\mu(dx)\right)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\left(\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(dx)\right)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt\right)\mu(dx) \\
&\quad \text{(by Fubini's theorem since } \left|\tfrac{e^{-ita}-e^{-itb}}{it}e^{itx}\right| = \left|\int_a^b e^{-itx}dx\right|\,\left|e^{itx}\right| \le b-a) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt\right)\mu(dx) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\frac{1}{it}\left(\cos[t(x-a)]-\cos[t(x-b)]\right)dt\right)\mu(dx) \\
&\quad +\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\frac{i}{it}\left(\sin[t(x-a)]-\sin[t(x-b)]\right)dt\right)\mu(dx)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\frac{1}{t}\left(\sin[t(x-a)]-\sin[t(x-b)]\right)dt\right)\mu(dx)\\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{0}^{T}\frac{1}{t}\left(\sin[t(x-a)]-\sin[t(x-b)]\right)dt\right)\mu(dx)\\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}[I(x-a,T)-I(x-b,T)]\mu(dx),
\end{aligned}
$$

where $I(a,T)$ is defined in Lemma 10.3.1 below. Again from Lemma 10.3.1, we get

$$
\begin{aligned}
\lim_{T\to\infty}\frac{1}{\pi}I(x-a,T)-I(x-b,T) &= (-\frac{1}{2})-(-\frac{1}{2})=0, \qquad x<a\\
&= 0-(-\frac{1}{2})=\frac{1}{2}, \qquad x=a\\
&= \frac{1}{2}-(-\frac{1}{2})=1, \qquad a<x<b\\
&= \frac{1}{2}-0=\frac{1}{2}, \qquad x=b\\
&= (\frac{1}{2})-(\frac{1}{2})=0, \qquad x>b.
\end{aligned}
$$

Note that from (3.2), for every $\theta$, we have

$$
|I(\theta,T)|\le\int_0^{\theta T}\frac{\sin v}{v}dv\le\int_0^{\infty}\frac{\sin v}{v}dv=I(1,\infty)=\frac{\pi}{2}.
$$

Using the dominated (or bounded) convergence theorem, we have

$$
\begin{aligned}
\lim_{T\to\infty}I(T) &= \int_{-\infty}^{\infty}\frac{1}{\pi}\lim_{T\to\infty}[I(x-a,T)-I(x-b,T)]\mu(dx)\\
&= \int_{(-\infty,a)\cup(b,\infty)}0\mu(dx)+\int_{(a,b)}1\mu(dx)+\int_a\frac{1}{2}\mu(dx)+\int_b\frac{1}{2}\mu(dx)\\
&= \mu(a,b)+\frac{1}{2}\mu(\{a,b\}). \quad \blacksquare
\end{aligned}
$$

We now prove a lemma, which was used in the above proof.

**Lemma 10.3.1** *We have* $\lim_{T\to\infty}\int_0^T\frac{\sin au}{u}du=\frac{\pi}{2}sgn\{a\}.$

*Proof.* Write

$$
I(a,T)=:\int_0^T\frac{\sin au}{u}du=\int_0^{aT}\frac{\sin v}{v}dv. \tag{3.2}
$$

Then, we have

$$
\lim_{T\to\infty}I(a,T)=sgn\{a\}\lim_{T\to\infty}I(1,T).
$$

This can be seen as follows:

(i) 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 $\lim_{T\to\infty}I(a,T)=0=sgn\{0\}\lim_{T\to\infty}I(1,T)$.

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 $T \to \infty$. ∎

We can imitate the proof in the last theorem to derive the theorem.

THEOREM **10.3.2**

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

*Proof.* Here are the main steps involved.

$$
\begin{aligned}
\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita} \psi(t) dt &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{\infty}^{\infty} e^{it(y-a)} d\mu(y) dt \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} \int_{-T}^{T} e^{it(y-a)} dt d\mu(y) \\
&= \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{\sin[(y-a)T]}{(y-a)T} d\mu(y) \\
&= \lim_{T\to\infty} \int_{(-\infty,a)} + \int_{\{a\}} + \int_{(a,\infty)} \frac{\sin[(y-a)T]}{(y-a)T} d\mu(y) \\
&= 0 + \mu(\{a\}) + 0 \\
&= \mu(\{a\}).
\end{aligned}
$$

∎

## One-to-one correspondence between d.f. and c.f.

THEOREM **10.3.3 (Uniqueness)** *Characteristic functions uniquely determines distribution functions. That is, there is a one-one correspondence between c.f.s and d.f.s.*

*Proof.* 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 $\psi$ 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, $\mu$ is a continuous measure. In fact, we can get a stronger result than this.

THEOREM **10.3.4** *If $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$, then $\mu$ 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 10.3.2 that $\mu(\{a,b\}) = 0$ for all $a < b$. Take $a = y$ and $b = y + h$ in Theorem 10.3.1 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

$$\int_{-\infty}^{\infty} \Big| \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t)\Big| dt = \int_{-\infty}^{\infty} \Big|\frac{1 - e^{-ith}}{ith} \psi(t)\Big| dt \le \int_{-\infty}^{\infty} \big|\psi(t)\big| dt < \infty.$$

Then

$$\begin{aligned}
\lim_{h\searrow 0} \frac{\mu(y, y+h)}{h} &= \lim_{h\searrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t)dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h\searrow 0} \frac{e^{-ity} - e^{-it(y+h)}}{ith} \psi(t)dt \\
&\qquad \text{(dominated convergence theorem)} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h\searrow 0} \frac{ite^{-it(y+h)}}{it} \psi(t)dt \qquad \text{(L'Hospital rule)} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \psi(t)dt.
\end{aligned}$$

This completes the proof. ∎

REMARK **10.3.1** *When $\int_{-\infty}^{\infty} |\psi(t)|dt < \infty$, we have the following interesting relationships:*

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} e^{ity} f(y)dy, \\
f(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \psi(t)dt.
\end{aligned}$$

*Therefore, if we know one relation, we know the other one straightaway. One can use these relationships to find some c.f.s easily.*

REMARK **10.3.2** *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 $\psi$ is not integrable

THEOREM **10.3.5** *If $P(X \in b + h\mathbf{Z}) = 1$, where $\mathbf{Z} = \{0, \pm 1, \pm 2, ....\}$, 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,

$$\begin{aligned}
\frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-itx} \psi(t)dt &= \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-it(b+jh)} \sum_{k=-\infty}^{\infty} e^{it(b+kh)} p_k dt \\
&= \frac{h}{2\pi} \sum_{k=-\infty}^{\infty} p_k \int_{-\pi/h}^{\pi/h} e^{it(k-j)h} dt \\
&= \frac{h}{2\pi} \left( \sum_{k\neq j} + \sum_{k=j} \right) p_k \int_{-\pi/h}^{\pi/h} e^{it(k-j)h} dt \\
&= \frac{h}{2\pi} \left( 0 + p_j \int_{-\pi/h}^{\pi/h} 1 dt \right) \\
&= p_j = P(X = jh) \\
&= P(X = x). \quad ∎
\end{aligned}$$

## 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 **10.4.1** *For any $a > 0$, we have*

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

*Proof.*

$$\begin{aligned}
\int_{-a}^{a}(1-\psi(t))dt &= 2a - \int_{-a}^{a} Ee^{itX}dt \\
&= 2a - E\left(\int_{-a}^{a} e^{itX}dt\right) \quad \text{(by Fubini theorem)} \\
&= 2a - E\left(\int_{-a}^{a} \cos(tX)dt\right) \\
&= 2a - E\left(\frac{2\sin(aX)}{X}\right) \\
&= 2aE\left(1 - \frac{\sin(aX)}{aX}\right) \\
&\geq 2aE\left(1 - \frac{\sin(aX)}{aX}\right) I\{|aX| > 2\} \\
&\geq 2aE\left(1 - \frac{1}{2}\right) I\{|aX| > 2\} \\
&\geq aP(|aX| > 2). \quad \blacksquare
\end{aligned}$$

REMARK **10.4.1** *When a is chosen to be very small, Lemma 10.4.1 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 **10.4.2** *Let $F_n$ be a sequence of d.f.s with c.f.s $\psi_n$. 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

$$\begin{aligned}
P\left(|X_n| > \frac{2}{a_0}\right) &\leq \left|\frac{1}{a_0}\int_{-a_0}^{a_0}(1-\psi_n(t))dt\right| \qquad \text{(Lemma 10.4.1)} \\
&\longrightarrow \left|\frac{1}{a_0}\int_{-a_0}^{a_0}(1-g(t))dt\right| \qquad \text{(dominated convergence theorem)} \\
&\leq \frac{1}{a_0}\int_{-a_0}^{a_0}|1-g(t)|dt \leq \frac{1}{a_0}\int_{-a_0}^{a_0}\varepsilon dt \leq \varepsilon/2.
\end{aligned}$$

Thus, $\exists N_0 > 0$ such that, for all $n > N_0$, one has $P\left(|X_n| > \frac{2}{a_0}\right) \leq \varepsilon$. $\blacksquare$

THEOREM **10.4.1 (Levy continuity theorem)**

*Assume that $X_n$ has d.f. $F_n$ and c.f. $\psi_n$ for $1 \leq n \leq \infty$.*

*(i) If $X_n \to_d X_\infty$, (i.e., $F_n \Longrightarrow F_\infty$), then $\psi_n(t) \to \psi_\infty(t)$ for all $t$.*

*(ii) 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 $\psi$ is the c.f. of $X$.*

*Proof.*

(i) The proof follows from the bounded convergence theorem.

(ii). Now suppose that $\psi_n(t) \to \psi(t)$, and that $\psi(t)$ is continuous at 0. From Lemma 10.4.2, $F_n$ 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 10.3.3, we get $\tilde{F} = F$. This shows that $F$ is the only possible weak limit of the $F_n$. Therefore,

$$F_n \Longrightarrow F. \quad \blacksquare$$

The following corollary is immediate.

Corollary **10.4.1 (Levy continuity theorem)**

*$X_n \to_d X$ if and only if $\psi_{X_n}(t) \to \psi_X(t)$ for all $t$.*

## 10.5 Moments of r.v.s and derivatives of their c.f.s

## Relation between moments of a r.v. and derivatives of its c.f.

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 **10.5.1** *If $E|X|^n < \infty$, then $\psi^{(n)}(t)$ exists and is a uniformly continuous function given by*

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

*In particular,*

$$\psi^{(k)}(0) = i^k EX^k, \qquad k = 0, 1, ..., n.$$

*Proof.* Note that

$$\frac{\psi(t+h) - \psi(t)}{h} = \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF(x).$$

Using Lemma 10.5.1, the integrand is dominated by $|x|$. So the first derivative of $\psi(t)$ exists by the dominated convergence theorem, and is given by

$$\psi'(t) = \lim_{h\to 0} \frac{\psi(t+h) - \psi(t)}{h} = i \int_{-\infty}^{\infty} x e^{itx} dF(x).$$

The uniform continuity of $\psi'(t)$ follows from

$$|\psi'(t+\delta) - \psi'(t)| = \left| \int_{-\infty}^{\infty} x e^{itx} (e^{it\delta} - 1) dF(x) \right| \to 0$$

by the dominated convergence theorem. Therefore, the assertion is true for $n = 1$. The general case follows by induction. $\blacksquare$

A partial converse is given by the following theorem.

Theorem **10.5.2** *If $\psi^{(n)}(0)$ exists and is finite for some $n = 1, 2, ...,$ 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

$$
\begin{aligned}
-\psi^{(2)}(0) &= -\lim_{h\to 0}\frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2} \qquad (5.3)\\
&= 2\lim_{h\to 0} E\frac{1-\cos(hX)}{h^2}\\
&\geq 2E\lim_{h\to 0}\frac{1-\cos(hX)}{h^2} \qquad \text{(by Fatou's lemma)}\\
&= EX^2.
\end{aligned}
$$

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 10.5.1,

$$\psi^{(2j)}(t) = i^{2j}\int_{-\infty}^{\infty} x^{2j}e^{itx}dF(x) = (-1)^j E\left(X^{2j}e^{itX}\right),$$

So if $\psi^{(2(j+1))}(0) < \infty$, then

$$
\begin{aligned}
(-1)^{j+1}\psi^{(2(j+1))}(0) &= (-1)^{j+1}\lim_{h\to 0}\frac{\psi^{(2j)}(h)+\psi^{(2j)}(-h)-2\psi^{(2j)}(0)}{h^2}\\
&= 2(-1)^{2j+2}\lim_{h\to 0}\int_{-\infty}^{\infty}\frac{x^{2j}\left[1-\cos(hx)\right]}{h^2}dF(x)\\
&= 2\lim_{h\to 0} E\frac{X^{2j}\left[1-\cos(hX)\right]}{h^2}\\
&\geq 2E\lim_{h\to 0}\frac{X^{2j}\left[1-\cos(hX)\right]}{h^2} \qquad \text{(by Fatou's lemma)}\\
&= EX^{2(j+1)}.
\end{aligned}
$$

Therefore, $\psi^{(2(j+1))}(0) < \infty$ implies $EX^{2(j+1)} < \infty$. ∎

Remark **10.5.1**

*If $g^{(2)}(0)$ exists, then applying L'Hospital Rule, we can see that*

$$\lim_{h\to 0}\frac{g(h)+g(-h)-2g(0)}{h^2} \quad \text{exists and equals } g^{(2)}(0).$$

*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*

$$g(t) = a_0 + a_1 t + \frac{1}{2}a_2 t^2 + D(t)t^3, \qquad \text{as } t \to 0,$$

*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 $t_0$.*

Remark **10.5.2**

*However, if $g(t) = \psi(t)$ is a c.f., then*

$$\psi^{(2)}(0) \text{ exists} \qquad \Longleftrightarrow \qquad \lim_{h\to 0}\frac{\psi(h)+\psi(-h)-2\psi(0)}{h^2} \text{ exists},$$

*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 (5.3) in the proof of Theorem 10.5.2 above that $EX^2 < \infty$. Applying Theorem 10.5.1 again to obtain $\psi^{(2)}(0)$ exists.* ∎

Remark **10.5.3** *From Theorem 10.5.2, $\psi^{(n)}(0) < \infty$ 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**. Define $P(X = \pm n) = C/(2n^2 \log n), n = 2, 3, ....$ 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,

$$E|X| = C \sum_{n=2}^{\infty} \frac{1}{n \log n} = \infty. \quad \blacksquare$$

# 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 **10.5.3**

(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*

$$\begin{aligned} \psi(t) &= \sum_{k=0}^{n} (EX^k) \frac{(it)^k}{k!} + \theta \frac{2E|X|^{n+\delta}|t|^{n+\delta}}{n!} \qquad \textit{with } |\theta| \leq 1 \\ \psi(t) &= \sum_{k=0}^{n} a_k \frac{(it)^k}{k!} + o(t^n), \qquad \textit{as } t \to 0, \end{aligned}$$

*where $a_k = EX^k$ for $k = 0, 1, 2, ..., n$*

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

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

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

Remark **10.5.4** *Suppose that*

$$g(t) = \sum_{k=0}^{2} a_k \frac{(it)^k}{k!} + o(t^2) = a_0 + a_1 t + \frac{1}{2} a_2 t^2 + o(t^2), \qquad \textit{as } t \to 0, \tag{5.4}$$

*we can easily show that*

$$\begin{aligned} g(0) &= \lim_{t \to 0} g(t) = a_0 \\ g'(0) &= \lim_{t \to 0} \frac{g(t) - g(0)}{t} = \lim_{t \to 0} \frac{a_1 t + \frac{1}{2} a_2 t^2 + o(t^2)}{t} \\ &= \lim_{t \to 0} \left( a_1 + \frac{1}{2} a_2 t + o(t) \right) = a_1. \end{aligned}$$

*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 (5.4), as can be easily seen from the same example given in Remark 10.5.1.*

*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 of Theorem 10.5.3**

(1). The first equation in (1) follows directly from Lemma 10.5.2. Let us show the second one. By Taylor expansion and Theorem 10.5.1,

$$\psi(t) = \sum_{k=0}^{n-1} \frac{\psi^{(k)}(0)}{k!} t^k + \frac{\psi^{(n)}(\theta t)}{n!} t^n = \sum_{k=0}^{n} (EX^k) \frac{(it)^k}{k!} + R_n(t),$$

where $R_n(t) = t^n[\psi^{(n)}(\theta t) - \psi^{(n)}(0)]/n!$ with $\theta \in [0, 1]$. Note that

$$\frac{R_n(t)}{t^n} = \int_{-\infty}^{\infty} \frac{(ix)^n}{n!}(e^{i\theta tx} - 1) dF(x).$$

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

$$\psi(t) \quad = \quad a_0 + a_1(it) + \frac{a_2}{2}(it)^2 + o(t^2),$$

we have

$$\psi(t) + \psi(-t) = 2\left(a_0 + \frac{a_2}{2}(it)^2 + o(t^2)\right).$$

Thus,

$$-\lim_{h\to 0} \frac{\psi(h) + \psi(-h) - 2\psi(0)}{h^2} = \lim_{h\to 0} (a_2 + o(1)) = a_2. \quad ∎$$

## Application to Weak Law of Large Numbers

**Example**. (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 10.5.3,

$$\psi(t) = 1 + i(EX)t + o(t) = 1 + o(t).$$

Therefore, the c.f. of $\bar{X}$ is

$$\psi_{\bar{X}}(t) = \psi^n(t/n) = \left(1 + o(n^{-1})\right)^n \to 1$$

as $t \to 0$ by taking logarithms. Thus, $\bar{X} \to_d 0$, and so $\bar{X} \to_p 0$. 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): $\bar{X} \to 0$ a.s. If we are content with WLLN, we can do without assuming the existence of the first moment; see Theorem 10.5.4 below. ∎

In Theorem 10.5.1, 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 $\psi$ is closely connected with the weak law of large numbers (WLLN) for i.i.d. r.v.'s.

THEOREM **10.5.4** *Let $X, X_1, X_2, ...$ be i.i.d. r.v.'s with d.f. $F$. The following three statements are equivalent:*

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

(ii) $t[1 - F(t) - F(-t)] \longrightarrow 0, \qquad \int_{-t}^{t} xF\{dx\} \to \nu$ *as* $t \to \infty$.

(*or equivalently,* $tP(|X| > t) \to 0$ *and* $E[XI_{\{|X| \le t\}}] \to \nu$.)

(iii) *The WLLN holds:* $(X_1 + ...X_n)/n \to \nu$ *in probability.*

**Proof.** Omitted. ∎

REMARK **10.5.5** *It should be noted that $\nu$ is not necessarily the mean $EX$, which may not even exist here. For instance, from (ii) of Theorem 10.5.4, we can take $X$ to be symmetric around 0 and*

$$P(X > t) = \frac{1}{t \log t}, \qquad \text{as } t \to \infty.$$

The mean does not exist, but $\nu = 0$ and the WLLN holds here.

## Appendix: Several useful lemmas

LEMMA **10.5.1** *For $n = 0, 1, 2, ...$ and any real $t$,*

$$\left| e^{it} - 1 - it - \frac{(it)^2}{2!} - ... - \frac{(it)^n}{n!} \right| \le \min\left\{ \frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right\}$$

*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

$$\begin{aligned} e^{it} &= 1 + \left(e^{it} - 1\right) = 1 + i \int_0^t e^{is} ds \\ &= 1 + it + i^2 \int_0^t (t-s) e^{is} ds \\ &= ...... \\ &= 1 + it + \frac{(it)^2}{2!} + ... + \frac{(it)^n}{n!} + \frac{i^{n+1}}{n!} \int_0^t (t-s)^n e^{is} ds \end{aligned}$$

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

$$\left| \int_0^t (t-s)^n e^{is} ds \right| \le 2n \int_0^{|t|} |t-s|^{n-1} ds = 2|t|^n,$$

Using these relationships, we get

$$\left| e^{it} - 1 - it - \frac{(it)^2}{2!} - ... - \frac{(it)^n}{n!} \right| \leq \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right\}$$

The proof is complete. ∎

The next lemma follows directly from Lemma 10.5.1.

LEMMA **10.5.2** *For $n = 0, 1, 2, ...$ and any real $t$,*

$$\left| e^{it} - 1 - it - \frac{(it)^2}{2!} - ... - \frac{(it)^n}{n!} \right| \leq \frac{2|t|^{n+\delta}}{n!}, \qquad \textit{for every } \delta \in [0,1].$$

*Proof.* If $|t| \leq 1$, then $|t|^{n+1}/(n+1)! \leq |t|^{n+\delta}/n!$. If $|t| \geq 1$, then $2|t|^n/n! \leq 2|t|^{n+\delta}/n!$. The lemma follows from these and Lemma 10.5.1. ∎

## 10.6 When is a function a c.f.?

**Example**.

(a) 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 $\psi(t) = e^{itb}$.

(b) $\psi(t) = e^{-|t|^\alpha}$ is not a c.f. for $\alpha > 2$ since $\psi'(t) = e^{-|t|^\alpha}\alpha|t|^\alpha$ and $\psi''(t) = e^{-|t|^\alpha}$ $\left[(\alpha|t|^{\alpha-1})^2 + \alpha(\alpha-1)|t|^{\alpha-2})\right]$, which gives $\psi''(0) = 0$. Therefore, we get $\psi(t) = e^{itb} \neq e^{-|t|^\alpha}$.

(c) 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. ∎

## 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

$$\sup_x |P(T_n \le x) - \Phi(x)| \to 0 \qquad \text{as } n \to \infty.$$

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:

$$F_{X+Y}(t) = \int_{-\infty}^{\infty} F_Y(t-y)dF_X(y).$$

Let $V_T$ be the d.f. with a p.d.f. (i.e. inverse triangular d.f.)

$$v_T(x) = \frac{1-\cos(Tx)}{\pi T x^2},$$

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

$$\omega_T(t) = \left(1 - \frac{t}{T}\right) I\{|t| \le T\}.$$

The explicit form of $\omega_T(t)$ is of no importance. What matters is that $\omega_T(t)$ *vanishes* for $|t| \ge T$, since this eliminates all questions of convergence.

For any function $\Delta(x)$, we denote its convolution with $V_T(x)$ by

$$\Delta^T(t) \equiv \Delta \star V_T(t) := \int_{-\infty}^{\infty} \Delta(t-x)v_T(x)dx. \tag{7.5}$$

Our objective is to estimate the maximum of $|\Delta|$ in terms of the maximum of $|\Delta^T|$.

Lemma **10.7.1** *Let $F$ be a d.f. and $G$ a function such that $G(-\infty)=0$, $G(\infty)=1$, and $\sup_x |G'(x)| \le \lambda < \infty$. Put*

$$\Delta(x) = F(x) - G(x)$$

*Then*

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

*Proof.* Denote

$$\eta = \sup_x |\Delta(x)|, \qquad \eta_T = \sup_x |\Delta^T(x)|, \qquad h = \frac{\eta}{2\lambda}.$$

Then it suffices to show that

$$\eta \le 2\eta_T + \frac{24\lambda}{\pi T}.$$

Since $\Delta(x) \to 0$ as $x \to \pm\infty$, 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,

$$\begin{aligned}
\eta_T &=: \sup_x |\Delta^T(x)| \ge \Delta^T(x_0+h) = \int_{-\infty}^{\infty} \Delta(x_0+h-x)v_T(x)dx \\
&= \left(\int_{|x|\le h} + \int_{|x|>h}\right) \Delta(x_0+h-x)v_T(x)dx \\
&= \int_{|x|\le h} \Delta(x_0+h-x)v_T(x)dx + \int_{|x|>h} \Delta(x_0+h-x)v_T(x)dx.
\end{aligned} \tag{7.6}$$

We will get a lower bound for $\Delta(x_0 + h - x)$ in the first and second integral.

(a) In the first integral, we have $|x| \le h$, which implies that $h - x \ge 0$. Thus,

$$\begin{aligned}
\Delta(x_0 + h - x) &= \Delta(x_0) + [\Delta(x_0 + h - x) - \Delta(x_0)] \\
&= \eta + [F(x_0 + h - x) - F(x_0)] - [G(x_0 + h - x) - G(x_0)] \\
&\ge \eta - [G(x_0 + h - x) - G(x_0)] \qquad (\text{as } h - x \ge 0) \\
&= \eta - G'(x_0 + \theta(h - x))(h - x) \qquad (\text{where } 0 \le \theta \le 1) \\
&\ge \eta - \lambda(h - x) \\
&= \frac{\eta}{2} + \lambda x. \qquad (\text{as } \lambda h = \eta/2)
\end{aligned} \tag{7.7}$$

(b) In the second integral, we have $|x| > h$. We will use the following trivial bound

$$\Delta(x_0 + h - x) \ge -\Delta(x_0) = -\eta. \tag{7.8}$$

Puting (7.7) and (7.8) into (7.6), we get

$$\begin{aligned}
\eta_T &\ge \int_{|x|\le h} \left(\frac{\eta}{2} + \lambda x\right) v_T(x)dx + \int_{|x|>h} (-\eta)\, v_T(x)dx \\
&= \frac{\eta}{2} P(|V_T| \le h) - \eta P(|V_T| > h) \\
&= \frac{\eta}{2} [1 - P(|V_T| > h)] - \eta P(|V_T| > h) \\
&= \frac{\eta}{2} - \frac{3\eta}{2} P(|V_T| > h).
\end{aligned}$$

Now

$$\begin{aligned}
P(|V_T| > h) &= \int_{|x|>h} \frac{1 - \cos(Tx)}{\pi T x^2} dx = \int_{|x|>h} \frac{2\sin^2(Tx/2)}{\pi T x^2} dx \\
&= \int_{|x|>h} \frac{\sin^2(Tx/2)}{\pi (Tx/2)^2} d(Tx/2) = \int_{|y|>Th/2} \frac{\sin^2 y}{\pi y^2} dy \\
&= \frac{2}{\pi} \int_{y=Th/2}^{\infty} \frac{\sin^2 y}{y^2} dy \\
&\le \frac{2}{\pi} \int_{y=Th/2}^{\infty} \frac{1}{y^2} dy \\
&= \frac{4}{\pi T h},
\end{aligned}$$

we finally have

$$\eta_T \ge \frac{\eta}{2} - \frac{3\eta}{2} \times \frac{4}{\pi T h} = \frac{\eta}{2} - \frac{6\eta}{\pi T h} = \frac{\eta}{2} - \frac{12\lambda}{\pi T}.$$

Namely

$$\eta \le 2\eta_T + \frac{24\lambda}{\pi T}. \quad \blacksquare$$

LEMMA **10.7.2 (Esseen's Smoothing Lemma)** *Let $F$ be a d.f. with vanishing expectation and c.f. $\psi_F(t)$. Suppose that $F - G$ vanishes at $\pm\infty$ and that $G$ has a derivative $G'$ such that $\sup_x |G'(x)| \le \lambda$. Finally, suppose that $g$ has a continuously differentiable Fourier transform $\psi_G$ such that $\psi_G(0) = 1$ and $\psi'_G(0) = 0$. Then for any $T > 0$,*

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

*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

$$\begin{aligned}
\Delta^T(x) &:= V_T \star [F(x) - G(x)] \\
&= V_T \star F(x) - V_T \star G(x) \\
&= \int_{-\infty}^{x} \left(f^T(y) - g^T(y)\right) dy \\
&= \int_{-\infty}^{x} \left(\frac{1}{2\pi}\int_{-T}^{T} e^{-ity}\left(\psi_F(t) - \psi_G(t)\right) w_T(t)dt\right) dy \\
&= \frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{x} e^{-ity}\left(\psi_F(t) - \psi_G(t)\right) w_T(t)dydt \\
&\qquad \text{(by Fubini theorem, with justifications give below)} \\
&= \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\frac{\psi_F(t) - \psi_G(t)}{-it} w_T(t)dt + C. \qquad (7.9)
\end{aligned}$$

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 (7.9) 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 (7.9) also goes to 0 as $|x| \to \infty$ by the Riemann-Lebesgue lemma. Hence we must have $C = 0$. Then from Lemma 10.7.2 and (7.9), we have

$$\begin{aligned}
\sup_x |F(x) - G(x)| &\leq \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{\psi_F(t) - \psi_G(t)}{t}\right| w_T(t)dt + \frac{24\lambda}{\pi T} \\
&\leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\psi_F(t) - \psi_G(t)}{t}\right| dt + \frac{24\lambda}{\pi T}. \qquad \blacksquare
\end{aligned}$$

## 10.8 Characteristic functions and smoothness condition

### Lattice d.f. and nonlattice d.f.

DEFINITION **10.8.1** *If all points of increase of $F$ are among $b, b \pm h, b \pm 2h, ....,$ then we say that $F$ is a* **lattice d.f.** *with span $h$.*

The following two theorems give a characterization of *lattice* distribution.

THEOREM **10.8.1** *If $\lambda \neq 0$, the following three statements are equivalent:*

(*a*) $\psi(\lambda) = 1$.

(*b*) *$\psi(t)$ has period $\lambda$, i.e., $\psi(t + n\lambda) = \psi(t)$ for all $t$ and $n$.*

(*c*) *All points of increase of $F$ are among $0, \pm h, \pm 2h, ....,$ where $h = 2\pi/\lambda$.*

*Proof.* We shall show that $(c) \to (b) \to (a) \to (c)$.

If (c) is true and $F$ attributes weight $p_k$ to $kh$, $k = 0, \pm 1, \pm 2, ...$, then $\psi(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikht}$, which has period $2\pi/h = \lambda$. So (c) implies (b).

If (b) is true, by taking $n = 1$ and $t = 0$, we get $\psi(\lambda) = \psi(0) = 1$, which proves (a).

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. ∎

It is easy to deduce the next corollary by applying the last lemma to $X - b$.

COROLLARY **10.8.1** *If $\lambda \neq 0$, the following three statements are equivalent:*

(*a*) $\psi(\lambda) = e^{ib\lambda}$.

(*b*) *$\psi(t)$ satisfies $\psi(t + n\lambda) = \psi(t)e^{in\lambda b}$ for all $t$ and $n$.*

(*c*) *All points of increase of $F$ are among $b, b \pm h, b \pm 2h, ....,$ where $h = 2\pi/\lambda$.*

The following result is thus immediate. It states that any distribution is either lattice, or nonlattice, or degenerate.

THEOREM **10.8.2** *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.** 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.,

$$\limsup_{|t|\to\infty} |\psi(t)| < 1. \quad \blacksquare$$

The next theorem shows that strong nonlatticeness is stronger than latticeness.

THEOREM **10.8.3** *If a d.f. $F$ is strongly non-lattice, it is non-lattice. 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, p342), 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 $\{m_k, k \ge 1\}$ and $\{n_k, k \ge 1\}$. Then $n_k$'s are strictly increasing to $\infty$ and

$$\left|a - \frac{m_k}{n_k}\right| < \frac{1}{\sqrt{5}\ n_k^2}, \qquad k = 1, 2, ....$$

Now choosing $t_k = 2\pi n_k$, we get

$$\begin{aligned}
\psi(t_k) &= p_0 + p_1 e^{i2\pi n_k} + p_2 e^{i2\pi n_k a} \\
&= p_0 + p_1 + p_2 e^{i2\pi n_k(\frac{m_k}{n_k} + [a - \frac{m_k}{n_k}])} \\
&= 1 - p_2 + p_2 e^{i2\pi m_k} e^{i2\pi n_k [a - \frac{m_k}{n_k}]} \\
&= 1 - p_2 + p_2 e^{i2\pi(n_k a - m_k)} \\
&= 1 + p_2 \left(e^{i2\pi(n_k a - m_k)} - 1\right).
\end{aligned}$$

Therefore,

$$|\psi(t_k) - 1| \le 2\pi p_2 |n_k a - m_k| < \frac{2\pi p_2}{\sqrt{5}\ n_k} \to 0. \quad \blacksquare$$

## Absolutely continuity implies strong non-latticeness

THEOREM **10.8.4** *If the d.f. $F$ has a non-zero absolutely continuous component, 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 $c_k \ge 0$ $(k = 1, 2, 3)$, $c_1 + c_2 + c_3 = 1$, 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, $0 < c_1 \le 1$ 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,

$$\limsup_{|t|\to\infty} |\psi(t)| \le c_1 \limsup_{|t|\to\infty} |\psi_1(t)| + c_2 + c_3 = c_2 + c_3 < 1. \quad \blacksquare$$

## 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 [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 $\{x_{kj} = j/3^k, k = 1, 2, ..., j = 1, 2, ..., \} \cap [0, 1]$ 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

$$\psi(t) = Ee^{itX} = \sum_{k,j\ge 1} \exp\{itx_{kj}\} p_{kj} = \sum_{k,j\ge 1} \exp\{itj/3^k)\} p_{kj}$$

Choosing $t = t_m = 2\pi 3^m \to \infty$ as $m \to \infty$, we get

$$\begin{aligned}
\psi(t_m) &= \sum_{k,j\geq 1} \exp\{i(2\pi 3^m)j/3^k)\}p_{kj} \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \exp\{i(2\pi j 3^{m-k}))\}p_{kj} \\
&= \left(\sum_{k=1}^{m} + \sum_{k=m+1}^{\infty}\right)\sum_{j=1}^{\infty} \exp\{i(2\pi j 3^{m-k}))\}p_{kj} \\
&= \sum_{k=1}^{m}\sum_{j=1}^{\infty} p_{kj} + \sum_{k=m+1}^{\infty}\sum_{j=1}^{\infty} \exp\{i(2\pi j 3^{m-k}))\}p_{kj} \\
&\longrightarrow 1.
\end{aligned}$$

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).

```
I. Discrete
   1. Lattice (e.g. Binomial, Poisson, Negative Binomial, Geometric)
   2. Nonlattice
      (a) Strongly nonlattice
          (e.g. an empirical distribution function (edf) with pdf f,
           such as a bootstrap sample from a absolute continuous dist.)
      (b) 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.)

II. Absolutely continuous    (which must be nonlattice)

III. Singular   (which is weakly nonlattice)

IV. Mixtures of I--III.
     (e.g. any d.f. with a non-zero absolutely continuous component is
      strongly nonlattice.)
```

## 10.9 Exercises

1. Give an example where $\int_{-\infty}^{\infty} |\psi(t)|dt = \infty$, but a p.d.f. still exists.

2. Let $\psi(t)$ be a c.f. of some r.v. $X$. For $a < 0$, if $\psi^a(t)$ is also a c.f., then $X$ must be degenerate.

3. Find two dependent r.v.s $X, Y$ such that $\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t)$.

   (Hint: choose $X = Y$ and use c.f. to find one example.)

   *Remark*: It is known that if $X, Y$ are i.i.d. with the same d.f. $F$, then $X + Y$ has distribution $F * F$ (i.e. convolution of $F$ with $F$). This exercise illustrates that the same conclusion may still hold when $X, Y$ have the same d.f. $F$, but not independent.

4. (a) Find the c.f. $\psi_X(t)$ if $P(X = \pm\alpha) = 1/2$.

   (b) Show by iteration of an elementary trigomonetric identity that

$$\frac{\sin t}{t} = \frac{\sin t/2^n}{t/2^n} \prod_{j=1}^{n} \cos\frac{t}{2^j} \longrightarrow \prod_{j=1}^{\infty} \cos\frac{t}{2^j}.$$

   (c) Utilize (b) to state a result on convergence in law of sums of independent r.v.s.

5. Suppose that $X_n \Longrightarrow X$ and $X_n \sim N(0, \sigma_n^2)$. Show that $\sigma_n^2 \to \sigma^2 \in [0, \infty)$.

6. Show that if $X_n$ and $Y_n$ are independent for $1 \le n \le \infty$, $X_n \Longrightarrow X_\infty$ and $Y_n \Longrightarrow Y_\infty$, then $X_n + Y_n \Longrightarrow X + Y$.

7. Let $X_1, X_2, ...$ be independent and let $S_n = X_1 + ... + X_n$. Let $\psi_j$ be the c.f. of $X_j$, and suppose that $S_n \to S_\infty$ a.s. Then $S_\infty$ has c.f. $\psi_\infty(t) = \prod_{j=1}^{\infty} \psi_j(t)$.

8. Show that if $\psi(t) = 1 + o(t^2)$ as $t \to 0$, then $\psi(t) \equiv 1$, namely, $X = 0$ a.s.

9. **($F_n(x) \to F(x)$ does not imply $F_n'(x) \to F'(x)$)** Let $X_n$ hav d.f.

$$F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, \qquad x \in [0, 1].$$

   (a) Show that $F_n$ is indeed a d.f. and find its p.d.f. $f_n$.

   (b) Show that $F_n \Longrightarrow U[0, 1]$, but $f_n$'s do not converge.

   **Remark.** The example shows that convergence of d.f.s does not imply that converge of their p.d.f.s., which is referred to as local limit theory. Extra conditions will be required for the latter to hold.

10. Let $X_1, X_2, ...$ be i.i.d. Cauchy r.v.s. Show that $\bar{X} = n^{-1} \sum_{j=1}^{n} X_j$ is still a Cauchy r.v.

11. (*The limit of c.f.s may not be a c.f.*) Verify that the c.f. corresponding to a uniform d.f. on $(-n, n)$ is $\psi_n(t) = \sin nt/(nt)$, and that $\lim_n \psi_n(t)$ exists but for all $t$, but $\psi(t)$ is not a c.f. Consequently, $F_n \not\Longrightarrow$ a d.f.

12. Show by Levy continuity theorem that, as $n \to \infty$,

    (a) if $X_n \sim Bin(n, \lambda/n)$, then $X_n \Longrightarrow$ a Poisson d.f.

    (b) if $Y_n$ is geometric with parameter $p = \lambda/n$, then $Y_n/n \Longrightarrow$ an exponential d.f.

13. Show that if $X$ and $Y$ are independent and $X + Y$ and $X$ have the same distribution, then $Y = 0$ a.s.

14. If $F_n \Longrightarrow F$ and $G_n \Longrightarrow G$, show that $F_n * G_n \Longrightarrow F * G$. (Proof of this without using c.f. is possible but tedious.)

15. Prove that a real-valued c.f. $\psi$ satisfies $\psi(t + h) \le \psi(h)$ for $t > 0$ and sufficiently small $h > 0$.

16. Prove that

    (a) the convolution of two discrete d.f.s is discrete.

    (b) the convolution of a continuous d.f. with any d.f. is continuous.

(c) (optional) the convolution of an absolutely continuous d.f. with any d.f. is absolutely continuous.

REMARK **10.9.1** *Note that (b) and (c) are the basis for kernel or other type of smoothing techniques commonly used in statistics.*

# Chapter 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.

## 11.1 Central Limit Theorems (CLT)

### 11.1.1 CLT for i.i.d. r.v.s

When $X_1, ..., X_n$ are i.i.d., we get the following simple CLT.

THEOREM **11.1.1 (Levy theorem)** *Let $X_1, ..., X_n$ be i.i.d. r.v.'s with $EX_1 = 0$, and $\sigma^2 = EX_1^2 < \infty$. Let $F_n(x) = P\left(\sqrt{n}\bar{X}/\sigma \le x\right)$. Then*

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \to 0,$$

*Proof.* We provide two methods.

**Method 1.** Since $EX_1^2 < \infty$, $\psi(t)$ is twice differentiable and has the following Taylor expansion,

$$\psi(t) = \psi(0) + \psi'(0)t + \frac{1}{2}\psi''(0)t^2 + o(t^2) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

Then the c.f. of $F_n$ is

$$\psi_{\sqrt{n}\bar{X}/\sigma}(t) = \psi^n\left(\frac{t}{\sqrt{n}\sigma}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-t^2/2}. \quad \blacksquare$$

**Method 2.** The Lindeberg condition in Corollary 11.1.2 holds since

$$\frac{1}{B_n^2}\sum_{k=1}^n EX_k^2 I\{|X_k| \ge \epsilon B_n\} \quad = \quad \frac{EX_1^2 I\{|X_1| \ge \epsilon\sqrt{nEX_1^2}\}}{EX_1^2} \to 0. \quad \blacksquare$$

### 11.1.2 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 **11.1.2 (Lindeberg-Feller CLT)** *For each $n$, let $X_{n,k}$, $1 \le k \le n$, 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 $F_n(x) = P(\sum_{m=1}^n X_{n,k} \le x)$. Then the following two statements are equivalent.*

*(i) The Lindeberg condition holds:*

$$\forall \epsilon > 0: \quad \lim_{n\to\infty} \sum_{m=1}^n EX_{n,k}^2 I\{|X_{n,k}| \ge \epsilon\} = 0.$$

*(ii) (a)* $\max_{1\le m\le n} \sigma_{n,k}^2 \to 0$ *and (b)* $\sup_{x\in R} |F_n(x) - \Phi(x)| \to 0$.

REMARK **11.1.1** *Clearly, the Lindeberg condition implies that*

$$\forall \epsilon > 0: \quad \lim_{n\to\infty} \sum_{m=1}^n P(|X_{n,k}| \ge \epsilon) = 0,$$

*which further implies that*

$$\forall \epsilon > 0: \quad \sup_n P(|X_{n,k}| \ge \epsilon) \longrightarrow 0.$$

*That is, all the individual terms $X_{n,k}$ are uniformly small.*

*Proof.* "(i) $\Longrightarrow$ (ii)". We first prove (a) of (ii). For $1 \le k \le n$, we have

$$\begin{aligned} \sigma_{n,k}^2 &= EX_{n,k}^2 I\{|X_{n,k}| < \epsilon\} + EX_{n,k}^2 I\{|X_{n,k}| \ge \epsilon\} \\ &\le \epsilon^2 + EX_{n,k}^2 I\{|X_{n,k}| \ge \epsilon\}. \end{aligned}$$

Thus $\max_{1\le k\le n} \sigma_{n,k}^2 \le \epsilon^2 + \sum_{k=1}^n EX_{n,k}^2 I\{|X_{n,k}| \ge \epsilon\}$. Letting $n \to \infty$, the Lindeberg condition implies that $\max_{1\le k\le n} \sigma_{n,k}^2 \le 2\epsilon^2$. Since $\epsilon$ can be chosen to be arbitrarily small, we have

$$\max_{1\le m\le n} \sigma_{n,k}^2 \to 0.$$

We now prove (b) of (ii). Write $\psi_{n,k}(t) = Ee^{itX_{n,k}}$. It suffices to show that, $\forall t \in R$,

$$\prod_{k=1}^n \psi_{n,k}(t) \longrightarrow e^{-t^2/2} \qquad \Longleftrightarrow \qquad \sum_{k=1}^n \ln \psi_{n,k}(t) + t^2/2 \longrightarrow 0 \tag{1.1}$$

It suffices to show that, as $n \to \infty$, $\forall t \in R$,

$$\begin{aligned} \sum_{k=1}^n \ln \psi_{n,k}(t) - \sum_{k=1}^n (\psi_{n,k}(t) - 1) &\to 0, & (1.2) \\ \sum_{k=1}^n (\psi_{n,k}(t) - 1) + \frac{t^2}{2} &\to 0. & (1.3) \end{aligned}$$

Let us prove (1.2) first. From the inequality $|e^{it} - 1 - it| \le t^2/2$ for any real $t$, we have

$$|\psi_{n,k}(t) - 1| = \left|Ee^{itX_{n,k}} - 1 - itEX_{n,k}\right| \le E\left|e^{itX_{n,k}} - 1 - itX_{n,k}\right| \le \frac{1}{2}t^2 EX_{n,k}^2 = \frac{1}{2}t^2\sigma_{n,k}^2.$$

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 $|z|\le 1/2$, (1.2) follows from

$$\sum_{k=1}^n |\ln\psi_{n,k}(t)-(\psi_{n,k}(t)-1)| \le \sum_{k=1}^n |\psi_{n,k}(t)-1|^2 \le \quad o(1)\textstyle\sum_{k=1}^n |\psi_{n,k}(t)-1| = o(1).$$

Next let us prove (1.3). By using the inequality $\left|e^{it}-1-it-\frac{1}{2}(it)^2\right|\le \min\{t^2, \frac{1}{6}|t|^3\}$ for any real $t$, we have

$$\begin{aligned}
\left|\sum_{k=1}^n (\psi_{n,k}(t)-1)+\frac{t^2}{2}\right| &= \left|\sum_{k=1}^n E\left(e^{itX_{n,k}}-1-itX_{n,k}-\frac{1}{2}(itX_k)^2\right)\right| \\
&\le \sum_{k=1}^n E\min\left\{t^2X_{n,k}^2, \frac{1}{6}|tX_{n,k}|^3\right\} \\
&\le t^2\sum_{k=1}^n EX_{n,k}^2 I\{|X_k|\ge\epsilon\}+\frac{|t|^3\epsilon}{6}\sum_{k=1}^n E|X_{n,k}|^2 I\{|X_{n,k}|<\epsilon\} \\
&\le t^2\sum_{k=1}^n EX_{n,k}^2 I\{|X_k|\ge\epsilon\}+\frac{1}{6}|t|^3\epsilon.
\end{aligned}$$

Then (1.3) follows from this, the Lindeberg condition and by choosing $\epsilon$ arbitrarily small.

"(ii) $\Longrightarrow$ (i)". Assume that (ii) holds. First, part (b) of (ii) implies (1.1). Secondly, from the proceeding proof, we can see that that (1.2) is implied from part (a) of (ii). Putting these two together, we see that (1.3) still holds. In particular, the real part of the left hand side in (1.3) should tend to 0, i.e.,

$$\begin{aligned}
0 \quad \longleftarrow \quad & Re\left(\sum_{k=1}^n (\psi_{n,k}(t)-1)+\frac{t^2}{2}\right) \\
= \quad & \sum_{k=1}^n E\left(\cos(tX_{n,k})-1+\frac{1}{2}t^2X_{n,k}^2\right) \\
\ge \quad & \sum_{k=1}^n E\left(\cos(tX_{n,k})-1+\frac{1}{2}t^2X_{n,k}^2\right) I\{|X_{n,k}|\ge\epsilon\} \\
& \qquad\qquad (\text{as } \cos(y)-1+\tfrac{1}{2}y^2\ge 0) \\
\ge \quad & \sum_{k=1}^n E\left(\frac{1}{2}t^2X_{n,k}^2-2\right) I\{|X_{n,k}|\ge\epsilon\} \qquad (\text{as } \cos(y)\ge -1) \\
= \quad & \sum_{k=1}^n EX_{n,k}^2\left(\frac{1}{2}t^2-\frac{2}{X_{n,k}^2}\right) I\{|X_{n,k}|\ge\epsilon\} \\
\ge \quad & \left(\frac{t^2}{2}-\frac{2}{\epsilon^2}\right)\sum_{k=1}^n EX_{n,k}^2 I\{|X_{n,k}|\ge\epsilon\},
\end{aligned}$$

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. ∎

The next theorem is an easy consequence of Lindeberg-Feller CLT.

THEOREM **11.1.3 (Lyapunov CLT)** *For each $n$, let $X_{n,k}$, $1\le k\le n$, 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 $F_n(x)=P(\sum_{m=1}^n X_{n,k}\le x)$. If*

$$\sum_{k=1}^n E|X_{n,k}|^{2+\delta}\to 0, \tag{1.4}$$

*then*

$$\sup_{x\in R}|F_n(x)-\Phi(x)|\to 0.$$

*Proof.* The Lindeberg condition holds since

$$\begin{aligned}\sum_{k=1}^{n} EX_{n,k}^2 I\{|X_{n,k}| \geq \epsilon\} &\leq \frac{1}{\epsilon^\delta}\sum_{k=1}^{n} E|X_{n,k}|^{2+\delta} I\{|X_{n,k}| \geq \epsilon\} \\ &\leq \frac{1}{\epsilon^\delta}\sum_{k=1}^{n} E|X_{n,k}|^{2+\delta} \to 0. \quad \blacksquare\end{aligned}$$

**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 $var(X_j) = \sigma_j^2 < \infty$, $j = 1, ..., n$. Let

$$S_n = \sum_{j=1}^{n} X_j, \qquad B_n^2 = \sum_{j=1}^{n} \sigma_j^2.$$

and

$$F_n(x) = P\left(S_n/B_n \leq x\right).$$

Corollary **11.1.1 (Lindeberg-Feller CLT)** *The following two statements are equivalent.*

*(i) The Lindeberg condition holds: for every fixed* $\epsilon > 0$,

$$B_n^{-2}\sum_{k=1}^{n} EX_k^2 I\{|X_k| \geq \epsilon B_n\} \to 0, \qquad \text{as } n \to \infty$$

*(ii) (a)* $\max_{1\leq k\leq n}\left\{\sigma_k^2/B_n^2\right\} \to 0$ *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 11.1.2. ∎

Corollary **11.1.2 (Lyapunov CLT)** *If*

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n} E|X_k|^{2+\delta} \to 0, \tag{1.5}$$

*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 11.1.3. ∎

**Remarks.**

(1) Both the Lindeberg condition in Theorem 11.1.2 and the Lyapunov condition (1.5) in Theorem 11.1.3 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 $k \geq 2$. 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)

$$\sup_x |F_n(x) - \Phi(x)| \leq A\left(\Lambda_n(\epsilon) + \epsilon\right) \qquad \text{for any } \epsilon > 0,$$

where

$$\Lambda_n(\epsilon) = B_n^{-2}\sum_{k=1}^{n} EX_k^2 I\{|X_k| \geq \epsilon B_n\}.$$

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 11.1.1 is equivalent to

$$\Lambda_n'(\epsilon) := B_n^{-2} \sum_{k=1}^n EX_k^2 I\{|X_k| \ge \epsilon B_k\} \to 0,$$

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 $\delta > 0$,

$$\begin{aligned}
\Lambda_n'(\epsilon) &= B_n^{-2}\left(\sum_{\{k:B_k\le\delta B_n\}} + \sum_{\{k:B_k>\delta B_n\}}\right) EX_k^2 I\{|X_k| \ge \epsilon B_k\} \\
&\le B_n^{-2} \sum_{\{k:B_k\le\delta B_n\}} EX_k^2 + \sum_{k=1}^n EX_k^2 I\{|X_k| \ge \epsilon\delta B_n\} \\
&\le B_n^{-2} B_m^2 + \Lambda_n(\epsilon) \\
&\quad \text{(where } m = \sup\{k : B_k \le \delta B_n\}\text{. Note: } B_1^2 \le B_2^2 \le \ldots \le B_n^2.\text{)} \\
&= \delta^2 + \Lambda_n(\epsilon) \\
&\to \delta^2.
\end{aligned}$$

### 11.1.3 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 **11.1.4** *If $X, \{X_n, n \ge 1\}$ are i.i.d. r.v.s with non-degenerate d.f. $F$, then*

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{i=1}^n X_i - A_n \le x\right) = \Phi(x)$$

*for some $B_n > 0$ and $A_n$ iff*

$$\lim_{C\to\infty} \frac{P(|X| > C)}{C^{-2}EX^2 I\{|X| \le C\}} = 0;$$

*Moreover, $A_n, B_n$ may be chosen as*

$$\begin{aligned}
B_n &= \sup\{C : C^{-2}EX^2 I\{|X| \le C\} \ge \frac{1}{n}\}, \\
A_n &= \frac{n}{B_n} EXI\{|X| < B_n\}.
\end{aligned}$$

*Proof.* See Theorem 4, Chow and Teicher, p323. ∎

## 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 $X_1, ..., X_n$ be independent r.v.'s such that $EX_j = 0$ and $E|X_j|^{2+\delta} < \infty$ for some $0 < \delta \le 1$ $(j = 1, ..., n)$. Put

$$EX_j^2 = \sigma_j^2, \qquad B_n^2 = \sum_{j=1}^n \sigma_j^2, \qquad L_{n,\delta} = \frac{\sum_{j=1}^n E|X_j|^{2+\delta}}{B_n^{2+\delta}}.$$

In particular, in the i.i.d. case, where $EX_j^2 = \sigma^2$, we have

$$B_n^2 = n\sigma^2, \qquad L_{n,\delta} = \frac{nE|X_1|^{2+\delta}}{(n\sigma^2)^{(2+\delta)/2}} = \frac{C}{n^{\delta/2}}, \qquad L_{n,1} = \frac{C}{n^{1/2}}.$$

### 11.2.1 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 **11.2.1**

$$|\psi_n(t) - e^{-t^2/2}| \le 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/2} \qquad \textit{for } |t| < \tfrac{1}{2}L_{n,\delta}^{-1/(2+\delta)}. \tag{2.6}$$

**Remark:** In the i.i.d. case, $L_{n,\delta}^{-1/(2+\delta)} = Cn^{\delta/[2(2+\delta)]}$.

*Proof.* Note that

$$\begin{aligned}
\left|\psi_n(t) - e^{-t^2/2}\right| &= \left|\prod_{k=1}^n \psi_{n,k}(t) - e^{-t^2/2}\right| \\
&= \left|\exp\left\{\sum_{k=1}^n \ln\psi_{n,k}(t)\right\} - e^{-t^2/2}\right| \\
&= e^{-t^2/2}\left|\exp\left\{\sum_{k=1}^n \ln\psi_{n,k}(t) + \frac{t^2}{2}\right\} - 1\right| 
\end{aligned} \tag{2.7}$$

Using Theorem 10.5.3, we have

$$\begin{aligned}
\psi_{n,k}(t) &= 1 + itEX_{n,k} + \frac{1}{2}(it)^2EX_{n,k}^2 + \theta|t|^{2+\delta}E|X_{n,k}|^{2+\delta} \\
&= 1 - \frac{1}{2}t^2\sigma_{n,k}^2 + \theta|t|^{2+\delta}E|X_{n,k}|^{2+\delta},
\end{aligned}$$

where $\theta$ is a complex number with $|\theta| \le 1$. Noting that

$$|t|\sigma_{n,k} \le |t|\left(E|X_{n,k}|^{2+\delta}\right)^{1/(2+\delta)} \le |t|L_{n,\delta}^{1/(2+\delta)} < \frac{1}{2}.$$

Thus,

$$\begin{aligned}
|\psi_{n,k}(t) - 1| &\le \frac{1}{2}\times\frac{1}{4} + \frac{1}{2^{2+\delta}} < \frac{3}{8}, \\
|\psi_{n,k}(t) - 1|^2 &\le 2\left[\left(\frac{1}{2}t^2\sigma_{n,k}^2\right)^2 + \left(|t|^{2+\delta}E|X_{n,k}|^{2+\delta}\right)^2\right] \\
&= 2\left(\frac{1}{8}|t\sigma_{n,k}|^{2+\delta}|t\sigma_{n,k}|^{2-\delta} + \left(|t|^{2+\delta}E|X_{n,k}|^{2+\delta}\right)\left(|t|^{2+\delta}E|X_{n,k}|^{2+\delta}\right)\right) \\
&\le 2\left(\frac{1}{8} + \frac{1}{2^{2+\delta}}\right)|t|^{2+\delta}E|X_{n,k}|^{2+\delta} \\
&= \frac{3}{4}|t|^{2+\delta}E|X_k|^{2+\delta}.
\end{aligned}$$

By Taylor expansion of $\log(1+z)$, we can easily find that (see Appendix)

$$\log(1+z) = z + \frac{4}{5}\theta|z|^2, \qquad \text{where } |\theta| \le 1 \text{ and } |z| < \tfrac{3}{8}.$$

Hence, taking $z = \psi_{n,k}(t) - 1$, we get

$$\begin{aligned}\ln \psi_{n,k}(t) &= \ln\left(1 + [\psi_{n,k}(t) - 1]\right) = (\psi_{n,k}(t) - 1) + \frac{4}{5}\theta_k|\psi_{n,k}(t) - 1|^2 \\ &= -\frac{1}{2}t^2\sigma_{n,k}^2 + \frac{8}{5}\theta_k|t|^{2+\delta}E|X_{n,k}|^{2+\delta}, \qquad |\theta_k| \le 1,\end{aligned}$$

and thus

$$s =: \sum_{k=1}^{n} \ln \psi_{n,k}(t) + \frac{t^2}{2} = \sum_{k=1}^{n} \ln(1 + r_k) + \frac{t^2}{2} = \frac{8}{5}\theta|t|^{2+\delta}L_{n,\delta}, \qquad |\theta| \le \max_{1\le k\le n} |\theta_k| \le 1.$$

Clearly, $|s| \le 2/5$. So from (2.7) and the inequality $|e^z - 1| \le |z|e^{|z|}$ for every complex $z$, we get

$$\begin{aligned}\left|\psi_n(t) - e^{-t^2/2}\right| &\le e^{-t^2/2}|s|e^{|s|} \\ &\le e^{-t^2/2}\frac{8}{5}|t|^{2+\delta}L_{n,\delta}e^{2/5} \\ &\le 3|t|^{2+\delta}L_{n,\delta}e^{-t^2/2}. \quad \blacksquare\end{aligned}$$

Lemma **11.2.2**

$$|\psi_n(t) - e^{-t^2/2}| \le 16L_{n,\delta}|t|^{2+\delta}e^{-t^2/3} \qquad \textit{for } |t| \le \left(\frac{1}{36L_{n,\delta}}\right)^{1/\delta}. \tag{2.8}$$

*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. $|\psi_{n,j}(t)|^2$ and the variance $2\sigma_{n,j}^2$. Furthermore, by using the inequality $|a + b|^r \le 2^{r-1}(|a|^r + |b|^r)$, we have

$$E|X_{n,j}^s|^{2+\delta} \le 2^{1+\delta}\left(E|X_{n,j}|^{2+\delta} + E|Y_{n,j}|^{2+\delta}\right) \le 8E|X_{n,j}|^{2+\delta}.$$

Using Theorem 10.5.3 and the inequality $e^x \ge 1 + x$ for all real $x$, we have

$$\begin{aligned}|\psi_{n,j}(t)|^2 &= 1 + (it)EX_{n,j}^s - \frac{1}{2}t^2E(X_{n,j}^s)^2 + \theta|t|^{2+\delta}E|X_{n,j}^s|^{2+\delta} \\ &\le 1 - \sigma_{n,j}^2t^2 + 8|t|^{2+\delta}E|X_{n,j}|^{2+\delta} \\ &\le \exp\left\{-\sigma_{n,j}^2t^2 + 8|t|^{2+\delta}E|X_{n,j}|^{2+\delta}\right\},\end{aligned} \tag{2.9}$$

From the assumption that $|t|^\delta \le 1/(36L_{n,\delta})$, we have

$$\begin{aligned}|\psi_n(t)|^2 &= \prod_{j=1}^{n} |\psi_{n,j}(t)|^2 \le \exp\left\{-t^2 + 8|t|^{2+\delta}L_{n,\delta}\right\} \le \exp\left\{-t^2 + 8t^2/36\right\} \\ &= \exp\left\{-t^2 + 2t^2/9\right\} = \exp\left\{-7t^2/9\right\} \le \exp\{-2t^2/3\}.\end{aligned} \tag{2.10}$$

(a) If $|t| < (2L_{n,\delta}^{1/(2+\delta)})^{-1}$, then from Lemma 11.2.1,

$$|\psi_n(t) - e^{-t^2/2}| \le 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/2} \le 3L_{n,\delta}|t|^{2+\delta}e^{-t^2/3}.$$

(b) If $(2L_{n,\delta}^{1/(2+\delta)})^{-1} \le |t| \le (36L_{n,\delta})^{-1/\delta}$, then $(2^{2+\delta})^{-1} \le |t|^{2+\delta}$, i.e., $|t|^{2+\delta}L_{n,\delta} \ge 2^{-(2+\delta)} \ge 2^{-3} = 1/8$. Thus, then from (2.10),

$$|\psi_n(t) - e^{-t^2/2}| \le |\psi_n(t)| + e^{-t^2/2} \le 2\exp\{-t^2/3\} \le 16L_{n,\delta}|t|^{2+\delta}e^{-t^2/3}. \quad \blacksquare$$

If we take $\delta = 1$ above, we have

Corollary **11.2.1** *Let $X_1, ..., X_n$ be independent r.v.'s such that $EX_j = 0$ and $E|X_j|^3 < \infty$ ($j = 1, ..., n$). Put*

$$EX_j^2 = \sigma_j^2, \qquad B_n^2 = \sum_{j=1}^{n} \sigma_j^2, \qquad L_{n,1} = B_n^{-3/2}\sum_{j=1}^{n} E|X_j|^3.$$

*Let $\psi_n(t)$ be the c.f. of the random variable $B_n^{-1/2}\sum_{j=1}^n X_j$. Then*

$$(1) \quad |\psi_n(t) - e^{-t^2/2}| \le 3|t|^3 L_{n,1} e^{-t^2/2} \qquad \textit{for } |t| < \tfrac{1}{2L_{n,1}^{1/3}},$$
$$(2) \quad |\psi_n(t) - e^{-t^2/2}| \le 16|t|^3 L_{n,1} e^{-t^2/3} \qquad \textit{for } |t| \le \tfrac{1}{36L_{n,1}}. \quad \blacksquare$$

The corollary becomes even more apparent in the i.i.d. case.

COROLLARY **11.2.2** *Let $X_1, ..., X_n$ be i.i.d. r.v.'s. Let*

$$EX_1 = 0, \qquad EX_1^2 = \sigma^2 > 0, \qquad E|X_1|^3 < \infty, \qquad \rho = E|X_1|^3/\sigma^3.$$

*Let $\psi_n(t)$ be the c.f. of the random variable $\sigma^{-1}n^{-1/2}\sum_{j=1}^n X_j$. Then*

$$(1) \quad |\psi_n(t) - e^{-t^2/2}| \le 3|t|^3 \rho e^{-t^2/2} n^{-1/2} \qquad \textit{for } |t| < (2\rho^{1/3})^{-1} n^{1/6},$$
$$(2) \quad |\psi_n(t) - e^{-t^2/2}| \le 16|t|^3 \rho e^{-t^2/3} n^{-1/2} \qquad \textit{for } |t| \le (36\rho)^{-1} n^{1/2}. \quad \blacksquare$$

### 11.2.2 Berry-Esseen bounds

THEOREM **11.2.1 (Berry-Esseen bounds for independent r.v.'s)** *Let $X_1$, ..., $X_n$ be independent r.v.'s such that $EX_j = 0$ and $E|X_j|^{2+\delta} < \infty$ ($j = 1, ..., n$) for some $0 < \delta \le 1$. Put*

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

*Then for all $n$,*

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

*Proof.* In the Smoothing Lemma we set $T = (36L_{n,\delta})^{-1/\delta}$. Note that $\Phi'(x) = \phi(x) = (2\pi)^{-1/2}e^{-x^2/2} \le (2\pi)^{-1/2}$. Then we have

$$\begin{aligned}
\sup_x |F_n(x) - \Phi(x)| &\le \frac{2}{\pi}\int_0^T |t|^{-1}\left|\psi_n(t) - e^{-t^2/2}\right| dt + \frac{24\lambda}{\pi T} \\
&\le \pi^{-1}\int_0^T 16L_{n,\delta}t^{1+\delta}e^{-t^2/3}dt + \frac{24\lambda}{\pi}(36L_{n,\delta})^{1/\delta} \\
&\le C_\delta L_{n,\delta} + C_\delta L_{n,\delta}^{1/\delta}.
\end{aligned}$$

If $L_{n,\delta}^{1/\delta} \le 1$, then $\sup_x |F_n(x) - \Phi(x)| \le 2C_\delta L_{n,\delta}$ with $A = 2C_\delta$. On the other hand, if $L_{n,\delta}^{1/\delta} > 1$, then we can simply take $A = 1$. $\blacksquare$

In the i.i.d. case, Theorem 11.2.1 reduces to the following corollary.

COROLLARY **11.2.3 (Berry-Esseen bounds for i.i.d. r.v.'s)** *Let $X_1, ..., X_n$ be i.i.d. r.v.'s. Let $\delta \in (0, 1]$, and*

$$EX_1 = 0, \qquad EX_1^2 = \sigma^2 > 0, \qquad E|X_1|^{2+\delta} < \infty, \qquad \rho_\delta = E|X_1|^{2+\delta}/\sigma^{2+\delta}.$$

*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 **11.2.4** *Let $X_1, ..., X_n$ be i.i.d. r.v.'s. Let*

$$EX_1 = 0, \qquad EX_1^2 = \sigma^2 > 0, \qquad E|X_1|^3 < \infty, \qquad \rho = E|X_1|^3/\sigma^3.$$

*Then for all $n$,*

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \leq \frac{A\rho}{\sqrt{n}}.$$

**Remarks**.

(1). The constant $A$ in Theorem 11.2.4 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 Theorem 11.2.4 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 **11.2.1** *Let $P(X_i = \pm 1) = 1/2$, $i = 1, ..., n$. 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*

$$\begin{aligned} P(S_n = 0) &= P\left(\sum_{i=1}^n I\{X_i = -1\} = m\right) = \binom{n}{m}\frac{1}{2^n} = \frac{1}{2^n}\frac{n!}{m!^2} \\ &= \frac{\sqrt{2\pi n}(n/e)^n(1+o(1))}{\left(\sqrt{2\pi m}(m/e)^m(1+o(1))\right)^2} \cdot \frac{1}{2^n} \\ &= \frac{\sqrt{2}}{\sqrt{\pi n}}(1 + o(1)). \end{aligned}$$

*That is, the d.f. of $S_n$ has a jump of size $O(n^{-1/2})$ at 0, and can not be approximated by a continuous function at 0 with error size of less than $O(n^{-1/2})$. More precisely, we have*

$$\begin{aligned} F_n(0) &= P\left((S_n - ES_n)/\sqrt{var(S_n)} \leq 0\right) = P(S_n \leq 0) \\ &= \frac{1}{2}\left(P(S_n \leq 0) + P(S_n \geq 0)\right) \qquad \textit{(by symmetry)} \\ &= \frac{1}{2}\left(P(S_n \leq 0) + P(S_n > 0) + P(S_n = 0)\right) = \frac{1}{2}\left(1 + P(S_n = 0)\right) \\ &= \Phi(0) + \frac{1}{\sqrt{2\pi n}}(1 + o(1)). \end{aligned}$$

*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. $\Phi$ 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 $O(n^{-1/2})$. This is because we are approximating a distribution with jump size $O(n^{-1/2})$ 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 11.2.4 is not less than $(2\pi n)^{-1/2} \approx 0.4$. This is also true for other Berry-Esseen bounds.

### 11.2.3 A generalization of Berry-Esseen bounds

Here we shall give Berry-Esseen bounds assuming only second-order moments.

THEOREM **11.2.2** *Let $X_1$, ..., $X_n$ be independent r.v.'s such that $EX_j = 0$ and $E|X_j|^{2+\delta} < \infty$ ($j = 1, ..., n$). Put*

$$\Lambda_n(\epsilon) = B_n^{-2} \sum_{j=1}^{n} E|X_j|^2 I\{|X_j| > \epsilon B_n\},$$

$$\lambda_n(\epsilon) = B_n^{-3} \sum_{j=1}^{n} E|X_j|^3 I\{|X_j| \le \epsilon B_n\}.$$

*Then for all $n$ and $\epsilon > 0$,*

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \le A\left(\Lambda_n(\epsilon) + \lambda_n(\epsilon)\right). \quad \blacksquare$$

*Proof.* Omitted. $\blacksquare$

**Remarks.**

- From the definition, we have, for all $\epsilon > 0$,

$$\lambda_n(\epsilon) \le \epsilon B_n^{-2} \sum_{j=1}^{n} E|X_j|^2 I\{|X_j| \le \epsilon B_n\} \le \epsilon.$$

  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 $\epsilon \to 0$.

- On the other hand, we note that, for every $\delta \in (0, 1]$,

$$\Lambda_n(\epsilon) + \lambda_n(\epsilon) \le \frac{\sum_{j=1}^{n} E|X_j|^{2+\delta}}{B_n^{2+\delta}}.$$

  Then we derive the Berry-Esseen bounds

$$\sup_{x \in R} |F_n(x) - \Phi(x)| \le \frac{A \sum_{j=1}^{n} E|X_j|^{2+\delta}}{B_n^{2+\delta}},$$

  which, in turn, implies the Lynapnov CLT.

## 11.3 Non-Uniform Berry-Esseen Bounds

We have seen in Section 11.2 that the minimal condition for Berry-Esseen bounds of order $O(n^{-1/2})$ is the existence of the third moment. Also we have seen that higher moment conditions can not improve the error order $O(n^{-1/2})$ 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 $r \ge 2$, then we have, for $x > 0$,

$$\begin{aligned}
|F_n(x) - \Phi(x)| &= \max\{\Phi(x) - F_n(x), F_n(x) - \Phi(x)\} \\
&\le \max\{1 - F_n(x), 1 - \Phi(x)\} \\
&= \max\{P(\sqrt{n}\bar{X} > x), P(N(0,1) > x)\} \\
&\le \frac{1}{x^r} \max\{E|\sqrt{n}\bar{X}|^r, E|N(0,1)|^r\} \\
&\le \frac{C}{x^r},
\end{aligned}$$

where the last line follows from Rosenthal inequality (e.g., Petrov, 1995, p59):

$$\begin{aligned} E|\sqrt{n}\bar{X}|^r &= n^{-r/2}E|\sum_1^n X_k|^r \\ &\le n^{-r/2}C_r\left(\sum_1^n E|X_k|^r + \left(\sum_1^n EX_k^2\right)^{r/2}\right) \\ &\le n^{-r/2}C\left(n+n^{r/2}\right) \le C \end{aligned}$$

Similarly, it is also true for $x<0$. Therefore, for all $x$, we have

$$|F_n(x)-\Phi(x)| \le \frac{C}{(1+|x|)^r}.$$

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 **11.3.1** *Let $X_1,...,X_n$ be i.i.d. r.v.'s. Let*

$$EX_1=0, \qquad EX_1^2=\sigma^2>0, \qquad E|X_1|^3<\infty, \qquad \rho = E|X_1|^3/\sigma^3.$$

*and*

$$F_n(x) = P\left(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n X_j \le x\right).$$

*Then for all $x$ and $n$,*

$$|F_n(x)-\Phi(x)| \le \frac{A\rho}{\sqrt{n}}\frac{1}{(1+|x|^3)}. \quad ∎$$

For the independent case, we have

THEOREM **11.3.2** *Let $X_1,...,X_n$ be independent r.v.'s such that $EX_j=0$ and $E|X_j|^3<\infty$ $(j=1,...,n)$. Put*

$$EX_j^2=\sigma_j^2, \qquad B_n^2=\sum_{j=1}^n \sigma_j^2, \qquad L_n = B_n^{-3/2}\sum_{j=1}^n E|X_j|^3.$$

$$F_n(x) = P\left(B_n^{-1}\sum_{j=1}^n X_j \le x\right).$$

*Then for all $n$ and $x$,*

$$|F_n(x)-\Phi(x)| \le \frac{AL_n}{1+|x|^3}. \quad ∎$$

THEOREM **11.3.3** *Let $X_1,...,X_n$ be i.i.d. with $E|X_1|^r<\infty$, $r\ge 3$. Then for all $x$ and $n$,*

$$|F_n(x)-\Phi(x)| \le C_r\left(\frac{E|X_1|^3/\sigma^3}{\sqrt{n}} + \frac{E|X_1|^r/\sigma^r}{n^{(r-2)/2}}\right)\frac{1}{(1+|x|)^r}. \quad ∎$$

THEOREM **11.3.4** *Let $X_1,...,X_n$ be i.i.d. with $E|X_1|^{2+\delta}<\infty$, $\delta\in(0,1]$. Then for all $x$ and $n$,*

$$|F_n(x)-\Phi(x)| \le \frac{C_\delta E|X_1|^{2+\delta}}{\sigma^{2+\delta}n^{\delta/2}}\frac{1}{1+|x|^{2+\delta}}. \quad ∎$$

## 11.4 Edgeworth Expansions

The CLT can be established under second moment condition. Under higher moment condition, Berry-Esseen bounds give convergence rates to normality. For instance, when $E|X_1|^3<\infty$, the error in the normal approximation is of size $O(n^{-1/2})$. 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.

### 11.4.1 Heuristic argument for informal Edgeworth expansions

For simplicity, assume that $X, X_1, X_2, ...$ are i.i.d. r.v.s with $EX = 0$, $EX^2 = 1$, and $\rho = EX^3$. Let

$$F_n(x) = P(\sqrt{n}(\bar{X} - 0)/1 \le x) = P(\sqrt{n}\bar{X} \le x).$$

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 $\psi_X(t) = 1 + itEX + (it)^2 EX^2 + \frac{1}{6}(it)^3 EX^3 + o(t^3) = 1 - t^2 + \frac{1}{6}(it)^3\rho + o(t^3)$. Hence,

$$\psi_{\sqrt{n}\bar{X}}(t) = \psi_X^n(t/\sqrt{n}) = \left(1 - \frac{1}{2n}t^2 + \frac{1}{6n^{3/2}}(it)^3\rho + n^{-3/2}o(t^3)\right)^n$$

Hence, using $\ln(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} + ...$, we have

$$\begin{aligned}
\ln \psi_{\sqrt{n}\bar{X}}(t) &= n \ln\left(1 - \frac{1}{2n}t^2 + \frac{1}{6n^{3/2}}(it)^3\rho + n^{-3/2}o(t^3)\right) \\
&= n\left(-\frac{1}{2n}t^2 + \frac{1}{6n^{3/2}}(it)^3\rho + n^{-3/2}o(t^3) + ...\right) \\
&= -\frac{t^2}{2} + \frac{1}{6n^{1/2}}(it)^3\rho + n^{-1/2}o(t^3) + ...
\end{aligned}$$

Thus,

$$\begin{aligned}
\psi_{\sqrt{n}\bar{X}}(t) = \psi_X^n(t/\sqrt{n}) &= e^{-t^2/2}\exp\{\frac{1}{6n^{1/2}}(it)^3\rho + n^{-1/2}o(t^3)\} \\
&= e^{-t^2/2}\left(1 + \frac{1}{6n^{1/2}}(it)^3\rho + ...\right) \\
&= e^{-t^2/2} + \frac{\rho}{6n^{1/2}}(it)^3 e^{-t^2/2} + ...
\end{aligned}$$

Therefore, the "formal density" of $\sqrt{n}\bar{X}$ is

$$\begin{aligned}
f_{\sqrt{n}\bar{X}}(x) &= \frac{1}{2\pi}\int e^{-itx}\psi_{\sqrt{n}\bar{X}}(t)dt \\
&= \frac{1}{2\pi}\int e^{-itx}e^{-t^2/2}dt + \frac{\rho}{6n^{1/2}}\frac{1}{2\pi}\int e^{-itx}(it)^3 e^{-t^2/2}dt + ... \\
&= \phi(x) + \frac{\rho}{6n^{1/2}}H_3(x)\phi(x) + ...
\end{aligned}$$

Finally, we integrate the "density" to get the d.f.

$$\begin{aligned}
P(\sqrt{n}\bar{X} \le x) &= \int_{-\infty}^{x}\left(\phi(x) + \frac{\rho}{6n^{1/2}}H_3(x)\phi(x) + ...\right)dx \\
&= \Phi(x) - \frac{\rho}{6n^{1/2}}H_2(x)\phi(x) + ...
\end{aligned}$$

where the last line follows since

$$\begin{aligned}
\frac{d}{dx}H_2(x)\phi(x) &= H_2'(x)\phi(x) + H_2(x)(-x)\phi(x) = 2x\phi(x) + (-x^3 + x)\phi(x) \\
&= -(x^3 - 3x)\phi(x) = -H_3(x)\phi(x).
\end{aligned}$$

### 11.4.2 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

$$\phi^{(k)}(x) = (-1)^k H_k(x)\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(-it)^k e^{-t^2/2}e^{-itx}dt, \quad k \ge 0, \tag{4.11}$$

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 (4.11), we have, for $k \geq 0$,

$$\begin{aligned}
H_k(x)\phi(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}(it)^k e^{-t^2/2}e^{-itx}dt, \\
(it)^k e^{-t^2/2} &= \int_{-\infty}^{\infty} H_k(x)\phi(x)e^{itx}dt. \qquad (4.12)
\end{aligned}$$

That is, the Fourier transform of $H_k(x)\phi(x)$ is $(it)^k e^{-t^2/2}$.

### 11.4.3 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

$$\psi^{(k)}(0) = i^k EX^k, \qquad k = 0, 1, 2, ...$$

Define $\kappa(t) = \ln \psi(t) = \ln Ee^{itX}$ to be the cumulant generating function, and define the cumulants $\kappa_k$'s by

$$\kappa^{(k)}(0) = i^k \kappa_k \qquad k = 0, 1, 2, ....$$

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

$$\begin{aligned}
\kappa_1 &= \mu_1 = EX_1, \\
\kappa_2 &= \mu_2 - \mu_1^2 = E(X - \mu_1)^2, \\
\kappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 = E(X - \mu_1)^3, \\
\kappa_4 &= \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 \\
&= E(X - \mu_1)^4 - 3\left(E(X - \mu_1)^2\right)^2.
\end{aligned}$$

### 11.4.4 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 $O(n^{-1/2})$. 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 **11.4.1 (Edgeworth expansion for i.i.d. r.v.'s)** *Let $X_1, ..., X_n$ 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*

$$\sup_{x\in R}\left|F_n(x) - \Phi(x) + \frac{\mu_3}{6\sigma^3\sqrt{n}}H_2(x)\phi(x)\right| = o(n^{-1/2}).$$

*Proof.* In the Smoothing Lemma, we put

$$F(x) = F_n(x), \qquad \text{and} \qquad G(x) = \Phi(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}} H_2(x)\phi(x).$$

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

$$\Delta := \sup_x |F_n(x) - G(x)| \quad \leq \quad \frac{1}{\pi}\int_{-a\sqrt{n}}^{a\sqrt{n}} |t|^{-1}\left|\psi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - \psi_G(t)\right| dt + \frac{\epsilon}{\sqrt{n}}. \tag{4.13}$$

We partition the interval of integration into two parts:

$$\text{(i)} \quad \delta \leq \frac{|t|}{\sqrt{n}\sigma} \leq a/\sigma, \qquad\qquad \text{(ii)} \quad \frac{|t|}{\sqrt{n}\sigma} \leq \delta,$$

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 (4.13) 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 $\Delta$ as

$$\Delta \quad \leq \quad \int_{-a\sqrt{n}}^{a\sqrt{n}} |t|^{-1} e^{-t^2/2}\left|\exp\left\{n\xi\left(\frac{t}{\sigma\sqrt{n}}\right)\right\} - \left[1 + \frac{\mu_3}{6\sigma^3\sqrt{n}}(it)^3\right]\right| dt + \frac{\epsilon}{\sqrt{n}}, \tag{4.14}$$

where

$$\xi(t) = \ln\psi(t) + \frac{\sigma^2 t^2}{2}.$$

The integrand will be estimated using the following inequality: for any complex $\alpha$ and $\beta$,

$$\begin{aligned} |e^\alpha - 1 - \beta| \quad &\leq \quad \left|e^\alpha - e^\beta\right| + \left|e^\beta - 1 - \beta\right| \\ &\leq \quad \max\left\{e^{|\alpha|}, e^{|\beta|}\right\}\left(|\alpha - \beta| + \frac{1}{2}\beta^2\right). \end{aligned} \tag{4.15}$$

To see this, letting $\gamma = \max\{|\alpha|, |\beta|\}$, then

$$\begin{aligned} \left|e^\alpha - e^\beta\right| \quad &= \quad \left|\left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + ...\right) - \left(1 + \beta + \frac{\beta^2}{2!} + \frac{\beta^3}{3!} + ...\right)\right| \\ &= \quad \left|(\alpha - \beta) + \frac{1}{2!}(\alpha^2 - \beta^2) + \frac{1}{3!}(\alpha^3 - \beta^3) + ...\right| \\ &= \quad |\alpha - \beta|\left|1 + \frac{1}{2!}(\alpha + \beta) + \frac{1}{3!}(\alpha^2 + \alpha\beta + \beta^2) + ...\right| \\ &\leq \quad |\alpha - \beta|\left(1 + \frac{1}{2!}(2\gamma) + \frac{1}{3!}(3\gamma^2) + ...\right) \\ &= \quad |\alpha - \beta|\left(1 + \gamma + \frac{1}{2!}\gamma^2 + ...\right) \\ &= \quad |\alpha - \beta| e^\gamma, \end{aligned}$$

and

$$\left|e^{\beta}-1-\beta\right| = \left|\frac{\beta^2}{2!}+\frac{\beta^3}{3!}+...\right| \le \frac{\beta^2}{2}\left(1+|\beta|+\frac{|\beta|^2}{2!}+\frac{|\beta|^3}{3!}+...\right) \le \frac{\beta^2}{2}e^{|\beta|}.$$

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,

$$\xi(t) = \xi(0) + \xi'(0)t + \frac{1}{2}\xi''(0)t^2 + \frac{1}{6}\xi'''(0)t^3 + o(t^3) = \frac{1}{6}(it)^3\mu_3 + o(t^3).$$

We conclude that, there exists some small $\delta > 0$ such that

$$\left|\xi(t) - \frac{1}{6}\mu_3(it)^3\right| < \epsilon\sigma^3|t|^3 \qquad \text{for } |t| < \delta.$$

Here we can choose $\delta$ small enough so that

$$|\xi(t)| < \frac{1}{4}\sigma^2 t^2, \qquad \left|\frac{1}{6}\mu_3(it)^3\right| < \frac{1}{4}\sigma^2 t^2 \qquad \text{for } |t| < \delta.$$

Hence, using (4.15), the integrand in (4.14) 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 (4.14) is $o(n^{-1/2})$ since $\epsilon$ is arbitrary. Therefore, we have shown that $\Delta = o(n^{-1/2})$. ∎

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 11.4.1 can not be true of any lattice distribution. For a special example, see the remarks in Section 11.2.

Theorem **11.4.2 (Higher-order Edgeworth expansion for i.i.d. r.v.'s)** *Let $X_1, ..., X_n$ 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 $r \ge 3$. Furthermore, assume that*

$$\limsup_{|t|\to\infty} |\psi(t)| < 1, \qquad \textit{(Cramer's condition)}$$

*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 $\le r$, but not on $n$ and $r$.*

*Proof.* The proof of this theorem is very similar to that of Theorem 11.4.1. 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 $\kappa_j$ 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

$$\Delta \;:=\; \sup_x |F_n(x) - G(x)| \le \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\psi^n\left(\frac{t}{\sigma\sqrt{n}}\right) - \psi_G(t)}{t}\right| dt + \frac{\epsilon}{n^{(r-2)/2}}. \qquad (4.16)$$

We partition the interval of integration into two parts: $|t| \le \delta\sigma\sqrt{n}$ and $\delta\sigma\sqrt{n} \le |t| \le T$, where $\delta > 0$ is arbitrary, but fixed. First, consider $\delta\sigma\sqrt{n} \le |t| \le T$. From the Cramer's condition, the maximum of $|\psi(t)| < q < 1$ for $|t| \ge \delta$. Therefore, the contribution of the intervals $\delta\sigma\sqrt{n} \le |t| \le T$ to the integral in (4.13) is

$$\le \int_{\delta\sigma\sqrt{n}}^{T} t^{-1}\left[q^n + e^{-t^2/2}\left(1 + \left|\sum_{j=3}^{r}\frac{\kappa_j t^j}{\sigma^j j!}n^{-(j-2)/2}\right|\right)\right]dt$$

and this tends to zero more ripidly than any power of $1/n$.

Second, consider $|t| \le \delta\sigma\sqrt{n}$. Let

$$\xi(t) = \ln\psi(t) + \frac{\sigma^2 t^2}{2}, \qquad v^*(t) = 1 + \sum_{j=3}^{r}\frac{\kappa_j (it)^j}{\sigma^j j!}$$

Then we can rewrite $\Delta$ as

$$\Delta \quad \le \quad \int_{-T}^{T}|t|^{-1}e^{-t^2/2}\left|\exp\left\{n\xi\left(\frac{t}{\sigma\sqrt{n}}\right)\right\} - \left[1 + nv^*\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\right|dt + \frac{\epsilon}{n^{(r-2)/2}}, \tag{4.17}$$

..................

See Feller, page 535.

..................

The integrand will be estimated using the following inequality: for any complex $\alpha$ and $\beta$,

$$\begin{aligned}\left|e^\alpha - 1 - \beta - \dots - \frac{\beta^m}{m!}\right| \quad &\le \quad \left|e^\alpha - e^\beta\right| + \left|e^\beta - 1 - \beta - \dots - \frac{\beta^m}{m!}\right| \\ &\le \quad \max\left\{e^{|\alpha|}, e^{|\beta|}\right\}\left(|\alpha - \beta| + \frac{\beta^{m+1}}{(m+1)!}\right).\end{aligned} \tag{4.18}$$

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

$$|\xi(t) - v^*(t)| < \epsilon\sigma^r|t|^r \qquad \text{for } |t| < \delta.$$

We can further choose $\delta$ small enough so that

$$|\xi(t)| < \frac{1}{4}\sigma^2 t^2, \qquad |v^*(t)| < C\sigma^3|t|^3 \qquad \text{for } |t| < \delta.$$

Thus, if $|t| \le \delta\sigma\sqrt{n}$, 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

$$\left|n\xi\left(\frac{t}{\sqrt{n}\sigma}\right)\right| \le \frac{t^2}{4}, \qquad \left|nv^*\left(\frac{t}{\sqrt{n}\sigma}\right)\right| \le C\frac{|t|^3}{\sqrt{n}}$$

Hence, using (4.18), the integrand in (4.17) 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 (4.17) is $o(n^{-1/2})$ since $\epsilon$ is arbitrary. Therefore, we have shown that $\Delta = o(n^{-1/2})$. ∎

THEOREM **11.4.3 (One-term Edgeworth expansion for independent r.v.'s)** *Let $X_1, \dots, X_n$ be independent r.v.'s with c.f. $\psi_j$ 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. $n^{-1}\sum_{j=1}^{n} E|X_j|^3 < C_0, \qquad$ *where* $0 < C_0 < \infty$,

2. $cn < B_n^2 < Cn$, *where* $0 < c < C < \infty$,

3. $\left|\prod_{k=1}^{n} \psi_k(t)\right| = o(n^{-1/2})$ *uniformly in* $|t| > \delta > 0$.

*Then,*

$$\sup_{x \in R}\left|F_n(x) - \Phi(x) + \frac{\mu_{3n}}{6B_n^3} H_2(x)\phi(x)\right| = o(n^{-1/2}).$$

*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

$$\begin{aligned} \Delta &:= \sup_x |F_n(x) - G(x)| \\ &\leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\prod_{k=1}^{n}\psi_k\left(\frac{t}{B_n}\right) - \psi_G(t)}{t}\right| dt + \frac{\epsilon}{\sqrt{n}}. \end{aligned} \tag{4.19}$$

We partition the interval of integration into two parts: $|t| \leq \delta B_n$ and $\delta B_n \leq |t| \leq T$, where $\delta > 0$ is arbitrary, but fixed. First, consider $\delta B_n \leq |t| \leq T$. From the assumption, we see that the contribution of the intervals $\delta B_n \leq |t| \leq T$ to the integral in (4.19) is

$$\begin{aligned} &\leq \int_{\delta B_n}^{T} t^{-1}\left|\prod_{k=1}^{n}\psi_k(t)\right| dt + \int_{\delta B_n}^{T} t^{-1} e^{-t^2/2}\left(1 + \left|\frac{\mu_{3n}}{6B_n^3}\right|\right) dt \\ &\leq o(n^{-1/2})\int_{\delta B_n}^{T} t^{-1} dt + o(n^{-1/2}) \\ &\leq o(n^{-1/2})\ln(T/\delta B_n) + o(n^{-1/2}) \\ &= o(n^{-1/2}). \end{aligned}$$

Second, consider $|t| \leq \delta B_n$. Note that we can rewrite $\Delta$ as

$$\Delta \leq \int_{-T}^{T} |t|^{-1} e^{-t^2/2}\left|\exp\left\{\xi\left(\frac{t}{B_n}\right)\right\} - \left[1 + \frac{\mu_{3n}}{6}\left(\frac{it}{B_n}\right)^3\right]\right| dt + \frac{\epsilon}{\sqrt{n}}, \tag{4.20}$$

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 $\alpha$ and $\beta$,

$$\begin{aligned} |e^\alpha - 1 - \beta| &\leq \left|e^\alpha - e^\beta\right| + \left|e^\beta - 1 - \beta\right| \\ &\leq \max\left\{e^{|\alpha|}, e^{|\beta|}\right\}\left(|\alpha - \beta| + \frac{1}{2}\beta^2\right). \end{aligned} \tag{4.21}$$

Since $E|X_k|^3 < \infty$, 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

$$\left|\xi(t) - \frac{\mu_{3n}}{6n}(it)^3\right| < \frac{\epsilon B_n^3 |t|^3}{n^{3/2}} \qquad \text{for } |t| < \delta.$$

Here we can choose $\delta$ small enough so that

$$|\xi(t)| < \frac{B_n^2 t^2}{4n}, \qquad \left|\frac{\mu_{3n}}{6n}(it)^3\right| < \frac{B_n^2 t^2}{4n} \qquad \text{for } |t| < \delta.$$

Hence, using (4.21), the integrand in (4.20) 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 (4.20) is $o(n^{-1/2})$ since $\epsilon$ is arbitrary. Therefore, we have shown that $\Delta = o(n^{-1/2})$. ∎

THEOREM **11.4.4 (Higher-order Edgeworth expansion for independent r.v.'s)** *Let $X_1$, ..., $X_n$ be independent r.v.'s with c.f. $\psi_j$. Assume*

$(i)$ *(moment conditions)*

$$C_1 < n^{-1}\sum_{j=1}^n EX_j^2 < C_2, \qquad n^{-1}\sum_{j=1}^n E|X_j|^r < C_1, \quad (r \geq 3)$$

$(ii)$ *(smoothness condition)*

$$\left|\prod_{k=1}^n \psi_k(t)\right| = o(n^{-(r-2)/2}) \qquad \textit{uniformly in } ?????? > |t| > \delta > 0.$$

*Then,*

$$\sup_{x\in R}\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| = o(n^{-(r-2)/2}). \quad ∎$$

In the i.i.d. case, it reduces to the following theorem.

THEOREM **11.4.5 (Edgeworth expansion for i.i.d. r.v.'s)** *Let $X$, $X_1$, ..., $X_n$ be i.i.d. r.v.'s with c.f. $\psi$. Assume*

$(i)$ *(moment conditions:)* $E|X_j|^r < C, \quad (r \geq 3)$

$(ii)$ *(smoothness condition)*

$(a)$. *The d.f. of $X$ is nonlattice if $r = 3$.*

$(b)$. $\limsup_{|t|\to\infty} |\psi(t)| < 1$ *if $r \geq 3$.*

*Then,*

$$\sup_{x\in R}\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_n(x)\phi(x)\right| = o\left(n^{-(r-2)/2}\right). \quad ∎$$

## 11.5 Non-uniform Edgeworth expansions

THEOREM **11.5.1 (Non-uniform Edgeworth expansion for i.i.d. r.v.'s)** *Let $X$, $X_1$, ..., $X_n$ 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 $x$,*

$$\begin{aligned}
&\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| \\
\leq\ & C_r \frac{E|X|^r I\{|X| \geq \sigma\sqrt{n}(1+|x|)\}}{\sigma^r(1+|x|)^r n^{(r-2)/2}} \\
&+ C_r \frac{E|X|^{r+1} I\{|X| < \sigma\sqrt{n}(1+|x|)\}}{\sigma^{r+1}(1+|x|)^{r+1} n^{(r-1)/2}} \\
&+ C_r \left(\sup_{|t|\geq\delta} |\psi(t)| + \frac{1}{2n}\right)^n \frac{n^{r(r+1)/2}}{(1+|x|)^{r+1}},
\end{aligned}$$

*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 **11.5.2 (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*

$$\left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| \leq \frac{\epsilon\left(\sqrt{n}(1+|x|)\right)}{n^{(r-2)/2}(1+|x|)^r}.$$

Theorem **11.5.3 (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*

$$(1+|x|)^r \left|F_n(x) - \Phi(x) + \sum_{j=3}^{r-2} n^{-(j-2)/2} P_{nj}(x)\phi(x)\right| = o\left(\frac{1}{n^{(r-2)/2}}\right).$$

## 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.

### 11.6.1 Cramer conditioin

Let $X, X_1, ..., X_n$ be i.i.d. r.v.'s. Let the following *Cramer's condition* hold:

$$Ee^{tX} < \infty, \qquad \text{in } |t| < H \text{ for some constant } H > 0.$$

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 **11.6.1** *The following assertions are equivalent:*

(1) $Ee^{tX} < \infty$ *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) → (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) → (1)". This follows from $e^{tX} \leq e^{t|X|}$.
"(2) → (3)". This follows from

$$P(|X| \geq x) = P(a|X| \geq ax) \leq e^{-ax} Ee^{a|X|}, \qquad x, a > 0.$$

"(3) → (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 $0 \leq a < c$,

$$\begin{aligned} I_2 &= -\left[e^{ax}(1 - F(x))\right]\big|_0^{\infty} + \int_0^{\infty} (1 - F(x)) de^{ax} \\ &\leq [1 - F(0)] - \lim_{t\to\infty} be^{(a-c)t} + a\int_0^{\infty} be^{(a-c)x} dx \\ &< \infty. \end{aligned}$$

Similarly, we can show that $I_2 < \infty$. ∎

Part (3) of the above lemma states that the tail probability of $X$ decreases to zero exponentally fast under the Cramer's condition.

### 11.6.2 Conjugate (or associated) distributions

One of the key tools in large deviation theory is the so-called *conjugate* (or *associated*) distribution. Let $X, X_1, ..., X_n$ 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

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tX} dF(x), \qquad K_X(t) = \ln M_X(t).$$

Definition **11.6.1** *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,*

$$dG(x) = \frac{e^{sx} dF(x)}{\int_{-\infty}^{\infty} e^{sx} dF(x)} = \frac{e^{sx} dF(x)}{M_X(s)} = e^{sx-K_X(s)} dF(x). \quad \blacksquare$$

We shall provide several useful lemmas.

Lemma **11.6.2** *It is easy to see that m.g.f.'s and c.g.f.'s of $X$ and $Y$ are related by*

$$M_Y(t) = \frac{M_X(t+s)}{M_X(s)}, \qquad K_Y(t) = K_X(t+s) - K_X(s). \tag{6.22}$$

*In particular,*

$$EY = K_Y'(t)|_{t=0} = K_X'(s), \qquad Var(Y) = K_Y''(t)|_{t=0} = K_X''(s).$$

*Proof.* Omitted. $\blacksquare$

Remark **11.6.1** *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.* $\blacksquare$

Lemma **11.6.3** *Let $Y, Y_1, ..., Y_n$ be i.i.d. r.v.'s with a common d.f. $G$. Define*

$$F_n(x) = P\left(\sum_{i=1}^{n} X_i \le x\right), \qquad G_n(y) = P\left(\sum_{i=1}^{n} Y_i \le y\right)$$

*Then*

$$G_n(y) = \int_{-\infty}^{y} e^{sx-nK_X(s)} dF_n(x) = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{(e^{K_X(s)})^n} = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{M_X^n(s)} = \frac{\int_{-\infty}^{y} e^{sx} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)}$$

*or equivalently,*

$$dG_n(x) = e^{sx-nK_X(s)} dF_n(x) = \frac{e^{sx} dF_n(x)}{(e^{K_X(s)})^n} = \frac{e^{sx} dF_n(x)}{M_X^n(s)} = \frac{e^{sx} dF_n(x)}{\int_{-\infty}^{\infty} e^{sx} dF_n(x)}. \quad \blacksquare$$

*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 **11.6.4** *If we choose $\tau$ such that $K'_X(\tau) = z$, then we have*

$$\begin{aligned} P\left(\bar{X} > z\right) &= e^{n[K_X(\tau)-\tau K'_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right) & (6.23)\\ &= e^{-n[z\tau - K_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right) & (6.24)\\ &= e^{-n\sup_{t>0}[zt-K_X(t)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right), & (6.25)\end{aligned}$$

*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.* From (**??**),

$$\begin{aligned} P\left(\bar{X} > z\right) &= P\left(\sum_{i=1}^n X_i > nz\right)\\ &= \int_{nz}^\infty dP\left(\sum_{i=1}^n X_i \le x\right)\\ &= \int_{nz}^\infty e^{-sx+nK_X(s)} dP\left(\sum_{i=1}^n Y_i \le x\right)\\ &= e^{nK_X(s)} \int_{nz}^\infty e^{-sx} dP\left(\frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nVar(Y)}} \le \frac{x - nK'_X(s)}{\sqrt{nK''_X(s)}}\right)\\ &= e^{nK_X(s)} \int_{\frac{\sqrt{n}[z-K'_X(s)]}{\sqrt{K''_X(s)}}}^\infty e^{-s\left(nK'_X(s)+y\sqrt{nK''_X(s)}\right)} dP\left(T_n \le y\right)\\ &= e^{n[K_X(s)-sK'_X(s)]} \int_{\frac{\sqrt{n}[z-K'_X(s)]}{\sqrt{K''_X(s)}}}^\infty e^{-sy\sqrt{nK''_X(s)}} dP\left(T_n \le y\right),\end{aligned}$$

where

$$T_n = \frac{\sum_{i=1}^n Y_i - nEY}{\sqrt{nVar(Y)}}, \qquad y = \frac{x - nK'_X(s)}{\sqrt{nK''_X(s)}}.$$

If we choose $\tau$ such that $K'_X(\tau) = z$, then we have

$$P\left(\bar{X} > z\right) = e^{n[K_X(\tau)-sK'_X(\tau)]} \int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right). \quad ∎$$

### 11.6.3 Large Deviations

From Lemma 11.6.1, under Cramer's condition, $P(S_n > na)$ converges to 0 exponentially fast. More careful analysis can give a precise convergence rate.

THEOREM **11.6.1** *Let $X_1, X_2, ...$ 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,*

$$\lim_n P(S_n > na)^{1/n} = \lim_n P(\bar{X} > a)^{1/n} = e^{-\sup_{t>0}[at-K(t)]} =: e^{-\eta(a)}$$

*or equivalently,*

$$-\lim_n \frac{1}{n} \ln P(S_n > na) = \sup_{t>0}[at - K(t)] =: \eta(a)$$

*where $\eta(a) = \sup_{t>0}[at - K(t)]$.*

Remark **11.6.2** *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*

$$\begin{aligned} at - K(t) &= at - [K(0) + K'(0)t + \frac{1}{2}K''(0)t^2 + ...] \\ &= at - \frac{1}{2}\sigma^2 t^2 + o(t^2) = at\left(1 - \frac{1}{2}\sigma^2 t + o(t)\right) > 0 \end{aligned}$$

*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

$$P(S_n > na)^{1/n} = P(e^{tS_n} > e^{nat})^{1/n} \le \left(\frac{Ee^{tS_n}}{e^{nat}}\right)^{1/n} = \left(\frac{(Ee^{tX_1})^n}{e^{nat}}\right)^{1/n} = \frac{M(t)}{e^{at}} = e^{K(t)-at}$$

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

$$\begin{aligned} P\left(\bar{X} > a\right)^{1/n} &= e^{[K_X(\tau) - sK'_X(\tau)]}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right)\right)^{1/n} \\ &= e^{-\psi(a)}\left(\int_0^\infty e^{-\tau y\sqrt{nK''_X(\tau)}} dP\left(T_n \le y\right)\right)^{1/n} \\ &=: e^{-\psi(a)} A_n^{1/n}. \end{aligned}$$

It suffices to show that $A_n^{1/n} \ge 1$ as $n \to \infty$. To do this, note that

$$\begin{aligned} A_n &=: \int_0^\infty e^{-\sqrt{n}\tau_0 y} dP\left(T_n \le y\right) \qquad (\text{where } \tau_0 = \tau\sqrt{K''_X(\tau)}) \\ &=: \frac{\int_0^\infty e^{-\sqrt{n}\tau_0 y} dP\left(T_n \le y\right)}{\int_0^\infty dP\left(T_n \le y\right)} \int_0^\infty dP\left(T_n \le y\right) \\ &= E\left(e^{-\sqrt{n}\tau_0 T_n} | T_n > 0\right) P(T_n > 0) \\ &\ge \exp\left(\{-\sqrt{n}\tau_0 E[T_n | T_n > 0]\}\right) P(T_n > 0), \qquad (\text{by Jensen's inequality}) \end{aligned}$$

Now

$$E[T_n|T_n > 0] = \frac{E[T_n I(T_n > 0)]}{P(T_n > 0)} \le \frac{E|T_n|}{P(T_n > 0)} \le \frac{(ET_n^2)^{1/2}}{P(T_n > 0)} \le \frac{1}{P(T_n > 0)} \le \frac{1}{0.5 - \epsilon},$$

and therefore,

$$A_n^{1/n} \ge \exp\left(\frac{-\sqrt{n}\tau_0}{nP(T_n > 0)}\right) P^{1/n}(T_n > 0) \longrightarrow 1,$$

since $P(T_n > 0) \to 1/2$ and $P^{1/n}(T_n > 0) \to 1$. ∎

### 11.6.4 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

$$\frac{1 - F_n(x)}{1 - \Phi(x)} \to 1, \qquad \frac{F_n(-x)}{\Phi(-x)} \to 1, \tag{6.26}$$

when both $x$ and $n$ tend to $\infty$.

However, the statement (6.26) can not be true in general. For instance, if $X_1, ..., X_n$ are i.i.d. with $P(X_1 = 0) = P(X_1 = 1) = 1/2$, then $P(S_n > n) = 1 - F_n(\sqrt{n}) = 0$. Thus, for $|x| > \sqrt{n}$, the ratios in (6.26) 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 **11.6.2**

$$\frac{1 - F_n(x)}{1 - \Phi(x)} = 1 + O\left(\frac{x^3}{\sqrt{n}}\right). \quad \blacksquare$$

*Proof.* Note that

$$1 - F_n(x) \;=\; P\left(\sqrt{n}(\bar{X} - \mu)/\sigma > x\right) = P\left(\bar{X} > \mu + \frac{x\sigma}{\sqrt{n}}\right).$$

Then from (**??**), we have

$$1 - F_n(x) \;=\; e^{n[K_X(\tau) - \tau K'_X(\tau)]} \int_0^\infty e^{-\tau y \sqrt{nK''_X(\tau)}} dP\left(T \le y\right). \tag{6.27}$$

where $\tau$ satisfies

$$K'_X(\tau) = \mu + \frac{x\sigma}{\sqrt{n}}.$$

Let

$$\Delta(y) = P\left(T \le y\right) - \Phi(y).$$

We can approximate $P\left(T \le y\right)$ 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

$$K_X(\tau) \;=\; \mu\tau + \frac{1}{2}\sigma^2\tau^2 + \frac{1}{6}EX^3\tau^3 + ......,$$

we get

$$K'_X(\tau) = \mu + \sigma^2\tau + \frac{1}{2}EX^3\tau^2 + ...... = \mu + \frac{x\sigma}{\sqrt{n}}.$$

Thus

$$\tau \sim \frac{x}{\sigma\sqrt{n}} \qquad \text{if } \tfrac{x}{\sqrt{n}} \to 0.$$

Now

$$\begin{aligned} A_n &= e^{n[K_X(\tau) - \tau K'_X(\tau)]} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\tau y \sqrt{nK''_X(\tau)} - \frac{1}{2}y^2} dy \\ &= e^{n\left[K_X(\tau) - \tau K'_X(\tau) + \frac{1}{2}\tau^2 K''_X(\tau)\right]} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}\left(y + \tau\sqrt{nK''_X(\tau)}\right)^2} dy \\ &= e^{n\left[K_X(\tau) - \tau K'_X(\tau) + \frac{1}{2}\tau^2 K''_X(\tau)\right]} \left(1 - \Phi(\tau\sqrt{nK''_X(\tau)})\right). \end{aligned}$$

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 $x_\tau = \tau\sqrt{nK_X''(\tau)}$.

$$\begin{aligned}
\frac{x_\tau - x}{\sqrt{n}} &= \tau\sqrt{K_X''(\tau)} - \frac{K_X'(\tau)}{\sigma} \\
&= \sigma\tau\left(1 + \frac{(EX^3)}{2\sigma^2}\tau + O(\tau^2)\right) - \sigma\tau\left(1 + \frac{(EX^3)}{2\sigma^2}\tau + O(\tau^2)\right) \\
&= O\left(\tau^3\right).
\end{aligned}$$

Thus,

$$x_\tau - x = O\left(\sqrt{n}\tau^3\right) = O\left(\frac{x^3}{n}\right)$$

Now

$$\begin{aligned}
\log\left(\frac{1-\Phi(x_\tau)}{1-\Phi(x)}\right) &= \log\left(1-\Phi(x_\tau)\right) - \log\left(1-\Phi(x)\right) \\
&= \frac{-\phi(x_0)}{1-\Phi(x_0)}(x_\tau - x) \quad \text{where } x_0 \text{ is between } x \text{ and } x_\tau \\
&\sim -x_0(x_\tau - x) = O\left(\frac{x^4}{n}\right)
\end{aligned}$$

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,

$$\begin{aligned}
|B_n| &\le e^{n[K_X(\tau) - \tau K_X'(\tau)]}\int_0^\infty e^{-\tau y\sqrt{nK_X''(\tau)}}d|\Delta(y)| \\
&\le 2e^{n[K_X(\tau) - \tau K_X'(\tau)]}\sup_y|\Delta(y)| \\
&\le 4e^{-\frac{1}{2}n\sigma^2\tau^2 + nO(\tau^3)}\frac{E|X|^3}{(EX^2)^{3/2}\sqrt{n}} \\
&\le C\phi(x)\left[1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right]\frac{E|X|^3}{(EX^2)^{3/2}\sqrt{n}} \\
&\sim C\left(1-\Phi(x)\right)\left[1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right]\frac{x}{\sqrt{n}} \\
&\sim A_n O\left(\frac{x}{\sqrt{n}}\right).
\end{aligned}$$

Finally, we get

$$1 - F_n(x) = A_n O\left(1 + \frac{x}{\sqrt{n}}\right) = (1-\Phi(x))\left[1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right]. \quad \blacksquare$$

## 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

$$\begin{aligned}
P\left(\bar{X} > x\right) &= e^{n[K_X(\tau) - \tau K_X'(\tau)]}\int_0^\infty e^{-\tau y\sqrt{nK_X''(\tau)}}dP\left(T \le y\right) \\
&= e^{-n[\tau x - K_X(\tau)]}\int_0^\infty e^{-\tau y\sqrt{nK_X''(\tau)}}dP\left(T \le y\right) \\
&= e^{-\hat{w}^2/2}\int_0^\infty e^{-\hat{z}y}dP\left(T \le y\right).
\end{aligned}$$

where $K'_X(\tau) = x$ and

$$\hat{w} = \sqrt{2n[\tau x - K_X(\tau)]}sgn\{\tau\}, \qquad \hat{z} = \tau\sqrt{nK''_X(\tau)}.$$

**First order approximations.**

By the CLT, we can use the normal distribution to $\Phi(y)$ to approximate $P(T \le y)$, resulting in

$$P\left(\bar{X} > x\right) \quad = \quad e^{-n\hat{w}^2/2}\int_0^\infty e^{-\hat{z}y}d[\Phi(y)+\Delta(y)] =: A_n + B_n,$$

where $\Delta(y) = P(T \le y) - \Phi(y)$. First

$$\begin{aligned} A_n &= e^{-\hat{w}^2/2}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\hat{z}y-\frac{1}{2}y^2}dy \\ &= e^{-\hat{w}^2/2}e^{\frac{1}{2}\hat{z}^2}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(y+\hat{z})^2}dy \\ &= e^{-\hat{w}^2/2}e^{\frac{1}{2}\hat{z}^2}\left(1-\Phi(\hat{z})\right) \\ &= \phi(\hat{w})\frac{1-\Phi(\hat{z})}{\phi(\hat{z})} \\ &= \phi(\hat{w})M(\hat{z}). \end{aligned}$$

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,

$$P\left(\bar{X} > x\right) \quad = \quad \phi(\hat{w})\left(M(\hat{z}) + O\left(n^{-1/2}\right)\right).$$

**Remarks:**

(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:

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\phi(t) < 1-\Phi(t) < \frac{\phi(t)}{t} \quad \text{for } t > 1.$$

$$\frac{\phi(t)}{2t} < 1-\Phi(t) < \frac{\phi(t)}{t} \quad \text{for } t > \sqrt{2}.$$

(3). From (2), we can rewrite

$$\begin{aligned} P\left(\bar{X} > x\right) &= \phi(\hat{w})M(\hat{z})\left(1+\hat{z}O\left(n^{-1/2}\right)\right) \\ &= [1-\Phi(\hat{w})]\frac{M(\hat{z})}{M(\hat{w})}\left(1+\hat{z}O\left(n^{-1/2}\right)\right). \end{aligned}$$

In particular,

(a) if $|x-\mu| = O(n^{-1/2})$, then $\tau = O(n^{-1/2})$ 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 $P(T \le y)$:

$$E_1(y) = \Phi(y) - \frac{\kappa_3(\tau)}{6\sqrt{n}}H_3(y)\phi(y), \qquad \Delta_1(y) = P(T \le y) - E_1(y).$$

Then we have

$$\begin{aligned} P\left(\bar{X} > x\right) &= e^{-n\hat{w}^2/2}\int_0^\infty e^{-\hat{z}y}d[E_1(y)+\Delta_1(y)] \\ &= A_{n1} + B_{n1}, \end{aligned}$$

where $\Delta(y) = P(T \le y) - \Phi(y)$. First

$$\begin{aligned} A_n &= e^{-\hat{w}^2/2}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\hat{z}y-\frac{1}{2}y^2}dy \\ &= e^{-\hat{w}^2/2}e^{\frac{1}{2}\hat{z}^2}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(y+\hat{z})^2}dy \\ &= e^{-\hat{w}^2/2}e^{\frac{1}{2}\hat{z}^2}\left(1-\Phi(\hat{z})\right) \\ &= \phi(\hat{w})\frac{1-\Phi(\hat{z})}{\phi(\hat{z})} \\ &= \phi(\hat{w})M(\hat{z}). \end{aligned}$$

where $M(t) = \dfrac{1-\Phi(t)}{\phi(t)}$ 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,

$$P\left(\bar{X} > x\right) = \phi(\hat{w})\left(M(\hat{z}) + O\left(n^{-1/2}\right)\right).$$

**Remarks:**

(1). The Miller's ratio has the following asymptotic expansions:

$$1-\Phi(t)=\phi(t)\left(\frac{1}{t}-\frac{1}{t^3}+\frac{1\cdot 3}{t^5}-\frac{1\cdot 3\cdot 5}{t^7}+......\right), \quad \text{as } t\to\infty.$$

Proof of this can be seen by integration by parts iteratively,

$$1-\Phi(t)=\int_t^\infty \phi(x)dx=-\int_t^\infty \frac{d\phi(x)}{x}=\frac{\phi(t)}{t}+\int_t^\infty \frac{\phi(x)}{x^2}dx=......$$

(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

$$\begin{aligned} P\left(\bar{X}>x\right) &= \phi(\hat{w})M(\hat{z})\left(1+\hat{z}O\left(n^{-1/2}\right)\right) \\ &= \left[1-\Phi(\hat{w})\right]\frac{M(\hat{z})}{M(\hat{w})}\left(1+\hat{z}O\left(n^{-1/2}\right)\right). \end{aligned}$$

In particular,

(a) if $|x-\mu|=O(n^{-1/2})$, then $\tau=O(n^{-1/2})$ 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 $o(1)$.

## 11.8 Appendix: Some useful elementary inequalities

THEOREM **11.8.1** *For a complex $z$,*

1. $|e^z-1|\le |z|e^{|z|}$.
2. $|e^z-1|\le e^{|z|}-1$.
3. $|\ln(1+z)-z|\le \dfrac{|z|^2}{2(1-|z|)}$, *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.*

$$\begin{aligned} |e^z-1| &= \left|z+\frac{z^2}{2!}+\frac{z^3}{3!}+\frac{z^4}{4!}+\cdots\right| \\ &\le |z|\left|1+|z|+\frac{|z|}{2!}+\frac{|z|^3}{3!}+\frac{|z|^4}{4!}+\cdots\right| \\ &\le |z|e^{|z|}. \end{aligned}$$

$$\begin{aligned} |e^z-1| &= \left|z+\frac{z^2}{2!}+\frac{z^3}{3!}+\frac{z^4}{4!}+\cdots\right| \\ &\le \left||z|+\frac{|z|}{2!}+\frac{|z|^3}{3!}+\frac{|z|^4}{4!}+\cdots\right| \\ &\le e^{|z|}-1. \end{aligned}$$

$$\begin{aligned}
|\ln(1+z)-z| &= \left|\frac{z^2}{2}-\frac{z^3}{3}+\frac{z^4}{4}+\cdots\right| \\
&\le \frac{|z|^2}{2}\left|1+|z|+|z|^2+|z|^3+\cdots\right| \\
&\le \frac{|z|^2}{2(1-|z|)}. \quad \blacksquare
\end{aligned}$$

## 11.9 Exercises

1. Show that if $\{X_n, n \ge 1\}$ are independent r.v.s with $EX_n = 0$ and $|X_n| \le C_n$ a.s., $n \ge 1$, and $C_n = o(B_n)$, where $B_n^2 = \sum_{i=1}^n EX_n^2 \to \infty$, then $S_n/B_n \Longrightarrow N(0,1)$.

2. if $\{X_n, n \ge 1\}$ are independent r.v.s with
$$P(X_n = \pm n^\alpha) = \frac{1}{2n^\beta}, \qquad P(X_n = 0) = 1 - \frac{1}{n^\beta}, \quad 2\alpha > \beta - 1,$$
the Lindeberg condition holds when and only when $0 \le \beta < 1$.

3. **Failure of the Lindeberg condition does not preclude asymptotic normality.** Let $\{Y_n, n \ge 1\}$ be i.i.d. with $EY_n = 0$, $EY_n^2 = 1$. Let $\{Z_n, n \ge 1\}$ be independent with $P(Z_n = \pm n) = 1/(2n^2)$ and $P(Z_n = 0) = 1 - 1/n^2$ and $\{Z_n, n \ge 1\}$ are independet of $\{Y_n, n \ge 1\}$. Define $X_n = Y_n + Z_n$, $S_n = \sum_{i=1}^n X_n$. Show that the Lindeberg condition can not hold, but $S_n/\sqrt{n} \Longrightarrow N(0,1)$. Explain why this does not contravene Lindeberg-Feller Theorem.

4. Let $\{X_n, n \ge 1\}$ be independent with
$$P(X_n = \pm 1) = \frac{1}{2a}, \quad P(X_n = \pm n) = \frac{1}{2}\left(1 - \frac{1}{a}\right)\frac{1}{n^2}, \quad P(X_n = 0) = \left(1 - \frac{1}{a}\right)\left(1 - \frac{1}{n^2}\right),$$
where $n \ge 1,\ a > 1$. Again, $S_n/\sqrt{n} \Longrightarrow N(0,a)$ despite the Lindeberg condition being violated.

5. Let $\{Y_n, n \ge 1\}$ be i.i.d. r.v.s with finite variance $\sigma^2$ (say $\sigma^2 = 1$), and let $\{\sigma_n^2, n \ge 1\}$ be nonzero constants with $B_n^2 = \sum_{i=1}^n \sigma_i^2 \nearrow \infty$. Show that the weighted i.i.d. r.v.s $\{\sigma_n Y_n, n \ge 1\}$ obey the CLT, i.e.,
$$\frac{1}{B_n}\sum_{j=1}^n \sigma_j Y_j \Longrightarrow N(0,1)$$
if $EY_1 = 0$ and $\sigma_n = o(B_n)$.

6. Let $X_1, X_2, ...$ be independent and $S_n = X_1 + ... + X_n$. Show that if $|X_i| \le M$ and $\sum_{i=1}^\infty Var(X_i) = \infty$, then
$$\frac{S_n - ES_n}{\sqrt{Var(S_n)}} \Longrightarrow N(0,1).$$

7. Let $X_1, X_2, ...$ be i.i.d. r.v.s such that $P(X_i = \pm 1) = 1/2$. Show that
$$\sqrt{\frac{3}{n^3}}\sum_{k=1}^n kX_k \Longrightarrow N(0,1).$$

8. Let $X_1, X_2, ...$ be independent r.v.s with
$$P(X_i = \pm i) = \frac{1}{2} - \frac{1}{2i^2}, \qquad P(X_i = \pm i^2) = \frac{1}{2i^2}, \quad \text{for } i = 1, 2, ....$$
   (i) Find $\lim_{n\to\infty} Var(S_n)/n^3$;
   (ii) What is the limiting distribution of $S_n/n^{3/2}$?

9. Let $X_1, X_2, ...$ be i.i.d. r.v.s with $EX_i = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Put
$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2, \qquad T_n = \sqrt{n}(\bar{X} - \mu)/S_n.$$
Prove that $T_n \Longrightarrow N(0,1)$.

# Chapter 12

# Infinitely 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.

## 12.1 Definitions

DEFINITION **12.1.1** *A d.f. $F$ and its corresponding c.f. $\psi(t)$ are called* **infinitely divisible (i.d.)** *if for every positive integer $n$,*

- *there exists a c.f. $\psi_n(t)$ such that*
$$\psi(t) = (\psi_n(t))^n;$$
- *or equivalently, there exists a d.f. $F_n$ such that*
$$F = F_n^{*n}, \qquad \text{the } n\text{-fold convolution of the function } F_n.$$
- *or equivalently, there exist a r.v. $X \sim F$ with c.f. $\psi$ and i.i.d. r.v.s $X_{n1}, ..., X_{nn}$ with c.f. $\psi_n$ such that*
$$X =_d X_{n1} + ... + X_{nn}. \quad \blacksquare$$

## 12.2 Some examples

EXAMPLE **12.2.1**

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.*
$$\psi(t) = \exp\{i\mu t - \sigma^2 t^2/2\} = \left(\exp\{i(\mu/n)t - (\sigma^2/n)t^2/2\}\right)^n = (\psi_n(t))^n.$$
3. *Poisson d.f.*
$$\psi(t) = \exp\{\lambda(e^{it} - 1)\} = \left(\exp\{(\lambda/n)(e^{it} - 1)\}\right)^n = (\psi_n(t))^n.$$
4. *Compound Poisson d.f.: $S_N = X_1 + ... + X_N$, where $X \sim F$ and $N \sim Poisson(\lambda)$. 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.* $f(x) = \dfrac{a}{\pi}\dfrac{1}{a^2+x^2}$*:*

$$\psi(t) = \exp\{-a|t|\} = \left(\exp\{-(a/n)|t|\}\right)^n = \left(\psi_n(t)\right)^n .$$

6. *$\alpha-$stable d.f. (including Cauchy d.f.)*

7. *Gamma d.f. with p.d.f.* $f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$*:*

$$\psi(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha} = \left(\left(1 - \frac{it}{\beta}\right)^{-\alpha/n}\right)^n = \left(\psi_n(t)\right)^n .$$

   *Two special cases:*

   *(a) The $\chi^2$-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 $\chi^2$-distribution (with degree of freedom 1).*

   **Less trivial examples:**

8. *log-normal d.f.*
9. *Pareto d.f.*
10. *Student d.f.*

## 12.3 Some properties

THEOREM **12.3.1**

1. *Let $\psi(t)$ be an i.d. c.f. Then $\psi(t) \neq 0$ for every $t$.*
2. *If $\psi_i(t)$, $i = 1, 2, ..., m$, 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 $\{\psi^{(m)}(t); m = 1, 2, ...\}$ 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
$$|\psi_n(t)| = |\psi(t)|^{1/n} > (1-\epsilon)^{1/n} \to 1, \qquad \text{as } n \to \infty,$$
   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
$$1 - |\psi_n(2t)| \leq 8(1 - |\psi_n(t)|) < \epsilon.$$
   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,
$$|\psi(t)| = \left|\left(\psi_{N_0}(t)\right)^{N_0}\right| = |\psi_{N_0}(t)|^{N_0} > 0. \quad ∎$$

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,
$$|\psi(t)|^2 = |\,(\psi_{2n}(t))^{2n}\,|^2 = \left(|\psi_{2n}(t)|^2\right)^{2n},$$
therefore,
$$|\psi(t)| = \left(|\psi_{2n}(t)|^2\right)^n =: \left(\tilde{\psi}_n(t)\right)^n,$$
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 $\{\psi_n^{(m)}(t), m \geq 1\}$ 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 **12.3.1** *We can use Theorem 12.3.1 (part 1) to judge whether a c.f. is NOT i.d.*

*(a) 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.*

*(b) Let $P(X = \pm 1)$ with c.f. $f(t) = \cos t$. Similarly, it is NOT i.d.* ∎

## 12.4 Levy-Khintchine representation of infinitely divisible c.f.s

The key theorem in this section is the **"Levy-Khintchine" representation**.

Theorem **12.4.1 ("Levy-Khintchine" representation)** *A function $\psi(t)$ is an i.d.c.f. if and only if it admits the representation*
$$\psi(t) = \exp\left\{it\gamma + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x)\right\} =: e^{\eta(t)}, \tag{4.1}$$
*where $\gamma$ 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.)*

Remark **12.4.1**

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 (4.1) 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 (4.1) either. For the purpose of definiteness, we may assume that it is left-continuous and $G(-\infty) = 0$ from now on.*

### 12.4.1 Proof of necessity

THEOREM **12.4.2** *If $\psi(t)$ is i.d.c.f., then $\psi(t)$ has the representation (4.1).*

*Proof.* We start with $\psi(t) = (\psi_n(t))^n$. Let $F_n$ denote the d.f. corresponding to the c.f. $\psi_n$. Now as $0 < |\psi(t)| \le 1$, $\ln \psi(t)$ exists and is finite. Furthermore,

$$\psi^{1/n}(t) = \exp\left(\frac{1}{n}\ln\psi(t)\right) = 1 + \frac{1}{n}\ln\psi(t) + O(\frac{1}{n^2}) =: 1 + \frac{1}{n}\eta(t) + O(\frac{1}{n^2}),$$

from which we get

$$\begin{aligned}
\eta(t) &= \lim n\left(\psi^{1/n}(t) - 1 - O(\frac{1}{n^2})\right) \\
&= \lim_{n\to\infty} n\left(\psi^{1/n}(t) - 1\right) \\
&= \lim_{n\to\infty} n\left(\psi_n(t) - 1\right) = \lim_{n\to\infty}\int_{-\infty}^{\infty} n\left(e^{itx} - 1\right) dF_n(x) \\
&= \lim_{n\to\infty}\int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2} + \frac{itx}{1+x^2}\right) dF_n(x) \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) dF_n(x)\right\} \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty} n\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{x^2}{1+x^2}dF_n(x)\right\} \\
&= \lim_{n\to\infty}\left\{it\int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x) + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}d\int_{-\infty}^{x}\frac{ny^2}{1+y^2}dF_n(y)\right\} \\
&=: \lim_{n\to\infty}\left\{it\gamma_n + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG_n(x)\right\} \\
&=: \lim_{n\to\infty}\eta_n(t),
\end{aligned}$$

where

$$\eta_n(t) =: it\gamma_n + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG_n(x), \tag{4.2}$$

and

$$\gamma_n = \int_{-\infty}^{\infty}\frac{nx}{1+x^2}dF_n(x), \qquad G_n(x) = \int_{-\infty}^{x}\frac{ny^2}{1+y^2}dF_n(y).$$

We have shown that

$$\lim_{n\to\infty}\eta_n(t) = \eta(t),$$

where $\eta(t)$ is continuous at 0. From Lemma 12.4.3 below, we have that $\gamma_n \to \gamma$ and $G_n \Longrightarrow G$ for some constant $\gamma$ and some "nice" function $G$ as described in Theorem 12.4.1.

#### 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

$$A(y) = \left(1 - \frac{1}{y}\left(y - \frac{y^3}{3!} + \frac{y^5}{5!} - ...\right)\right)\frac{1+y^2}{y^2} = \left(\frac{1}{3!} - \frac{y^2}{5!} - ...\right)(1+y^2);$$

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

$$0 < c_1 \le A(y) \le c_2 < \infty, \qquad \text{for all } y.$$

Now define

$$\Lambda(x) = \int_{-\infty}^{x} A(y)dG(y), \qquad \text{and} \qquad \Lambda_n(x) = \int_{-\infty}^{x} A(y)dG_n(y).$$

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 **12.4.1** *We have*

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{itx} d\Lambda(x) &= \eta(t) - \frac{1}{2}\int_0^1 [\eta(t+h) + \eta(t-h)]dh =: \lambda(t), \\
\int_{-\infty}^{\infty} e^{itx} d\Lambda_n(x) &= \eta_n(t) - \frac{1}{2}\int_0^1 [\eta_n(t+h) + \eta_n(t-h)]dh =: \lambda_n(t).
\end{aligned}$$

*Proof.*

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{itx} d\Lambda(x) &= \int_{-\infty}^{\infty} e^{itx}\left(1 - \frac{\sin x}{x}\right)\frac{1+x^2}{x^2}dG(x) \\
&= \int_{-\infty}^{\infty}\int_0^1 e^{itx}(1-\cos hx)\frac{1+x^2}{x^2}dhdG(x) \\
&\quad \text{(by Fubini's theorem since the integrand is bounded} \\
&\qquad \text{continuous in } [0,1]\times(-\infty,\infty)) \\
&= \int_0^1\int_{-\infty}^{\infty} e^{itx}(1-\cos hx)\frac{1+x^2}{x^2}dG(x)dh \\
&= \int_0^1 \left(\eta(t) - \frac{1}{2}[\eta(t+h)+\eta(t-h)]\right)dh \\
&= \eta(t) - \frac{1}{2}\int_0^1 [\eta(t+h)+\eta(t-h)]dh \\
&= \lambda(t). \quad \blacksquare
\end{aligned}$$

LEMMA **12.4.2** *There is a 1-to-1 correspondence between the functions $\eta(t)$ given by (4.1) and the pair $(\gamma, G)$, where $\gamma$ 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

$$G(x) = \int_{-\infty}^{x} \frac{d\Lambda(y)}{A(y)}. \tag{4.3}$$

Finally, $\psi$ and $G$ together uniquely defines $\gamma$. $\blacksquare$

LEMMA **12.4.3**

*(i) If $\gamma_n \to \gamma$ and $G_n \Longrightarrow G$, then $\eta_n(t) \to \eta(t) = (\gamma, G)$.*

*(ii) If $\eta_n(t) \to \eta(t)$ and $\eta(t)$ is continuous at 0, then there exist some real constant $\gamma$ 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.*

(i) This part is trivial.

(ii) 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

12.3.1, 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 $\gamma$ such that $\lim_n \gamma_n = \gamma$; and from part (i), we know $\psi = (\gamma, G)$. ∎

### 12.4.2 Proof of sufficiency

Theorem **12.4.3** *If $\psi(t)$ has the representation (4.1), then $\psi(t)$ is i.d.c.f.*

*Proof.* Denote the integral in (4.1) 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

$$\psi(t) = e^{it\gamma} \exp\{I(t)\} = e^{it\gamma} \exp\{I_+(t)\} \exp\{I_-(t)\} \exp\{-\frac{t^2}{2}[G(0+) - G(0-)]\}.$$

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

$$\exp\{I_+(t)\} = \lim_m \exp\{I_+^{1/m}(t)\},$$

where

$$\begin{aligned} I_+^{\epsilon}(t) &= \int_{\epsilon}^{1/\epsilon} g(t,x) dG(x) \\ &= \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1 - \frac{it\xi_k}{1+\xi_k^2}\right) \frac{1+\xi_k^2}{\xi_k^2} [G(x_{k+1}) - G(x_{k+1})] \\ &\qquad (\epsilon = x_0 < x_1 < \ldots < x_n = 1/\epsilon, \quad x_k \leq \xi_k < x_{k+1}) \\ &=: \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1 - \frac{it\xi_k}{1+\xi_k^2}\right) \lambda_{nk} \end{aligned}$$

$$
\begin{aligned}
&=: \lim_n \sum_{k=0}^{n-1} \left(e^{it\xi_k} - 1\right) \lambda_{nk} - it\frac{\xi_k \lambda_{nk}}{1+\xi_k^2} \\
&=: \lim_n \sum_{k=0}^{n-1} \left(ita_{nk} + (e^{it\xi_k} - 1)\lambda_{nk}\right) \\
&=: \lim_n \sum_{k=0}^{n-1} T_{nk},
\end{aligned}
$$

where

$$
\lambda_{nk} = \frac{1+\xi_k^2}{\xi_k^2}[G(x_{k+1}) - G(x_{k+1})], \qquad a_{nk} = -\frac{\xi_k \lambda_{nk}}{1+\xi_k^2}
$$

Note that $\exp(T_{nk}) = \exp\left(ita_{nk} + (e^{it\xi_k} - 1)\lambda_{nk}\right)$ 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

$$
\lim_n \prod_{k=0}^{n-1} \exp\{T_{nk}\} = \lim_n \exp\{\sum_{k=0}^{n-1} T_{nk}\} = \exp\{I_+^\epsilon(t)\}.
$$

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. ∎

### 12.4.3 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

$$
\gamma_n = \int_{-\infty}^{\infty} \frac{nx}{1+x^2} dF_n(x) \to \gamma, \qquad G_n(x) = \int_{-\infty}^{x} \frac{ny^2}{1+y^2} dF_n(y) \to G(x).
$$

Example **12.4.1** *Give the Levy-Khintchine representation for the following r.v.'s*

1. *$N(\mu, \sigma^2)$,*
2. *Poisson$(\lambda)$,*
3. *Degenerate r.v.: $P(X = C) = 1$.*
4. *Cauchy r.v.: $f(x) = \dfrac{a}{\pi} \dfrac{1}{a^2 + x^2}$ and $\psi(t) = e^{-a|t|}$.*

**Solution.** 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,

$$
\gamma = \mu, \qquad G(x) = \sigma^2 \delta_0(x) = \sigma^2 I\{x > 0\}.
$$

2. For Poisson($\lambda$) d.f., we have

$$
\begin{aligned}
\eta(t) &= \lambda(e^{it}-1) = \lambda(e^{itx}-1)\big|_{x=1} \\
&= \left(e^{itx}-1-\frac{itx}{1+x^2}+\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}\Big|_{x=1} \\
&= it\frac{\lambda}{2}+\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}\Big|_{x=1} \\
&= it\frac{\lambda}{2}+\int_{-\infty}^{\infty}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\frac{\lambda}{2}d\delta_1(x).
\end{aligned}
$$

So

$$\gamma = \frac{\lambda}{2}, \qquad G(x) = \frac{\lambda}{2}\delta_1(x) = \frac{\lambda}{2}I\{x>1\}.$$

3. For degenerate d.f. $X = C$ a.s., we have

$$\eta(t) = itC.$$

So

$$\gamma = C, \qquad G(x) \equiv 0. \quad \blacksquare$$

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,

$$\gamma = 0, \qquad G(x) = \frac{a}{\pi}\int_{-\infty}^{x}\frac{1}{1+y^2}dy. \quad \blacksquare$$

## 12.5 Levy formula

We now give another very useful representation: **Levy formula**. Let us rewrite the Levy-Khintchine representation as

$$
\begin{aligned}
\psi(t) &= \exp\left\{it\gamma+\left(\int_{\{x=0\}}+\int_{\{|x|>0\}}\right)\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x)\right\} \\
&= \exp\left\{it\gamma-\frac{1}{2}\sigma^2t^2+\int_{\{|x|>0\}}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)dL(x)\right\}.
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma^2 &= G(0+)-G(0-), \\
dL(x) &= \frac{1+x^2}{x^2}dG(x), \qquad x \neq 0.
\end{aligned}
$$

Several remarks about $L(x)$ are given below.

1. $L(x)$ is defined on $\mathcal{R} - \{0\}$. It is easy to see that

$$\begin{aligned} L(x) &= C_1 + \int_{-\infty}^{x} \frac{1+y^2}{y^2} dG(y), \qquad \text{if } x < 0, \\ & \quad C_2 - \int_{x}^{\infty} \frac{1+y^2}{y^2} dG(y), \qquad \text{if } x > 0, \end{aligned}$$

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

$$\lim_{x \to -\infty} L(x) = C_1, \qquad \lim_{x \to \infty} L(x) = C_2. \tag{5.4}$$

2. Note that, for every finite $\delta > 0$, we have

$$\int_{0<|x|<\delta} x^2 dL(x) = \int_{0<|x|<\delta} \left(1 + x^2\right) dG(x) \leq \left(1 + \delta^2\right) \int_{0<|x|<\delta} dG(x) < \infty,$$

3. In view of (5.4), 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 $x \neq 0$. 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

$$L((-\epsilon, \epsilon)^c) = \int_{|x|>\epsilon} dL(x) = \int_{|x|>\epsilon} \frac{1+x^2}{x^2} dG(x) \leq \left(1 + \epsilon^{-2}\right) \int_{|x|>\epsilon} dG(x) < \infty.$$

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 **12.5.1 (Levy formula)** *A function $\psi(t)$ is an i.d.c.f. if and only if it admits the following (unique) representation*

$$\psi(t) = \exp\left\{ it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{|x|>0} \left( e^{itx} - 1 - \frac{itx}{1+x^2} \right) dL(x) \right\}, \tag{5.5}$$

*where $\gamma$ is a real constant, $\sigma^2$ is a non-negative constant, and the function $L$ is non-decreasing on the intervals $(-\infty, 0)$ and $(0, \infty)$, and satisfies*

$$\int_{0<|x|<\delta} x^2 dL(x) < \infty, \qquad \textit{for every finite } \delta > 0. \ \blacksquare$$

REMARK **12.5.1** *An alternative Levy formula takes the following form:*

$$\begin{aligned} \psi(t) &= \exp\left\{ it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{|x|>0} \left( e^{itx} - 1 - itxI\{|x| < 1\} \right) dL(x) \right\}. \\ &= \exp\left\{ it\gamma - \frac{1}{2}\sigma^2 t^2 + \int_{0<|x|<1} \left( e^{itx} - 1 - itx \right) dL(x) + \int_{|x|\geq 1} \left( e^{itx} - 1 \right) dL(x) \right\}. \end{aligned}$$

*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.* $\blacksquare$

Example **12.5.1** *Redo Example 12.4.1.*

**Solution.** (1) For Normal d.f., we have $\eta(t) = it\mu - \frac{1}{2}\sigma^2 t^2 + 0$ So

$$\gamma = \mu, \qquad \sigma^2 = \sigma^2, \qquad L(x) = 0.$$

The other two cases are left as exercises. ∎

## 12.6 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 **12.6.1 (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*

$$\psi(t) = \exp\left\{ it\gamma + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - itx\right) \frac{1}{x^2} dK(x) \right\}, \tag{6.6}$$

*where $\gamma$ is a real constant, the function $K$ is a bounded non-decreasing function.* ∎

## 12.7 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:

- If $E|X_1| < \infty$, then $\bar{X} \to_d \mu$, whose c.f is $e^{it\mu}$.
- If $E|X_1| < \infty$, then $\sqrt{n}(\bar{X} - \mu)/\sigma \to_d N(0,1)$, whose c.f is $e^{-t^2/2}$
- If $X_{nk}$ are i.i.d. Bin$(n, p = \lambda/n)$, then $\sum X_{nk} \to_d Poisson(\lambda)$, whose c.f. is $e^{\lambda(e^{it}-1)}$.
- If $X_k$ are i.i.d. 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 **12.7.1** *Let $\sum X_{nk}$ be independent r.v.s satisfying the following infinitesimal condition:*

$$\max_{1 \le k \le n} P(|X_{nk}| > \epsilon) \to 0, \qquad \textit{as } n \to \infty.$$

*Then,*

$$\{\textit{all limiting d.f.s of } \textstyle\sum X_{nk}\} = \{\textit{ all i.d. d.f.s}\}$$

## 12.8 Exercises

1. If $\psi$ is an i.d. c.f., then $\psi^\lambda$ is a c.f. for every $\lambda \ge 0$.

   *Proof.* Note that $\psi(t) = (\psi_n(t))^n$ for all $n \ge 1$.

   First consider $\lambda$ to be a rational number $j/k$. We choose $n = k$ above to get $(\psi(t))^{j/k} = (\psi_n(t))^{nj/k} = (\psi_k(t))^j$, which is a c.f. (a convolution of $j$ components).

   Next if $\lambda$ is irrational, it can be approximated by a sequence of rational numbers. The result follows by taking the limit and applying Levy Continuity Theorem. ∎

# REFERENCES


Chow, Y. S. and Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

Chung, K. L. (1974) *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

Durrett, R. (1991). *Probability: Theory and Examples.* Duxbury Press.

Grimmett, G. R. and Stirzaker, D. R. (1991). *Probability and Random Processes*, 2nd ed. Oxford Scientific Publications.

Petrov, V. V. (1975). Sums of independent random variables. Springer-Verlag, New York, Heidelberg, Berlin,

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley and Sons, New York.