# MATH1312: Lecture Note on Probability Theory and Mathematical Statistics

Shixiao W. Jiang

Institute of Mathematical Sciences, ShanghaiTech University, Shanghai 201210, China

jiangshx@shanghaitech.edu.cn

2024 年 11 月 28 日

# Contents

# Chapter 1

# Limit Theorems

## 1.1 Convergence in Probability

See the introduction to measure theory in Chapter 1, including the reference website:

https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/02%3A_Probability_Spaces/2.03%3A_Probability_Measures

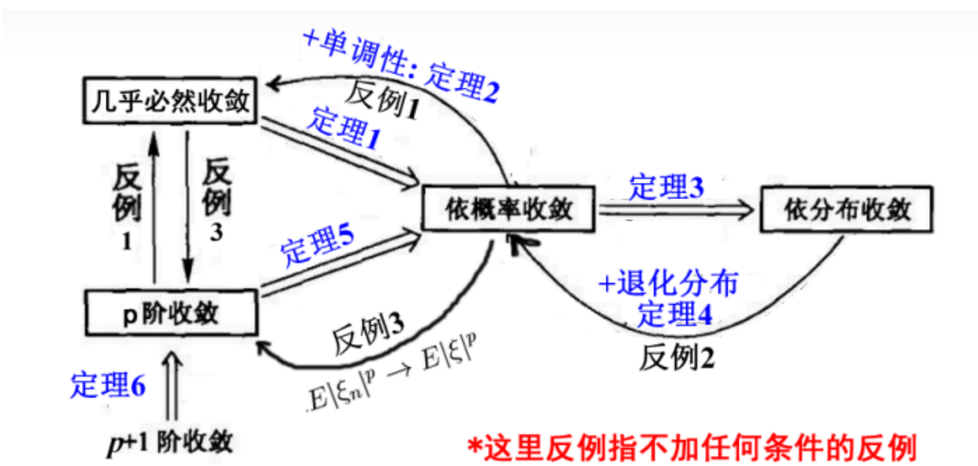


Figure 1.1: The relation among various of convergence.

### 1.1.1 Definitions

**Definition 1.1.1** *(**Almost Surely Convergence**). If there exists $A \in \mathcal{F}$ such that $P(A) = 0$ ($A$ is a zero measure set) and $\forall w \in A^c$, $\lim_{n\to\infty} X_n(w) = X(w)$, then $X_n \overset{a.s}{\to} X (n \to \infty)$. Then $X_n$ is said to be convergent to $X$ almost surely or almost everywhere.*

**Definition 1.1.2** *(**Convergence with Probability 1**). If $P(\lim_{n\to\infty} X_n = X) = 1$, then we say $X_n$ converges to $X$ with probability 1, denoted by $X_n \overset{w.p.1}{\longrightarrow} X$.*

**Remark 1.1.3** *The above two definitions are exactly equivalent. In addition, one more equivalent definition can be found in equation (1.37).*

**Definition 1.1.4** *(**Stochastic Convergence**). The sequence $\{X_n\}$ of random variables is called stochastically convergent to zero if for every $\epsilon > 0$ the relation*

$$\lim_{n\to\infty} P\left(|X_n| > \epsilon\right) = 0 \tag{1.1}$$

*is satisfied.*

**Definition 1.1.5** *(**Converge in probability**). If for any $\epsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$, then $X_n$ is said to be convergent in probability to $X$. We denote by $X_n \xrightarrow{P} X, n \to \infty$.*

**Remark 1.1.6** *The above two definitions are completely equivalent but only with terminologies different. Nowadays, we use terminology "convergence in probability" and rarely see terminology "stochastic convergence" (as I understand).*

**Definition 1.1.7** *(**Weak convergence of distribution functions**). The sequence $\{F_n(x)\}$ of distribution functions of random variables $\{X_n\}$ is called weakly convergent to $F(x)$, denoted by $F_n \overset{w}{\to} F$, if there exists a non-decreasing and non-negative function $F(x)$, which mat not be a distribution function, such that at every continuity point of $F(x)$, the relation*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

*is satisfied.*

**Definition 1.1.8** *(**Converge in distribution of random variables**). Let $F_n$ and $F$ be distribution functions of $X_n$ and $X$, respectively. If $F_n \overset{w}{\to} F$, then we say $\{X_n\}$ is convergent in distribution to $X$, denoted by $X_n \overset{d}{\to} X$. The distribution function $F(x)$ is called the limit distribution function.*

**Example 1.1.9** *$F$ may not be a distribution function. Consider the sequence $\{X_n\}$ of random variables with the one-point distribution given by $P(X_n = n) = 1, n = 1, 2, \cdots$. The distribution function $F_n(x)$ of $X_n$ is*

$$F_n(x) = \begin{cases} 0 & \textit{for} \quad x \leq n, \\ 1 & \textit{for} \quad x > n, \end{cases}$$

*We have for $\forall x \in \mathbb{R}$,*

$$\lim_{n\to\infty} F_n(x) = 0.$$

*Thus $F(x) \equiv 0$. The sequence $\{F_n(x)\}$ is not convergent to a distribution function.*

Now let us review some properties of a CDF in the previous chapter.

**Theorem 1.1.10** *The single-valued function $F(x)$ is a distribution function if and only if it is non-decreasing, continuous at least from the left, and satisfies $F(-\infty) = 0$, $F(+\infty) = 1$.*

**Proposition 1.1.11** *The set of points of discontinuity is at most countable for a CDF $F(x)$. Hence a CDF is almost everywhere continuous.*

**Definition 1.1.12** *(**Norm on Probability Space**). Let $p \in (1, +\infty)$ and $L^p(\Omega, \mathcal{F}, P) := \{X : E|X|^p < +\infty\}$, where $\|X\|_p := (E|X|^p)^{\frac{1}{p}}$. Then $\|\cdot\|_p$ is a norm on $L^p(\Omega, \mathcal{F}, P)$, satisfying non-negativity, homogeneity and triangle inequality.*

**Definition 1.1.13** *($L^p$ **convergence**). Let $\{X_n; X\} \subset L^p(\Omega, \mathcal{F}, P)$. If $\lim\limits_{n\to\infty} \|X_n - X\|_p = 0$, then $\{X_n\}$ is said to be $L^p$ convergent to $X$, denoted by $X_n \xrightarrow{L^p} X$.*

### 1.1.2 Appendix for Weak Convergence

See my local folder jiaoxue limit theorem for convergence for reference.

为什么分布函数的收敛叫弱收敛?

**We now give an intuition why the convergence of distribution functions is called the weak convergence.**

**Proposition 1.1.14** *(**The Portmanteau Theorem**). The following statements are equivalent.*
*(1). $X_n \xrightarrow{d} X$.*
*(2). $E(h(X_n)) \to E(h(X))$ for all continuous functions $h : \mathbb{R}^d \to \mathbb{R}$ that are nonzero only on a closed and bounded set.*
*(3). $E(h(X_n)) \to E(h(X))$ for all **bounded** continuous functions $h : \mathbb{R}^d \to \mathbb{R}$.*
*(4). $E(h(X_n)) \to E(h(X))$ for all bounded measurable functions $h : \mathbb{R}^d \to \mathbb{R}$ for which $P(X \in \{x : h$ is continuous at $x\}) = 1$.*

上述性质中(1)和(3)等价，从(3)的角度就理解为什么依分布收敛叫弱收敛，因为在(3)中可以 理解有界连续函数类$h$是试验函数(test function).

**Definition 1.1.15** *For random variables $X_n \in \mathbb{R}$ and $X \in \mathbb{R}$, $X_n$ converges in distribution to $X$,*

$$X_n \xrightarrow{d} X,$$

*if for all $x$ such that $x \mapsto P(X \le x)$ is continuous,*

$$P(X_n \le x) \to P(X \le x) \text{ as } n \to \infty.$$

**Definition 1.1.16** *For metric space-valued random variables $X_n$ and $X$, $X_n$ converges in distribution to $X$ if for all bounded continuous $h$*

$$E[h(X_n)] \to E[h(X)] \text{ as } n \to \infty.$$

*Note that boundedness of $h$ in the Portmanteau theorem is important.*

其实也有参考书中，直接把(3)作为定义，再证明定义(1)和(3)的等价性，注意定义中$h$有界很重要.

### 1.1.3 Theorems

All the followings correspond to theorems and counter-examples in Fig. 1.1.

**Theorem 1.1.17** *(**Theorem 1**). If $\xi_n \xrightarrow{\text{a.s.}} \xi$, then $\xi_n \xrightarrow{P} \xi$ for $n \to \infty$.*

**Proof.** We have

$$\xi_n \xrightarrow{a.s} \xi \Leftrightarrow \forall \epsilon > 0, P(\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} [|\xi_n - \xi| \geq \epsilon]) = 0.$$

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \epsilon) \leq \lim_{k\to\infty} P(\bigcup_{n=k}^{\infty} |\xi_n - \xi| \geq \epsilon) = P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} |\xi_n - \xi| \geq \epsilon) = 0.$$

Thus, $\xi_n \xrightarrow{P} \xi, n \to \infty$. ■

**Remark 1.1.18** *(**Theorem 2**). I did not check the proof for Theorem 2.*

参考 概率论复习笔记(9) – 几种收敛的关系 - 知乎

**Theorem 1.1.19** *(**Theorem 3**). $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.* **[there is another proof using Lévy theorem (see Appendix).]**

**Proof.** We first prove two Lemmas.

① $P(X+Y \leq a+b) \leq P(\{X \leq a\} \cup \{Y \leq b\}) \leq P(X \leq a) + P(Y \leq b)$ since $\{X+Y \leq a+b\} \subset \{X \leq a\} \cup \{Y \leq b\}$ and $P(A \cup B \subset \Omega) = P(A) + P(B) - P(AB)$.

② $P(X+Y \leq a+b) \geq P(X \leq a$ and $Y \leq b)$ since $\{X+Y \leq a+b\} \supset \{X \leq a\}$ and $(\cap)\{Y \leq b\}$.

1. $F(x)$ is right continuous at $x_0$.

$$\begin{aligned} F_n(x_0) = P(X_n \leq x_0) &= P(X_n - X + X \leq x_0 - \epsilon + \epsilon) \\ &\leq P(X_n - X \leq -\epsilon) + P(X \leq x_0 + \epsilon) \\ &\leq P(|X_n - X| \geq \epsilon) + P(X \leq x_0 + \epsilon) \\ &\overset{n\to\infty}{\to} 0 + P(X \leq x_0 + \epsilon) \\ &\underset{\text{cont. at } x_0}{\overset{\epsilon\to 0}{\longrightarrow}} F(x_0). \quad \lim_{n\to\infty} F_n(x_0) \leq F(x_0). \end{aligned}$$

2. $F(x)$ is left continuous at $x_0$.

$$\begin{aligned} F_n(x_0) = P(X_n \leq x_0) &= P(X_n - X + X \leq x_0 - \epsilon + \epsilon) \\ &\geq P(X_n - X \leq \epsilon \text{ and } X \leq x_0 - \epsilon) \\ &\geq P(X \leq x_0 - \epsilon) - P(X_n - X > \epsilon) \quad (P(A \text{ and } B) + P(B^c) \geq P(A)) \\ &\geq P(X \leq x_0 - \epsilon) - P(|X_n - X| \geq \epsilon) \\ &\overset{n\to\infty}{\to} P(X \leq x_0 - \epsilon) - 0 \\ &\underset{\text{cont. at } x_0}{\overset{\epsilon\to 0}{\longrightarrow}} F(x_0). \end{aligned}$$

Thus $\lim_{n\to\infty} F_n(x_0) = F(x_0)$. ■

**Theorem 1.1.20** *(**Theorem 3 and 4**). $X_n \xrightarrow{P} C$ $(n \to \infty)$ if and only if $X_n \xrightarrow{d} C$ $(n \to \infty)$ for constant $C$.*

**Proof.** See proof later in Theorem 1.3.4. ■

**Theorem 1.1.21** *(**Theorem 5**). $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{P} X$ and $E|X_n|^p \to E|X|^p$.*

**Proof.**

$$P(|X_n - X| \geq \epsilon) \overset{Chebyshev}{\leq} \frac{1}{\epsilon^p} E|X_n - X|^p = \frac{\|X_n - X\|_p^p}{\epsilon^p} \to 0, \quad n \to \infty.$$

Thus $X_n \overset{P}{\to} X$.

$$|\|X_n\|_p - \|X\|_p| \leq \|X_n - X\|_p \to 0, \quad n \to \infty.$$

Thus $E|X_n|^p \to E|X|^p$. ■

**Theorem 1.1.22** *(**Theorem 6**).* $\xi_n \xrightarrow{L^{p+1}} \xi \Rightarrow \xi_n \xrightarrow{L^p} \xi$.

**Proof.** Young or Cauchy-Schwarz Inequality. ■

### 1.1.4 Appendix for Proof of Convergence in Probability to Convergence in Distribution

We now give another proof for $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.

See references in folder limit theorem for convergence.

See my local folder jiaoxue limit theorem for convergence for reference.

See website on https://www.statlect.com/asymptotic-theory/Slutsky-theorem, equivalently Taboga, Marco (2021). "Slutsky's theorem", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing.

We first provide two important theorems in probability theory.

**Theorem 1.1.23** *(Continuous Mappling Theorem or CMT). Let $g$ be continuous on a set $B$ such that $P(x \in B) = 1$. Then*
*(1) $X_n \xrightarrow{P} X$ implies $g(X_n) \xrightarrow{P} g(X)$.*
*(2) $X_n \xrightarrow{d} X$ implies $g(X_n) \xrightarrow{d} g(X)$.*
*(3) $X_n \xrightarrow{a.s.} X$ implies $g(X_n) \xrightarrow{a.s.} g(X)$.*

**Theorem 1.1.24** *(Slutsky's Theorem). Slutsky's theorem concerns the convergence in distribution of the transformation of two sequences of random vectors, one converging in distribution and the other converging in probability to a constant.*
*(1) (Joint Convergence) Let $X_n$ and $Y_n$ be two sequences of random vectors. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$, where $c$ is a constant, then*

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}.$$

*(2) (Continuous Mapping Convergence) Let $g(x, y)$ be a continuous function. Then,*

$$g(X_n, Y_n) \xrightarrow{d} g(X, c).$$

*(3) (Sum and Product Convergence) Above Slutsky's theorem implies that*

$$X_n + Y_n \xrightarrow{d} X + c,$$
$$X_n Y_n \xrightarrow{d} cX.$$

*(4) If $X_n \xrightarrow{d} X$ and $X_n - Z_n \xrightarrow{d} 0$, then*

$$Z_n \xrightarrow{d} X.$$

I now show a proof for $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.

**Proof.** We have $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$. For a continuous and bounded function $g$, we have from Continuous Mapping Theorem that,

$$\lim_{n\to\infty} P(|g(X_n) - g(X)| \geq \epsilon) = 0.$$

Let $|g(x)| \leq M$ for all $x \in \mathbb{R}$. Then

$$\begin{aligned} E|g(X_n) - g(X)| &= \int |g(X_n) - g(X)| dF(X_n, X) \\ &= \int_{|g(X_n)-g(X)|<\epsilon} |g(X_n) - g(X)| dF + \int_{|g(X_n)-g(X)|\geq\epsilon} |g(X_n) - g(X)| dF \\ &\leq \epsilon \int dF + 2MP(|g(X_n) - g(X)| \geq \epsilon) \\ &= \epsilon + 2MP(|g(X_n) - g(X)| \geq \epsilon) \to \epsilon. \end{aligned}$$

Based on the arbitrary of $\epsilon$, we have $\lim_{n\to\infty} E|g(X_n) - g(X)| = 0$, which implies that

$$Eg(X_n) \to Eg(X),$$

for any bounded and continuous $g$. Thus, $X_n \xrightarrow{d} X$. ■

In fact, there is even one more proof from website which seems correct.

**Proof.** For any continuous and bounded function $f$, by Continuous Mapping Theorem, we have

$$X_n \xrightarrow{P} X \;\Rightarrow\; f(X_n) \xrightarrow{P} f(X).$$

By Dominate Convergence Theorem, the online states that

$$Ef(X_n) \to E(f(X)). \tag{1.2}$$

(I am not for sure this step. For DCT, we need three conditions, (i) $|f(X_n)| \leq M$, *a.s.* for all $n$, (ii) $f(X_n) \xrightarrow{P} f(X)$ or $f(X_n) \xrightarrow{a.s.} f(X)$, (iii) $E(M) = M < \infty$. I am not for sure if the convergence in probability in (ii) is already enough for the DCT.) (It seems answered on a forum but I have not found book or paper. https://math.stackexchange.com/questions/206851/generalisation-of-dominated-convergence-theorem# or https://math.stackexchange.com/questions/3374830/dominated-convergence-theorem-with-almost-surely-replaced-by-convergence-in-p) Then one can obtain the convergence in distribution based on the Portmanteau Theorem, that is, $X_n$ converges in distribution to $X$ if equation (1.2) holds for all bounded continuous $f$. The other last step, one can take $f(x) = e^{itx}$ which is a bounded and continuous function. Then based on Levy-Cramer Theorem, the convergence of Ch.f.s is equivalent to the convergence of distribution functions. ■

### 1.1.5 Counter-Examples

**Example 1.1.25** ***(Counter-Example 1).*** *$L^p$ convergence $\Rightarrow$ a.e. convergence?*

*No. For simplicity, consider the interval $\Omega = [0,1]$ and construct a sequence of sets $A_n$ such that the measures $m(A_n)$ tend to 0 but every point belongs to infinitely many $A_n$. For example $A_1 = [0,1/2]$, $A_2 = [1/2,1]$, $A_3 = [0,1/4]$,...,$A_6 = [3/4,1]$, $A_7 = [0,1/8]$,.... If $f_n$ is the indicator function of $A_n$, that is $f_n(x) = 1$ if $x \in A_n$ and $f_n(x) = 0$ else, then $f_n \to 0$ in all $L^p([0,1])$ because $\|f_n\|_p = [1^p \cdot m(A_n) + 0^p \cdot (1 - m(A_n))]^{1/p} \to 0$ but there is no $x \in [0,1]$ with $f_n(x) \to 0$.*

*See reference in https://math.stackexchange.com/questions/689664/lp-convergence-a-e-convergence*

**Example 1.1.26** ***(Counter-Example 2).*** $X_n \stackrel{d}{\to} X \nRightarrow X_n \stackrel{P}{\to} X$.

*Let* $X_n = -X, n = 1, 2, \cdots$ *and let* $X_n$ *and* $X$ *have the following distributions.*

$$\begin{array}{ccc} X & -1 & +1 \\ P & \frac{1}{2} & \frac{1}{2} \\ X_n & -1 & +1 \\ P & \frac{1}{2} & \frac{1}{2} \end{array}$$

*Then we have*

$$X_n \stackrel{d}{\to} X.$$

*However,*

$$P(|X_n - X| > \epsilon) = P(|-2X| > \epsilon) = P(|X| > \frac{\epsilon}{2}) = 1 \nrightarrow 0,$$

*when* $\epsilon < 2$, $X_n \stackrel{P}{\nrightarrow} X$.

**Example 1.1.27** ***(Counter-Example 3).*** $X_n \stackrel{a.s.}{\to} X$ *or* $X_n \stackrel{P}{\to} X \nRightarrow X_n \stackrel{L^p}{\to} X$.

*Let* $\Omega = [0, 1], \xi(w) \equiv 0, P(\xi = 0) = 1$.

$$\xi_n(w) = \begin{cases} n^{\frac{1}{p}} & 0 \le w \le \frac{1}{n}, \\ 0 & \frac{1}{n} \le w \le 1. \end{cases}$$

$$P(\xi_n = n^{\frac{1}{p}}) = \frac{1}{n},$$

$$P(\xi_n = 0) = 1 - \frac{1}{n}.$$

$$\forall w \in \Omega, \ \lim_{n\to\infty} \xi_n(w) \to \xi(w), \ \xi_n \stackrel{a.s}{\to} \xi.$$

$$\forall \epsilon > 0, \ P(|\xi_n(w) - \xi(w)| \ge \epsilon) \le \frac{1}{n} \Rightarrow \xi_n \stackrel{P}{\to} \xi.$$

*However, we see that*

$$E|\xi_n - \xi|^p = (n^{\frac{1}{p}})^p \cdot \frac{1}{n} = 1 \ne 0.$$

### 1.1.6 Appendix for Other Convergence

There are some other useful relationships between convergence in probability theory and measure theory.

Convergence in KL divergence $\Rightarrow$ Convergence in total variation $\Rightarrow$ strong convergence of measure $\Rightarrow$ weak convergence, where

i. $\mu_n \stackrel{TV}{\longrightarrow} \mu$ means $\lim_{n\to\infty} \|\mu_n - \mu\|_{TV} = 0$, where

$$\|\mu_n - \mu\|_{TV} = \sup_{\|f\|_\infty \le 1} \{\int f d\mu_n - \int f d\mu\},$$

which also equals

$$\|\mu_n - \mu\|_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)|.$$

ii. $\mu_n \to \mu$ strongly if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$, $\forall A \in \mathcal{F}$.

## 1.2 Preliminary Remarks

1. Theorems of de Moivre-Laplace, Lindeberg-Lévy, Lapunov and Lindeberg-Feller.

2. Laws of large numbers.

3. Modern theory of limit distributions for sums of independent random variables has developed greatly during 1900-1950 due mainly to Khintchin, Gnedenko, Kolmogorov and Lévy.

4. For dependent random variables, the convergence of a sequence of distribution functions is also interesting. See Markov, Bernstein.

## 1.3 Stochastic Convergence (Convergence in Probability)

### 1.3.1 Part A. Example

**Example 1.3.1** *The random variable $Y_n$ can take on the values*

$$0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1$$

*and its probability function is given by the formula*

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r}\frac{1}{2^n} \qquad (r = 0, 1, \cdots, n). \tag{1.3}$$

*Consider the random variable $X_n$ defined by the formula*

$$X_n = Y_n - \frac{1}{2}. \tag{1.4}$$

*Thus $X_n$ can take on the values*

$$-\frac{1}{2}, \frac{2-n}{2n}, \frac{4-n}{2n}, \cdots, \frac{n-4}{2n}, \frac{n-2}{2n}, \frac{1}{2}.$$

*The probability function of $X_n$ is given by the formula*

$$P\left(X_n = \frac{2r-n}{2n}\right) = \binom{n}{r}\frac{1}{2^n}.$$

*Let $n = 2$. The random variable $X_2$ can take on the values*

$$-0.5, 0, 0.5$$

*with the respective probabilities*

$$\frac{1}{4}, \frac{1}{2}, \frac{1}{4}.$$

*Let $\epsilon = 0.3$, then*

$$P(|X_2| > 0.3) = P\left(X_2 = -\frac{1}{2}\right) + P\left(X_2 = \frac{1}{2}\right) = 0.5.$$

*Let $n = 5$. The random variable $X_5$ can take on the values*

$$-0.5, -0.3, -0.1, 0.1, 0.3, 0.5$$

*with the respective probabilities*

$$\frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32}.$$

*Hence*

$$P\left(|X_5| > 0.3\right) = 0.0625.$$

*Let $n = 10$. The random variable $X_{10}$ can take on the values*

$$-0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5$$

*with the respective probabilities*

$$\frac{1}{1024}, \frac{10}{1024}, \frac{45}{1024}, \frac{120}{1024}, \frac{210}{1024}, \frac{252}{1024}, \frac{210}{1024}, \frac{120}{1024}, \frac{45}{1024}, \frac{10}{1024}, \frac{1}{1024}.$$

*Hence*

$$P\left(|X_{10}| > 0.3\right) \cong 0.02,$$

*which is very small.*
*We will prove* $\lim_{n\to\infty} P\left(|X_n| > 0.3\right) = 0$.

### 1.3.2 Part B. Theory

**Definition 1.3.2** *The sequence $\{X_n\}$ of random variables is called stochastically convergent to zero if for every $\epsilon > 0$ the relation*

$$\lim_{n\to\infty} P\left(|X_n| > \epsilon\right) = 0 \tag{1.5}$$

*is satisfied.*

**Remark 1.3.3** *If $\{X_n\}$ is stochastically convergent to zero , it does not follow that for every $\epsilon > 0$, we can find a a finite $n_0$ such that for all $n > n_0$ the relation $|X_n| < \epsilon$ will be satisfied. It follows only that the probability of the event $\{|X_n| \geq \epsilon\}$ tends to zero as $n \to \infty$.*

**Theorem 1.3.4** *Let $F_n(x)$ be the distribution function of the random variable $X_n$. The sequence $\{X_n\}$ is stochastically convergent to zero if and only if the sequence $\{F_n(x)\}$ satisfies the relation*

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0 & \text{for} \quad x \leq 0, \\ 1 & \text{for} \quad x > 0, \end{cases} \tag{1.6}$$

*at every continuity point.*

**Proof.**
$(\Rightarrow)$. $\{X_n\}$ is stochastically convergent to zero. $\lim_{n\to\infty} P(|X_n| > \epsilon) = 0$.

$$P(X_n < -\epsilon) = F_n(-\epsilon) \to 0 \Rightarrow F_n(-x) \to 0, \quad \text{for } \forall x > 0.$$

$$P(X_n > \epsilon) = 1 - F_n(\epsilon) - P(X_n = \epsilon) \to 0.$$

Since $\forall\epsilon > 0$ there exists $0 < \epsilon_1 < \epsilon$, we have $P(X_n = \epsilon > \epsilon_1) \leq P(|X_n| > \epsilon_1) \to 0$. Thus

$$1 - F_n(\epsilon) \to 0 \Rightarrow 1 - F_n(x) \to 0, \quad \text{for } \forall x > 0.$$

$(\Leftarrow)$. $\forall\epsilon > 0$,

$$\lim_{n\to\infty} P(X_n < -\epsilon) = \lim_{n\to\infty} F_n(-\epsilon) = 0.$$

$$\lim_{n\to\infty} P(X_n > \epsilon) \leq \lim_{n\to\infty} P(X_n \geq \epsilon) = \lim_{n\to\infty} [1 - F_n(\epsilon)] = 0.$$

■

**Remark 1.3.5** *$F(x)$ correspond to the random variable $X$ with a one-point distribution such that $P(X = 0) = 1$. $F(x)$ is continuous at every point $x \neq 0$ so that $F_n(x) \to F(x)$ at $x \neq 0$.*

**Remark 1.3.6** *We stress the fact that at the discontinuity point of $F(x)$, that is, at the point $x = 0$, the sequence $\{F_n(0)\}$ may not converge to $F(0)$.*

**Remark 1.3.7** *Let $X_n \xrightarrow{d} c \neq 0$. We can consider $\{Y_n\} = \{X_n - c\}$. Then $Y_n \xrightarrow{d} 0$ so that $\{Y_n\}$ is stochastically convergent to zero. The theorem holds.*

**Remark 1.3.8** *Let $X_n \xrightarrow{P} X \neq 0$ ($X$ is a random variable). Then we have $X_n \xrightarrow{d} X$. However, the inverse may not hold, that is, the theorem does not hold (see Example 1.1.26).*

## 1.4 Bernoulli's Law of Large Numbers

Denote by $\{Y_n\}$ the sequence of random variables with probability functions given by the formula

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r} p^r (1-p)^{n-r},$$

where $0 < p < 1$ and $r$ can take on the values $0, 1, 2, \cdots, n$. Further denote

$$X_n = Y_n - p. \tag{1.7}$$

**Theorem 1.4.1** *The sequence of random variables $\{X_n\}$ given by (1.7) is stochastically convergent to 0, that is, for any $\epsilon > 0$ we have*

$$\lim_{n\to\infty} P(|X_n| > \epsilon) = 0. \tag{1.8}$$

**Proof.** We compute

$$E(X_n) = 0.$$

$$\sigma_n = \sqrt{Var(X_n)} = \sqrt{p(1-p)/n}.$$

By Chebyshev inequality

$$P(|X_n| > \epsilon) \leq \frac{Var(X_n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} < \frac{1}{n\epsilon^2} \to 0, \quad \text{for} \quad \forall\epsilon > 0.$$

■

## 1.5 The Convergence of A Sequence of Distribution Functions

### 1.5.1 Part A. Example

**Definition 1.5.1** *The sequence* $\{F_n(x)\}$ *of distribution functions of the random variables* $\{X_n\}$ *is called (weakly) convergent, if there exists a distribution function* $F(x)$ *such that,* ***at every continuity point of*** $F(x)$*, the relation*

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \textit{(at any continuity point } x\textit{)} \tag{1.9}$$

*is satisfied.*

**Remark 1.5.2** *It is not required that* $\{F_n(x)\}$ *converge to* $F(x)$ *at the discontinuity points of* $F(x)$*. See example in Section 1.3.* $F_n(0)$ *is not convergent to* $F(0)$*.*
*Consider the subsequence of* $\{F_n(0)\}$ *with only* $n = 2k+1$*.* $X_{2k+1}$ *can take on the values*

$$-\frac{1}{2}, \frac{2-(2k+1)}{2(2k+1)}, \frac{4-(2k+1)}{2(2k+1)}, \cdots, \frac{2k+1-4}{2(2k+1)}, \frac{2k+1-2}{2(2k+1)}, \frac{1}{2}.$$

*For every* $k$*, half of these terms are less than zero, the other half greater than zero.* $P(X_{2k+1} < 0) = F_{2k+1}(0) = 0.5$*.*

$$\lim_{k\to\infty} F_{2k+1}(0) = 0.5 \neq 0 = F(0).$$

**Remark 1.5.3** *Recall that it may happen that a sequence of distribution functions converges to a function that is not a distribution function.*

**Example 1.5.4** *For example,*

$$F_n(x) = \begin{cases} 0 & \textit{for} \quad x \leq n, \\ 1 & \textit{for} \quad x > n, \end{cases}$$

$$\lim_{n\to\infty} F_n(x) = F(x) \equiv 0 \qquad (-\infty < x < \infty).$$

**Remark 1.5.5** *Let* $a < b$ *be two continuity points of the limit distribution function* $F(x)$*. We have*

$$P(a \leq X_n < b) = F_n(b) - F_n(a). \tag{1.10}$$

*Since* $\lim_{n\to\infty} F_n(a) = F(a)$ *and* $\lim_{n\to\infty} F_n(b) = F(b)$*, we have*

$$\lim_{n\to\infty} P(a \leq X_n < b) = F(b) - F(a). \tag{1.11}$$

### 1.5.2 Part B. Weak Convergence for High Dimensional Distributions

**Definition 1.5.6** *The sequence of distribution functions* $\{F_n(x_1, \cdots, x_k)\}$ *of random vectors* $(X_{n1}, X_{n2}, \cdots, X_{nk})$ *is (weakly) convergent if there exists a distribution function* $F(x_1, \cdots, x_k)$ *such that at every one of its continuity points*

$$\lim_{n\to\infty} F_n(x_1, x_2, \cdots, x_k) = F(x_1, x_2, \cdots, x_k). \tag{1.12}$$

## 1.6 The Riemann-Stieltjes Integral

There is one-to-one map between a CDF $F(x)$ and a Ch.f. $\phi(x)$. To prove this, we need some backgrounds.

上确界，全变差

**Definition 1.6.1** *Let $F(x)$ be a function defined on the interval $[a, b]$, which can be either finite or infinite. Let us take a partition of the interval $[a, b]$ with the points*

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

*and form the sum*

$$T = \sum_{k=0}^{n-1} |F(x_{k+1}) - F(x_k)|.$$

*The value of $T$ may depend on the number $n$ and on the partition into subintervals. The **least upper bound** of the values of $T$ is called the **total absolute variation** of the function $F(x)$ in the interval $[a, b]$.*

$$\textit{Total variation} = \sup_{n, x_k} \quad T_{n, x_k}.$$

**Definition 1.6.2** *If the total absolute variation of $F(x)$ in $[a, b]$ is finite, we say that $F$ is a function of bounded variation on the interval $[a, b]$. The set of all such functions is denoted by $BV([a, b])$.*

**Proposition 1.6.3** *Every non-decreasing bounded function is of bounded variation.*

$$T = \sum_{k=0}^{n-1} [F(x_{k+1}) - F(x_k)] = F(b) - F(a) < \infty.$$

*Every distribution function $F(x)$ is a function of bounded variation.*

$$T = F(+\infty) - F(-\infty) = 1.$$

We now introduce Stieltjes integral.

**Definition 1.6.4** *Given a function $g(x)$ and a function $F(x)$ in a finite interval $[a, b]$. Let us form a partition of the interval $[a, b]$ into $n$ parts with the points*

$$a = x_0 < x_1 < \cdots < x_n = b.$$

*Consider the sum*

$$S = \sum_{k=0}^{n-1} g(x'_k)[F(x_{k+1}) - F(x)], \tag{1.13}$$

*where $x'_k$ is an arbitrary point in the $k$th interval $(x_k, x_{k+1})$. If as $n \to \infty$ and $\max(x_{k+1} - x_k) \to \infty$ the sum $S$ tends to a finite limit $I$ independent of the choice of the points $x'_k$ and the partition of the interval $[a, b]$. This limit is called the Stieltjes integral of the function $g(x)$ with respect to the function $F(x)$. We denote the integral as*

$$I = \lim_{\substack{n \to \infty \\ \Delta x \to 0}} S = \int_a^b g(x) dF(x). \tag{1.14}$$

**Remark 1.6.5** *The Stieltjes integral is a generalization of the Riemann integral, since for $F(x) = x$, (1.14) represents the Riemann integral.*

**Remark 1.6.6** *When the interval of integration is infinite, we define the improper Stieltjes integral as the limit of a sequence of proper Stieltjes integrals. Thus, if*

$$\lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b g(x)dF(x)$$

*exists as $a$ and $b$ tend to $-\infty$ and $+\infty$, respectively, this limit is called the improper Stieltjes integral of the function $g(x)$ with respect to the function $F(x)$.*

**Proposition 1.6.7**

1. *If $c$ and $l$ are constants, then*

$$\int_a^b cg(x)d[lF(x)] = cl\int_a^b g(x)dF(x).$$

2. *If the integrals on the RHS exists, then the integrals on the LHS exist and*

$$\int_a^b [g_1(x) + g_2(x)]dF(x) = \int_a^b g_1(x)dF(x) + \int_a^b g_2(x)dF(x),$$

$$\int_a^b g(x)d[F_1(x) + F_2(x)] = \int_a^b g(x)dF_1(x) + \int_a^b g(x)dF_2(x).$$

*This is satisfied for an arbitrary finite number of functions $g_i(x)$ and $F_i(x)$.*

3. *If $a < b < c$ and all three integrals*

$$\int_a^b g(x)dF(x), \int_a^c g(x)dF(x), \int_c^b g(x)dF(x)$$

*exist, the equation*

$$\int_a^b g(x)dF(x) = \int_a^c g(x)dF(x) + \int_c^b g(x)dF(x)$$

*holds. This is satisfied for an arbitrary finite number of points $a < c_1 < c_2 < \cdots < c_n < b$.*

**Remark 1.6.8** *In the theory of real functions it is proved that if $g(x)$ is continuous and bounded over the real axis and $F(x)$ is a function of bounded variation, both proper and improper Stieltjes integrals exist. However, Stieltjes integral may not exist when $g(x)$ is not bounded.*

**Corollary 1.6.9** *If $F(x)$ is the distribution function of a random variable of the continuous type with the density function $f(x)$,*

$$\int_a^b g(x)dF(x) = \int_a^b g(x)f(x)dx, \tag{1.15}$$

*which reduces to the Riemann integral.*

Suppose that $F(x)$ is the distribution function of a random variable of the discrete type with jump points $x'_k$ and jumps $p_k (k = 1, 2, \cdots)$. Then $F(x)$ has the form

$$F(x) = \sum_{x'_k < x} [F(x'_k + 0) - F(x'_k)].$$

By (1.13) and (1.14), we obtain

$$\int_a^b g(x)dF(x) = \sum_k g(x'_k)p_k. \tag{1.16}$$

**Proposition 1.6.10** *If the expected value of $Y = g(X)$ exists, then*

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)dF(x), \tag{1.17}$$

*where $F(x)$ denotes the distribution function of $X$.*

**Example 1.6.11** *Setting $g(x) = x^r$, we obtain the general expression for the moment of the $r$th order*

$$m_r = E(x^r) = \int_{-\infty}^{+\infty} x^r dF(x).$$

**Example 1.6.12** *$g(x) = e^{itx}$, the characteristic function*

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x).$$

**Proposition 1.6.13** *Let $F_X(x), F_Y(y), F_Z(z)$ be distribution functions of random variables $X, Y, Z$, then*

1. *If $Z = X + Y$,*

$$F_Z(z) = \int_{-\infty}^{+\infty} F_Y(z-x)dF_X(x) = \int_{-\infty}^{+\infty} F_X(z-y)dF_Y(y). \tag{1.18}$$

2. *If $Z = X - Y$,*

$$F_Z(z) = \int_{-\infty}^{+\infty} F_X(z+y)dF_Y(y). \tag{1.19}$$

3. *If $Z = XY$ and $P(X = 0) = P(Y = 0) = 0$,*

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{0} \left[1 - F_Y\left(\frac{z}{x}\right)\right] dF_X(x) + \int_{0}^{+\infty} F_Y\left(\frac{z}{x}\right) dF_X(x) \\ &= \int_{-\infty}^{0} \left[1 - F_X\left(\frac{z}{y}\right)\right] dF_Y(y) + \int_{0}^{+\infty} F_X\left(\frac{z}{y}\right) dF_Y(y). \end{aligned} \tag{1.20}$$

4. *If $Z = \frac{X}{Y}$ and $P(Y = 0) = 0$,*

$$F_Z(z) = \int_{-\infty}^{0} \left[1 - F_X(zy)\right] dF_Y(y) + \int_{0}^{+\infty} F_X(zy) dF_Y(y). \tag{1.21}$$

## 1.7 The Lévy-Cramér Theorem

### 1.7.1 Lemmas Before the Lévy-Cramér Theorem

非负渐升列积分定理，单调有界必收敛

**Theorem 1.7.1** ***(Beppo Lévy Monotone Convergence Theorem)**. Let $\{f_n\}_n$ be a sequence of measurable, non-negative functions on a measurable set $E$. If $\{f_n\}$ is monotonically increasing a.e. on $E$*

$$f_1 \leq f_2 \leq \cdots \leq f_k(x) \leq \cdots$$

*satisfying*

$$\lim_{k\to\infty} f_k(x) = f(x), \ \text{a.e. } x \in E, \quad \text{and} \quad \int_E f(x)dx < \infty,$$

*then*

$$\lim_{k\to\infty} \int_E f_k(x)dx = \int_E f(x)dx.$$

*This means that the order of integral and limit can be interchanged.*

**Proof.** Since $f_k \le f_{k+1} \le f$, we have

$$\int_E f_k dx \le \int_E f_{k+1} dx \le \int_E f dx < \infty.$$

Thus $\lim\limits_{k\to\infty} \int_E f_k dx$ is well defined such that

$$\lim_{k\to\infty} \int_E f_k dx \le \int_E f dx.$$

■

To prove the opposite inequality, we need more background and knowledge.

**Definition 1.7.2** *A measurable function $f : E \to \mathbb{R}^n$ is simple (or a simple function) if the size of the range $|f(E)|$ is finite and assumes only a finite number of values.*

In particular, $f(x) = \sum_{i=1}^p c_i \chi_{A_i}(x)$, $\bigcup\limits_{i=1}^p A_i = \mathbb{R}^n$, $A_i \bigcap\limits_{i\ne j} A_j = \phi$. Thus any simple function is a linear combination of finitely many indicator functions. Moreover, the integral can be defined as

$$\int_E f(x)dx = \sum_{i=1}^p c_i m(E \cap A_i),$$

where $m(\cdot)$ is the Lebesgue measure.

**Lemma 1.7.3** *Let $\{E_k \subset \mathbb{R}^n\}$ be a sequence of increasing measurable sets. Let $f(x)$ be a non-negative measurable simple function over $\mathbb{R}^n$. Then*

$$\int_E f(x)dx = \lim_{k\to\infty} \int_{E_k} f(x)dx, \quad E = \bigcup_{k=1}^\infty E_k.$$

Proof. Let $f(x)$ takes on values of $c_i (i = 1, \cdots, p)$ on sets $A_i (i = 1, \cdots, p)$. Then

$$\begin{aligned}
\lim_{k\to\infty} \int_{E_k} f(x)dx &= \lim_{k\to\infty} \sum_{i=1}^p c_i m(E_k \cap A_i) \\
&= \sum_{i=1}^p c_i \lim_{k\to\infty} m(E_k \cap A_i) \\
&= \sum_{i=1}^p c_i m(E \cap A_i) \\
&= \int_E f(x)dx.
\end{aligned}$$

**Definition 1.7.4** *Let $f(x)$ be a non-negative measurable function on $E \subset \mathbb{R}^n$. Define*

$$\int_E f(x)dx = \sup_{\substack{h(x)\le f(x)\\ x\in E}} \left\{ \int_E h(x) : h(x) \text{ is a non-negative measurable simple function over } \mathbb{R}^n \right\}.$$

*If $\int_E f(x)dx < +\infty$, $f(x)$ is said to be integrable over $E$.*

Now we continue to prove the Beppo Lévy Monotonic Convergence Theorem 1.7.1.

**Proof.** Let $0 < c < 1$ and $h(x)$ be a non-negative measurable simple function over $\mathbb{R}^n$ with $h(x) \le f(x)$ a.e. $x \in E$. Let $f_k \nearrow f$ and

$$E_k = \{x \in E : f_k(x) \ge ch(x)\}, (k = 1, 2, \cdots).$$

Then $E_k$ is an increasing measurable set such that

$$\lim_{k\to\infty} E_k = E.$$

Using lemma,

$$\lim_{k\to\infty} c\int_{E_k} h(x)dx = c\int_E h(x)dx.$$

Then

$$\int_E f_k(x)dx \geq \int_{E_k} f_k(x)dx \geq \int_{E_k} ch(x)dx = c\int_{E_k} h(x)dx.$$

$$\lim_{k\to\infty}\int_E f_k(x)dx \geq \lim_{k\to\infty} c\int_{E_k} h(x)dx = c\int_E h(x)dx.$$

Let $c \to 1$,

$$\lim_{k\to\infty}\int_E f_k(x)dx \geq \int_E h(x)dx.$$

Using definition of $\int_E f(x)dx$,

$$\lim_{k\to\infty}\int_E f_k(x)dx \geq \int_E f(x)dx.$$

■

**Theorem 1.7.5** ***(Fatou's Lemma)***. *Let $f_k(x)$ be non-negative measurable functions on $E \subset \mathbb{R}^n$, then*

$$\int_E \liminf_{k\to\infty} f_k(x)dx \leq \liminf_{k\to\infty}\int_E f_k(x)dx. \tag{1.22}$$

*In above formula, we can denote $\underline{\lim}_{k\to\infty}$ by $\liminf_{k\to\infty}$.*

**Proof.** Let $g_k(x) = \inf\{f_j(x) : j \geq k\}$, then

$$g_k(x) \leq g_{k+1}(x), (k = 1, 2, \cdots)$$

and

$$\liminf_{k\to\infty} f_k(x) = \lim_{k\to\infty} g_k(x).$$

$$\int_E \liminf_{k\to\infty} f_k(x)dx = \int_E \lim_{k\to\infty} g_k(x)dx = \lim_{k\to\infty}\int_E g_k(x)dx = \liminf_{k\to\infty}\int_E g_k(x)dx \leq \liminf_{k\to\infty}\int_E f_k(x)dx.$$

■

一般可测函数的积分

**Definition 1.7.6** *Let $f(x)$ be a measurable function on $E \subset \mathbb{R}^n$. If $\int_E f^+(x)dx < \infty$, $\int_E f^-(x)dx < \infty$, then $f(x)$ is said to be Lebesgue integrable over $E$ and*

$$\int_E f(x)dx = \int_E f^+(x)dx - \int_E f^-(x)dx.$$

*Here, $f^+(x) = \max\{f(x), 0\} \geq 0$, $f^-(x) = -\min\{f(x), 0\} \geq 0$. All integrable functions on $E$ are denoted by $L(E)$.*

**Theorem 1.7.7** ***(Lebesgue Dominated Convergence Theorem)***. *Let* $f_k \in L(E), (k = 1, 2, \cdots)$, *such that*

$$\lim_{k\to\infty} f_k(x) = f(x) \quad a.e. \quad x \in E.$$

*If there exists an integrable function* $F(x)$ *over* $E$ *such that,*

$$|f_k(x)| \le F(x) \quad a.e. \quad x \in E, \quad (k = 1, 2, \cdots).$$

*Then*

$$\lim_{k\to\infty} \int_E f_k(x)dx = \int_E f(x)dx. \tag{1.23}$$

*Here* $F(x)$ *is called the control function of* $\{f_k(x)\}$.

$F(x)$称为$\{f_k(x)\}$ 控制函数

**Proof.** $f(x)$ is measurable over $E$. Since $|f_k(x)| \le F(x)$ (a.e. $x \in E$), then $|f(x)| \le F(x)$ (a.e. $x \in E$). Thus $f(x)$ is integrable over $E$. Set

$$g_k(x) = |f_k(x) - f(x)|, \quad (k = 1, 2, \cdots)$$

then $g_k(x) \in L(E)$ and $0 \le g_k(x) \le 2F(x)$ a.e. $x \in E$, $(k = 1, 2, \cdots)$.
Using Fatou's Lemma,

$$\int_E \lim_{k\to\infty}(2F(x) - g_k(x))dx \le \liminf_{k\to\infty} \int_E (2F(x) - g_k(x))dx.$$

Since $F(x)$ and $\{g_k(x)\}$ are all integrable,

$$\int_E 2F(x)dx - \int_E \lim_{k\to\infty} g_k(x)dx \le \int_E 2F(x)dx - \limsup_{k\to\infty} \int_E g_k(x)dx.$$

Notice that $\lim\limits_{k\to\infty} g_k(x) = \lim\limits_{k\to\infty} |f_k(x) - f(x)| = 0$ a.e. $x \in E$. Thus,

$$0 \le \limsup_{k\to\infty} \int_E g_k(x)dx \le 0.$$

Last,

$$|\int_E f_k(x)dx - \int_E f(x)dx| \le \int_E |f_k(x) - f(x)|dx = \int_E g_k(x)dx \to 0 \quad \text{as} \quad k \to \infty.$$

■

**Remark 1.7.8** *In fact, from above we have stronger conclusion*

$$\lim_{k\to\infty} \int_E |f_k(x) - f(x)|dx = 0. \tag{1.24}$$

**Corollary 1.7.9** ***(Bounded Convergence Theorem)***. *Let* $\{f_k(x)\}$ *be a sequence of measurable functions.* $m(E) < +\infty$. *For a.e.* $x \in E$, $\lim\limits_{k\to\infty} f_k(x) = f(x)$, $|f_k(x)| < M$, $(k = 1, 2, \cdots)$. *Then* $f \in L(E)$ *and*

$$\lim_{k\to\infty} \int_E f_k(x) = \int_E f(x). \tag{1.25}$$

**Proof.** Set the bounded function $F(x) = M$ in Dominated Convergence Theorem. ■

### 1.7.2 Dominated Convergence Theorem in Measure Theory

When a.s. convergence implies $L^1$ convergence. Monotone convergence (MCT), Dominated convergence (DCT).

Let $X$ and $\{X_n\}$ be random variables on the same probability space $(\Omega, \mathcal{F}, P)$. If $X_n(\omega) \to X(\omega)$ for each $\omega \in \Omega$ (a.s. convergence), does it follow that $E[X_n] \to E[X]$? That is, may we exchange expectation and limits in th equation,

$$\lim_{n\to\infty} E[X_n] \xrightarrow{?} E[\lim_{n\to\infty} X_n]?$$

As we know, we cannot always interchange such order as seen from counter-example 1.1.27. However, when we have some resctriction for $X_n$, we can do interchange. We see the following corresponding theorems in probability theory.

**Theorem 1.7.10** ***(Monotone Convergence Theorem or MCT)***. *Let $X$ and $X_n \geq 0$ be random variables (not necessarily simple) for which $X_n(\omega) \nearrow X(\omega)$ for each $\omega \in \Omega$. Then*

$$\lim_{n\to\infty} E[X_n] = E[\lim_{n\to\infty} X_n] = EX.$$

*If $E|X| < \infty$, then also $E|X_n - X| \to 0$.*

**Theorem 1.7.11** ***(Fatou's Lemma)***. *Let $X_n \geq 0$ be random variables. Then*

$$E[\lim_{n\to\infty} \inf X_n] \leq \lim_{n\to\infty} \inf E[X_n].$$

**Theorem 1.7.12** ***(Dominate Convergence Theorem or DCT)***. *Let $X$ and $X_n$ be random variables (not necessarily simple or positive) for which $P(\lim_{n\to\infty} X_n = X) = 1$ (a.s.) [or $X_n \xrightarrow{P} X$ (in probability) (not for sure?)], and suppose that $P(|X_n| \leq Y) = 1$ a.s. for all $n$ and for some integrable random variable $Y$ with $EY < \infty$. Then*

$$\lim_{n\to\infty} E[X_n] = E[\lim_{n\to\infty} X_n] = EX.$$

*That is, above equality holds if $\{X_n\}$ is "dominated" by $Y \in L^1$. Moreover, $E|X_n - X| \to 0$.*

### 1.7.3 Main Theorem

Go back to Lévy-Cramér Theorem.

**Theorem 1.7.13** *If the sequence $\{F_n(x)\}(n = 1, 2, \cdots)$ of distribution functions is convergent to the distribution function $F(x)$, then the corresponding sequence of characteristic functions $\{\phi_n(t)\}$ converges at every point $t$ $(-\infty < t < \infty)$ to the function $\phi(t)$ which is the characteristic function of the limit distribution function $F(x)$, and the convergence to $\phi(t)$ is uniform with respect to $t$ in every finite interval on the $t$-axis.*

**Proof.**

$$\phi_n(t) = \int_{-\infty}^{+\infty} e^{itx} dF_n(x), \quad \phi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x).$$

Let $a < 0$ and $b > 0$ be continuity points of $F(x)$, we have

$$\begin{aligned}\phi_n(t) &= \int_{-\infty}^{a} e^{itx} dF_n(x) + \int_{a}^{b} e^{itx} dF_n(x) + \int_{b}^{+\infty} e^{itx} dF_n(x) \\ &= I_{n1} + I_{n2} + I_{n3}, \\ \phi(t) &= \int_{-\infty}^{a} e^{itx} dF(x) + \int_{a}^{b} e^{itx} dF(x) + \int_{b}^{+\infty} e^{itx} dF(x) \\ &= I_1 + I_2 + I_3.\end{aligned} \tag{1.26}$$

Consider the difference

$$I_{n2} - I_2 = \int_a^b e^{itx} dF_n(x) - \int_a^b e^{itx} dF(x).$$

Integrating by parts, we obtain

$$I_{n2} - I_2 = e^{itx} \left\{ [F_n(x)]_a^b - [F(x)]_a^b \right\} - it \int_a^b [F_n(x) - F(x)] \, e^{itx} dx.$$

Hence,

$$|I_{n2} - I_2| \leq |F_n(b) - F(b)| + |F_n(a) - F(a)| + |t| \int_a^b |F_n(x) - F(x)| dx.$$

For $\forall \epsilon > 0$, since $a$ and $b$ are continuity points of $F(x)$, we have

$$|F_n(b) - F(b)| < \frac{\epsilon}{9}, \quad |F_n(a) - F(a)| < \frac{\epsilon}{9}.$$

Using Lebesgue Dominated Convergence Theorem, $|F_n(x) - F(x)| \leq 2$ is uniformly bounded in every interval, we have

$$\lim_{n\to\infty} \int_a^b |F_n(x) - F(x)| dx = \int_a^b \lim_{n\to\infty} |F_n(x) - F(x)| dx = 0.$$

For some fixed $t$ satisfying $T_1 < t < T_2$, set $K(t) = \max(|T_1|, |T_2|)$. Then for sufficiently large $n$ and all $t$, we have

$$|t| \int_a^b |F_n(x) - F(x)| dx \leq K(t) \int_a^b |F_n(x) - F(x)| dx \leq \frac{\epsilon}{9}. \tag{1.27}$$

Thus we obtain

$$|I_{n2} - I_2| < \frac{\epsilon}{3}. \tag{1.28}$$

Now consider the difference

$$I_{n1} - I_1 = \int_{-\infty}^{a} e^{itx} dF_n(x) - \int_{-\infty}^{a} e^{itx} dF(x).$$

We have

$$|I_{n1} - I_1| \leq \int_{-\infty}^{a} dF_n(x) + \int_{-\infty}^{a} dF(x) = F_n(a) + F(a).$$

Thus if $a$ is sufficiently large in absolute value, and $a$ is the continuity of $F(x)$,

$$F(a) < \frac{\epsilon}{6}, \quad F_n(a) \sim F(a) < \frac{\epsilon}{6}.$$

Thus

$$|I_{n1} - I_1| < \frac{\epsilon}{3}$$

for all $t$ and sufficiently large $n$. Similarly, $|I_{n3} - I_3| < \frac{\epsilon}{3}$. ■

**Theorem 1.7.14** ***(Helly's Selection Theorem)***. *For every sequence $\{F_n(x)\}$ of distribution functions, there exists a subsequence $\{F_{n_k}(x)\}$ and a non-decreasing right continuous function $F_\infty$ such that $F_{n_k}(x) \to F_\infty(x)$ as $k \to \infty$ at all continuity points $x$ of $F_\infty$, that is $F_{n_k}(x) \stackrel{w}{\to} F_\infty(x)$.* ***Moreover, $F_\infty$ is a distribution function if and only if $\{F_n\}$ is tight. (see Slutsky Theorem Apr5.pdf)***

**Theorem 1.7.15** *If the sequence of characteristic functions $\{\phi_n(t)\}$ converges at every point $t$ $(-\infty < t < \infty)$ to a function $\phi(t)$ continuous in some interval $|t| < \tau$, then the sequence $\{F_n(x)\}$ converges to $F(x)$ corresponding to $\phi(t)$.*

**Proof.** The proof has two parts. **Part 1**.
First, we select a subsequence by Helly's Selection Theorem, $F_{n_k} \to F$ as $k \to \infty$. It does not, however, follow from the Helly's theorem that $F(x)$ is a distribution function. We have $0 \le F(x) \le 1$, but we do not know if $F(-\infty) = 0$ and $F(\infty) = 1$. We now prove them.
Suppose that

$$\alpha = F(+\infty) - F(-\infty) < 1. \tag{1.29}$$

Since $\phi_n(t) \to \phi(t)$ and $\phi_n(0) = 1$, we have $\phi(0) = 1$. By the assumption that the function $\phi(t)$ is continuous, it follows that in some neighborhood of the origin $t = 0$ it will differ little from 1. Thus for sufficiently small $\tau$ we have the inequality

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} \phi(t)dt\right| > 1 - \frac{\epsilon}{2} > \alpha + \frac{\epsilon}{2}, \tag{1.30}$$

where the number $\epsilon$ is chosen in such a way that $\alpha + \epsilon < 1$. Since the subsequence $F_{n_k}(x)$ are CDFs, we know

$$F_{n_k}(+\infty) - F_{n_k}(-\infty) = 1.$$

We can choose sufficiently large $a > \frac{4}{\epsilon\tau}$ such that $a$ and $-a$ are continuity points of $F(x)$, and a number $K$ such that for $k > K$,

$$k \to \infty, F_{n_k}(x) \to F(x) \quad \Rightarrow \quad F_{n_k}(a) - F(a) < \frac{\epsilon}{16}, \quad F(-a) - F_{n_k}(-a) < \frac{\epsilon}{16}, (k \to \infty).$$

$$F(+\infty) \text{ exists} \quad \Rightarrow \quad F(a) - F(+\infty) < \frac{\epsilon}{16}, \quad F(-\infty) - F(-a) < \frac{\epsilon}{16}, (k \to \infty).$$

$\Rightarrow$

$$\underbrace{F_{n_k}(a) - F_{n_k}(-a)}_{:=\alpha_k} - F(+\infty) + F(-\infty) < \frac{\epsilon}{4}.$$

Then

$$\alpha_k := F_{n_k}(a) - F_{n_k}(-a) < F(+\infty) - F(-\infty) + \frac{\epsilon}{4} = \alpha + \frac{\epsilon}{4}.$$

On the other hand, since $\phi_n(t) \to \phi(t)$ in $|t| < \tau$, it follows from (1.30) that there exists sufficiently large $k$,

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} \phi_{n_k}(t)dt\right| > \alpha + \frac{\epsilon}{2}. \tag{1.31}$$

We now show that this inequality is not satisfied. Indeed, we have

$$\int_{-\tau}^{\tau} \phi_{n_k}(t)dt = \int_{-\tau}^{\tau}\left[\int_{-\infty}^{+\infty} e^{itx} dF_{n_k}(x)\right] dt \overset{\text{Fubini since } |\iint| < 2\tau}{=} \int_{-\infty}^{+\infty}\left[\int_{-\tau}^{\tau} e^{itx}dt\right] dF_{n_k}(x).$$

Since

$$\begin{aligned}
\left|\int_{-\tau}^{\tau} e^{itx}dt\right| &= \left|\left[\frac{e^{itx}}{ix}\right]\Big|_{t=-\tau}^{\tau}\right| \\
&= \frac{|(\cos\tau x + i\sin\tau x) - (\cos\tau x - i\sin\tau x)|}{|x|} \\
&= \frac{|2\sin\tau x|}{|x|} \\
&\leq \frac{2}{|x|} \\
&< \frac{2}{a}, \quad \text{for } |x| > a.
\end{aligned}$$

Then

$$\begin{aligned}
&\left|\int_{-\infty}^{+\infty}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x)\right| \\
\leq&\left|\int_{|x|\leq a}\left[\int_{-\tau}^{+\tau} e^{itx}dt\right]dF_{n_k}(x)\right| + \left|\int_{|x|>a}\left[\int_{-\tau}^{+\tau} e^{itx}dt\right]dF_{n_k}(x)\right| \\
\leq&\left|\int_{|x|\leq a} 2\tau dF_{n_k}(x)\right| + \left|\int_{|x|>a}\frac{2}{a}dF_{n_k}(x)\right| \\
=&2\tau\alpha_k + \frac{2}{a}.
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\phi_{n_k}(t)dt\right| &\leq \frac{1}{2\tau}\left|\int_{-\infty}^{+\infty}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x)\right| \\
&\leq \frac{1}{2\tau}(2\tau\alpha_k + \frac{2}{a}) = \alpha_k + \frac{1}{a\tau} \leq \alpha_k + \frac{\epsilon}{4} \\
&< \alpha + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \alpha + \frac{\epsilon}{2}.
\end{aligned}$$

This is a contradiction to equation (1.31). Hence $F(x)$ is a distribution function. From Theorem 1.7.13, it follows that $\phi(t)$ is its characteristic function.

**Part 2**.

We last prove that not only $\{F_{n_k}(x)\}$, but the whole $\{F_n(x)\}$ converges to $F(x)$. If this were not so there would be another subsequence $\{\tilde{F}_{n_k}(x)\}$ convergent to a limit function $\tilde{F}(x) \neq F(x)$. One can show $\tilde{F}(x)$ is also a distribution function and moreover, $\tilde{F}(x)$ has the same characteristic function as $F(x)$. Hence by Lévy Theorem **??**, $\tilde{F}(x) \equiv F(x)$. Thus $\{F_n(x)\} \to F(x)$. ■

**Theorem 1.7.16** ***(Lévy-Cramér).*** *The conclusions from both Theorem 1.7.15 and Theorem 1.7.13. The convergence of CDFs ⇔ the convergence of Ch.f.s.*

**Remark 1.7.17** *Theorem 1.7.16 remains true if we assume the continuity of the limit function $\phi(t)$ only at the point $t = 0$.*

**Remark 1.7.18** *In general, we cannot replace the convergence at every point $t$ in $(-\infty, +\infty)$ by convergence in some interval on the $t$-axis containing the origin. (In order to determine the distribution, we need information of $\phi(t)$ for $-\infty < t < \infty$. If not,* ***Part 2*** *proof may fail.)*

## 1.8 The de Moivre-Laplace Theorem

### 1.8.1 Part A

$\{X_n\}$ is a sequence of binomial random variables. For every $n$, $X_n$ takes on the values $0, 1, \dots, n$,

$$P(X_n = r) = C_n^r p^r q^{n-r},$$

where $0 < p < 1$ and $q = 1 - p$. The moments are

$$EX_n = np, \quad Var(X_n) = npq.$$

Consider $\{Y_n\}$ of standardized random variables

$$Y_n = \frac{X_n - np}{\sqrt{npq}}.$$

**Theorem 1.8.1** *(de Moivre-Laplace theorem). Let $\{F_n(y)\}$ be the sequence of distribution functions of $Y_n$ above. If $0 < p < 1$, then for every $y$ we have the relation*

$$\lim_{n\to\infty} F_n(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{s^2}{2}} ds.$$

**Proof.** The Ch.f. of $X_n$ is

$$\phi_X(t) = Ee^{itX_n} = \sum_r e^{itr} C_n^r p^r q^{n-r} = (q + pe^{it})^n.$$

Notice that if $Y = aX + b$, then $\phi_Y(t) = e^{ibt}\phi_X(at)$. Then the Ch.f. of $Y_n$ is

$$\begin{aligned}
\phi_X(t) &= \exp\left(-\frac{np}{\sqrt{npq}} it\right)\left(q + pe^{\frac{it}{\sqrt{npq}}}\right)^n \\
&= \left(q\exp(-\frac{pit}{\sqrt{npq}}) + p\exp(\frac{qit}{\sqrt{npq}})\right)^n.
\end{aligned}$$

Let us expand $e^{iz}$ in the neighborhood of $z = 0$ with the Peano remainder,

$$\begin{aligned}
p\exp(\frac{qit}{\sqrt{npq}}) &= p\left(1 + \frac{qit}{\sqrt{npq}} - \frac{1}{2}\frac{q^2t^2}{npq} + o(\frac{t^2}{n})\right) \\
&= p + it\sqrt{\frac{pq}{n}} - \frac{1}{2}\frac{qt^2}{n} + o(\frac{t^2}{n}).
\end{aligned}$$

$$\begin{aligned}
q\exp(-\frac{pit}{\sqrt{npq}}) &= q\left(1 - \frac{pit}{\sqrt{npq}} - \frac{1}{2}\frac{p^2t^2}{npq} + o(\frac{t^2}{n})\right) \\
&= q - it\sqrt{\frac{pq}{n}} - \frac{1}{2}\frac{pt^2}{n} + o(\frac{t^2}{n}),
\end{aligned}$$

where

$$\lim_{n\to\infty} n \cdot o(\frac{t^2}{n}) = \lim_{n\to\infty} o(1) = 0.$$

Then the Ch.f.

$$\begin{aligned}
\phi_Y(t) &= (p + q - \frac{1}{2}\frac{(p+q)t^2}{n} + o(\frac{t^2}{n}))^n = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n \\
&\to e^{-\frac{t^2}{2}}.
\end{aligned}$$

The Ch.f. converges to that of a standard normal distribution. By Lévy-Cramér theorem, we have the convergence of the distribution function for every $y$, since the limit Gaussian distribution function has no discontinuity points. ■

We see that

$$\begin{aligned}\lim_{n\to\infty} P(y_1 &< Y_n < y_2) = \lim_{n\to\infty}[F_n(y_2) - F_n(y_1)] = \frac{1}{\sqrt{2\pi}}\int_{y_1}^{y_2} e^{-\frac{s^2}{2}}\,ds \\ &= \lim_{n\to\infty} P(y_1 < \frac{X_n - np}{\sqrt{npq}} < y_2) \\ &= \lim_{n\to\infty} P(y_1\sqrt{npq} + np < X_n < y_2\sqrt{npq} + np) \\ &:= \lim_{n\to\infty} P(x_1 < X_n < x_2).\end{aligned}$$

We say that $X_n$ has an **asymptotically normal distribution** $N(np, npq)$.

Replacing $y_1$ and $y_2$ with

$$y_1 + \frac{1}{2\sqrt{npq}} \text{ and } y_2 - \frac{1}{2\sqrt{npq}},$$

we get a somewhat better approximation.

**Example 1.8.2** *We throw a coin $n = 100$ times. $1 =$head and $0 =$tail. Let $p = q = 0.5$. What is the probability that heads will appear more than 50 times and less than 60 times?*
*Solution. We have*

$$EX_n = np = 50, \quad Var(X_n) = npq = 25.$$

*Then*

$$\begin{aligned}P(50 &< X_n < 60) = P(\frac{50-50}{\sqrt{25}} < \frac{X_n - 50}{\sqrt{25}} < \frac{60-50}{\sqrt{25}}) \\ &= P(0 < Y_n < 2) \cong P(0 + \frac{1}{2\sqrt{25}} < Y_n < 2 - \frac{1}{2\sqrt{25}}) \\ &= P(0.1 < Y_n < 1.9) = \frac{1}{\sqrt{2\pi}}\int_{0.1}^{1.9} e^{-\frac{s^2}{2}}\,ds \approx 0.4315.\end{aligned}$$

### 1.8.2 Part B

From the de Moivre-Laplace limit theorem, we obtain an analogous theorem for

$$U_n = \frac{X_n}{n},$$

with $EU_n = p, Var(U_n) = \frac{pq}{n}$. We have

$$Z_n = \frac{U_n - p}{\sqrt{\frac{pq}{n}}} = \frac{X_n - np}{\sqrt{npq}} = Y_n.$$

Then

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{s^2}{2}}\,ds.$$

For $z_1 < z_2$, we have.

$$\begin{aligned}\lim_{n\to\infty} P(z_1 &< Z_n < z_2) = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{s^2}{2}}\,ds \\ &= \lim_{n\to\infty} P(z_1 < \frac{U_n - p}{\sqrt{\frac{pq}{n}}} < z_2) \\ &= \lim_{n\to\infty} P(u_1 = z_1\sqrt{\frac{pq}{n}} + p < U_n < z_2\sqrt{\frac{pq}{n}} + p = u_2).\end{aligned}$$

We say that $U_n$ has an **asymptotically normal distribution** $N(p, \frac{pq}{n})$.

**Example 1.8.3** *IBM cards correspond to the workers. Of the workers 20% are minors and 80% adults. Before choosing the next card, we always return the first one to the box, so that the probability of selecting the card corresponding to a minor remains 0.2. We observe n cards in this manner. What value should n have in order that the probability will be 0.95 that the frequency of cards corresponding to minors lies between 0.18 and 0.22?*

*Denote the frequency of the appearance of the card corresponding to a minor by $U_n$. We then have*

$$EU_n = 0.2, \quad Var(U_n) = \frac{pq}{n} = \frac{0.16}{n}.$$

*Consider the probability*

$$\begin{aligned} P(0.18 &< U_n < 0.22) = P(\frac{-0.02}{\frac{0.4}{\sqrt{n}}} < \frac{U_n - 0.2}{\frac{0.4}{\sqrt{n}}} < \frac{0.02}{\frac{0.4}{\sqrt{n}}}) \\ &= P(-0.05\sqrt{n} < \frac{U_n - 0.2}{0.4}\sqrt{n} < 0.05\sqrt{n}) \cong 0.95. \end{aligned}$$

*From tables of the normal distribution we obtain* $0.05\sqrt{n} \cong 1.96$; *consequently* $n \cong 1537$.

## 1.9 The Lindeberg-Lévy Theorem

### 1.9.1 Part A

The de Moivre-Laplace theorem is, as we shal see later, a particular case of a more general limit theorem, namely, the Lindeberg-Lévy Theorem.

Consider a sequence $\{X_n\}$ of i.i.d. random variables whose second order moment exists. For every $k$,

$$EX_k = m, \quad Var(X_k) = \sigma^2.$$

Consider

$$Y_n = X_1 + \ldots + X_n$$

We have

$$EY_n = nm, \quad Var(Y_n) = n\sigma^2.$$

Let

$$Z_n = \frac{Y_n - nm}{\sqrt{n}\sigma}.$$

Then we have the following Central Limit Theorem.

**Theorem 1.9.1** *(Lindeberg-Lévy Theorem). If $X_1, X_2, \ldots$ are independent, identically distributed random variables, whose standard deviation $\sigma \neq 0$ exists, then the distribution functions $\{F_n(z)\}$ of $Z_n = \frac{Y_n - nm}{\sqrt{n}\sigma}$, satisfies, for every $z$, the equality*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{s^2}{2}} ds.$$

**Proof.** Let us write

$$Z_n = \frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n}(X_k - m).$$

All $X_k - m$ are i.i.d., hence the same Ch.f. $\phi_X(t)$. Thus $\frac{X_k - m}{\sqrt{n}\sigma}$ has Ch.f. $\phi_X(\frac{t}{\sqrt{n}\sigma})$ and $\phi_Z(t)$ of $Z_n$ is

$$\phi_Z(t) = \left[\phi_X\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n.$$

Since we have

$$E(X_k - m) = 0, \quad Var(X_k - m) = \sigma^2,$$

we can expand $\phi_X(t)$ in a neighborhood of $t = 0$ according to the MacLaurin formula:

$$\phi_X(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2).$$

Then

$$\begin{aligned}\phi_Z(t) &= \left[1 - \frac{1}{2}\sigma^2\left(\frac{t}{\sqrt{n}\sigma}\right)^2 + o(\frac{t^2}{n})\right]^n \\ &= \left[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right]^n \\ &\longrightarrow e^{-\frac{t^2}{2}},\end{aligned}$$

which is the Ch.f. of the standard normal distribution. By Lévy-Cramér theorem (Ch.f. is a one-to-one map to CDF), we prove the Lindeberg-Lévy Theorem. ■

### 1.9.2 Part B

For $z_1 < z_2$, we have

$$\begin{aligned}\lim_{n\to\infty} P(z_1 &< Z_n < z_2) = \lim_{n\to\infty}[F_n(z_2) - F_n(z_1)] = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{s^2}{2}}ds \\ &= \lim_{n\to\infty} P(z_1 < \frac{Y_n - nm}{\sqrt{n}\sigma} < z_2) \\ &= \lim_{n\to\infty} P(z_1\sqrt{n}\sigma + nm < Y_n < z_2\sqrt{n}\sigma + nm).\end{aligned}$$

Let $y_1 = z_1\sqrt{n}\sigma + nm, y_2 = z_2\sqrt{n}\sigma + nm$. We say that $Y_n$ has an **asymptotically normal distribution** $N(nm, \sigma^2 n)$.

When a sum of random variables has an asymptotically normal distribution, we say that it satisfies the central limit theorem (CLT).

**Example 1.9.2** *$X_n$ are independent and each of them has the Poisson distribution given by*

$$P(X_n = r) = \frac{2^r}{r!}e^{-2} \quad (r = 0, 1, 2, \cdots) \quad \lambda = 2.$$

*Find the probability that the sum $Y_{100} = X_1 + \cdots + X_{100}$ is greater than 190 and less than 210.*
*Solution. Notice that*

$$EY_n = 100\lambda = 200, \quad Var(Y_n) = 100\lambda = 200.$$

*Then $Y_{100} \sim N(200, 200)$. Thus*

$$P(190 < Y_{100} < 210) = P(\frac{-10}{10\sqrt{2}} < \frac{Y_{100} - 200}{10\sqrt{2}} < \frac{10}{10\sqrt{2}}) \cong 0.52.$$

### 1.9.3 Part C

From the above Lindeberg-Lévy Theorem, we have the following analogous theorem.

**Theorem 1.9.3** *Suppose that $X_1, X_2, \cdots$ are i.i.d. with standard deviation $\sigma \neq 0$. Let $U_n$ be*

$$U_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

*Let $F_n(v)$ be the distribution function of*

$$V_n = \frac{U_n - EU_n}{\sqrt{Var(U_n)}}.$$

*Then*

$$\lim_{n\to\infty} F_n(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{-\frac{s^2}{2}} ds.$$

**Proof.** We see that $EU_n = m, Var(U_n) = \frac{\sigma^2}{n}$. We have

$$V_n = \frac{U_n - m}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}} = Z_n.$$

For $z_1 < z_2$, we have.

$$\begin{aligned}
\lim_{n\to\infty} P(v_1 &< V_n < v_2) = \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} e^{-\frac{s^2}{2}} ds \\
&= \lim_{n\to\infty} P(v_1 < \frac{U_n - m}{\sqrt{\frac{\sigma^2}{n}}} < v_2) \\
&= \lim_{n\to\infty} P(v_1 \frac{\sigma}{\sqrt{n}} + m < U_n < v_2 \frac{\sigma}{\sqrt{n}} + m) \\
&:= \lim_{n\to\infty} P\left(u_1 < U_n < u_2\right).
\end{aligned}$$

We say that $U_n$ has an **asymptotically normal distribution** $N(m, \frac{\sigma^2}{n})$. ■

**Example 1.9.4** *Let $\{X_n\}$ be i.i.d. with uniform distribution with pdf*

$$f(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & else. \end{cases}$$

*We know that $m = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$. Consider*

$$Y_n = \frac{X_1 + \cdots + X_n}{n}.$$

*For $n = 48$, compute the probability that $Y_n$ is less than 0.4.*

$$\begin{aligned}
P(Y_n &< 0.4) = P(\frac{Y_n - \frac{1}{2}}{\sqrt{\frac{1}{12}}/\sqrt{48}} < \frac{0.4 - \frac{1}{2}}{\sqrt{\frac{1}{12}}/\sqrt{48}}) \\
&= P(\xi < -2.4) = 0.0082.
\end{aligned}$$

*As we see, although the random variable $X_k$ $(k = 1, 2, \cdots)$ have a uniform distribution in the interval $[0,1]$, their arithmetic mean has, for large $n$, approximately a distribution in which values that are less than $m = 0.5$ by more than 0.1 appear extremely rarely.*

**Example 1.9.5** *Let* $\{X_n\}$ *be i.i.d. Each of them can take on the values* $k = 0, 1, 2, \dots, 9$ *with* $P(X_n = k) = 0.1$ *for ever* $k$*. Then*

$$\begin{aligned} m &= 4.5, \\ \sigma^2 &= EX_n^2 - (EX_n)^2 = \frac{1}{10}\sum_{k=0}^{9} k^2 - m^2 = 28.5 - 20.25 = 8.25, \\ \sigma &= 2.87. \end{aligned}$$

*Consider*

$$Y_{100} = \frac{X_1 + \cdots + X_{100}}{100}.$$

*What is the probability that* $Y_{100}$ *will exceed 5?*

$$\begin{aligned} P(Y_{100} &> 5) = P(\frac{Y_{100} - 4.5}{\frac{2.87}{\sqrt{100}}} > \frac{5 - 4.5}{\frac{2.87}{\sqrt{100}}}) = P(\xi > 1.74) \\ &\approx 0.041. \end{aligned}$$

### 1.9.4 Part D

If their moment of the second order does not exist, CLT may not be satisfied.

**Example 1.9.6** *Let* $\{X_k\}$ *be i.i.d. Cauchy distribution with pdf*

$$f(y) = \frac{1}{\pi}\frac{1}{1+y^2}.$$

*The Ch.f. function*

$$\phi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{ity}\frac{1}{1+y^2}dy.$$

*To fund* $\phi(t)$ *consider first the pdf density*

$$f_1(y) = \frac{1}{2}e^{-|y|}.$$

*The reader may verify that the above expression is a density. The Ch.f. of the random variable with the density is*

$$\begin{aligned} \phi_1(t) &= \frac{1}{2}\int_{-\infty}^{\infty} e^{ity}e^{-|y|}dy = \frac{1}{2}\int_{-\infty}^{\infty}(\cos ty + i\sin ty)e^{-|y|}dy \\ &= \int_0^{\infty}\cos(ty)e^{-y}dy. \end{aligned}$$

*Integrating by parts, we obtain*

$$\begin{aligned} \int_0^{\infty}\cos(ty)e^{-y}dy &= -\cos(ty)e^{-y}|_{y=0}^{\infty} - t\int_0^{\infty}\sin(ty)e^{-y}dy \\ &= 1 - t\int_0^{\infty}\sin(ty)e^{-y}dy. \end{aligned}$$

*Similarly,*

$$\begin{aligned} \int_0^{\infty}\sin(ty)e^{-y}dy &= -\sin(ty)e^{-y}|_{y=0}^{\infty} + t\int_0^{\infty}\cos(ty)e^{-y}dy \\ &= t\int_0^{\infty}\cos(ty)e^{-y}dy. \end{aligned}$$

*Hence we obtain*

$$\int_0^\infty \cos(ty)e^{-y}dy = 1 - t^2 \int_0^\infty \cos(ty)e^{-y}dy.$$

*Finally, we obtain*

$$\phi_1(t) = \int_0^\infty \cos(ty)e^{-y}dy = \frac{1}{1+t^2}.$$

*The Ch.f. is absolutely integrable over the interval* $(-\infty, \infty)$*. Hence by (??) its corresponding density is*

$$f_1(y) = \frac{1}{2}e^{-|y|} = \frac{1}{2\pi}\int_{-\infty}^\infty \frac{e^{-ity}}{1+t^2}dt.$$

*Thus we obtain*

$$e^{-|y|} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{e^{-ity}}{1+t^2}dt.$$

*Changing* $e^{-ity}$ *into* $e^{ity}$ *under the integral sign and changing the roles of* $t$ *and* $y$*, we obtain*

$$e^{-|t|} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{e^{ity}}{1+y^2}dy.$$

*Thus we obtain*

$$\phi(t) = \frac{1}{\pi}\int_{-\infty}^\infty e^{ity}\frac{1}{1+y^2}dy = e^{-|t|}.$$

*Then the Ch.f. of*

$$Y_n = \frac{X_1 + \cdots + X_n}{n},$$

*is*

$$\phi_{Y_n}(t) = \left(e^{-\frac{|t|}{n}}\right)^n = e^{-|t|}.$$

$Y_n$ *also has the Cauchy distriubtion for arbitrary* $n$*, which does not have an asymptotic normal distribution. (Note that Cauchy distribution does not have a standard deviation.)*

### 1.9.5 Part E

Let the random variables $X_k$ $(k = 1, 2, \ldots)$ satisfy the assumptions of Lindeberg-Lévy Theorem and let $EX_k = 0$. Consider for every $n$ the partial sums

$$S_j = \sum_{k=1}^j X_k \quad (j = 1, 2, \ldots, n).$$

Erdös and Kac [1,2] have found the limit distributions for the sequences of random variables

$$\left\{\max_{1\le j\le n}\frac{S_j}{\sqrt{n}}\right\}, \quad \left\{\max_{1\le j\le n}\frac{|S_j|}{\sqrt{n}}\right\}, \quad \left\{\frac{1}{n^2}\sum_{j=1}^n S_j^2\right\}, \quad \left\{\frac{1}{n^{3/2}}\sum_{j=1}^n |S_j|\right\}.$$

These papers began a seris of fruitful investigations concerning the limit distributions of a large class of functionals definds on the vectors $(S_1, \ldots, S_n)$,even with much more general assumptions concerning the random variables $X_k$ than those considered here. We shall not discuss these results. The reader can find them in the papers of Erdös and Kac [1,2], Donsker [1], Prohorov, Skorohod, Spitzer, Baxter and Donsker, Varadarjan, Lamperti, Bartoszynski, and Billingsley.

### 1.9.6 Part F. Substitution of Sample Variance

For the CLT, we know that $Z_n = \sqrt{n}(\overline{X} - \mu)/\sigma$ is approximately $N(0,1)$. However, we rarely know $\sigma$. We may estimate $\sigma^2$ by unbiased sample variance $S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. This raises the following question: if we replace $\sigma$ with $S_n$, is the central limit theorem still true? The answer is yes.

**Theorem 1.9.7** *Assume the same conditions as the CLT. Then*

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S_n^*} \xrightarrow{d} N(0,1).$$

You might wonder, how accurate the normal approximation is. The answer is given in the Berry-Esseen theorem.

**Theorem 1.9.8** *(the Berry-Esseen inequality). Suppose that* $E\left|X_1\right|^3 < \infty$. *Then*

$$\sup_z |P(Z_n \le z) - \Phi(z)| \le \frac{33}{4}\frac{E\left|X_1 - X\right|^3}{\sqrt{n}\sigma^3}.$$

## 1.10 The Lapunov Theorem

The distribution of a sum of independent random variables may not converge to the normal distribution, if the terms do not have the same distribution, even if all the random variables have standard deviations.

### 1.10.1 Part A

We now provide the Lapunov theorem, which gives a **sufficient condition** for a sum of independent random variables to have a limiting normal distribution. Consider a sequence $\{X_k\}$ of independent random variables whose moments of the third order exist.

**Theorem 1.10.1** *(Lapunov Theorem). Let* $\{X_k\}$ $(k = 1, 2, \ldots)$ *be a sequence of independent random variables whose moments of the third order exist, and let* $m_k, \sigma_k \ne 0, a_k,$ *and* $b_k$ *denote the expected value, standard deviation, central moment of the third order, adn the absolute central moment of the third order of* $X_k$, *respectively. Furthermore, let*

$$B_n = \sqrt[3]{\sum_{k=1}^{n} b_k}, \quad C_n = \sqrt{\sum_{k=1}^{n} \sigma_k^2}.$$

*If the relation*

$$\lim_{n\to\infty} \frac{B_n}{C_n} = 0$$

*is satisfied, the sequence* $\{F_n(z)\}$ *of the distribution functions of the random variables* $Z_n$, *defined as*

$$Z_n = \frac{\sum_{k=1}^{n}(X_k - m_k)}{C_n}, \tag{1.32}$$

*satisfies, for every* $z$, *the relation*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{s^2}{2}}\, ds. \tag{1.33}$$

The proof is ignored since it is too long.

### 1.10.2 Part B

We present the theorem of Lindeberg-Feller, giving a **necessary and sufficient condition**.

(Lindeberg-Feller Theorem). Let $\{X_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent random variables whose variances exist, and let $G_k(x), m_k, \sigma_k \neq 0$ denote, respectively, the distribution function, the expected value, standard deviation of the random variable $X_k$, and let $F_n(z)$ denote the distribution function of the standardized random variable $Z_n$ given by formula (1.32).

Then the relations

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{\sigma_k}{C_n} = 0, \qquad \lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{s^2}{2}} ds,$$

hold if and only if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \int_{|x-m_k|>\varepsilon C_n} (x - m_k)^2 dG_k(x) = 0.$$

Let $\{X_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent, uniformly bounded random variables, that is, there exists a constant $a > 0$ such that for every $k$,

$$P(|X_k| \le a) = 1,$$

and suppose that $Var(X_k) \neq 0$ for every $k$. Then a **necessary and sufficient condition** for relation (1.33) to hold is

$$\lim_{n\to\infty} C_n^2 = \infty.$$

## 1.11 The Gnedenko Theorem

The sequence of pdf's (in continous case) and the sequence of pmf's (in discrete case) may not converge to the corresponding limit pdf or pmf (see Problems 6.25 and 6.26 in Fisz book). We need more conditions to interchange limitation lim and derivative $\frac{d}{dx}$,

$$\lim_{n\to\infty} F_n(x) = F(x) \overset{\text{interchange}}{\longrightarrow} \frac{d}{dx}F(x) = \frac{d}{dx}\lim_{n\to\infty} F_n(x)? = \lim_{n\to\infty}\frac{d}{dx}F_n(x).$$

Here we present a case where a local limit theorem holds true.

**Theorem 1.11.1** *(Gnedenko). Suppose that the independent and equally distributed random variables $X_i$ $(i = 1, 2, \ldots)$ of the discrete type can take on with positive probability only integer values, and let $E(X_i) = m$ and $Var(X_i) = \sigma^2 > 0$. Then the relation*

$$\lim_{n\to\infty} \left[\sigma\sqrt{n}P_n(k) - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_{nk}^2}{2}\right)\right] = 0,$$

*where*

$$P_n(k) = P(\sum_{k=1}^{n} X_i = k),$$

$$z_{nk} = \frac{k - nm}{\sigma\sqrt{n}},$$

*is satisfied uniformly w.r.t. $k$ in the interval $(-\infty < k < \infty)$ if and only if the maximum span of the distribution of $X_i$ is equal to one.*

**Remark 1.11.2** *A particular case of this theorem is the local limit theorem of de Moivre-Laplace when $X_i$ can take the values 0 and 1, with with probabilities $1-p$ and $p$, respectively.*

## 1.12 Poisson's Chebyshev's and Khintchin's laws of large number

### 1.12.1 Part A. Chebyshev Theorem

History. The Bernoulli law of large numbers, historically the oldest, is only a particular case of more general theorems which are known under the common name of laws of large numbers.

Consider $\{X_k\}$ $(k = 1, 2, \dots)$ the only assumption is that for every $k$, the first two moments exist,

$$EX_k = m_k, \quad E(X_k - m_k)^2 = \sigma_k^2.$$

$X_k$ may or may not be independent.

By Chebyshev inequalities, we have for $\forall k$, $\forall \varepsilon > 0$,

$$P(|X_k - m_k| > \varepsilon) \le \frac{\sigma_k^2}{\varepsilon^2}.$$

If **the Markov condition**

$$\lim_{k\to\infty} \sigma_k^2 = 0, \tag{1.34}$$

then

$$\lim_{k\to\infty} P(|X_k - m_k| > \varepsilon) = 0.$$

**Theorem 1.12.1** *(Chebyshev theorem). Let $\{X_k\}$ be an arbitrary sequence of random variables with variance $\sigma_k^2$. If the Markov condition (1.34) is satisfied, the sequence $\{X_k - m_k\}$ is stochastically convergent to zero.*

**Corollary 1.12.2** *(Corollary of Chebyshev theorem). Let $\{X_k\}$ be a sequence of random variables pairwise uncorrelated and let $EX_k = m_k$ and $Var(X_k) = \sigma_k^2$. If condition*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 = 0, \tag{1.35}$$

*is satisfied, then*

$$\left\{ Y_n - \frac{m_1 + m_2 + \cdots + m_n}{n} \right\} \quad (n = 1, 2, \cdots)$$

*is stochastically convergent to zero.*

**Proof.** Let

$$Y_n - \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

We have

$$EY_n = \frac{m_1 + m_2 + \cdots + m_n}{n}.$$

Since $X_i$ are pairwise uncorrelated, we have

$$Var(Y_n) = Cov(\frac{X_1 + X_2 + \cdots + X_n}{n}, \frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2.$$

Since $Var(Y_n) \to 0$, then by Chebyshev Theorem, we have $Y_n - EY_n$ is stochastically convergent to zero. ■

### 1.12.2 Part B. Poisson law of large number

We consider sum of $n$ independent random varibles $X_k (k = 1, 2, \ldots)$ with the zero-one distribution, where

$$P(X_k = 0) = 1 - p_k, \quad P(X_k = 1) = p_k.$$

Since $Var(X_k) = p_k(1 - p_k) \leq 1/4$, Condition (1.35) is satisfied. Then we have the Poisson law of large numbers following the corollary of the Chebyshev theorem.

**Theorem 1.12.3** *(Poisson law of large numbers). If the random variables $Y_n$ is the arithmetic mean of the random variables $X_k$ in the Poisson scheme,*

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

*then the sequence*

$$\left\{ Y_n - \frac{p_1 + p_2 + \cdots + p_n}{n} \right\} \quad (n = 1, 2, \ldots)$$

*is stochastically convergent to 0.*

### 1.12.3 Part C. Chebyshev law of large numbers

Consider the pairwise uncorrelated $X_k$ have the same expected value and the same standard deviation. For every $k$, we write

$$EX_k = m, \quad Var(X_k) = \sigma^2.$$

If $Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$, we have

$$EY_k = m, \quad Var(Y_k) = \frac{\sigma^2}{k}.$$

Thus,

$$\lim_{k \to \infty} Var(Y_k) = 0.$$

**Theorem 1.12.4** *Let $\{X_k\}$ be pairwise uncorrelated with the same expected value and the same standard deviation, and let $Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$. Then $\{Y_n\}$ is stochastically convergent to the common expected value $m$.*

**Remark 1.12.5** *Bernoulli law is a special case of the Chebyshev law of large numbers.*

### 1.12.4 Part D. Khintchin's law of large numbers

In all above, the variances are assumed to exist. For the NEXT one, no assumption is made about the existence of the variances.

**Theorem 1.12.6** *(Khintchin's law of large numbers). Let $\{X_k\}$ be independent random variables with the same distribution and with expected value $EX_k = m$. Then $Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$ is stochastically convergent to $m$.*

**Proof.** Let $\phi(t)$ be Ch.f. of $X_k$. By independence of $X_k$, the Ch.f. of $Y_n$ is

$$\left[\phi(\frac{t}{n})\right]^n.$$

We can expand $\phi(t)$ around $t = 0$ according to Maclaurin,

$$\phi(t) = 1 + mit + o(t).$$

Then

$$\left[\phi(\frac{t}{n})\right]^n = \left[1 + \frac{mit}{n} + o(\frac{t}{n})\right]^n \to e^{mit},$$

which is the Ch.f. of the one-point distribution such that

$$P(Y = m) = 1.$$

By Lévy-Cramér theorem, $\{F_n(y)\}$ of distribution function of $Y_n$ converges to the distribution function of $Y$. By the equivalence of convergence in distribution and convergence in probability when $Y$ is a constant, we see that $\{Y_n\}$ is stochastically convergent to $m$. ■

**Example 1.12.7** *For a Cauchy distribution with Ch.f.* $\phi(t) = e^{-|t|}$. *The expected value does not exist. Thus the law of large numbers does not apply to* $\{Y_n\}$.

### 1.12.5 Part E. The strong law of large numbers

#### Definition

The laws of large numbers considered until now state that under certain conditions the sequence $\{Z_n\}$ of random variables defined by the formula

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k - c_n, \tag{1.36}$$

where $c_n = \frac{1}{n}\sum_{k=1}^{n} EX_k$ and the random variables $X_k$ are independent, is stochastically convergent to zero. Thus for arbitrary $\varepsilon > 0$ and $\eta > 0$ we can find an $N$ such that, for $n > N$, we have $P(|Z_n| > \varepsilon) < \eta$. It does not follow, however, that for arbitrary $\varepsilon > 0$ and $\eta > 0$ we can find an $N$ such that

$$P(\sup_{n \geq N} |Z_n| > \varepsilon) < \eta. \tag{1.37}$$

We observe that relation (1.37) implies that the probability of occurrence of the inequality $|Z_n| > \varepsilon$ for at least one value $n \geq N$ is smaller than $\eta$; thus, instead of the probability of one event $(|Z_n| > \varepsilon)$ we have here the probability of an alternative of events

$$(|Z_N| > \varepsilon) \cup (|Z_{N+1}| > \varepsilon) \cup (|Z_{N+2}| > \varepsilon) \cup \cdots$$

We show in the following Appendix that relation (1.37) is equivalent to the relation

$$P(\lim_{n\to\infty} Z_n = 0) = 1. \tag{1.38}$$

If (1.38) holds, we say that the sequence $\{Z_n\}$ is convergent to zero almost everywhere or almost surely.

So far, the laws of large numbers considered in the previous sections are called **weak laws of large numbers** since in the conclusion **we only arrive at the convergence in probability instead of the convergence almost surely**.

We say that the sequence $\{X_k\}$ $(k = 1, 2, \dots)$ of random variables obeys the strong law of large numbers if there exists a sequence of constants $\{c_n\}$ $(n = 1, 2, \dots)$ such that, for the random variables $Z_n$ defined by formula (1.36), relation (1.37) holds for all $\varepsilon > 0$ and $\eta > 0$.

### Theorem

It is important for the solution of the problem of necessary and sufficient conditions for the validity of the strong law of large numbers for a sequence of independent random variables. Detailed information on the present state of investigations in this field can be found in the monograph by Loeve [1] and in the paper of Chung [1]. The most advanced results have been obtained by Prohorov. We shall present the theorem of Kolmogorov [2] giving sufficient conditions for the validity of the strong law of large numbers. The proof of this theorem is based on a generalization of the Chebyshev inequality which was proved by Kolmogorov [1]. In the Fisz book, also proved is the Borel-Cantelli lemma (Borel [1], Cantelli [1]), which is used in the proof of the theorem of Kolmogorov [7], stating that for a sequence of *independent, identically distributed* random variables the existence of the expected value is a necessay and sufficient condition for the validity of the strong law of large numbers.

E. Kolmogorov [7] proved the following theorem concerning the validity of the strong law of large numbers for identically distributed random variables; it is called the **Kolmogorov law of large numbers**.

**Theorem 1.12.8** *Let $\{Y_i\}$ $(i = 1, 2, \dots)$ be a sequence of independent random variables with the same distribution function $F(y)$. Then the relation*

$$P\left[\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} Y_k - c\right) = 0\right] = 1,$$

*holds for some $c$ if and only if the expected value $E(Y)$ of a random variable $Y$ with the distribution function $F(y)$ exists; here $c = EY$.*

### Appendix

We now prove that relations (1.37) and (1.38) are equivalent.

Denote by $A_N$ the event that $\sup_{n\geq N} |Z_n| > \varepsilon$, where $\varepsilon > 0$, and by $A$ the product of the events $A_N$, that is,

$$A = \bigcap_N A_N.$$

We observe that for every $N$,

$$A_{N+1} \subset A_N,$$

and hence we have

$$P(A) = \lim_{N\to\infty} P(A_N). \tag{1.39}$$

The event $\overline{A_N}$, the complement of the event $A_N$, occurs if and only if, for every $n \geq N$, we have the relation $|Z_n| \leq \varepsilon$. Thus we have for every $N$,

$$\overline{A_{N+1}} \supset \overline{A_N},$$

hence

$$P(\overline{A}) = P(\sum_N \overline{A_N}) = \lim_{N\to\infty} P(\overline{A_N}). \tag{1.40}$$

Suppose now that relation (1.37) is not satisfied. Then there exist $\varepsilon > 0$ and $\eta > 0$ such that for every $N$,

$$P(A_N) \geq \eta.$$

From the last relation and from relation (1.39) we obtain the inequality

$$P(A) \geq \eta > 0, \tag{1.41}$$

from which it follows that relation (1.38) is not satisfied; for if it were satisfied, then for every $\varepsilon > 0$ the probability would be zero that for every $N$ there exists an $n \geq N$ such that $|Z_n| > \varepsilon$. Hence $P(A) = 0$, in contradiction to (1.41).

Suppose, now, that relation (1.38) is not satisfied. Then there exist $\varepsilon > 0$ and $\eta > 0$ such that the probability of occurrence of the event $\overline{A_N}$ is smaller than $1-\eta$ for every $N$, or

$$P(\overline{A}) < 1 - \eta. \tag{1.42}$$

It follows from the last inequality that relation (1.37) is not satisfied; for if it were satisfied, then for any $\varepsilon > 0$ and $\eta > 0$ there would exist an $N$ such that $P(\overline{A_N}) \geq 1-\eta$, so that from the fact that the sequence $\{\overline{A_n}\}$ is nondecreasing and from formula (1.40) we would obtain $P(\overline{A}) \geq 1-\eta$, in contradiction to (1.42).

The equivalence of relations (1.37) and (1.38) is proved.

At the end of this section we give an example of a sequence of random variables which converges to zero stochastically but does not converge to zero almost everywhere (see also Problem 6.38).

**Example 1.12.9** *Let us consider the sequence* $\{Z_n\}$ $(n = 1, 2, \ldots)$ *of independent random variables, where*

$$\begin{aligned} P(Z_n &= 1) = \frac{1}{n}, \\ P(Z_n &= 0) = 1 - \frac{1}{n}. \end{aligned} \tag{1.43}$$

*The sequence* $\{Z_n\}$ *converges to zero stochastically, since from the equality* $P(|Z_n| > \varepsilon) = P(Z_n = 1)$*, which holds for every* $0 < \varepsilon < 1$*, we obtain, for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n| > \varepsilon) = \lim_{n\to\infty} \frac{1}{n} = 0.$$

*However, the considered sequence* $\{Z_n\}$ *does not satisfy relation (1.38); for, denoting by* $A_n$ *the event* $(Z_n = 1)$*, it follows from (1.43) that*

$$\sum_{n=1}^{\infty} P(A_n) = \infty.$$

*From the independence of the* $A_n$ *and from the Borel-Cantelli lemma it follows that the probability that an infinite number of the* $A_n$ *will occur equals one; hence with probability one there will exist a subsequence of the sequence* $\{Z_n\}$ *which is not convergent to zero. This obviously contradict relation (1.38).*

## 1.13 Multi-Dimensional Limit Distribution and The Delte Method

**Theorem 1.13.1** *Multivariate Central Limit Theorem (MCLT). If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. $k \times 1$ random vectors with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $Cov(\mathbf{X}, \mathbf{X}) = \boldsymbol{\Sigma}$, then*

$$\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_k(\mathbf{0}, \boldsymbol{\Sigma}),$$

*where the sample mean*

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i,$$

*with*

$$\mathbf{X}_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{ik} \end{pmatrix} \in \mathbb{R}^k, \quad \boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}.$$

If $Y_n$ has a limiting Normal distribution then the delta method allows us to find the limiting distribution of $g(Y_n)$ where $g$ is any smooth function.

**Theorem 1.13.2** *(The Delta Method). Suppose that*

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \xrightarrow{d} N(0,1),$$

*and that $g$ is a differentiable function such that $g'(\mu) \neq 0$. Then*

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\,\sigma} \xrightarrow{d} N(0,1).$$

*In other words, $Y_n \approx N(\mu, \frac{\sigma^2}{n})$ implies that $g(Y_n) \approx N(g(\mu), (g'(\mu))^2\frac{\sigma^2}{n})$.*

**Consider what if $g'(\mu) = 0$? As I guess, when $g(x) = x^2$ then the limit distribution should be related to $\chi^2$ distribution. For other functions, one may follow the derivation of $\chi^2$ distribution.**

There is also a multivariate version of the delta method.

**Theorem 1.13.3** *(The Multivariate Delta Method). Suppose that $Y_n = (Y_{n1}, ..., Y_{nk})$ is a sequence of random vectors such that*

$$\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \Sigma).$$

*Let $g : \mathbb{R}^k \to \mathbb{R}$ and let*

$$\nabla g(y) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \vdots \\ \frac{\partial g}{\partial y_k} \end{pmatrix}.$$

*Let $\nabla_\mu$ denote $\nabla g(y)$ evaluated at $y = \mu$ and assume that the elements of $\nabla_\mu$ are nonzero. Then*

$$\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{d} N(0, \nabla_\mu^\top \Sigma \nabla_\mu).$$

**Example 1.13.4** *Let*

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix},$$

*be i.i.d. random vectors with mean* $\mu = (\mu_1, \mu_2)^\top$ *and variance* $\Sigma$*. Let*

$$\overline{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}, \quad \overline{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i},$$

*and define* $Y_n = \overline{X}_1 \overline{X}_2$*. Thus,* $Y_n = g(\overline{X}_1, \overline{X}_2)$ *where* $g(s_1, s_2) = s_1 s_2$*. By the central limit theorem,*

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \overline{X}_2 - \mu_2 \end{pmatrix} \xrightarrow{d} N(0, \Sigma).$$

*Now*

$$\nabla g(s) = \begin{pmatrix} \frac{\partial g}{\partial s_1} \\ \frac{\partial g}{\partial s_2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_1 \end{pmatrix},$$

*and so*

$$\nabla_\mu^\top \Sigma \nabla_\mu = \begin{pmatrix} \mu_2 & \mu_1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix} = \mu_2^2 \sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}.$$

*Therefore,*

$$\sqrt{n}\left(\overline{X}_1 \overline{X}_2 - \mu_1\mu_2\right) \xrightarrow{d} N(0, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}).$$