Summary of Distribution Theory

Jinhong Du

December 8, 2019

Contents

Chapter 1

Basic of Distribution Theory

1.1 Probability Space

1.1.1 Sample Space

1.1.2 σ-Algebra Field

Definition. **Algebra Field**

Given a sample space Ω, the *algebra field* is a collection of events such that

(i) $\Omega \in \mathcal{F}$;

(ii) if $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$;

(iii) if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

Also, (ii) and (iii) implies that if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

Definition. **σ-Algebra Field**

Given a sample space Ω, the σ-*algebra field* is a collection of events such that

(i) $\Omega \in \mathcal{F}$;

(ii) if $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$;

(iii) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Also, (ii) and (iii) implies that if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

1.1.3 Probability Measure

Definition. **Probability Measure**

Given a sample space Ω and the σ-algebra field $\mathcal{F}$ on Ω, then the *probability measure* $\mathbb{P} : \mathcal{F} \to [0, 1]$ is defined as

(i) $0 \le \mathbb{P}(A) \le 1$ for all $A \in \mathcal{F}$;

(ii) $\mathbb{P}(\Omega) = 1$;

(iii) (σ-additivity) if $A_1, A_2, \ldots \in \mathcal{F}$, $A_i \cap A_j = \varnothing$ for $i \ne j$, then $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

Properties

If $A_n \uparrow A$ as $n \to \infty$, i.e., $A_1 \subseteq A_2 \subseteq \cdots$, then $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ as $n \to \infty$.

1.2 Random Variables

1.2.1 Definition

Definition. **Random Variable**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the *random variable* is a function $X : \Omega \to \mathbb{R}$ such that X is measurable with respect to $\mathcal{F}$, i.e., $\{\omega : X(\omega) \le t\} \in \mathcal{F}$ for all $t \in \mathbb{R}$.
The probability that X takes on a value in a measurable set $A \subseteq \mathbb{R}$ is written as

$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

1.2.2 Discrete Randon Variables

1.2.3 Continuous Randon Variables

Theorem. **Continuity of p.d.f.**

Let F be the c.d.f. of a random vairable X. Then X has a p.d.f. continuous on $\mathbb{R} \setminus A$ if and only if F is continuous on $\mathbb{R}$ and continuously differentiable on the intervales between x_i's where $A = \{x - 1, x_2, \ldots\}$ is a set of isolated points.

1.2.4 Singular Randon Variables

1.3 Distribution Function

1.3.1 Definition

Definition. **Cumulative Distribution Function**

The *cumulative distribution function (c.d.f.)* of a random variable X defined in $(\Omega, \mathcal{F}, \mathbb{P})$ is given by

$$F(t) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\})$$

for $t \in \mathbb{R}$.

1.3.2 Properties

(1) (non-decreasing) if $u \leq v$, then $F(u) \leq F(v)$;

(2) (bounded) $\lim_{u \to +\infty} F(u) = 1$, $\lim_{u \to -\infty} F(u) = 0$;

(3) (right continuous left limit) F is right continuous with left limit.

Reversely, if there is a function G that satisfies these three conditions, then G is a c.d.f. for some random variable.

Theorem. **Continuity of c.d.f.**

Let F be the c.d.f. of a random variable X. Then F is continuous if and only if $F(X) \sim \text{Uniform}(0, 1)$.

1.4 Inverse Function

1.4.1 Definition

Definition. **Inverse Function**

Let F be the c.d.f. of a random variable X. Then the inverse function is defined as

$$F^{-}(u) = \inf_{x}\{x \in \mathbb{R} : F(x) \geq u\}$$

for $0 < u < 1$. If F is continuous and strictly increasing, then $F^{-} = F^{-1}$.

1.4.2 Properties

1. $x < F^{-}(u) \iff F(x) < u$, $x \geq F^{-}(u) \iff F(x) \geq u$;
2. F^{-} is non-decreasing;

3. $F(F^{-}(u)) \geq u$;
4. F^{-} is left continuous with right limit.
5. $\mathbb{P}(F^{-}(U) \leq x) = \mathbb{P}(F(x) \geq U) = F(x)$ for $U \sim \text{Uniform}(0,1)$.

Theorem. **Baysian Confidence Interval**

Let F^{-} be the inverse function of a random variable X and $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $0 < \alpha_1 + \alpha_2 < 1$, then

$$\mathbb{P}(X \in [F^{-}(\alpha_1), F^{-}(1-\alpha_2)]) \geq 1 - \alpha_1 - \alpha_2.$$

1.5 Representing Function

1.5.1 Definition

Definition. **Representing Function**

$R : (0,1) \to \mathbb{R}$ is a representing function for a random variable X if

(i) R is non-decreasing;

(ii) $R(U) \stackrel{D}{=\!=} X$ where $U \sim \text{Uniform}(0,1)$.

1.5.2 Properties

Theorem. **Transformation of Representing Function**

Let R be a representing function for a random variable X and $Y = T(X)$. If T is non-decreasing, then $T \circ R$ is a representing function for Y.

Theorem. **Representing Function And Inverse Function**

Let F be the c.d.f. of a random variable X and $R : (0,1) \to \mathbb{R}$. Then R is a representing function for X if and only if $F^{-}(u) \leq R(u) \leq F^{-}(u+)$ for all $0 < u < 1$. Furthermore,

(i) if u is a continuous point of F^{-}, then $R(u) = F^{-}(u)$;

(ii) if u is a discontinuous point of F^{-}, then $R(u) \in [F^{-}(u), F^{-}(u+)]$;

(iii) if R is left continuous, then $R = F^{-}$.

1.6 Quantile Function

1.6.1 Definition

Definition. **Quantile**

Let F be the c.d.f. of X. For $0 < \alpha < 1$, the αth quantile of X is any $x \in \mathbb{R}$ such that

$$F(x) = \mathbb{P}(X \leq x) \geq \alpha,$$

and

$$1 - F(x-) = \mathbb{P}(X \geq x) \geq 1 - \alpha.$$

Definition. **Quantile Function**

Let F be the c.d.f. of a random variable X. The quantile function of X is a defined as $Q : (0, 1) \to \mathbb{R}$ such that $Q(\alpha)$ is a quantile of F for all $0 < \alpha < 1$.

Furthermore, if F is continuous, then Q is unique.

1.6.2 Properties

Theorem. **Quantile Function And Representing Function**

Q is a quantile function of a random variable X if and only if Q is a representing function of X.

Chapter 2

Distribution Theory

2.1 Transformation of Random Variables

2.2 Order Statistics

2.3 Expectation

2.3.1 Inequalities

Markov Inequality

For a nonnegative random variable X and $c > 0$, if $\mu = \mathbb{E}X < \infty$, then

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}$$

Chebeyshev's Inequality

For a random variable X and $c > 0$, if $\mu = \mathbb{E}X < \infty$ and $\sigma^2 = \mathbb{E}|X - \mu|^2 < \infty$, then

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Chernoff's Inequality

For a random variable X, $c > 0$ and $t > 0$, then

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\mathbb{E}e^{tX}}{e^{tc}}.$$

2.4 Conditional Expectation

2.4.1 Inequalities

Jensen's Inequality

Young's Product Inequality

Holder's Inequality

Norm Inequality

2.5 Copula

Chapter 3

Tool

3.1 Lebesgue Dominated Convergence Theorem

3.2 Fubini Theorem

3.3 Fatou's Lemma

3.4 Uniformly integrablity

Chapter 4

Description of Random Variables

4.1 Moment Generating Function

4.1.1 Definition

Definition. **Moment Generating Function**

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto \mathbb{R}$, the *moment generating function (m.g.f.)* of X is given by

$$m(t) = \mathbb{E}e^{tX}, \qquad t \in \mathbb{R},$$

which is always well-defined and quasi-integrable.

There are other generating functions that are similar to the m.g.f, such as the following functions.

Definition. **Complex Moment Generating Function**

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto \mathbb{R}$, the *moment generating function (m.g.f.)* of X is given by

$$m(z) = \mathbb{E}e^{zX}, \qquad z \in \mathbb{C}.$$

Definition. **Cumulative Generating Function**

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto \mathbb{R}$, the *cumulative generating function* of X is given by

$$k(t) = \log \mathbb{E}e^{tX},$$

which is always well-defined and quasi-integrable.

However, it is more common to use the moment generating function. Later we will see that the m.g.f. generates the moments, and the moments can (in certain cases) be used to recover the MGF.

4.1.2 Properties

Convexity

Theorem. Convexity of m.g.f.

Let $m(t)$ be the m.g.f. of a random variable X. Then

$$B = \{t \in \mathbb{R} : m(t) < \infty\}$$

is a (possibly degenerate) interval containing 0. Moreover, $m(t)$ is convex on B.

Proof. Since $m(0) = 1 < \infty$, $0 \in B$. Since e^x is convex on $\mathbb{R}$, $\forall\ t_1, t_2 \in B$, $\forall\ \alpha \in [0, 1]$, we have

$$e^{[\alpha t_1 + (1-\alpha)t_2]X(\omega)} \leq \alpha e^{t_1 X(\omega)} + (1-\alpha)e^{t_2 X(\omega)}.$$

Taking expectation yields

$$m(\alpha t_1 + (1-\alpha)t_2) \leq \alpha m(t_1) + (1-\alpha)m(t_2).$$

■

Infinite Differentiability

Theorem. Infinite Differentiability of m.g.f.

Let $m(t)$ be the m.g.f. of a random variable X and

$$B = \{t \in \mathbb{R} : m(t) < \infty\}.$$

Let $\mathring{B}$ be an interior set of B. If $\mathring{B} \neq \varnothing$, then $\forall\ k \in \mathbb{N}$, $X^k e^{tX}$ is integrable and

$$m^{(k)}(t) = \mathbb{E}(X^k e^{tX}).$$

In particular, if $0 \in \mathring{B}$, then $\forall\ k \in \mathbb{N}$

$$m^{(k)}(0) = \mathbb{E}X^k.$$

This property shows that we can get the moments of X from its m.g.f..

Series Expansion

Theorem. **Series Expansion of m.g.f.**

Let $m(t)$ be the m.g.f. of a random variable X. If either one of the following holds,

(1) X^k is integrable for any $k \in \mathbb{N}$ and the series $\sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!}$ converges when $|t| < R$, where $R > 0$ is the radius of convergence;

(2) $R^* = \sup\{t : m(t) < \infty, m(-t) < \infty\} > 0$.

then $R = R^*$ and $m(t) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!}$ for all $|t| < R$.

Proof. (2) $\implies$ (1)

Since $\forall\ t \in (0, R)$, $m(t) + m(-t) < \infty$, we have

$$\mathbb{E}\left|\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right| \le \mathbb{E}\left(\sum_{k=0}^{\infty}\left|\frac{t^k X^k}{k!}\right|\right) = \mathbb{E}e^{|tX|} \le \mathbb{E}e^{tX} + \mathbb{E}e^{-tX} = m(t) + m(-t) < \infty,$$

i.e. $\mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right) < \infty$. By Fubini's Theorem,

$$\sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!} = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right) < \infty.$$

Therefore, $\mathbb{E}\frac{t^k X^k}{k!} < \infty$, i.e. X^k is integrable for any $k \in \mathbb{N}$. Also, $R \le R^*$.

(1) $\implies$ (2)

Let $t < R$ such that $\sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!}$ and $\sum_{k=0}^{\infty} \frac{(-t)^k \mathbb{E}X^k}{k!}$ converge. Then by Fubini's Theorem,

$$m(t) + m(-t) = \mathbb{E}e^{tX} + \mathbb{E}e^{-tX} = 2\mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^{2k} X^{2k}}{(2k)!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!} + \sum_{k=0}^{\infty} \frac{(-t)^k \mathbb{E}X^k}{k!} < \infty.$$

Therefore, $R^* \le R$.

Therefore, if one of (1) and (2) holds, the other also holds. So $R^* = R$ and $m(t) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}X^k}{k!}$ for all $|t| < R$. ■

This property shows that in certain condition, we can recover the m.g.f. of X if we know all of its moments.

4.1.3 Uniqueness

Theorem. **Uniqueness of m.g.f.**

Let $m_X(t)$ and $m_Y(t)$ be the m.g.f. of random variables X and Y. If there exists $a, b \in \mathbb{R}$ and $a < b$, s.t. $m_X(t) = m_Y(t) < \infty$, $\forall t \in (a, b)$, then $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$.

This theorem shows that if $m(t) < \infty$ in a non-degenerate interval, it can uniquely determine the distribution function. However, for some random variables, the m.g.f. may be finite only at zero. So the converse statement $F_X(x) = F_Y(x),\ x \in \mathbb{R} \quad \Longrightarrow \quad m_X(t) = m_Y(t),\ t \in [a, b]$ is wrong. It also means that even if $\mathbb{E}X^k = \mathbb{E}Y^k$ for all $k = 1, 2, \ldots$, F_X may be different from F_Y.

4.2 Characteristic Function

4.2.1 Definition

Definition. **Characteristic Function**

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto \mathbb{R}$, the *characteristic function (ch.f.)* of X is given by

$$\phi(t) = \mathbb{E}e^{itX}, \qquad t \in \mathbb{R}.$$

4.2.2 Properties

Bounded

$$|\phi(t)| = |\mathbb{E}e^{itX}| \le \mathbb{E}|e^{itX}| = 1$$

Continuity

$\phi(t)$ is continuous with respect to t, i.e., $\forall t_0 \in \mathbb{R}$, $\lim_{t \to t_0} \phi(t) = \phi(t_0)$.

Ch.f. of Sum of Independent Variables

4.2.3 Useful Tools

Levy's Inversion Formula

Theorem. **Levy's Inversion Formula**

$$\lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) \mathrm{d}t = \mathbb{P}(a < X < b) + \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2}.$$

Proof. By mean value theorem,

$$e^{-ita} - e^{-itb} = (b - a)(-it)e^{-itc}$$

for some $c \in [a, b]$. Then $\left|\frac{e^{-ita} - e^{-itb}}{it} e^{itX}\right| \le \left|\frac{e^{-ita} - e^{-itb}}{it}\right| = (b - a)|e^{-itc}| \le b - a$. So, $\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itX} \mathrm{d}t < \infty$.

By Fubini's Theorem,

$$\begin{aligned}
&\frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \mathbb{E}e^{itX} \mathrm{d}t \\
=&\frac{1}{2\pi} \mathbb{E}\left(\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itX} \mathrm{d}t\right) \\
=&\frac{1}{2\pi} \mathbb{E}\left(\int_{-c}^{c} \frac{e^{it(X-a)} - e^{it(X-b)}}{it} \mathrm{d}t\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\mathbb{E}\left(\int_{-c}^{c}\frac{\sin[it(X-a)]-\sin[it(X-b)]}{t}\mathrm{d}t\right)\\
&= \frac{1}{\pi}\mathbb{E}\left(\int_{0}^{c(X-a)}\frac{\sin t}{t}\mathrm{d}t-\int_{0}^{c(X-b)}\frac{\sin t}{t}\mathrm{d}t\right)\\
&= \frac{1}{\pi}\mathbb{E}\left(\int_{c(X-b)}^{c(X-a)}\frac{\sin t}{t}\mathrm{d}t\right).
\end{aligned}
$$

Since

$$
\lim_{c\to\infty} J_c(u) = \lim_{c\to\infty}\frac{1}{\pi}\int_{c(u-b)}^{c(u-a)}\frac{\sin t}{t}\mathrm{d}t = \begin{cases} 1 & \text{, if } a<u<b \\ \frac{1}{2} & \text{, if } u=a \text{ or } u=b\,, \\ 0 & \text{, if } u<a \text{ or } u>b \end{cases}
$$

$$
\sup_{u,v}\int_{u}^{v}\frac{\sin t}{t}\mathrm{d}t = \int_{-\pi}^{\pi}\frac{\sin t}{t}\mathrm{d}t < \infty
$$

and the dominated function is given by $\frac{\sin t}{t}\mathbb{1}_{(-\pi,\pi)}$ and by LDCT, we have

$$
\begin{aligned}
\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)\mathrm{d}t &= \lim_{c\to\infty}\mathbb{E}J_c(u)\\
&= \mathbb{E}(\lim_{c\to\infty}J_c(u))\\
&= \mathbb{P}(a<X<b)+\frac{\mathbb{P}(X=a)+\mathbb{P}(X=b)}{2}.
\end{aligned}
$$

■

Smoothing

Theorem. Smoothed Density

Let X be a real random variable with characteristic function $\phi(t)$ (X may not have a density function). Let $Z \sim N(0,1)$ be independent of X. For each $\sigma > 0$, the random variable $X_\sigma = X + \sigma Z$ has density f_σ given by

$$
f_\sigma(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t
$$

for $x \in \mathbb{R}$.

Proof. Since

$$
\phi_X(t) = \mathbb{E}e^{itX} = \int_{\mathbb{R}} e^{itx}\mathrm{d}F_X(x),
$$

we have

$$
\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itz}\phi_X(t)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-itz}e^{itx}e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}F_X(x)\mathrm{d}t
$$

$$
\begin{aligned}
&\overset{Fubini}{=\!=\!=} \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{\frac{\sigma^2}{2}\left[t-\frac{i(x-z)}{\sigma^2}\right]^2-\frac{(x-z)^2}{2\sigma^2}} \mathrm{d}t\mathrm{d}F_X(x)\\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} e^{-\frac{(z-x)^2}{2\sigma^2}}\mathrm{d}F_X(x)\\
&= \int_{\mathbb{R}} f_{\sigma Z}(z-x)\mathrm{d}F_X(x)\\
&= f_\sigma(z)
\end{aligned}
$$

■

4.2.4 Inverse Formula for Densities

Theorem. L

et $\phi_X(t)$ be the ch.f. of a random variable X. If $\int_{\mathbb{R}} |\phi(t)|\mathrm{d}t < \infty$, then X has a density function

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \phi(t)e^{-itx}\mathrm{d}t.$$

Also, $f(x)$ is bounded and continuous.

Proof. Method 1

From Levy's Inverse Theorem, we have

$$\lim_{b\to a^+}\frac{1}{2\pi}\int_{\mathbb{R}}\frac{e^{-iat}-e^{-ibt}}{it(b-a)}\phi(t)\mathrm{d}t = \lim_{b\to a^+}\frac{F(b)-F(a)}{b-a},$$

By mean value theorem, $\left|\frac{e^{-iat}-e^{-ibt}}{it(b-a)}\right| \le 1$. By LDCT, we have

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-iat}\phi(t)\mathrm{d}t = f(a).$$

Method 2

Let $X_\sigma = X + \sigma Z$ where $\sigma > 0$ and $Z \sim N(0,1)$ is independent of X. Then the p.d.f. for X_σ is given by $f_\sigma(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(x)e^{-\frac{\sigma t^2}{2}}\mathrm{d}t$. Define $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}}\phi(t)e^{-itx}\mathrm{d}t$. Since $\left|e^{-itx}\phi(x)e^{-\frac{\sigma t^2}{2}}\right| \le |\phi(x)|$, by LDCT,

$$\lim_{\sigma\to} f_\sigma(x) = f(x),\ \forall\ x\in\mathbb{R}.$$

Since

$$\sup_x |f_\sigma(x) - f_0(x)| \le \frac{1}{2\pi}\int_{\mathbb{R}}|\phi(t)|\cdot(1-e^{-\frac{\sigma^2 t^2}{2}})\mathrm{d}t.$$

By LDCT, $\lim_{\sigma\to 0}\frac{1}{2\pi}\int_{\mathbb{R}}|\phi(t)|\cdot(1-e^{-\frac{\sigma^2 t^2}{2}})\mathrm{d}t = 0$. So $f_\sigma(x)\to f(x)$ uniformly as $\sigma\to 0$. In particular, $f_\sigma(x)$ is real and nonnegative, so is $f(x)$. ■

4.2.5 Uniqueness

Theorem. **Uniqueness of ch.f.**

Let $\phi_X(t)$ and $\phi_Y(t)$ be the ch.f. of random variables X and Y. Then

$$\phi_X(t) = \phi_Y(t),\ \forall t \in \mathbb{R} \quad \Longleftrightarrow \quad F_X(x) = F_Y(x),\ \forall x \in \mathbb{R}$$

Proof. $\Longleftarrow$ If $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$, then X and Y have the same probability distribution and hence $\phi_X(t) = \phi_Y(t)$, $\forall t \in \mathbb{R}$.

$\Longrightarrow$

Method 1

Since $\phi_X(t) = \phi_Y(t),\ \forall t \in \mathbb{R}$, by Levy's Inversion Formula, we have

$$\mathbb{P}(a < X < b) + \frac{\mathbb{P}(X = a) + \mathbb{P}(X = b)}{2} = \mathbb{P}(a < Y < b) + \frac{\mathbb{P}(Y = a) + \mathbb{P}(Y = b)}{2}.$$

If we choose a and b to be the continuous point of F_X, then $\mathbb{P}(a < X < b) = \mathbb{P}(a < Y < b)$. Let $a \to -\infty$, then $\mathbb{P}(X < b) = \mathbb{P}(Y < b)$. Since distribution functions have left limits, we have $\mathbb{P}(X \le b) = \mathbb{P}(Y \le b)$.

Method 2

Consider $X_\sigma = X + \sigma Z$ and $Y_\sigma = Y + \sigma Z$ where $\sigma > 0$ and $Z \sim N(0, 1)$ is independent of X and Y. Then X_σ and Y_σ has density

$$f_\sigma(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi_X(t)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi_Y(t)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t.$$

Hence for all continuous and bounded g, $\mathbb{E}g(X_\sigma) = \mathbb{E}g(Y_\sigma)$. Let $\sigma \to 0$, by LDCT we have $\mathbb{E}g(X) = \mathbb{E}g(Y)$.

Let $g_{n,x}(t) = \begin{cases} 1 & \text{, if } t \le x \\ -n(t-x)+1 & \text{, if } x < t < x + \frac{1}{n} \\ 0 & \text{, if } t \ge x + \frac{1}{n} \end{cases}$ and $g_x(t) = \mathbb{1}_{t \le x}$, then $\lim\limits_{n\to\infty} g_{n,x}(t) = g_x(t)$, $\forall t \in \mathbb{R}$. Since $\mathbb{E}g_{n,x}(X) = \mathbb{E}g_{n,x}(Y)$, let $n \to \infty$, by LDCT we have $F_X(x) = F_Y(x)$. ∎

4.2.6 Convergence

Theorem. **Convergence of ch.f.**

Let $\phi_n(t)$ and $\phi(t)$ be the ch.f. of random variables X_n $(n = 1, 2, \ldots)$ and X. Then

$$\phi_n(t) \to \phi(t),\ \forall\ t \in \mathbb{R} \qquad \text{if and only if} \qquad X_n \xrightarrow{D} X.$$

Proof. $\implies$

Method 1. Let $X_\sigma = X + \sigma Z$ and $X_{n,\sigma} = X_n + \sigma Z$ where $\sigma > 0$ and $Z \sim N(0,1)$ is independent of X and X_n. From Inverse Formula for densities, we have

$$f_{n,\sigma}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi_n(x)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t$$
$$f_\sigma(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(x)e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t.$$

Since $|f_{n,\sigma}(t) - f_\sigma(t)| \le \displaystyle\int_{\mathbb{R}} |\phi_n(t) - \phi(t)|e^{-\frac{\sigma^2 t^2}{2}}\mathrm{d}t$ and $|\phi_n(t) - \phi(t)|e^{-\frac{\sigma^2 t^2}{2}} \le 2e^{-\frac{\sigma^2 t^2}{2}}$, by LDCT we have

$$\lim_{n\to\infty} |f_{n,\sigma}(t) - f_\sigma(t)| = 0.$$

Then $f_{n,\sigma} \to f_\sigma$ and so $X_{n,\sigma} \xrightarrow{D} X_\sigma$.

For all $g : \mathbb{R} \to \mathbb{R}$ which is bounded and Lipschitz continuous,

$$\begin{aligned}
|\mathbb{E}g(X_n) - \mathbb{E}g(X)| &\le |\mathbb{E}g(X_n) - \mathbb{E}g(X_{n,\sigma})| + |\mathbb{E}g(X_{n,\sigma}) - \mathbb{E}g(X_\sigma)| + |\mathbb{E}g(X_\sigma) - \mathbb{E}g(X)| \\
&\le \mathbb{E}(L\sigma|Z|) + |\mathbb{E}g(X_{n,\sigma}) - \mathbb{E}g(X_\sigma)| + \mathbb{E}(L\sigma|Z|) \\
&= 2L\sigma\mathbb{E}|Z| + +|\mathbb{E}g(X_{n,\sigma}) - \mathbb{E}g(X_\sigma)| \to 0
\end{aligned}$$

as $\sigma \to 0$. Thus, $X_n \xrightarrow{D} X$.

Method 2. From similar argument above, we have $X_{n,\sigma} \xrightarrow{D} X_\sigma$. Since

$$\mathbb{P}(X \le x) \le \mathbb{P}(X_\sigma \le x + \delta) + \mathbb{P}(\sigma|Z| \ge \delta)$$
$$\mathbb{P}(X_n \le x) \le \mathbb{P}(X_{n,\sigma} \le x + \delta) + \mathbb{P}(\sigma|Z| \ge \delta),$$

we have

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(X_n \le x) &\le \limsup_{n\to\infty} \mathbb{P}(X_{n,\sigma} \le x + \delta) + \mathbb{P}(\frac{1}{k}|Z| \ge \delta) \\
&= \mathbb{P}(X_\sigma \le x + \delta) + \mathbb{P}(\frac{1}{k}|Z| \ge \delta) \\
&\le \mathbb{P}(X \le x + 2\delta) + 2\mathbb{P}(\frac{1}{k}|Z| \ge \delta).
\end{aligned}$$

Let $\sigma \to 0$,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \le x) \le leq\mathbb{P}(X \le x + 2\delta)$$

. Similar, we will have

$$\mathbb{P}(X \le x + 2\delta) \le \liminf_{n\to\infty} \mathbb{P}(X_n \le x).$$

Assume $x \in C_{F_X}$, let $\delta \to 0$, we have

$$\mathbb{P}(X \le x) \le \liminf_{n\to\infty} \mathbb{P}(X_n \le x) \le \limsup_{n\to\infty} \mathbb{P}(X_n \le x)w \le leq\mathbb{P}(X \le x),$$

i.e. $\lim\limits_{n\to\infty} F_n(x) = F(x)$, i.e., $X_n \xrightarrow{D} X$.

$\Longleftarrow$ By the Characterization of Weak Convergence, let $f(x) = e^{itx}$, we have $\mathbb{E}e^{itX_n} \to \mathbb{E}e^{itX}$ as $n \to \infty$. $\blacksquare$

Chapter 5

Convergence of Random Variables

5.1 Types of Convergence

5.1.1 Convergence in Distribution

Definition. **Weak Convergence/Convergence in Distribution**

Let $X_1, X_2, \ldots$ and X be real random variables with respect to distributions $F_1, F_2, \ldots$ and F. If

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad \forall\ x \in C_F,$$

where $C_F = \{x \in \mathbb{R} : F \text{ is continuous at } x\} = \{x \in \mathbb{R} : \mathbb{P}(X = x) = 0\}$, then X_n converges to X in distribution, denoted by $X_n \xrightarrow{D} X$, as $n \to \infty$.

Theorem. **Convergence in Inverse Distribution**

Let $X_1, X_2, \ldots$ and X be real random variables with left-continuous inverse distribution functions $F_1^-, F_2^-, \ldots$ and F^- respectively. Then

$$X_n \xrightarrow{D} X \qquad \text{if and only if} \qquad \lim_{n\to\infty} F_n^-(u) = F^-(u) \text{ for all } u \in C_{F^-},$$

where C_{F^-} is the set of continuous points of F^-.

Proof. First we have the following fact.

Let B be either $\mathbb{R}$ or $(0,1)$. Let f be a non-decreasing mapping from B into $\mathbb{R}$. Then $\mathcal{D}_f = \{x \in B : f \text{ is discontinuous at } x\}$ is at most countable and $\mathcal{D}_f^c$ is a dense set. To see this, first look at a bounded interval $[n, n+1)\ \forall\ n \in \mathbb{Z}$. If $f(n+1) \neq f(n)$, let D_n denote the set of points at which f has a discontinuity. For each positive integer m, let $D_{n,m}$ denote the set of points $x \in [n, n+1)$ such that f has a jump of at least $\frac{1}{m}(f(n+1) - f(n))$ at x and let $N_{n,m}$ denote the number of elements in $D_{n,m}$. Note that

$$D_n = \bigcup_{m=1}^{\infty} D_{n,m},$$

we have $N_{n,m} \le m$. It follows that the number of points of discontinuity is bounded by $\sum_n \sum_{m=1}^{\infty} m$. The result follows. Then $\mathcal{D}_F$ and $\mathcal{D}_{F^-}$ are both countable.

$\Longrightarrow$

Suppose that $u \in (0,1)$ and that w is a continuity point of F with $F^-(u) > w$. Since

$$F^-(u) > w \quad \Longleftrightarrow \quad F(w) < u$$

and $\lim_{n\to\infty} F_n(x) = F(x)$ for all continuity points x of F, we have $\lim_{n\to\infty} F_n(w) < u$. Therefore, $\exists\ N_1 = N_1(w) \in \mathbb{Z}_+$, s.t. $\forall\ n > N_1$, $F_n(w) < u$, i.e. $F_n^-(u) > w$.

Suppose that $u \in (0,1)$ and that y is a continuity point of F with $F^-(u+) < y$. Since

$$F^-(u) \le y \quad \Longleftrightarrow \quad F(y) \ge u$$

Similarly, we have $\lim_{n\to\infty} F_n(y) \ge u$. Therefore, $\exists\ N_2 = N_2(y) \in \mathbb{Z}_+$, s.t. $\forall\ n > N_2$, $F_n(y) \ge u$, i.e. $F_n^-(u) \le y$.

Since the set $\mathcal{D}_F$ of jumps of the distribution function F is at most countable, the complement of countable set $\mathbb{R} \setminus \mathcal{D}_F$ is dense. $\forall\ \epsilon > 0$, which makes $w = F^-(u) - \epsilon$, $y = F^-(u+) + \epsilon \in \mathbb{R} \setminus \mathcal{D}_F$, $\exists\ N = \max\{N_1, N_2\}$, s.t. $\forall\ n > N$,

$$F^-(u) - \epsilon \le F_n^-(u) \le F^-(u+) + \epsilon.$$

Since $\mathbb{R} \setminus \mathcal{D}_F$ is dense, such ϵ can be arbitrarily small. Let $\epsilon \to 0$, we have

$$F^-(u) \le \liminf_n F_n^-(u) \le \limsup_n F_n^-(u) \le F^-(u+),$$

which implies

$$\lim_n F_n^-(u) = F^-(u) \text{ for all continuity points } u \text{ of } F^-.$$

$\Longleftarrow$

Suppose that $x \in \mathbb{R}$ and that w' is a continuity point of F^- with $F(x) < w'$. Since

$$F(u) < w' \quad \Longleftrightarrow \quad F^-(w') > x$$

and $\lim_{n\to\infty} F_n^-(x) = F^-(x)$ for all continuity points u of F^-, we have $\lim_{n\to\infty} F_n^-(w') > x$. Therefore, $\exists\ N_3 = N_3(w') \in \mathbb{Z}_+$, s.t. $\forall\ n > N_3$, $F_n^-(w') > x$, i.e. $F_n(x) < w'$.

Suppose that $x \in \mathbb{R}$ and that y' is a continuity point of F^- with $F(x-) \ge y'$. Since

$$F(x) \ge y' \quad \Longleftrightarrow \quad F^-(y') \le x$$

Similarly, we have $\lim_{n\to\infty} F_n^-(y') \leq x$. Therefore, $\exists\ N_4 = N_4(y') \in \mathbb{Z}_+$, s.t. $\forall\ n > N_4$, $F_n^-(y') \leq x$, i.e. $F_n(x) \geq y'$.

Since the set $\mathcal{D}_{F^-}$ of jumps of the distribution function F^- is at most countable, the complement of countable set $(0,1)\backslash\mathcal{D}_{F^-}$ is dense. $\forall\ \epsilon' > 0$, which makes $w' = F(x)+\epsilon$, $y' = F(x-)-\epsilon \in (0,1)\backslash\mathcal{D}_{F^-}$, $\exists\ N' = \max\{N_3, N_4\}$, s.t. $\forall\ n > N'$,

$$F(x-) - \epsilon' \leq F_n(x) \leq F(x) + \epsilon'.$$

Since $(0,1) \setminus \mathcal{D}_F$ is dense, such ϵ' can be arbitrarily small. Let $\epsilon' \to 0$, we have

$$F(x-) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x),$$

which implies

$$\lim_n F_n(x) = F(x) \text{ for all continuity points } x \text{ of } F.$$

■

Theorem. Characterization of Weak Convergence

Let $X_1, X_2, \ldots$ and X be real random variables. The following are equivalent.

(1) $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded and X-continuous $f : \mathbb{R} \to \mathbb{R}$ i.e. $\mathbb{P}(X \in D_f) = 0$ where D_f is the set of discontinuous points of f.

(2) $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded and continuous $f : \mathbb{R} \to \mathbb{R}$.

(3) $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded and Lipschitz continuous $f : \mathbb{R} \to \mathbb{R}$.

Proof. $\Longrightarrow$ From Sckorohod Representing Theorem, $\exists\ Y_n \xrightarrow{D} X_n$, $Y \xrightarrow{D} X$ and $Y_n \to Y$ a.s.. Then for f which is bounded and X-continuous, it is also Y-continuous and $f(Y_n) \to f(Y)$ a.s.. Therefore, $f(Y_n) \xrightarrow{D} f(Y)$ and $\mathbb{E}f(X_n) = \mathbb{E}f(Y_n) \to \mathbb{E}f(Y) = \mathbb{E}f(X)$ as $n \to \infty$.

$\Longleftarrow$ $\forall\ x \in C_F$, then $f(t) = \mathbb{1}_{t \leq x}$ is X-continuous. Therefore, $F_n(x) = \mathbb{E}f(X_n) \to \mathbb{E}f(X) = F(x)$ as $n \to \infty$. ■

Theorem. The Mapping Theorem for Convergence In Distribution

Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If $X_n \xrightarrow{D} X$ as $n \to \infty$ and $g : \mathbb{R} \to \mathbb{R}$ is X-continuous, then $h(X_n) \xrightarrow{D} h(X)$.

5.1.2 Convergence in Probability

Definition. **Convergence in Probability**

Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If for any $\epsilon > 0$,

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

then X_n converges to X in probability, denoted by $X_n \xrightarrow{\mathbb{P}} X$, as $n \to \infty$.

5.1.3 Convergence Almost Surely

Definition. **Convergence Almost Surely**

Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If

$$\mathbb{P}(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1,$$

then X_n converges to X almost surely/almost everywhere/with probability one, denoted by $X_n \xrightarrow{a.s.} X$ or $X_n \xrightarrow{a.s.} X$, as $n \to \infty$.

Theorem. **Borel Cantelli Implies Almost Sure Convergence**

Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If $\forall\ \epsilon > 0$ and $A_n(\epsilon) = \{\omega \in \Omega : \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}\}$, $\sum_{n=1}^{\infty} \mathbb{P}(A_n(\epsilon)) < \infty$, then $X_n \xrightarrow{a.s.} X$ as $n \to \infty$.

Proof.

$$\begin{aligned}
\mathbb{P}(\lim_{n\to\infty} X_n = X) &= 1 - \mathbb{P}\left(\bigcup_{m=1}^{\infty} \limsup_n A_n(\frac{1}{m})\right) \\
&\geq 1 - \sum_{m=1}^{\infty} \mathbb{P}\left(\limsup_n A_n(\frac{1}{m})\right) \\
&= 1
\end{aligned}$$

where the last inequality comes from Borel Cantelli Lemma. ■

Theorem. **The Mapping Theorem for Convergence Almost Surely**

Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If $X_n \xrightarrow{a.s.} X$ as $n \to \infty$ and $g : \mathbb{R} \to \mathbb{R}$ is X-continuous, then $h(X_n) \xrightarrow{a.s.} h(X)$.

Proof. If $\omega \in \Omega$ such that $X_n(\omega) \to X(\omega)$ and $X(\omega) \in C_h$, then $h(X_n) \to h(X)$. Hence,

$$\{\omega : h(X_n) \nrightarrow h(X)\} \subset \{\omega : X_n \nrightarrow X\} \cup \{\omega : X(\omega) \notin C_h\}.$$

Therefore,

$$0 \leq \mathbb{P}(\{\omega : h(X_n) \nrightarrow h(X)\}) \leq \mathbb{P}(\{\omega : X_n \nrightarrow X\}) + \mathbb{P}(\{\omega : X(\omega) \notin C_h\}) = 0$$

i.e., $\mathbb{P}(\{\omega : h(X_n) \nrightarrow h(X)\}) = 0$. ■

5.2 Properties

Theorem. **Slutsky's Theorem**

Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{\mathbb{P}} (Y)$, then

(i) $X_n + Y_n \xrightarrow{D} X + c$;

(ii) $X_n Y_n \xrightarrow{D} cX$;

(iii) $\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}$.

Proof. It suffices to show that $\mathbb{E}[g(X_n, Y_n)] \to \mathbb{E}[g(X, c)]$ for any bounded and Lipschitz continuous $g : \mathbb{R}^2 \to \mathbb{R}$. Suppose that g is bounded by C with Lipschitz norm L. Since

$$|\mathbb{E}[g(X_n, Y_n)] - \mathbb{E}[g(X, c)]| \leq I_n + II_n,$$

where $I_n = |\mathbb{E}[g(X_n, Y_n)] - \mathbb{E}[g(X_n, c)]|$ and $II_n = |\mathbb{E}[g(X_n, c)] - \mathbb{E}[g(X, c)]|$. One has $I_n \to 0$ since $I_n \leq \mathbb{E}[\min\{L|Y_n - c|, C\}]$ and $Y_n \xrightarrow{\mathbb{P}} Y$. Moreover, $II_n \to 0$ since $X_n \xrightarrow{D} X$ and $g(\cdot, c) : \mathbb{R} \to \mathbb{R}$ is bounded and continuous. Then $(X_n, Y_n) \xrightarrow{D} (X, c)$. By choosing (X, c)-continuous function $h(x, c) = x + c$, cx and $\frac{x}{c}$, the results follow. ∎

5.3 Relationship Between Different Convergence

5.3.1 $X_n \xrightarrow{L^1} X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X$

Proof. By Markov Inequality, $\forall\ \epsilon > 0$,

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}|X_n - X|}{\epsilon} \to 0$$

as $n \to \infty$ since $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$. Therefore, $X_n \xrightarrow{\mathbb{P}} X$. ■

5.3.2 $X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{\mathbb{P}} X$

Proof. $\forall\ \epsilon > 0$, since $X_n \xrightarrow{a.s.} X$, we have $\lim_{n\to\infty} \mathbb{1}_{\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}} = 0$ a.s..

By Fatou's Lemma, we have

$$0 \leq \limsup_{n\to\infty} \mathbb{E}\mathbb{1}_{\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}} \leq \mathbb{E} \limsup_{n\to\infty} \mathbb{1}_{\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}} = 0,$$

i.e. $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0$, i.e. $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$. ■

5.3.3 $X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{D} X$

Proof. Since $X_n \xrightarrow{a.s.} X$ and for all $x \in C_{F_X}$ $h(t) = \mathbb{1}_{t \leq x}$ is X-continuous, we have $\mathbb{1}_{X_n \leq x} \xrightarrow{a.s.} \mathbb{1}_{X \leq x}$. Then by LDCT, $\mathbb{E}\mathbb{1}_{X_n \leq x} \to \mathbb{E}\mathbb{1}_{X \leq x}$, i.e., $F_{X_n}(x) \to F_X(x)$ for all $x \in C_{F_X}$. ■

5.3.4 $X_n \xrightarrow{\mathbb{P}} X \Longrightarrow X_n \xrightarrow{D} X$

Proof. For $x \in C_{F_X}$ and $\epsilon > 0$, since

$$\begin{aligned}
\mathbb{P}(X_n \leq x) &\leq \mathbb{P}(X \leq x + \epsilon, X_n - X \leq -\epsilon) \\
&\leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(X_n - X \leq -\epsilon) \\
&\leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X - X_n| \geq \epsilon),
\end{aligned}$$

we have

$$\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x) \geq -\mathbb{P}(x < X \leq x + \epsilon) - \mathbb{P}(|X_n - X| \geq \epsilon).$$

Then

$$\liminf_{n\to\infty} [\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x)] \geq -\mathbb{P}(x < X \leq x + \epsilon).$$

Let $\epsilon \to 0^+$, we have $\liminf_{n\to\infty} [\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x)] \geq 0$.

Also,

$$\mathbb{P}(X \leq x - \epsilon) \leq \mathbb{P}(X_n \leq x, X - X_n \leq -\epsilon)$$

$$
\begin{aligned}
&\leq \mathbb{P}(X_n \leq x) + \mathbb{P}(X - X_n \leq -\epsilon) \\
&\leq \mathbb{P}(X_n \leq x) + \mathbb{P}(|X - X_n| \geq \epsilon),
\end{aligned}
$$

so

$$\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x) \leq \mathbb{P}(x - \epsilon < X \leq x) + \mathbb{P}(|X_n - X| \geq \epsilon).$$

Then

$$\limsup_{n\to\infty}[\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x)] \leq \mathbb{P}(x - \epsilon < X \leq x).$$

Let $\epsilon \to 0^+$, we have $\limsup_{n\to\infty}[\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x)] \leq 0$.

Therefore,

$$\lim_{n\to\infty}[\mathbb{P}(X \leq x) - \mathbb{P}(X_n \leq x)] = 0,$$

i.e., $\lim_{n\to\infty} F_n(x) = F(x)$ and $X_n \xrightarrow{D} X$.

■

5.3.5 $\phi_n \xrightarrow{L^1} \phi \Longrightarrow f_n \xrightarrow{L^1} f$

Theorem. Convergence of ch.f. Impies Convergence of Density

Let $\phi_n(t)$ and $\phi(t)$ be integrable ch.f. of random variables X_n $(n = 1, 2, \ldots)$ and X. If $\int_{\mathbb{R}} |\phi_n(t) - \phi(t)| \mathrm{d}t \to 0$, as $n \to \infty$, then X_n and X have bounded continuous densities f_n and f with respect to Lebesgue measure measure on $\mathbb{R}$, $f_n \to f$ uniformly and $f_n \xrightarrow{L^1} f$.

Proof. By the inversion theorem, X_n and X have bounded continuous densities

$$
\begin{aligned}
f_n(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_n(t) \mathrm{d}t, \\
f(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t) \mathrm{d}t.
\end{aligned}
$$

Since $\forall\ x \in \mathbb{R}$,

$$
\begin{aligned}
|f_n(x) - f(x)| &= \frac{1}{2\pi} \left| \int_{\mathbb{R}} e^{-itx} (\phi_n(t) - \phi(t)) \mathrm{d}t \right| \\
&\leq \frac{1}{2\pi} \int_{\mathbb{R}} \left| e^{-itx} (\phi_n(t) - \phi(t)) \right| \mathrm{d}t \\
&\leq \frac{1}{2\pi} \int_{\mathbb{R}} |\phi_n(t) - \phi(t)| \, \mathrm{d}t \\
&\to 0
\end{aligned}
$$

as $n \to \infty$, we have

$$\lim_{n\to\infty} \sup_{x} |f_n(x) - f(x)| = 0.$$

Since $|f_n(x)-f(x)| \le f_n(x)+f(x) \to 2f(x)$ as $n \to \infty$ and $\int_{\mathbb{R}} [f_n(x)+f(x)]\mathrm{d}x = 2 \to \int_{\mathbb{R}} 2f(x)\mathrm{d}x$, we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f_n(x) - f(x)|\mathrm{d}x = 0$$

by Sandwich Theorem. ■

Chapter 6

Limit Theorems

6.1 The Law of Large Nnumbers

Theorem. The Weak Law of Large Nnumbers

Let $X_1, X_2, \ldots$ be independent identically distributed integrable random variable with mean $\mathbb{E}X_k = \mu < \infty$, and $S_n = \sum_{k=1}^{n} X_k$ for $n \geq 1$. Then, as $n \to \infty$,

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

Proof. Since $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$ means $\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$. It is equivalent to

$$\mathbb{P}(\frac{S_n}{n} \leq x) \to \begin{cases} 1 & \text{, if } x > \mu \\ 0 & \text{, if } x < \mu \end{cases},$$

i.e., $\frac{S_n}{n} \xrightarrow{D} \mu$. Thus, we just need to show that $\frac{S_n}{n} \xrightarrow{D} \mu$.

The ch.f. of $\overline{X}_n$ is given by

$$\phi_n(t) = \mathbb{E}e^{it\overline{X}_n} = \mathbb{E}e^{i\frac{t}{n}\sum_{k=1}^{n} X_k} = \phi(\frac{t}{n})^n$$

where ϕ is the ch.f. of X_k. Since $\phi'(0) = \mathbb{E}(iX_k e^{itX_k})|_{t=0} = i\mathbb{E}X_k = i\mu$, we have

$$\begin{aligned}\phi_n(t) &= \left(1 + \frac{it}{n}\phi'(0) + o(\frac{t}{n})\right)^n \\ &= \left(1 + \frac{t\mu + o(1)}{n}\right)^n \\ &\to e^{it\mu}\end{aligned}$$

as $n \to \infty$. Therefore, $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$. ∎

Theorem. The Strong Law of Large Nnumbers

Let $X_1, X_2, \ldots$ be independent identically distributed integrable random variable with mean $\mathbb{E}X_k = \mu < \infty$, and $X_n = \frac{1}{n}\sum_{k=1}^{n} x_k$ for $n \geq 1$. Then, as $n \to \infty$,

$$\overline{X}_n \xrightarrow{a.s.} \mu.$$

6.2 The central limit theorem

Theorem. **The Central Limit Theorem**

Let $X_1, X_2, \ldots$ be independent identically distributed random variable with mean $\mathbb{E}X_i = \mu$ and variance $Var(X_i) = \sigma^2 < \infty$, and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n x_i$ for $n \geq 1$. Then, as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Proof. The ch.f. of $\sqrt{n}(\overline{X}_n - \mu)$ is given by

$$\phi_n(t) = \mathbb{E}e^{it\sqrt{n}(\overline{X}_n - \mu)} = \mathbb{E}e^{i\frac{t}{\sqrt{n}}\sum_{k=1}^n (X_k - \mu)} = \phi(\frac{t}{\sqrt{n}})^n$$

where ϕ is the ch.f. of $X_k - \mu$. Since

$$\phi'(0) = \mathbb{E}[i(X_k - \mu)e^{it(X_k - \mu)}]|_{t=0} = i\mathbb{E}(X_k - \mu) = 0$$
$$\phi^{(2)}(0) = -\mathbb{E}[(X_k - \mu)^2 e^{it(X_k - \mu)}]|_{t=0} = -\mathbb{E}[(X_k - \mu)^2] = -\sigma^2$$

we have

$$\begin{aligned}\phi_n(t) &= \left(1 + \frac{t}{\sqrt{n}}\phi'(0) + \frac{t^2}{2n}\phi^{(2)}(0) + o(\frac{t^2}{n})\right)^n \\ &= \left(1 - \frac{t^2\sigma^2 + o(1)}{2n}\right)^n \\ &\to e^{-\frac{it^2\sigma^2}{2}}\end{aligned}$$

as $n \to \infty$. Therefore, $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$. ■

Theorem. **The Central Limit Theorem for Uniform Order Statistics**

Let $U_1, \ldots, U_n \overset{iid}{\sim} \text{Uniform}(0,1)$ and $U_{(1)}, \ldots, U_{(n)}$ be the corresponding order statistics. Let $m = \frac{n}{2}$, then

$$\sqrt{n}(U_{(m)} - \tfrac{1}{2}) \xrightarrow{D} N(0, \tfrac{1}{4}).$$

Proof. Since $U_{(m)} = \frac{S_m}{S_{n+1}}$ where $S_k = \sum_{i=1}^k Z_i$ and $Z_1, \ldots Z_{n+1} \overset{iid}{\sim} \text{Exp}(1)$, we have

$$\begin{aligned}\sqrt{n}(U_{(m)} - \tfrac{1}{2}) &= \sqrt{n}(\frac{S_m}{S_{n+1}} - \tfrac{1}{2}) \\ &= \sqrt{n}\frac{S_m - \frac{1}{2}S_{n+1}}{S_{n+1}} \\ &= \frac{\sqrt{n}}{2}\frac{\sum_{i=1}^m (X_i - X_{m+i})}{S_{n+1}}\end{aligned}$$

$$= \frac{n}{2\sqrt{2}S_{n+1}} \cdot \frac{\sum_{i=1}^{m}(X_i - X_{m+i})}{\sqrt{\frac{n}{2}}}$$

By LLN, $\frac{2\sqrt{2}S_{n+1}}{n+1} \xrightarrow{\mathbb{P}} 2\sqrt{2}$. By Slutsky's Lemma, we have $\frac{n+1}{2\sqrt{2}S_{n+1}} \xrightarrow{\mathbb{P}} \frac{1}{2\sqrt{2}}$ and $\frac{n}{2\sqrt{2}S_{n+1}} \xrightarrow{D} \frac{1}{2\sqrt{2}}$.

Since $X_1 - X_{m+1}, \ldots, X_m - X_n$ are independent with mean 0 and variance 2, we have $\frac{\sum_{i=1}^{m}(X_i - X_{m+i})}{\sqrt{\frac{n}{2}}} \xrightarrow{D} N(0,2)$. By Slutsky's Lemma, we have

$$\sqrt{n}(U_{(m)} - \tfrac{1}{2}) \xrightarrow{D} N(0, \tfrac{1}{4}).$$

■

Theorem. Lindeberg's Feller CLT

For $n \in \mathbb{N}_+$, let $X_{n1}, X_{n2}, \ldots, X_{nn}$ be independent random variable with mean $\mathbb{E}X_{nk} = \mu_{nk}$ and variance $Var(X_{nk}) = \sigma_{nk}^2 < \infty$, $S_n = \sum_{k=1}^{n}(X_{nk} - \mu_{nk})$ and $Var(S_n) = v_n^2 = \sum_{k=1}^{n} \sigma_{nk}^2$.

(1) If $\forall\ \epsilon > 0$, $\frac{1}{v_n^2}\sum_{k=1}^{n} \mathbb{E}(X_{nk} - \mu_{nk})^2 \mathbb{1}_{|X_{nk}-\mu_{nk}| \geq \epsilon v_n} \to 0$, then as $n \to \infty$,

$$\frac{S_n}{v_n} \xrightarrow{D} N(0,1).$$

(2) If as $n \to \infty$, $\frac{S_n}{v_n} \xrightarrow{D} N(0,1)$ and $\frac{\max_k \sigma_{nk}^2}{v_n^2} \to 0$, then $\forall\ \epsilon > 0$, $\frac{1}{v_n^2}\sum_{k=1}^{n} \mathbb{E}(X_{nk} - \mu_{nk})^2 \mathbb{1}_{|X_{nk}-\mu_{nk}| \geq \epsilon v_n} \to 0$.

Proof. (1) Since $\forall\ \epsilon > 0$, $\frac{1}{v_n^2}\sum_{k=1}^{n} \mathbb{E}(X_{nk} - \mu_{nk})^2 \mathbb{1}_{|X_{nk}-\mu_{nk}| \geq \epsilon v_n} \to 0$, we have

$$\max_k \frac{\sigma_{nk}^2}{v_n^2} \leq \epsilon^2 + \frac{1}{v_n^2}\mathbb{E}(X_{nk} - \mu_{nk})^2 \mathbb{1}_{|X_{nk}-\mu_{nk}| \geq \epsilon v_n} \to 0.$$

by letting $n \to \infty$ and $\epsilon \to 0^+$.

Since

$$e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!} = \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} \mathrm{d}s = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is} - 1)\mathrm{d}s$$

$$\left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} \mathrm{d}s\right| \leq \frac{|x|^{n+1}}{(n+1)!}$$

$$\left|\frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is} - 1)\mathrm{d}s\right| \leq \frac{2|x|^n}{n!},$$

we have

$$\left|e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!}\right| \leq \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\}.$$

Obeserve that for every real $x > 0$, $|e^{-x} - 1 + x| \leq \frac{x^2}{2}$. Moreover, for complex z_j and w_j with $|z_j| < 1$ and $|w_j| < 1$, $\left|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j\right| \leq \sum_{j=1}^n |z_j - w_j|$. Then, for any $\epsilon > 0$,

$$
\begin{aligned}
&\left|\mathbb{E}e^{\frac{it(X_{nk}-\mu_{nk})}{v_n}} - e^{-\frac{t^2\sigma_{nk}^2}{2v_n^2}}\right| \\
\leq&\left|\mathbb{E}\left(1 + it(X_{nk} - \mu_{nk}) - \frac{t^2(X_{nk} - \mu_{nk})^2}{2v_n^2}\right) - \left(1 - \frac{t^2\sigma_{nk}^2}{2v_n^2}\right)\right| \\
&\quad + \mathbb{E}\left[\min\left\{\frac{t^2(X_{nk} - \mu_{nk})^2}{v_n^2}, \frac{|t(X_{nk} - \mu_{nk})|^3}{6v_n^3}\right\}\right] + \frac{t^4\sigma_{nk}^4}{8v_n^4} \\
\leq&\frac{t^2}{v_n^2}\mathbb{E}\left[X_{nk}^2 \mathbb{1}_{\{|X_{nk}-\mu_{nk}| \geq \epsilon v_n\}}\right] + \frac{|t|^3}{6v_n^3}\mathbb{E}\left[|X_{nk} - \mu_{nk}|^3 \mathbb{1}_{\{|X_{nk}-\mu_{nk}| \leq \epsilon v_n\}}\right] + \frac{t^4\sigma_{nk}^2}{8v_n^4} \\
\leq&\frac{t^2}{v_n^2}\mathbb{E}\left[(X_{nk} - \mu_{nk})^2 \mathbb{1}_{\{|X_{nk}-\mu_{nk}| \geq \epsilon v_n\}}\right] + \frac{|t|^3\epsilon\sigma_{nk}^2}{6v_n^3} + \frac{t^4\sigma_{nk}^2}{v_n^2} \max_j \frac{\sigma_j^2}{v_n^2}.
\end{aligned}
$$

For any fixed t,

$$
\begin{aligned}
&\left|\mathbb{E}e^{\frac{itS_n}{v_n}} - e^{-\frac{t^2}{2}}\right| \\
=&\left|\prod_{k=1}^n \mathbb{E}e^{\frac{it(X_{nk}-\prod_{k=1}^n \mu_{nk})}{v_n}} - e^{-\frac{t^2\sigma_{nk}^2}{2v_n^2}}\right| \\
\leq&\sum_{k=1}^n \left|\mathbb{E}e^{\frac{it(X_{nk}-\mu_{nk})}{v_n}} - e^{-\frac{t^2\sigma_{nk}^2}{2v_n^2}}\right| \\
\leq&\sum_{k=1}^n \left[\frac{t^2}{v_n^2}\mathbb{E}\left[(X_{nk} - \mu_{nk})^2 \mathbb{1}_{\{|X_{nk}-\mu_{nk}| \geq \epsilon v_n\}}\right] + \frac{|t|^3\epsilon\sigma_{nk}^2}{6v_n^3} + \frac{t^4\sigma_{nk}^2}{v_n^2} \max_j \frac{\sigma_j^2}{v_n^2}\right] \\
\leq&\frac{t^2}{v_n^2}\sum_{k=1}^n \mathbb{E}\left[(X_{nk} - \mu_{nk})^2 \mathbb{1}_{\{|X_{nk}-\mu_{nk}| \geq \epsilon v_n\}}\right] + \epsilon|t|^3 + t^4 \max_k \frac{\sigma_{nk}^2}{v_n^2} \\
\to&\epsilon|t|^3
\end{aligned}
$$

as $n \to \infty$. Let $\epsilon \to 0^+$, then $\mathbb{E}e^{\frac{itS_n}{v_n}} \to e^{-\frac{t^2}{2}}$ for all $t \in \mathbb{R}$. Therefore, $\frac{S_n}{v_n} \xrightarrow{D} N(0,1)$.

(2) Omitted. ∎