Statistics 553 : Asymptotic Tools

David R. Hunter, PSU

Summary Notes of the Course written by Jaeho, Chang

Contents

1 Mathematical and Statistical Preliminaries

1.1 Limit superior and inferior

Definition 1.1. Let S be a set of real-valued elements. We say α is a supremum of S if

$$\forall \varepsilon > 0 \;\; \exists x \in S \;\; \textit{such that} \;\; \alpha - \varepsilon < x \leq \alpha.$$

Note that $-\sup(-S) = \inf\ S$.

Definition 1.2. $\liminf, \limsup$

$$\liminf_{n\to\infty} a_n \overset{def}{=} \lim_{n\to\infty} \inf_{k\geq n} a_k$$
$$\limsup_{n\to\infty} a_n \overset{def}{=} \lim_{n\to\infty} \sup_{k\geq n} a_k$$

Lemma 1.1. *Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ be real-valued sequences. Then,*

- $\liminf_n a_n \leq \limsup_n a_n$
- $\exists \lim_n a_n < \infty$ *iff* $\liminf_n a_n = \limsup_n a_n$
- *If* $a_n \leq b_n \ \forall n$ *then* $\liminf_n a_n \leq \liminf_n b_n$ *and* $\limsup_n a_n \leq \limsup_n b_n$.
- $\liminf_n a_n = -\limsup_n(-a_n)$

Exercise 1.8

Let $X \sim Bernoulli(0.5)$. Find examples.

(a) If $a_n = 1 - \frac{1}{n}$ and $a = 1$, then $a_n \underset{n\to\infty}{\to} a$, but $\limsup_n F(a_n) \neq F(a)$.

(b) If $a_n = 1 + \left(-\frac{1}{n}\right)^n$ and $a = 1$, then $a_n \underset{n\to\infty}{\to} a$ and $\limsup_n F(a_n) = F(a)$ but $\nexists \lim_n F(a_n)$.

(c) Because F is right-continuous.

1.2 Differentiability, Taylor's Theorem

Theorem 1.1. *If f has d derivatives at a, then*

$$f(x) = \sum_{i=0}^{d} \frac{f^{(i)}(a)(x-a)^i}{i!} + r_d(x,a)$$

where $\frac{r_d(x,a)}{(x-a)^d} \to 0$ as $x \to a$. For example, if $f^{(d+1)}$ exists on the closed interval from x to a,

$$r_d(x,a) = \int_a^x \frac{f^{(d+1)}(t)(x-t)^d}{d!} dt \tag{1}$$
$$r_d(x,a) = \frac{f^{(d+1)}(c)(x-a)^{d+1}}{(d+1)!}$$

where c exists between x and a. (1) can also be expressed as

$$\frac{(x-a)^d}{(d-1)!}\int_0^1 [f^{(d)}\{xt+a(1-t)\}-f^{(d)}(a)](1-t)^{d-1}dt.$$

Theorem 1.2. *l'Hôpital's rule*
Let $\exists\, f', g'$ on $N(c,\delta)\backslash\{c\}$ for some $c\in\mathbb{R}, \delta>0$. If $\lim\limits_{x\to c} f(x)=\lim\limits_{x\to c} g(x)=0$ or $\lim\limits_{x\to c} f(x)=\lim\limits_{x\to c} g(x)=\infty$ then

$$\lim_{x\to c}\frac{f(x)}{g(x)}=\lim_{x\to c}\frac{f'(x)}{g'(x)}$$

if RHS exists.

1.3 Order Notation

Definition 1.3. Asymtotic Equivalence
We say that the sequence of real numbers $a_1, a_2, \cdots$ is asymptotically equivalent to the sequence $b_1, b_2, \cdots$, written $a_n \sim b_n$, if $\frac{a_n}{b_n}\to 1$ as $n\to\infty$. That is,

$$\left|1-\frac{a_n}{b_n}\right| \underset{n\to\infty}{\to} 0$$

Example 1.22

$$n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$$

Definition 1.4. We write $a_n = o(b_n)$ as $n\to\infty$ if $\frac{a_n}{b_n}\to 0$ as $n\to\infty$.

Definition 1.5. We write $a_n = O(b_n)$ as $n\to\infty$ if there exist $M>0$ and $N>0$ such that $\left|\frac{a_n}{b_n}\right|<M$ for all $n>N$.

For example, (1) may be rewritten as

$$f(x)=\sum_{i=0}^{d}\frac{f^{(i)}(a)(x-a)^i}{i!}+\frac{\{f^{(d+1)}(a)+o(1)\}(x-a)^{d+1}}{(d+1)!}$$

as $x\to a$.

Lemma 1.2.
Let f, g be real-valued functions and $f_1 = O(g_1)$ and $f_2 = O(g_2)$. Then,

(a) $f_1 f_2 = O(g_1 g_2)$

(b) $f_1 + f_2 = O(g_1 + g_2)$

(c) $fO(g) = O(fg)$

If $f = o(F)$ and $g = o(G)$, then

(a) $fg = o(FG)$

(b) *if $f = o(g)$, then $f = o(G)$.*

Theorem 1.3.
Let a_n,$b_n \to \infty$ as $n \to \infty$ and $a_n = o(b_n)$. If f is convex and $f(x) \to \infty$ as $x \to \infty$, then

$$f(a_n) = o[f(b_n)].$$

Proof. Pick a constant c from the domain of f. Since f is convex,

$$\begin{aligned} f\left(\frac{a_n - c}{b_n - c}b_n + \frac{b_n - a_n}{b_n - c}c\right) &\leq \frac{a_n - c}{b_n - c}f(b_n) + \frac{b_n - a_n}{b_n - c}f(c) \\ \Leftrightarrow \quad \frac{f(a_n)}{f(b_n)} &\leq \frac{a_n - c}{b_n - c} + \frac{b_n - a_n}{b_n - c}\frac{f(c)}{f(b_n)}. \end{aligned} \tag{2}$$

It is obvious that $\frac{a_n - c}{b_n - c} \to 0$, $\frac{b_n - a_n}{b_n - c} \to 1$ and $f(b_n) \to \infty$ as $n \to \infty$. Therefore taking $\liminf_n$ and $\limsup_n$ on (2) obtains

$$0 \leq \liminf_{n\to\infty} \frac{f(a_n)}{f(b_n)} \leq 0 \ \ \& \ \ 0 \leq \limsup_{n\to\infty} \frac{f(a_n)}{f(b_n)} \leq 0.$$

Therefore $f(a_n) = o[f(b_n)]$. □

Excercise 1.18

Suppose that $a_n \sim b_n$ and $c_n \sim d_n$. Then,

(a) $a_n c_n \sim b_n d_n$

(b) $|a_n| + |c_n| \sim |b_n| + |d_n|$

Proof of (b). By assumption,

$$\begin{aligned} &\forall \varepsilon > 0 \ \ \exists N_1, N_2 \ \ \textit{such that if} \ \ n > N_1 \ \ \textit{and} \ \ n > N_2 \ \ \textit{then} \\ &\varepsilon > \left|\frac{a_n}{b_n} - 1\right| \geq \left|\frac{a_n}{b_n}\right| - 1 \ \ \& \ \ \varepsilon > \left|1 - \frac{a_n}{b_n}\right| \geq 1 - \left|\frac{a_n}{b_n}\right| \\ &\textit{which means} \ \ \left|\frac{a_n}{b_n}\right| \in N(1, \varepsilon) \ \ \textit{and similarly} \ \ \left|\frac{c_n}{d_n}\right| \in N(1, \varepsilon). \\ &\textit{Let } N = \max(N_1, N_2) \ \ \textit{then if} \ n > N, \\ &\frac{|a_n| + |c_n|}{|b_n| + |d_n|} = \frac{\left|\frac{a_n}{b_n}\right| |b_n| + \left|\frac{c_n}{d_n}\right| |d_n|}{|b_n| + |d_n|} \in N(1, \varepsilon) \end{aligned} \tag{3}$$

because (3) is just an interpolation of $\left|\frac{a_n}{b_n}\right|$ and $\left|\frac{c_n}{d_n}\right|$. Therefore, the result follows. □

1.4 Multivariate Extensions

Definition 1.6. Differentiability
Suppose that $\mathbf{f} : \mathbf{U} \to \mathbb{R}^l$, where $\mathbf{U} \in \mathbb{R}^k$ is open. For $\mathbf{a} \in \mathbf{U}$, suppose there exists an $l \times k$ matrix $\mathbf{J_f}(\mathbf{a})$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - \mathbf{J_f}(\mathbf{a})\mathbf{h}}{||\mathbf{h}||} = \mathbf{0}. \tag{4}$$

Then $\mathbf{J_f}(\mathbf{a})$ is unique and we call it Jacobian matrix of $\mathbf{f}(\mathbf{x})$ at $\mathbf{a}$. $\mathbf{J_f}(\mathbf{x})$ may be called the derivative of $\mathbf{f}(\mathbf{x})$.

Definition 1.7. Let $g(\mathbf{x})$ be a real-valued function defined on a neighborhood of $\mathbf{a}$ in $\mathbf{R}^k$. For $1 \le i \le k$, let $\mathbf{e}_i$ denote the i^{th} standard basis vector in $\mathbf{R}^k$. We define the i^{th} partial derivative of $g(\mathbf{x})$ at $\mathbf{a}$ to be :

$$\left.\frac{\partial g(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{a}} = g_{x_i}(\mathbf{a}) \overset{def}{=} \lim_{h\to 0} \frac{g(\mathbf{a}+h\mathbf{e}_i) - g(\mathbf{a})}{h} \tag{5}$$

if RHS of (5) exists.

Theorem 1.4. *Suppose* $\mathbf{f}(\mathbf{x})$ *is differentiable at* $\mathbf{a}$ *in the sense of (4). Define the gradient matrix* $\boldsymbol{\nabla}\mathbf{f}(\mathbf{a})$ *to be* $\mathbf{J_f}(\mathbf{a})^\top$*. Then*

$$\boldsymbol{\nabla}\mathbf{f}(\mathbf{a}) = \left.\begin{pmatrix} f_{1x_1}(\mathbf{x}) & \cdots & f_{lx_1}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ f_{1x_k}(\mathbf{x}) & \cdots & f_{lx_k}(\mathbf{x}) \end{pmatrix}\right|_{\mathbf{x}=\mathbf{a}} \quad ;\ k \times l \text{ matrix.} \tag{6}$$

The converse of Theorem 1.4 is not true, in the sense that the existence of partial derivatives of a function does not guarantee the differentiability of that function.

Exercise 1.31

Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $f(x,y) = \frac{xy}{x^2+y^2} I(xy \neq 0)$. Then

(a) since $f(x,0) = 0$, $f_x(0,0) = 0$.

(b) Likewise, $f_y(0,0) = 0$.

But $f(x,x) = \frac{1}{2} \neq f(0,0)$ so f is not differentiable.

Excercise 1.41

Let $X_1, \cdots, X_n \overset{ind}{\sim} (\mu_i, \sigma_i^2)$ where $\sigma_i^2 < \infty$ for $i = 1, \cdots, n$. Let

$$S_k = \sum_{i=1}^{k} (X_i - \mu_i) \tag{7}$$

for $k = 1, \cdots, n$.

(a) *Proof.* Let

$$A_k = \{\omega \in \Omega\,; |S_k(\omega)| \geq a, |S_j(\omega)| < a \ \forall j < k\} = \{\cap_{i=1}^{k-1}(|S_i| < a)\} \cap (|S_k| \geq a) \tag{8}$$
$$= \{\cap_{i=1}^{k-1}(\sim B_i)\} \cap B_k \quad and \quad A_1 = (|S_1| \geq a) = B_1. \tag{9}$$

It is trivial that $A_1 \cap A_2 = \emptyset$, and

$$A_k \cap A_{k+1} = \{(B_k \sim \cup_{i=1}^{k-1} B_i) \cap B_{k+1}\} \cap \{(B_k \sim \cup_{i=1}^{k-1} B_i) \sim (\cup_{i=1}^{k} B_i)\} = \emptyset \tag{10}$$

for any k, so $A_i \cap A_j = \emptyset$ for any $i \neq j$. This leads to

$$\sum_{i=1}^{n} EI(A_i) = EI(\cup_{i=1}^{n} A_i) = EI(\cup_{i=1}^{n} B_i) = EI[\sim \cap_{i=1}^{n}(\sim B_i)] = P\left(\max_{1\leq i\leq n}|S_i| \geq a\right). \tag{11}$$

Now, $E\{I(A_k)S_k^2\} \geq a^2 EI(A_k)$ because

$$I(A_k) \leq \frac{S_k^2}{a^2} I(A_k). \tag{12}$$

Finally we get the results :

$$a^2 P\left(\max_{1\leq i\leq n}|S_i| \geq a\right) \leq \sum_{i=1}^{n} E\{I(A_k)S_k^2\}. \tag{13}$$

□

(b) *Proof.* Since A_is are mutually disjoint, $\sum_{i=1}^{n} I(A_i) \leq 1$, which leads to ;

$$S_n^2 \geq \sum_{i=1}^{n} S_n^2 I(A_i), \quad therefore \quad E(S_n^2) \geq \sum_{i=1}^{n} E[I(A_i)\{S_i^2 + 2S_i(S_n - S_i) + (S_n - S_i)^2\}]$$
$$\geq \sum_{i=1}^{n} E[I(A_i)\{S_i^2 + 2S_i(S_n - S_i)\}] \tag{14}$$

□

(c) Since $I(A_i)S_i$ and $S_n - S_i$ are independent for $n \neq i$ and $E(S_n - S_i) = 0$, combining (13) and (14) obtains $a^2 P\left(\max_{1\leq i\leq n}|S_i| \geq a\right) \leq \sum_{i=1}^{n} E\{I(A_i)S_i^2\} \leq E(S_n^2) = V(S_n)$.

Exercise 1.43

(a) Since

$$\cos x - 1 = r_0(x, 0) = -\int_0^x \sin u du \quad and \quad \sin x = 0 + \int_0^x \cos u du, \tag{15}$$

$$|e^{it} - 1| = \left|\int_0^t (-\sin u + i\cos u) du\right| \leq \int_0^t |-\sin u + i\cos u| du = t \leq |t|. \tag{16}$$

(b) Since

$$\cos x - 1 = \int_0^x (x-u)\cos u du \quad and \quad \sin x = x - \int_0^x (x-u)\sin u du, \tag{17}$$

$$|e^{it} - 1 - it| = \left|\int_0^t (t-u)(\cos u - i\sin u)du\right| \le \int_0^t (t-u)du = \frac{t^2}{2}. \tag{18}$$

(c)

$$\left|e^{it} - 1 - it + \frac{t^2}{2}\right| = \left|\int_0^t \frac{-ie^{iu}(t-u)^2}{2}du\right| \le \int_0^t \frac{|t-u|^2}{2}du \le \frac{|t|^3}{6} \tag{19}$$

(d) Since

$$i\sin t = it - i\int_0^t (t-u)\sin u du \quad and \quad \cos t = 1 - \frac{t^2}{2} + \int_0^t \frac{(t-u)^2\sin u}{2}du \tag{20}$$

$$= 1 - \frac{t^2}{2} - \frac{(t-u)^2}{2}\cos u\Big|_0^t - \int_0^t (t-u)\cos u du = 1 - \int_0^t (t-u)\cos u du, \tag{21}$$

$$\left|e^{it} - 1 - it + \frac{t^2}{2}\right| = \left|\frac{t^2}{2} - \int_0^t (t-u)(\cos u - i\sin u)du\right| \le \frac{t^2}{2} + \int_0^t (t-u)du = t^2. \tag{22}$$

2 Weak Convergence

2.1 Modes of Convergence

2.1.1 Probabilistic Order Notation

Definition 2.1. We write $X_n = o_P(Y_n)$ if $\frac{X_n}{Y_n} \xrightarrow{P} 0$ as $n \to \infty$.

For example, if $X_n \xrightarrow{P} \theta$ then $X_n - \theta = o_P(1)$.

Definition 2.2. We write $X_n = O_P(Y_n)$ if for any $\varepsilon > 0$ there exist M and N such that

$$P\left(\left|\frac{X_n}{Y_n}\right| < M\right) > 1 - \varepsilon \quad \forall n > N. \tag{23}$$

If $X_n = O_P(1)$, we say that X_n is *bounded in probability.*

Theorem 2.1. *Suppose that $X_n \xrightarrow{P} \theta_0$ for a sequence of random variables $X_1, X_2, \cdots$ and a constant θ_0. Furthermore, suppose that f has d derivatives at the point θ_0. Then there is a random variable Y_n such that*

$$f(X_n) = \sum_{i=0}^{d-1} \frac{f^{(i)}(\theta_0)(X_n - \theta_0)^i}{i!} + \frac{\{f^{(d)}(\theta_0) + Y_n\}(X_n - \theta_0)^d}{d!} \tag{24}$$

and $Y_n = o_P(1)$ as $n \to \infty$.

2.1.2 Convergence in Distribuion (Law)

Any distribution function $F(x)$ is nondecreasing and right-continuous, and it has limits 0 and 1 for each cases $x \to -\infty$ and $x \to \infty$.

Definition 2.3. Suppose that X has distribution function $F(x)$ and that X_n has distribution function $F_n(x)$. We say X_n converges in distribution to X, written $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{\mathcal{L}} X$, if $F_n(x) \to F(x)$ as $n \to \infty$ for all x at which $F(x)$ is continuous.

For example, for $X_1, \cdots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$ with finite σ^2 we have $\bar{X}_n - \mu = O_P\left(\frac{1}{\sqrt{n}}\right)$. $(\because)$ $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$ as $n \to \infty$.

Theorem 2.2. *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.*

Theorem 2.3. *$X_n \xrightarrow{d} c$ iff $X_n \xrightarrow{P} c$.*

The reverse of Theorem 2.3 is true when X_n and X are defined on the same sample space for every n.

2.1.3 Convergence in Mean

Definition 2.4. Let a be a postive constant. We say that X_n converges in a^{th} mean to X, written $X_n \xrightarrow{a} X$, if

$$E|X_n - X|^a \to 0 \quad as \quad n \to \infty. \tag{25}$$

Especially, we denote the case when $a = 2$ in the above expression by $X_n \xrightarrow{qm} X$.

Theorem 2.4.
(a) For a constant c, $X_n \xrightarrow{qm} c$ iff $E(X_n) \to c$ and $V(X_n) \to 0$.
(b) For fixed $a > 0$, $X_n \xrightarrow{a} X$ implies $X_n \xrightarrow{P} X$.

Excercise 2.2

(a) If $X_n \xrightarrow{d} X$ then $X_n = O_P(1)$.

Proof. $\forall \varepsilon > 0$, $\exists\, a, b$ such that $F(a) < \frac{\varepsilon}{4}$ and $F(b) > 1 - \frac{\varepsilon}{4}$. Since $X_n \xrightarrow{d} X$, $\exists\, N_1, N_2$ such that if $n > \max(N_1, N_2)$ then

$$|F_n(a) - F(a)| < \varepsilon/4$$
$$|F_n(b) - F(b)| > 1 - \varepsilon/4$$

which implies that $F_n(a) < \varepsilon/2$ and $F_n(b) > 1 - \varepsilon/2$. Let $M = \max(|a|, |b|)$ then $P(|X_n| \leq M) \geq F_n(b) - F_n(a) > 1 - \varepsilon$. $\square$

(b) If $X_n = O_P(1)$ and $Y_n \xrightarrow{P} 0$, then $X_n Y_n \xrightarrow{P} 0$.

Proof. For any $\varepsilon > 0$, $\exists\, M > 0$ and $\exists\, N_1, N_2$ such that if $n > \max(N_1, N_2)$ then $P(|X_n| < M) > 1 - \varepsilon/2$ and $P(|Y_n| < \varepsilon/M) > 1 - \varepsilon/2$.

$$
\begin{aligned}
P(|X_nY_n| < \varepsilon) \geq & P(|X_n| < M, |Y_n| < \varepsilon/M) \\
& = P(|X_n| < M) + P(|Y_n| < \varepsilon/M) - P(|X_n| < M \textit{ or } |Y_n| < \varepsilon/M) > 1 - \varepsilon
\end{aligned}
$$

□

Exercise 2.3 (a)

Let $X_n \sim B(n, \lambda/n)$ and $X \sim Pois(\lambda)$ then $X_n \xrightarrow{d} X$.

Proof. It suffices to prove that $p_{X_n}(x) \to p_X(x)$ for any $x = 0, 1, 2, \cdots$.

$$
p_{X_n}(x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x(1-\lambda/n)^n}{x!}\left(1 - \frac{\lambda}{n}\right)^{-x} \prod_{i=0}^{x-1}\left(1 - \frac{i}{n}\right) \tag{26}
$$

□

Excercise 2.4 (c)

Let $X_1, \cdots, X_n \overset{iid}{\sim} Exp(1)$ with $F(t) = 1 - e^{-t}$. Then $Y_n \xrightarrow{d} Y$ where $Y_n \sim Bin(n, F(c/n))$ and $Y \sim Pois(c)$ for $c > 0$.

Proof.

$$
\begin{aligned}
P_{Y_n}(y) = & \frac{n!}{y!(n-y)!}\left(1 - e^{-c/n}\right)^y (e^{-c/n})^{n-y} = \frac{e^{-c}}{y!} \times \frac{n!(e^{c/n} - 1)^y}{(n-y)!} = \frac{e^{-c}}{y!}\prod_{i=0}^{y-1}\{(n-i)(e^{c/n} - 1)\} \\
= & \frac{e^{-c}}{y!}\prod_{i=0}^{y-1}\left[(n-i)\left\{\frac{c}{n} + o\left(\frac{1}{n}\right)\right\}\right] = \frac{e^{-c}}{y!}\prod_{i=0}^{y-1}\{c + o(1)\} \quad as \quad n \to \infty
\end{aligned}
$$

because $(n-i)o(1/n) = o(1) + o(1/n) = o(1)$ as n goes to ∞. □

Exercise 2.8

If there exists M such that $P(|X_n| < M) = 1$ for all n, then $X_n \xrightarrow{P} c$ implies $X_n \xrightarrow{qm} c$.

Proof. For any $\varepsilon > 0$ and M given in assumption, $\exists\, N$ such that if $n > N$ then $P(|X_n - c| > \varepsilon) < \varepsilon$. Next,

$$
\begin{aligned}
(X_n - c)^2 = & (X_n - c)^2 I(|X_n - c| > \varepsilon) + (X_n - c)^2 I(|X_n - c| \leq \varepsilon) \\
& \leq (M + |c|)^2 I(|X_n - c| > \varepsilon) + \varepsilon^2 I(|X_n - c| \leq \varepsilon)
\end{aligned} \tag{27}
$$

because $EI(|X_n| < M) = 1 \Leftrightarrow I(|X_n| < M) = 1$, which leads to

$$
|X_n - c|^2 \leq (|X_n| + |c|)^2 I(|X_n| < M) < (M + |c|)^2. \tag{28}
$$

Now, taking expectations on (27) obtains

$$
\begin{aligned}
E(X_n - c)^2 \leq & (M + |c|)^2 P(|X_n - c| > \varepsilon) + \varepsilon^2 P(|X_n - c| \leq \varepsilon) \\
& < (M + |c|)^2\varepsilon + \varepsilon^2.
\end{aligned} \tag{29}
$$

□

2.2 Consistent Estimates of the Mean

Theorem 2.5. *WLLN*
Suppose that $X_1, X_2, \cdots \overset{iid}{\sim} f(\cdot)$ *with finite mean* μ. *Then* $\bar{X}_n \xrightarrow{P} \mu$.

2.2.1 Independent but not Identically Distributed Variables

Let $X_1, X_2, \cdots$ have the same mean but different variances. Since $\bar{X}_n$ is unbiased, if $V(\bar{X}_n)$ tends to 0 as $n \to \infty$, it is consistent. That is,

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{if} \quad Var(\bar{X}_n)n^2 = \sum_{i=1}^{n} \sigma_i^2 = o(n^2). \tag{30}$$

Alternative for sample mean :

$$\hat{\mu}_n = \frac{\sum_{i=1}^n c_i X_i}{\sum_{i=1}^n c_i} \tag{31}$$

for some sequence of positive constants $c_1, c_2, \cdots$. $\hat{\mu}_n$ is unbiased so we now investigate if its variance tends to zero as $n \to \infty$.

$$V(\hat{\mu}_n) = \frac{\sum_{i=1}^n c_i^2 \sigma_i^2}{(\sum_{i=1}^n c_i)^2} := \sum_{i=1}^{n} \gamma_i^2 \sigma_i^2 \tag{32}$$

The minimizer of $V(\hat{\mu}_n)$ with respect to $\gamma_1, \cdots, \gamma_{n-1}, \gamma_n (= 1 - \sum_{i=1}^{n-1} \gamma_i)$ obtains the equations :

$$\gamma_i \sigma_i^2 = \gamma_n \sigma_n^2, \quad i \in \{1, \cdots, n-1\}, \tag{33}$$

which means $c_i = \frac{c_n \sigma_n^2}{\sigma_i^2}$. Therefore

$$\hat{\mu}_n = \frac{\sum_{i=1}^n \frac{X_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \tag{34}$$

whose variance is

$$V(\hat{\mu}_n) = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \tag{35}$$

2.2.2 Identically Distributed but not Independent Variables

Suppose that $X_1, X_2, \cdots$ have the same mean and variance, say μ and σ^2, but that they are not necessarily independent. We still have $E(\bar{X}_n) = \mu$ so it's consistent if its variance tends to zero.

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Var(X_i, X_j) = \frac{\sigma^2}{n} + \frac{2}{n^2} \sum\sum_{i<j} Var(X_i, X_j) \tag{36}$$

Now if we add the 'exchangeable' assumption, $Var(X_i, X_j) = Var(X_1, X_2)\ \forall i \neq j$. Therefore (36) reduces to

$$V(\bar{X}_n) = \frac{\sigma^2}{n} + \frac{n-1}{n} Var(X_1, X_2) \tag{37}$$

which implies that $Var(\bar{X}_n) \to Var(X_1, X_2)$ as $n \to \infty$.

Definition 2.5. The sequence $X_1, X_2, \cdots$ is said to be stationary if, for a fixed $k \geq 0$, the joint distribution of $(X_i, \cdots, X_{i+k})$ is the same no matter what positive value of i is chosen.

Lemma 2.1. *Under stationary sequence assumption,*

$$(36) = \frac{\sigma^2}{n} + \frac{2}{n^2} \sum_{k=1}^{n-1} (n-k) Var(X_1, X_{1+k}) \tag{38}$$

and this tends to 0 as $n \to \infty$ if $\sigma^2 < \infty$ and $Var(X_1, X_{1+k}) \to 0$ as $k \to \infty$.

Proof. $\forall \varepsilon > 0$ take K such that if $k > K$ then $|Var(X_1, X_{1+k})| < \varepsilon$. Then

$$\begin{aligned} |(38)| &\leq \frac{\sigma^2}{n} + \frac{2}{n^2} \sum_{k=1}^{n-1} (n-k) |Var(X_1, X_{1+k})| \\ &= \frac{\sigma^2}{n} + \frac{2}{n} \sum_{k=1}^{K} |Var(X_1, X_{1+k})| + \frac{2}{n} \sum_{k=K+1}^{n-1} |Var(X_1, X_{1+k})| \\ &< \frac{\sigma^2}{n} + \frac{2}{n} \sum_{k=1}^{K} |Var(X_1, X_{1+k})| + \frac{2\varepsilon}{n}(n-K-1). \end{aligned} \tag{39}$$

(39) tends to 2ε as $n \to \infty$ and obtains the result. □

Definition 2.6. For $m \geq 0$, the sequence $X_1, X_2, \cdots$ is called m-dependent if the random vectors $(X_1, \cdots, X_i)$ and $(X_j, X_{j+1}, \cdots)$ are independent whenever $j - i > m$.

Any stationary m-dependent sequence with finite variance is consistent. Also, any independent sequence is 0-dependent.

Exercise 2.11

Since $V(\bar{X}_n) = \sum_{i=1}^n \frac{\sigma_i^2}{n^2}$ and $V(\delta_n) = \frac{1}{\sum_{i=1}^n \sigma_i^{-2}}$,

$$\frac{V(\bar{X}_n)}{V(\delta_n)} = \frac{\sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n \sigma_i^{-2}}{n^2} = \frac{\sum\sum_{1 \leq i,j \leq n} \frac{\sigma_i^2}{\sigma_j^2}}{n^2} = \frac{n + \sum\sum_{i \neq j} \frac{\sigma_i^2}{\sigma_j^2}}{n^2} > \frac{n + n(n-1)}{n^2} \tag{40}$$

so $\liminf_n V(\bar{X}_n)/V(\delta_n) > 1$.

Exercise 2.13

Let $Y_1, Y_2, \cdots \overset{iid}{\sim} (\mu, \sigma^2)$ and $\sigma^2 < \infty$. Let δ_n be defined as (34). From the definition of X_k, $E(X_k) = \mu$ and $V(X_k) = \frac{\sigma^2}{k}$. Since Y_i's are independent, so does X_i's.

(a) Since $V(\bar{X}_n) = \frac{\sigma^2}{n^2} \sum_{k=1}^n \frac{1}{k}$ and $Var(\delta_n) = \frac{1}{\sum_{i=1}^n \frac{i^2}{\sigma^2}}$, both are consistent estimators for μ.

2.3 Convergence of Transformed Sequences

2.3.1 Continuous Transformation : The Univariate Case

Theorem 2.6. *Let f be a continuous function. Then*

(a) *If $X_n \xrightarrow{P} X$, then $f(X_n) \xrightarrow{P} f(X)$.*

(b) *If $X_n \xrightarrow{d} X$, then $f(X_n) \xrightarrow{d} f(X)$.*

Theorem 2.7. *Helley-Bray theorem*
*$X_n \xrightarrow{d} X$ iff $Eg(X_n) \to Eg(X)$ for **all** bounded continuous real-valued function g.*

Proof. Theorem 2.6 (b)
It sufficies to show that $Eg\{f(X_n)\} \to E\{f(X)\}$. Since f is continuous, $g \circ f$ is also bounded and continuous. Therefore we get the result. □

2.3.2 Multivariate Extensions

Definition 2.7. For $a > 0$ $\mathbf{X}_n$ converges in a^{th} mean to $\mathbf{X}$ if

$$E||\mathbf{X}_n - \mathbf{X}||^a \to 0 \tag{41}$$

as $n \to \infty$.

Exercise 2.15

If $X_n \xrightarrow{d} X$ and g is a bounded and continuous real-valued function, then

$$Eg(X_n) \to Eg(X). \tag{42}$$

Proof. For any $\varepsilon > 0$, $\exists\, b < c$ such that $F(b) < \varepsilon$ and $F(c) > 1 - \varepsilon$. Since g is bounded and continuous, $\exists\, \delta$ such that $|g(x) - g(y)| < \varepsilon$ whenever $|x - y| < \delta$, and $\exists$ finite set of real numbers $b = t_0 < t_1 < \cdots < t_m = c$ such that

(1) Each t_i is a point of continuity of F.

(2) $F(t_0) < \varepsilon$ and $F(t_m) > 1 - \varepsilon$

(3) For $i = 1, \cdots, m$, $|g(x) - g(t_i)| < \varepsilon$ for any $x \in [t_{i-1}, t_i]$.

(a) Define

$$h(x) = g(t_i) I_{(t_{i-1}, t_i]}(x). \tag{43}$$

Then,

$$|Eh(X_n) - Eh(X)| = \left| \sum_{i=1}^{m} g(t_i)\{F_n(t_i) - F_n(t_{i-1}) - F(t_i) + F(t_{i-1})\} \right|. \tag{44}$$

If $n > N := \max(N_1, \cdots, N_m)$ such that $n > N_i \quad \rightarrow \quad |F_n(t_i) - F(t_i)| < \frac{\varepsilon}{2m \sup_x |g(x)|}$,

$$(44) \leq \sum_{i=1}^{m} |g(t_i)\{F_n(t_i) - F(t_i)\}| + \sum_{i=1}^{m} |g(t_i)\{F(t_{i-1}) - F_n(t_{i-1})\}| < \varepsilon. \tag{45}$$

(b) Note that

$$|Eg(X_n) - Eg(X)| \leq |Eg(X_n) - Eh(X_n)| + |Eh(X_n) - Eh(X)| + |Eh(X) - Eg(X)|. \tag{46}$$

Since $|g(x) - h(x)| < \varepsilon$ for any $x \in (t_{i-1}, t_i]$ and $i = 1, \cdots, m$,

$$|Eg(X_n) - Eh(X_n)| I_{[b,c]}(X_n) \leq \varepsilon \cdot 1. \tag{47}$$

Furthermore, since $F_n(b) \rightarrow F(b) < \varepsilon$ and $F_n(c) \rightarrow F(c) > 1 - \varepsilon$, $\exists\, N'$ such that if $n > N'$, then $P(X_n \leq b \text{ } or \text{ } X_n > c) < 4\varepsilon$.

$$(\because) \quad P(X_n \leq b \text{ } or \text{ } X_n > c) \leq F_n(b) + 1 - F_n(c) < 4\varepsilon. \tag{48}$$

$$(\therefore) \quad |Eg(X_n) - Eh(X_n)| I(X_n \leq b \text{ } or \text{ } X_n > c) < 4\varepsilon \sup_x g(x) \quad and \tag{49}$$

$$|Eg(X) - Eh(X)| I(X \leq b \text{ } or \text{ } X > c) < 2\varepsilon \sup_x g(x) \tag{50}$$

$$(\therefore) \quad |Eg(X_n) - Eg(X)| < 6\varepsilon \sup_x g(x) + \varepsilon \quad for \quad n > \max(N, N'). \tag{51}$$

□

Exercise 2.16

(a) Since F is continuous on t, $\forall\, \varepsilon > 0\, \exists\, \delta_0$ such that $|F(x) - F(y)| < \varepsilon$ whenever $|x - y| < \delta_0$ where x, y are points of continuity. If we take $0 < \delta < \delta_0$, then results follows.

(b) Define $g_1, g_2\, ; \mathbb{R} \rightarrow [0, 1]$ such that

$$g_1(x) = \begin{cases} 0 & if \quad x > t + \delta \\ 1 & if \quad x \leq t \\ -\frac{1}{\delta}(x - t - \delta) & o.w. \end{cases}, \quad g_2(x) = \begin{cases} 0 & if \quad x > t \\ 1 & if \quad x \leq t - \delta \\ -\frac{1}{\delta}(x - t) & o.w. \end{cases}. \tag{52}$$

Then $g_1(x) = g_2(x - \delta) = 1$ for all $x \leq t$ and $g_1(x + \delta) = g_2(x) = 0$ for all $x > t$. Since

$$I(x \leq t - \delta) \leq g_2(x) \leq I(x \leq t) \leq g_1(x) \leq I(x \leq t + \delta), \tag{53}$$

$$Eg_2(X_n) \leq F_n(t) \leq Eg_1(X_n). \tag{54}$$

and $\exists\, \delta > 0$ so

$$F(t) - \varepsilon < F(t - \delta) \leq Eg_2(X) \leq F(t) \leq Eg_1(X) \leq F(t + \delta) < F(t) + \varepsilon \tag{55}$$

$\forall\, \varepsilon > 0$. Therefore,

$$Eg_2(X_n) - Eg_1(X) \le F_n(t) - F(t) \le Eg_1(X_n) - Eg_2(X). \tag{56}$$

Since $Eg_j(X_n) \to Eg_j(X)$ for $j = 1, 2$,

$$-2\varepsilon < Eg_2(X) - Eg_1(X) \le \limsup_{n\to\infty}\{F_n(t) - F(t)\} \le Eg_1(X) - Eg_2(X) < 2\varepsilon. \tag{57}$$

Exercise 2.17

Let $Z \sim N(0,1)$ and

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1-\frac{1}{n} \\ 1-\frac{1}{n} & 1 \end{pmatrix}\right]. \tag{58}$$

Then $(X_n, Y_n)^\top \xrightarrow{d} (Z, Z)^\top$ and for $S = \{(x,y) \in \mathbb{R}^2 | x = y\}$ every point is a continuity point of $(Z,Z)^\top$. But $P[(X_n, Y_n) \in S](=0) \neq P[(Z,Z) \in S](=1)$.

Exercise 2.21

Counter-example which argues that $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{P} \mathbf{Y}$ but NOT $\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$.

Proof. Let $\mathbf{X}_n \sim N_2(\mathbf{0}_2, 1/n \times \mathbf{I}_2)$, $\mathbf{Y}_n := -\mathbf{X}_n$. Then $\mathbf{X}_n \xrightarrow{d} \mathbf{0_2}$ and $\mathbf{Y} \xrightarrow{P} \mathbf{0_2}$, however $(\mathbf{X}_n, \mathbf{Y}_n)^\top$ does not converge in law to $\mathbf{0}_4$ but $\mathbf{0}_2$ becuase $\mathbf{Y}_n \stackrel{d}{=} \mathbf{X}_n$. □

Exercise 2.22

(a) Define B_δ as

$$B_\delta = \left\{\mathbf{x} \in \mathbb{R}^k \mid \exists\, \mathbf{y} \in \mathbb{R}^k : \ ||\mathbf{x} - \mathbf{y}|| < \delta,\, ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})|| > \varepsilon\right\} \tag{59}$$

for fixed arbitrary $\varepsilon > 0$ and any $\delta > 0$. Now, we can infer that

$$\Pr\left(\big|\big|\mathbf{f}(\mathbf{X}_n) - \mathbf{f}(\mathbf{X})\big|\big| > \varepsilon\right) \le \Pr\left(||\mathbf{X}_n - \mathbf{X}|| \ge \delta\right) + \Pr(\mathbf{X} \in B_\delta) \tag{60}$$

because the RHS of (60) is always true given the LHS of (60) in the view of sets. On the RHS, the first term converges to zero as $n \to \infty$ for any $\delta > 0$, by the definition of convergence in probability of $\mathbf{X}_n$. The second term converges to zero as $\delta \to 0$ since the set B_δ shrinks to $\emptyset$. Therefore, the conclusion is that

$$\lim_{n\to\infty} \Pr\left(\big|\big|\mathbf{f}(\mathbf{X}_n) - \mathbf{f}(\mathbf{X})\big|\big| > \varepsilon\right) = 0. \tag{61}$$

(b) Since $f(\mathbf{X}_n) = \mathbf{1}^\top \mathbf{X}_n$, if we take $\mathbf{X}_n = (X, X)^\top$ and $\mathbf{X} = (X, -X)^\top$ where $X \sim N(0,1)$, it is trivial that $f(\mathbf{X}_n) \xrightarrow{d} 2X \sim N(0,4)$ but $f(\mathbf{X}) = 0$.

Excercise 2.23

Proof. Put F_n, F, G_n, G as distribution functions of X_n, X, Y_n, Y.

$$\forall\, \varepsilon > 0 \;\; \exists\, N_1, N_2 \;\; such \;\; that \;\; if \;\; n > \max(N_1, N_2), \; then \tag{62}$$
$$|F_n(x) - F(x)| < \sqrt{\varepsilon + 1} - 1 \;\; and \;\; |G_n(y) - G(y)| < \sqrt{\varepsilon + 1} - 1 \tag{63}$$

for x, y on which F, G are continuous each. Then,

$$|F_n(x)G_n(y) - F(x)G(y)| = |\{F_n(y) - F(y)\}\{G_n(y) - G(y)\} + F(x)\{G_n(y) - G(y)\} \tag{64}$$
$$+ G(y)\{F_n(x) - F(x)\}| \leq (\sqrt{\varepsilon + 1} - 1)^2 + 2\sqrt{\varepsilon + 1} - 2 = \varepsilon \tag{65}$$

for $(x, y) \in \{(a, b) \in \mathbb{R}^2 : F(a) \; and \; G(b) \; is \; continuous.\}$.
Since $F(x)G(y) = P(X \leq x, Y \leq y)$, the result follows. □

Exercise 2.24 (Slutsky's Theorem)

For some $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, denote $a_i \leq b_i$ for all $i = 1, \cdots, k$ by $\mathbf{a} \leq \mathbf{b}$.

(a) Let $\mathbf{V}_n$ and $\mathbf{W}_n$ be k-dimensional random vectors on the same sample space.

$$If \;\; \mathbf{V}_n \xrightarrow{d} \mathbf{V} \;\; and \;\; \mathbf{W}_n \xrightarrow{P} \mathbf{0}, \;\; then \;\; \mathbf{V}_n + \mathbf{W}_n \xrightarrow{d} \mathbf{V}. \tag{66}$$

Proof. Since $P(A \sim B) \geq P(A) - P(B)$ for any set A, B,

$$P(\mathbf{V}_n \leq \mathbf{a} - \boldsymbol{\varepsilon}) - P(||\mathbf{W}_n|| \geq \varepsilon) \leq P(\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a}) \leq$$
$$P(\mathbf{V}_n \leq \mathbf{a} + \boldsymbol{\varepsilon}) + P(||\mathbf{W}_n|| \geq \varepsilon) \tag{67}$$

because

$$(\mathbf{V}_n \leq \mathbf{a} - \boldsymbol{\varepsilon}) \sim (||\mathbf{W}_n|| \geq \epsilon) \subset (\mathbf{V}_n \leq \mathbf{a} - \boldsymbol{\varepsilon}) \sim (\mathbf{W}_n \geq \epsilon) \subset (\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a}). \tag{68}$$

And $(\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a}) \subset (\mathbf{V}_n \leq \mathbf{a} + \boldsymbol{\varepsilon}) \cup (||\mathbf{W}_n|| \geq \varepsilon)$ since one of two statements in the RHS must be true given that $(\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a})$. Taking $\liminf$ and $\limsup$ on both sides of (67) obtains

$$F(\mathbf{a} - \boldsymbol{\varepsilon}) \leq \liminf_{n \to \infty} P(\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a}) \leq \limsup_{n \to \infty} P(\mathbf{V}_n + \mathbf{W}_n \leq \mathbf{a}) \leq F(\mathbf{a} + \boldsymbol{\varepsilon}) \tag{69}$$

for any $\varepsilon > 0$. □

(b) Use (a).

3 Strong Convergence

3.1 Strong Consistency Defined

If $X_n(\omega_0) \to X(\omega_0)$ as $n \to \infty$, $\omega_0 \in \{\omega \in \Omega\,; X_n(\omega) \to X(\omega)\}$.

Definition 3.1. Let X and $\{X_i\}_{i=1}^{\infty}$ be defined on Ω. If $P(\{\omega \in \Omega\,; X_n(\omega) \to X(\omega)\}) = 1$, then X_n is said to converge almost surely to X, denoted $X_n \overset{a.s.}{\to} X$ or $X_n \to X$ w.p. 1.

3.1.1 Strong Consistency vs. Consistency

Let X and $\{X_i\}_{i=1}^{\infty}$ be defined on Ω. For given $\varepsilon > 0$ and n, define

$$A_n = \{\omega \in \Omega\,; |X_k(\omega) - X(\omega)| < \varepsilon \ \ \forall k \geq n\} \ \ \textit{and} \tag{70}$$

$$B_n = \{\omega \in \Omega\,; |X_n(\omega) - X(\omega)| < \varepsilon\}. \tag{71}$$

Then

(a) $A_n \subset B_n$ and $P(A_n) \leq P(B_n)$.

(b) If $P(A_n) \to 1$, then $P(B_n) \to 1$.

Theorem 3.1. *$P(A_n) \to 1$ $\forall \varepsilon > 0$ iff $X_n \overset{a.s.}{\to} X$.*

Corollary 3.1. *If $X_n \overset{a.s.}{\to} X$, then $X_n \overset{P}{\to} X$.*

The converse of Corollary 3.1 is not true(see pg.72.).

Example 3.5

Take Ω to be $(0, 1]$ and for any interval $J = (a, b] \subset \Omega$, take $P(J) = b - a$. Let

$$J_{m^2+1} \ \ \textit{through} \ \ J_{(m+1)^2} = \left(0, \frac{1}{2m+1}\right], \cdots, \left(\frac{2m}{2m+1}, 1\right] \tag{72}$$

where $m = 0, 1, \cdots$. This implies that

$$P(J_n) = \frac{1}{2m+1} \ \ \textit{where} \ \ m = \lfloor \sqrt{n-1} \rfloor. \tag{73}$$

Define $X_n = I(J_n)$ and take $\varepsilon \in (0, 1)$, then $P(|X_n| < \varepsilon) = P\{I(J_n) < \varepsilon\} = 1 - P\{I(J_n) \geq \varepsilon\} = 1 - P\{I(J_n) = 1\} = 1 - P(J_n) = 2m/(2m+1)$. Since this probability tends to 1 as $n \to \infty$, $X_n \overset{P}{\to} 0$.

However, although $I(J_n) \overset{P}{\to} 0$, J_ns repeatedly cover the entire interval $(0, 1]$ so

$$\{\omega \in \Omega = (0, 1]\,;\, X_n \to 0\} = \emptyset. \tag{74}$$

Therefore, it is not true that $X_n \overset{a.s.}{\to} 0$.

3.1.2 Multivariate Extensions

Definition 3.2. $\mathbf{X}_n$ is said to converge a.s. to $\mathbf{X}$ if

$$P(\omega \in \Omega\,; \mathbf{X}_n \to \mathbf{X}) = 1 \tag{75}$$

or $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(||\mathbf{X}_n - \mathbf{X}|| < \varepsilon \;\; \forall k > n) = P\left(\liminf_{n\to\infty}\{||\mathbf{X}_n - \mathbf{X}|| < \varepsilon\}\right) = 1. \tag{76}$$

Theorem 3.2.

(a) Let $\mathbf{f} : S \to \mathbb{R}^l$ *is a continuous function defined on some subset* $S \subset \mathbb{R}^k$*, and the range of* $\mathbf{X}$ *and of each* $\mathbf{X}_n$ *is contained in* S *w.p. 1. If* $\mathbf{X}_n \stackrel{a.s.}{\to} \mathbf{X}$*, then* $\mathbf{f}(\mathbf{X}_n) \stackrel{a.s.}{\to} \mathbf{f}(\mathbf{X})$*.*

(b) $\mathbf{X}_n \stackrel{a.s.}{\to} \mathbf{X}$ *iff* $X_{n,j} \stackrel{a.s.}{\to} X_j \; \forall j$*.*

Summary Diagram

$$\begin{array}{ccccccccccc} & & \mathbf{X}_n \stackrel{qm}{\to} \mathbf{X} & & & & & & \mathbf{X}_n \stackrel{qm}{\to} \mathbf{c} & & \\ & & \Downarrow & & & & & & \Downarrow & & \\ \mathbf{X}_n \stackrel{a.s.}{\to} \mathbf{X} & \Rightarrow & \mathbf{X}_n \stackrel{P}{\to} \mathbf{X} & \Rightarrow & \mathbf{X}_n \stackrel{d}{\to} \mathbf{X} & \quad & \mathbf{X}_n \stackrel{a.s.}{\to} \mathbf{c} & \Rightarrow & \mathbf{X}_n \stackrel{P}{\to} \mathbf{c} & \Rightarrow & \mathbf{X}_n \stackrel{d}{\to} \mathbf{c} \end{array} \tag{77}$$

Excercise 3.1

Let $S = \{\omega \in \Omega\,; X_n(\omega) \to X(\omega)\}$.

(a) With A_n defined as in (70), show that

$$\omega_0 \in \cup_{n=1}^{\infty} A_n \;\; \forall \varepsilon > 0 \quad \Leftrightarrow \quad \omega_0 \in S. \tag{78}$$

Proof. $(\Leftarrow)$

$$\omega_0 \in S \quad \Rightarrow \quad \forall \varepsilon > 0 \;\; \exists\, N \;\; such \;\; that \;\; |X_n(\omega_0) - X(\omega_0)| < \varepsilon \;\; \forall n \geq N$$
$$(\therefore) \;\; \omega_0 \in A_N \subset \cup_{n=1}^{\infty} A_n.$$

$(\Rightarrow)$

$$\forall\, n, \varepsilon > 0 \;\; if \;\; a \in A_n, \;\; then \;\; a \in A_{n+1}.$$
$$(\therefore) \;\; if \;\; \omega_0 \in \cup_{n=1}^{\infty} A_n \left(= \lim_{n\to\infty} A_n\right), then \;\; \omega_0 \in S.$$

□

(b) Prove Theorem 3.1.

Proof. $(\Rightarrow)$
By assumption, $\forall\, \varepsilon > 0$, $\lim_n P(A_n) = 1$. And this equals $P(\cup_{n=1}^{\infty} A_n) = P(S)$ by (78) and the continuity of probability measure.
$(\Leftarrow)$
Since $P(S) = 1$ and by (78), $1 = P(S) \leq P(\cup_{n=1}^{\infty} A_n) = \lim_n P(A_n)$ by the continuity of probability measure. □

Exercise 3.2

(a) $X_n \stackrel{a.s.}{\to} X$ but not $X_n \stackrel{qm}{\to} X$
Ex. Let $X_n = nI(U < 1/n^2)$ where $U \sim U(0,1)$. Then $\sum_{i=1}^{\infty} P(|X_i| > \varepsilon) \leq \sum_{i=1}^{\infty} P(U < 1/i^2) = \sum_{i=1}^{\infty} 1/i^2 < \infty$, so $P(X_n \to 0) = 1$ but $EX_n^2 = 1$.

(b) $X_n \stackrel{qm}{\to} X$ but not $X_n \stackrel{a.s.}{\to} X$
Ex. Consider the Example 3.5. It is clear that $X_n \stackrel{qm}{\to} 0$ because $EX_n^2 = P(J_n) \to 0$, but not $X_n \stackrel{a.s.}{\to} 0$.

3.2 The Strong Law of Large Numbers

Theorem 3.3. *SLLN : Suppose that $\mathbf{X}_1, \mathbf{X}_2, \cdots$ are iid and have finite mean $\boldsymbol{\mu}$. Then $\bar{\mathbf{X}}_n \stackrel{a.s.}{\to} \boldsymbol{\mu}$.*

Lemma 3.1. *If $\sum_{i=1}^{\infty} P(||\mathbf{X}_i - \mathbf{X}|| > \varepsilon) < \infty$ for any $\varepsilon > 0$, then $\mathbf{X}_n \stackrel{a.s.}{\to} \mathbf{X}$.*

Proof. By the countable subadditivity of probability measure,

$$\{P(||\mathbf{X}_k - \mathbf{X}|| > \varepsilon \ \ \forall k \geq n) =\} P\left(\cup_{i=n}^{\infty} C_i\right) \leq \sum_{i=n}^{\infty} P(C_i) \left\{= \sum_{i=1}^{\infty} P(C_i) - \sum_{i=1}^{n-1} P(C_i)\right\} \quad (79)$$

where $C_i = \{\omega \in \Omega\,;\, ||\mathbf{X}_i - \mathbf{X}|| > \varepsilon\}$. The last term in (79) tends to 0 as $n \to \infty$. □

Theorem 3.4. *$\mathbf{X}_n \stackrel{P}{\to} \mathbf{X}$ iff each subsequence $\mathbf{X}_{n_1}, \mathbf{X}_{n_2}, \cdots$ contains a further subsequence that converges a.s. to $\mathbf{X}$.*

Theorem 3.5. *Kolmogorov's SLLN*
Suppose that $Y_1, Y_2, \cdots$ are independent with mean μ_i and

$$\sum_{i=1}^{\infty} \frac{V(Y_i)}{i^2} < \infty. \quad (80)$$

Then $\bar{Y}_n - \bar{\mu}_n \stackrel{a.s.}{\to} 0$.

Lemma 3.2. *iid case*
Suppose that $X_1, X_2, \cdots$ are iid and have finite mean μ. Define $Y_i = X_i I(|X_i| \leq i)$. Then

$$\sum_{i=1}^{\infty} \frac{V(Y_i)}{i^2} < \infty \quad (81)$$

and $\bar{X}_n - \bar{Y}_n \stackrel{a.s.}{\to} 0$.

Proof. of Theorem 3.3

If we define $Y_i = X_i I(|X_i| \le i)$, then from above theorem and lemma, $\bar{Y}_n \overset{a.s.}{\to} \mu$. Also, we obtain $\bar{X}_n - \bar{Y}_n \overset{a.s.}{\to} 0$. So,

$$\bar{X}_n = \bar{Y}_n + (\bar{X}_n - \bar{Y}_n) \overset{a.s.}{\to} \mu. \tag{82}$$

□

Exercise 3.3 (Borel-Cantelli Lemma)

Let

$$B_n \; i.o. := \{\omega \in \Omega \, ; \forall n \;\; \exists k \;\; such \;\; that \;\; \omega \in B_k \;\; whenever \;\; k \ge n.\}. \tag{83}$$

If $\sum_{i=1}^n P(B_i) < \infty$, then $P(B_n \; i.o.) = 0$.

Proof. Since there exists N such that $\sum_{i=N}^\infty P(B_i) < \varepsilon$,

$$P(B_n \; i.o.) \le P(\cup_{i=N}^\infty B_i) < \varepsilon. \tag{84}$$

□

Exercise 3.4 (SLLN)

Let $X_1, X_2, \cdots$ are iid and have $EX_1^4 < \infty$.

(a) Assume that $EX_1 = 0$.

$$E(X_1 + \cdots + X_n)^4 = \sum_{k_1, \cdots, k_n \; ; \; \sum_{i=1}^n k_i = 4} \begin{pmatrix} 4 \\ k_1 \; \cdots \; k_n \end{pmatrix} E(X_1^{k_1} \cdots X_n^{k_n}) \tag{85}$$

$$= {}_n C_2 \, {}_4C_2 E(X_1^2 X_2^2) + nE(X_1^4) = 3n(n-1)E(X_1^2)^2 + nE(X_1^4) \tag{86}$$

(b) $P(|\bar{X}_n| > \varepsilon) < \frac{E(\bar{X}_n^4)}{\varepsilon^4} = \frac{E(X_1 + \cdots + X_n)^4}{\varepsilon^4 n^4} \to 0$ as $n \to \infty$. Therefore, $\bar{X}_n \overset{P}{\to} 0$.

(c) $\sum_{k=1}^n P(|\bar{X}_k| > \varepsilon) < \sum_{k=1}^n \frac{E(\bar{X}_k^4)}{\varepsilon^4} = \sum_{k=1}^n \frac{3k(k-1)(EX_1^2)^2 + kEX_1^4}{k^4 \varepsilon^4}$. Since $\sum_{i=1}^\infty i^{-2} < \infty$, $\limsup_n \sum_{k=1}^n P(|\bar{X}_k| > \varepsilon)$ is bounded, and by the Lemma 3.1, $\bar{X}_n \overset{a.s.}{\to} 0$.

Exercise 3.5

(a) *Proof.*

$$c^2 \sum_{i=1}^\infty \frac{1}{i^2} I(|c| \le i) = |c|^2 I(|c| \le 1) + \sum_{i=2}^\infty \frac{c^2}{i^2} I(|c| \le i) \le 1 + \int_1^\infty \frac{c^2}{x^2} I(|c| \le x) dx \tag{87}$$

For the last term in (87),

$$1 + \int_1^\infty \frac{c^2}{x^2} I(|c| \le x) dx = 1 + \int_{\max(1,|c|)}^\infty \frac{c^2}{x^2} dx = 1 + \frac{c^2}{\max(1, |c|)} \le 1 + |c| \tag{88}$$

□

(b) *Proof.* Let Y_i be defined as Lemma 3.2. Since $V(Y_i)/i^2 \le E(Y_i)^2/i^2 = E\frac{X_1^2}{i^2}I(|X_1| \le i)$ and from (a) we know that

$$\sum_{i=1}^{\infty} \frac{X_1^2}{i^2} I(|X_1| \le i) \le 1 + |X_1|, \tag{89}$$

taking the expecation on both sides obtains the result. □

Exercise 3.6

(a) *Proof.* $|X_n - Y_n| = |X_n - X_n I(|X_n| \le n)| = |X_n| I(|X_n| > n)$, so for $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P(|X_n - Y_n - 0| > \varepsilon) = \sum_{n=1}^{\infty} P(|X_1| I(|X_1| > n) > \varepsilon) \le \sum_{n=1}^{\infty} P(|X_1| > n) < E|X_1| < \infty \tag{90}$$

□

(b) *Proof.* If $a_n \to c$ as $n \to \infty$, then $\bar{a}_n \to c$. Since $X_n - Y_n \overset{a.s.}{\to} 0$,

$$1 = P(\omega\,;\, X_n - Y_n \to 0) \le P(\omega\,;\, \bar{X}_n - \bar{Y}_n \to 0). \tag{91}$$

□

Exercise 3.7 (Proof of Theorem 3.5)

(a) *Proof.* Let $S_n = \sum_{i=1}^{n}(X_i - EX_i)$, then $Y_k = \max_{2^{k-1} \le n < 2^k} \frac{|S_n|}{n} \le \max_{2^{k-1} \le n < 2^k} \frac{|S_n|}{2^{k-1}} \le \max_{1 \le n \le 2^k} \frac{|S_n|}{2^{k-1}}$. By the Kolmogorov's inequality,

$$P(Y_k \ge \varepsilon) \le P\left(\max_{1 \le n \le 2^k} |S_n| \ge 2^{k-1}\varepsilon\right) \le \frac{V(\bar{X}_{2^k} - \mu)}{4^{k-1}\varepsilon^2} = \frac{\sum_{i=1}^{2^k} V(X_i)}{4^{k-1}\varepsilon^2} \tag{92}$$

□

(b) *Proof.* Since

$$\sum_{k=1}^{\infty} P(|Y_k| \ge \varepsilon) \le \sum_{k=1}^{\infty}\sum_{i=1}^{2^k} \frac{V(X_i)}{4^{k-1}\varepsilon^2} = \sum_{i=1}^{\infty}\sum_{k=\lceil \log_2 i \rceil}^{\infty} \frac{V(X_i)}{4^{k-1}\varepsilon^2} \tag{93}$$

and

$$\sum_{i=1}^{\infty} \frac{V(X_i)}{i^2} < \infty \quad \& \quad \sum_{k=\lceil \log_2 i \rceil}^{\infty} \frac{1}{4^k} = \frac{4^{-\lceil \log_2 i \rceil}}{1 - 1/4} = \frac{4^{-\lceil \log_2 i \rceil + 1}}{3} \le \frac{4}{3i^2}, \tag{94}$$

$$\sum_{i=1}^{\infty}\sum_{k=\lceil \log_2 i \rceil}^{\infty} \frac{V(X_i)}{4^{k-1}\varepsilon^2} \le \sum_{i=1}^{\infty} \frac{16V(X_i)}{3i^2\varepsilon^2} < \infty. \tag{95}$$

Therefore $Y_k \overset{a.s.}{\to} 0$, which implies that $|\bar{X}_n - \mu| \overset{a.s.}{\to} 0$ □

3.3 The Dominated Convergence Theorem

3.3.1 Moments Do Not Always Converge

Example 3.14

Let $U \sim U(0,1)$ and define

$$X_n = nI\left(U < \frac{1}{n}\right), \quad \frac{X_n}{n} \sim Bern(n^{-1}). \tag{96}$$

Then $P(X_n > \varepsilon) \leq P[I(U < 1/n) > \varepsilon] < P(U < 1/n)/\varepsilon = \frac{1}{n\varepsilon}$ so $X_n \xrightarrow{P} 0$. However, $E(X_n) = 1$ for all n, so $EX_n \nrightarrow 0$.

Example 3.15

Let X_n be a contaminated standard normal distribution with mixture cdf

$$F_n(x) = \left(1 - \frac{1}{n}\right)\Phi(x) + \frac{1}{n}G_n(x). \tag{97}$$

Here, $X_n \xrightarrow{d} N(0,1)$ but can choose G_n such that $EX_n \to 0$.

Corollary 3.2. *If $X_1, X_2, \cdots$ are uniformly bounded (i.e. there exists M such that $|X_n| < M$ for all n) and $X_n \xrightarrow{d} X$, then $EX_n \to EX$.*

3.3.2 Quantile Functions and the Skorohod Representation Theorem

Definition 3.3. If $F(x)$ is a cdf, then we define the quantile function

$$F^- : (0,1) \to \mathbb{R}, \quad F^-(q) := \inf\{x \in \mathbb{R}\,;\, F(x) \geq q\} \tag{98}$$

Lemma 3.3. *$q \leq F(x)$ iff $F^-(q) \leq x$.*

Proof.
($\Rightarrow$) Suppose that there exists x_0 such that $q \leq F(x_0)$. By definition, $F^-(q) \leq x_0$.
($\Leftarrow$) Suppose that there exists x_0 such that $\inf\{x\,;\, q \leq F(x)\} \leq x_0$ and since F is a non-decreasing function, $q \leq F(x_0)$. □

Corollary 3.3. *If X is a random variable with F and $U \sim U(0,1)$, then X and $F^-(U)$ have the same distribution.*

Proof. By the lemma above,

$$P(U \leq F(x)) = P(F^-(U) \leq x) = F(x). \tag{99}$$

□

Theorem 3.6. *(Skorohod Representation Theorem)*

Assume $F, F_1, F_2, \cdots$ are cdfs and $F_n(x) \to F(x)$ as $n \to \infty$ for the points of continuity x. Then there exists $Y_1, Y_2, \cdots$ such that

1. *$P(Y_n \le t) = F_n(t)$ for all n and $P(Y \le t) = F(t)$.*
2. *$Y_n \overset{a.s.}{\to} Y$.*

Theorem 3.7. *(Dominated Convergence Theorem)*

If for a nonnegative Z, $|X_n| \le Z$ for all n and $EZ < \infty$, then $X_n \overset{d}{\to} X$ implies that $EX_n \to EX$.

Proof. Use the Skorohod Representation Theorem and contruct a sequence $\{Y_i\}_{i=1}^{\infty}$ such that $Y_n \overset{d}{=} X_n$, $Y \overset{d}{=} X$ and $Y_n \overset{a.s.}{\to} Y$. Let $Z' = \sup_n |Y_n| + W$ such that $Z' \overset{d}{=} Z$ where $W := Z - \sup_n |X_n|$. Then, we have $|Y_n| \le Z'$, $EZ' < \infty$. Since $EX_n = EY_n$ for all n and $EX = EY$, it suffices to prove that $EY_n \to EY$. By Fatou's Lemma,

$$E \liminf_n |Y_n| \le \liminf_n E|Y_n| \quad \text{and} \quad EZ' - E \limsup_n |Y_n| \le EZ' - \limsup_n E|Y_n| \tag{100}$$

$$(\therefore) \quad E \liminf_n |Y_n| \le \liminf_n E|Y_n| \le \limsup_n E|Y_n| \le E \limsup_n |Y_n|. \tag{101}$$

Since $Y_n \overset{a.s.}{\to} Y$, $P(\lim_n Y_n = Y) = 1$, which leads to $\lim_n E|Y_n| = E|Y|$. □

Exercise 3.9

Proof. Since f is a non-decreasing function, for any point of discontinuity x,

$$f(x^-) < f(x^+) \tag{102}$$

where $x^- = x - \varepsilon$ and $x^+ = x + \varepsilon$ for $\varepsilon > 0$. If we define $D = \{x \in \mathbb{R}; f(x) \ \textit{is discontinuous}\}$, for each $x \in D$ there exists a rational number $r(x)$ such that

$$f(x^-) < r(x) < f(x^+). \tag{103}$$

On the other hand, since $f(x_1^+) \le f(x_2^-)$ for $x_1 < x_2$, we can say that

$$x_1 \ne x_2 \quad for \quad x_1, x_2 \in D \quad \Rightarrow \quad r(x_1) \ne r(x_2) \tag{104}$$

so if we define $g : D \to \mathbb{Q}$ as $g(x) = r(x)$, g is an 1-1 function. Therefore, D is equivalent with the subset of $\mathbb{Q}$. □

Exercise 3.10 (Skorohod Representation Theorem)

(a) *Proof.* Let $Y_n(\omega) := F_n^-(\omega)$ and $Y(\omega) := F^-(\omega)$. Note that $U(\omega) \sim U(0,1)$. Since a point of continuity x_0 for F exists, there exists δ such that

$$Y(\omega) - \delta < x_0 < Y(\omega) \ \ and \ \ 0 < \omega - F(x_0). \tag{105}$$

Since $F_n(x_0) \to F(x_0)$, there exists N_1 such that if $n > N_1$, then $0 < |F_n(x_0) - F(x_0)| < \omega - F(x_0)$ $(= \varepsilon_1)$, which leads to

$$F_n(x_0) < \omega, \ \ therefore \ \ Y(\omega) - \delta < x_0 < Y_n(\omega). \tag{106}$$

□

(b) *Proof.* There exists a point of continuity of $F(x)$, say x_1, such that

$$Y(\omega + \varepsilon) < x_1 < Y(\omega + \varepsilon) + \delta \ \ and \ \ 0 < F(x_1) - \omega. \tag{107}$$

Since there is N_2 such that if $n > N_2$, then $|F_n(x_1) - F(x_1)| < F(x_1) - \omega$ $(= \varepsilon_1)$. These lead to

$$F_n(x_1) > \omega, \ \ therefore \ \ Y(\omega + \varepsilon) + \delta > x_1 > Y_n(\omega). \tag{108}$$

□

(c) *Proof.* Let $n > \max(N_1, N_2)$, then

$$Y(\omega) - \delta < Y_n(\omega) < Y(\omega + \varepsilon) + \delta. \tag{109}$$

Therefore, $Y(\omega) - \delta < \liminf_n Y_n(\omega) \leq \limsup_n Y_n(\omega) < Y(\omega + \varepsilon) + \delta$ and obtains the result as $\delta, \varepsilon \to 0$. □

(d) *Proof.* Since the P-measure of the points of discontinuity for $F(x)$ is zero, and by (c),

$$\begin{aligned} &P(\omega \in (0,1)\,; Y_n(\omega) \to Y(\omega)) = \\ &\qquad P(\omega \in (0,1)\,;\ F(\omega) \ \ is \ \ discontinuous, Y_n(\omega) \to Y(\omega)) + \\ &\qquad P(\omega \in (0,1)\,;\ F(\omega) \ \ is \ \ continuous, Y_n(\omega) \to Y(\omega)) = 0 + 1. \end{aligned} \tag{110}$$

□

Therefore, $Y_n \overset{a.s.}{\to} Y$.

Exercise 3.11 (Fatou's Lemma)

Proof. First, $\inf_{k \geq n} |X_k|$ is a non-decreasing sequence and $|X_n| \geq \inf_{k \geq n} |X_k|$ for any n so $E|X_n| \geq E \inf_{k \geq n} |X_k|$. Also, by definition,

$$\lim_{n \to \infty} \inf_{k \geq n} |X_k| = \liminf_{n \to \infty} |X_n| \ \ for \ \ any \ \ \omega \in \Omega, \tag{111}$$

$$(\therefore) \quad \lim_{n \to \infty} E \inf_{k \geq n} |X_k| = E \liminf_{n \to \infty} |X_k| \tag{112}$$

by the monotone convergence property of the expectation. Therefore,

$$\liminf_n E|X_n| \geq \liminf_n E \inf_{k \geq n} |X_k| = E \liminf_n |X_n|. \tag{113}$$

□

Exercise 3.12 (Uniform Integrability)

(a) *Proof.*

$$|A_n + B_n|I(|A_n + B_n| \geq \alpha) \leq (|A_n| + |B_n|)I(|A_n| \geq \alpha/2) + (|A_n| + |B_n|)I(|B_n| \geq \alpha/2) \tag{114}$$

$$\leq 2|A_n|I(|A_n| \geq \alpha/2) + 2|B_n|I(|B_n| \geq \alpha/2). \tag{115}$$

Therefore,

Figure 1: Exercise 3.12 (a) ; yellow-colored surface is always above the other.

$$0 \leq \sup_n E[|A_n + B_n|I(|A_n + B_n| \geq \alpha)] \leq o(1) \quad as \quad \alpha \to \infty \tag{116}$$

by the uniform integrability of A_n and B_n. □

(b) *Proof.* By the Skorohod Representation Theorem, there exist $Z_n \overset{d}{=} Y_n$, $Z \overset{d}{=} Y$ and $Z_n \overset{a.s.}{\to} Z$. By the Fatou's Lemma, $E|Z| \leq \liminf_n E|Z_n|$. Since $\sup_n E|Z_n|I(|Z_n| \geq \alpha) \to 0$, if α is large enough that the supremum is less than 1, then

$$E|Z_n| \leq \alpha P(\Omega) + 1 = \alpha + 1 < \infty. \tag{117}$$

□

So $E|Z| < \infty$, therefore $E|Z|I(|Z| \geq \alpha) \overset{\alpha\to\infty}{\to} 0$. Also, (a) implies that since $X_n = |Z_n - Z| \leq |Z_n| + |Z|$, $X_1, X_2, \cdots$ is an uniformly integrable sequence.

(c) *Proof.* Since

$$EX_n = EX_nI(|X_n| \geq \alpha) + EX_nI(|X_n| < \alpha), \tag{118}$$

the LHS tends to zero as $\alpha \to \infty$ and $X_nI(|X_n| < \alpha)$ is always bounded above by fixed α and $X_n \overset{d}{\to} 0$, $EX_n \to 0$ by the DCT. □

Exercise 3.13

Proof. Since

$$|Y_n|I(|Y_n| \geq \alpha) \leq |Y_n|\frac{|Y_n|^\varepsilon}{\alpha^\varepsilon} = \frac{|Y_n|^{\varepsilon+1}}{\alpha^\varepsilon}, \tag{119}$$

$$(\therefore) \quad 0 \leq \sup_n |Y_n|I(|Y_n| \geq \alpha) \leq \sup_n \frac{|Y_n|^{\varepsilon+1}}{\alpha^\varepsilon} < \infty. \tag{120}$$

Taking $\liminf_\alpha$ and $\limsup_\alpha$ on both sides obtains the results. □

Exercise 3.14
Let $g(X,\alpha) := XI(X<\alpha)$. Then

$$Eg(|Z|, \lceil\alpha\rceil - 1) \leq Eg(|Z|,\alpha) \leq Eg(|Z|,\lceil\alpha\rceil) = E|Z| - \sum_{i=\lceil\alpha\rceil}^{\infty} E|Z|I_{[i-1,i)}(|Z|) \tag{121}$$

where both bounds tends to $E|Z|$ as $\alpha \to \infty$ because $E|Z| < \infty$. Therefore, $Eg(|Z|,\alpha) \to E|Z|$, that is, $E|Z|I(|Z| \geq \alpha) = E|Z| - Eg(|Z|,\alpha)$ tends to zero. Since $E|Z| \geq E|Y_n|$, ???

4 Central Limit Theorems

4.1 Characteristic Functions and Normal Distributions

4.1.1 Inversion and the Uniqueness Theorem

A characteristic function φ uniquely determines the measure μ it comes from. This fundamental fact will be derived by means of an inversion formula through which μ can in principle be recovered from φ. Define

$$S(T) = \int_0^T \frac{\sin x}{x} dx, \quad T \geq 0.$$

Theorem 4.1. $S(T) \overset{T\to\infty}{\to} \frac{\pi}{2}$.

Proof. Since

$$\begin{aligned}\int_0^T \frac{\sin x}{x} dx = \int_0^T \sin x \int_0^\infty e^{-ux} du dx = \int_0^\infty \int_0^T e^{-ux} \sin x dx du \\ = \int_0^\infty \frac{1 - e^{-uT}(u\sin T + \cos T)}{1+u^2} du = \frac{\pi}{2} - \int_0^\infty \frac{e^{-uT}(u\sin T + \cos T)}{1+u^2} du \\ = \frac{\pi}{2} - \int_0^\infty \frac{e^{-z}(z/T\sin T + \cos T)}{1+z^2/T^2} dz \to \frac{\pi}{2}\end{aligned}$$

as $T \to \infty$ by LDCT. $\square$

Also,

$$\int_0^T \frac{\sin t\theta}{t} dt = sgn(\theta) \cdot S(T|\theta|), \quad T \geq 0. \tag{122}$$

Theorem 4.2. *If the probability measure μ has characteristic function φ, and if $\mu\{a\} = \mu\{b\} = 0$, then*

$$\mu(a,b] = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt. \tag{123}$$

Distinct measures cannot have the same φ.

Proof. Define $I_T = \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)$ then by Fubini,

$$I_T = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-T}^{T}\frac{e^{-it(x-a)}-e^{-it(x-b)}}{it}dt\mu(dx)$$

and by (122),this equals

$$I_T = \int_{-\infty}^{\infty}\left[\frac{sgn(x-a)}{\pi}S(T|x-a|)-\frac{sgn(x-b)}{\pi}S(T|x-b|)\right]\mu(dx)$$

whose integrand is bounded and converges as $T \to \infty$ to the function

$$\psi_{a,b}(x) = \frac{I_{\{a,b\}}(x)}{2} + I_{(a,b)}(x).$$

Thus, $I_T \to \int \psi_{a,b}d\mu$ which implies the conclusion when $\mu\{a\} = \mu\{b\} = 0$. □

Lemma 4.1. *Suppose that $\varphi \in L^1$. Then F has derivative*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi(t)dt.$$

Proof. Since

$$\frac{|e^{-itb}-e^{-ita}|}{|it|} = \frac{|e^{it(b-a)}-1|}{|t|} \leq |b-a|,$$

$$\mu(a,b) \leq (b-a)\int_{-\infty}^{\infty}|\varphi(t)|dt.$$

By (123),

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}-e^{-it(x+h)}}{ith}\varphi(t)dt \tag{124}$$

and the integrand of (124) is bounded by $|\varphi|$ and goes to $e^{itx}\varphi(t)$ as $h \to 0$. Therefore, the result follows. □

4.1.2 The Continuity Theorem

Remark

- A complex number c is given by $c = a + ib$ where $a, b \in \mathbb{R}$.
- $c\bar{c} = a^2 + b^2$ and $\Re(c) = \frac{1}{2}(c+\bar{c})$ & $\Im(c) = \frac{1}{2i}(c-\bar{c})$.
- $d/c = d \times \frac{\bar{c}}{c\bar{c}} = d\frac{a-ib}{a^2+b^2}$
- Points on the unit circle are now given by the complex numbers $\cos\theta + i\sin\theta$.

- Since

$$\begin{aligned} e^{ix} &= 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \frac{(ix)^8}{8!} + \cdots \\ &= 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \frac{x^8}{8!} + \cdots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right) \\ &= \cos x + i \sin x, \end{aligned} \tag{125}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \textit{and} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}. \tag{126}$$

Definition 4.1. For $\mathbf{X}$, we define the characteristic function $\varphi_{\mathbf{X}} : \mathbb{R}^k \to \mathbb{C}$ by

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\exp(i\mathbf{t}^\top\mathbf{X}) = E\cos(\mathbf{t}^\top\mathbf{X}) + iE\sin(\mathbf{t}^\top\mathbf{X}), \tag{127}$$

and the characteristic functions have properties of mgf.

Theorem 4.3. *$\mathbf{X}_1 \overset{d}{=} \mathbf{X}_2$ iff $\varphi_{\mathbf{X}_1}(\mathbf{t}) = \varphi_{\mathbf{X}_2}(\mathbf{t})$ for all $\mathbf{t}$.*

Theorem 4.4. *$\mathbf{X}_n \overset{d}{\to} \mathbf{X}$ iff $\varphi_{\mathbf{X}_n}(\mathbf{t}) \to \varphi_{\mathbf{X}}(\mathbf{t})$ for all $\mathbf{t}$.*

4.1.3 Moments

Lemma 4.2. *If $E||\mathbf{X}|| < \infty$, then $\boldsymbol{\nabla}\varphi_{\mathbf{X}}(\mathbf{0}) = iE(\mathbf{X})$.*

Proof. Let $\mathbf{e}_j = \{I(i=j)\}_{i=1}^k$ then

$$\frac{\varphi_{\mathbf{X}}(\mathbf{t} + h\mathbf{e}_j) - \varphi_{\mathbf{X}}(\mathbf{t})}{h} = E\left[\exp(i\mathbf{t}^\top\mathbf{X})\frac{\exp(ihX_j) - 1}{h}\right]. \tag{128}$$

Note that

$$\left|\exp(i\mathbf{t}^\top\mathbf{X})\frac{\exp(ihX_j) - 1}{h}\right| = \left|\int_0^{X_j} \exp(iht)dt\right| \le |X_j| \tag{129}$$

so if $E|X_j| < \infty$, then the DCT implies that

$$\frac{\partial}{\partial t_j}\varphi_{\mathbf{X}}(\mathbf{t}) = E\lim_{h\to 0}\left[\exp(i\mathbf{t}^\top\mathbf{X})\frac{\exp(ihX_j) - 1}{h}\right] = iE[X_j\exp(i\mathbf{t}^\top\mathbf{X})]. \tag{130}$$

□

Lemma 4.3. *If $E||\mathbf{X}|| < \infty$, then $\boldsymbol{\nabla}\varphi_{\mathbf{X}}(\mathbf{0}) = iE\mathbf{X}$.*

Lemma 4.4. *If $E\mathbf{X}^\top\mathbf{X} < \infty$, then $\boldsymbol{\nabla}^2\varphi_{\mathbf{X}}(\mathbf{0}) = -E\mathbf{X}\mathbf{X}^\top$.*

4.1.4 The Multivariate Normal Distribution

Definition 4.2. Let Σ be any symmetric, nonnegative definite, $k \times k$ matrix and let $\boldsymbol{\mu}$ be any vector in $\mathbb{R}^k$. Then the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ is defined to be the distribution with characteristic function

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left(i\mathbf{t}^\top \boldsymbol{\mu} - \frac{\mathbf{t}^\top \Sigma \mathbf{t}}{2}\right). \tag{131}$$

Definition 4.3. A square matrix Q is orthogonal if Q^{-1} exists and is equal to $Q^\top$.

Lemma 4.5. *If A is a symmetric $k \times k$ matrix, then there exists an orthogonal matrix Q such that $QAQ^\top$ is diagonal.*

4.1.5 Asymptotic Normality

Theorem 4.5. *CLT*
If $\mathbf{X}_1, \mathbf{X}_2, \cdots$ are iid with mean $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance Σ, where Σ has finite entries, then

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_k(\mathbf{0}, \Sigma). \tag{132}$$

Example 4.11

Suppose that $X_1, X_2, \cdots$ are iid with mean μ and $V(X_1) = \sigma^2$ and $V[(X_1-\mu)^2] = \tau^2 < \infty$. Define

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}. \tag{133}$$

Without loss of generality, we assume that $\mu = 0$. By the CLT,

$$\sqrt{n}\left(\frac{\sum_{i=1}^n X_i^2}{n} - \sigma^2\right) \xrightarrow{d} N(0, \tau^2). \tag{134}$$

Furthermore, the CLT and the WLLN imply $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, \sigma^2)$ and $\bar{X}_n \xrightarrow{P} 0$, so Slutsky's theorem implies that $\sqrt{n}\bar{X}_n^2 \xrightarrow{P} 0$. Therefore, since

$$\sqrt{n}(S_n^2 - \sigma^2) = \sqrt{n}\left(\frac{\sum_{i=1}^n X_i^2}{n} - \sigma^2\right) - \sqrt{n}\bar{X}_n^2, \tag{135}$$

$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \tau^2)$.

4.1.6 The Cramér-Wold Theorem

Theorem 4.6. $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ *iff* $\mathbf{a}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{a}^\top \mathbf{X}$ *for all* $\mathbf{a} \in \mathbb{R}^k$.

Exercise 4.1

(a) *Proof.* Consider the DE

$$\frac{d}{dt}\varphi_Y(t) = -t\sigma^2\varphi_Y(t). \ \ Then \ \ \frac{y'(t)}{y(t)} = -t\sigma^2 \ \ given \ \ that \ \ y(t) \neq 0. \tag{136}$$

$$(\therefore) \ \ y(t) = \exp(-t^2\sigma^2/2). \tag{137}$$

□

(b) *Proof.* Let $\mathbf{X} := \boldsymbol{\mu} + Q^\top \mathbf{Y}$ where $\mathbf{Y} \sim N(\mathbf{0}, Q\Sigma Q^\top)$. Then

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\exp\{i\mathbf{t}^\top(\boldsymbol{\mu} + Q^\top \mathbf{Y})\} = \exp\left(i\mathbf{t}^\top\boldsymbol{\mu} - \frac{\mathbf{t}^\top\Sigma\mathbf{t}}{2}\right). \tag{138}$$

□

Exercise 4.2

(a) *Proof.*

$$E\exp\{i(\mathbf{X}-\mathbf{a})^\top\mathbf{Y}\} = EE[\exp\{i\mathbf{Y}^\top(\mathbf{X}-\mathbf{a})\}|\mathbf{Y}] = E\exp(-i\mathbf{a}^\top\mathbf{Y})\varphi_{\mathbf{X}}(\mathbf{Y}) \tag{139}$$

□

(b) *Proof.*

$$Ef_{\mathbf{Y}}(\mathbf{s}-\mathbf{X}) = \int_{\mathbf{X}} f(\mathbf{X})f_{\mathbf{Y}}(\mathbf{s}-\mathbf{X})d\mathbf{X} = f * f(\mathbf{s}) = f_{\mathbf{X}+\mathbf{Y}}(\mathbf{s}). \tag{140}$$

□

(c) *Proof.*

$$f_{\mathbf{X}+\mathbf{Y}}(\mathbf{s}) = Ef_{\mathbf{Y}}(\mathbf{s}-\mathbf{X}) = (2\pi\sigma^2)^{-\frac{k}{2}}E\exp\{-(\mathbf{s}-\mathbf{X})^\top(\mathbf{s}-\mathbf{X})/2\sigma^2\} \tag{141}$$

$$= (2\pi\sigma^2)^{-\frac{k}{2}}E\varphi_{\mathbf{Y}}\left(\frac{\mathbf{X}}{\sigma^2} - \frac{\mathbf{s}}{\sigma^2}\right) \tag{142}$$

because $\mathbf{Y} \sim N_k(\mathbf{0}, \sigma^2 I)$. □

Exercise 4.3

Proof. ($\Rightarrow$)

Since $\varphi_{\mathbf{a}^\top\mathbf{X}_n}(1) \to \varphi_{\mathbf{a}^\top\mathbf{X}}(1)$ for all $\mathbf{a}$, $\mathbf{a}^\top\mathbf{X}_n \xrightarrow{d} \mathbf{a}^\top\mathbf{X}$.

($\Leftarrow$)

Take $\mathbf{a} = \mathbf{e}_j$ for each $j = 1, \cdots, k$. Then $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$. □

Exercise 4.4

Since Σ is symmetric and nonnegative definite, there exists an orthogonal Q such that $Q\Sigma Q^\top = \Lambda$ is diagonal. Since there exists $\Lambda^{\frac{1}{2}}$ such that $\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}} = \Lambda$, $\Sigma = Q^\top \Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}} Q = A^\top A$. Since $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$, we can say that

$$\Lambda^{-\frac{1}{2}} Q(\mathbf{X} - \boldsymbol{\mu}) \sim N(\mathbf{0}, I). \tag{143}$$

Therefore, the result follows.

Exercise 4.5

(a) Since $EY_n = 0$ and $Y_n \overset{d}{=} Y_n^+ - Y_n^-$ where $Y_n^- = -Y_n I(Y_n \leq 0)$, $EY_n^+ = EY_n^-$. Also, since $Y_n^- = -\sqrt{n}(\bar{X}_n - 1)I(\bar{X}_n \leq 1) = \sqrt{n}(n - \sum_{i=1}^n X_i)I(\sum_{i=1}^n X_i \leq n)/n$,

$$EY_n^- = \frac{e^{-n}\sqrt{n}}{n}\sum_{0\leq y\leq n}\left(n\frac{n^y}{y!} - y\frac{n^y}{y!}\right) = \frac{e^{-n}n^{n+.5}}{n!} = EY_n^+. \tag{144}$$

(b) Since $Y_n \xrightarrow{d} N(0,1)$ and $E|Y_n|^{1+\varepsilon} < \infty$ for $\varepsilon = 1$ and any n, $EY_n \to 0$. Likewise, since $|\cdot|$ is a continuous function, $|Y_n| \xrightarrow{d} |N(0,1)|$ and $E|Y_n| \to E|N(0,1)| = 2/\sqrt{2\pi}$.

$$|Y_n| = Y_n^+ + Y_n^- \Rightarrow E|Y_n| = 2EY_n^+ \Rightarrow EY_n^+ \to \frac{1}{\sqrt{2\pi}} \tag{145}$$

$$(\therefore) \quad \frac{\sqrt{2\pi}n^{n+.5}e^{-n}}{n!} \to 1. \tag{146}$$

Exercise 4.6

Suppose that $X_1, X_2, \cdots$ are independent and identically distributed and have finite mean μ. Then $\bar{X}_n \xrightarrow{P} \mu$.

Proof. Since

$$\begin{aligned}\varphi_{\bar{X}_n}(t) &= \left\{\varphi_{X_1}\left(\frac{t}{n}\right)\right\}^n = \left(\int_{\mathbb{R}} e^{\frac{itx}{n}}dF(x)\right)^n \\ &= \left(1 + \frac{it\mu}{n} - \int_{\mathbb{R}}\int_0^x \frac{t^2u^2(x-u)}{n^2}dudF(x)\right)^n = \left\{1 + \frac{it\mu}{n} + o\left(\frac{1}{n}\right)\right\}^n\end{aligned}$$

as $n \to \infty$ by DCT, $\varphi_{\bar{X}_n}(t) \to e^{it\mu}$. □

Exercise 4.7

First, the Cramér-Wold Theorem states that

$$\mathbf{X}_n \xrightarrow{d} \mathbf{X} \quad \iff \quad \mathbf{a}^\top\mathbf{X}_n \xrightarrow{d} \mathbf{a}^\top\mathbf{X} \;\; for \;\; all \;\; \mathbf{a} \in \mathbb{R}^k. \tag{147}$$

And the CLT states that $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. Since $Y_i := \mathbf{a}^\top\mathbf{X}_i$ is an univariate sequence of iid random variables with the expectation $\mathbf{a}^\top\boldsymbol{\mu}$ and the covariance matrix $\mathbf{a}^\top\Sigma\mathbf{a}$ for any $\mathbf{a} \in \mathbb{R}^k$, $\sqrt{n}(\bar{Y}_n - \mathbf{a}^\top\boldsymbol{\mu}) \xrightarrow{d} N(0, \mathbf{a}^\top\Sigma\mathbf{a})$. Therefore, $\mathbf{a}^\top\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{a}^\top N(\mathbf{0}, \Sigma)$.

4.2 The Lindeberg-Feller Central Limit Theorem

4.2.1 The Lindeberg and Lyapunov Conditions

Theorem 4.7. *For $X_1, X_2, \cdots$ such that $EX_n = \mu_n$ and $VX_n = \sigma_n^2 < \infty$, define $Y_n = X_n - \mu_n$, $T_n = \sum_{i=1}^n Y_i$, $s_n^2 = V(T_n) = \sum_{i=1}^n \sigma_i^2$. If*

$$\exists\, \delta > 0 \quad such \quad that \quad \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|Y_i|^{2+\delta} \to 0 \quad as \quad n \to \infty, \tag{148}$$

then

$$for \quad each \quad \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^n E[Y_i^2 I(|Y_i| \geq \varepsilon s_n)] \to 0 \quad as \quad n \to \infty. \tag{149}$$

That is, the Lyapunov condition implies the Lindeberg condition.

Proof. Fix $\varepsilon > 0$, since

$$\frac{Y_i^2}{s_n^2} I(|Y_i| \geq \varepsilon s_n) \leq \frac{|Y_i|^{2+\delta}}{\varepsilon^\delta s_n^{2+\delta}}, \quad \frac{1}{s_n^2} \sum_{i=1}^n E[Y_i^2 I(|Y_i| \geq \varepsilon s_n)] \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{i=1}^n E|Y_i|^{2+\delta}. \tag{150}$$

It is clear that the RHS of the inequality tends to 0 as $n \to \infty$, which proves the conclusion. $\square$

Example 4.14

Let $X_n \sim Bern(p_n)$, so that $Y_n = X_n - p_n$ and $\sigma_n^2 = p_n(1-p_n)$. First, take $\delta = 1$ then

$$\frac{1}{s_n^3} \sum_{i=1}^n E|Y_i|^3 = \frac{1}{s_n^3} \sum_{i=1}^n \{p_i^3(1-p_i) + p_i(1-p_i)^3\} = \frac{1}{s_n^3} \sum_{i=1}^n \sigma_i^2 \{p_i^2 + (1-p_i)^2\} \leq \frac{1}{s_n^3} \sum_{i=1}^n \sigma_i^2. \tag{151}$$

Therefore, if $\frac{s_n^2}{s_n^3} \to 0$ as $n \to \infty$, that is, if $s_n \to \infty$, then the Lyapunov condition is satisfied.

4.2.2 Independent and Identically Distributed Variables

Example 4.15

Suppose that $X_1, X_2, \cdots$ are iid with mean μ and variance $0 < \sigma^2 < \infty$. Then $s_n^2 = n\sigma^2$. Fix $\varepsilon > 0$, then the Lindeberg condition is satisfied that for each $\varepsilon > 0$,

$$\frac{1}{\sigma^2} E[Y_1^2 I(|Y_1| \geq \varepsilon\sigma\sqrt{n})] \to 0 \quad as \quad n \to \infty. \tag{152}$$

Proof. Since $P(|Y_1| \geq \varepsilon\sigma\sqrt{n}) \leq \frac{EY_1^2}{\varepsilon^2\sigma^2 n} \to 0$ as $n \to \infty$, $Z_n \xrightarrow{P} 0$ as $n \to \infty$ where $Z_n := Y_1^2 I(|Y_1| \geq \varepsilon\sigma\sqrt{n})$. Also, since $|Z_n| \leq Y_1^2$ and $EY_1^2 < \infty$, $EZ_n \to 0$ by DCT. $\square$

4.2.3 Triangular Arrays

We assume that for each n, $X_{n1}, \cdots, X_{nn}$ are independent and $EX_{ni} = \mu_{ni}$, $V(X_{ni}) = \sigma_{ni}^2 < \infty$. Let

$$Y_{ni} = X_{ni} - \mu_{ni} \tag{153}$$

$$T_n = \sum_{i=1}^{n} Y_{ni} \tag{154}$$

$$s_n^2 = V(T_n) = \sum_{i=1}^{n} \sigma_{ni}^2. \tag{155}$$

Theorem 4.8. *Lindeberg-Feller CLT*

$$for \quad each \quad \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^{n} E[Y_{ni}^2 I(|Y_{ni}| \geq \varepsilon s_n)] \to 0 \quad as \quad n \to \infty \tag{156}$$

if and only if

(a)

$$\frac{T_n}{s_n} \xrightarrow{d} N(0,1) \tag{157}$$

and

(b)

$$\frac{1}{s_n^2} \max_{i \leq n} \sigma_{ni}^2 \to 0 \tag{158}$$

as $n \to \infty$.

Example 4.17

Suppose $X_n \sim Bin(n, p_n)$. We claim that

$$\frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow{d} N(0,1)$$

whenever $np_n(1-p_n) \to \infty$ as $n \to \infty$.

Proof. Let $Y_{n1}, \cdots, Y_{nn}$ be iid and defined as

$$Y_{ni} = B_n - EB_n (= B_n - p_n) \quad where \quad B_n \sim Bernoulli(p_n)$$

then $X_n \overset{d}{=} np_n + \sum_{i=1}^{n} Y_{ni}$. Note that $V(Y_{ni}) = np_n(1-p_n) = s_n^2$. □

So the Lindeberg condition says that for any $\varepsilon > 0$,

$$\frac{1}{s_n^2}\sum_{i=1}^{n} E\left[Y_{ni}^2 I\left\{|Y_{ni}| \geq \varepsilon s_n\right\}\right] \to 0. \tag{159}$$

Since $|Y_{ni}| \leq 1$, if $\varepsilon s_n > 1$, then the LHS of (159) is always zero. Thus, since $np_n(1-p_n) = s_n^2 \to \infty$ as $n \to \infty$, the asymptotic normality follows.

Corollary 4.1. *(157) holds if the triangular array satisfies the Lyapunov condition.*

Exercise 4.8

(a) *Proof.* Note that

$$|a_1a_2 - b_1b_2| = |a_1a_2 - a_1b_2 + a_1b_2 - b_1b_2| \leq |a_1||a_2 - b_2| + |b_2||a_1 - b_1|$$

and since $|a_i| \leq 1$ and $|b_i| \leq 1$, the case when $n = 2$ follows. Now, let us show that

$$|a_1\cdots a_{n-1} - b_1\cdots b_{n-1}| \leq \sum_{i=1}^{n-1}|a_i - b_i| \quad \Rightarrow \quad |a_1\cdots a_n - b_1\cdots b_n| \leq \sum_{i=1}^{n}|a_i - b_i|.$$

Let $\alpha_n = a_1\cdots a_n$ and $\beta_n = b_1\cdots b_n$. Then

$$\begin{aligned} &|\alpha_{n-1}a_n - \beta_{n-1}b_n| = |\alpha_{n-1}a_n - \alpha_{n-1}b_n + \alpha_{n-1}b_n - \beta_{n-1}b_n| \\ \leq |\alpha_{n-1}||a_n - b_n| + |b_n||\alpha_{n-1} - \beta_{n-1}| &\leq |a_n - b_n| + |\alpha_{n-1} - \beta_{n-1}| \leq \sum_{i=1}^{n}|a_i - b_i|. \end{aligned}$$

So by the mathematical induction, the result follows. □

(b) *Proof.* Since we know the results in parts (b) and (c) in Exercise 1.43 ;

$$\begin{aligned} &\left|e^{ity/s} - 1 - \frac{ity}{s} + \frac{t^2y^2}{2s^2}\right|\{I(|y/s| < \varepsilon) + I(|y/s| \geq \varepsilon)\} \\ \leq \frac{|ty|^3}{s^3}I(|y/s| < \varepsilon) + \frac{t^2y^2}{s^2}I(|y/s| \geq \varepsilon) &< \frac{\varepsilon|t|^3y^2}{s^2} + \frac{t^2y^2}{s^2}I(|y/s| \geq \varepsilon), \end{aligned}$$

for $s > 0$,

$$\begin{aligned} \left|\varphi_{Y_{ni}}(t/s_n) - 1 + \frac{t^2\sigma_{ni}^2}{2s_n^2}\right| &= \left|E\left[\exp(itY_{ni}/s_n) - 1 - \frac{itY_{ni}}{s_n} + \frac{t^2Y_{ni}^2}{2s_n^2}\right]\right| \\ &\leq E|\cdot| \leq \frac{\varepsilon|t|^3\sigma_{ni}^2}{s_n^2} + \frac{t^2}{s_n^2}E\left[Y_{ni}^2 I(|Y_{ni}| \geq \varepsilon s_n)\right]. \end{aligned}$$

□

(c) *Proof.* Let's show that (156) implies (158). Since

$$\frac{\sigma_{ni}^2}{s_n^2} = \frac{EY_{ni}^2}{s_n^2}\{I(|Y_{ni}| \geq \varepsilon s_n) + I(|Y_{ni}| < \varepsilon s_n)\} < \varepsilon^2 + \frac{E\left[Y_{ni}^2 I(|Y_{ni}| \geq \varepsilon s_n)\right]}{s_n^2},$$

$\frac{\sigma_{ni}^2}{s_n^2} < \varepsilon^2$ as $n \to \infty$ for any $\varepsilon > 0$. □

(d) *Proof.* To use (a), first we check that $|\varphi_{Y_{ni}}(t/s_n)| \leq E|\exp(itY_{ni}/s_n)| = 1$ and take n large enough so $t^2 \max_i \frac{\sigma_{ni}^2}{s_n^2} \leq 1$, then

$$\left|\prod_{i=1}^n \varphi_{Y_{ni}}(t/s_n) - \prod_{i=1}^n \left(1 - \frac{t^2\sigma_{ni}^2}{2s_n^2}\right)\right| \leq \sum_{i=1}^n \left|\varphi_{Y_{ni}}(t/s_n) - 1 + \frac{t^2\sigma_{ni}^2}{2s_n^2}\right|,$$

which obtains the result through (b). □

(e) *Proof.* Since

$$|1 + x - e^x| = \left|\int_x^0 e^u(x-u)du\right| \leq \int_x^0 |x-u|du = \frac{x^2}{2} \leq x^2$$

for $x \leq 0$,

$$\left|\prod_{i=1}^n \left(1 - \frac{t^2\sigma_{ni}^2}{2s_n^2}\right) - \prod_{i=1}^n \exp\left(-\frac{t^2\sigma_{ni}^2}{2s_n^2}\right)\right| \leq \sum_{i=1}^n \left|1 - \frac{t^2\sigma_{ni}^2}{2s_n^2} - \exp\left(-\frac{t^2\sigma_{ni}^2}{2s_n^2}\right)\right|$$
$$\leq \frac{t^4}{4s_n^4}\sum_{i=1}^n \sigma_{ni}^4 \leq \frac{t^4}{4s_n^4}\max_{1\leq i\leq n}\sigma_{ni}^2 \sum_{i=1}^n \sigma_{ni}^2 = \frac{t^4}{4s_n^2}\max_{1\leq i\leq n}\sigma_{ni}^2$$

for n large enough so $t^2 \max_i \frac{\sigma_{ni}^2}{s_n^2} \leq 1$. □

(f) *Proof.* Using triangular inequalities with (d) and (e), we obtain

$$\left|\prod_{i=1}^n \varphi_{Y_{ni}}(t/s_n) - \prod_{i=1}^n \exp\left(-\frac{t^2\sigma_{ni}^2}{2s_n^2}\right)\right| \leq \varepsilon|t|^3 + \frac{t^2}{s_n^2}\sum_{i=1}^n EY_{ni}^2 I(|Y_{ni}| \geq \varepsilon s_n) + \frac{t^4}{4s_n^2}\max_{1\leq i\leq n}\sigma_{ni}^2 \tag{160}$$

and composing LF condition and (c) gets the conclusion. This leads to

$$\frac{\sum_{i=1}^n Y_{ni}}{s_n} \xrightarrow{d} N(0,1). \tag{161}$$

□

Exercise 4.9

(a) *Proof.* First, $|e^{it}-1| \le |t|$ and $|e^{it}-1| \le 2$ so $|e^{it}-1| \le 2\min(|t|,1)$. Therefore,

$$|\varphi_{Y_{ni}}(t/s_n) - 1| \le E|\exp(itY_{ni}/s_n) - 1| \le 2E\min(1, |tY_{ni}/s_n|) \le 2|t|E\left|\frac{Y_{ni}}{s_n}\right|$$
$$\le 2|t|\varepsilon P(|Y_{ni}| < \varepsilon s_n) = 2|t|\varepsilon - 2|t|\varepsilon P(|Y_{ni} \ge \varepsilon s_n) \le 2|t|\varepsilon + 2P(|Y_{ni}| \ge \varepsilon s_n). \quad (162)$$

Taking the maximum on both sides obtains the upper bound :

$$\max_{1\le i\le n} |\alpha_{ni}| \le 2|t|\varepsilon + 2\max_{1\le i\le n} P(|Y_{ni}| \ge \varepsilon s_n) \le 2|t|\varepsilon + 2\max_{1\le i\le n} \frac{\sigma_{ni}^2}{\varepsilon^2 s_n^2} \quad (163)$$

for any $\varepsilon > 0$, which tends to zero as $n \to \infty$. □

(b) *Proof.* Since

$$|\varphi_{Y_{ni}}(t/s_n) - 1| = |E\exp(itY_{ni}/s_n) - itEY_{ni}/s_n - 1| \le \frac{t^2\sigma_{ni}^2}{2s_n^2} \le \frac{t^2}{2} \quad (164)$$

and $|\alpha_{ni}|^2 \le |\alpha_{ni}|\max_i |\alpha_{ni}|$, of which upper bound tends to zero as $n \to \infty$, we can obtain the conclusion. □

(c) *Proof.* For $|z| \le 1$, $\Re(z) \le 1$ so

$$|e^{z-1}| = |\exp\{\Re(z) - 1\}||\exp\{i\Im(z)\}| = |\exp\{\Re(z) - 1\}| \le 1. \quad (165)$$

Therefore, since $|1+\alpha_{ni}| = |\varphi_{Y_{ni}}(t/s_n)| \le 1$ and $|\exp(\alpha_{ni})| = |\exp\{\varphi_{Y_{ni}}(t/s_n) - 1\}| \le 1$,

$$\left|\prod_{i=1}^n \exp(\alpha_{ni}) - \prod_{i=1}^n (1+\alpha_{ni})\right| \le \sum_{i=1}^n |\exp(\alpha_{ni}) - 1 - \alpha_{ni}|. \quad (166)$$

Next, since $|e^z - 1 - z| = \left|\frac{e^c z^2}{2}\right| \le \frac{e^{0.5}|z|^2}{2} \le |z|^2$ for some c and any z such that $|c| \le |z| \le 1/2$, the conclusion follows. □

(d) *Proof.* Since $\sum_{i=1}^n Y_{ni}/s_n \xrightarrow{d} N(0,1)$,

$$\left|\prod_{i=1}^n (1+\alpha_{ni}) - \exp\left(-\frac{t^2}{2}\right)\right| \to 0 \quad as \quad n \to \infty. \quad (167)$$

Combining this with (c) obtains

$$\limsup_{n\to\infty} \left|\prod_{i=1}^n e^{\alpha_{ni}} - e^{-\frac{t^2}{2}}\right| \le \limsup_{n\to\infty} \sum_{i=1}^n |\alpha_{ni}|^2 = 0 \quad (168)$$

which means that $|\exp(\sum_{i=1}^n \alpha_{ni})| = \exp\{\sum_{i=1}^n \Re(\alpha_{ni})\} \to \exp(-t^2/2)$ and since $\exp(\cdot)$ is a continuous function, the result follows. □

(e) **

Proof. Since for some real X such that $E(X) = 0$ and $V(X) = \sigma^2 < \infty$,

$$E(\cos X - 1) = E\left\{\frac{(Y_n - 1)X^2}{2}\right\} = o_P(1) - \frac{\sigma^2}{2}$$
$$E(\cos X - 1 + X^2/2) = E(\cos X - 1) + \frac{\sigma^2}{2} = o_P(1)$$

where $Y_n = o_P(1)$ as $n \to \infty$. Therefore, the result follows. $\square$

(f) *Proof.* Take t large enough so $\frac{t^2}{2} > \frac{2}{\varepsilon^2}$ for $\varepsilon > 0$. Since

$$\frac{-2Y_{ni}^2}{\varepsilon^2 s_n^2} I(|Y_{ni}| \geq \varepsilon s_n) \leq -2 \leq \cos\frac{tY_{ni}}{s_n} - 1 \tag{169}$$
$$\frac{t^2 Y_{ni}^2}{2s_n^2} I(|Y_{ni}| \geq \varepsilon s_n) \leq \frac{t^2 Y_{ni}^2}{2s_n^2} \tag{170}$$

and by (e), the proof ends. $\square$

Exercise 4.10

Consider iid $X_1, \cdots, X_n$ with a pdf $f_X(x) = (2+\delta)x^{-3-\delta}, x > 1$ and $\delta > 0$. Then

$$E(X_1) = \int_1^\infty (2+\delta)x^{-2-\delta}dx = \frac{2+\delta}{1+\delta}$$
$$E(X_1^2) = \int_1^\infty (2+\delta)x^{-1-\delta}dx = \frac{2+\delta}{\delta}$$

but all of its higher moments does not exists. So, for any $\delta > 0$, $E|Y_{ni}|^{2+\delta} = \infty$.

Exercise 4.11

Proof. Now since $\frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow{d} N(0,1)$ and $\frac{\max_{i\leq n}\sigma_{ni}^2}{s_n^2} = \frac{p_n(1-p_n)}{np_n(1-p_n)} = \frac{1}{n} \to 0$ for $n \to \infty$, we need to show that if $np_n(1-p_n) \to \infty$,

$$\frac{1}{np_n(1-p_n)}\sum_{i=1}^n E|Y_{ni}|^2 I\{|Y_{ni}| \geq \varepsilon\sqrt{np_n(1-p_n)}\} \to 0$$
$$\Leftrightarrow \frac{1}{p_n(1-p_n)} E|Y_{n1}|^2 I\{|Y_{n1}| \geq \varepsilon\sqrt{np_n(1-p_n)}\} \to 0$$

as $n \to \infty$ for any $\varepsilon > 0$. If $np_n(1-p_n) \nrightarrow \infty$, there exists N such that $\varepsilon\sqrt{np_n(1-p_n)} < \frac{1}{2}$ whenever $n > N$. This means that

$$|Y_{n1}|^2 I\{|Y_{n1}| \geq \varepsilon\sqrt{np_n(1-p_n)}\} \geq |Y_{n1}|^2 I(|Y_{n1}| \geq 1/2) \geq \frac{|Y_{n1}|}{2} I(|Y_{n1}| \geq 1/2)$$

whose expectation obtains the lower bound $\frac{p_n(1-p_n)}{2}$. So,

$$\frac{E|Y_{n1}|^2 I\{|Y_{n1}| \geq \varepsilon\sqrt{np_n(1-p_n)}\}}{p_n(1-p_n)} \geq \frac{1}{2}$$

and it is clear that $np_n(1-p_n)$ must not be bounded w.r.t. n. $\square$

Exercise 4.12

(a) *Proof.* Let $A_n := \max_{1\le i\le n} a_{ni}^2$ and for any $\varepsilon > 0$,

$$I\{a_{ni}(X_i-\mu)\ge \varepsilon s_n\} \le I\{A_n(X_i-\mu)^2 \ge \varepsilon^2 s_n^2\} \tag{171}$$

where $s_n^2 := V(T_n) = \sigma^2\sum_{i=1}^n a_{ni}^2$. Then,

$$Y_i := (X_i-\mu)^2 I\{A_n(X_i-\mu)^2\ge\varepsilon^2 s_n^2\} \tag{172}$$

which is an i.i.d. sequence. Thus,

$$\frac{1}{s_n^2}\sum_{i=1}^n Ea_{ni}^2(X_i-\mu)^2 I\{a_{ni}(X_i-\mu)\ge\varepsilon s_n\} \le \frac{1}{s_n^2}\sum_{i=1}^n a_{ni}^2 EY_i = \frac{EY_1}{\sigma^2}. \tag{173}$$

Since $|Y_1| \le (X_1-\mu)^2$, $\frac{E|Y_1|}{\sigma^2} \le 1$ and $Y_1 \xrightarrow{P} 0$ as $n\to\infty$, the LF condition was satisfied. □

Exercise 4.15

(a)

$$\begin{aligned}
P(X_1+X_2+X_3\le 1) &= \int_{\{\mathbf{x};\mathbf{1}^\top\mathbf{x}\le 1\}} 8x_1x_2x_3dx_1dx_2dx_3\\
&= \int_0^1\int_0^{1-x_1}\int_0^{1-x_1-x_2} 8x_1x_2x_3dx_3dx_2dx_1 = \int_0^1\int_0^{1-x_1} 4x_1x_2(1-x_1-x_2)^2dx_2dx_1\\
&= \frac{1}{3}\int_0^1 (x_1-4x_1^2+6x_1^3-4x_1^4+x_1^5)dx_1 = \frac{1}{90}\approx 0.0111.
\end{aligned}$$

(b) By CLT, $\sqrt{3}(\bar{X}_3-2/3)\ \dot\sim\ N(0,1/18)$.

$$(\therefore)\quad P(X_1+X_2+X_3\le 1) = P(\sqrt{3}) = P\left[\sqrt{3}\frac{\bar{X}_3-2/3}{\sqrt{1/18}} \le -\sqrt{3}\frac{\sqrt{18}}{3}\right] \asymp \Phi(-\sqrt{6})\approx 0.00715.$$

(c) $\bar{Z}_n \approx 0.12$, $\hat{V}(Z)\approx 0.011$. Theoretically,

$$V(Z) = P(X_1+X_2+X_3\le 1) - P(X_1+X_2+X_3\le 1)^2 = 0.01098765.$$

(d) $\bar{Z}_n = 0.139$, $\hat{V}(Z)\approx 0.12$.

Exercise 4.16

(a)

$$X_n = \begin{cases} 1 & w.p. \ \frac{1-2^{-n}}{2} \\ -1 & w.p. \ \frac{1-2^{-n}}{2} \\ 2^k & w.p. \ 2^{-k} \ \forall k > n \end{cases}$$

Therefore, $E(X_n^j) = \frac{1-2^{-n}}{2}\{1 + (-1)^j\} + \sum_{k>n} 2^{k(j-1)} = \infty$ for any positive integers j, n.

(b) ***

Exercise 4.17

(a) Let X_{nk} be the count of drawings until the k^{th} distinct one appears given $k-1$ distinct coupons. Then,

$$P(X_{nk} = x) = \left(\frac{k-1}{n}\right)^{x-1}\left(1 - \frac{k-1}{n}\right), \quad x = 1, 2, \cdots .$$

Therefore, $EX_{nk} = \frac{n}{n-k+1}$ and $V(X_{nk}) = \frac{n(k-1)}{(n-k+1)^2}$.

(b) First, $m_n = \sum_{k=1}^{n} \frac{n}{n-k+1}$ and $\tau_n^2 = \sum_{k=1}^{n} \frac{n(k-1)}{(n-k+1)^2}$. Now, let's check Lyapunov's condition with $\delta = 2$. Then,

$$\frac{\sum_{k=1}^{n} E|X_{nk} - E(X_{nk})|^4}{\tau_n^4} =$$