# Solution (Suggestion) to STATS 310 (MATH 230) Lecture Notes by Sourave Chatterjee

Neng-Tai Chiu, Chia-Cheng Hao, Yan-Wei Su

September 26, 2024

**Abstract**

It's the work accompanying Prof. Shue's Probability theory class in 2022 spring. We hope to cover exercises in Prof. Chatterjee's STATS 310 notes [Cha] from chapter 7 onward. Special thanks go to Pingbang Hu [Hu] who made the LaTeXtemplate.

## Contents

# 5 Chapter5: Random variables

## 5.1 Definition

**Exercise.** Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any $A \in \sigma(X_1, X_2, \dots)$ and any $\epsilon > 0$, show that there is some $n \geq 1$ and some $B \in \sigma(X_1, \dots, X_n)$ such that $\mathbb{P}(A \Delta B) < \epsilon$. (Hint: Use Theorem 1.2.6.)

**Answer.** (Hao) We are going to show that $\mathcal{A} = \{A | \exists N \in \mathbb{N} \text{ s.t. } A \in \sigma(X_1, \dots, X_N)\}$ is an algebra that generate $\sigma(X_1, \dots)$. Then, Theorem 1.2.6 gives the desired result for the exercise.

First, let's show $\mathcal{A}$ is an algebra. ($\Omega \in \mathcal{A}$): $\Omega \in \sigma(X_1) \subset \mathcal{A}$. (Closed under taking complement): $A \in \mathcal{A}$ implies $\exists n \in \mathbb{N}, A \in \sigma(X_1, \dots, X_n)$, which means $A^c \in \sigma(X_1, \dots, X_n) \subset \mathcal{A}$. (Closed under finite union): If $A_1, \dots, A_k \in \mathcal{A}$ for some finite $k \in \mathbb{N}$, there exists $N_k \in \mathbb{N}$ s.t. $A_1, \dots, A_k \in \sigma(X_1, \dots, X_{N_k})$. Hence, $\cup_{i=1}^{k} A_i \in \sigma(X_1, \dots, X_{N_k}) \subset \mathcal{A}$, and we conclude that $\mathcal{A}$ is an algebra.

Second, $\sigma(\mathcal{A}) = \sigma(X_1, \dots)$. ($\subseteq$): Since $\mathcal{A} \subset \sigma(X_1, \dots)$. ($\supseteq$): By definition in this section, $\sigma(X_1, \dots) = \sigma(\cup_{n=1}^{\infty} \sigma(X_n))$. Then, due to $\sigma(X_n) \subset \sigma(X_1, \dots, X_n) \subset \mathcal{A}, n \in \mathbb{N}$, the $\sigma$-algebra generated by $\mathcal{A}$, $\sigma(\mathcal{A})$ contains $\sigma(X_1, \dots)$.

# 6 Chapter6: Expectation, variance, and other functionals

## 6.1 Expected value

**Exercise.** 6.1.1. TBD

**Answer.** (Hao) TBD

# 7 Chapter7: Independece

## 7.1 Definition

**Exercise.** 7.1.3. Show that a collection of events $\{A_i\}_{i \in I}$ are independent if and only if for any finite $J \subseteq I$,

$$\mathbb{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} A_j.$$

---

**Answer.** (Neng-Tai) For any $B_i \in \{A_i, A_i^c\}$. Where $j \in J$ finite index, $|J| = m + n \in \mathbb{N}$. Let $J = J_1 \bigcup J_2$, where $J_1 = \{j | B_j = A_j\}$, $J_2 = \{j | B_j = A_j^c\}$, and $|J_1| = m, |J_2| = n$,

$$\begin{aligned}
\mathbb{P}(\bigcap_{j \in J} B_j) &= \mathbb{P}(\bigcap_{j \in J_1} A_j \cap \bigcap_{j \in J_2} A_j^c) = \mathbb{P}(\bigcap_{j \in J_1} A_j \cap (\Omega \backslash \bigcup_{j \in J_2} A_j)) \\
&= \mathbb{P}(\bigcap_{j \in J_1} A_j) - \mathbb{P}(\bigcap_{j \in J_1} A_j \cap (\bigcup_{j \in J_2} A_j)) \qquad \text{(Disjoint additivity)} \\
&= \prod_{j \in J_1} \mathbb{P}(A_j) - \mathbb{P}(\bigcup_{k \in J_2} (A_k \bigcap_{j \in J_1} A_j)) \\
&= \prod_{j \in J_1} \mathbb{P}(A_j) - \left[ \sum_{l=1}^{n} \sum_{\{j_1, \ldots, j_l\} \subseteq J_2} (-1)^{l-1} \mathbb{P}((\bigcap_{j \in \{j_1, \ldots, j_l\}} A_j) \cap \bigcap_{j \in J_1} A_j) \right] \\
&\qquad \text{(Includsion-exclusion principle)} \\
&= \prod_{j \in J_1} \mathbb{P}(A_j) \left[ 1 - \sum_{l=1}^{n} \sum_{\{j_1, \ldots, j_l\} \subseteq J_2} (-1)^{l-1} \prod_{j \in \{j_1, \ldots, j_l\}} \mathbb{P}(A_j) \right] \\
&= \prod_{j \in J_1} \mathbb{P}(A_j) \prod_{j \in J_2} (1 - \mathbb{P}(A_j)) = \prod_{j \in J} \mathbb{P}(B_j) \qquad \blacksquare
\end{aligned}$$

**Exercise.** 7.1.5. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined on the same probability space. If $X_n$ is independent of $\sigma(X_1, ..., X_n)$ for each $n$, prove that the whole collection is independent.

**Answer.** 7.1.5. (Neng-Tai)
Need to show:
For any finite $J \subseteq \mathbb{N}$, $\mathbb{P}(\bigcap_{i \in J} B_i) = \prod_{i \in J} \mathbb{P}(B_i)$, where $B_i \in \sigma(X_i)$.
To show this, we perform induction on the cardinality of $J$.
For $|J| = 2$, let $J = \{i_1, i_2\}$, where $i_2 > i_1 \geq 1$. Since $\sigma(X_{i_2})$ and $\sigma(X_1, X_2, ..., X_{i_1}, ..., X_{i_2 - 1})$ are independent, and observe that $\sigma(X_{i_1}) \subseteq \sigma(X_1, X_2, ..., X_{i_1}, ..., X_{i_2-1}))$,
given any $B_{i_2} \in \sigma(X_{i_2})$ and $B_{i_1} \in \sigma(X_{i_1}) \subseteq \sigma(X_1, X_2, ..., X_{i_1}, ..., X_{i_2-1}))$,
we have $\mathbb{P}(B_{i_1} \cap B_{i_2}) = \mathbb{P}(B_{i_1})\mathbb{P}(B_{i_2})$.
Suppose the statement holds for $|J| = n$, where $n \geq 2$.
For $|J| = n + 1$, let $J = \{i_1, i_2, ..., i_{n+1}\}$, where $1 \leq i_1 < i_2 < ... < i_{n+1}$.
Given any $B_{i_k} \in \sigma(X_{i_k})$, where $k \in \{1, ..., n+1\}$, observe that:

$$B_{i_k} \in \sigma(X_{i_k}) \subseteq \sigma(X_1, X_2, ..., X_{i_{n+1}-1}), \forall k \in \{1, ..., n\} \Longrightarrow \bigcap_{k=1}^{n} B_{i_k} \in \sigma(X_1, X_2, ..., X_{i_{n+1}-1}).$$

Since $\sigma(X_{i_{n+1}})$ and $\sigma(X_1, X_2, ..., X_{i_{n+1}-1})$ are independent sub $\sigma-$ algebras,
$\implies \mathbb{P}(\bigcap_{k=1}^{n} B_{i_k} \cap B_{i_{n+1}}) = \mathbb{P}(\bigcap_{k=1}^{n} B_{i_k})\mathbb{P}(B_{i_{n+1}})$,
By induction hypothesis, the statement holds for $|J| = n$, hence $\mathbb{P}(\bigcap_{k=1}^{n} B_{i_k}) = \prod_{k=1}^{n} \mathbb{P}(B_{i_k})$
$\implies \mathbb{P}(\bigcap_{k=1}^{n+1} B_{i_k}) = \prod_{k=1}^{n+1} \mathbb{P}(B_{i_k})$. ■

**Exercise.** 7.1.6. If $X_1, \cdots, X_n$ are independent random variables and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a measurable function, show that the law of $f(X_1, \cdots, X_n)$ is determined by the laws of $X_1, \cdots, X_n$.

**Answer.** First I show that the law of $(X_1, \cdots, X_n)$ can be determined by the law of $X_i$. Let $\mu$ be the law of $(X_1, \cdots, X_n)$, and let $\mu_i$ be the law of $X_i$. Then for any $A_1 \times \cdots \times A_n$ where $A_i \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mu(A_1 \times \cdots \times A_n) &= \mathbb{P}\big(\{(X_1, \cdots, X_n) \in A_1 \times \cdots \times A_n\}\big) \\
&= \mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i \in A_i\}\Big) \\
&= \prod_{i=1}^{n} \mathbb{P}\big(\{X_i \in A_i\}\big) && \text{(By independence of } X_i\text{s)} \\
&= \prod_{i=1}^{n} \mu_i(A_i) && \text{(law of } X_i\text{s)} \\
&= \mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n),
\end{aligned}$$

where $\mu_1 \times \cdots \times \mu_n$ is a measure on $(\mathbb{R}^n, \mathcal{F})$ that agrees with $\prod_{i=1}^{n} \mu_i(A_i)$ on the set $\mathcal{P} = \{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}(\mathbb{R})\}$. (It exists by Proposition 3.1.1.)
Since $\mathcal{P}$ is closed under finite intersection, it is a $\pi$-system. Also, there is a sequence of subsets $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$ in $\mathcal{P}$ increasing to $\mathbb{R}^n$. Hence by Theorem 1.3.6, $\mu = \mu_1 \times \cdots \times \mu_n$ on $\sigma(\mathcal{P}) = \mathcal{F}$ since they agree on $\mathcal{P}$. Hence the law of $(X_1, \cdots, X_n)$ is exactly the product measure $\mu_1 \times \cdots \times \mu_n$.
Since $f : \mathbb{R}^n \to \mathbb{R}$ is measurable, the law of $f(X_1, \cdots, X_n)$, $\nu$, can be wrote as

$$\begin{aligned}
\nu(S) &= \mathbb{P}(\{f(X_1, \cdots, X_n) \in S\}) \\
&= \mathbb{P}\big(\{(X_1, \cdots, X_n) \in \{f \in S\}\}\big) \\
&= \mu(\{f \in S\}) \\
&= \mu_1 \times \cdots \times \mu_n(\{f \in S\}).
\end{aligned}$$

for any $S \in \mathcal{B}(\mathbb{R})$. ■

**Exercise.** 7.1.7. If $\{X_i\}_{i \in I}$ is a collection of independent random variables and $\{A_i\}_{i \in I}$ is a collection of measurable subsets of $\mathbb{R}$, show that the events $\{X_i \in A_i\}_{i \in I}$ are independent.

**Answer.** (Hao)
A random variable is a measurable function, and $A_i \in B(\mathbb{R})$, so the pullback of $A_i$ will be in the $\sigma(X_i)$, i.e $\sigma(X_i^{-1}(A_i)) \subset \sigma(X_i)$. Hence the statement is a direct implication that $\{\sigma(X_i)\}$ are independent based on the assumption. ■

**Exercise.** 7.1.8. If $\{F_i\}_{i\in I}$ is a family of cumulative distribution functions, show that there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent random variables $\{X_i\}_{i\in I}$ defined on $\Omega$ such that for each $i$, $F_i$ is the C.D.F. of $X_i$.(Hint: use product spaces.)

**Answer.** 7.1.8(Neng-Tai)
For any $i \in I$, let $\Omega_i = (0,1)$, $\mathcal{B}_i =$ the restriction of $\mathcal{B}(\mathbb{R})$ to $(0,1)$ and $\lambda_i$ be the restriction of Lebesgue measure to $(0,1)$, then $(\Omega_i, \mathcal{B}_i, \lambda_i)$ is a probability space.
For $\omega \in \Omega_i$, define $Y_i(\omega) \coloneqq \inf\{t \in \mathbb{R} : F_i(t) \geq \omega\}$, where $F_i's$ are the given cumulative distribution functions. By Proposition 5.2.2, $\forall i \in I, Y_i$ is measurable and $F_i$ is the C.D.F of $Y_i \implies F_i(t) = \lambda_i(Y_i(\omega) \leq t)$.
Set $\Omega = \prod_{i\in I}\Omega_i$, and the product $\sigma$-algebra be $\mathcal{F} = \prod_{i\in I}\mathcal{B}_i$, then exist a unique probability measure $\mathbb{P} = \prod_{i\in I}\lambda_i$ on $(\Omega, \mathcal{F})$.
Now we will begin to construct the required independent random variables $X_i$. Let $p_i : \Omega \to \Omega_i$ be the canonical projection defined by $p_i(\omega) \coloneqq \omega_i$, where $\omega = \prod_{j\in I}\omega_j$. Observe that for any $\omega_i \in \mathcal{B}_i$, $p_i^{-1}(\omega_i) = \prod_{j\in I\setminus\{i\}}\Omega_j \times \omega_i \in \mathcal{F}$, hence $p_i$ is measurable.
Then we can define $X_i : \Omega \to \mathbb{R}$ by $X_i(\omega) \coloneqq Y_i \circ p_i(\omega) \implies X_i$ is measurable.

Claim: $F_i$ is the C.D.F of $X_i$.
Proof. Let $G_i(t)$ be the C.D.F of $X_i$, then

$$\begin{aligned}
G_i(t) &= \mathbb{P}(X_i(\omega) \leq t) = \mathbb{P}(p_i^{-1} \circ Y_i^{-1}(-\infty, t]) \\
&= \mathbb{P}(\prod_{j\in I\setminus\{i\}} \Omega_j \times Y_i^{-1}(-\infty, t]) \\
&= \prod_{j\in I\setminus\{i\}} \lambda_j(\Omega_j) \times \lambda_i(Y_i^{-1}(-\infty, t]) \\
&= \lambda_i(Y_i^{-1}(-\infty, t]) = F_i(t)
\end{aligned}$$

□

Claim: $X_i$'s are independent random variables.
Proof: Let $J \subseteq I$ be a finite set, $\{A_i : A_i \in \sigma(X_i)\}_{i\in J} \subseteq \Omega$.
Need to show $\mathbb{P}(\bigcap_{i\in J} A_i) = \prod_{i\in J}\mathbb{P}(A_i)$.
Observe that for any $A_i \in \sigma(X_i)$, $A_i = \prod_{s\in I\setminus\{i\}}\Omega_s \times \omega_i$, where $\omega_i \in \sigma(Y_i)$.

Hence,

$$\begin{aligned}
\mathbb{P}(\bigcap_{i\in J} A_i) &= \mathbb{P}\Big(\bigcap_{i\in J}(\prod_{j\in I\setminus\{i\}}\Omega_j \times \omega_i)\Big)\\
&= \mathbb{P}\Big((\prod_{s\in I\setminus J}\Omega_s)\times(\prod_{i\in J}\omega_i)\Big)\\
&= \Big(\prod_{s\in I\setminus J}\lambda_s(\Omega_s)\Big)\Big(\prod_{i\in J}\lambda_i(\omega_i)\Big) = \prod_{i\in J}\lambda_i(\omega_i)\\
&= \prod_{i\in J}\mathbb{P}(\prod_{s\in I\setminus\{i\}}\Omega_s\times\omega_i) = \prod_{i\in J}\mathbb{P}(A_i)
\end{aligned}$$

■

**Exercise.** 7.1.9. If $\mathcal{A}$ is a $\pi$-system of sets and $B$ is an event such that $B$ and $A$ are independent for every $A \in \mathcal{A}$, show that $B$ and $\sigma(\mathcal{A})$ are independent.

**Answer.** $\mathcal{A}$ is a $\pi$-system. B is the set that is independent to all elements in $\mathcal{A}$.
The strategy is to construct a larger $\lambda$-system $\mathcal{L}$ containing $\mathcal{A}$, with its elements are all independent to $B$. Since Dynkin's $\pi-\lambda$ theorem tells us that $\sigma(\mathcal{A}) \subset \mathcal{L}$, then the proof would be completed.
Consider the set that collects all the event that is independent to $B$ in $\mathcal{F}$:

$$\mathcal{L} = \{C \in \mathcal{F} : \mathbb{P}(B\cap C) = \mathbb{P}(B)\mathbb{P}(C)\}.$$

Check $\mathcal{L}$ is a $\lambda$-system. First $\Omega \in \mathcal{L}$ since $\mathbb{P}(\Omega\cap B) = \mathbb{P}(B) = \mathbb{P}(\Omega)\mathbb{P}(B)$. Secondly if $L \in \mathcal{L}$ then $\mathbb{P}(L\cap B) = \mathbb{P}(B)\mathbb{P}(L)$. Since $\mathbb{P}(L^c\cap B) = \mathbb{P}(B\setminus B\cap L) = \mathbb{P}(B) - \mathbb{P}(B\cap L)$ by $(B\cap L)\cup(B\cap L^c)$ is a disjoint union, we then have $\mathbb{P}(L^c\cap B) = (1-\mathbb{P}(L))\mathbb{P}(B) = \mathbb{P}(L^c)\mathbb{P}(B)$. So $L^c \in \mathcal{L}$. Lastly, suppose $L_1, L_2, \cdots$ is a sequence of disjoint sets in $\mathcal{L}$, then

$$\begin{aligned}
\mathbb{P}(\bigcup_{n=1}^{\infty} L_n\cap B) &= \mathbb{P}(\bigcup_{n=1}^{\infty}(L_n\cap B))\\
&= \sum_{n\in\mathbb{N}}\mathbb{P}(L_n\cap B) \quad ((L_n\cap B)\subset L_n \text{ and } L_n\text{s are disjoint})\\
&= \sum_{n\in\mathbb{N}}\mathbb{P}(L_n)\mathbb{P}(B) \quad (L_n\in\mathcal{L})\\
&= \mathbb{P}(B)\mathbb{P}(\bigcup_{n=1}^{\infty}L_n) \quad (L_n \text{ disjoint}).
\end{aligned}$$

So $\mathcal{L}$ forms a $\lambda$-system. Since all the elements $A \in \mathcal{A}$ are independent to $B$, we have $\mathcal{A} \subset \mathcal{L}$.
Now by Dynkin's $\pi-\lambda$ theorem, $\sigma(\mathcal{A}) \subset \mathcal{L}$ since $\mathcal{A}$ is a $\pi$-system. Hence $\sigma(\mathcal{A})$ is independent to $B$. ■

**Exercise.** 7.1.10. If $\{X_i\}_{i\in I}$ is a collection of independent random variables, then show that for any disjoint subsets $J, K \subseteq I$, the $\sigma$-algebras generated by $\{X_i\}_{i\in J}$ and $\{X_i\}_{i\in K}$ are independent.

**Answer.** $\{X_i\}_{i\in I}$ is a collection of independent random variables. Let $J, K \subseteq I$ be disjoint. I want to show that $\sigma(\{X_i\}_{i\in J})$ and $\sigma(\{X_i\}_{i\in K})$ are independent.
Define $\mathcal{P}_J$, $\mathcal{P}_K$ as follows:

$$\mathcal{P}_J = \Big\{ \bigcap_{i=1}^{n} \{X_{j_i} \in A_i\} : A_i \in \mathcal{B}(\mathbb{R}), \{j_1, \cdots, j_n\} \subset J \Big\}.$$

$$\mathcal{P}_K = \Big\{ \bigcap_{i=1}^{n} \{X_{k_i} \in A_i\} : A_i \in \mathcal{B}(\mathbb{R}), \{k_1, \cdots, k_n\} \subset K \Big\}.$$

They're $\pi$-systems since the intersection is closed in $\mathcal{P}_J$ ($\mathcal{P}_K$):

$$\left(\bigcap_{n=1}^{N_1} \{X_{i_n} \in A_n\}\right) \bigcap \left(\bigcap_{n=1}^{N_2} \{X_{j_n} \in B_n\}\right) = \bigcap_{n=1}^{N} \{X_{k_n} \in C_n\}$$

for some $\{k_1, \cdots, k_N\} \subseteq \{i_1, \cdots, i_{N_1}\} \cup \{j_1, \cdots, j_{N_2}\}$ and some Borel sets $C_n$.
We can also see that $\{X_j \in A\} \in \mathcal{P}_J$ for any $j \in J$ and $A \in \mathcal{B}(\mathbb{R})$, hence

$$\bigcup_{j\in J} \sigma(X_j) = \bigcup_{j\in J} \Big\{ \{X_j \in A\} : A \in \mathcal{B}(\mathbb{R}) \Big\} \subset \mathcal{P}_J.$$

Thus $\sigma(\{X_j\}_{j\in J}) \subseteq \sigma(\mathcal{P}_J)$. Similarly $\sigma(\{X_k\}_{k\in K}) \subseteq \sigma(\mathcal{P}_K)$.
Let $\cap_{n=1}^{N_1} \{X_{j_n} \in A_n\} \in \mathcal{P}_J$ and $\cap_{n=1}^{N_2} \{X_{k_n} \in B_n\} \in \mathcal{P}_K$, then by independence of $\{\sigma(X_i)\}_{i\in I}$:

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{n=1}^{N_1} \{X_{j_n} \in A_n\} \cap \bigcap_{n=1}^{N_2} \{X_{k_n} \in B_n\}\right) &= \prod_{n=1}^{N_1} \mathbb{P}(\{X_{j_n} \in A_n\}) \times \prod_{n=1}^{N_2} \mathbb{P}(\{X_{k_n} \in B_n\}) \\
&= \mathbb{P}\left(\bigcap_{n=1}^{N_1} \{X_{j_n} \in A_n\}\right) \mathbb{P}\left(\bigcap_{n=1}^{N_2} \{X_{k_n} \in B_n\}\right).
\end{aligned}$$

Hence every elements in $\mathcal{P}_J$ are independent to all elements in $\mathcal{P}_K$. By Exercise 7.1.9 this implies $\sigma(\mathcal{P}_J)$ is independent to $\sigma(\mathcal{P}_K)$. Since $\sigma(\{X_j\}_{j\in J}) \subseteq \sigma(\mathcal{P}_J)$ and $\sigma(\{X_k\}_{k\in K}) \subseteq \sigma(\mathcal{P}_K)$, $\sigma(\{X_j\}_{j\in J})$ is independent to $\sigma(\{X_k\}_{k\in K})$. ■

**Exercise.** 7.1.11. Let $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ be two collections of random variables defined on the same probability space. Suppose that for any finite $F \subseteq I$ and $G \subseteq J$, the collections $\{X_i\}_{i\in F}$ and $\{Y_j\}_{j\in G}$ are independent. Then prove that $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ are independent.

**Answer.** $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ are two collections of random variables defined on the same probability space. For any finite $F \subseteq I$ and $G \subseteq J$ the collections $\{X_i\}_{i\in F}$ and $\{Y_j\}_{j\in G}$ are independent. I'm going to prove that $\sigma(\{X_i\}_{i\in I})$ and $\sigma(\{Y_j\}_{j\in J})$ are independent.
Similar to Exercise 7.1.10, define the collection $\mathcal{P}_I$ and $\mathcal{P}_J$ as

$$\mathcal{P}_X = \Big\{ \bigcap_{f\in F} \{X_f \in A_f\} : A_f \in \mathcal{B}(\mathbb{R}),\, F \subseteq I,\, |F| < \infty \Big\}.$$
$$\mathcal{P}_Y = \Big\{ \bigcap_{g\in G} \{Y_g \in A_g\} : A_g \in \mathcal{B}(\mathbb{R}),\, G \subseteq J,\, |G| < \infty \Big\}.$$

Then they're $\pi$-systems. Since $\{X_i \in A\} \in \mathcal{P}_X$ for all $A \in \mathcal{B}(\mathbb{R})$ and $i \in I$,

$$\bigcup_{i\in I} \sigma(X_i) = \bigcup_{i\in I} \Big\{ \{X_i \in A\} : A \in \mathcal{B}(\mathbb{R}) \Big\} \subset \mathcal{P}_X.$$

Thus $\sigma(\{X_i\}_{i\in I}) \subseteq \sigma(\mathcal{P}_X)$, so does $\sigma(\{Y_j\}_{j\in J}) \subseteq \sigma(\mathcal{P}_Y)$.
Let $S_X \in \mathcal{P}_X$, then $S_X = \bigcap_{f\in F}\{X_f \in A_f\} \in \sigma(\{X_f\}_{f\in F})$ for some finite $F \subseteq I$ and some Borel sets $A_f$. If $S_Y \in \mathcal{P}_Y$ we also have $S_Y \in \sigma(\{Y_g\}_{g\in G})$ for some finite $G \subseteq J$. Since $\sigma(\{X_f\}_{f\in F})$ and $\sigma(\{Y_g\}_{g\in G})$ are independent for any finite $F \subseteq I$ and $G \subseteq J$, $S_X$ and $S_Y$ are independent.
Then by Exercise 7.1.9, all elements in $\mathcal{P}_X$ are independent to all elements in $\mathcal{P}_Y$ implies $\sigma(\mathcal{P}_X)$ and $\sigma(\mathcal{P}_Y)$ are independent. Hence $\sigma(\{X_i\}_{i\in I})$ and $\sigma(\{Y_j\}_{j\in J})$ are independent. ■

## 7.2 Expectation of a product under independence

**Exercise.** 7.2.2. If $X_1, X_2, \ldots, X_n$ are independent integrable random variables, show that the product $X_1X_2X_n$ is also integrable and

$$\mathbb{E}(X_1X_2\cdots X_n) = \mathbb{E}(X_1)\mathbb{E}(X_2)\cdots\mathbb{E}(X_n).$$

**Answer.** (Hao)
One method is to follow the procedure in the proposition 7.2.1 but do it for $n$ random variables. There is yet another way that we can do this, induction. In order to make induction work, we first have to show that the product of two random variable is still a random variable, so we can make induction w.r.t. the number of random variables $n$. In other words, we first have to show that if $X, Y$ are both measurable function, then $XY$ is also a measurable function. Fortunately, we can refer to exercise 2.1.5 to see that $X + Y$ and $X^2$ are also measurable. Hence

$$XY = \frac{1}{4}(X+Y)^2 - (X-Y)^2$$

is measurable, i.e $XY$ is a random variable. We can now proceed with the induction.

The base case $n = 2$ is shown in the proposition 7.2.1. Suppose the statement of this exercise is correct up to $n = m-1, m \in \mathbb{N}$ (induction hypothesis), and that $X_1X_2\dots X_{m-1}$ is integrable and is a random variable. Then for $n = m$, we may consider the two random variable $X_1X_2\dots X_{m-1}$ and $X_m$. Again, by the proposition 7.2.1 we get the result that $X_1X_2\dots X_m$ is also integrable and

$$\mathbb{E}(X_1X_2\dots X_m) = \mathbb{E}(X_1X_2\dots X_{m-1})\mathbb{E}(X_m) = \mathbb{E}(X_1)\mathbb{E}(X_2)\cdots\mathbb{E}(X_n).$$

The first equality works because we treat $X_1X_2\dots X_{m-1}$ as one integrable random variable; the second equality works because of our induction hypothesis.

Hence, the statement of this exercise is true for all $n \in \mathbb{N}$.

■

**Exercise.** 7.2.3. If $X$ and $Y$ are independent integrable random variables, show that $\mathrm{Cov}(X, Y) = 0$. In other words, independent random variables are uncorrelated.

**Answer.** Suppose $X, Y$ are two independent random variables on the same probability space, then $\mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) = \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)) = 0$ since $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ by independence.

■

**Exercise.** 7.2.4
Give an example of a pair of uncorrelated random variables that are not independent.

**Answer.** (Hao)
Set $X \sim \mathrm{Unif}[-1, 1],\ Y = X^2$. Then,

$$\mathbb{E}(XY) = \mathbb{E}(X^3) = \int_{-1}^{1} x^3 \frac{1}{2} dx = 0 = 0 \cdot \frac{2}{3} = (\int_{-1}^{1} x dx) \cdot (\int_{0}^{1} y \frac{1}{\sqrt{y}} dy) = \mathbb{E}(X)\mathbb{E}(Y)$$

However, $X$ and $Y$ are clearly not independent because

$$\mathbb{P}(X \in [-0.5, 0.5], Y \in [1/\sqrt{2}, 1]) = 0 \neq \frac{1}{2} \cdot 1 = \mathbb{P}(X \in [-0.5, 0.5])\mathbb{P}(Y \in [1/\sqrt{2}, 1])$$

■

**Exercise.** 7.2.5 Give an example of three random variables $X_1, X_2, X_3$ that are pairwise independent but not independent.

**Answer.** (Hao)

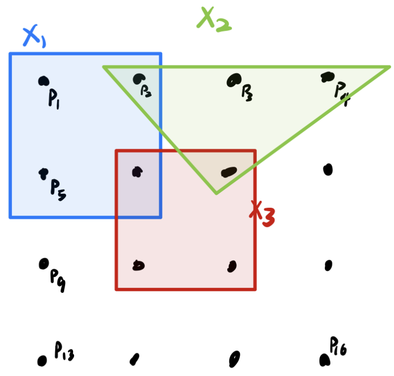


Figure 1: An example that is pairwise independent but not independent

The example is an discrete probability space, $\Omega = \{p_i\}_{i=1}^{16}$, $\mathcal{F} = \sigma(\Omega)$. The probability measure is $\mathbb{P}(A) = \frac{||A||}{16}$. In other words, counting how much points $p_i$ are selected. Let's define

$$X_1 = 1_{\{p_1, p_2, p_5, p_6\}}$$
$$X_2 = 1_{\{p_2, p_3, p_4, p_6\}}$$
$$X_3 = 1_{\{p_6, p_7, p_{10}, p_{11}\}}$$

From exercise 7.1, we know that we only have to check the case with $A_i = \{X_i = 1\}$. For example, we take $X_1$ and $X_2$

$$\mathbb{P}(X_1 = 1 \cap X_2 = 1) = \mathbb{P}(\{p_2\}) = \frac{1}{16} = (\frac{1}{4})^2 \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)$$

Hence, $\{X_i\}_{i=1}^3$ is pairwise independent. However, the collection $\{X_i\}_{i=1}^3$ is not independent because

$$\mathbb{P}(\bigcap_{i=1}^{3} X_i = 1) = 0 \neq (\frac{1}{4})^3 = \prod_{i=1}^{3} \mathbb{P}(X_i = 1)$$

■

## 7.3 The second Borel-Cantelli lemma

## 7.4 The Kolmogorov zero-one law

**Exercise.** 7.4.2. If $\mathcal{G}$ is a trivial $\sigma$-algebra for a probability measure $\mathbb{P}$, show that any random variable $X$ that is measurable with respect to $\mathcal{G}$ must be equal to a constant almost surely.

**Answer.** Since $\mathbb{P}\big(\cup_{n\in\mathbb{Z}}\{X\in(n,n+1]\}\big)=\sum_{n\in\mathbb{N}}\mathbb{P}(X\in(n,n+1])=1$ and $X$ is $\mathcal{G}$-measurable, there is a unique integer $m$ s.t. $\mathbb{P}(X\in(m,m+1])=1$. Set $I_1=(m,m+1]$, construct $I_k$ for $k\geq 2$ by following: suppose $I_{k-1}=(a,b]$,

■

**Exercise.** 7.4.3. If $\{X_n\}_n^\infty$ is a sequence of independent random variables, show that the random variables $\limsup X_n$ and $\liminf X_n$ are equal to constants almost surely.

**Answer.** Let $a\in\mathbb{R}$. Since $\bigcup_{k\geq n}\{X_k>a\}$ decreases as $n\to\infty$,

$$\Big\{\limsup_{n\to\infty}X_n>a\Big\}=\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n}\{X_k>a\}=\bigcap_{n\geq N}\bigcup_{k\geq n}\{X_k>a\}\quad\forall N\in\mathbb{N}.$$

Hence $\{\limsup_{n\to\infty}X_n>a\}\in\sigma(X_N,X_{N+1},\cdots)$ for all $N\in\mathbb{N}$. That is, $\limsup_{n\to\infty}X_n$ is $\bigcap_{n\in\mathbb{N}}\sigma(X_n,X_{n+1},\cdots)$-measurable. Therefore by EXERCISE 7.4.2, $\limsup_{n\to\infty}X_n$ equals a constant almost surely. For $\liminf_{n\to\infty}X_n$, since $\bigcup_{k\geq n}\{X_k<a\}$ decreases as $n\to\infty$,

$$\Big\{\liminf_{n\to\infty}X_n<a\Big\}=\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n}\{X_k<a\}=\bigcap_{n\geq N}\bigcup_{k\geq n}\{X_k<a\}\quad\forall N\in\mathbb{N}.$$

Then the same reason holds. Hence $\liminf_{n\to\infty}X_n$ equals a constant almost surely, too.

■

**Exercise.** 7.4.4. Let $\{X_n\}_{n=1}^\infty$ be a sequence of independent random variables. Let $S_n:=X_1+\cdots+X_n$, and let $\{a_n\}_{n=1}^\infty$ be a sequence of constants increasing to infinity. Then show that $\limsup S_n/a_n$ and $\liminf S_n/a_n$ are constants almost surely.

**Answer.** *Proof*. Let $S_{k+1,n}=\sum_{i=k+1}^n X_i$, denote $S_{1,k}=S_k$. Claim:

$$\limsup_{n\to\infty}\frac{S_n}{a_n}=\limsup_{n\to\infty}\frac{S_{k+1,n}}{a_n}\quad\text{and}\quad\liminf_{n\to\infty}\frac{S_n}{a_n}=\liminf_{n\to\infty}\frac{S_{k+1,n}}{a_n}$$

for any $k\in\mathbb{N}$.
(Proof of claim) Let $k\in\mathbb{N}$. Since $a_n$ increases to infinity, there is a sufficiently large $N>k$ s.t. $a_n>0\ \forall n\geq N$. Suppose $S_k=\sum_{i=1}^k X_i\geq 0$, then for any $m\geq n\geq N$,

$$\frac{S_m}{a_m}\leq\frac{S_{1,k}}{a_n}+\frac{S_{k+1,m}}{a_m}$$

Hence

$$\sup_{m\geq n}\left\{\frac{S_m}{a_m}\right\}\leq\frac{S_k}{a_n}+\sup_{m\geq n}\left\{\frac{S_{k+1,m}}{a_m}\right\}\text{ and }\inf_{m\geq n}\left\{\frac{S_m}{a_m}\right\}\leq\frac{S_k}{a_n}+\inf_{m\geq n}\left\{\frac{S_{k+1,m}}{a_m}\right\}.$$

Take $n \to \infty$. Since $a_n \uparrow \infty$, $S_k/a_n \downarrow 0$. Thus

$$\limsup_{n\to\infty} \frac{S_n}{a_n} \le \limsup_{n\to\infty} \frac{S_{k+1,n}}{a_n} \text{ and } \liminf_{n\to\infty} \frac{S_n}{a_n} \le \liminf_{n\to\infty} \frac{S_{k+1,n}}{a_n}.$$

For the other direction, since $a_m > 0$ when $m \ge N$,

$$\frac{S_m}{a_m} = \frac{S_k}{a_m} + \frac{S_{k+1,m}}{a_m} \ge \frac{S_{k+1,m}}{a_m}$$

Thus we have

$$\limsup_{n\to\infty} \frac{S_n}{a_n} \ge \limsup_{n\to\infty} \frac{S_{k+1,n}}{a_n} \text{ and } \liminf_{n\to\infty} \frac{S_n}{a_n} \ge \liminf_{n\to\infty} \frac{S_{k+1,n}}{a_n}.$$

Hence the claim holds for $S_k \ge 0$.

If $S_k < 0$, then we can check for any $m \ge n \ge N$,

$$\begin{aligned} \frac{S_m}{a_m} &\ge \frac{S_k}{a_n} + \frac{S_{k+1,m}}{a_m}; \\ \frac{S_m}{a_m} &= \frac{S_k}{a_m} + \frac{S_{k+1,m}}{a_m} \le \frac{S_{k+1,m}}{a_m} \end{aligned}$$

Then repeat the process we still have the same conclusion. $\square$
With this fact, we have

$$\begin{aligned} \left\{ \limsup_{n\to\infty} \frac{S_n}{a_n} > a \right\} &= \bigcap_{k\in\mathbb{N}} \left\{ \limsup_{n\to\infty} \frac{S_{k+1,n}}{a_n} > a \right\} \\ \left\{ \liminf_{n\to\infty} \frac{S_n}{a_n} < a \right\} &= \bigcap_{k\in\mathbb{N}} \left\{ \liminf_{n\to\infty} \frac{S_{k+1,n}}{a_n} < a \right\} \end{aligned}$$

Check: $\limsup_{n\to\infty} S_{k+1,n}/a_n$ is $\sigma(X_k, X_{k+1}, \cdots)$-measurable.
For any $n > k$, $S_{k+1,n} = f(X_{k+1}, \cdots, X_n)$ where $f : (x_1, \cdots, x_{n-k}) \mapsto x_1 + \cdots + x_{n-k}$. Since $f$ is continuous and thus measurable, by Exercise 5.1.1. we know that $S_{k+1,n}$ is $\sigma(X_{k+1}, \cdots, X_n)$-measurable, and thus $\sigma(X_{k+1}, \cdots)$-measurable. This implies that $\limsup_{n\to\infty} S_{k+1,n}/a_n$ is $\sigma(X_k, X_{k+1}, \cdots)$-measurable.
Hence

$$\left\{ \limsup_{n\to\infty} \frac{S_n}{a_n} > a \right\} \in \bigcap_{k\in\mathbb{N}} \sigma(X_k, X_{k+1}, \cdots) = \mathcal{T}.$$

Since $\mathcal{T}$ is a trivial sigma algebra and $\limsup_{n\to\infty} S_n/a_n$ is $\mathcal{T}$-measurable, $\limsup_{n\to\infty} S_n/a_n$ is a constant almost surely by Exercise 7.4.2.

$\blacksquare$

## 7.5 Zero-one laws for i.i.d. random variables

**Exercise.** 7.5.5. Suppose that we have $m$ boxes, and an infinite sequence of balls are dropped into the boxes independently and uniformly at random. Set this up as a problem in measure-theoretic probability, and prove that with probability one, each box has the maximum number of balls among all boxes infinitely often.

*Proof.* Let $\Omega = \{1, \cdots, m\}^{\mathbb{N}}$ be the sample space. Define the random variables $\{X_n\}_{n\in\mathbb{N}}$ on $\Omega$ such that $X_n(\omega) = \omega(n) = \omega_n$ for all $\omega$ and $n$. (Simply $\{X_n\}_{n\in\mathbb{N}}(\omega) = \omega$)
Set $k \in [m]$, and consider the subset $E \subseteq \Omega$:

$$E = \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \{\omega \in \Omega : k \text{ is the maximal appearance in } (\omega_1, \cdots, \omega_n)\},$$

and denotes the event $E_n = \{\omega \in \Omega : k \text{ is the maximal appearance in } \omega_1, \cdots, \omega_n\}$.

Since $E = \bigcap_{N\geq M} \bigcup_{n\geq N} E_n$ for any $M \in \mathbb{N}$, $E$ is permutation invariant. Hence if we define $f = \mathbb{1}_E$, then $\mathbb{1}_E\big((\omega_i)_{i\in\mathbb{N}}\big) = \mathbb{1}_E\big((\omega_{\sigma(i)})_{i\in\mathbb{N}}\big)$ since $\omega \in E$ is equivalent to $\omega_{\sigma(i)i\in\mathbb{N}} \in E$ for any permutation $\sigma$. This also means $f(\{X_i\}_{i\in\mathbb{N}}) = f(\{X_{\sigma(i)}\}_{i\in\mathbb{N}})$ on $\Omega$, then by COROLLARY 7.5.4, $f(\{X_i\}_{i\in\mathbb{N}})$ is a constant almost surely.

Since $\{f(X_1, X_2, \cdots) = 1\} = \{(X_1(\omega), X_2(\omega), \cdots) \in E\} = E$, $\mathbb{P}(E) \in \{0, 1\}$. Moreover, since $\mathbb{P}\Big(\bigcup_{n\geq N} E_n\Big) \geq \mathbb{P}(E_N)$ for any $N \in \mathbb{N}$, and

$$\begin{aligned}
\mathbb{P}(E_N) &= \mathbb{P}\big(\{\omega \in \Omega : k \text{ is the maximal appearance in } (\omega_1, \cdots, \omega_N)\}\big) \\
&= 1 - \mathbb{P}\Big(\bigcup_{l\neq k}\{\omega : l \text{ is the maximal appearance in } (\omega_1, \cdots, \omega_N)\}\Big) \\
&\geq 1 - \sum_{l\neq k} \mathbb{P}\big(\{\omega : l \text{ is the maximal appearance in } (\omega_1, \cdots, \omega_N)\}\big) \\
&= 1 - (m-1)\mathbb{P}(E_N),
\end{aligned}$$

we have $\mathbb{P}\Big(\bigcup_{n\geq N} E_n\Big) \geq \mathbb{P}(E_N) \geq 1/m$ for any $N$.
Hence $\mathbb{P}(E) = \lim_{N\to\infty} \mathbb{P}\Big(\bigcup_{n\geq N} E_n\Big) \geq 1/m \neq 0$, i.e. $\mathbb{P}(E) = 1$. So this event happens with probability one. ∎

## 7.6 Random vectors

**Exercise.** 7.6.1. Prove that the c.d.f. of a random vector uniquely characterizes its law.

*Proof.* Claim that $\mathcal{R} = \{(-\infty, b_1] \times \cdots \times (-\infty, b_n] : b_i \in \mathbb{R}\}$ generates $\mathcal{B}_{\mathbb{R}^n}$. Since $\mathcal{B}_{\mathbb{R}^n} = \sigma\big(\big\{\{x \in \mathbb{R}^n : x_i \in E\} : E \in \mathcal{B}_{\mathbb{R}}, i \in [n]\big\}\big)$, it's clear that $\sigma(\mathcal{R}) \subseteq \mathcal{B}_{\mathbb{R}^n}$. For the other direction, since the collection $\{(-\infty, b] : b \in \mathbb{R}\}$ generates $\mathcal{B}_{\mathbb{R}}$, for any $i \in [n]$ and

$E \in \mathcal{B}_{\mathbb{R}^n}$ we have $\{x_i \in E\} \in \sigma\big(\big\{\{x_i \in (-\infty, b_i]\} : b_i \in \mathbb{R}\big\}\big)$, and thus $\{x_i \in E\} \in \sigma(\mathcal{R})$. Therefore, $\mathcal{B}_{\mathbb{R}^n} \subseteq \sigma(\mathcal{R})$.

Suppose $X, Y$ are two random vectors that have the same c.d.f $F$, then $\mu_X = \mu_Y$ on $\mathcal{R}$. Since $\mathcal{R}$ is a $\pi$-system (it's closed under finite intersection), by THEOREM 1.3.6 we have $\mu_X = \mu_Y$ on $\sigma(\mathcal{R}) = \mathcal{B}_{\mathbb{R}^n}$. Hence the law corresponding to $F$ is unique. ■

**Exercise.** 7.6.2. If $X_1, \cdots, X_n$ are independent random variables with laws $\mu_1, \cdots, \mu_n$, then show that the law of the random vector $(X_1, \cdots, X_n)$ is the product measure $\mu_1 \times \cdots \times \mu_n$.

**Answer.** Let $\mu$ be the law of random vector $(X_1, \cdots, X_n)$. Let $\mathcal{R} = \{E_1 \times \cdots \times E_n : E_i \in \mathcal{B}_{\mathbb{R}}\}$. Since $X_1, \cdots, X_n$ are independent,

$$\begin{aligned}
\mu(E_1 \times \cdots \times E_n) &= \mathbb{P}(X_1 \in E_1, \cdots, X_n \in E_n) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in E_i) \\
&= \prod_{i=1}^{n} \mu_i(E_i) = \mu_1 \times \cdots \times \mu_n(E_1 \times \cdots \times E_n).
\end{aligned}$$

Hence $\mu$ and $\mu_1 \times \cdots \times \mu_n$ agrees on $\mathcal{R}$. Since $\mathcal{R}$ is a $\pi$-system that generates $\mathcal{B}_{\mathbb{R}^n}$, by THEOREM 1.3.6, $\mu = \mu_1 \times \cdots \times \mu_n$ on $\mathcal{B}_{\mathbb{R}^n}$. Therefore $\mu_1 \times \cdots \times \mu_n$ is the law of $(X_1, \cdots, X_n)$.

■

**Exercise.** 7.6.3. If $X_1, \cdots, X_n$ are independent random variables, show that the cumulative distribution function and the characteristic function of the random vector $(X_1, \cdots, X_n)$ can be written as products of one-dimensional distribution functions and characteristic functions.

**Answer.** Let $F_i$ be the c.d.f of $X_i$ and $F$ be the c.d.f of random vector $(X_1, \cdots, X_n)$. Since $X_1, \cdots, X_n$ are independent, by EXERCISE 7.6.2 we have $F(x_1, \cdots, x_n) = \mu_1 \times \cdots \times \mu_n((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = \prod_{i=1}^n \mu_i((-\infty, x_i]) = \prod_{i=1}^n F_i(x_i)$.

Claim: If $X_1, \cdots, X_n$ are independent, let $t \in \mathbb{R}$, then the complex-valued random variables $\{Y_k = \exp(itX_k)\}_{k=1}^n$ are independent.
(Proof of Claim) Since $x \mapsto \exp(itx)$ is continuous, it is measurable. Hence $\{x : \exp(itx) \in E\} \in \mathcal{B}_{\mathbb{R}}$ for any $E \in \mathcal{B}_{\mathbb{C}}$. Thus for any $E_1, \cdots, E_n \in \mathcal{B}_{\mathbb{C}}$, $\mathbb{P}(Y_1 \in E_1, \cdots, Y_n \in E_n) = \mathbb{P}(X_1 \in O_1, \cdots, X_n \in O_n)$ for some Borel sets $O_1, \cdots, O_n$ in $\mathbb{R}$. By independence of $X_1, \cdots, X_n$, this equals $\prod_{i=1}^n \mathbb{P}(X_i \in O_i) = \prod_{i=1}^n \mathbb{P}(Y_i \in E_i)$. Hence $Y_1, \cdots, Y_n$ are independent. □

By claim and independence, $\phi(t_1, \cdots, t_n) = \mathbb{E}\big(\prod_{i=1}^n e^{it_iX_i}\big) = \prod_{i=1}^n \mathbb{E}(e^{it_iX_i}) = \prod_{i=1}^n \phi_i(t_i)$.

■

**Exercise.** 7.6.4. If $(X_1, \cdots, X_n)$ are independent random variables, and each has a probability density function, show that $(X_1, \cdots, X_n)$ also has a p.d.f. and it is given by a product formula.

**Answer.** By EXERCISE 7.6.2, since $X_1, \cdots, X_n$ are independent, the random vector $(X_1, \cdots X_n)$ has the law $\mu_1 \times \cdots \times \mu_n$. Define $f : \mathbb{R}^n \to \mathbb{R}$ as $f(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n)$, then $f$ is measurable since $f_i$ are all measurable. Let $A_1, \cdots, A_n \in \mathcal{B}_{\mathbb{R}}$ and $\mu = \mu_1 \times \cdots \times \mu_n$, then by Fubini's theorem,

$$\begin{aligned}
\mu(A_1 \times \cdots \times A_n) &= \prod_{i=1}^n \mu_i(A_i) \\
&= \int_{A_1} f_1 d\mu_1 \times \cdots \times \int_{A_n} f_n d\mu_n \\
&= \int_{A_1} \cdots \int_{A_n} f_1(x_1) \cdots f_n(x_n) d\mu_n(x_n) \cdots d\mu_1(x_1) \\
&= \int_{A_1 \times \cdots \times A_n} f(x_1, \cdots, x_n) d\mu(x_1, \cdots, x_n) \qquad \text{(Fubini's Theorem)}
\end{aligned}$$

This means $\mu$ and $\nu$ agrees on $\mathcal{R} = \{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$ where $\nu(E) = \int_E f d\mu$ is the measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ defined by $f$. Moreover, $\mathcal{R}$ is a $\pi$-system that generates $\mathcal{B}_{\mathbb{R}^n}$, therefore by THEOREM 1.3.6 again, $\nu = \mu$ on all the Borel sets of $\mathbb{R}^n$. Since $\nu$ is exactly the law of $(X_1, \cdots, X_n)$, $f = f_1 \cdots f_n$ is thus the p.d.f of $(X_1, \cdots, X_n)$.

■

**Exercise.** 7.6.5. Let $X$ be an $\mathbb{R}^n$-valued random vector and $U \subseteq \mathbb{R}^n$ be an open set such that $\mathbb{P}(X \in U) = 1$. Suppose that the c.d.f. $F$ of $X$ is $n$ times differentiable in $U$. Let

$$f(x_1, \cdots, x_n) = \frac{\partial^n F}{\partial x_1 \cdots \partial x_n}$$

for $(x_1, \cdots, x_n) \in U$, and let $f = 0$ outside $U$. Prove that $f$ is the p.d.f. of $X$.

**Answer.** Let $F$ be the c.d.f. of random vector $X$. Denotes its partial derivative $\frac{\partial^m F}{\partial x_1 \cdots \partial x_m}$ as $\partial^m F$, and set $f = \partial^n F$ on $U$, $f = 0$ otherwise. First we show that $\mathbb{P}(X \in R) = \int_R f d^n x$ for any $R = (a_1, b_1] \times \cdots \times (a_n, b_n] \subseteq U$.

---

Given real numbers $a_1 < b_1, \cdots, a_n < b_n$, let $R_1, \cdots, R_n$ be the rectangles $(a_1, b_1], (a_1, b_1] \times (a_2, b_2], \cdots, (a_1, b_1] \times \cdots \times (a_n, b_n] = R$. Claim:

$$\mathbb{P}(X \in R_{n-1} \times (-\infty, b_n]) = \int_{R_{n-1}} \partial^{n-1} F(x_1, \cdots, x_{n-1}, b_n) d^{n-1}x.$$

Prove this by induction on $k = 1$ to $n-1$. For $k = 1$,

$$\begin{aligned}
&\mathbb{P}(X \in R_1 \times (-\infty, b_2] \times \cdots \times (-\infty, b_n]) \\
&= \mathbb{P}(X \le b_1, X \le b_2, \cdots, X_n \le b_n) - \mathbb{P}(X \le a_1, X_2 \le b_2, \cdots, X_n \le b_n) \\
&= F(b_1, b_2, \cdots, b_n) - F(a_1, b_2, \cdots, b_n) \\
&= \int_{(a_1, b_1]} \partial_{x_1} F(x_1, b_2, \cdots, b_n) dx_1 \\
&= \int_{R_1} \partial^1 F(x_1, b_2, \cdots, b_n) dx_1.
\end{aligned}$$

Now suppose for any $m < n-1$,

$$\mathbb{P}(X \in R_m \times (-\infty, b_{m+1}] \times \cdots \times (-\infty, b_n]) = \int_{R_m} \partial^m F(x_1, \cdots, x_m, b_{m+1}, \cdots, b_n) d^m x.$$

Then by induction hypothesis, when $k = n-1$,

$$\begin{aligned}
&\mathbb{P}(X \in R_{n-1} \times (-\infty, b_n]) \\
&= \mathbb{P}(X \in R_{n-2} \times (-\infty, b_{n-1}] \times (-\infty, b_n]) - \mathbb{P}(X \in R_{n-2} \times (-\infty, a_{n-1}] \times (-\infty, b_n]) \\
&= \int_{R_{n-2}} \big(\partial^{n-2} F(x_1, \cdots, x_{n-2}, b_{n-1}, b_n) - \partial^{n-2} F(x_1, \cdots, x_{n-2}, a_{n-1}, b_n)\big) d^{n-2}x \\
&= \int_{R_{n-2}} \Big( \int_{(a_{n-1}, b_{n-1}]} \partial_{x_{n-1}} \partial^{n-2} F(x_1, \cdots, x_{n-1}, b_n) dx_{n-1} \Big) d^{n-2}x \\
&= \int_{R_{n-1}} \partial^{n-1} F(x_1, \cdots, x_{n-1}, b_n) d^{n-1}x.
\end{aligned}$$

Hence the hypothesis is true for 1 to $n-1$, and thus the claim is true. Then

$$\begin{aligned}
\mathbb{P}(X \in R) &= \mathbb{P}(X \in R_{n-1} \times (-\infty, b_n]) - \mathbb{P}(X \in R_{n-1} \times (-\infty, a_n]) \\
&= \int_{R_{n-1}} \int_{(a_n, b_n]} \partial_{x_n} \partial^{n-1} F(x_1, \cdots, x_n) dx_n d^{n-1}x \\
&= \int_{R_n} (\partial^n F) d^n x = \int_{R_n} f d^n x.
\end{aligned}$$

Hence for any rectangle $R \subseteq U$ we have $\mathbb{P}(X \in R) = \int_R f d^n x$. $\square$

With this fact, let $\mathcal{R} = \{R \subseteq U : R = (a_1, b_1] \times \cdots \times (a_n, b_n]\}$, then $\mathcal{R}$ is a $\pi$-system since $\mathcal{R}$ is closed under finite intersection. Also, $\mathcal{R}$ generates $\mathcal{B}_U$. (Clearly every sets in $\mathcal{R}$ can be wrote as a countable union of open subsets in $U$, thus $\sigma(\mathcal{R}) \subseteq \mathcal{B}_U$. On the other hand, since $\mathbb{R}^n$ is separable, every open set in $U$ can be wrote as a countable union of elements in $\mathcal{R}$, hence $\mathcal{B}_U \subseteq \sigma(\mathcal{R})$.)

By EXERCISE 2.3.5., $\int_A f d^n x$ defines a measure on $(U, \mathcal{B}_U)$, say $\nu(A)$, and it agrees with the probability measure $\mu_X(A) = \mathbb{P}(X \in A)$ on $\mathcal{R}$. Then by THEOREM 1.3.6., since $\sigma(\mathcal{R}) = \mathcal{B}_U$, $\nu = \mu_X$ on $\mathcal{B}_U$. Hence $\mathbb{P}(X \in A) = \int_A f d^n x$ for all $A \in \mathcal{B}_U$.

Finally, for any $E \in \mathcal{B}_{\mathbb{R}^n}$, if $E \subseteq U$ then clearly $\int_E f d^n x = \mathbb{P}(X \in E)$. If $E \subseteq U^c$ then $\int_E f d^n x = 0 = \mathbb{P}(X \in E)$ since $\mathbb{P}(X \in E) \leq \mathbb{P}(X \in U^c) = 0$. Otherwise, $\int_E f d^n x = \int_{E \cap U} f d^n x = \mathbb{P}(X \in E \cap U) = \mathbb{P}(X \in E) - \mathbb{P}(X \in E \cap U^c) = \mathbb{P}(X \in E)$. Hence $\mathbb{P}(X \in E) = \int_E f d^n x$ for all Borel sets $E$ in $\mathbb{R}^n$, this means $f$ is the p.d.f. of random vector $X$.

■

**Exercise.** 7.6.6 Prove that the covariance matrix of any random vector is a positive semi-definite matrix.

**Answer.** Let $X \in \mathbb{R}^n$be a random vector, then the covariance matrix of $X$ is $E((X - \bar{X})(X - \bar{X})^T)$.
For any $u \in \mathbb{R}^n$, $u^T E((X - \bar{X})(X - \bar{X})^T)u = E(|u^T(X - \bar{X})|^2) \geq 0$.

■

**Exercise.** 7.6.7. Given $\mu \in \mathbb{R}^n$, and $\Sigma$ is a strictly positive definite matrix of order $n$. Show that the formula

$$f(x) = \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$$

is indeed a p.d.f. of a probability law on $\mathbb{R}^n$, and that this law has mean vector $\mu$ and a covariance matrix $\Sigma$.

**Answer.** *Proof.*(Neng-Tai)
(i). Show that $f(x)$ is a p.d.f on $\mathbb{R}^n$.
Need to show $\int_{\mathbb{R}^n} f(x)\lambda_{\mathbb{R}^n}(dx) = 1$, where $\lambda_{\mathbb{R}^n}$ is the Lebesgue measure on $\mathbb{R}^n$.
Since $\Sigma$ is positive definite, exist a unique square root of $\Sigma$ that is also positive definite, denote as $\Sigma^{\frac{1}{2}}$.

Define $g(x) = \Sigma^{\frac{1}{2}}(x) + \mu$, by the translation invariant property and the linear transformation property of Lebesgue measure, the pushforward measure $g_*\lambda_{\mathbb{R}^n}(x) = det(\Sigma^{-\frac{1}{2}})\lambda_{\mathbb{R}^n}(x)$, then

$$\begin{aligned}
\int_{\mathbb{R}^n} f \circ g(x)\lambda_{\mathbb{R}^n}(dx) = \int_{\mathbb{R}^n} f(x) g_*\lambda_{\mathbb{R}^n}(dx) &= det(\Sigma^{-\frac{1}{2}}) \int_{\mathbb{R}^n} f(x)\lambda_{\mathbb{R}^n}(dx) \\
&= det(\Sigma^{\frac{1}{2}})^{-1} \int_{\mathbb{R}^n} f(x)\lambda_{\mathbb{R}^n}(dx) \\
&= ((det\Sigma)^{\frac{1}{2}})^{-1} \int_{\mathbb{R}^n} f(x)\lambda_{\mathbb{R}^n}(dx) \qquad (1)
\end{aligned}$$

Also,

$$\begin{aligned}
\int_{\mathbb{R}^n} f \circ g(x)\lambda_{\mathbb{R}^n}(dx) &= \int_{\mathbb{R}^n} f(\Sigma^{\frac{1}{2}}(x) + \mu)\lambda_{\mathbb{R}^n}(dx) \\
&= \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} \int_{\mathbb{R}^n} exp(-\frac{1}{2}(\Sigma^{\frac{1}{2}}x)^T\Sigma^{-1}(\Sigma^{\frac{1}{2}}x))\lambda_{\mathbb{R}^n}(dx) \\
&= \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} \int_{\mathbb{R}^n} exp(-\frac{1}{2}\|x\|^2)\lambda_{\mathbb{R}^n}(dx) \\
&= \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} exp(-\frac{1}{2}(\Sigma_{i=1}^n x_i^2))\lambda_{\mathbb{R}}(dx_1)\dots\lambda_{\mathbb{R}}(dx_n) \\
&= \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \prod_{i=1}^n exp(-\frac{1}{2}(x_i^2))\lambda_{\mathbb{R}}(dx_1)\dots\lambda_{\mathbb{R}}(dx_n) \\
&= \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}}(2\pi)^{n/2} \\
&= \frac{1}{(det\Sigma)^{1/2}} \qquad (2)
\end{aligned}$$

Combine (1) and (2) $\Longrightarrow \int_{\mathbb{R}^n} f(x)\lambda_{\mathbb{R}^n}(dx) = 1$ $\square$

(ii). Show that this law has mean vector $\mu$ and a covariance matrix $\Sigma$.

Let $Z \in \mathbb{R}^n$ be a standard normal random vector, write $Z = (z_1, ..., z_n)^T$, where $z_i$'s are i.i.d., and $z_i \sim N(0, 1)$, then $E(Z) = 0$, and $E((Z - \bar{Z})(Z - \bar{Z})^T) = I_n$.
Denote the p.d.f of $Z$ as $f_Z$, since $z_1, ...z_n$ are i.i.d normal,
$f_Z = \prod_{i=1}^n f_{z_i} = \prod_{i=1}^n (\frac{1}{\sqrt{2\pi}} exp(\frac{z_i^2}{2})) = (2\pi)^{-\frac{n}{2}} exp(-\frac{\|Z\|^2}{2})$
$\Longrightarrow$there is a probability law of $Z := v_Z(A) = \int_A f_Z(z)\lambda(dz)$, where $A \in \mathcal{B}(\mathbb{R}^n)$.

Since $\Sigma$ is positive definite, exist a unique square root of $\Sigma$ that is also positive definite, denote as $\Sigma^{\frac{1}{2}}$. Define $g(Z) = \Sigma^{\frac{1}{2}}(Z) + \mu$, and let $X \in \mathbb{R}^n$ be a random vector, where $X = g(Z)$. Then the probability law of $X := v_X = g_*v_Z$, we want to show

$g_* v_Z(\mathcal{T}) = \int_{\mathcal{T}} f(x)\lambda(dx)$, $\mathcal{T} \in \mathcal{B}(R^n)$.

Before proceeding, recall the linear transformation property of Lebesgue measure $\Longrightarrow$ ${g^{-1}}_* \lambda = det(\Sigma^{\frac{1}{2}})\lambda = (det\Sigma)^{\frac{1}{2}}\lambda$.

Suppose $g(\mathcal{T}') = \mathcal{T}$,

$$\begin{aligned}
g_* v_Z(\mathcal{T}) = v_Z(\mathcal{T}') &= \int_{\mathcal{T}'} f_Z(z)\lambda(dz) = \frac{1}{(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}'} f_Z(z){g^{-1}}_*\lambda(dz) \\
&= \frac{1}{(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}} f_Z \circ g^{-1}(x)\lambda(dx) \\
&= \frac{1}{(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}} f_Z(\Sigma^{-\frac{1}{2}}(x-\mu))\lambda(dx) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}} exp(-\frac{1}{2}(\Sigma^{-\frac{1}{2}}(x-\mu))^T(\Sigma^{-\frac{1}{2}}(x-\mu)))\lambda(dx) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}} exp(-\frac{1}{2}((x-\mu)^T(\Sigma^{-\frac{1}{2}})^T\Sigma^{-\frac{1}{2}}(x-\mu))\lambda(dx) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}(det\Sigma)^{\frac{1}{2}}}\int_{\mathcal{T}} exp(-\frac{1}{2}((x-\mu)^T\Sigma^{-1}(x-\mu))\lambda(dx) \\
&= \int_{\mathcal{T}} f(x)\lambda(dx)
\end{aligned}$$

Hence, $\int_{\mathcal{T}} f(x)\lambda(dx)$ is the probability law of $X$, where $X = \Sigma^{\frac{1}{2}}Z + \mu$.
Then the mean and covariance of $X$ is given by

$$\begin{aligned}
E(X) &= E(\Sigma^{\frac{1}{2}}(Z) + \mu) = \Sigma^{\frac{1}{2}}E(Z) + \mu = \mu \\
Cov(X) &= E((X-\bar{X})(X-\bar{X})^T) = E((\Sigma^{\frac{1}{2}}Z)(\Sigma^{\frac{1}{2}}Z)^T) \\
&= \Sigma^{\frac{1}{2}}E(ZZ^T)\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}I_n\Sigma^{\frac{1}{2}} \\
&= \Sigma
\end{aligned}$$

■

**Exercise.** 7.6.8
Let $X \sim N_n(\mu, \Sigma)$, and let $m$ be a positive integer $\leq n$. Show that for any $a \in \mathbb{R}^m$ and any $m \times n$ matrix $A$ of full rank, $AX + a \sim N_m(a + A\mu, A\Sigma A^T)$.

**Answer.** $E(AX + a) = AE(X) + a = A\mu + a$.
$E((AX + a - \bar{AX + a})(AX + a - \bar{AX + a})^T) = E((AX - A\mu)(AX - A\mu)^T) = AE((X - \mu)(X-\mu)^T)A^T = A\Sigma A^T$.

■

## 7.7 Convolutions

**Exercise.** 7.7.2 If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, prove that $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

**Answer.** (Hao)
Since $X_1, X_2$ are two independent variables and $X_2$ have probability density functions $f_{X_2}(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \frac{-(x-\mu_2)^2}{2\sigma_2^2}$, $f_{X_1}(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \frac{-(x-\mu_1)^2}{2\sigma_1^2}$, respectively. Then, proposition 7.7.1 gives that the density of $Z = X_1 + X_2$ as follows

$$
\begin{aligned}
f_Z(z) = \mathbb{E} f_{X_2}(z - X_1) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \frac{-((z-x)-\mu_2)^2}{2\sigma_2^2} f_{X_1}(x) dx \\
&= \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(\sigma_1\sigma_2)^2}} \exp \frac{-1}{2\sigma_1^2\sigma_2^2} \left[ x\sqrt{\sigma_1^2+\sigma_2^2} - \frac{(\sigma_1^2(z-\mu_2)+\sigma_2^2\mu_1)}{\sqrt{\sigma_1^2+\sigma_2^2}} \right]^2 dx}_{\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}} \\
&\quad \cdot \frac{1}{\sqrt{2\pi}} \exp \frac{-1}{2(\sigma_1^2+\sigma_2^2)} \left[ z - (\mu_1+\mu_2) \right]^2 \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp \frac{-1}{2(\sigma_1^2+\sigma_2^2)} \left[ z - (\mu_1+\mu_2) \right]^2
\end{aligned}
$$

Hence, $Z = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

■

**Exercise.** 7.7.3 As a consequence of the above exercise, prove that any linear combination of independent normal random variables is normal with the appropriate mean and variance.

**Answer.** (Hao)
By exercise 5.5.2, If $X \sim N(\mu, \sigma^2)$, then for any $a, b \in \mathbb{R}$, $aX + b \sim N(a\mu + b, a^2\sigma^2)$. Hence, suppose that $X_i \sim N(\mu_i, \sigma_i^2), i = 1 \dots n$. The independence still hold after scalar multiplication of $a_i$, so the linear combination $a_1X_1 + a_2X_2 + \cdots + nX_n \sim N(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2)$

■

# 8 Convergence of Random Variables

## 8.1 Four notions of convergence

## 8.2 Interrelations between the four notions

**Exercise.** 8.2.2
Give a counterexample to show that convergence in probability does not imply almost sure convergence.

**Answer.** Define a sequence of functions on $\{0,1\}$ by $f_n = \frac{1}{n}\mathbb{1}_{\{0\}} + (1-\frac{1}{n})\mathbb{1}_{\{1\}}$, then there exists a sequence of independent random variables $\{X_n\}_{\mathbb{N}}$ on $(\{0,1\}^{\mathbb{N}}, \mathcal{F}, \mathbb{P})$ s.t. $f_n$ is the p.m.f. of $X_n$ (See EXERCISE 7.1.8.). Then for any $\epsilon > 0$, $\mathbb{P}(|X_n - 1| > \epsilon) = 1/n$ hence $\lim_{n\to\infty} \mathbb{P}(|X_n - 1| > \epsilon) = 0$ (converge to 1 in probability). On the other hand, $\mathbb{P}(\{|X_n - 1| > \epsilon \text{ i.o.}\}) = 1$ since $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - 1| > \epsilon) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$ and by the second Borel-Cantelli lemma. Thus $X_n \not\to 1$ a.s. ■

**Exercise.** 8.2.5
Show that the above proposition(proposition 8.2.4) is not valid if we demanded that $F_{X_n}(t) \to F_X(t)$ for all $t$, instead of just the continuity points of $F_X$.

**Answer.** (In class) We can consider $X_n = X + 1/n$. Then, $F_{X_n}(x) = F_X(x - 1/n)$, which does not necessarily approach $F_X(x)$(cdf only guarantees continuous on the right). For example, let

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.5, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases} \tag{1}$$

Then $F_{X_n}(1) = 0.5$ for all $n$, so $\lim_{n\to\infty} F_{X_n}(1) = 0 \neq 1 = F_X(1)$. ■

**Exercise.** 8.2.6
If $X_n \to c$ in distribution, where c is a constant, show that $X_n \to c$ in probability.

**Answer.** (In class) This is like the converse statement of proposition 8.2.4. The goal of the exercise is to show that $\mathbb{P}(|X_n - c| < d) \to 0$ as $n \to \infty$, for all $d > 0$. The condition $X_n \to c$ in distribution implies given any $\epsilon > 0$, $|F_{X_n}(t) F_c(t)| < \epsilon$, for $t$ is any continuous point of $F_c$. Moreover, since $F_c(t) = \begin{cases} 1, t \ge c \\ 0, t < c \end{cases}$, $t = c$ is the only discontinuous point of $c$. Hence, given any $d > 0$ and $\epsilon > 0$, we have $N_d \in \mathbb{N}$ such that

$$|F_{X_n}(c+d) - F_c(c+d)| = |F_{X_n}(c+d) - 1| < \epsilon/2$$

, and
$$|F_{X_n}(c-d) - F_c(c-d)| = |F_{X_n}(c-d) - 0| < \epsilon/2$$
, for all $n > N_d$. As a result,
$$\begin{aligned} |\mathbb{P}(|X_n - c| < d)| &= 1 - \mathbb{P}(|X_n - c| \geq d) \\ &= 1 - F_{X_n}(c+d) + F_{X_n}(c-d) \\ &< |1 - F_{X_n}(c+d)| + |0 - F_{X_n}(c-d)| \\ &< \epsilon \end{aligned} \tag{2}$$
, for all $n > N_d$. ■

**Exercise.** 8.2.10
Take any $p > 0$. Give counterexamples to show that almost sure convergence does not imply $L^p$ convergence, and $L^p$ convergence does not imply almost sure convergence.

**Answer.** For the counterexample showing convergence in $L^p$ does not imply almost sure convergence, we reuse the answer of exercise 8.2.2, and check that $|X_n| < 1$ so proposition 8.2.9 gives $X_n \to X$ in $L^p$, but in exercise 8.2.2 we had showed that $X_n$ does not converge to $X$ a,s.

For the counterexample showing $X_n \overset{L^p}{\not\to} X$, we consider $X_n$ on $\Omega = [0, 1]$ with uniform density.
$$X_n(\omega) = \begin{cases} (\frac{1}{2}n\omega)^{1/p}, 0 \leq \omega < \frac{1}{2n} \\ [\frac{1}{2n} - \omega]^{1/p}, \frac{1}{2n} \leq \omega < 1/n \\ 0, 1/n < \omega \leq 1 \end{cases} \tag{3}$$
Then, it is clear that $X_n$ converge to $X = 0$ point wisely, so $X_n \xrightarrow{a.s} X$. However,
$$\mathbb{E}[||X_n - 0||^p] = \int_0^1 ||X_n(\omega)||^p d\omega = 1 \tag{4}$$
for all $n$, so $X_n \overset{L^p}{\not\to} X$. ■

## 8.3 Uniform integrability

## 8.4 The weak law of large numbers

**Exercise.** 8.4.3 (An occupancy problem)
Let $n$ balls be dropped uniformly and independently at random into $n$ boxes. Let $N_n$ be the number of empppty boxes. Prove that $N_n/n \to e^{-1}$ in probability as $n \to \infty$ (Hint: Write $N_n$ as a sum of indicator variables.)

**Answer.** (In class) Define $X_i$ as indicator if the ith box is empty or not, i.e $X_i = 1$ if the box is empty, otherwise it is 0. In addition, $\mathbb{P}(X_i = 0) = 1 - (\frac{n-1}{n})^n, \mathbb{P}(X_i = 1) = (\frac{n-1}{n})^n$. With all the information, we can derive that $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = (\frac{n-1}{n})^n = (1 + \frac{-1}{n})^n \to e^{-1}$. That implies $\lim_{n\to\infty} \mathbb{P}(|P_n - E(X_1)| < \epsilon) < \lim_{n\to\infty} \mathbb{P}(|P_n - e^{-1}| + |e^{-1} - E(X_1)| < \epsilon) = \lim_{n\to\infty} \mathbb{P}(|P_n - e^{-1}| < \epsilon)$

Moving on, if we want to estimate the expectation of $N_n = \sum X_i$ by proposition 8.4.1, we have to calculate the variance of $X_i$ (which is $\sigma_{ii}$) and the correlation of $X_i, X_j$, $\sigma_{ij}$. For $\sigma_{ii}$, there are $n$ of them and $\mathbb{E}(X_i^2) = \mathbb{P}(X_i = 1) = (1 + \frac{-1}{n})^n$, so $\sigma_{ii} = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = (1 + \frac{-1}{n})^n(1 - (1 + \frac{-1}{n})^n) < 1$.

On the other hand, the number of $\sigma_{ij}, i \neq j$ we have to consider is $C(n, 2)$, which has the order $O(n^2)$, so we have to make sure $\sigma_{ij} \to 0$ as $n$ increase. Fortunately, it is really the case. We know that $\mathbb{E}(X_iX_j) = (\frac{n-2}{n})^n$, so $|\sigma_{ij}| = |(\frac{n-2}{n})^n - (\frac{n-1}{n})^{2n}| < |(\frac{n-2}{n})^n - e^{-2}| + |(e^{-1})^2 - (\frac{n-1}{n})^{2n}| = \epsilon(n)$, where $\lim_{n\to\infty} \epsilon(n) = 0$

Then by 8.4.1

$$\mathbb{P}\left(|\frac{1}{n}N_n - e^{-1}| < d\right) \leq \frac{1}{d^2n^2}\left[n + C(n, 2)\epsilon(n)\right]$$

, which has the order $O(n^{-1})$, so indeed $\frac{1}{n}N_n \to e^{-1}$ in probability.

■

**Exercise.** 8.4.4 (Coupon collector's problem). Suppose that there are $n$ types of coupons, and a collector wants to obtain at least one of each type. Each time a coupon is bought, it is one of the $n$ types with equal probability. Let $T_n$ be the number of trials needed to acquire all $n$ types. Prove that $T_n/(n \log n) \to 1$ in probability as $n \to \infty$. (Hint: Let $\tau_k$ be the number of trials needed to acquire $k$ distinct types of coupons. Prove that $\tau_k - \tau_{k-1}$ are independent geometric random variables with different means, and $T_n$ is the sum of these variables.)

**Answer.** Let $\tau_k$ be the number of trials needed to collect $k$ distinct types. Then $X_k = \tau_k - \tau_{k-1}$ is the number of trials needed to collect $k$ distinct types once we have $k - 1$ distinct types. Then $T_n = X_1 + \cdots + X_n$, and $X_k$ are independent since each trials are independent. Moreover, $X_k$ has p.d.f.

$$\mathbb{P}(X_k = x) = \mathbb{P}(\tau_k - \tau_{k-1} = x) = \Big(\frac{k-1}{n}\Big)^{x-1}\Big(1 - \frac{k-1}{n}\Big),\ x \in \mathbb{N}.$$

By Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}\Big(\Big|\frac{T_n}{n\log n}-\frac{\mathbb{E}(T_n)}{n\log n}\Big|\geq\epsilon\Big) &\leq \frac{1}{(n\log n)^2\epsilon^2}\mathbb{E}\Big(\sum_{k=1}^{n}X_k-\mathbb{E}X_k\Big)^2\\
&=\frac{1}{(n\log n)^2\epsilon^2}\sum_{k=1}^{n}\mathbb{E}(X_k-\mathbb{E}X_k)^2 \quad \text{(Independence)}\\
&=\frac{\sum_{k=1}^{n}\mathrm{Var}(X_k)}{(n\log n)^2\epsilon^2}.
\end{aligned}$$

One can calculate that $\mathrm{Var}(X_k)=(\frac{k-1}{n})/(1-\frac{k-1}{n})^2$. Hence

$$\begin{aligned}
\sum_{k=1}^{n}\mathrm{Var}(X_k)=\sum_{t=1}^{n-1}\frac{\frac{t}{n}}{(1-\frac{t}{n})^2} &\leq \int_0^{n-1}\frac{\frac{t}{n}}{(1-\frac{t}{n})^2}dt+(n^2-n) \quad \text{(Monotonically increase)}\\
&=n\int_{\frac{1}{n}}^{1}\frac{1-u}{u^2}du+n^2-n \qquad (u=1-\frac{t}{n})\\
&=2(n^2-n)-n\log(n).
\end{aligned}$$

So the tail bound reads

$$\frac{\sum_{k=1}^{n}\mathrm{Var}(X_k)}{(n\log n)^2\epsilon^2}\leq\frac{2(n^2-n)-n\log(n)}{(n\log n)^2\epsilon^2}\to 0 \text{ as } n\to\infty.$$

Thus $T_n/(n\log n)\to\mathbb{E}(T_n)/(n\log n)$ in probability. Also, one can verify

$$\begin{aligned}
&\mathbb{E}(T_n)=\sum_{k=1}^{n}\mathbb{E}(X_k)=\sum_{k=1}^{n}\frac{1}{1-\frac{k-1}{n}}\leq\int_0^{n-1}\frac{1}{1-\frac{t}{n}}dt+n,\\
&\text{and } \mathbb{E}(T_n)\geq\int_0^{n-1}\frac{1}{1-\frac{t}{n}}dt
\end{aligned}$$

by monotonicity of the summation. Hence we have $n\log(n)\leq\mathbb{E}(T_n)\leq n\log(n)+n$. Divide all of them by $n\log n$ and let $n\to\infty$, then we have $\lim_{n\to\infty}\mathbb{E}(T_n)/(n\log n)=1$. That is, $T_n/(n\log n)\to 1$ in probability. ■

**Exercise.** 8.4.5 (Erdős–Rényi random graphs)
Define an undirected random graph on $n$ vertices by putting an edge between any two vertices with probability $p$ and excluding the edge with probability $1-p$, all edges independent. This is known as the Erdős–Rényi $G(n,p)$ random graph. First, formulate the model in the measure theoretic framework using independent Bernoulli random variables. Next, show that if $T_{n,p}$ is the number of triangles in this random graph, then $T_{n,p}/n^3\to p^3/6$ in probability as $n\to\infty$, if $p$ remains fixed.

**Answer.** (Hao) First of all, let's go through some basic setup. We understand that each triangle of the graph can be represented by three edges, so we can consider the $X_i$ is the 1 iff all three edges are put. $X_i = 1 * \mathbb{1}_{X_i=1} + 0 * \mathbb{1}_{X_i=0}$, and $\mathbb{P}(X_i = 1) = p^3$. The number of different triangles is $C(n, 3)$, and $T_{n,p} = \sum X_i, i = 1, 2, \ldots C(n, 3)$.

There is an intuitive way and a rigorous way. The intuitive way goes as follows: Since each $X_i$ as an expectation of $p^3$ and there are in total around $n^3/6$ of triangles. Hence, on average $T_{n,p}/n^3$ is around $p^3/6$.

Of course, we still need to make sure that it converge in probability. For this matter, we should estimate the variance of $X_i$ and the correlation between $X_i, X_j$, i.e we calculate $\mathbb{E}[(X_i - p^3)(X_j - p^3)]$. There are two cases need to be determined, that is whether the two triangles share the same side or not. Notice that two triangles can't share two same sides.

The variance of $X_i$(i.e $\sigma_{ii}$) can be easily verified to be $p^3 - p^6$, and there are $C(n, 3)$ of them. Next we consider correlation. The case where two triangles do not share the same side is easy, they are independent by our model setup, so the correlation is 0. For the second case, suppose the ith and jth triangle share one side, which in total $C(n, 4)$ (see Figure **??**)of them and $\mathbb{P}(X_iX_j = 1) = p^5$, so

$$\sigma_{ij} = \mathbb{E}[(X_i - p^3)(X_j - p^3)] = \mathbb{E}(X_iX_j) - 2p^6 + p^6 = p^5 - p^6$$

Hence, by the methods employed in 8.4.1

$$\mathbb{P}\left(|\frac{1}{C(n,3)}T_{n,p} - p^3| < \epsilon\right) \leq \frac{1}{\epsilon^2 C(n,3)^2}\left[(p^5 - p^6)C(n,4) + (p^3 - p^6)C(n,3)\right]$$

The right hand side is of order $O(n^{-2})$, so indeed $\frac{1}{C(n,3)}T_{n,p} \xrightarrow{p} p^3$. Also, the fact that $C(n, 3)/(n^3/6) \to 1$ give the result.

■

## 8.5 The strong law of large numbers

**Exercise.** 8.5.2. Using Exercise 7.3.2, show that if $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables such that $\mathbb{E}|X_1| = \infty$, then

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \text{ has a finite limite as } n \to \infty\Big) = 0.$$

**Answer.** Recall EXERCISE 7.3.2.: if $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of random variables s.t. $\mathbb{E}|X_1| = \infty$, then $\mathbb{P}(|X_n| > n \text{ i.o.}) = 1$. Now since

$$\bigcap_{N\in\mathbb{N}} \bigcup_{n>N} \{|X_n| > n\} \subseteq \bigcap_{N\in\mathbb{N}} \bigcup_{n>N} \{|X_n| > \epsilon\}$$

for some $\epsilon \in (0,1)$, clearly

$$\mathbb{P}\Big(\bigcup_{\epsilon>0} \bigcap_{N\in\mathbb{N}} \bigcup_{n>N} \{|X_n| > \epsilon\}\Big) = 1,$$

i.e. $\lim_{n\to\infty} X_n \neq 0$ a.e., therefore $\lim_{n\to\infty} \sum_{i=1}^n X_i/n$ doesn't converge to finite a.e. ■

**Exercise.** 8.5.3. If $X_1, X_2, \cdots$ are i.i.d. random variables with $\mathbb{E}(X_1) = \infty$, show that $n^{-1}\sum_{i=1}^n X_i \to \infty$ a.s.

**Answer.** In the sense of Lebesgue integral, $\mathbb{E}(X) = \infty$ if and only if $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) < \infty$. Since $X^+$ is measurable and nonnegative, there exists a sequence of simple functions $\{Y_m\}_{m\in\mathbb{N}}$ on $\Omega$ such that $Y_m \uparrow X$ as $m \to \infty$ with each $Y_m \leq m$ on $\Omega$. (For example $Y_m = X^+\mathbb{1}_{\{X^+<m\}} + m\mathbb{1}_{\{X^+\geq m\}}$.) Then $\mathbb{E}(Y_m) < \infty$ for all $m \in \mathbb{N}$. Furthermore, by monotone convergence theorem,

$$\lim_{m\to\infty} \int Y_m d\mathbb{P} = \int X^+ d\mathbb{P} = \mathbb{E}(X^+) = \infty,$$

i.e. $\mathbb{E}(Y_m) \to \infty$. Now we claim that $\lim_{n\to\infty} \sum_{i=1}^n X_i^+/n = \infty$ a.s.
(Proof of claim) Let $L \in \mathbb{N}$, then there exists $m \in \mathbb{N}$ such that $\mathbb{E}(Y_m) > L$. Also,

$$\begin{aligned}
\Big\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n Y_m^{(i)} = \mathbb{E}(Y_m)\Big\} &\subseteq \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \Big\{\frac{1}{n}\sum_{i=1}^n Y_m^{(i)} > L\Big\} \quad (\mathbb{E}(Y_m) > L) \\
&\subseteq \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \Big\{\frac{1}{n}\sum_{i=1}^n X_i^+ > L\Big\} \quad (X^+ \geq Y_m \text{ on } \Omega) \\
&\subseteq \Big\{\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^+ \geq L\Big\}.
\end{aligned}$$

But by SLLN, LHS $= \Omega$ almost surely for any $m \in \mathbb{N}$, that means

$$\Big\{\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^+ \geq L\Big\} = \Omega \; a.s. \; \forall L \in \mathbb{N}.$$

Hence $\liminf_{n\to\infty}\sum_{i=1}^n X_i^+/n = \infty$ a.s., which completes the proof.$\square$
Since $\mathbb{E}(X^-) < \infty$, $\lim_{n\to\infty}\sum_{i=1}^n X_i^-/n = \mathbb{E}(X^-)$ a.s. by SLLN. Hence

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i = \lim_{n\to\infty}\Big(\frac{1}{n}\sum_{i=1}^n X_i^+ - \frac{1}{n}\sum_{i=1}^n X_i^-\Big) = \infty - \mathbb{E}(X^-) = \infty \ a.s.$$

$\blacksquare$

**Exercise.** 8.5.7. (SLLN under bounded fourth moment). Let $\{X_n\}_{n=1}^\infty$ be a sequence of independent random variables with mean zero and uniformly bounded fourth moment. Prove that $n^{-1}\sum_{i=1}^n X_i \to 0$ a.s. (Hint: Use a fourth moment version of Chebychev's inequality.)

**Answer.** Let $S_n = \sum_{i=1}^n X_i$. To prove $\mathbb{P}(\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\{|S_n/n| < \epsilon\}) = 1$ for any $\epsilon > 0$, it's suffice to prove that $\mathbb{P}(\big\{|S_n/n| > \epsilon \text{ i.o.}\big\}) = 0$ for any $\epsilon > 0$.
Let $\epsilon > 0$, $n \in \mathbb{N}$. Since $\{X_i\}_{i\in\mathbb{N}}$ are independent and $\mathbb{E}(X_i) = 0$,

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| > \epsilon\Big) \leq \frac{\mathbb{E}(S_n^4)}{n^4\epsilon^4} = \frac{1}{n^4\epsilon^4}\Big(\sum_{i=1}^n \mathbb{E}(X_i^4) + \sum_{i\neq j}\mathbb{E}(X_i^2)\mathbb{E}(X_j^2)\Big).$$

One can check the above equation:

$$\begin{aligned}\mathbb{E}(S_n^4) =& \sum_{i=1}^n \mathbb{E}(X_i^4) + \sum_{\substack{i,j,k \\ \text{all distinct}}}^{3\binom{n}{3}\frac{4!}{2!}} \mathbb{E}(X_i^2X_jX_k) + \sum_{i\neq j}^{2\binom{n}{2}\frac{4!}{3!}} \mathbb{E}(X_i^3X_j) \\ &+ \sum_{i\neq j}^{\binom{n}{2}\frac{4!}{2!2!}} \mathbb{E}(X_i^2X_j^2) + \sum_{\substack{i,j,k,l \\ \text{all distinct}}}^{\binom{n}{4}4!} \mathbb{E}(X_iX_jX_kX_l).\end{aligned}$$

Let $M$ be the uniform bound of $\mathbb{E}(X_i^4)$, then by Jenson's inequality we have $\mathbb{E}(X_i^2) \leq \sqrt{M}$. So the above equation becomes

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| > \epsilon\Big) \leq \frac{M}{n^4\epsilon^4}\Big(n + \binom{n}{2}\frac{4!}{2!2!}\Big) = \frac{3M(n^2-5n)}{n^4\epsilon^4} \leq \frac{3M}{n^2\epsilon^4}.$$

Hence $\sum_{n\in\mathbb{N}}\mathbb{P}(|S_n/n| > \epsilon) \leq C(\epsilon)\sum_{n\in\mathbb{N}} 1/n^2 < \infty$. Then by Borel-Cantelli lemma we have $\mathbb{P}(|S_n/n| > \epsilon \text{ i.o.}) = 0$, which implies $\mathbb{P}(\bigcup_{N\in\mathbb{N}}\bigcap_{n\geq N}\{|S_n/n| < \epsilon\}) = 1$. Since $\epsilon > 0$ is arbitrary, $\mathbb{P}(\lim_{n\to\infty} S_n/n = 0) = 1$. $\blacksquare$

**Exercise.** 8.5.8. (Random matrices). Let $\{X_{ij}\}_{1\leq i\leq j\leq\infty}$ be a collection of i.i.d. random variables with mean zero and all moments finite. Let $X_{ji} = X_{ij}$ if $j > i$. Let $W_n$ be the $n \times n$ symmetric random matrix whose $(i,j)$th entry is $n^{-1/2}X_{ij}$. A matrix like $W_n$ is called a Wigner matrix. Let $\lambda_{n,1} \geq \cdots \geq \lambda_{n,n}$ be the eigenvalues of $W_n$, repeated by multiplicities. For any integer $k \geq 1$, show that

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_{n,i}^{k} - \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\lambda_{n,i}^{k}\Big) \to 0 \ a.s. \text{ as } n \to \infty.$$

**Answer.** Rewrite $\sum_{i=1}^{n}\lambda_{n,i}^{k} = tr(W_n^k) = \sum_{i=1}^{n} W_{ii}^k$, the diagonal sum of the $k$th power of $W_n$. Fix $k \in \mathbb{N}$, $\epsilon > 0$. For each $n \in \mathbb{N}$, we're going to show the tail probability at least

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k - \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k\Big)\Big| \geq \epsilon\Big) \sim \frac{\text{const.}}{\epsilon^2 n^2}.$$

Then by Borel-Cantelli lemma we'll have

$$\mathbb{P}\Big(\bigcap_{N\in\mathbb{N}}\bigcup_{n>N}\Big\{\Big|\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k - \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k\Big)\Big| \geq \epsilon\Big\}\Big) = 0,$$

i.e.

$$\mathbb{P}\Big(\bigcup_{N\in\mathbb{N}}\bigcap_{n>N}\Big\{\Big|\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k - \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} W_{ii}^k\Big)\Big| < \epsilon\Big\}\Big) = 1.$$

Since $\epsilon > 0$ is arbitrary, we thus proved that $\sum_{i=1}^{n} W_{ii}^k/n - \mathbb{E}\big(\sum_{i=1}^{n} W_{ii}^k/n\big) \to 0$ a.s. To estimate the tail bound, we evaluate $\sum_{i,j}\sigma_{ij}$ where $\sigma_{ij} = \mathbb{E}(W_{ii}^k W_{jj}^k)$ in the following part.

**(Covariance estimation)** Let $k$ be an integer, observe that

$$W_{ii}^k = n^{-k/2}\sum_{(j_1,\cdots,j_{k-1})\in[n]^{k-1}} X_{i,j_1}X_{j_1,j_2}\cdots X_{j_{k-1}i}$$

acts like a path $\mathbf{p}_i$ of length $k$ starts from $i$ and go back to $i$. In the following, I denotes $(V_{\mathbf{p}_i}, E_{\mathbf{p}_i})$ to be the graph covered by $\mathbf{p}_i$ excluding the end vertex $i$.

**(A) For $k$ is odd.**

(Expectation part)

**Observation 1.** *Let $\mathbf{p}_i$ be a path of length $k$ that starts and ends at some $i \in [n]$. If there is an edge passed by $\mathbf{p}_i$ exactly once (maybe odd number times?), then there exists loop or cycle in $(\{i\} \cup V_{\mathbf{p}_i}, E_{\mathbf{p}_i})$. Hence $|\{i\} \cup V_{\mathbf{p}_i}| \leq |E_{\mathbf{p}_i}|$.*

Fix $i \in [n]$. If $k$ is odd and $\mathbb{E}(X_{\mathbf{p}_i}) \neq 0$, then $|E_{\mathbf{p}_i}| \leq (k-1)/2$. Suppose $\mathbf{p}_i$ satisfies $|E_{\mathbf{p}_i}| = (k-1)/2$ and $\mathbb{E}(X_{\mathbf{p}_i}) \neq 0$, then $|V_{\mathbf{p}_i}| \leq (k-3)/2$ (to be confirmed). Let $D_k$ be the number of patterns that such path can have, then we have

$$\begin{aligned}\sum_{\mathbf{p}_i} \mathbb{E}(X_{\mathbf{p}_i}) &= M_2^{(k-3)/2} M_3 D_k (n-1)\cdots(n-(k-3)/2) + O(n^{(k-3)/2-1}) \\ &= M_2^{(k-3)/2} M_3 D_k n^{(k-3)/2} + O(n^{(k-3)/2-1}).\end{aligned}$$

Hence

$$\mathbb{E}(W_{ii}^k) = \frac{1}{n^{(k/2)}} \sum_{\mathbf{p}_i} \mathbb{E}(X_{\mathbf{p}_i}) = M_2^{(k-3)/2} M_3 D_k \frac{1}{n\sqrt{n}} + O\Big(\frac{1}{n^2\sqrt{n}}\Big).$$

(Covariance part)
Now we need to evaluate $\mathbb{E}(W_{ii}^k W_{jj}^k) = n^{-k} \sum_{\mathbf{p}_i,\mathbf{p}_j} \mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j})$. For $\mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) \neq 0$, $|E_{\mathbf{p}_i} \cup E_{\mathbf{p}_j}|$ is at most $k$. Suppose $\mathbf{p}_i, \mathbf{p}_j$ satisfy $|E_{\mathbf{p}_i} \cup E_{\mathbf{p}_j}| = k$ and $\mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) \neq 0$.

If $k$ is odd and $i \neq j$, then $(\{i\} \cup V_{\mathbf{p}_i}, E_{\mathbf{p}_i}), (\{j\} \cup V_{\mathbf{p}_j}, E_{\mathbf{p}_j})$ is connected. (If not, then $E_i \cap E_j = \emptyset$ and $|E_k| \leq (k-1)/2$ for $k = i, j$. But $|E_i \cup E_j| = k-1$. $\rightarrow\leftarrow$) Also, $|E_k| > (k-1)/2$ for some $k \in \{i,j\}$, thus there is $e \in E_k$ which is passed by $\mathbf{p}_k$ only once. By *Observation 1*, $|\{i,j\} \cup V_i \cup V_j| \leq |E_i \cup E_j|$, thus $|(V_i \cup V_j) \setminus \{i,j\}| \leq k-2$. Hence

$$\sum_{\mathbf{p}_i,\mathbf{p}_j} \mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) = C(n-2)\cdots(n-(k-1)) + O(n^{k-3}) = M_2^k n^{k-2} + O(n^{k-3}).$$

Thus we have

$$\mathbb{E}(W_{ii}^k W_{jj}^k) = C\frac{1}{n^2} + O\Big(\frac{1}{n^3}\Big) \qquad \text{for } i \neq j$$

If $k$ is odd and $i = j$, then there's one path with $|E_{\mathbf{p}_i}| > (k-1)/2$, and by *Observation 1* again, $|\{i\} \cup V_{\mathbf{p}_i} \cup V_{\mathbf{p}'_i}| \leq k$, thus $|V_{\mathbf{p}_i} \cup V_{\mathbf{p}'_i}| \leq k-1$:

$$\sum_{\mathbf{p}_i,\mathbf{p}'_i} \mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}'_i}) = C'(n-1)\cdots(n-(k-1)) + O(n^{k-2}) = C' n^{k-1} + O(n^{k-2}).$$

Hence

$$\mathbb{E}[(W_{ii}^k)^2] = C'\frac{1}{n} + O\Big(\frac{1}{n^2}\Big) \qquad \text{for } i = j.$$

($k$ is odd summary) If $k$ is odd, then

$$\begin{aligned}\sum_{i,j} \sigma_{ij}^2 &= \sum_{i=1}^n \mathbb{E}[(W_{ii}^k)^2] + \sum_{i \neq j} \mathbb{E}(W_{ii}^k W_{jj}^k) - n^2\Big(\mathbb{E}(W_{ii}^k)\Big)^2 \\ &= C' + O\Big(\frac{1}{n}\Big) + (n^2-n)\Big(C\frac{1}{n^2} + O\Big(\frac{1}{n^3}\Big)\Big) - \Big(A_k^2\frac{1}{n} + O\Big(\frac{1}{n^2}\Big)\Big) \\ &= C' + C + O\Big(\frac{1}{n}\Big).\end{aligned}$$

**(B) For $k$ is even.**

(Expectation part)

If $k$ is even, let $i \in [n]$, and $\mathbf{p}_i$ be the path with both ends are $i$. Then $\mathbb{E}(X_{\mathbf{p}_i}) \neq 0$ only if $|E_{\mathbf{p}_i}| \leq k/2$. Suppose $\mathbf{p}_i$ satisfies $|E_{\mathbf{p}_i}| = k/2$ and $\mathbb{E}(X_{\mathbf{p}_i}) \neq 0$, then $|V_{\mathbf{p}_i}| \leq k/2$. Hence

$$\begin{aligned}\sum_{\mathbf{p}_i} \mathbb{E}(X_{\mathbf{p}_i}) &= C_{k,1} M_2^{k/2} \times (n-1)\cdots(n-k/2) && (\frac{k}{2} \text{ edges and } \frac{k}{2} \text{ vertices}) \\ &+ C_{k,2} M_2^{k/2} \times (n-1)\cdots(n-(k/2-1)) && (\frac{k}{2} \text{ edges and } \frac{k}{2}-1 \text{ vertices}) \\ &+ C_{k,3} M_2^{k/2-2} M_4 \times (n-1)\cdots(n-(k/2-1)) && (\frac{k}{2}-1 \text{ edges and } \frac{k}{2}-1 \text{ vertices}) \\ &+ O(n^{k/2-2})\end{aligned}$$

Where $C_{k,1}, C_{k,2}, C_{k,3}$ are defined to be the numbers of patterns of $k$-length paths $\mathbf{p}$ that covers the graph $(V, E) = (\frac{k}{2}+1, \frac{k}{2}), (\frac{k}{2}+1, \frac{k}{2}-1), (\frac{k}{2}, \frac{k}{2}-1)$, with no edges are passed only once, respectively. So they can be wrote as

$$\sum_{\mathbf{p}_i} \mathbb{E}(X_{\mathbf{p}_i}) = C_{k,1} M_2^{k/2} n^{k/2} - C_{k,1} M_2^{k/2} \Big( S_{k/2} - \frac{C_{k,2}}{C_{k,1}} - \frac{C_{k,3} M_4}{C_{k,1} M_2^2} \Big) n^{k/2-1} + O(n^{k/2-2}),$$

where $S_{k/2} = 1 + 2 + \cdots + k/2$. And $\mathbb{E}(W_{ii}^k) = n^{-k/2} \sum_{\mathbf{p}_i} \mathbb{E}(X_{\mathbf{p}_i})$.

(Covariance part)

($k$ is even) $\mathbb{E}(W_{ii}^k W_{jj}^k) = n^{-k} \sum_{\mathbf{p}_i, \mathbf{p}_j} \mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j})$. $\mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) \neq 0$ only if $|E_{\mathbf{p}_i} \cup E_{\mathbf{p}_j}| \leq k$.

**Observation 2.** *When $k$ is even, $|V_{\mathbf{p}_i} \cup V_{\mathbf{p}_j}| \geq k-1 \Rightarrow |E_{\mathbf{p}_k}| \leq k/2$ for $k = i, j$.*

*Check.* Suppose $|E_{\mathbf{p}_i}| > k/2$, then there exists $e \in E_i$ passed by $\mathbf{p}_i$ only once since $\mathbf{p}_i$ is of length $k$. Then $\mathbf{p}_j$ must pass $e$ otherwise $\mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) = 0$. Hence $(\{i,j\} \cup V_i \cup V_j, E_i \cup E_j)$ is a graph of $k$ edges with cycle or loop exists by *Observation 1*. Hence $|\{i,j\} \cup V_i \cup V_j| \leq |E_i \cup E_j| = k$, i.e. $|V_i \cup V_j| \leq k-2$. $\square$

($k$ edges and $k$ extra vertices) In this case, there are $k+2$ vertices(with $i, j$) in this graph. This implies $(\{i\} \cup V_i, E_i)$ and $(\{j\} \cup V_j, E_j)$ cannot be connected, otherwise there are at most $k+1$ vertices in total. Hence $|E_i| = |E_j| = k/2$ and $|V_i| = |V_j| = k/2$ since $|V| \leq |E|$. To have nonzero expectation, there are $C_{k,1}$ such patterns for each $i, j$, thus there are total $(C_{k,1})^2$ patterns, and hence $(C_{k,1})^2 (n-2)\cdots(n-(k+1))$ choices.

($k$ edges and $k-1$ extra vertices) By the observation, if $|V_i \cup V_j| = k-1$, since $|V_i| \leq |E_i|$, either $|E_i| = |E_j| = k/2$ with $|E_i \cap E_j| = 0$ and $|V_i| = |V_j| = k/2$ with $|V_i \cap V_j| = 1$, or $|E_i| = |E_j| = k/2$ with $|E_i \cap E_j| = 0$ and $\{|V_i|, |V_j|\} = \{k/2, k/2-1\}$. There are $(C_{k,1})^2$ patterns in the first case, and $2C_{k,1}C_{k,2}$ patterns in the second case. Hence in total $((C_{k,1})^2 + 2C_{k,1}C_{k,2})(n-2)\cdots(n-k)$ choices.

($k-1$ edges and $k-1$ extra vertices) By the observation, since $|V_i \cup V_j| = k-1$, either $\{|E_i|, |E_j|\} = \{k/2, k/2-1\}$ with $|E_i \cap E_j| = 0$, or $|E_i| = |E_j| = k/2$ with $|E_i \cap E_j| = 1$. But the latter one is impossible since if $E_i \cap E_j \neq \emptyset$, $(\{i\} \cup V_i, E_i)$ and $(\{j\} \cup V_j, E_j)$ are connected and has at most $k-1+1 = k$ total vertices, i.e. $k-2 \neq |V_i \cup V_j|$ extra vertices. As for the former one, they must be $\{|V_i|, |V_j|\} = \{k/2, k/2-1\}$ and $V_i \cap V_j = \emptyset$. Use the previous notation, there are $2C_{k,1}C_{k,3}$ such patterns, thus we have in total $2C_{k,1}C_{k,3}(n-2)\cdots(n-k)$ choices.

In summary,

$$\begin{aligned}\sum_{\mathbf{p}_i,\mathbf{p}_j} \mathbb{E}(X_{\mathbf{p}_i} X_{\mathbf{p}_j}) = {} & (C_{k,1})^2 M_2^k (n-2)\cdots(n-(k+1)) && \text{(case 1)} \\ & + ((C_{k,1})^2 + 2C_{k,1}C_{k,2}) M_2^k (n-2)\cdots(n-k) && \text{(case 2)} \\ & + 2C_{k,1}C_{k,3} M_2^{k-2} M_4 (n-2)\cdots(n-k) && \text{(case 3)} \\ & + O(n^{k-2}) \end{aligned}$$

i.e.

$$\begin{aligned}\mathbb{E}(W_{ii}^k W_{jj}^k) = {} & (C_{k,1})^2 M_2^k + (C_{k,1})^2 M_2^k \Big(1 + 2\frac{C_{k,2}}{C_{k,1}} + 2\frac{C_{k,3}M_4}{C_{k,1}M_2^2} - (S_{k+1}-1)\Big)\frac{1}{n} \\ & + O\Big(\frac{1}{n^2}\Big)\end{aligned}$$

Then the covariance $\sigma_{ij}^2 = \mathbb{E}(W_{ii}^k - \mathbb{E}(W_{ii}^k))(W_{jj}^k - \mathbb{E}(W_{jj}^k)) = \mathbb{E}(W_{ii}^k W_{jj}^k) - (\mathbb{E}(W)_{ii}^k)^2$ is

$$\begin{aligned}\sigma_{ij}^2 &= \mathbb{E}(W_{ii}^k W_{jj}^k) - (\mathbb{E}(W)_{ii}^k)^2 \\ &= (C_{k,1})^2 M_2^k (2S_{k/2} - S_{k+1} + 2)\frac{1}{n} + O\Big(\frac{1}{n^2}\Big) \\ &= O\Big(\frac{1}{n^2}\Big),\end{aligned}$$

since in the first term of RHS, $2S_{k/2} - S_{k+1} + 2 < 0$ for all even $k$. So the total sum $\sum_{i,j} \sigma_{ij}^2 = n^2 \times O(1/n^2) = O(1)$. □

By WLLN, the tail bound is $\big(\sum_{i,j}\sigma_{ij}^2\big)/n^2\epsilon^2$, and in both **(A)** and **(B)** we've shown that $\big(\sum_{i,j}\sigma_{ij}^2\big)/n^2\epsilon^2 \leq O(1/n^2\epsilon^2)$. Follow by Borel-Cantelli lemma we thus have proved that $\sum_{i=1}^n W_{ii}^k/n - \mathbb{E}\big(\sum_{i=1}^n W_{ii}^k/n\big) \to 0$ a.s. (As the statement in the first paragraph). ■

**Exercise.** 8.5.9. If all the random graphs in EXERCISE 8.4.5 are defined on the same probability space, show that the convergence is almost sure.

**Answer.** Exercise 8.4.5. has shown that for any $\epsilon > 0$,

$$\mathbb{P}\Big(\Big|\frac{T_{n,p}}{\binom{n}{3}} - p^3\Big| \geq \epsilon\Big) \leq \frac{6}{n^6\epsilon^2}\Big(\binom{n}{4}\big(p^5 - (p^3)^2\big) + \binom{n}{3}\big(p^3 - (p^3)^2\big)\Big)$$
$$= O(n^{-2}).$$

Let $\{|T_{n,p}/\binom{n}{3} - p^3| \geq \epsilon\} = A_{n,\epsilon}$, then $\sum_{n\in\mathbb{N}} \mathbb{P}(A_{n,\epsilon}) < \infty$. By Borel-Cantelli lemma we then have $\mathbb{P}(\cap_{N\in\mathbb{N}} \cup_{n>N} A_{n,\epsilon}) = 0$, i.e.

$$\mathbb{P}\Big(\bigcup_{N\in\mathbb{N}}\bigcap_{n>N}\Big\{\Big|\frac{T_{n,p}}{\binom{n}{3}} - p^3\Big| < \epsilon\Big\}\Big) = 1,$$

hence

$$\mathbb{P}\Big(\bigcap_{\epsilon>0}\bigcup_{N\in\mathbb{N}}\bigcap_{n>N}\Big\{\Big|\frac{T_{n,p}}{\binom{n}{3}} - p^3\Big| < \epsilon\Big\}\Big) = 1.$$

Which means $T_{n,p}/\binom{n}{3} \to p^3$ a.s. ■

## 8.6 Tightness and Helly's selection theorem

**Exercise.** 8.6.2. If $X_n \to X$ in distribution, show that $\{X_n\}_{n\geq 1}$ is a tight family.

**Answer.** Suppose $X_n \to X$ in distribution, and $X$ has c.d.f. $F$. Given $\epsilon > 0$, there exists $K > 0$ s.t. $1 - F(K) + F(-K) < \epsilon/3$. There is also a $N \in \mathbb{N}$ s.t. $|F_n(K) - F(K)| < \epsilon/3$ and $|F_n(-K) - F(-K)| < \epsilon/3$ whenever $n > N$, by $X_n \to X$ in distribution. Hence we have

$$1-F_n(K)+F_n(-K) \leq (1-F(K)+F(-K))+|F_n(K)-F(K)|+|F_n(-K)-F(-K)| < \epsilon.$$

That is, $\mathbb{P}(|X_n| > K) < \epsilon$ whenever $n > N$. For $m \in [N]$, let $K_m$ be the positive number s.t. $\mathbb{P}(|X_m| > K_m) < \epsilon$, and let $K' = \max\{K_1, \cdots, K_m, K\} + 1$. Since $\mathbb{P}(|X| > t)$ is monotonically decreasing, $\mathbb{P}(|X_n| \geq K') < \epsilon$ for all $n$. Thus $\sup_n \mathbb{P}(|X_n| \geq K') \leq \epsilon$, i.e. $\{X_n\}$ is a tight family. ■

**Exercise.** 8.6.3. If $\{X_n\}_{n\geq 1}$ is a tight family and $\{c_n\}_{n\geq 1}$ is a sequence of constants tending to 0, show that $c_nX_n \to 0$ in probability.

**Answer.** Suppose $\{X_n\}$ is a tight family and $\{c_n\}$ is a sequence of constants tending to 0. Fix $\delta > 0$. Given $\epsilon > 0$, then there exists $K > 0$ s.t. $\mathbb{P}(|X_n| \geq K) \leq \epsilon$, and hence a $N \in \mathbb{N}$ s.t. $|c_n|K < \delta$ whenever $n > N$, i.e.

$$\mathbb{P}(|c_nX_n| \geq \delta) \leq \mathbb{P}(|c_nX_n| \geq |c_n|K) \leq \epsilon$$

whenever $n > N$. That is, $\lim_{n\to\infty} \mathbb{P}(|c_nX_n| \geq \delta) \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, we have $\lim_{n\to\infty} \mathbb{P}(|c_nX_n| \geq \delta) = 0$. Again, since $\delta > 0$ is arbitrary, we have $c_nX_n \to 0$ in probability. ■

## 8.7 An alternative characterization of weak convergence

## 8.8 Inversion formulas

## 8.9 Levy's continuity theorem

**Exercise.** 8.9.2. If a sequence of characteristic functions $\{\phi_n\}_{n\geq 1}$ converges pointwise to a characteristic function $\phi$, prove that the convergence is uniform on any bounded interval.

**Answer.** Let $\{X_n\}$ be the sequence of random variables that their characteristic functions $\{\phi_n\}$ converge pointwisely to $X$'s characteristic function $\phi$. By Levy's continuity theorem (THEOREM 8.9.1.), $X_n \to X$ in distribution. Then by EXERCISE 8.6.2., $\{X_n\}$ is a tight family. Hence, given $\epsilon \in (0,1)$, there exists a number $K > 0$ s.t. $\mathbb{P}(|X_n| \geq K) \leq \epsilon$ for all $n$ and $\mathbb{P}(|X| \geq K) \leq \epsilon$. With this $K$, take $\Delta = \epsilon/K$, then for any $|\delta| \in (0, \Delta)$,

$$\begin{aligned}
|\phi_n(a) - \phi_n(a+\delta)| &= |\mathbb{E}(e^{iaX_n} - e^{i(a+\delta)X_n})| \\
&\leq \mathbb{E}\big(|1 - e^{i\delta X_n}|; |X_n| \geq K\big) + \mathbb{E}\big(|1 - e^{i\delta X_n}|; |X_n| < K\big) \\
&\leq 2\epsilon + 2(1-\cos(\epsilon)) < 4\epsilon.
\end{aligned}$$

The last line follows by $|1 - e^{i\delta X_n}| = 2(1-\cos(\delta X_n))$. Also, $|\phi(a) - \phi(a+\delta)| < 4\epsilon$ for the same reason.

Now for the given bounded interval $I = (p, q)$, choose the centers of open covering $\{a_i\}_{i=1}^N$ as $p, p+\Delta, \cdots, p+(N-1)\Delta$, where $N = \lceil \frac{|I|}{\Delta} \rceil$. Then we can cover $I$ by intervals $I_\Delta(a_i) = (a_i - \Delta, a_i + \Delta)$, i.e. $I \subseteq \cup_{i=1}^N I_\Delta(a_i)$. Further, we are able to take $M \in \mathbb{N}$ such that $n > M \Rightarrow |\phi(a_i) - \phi_n(a_i)| < \epsilon$ for all $a_i$. Now fix $n > M$. For any $t \in I$, there is some $a \in \{a_i\}_{i=1}^N$ s.t. $|t - a| < \Delta$, and hence

$$|\phi(t) - \phi_n(t)| \leq |\phi(t) - \phi(a)| + |\phi_n(a) - \phi_n(t)| + |\phi(a) - \phi_n(a)| < 4\epsilon + 4\epsilon + \epsilon = 9\epsilon.$$

That is, $\|\phi - \phi_n\|_{\infty, I} \leq 9\epsilon$ if $n > M$. Since $\epsilon$ is arbitrary, $\phi_n \to \phi$ uniformly on $I$. ■

**Exercise.** 8.9.3. If a sequence of characteristic functions $\{\phi_n\}_{n\geq 1}$ converges pointwise to some function $\phi$, and $\phi$ is continuous at zero, prove that $\phi$ is also a characteristic function.

**Answer.** Let $\{\phi_n\}$ be the sequence of characteristic functions that converge pointwisely to $\phi$. Clearly $\phi(0) = 1$ since $\phi_n(0) = 1$ for all $n$. Also $\phi$ is continuous at 0, for given $\epsilon > 0$ there is a number $a > 0$ s.t. $|\phi(0) - \phi(s)| \leq \epsilon/2$ if $|s| < a$. Hence

$$\frac{1}{a}\int_{-a}^{a} (1 - \phi(s))ds \leq \frac{\epsilon}{2} \cdot 2 = \epsilon.$$

Since $\phi_n \to \phi$ pointwisely and $|\phi_n| \leq 1$, by LDCT we have

$$\lim_{n\to\infty} \frac{1}{a}\int_{-a}^{a}(1-\phi_n(s))ds = \frac{1}{a}\int_{-a}^{a}(1-\phi(s))ds \leq \epsilon.$$

Hence $\limsup_{n\to\infty} \mathbb{P}(|X_n| \geq t) \leq \epsilon$ by letting $t = 2/a$ and PROPOSITION 6.4.4. This implies there is a $N \in \mathbb{N}$ s.t. $n > N \Rightarrow \mathbb{P}(|X_n| \geq t) < 2\epsilon$. Then there exists a larger $K_N$ s.t. $\mathbb{P}(|X_n| \geq K_N) < 2\epsilon$ for all $n$, i.e. $\{X_n\}$ is a tight family.

Since $\{X_n\}$ is a tight family, by Helly's selection theorem, there is a subsequence $\{X_{k(n)}\}$ of $\{X_n\}$ s.t. $X_{k(n)} \to Y$ in distribution for some $Y$. Then by Levy's continuity theorem, $\phi_{k(n)} \to \phi_Y$ pointwisely. But $\phi_n \to \phi$ pointwisely, i.e. every subsequence of $\{\phi_n\}$ converges to $\phi$ pointwisely. Hence $\phi = \phi_Y$ is a characteristic function. ■

## 8.10 The central limit theorem for i.i.d. sums

**Exercise.** 8.10.5 Give a counterexample to show that the i.i.d. assumption in Theorem 8.10.1 cannot be replaced by the assumption of identically distributed and pairwise independent.

**Answer.** (c.f. Durrett's textbook Example 4.5.) Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with $\mathbb{P}(X_1 = 1) = 1/2$, $\mathbb{P}(X_1 = -1) = 1/2$. Define a set of random variables

$$\{Y_i\}_{i=1}^{2^n} = \Big\{X_1 \prod_{j\in S} X_j : S \subseteq \{2, \cdots, n+1\}\Big\}.$$

Then $Y_i$ are identically distributed (as same as $X_1$) and pairwise independent. Also,

$$S_{2^n} = \sum_{i=1}^{2^n} Y_i = X_1(1+X_2)\cdots(1+X_{n+1}),$$

and $\mathbb{P}(S_{2^n} = 2^n) = 1/2^n$, $\mathbb{P}(S_{2^n} = -2^n) = 1/2^n$, $\mathbb{P}(S_{2^n} = 0) = 1 - 2/2^n$ by i.i.d. of $\{X_i\}$. Therefore

$$\mathbb{E}\exp\left(it\frac{S_{2^n}}{\sqrt{2^n}}\right) = 1 - \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}}\cos(t2^{n/2}) \to 1 \text{ as } n \to \infty.$$

That is, by Levy's continuity theorem, $S_{2n}/\sqrt{2^n}$ doesn't converge weakly to the standard normal distribution. ∎

**Exercise.** 8.10.6. Let $X_n \sim Bin(n,p)$. Prove a central limit theorem for $X_n$.

**Answer.** Consider $\{Y_i\}_{i=1}$ be a sequence of i.i.d. $Ber(p)$ random variables. Then $S_n = \sum_{i=1}^n Y_i \sim Bin(n,p)$. By CLT we immediately have $(S_n - np)/\sqrt{np(1-p)}$ converges weakly to the standard normal distribution as $n \to \infty$. ∎

**Exercise.** 8.10.7. Let $X_n \sim Gamma(n, \lambda_n)$. Prove a central limit theorem for $X_n$.

**Answer.** Consider $\{Y_i\}_{i=1}$ be a sequence of i.i.d. $Exp(\lambda)$ random variables. Then $S_n = \sum_{i=1}^n Y_i \sim Gamma(n,\lambda)$. By CLT we immediately have $(S_n - n/\lambda)/\sqrt{n/\lambda^2}$ converges weakly to the standard normal distribution as $n \to \infty$. ∎

**Exercise.** 8.10.8. Suppose that $X_n \sim Gamma(n, \lambda_n)$, where $\{\lambda_n\}_{n=1}^\infty$ is any sequence of positive constants. Prove a CLT for $X_n$.

**Answer.** Here we need to recall the characteristic function of $Gamma(n,\lambda_n)$ random variable. Let $t\in\mathbb{R}$ and $\gamma_1(x)=(\lambda_n-it)x$ for $x\in[0,R]$ be a curve in complex plane. Then we can write

$$\int_0^\infty x^{n-1}e^{-(\lambda_n-it)x}dx=\frac{1}{(\lambda_n-it)^n}\lim_{R\to\infty}\int_{\gamma_1}z^{n-1}e^{-z}dz.$$

Now define $\gamma_2(x)=x$ on $[0,R]$ and $\gamma_3(x)=R+x(\lambda_n-1-it)$ for $x\in[0,R]$. Let $f(z)=z^{n-1}e^{-z}$. Then by Cauchy integral theorem we have

$$\int_{\gamma_1}f(z)dz-\int_{\gamma_3}f(z)dz-\int_{\gamma_2}f(z)dz=0,$$

where

$$\int_{\gamma_3}f(z)dz=e^{-R}(\lambda_n-1-it)\int_0^R(R+(\lambda_n-1-it)x)^{n-1}e^{-(\lambda_n-1)x}e^{-itx}dx$$

tends to 0 as $R\to\infty$. Therefore $\lim_{R\to\infty}\int_{\gamma_1}=\lim_{R\to\infty}\int_{\gamma_2}$, hence.

$$\int_0^\infty x^{n-1}e^{-(\lambda_n-it)x}dx=\frac{1}{(\lambda_n-it)^n}\int_0^\infty x^{n-1}e^{-x}dx=\frac{\Gamma(n)}{(\lambda_n-it)^n}.$$

Thus the characteristic function of $Gamma(n,\lambda_n)$ random variable is

$$\mathbb{E}e^{itX}=\frac{\lambda_n}{(\lambda_n-it)^n}=\Big(1-i\frac{t}{\lambda_n}\Big)^{-n}.$$

Now let $Y_n=(X_n-a_n)/b_n$, then

$$\mathbb{E}e^{itY_n}=e^{-i\frac{a_n}{b_n}t}\Big(1-i\frac{t}{b_n\lambda_n}\Big)^{-n}.$$

One can find that by taking logarithm and doing Taylor expansion, if we set $a_n=n/\lambda_n$ and $b_n=\sqrt{n}/\lambda_n$, then

$$\log\mathbb{E}(e^{itY_n})=-i\sqrt{n}t-n\Big(-i\frac{t}{\sqrt{n}}+\frac{t^2}{2n}+i\frac{t^3}{3n\sqrt{n}}+o\Big(\frac{t^3}{n^{3/2}}\Big)\Big)\to-\frac{t^2}{2}\text{ as }n\to\infty.$$

That is, $\mathbb{E}e^{itY_n}\to e^{-\frac{t^2}{2}}$ pointwisely as $n\to\infty$. By Levy's continuity theorem, we thus have $\frac{Y_n-(n/\lambda_n)}{\sqrt{n}/\lambda_n}$ converges to standard normal r.v. in distribution. ■

**Exercise.** 8.10.10. Let $X_1,X_2,\cdots$ be a sequence of i.i.d. random variables. For each $i\geq 2$, let

$$Y_i=\begin{cases}1 & \text{if } X_i\geq\max\{X_{i-1},X_{i+1}\}\\ 0 & \text{if not.}\end{cases}$$

In other words, $Y_i$ is 1 if and only if the original sequence has a local maximum at $i$. Prove a central limit theorem $\sum_{i=2}^n Y_i$.

**Answer.** Since $X_i$ are i.i.d., one can find that $\{Y_i\}_{i=2}$ is a stationary 2-dependent sequence of r.v. Hence we can apply CLT for stationary $m$-dependent sequence by setting

$$\sigma^2 = \text{Var}(Y_2) + 2\text{Cov}(Y_2, Y_3) + 2\text{Cov}(Y_2, Y_4).$$

Where $\mathbb{E}(Y_2) = \mathbb{E}(Y_2^2) = \mathbb{P}(X_2 = \max\{X_1, X_2, X_3\}) = 1/3$, and

$$\begin{aligned}
\mathbb{E}(Y_2Y_3) &= \mathbb{P}(Y_2 = \max\{X_1, X_2, X_3\}, Y_3 = \max\{X_2, X_3, X_4\}) = 0 \\
\mathbb{E}(Y_2Y_4) &= \mathbb{P}(Y_2 = \max\{X_1, X_2, X_3\}, Y_4 = \max\{X_3, X_4, X_5\}) \\
&= \frac{2 \cdot (1 \cdot 1 \cdot 2! + 1 \cdot 3!)}{5!} = \frac{2}{15},
\end{aligned}$$

therefore $\sigma^2 = 2/45$. And $\frac{\sum_{i=2}^{n}(Y_i - \frac{1}{3})}{\sqrt{2n/45}}$ converges weakly to standard normal distribution as $n \to \infty$ by THEOREM 8.10.9. ■

## 8.11 The Lindeberg–Feller central limit theorem

**Exercise.** 8.11.4. Suppose that $X_n \sim Bin(n, p_n)$, where $\{p_n\}_{n=1}^{\infty}$ is a sequence of constants such that $np_n(1-p_n) \to \infty$. Prove a CLT for $X_n$.

**Answer.** For each $n$, define $S_n = \sum_{i=1}^{n} Y_{p_n,i}$ where $\{Y_{p_n,i}\}_{i=1}^{n}$ are i.i.d. $Ber(p_n)$ random variables. It's suffice to prove CLT for $S_n$ since $S_n \sim Bin(n, p_n)$ has the same distribution to $X_n$. Now, we prove the Lindeberg's condition for $\{Y_{p_n,i}\}_{i=1}^{n}$. Note that $\text{Var}(S_n) = n\text{Var}(Y_{p_n,1}) = np_n(1-p_n)$. Given $\epsilon > 0$ and $n \in \mathbb{N}$, by Cauchy-Schwarz and Markov inequality,

$$\begin{aligned}
\mathbb{E}\Big(\frac{(Y_{p_n,1} - p_n)^2}{np_n(1-p_n)}; \frac{|Y_{p_n,1} - p_n|}{\sqrt{np_n(1-p_n)}} \ge \epsilon\Big)^2 &\le \mathbb{E}\Big(\frac{(Y_{p_n,1} - p_n)^4}{n^2p_n^2(1-p_n)^2}\Big)\mathbb{P}\Big(\frac{(Y_{p_n,1} - p_n)^2}{np_n(1-p_n)} \ge \epsilon^2\Big) \\
&\le \frac{p_n^3 + (1-p_n)^3}{n^2p_n(1-p_n)} \times \frac{1}{n\epsilon^2}.
\end{aligned}$$

Since $np_n(1-p_n) \to \infty$ as $n \to \infty$,

$$\sum_{i=1}^{n} \mathbb{E}\Big(\frac{(Y_{p_n,i} - p_n)^2}{np_n(1-p_n)}; \frac{|Y_{p_n,i} - p_n|}{\sqrt{np_n(1-p_n)}} \ge \epsilon\Big) \le \frac{1}{\epsilon}\sqrt{\frac{p_n^3 + (1-p_n)^3}{np_n(1-p_n)}} \to 0 \text{ as } n \to \infty.$$

That is, $\{Y_{p_n,i}\}_{i=1}^{n}$ satisfies the Lindeberg's condition. In fact, one can simply see that $\mathbb{P}\big(|Y_{p_n,1} - p_n| \ge \epsilon\sqrt{np_n(1-p_n)}\big) = 0$ as $n \to \infty$ since $|Y_{p_n,1} - p_n|$ is bounded. Now by Lindeberg-Feller's CLT, we have $\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}}$ converges weakly to the standard normal distribution. ■

**Exercise.** 8.11.5. Let $X_1, X_2, \cdots$ be a sequence of uniformly bounded independent random variables, and let $S_n = \sum_{i=1}^n X_i$. If $\mathrm{Var}(S_n) \to \infty$, show that $S_n$ satisfies a central limit theorem.

**Answer.** Let $\mu_i = \mathbb{E}(X_i)$ and $s_n^2 = \mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(X_i)$. Since $\{X_i\}_{i=1}^n$ is uniformly bounded, clearly $|X_i - \mu_i|$ is also uniformly bounded, i.e. there is some $M > 0$ s.t. $M \geq |X_i - \mu_i|$ for all $i \in \mathbb{N}$. Now given $\epsilon > 0$, by the fact that $s_n^2 \to \infty$, we choose $N \in \mathbb{N}$ s.t. $s_n^2 > M/\epsilon$ whenever $n > N$. Therefore, for any $n > N$, $\mathbb{P}(|X_i - \mu_i| \geq \epsilon s_n^2) = 0$ for all $i$, i.e. $\mathbb{E}((X_i - \mu_i)^2; |X_i - \mu_i| \geq \epsilon s_n) = 0$ for all $i$, and thus the Lindeberg's condition holds. By Lindeberg-Feller's CLT again, we now have $\frac{S_n - \sum_{i=1}^n \mu_i}{s_n}$ converges weakly to the standard normal distribution as $n \to \infty$. ■

# Appendix

## A Additional Proofs

### A.1 Proof of ??

See https://en.wikipedia.org/wiki/Mass%E2%80%93energy_equivalence.

## References

[Cha] Sourav Chatterjee. *STATS 310 (MATH 230) Lecture notes (ongoing,to be updated)*. URL: https://souravchatterjee.su.domains//stats310notes.pdf.

[Hu] Pingbang Hu. *Academic-Template*. URL: https://github.com/sleepymalc/Academic-Template.git.