Notes on Probability Theory

Jiezheng Wei*

February 25, 2026

This is my notes for the lecture "Probability Theory" provided by Professor Claudio Landim. I add some content based on my own demands. All errors are my own.

1 Introduction

- The setup of probability theory.

  The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space, or a probability measure space, or generally just a measure space, in which

  - ⋆ Ω: an abstract space.
  - ⋆ $\mathcal{F}$: a σ-algebra[1].
  - ⋆ $\mathbb{P}$: the probability measure.

    The probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a set function

    (i) $\mathbb{P}(\varnothing) = 0$.

    (ii) For a sequence $(A_i)_{i\geq 1}$ of mutually disjoint sets, i.e. $A_j \cap A_k = \varnothing$ for all $j \neq k$, we have $\mathbb{P}(\cup_i A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

    (iii) $\mathbb{P}(\Omega) = 1$.

*School of Economics, Singapore Management University, E-mail: jzwei.2022@phdecons.smu.edu.sg

[1]A collection $\mathcal{F}$ of subsets of Ω is a σ-algebra if (i) $\Omega \in \mathcal{F}$; (ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$; (iii) if $A_i \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition 1 (random variable)**.** A real-valued random variable is a mapping $X : \Omega \to \mathbb{R}$, and X is $\mathcal{B}(\mathbb{R})$-measurable where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra. That is, $\forall A \in \mathcal{B}(\mathbb{R})$, $X^{-1}(A) \in \mathcal{F}$.

Anecdotal remark. Two legendary probability theorists, Joe Doob and William Feller, had a story on the name of this measurable function. Doob wanted to call "random variable" as "chance variable", Feller wanted to call it "random variable". In order to be consistent in their textbooks, they tossed a coin and Feller won, see Snell (1997).

Definition 2 (Probability distribution measures)**.** For X a $\mathbb{R}$-valued random variable, the distribution measure μ_X is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined as

$$\begin{aligned}\mu_X(A) = \mathbb{P}[X^{-1}(A)] &= \mathbb{P}\{\omega : X(\omega) \in A\} \\ &= \mathbb{P}(X \in A)\end{aligned}$$

That is, the probability distribution measure is the pushforword measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ of the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$.

So X transform the probability space $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$.

Remark: understanding the necessities of "measurability" of X in the definition of the random variable.

Definition 3 (Distribution functions)**.** Let X be a $\mathbb{R}$-valued random variable, its distribution function $F_X : \mathbb{R} \to [0, 1]$ is defined as

$$F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

Proposition 1.1. The distribution function has the following properties.

(i) F_X is monotone, i.e., $x \leq y \implies F_X(x) \leq F_X(y)$.

(ii) F_X is càdlàg.

(iii) $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$.

Remark: a function $F : \mathbb{R} \to [0, 1]$ satisfying the three properties in Proposition 1.1 is a distribution function.

- (Optional) A little more about collection of subsets. Let Ω denote the whole space, it can the the sample space, or the transformed space in which a random variable takes value. This part is based on Chapter 1 and Appendix A of Durrett (2019).

  Definition 4 (semialgebra)**.** A collection of subsets, say $\mathcal{S}$, is a semialgebra if (i) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$; (ii) $A \in \mathcal{S} \implies A^c$ is a finite disjoint union of sets in $\mathcal{S}$.

  Definition 5 (algebra)**.** A collection of subsets, say $\mathcal{A}$, is an algebra if (i) $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$; (ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$.

  Remark 1: clearly an algebra is a semialgebra, a σ-algebra is a algebra.

  Remark 2: the algebra can be equivalently defined as (i) closed under union; (ii) closed under complement.

  Lemma 1.1. Let $\mathcal{S}$ be a semialgebra, then $\bar{\mathcal{S}} = \{\text{finite disjoint unions of sets in } \mathcal{S}\}$ is a algebra called algebra generated by $\mathcal{S}$.

  Proof. (i) Let $A, B \in \bar{\mathcal{S}}$, then $A = \sum_{i=1}^{n_S} S_i$ for $S_i \in \mathcal{S}$ and $B = \sum_{i=1}^{n_T} T_i$ for $T_i \in \mathcal{S}$, then $A \cap B = \sum_{i=1}^{n_S} \sum_{j=1}^{n_T} S_i \cap T_j$, note $S_i \cap T_j \in \mathcal{S}$, so $A \cap B \in \bar{S}$; (ii) let $A \in \bar{\mathcal{S}}$, then $A = \sum_{i=1}^{n_S} S_i$ where $S_i \in \mathcal{S}$, and $A^c = \cap_{i=1}^{n_S} S_i^c \in \bar{\mathcal{S}}$ by the reasoning that S_i^c can be written as finite disjoint union of sets in $\mathcal{S}$, so $S_i^c \in \bar{\mathcal{S}}$ and (i). $\square$

  Definition 6 (π-system)**.** A collection of subsets $\mathcal{P}$, is a π-system if $A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$.

  Definition 7 (λ-system)**.** A collection of subsets $\mathcal{L}$, is a λ-system if (i) $\Omega \in \mathcal{L}$; (ii) $A, B \in \mathcal{L}, A \subset B \implies B \setminus A \in \mathcal{L}$; (iii) if $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$.

  Theorem 1.1 (Dynkin's π-λ theorem)**.** Let $\mathcal{P}$ be a π-system and $\mathcal{L}$ be a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

  Proof. Step 1. If $\ell(\mathcal{P})$ is the smallest λ-system containing $\mathcal{P}$, then $\ell(\mathcal{P})$ is a σ-algebra.

  Note: a λ-system that is closed under intersection is a σ-algebra. Because (i) $\Omega \in \mathcal{L}$ is given; (ii)$\Omega \in \mathcal{L}$, $A \in \mathcal{L}$ and $A \subset \Omega$, so $A^c = \Omega \setminus A \in \mathcal{L}$; (iii) first we realize $\mathcal{L}$ is closed under finite union for $A_1 \cup A_2 = (A_1^c \cap A_2^c)^c$. Thus let $A_i \in \mathcal{L}$ and $B_n = \cup_{i=1}^n A_i$, so $B_n \in \mathcal{L}$ and $B_n \uparrow \cup_{i=1}^\infty A_i$, so $\cup_{i=1}^\infty A_i \in \mathcal{L}$.

Then it's sufficient to show $\ell(\mathcal{P})$ is closed under intersection. Define $\mathcal{G}_A = \{B : A \cap B \in \ell(\mathcal{P})\}$. We show

(a) If $A \in \ell(\mathcal{P})$, then $\mathcal{G}_A$ is a λ-system.

(i) $\Omega \cap A = A \in \ell(\mathcal{P})$, so $\Omega \in \mathcal{G}_A$; (ii) Let $B, C \in \mathcal{G}_A$ and $B \subset C$, then $(C \setminus B) \cap A = (C \cap A) \setminus (B \cap A) \in \ell(\mathcal{P})$ for $C \cap A \in \ell(\mathcal{P})$, $B \cap A \in \ell(\mathcal{P})$, $(B \cap A) \subset (C \cap A)$ and the fact that $\ell(\mathcal{P})$ is a λ-system, so $C \setminus B \in \mathcal{G}_A$; (iii) Let $B_n \in \mathcal{G}_A, \forall n$ and $B_n \uparrow B$, so we have $(B_n \cap A) \uparrow (B \cap A)$ and $B_n \cap A \in \ell(\mathcal{P})$ for all n, so $B \cap A \in \ell(\mathcal{P})$ for $\ell(\mathcal{P})$ being the λ-system.

(b) If $A, B \in \ell(\mathcal{P})$, then $A \cap B \in \ell(\mathcal{P})$.

If $A \in \mathcal{P} \subset \ell(\mathcal{P})$, then $\mathcal{P} \subset \mathcal{G}_A$ for $\mathcal{P}$ is a π-system, thus (a) implies $\ell(\mathcal{P}) \subset \mathcal{G}_A$. So we have, if $A \in \mathcal{P}$ and $B \in \ell(\mathcal{P})(\subset \mathcal{G}_A)$, then $A \cap B \in \ell(\mathcal{P})$ by definition of $\mathcal{G}_A$.

By the above argument interchanging the role of A and B, we have, if $A \in \ell(\mathcal{P})$ and $B \in \mathcal{P}$, then $A \cup B \in \ell(\mathcal{P})$. This implies $\mathcal{P} \subset \mathcal{G}_A$ by the arbitrariness of B. So $\ell(\mathcal{P}) \subset \mathcal{G}_A$.

Therefore take $A \in \ell(\mathcal{P})$ and $B \in \ell(\mathcal{P}) \subset \mathcal{G}_A$, then $A \cap B \in \ell(\mathcal{P})$ by the definition of $\mathcal{G}_A$. This finish the proof of the $\pi - \lambda$ theorem.

<u>Step 2</u>. Clearly, a σ-algebra is a λ-system, so $\sigma(\mathcal{P}) \subset \ell(\mathcal{P}) \subset \mathcal{L}$. $\square$

Definition 8 (A measure on algebra)**.** A set function μ is a measure on algebra $\mathcal{A}$, $\mu : \mathcal{A} \to \mathbb{R}$ if

(i) $\mu(A) \geq \mu(\varnothing) = 0$ for all $A \in \mathcal{A}$.

(ii) if $A_i \in \mathcal{A}$ are disjoint and their union is in $\mathcal{A}$, then $\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$.

Definition 9 (σ-finiteness of a measure)**.** A measure is σ-finite if $\exists A_i \in \mathcal{A}$ and $\mu(A_i) < \infty$ for $i \in \{1, 2, \cdots\}$, and $\cup_{i=1}^\infty A_i = \Omega$.

Theorem 1.2. Let $\mathcal{S}$ be a semialgebra and let $\mu : \mathcal{S} \to \mathbb{R}_+$ with $\mu(\varnothing) = 0$. Suppose

(i) μ is finite-additive. $\mu(\sum_{i=1}^n S_i) = \sum_{i=1}^n \mu(S_i)$.

(ii) μ is σ-subadditive. $\mu(\sum_{i=1}^\infty S_i) \leq \sum_{i=1}^\infty \mu(S_i)$.

where $\sum$ for sets means disjoint union. Then

(a) μ has a unique extension $\bar{\mu}$ that is a measure on $\bar{\mathcal{S}}$.

(b) If μ is σ-finite measure on an algebra $\mathcal{A}$, then there is a unique extension v that is a measure on $\sigma(\mathcal{S})$.

Remark. The statement (b) is actually the Carathéodory extension theorem.

Proof. (a) Recall Lemma 1.1 that $\bar{\mathcal{S}}$ is the collection of finite disjoint union of set in $\mathcal{S}$. For $A \in \bar{\mathcal{S}}$, suppose $A = \sum_{i=1}^{n_B} B_i$ for $B_i \in \mathcal{S}$, define $\bar{\mu}(A) = \sum_{i=1}^{n_B} \mu(B_i)$.

To check whether or not $\bar{\mu}$ is well defined. Suppose meanwhile $A = \sum_{i=1}^{n_C} C_i$ for $C_i \in \mathcal{S}$. We need to check $\sum_{i=1}^{n_B} \mu(B_i) = \sum_{i=1}^{n_C} \mu(C_i)$. Note $B_i = \sum_{j=1}^{n_C}(B_i \cap C_j)$ and $C_j = \sum_{i=1}^{n_B}(C_j \cap B_i)$ so

$$\sum_{i=1}^{n_B} \mu(B_i) = \sum_{i=1}^{n_B} \sum_{j=1}^{n_C} \mu(B_i \cap C_j) = \sum_{j=1}^{n_C} \sum_{i=1}^{n_B} \mu(B_i \cap C_j) = \sum_{j=1}^{n_c} \mu(C_j)$$

This establish the well-definedness of $\bar{\mu}$.

It's trivial that $\bar{\mu}$ is additive: if $A, B \in \bar{\mathcal{S}}$ and they are disjoint, then $\bar{\mu}(A + B) = \bar{\mu}(A) + \bar{\mu}(B)$. What we need to check is the σ-additivity by recalling Definition 8 of a measure on algebra. Let $(B_i)_i$ be a sequence of disjoint sets and $B_i \in \bar{\mathcal{S}}$ for $i = 1, 2, \cdots, \infty$. Thus $B_i = \sum_{j=1}^{n_i} S_{i,j}$ where $S_{i,j} \in \mathcal{S}$. Note

$$\sum_{i=1}^{\infty} \bar{\mu}(B_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} \mu(S_{i,j})$$

Note the RHS of the above is still a countable sum and $\sum_{i=1}^\infty B_i = \sum_{i=1}^\infty \sum_{j=1}^{n_i} S_{i,j}$, so without loss of generality we can assume $B_i \in \mathcal{S}$ by the trick that we replace B_i by $S_{i,j}$. Let $A = \sum_{i=1}^\infty B_i$ and $A \in \bar{S}$, so $A = \sum_{j=1}^{n_T} T_j$ where $T_j \in \mathcal{S}$ and $T_j = \sum_{i=1}^\infty (T_j \cap B_i)$, so (ii) implies

$$\bar{u}(A) = \sum_{j=1}^{n_T} \mu(T_j) \leq \sum_{j=1}^{n_T} \sum_{i=1}^{\infty} \mu(T_j \cap B_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{n_T} \mu(T_j \cap B_i) = \sum_{i=1}^{\infty} \mu(B_i)$$

where the last equality is due to (i). To prove the opposite inequality, let $A_n = \sum_{i=1}^n B_i$ and $C_n = A \cap A_n^c$. Note $A_n \in \bar{\mathcal{S}}$, so $A_n^c \in \bar{\mathcal{S}}$ and we have $C_n = A \cap A_n^c \in \bar{\mathcal{S}}$ for $\bar{\mathcal{S}}$ being an algebra. So

$$\bar{\mu}(A) = \bar{\mu}(B_1) + \bar{\mu}(B_2) + \cdots + \bar{\mu}(B_n) + \bar{\mu}(C_n) \geq \sum \bar{\mu}(B_1) + \bar{\mu}(B_2) + \cdots + \bar{\mu}(B_n)$$

letting $n \to \infty$ gives $\bar{\mu}(A) \geq \sum_{i=1}^\infty \bar{\mu}(B_i)$. This finishes the proof that $\bar{\mu}$ is σ-additive. And (b) relies on Carathéodory extension theorem. $\square$

Lemma 1.2. Let $\mathcal{P}$ be a π-system, if v_1, v_2 are measures on σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ where $\sigma(\mathcal{P}) \subset \mathcal{F}_1$ and $\sigma(\mathcal{P}) \subset \mathcal{F}_2$. And v_1, v_2 agree on $\mathcal{P}$. Moreover if $\exists$ a sequence $A_n \in \mathcal{P}$ with $A_n \uparrow \Omega$, then v_1 and v_2 agree on $\sigma(\mathcal{P})$.

Proof. Take $A \in \mathcal{P}$, then $v_1(A) = v_2(A)$ and define

$$\mathcal{L}_A = \{B \in \sigma(\mathcal{P}) : v_1(A \cap B) = v_2(A \cap B)\}$$

we claim that $\mathcal{L}_A$ is a λ-system. (i) $\Omega \in \mathcal{L}_A$, trivial; (ii) suppose $B, C \in \mathcal{L}_A$ and $C \subset B$, then

$$\begin{aligned} v_1(A \cap (B \setminus C)) = v_1\left((A \cap B) \setminus (A \cap C)\right) &\overset{(1)}{=} v_1(A \cap B) - v_1(A \cap C) \\ &\overset{(2)}{=} v_2(A \cap B) - v_2(A \cap C) = v_2\left(A \cap (B \setminus C)\right) \end{aligned}$$

where step (1) is due to v_1 is also a measure on $\sigma(\mathcal{P})$ and step (2) is due to $B, C \in \mathcal{L}_A$. (iii) Let $D_n \in \mathcal{L} - A$ and $D_n \uparrow D$, so

$$v_1(A \cap D) = \lim_{n \to \infty} v_1(A \cap D_n) = \lim_{n \to \infty} v_2(A \cap D_n) = v_2(A \cap D)$$

where the first and the last equality is due to that $A \cap D_n$ is monotone. This finish the proof that $\mathcal{L}_A$ is a λ-system.

Since $\forall E \in \mathcal{P}$, $A \cap E \in \mathcal{P}$, so $v_1(A \cap E) = v_2(A \cap E)$, implying $E \in \mathcal{L}_A$. Thus $\mathcal{P} \subset \mathcal{L}_A$, and π-λ theorem tells $\sigma(\mathcal{P}) \subset \mathcal{L}_A$ for any $A \in \mathcal{P}$. This gives us $\sigma(\mathcal{P}) \subset \mathcal{L}_{A_n}$ for all n. So for any $F \in \mathcal{L}$,

$$v_1(F \cap A_n) = v_2(F \cap A_n)$$

Letting $n \to \infty$ and monotonicity of $(F \cap A_n) \uparrow F \cap \Omega = F$ gives $v_1(F) = v_2(F)$ for any $F \in \sigma(\mathcal{P})$. This completes the proof that v_1 and v_2 will agree on $\sigma(\mathcal{P})$. $\square$

Lemma 1.3. A measure μ (not necessarily σ-finite) defined on an algebra $\mathcal{A}$ has an extension to $\sigma(\mathcal{A})$.

Proof. This proof not only proves, but also clarifies math concepts.

If $E \subset \Omega$, define the outer measure

$$\mu^*(E) = \inf_{\{A_i\}} \sum_i \mu(A_i), \quad \text{where } E \subset \cup_i A_i, \text{ where } A_i \in \mathcal{A}, \forall i$$

Suppose v is a measure that agrees with μ on $\mathcal{A}$, and moreover defined for E, then

$$v(E) \leq v(\cup_i A_i) \leq \sum_i v(A_i) = \sum_i \mu(A_i)$$

so $\mu^*(E)$ is an upper bound on the measure of E. Intuitively, measurable sets are ones for which the upper bound is tight. E is measurable if

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \cap E^c), \quad \forall F \subset \Omega$$

Immediately, we have

(i) Monotonicity. If $E \subset F$, then $\mu^*(E) \leq \mu^*(F)$

(ii) Subadditivity. If $F \subset$

□

Theorem 1.3 (Carathéodory extension theorem)**.** Let μ be a σ-finite measure on an algebra $\mathcal{A}$. Then μ has an unique extension to $\sigma(\mathcal{A})$.

Proof. (1) Existence.

(2) Let's prove uniqueness first. □

- Then we go back to our probability theory.

Theorem 1.4. There is one-one correspondence between probability distribution measure and the distribution function. That is,

(1) If μ is a probability distribution measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then $F(x) \equiv \mu((-\infty, x])$ is a distribution function.

(2) If $F(x)$ is a distribution function, then there exists a probability distribution measure μ that generates F.

Proof. (1) by the Proposition 1.1.

(2) The proof relies on Carathéodory extension theorem.

Define $\mu((a,b]) = F(b) - F(a)$, consider the collection $\mathcal{S} = \{(a,b] : -\infty \leq a < b \leq \infty\}$. Then $\mu : \mathcal{S} \to [0,1]$ and $\mathcal{S}$ is a semi-algebra.

And μ is σ-additive $\square$

Remark. In some books and papers, it is by default assumed that the sample space and support space of X are the same, of course this can be done by treating $X(\omega) = \omega$. Moreover, Theorem 1.4 tells us it loses nothing to directly treat the support space of X as its sample space, and the distribution probability measure (pushforward measure) to be the probability measure.

Definition 10 (Discrete random variables)**.** X is a discrete random variable *iff* $\exists$ a countable set $\mathscr{B} \subset \mathbb{R}$, s.t. $\mathbb{P}(X \in \mathscr{B}) = 1$.

Definition 11 (Discrete distribution functions)**.** The distribution function F is discrete *iff* $\exists$ countable sets $\mathscr{B} = \{x_j : x_j \in \mathbb{R}\}$ and $\{P_j\}$ with $\sum_j P_j = 1$, s.t. $F(x) = \sum_{j: x_j \leq x} P_j$. (Check that F in this definition is indeed a distribution function, i.e. check the 3 properties).

Exercise. If a random variable X is discrete, then its distribution function is discrete.

- Completion of a probability space.

We first introduce the completeness concept of a probability space, then proceed the development of the idea.

Definition 12 (Complete probability space)**.** A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *complete* if $\forall A \subset F$ where $F \in \mathcal{F}$ and $\mathbb{P}(F) = 0$, $A \in \mathcal{F}$. And if so, we also say $\mathcal{F}$ is $\mathbb{P}$-complete or just complete if it's clear what the underlying $\mathbb{P}$ is.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Call N a $\mathbb{P}$-*null* set if $N \in \mathcal{F}$ and $\mathbb{P}(N) = 0$. A set $B \subset \Omega$ is said to be $\mathbb{P}$-*negligible* if there exists a $\mathbb{P}$-null set N s.t. $B \subset N$, note here B is not necessarily measurable. Let $\mathcal{N} = \{B \subset \Omega : B \text{ is } \mathbb{P}\text{-negligible}\}$. Clearly in these terminologies, $\mathcal{F}$ is $\mathbb{P}$-complete if every $\mathbb{P}$-negligible set belongs to $\mathcal{F}$.

If $\mathcal{F}$ is not complete, define its $\mathbb{P}$-completion as follows.

$$\bar{\mathcal{F}} = \{A \cup M : A \in \mathcal{F}, M \in \mathcal{N}\}$$

Proposition 1.2. $\bar{\mathcal{F}}$ is a σ-algebra.

Proof. $\bar{\mathcal{F}}$ is a σ-algebra:

(i) Note $\Omega = \Omega \cup M$ for any $M \in \mathcal{N}$.

(ii) Suppose $A \cup M \in \bar{\mathcal{F}}$, then $(A \cup M)^c = A^c \cap M^c$. Note M is negligible, so $\exists$ a $\mathbb{P}$-null set $N \in \mathcal{F}$, s.t. $M \subset N$, so $N^c \subset M^c$

$$\begin{aligned}(A \cup M)^c = A^c \cap M^c &= A^c \cap (N^c \cup (M^c \setminus N^c)) \\ &= (A^c \cap N^c) \cup (A^c \cap (M^c \setminus N^c)) \\ &= (A^c \cap N^c) \cup (A^c \cap M^c \cap N)\end{aligned}$$

note $(A^c \cap N^c) \in \mathcal{F}$ and $(A^c \cap M^c \cap N) \in \mathcal{N}$.

(iii) Let $A_1 \cup M_1, A_2 \cup M_2, \dots \in \bar{\mathcal{F}}$, then

$$\bigcup_{n=1}^{\infty} (A_n \cup M_n) = (\bigcup_{n=1}^{\infty} A_n) \cup (\bigcup_{n=1}^{\infty} M_n)$$

clearly $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, suppose $M_n \subset N_n$ with $N_n \in \mathcal{F}$ and $\mathbb{P}(N_n) = 0$. So

$$\bigcup_{n=1}^{\infty} M_n \subset \bigcup_{n=1}^{\infty} N_n$$

thus $\mathbb{P}(\bigcup_{n=1}^{\infty} M_n) \leq \mathbb{P}(\bigcup_{n=1}^{\infty} N_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(N_n) = 0$, so $\bigcup_{n=1}^{\infty} M_n \in \mathcal{N}$.

□

We can define a new probability measure over $\bar{\mathcal{F}}$ as $\bar{\mathbb{P}} : \bar{\mathcal{F}} \to [0, 1]$ by

$$\bar{\mathbb{P}}(A \cup M) = \mathbb{P}(A), \quad A \in \mathcal{F}, M \in \mathcal{N}$$

Proposition 1.3. $\bar{\mathbb{P}}$ is an extension of $\mathbb{P}$.

Proof. We need to show (a) $\bar{\mathbb{P}}$ is a probability measure on $\bar{\mathcal{F}}$; (b) $\mathcal{F} \subset \bar{\mathcal{F}}$ and $\bar{\mathbb{P}}(A) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$. For (a), let's check the three requirements.

(i) $\bar{\mathbb{P}}(A \cup M) = \mathbb{P}(A) \geq 0$.

(ii) Let $\{A_k \cup M_k\}_{k \geq 1}$ be disjoint sets in $\bar{\mathcal{F}}$, so $\{A_k\}_{k \geq 1}$ are also disjoint, thus

$$\begin{aligned}\bar{\mathbb{P}}\left[\bigcup_{n=1}^{\infty}(A_n \cup M_n)\right] &= \bar{\mathbb{P}}\left[(\bigcup_{n=1}^{\infty} A_n) \cup (\bigcup_{n=1}^{\infty} M_n)\right] \overset{(\star)}{=} \mathbb{P}(\bigcup_{n=1}^{\infty} A_n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \bar{\mathbb{P}}(A_n \cup M_n)\end{aligned}$$

where equality $(\star)$ is due to $\bigcup_{n=1}^{\infty} M_n \in \mathcal{N}$.

(iii) $\mathbb{P}^*(\Omega) = \mathbb{P}^*(\Omega \cup M) = \mathbb{P}(\Omega) = 1$ for all $M \in \mathcal{N}$.

For (b), we note $\varnothing \in \mathcal{N}$. Therefore $\forall A \in \mathcal{F}$, $A = A \cup \varnothing \in \bar{\mathcal{F}}$, we have $\mathcal{F} \subset \bar{\mathcal{F}}$ and $\bar{\mathbb{P}}(A) = \bar{\mathbb{P}}(A \cup \varnothing) = \mathbb{P}(A)$. □

Proposition 1.4. $\bar{\mathcal{F}}$ is $\bar{\mathbb{P}}$-complete, so indeed every σ-algebra has a completion.

Proof. For the probability space $(\Omega, \bar{\mathcal{F}}, \bar{\mathbb{P}})$, we need to show for every $\bar{\mathbb{P}}$-negligible set M, $M \in \bar{\mathcal{F}}$. The idea of proof is to show M is also $\mathbb{P}$-negligible.

Suppose $\exists N \in \bar{\mathcal{F}}$, s.t. $M \subset N$ and $\bar{\mathbb{P}}(N) = 0$. Let's write $N = A' \cup M'$ where $A' \in \mathcal{F}$ and M' is $\mathbb{P}$-negligible. And we have $\bar{\mathbb{P}}(A' \cup M') = \mathbb{P}(A') = 0$. Note M' is $\mathbb{P}$-negligible, so $\exists N' \in \mathcal{F}$, s.t. $M' \subset N'$ and $\mathbb{P}(N') = 0$. Clearly $(A' \cup N') \in \mathcal{F}$, thus its $\mathbb{P}$-probability

$$\mathbb{P}(A' \cup N') \leq \mathbb{P}(A') + \mathbb{P}(N') = 0$$

thus $(A' \cup N')$ is a $\mathbb{P}$-null set, and note $M \subset N = (A' \cup M') \subset (A' \cup N')$. Thus M is a $\mathbb{P}$-negligible set. Note $\varnothing \in \mathcal{F}$, so $M = \varnothing \cup M \in \bar{\mathcal{F}}$. □

We have shown $(\Omega, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ is a complete extension of $(\Omega, \mathcal{F}, \mathbb{P})$, and there might be other complete extensions $(\Omega, \mathcal{F}^*, \mathbb{P}^*)$ which means $\mathcal{F} \subset \mathcal{F}^*$ and $\mathbb{P}^*|_{\mathcal{F}} = \mathbb{P}$ on $\mathcal{F}$. We note complete extensions can be different.

Proposition 1.5. Let $(\Omega, \mathcal{F}_1^*, \mathbb{P}_1^*)$ and $(\Omega, \mathcal{F}_2^*, \mathbb{P}_2^*)$ be two complete extensions of $(\Omega, \mathcal{F}, \mathbb{P})$. Then it's possible that $\mathbb{P}_1^*$ and $\mathbb{P}_2^*$ don't agree with each other on $\mathcal{F}_1^* \cap \mathcal{F}_2^*$.

Proof. To be added. $\square$

And there is a reason why we consider the completion procedure of $\bar{\mathcal{F}}$.

Proposition 1.6. $(\Omega, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ is the minimum complete extension.

Proof. To be added. $\square$

- Kolmogorov extension theorem.

2 Independence

This section is where probability theory differs from measure theory. As P37 of Durrett (2019) says: "Measure theory ends and probability theory begins with the definition of independence."

- Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $A \in \mathcal{F}$, we call A an *event*.

Definition 13 (Independence of events)**.** Events $A_1, A_2, \cdots, A_N$ are independent if for any $\{n_1, \cdots, n_p\} \subset \{1, \cdots, N\}$ with $n_i \neq n_j$ if $i \neq j$, we have

$$\mathbb{P}(\cap_{j=1}^p A_{n_j}) = \prod_{j=1}^p \mathbb{P}(A_{n_j})$$

Definition 14 (Independence of random variables)**.** Random variables $X_1, X_2, \cdots, X_N$ are independent if for any $B_1, B_2, \cdots, B_N \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}\left[\cap_{j=1}^N \{X_j \in B_j\}\right] = \prod_{j=1}^N \mathbb{P}(X_j \in B_j)$$

Moreover, we can extend the concept to suit a possibly continuously indexed set of random variables (a continuous-time stochastic process). A family of random variables $\{X_\alpha : \alpha \in I\}$ are independent if $\forall N$ and $\forall \{\alpha_1, \cdots, \alpha_N\} \subset I$, $X_{\alpha_1}, \cdots, X_{\alpha_N}$ are independent.

Lemma 2.1. If $\{X_1, \cdots, X_N\}$ are independent, then $\forall p < N$, $\{X_{n_1}, \cdots, X_{n_p}\}$ is independent where $n_j \neq n_k$ for $j \neq k$.

Proof. The lemma is proved by noting

$$\mathbb{P}\left[\cap_{j=1}^{p}\{X_{n_j} \in B_j\}\right] \overset{(\star)}{=} \mathbb{P}\left[\cap_{i=1}^{N}\{X_i \in A_i\}\right] = \Pi_{i=1}^{N}\mathbb{P}(X_i \in A_i) = \Pi_{j=1}^{p}\mathbb{P}(X_j \in B_j)$$

where step $(\star)$ is by defining A_i as

$$A_i = \begin{cases} B_j & \text{if } i = n_j \\ \mathbb{R} & \text{if } i \notin \{n_1, \cdots, n_p\} \end{cases}$$

$\square$

Definition 15 (Random vector)**.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the measurable space $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, then $X : \Omega \to \mathbb{R}^N$ is a random vector if $\forall A \in \mathcal{B}(\mathbb{R}^N)$, $X^{-1}(A) \in \mathcal{F}$.

Remark. Clearly we can define distribution function and probability distribution measure for random vectors. Let $X = (X_1, \cdots, X_N)^\top$ be a random vector, then the distribution function of X is $F_X : \mathbb{R}^N \to [0, 1]$ and $F_X(x) = \mathbb{P}(X_1 \leq x_1, \cdots, X_N \leq x_N)$ for $x = (x_1, \cdots, x_N)^\top$. And μ_X on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ is $\mu_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A)$.

Lemma 2.2. Let's denote a random vector by $X = (X_1, \cdots, X_N)^\top$ and X_j has distribution function F_{X_j} and probability distribution function μ_{X_j}. Then the following three is equivalent

(i) $X_1, \cdots, X_N$ are independent *iff* $F_X(x_1, \cdots, x_N) = \prod_{i=1}^{N} F_{X_i}(x_i)$ where $x_i \in \mathbb{R}$.

(ii) $X_1, \cdots, X_N$ are independent *iff* $\mu_X(B_1 \times \cdots \times B_N) = \prod_{i=1}^{N} \mu_{X_i}(B_i)$ where B_i's are Borel sets.

Proof. Left as an exercise. Some extension from $(-\infty, x_j]$ to any Borel sets. $\square$

Theorem 2.1. Let $X_1, \cdots, X_N$ be independent real-valued random variables, and $f_1, f_2, \cdots, f_N : \mathbb{R} \to \mathbb{R}$ are Borel measurable functions, then $f_1(X_1), \cdots, f_N(X_N)$ are independent.

Proof. Let $B_1, \cdots, B_N$ denote Borel sets, so

$$\begin{aligned} \mathbb{P}\left[\bigcap_{i=1}^{N}\{f_i(X_i) \in B_i\}\right] &= \mathbb{P}\left[\bigcap_{i=1}^{N}\{X_i \in f_i^{-1}(B_i)\}\right] \overset{(\star)}{=} \prod_{i=1}^{N}\mathbb{P}\left[\{X_i \in f_i^{-1}(B_i)\}\right] \\ &= \prod_{i=1}^{N}\mathbb{P}\left[\{f_i(X_i) \in B_i\}\right] \end{aligned}$$

where step $(\star)$ is because $f_i^{-1}(B_i)$'s are Borel sets. $\square$

Theorem 2.2. Let $X_1, \cdots, X_N$ be independent real-valued random variables, $1 \le n_1 \le n_2 \le \cdots \le n_p = N$ and Borel measurable functions

$$\begin{aligned} &f_1 : \mathbb{R}^{n_1} \to \mathbb{R} \\ &f_j : \mathbb{R}^{n_j - n_{j-1}} \to \mathbb{R}, \text{ for } 2 \le j \le p \end{aligned}$$

then $f_1(X_1, \cdots, X_{n_1}), f_2(X_{n_1+1}, \cdots, X_{n_2}), \cdots, f_p(X_{n_{p-1}+1}, \cdots, X_{n_p})$ are independent.

Proof. Left as exercise. $\square$

Theorem 2.3. On probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let X and Y be independent and integrable which means $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$, then we have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Proof. $\square$

-

3 Applications of Independence

- Weak law of large numbers.

Theorem 3.1. Suppose $\{X_j, j \ge 1\}$ are *i.i.d.*, defined on the same $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}X_1^2 < \infty$, $\mathbb{E}|X_1| < \infty$, $m \equiv \mathbb{E}X_1$. T

4 Large Deviation for *i.i.d.* RVs, Upper Bound

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define a sequence of *iid* r.v.'s $(X_n)_{n\geq 1}$.

We assume for any $\lambda > 0$, $\mathbb{E}e^{\lambda X_1} < \infty$

5 Large Deviation, Lower Bound

6 Convergence of Random Variables.

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we consider a sequence of r.v.'s $(X_n)_{n\geq 1}$ defined on the probability space. Usually we let X denote the limit of the sequence.

- Definitions. The first perception about the notion of convergence of random variables might be $X_n \to X$ if $X_n(\omega) \to X(\omega), \forall \omega \in \Omega$, but this is too demanding and exclude the allowance of randomness in closeness.

  Definition 16 (Almost sure convergence). $X_n \overset{a.s.}{\to} X$ *iff* $\exists A \in \mathcal{F}, \mathbb{P}(A) = 1$, s.t. $\forall \omega \in \Omega$, $X_n(\omega) \to X(\omega)$.

  Remark 1. $X_n(\omega) \to X(\omega) \Leftrightarrow \forall \varepsilon > 0, \exists n_0$, s.t. $n \geq n_0, |X_n(\omega) - X(\omega)| \leq \varepsilon$. It's equivalent to say

  $$\omega \in \cap_{k\geq 1} \cup_{n_0 \geq 1} \cap_{n\geq n_0} \{\omega' : |X_n(\omega') - X(\omega')| \leq \frac{1}{k}\} \in \mathcal{F}$$

  because indeed $\{|X_n - X| \leq \frac{1}{k}\} \in \mathcal{F}$. Therefore $X_n \overset{a.s.}{\to} X \Leftrightarrow$

  $$\begin{aligned}
  & \mathbb{P}\{\cap_{k\geq 1} \cup_{n\geq 1} \cap_{m\geq n} |X_m - X| \leq \frac{1}{k}\} = 1 & (1)\\
  \Leftrightarrow \quad & \mathbb{P}\{\cup_{n\geq 1} \cap_{m\geq n} |X_m - X| \leq \frac{1}{k}\} = 1, \forall k \geq 1 & (2)\\
  \Leftrightarrow \quad & \lim_{n\to\infty} \mathbb{P}\{\cap_{m\geq n} |X_m - X| \leq \frac{1}{k}\} = 1, \forall k \geq 1 & (3)
  \end{aligned}$$

  Proposition 6.1. We have

  $$\begin{aligned}
  & X_n \overset{a.s.}{\to} X \\
  \Leftrightarrow \quad & \lim_{n\to\infty} \mathbb{P}(\cap_{m\geq n}\{|X_m - X| \leq \varepsilon\}) = 1, \forall \varepsilon > 0 \\
  \Leftrightarrow \quad & \lim_{n\to\infty} \mathbb{P}(\cup_{m\geq n}\{|X_m - X| > \varepsilon\}) = 0, \forall \varepsilon > 0
  \end{aligned}$$

Definition 17 (Convergence in probability)**.** $X_n \xrightarrow{p} X$ *iff* $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

Lemma 6.1. If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{\mathbb{P}} X$.

Proof. $\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(\cup_{m\geq n}\{|X_m - X| > \varepsilon\}) \to 0$ due to $X_n \xrightarrow{a.s.} X$. □

Lemma 6.2 (Subsequence characterization of convergence in probability)**.** We have following results.

(i) If $X_n \xrightarrow{\mathbb{P}} X$, then $\exists \{n_k\}_{k\geq 1}$, s.t. $X_{n_k} \xrightarrow{a.s.} X$.

(ii) If for any subsequence $(n_k)_{k\geq 1}$ of $(X_n)_{n\geq 1}$, there exists a subsubsequence $(n_{k_\ell})_{\ell\geq 1}$, s.t. $X_{n_{k_\ell}} \xrightarrow{\mathbb{P}} X$ as $\ell \to \infty$, then $X_n \xrightarrow{\mathbb{P}} X$.

Proof. (i) Convergence in probability implies $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. Then $\varepsilon = 1/2^k$ gives $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > 1/2^k) = 0$ means $\exists n_k$, s.t. $n \geq n_k$, $\mathbb{P}(|X_n - X| > 1/2^k) \leq 1/2^k$. Then we need to show $X_{n_k} \xrightarrow{a.s.} X$. It's equivalent to show

$$\lim_{k\to\infty} \mathbb{P}(\cup_{l\geq k}\{|X_{n_l} - X| > \varepsilon\}) = 0, \forall \varepsilon > 0$$

It's equivalent to show $\forall \varepsilon > 0$, $\forall \delta > 0$, $\exists k_0$, s.t. for $k \geq k_0$, $\mathbb{P}(\cup_{l\geq k}\{|X_{n_l} - X| > \varepsilon\} \leq \delta$. We choose k_0 such that

$$\frac{1}{2^{k_0}} < \varepsilon, \quad \frac{1}{2^{k_0-1}} < \delta$$

Therefore

$$\begin{aligned} &\mathbb{P}(\cup_{l\geq k}\{|X_{n_l} - X| > \varepsilon\}) \leq \mathbb{P}(\cup_{l\geq k}\{|X_{n_k} - X| > 1/2^l\}) \\ &\leq \sum_{l=k}^{\infty} \mathbb{P}(\{|X_{n_l} - X| > 1/2^l\}) \leq \frac{1}{2^{k-1}} \leq \frac{1}{2^{k_0-1}} \leq \delta. \end{aligned}$$

so we finish the proof of (i).

(ii) □

Definition 18 (Convergence in L^p)**.** $0 < p < \infty$, $X_n \xrightarrow{L^p} X$ *iff* $\lim_{n\to\infty} \mathbb{E}|X_n - X|^p \to 0$.

Lemma 6.3. If $X_n \xrightarrow{L^p} X$ for $p \in (0, \infty)$, then $X_n \xrightarrow{\mathbb{P}} X$.

Proof. By Chebyshev, $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \to 0$. □

Theorem 6.1. If $X_n \xrightarrow{\mathbb{P}} X$, $\exists Y$, s.t. $|X_n| \leq Y, \forall n$, and $\mathbb{E}(Y^p) < \infty$, then $X_n \xrightarrow{L^p} X$.

Proof. $X_n \xrightarrow{p} X$ implies $\exists$ a subsequence $(n_k)_{k\geq 1}$, $X_{n_k} \xrightarrow{a.s.} X$. So $X_{n_k}(\omega) \to X(\omega)$, note $|X_{n_k}| \leq Y$, therefore $|X| \leq Y$. Consider a decomposition

$$\begin{aligned} \mathbb{E}|X_n - X|^p &= \mathbb{E}(|X_n - X|^p \cdot 1_{|X_n - X| \leq \varepsilon} + |X_n - X|^p \cdot 1_{|X_n - X| > \varepsilon}) \\ &\leq \varepsilon^p + \int_{A_n} |X_n - X|^p \mathrm{d}\mathbb{P} \to \varepsilon^p + 0 \end{aligned}$$

where $A_n = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ and $\mathbb{P}(A_n) \to 0$, note $|X_n - X|^p \leq 2^p Y^p$, thus $\mathbb{E}|X_n - X|^p \leq 2^p \mathbb{E}(Y^p) < \infty$. Letting $\varepsilon \to 0$ gives the result. □

- Examples.

Example 1 ($\xrightarrow{\mathbb{P}} \not\Longrightarrow \xrightarrow{L^p}$). Let the probability triplet be $([0,1), \mathcal{B}[0,1), \lambda)$ where λ is the Lebesgue measure.

Example 2 ($\xrightarrow{L^p} \not\Longrightarrow \xrightarrow{a.s.}$)

7 The Borel-Cantelli lemma.

- Limits of sets.

Let $(E_n)_{n\geq 1}$ be subsets of Ω in $\mathcal{F}$.

$$\limsup E_n = \cap_{n\geq 1} \cup_{m\geq n} E_m \tag{4}$$

$$\liminf E_n = \cup_{n\geq 1} \cap_{m\geq n} E_m \tag{5}$$

Theorem 7.1 (Fatou's Lemma)**.** We have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $A_n \in \mathcal{F}$, $\forall n$

$$\mathbb{P}(\liminf_n A_n) \leq \liminf_n \mathbb{P}(A_n) \leq \limsup_n \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n)$$

Consequently if $A_n \to A$, then $\mathbb{P}(A_n) \to \mathbb{P}(A)$

Proof. We note

$$\mathbb{P}(\liminf_n A_n) = \mathbb{P}(\cup_{n\geq 1} \cap_{m\geq n} A_m) = \lim_{n\to\infty} \mathbb{P}(\cap_{m\geq n} A_m) \leq \liminf_n \mathbb{P}(A_n)$$

where the last equality is due to $\cap_{m\geq n} A_m$ is increasing in n and the last inequality is due to $\mathbb{P}(\cap_{m\geq n} A_m) \leq \mathbb{P}(A_n)$ for all n. And similarly

$$\limsup_n \mathbb{P}(A_n) \leq \lim_n \mathbb{P}(\cup_{m\geq n} A_m) = \mathbb{P}(\cap_{n\geq 1} \cup_{m\geq n} A_m) = \mathbb{P}(\limsup_n A_n)$$

Thus the Fatou's lemma is proved.

Note if $A_n \to A$, then $\liminf_n A_n = \limsup_n A_n$, so by Fatou's lemma, $\mathbb{P}(\liminf_n A_n) = \liminf_n \mathbb{P}(A_n) = \limsup_n \mathbb{P}(A_n) = \mathbb{P}(\limsup_n A_n) = \mathbb{P}(A)$. $\lim_n \mathbb{P}(A_n) = \mathbb{P}(A)$. □

- The main result.

Lemma 7.1 (Borel-Cantelli lemma)**.** (a) If $\sum_n \mathbb{P}(E_n) < \infty$, then $\mathbb{P}(E_n \text{ i.o.}) = 0$.

Or equivalently

$$\mathbb{P}\left[(\limsup E_m)^c\right] = \mathbb{P}\left[\liminf E_m^c\right] = 1$$

(b) If $\sum_n \mathbb{P}(E_n) = \infty$ and E_n's are independent, then $\mathbb{P}(E_n \text{ i.o.}) = 1$.

Proof. □

Remark: (1) One can use the first part, (a) of the Borel-Cantelli lemma to show a sufficient condition for the strong law of large number (SLLN) is $\mathbb{E}(|X_1|^4) < \infty$.

(2) One can use the second claim of the Borel-Cantelli lemma to show the necessary condition of SLLN is $\mathbb{E}|X_1| < \infty$.

(3) This is the first example of 0-1 law.

8 Weak Convergence 1: Helly's Theorem, Tightness, and Prokhorov's Theorem

- The definition of weak convergence.

On Borel-σ-algebra $(\mathbb{R}, \mathcal{B})$ are defined a sequence of probability measures $(\mu_n)_{n\geq 1}$ and a probability measure μ. One may consider defining μ_n "in some sense" converges to μ via sup of "probabilities difference" tending to zero, say if

$$\|\mu_n - \mu\|_{\text{TV}} \equiv \sup_{A\in\mathcal{B}} |\mu_n(A) - \mu(A)| \to 0$$

However, one will soon realize that it is not a good definition. For example, $\mu_n(A) = \delta_{x_n}(A)$ and $\mu(A) = \delta_x(A)$ with $x_n \neq x$, let $x_n \to x$, then it's quite embarrassing to find

$$\sup_{A\in\mathcal{B}} |\mu_n(A) - \mu(A)| = 1$$

One may think the issue can be due to $\mathcal{B}$ which is too large to lower down the "sup", however replacing $\mathcal{B}$ by a much smaller class $\mathcal{I}$ which consists of intervals doesn't solve the definition problem. Therefore, we must abandon "sup" in the definition and consider a very weak version of convergence.

Definition 19 (Weak convergence)**.** A sequence of probability measures[2] μ_n weakly converges to μ, written as $\mu_n \xrightarrow{d} \mu$ or $\mu_n \rightsquigarrow \mu$ if for all intervals of type $(a, b]$, $\mu_n(a, b] \to \mu(a, b]$ and $\mu(\{a\}) = \mu(\{b\}) = 0$.

<u>Remark</u>. In many books or papers, different conventions or notations may be adopted. In summary, "converge in distribution" ($\xrightarrow{d}$), "converge in law" ($\xrightarrow{\mathcal{L}}$), and "weak converge" ($\Rightarrow$ or $\rightsquigarrow$), written for probability (distribution) measures, distribution functions, random variables, all are of the same meaning.

Theorem 8.1. $\mu_n \rightsquigarrow \mu$ *iff* $F_n(x) \to F(x)$ for all continuity points x's of F, where $F_n = \mu_n(-\infty, x]$ is the distribution function of μ_n and so is F for μ.

Theorem 8.2. If $F(x)$ is a distribution function, then the set of discontinuity points of F is at most countable.

[2] For probability measures not defined on $\mathbb{R}$, one can consider probability distribution measures.

Definition 20. We define $X_n \overset{\mathcal{L} \text{ or } d}{\to} X$, if $F_{X_n} \to F_X$. Note X_n and X don't need to be in the same probability space.

- We denote the family of distribution functions[3] by $\mathcal{N}$, and define the family of generalized distribution functions[4] as $\mathcal{M}$.

Theorem 8.3 (Helly's selection theorem)**.** $(F_n)_{n\geq 1}$ is a sequence of distribution functions, then $\exists G \in \mathcal{M}$ and $\exists$ a subsequence $(n_k)_{k\geq 1}$, s.t. $F_{n_k} \overset{w}{\to} G$, i.e. $F_{n_k}(x) \to G(x)$, for all $x \in \{t : G(t-) = G(t)\}$.

Proof. To be complete. □

<u>Remark</u>. We provide an example where G indeed belongs to $\mathcal{M}$ and $\mathcal{M}$ cannot be replaced by $\mathcal{N}$. Note if $F_n(x) = \delta_n(-\infty, x]$, then $F_n \to 0$, and if $F_n = \delta_{-n}(-\infty, x]$, then $F_n \to 1$.

Theorem 8.4. For a sequence of distribution functions $(F_n)_{n\geq 1}$, if its any weakly convergent subsequence $(n_k)_{k\geq 1}$ weakly converges to the same limit G_0, then $F_n \overset{d}{\to} G_0$.

- How to make sure the limit G is indeed a distribution function, not a generalized distribution function? We need the concept of tightness.

Definition 21 (Tightness of a set of probability measures)**.** A set of distribution measures $\{\mu_\alpha, \alpha \in I\}$, i.e. probability measures on $\mathbb{R}$, is tight if given any $\varepsilon > 0$, there exists a interval $[a_\varepsilon, b_\varepsilon]$ such that $\mu_\alpha([a_\varepsilon, b_\varepsilon]) \geq 1 - \varepsilon$ for all $\alpha \in I$.

Theorem 8.5. The sequence $(F_n)_{n\geq 1}$ is tight $\Leftrightarrow$ for any subsequence $(n_k)_{k\geq 1}$ which satisfies $F_{n_k} \overset{d}{\to} G \in \mathcal{M}$, then $G \in \mathcal{N}$.

Proof. □

[3] A generic element $F : \mathbb{R} \to [0, 1]$ of $\mathcal{N}$ satisfies 3 conditions: (1) monotone: $x \leq y \Rightarrow F(x) \leq F(y)$; (2) F is càdlàg; (3) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

[4] A generic element $G : \mathbb{R} \to [0, 1]$ of $\mathcal{M}$ satisfies 3 conditions: (1) & (2) the same as $\mathcal{N}$; (3) $\lim_{x\to\infty} F(x) \leq 1$ and $\lim_{x\to-\infty} F(x) \geq 0$.

In conclusion, if we want a nice statement such as a sequence of distribution functions $(F_n)_{n\geq 1}$ weakly converges to $G \in \mathcal{N}$. We need

(1) to show the "subsequence" distribution function limit is unique.

(2) to show the sequence (F_n) is tight.

- Prokhorov's Theorem.

  Given a metric space $\mathcal{X}$ with Borel-σ-algebra $\mathcal{A}$, say the measurable space $(\mathcal{X}, \mathcal{A})$, we consider probability measures on it. This is related to the Blackwell's idea on statistical experiment and later the Le Cam distance of experiments. This part of notes is based on Chapter 5 of Billingsley (1999), also see Theorem 36 of Pollard (2002).

  Definition 22 (Relatively compact)**.** Let Π be a family of probability measures on $(\mathcal{X}, \mathcal{A})$, then Π is called *relatively compact* if every sequence in Π contains a weakly convergent subsequence. That is, $\forall(\mathbb{P}_n)_{n\geq 1}$ in Π, $\exists(\mathbb{P}_{n_k})_{k\geq 1}$ s.t. $\mathbb{P}_{n_k} \rightsquigarrow \mathbb{Q}$ where $\mathbb{Q}$ is a Probability measure on $(\mathcal{X}, \mathcal{A})$ but not necessarily in Π.

  Remark. When we say the relative compactness of a sequence of probability measures $(\mathbb{P}_n)_{n\geq 1}$, it should be clear that this means every subsequence $(\mathbb{P}_{n_k})_{k\geq 1}$ contains a subsubsequence $(\mathbb{P}_{n_{k(m)}})_{m\geq 1}$, s.t. $\mathbb{P}_{n_{k(m)}} \rightsquigarrow \mathbb{Q}$ for some probability measure $\mathbb{Q}$.

  Definition 23 (Tightness)**.** The family Π is *tight* if given any $\varepsilon > 0$, $\exists$ a compact set K, s.t. $\mathbb{P}(K) > 1 - \varepsilon$ for every $\mathbb{P}$ in Π.

  Theorem 8.6 (Prokhorov's Theorem)**.** We have the following

  (i) If Π is tight, then Π is relatively compact.

  (ii) Suppose $\mathcal{X}$ is separable and complete (a Polish space), then Π is tight *iff* Π is relatively compact.

9 Weak convergence 2.

- There are two ways of defining weak convergence, one is by the pointwise convergence of distribution functions for continuity points, the other is via integral (expectation). Its underlying reason is the Helly-Bray theorem as follows.

Theorem 9.1 (Helly-Bray theorem). Let $(\mu_n)_{n\geq 1}$ and μ be probability measures on $\mathbb{R}$ or probability distribution measures, then $\mu_n \xrightarrow{w} \mu \Leftrightarrow \forall f \in C_b(\mathbb{R}), \int f d\mu_n \to \int f d\mu$, or we say $\mathbb{E}(f) = \mathbb{E}_n(f)$

Note: bounded + continuous function + compact domain $\Rightarrow$ uniformly continuous.

Proof. □

Note. For textbook expositions, see Lemma 2 in Chapter 8 of Chow and Teicher (1997), and Theorem 12.7 in Mörters and Peres (2010).

- One can show the following.

  (1) If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$.

  (2) If $X_n \xrightarrow{d} c$ where c is a constant, then $X_n \xrightarrow{\mathbb{P}} c$.

  Following (3) to (5) are Slutsky's theorem.

  (3) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $X_n + Y_n \xrightarrow{d} X + c$.

  (4) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $X_n Y_n \xrightarrow{d} cX$.

  (5) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$.

- Skorohod Representation Theorem.

  Theorem 9.2 (Skorohod representation theorem). If $X_n \xrightarrow{d} X$, then $\exists (Y_n)_{n\geq 1}$ and Y defined in a new probability space, s.t. $X_n \stackrel{d}{=} Y_n$, $X \stackrel{d}{=} Y$, and $Y_n \xrightarrow{a.s.} Y$.

10 Convergence of elements in function space.

10.1 The space $C[0,1]$.

10.2 The space $D[0,1]$.

11 Characteristic function.

- The content of this bullet point is based on Chapter 8 of Chow and Teicher (1997).

The characteristic function for r.v. X is defined as

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int e^{itx}\mathrm{d}F(x) \tag{6}$$

$$= \mathbb{E}\cos(tX) + i\mathbb{E}\sin(tX) = \int \cos(tx)\mathrm{d}F(x) + i\int \sin(tx)\mathrm{d}F(x) \tag{7}$$

It is called characteristic function because it's uniquely determined by F.

Theorem 11.1 (Lévy inversion formula)**.** If X is a r.v. with c.f. φ, then $\forall -\infty < a < b < \infty$,

$$\lim_{C\to\infty}\frac{1}{2\pi}\int_{-C}^{C}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\mathrm{d}t = \mathbb{P}(a<X<b) + \frac{\mathbb{P}(X=a)+\mathbb{P}(X=b)}{2} \tag{8}$$

Proof. For $C>0$, set

$$\begin{aligned}
I(C) = \frac{1}{2\pi}\int_{-C}^{C}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\mathrm{d}t &= \frac{1}{2\pi}\int_{-C}^{C}\frac{e^{-ita}-e^{-itb}}{it}\mathbb{E}(e^{itX})\mathrm{d}t \\
&= \frac{1}{2\pi}\mathbb{E}\int_{-C}^{C}\frac{e^{it(X-a)}-e^{it(X-b)}}{it}\mathrm{d}t \quad \text{by Fubini's theorem and integrand is bdd } \forall\omega \\
&= \frac{1}{2\pi}\mathbb{E}\int_{-C}^{C}\frac{\cos t(X-a)-\cos t(X-b)}{it} + \frac{\sin t(X-a)-\sin t(X-b)}{t}\mathrm{d}t
\end{aligned}$$

note $\frac{\cos t(X-a)-\cos t(X-b)}{it}$ is a odd function and $\frac{\sin t(X-a)-\sin t(X-b)}{t}$ is an even function, so

$$\begin{aligned}
I(C) &= \frac{1}{\pi}\mathbb{E}\int_0^C\frac{\sin t(X-a)-\sin t(X-b)}{t}\mathrm{d}t \\
&= \frac{1}{\pi}\mathbb{E}\left[\int_0^C\frac{\sin t(X-a)}{(X-a)t}\mathrm{d}(X-a)t - \int_0^C\frac{\sin t(X-b)}{(X-b)t}\mathrm{d}(X-b)t\right] \\
&= \frac{1}{\pi}\mathbb{E}\left[\int_0^{C(X-a)}\frac{\sin t}{t}\mathrm{d}t - \int_0^{C(X-b)}\frac{\sin t}{t}\mathrm{d}t\right] = \mathbb{E}J_C(X)
\end{aligned}$$

where

$$J_C(u) = \frac{1}{\pi}\int_{C(u-b)}^{C(u-a)}\frac{\sin t}{t}\mathrm{d}t \xrightarrow{C\to\infty} \begin{cases} 1 & a<u<b \\ \frac{1}{2} & u=a,b \\ 0 & u<a, u>b \end{cases}$$

for $\int_0^\infty \frac{\sin t}{t}\mathrm{d}t = \frac{\pi}{2}$ and the tail integral is zero, this is known as the Dirichlet integral. Since $|J_C| \le 2$ for all u and C, then by dominated convergence theorem.

$$\lim_{C\to\infty} I(C) = \lim_{C\to\infty}\mathbb{E}J_C(X) = \mathbb{E}\lim_{C\to\infty}J_C(X) = \mathbb{E}1_{\{a<X<b\}} + \frac{1}{2}\mathbb{E}1_{\{X=a \text{ or } b\}}$$

This completes the proof. $\square$

Corollary 11.1. There is a one-to-one correspondence between distribution functions on $\mathbb{R}$ and their characteristic functions.

Proof. $\square$

Corollary 11.2. If $\int_{-\infty}^{\infty} |\varphi(t)| \mathrm{d}t < \infty$, then for $-\infty < a < b < \infty$,

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

Then we list all the properties of characteristic functions.

Proposition 11.1 (Some basic properties)**.** We have the following

(1) $\varphi_X(0) = 1$.

(2) $|\varphi_X(t)| \leq 1$.

(3) $\varphi_X(t)$ is positive semidefinite, meaning that $\forall n \in \mathbb{N}$, $\forall \xi \in \mathbb{C}^n$, we have

$$\sum_{i=1}^{n} \xi_i \varphi_X(t_i - t_j) \bar{\xi}_j \geq 0.$$

where $t_1, \cdots, t_n$ are any real numbers and $\xi_1, \cdots, \xi_n$ are any complex numbers.

(4) $\varphi_X(-t) = \overline{\varphi_X(t)}$.

(5) $\varphi_X(t)$ is uniformly continuous.

(6) If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

(7) If $X \sim \mathcal{N}(0, 1)$, then $\varphi_X(t) = e^{-t^2/2}$.

(8) If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $\varphi_Y(t) = e^{it\mu - \sigma^2 t^2/2}$.

(9) If $X \sim \mathrm{Poisson}(\lambda)$, then $\varphi_X(t) = \exp[\lambda(e^{it} - 1)]$.

(10) If $X \sim \mathrm{Poisson}(a, \lambda)$, that is $\mathbb{P}(X = ak) = \frac{\lambda^k}{k!} e^{-\lambda}$, then $\varphi_X(t) = \exp[\lambda(e^{iat} - 1)]$.

(11) If $X_j \sim \mathrm{Poisson}(a_j, \lambda_j)$ for $1 \leq j \leq l$, then $\exp[\sum_{j=1}^{l} \lambda_j (e^{ia_j t} - 1)]$.

(12) $1 - \mathrm{Re}(\varphi(t)) \leq 4[1 - \mathrm{Re}(\varphi(t/2))]$

Proof. (1) (6) are omitted.

(2) $|\varphi_X(t)| = |\mathbb{E}(e^{itX})| = |\mathbb{E}(\cos(tX) + i\sin(tX))| \leq \mathbb{E}|\cos(tX) + i\sin(tX)| \leq \mathbb{E}1 = 1.$

(3)

(4) $\varphi_X(-t) = \mathbb{E}(e^{-itX}) = \mathbb{E}(\cos(tX)) - i\mathbb{E}(\sin(tX)) = \overline{\mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))} = \overline{\varphi_X(t)}$.

(5) Note

$$|\varphi_X(t+\Delta t) - \varphi_X(t)| = \left|\mathbb{E}(e^{i(t+\Delta t)X} - e^{itX})\right| = \left|\mathbb{E}e^{itX}(e^{i\Delta tX} - 1)\right|$$

(9) $\varphi_X(t) = \mathbb{E}(e^{itX}) = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = e^{\lambda(e^{it}-1)}$.

(10) The same as (9).

(11) $\varphi_{X_1+\cdots+X_l}(t) = \Pi_{j=1}^{l}\exp[\lambda_j(e^{ia_jt} - 1)] = \exp[\sum_{j=1}^{l}\lambda_j(e^{ia_jt} - 1)]$ □

- Main reuslts.

We have

$$F \leftrightarrow \mu \quad \& \quad \mathbb{P} \leftrightarrow \varphi_X.$$

F: distribution function.

μ: distribution measure, pushforward measure

$\mathbb{P}$: probability measure

φ_X: characteristic function.

Theorem 11.2. Suppose μ and ν are two distribution measures, then

$$\varphi_\mu = \varphi_\nu \iff \mu = \nu.$$

Proof. $\Longleftarrow$ part is trivial. What we need to show is $\varphi_\mu = \varphi_\nu \implies \mu = \nu$.

Step 1. Let $C_0(\mathbb{R})$ denote the set of functions that are continuous with compact support. If we can show

$$\int f\mu(\mathrm{d}x) = \int fv(\mathrm{d}x), \quad \forall f \in C_0(\mathbb{R})$$

then it is sufficient to derive $\mu = v$ which is trivial by contradiction. We need trigonometric polynomial for the proof.

Definition 24 (Trigonometric Polynomial)**.** A function $P : \mathbb{R} \to \mathbb{C}$ is called a trigonometric polynomial if $\exists N \geq 1, C_{-N}, \cdots, C_N$, s.t.

$$P(t) = \sum_{j=-N}^{N} C_j e^{itj}$$

Let's recall a theorem from analysis.

Theorem 11.3. Let a function $F : \mathbb{R} \to \mathbb{C}$ be continuous and periodic, then there exists a trigonometric polynomial P, s.t. $\forall \varepsilon > 0$,

$$\sup_{t \in \mathbb{R}} |F(t) - P(t)| \leq \varepsilon$$

□

Theorem 11.4. $\exists 0 < K < \infty$, s.t. $\forall a > 0, \forall \mu$

$$\mu([-\frac{1}{a}, \frac{1}{a}]^c) \leq \frac{K}{a} \int_0^a [1 - \mathrm{Re}(\varphi_\mu(t))] \mathrm{d}t \tag{9}$$

Corollary 11.3. Suppose $(X_n)_{n \geq 1}$ with μ_n and φ_n satisfying

(1) $\exists \delta > 0$, $\varphi_n(t) \to \varphi(t)$ for all $t \in [-\delta, \delta]$.

(2) $\varphi(t)$ is continuous at $t = 0$.

Then the sequence $(X_n)_{n \geq 1}$ is tight.

12 Lévy continuity theorem.

- The main results.

Theorem 12.1 (Glivenko)**.** If φ_n and φ are the characteristic functions of probability distributions $\mathbb{P}_n$ and $\mathbb{P}$ respectively, for each $n \in \mathbb{N}$, then $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}^d \implies \mathbb{P}_n \to \mathbb{P}$ weakly as $n \to \infty$.

Theorem 12.2 (The Lévy continuity theorem)**.** Let $(X_n)_{n\geq 1}$ be a sequence of random variables, if

(1) $\varphi_{X_n}(t) \to \varphi(t)$ for all $t \in \mathbb{R}$.

(2) $\varphi(t)$ is continuous at $t = 0$.

Then X_n converges in distribution to some random variable X whose characteristic function is $\varphi(t)$. That is to say, there exist a probability measure μ on $\mathbb{R}$ s.t. $\mu_n \xrightarrow{w} \mu$ where μ_n is the probability distribution measure of X_n.

Recall the conclusion in " Weak convergence on $\mathbb{R}$, part 1. ". We need to show (μ_n) is tight and the uniqueness of subsequence limit.

Note: if $\mu_n \xrightarrow{w} \mu$, then $\varphi_{\mu_n} \to \varphi_\mu$

A very useful formula is as follows, very useful to show the convergence in distribution via characteristic functions.

Proposition 12.1. A random variable X, suppose $\mathbb{E}(|X|^k) < \infty$ for an integer $k \geq 1$, then

$$\varphi_X(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!}\mathbb{E}[X^j] + \frac{t^k}{k} r_k(t) \tag{10}$$

Then $|r_k(t)| \leq 4\mathbb{E}[|X|^k]$ for all $t \in \mathbb{R}$, and $\lim_{t\to 0} r_k(t) = 0$.

Proof. To be added. □

We can apply the above proposition and Lévy continuity theorem to show a central limit theorem stated as follows.

Corollary 12.1 (CLT)**.** Suppose $(X_i, i = 1, \cdots, n)$ are i.i.d. random variables with $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(X_i^2) < \infty$ and variance σ^2, then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Proof. To be added. □

13 Weak Law of Large Numbers.

- We have already shown via Markov inequality that suppose $(X_i, i = 1, \cdots, n)$ are i.i.d random variables with $\mathbb{E}X_i = \mu$ and $\mathbb{E}|X_i^2| < \infty$, then $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P/L^2} \mu$. In this section we will relax the condition.

Theorem 13.1 (WLLN)**.** $X_1, X_2, \cdots, X_n$ i.i.d. with $\mathbb{E}|X_1| < \infty$ and let $\mathbb{E}(X_1) = \mu$, then

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

Proof 1 (Truncation). □

Proof 2 (Characteristic function). □

- Next let's mention the Kolmogorov inequality, also known as the "maximal inequality".

Theorem 13.2 (Kolmogorov)**.** Suppose $X_1, \cdot, X_n$ are independent wih $\mathbb{E}(X_j) = 0$ and $\mathbb{E}(X_j^2) = \sigma_j^2$, and let $S_n = \sum_{j=1}^n X_j$ denote them partial sum, then for all $\varepsilon > 0$,

$$\mathbb{P}(\max_{1\le k\le n} |S_k| > \varepsilon) < \frac{1}{\varepsilon^2}\sum_{j=1}^n \sigma_j^2$$

Proof. The proof strategy is to translate maximal of sum to hitting times. □

The next inequality is due to Lévy.

Theorem 13.3 (Lévy)**.** Let $X_1, X_2, \cdots, X_n$ be independent random variables, suppose for some $a > 0$, there exists $\delta \in (0, 1)$ such that

$$\mathbb{P}(|X_j + \cdots + X_n| \ge \frac{a}{2}) \le \delta. \qquad \forall 1 \le j \le n$$

then we have

$$\mathbb{P}(\max_{1\le j\le n} |S_j| \ge a) \le \frac{\delta}{1-\delta}$$

Proof strategy: Define $T = \min\{j : S_j \geq a\}$,

$$\begin{aligned}
&\mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq a, |S_n| < a/2) + \mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq a, |S_n| \geq a/2) \\
&\leq \mathbb{P}(T \leq n, |S_n| \leq a/2) + \delta = \sum_{j=1}^{n-1} \mathbb{P}(T = j, |S_n| \leq a/2) + \delta \\
&\leq \sum_{j=1}^{n-1} \mathbb{P}(T = j, |S_n - S_j| \geq a/2) + \delta = \sum_{j=1}^{n-1} \mathbb{P}(T = j)\mathbb{P}(|S_n - S_j| \geq a/2) + \delta \\
&\leq \delta\mathbb{P}(T \leq n-1) + \delta \leq \delta\mathbb{P}(T \leq n) + \delta.
\end{aligned}$$

14 Series of Random Variables.

- The main result.

Theorem 14.1. Let $X_1, X_2, \cdots$, be independent random variables, and $S_n = \sum_{j=1}^n X_j$ whose distribution function is denoted by F_{S_n}, then

(1) $F_{S_n} \xrightarrow{d} F \Leftrightarrow \exists S$, $S_n \xrightarrow{p} S$ and the distribution function of S is F.

(2) $S_n \xrightarrow{p} S \Leftrightarrow S_n \xrightarrow{a.s.} S$.

We need two lemmas to prove the theorem.

Lemma 14.1 (Cauchy in probability $\Leftrightarrow$ convergence in probability)**.** For $(X_n)_{n\geq 1}$,

$$X_m - X_n \xrightarrow{p} 0 \text{ as } m, n \to \infty \Leftrightarrow \exists \text{ r.v. } X, X_n \xrightarrow{p} X. \tag{11}$$

Proof. $\Rightarrow$ (harder): $\forall \varepsilon > 0$, $\lim_{m,n} \mathbb{P}(|X_m - X_n| > \varepsilon) = 0$. Let $\varepsilon = 1/2^k$, so

$$\exists n_k, \text{ s.t. } m, n \geq n_k, \mathbb{P}(|X_m - X_n| > 1/2^k) < 1/2^k$$

we can let $n_k < n_{k+1}$ and $n_k \to \infty$ as k increases, therefore

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 1/2^k) < 1/2^k$$

Thus

$$\mathbb{P}(\cup_{j\geq k}\{|X_{n_{j+1}} - X_{n_j}| > 1/2^j\}) < 1/2^{k-1}$$

note the set $\cup_{j\geq k}\{|X_{n_{j+1}} - X_{n_j}| > 1/2^j\}$ is decreasing as k increases, so we have

$$A^c := \cap_{k=1}^{\infty} \cup_{j\geq k} \{|X_{n_{j+1}} - X_{n_j}| > 1/2^j\}, \quad \mathbb{P}(A^c) = 0, \quad \mathbb{P}(A) = 1$$

so $A = \cup_{k=1}^{\infty} \cap_{j\geq k} \{|X_{n_{j+1}} - X_{n_j}| \leq 1/2^j\}$, for any $\omega \in A$, we have $X_{n_j}(\omega) \to X(\omega)$ $\forall j \geq 1$ where $X(\omega)$ defines the limit in the statement. The reason is because X_{n_j} is Cauchy in j, for integers $p > q \geq 1$,

$$\text{Fix } \omega, \quad |X_{n_p} - X_{n_q}| = |\sum_{j=q}^{p-1}(X_{n_{j+1}} - X_{n_j})| \leq \sum_{j=q}^{p-1}|X_{n_{j+1}} - X_{n_j}| \leq 1/2^q$$

so $|X_{n_p}(\omega) - X_{n_q}(\omega)| \to 0$ if q large, so $X_{n_j}(\omega)$ is Cauchy in j. This means $X_{n_j} \xrightarrow{a.s.} X$ and therefore $X_{n_j} \xrightarrow{p} X$. So $\forall \varepsilon, \delta > 0$, $X_{n_j} \xrightarrow{p} X$ gives

$$\exists j_0 \in N_+, \text{s.t. } j > j_0, \mathbb{P}(|X_{n_j} - X| > \varepsilon/2) < \delta/2$$

Cauchy in probability gives,

$$\exists n_0 \in N_+, \text{s.t. } m, n > n_0, \mathbb{P}(|X_n - X_m| > \varepsilon/2) < \delta/2$$

Thus for $n \geq n_0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(|X_n - X_{n_j}| > \varepsilon/2) + \mathbb{P}(|X_{n_j} - X| > \varepsilon/2) \leq \delta/2 + \delta/2 = \delta$$

if we choose a very large j s.t. $n_j > n_0$ and $j > j_0$.

$\Leftarrow$: $X_n \xrightarrow{p} X$ means $\forall \varepsilon > 0, \delta > 0$, $\exists n_0$, s.t. $n > n_0$, $\mathbb{P}[|X_n - X| > \varepsilon/2] < \delta/2$. Note

$$\mathbb{P}[|X_m - X_n| > \varepsilon] \leq \mathbb{P}[|X_m - X| > \varepsilon/2] + \mathbb{P}[|X_n - X| > \varepsilon/2] \leq \delta$$

and this holds with $m, n > n_0$. So $X_m - X_n \xrightarrow{p} 0$ as $m, n \to \infty$. ☐

Note: one can also show that Cauchy a.s. $\Leftrightarrow$ convergence a.s.

Lemma 14.2. For a sequence of random variables $(X_n)_{n\geq 1}$, if

$$\lim_{m,n\to\infty} \mathbb{P}[\max_{m<k\leq n} |X_k - X_m| > \varepsilon] = 0, \quad \forall \varepsilon > 0$$

then $\exists$ a random variable X, s.t. $X_n \xrightarrow{a.s.} X$.

Proof. A proof to be completed. ☐

Then let's go back to the main result of this section.

Proof of Theorem 14.1. T ☐

15 Kolmogorov's Three Series Theorem

Let $(X_n)_{n\geq 1}$ be independent r.v., let $S_n = \sum_{j=1}^n X_j$.

- Main Results.

  Theorem 15.1 (1-series theorem)**.** Suppose $\mathbb{E}(X_j) = 0, \forall j$, $\sigma_j^2 = \mathbb{E}(X_j^2) < \infty$, and $\sum_{j\geq 1} \sigma_j^2 < \infty$, then S_n converges a.s.

  We provide two proofs.

  Proof 1. □

  Proof 2. □

  Theorem 15.2 (2-series theorem)**.** Let $EX_j = \mu_i$, $\sigma_j^2 = \mathrm{Var}(X_j)$, then if

  (1) $\sum_{j=1}^n \mu_j$ converges;

  (2) $\sum_{j\geq 1} \sigma_j^2 < \infty$.

  then S_n converges a.s.

  Proof. Simple corollary of Theorem 15.1. □

  Theorem 15.3 (3-series theorem)**.** Given a constant $c > 0$, let $Y_j = X_j 1_{\{|X_j|\geq c\}}$ and $EY_j = \mu_i$, $\sigma_j^2 = \mathrm{Var}(Y_j)$, then if

  (a) $\sum_{j=1}^\infty \mathbb{P}(X_j \neq Y_j) < \infty$. (Borel-Cantelli Lemma)

  (b) $\sum_{j=1}^n \mu_j$ converges;

  (c) $\sum_{j\geq 1} \sigma_j^2 < \infty$.

  Then $S_n = \sum_{j=1}^n X_j$ converges a.s.

  Proof. The proof contains the intuition of the theorem. Since by Borel-Cantelli lemma,

  $$\mathbb{P}(X_j \neq Y_j \text{ i.o.}) = 0 \Leftrightarrow \mathbb{P}(\text{ for some } n_0, n \geq n_0, X_n = Y_n) = 1$$

  which implies $\sum_n X_n$ converges *iff* $\sum_n Y_n$ converges a.s., and by the 2-series theorem, $\sum_n Y_n$ converges. Thus S_n converges a.s. □

Applications: suppose $(X_j)_{j\geq 1}$ are independent and $\mathbb{P}(X_j = 1) = \mathbb{P}(X_j = -1) = 1/2$, then we would like to explore if the series $\sum_{j=1}^n \frac{X_j}{j}$ converges. This question is somewhat interesting since we know $\sum_j \frac{1}{j}$ diverges and $\sum_{j=1}^\infty \frac{(-1)^{j+1}}{j} = \log 2$.

Theorem 15.4 (converse of 3-series theorem)**.** Suppose $(X_j)_{j\geq 1}$ independent, $S_n = \sum_{j=1}^n X_j$ a.s. converges, then $\forall A > 0$, define $Y_j = X_j 1_{|X_j|\leq A}$, then the three series satisfy:

(a) $\sum_{j=1}^\infty \mathbb{P}(X_j \neq Y_j) < \infty$.

(b) $\sum_{j=1}^n \mathbb{E}(Y_j)$ converges;

(c) $\sum_{j\geq 1} \mathrm{Var}(Y_j) < \infty$.

Lemma 15.1. Suppose $(Y_j)_{j\geq 1}$ are independent, and

(a) $|Y_j| \leq C$ for some constant C.

(b) $\mathbb{E}(Y_j) = 0$

(c) $\sum_{j=1}^n Y_j$ converges a.s. as $n \to \infty$.

Then $\sum_{j=1}^\infty \mathrm{Var}(Y_j) < \infty$.

Proof. Let $S_n = \sum_{j=1}^n Y_j$. For $\ell > 0$, define

$$A_n^\ell = \{|S_1| \leq \ell, |S_2| \leq \ell, \cdots, |S_n| \leq \ell\}, \quad A_0^\ell = \Omega$$

Step 1: we claim: $\exists \ell > 0$. s.t. $\forall \delta \in (0,1), \forall n$, $\mathbb{P}(A_n^\ell) \geq \delta$.

Proof of the claim: Because A_n^ℓ is decreasing as n increases, we actually need to show $\mathbb{P}(|S_n| \leq \ell, \forall n) \geq \delta$. Note $S_n \overset{a.s.}{\to} S$, so $\forall \varepsilon > 0$, $\exists n_0 > 0$, s.t. $n \geq n_0$, $|S_n - S| \leq \varepsilon$ a.s. Using probability, it implies that

$$\mathbb{P}(\cap_{k\geq 1} \cup_{n\geq 1} \cap_{p\geq n} \{|S_p - S| \leq \frac{1}{k}\}) = 1$$

Thus if we take $k = 1$, we get $\mathbb{P}(\cup_{n\geq 1} \cap_{p\geq n} \{|S_p - S| \leq 1\}) = 1$, since $\cap_{p\geq n}\{|S_p - S| \leq 1\}$ is increasing in n, then

$$\mathbb{P}(\cup_{n\geq 1} \cap_{p\geq n} \{|S_p - S| \leq 1\}) = \lim_{n\to\infty} \mathbb{P}(\cap_{p\geq n}\{|S_p - S| \leq 1\}) = 1$$

so $\forall\delta \in (0, 1/3)$, $\exists n'$, s.t.

$$\mathbb{P}(\cap_{p\geq n'}\{|S_p - S| \leq 1\}) \geq 1-\delta. \tag{12}$$

Note $\lim_{A\to\infty}\mathbb{P}(|S| > A) = 0$, so $\exists A$, s.t. $\mathbb{P}(|S| > A) < \delta$ which means

$$\mathbb{P}(|S| \leq A) \geq 1-\delta \tag{13}$$

Therefore[5] by (12) and (13),

$$\mathbb{P}(\cap_{p\geq n'}\{|S_p| \leq 1 + A\}) \geq \mathbb{P}(|S| \leq A, \cap_{p\geq n'}\{|S_p - S| \leq 1\}) \geq 1 - 2\delta \tag{14}$$

And again, note $\lim_{B\to\infty}\mathbb{P}(|S_j| > B) = 0$ for $1 \leq j \leq n'$. So $\exists B_j > 0$, s.t.

$$\mathbb{P}(|S_j| > B_j) \leq \frac{\delta}{n'}$$

let $B = \max_{1\leq j\leq n'} B_j$, therefore $\mathbb{P}(|S_j| > B) \leq \delta/n'$,

$$\mathbb{P}(\max_{1\leq j\leq n'}|S_j| > B) \leq \sum_{j=1}^{n'}\mathbb{P}(|S_j| > B) \leq \delta \tag{15}$$

Let $K = \max(B, 1 + A)$, by (14) and (15), we have

$$\mathbb{P}(|S_j| \leq K, \forall j \geq 1) \geq 1 - 3\delta$$

The K is the ℓ we want, and $1 - 3\delta$ can be any value in $(0, 1)$.

Step 2: then we proceed to prove the lemma. Let's eliminate the subscript ℓ, let $A_n = A_n^\ell$ if no confusion. Let $S_0 = 0$ and $\sigma_j^2 = \mathrm{Var}(Y_j)$. For $n \geq 1$,

$$\begin{aligned}
\int_{A_{n-1}} S_n^2 \mathrm{d}\mathbb{P} = \int_{A_{n-1}}(S_{n-1}+Y_n)^2\mathrm{d}\mathbb{P} &= \int_{A_{n-1}} S_{n-1}^2\mathrm{d}\mathbb{P} + 2\int 1_{A_{n-1}}S_{n-1}Y_n\mathrm{d}\mathbb{P} + \mathbb{P}(A_{n-1})\sigma_n^2 \\
&= \int_{A_{n-1}} S_{n-1}^2\mathrm{d}\mathbb{P} + 2\mathbb{E}(Y_n)\int 1_{A_{n-1}}S_{n-1}\mathrm{d}\mathbb{P} + \mathbb{P}(A_n)\sigma_n^2 \\
&= \int_{A_{n-1}} S_{n-1}^2\mathrm{d}\mathbb{P} + \mathbb{P}(A_n)\sigma_n^2 \geq \int_{A_{n-1}} S_{n-1}^2\mathrm{d}\mathbb{P} + \delta\sigma_n^2
\end{aligned} \tag{16}$$

Note $A_n \subseteq A_{n-1}$,

$$\int_{A_{n-1}} S_n^2\mathrm{d}\mathbb{P} = \int_{A_{n-1}\setminus A_n}(S_{n-1}+Y_n)^2\mathrm{d}\mathbb{P} + \int_{A_n} S_n^2\mathrm{d}\mathbb{P}$$

[5] The 1st inequality is because if $|S| \leq A, |S_p - S| \leq 1$, then $|S_p| \leq 1 + A$; The 2nd inequality is due to the fact that if $\mathbb{P}(A) \geq 1-\delta$, $\mathbb{P}(B) \geq 1-\delta$, then $1 - \mathbb{P}(A\cap B) = \mathbb{P}(A^c \cup B^c) \leq \mathbb{P}(A^c) + \mathbb{P}(B^c) \leq 2\delta$.

$$\leq (\ell + C)^2\mathbb{P}(A_{n-1} \setminus A_n) + \int_{A_n} S_n^2 \mathrm{d}\mathbb{P} \tag{17}$$

Therefore by (16) and (17),

$$\delta\sigma_n^2 \leq (\ell + C)^2\mathbb{P}(A_{n-1} \setminus A_n) + \int_{A_n} S_n^2 \mathrm{d}\mathbb{P} - \int_{A_{n-1}} S_{n-1}^2 \mathrm{d}\mathbb{P}$$
$$\sum_{n=1}^{p} \delta\sigma_n^2 \leq (\ell + C)^2\mathbb{P}(\Omega \setminus A_n) + \int_{A_p} S_p^2 \mathrm{d}\mathbb{P} \leq (\ell + C)^2 + \ell^2, \forall p$$

Let $p \to \infty$, this proves $\sum_{j=1}^{\infty} \mathrm{Var}(Y_j) < \infty$. □

Proof of Theorem 15.4. We prove in the order, first (a), then (c), finally (b).

(a) Let's fix A in the definition of $Y_j = X_j 1_{|X_j| \leq A}$, $\sum_{j=1}^{n} X_j$ converges, therefore $\exists n_0(\omega)$, s.t. $n \geq n_0(\omega)$, $|X_n(\omega)| \leq A$, which means $X_j(\omega) = Y_j(\omega)$ if $j \geq n_0(\omega)$, so $\omega \notin \{X_j \neq Y_j, \text{i.o.}\}$, therefore $\mathbb{P}\{X_j \neq Y_j, \text{i.o.}\} = 0$. If $\sum_{j=1}^{\infty} \mathbb{P}(X_j \neq Y_j) = \infty$, note $\{X_j \neq Y_j\}$ are independent across j, then by the Borel-Cantelli lemma, $\mathbb{P}\{X_j \neq Y_j, \text{i.o.}\} = 1$, a contradiction. So $\sum_{j=1}^{\infty} \mathbb{P}(X_j \neq Y_j) < \infty$.

(b) Step 1: We observe that if $Z_n \overset{a.s.}{\to} Z$, and $Z_n' \overset{d}{=} Z_n$, then $Z_n' \overset{a.s.}{\to} Z'$ for some r.v. Z'. This is because $Z_n \overset{a.s.}{\to} Z$ *iff* Z_n Cauchy a.s., which means $\forall \varepsilon > 0, \exists n_0$, $m, n \geq n_0$, $|Z_m - Z_n| < \varepsilon$ a.s.

$$\mathbb{P}(\cap_{k\geq 1} \cup_{n \geq 1} \cap_{p,q \geq n} \{|Z_p - Z_q| < \frac{1}{k}\}) = 1 \quad \forall k$$

Therefore we can imply,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\cap_{p,q\geq n}\{|Z_p - Z_q| < \frac{1}{k}\}) = \lim_{n\to\infty} \mathbb{P}(\cup_{l \geq n} \cap_{n \leq p,q \leq l} \{|Z_p - Z_q| < \frac{1}{k}\}) \\
&= \lim_{n\to\infty} \lim_{l\to\infty} \mathbb{P}(\cap_{n\leq p,q\leq l}\{|Z_p - Z_q| < \frac{1}{k}\}) \\
&= \lim_{n\to\infty} \lim_{l\to\infty} \mathbb{P}(\cap_{n\leq p,q\leq l}\{|Z_p' - Z_q'| < \frac{1}{k}\}) \quad \forall k
\end{aligned}$$

Reverse the process, we will obtain that Z_n' Cauchy a.s., so Z_n' converges a.s.

Step 2: we prove that $\sum_j Y_j$ converges a.s.

By (1), $\mathbb{P}(X_j \neq Y_j \text{ i.o.}) = 0$, so let $A = \{\omega : \exists n_0(\omega), \text{ s.t. } n \geq n_0, X_n(\omega) = Y_n(\omega)\}$, $\mathbb{P}(A) = 1$. Thus for $\omega \in A$, $\sum_j X_j(\omega)$ converges *iff* $\sum_j Y_j(\omega)$ converges. Note

$B = \{\omega \in \Omega : \sum_j X_j(\omega) \text{ converges}\}$, $\mathbb{P}(B) = 1$, then $A \cap B \subseteq C = \{\omega \in \Omega : \sum_j Y_j \text{ converges}\}$, and $\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \geq 1 - 0 - 0 = 1$. Thus $\mathbb{P}(C) = 0$.

Step 3: Suppose $Y_j' \overset{d}{=} Y_j$ and $(Y_j')_j$ are independent across j, so $\sum_j Y_j'$ converges a.s. And define $W_j = Y_j - Y_j'$, thus $\mathbb{E}W_j = 0$, $(W_j)_j$ are independent, $|W_j| \leq 2C$, $\sum_j W_j$ converges a.s. By the Lemma 15.1, $\sum_{j=1}^{\infty} \mathrm{Var}(W_j) < \infty$, and $\sum_{j=1}^{\infty} \mathrm{Var}(Y_j) = \frac{1}{2} \sum_{j=1}^{\infty} \mathrm{Var}(W_j) < \infty$.

(b) Let $Z_i = Y_i - \mathbb{E}(Y_i)$, note $|Y_i| \leq C$ so $\mathbb{E}(Y_i)$ is well defined. Note $\sum_j \mathrm{Var}(Z_j) = \sum_j \mathrm{Var}(Y_j) < \infty$ by (c), $\sum_j Z_j$ converges a.s. due to the one-series theorem, note $\sum_j Y_j$ converges a.s. by proof of (c). It can by implied that $\sum_j \mathbb{E}(Y_j)$ converges. $\square$

16 Strong Law of Large Numbers

- Results

Lemma 16.1. Suppose $(x_k)_{k\geq 1}$ a deterministic sequence, $x_k \in \mathbb{R}$. And let $(a_k)_{k\geq 1}$ be a deterministic sequence satisfying $a_1 > 0$, $a_k \leq a_{k+1}$ and $a_k \uparrow \infty$. If $\sum_{k=1}^{n} \frac{x_k}{a_k}$ converges, then $\frac{x_1 + \cdots + x_n}{a_n} \to 0$.

Proof. Let $b_n = \sum_{k=1}^{n} x_k/a_k$, then we have say $b_n \to b$, which means $\exists n_0$ s.t. $\forall n \geq n_0$, $|b_n - b| \leq \varepsilon$. Note

$$a_n(b_n - b_{n-1}) = x_n, \quad b_0 \equiv 0$$

Thus

$$\begin{aligned} \frac{x_1 + \cdots + x_n}{a_n} &= \frac{1}{a_n}\left(\sum_{k=1}^{n} a_k b_k - \sum_{k=1}^{n} a_k b_{k-1}\right) = \frac{1}{a_n}\left(\sum_{k=1}^{n} a_k b_k - \sum_{k=1}^{n-1} a_{k+1} b_k\right) \\ &= \frac{1}{a_n}\left(a_n b_n - \sum_{k=1}^{n-1}(a_{k+1} - a_k) b_k\right) = b_n - \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k \end{aligned}$$

Note

$$\left|\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} b_k - b\right| = \left|\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n}(b_k - b) + \frac{a_1}{a_n} b\right|$$

$$
\begin{aligned}
&\le \left|\sum_{k=1}^{n_0-1} \frac{a_{k+1}-a_k}{a_n}(b_k-b)\right| + \left|\sum_{k=n_0}^{n} \frac{a_{k+1}-a_k}{a_n}(b_k-b)\right| + \frac{a_1}{a_n}|b| \\
&\le \frac{1}{a_n}\left|\sum_{k=1}^{n_0-1}(a_{k+1}-a_k)(b_k-b)\right| + \frac{\varepsilon}{2}(1-\frac{a_{n_0}}{a_n}) + \frac{a_1}{a_n}|b| \\
&= \frac{1}{a_n}\left(|b| - |a_{n_0}|\varepsilon/2 + \sum_{k=1}^{n_0-1}|a_{k+1}-a_k||b_k-b|\right) + \frac{\varepsilon}{2} \\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}
$$

for $n \ge n_2$, where n_1 is defined through

$$
\begin{aligned}
&n_2 = n_0 \vee n_1 \\
&\exists n_1 \text{ s.t. } \forall n \ge n_1, \frac{1}{a_n}\left||b| - |a_{n_0}|\varepsilon/2 + \sum_{k=1}^{n_0-1}|a_{k+1}-a_k||b_k-b|\right| \le \frac{\varepsilon}{2}
\end{aligned}
$$

for $a_n \to \infty$ as $n \to \infty$. □

Theorem 16.1 (Strong LLN)**.** Let $(X_j)_{j\ge1}$ be *i.i.d.* with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}|X_1| < \infty$, then

$$
\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} 0 \tag{18}
$$

Proof. Claim 1: □

17 The Law of Large Numbers, Part 2.

18 The Central Limit Theorem, Part 1.

- Introduction.

We have already shown the simplest version of the CLT by characteristic function method (Lévy's Continuity Theorem): suppose $X_1, \cdots, X_n$ are *i.i.d.* random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(X_1^2) = \sigma^2$, then

$$
\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2).
$$

then we would like to ask " can we relax the *i.i.d.* assumption?" Let's rephrase.

Suppose $S_n^2 \equiv \sum_{j=1}^n \sigma_j^2$, then preassumbly we expect to have

$$\frac{X_1 + \cdots + X_n}{S_n} \xrightarrow{d} \mathcal{N}(0,1) \tag{19}$$

Then what is the requirement, or we say a sufficient condition, such that the central limit theorem (19) holds. The answer is Lindeberg condition stated as follows.

Definition 25 (The Lindeberg condition)**.** Set μ_j to be the distribution probability measure of X_j, $j = 1, \cdots, n$, the Lindeberg condition is

$$\frac{1}{S_n^2} \sum_{j=1}^n \int_{|x| \geq \varepsilon S_n} x^2 \mathrm{d}\mu_j \to 0, \quad \forall \varepsilon > 0. \tag{20}$$

The main result is formalized as follows.

Theorem 18.1 (Lindeberg CLT)**.** Let $(X_j)_{j\geq 1}$ be independent and $\mathbb{E}(X_j) = 0, \forall j$, $\mathbb{E}X_j^2 = \sigma_j^2 < \infty$. Set $S_n^2 = \sum_{j=1}^n \sigma_j^2 \to \infty$. If the Lindeberg condition (20) holds, then

$$\frac{X_1 + \cdots + X_n}{S_n} \xrightarrow{d} \mathcal{N}(0,1)$$

Proof. Let's first show some consequences of the Lindeberg condition (20).

Claim 1. Then Lindeberg condition implies $\max_{1\leq j \leq n} \frac{\sigma_j^2}{S_n^2} \to 0$ as $n \to \infty$.

Note $\forall \varepsilon > 0$, and $\forall 1 \leq j \leq n$,

$$\begin{aligned} \sigma_j^2 = \int x^2 \mathrm{d}\mu_j &= \int_{|x|<\varepsilon S_n} x^2 \mathrm{d}\mu_j + \int_{|x|\geq \varepsilon S_n} x^2 \mathrm{d}\mu_j \\ &\leq \varepsilon^2 S_n^2 + \sum_{j=1}^n \int_{|x|\geq\varepsilon S_n} x^2 \mathrm{d}\mu_j \end{aligned}$$

Thus $\lim_{n\to\infty} \max_j \frac{\sigma_j^2}{S_n^2} \leq \varepsilon^2$. Letting $\varepsilon \to 0$ gives the claim 1.

Claim 2. Let $X_{n,j} = \frac{X_j}{S_n}$, then the Lindeberg condition implies $\sum_{j=1}^n \mathbb{P}(|X_{n,j}| > \varepsilon) \to 0$ as $n \to \infty$.[6]

[6]Note the statement $\frac{X_1+\cdots+X_n}{S_n} \xrightarrow{d} \mathcal{N}(0,1)$ is now actually $\sum_j X_{n,j} \xrightarrow{d} \mathcal{N}(0,1)$, and $\sum_{j=1}^n \mathbb{P}(|X_{n,j}| > \varepsilon) \to 0$ is a very strong result because $\mathbb{P}(|X_{n,j}| > \varepsilon) \to 0$ means $X_{n,j} \xrightarrow{\mathbb{P}} 0$. The claim 2 goes beyond that, which means each summand is really very small.

The proof just applies Chebyshev inequality,

$$\sum_{j=1}^{n} \mathbb{P}(|X_{n,j}| > \varepsilon) = \sum_{j=1}^{n} \int_{|X_j|>\varepsilon S_n} \mathrm{d}\mu_j \leq \sum_{j=1}^{n} \int_{|x|>\varepsilon S_n} \frac{x^2}{\varepsilon^2 S_n^2} \mathrm{d}\mu_j$$
$$= \frac{1}{\varepsilon^2} \frac{1}{S_n^2} \sum_{j=1}^{n} \int_{|x|>\varepsilon S_n} x^2 \mathrm{d}\mu_j \to 0$$

Then let's go back to the the proof of CLT. The strategy is again by characteristic function (the Lévy continuity theorem).

We would like to show for any $|t| \leq T$,

$$\varphi_{\sum_{j=1}^{n} X_{n,j}}(t) \to \varphi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$$

which is equivalent to

$$\log \varphi_{\sum_{j=1}^{n} X_{n,j}}(t) + \frac{t^2}{2} \to 0.$$

Note

$$\log \varphi_{\sum_{j=1}^{n} X_{n,j}}(t) + \frac{t^2}{2} = \sum_{j=1}^{n} \log \varphi_{X_{n,j}}(t) + \frac{t^2}{2} = \sum_{j=1}^{n} \log[1 + (\varphi_{X_{n,j}}(t) - 1)] + \frac{t^2}{2}$$
$$\leq \sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1| + \frac{t^2}{2} + c_0 \sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1|^2 \tag{21}$$

where the last inequality is due to $|\log(1+z) - z| \leq c_0|z|^2$. To prove (21), we first show that for any $|t| \leq T$,

$$\sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1|^2 \to 0 \tag{22}$$

Note

$$\sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1|^2 \leq \max_j |\varphi_{X_{n,j}}(t) - 1| \cdot \sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1|$$

If we have

$$\lim_{n\to\infty} \sup_{|t|\leq T} \max_{1\leq j\leq n} |\varphi_{X_{n,j}}(t) - 1| \to 0 \tag{23}$$
$$\sup_{|t|\leq T} \sum_{j=1}^{n} |\varphi_{X_{n,j}}(t) - 1| \leq C_T \tag{24}$$

then (22) is proved. Note

$$\begin{aligned}
\sup_{|t|\le T}\max_j |\varphi_{X_{n,j}}(t)-1| &= \sup_{|t|\le T}\max_j \left|\int (e^{itx/S_n}-1-\frac{itx}{S_n})\mathrm{d}\mu_j\right| \\
&\le \sup_{|t|\le T}\max_j \int \left|e^{itx/S_n}-1-\frac{itx}{S_n}\right|\mathrm{d}\mu_j \\
&\le \sup_{|t|\le T}\max_j \int c_0\left|\frac{tx}{S_n}\right|^2\mathrm{d}\mu_j \\
&\le c_0T^2\max_j\frac{\sigma_j^2}{S_n^2}
\end{aligned}$$

where the second inequality is due to $|e^{iz}-1-iz|\le c_0|z|^2$, so (23) is shown by the Claim 1. And similarly,

$$\begin{aligned}
\sup_{|t|\le T}\sum_{j=1}^n |\varphi_{X_{n,j}}(t)-1| &= \sup_{|t|\le T}\sum_{j=1}^n \left|\int (e^{itx/S_n}-1-\frac{itx}{S_n})\mathrm{d}\mu_j\right| \\
&\le \sup_{|t|\le T}\sum_{j=1}^n \int \left|e^{itx/S_n}-1-\frac{itx}{S_n}\right|\mathrm{d}\mu_j \\
&\le \sup_{|t|\le T}\sum_{j=1}^n \int c_0\left|\frac{t^2x^2}{S_n^2}\right|\mathrm{d}\mu_j \\
&\le c_0T^2 = C_T
\end{aligned}$$

Then we show in (21),

$$\sum_{j=1}^n |\varphi_{X_{n,j}}(t)-1|+\frac{t^2}{2}\to 0 \tag{25}$$

Note

$$\sum_{j=1}^n |\varphi_{X_{n,j}}(t)-1|+\frac{t^2}{2} = \sum_{j=1}^n\left|\int (e^{itx/S_n}-1-\frac{itx}{S_n}+\frac{t^2x^2}{2S_n^2})\mathrm{d}\mu_j-\frac{t^2}{2}\right|+\frac{t^2}{2}$$

□

<u>Remark 1</u>. (Lyapounov theorem) Let's assume all the previous setup holds, if $\exists\delta>0$ s.t.

$$\lim_{n\to\infty}\frac{1}{S_n^{2+\delta}}\sum_{j=1}^n\int |x|^{2+\delta}\mathrm{d}\mu_j = 0$$

then the Lindeberg condition holds.

19 Central Limit Theorem, Part 2.

20 Infinitely divisible law.

- General ideas.

Consider an array of independent random variables $(X_{n,j})_{1\le j\le k_n}$ with $k_n \to \infty$. The question we would like to ask is

$$S_n = \sum_{j=1}^{k_n} X_{n,j} \qquad S_n - A_n \xrightarrow{d} \text{what limit?}$$

for some deterministic sequence of real numbers $(A_n)_{n\ge 1}$.

It turns out the concept of being "uniformly negligible" is very important, which is illustrated as follows.

Definition 26 (uniformaly negligible)**.** The array $(X_{n,j})_{1\le j\le k_n}$ is uniformly negligible if

$$\max_{1\le j\le k_n} \mathbb{P}[|X_{n,j}| > \varepsilon] \to 0 \quad \forall \varepsilon > 0$$

- The Poisson limit.

Let $k_n = n$, and $X_{n,1}, \cdots, X_{n,n}$ are *i.i.d.*, and $\mathbb{P}(X_{n,j} = 1) = p_n, \mathbb{P}(X_{n,j} = 0) = 1 - p_n$ with $p_n \to 0$. Assume that $(X_{n,k_n=n})$ is uniformly negligible, which means $p_n \to 0$. Then we would like to know what conditions are required to make sure $S_n = \sum_{j=1}^{n} X_{n,j}$ converge in distribution.

Observation: $\mathbb{P}(S_n \in \mathbb{N}_0) = 1 \iff \mathbb{P}(S_n = k) \to q(k)$ for some function $q(\cdot)$ and $\forall k$.

$$\mathbb{P}(S_n = 0) = (1 - p_n)^n = (1 - \frac{np_n}{n})^n \overset{a_n = np_n}{=} (1 - \frac{a_n}{n})^n \to e^{-\lambda} \iff a_n \to \lambda \in [0, \infty).$$

(The above $\iff$ should be proved). Then one condition should be $a_n \to \lambda$. Fortunately, this is all we need.

Proposition 20.1. S_n converges in distribution $\iff np_n \to \lambda, \lambda \in [0, \infty)$. And in this case, the limit $S \sim \text{Poisson}(\lambda)$.

Proof. Recall if $X \sim \text{Poisson}(\lambda)$, then $\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \forall k \in \mathbb{N}_0$ for $\lambda > 0$. If $\lambda = 0$, then $\mathbb{P}(X = 0) = 1$.

$(\Longleftarrow)$: $np_n \to \lambda$. For $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(S_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} &= \frac{1}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}(np_n)^k(1-p_n)^{-k}(1-\frac{np_n}{n})^n \\ &\to \frac{\lambda^k}{k!}e^{-\lambda}.\end{aligned}$$

For $\lambda = 0$, $\mathbb{P}(S_n = 0) = (1-p_n)^n = (1-\frac{np_n}{n})^n \to e^{-\lambda} = 1$.

$(\Longrightarrow)$: S_n converges in distribution $\implies \mathbb{P}(S_n = k) = q(k)$ for all k.

So $\mathbb{P}(S_n = 0) = (1-p_n)^n$ converges, so $np_n \to \lambda \in [0, \infty)$. □

Remark 2. We have an example of limiting distribution sum of *i.i.d.* r.v.'s,

$$\sum_{j=1}^n X_{n,j} \xrightarrow{d} \text{Poisson}$$

The limit is not a Gaussian but a Poisson. This raises several questions.

(1) For an array $(X_{n,k_n})_{n\geq 1}$ of *i.i.d.* random variables, what are possible limits of $\sum_{j=1}^{k_n} X_{n,j}$?

(2) Can relax the condition *i.i.d.* to just independent?

- Infinitely divisible law.

Definition 27 (Infinitely divisible law)**.** X has an infinitely indivisible law if $\forall k \geq 2$, $\exists y_1^{(k)}, y_2^{(k)}, \cdots, y_k^{(k)}$ *i.i.d.*, s.t. $X \overset{d}{=} y_1^{(k)} + y_2^{(k)} + \cdots + y_k^{(k)}$

Proposition 20.2. If $X_{n,1}, \cdots, X_{n,n}$ are *i.i.d.*, let $S_n = \sum_{j=1}^n X_{n,j}$, and if $S_n \xrightarrow{d} S$, then S has infinitely divisible law.

Proof. Note $S_{2n} \xrightarrow{d} S$, so

$$\begin{aligned}S_{2n} &= (X_{2n,1} + \cdots + X_{2n,n}) + (X_{2n,n+1} + \cdots + X_{2n,2n}) \\ &= Y_n^{(1)} + Y_n^{(2)}\end{aligned}$$

We claim that $Y_n^{(1)}$ is tight for the following reason. For $a > 0$,

$$[\mathbb{P}(Y_n^{(1)} \geq a)]^2 = \mathbb{P}(Y_n^{(1)} \geq a) \cdot \mathbb{P}(Y_n^{(2)} \geq a) = \mathbb{P}(Y_n^{(1)} \geq a, Y_n^{(2)} \geq a) \leq \mathbb{P}(S_{2n} > a)$$

and S_{2n} should be tight according to Theorem 8.5. So $\exists a_\varepsilon$ s.t. $\limsup_n \mathbb{P}(S_{2n} > a) \leq \varepsilon^2$. Therefore $\limsup_n \mathbb{P}(Y_n^{(1)} > a_\varepsilon) \leq \varepsilon$. The same argument applies for $\limsup_n \mathbb{P}(Y_n^{(1)} \leq -a_\varepsilon) \leq \varepsilon$. So $\exists a'_\varepsilon$, s.t. $\limsup_n \mathbb{P}(|Y_n^{(1)}| \geq a'_\varepsilon) \leq 2\varepsilon$. Therefore we know both $Y_n^{(1)}$ and $Y_n^{(2)}$ are tight.

Therefore by Theorem 8.3 and Theorem 8.5, $\exists$ subsequence n', s.t. $Y_{n'}^{(1)} \xrightarrow{\mathcal{L}} Y^{(1)}$ and a subsubsequence n'' based on n' s.t. $Y_{n''}^{(2)} \xrightarrow{d} Y^{(2)}$, and of course $Y_{n''}^{(1)} \xrightarrow{d} Y^{(1)}$. Clearly, $Y^{(1)} \overset{d}{=} Y^{(2)}$ and $Y_{n''}^{(1)} + Y_{n''}^{(2)} \xrightarrow{d} S$. So $S \overset{d}{=} Y^{(1)} + Y^{(2)}$ and $Y^{(1)}$, $Y^{(2)}$ can be chosen to be independent. $Y_{n''}^{(1)}$ and $Y_{n''}^{(2)}$ are independent. For example, let $Y^{(2)}$ be a random sample of size 1 from distribution of $Y^{(1)}$.

The above argument can be extended from 2 to n. $\square$

Proposition 20.3. X has an infinitely divisible law $\iff$ $\exists$ characteristic function φ_k, s.t. $\varphi_X(t) = [\varphi_k(t)]^k, \forall t \in \mathbb{R}, \forall k \geq 1$.

- Examples.

Example 20.1. (a) $\mathbb{P}(X = b) = 1$, then X has an infinitely divisible law, $X_k \overset{a.s.}{=} b/k$.

(b) $X \sim \mathcal{N}(0, \sigma^2)$, then X has an infinitely divisible law. $X_k \sim \mathcal{N}(0, \sigma^2/k)$.

(c) $X \sim \text{Poisson}(\lambda)$, then X has an infinitely divisible law. $X_k \sim \text{Poisson}(\lambda/k)$.

(d) Let $X_1, X_2, \cdots$ *i.i.d.* from X, and an independent $N \sim \text{Poisson}(\lambda)$. Define $Z = \sum_{i=1}^N X_i$, then Z has an infinitely divisible law.

$$\begin{aligned}
\mathbb{E}(e^{itZ}) &= \sum_{n=0}^{\infty} \mathbb{E}(e^{it(X_1+\cdots+X_n)}|N = n)\mathbb{P}(N = n) = \sum_{n=0}^{\infty} \mathbb{E}(e^{it(X_1+\cdots+X_n)})\mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} [\varphi_X(t)]^n \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=0}^{\infty} \frac{[\varphi_X(t)\lambda]^n}{n!} e^{-\lambda} = e^{-\lambda} e^{\lambda \varphi_X(t)} \\
&= e^{\lambda[\varphi_X(t)-1]}
\end{aligned}$$

So $\varphi_k(t) = e^{\lambda/k[\varphi_X(t)-1]}$, and $Z_k = \sum_{i=1}^{N'} X_i$ where $N' \sim \text{Poisson}(\lambda/k)$ and N' is independent of all other mentioned random variables.

21 Accompanying laws

Let $(X_{n,j})_{1\le j\le k_n}$, $k_n \to \infty$, be an array of *independent* and *uniformly negligible* random variables. The question in mind is: what it the sufficient conditions ensuring that $\exists$ a deterministic sequence of real numbers $(A_n)_{n\ge 1}$, s.t. $\sum_{j=1}^{k_n} X_{n,j} - A_n \xrightarrow{d} S$, for some limit S.

Let's review something and define some notations. Denote the "Poisson sum" by $X^{[\lambda]} = \sum_{i=1}^{N} X_i$, where $X_1, X_2, \cdots$ are *i.i.d.* from X, $N \sim \text{Poisson}(\lambda)$ and N is independent of X_i's. Then $X^{[\lambda]}$ has an infinitely divisible law and $\varphi_{X^{[\lambda]}}(t) = \exp(\lambda[\varphi_X(t) - 1])$.

- Before we start, let's mention two remarks.

  Remark 1. *Uniform negligibility* doesn't require the existence of the first moment.

  Remark 2. A frequently used result is as follows.

$$\lim_{n\to\infty} \max_{1\le j\le k_n} \mathbb{P}(|X_{n,j}| > \varepsilon) = 0 \iff \lim_{n\to\infty} \sup_{|t|\le T} \max_{1\le j\le k_n} |\varphi_{X_{n,j}} - 1| = 0, \forall T > 0.$$

- The main theorem.

  Theorem 21.1 (The Accompanying Law). Let $(X_{n,j})_{1\le j\le k_n}$ be an array of *independent* and *uniformly negligible* (abbreviated as U.N. hereafter) random variables. Define

$$\begin{aligned}
a_{n,j} &= \mathbb{E}\left(X_{n,j}1_{\{|X_{n,j}|\le 1\}}\right) = \int_{|x_{n,j}|\le 1} x_{n,j}\,\mathrm{d}\mu_{n,j} \\
\tilde{X}_{n,j} &= X_{n,j} - a_{n,j} \\
\tilde{Y}_{n,j} &= \tilde{X}_{n,j}^{[1]} = (X_{n,j} - a_{n,j})^{[1]} \\
Y_{n,j} &= \tilde{Y}_{n,j} + a_{n,j}
\end{aligned} \tag{26}$$

  where $\mu_{n,j}$ is the distribution measure of the random variable $X_{n,j}$, then

$$\exists (A_n)_{n\ge 1}, \text{s.t. } \sum_{j=1}^{k_n} X_{n,j} - A_n \xrightarrow{d} S. \iff \exists (A_n)_{n\ge 1}, \text{s.t. } \sum_{j=1}^{k_n} Y_{n,j} - A_n \xrightarrow{d} S.$$

  And the above A_n in both sides are the same.

  The proof needs lemmas.

  Lemma 21.1. $(X_{n,j})_{1\le j\le k_n}$ is independent and U.N. $\implies \lim_{n\to\infty} \max_{1\le j\le k_n} |a_{n,j}| = 0$.

Proof. For $\delta > 0$, which should be small for the proof, we have

$$
\begin{aligned}
|a_{n,j}| = |\int_{|x_{n,j}|\le 1} x_{n,j}\mathrm{d}\mu_{n,j}| &\le \int_{|x_{n,j}|\le 1, |x_{n,j}|>\delta} |x_{n,j}|\mathrm{d}\mu_{n,j} + \int_{|x_{n,j}|\le\delta} |x_{n,j}|\mathrm{d}\mu_{n,j} \\
&\le \int_{|x_{n,j}|>\delta} \mathrm{d}\mu_{n,j} + \delta \le \mathbb{E}(|X_{n,j}| > \delta) + \delta
\end{aligned}
$$

so $\max_{1\le j\le k_n} |a_{n,j}| \le \max_{1\le j\le k_n} \mathbb{E}(|X_{n,j}| > \delta)$ and taking the limit on n and letting $\delta \to 0$ gives the result. □

Corollary 21.1. $(X_{n,j})_{1\le j\le k_n}$ is independent and U.N. $\implies$ $(\tilde{X}_{n,j})_{1\le j\le k_n}$ is independent and U.N.

Proof. By the above lemma, $\forall\varepsilon > 0$, $\exists n_0$, s.t. $n > n_0$, $\max_{1\le j\le k_n} |a_{n,j}| \le \varepsilon/2$. So for $n > n_0$,

$$
\begin{aligned}
\mathbb{P}(|\tilde{X}_{n,j}| > \varepsilon) = \mathbb{P}(|X_{n,j} - a_{n,j}| > \varepsilon) &\le \mathbb{P}(|X_{n,j}| > \varepsilon - |a_{n,j}|) \\
&\le \mathbb{P}(|X_{n,j}| > \varepsilon/2)
\end{aligned}
$$

Taking both sides $\lim_{n\to\infty} \max_{1\le j\le k_n}$ gives the result. □

Lemma 21.2. Suppose $(X_{n,j})_{1\le j\le k_n}$ independent and U.N., $|a_{n,j}| \le \frac{1}{4}$. Let $\tilde{a}_{n,j} := \mathbb{E}\tilde{X}_{n,j}1_{|\tilde{X}_{n,j}|\le 1} = \int_{|x|\le 1} x\mathrm{d}\mu_{\tilde{X}_{n,j}}$ where $\mu_{\tilde{X}_{n,j}}$ is the distribution measure of $\tilde{X}_{n,j}$. Then $\exists c_0 < \infty$, s.t. $|\tilde{a}_{n,j}| \le c_0\mathbb{P}(|\tilde{X}_{n,j}| \ge \frac{1}{2})$.

Comment: note $a_{n,j} \to 0$ and $X_{n,j} \xrightarrow{\mathbb{P}} 0$, so the lemma is the reason why we introduce $\tilde{X}$.

Proof.

$$
\begin{aligned}
|\mathbb{E}(\tilde{X}_{n,j}1_{|\tilde{X}_{n,j}|\le 1})| &= |\mathbb{E}(X_{n,j}1_{|\tilde{X}_{n,j}|\le 1}) - a_{n,j}\mathbb{P}(|\tilde{X}_{n,j}| \le 1)| \\
&= |\mathbb{E}(X_{n,j}1_{|\tilde{X}_{n,j}|\le 1}) - a_{n,j} + a_{n,j}\mathbb{P}(|\tilde{X}_{n,j}| > 1))| \\
&\le |\mathbb{E}(X_{n,j}1_{|\tilde{X}_{n,j}|\le 1}) - a_{n,j}| + \mathbb{P}(|\tilde{X}_{n,j}| \ge \frac{1}{2})
\end{aligned}
$$

in which

$$
|\mathbb{E}(X_{n,j}1_{|\tilde{X}_{n,j}|\le 1}) - a_{n,j}| = |\mathbb{E}[X_{n,j}(1_{|\tilde{X}_{n,j}|\le 1} - 1_{|X_{n,j}|\le 1})]|
$$

$$= |\mathbb{E}[X_{n,j}(1_{|\tilde{X}_{n,j}|\le 1, |X_{n,j}|>1} - 1_{|X_{n,j}|\le 1, |\tilde{X}_{n,j}|>1})]|$$
$$\le \mathbb{E}|X_{n,j}(1_{|\tilde{X}_{n,j}|\le 1, |X_{n,j}|>1})| + \mathbb{P}(|\tilde{X}_{n,j}| \ge \frac{1}{2})$$

in which

$$\begin{aligned}\mathbb{E}|X_{n,j}(1_{|\tilde{X}_{n,j}|\le 1, |X_{n,j}|>1})| &= \mathbb{E}|(\tilde{X}_{n,j} + a_{n,j})(1_{|\tilde{X}_{n,j}|\le 1, |X_{n,j}|>1})| \\ &\le \frac{5}{4}\mathbb{P}(|X_{n,j}| > 1) \le \frac{5}{4}\mathbb{P}(|\tilde{X}_{n,j}| > 1 - |a_{n,j}|) \\ &\le \frac{5}{4}\mathbb{P}(|\tilde{X}_{n,j}| > \frac{3}{4}) \le \frac{5}{4}\mathbb{P}(|\tilde{X}_{n,j}| \ge \frac{1}{2})\end{aligned}$$

Combine all the inequalities together, we obtain the wanted result. □

Lemma 21.3. Let $B_n = \sum_{j=1}^{k_n} a_{n,j}$

(a) $\sum_{j=1}^{k_n} X_{n,j} - A_n \xrightarrow{d} S \iff \sum_{j=1}^{k_n} \tilde{X}_{n,j} + B_n - A_n \xrightarrow{d} S$.

(b) $\sum_{j=1}^{k_n} Y_{n,j} - A_n \xrightarrow{d} S \iff \sum_{j=1}^{k_n} \tilde{Y}_{n,j} + B_n - A_n \xrightarrow{d} S$.

By this lemma, to prove Theorem 21.1 is equivalent to prove the tilde version of it.

Now we turn back to the proof of the accompanying law.

Proof of the tilde version of Theorem 21.1. We show the statement in the following steps.

[Step 0] If $\varphi_{\sum_j \tilde{X}_{n,j} - A_n}(t) - \varphi_{\sum_j \tilde{Y}_{n,j} - A_n}(t) \to 0, \forall t, \forall A_n$, then the theorem is proved.

Reasoning: if $\sum_{j=1}^{k_n} \tilde{X}_{n,j} - A_n \xrightarrow{d} S$, then $\varphi_{\sum_j \tilde{X}_{n,j} - A_n}(t) \to \varphi_S(t)$, thus $\varphi_{\sum_j \tilde{Y}_{n,j} - A_n}(t) \to \varphi_S(t)$, by Lévy continuity theorem, $\sum_j \tilde{Y}_{n,j} - A_n \xrightarrow{d} S$.

[Step 1] To show Step 0, it's sufficient to show $\forall T > 0, \exists c_T < \infty$, s.t.

$$\sum_{j=1}^{k_n} |\varphi_{\tilde{X}_{n,j}}(t) - 1| \le c_T, \quad \forall |t| \le T \tag{27}$$

Reasoning: recall Example 20.1, $\forall T > 0$, and $\forall |t| \le T$.

$$\log \varphi_{\sum_j \tilde{X}_{n,j} - A_n}(t) - \log \varphi_{\sum_j \tilde{Y}_{n,j} - A_n}(t) = \log \varphi_{\sum_j \tilde{X}_{n,j}}(t) - itA_n - \log \varphi_{\sum_j \tilde{Y}_{n,j}}(t) + itA_n$$
$$= \sum_{j=1}^{k_n} (\log \varphi_{\tilde{X}_{n,j}} - [\varphi_{\tilde{X}_{n,j}} - 1]) \le \sum_{j=1}^{k_n} |\log(1 + \varphi_{\tilde{X}_{n,j}} - 1) - (\varphi_{\tilde{X}_{n,j}} - 1)|$$

$$\leq \sum_{j=1}^{k_n} |\varphi_{\tilde{X}_{n,j}} - 1|^2 \leq \max_{1\leq j\leq k_n} |\varphi_{\tilde{X}_{n,j}}(t) - 1| \cdot \sum_{j=1}^{k_n} |\varphi_{\tilde{X}_{n,j}}(t) - 1|$$

note $\tilde{X}_{n,j}$ is U.N., so $\sup_{|t|\leq T} \max_{1\leq j\leq k_n} |\varphi_{\tilde{X}_{n,j}}(t) - 1| \to 0$. So the object above is converging to zero by the condition (27).

[Step 2] To show Step 1, it's sufficient to show

(a) $\exists c_0 < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\leq 1}) \leq c_0$, $\forall n \geq 1$.

(b) $\forall \delta > 0$, $\exists c_\delta < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta) \leq c_\delta$, $\forall n \geq 1$.

Reasoning: Given $T > 0$,

$$\begin{aligned}
\sum_{j=1}^{k_n} |\mathbb{E}(e^{it\tilde{X}_{n,j}} - 1)| &= \sum_{j=1}^{k_n} |\mathbb{E}(e^{it\tilde{X}_{n,j}} - 1)(1_{|\tilde{X}_{n,j}|\leq 1} + 1_{|\tilde{X}_{n,j}|>1})| \\
&\leq \sum_{j=1}^{k_n} |\mathbb{E}(e^{it\tilde{X}_{n,j}} - 1)1_{|\tilde{X}_{n,j}|\leq 1}| + 2\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| > 1)
\end{aligned}$$

in which

$$\begin{aligned}
\sum_{j=1}^{k_n} |\mathbb{E}(e^{it\tilde{X}_{n,j}} - 1)1_{|\tilde{X}_{n,j}|\leq 1}| &\leq \sum_{j=1}^{k_n} |\mathbb{E}(e^{it\tilde{X}_{n,j}} - it\tilde{X}_{n,j} - 1)1_{|\tilde{X}_{n,j}|\leq 1}| + \sum_{j=1}^{k_n} |\mathbb{E}it\tilde{X}_{n,j}1_{|\tilde{X}_{n,j}|\leq 1}| \\
&\leq \sum_{j=1}^{k_n} |\mathbb{E}(c|it\tilde{X}_{n,j}|^2 1_{|\tilde{X}_{n,j}|\leq 1})| + T\sum_{j=1}^{k_n} |\mathbb{E}\tilde{X}_{n,j}1_{|\tilde{X}_{n,j}|\leq 1}| \\
&\leq cT^2 \sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\leq 1}) + T\sum_{j=1}^{k_n} |\tilde{a}_{n,j}| \\
&\leq c_0 cT^2 + c'T\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \frac{1}{2}) \leq c_0 cT^2 + c' c_{1/2} T
\end{aligned}$$

by the above conditions and Lemma 21.2 (where the condition $|\tilde{a}_{n,j}| \leq \frac{1}{4}$ is automatically satisfied if n large due to Lemma 21.1). So we obtain the condition in Step 1.

[Step 3] We show that $\tilde{Y}_{n,j} - A_n \xrightarrow{d} S$ or $\tilde{X}_{n,j} - A_n \xrightarrow{d} S$ implies the conditions (a) and (b) in Step 2.

$*$ Here we first show $\tilde{Y}_{n,j} - A_n \xrightarrow{d} S \implies$ conditions (a) and (b).

(Step 1-$\tilde{Y}$) We show that $\forall T > 0$, $\exists c_T$, s.t.

$$\sum_{j=1}^{k_n} \mathbb{E}[(1-\cos(t\tilde{X}_{n,j}))] \le c_T, \quad \forall |t| \le T \tag{28}$$

Note $1-\cos(2x) \le 4(1-\cos x)$, therefore we can extend to arbitrary T.

(Step 2-$\tilde{Y}$) Let's prove.

One question: how do we know the characteristic function is continuous?

$**$ Then we show $\tilde{X}_{n,j} - A_n \xrightarrow{d} S \implies$ conditions (a) and (b).

Remark 1. If $X \stackrel{d}{=} X'$, X and X' are independent, then $\varphi_{X-X'}(t) = |\varphi_X(t)|^2$ and consequently it's real-valued.

Reasoning: $\varphi_{X-X'}(t) = \mathbb{E}e^{it(X-X')} = \varphi_X(t)\mathbb{E}(\overline{e^{itX}}) = \varphi_X(t)\overline{\varphi_X(t)} = |\varphi_X(t)|^2$.

Remark 2. In this remark, we adopt the same notation in the remark 1. If $Z_n \xrightarrow{d} S$, then $Z_n - Z_n' \xrightarrow{d} S - S'$.

Reasoning: $\varphi_{Z_n - Z_n'}(t) = |\varphi_{Z_n}(t)|^2 \to |\varphi_S(t)|^2 = \varphi_{S-S'}(t)$. Then by Lévy continuity theorem.

Remark 3. Note $\sum_j \tilde{X}_{n,j} - A_n \xrightarrow{d} S$, thus $(\sum_j \tilde{X}_{n,j} - A_n) - (\sum_j \tilde{X}_{n,j} - A_n)' \xrightarrow{d} S - S'$

Define $U_{n,j} = \tilde{X}_{n,j} - \tilde{X}_{n,j}'$. Therefore

$$\sum_{j=1}^{k_n} U_{n,j} \xrightarrow{d} S - S'$$

and $(U_{n,j})_{1\le j\le k_n}$ are independent.

Proof strategy: we show the (a) and (b) conditions in step 2 hold for $U_{n,j}$ and then show that they hold for $\tilde{X}_{n,j}$

Note it's sufficient to show:

(a) $\exists c_0 < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) \le c_0$, $\forall n \ge n_0$.

(b) $\forall \delta > 0$, $\exists c_\delta < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \ge \delta) \le c_\delta$, $\forall n \ge n_0$.

[Step X-1] Note $\sum_{j=1}^{k_n} U_{n,j} \xrightarrow{d} S - S'$, so

$$\varphi_{\sum_j U_{n,j}}(t) \to \varphi_{S-S'}(t), \text{ uniformly compact} \tag{29}$$

note $\varphi_{S-S'}(t) = 0$, so $\exists t_0 > 0$, s.t. $\varphi_{S-S'}(t) \geq 1/2$ for all $|t| \leq t_0$. (note $\varphi_{S-S'}(t)$ is real-valued, so the inequality is well defined)

Accordingly, the above and the uniform convergence (29) imply that $\exists n_0 \geq 1$, s.t. $\forall n \geq n_0$, $\varphi_{\sum_j U_{n,j}}(t) \geq \frac{1}{4}$ for all $|t| \leq t_0$. So

$$\Pi_{j=1}^{k_n} \varphi_{U_{n,j}}(t) \geq \frac{1}{4} \implies \Pi_{j=1}^{k_n} e^{\varphi_{U_{n,j}}(t)-1} \geq \frac{1}{4}$$

By employing $y \leq e^{y-1}$ for $y \leq 1$ and the fact that $\varphi_{U_{n,j}}(t) \leq 1$. So taking log gives

$$\sum_{j=1}^{k_n} 1 - \varphi_{U_{n,j}}(t) \leq \log 4$$

And $\varphi_{U_{n,j}}(t)$ is real-valued, thus $\exists c_0$

$$\sum_{j=1}^{k_n} \mathbb{E}\left(1 - \cos(tU_{n,j})\right) \leq c_0, \quad \forall n \geq n_0, \forall |t| \leq t_0$$

since n_0 is a finite number, the conclusion can be extended to $n \geq 1$. Moreover, by applying the inequality $1 - \cos 2x \leq 4(1 - \cos x)$, t_0 can be extended to any positive real number T. Finally, we have $\forall T > 0$, $\exists c_T$, s.t.

$$\sum_{j=1}^{k_n} \mathbb{E}\left(1 - \cos(tU_{n,j})\right) \leq c_T, \quad \forall n \geq 1, \forall |t| \leq T \tag{30}$$

so conditions (a) and (b) for $U_{n,j}$. We then show conditions (a) and (b) hold for $\tilde{X}_{n,j}$.

[Step X-2] Suppose

(b-U) $\forall \delta > 0$, $\exists c_\delta < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{P}(|U_{n,j}| \geq \delta) \leq c_\delta$, $\forall n \geq 1$.

then we need to show

(b-$\tilde{X}$) $\forall \delta > 0$, $\exists c_\delta < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta) \leq c_\delta$, $\forall n \geq 1$.

Proof: let $\sum_{j=1}^{k_n} \mathbb{P}(|U_{n,j}| \geq \frac{\delta}{2}) \leq c_\delta, \forall n \geq 1$. Note

$$c_\delta \geq \sum_{j=1}^{k_n} \mathbb{P}(|U_{n,j}| \geq \frac{\delta}{2}) = \sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j} - \tilde{X}'_{n,j}| \geq \frac{\delta}{2}) \geq \sum_{j=1}^{k_n} \mathbb{P}\left(|\tilde{X}'_{n,j}| \geq \delta, |\tilde{X}_{n,j} - \tilde{X}'_{n,j}| \geq \frac{\delta}{2}\right)$$
$$\geq \sum_{j=1}^{k_n} \mathbb{P}\left(|\tilde{X}'_{n,j}| \geq \delta, |\tilde{X}_{n,j}| < \frac{\delta}{2}\right)$$

$$= \sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta)\mathbb{P}(|\tilde{X}_{n,j}| < \frac{\delta}{2}) \tag{31}$$

Note $\tilde{X}_{n,j}$ is uniformly negligible which means

$$\max_{1\leq j\leq k_n} \mathbb{P}\left(|\tilde{X}_{n,j}| \geq \frac{\delta}{2}\right) \to 0, \quad \text{as } n \to \infty$$

So $\exists n_0$, s.t. $n \geq n_0$, $\max_{1\leq j\leq k_n} \mathbb{P}\left(|\tilde{X}_{n,j}| \geq \frac{\delta}{2}\right) < \frac{1}{2}$, which implies $\mathbb{P}\left(|\tilde{X}_{n,j}| < \frac{\delta}{2}\right) \geq \frac{1}{2}, \forall 1 \leq j \leq k_n, \forall n \geq n_0$. Going back to (31)

$$\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta)\mathbb{P}(|\tilde{X}_{n,j}| < \frac{\delta}{2}) \geq \frac{1}{2} \sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta)$$

so we conclude

$$\sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \geq \delta) \leq 2c_\delta, \forall n \geq n_0$$

since n_0 is finite, the result is easy to be extended to $n \geq 1$ by redefining c_δ. *QED.*

And suppose

(a-U) $\exists c_0 < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(U_{n,j}^2 1_{|U_{n,j}|\leq 1}) \leq c_0$, $\forall n \geq 1$.

and we need to show

(a-$\tilde{X}$) $\exists c_0 < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\leq 1}) \leq c_0$, $\forall n \geq 1$.

Proof: It can be shown from (30) that $\forall m > 0, \exists c_m < \infty$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(U_{n,j}^2 1_{|U_{n,j}|\leq m}) \leq c_m$, $\forall n \geq 1$.

$$\sum_{j=1}^{k_n} \mathbb{E}\left(1 - \cos(tU_{n,j})\right) \leq c_T, \forall n \geq 1, \forall |t| \leq T$$
$$\Longrightarrow \sum_{j=1}^{k_n} \mathbb{E}\left(1 - \cos(tU_{n,j}) 1_{|tU_{n,j}|\leq \frac{\pi}{4}}\right) \leq c_T, \forall n \geq 1, \forall |t| \leq T$$

note $\exists a > 0$, s.t. $1 - \cos x \geq ax^2$ for $|x| \leq \frac{\pi}{4}$, then

$$at^2 \sum_{j=1}^{k_n} \mathbb{E}\left(U_{n,j}^2 1_{|tU_{n,j}|\leq \frac{\pi}{4}}\right) \leq c_T$$

letting $t = T = \frac{\pi/4}{m}$ gives

$$\sum_{j=1}^{k_n} \mathbb{E}\left(U_{n,j}^2 1_{|U_{n,j}|\le m}\right) \le c_m$$

Here we select $m = 2$, so we have

$$\begin{aligned}
c_2 &\ge \sum_{j=1}^{k_n} \mathbb{E}(U_{n,j}^2 1_{|U_{n,j}|\le 2}) = \sum_{j=1}^{k_n} \mathbb{E}\left((\tilde{X}_{n,j} - \tilde{X}'_{n,j})^2 1_{|\tilde{X}_{n,j}-\tilde{X}'_{n,j}|\le 2}\right) \\
&\ge \sum_{j=1}^{k_n} \mathbb{E}\left((\tilde{X}_{n,j} - \tilde{X}'_{n,j})^2 1_{|\tilde{X}_{n,j}|\le 1, |\tilde{X}_{n,j}-\tilde{X}'_{n,j}|\le 2}\right) \ge \sum_{j=1}^{k_n} \mathbb{E}\left((\tilde{X}_{n,j} - \tilde{X}'_{n,j})^2 1_{|\tilde{X}_{n,j}|\le 1} 1_{|\tilde{X}'_{n,j}|\le 1}\right) \\
&= \sum_{j=1}^{k_n} \mathbb{E}\left(1_{|\tilde{X}'_{n,j}|\le 1} \mathbb{E}[(\tilde{X}_{n,j} - \tilde{X}'_{n,j})^2 1_{|\tilde{X}_{n,j}|\le 1} | \tilde{X}'_{n,j}]\right) \ge \sum_{j=1}^{k_n} \mathbb{E}\left(1_{|\tilde{X}'_{n,j}|\le 1} \inf_{|y|\le 1} \mathbb{E}(\tilde{X}_{n,j} - y)^2 1_{|\tilde{X}_{n,j}|\le 1}\right) \\
&\ge \sum_{j=1}^{k_n} \mathbb{E}\left(1_{|\tilde{X}'_{n,j}|\le 1} \inf_{|y|\le 1} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1} - 2y\tilde{X}_{n,j} 1_{|\tilde{X}_{n,j}|\le 1})\right) \\
&= \sum_{j=1}^{k_n} \mathbb{P}(|\tilde{X}_{n,j}| \le 1) \left(\mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) + \inf_{|y|\le 1} -2y\mathbb{E}(\tilde{X}_{n,j} 1_{|\tilde{X}_{n,j}|\le 1})\right) \\
&\ge \sum_{j=1}^{k_n} \frac{1}{2}\left(\mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) - 2|\mathbb{E}(\tilde{X}_{n,j} 1_{|\tilde{X}_{n,j}|\le 1})|\right), \text{ for } n \ge n_0 \text{ same argument as above,} \\
&= \sum_{j=1}^{k_n} \frac{1}{2}\mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) - \sum_{j=1}^{k_n} |\tilde{a}_{n,j}| \ge \frac{1}{2}\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) - \sum_{j=1}^{k_n} c'\mathbb{P}(|\tilde{X}_{n,j}| \ge \frac{1}{2}) \\
&\ge \frac{1}{2}\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) - c' \cdot c'', \text{ by (b-}\tilde{X}).
\end{aligned}$$

Therefore $\exists c_0$, s.t. $\sum_{j=1}^{k_n} \mathbb{E}(\tilde{X}_{n,j}^2 1_{|\tilde{X}_{n,j}|\le 1}) \le c_0$, $\forall n \ge n_0$, and it's easy to extend the result to $\forall n \ge 1$ by redefining c_0, noting n_0 is a finite number.

Finally, we complete the very long proof of the accompanying law. □

22 Lévy-Khintchine Theorem.

- The intuition.

Definition 28 (Lévy measures)**.** A measure λ defined on $\mathcal{B}(\mathbb{R})$ is a Lévy measure if

(a) $\forall \delta > 0$, $\exists c_\delta < \infty$, s.t. $\int_{|x|\ge\delta} \lambda(\mathrm{d}x) \le c_\delta$.

(b) $\int_{|x|\le 1} x^2\lambda(\mathrm{d}x) < \infty$.

Comment: $\lambda(x)$ can approach ∞ as $x \to 0$ but not as fast as $\frac{1}{x^2}$, and λ is integrable on the tails of its support.

Proposition 22.1. The above (a) and (b) requirements are equivalent to one single requirement as follows.

$$\int_{\mathbb{R}} \frac{x^2}{1+x^2}\lambda(\mathrm{d}x) < \infty$$

Proof. Left as exercise. □

Definition 29 (A useful class of function $\theta(x)$)**.** We introduction a class of function that will be frequently used.

(a) θ is a bounded and continuous function.

(b) $\theta(x) \sim x$.

(c) $|\theta(x) - x| \le c|x|^3$ for $|x| \le 1$.

Examples:

$$\theta(x) = \begin{cases} x & |x| \le 1 \\ \mathrm{sign}(x) & |x| > 1 \end{cases}$$

$$\theta(x) = \frac{x}{1+x^2}$$

- The theorem.

Theorem 22.1 (Lévy-Khintchine theorem)**.** Let λ be a Lévy measure, define a function

$$\varphi(t) = \exp\left\{\int [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) + ibt - \frac{\sigma^2}{2}t^2\right\} = \varphi_{\lambda,b,\sigma^2}(t)$$

(a) $\varphi(t)$ is the characteristic function of a infinitely divisible random variable X.

(b) If X has a infinitely divisible law, then $\exists(\lambda, b, \sigma^2)$ s.t. $\varphi_X(t) = \varphi_{\lambda,b,\sigma^2}(t)$.

(c) If $\varphi_{\lambda_1,b_1,\sigma_1^2}(t) = \varphi_{\lambda_2,b_2,\sigma_2^2}(t), \forall t,$ and $\lambda_1(\{0\}) = \lambda_2(\{0\}) = 0$, then $(\lambda_1, b_1, \sigma_1^2) = (\lambda_2, b_2, \sigma_2^2)$.

Proof. (a) 3 steps.

Step 1. We claim that $\varphi_\delta(t) = \exp\left\{\int_{|x|\geq\delta}[e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) + ibt\right\}$ for some $\delta > 0$ is the characteristic function of an infinitely divisible random variable.

Proof. Define a measure on $\{x \in \mathbb{R} : |x| \geq \delta\}$ via differential form $\lambda_\delta(\mathrm{d}x) = \frac{\lambda(\mathrm{d}x)}{\lambda((-\delta,\delta)^c)}$. Note that $\lambda\left((-\delta,\delta)^c\right)$ is bounded, then $\lambda_\delta(\mathrm{d}x)$ is a probability measure on $(-\delta,\delta)^c$. So

$$\begin{aligned}\varphi_\delta(t) &= \exp\left\{a\int_{|x|\geq\delta}[e^{itx} - 1]\lambda_\delta(\mathrm{d}x) + ibt - it\int_{|x|\geq\delta}\theta(x)\lambda(\mathrm{d}x)\right\}\\ &= \exp\left\{a\int_{|x|\geq\delta}[e^{itx} - 1]\lambda_\delta(\mathrm{d}x) + i(b - c_0)t\right\}\end{aligned}$$

where $a = \lambda((-\delta,\delta)^c)$ and $c_0 = \int_{|x|\geq\delta}\theta(x)\lambda(\mathrm{d}x)$, c_0 is a constant given δ for λ is Lévy measure (integrable on $(-\delta,\delta)^c$) and θ is bounded and continuous.

Let's define X to be a random variable with distribution measure λ_δ, thus

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{|x|\geq\delta} e^{itx}\lambda_\delta(dx)$$

and let

$$Y = X^{[a]} + (b - c_0)$$

then we find the characteristic function of Y is (see Example 20.1 (d))

$$\begin{aligned}\varphi_Y(t) &= \exp\left\{a(\varphi_X(t) - 1)\right\} \cdot \exp\left(it(b - c_0)\right)\\ &= \exp\left\{a\int_{|x|\geq\delta}(e^{itx} - 1)\lambda_\delta(\mathrm{d}x) + it(b - c_0)\right\}\end{aligned}$$

the same as $\varphi_\delta(t)$, and Y indeed is a infinitely divisible random variable. (Sum of independent I.D.L.'s is I.D.L.)

Step 2. Generalize Step 1. The function $\varphi(t) = \exp\left\{\int[e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) + ibt\right\}$ is the characteristic function of an infinitely divisible random variable.

Proof: Note it is enough to show that $\varphi_\delta(t) \to \varphi(t)$ as $\delta \to 0$ uniformly over $|t| \leq T$. Note $\varphi_\delta(0) = \varphi(0) = 1$ and $\varphi(t)$ is continuous at $t = 0$. Then by Lévy continuity theorem, $\varphi(t)$ is the characteristic function of some random variable.

Let c denote a universal positive constant that may change from time to time and even within a line, we use its "with-subscript" version c_q to denote the dependence of c on q. First we note

$$\sup_{|t|\le T} |\varphi(t)| \le C_T$$

for some constance C_T depending on T. This is because

$$\begin{aligned}
&\left|\int [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x)\right| \\
&\le \left|\int_{|x|\le 1} [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x)\right| + \left|\int_{|x|>1} [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x)\right| \\
&\le ct^2 \int_{|x|\le 1} x^2\lambda(\mathrm{d}x) + (2 + c|t|)\int_{|x|>1} \lambda(\mathrm{d}x) \le c_t
\end{aligned}$$

so $|\varphi(t)| \le c_t$ and $\sup_{|t|\le T} |\varphi(t)| \le C_T$. Thus $\forall |t| \le T$,

$$|\varphi_\delta(t) - \varphi(t)| = |\varphi(t)| \left|\frac{\varphi_\delta(t)}{\varphi(t)} - 1\right| \le C_T \left|\frac{\varphi_\delta(t)}{\varphi(t)} - 1\right|$$

and

$$\left|\frac{\varphi_\delta(t)}{\varphi(t)} - 1\right| = \left|\exp\left\{-\int_{|x|<\delta} [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x)\right\} - 1\right|$$

note

$$\int_{|x|<\delta} [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) \le ct^2 \int_{|x|<\delta} x^2\lambda(\mathrm{d}x) \to 0, \text{ as } \delta \to 0.$$

hence $|\varphi_\delta(t) - \varphi(t)| \to 0$ as $\delta \to 0$. Thus $\varphi(t)$ is the characteristic function of some random variable X. Then let's show X is an I.D.L., let

$$\varphi_n(t) = \exp\left\{\int [e^{itx} - 1 - it\theta(x)]\frac{\lambda(\mathrm{d}x)}{n} + i\frac{b}{n}t\right\}$$

clearly, $\frac{\lambda(\cdot)}{n}$ is again a Lévy measure, $\frac{b}{n}$ is a constant, by previous argument, $\varphi_n(t)$ is the characteristic function of some random variable and $[\varphi_n(t)]^n = \varphi(t)$, so X is indeed an I.D.L.

Step 3. Given the X in Step 2, X is an infinitely divisible random variable with characteristic function φ. Define

$$V = X + Z$$

where $Z \sim \mathcal{N}(0, \sigma^2)$ and Z is independent of X. Since $\varphi_Z(t) = e^{-\frac{\sigma^2}{2}t^2}$, one can show

$$\phi_V(t) = \exp\left\{\int [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) + ibt - \frac{\sigma^2}{2}t^2\right\}$$

and V has the infinitely divisible law.

(b) Left to the next lecture.

(c) _Step 1._ Let's show $\sigma_1 = \sigma_2$ first.

Take log on both sides of $\varphi_{\lambda_1, b_1, \sigma_1^2}(t) = \varphi_{\lambda_2, b_2, \sigma_2^2}(t), \forall t$, and divide both sides by t^2,

$$\frac{1}{t^2}\int [e^{itx} - 1 - it\theta(x)]\lambda_1(\mathrm{d}x) + i\frac{b_1}{t} - \frac{\sigma_1^2}{2} = \frac{1}{t^2}\int [e^{itx} - 1 - it\theta(x)]\lambda_2(\mathrm{d}x) + i\frac{b_2}{t} - \frac{\sigma_2^2}{2}$$

letting $t \to \infty$ is enough to kill $i\frac{b}{t}$ term and note with $\delta \le 1$ (this condition is automatically satisfied as we require $\delta \to 0$),

$$\begin{aligned}
&\frac{1}{t^2}\int [e^{itx} - 1 - it\theta(x)]\lambda_1(\mathrm{d}x) \\
=&\frac{1}{t^2}\left[\int_{|x|\le\delta} [e^{itx} - 1 - it\theta(x)]\lambda_1(\mathrm{d}x) + \int_{|x|>\delta} [e^{itx} - 1 - it\theta(x)]\lambda_1(\mathrm{d}x)\right] \\
\le&\frac{1}{t^2}\left[\int_{|x|\le\delta} ct^2x^2\lambda_1(\mathrm{d}x) + \int_{|x|>\delta} (2 + c|t|)\lambda_1(\mathrm{d}x)\right] \\
=&c\int_{|x|\le\delta} x^2\lambda_1(\mathrm{d}x) + \frac{(2 + c|t|)}{t^2}\int_{|x|>\delta} \lambda_1(\mathrm{d}x)
\end{aligned}$$

let $t \to \infty$ will kill the second term and let $\delta \to 0$ will kill the first term. So this implies $\sigma_1 = \sigma_2$. (Usually we take $\sigma \ge 0$.)

Step 2. Let's show $\lambda_1 = \lambda_2$.

Let's define

$$A(t) = \int [e^{itx} - 1 - it\theta(x)]\lambda(\mathrm{d}x) + ibt$$

therefore

$$\frac{A(t+s) + A(t-s)}{2} - A(t) = \int \left[\frac{e^{i(t+s)x} + e^{i(t-s)x}}{2} - e^{itx}\right]\lambda(\mathrm{d}x)$$

Since we have shown $\sigma_1 = \sigma_2$, then $\varphi_{\lambda_1, b_1, \sigma_1^2}(t) = \varphi_{\lambda_2, b_2, \sigma_2^2}(t), \forall t$ gives

$$\int \left[\frac{e^{i(t+s)x} + e^{i(t-s)x}}{2} - e^{itx}\right]\lambda_1(\mathrm{d}x) = \int \left[\frac{e^{i(t+s)x} + e^{i(t-s)x}}{2} - e^{itx}\right]\lambda_2(\mathrm{d}x)$$

$$\Longrightarrow \int e^{itx}\left[\frac{e^{isx}+e^{-isx}}{2}-1\right]\lambda_1(\mathrm{d}x) = \int e^{itx}\left[\frac{e^{isx}+e^{-isx}}{2}-1\right]\lambda_2(\mathrm{d}x)$$
$$\Longrightarrow \int e^{itx}(\cos(sx)-1)\lambda_1(\mathrm{d}x) = \int e^{itx}(\cos(sx)-1)\lambda_2(\mathrm{d}x)$$

Let's define

$$\mathbb{P}_1(\mathrm{d}x) = \frac{[1-\cos(sx)]\lambda_1(\mathrm{d}x)}{\int[1-\cos(sx)]\lambda_1(\mathrm{d}x)}$$
$$\mathbb{P}_2(\mathrm{d}x) = \frac{[1-\cos(sx)]\lambda_1(\mathrm{d}x)}{\int[1-\cos(sx)]\lambda_2(\mathrm{d}x)}$$

Then by Theorem 11.2

so we have

$$[1-\cos(sx)]\lambda_1(\mathrm{d}x) = [1-\cos(sx)]\lambda_2(\mathrm{d}x), \quad \forall x$$

so we have $\lambda_1 = \lambda_2$.

Then $b_1 = b_2$ easily follows from $\sigma_1 = \sigma_2$ and $\lambda_1 = \lambda_2$.

□

23 The One-dimensional Central Limit Problem, I

Suppose we have an array of *i.i.d.* r.v.'s $(X_{n,j})$ where $1 \le j \le k_n$ with $k_n \to \infty$ as $n \to \infty$. What are the sufficient condition(s) s.t.

$$\sum_{j=1}^{k_n} X_{n,j} - A_n \xrightarrow{d} S \tag{32}$$

for some sequence A_n and some limit S?

Recall (26) in the section on accompany law,

24 The One-dimensional Central Limit Problem, II

We should prove the following theorem.

Theorem 24.1.

25 Conditional Expectation.

- Definitions.

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, given $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$.

Definition 30 (Conditional probability)**.** For any $B \in \mathcal{F}$, define the conditional probability given A as

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

Moreover, if we let $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot|A)$, then $\mathbb{P}_A$ is a well-defined probability measure on the measurable space $(\Omega, \mathcal{F})$. (Show it).

Definition 31 (Conditional expectation given a set)**.** Let $f : \Omega \to \mathbb{R}$ and $f \in L^1(\mathbb{P})$ which means $\int |f| \mathrm{d}\mathbb{P} < \infty$, then the conditional expectation is defined as

$$\mathbb{E}(f|A) = \mathbb{E}_A(f) = \int f \mathrm{d}\mathbb{P}_A = \frac{\int f 1_A \mathrm{d}\mathbb{P}}{\mathbb{P}(A)} = \frac{\int_A f \mathrm{d}\mathbb{P}}{\mathbb{P}(A)}$$

<u>Remark 1</u>. For $B \in \mathcal{F}$, let $f = 1_B$, then as usual,

$$\mathbb{E}(1_B|A) = \frac{\int_A 1_B \mathrm{d}\mathbb{P}}{\mathbb{P}(A)} = \frac{\int_{A\cap B} \mathrm{d}\mathbb{P}}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(A|B)$$

<u>Remark 2</u>. The conditional expectation given a measurable set is a non-random constant.

- Let $(A_j)_{j\geq 1}$ be a "measurable" partition of Ω which means $A_j \in \mathcal{F}$, $\cup_{j\geq 1} A_j = \Omega$ for all j, $A_i \cap A_j = \varnothing$ for any $i \neq j$, and $\mathbb{P}(A_j) > 0$ for every j. Then we can show for $B \in \mathcal{F}$

$$\mathbb{P}(B) = \sum_{j\geq 1} \mathbb{P}(B|A_j)\mathbb{P}(A_j)$$
$$\mathbb{E}(f) = \sum_{j\geq 1} \mathbb{E}(f|A_j)\mathbb{P}(A_j)$$

Let's define a σ-algebra $\mathcal{G}$ generated by the measurable partition, that is, $\mathcal{G} = \sigma(A_j : j \geq 1)$. Clearly, $\mathcal{G} \subset \mathcal{F}$.

Lemma 25.1. $B \in \mathcal{G} \iff \exists M \subset \mathbb{N}$ s.t. $B = \cup_{j\in M} A_j$.

Proof. $\square$

Then we have

Lemma 25.2. For a function $h : \Omega \to \mathbb{R}$, then h is $\mathcal{G}$-measurable *iff* h is constant on A_j.

Proof.

$$h(\omega) = \sum_{x \in B_0} x \cdot 1_{\{h^{-1}(x)\}}(\omega)$$

$\square$

Definition 32 (Conditional expectation given a "special" σ-algebra)**.** Let f be $\mathcal{F}$-measurable and $f \in L^1(\mathbb{P})$. Define the conditional expectation given $\mathcal{G}$ as a random variable $\mathbb{E}[f|\mathcal{G}] : \Omega \to \mathbb{R}$

$$\mathbb{E}[f|\mathcal{G}](\omega) = \sum_{j \geq 1} \mathbb{E}[f|A_j] \cdot 1_{A_j}(\omega)$$

Remark. The conditional expectation given a σ-algebra is a random variable.

Proposition 25.1. We have

(i) $\mathbb{E}(f|\mathcal{G})$ is $\mathcal{G}$-measurable.

(ii) $\forall B \in \mathcal{G}$, $\int_B \mathbb{E}(f|\mathcal{G}) \mathrm{d}\mathbb{P} = \int_B f \mathrm{d}\mathbb{P}$

(iii) Let $h : \Omega \to \mathbb{R}$ be $\mathcal{G}$-measurable and bounded, so fh is integrable. We have

$$\int fh \mathrm{d}\mathbb{P} = \int \mathbb{E}(f|\mathcal{G}) h \mathrm{d}\mathbb{P}$$

Proof. (i)

(ii) We show $\mathbb{E}(f|\mathcal{G}) \in L^1(\mathbb{P})$.

$$\begin{aligned}
\int |\mathbb{E}(f|\mathcal{G})| \mathrm{d}\mathbb{P} &= \int \left| \sum_{j \geq 1} \mathbb{E}(f|A_j) \cdot 1_{A_j} \right| \mathrm{d}\mathbb{P} \\
&\leq \int \sum_{j \geq 1} \left| \mathbb{E}(f|A_j) \cdot 1_{A_j} \right| \mathrm{d}\mathbb{P}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\geq 1} \int \left|\mathbb{E}(f|A_j)\cdot 1_{A_j}\right| \mathrm{d}\mathbb{P} \quad \text{(Tonelli's theorem)} \\
&= \sum_{j\geq 1} |\mathbb{E}(f|A_j)|\mathbb{P}(A_j) = \sum_{j\geq 1} \left|\frac{\int_{A_j} f \mathrm{d}\mathbb{P}}{\mathbb{P}(A_j)}\right| \mathbb{P}(A_j) \\
&\leq \sum_{j\geq 1} \int_{A_j} |f|\mathrm{d}\mathbb{P} = \int |f|\mathrm{d}\mathbb{P} < \infty
\end{aligned}
$$

Moreover, $h\mathbb{E}(f|\mathcal{G})$ is also integrable.

(iii) we proceed to prove the equation, which is decomposed into two steps. First, we claim if $h = 1_B$ for $B \in \mathcal{G}$, the statement holds.

$$
\begin{aligned}
\int 1_B \mathbb{E}(f|\mathcal{G})\mathrm{d}\mathbb{P} = \int 1_B \sum_{j\geq 1} \mathbb{E}(f|A_j)1_{A_j}\mathrm{d}\mathbb{P} &\overset{(*)}{=} \int \sum_{j\in M_B} \mathbb{E}(f|A_j)1_{A_j}\mathrm{d}\mathbb{P} \\
&= \sum_{j\in M_B} \int 1_{A_j}\mathbb{E}(f|A_j)\mathrm{d}\mathbb{P} \quad \text{(by DCT)} \\
&= \sum_{j\in M_B} \frac{\int_{A_j} f\mathrm{d}\mathbb{P}}{\mathbb{P}(A_j)}\mathbb{P}(A_j) = \sum_{j\in M_B} \int_{A_j} f\mathrm{d}\mathbb{P} = \int 1_B f \mathrm{d}\mathbb{P}
\end{aligned}
$$

where step $(*)$ is from Lemma 25.1, and the last step is again Tonelli's theorem. Finally, if h is bounded and $\mathcal{G}$-measurable, then by Lemma 25.2

$$
h = \sum_{j\geq 1} h_j \cdot 1_{A_j}
$$

for constants h_j and $A_j \in \mathcal{G}$. So by the first claim, we have

$$
\begin{aligned}
\int h\mathbb{E}(f|\mathcal{G})\mathrm{d}\mathbb{P} = \int \sum_{j\geq 1} h_j \cdot 1_{A_j}\mathbb{E}(f|\mathcal{G})\mathrm{d}\mathbb{P} &\overset{\star}{=} \sum_{j\geq 1} \int h_j \cdot 1_{A_j}\mathbb{E}(f|\mathcal{G})\mathrm{d}\mathbb{P} \\
&= \sum_{j\geq 1} h_j \int 1_{A_j} f \mathrm{d}\mathbb{P} = \int \sum_{j\geq 1} h_j \cdot 1_{A_j} f\mathrm{d}\mathbb{P} \\
&= \int hf\mathrm{d}\mathbb{P}
\end{aligned}
$$

where step $(\star)$ is by DCT.

Remark. A road map here will be helpful.

Definition 33 (Conditional expectation given "any" sub σ-algebra)**.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let f be $\mathcal{F}$-measurable and $f \in L^1(\mathbb{P})$, let $\mathcal{H}$ be a σ-algebra with $\mathcal{H} \subset \mathcal{F}$, then the conditional expectation $\mathbb{E}(f|\mathcal{H})$ is a random variable $\mathbb{E}(f|\mathcal{H}) : \Omega \to \mathbb{R}$ s.t.

(i) $\mathbb{E}(f|\mathcal{H})$ is $\mathcal{H}$-measurable.

(ii) $\forall B \in \mathcal{H}$, $\int_B \mathbb{E}(f|\mathcal{H})\mathrm{d}\mathbb{P} = \int_B f\mathrm{d}\mathbb{P}$.

Theorem 25.1 (Existence and uniqueness of the conditional expectation)**.** In the description of Definition 33, $\mathbb{E}(f|\mathcal{H})$ uniquely[7] exists.

Proof. (Existence.) Define a measure $\mathbb{Q}$ on $(\Omega, \mathcal{H})$ by

$$\mathbb{Q}(A) = \int_A f\mathrm{d}\mathbb{P}$$

then $\mathbb{Q}$ is indeed a measure. Note $\mathbb{P}$ is also a measure on $(\Omega, \mathcal{H})$, and $\mathbb{Q} \ll \mathbb{P}$. So by Randon-Nikodym theorem, there exists a $\mathcal{H}$-measurable function h s.t.[8]

$$\mathbb{Q}(A) = \int_A h\mathrm{d}\mathbb{P}$$

Clearly, h satisfies the definition of $\mathbb{E}(f|\mathcal{H})$. First, $h \in \mathcal{H}$; second, $\int_A h\mathrm{d}\mathbb{P} = \int_A f\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{H}$.

(Uniqueness.) Suppose h_1, h_2 both satisfy the definition of the conditional expectation, then one can obtain that both h_1 and h_2 are $\mathcal{H}$-measurable and

$$\int_A h_1\mathrm{d}\mathbb{P} = \int_A h_2\mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{H}$$

so $h_1 = h_2$ a.s. □

<u>Remark</u>. In the existence proof, can we say that $f = h$? No, because f is $\mathcal{F}$-measurable, while h is $\mathcal{H}$-measurable, they can be different. □

Proposition 25.2. $\mathbb{E}(f|\mathcal{H})$ satisfies the following properties

(i) $\mathbb{E}(f|\mathcal{H}) \in L^1(\mathbb{P})$.

(ii) If additionally f is $\mathcal{H}$-measurable, then $\mathbb{E}(f|\mathcal{H}) = f$ a.s. (Tautology)

[7] The uniqueness is in the "up to a null set" sense.

[8] The function h is usually written as $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ in this context.

Remark 1. A decomposition.

$$f = (f - \mathbb{E}(f|\mathcal{H})) + \mathbb{E}(f|\mathcal{H})$$

then $\mathbb{E}(f|\mathcal{H})$ is $\mathcal{H}$-measurable and $(f - \mathbb{E}(f|\mathcal{H})) \perp \mathcal{H}$, where a random variable $X \perp \mathcal{H}$ is defined as

Remark 2. Two notations.

- $\star$ For $A \in \mathcal{F}$, then $\mathbb{P}(A|\mathcal{H}) = \mathbb{E}(1_A|\mathcal{H})$.
- $\star$ $\sigma(X)$ is the smallest σ-algebra s.t. X is measurable, i.e.

$$\sigma(X) = \sigma(X^{-1}(B) : B \in \mathcal{B}(\mathbb{R}))$$

and we write $\mathbb{E}(f|\sigma(X))$ as $\mathbb{E}(f|X)$. Similarly,

$$\mathbb{E}(f|X_1, \cdots, X_n) = \mathbb{E}(f| \vee_{i=1}^n \sigma(X_i))$$

26 Properties of Conditional Expectation

Notations

$\mathbb{N} = \{1, 2, 3, \cdots\}, \quad \mathbb{N}_0 = \{0, 1, 2, 3, \cdots\}.$

References

Billingsley, P. (1999): *Convergence of Probability Measures*. John Wiley & Sons, 2 edn.

Chow, Y. S., and H. Teicher (1997): *Probability Theory: Independence, Interchangeability, Martingales*. Springer.

Durrett, R. (2019): *Probability: Theory and Examples*, vol. 49. Cambridge university press, 5 edn.

Mörters, P., and Y. Peres (2010): *Brownian motion*, vol. 30. Cambridge University Press.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Snell, J. L. (1997): "A Conversation with Joe Doob," *Statistical Science*, 12(4), 301–311.