UNIVERSITAT DE BARCELONA
Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball Final de Grau

From Probability to Convergence: Exploring the Foundations of the Central Limit Theorem

Autora: Aina Nadal Gavaldà

Director: Dr. David Márquez
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 17 de gener de 2024

Abstract

This thesis embarks on a mathematical journey, delving into the depths of probability theory and a comprehensive exploration of the Central Limit Theorem (CLT). The study begins with fundamental concepts such as characteristic functions and basic probability theory, continuing with the laws of large numbers. Afterwards, the narrative progresses to focus on various versions of the CLT. Essential theorems, including De Moivre-Laplace's, Lindeberg-Lévy's, and Lyapunov's, are studied, offering insights into the universal significance of the CLT. The journey concludes with a glimpse of the constraints of this impactful theorem, including the convergence of the compound Poisson distribution. This work contributes to a nuanced understanding of probability theory as well as serves as a guide through the elegance and applicability of the CLT.

Notation: Looking for a concise text, the following notation has been used: i.e. stands for "that is", r.v. means "random variable" and s.t. stands for "such that".

2020 Mathematics Subject Classification. 60F05, 60B10, 60E05, 60E10, 62H05, 62H10

Acknowledgment

As I pen down these words of gratitude, my heart is filled with appreciation for the support and care, that have illuminated my journey towards completing this thesis, which I see as the culmination of my Bachelor of Mathematics.

First and foremost, a special acknowledgment is for my thesis tutor, whose attention, time, and mentoring have been invaluable these past few months. Your guidance has been instrumental, in shaping the trajectory of this work and contributing to my development as a scholar.

I am deeply thankful to my family, the source of unending support and strength. To my dear mommy, whose love has been my anchor. To my daddy, for showing me resilience and perseverance. To my brother, for being the person who I look up to. To my boyfriend, whose encouragement has been a constant presence. To my friends who began being classmates, your understanding and companionship have been a light in the often challenging academic landscape. Thank you for standing by me and enriching these years with laughter and shared dreams.

I express my deepest gratitude to all the professors who have crossed my academic path, from the early days of high school to the present moment. In particular, I recall my initial encounter with probability theory, a subject that did not immediately capture my appreciation. Through the guidance and passion of my professors, I not only came to appreciate probabilities but developed a genuine love for the subject. This, to me, adds a poetic dimension to this thesis theme.

This journey would not have been possible without the warmth and wisdom of those who supported and believed in me. To each one of you, thank you and lots of love.

Gràcies i molt d'amor,
Aina.

Contents

Preface

In the vast landscape of mathematics, probability theory emerges as a lens through which uncertainty and randomness are explored. At the heart of this field lies the Central Limit Theorem (CLT), a transformative concept with far-reaching implications. This introduction sets the stage for our exploration, weaving a narrative that transitions from foundational probability concepts to the intricacies of convergence.

The Central Limit Theorem asserts that, under specific conditions, the distribution of the standardized sample mean converges to a standard normal distribution. Remarkably, this convergence holds true even when the original variables exhibit non-normal distributions. This theorem is crucial in probability theory, as it signifies that probabilistic and statistical methods that work for normal distributions can be extended to tackle problems involving diverse distribution types.

Chapter 1 provides an immersive introduction to probability theory, laying the groundwork for further discussions on the convergence of random variables. As we deepen into the nuances of characteristic functions in *Chapter 2*, we set the stage for a more advanced exploration of the CLT.

The Law of Large Numbers, unfolded in *Chapter 3*, introduces essential inequalities and theorems, preparing the stage for our in-depth exploration of weak convergence in *Chapter 4*. Here, we search through finite measures, probabilities, and convolutions, laying a profound foundation for comprehending the various versions of the Central Limit Theorem explored in *Chapter 5*.

The latter chapter, the pinnacle of our journey, unravels classical theorems and their multidimensional extensions, illustrating the universal significance of the CLT. Essential theorems will be included, such as De Moivre-Laplace's, Lindeberg-Lévy's, and Lyapunov's. This intellectual voyage concludes with a glimpse of the constraints and limitations of this impactful theorem, offering insights into the convergence of the compound Poisson distribution.

As the pages turn, the intent is not merely to present theorems but to provide a lens through which the significance and applicability of the CLT become apparent. This journey, from the basics of probability to the intricate details of the CLT, aims to provide a solid foundation for researchers, academics, and enthusiasts alike.

Before we embark on this journey, I extend gratitude to you, dear reader, for dedicating your time and attention to this work. Allow me a moment of cheeriness; a joke that encapsulates the essence of the CLT:

"Why did the data point throw a party for the Central Limit Theorem? Because it wanted to show its friends that even in a wild and diverse crowd, when you gather enough of them, they all tend to behave like normal distributions!"

Chapter 1

Introduction to probability and convergence

In this chapter, we will cover a concise introduction to the probability theory. We will study fundamental concepts and connect them to the convergence of random variables and see the most common ones. This chapter is very much extracted from *([1], Chapter 7)* and *([2], Chapter 9)*.

1.1 Introduction to probability theory

First of all, recall the next crucial theorem. The interpretation is that a pushforward measure is defined by transferring a measure from one measurable space to another one, using a measurable function.

Theorem 1.1.1. ***(Pushforward Measure Theorem).*** *Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be measurable spaces, where $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras[1] on $\mathcal{A}$ and $\mathcal{B}$, respectively, and let $f : X \longrightarrow Y$ be a measurable function. If μ is a measure on $(X, \mathcal{A})$, then the pushforward measure $f_*\mu$ is*

$$(f_*\mu)(B) := \mu(f^{-1}(B)), \; \forall B \in \mathcal{B}. \tag{1.1.1}$$

Proof. The proof can be found in (*[2], Proposition 6.9, Chapter 6*). ∎

Let $(\Omega, \mathcal{A}, P)$ be a **probability space**, where Ω is the sample space, $\mathcal{A}$ denotes a σ-algebra on Ω and P is a probability measure on $(\Omega, \mathcal{A})$. The **probability** of an event $B \in \mathcal{A}$ is defined by

$$P(B) := \int_B dP(\omega) = \int_\Omega \mathbb{1}_B(\omega) \, dP(\omega). \tag{1.1.2}$$

[1] $\mathcal{A}$ is a σ-**algebra** on Ω if it is a collection of subsets of Ω such that: $\Omega \in \mathcal{A}$, it is closed under complementation and it is closed under countable unions. It is also known as σ-field.

P satisfies summation to unity, i.e. $P(\Omega) = 1$, and σ-additivity. The **union bound** (or Boole's inequality) states that if $B_l \in \mathcal{A},\ l \in \{1, ..., n\}$ is a condition of events,

$$P\left(\bigcup_{l=1}^{n} B_l\right) \leq \sum_{l=1}^{n} P(B_l). \tag{1.1.3}$$

A **random variable** (r.v.) X is a real-valued measurable function on $(\Omega, \mathcal{A})$, i.e. $X : \Omega \longrightarrow \mathbb{R}$. Recall that X is called measurable if

$$X^{-1}(A) := \{\omega \in \mathcal{A} : X(\mathcal{A}) \in A\} \subset A,$$

for all Borel measurable subsets $A \in \mathbb{R}$, i.e. $A \in \mathcal{A}(\mathbb{R})$. The **law** of X, P_X, is

$$P_X(B) := P(X^{-1}(B)), \tag{1.1.4}$$

for every $B \in \mathcal{B}(\mathbb{R})$, which is exactly the pushforward measure of P using X. The **distribution function** $F : \mathbb{R} \longrightarrow [0, 1]$ of X, $F := F_X$, is given by

$$F(t) := P\{X \leq t\}, \tag{1.1.5}$$

for $t \in \mathbb{R}$. F must be non-decreasing, right-continuous, and the limits at infinity such that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Consider $a, b \in \mathbb{R}$. A random variable X has a **probability density function** $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ associated if

$$P\{a < X \leq b\} = \int_a^b f(t)\, dt, \tag{1.1.6}$$

for every $a < b$, then $f = \frac{\delta}{\delta t} F(t)$. f is non-negative everywhere and satisfies integration to unity, i.e. $\int_{-\infty}^{+\infty} f(x) = 1$. If X is integrable, the **expectation** or mean of X is

$$E(X) := \int_\Omega X(\omega)\, dP(\omega) = \int_\mathbb{R} x P_X(dx). \tag{1.1.7}$$

X is **integrable** with relation to P if the previous integral is well-defined and finite. If X has p^{th}-order **absolute moment**, i.e. $E(|X|^p) < \infty$, the p^{th}-order **moment** of X is $E(X^p)$, for $p > 0$. If $E(X^2) < \infty$, the quantity $E[(X - E(X))^2] = E(X^2) - [E(X)]^2$ is the **variance**. Let X, Y be r.v. on a probability space $(\Omega, \mathcal{A}, P)$. The **Holder's Inequality** states that for $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then

$$|E[XY]| \leq (E[|X|^p])^{\frac{1}{p}} (E[|Y|^q])^{\frac{1}{q}} \tag{1.1.8}$$

The function $t \longmapsto P\{|X| \geq t\}$ is the **tail** of X and it can be estimated by:

Theorem 1.1.2. ***(Markov's Inequality).*** *Let X be a random variable with $E(|X|) < \infty$, then*

$$P\{|X| \geq t\} \leq \frac{E(|X|)}{t},\ \forall t > 0. \tag{1.1.9}$$

Proof. Observe that $P\{|X| \geq t\} = E[\mathbb{1}_{\{|X|\geq t\}}]$ and $t\mathbb{1}_{\{|X|\geq t\}} \leq |X|$. Therefore, $tP\{|X| \geq t\} = tE[\mathbb{1}_{\{|X|\geq t\}}] = E[t\mathbb{1}_{\{|X|\geq t\}}] \leq E(|X|)$, as we wanted. ∎

Consequently, if $p > 0$, $P\{|X| \geq t\} = P\{|X|^p \geq t^p\} \leq t^{-p}E(|X|^p)$, for every $t > 0$. If $p = 2$ this is called the **Chebyshev's inequality**.

A random vector $\mathbf{X} = [X_1, ..., X_n]^T$ is a collection of n r.v. on a probability space $(\Omega, \mathcal{A}, P)$. The next definitions are similar to the univariant case. The **joint distribution function** F of $\mathbf{X}$ is defined as

$$F(t_1, ..., t_n) := P\{X_1 \leq t, ..., X_n \leq t\}, \; t_1, ..., t_n \in \mathbb{R}. \tag{1.1.10}$$

A random vector $\mathbf{X}$ has a joint probability density $f : \mathbb{R}^n \longrightarrow [0, 1]$ if

$$P\{\mathbf{X} \in D\} := \int_D f(t_1, ..., t_n)\, dt_1....dt_n, \tag{1.1.11}$$

for every $D \in \mathcal{B}(\mathbb{R})$. The expectation of $\mathbf{X}$ is $E(\mathbf{X}) = [E(X_1), ..., E(X_n)]^T \in \mathbb{R}^n$.

A collection of random variables $X_1, ..., X_n$ is (stochastically) **independent** if for all $t_1, ..., t_n \in \mathbb{R}$,

$$P\{X_1 \leq t_1, ..., X_n \leq t_n\} = \prod_{l=1}^{n} P\{X_l \leq t_l\}. \tag{1.1.12}$$

If they are independent r.v. with $E(|X_l|) < \infty$, for every $l \in \{1, ..., n\}$ then they satisfy that

$$E\left[\prod_{l=1}^{n} X_l\right] = \prod_{l=1}^{n} E(X_l) \tag{1.1.13}$$

and if they also have a joint density function f, then $f(t_1, ..., t_n) = f_1(t_1)...f(t_n)$, where $f_1, ..., f_n$ are the density functions of $X_1, ..., X_n$. A collection of independent r.v. that all have the same distribution are **independent identically distributed**.

Theorem 1.1.3. ***(Jensen's inequality).****: Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a convex function and $X \in \mathbb{R}^n$ be a random variable with $E(|X|) < \infty$ and $E[|f(X)|] < \infty$, then*

$$f(E[X]) \leq E[f(X)]. \tag{1.1.14}$$

Proof. The proof can be found in *([1], Theorem 7.9, Chapter 7).* ∎

1.2 Convergence of random variables

Once the concept of random variables is introduced, it is only natural to consider sequences of r.v., in particular their limit and convergence. Recall that if $\{A_n, n \geq 1\}$ is a sequence of events of $\mathcal{A}$, then $P\{\limsup_n A_n\} \geq \limsup_n P(A_n)$ and $P\{\liminf_n A_n\} \leq \liminf_n P(A_n)$.

Lemma 1.2.1. *The Borel-Cantelli's Lemmas state the following:*

1) ***(First Borel-Cantelli's Lemma)**. Let $\{A_n, n \geq 1\}$ be a sequence of events of $\mathcal{A}$. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P\{\limsup_n A_n\} = 0$.*

2) ***(Second Borel-Cantelli's Lemma)**. Let $\{A_n, n \geq 1\}$ be a sequence of independent events of $\mathcal{A}$. If $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P\{\limsup_n A_n\} = 1$.*

The next modes of convergence are the most common and important ones.

Definition 1.2.2. *A sequence of r.v. $\{X_n, n \geq 1\}$ converges **almost surely** (a.s.) to X if there is a set $N \in \mathcal{A}$ with probability zero s.t. $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, $\forall \omega \notin N$. This is denoted $X_n \xrightarrow{a.s.} X$.*

The a.s. convergence is the most similar to pointwise convergence of a sequence of functions. A useful tool might be the following.

Proposition 1.2.3. *A sequence of r.v. $\{X_n, n \geq 1\}$ converges a.s. to X if and only if*

$$\lim_{m\to\infty} P\left\{\sup_{n\geq m} |X_n - X| \leq \epsilon\right\} = 1, \ \forall \epsilon > 0. \tag{1.2.1}$$

Definition 1.2.4. *A sequence of r.v. $\{X_n, n \geq 1\}$ converges **in probability** to X if*

$$\lim_{n\to\infty} P\{|X_n - X| \geq \epsilon\} = 0, \ \forall \epsilon > 0. \tag{1.2.2}$$

or, equivalently, $\lim_{n\to\infty} P\{|X_n - X| \leq \epsilon\} = 1$. This is denoted $X_n \xrightarrow{P} X$.

The probability function is more relevant and measures how close or distant the sequence of random variables and the possible limit r.v. are. For $p \in [1, \infty]$, define $L^p(\Omega, \mathcal{A}, P)$ as the set of r.v. with finite p^{th}-order moment.

Definition 1.2.5. *Let $\{X_n, n \geq 1\}$ be a sequence of r.v. of $L^p(\Omega, \mathcal{A}, P)$. This sequence converges in L^p**-norm** to a random variable X with finite p^{th}-order moment, if*

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0. \tag{1.2.3}$$

This is denoted $X_n \xrightarrow{L^p} X$. If $p = 1$ it is called convergence in mean and if $p = 2$ it is named convergence in mean square.

Some relevant relations between convergences are the next ones.

Theorem 1.2.6. *Let $\{X_n, n \geq 1\}$ be a sequence of r.v. Let X be a r.v. Then*

1) *If $X_n \xrightarrow{a.s.} X$, as $n \to \infty$, then also $X_n \xrightarrow{P} X$, as $n \to \infty$.*

2) *If $X_n \xrightarrow{P} X$, as $n \to \infty$, then there exists a subsequence $\{X_{n_k}, k \geq 1\}$ of random variables such that converges almost surely to X, i.e. $X_{n_k} \xrightarrow{a.s.} X$, as $k \to \infty$.*

Proposition 1.2.7. *Let $\{X_n, n \geq 1\}$ be a sequence of r.v. such that $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{P} X$, as $n \to \infty$.*

Chapter 2

Characteristic functions

In this chapter, we explore the use and properties of characteristic functions, introduced by Paul-Lévy. This will be fundamental to simplify the analysis of the Weak Convergence of probabilities *(Chapter 4)* and prove versions of the Central Limit Theorem *(Chapter 5)*. This chapter is essentially from *([2], Chapter 9)*.

Definition 2.0.1. *Let μ be a probability in $\mathbb{R}$. The* ***characteristic function*** *of μ is the map $\varphi_\mu : \mathbb{R} \longrightarrow \mathbb{C}$, defined by*

$$\varphi_\mu(t) := \int_{\mathbb{R}} e^{itx}\ \mu(dx) = \int_{\mathbb{R}} \cos(tx)\ \mu(dx) + i \int_{\mathbb{R}} \sin(tx)\ \mu(dx). \tag{2.0.1}$$

φ_μ is well-defined since the sine and cosine are continuous and bounded.

Definition 2.0.2. *Let X be a random variable. The* ***characteristic function*** *of X is*

$$\varphi_X(t) := \int_{\mathbb{R}} e^{itx}\ P_X(dx) = E(e^{itX}). \tag{2.0.2}$$

Definition 2.0.3. *Let μ be a probability in $\mathbb{R}^n$. The characteristic function of μ is the map $\varphi_\mu : \mathbb{R}^n \longrightarrow \mathbb{C}$, defined by*

$$\varphi_\mu(t) := \int_{\mathbb{R}^n} e^{i<t,x>}\ \mu(dx). \tag{2.0.3}$$

The characteristic function of a random vector $\boldsymbol{X} = (X_1, ..., X_n)$ is the characteristic function of its distribution function.

2.1 Fundamental properties of characteristic functions

From now on, let μ be a probability in $\mathbb{R}^n$. The following properties are basic.

1) Trivially, $\varphi_\mu(0) = 1$.

2) $|\varphi_\mu(t)| \leq 1$, for every $t \in \mathbb{R}^n$.

Proof. A consequence of the identity $|e^{i<t,x>}| = 1$. ■

3) $\varphi_\mu(-t) = \overline{\varphi_\mu(t)}$.

Proof. Note that

$$\varphi_\mu(-t) = \int_{\mathbb{R}^n} e^{-i<t,x>} \mu(dx) = \int_{\mathbb{R}^n} \overline{e^{i<t,x>}} \mu(dx) = \overline{\int_{\mathbb{R}^n} e^{i<t,x>} \mu(dx)}.$$

■

4) φ_μ is a uniformly continuous function.

Proof. Consider $s, t \in \mathbb{R}^n$, then

$$\left|\varphi_\mu(t) - \varphi_\mu(s)\right| = \left|\int_{\mathbb{R}^n} (e^{i<t,x>} - e^{i<s,x>}) \mu(dx)\right| \leq \int_{\mathbb{R}^n} \left|e^{i<t-s,x>} - 1\right| \mu(dx).$$

Since $\left|e^{i<t-s,x>} - 1\right| \leq 2$ and the last integrand converges to 0 as $|t-s| \longrightarrow 0$, by the Dominated Convergence Theorem, we get the desired result. ■

5) Let $\mathbf{X}$ be a random n-dimensional vector, A a $m \times n$ matrix, $b \in \mathbb{R}^m$. Then,

$$\varphi_{A\mathbf{X}+b}(t) = e^{i<t,b>} \varphi_{\mathbf{X}}(A^* t), \ \forall t \in \mathbb{R}^m. \tag{2.1.1}$$

Proof. Observe that

$$\varphi_{A\mathbf{X}+b}(t) = E(e^{i<t,A\mathbf{X}+b>}) = e^{i<t,b>} E(e^{i(A^*t)^*\mathbf{X}}) = e^{i<t,b>} \varphi_{\mathbf{X}}(A^*t).$$

■

6) **(Fundamental property of injectivity).** If μ_1 and μ_2 are probabilities in $\mathbb{R}^n$ satisfying that $\varphi_{\mu_1} = \varphi_{\mu_2}$, then $\mu_1 = \mu_2$.

Proof. Assume $n = 1$. Observe the following. Fix an interval $[-T, T]$, then by the Stone-Weierstrass' Theorem, the finite linear combinations of $e^{\frac{i\pi kx}{T}}$, $k \in \mathbb{Z}$, are dense in the set of continuous functions in $[-T, T]$. These combinations form an algebra of functions. Therefore, we want to show that

$$\int_{\mathbb{R}} f \, d\mu_1 = \int_{\mathbb{R}} f \, d\mu_2, \tag{2.1.2}$$

for every real continuous function f with a compact image. Take $\epsilon > 0$ and $T > 0$ satisfying that the image of f is in $[-T, T]$ and also $\mu_1([-T,T]^c) \leq \epsilon$, $\mu_2([-T,T]^c) \leq \epsilon$. By the previous observation, there exists a function

$$\hat{f}(x) = \sum_{j=1}^{m} a_j \exp\left(\frac{i\pi k_j x}{T}\right),$$

with $a_j \in \mathbb{C}$ and $\sup_{|x|\leq T} \left| f(x) - \hat{f}(x) \right| < \epsilon$. By hypothesis,

$$\int_{\mathbb{R}} \hat{f}\, d\mu_1 = \int_{\mathbb{R}} \hat{f}\, d\mu_2.$$

On the other hand, since $\hat{f}$ is periodic with period $2T$, then

$$\|\hat{f}\|_\infty = \sup_{x\in[-T,T]} \left| \hat{f}(x) \right| \leq \epsilon + \|f\|_\infty.$$

Consequently,

$$\begin{aligned}\left| \int_{\mathbb{R}} f\, d\mu_1 - \int_{\mathbb{R}} f\, d\mu_2 \right| &\leq \left| \int_{\mathbb{R}} f\, d\mu_1 - \int_{\mathbb{R}} \hat{f}\, d\mu_1 \right| + \left| \int_{\mathbb{R}} \hat{f}\, d\mu_2 - \int_{\mathbb{R}} f\, d\mu_2 \right| \\ &\leq \epsilon\, (\mu_1([-T,T]) + \mu_2([-T,T])) + 2\epsilon\, (\epsilon + 2\|f\|_\infty) \\ &\leq 2\epsilon(1 + \epsilon + 2\|f\|_\infty),\end{aligned}$$

and taking $\epsilon \to 0$, we have proved (2.1.2). Finally, we are done.[1] ∎

7) Let $\mathbf{X} = (X_1, ..., X_n)$ be a random vector. The random variables $X_1, ..., X_n$ are independent if and only if $\varphi_{\mathbf{X}}(t_1, ..., t_n) = \varphi_{X_1}(t_1)...\varphi_{X_n}(t_n)$.

Proof. Recall that $X_1, ..., X_n$ are independent if and only if it is satisfied the equality $P \circ \mathbf{X}^{-1} = P \circ X_1^{-1} \times ... \times P \circ X_n^{-1}$.[2] Equivalently, using the *Property 6)*, the corresponding characteristic functions must be equal. The characteristic function of $P \circ \mathbf{X}^{-1}$ is $\varphi_{\mathbf{X}}(t_1, ..., t_n)$ and of $P \circ X_1^{-1} \times ... \times P \circ X_n^{-1}$ is the following

$$\begin{aligned}&\int_{\mathbb{R}^n} e^{i<t,\mathbf{X}>}\, (P \circ X_1^{-1})(dx_1)...(P \circ X_n^{-1})(dx_n) \\ &= \left(\int_{\mathbb{R}} e^{it_1 X_1}\, (P \circ X_1^{-1})(dx_1) \right) ... \left(\int_{\mathbb{R}} e^{it_n X_n}\, (P \circ X_n^{-1})(dx_n) \right) \\ &= \varphi_{X_1}(t_1)...\varphi_{X_n}(t_n),\end{aligned}$$

which is exactly what we wanted. ∎

[1] Note that $\epsilon > 0$ is arbitrary.

[2] The used notation is reviewed in *Chapter 1*.

8) If $X_1, ..., X_n$ are independent random variables, then

$$\varphi_{X_1+...+X_n}(t) = \varphi_{X_1}(t)...\varphi_{X_n}(t). \tag{2.1.3}$$

Proof. By hypothesis,

$$\varphi_{X_1+...+X_n}(t) = E(e^{it(X_1+...+X_n)}) = E(e^{itX_1})...E(e^{itX_n}) = \varphi_{X_1}(t)...\varphi_{X_n}(t).$$

∎

2.2 Examples of characteristic functions

The following example will be crucial from now on.

1) **Standard Normal distribution:** The characteristic function of a probability with distribution $N(0,1)$ is

$$\varphi(t) = e^{\frac{-t^2}{2}}. \tag{2.2.1}$$

Let μ be the standard normal distribution $N(0,1)$. If $\alpha(t) = \int_{\mathbb{R}} e^{\frac{-x^2}{2}} \cos(tx)\ dx$ and $\beta(t) = \int_{\mathbb{R}} e^{\frac{-x^2}{2}} \sin(tx)\ dx$, then

$$\varphi_\mu(t) = (\alpha(t) + i\beta(t))\frac{1}{\sqrt{2\pi}}.$$

It is trivial that $\beta(t) = 0$. On the other hand, observe that

$$\left|\frac{\partial f}{\partial t}\right| = \left|-x\cos(tx)e^{\frac{-x^2}{2}}\right| \leq |x|e^{\frac{-x^2}{2}}$$

and also $\int_{\mathbb{R}} |x|e^{\frac{-x^2}{2}}\ dx < \infty$. Note that $\alpha(t)$ is differentiable and applying Integration by parts, then

$$\begin{aligned}\alpha'(t) &= -\int_{\mathbb{R}} xe^{\frac{-x^2}{2}} \sin(tx)\ dx \\ &= \left[e^{\frac{-x^2}{2}} \sin(tx)\right]_{-\infty}^{+\infty} - \int_{\mathbb{R}} e^{\frac{-x^2}{2}} t\cos(tx)\ dx = -t\alpha(t).\end{aligned}$$

Since $\alpha(0) = \sqrt{2\pi}$, then $\alpha(t) = \sqrt{2\pi}e^{\frac{-t^2}{2}}$. Finally, $\varphi(t) = e^{\frac{-t^2}{2}}$, as we desired.

2) Other interesting examples can be found in *([2], Chapter 9)*.

Chapter 3

Law of Large Numbers

Random events can be quite irregular and unpredictable. Nevertheless, stability may be guaranteed when the same event is repeated several times, depending on the hypothesis. In this chapter, we will study the behavior of the partial sum $S_n = X_1 + ... + X_n$ of a sequence of independent r.v $\{X_n, n \geq 1\}$. We will state different versions of the Weak and Strong Law of Large Numbers. The versions depend on the conditions of the r.v. $\{X_n, n \geq 1\}$. We will also study previous results such as Kolmogorov's Inequalities and Kronecker's Lemma, to complete all the proofs. For a further analysis of the applications of the Laws of Large Numbers, we suggest *([3], Chapter 10)*.

3.1 Weak Law of Large Numbers

The Weak Law of Large Numbers states that the average of many observations will eventually be the population mean since the sample size can be increased.

Proposition 3.1.1. *Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed random variables. Suppose $E(X_1) = \mu$, $E\left(X_1^2\right) < \infty$. Then,*

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{L^2} \mu. \tag{3.1.1}$$

Proof. Observe that

$$E\left[\left|\frac{S_n}{n} - \mu\right|^2\right] = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var\left(S_n\right) = \frac{1}{n} Var\left(X_1\right) \overset{n\to\infty}{\longrightarrow} 0,$$

where in the first equality we have used that

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} E\left(S_n\right) = E\left(X_1\right) = \mu,$$

and the other identities and basic properties of independent random variables. ∎

Corollary 3.1.2. *With the same hypothesis, then* $\frac{S_n}{n} \xrightarrow[n\to\infty]{P} \mu$.[1]

Now we may study another result for convergence in probability.

Theorem 3.1.3. *Let* $\{X_n, n \geq 1\}$ *be a sequence of independent random variables such that there exists a sequence* $\{b_n\}_{n\geq 1} \nearrow \infty$ *that satisfies*

(a) $\sum_{i=1}^{n} P\{|X_i| > b_n\} \to 0$*, as* $n \to \infty$

(b) $\sum_{i=1}^{n} \frac{1}{b_n^2} E\left[|X_i|^2 \mathbb{1}_{\{|X_i| \leq b_n\}}\right] \to 0$*, as* $n \to \infty$.

Then $\frac{S_n - a_n}{b_n} \xrightarrow{P} 0$*, as* $n \to \infty$*, where* $a_n = \sum_{n=1}^{\infty} E\left[X_i \mathbb{1}_{\{|X_i| \leq b_n\}}\right]$.

Proof. Define the sequence $Y_{n_j} := X_j \mathbb{1}_{\{|X_j| \leq b_n\}}$ with $1 \leq j \leq n$. Since $\{X_n, n \geq 1\}$ are independent r.v., then $\{Y_{n_j}, 1 \leq j \leq n\}$ are too. Consider $T_n := \sum_{j=1}^{n} Y_{n_j}$. First, we want to see that if *(a)* is true, then $P\{S_n \neq T_n\} \underset{n\to\infty}{\to} 0$. Note that

$$\begin{aligned} P\{S_n \neq T_n\} &\leq P\left\{Y_{n_j} \neq X_j, \text{ for some } j \in \{1, ..., n\}\right\} \\ &\leq \sum_{j=1}^{n} P\left\{Y_{n_j} \neq X_j\right\} = \sum_{j=1}^{n} P\{|X_j| > b_n\} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Secondly, we see that $\frac{T_n - a_n}{b_n} \xrightarrow[n\to\infty]{\text{P}} 0$. By Chebyshev's Inequality, for all $\epsilon > 0$,

$$\begin{aligned} P\left\{\left|\frac{T_n - a_n}{b_n}\right| > \epsilon\right\} &\leq \frac{1}{\epsilon^2 b_n^2} E\left[|T_n - a_n|^2\right] = \frac{1}{\epsilon^2 b_n^2} Var(T_n) \\ &\leq \frac{1}{\epsilon^2 b_n^2} \sum_{j=1}^{n} Var(Y_{n_j}) \leq \frac{1}{\epsilon^2 b_n^2} \sum_{j=1}^{n} E\left(Y_{n_j}^2\right) \longrightarrow 0, \ \forall \epsilon > 0, \end{aligned}$$

where $E\left[|T_n - a_n|^2\right] = E\left[|T_n - E(T_n)|^2\right] = Var(T_n)$ and $E\left(Y_{n_j}^2\right) \to 0$, using *(b)*. Lastly, we show that if $P\{S_n \neq T_n\} \xrightarrow[n\to\infty]{} 0$ and $\frac{T_n - a_n}{b_n} \xrightarrow{\text{P}} 0$, then $\frac{S_n - a_n}{b_n} \xrightarrow{\text{P}} 0$. Notice that

$$\begin{aligned} P\left\{\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right\} &= P\left\{\left|\frac{S_n - a_n}{b_n}\right| > \epsilon, S_n = T_n\right\} + P\left\{\left|\frac{S_n - a_n}{b_n}\right| > \epsilon, S_n \neq T_n\right\} \\ &\leq P\left\{\left|\frac{T_n - a_n}{b_n}\right| > \epsilon\right\} + P\{S_n \neq T_n\} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

∎

[1] Remember that the L^2-norm implies convergence in probability *(Chapter 1)*.

3.2 Kolmogorov's Inequalities

Kolmogorov's Inequalities provide a bound on the probability that the partial sum S_n surpasses some specified bound. We will see two Kolmogorov's Inequalities, which will be necessary to prove Kolmogorov's Three-Series Theorem.[2]

Proposition 3.2.1. ***(First Kolmogorov's Inequality).*** *Let $\{X_n, n \geq 1\}$ be a sequence of independent r.v., centered and with finite second-order moment, for all $n \geq 1$. Then,*

$$P\left\{\max_{1\leq j\leq n} |S_j| > \epsilon\right\} \leq \frac{\sigma^2(S_n)}{\epsilon^2}, \ \forall \epsilon > 0. \tag{3.2.1}$$

Proof. If $n = 1$, the statement is exactly Chebyshev's Inequality.

Fix $n > 1$ and $\epsilon > 0$, then $M_0 = \Omega$, $M_k = \max_{1\leq j\leq n} |S_j| \leq \epsilon$, for all $n \geq 1$. Observe that $M_n \subset M_{n-1} \subset ... \subset M_1 \subset M_0$. Consider the disjoint sets $A_k = M_{k-1} - M_k = \{|S_j| \leq \epsilon, j = 1, ..., k-1, |S_j| > \epsilon\}$. Therefore, define $A = \dot{\bigcup}_{k=1}^{n} A_k = \{\max_{1\leq j\leq n} |S_j| > \epsilon\}$. We want to prove that $P(A) \leq \frac{\sigma^2(S_n)}{\epsilon^2}$, or, equivalently, $\sigma^2(S_n) \geq \epsilon^2 P(A)$. Since X_n are centered, then S_n are too, and using the definition of expectation,

$$\begin{aligned}\sigma^2(S_n) = E(S_n^2) &\geq \int_A (S_n)^2\, dP = \sum_{k=1}^{n} \int_{A_k} (S_n)^2\, dP \\ &= \sum_{k=1}^{n} \int_{A_k} (S_n - S_k + S_k)^2\, dP = \sum_{k=1}^{n} \int_{A_k} [(S_n - S_k) + S_k]^2\, dP.\end{aligned} \tag{3.2.2}$$

Solving the square, then $[(S_n - S_k) + S_k]^2 = (S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2$. Observe that the first and the third values are always positive and the second one satisfies the following

$$\int_{A_k} 2(S_n - S_k)S_k\, dP = 2E[\mathbb{1}_{A_k} S_k(S_n - S_k)] = 2E[\mathbb{1}_{A_k} S_k]E[S_n - S_k] = 0,$$

using that $\mathbb{1}_{A_k} S_k$ and $S_n - S_k$ are independent and also $E[S_n - S_k] = 0$ since $S_n - S_k$ are centered. Plugging it into (3.2.2) and omitting the first value, then

$$\sum_{k=1}^{n} \int_{A_k} [(S_n - S_k) + S_k]^2\, dP \geq \sum_{k=1}^{n} \int_{A_k} S_k^2\, dP \geq \epsilon^2 \sum_{k=1}^{n} P(A_k) = \epsilon^2 P(A),$$

where it has been used that S_k^2 is greater than ϵ^2 in A_k and the definition of A. Finally, $\sigma^2(S_n) \geq \epsilon^2 P(A)$, as we desired. ∎

[2] Remember the notation $\sigma^2(X) := Var(X)$.

Proposition 3.2.2. ***(Second Kolmogorov's Inequality).*** *Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables such that $E(|X_n|) < \infty, \forall n \geq 1$ and there exists a set $A > 0$ such that $|X_n - E[X_n]| \leq A$ a.s., $\forall n \geq 1$. Then,*

$$P\left\{\max_{1\leq j\leq n} |S_j| \leq \epsilon\right\} \leq \frac{(4\epsilon + 2A)^2}{\sigma^2(S_n)}. \tag{3.2.3}$$

Proof. Let $M_0 = \Omega$, $M_k = \{\max_{1\leq j\leq k} |S_j| \leq \epsilon\}$, for every $k \geq 1$. Note that $M_n \subset M_{n-1} \subset ... \subset M_1 \subset M_0$. Consider also the disjoint sets $A_k = M_{k-1} - M_k = \{|S_j| \leq \epsilon, j = 1, ..., k-1, |S_j| > \epsilon\}$. Notice $M_{k+1} = M_k \setminus A_{k+1}$. Therefore, define $A := \dot{\bigcup}_{k=1}^n A_k = \{\max_{1\leq j\leq n} |S_j| > \epsilon\} = \Omega \setminus M_n$. Let $X'_j = X_j - E(X_j)$, for every $j \geq 1$, which are centered and $|X'_j| = |X_j - E(X_j)| \leq A$.[3] Consider the partial sums $S'_0 = 0, S'_j = \sum_{k=1}^j X'_k$. Finally, $a_0 = 0$, $a_k = \frac{1}{P(M_k)} \int_{M_k} S'_k \, dP$.[4] We want to prove $P(M_n) \leq \frac{(4\epsilon+2A)^2}{\sigma^2(S_n)}$, equivalently, $P(M_n)\sigma^2(S_n) \leq (4\epsilon + 2A)^2$. Consider:

$$I := \int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2 \, dP = \int_{M_k \setminus A_{k+1}} (S'_k - a_k + a_k - a_{k+1} + X'_{k+1})^2 \, dP = I_1 - I_2, \tag{3.2.4}$$

where

$$I_1 := \int_{M_k} (S'_k - a_k + a_k - a_{k+1} + X'_{k+1})^2 \, dP,$$
$$I_2 := \int_{A_{k+1}} (S'_k - a_k + a_k - a_{k+1} + X'_{k+1})^2 \, dP.$$

The procedure will be to find a lower bound of I by finding a lower bound of I_1 and an upper bound of I_2. Afterwards, we will arrive at the desired result by operating. We will begin with an upper bound for I_2. Notice the next inequalities,

$$\begin{aligned} |S'_k - a_k| &= \left| S_k - E(S_k) - \frac{1}{P(M_k)} \int_{M_k} (S_k - E(S_k)) \, dP \right| \\ &\leq \left| S_k - E(S_k) - \frac{1}{P(M_k)} P(M_k)(\epsilon - E(S_k)) \right| \leq |S_k| + \epsilon \end{aligned}$$

and

$$\begin{aligned} &|a_k - a_{k+1}| = \left| \frac{1}{P(M_k)} \int_{M_k} (S_k - E(S_k)) \, dP - \frac{1}{P(M_{k+1})} \int_{M_{k+1}} (S_{k+1} - E(S_{k+1})) \, dP \right| \\ &= \left| \frac{1}{P(M_k)} \int_{M_k} (S_k - E(S_k))dP - \frac{1}{P(M_{k+1})} \int_{M_{k+1}} (S_k - E(S_k) + X_{k+1} - E(X_{k+1}))dP \right| \\ &\leq \left| \frac{1}{P(M_k)} \int_{M_k} (\epsilon - E(S_k)) \, dP - \frac{1}{P(M_{k+1})} \int_{M_{k+1}} (\epsilon - E(S_k) + A) \, dP \right| \\ &\leq \epsilon + \epsilon + A = 2\epsilon + A. \end{aligned}$$

[3]Since $|X_n - E[X_n]| \leq A$, then also $E[|X_n - E[X_n]|^2] \leq A^2$, so X'_n has bounded second-order moment and therefore it is not necessary to remark it.

[4]Note that it is sufficient that $P(M_k) > 0$, otherwise the result is trivial.

Plugging these into I_2, then

$$\begin{aligned} I_2 &\leq \int_{A_{k+1}} (|S'_k - a_k| + |a_k - a_{k+1}| + |X'_{k+1}|)^2 \, dP \\ &\leq \int_{A_{k+1}} (|S_k| + \epsilon + 2\epsilon + A + |X'_{k+1}|)^2 \, dP \leq (4\epsilon + 2A)^2 P(A_{k+1}), \end{aligned}$$

where in the last step it has been used that $|S_k| \leq \epsilon$ and $|X'_{k+1}| \leq \epsilon$, by definition, and also $\int_{A_{k+1}} dP = P(A_{k+1})$. We will find now a lower bound for I_1. Developing the notable product,

$$\begin{aligned} &([S'_k - a_k] + [a_k - a_{k+1}] + X'_{k+1})^2 \\ &\qquad = (S'_k - a_k)^2 + (a_k + a_{k+1})^2 + (X'_{k+1})^2 + 2(S'_k - a_k)(a_k - a_{k+1}) \\ &\qquad\quad + 2(S'_k - a_k)X'_{k+1} + 2(a_k - a_{k+1})X'_{k+1} \end{aligned}$$

and taking expectations, observe that

$$\begin{aligned} &E[\mathbb{1}_{M_k}(S'_k - a_k)(a_k - a_{k+1})] = (a_k - a_{k+1}) \int_{M_k} S'_k \, dP - P(M_k) \int_{M_k} S'_k \, dP = 0 \,; \\ &E[\mathbb{1}_{M_k}(S'_k - a_k)X'_{k+1}] = 0 \,;\; E[\mathbb{1}_{M_k}(a_k - a_{k+1})X'_{k+1}] = 0 \,; \end{aligned}$$

since X'_k are independent and centered, for all $k \geq 1$, by hypothesis. Substituting all these in I_1 and using that $(a_k - a_{k+1})^2$ is positive, then

$$\begin{aligned} I_1 &\geq \int_{M_k} (S'_k - a_k)^2 \, dP + \int_{M_k} (X'_{k+1})^2 \, dP \\ &= \int_{M_k} (S'_k - a_k)^2 \, dP + P(M_k)\sigma^2(X_{k+1}), \end{aligned}$$

using that X'_k are independent and centered and therefore $E[(X'_k)^2] = \sigma^2(X_{k+1})$. Plugging these inequalities in (3.2.4) and since $-I_2 \geq -(4\epsilon + 2A)^2 P(A_{k+1})$, then

$$\begin{aligned} I &= \int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2 \, dP = I_1 + (-I_2) \\ &\geq \left(\int_{M_k} (S'_k - a_k)^2 \, dP \right) + P(M_k)\sigma^2(X_{k+1}) - (4\epsilon + 2A)^2 P(A_{k+1}). \end{aligned}$$

Taking sums, consider the series

$$\begin{aligned} Y_n &:= \sum_{k=0}^{n-1} \int_{M_{k+1}} (S'_{k+1} - a_{k+1})^2 \, dP - \int_{M_k} (S'_k - a_k)^2 \, dP \\ &\geq \sum_{k=0}^{n-1} P(M_k)\sigma^2(X_{k+1}) - (4\epsilon + 2A)^2 P(A_{k+1}) \\ &\geq P(M_n)\sigma^2(S_n) - (4\epsilon + 2A)^2 P(A), \end{aligned} \tag{3.2.5}$$

using in the last step that $P(M_n) \le P(M_k),\ \forall k \in \{1, ..., n-1\}$ (by construction) and the definition of A and S_n. On the other hand, Y_n is a telescoping series, therefore $Y_n = \int_{M_n} (S_n' - a_n)^2\, dP$, since $S_0' = 0, a_0 = 0$. Operating, then

$$\begin{aligned} Y_n &= \int_{M_n} \left[S_n - E(S_n) - \frac{1}{P(M_n)} \int_{M_n} (S_n - E(S_n)) \right]^2 dP \\ &\le \int_{M_n} (|S_n| + \epsilon)^2\, dP \le 4\epsilon^2 P(M_n), \end{aligned} \tag{3.2.6}$$

using the bound found for I_2. Finally, by (3.2.5) and (3.2.6),

$$4\epsilon^2 P(M_n) \ge P(M_n)\sigma^2(S_n) - (4\epsilon + 2A)^2 P(A),$$

equivalently and using that $A = \Omega \setminus M_n$,

$$\begin{aligned} P(M_n)\sigma^2(S_n) &\le 4\epsilon^2 P(M_n) + (4\epsilon + 2A)^2 P(\Omega \setminus M_n) \\ &\le (4\epsilon + 2A)^2 [P(M_n) + P(\Omega \setminus M_n)] = (4\epsilon + 2A)^2. \end{aligned}$$

■

Finally, we can prove Kolmogorov's Three-Series Theorem, which provides a condition for the a.s. convergence of an infinite series of r.v.

Theorem 3.2.3. ***(Kolmogorov's Three-Series Theorem).*** *Let $\{X_n, n \ge 1\}$ be a sequence of independent r.v. Let the set $A > 0$ and the sequence $Y_n = X_n \mathbb{1}_{\{|X_n| \le A\}}, \forall n \ge 1$. Then, the series $\sum_{n=1}^{\infty} X_n$ converges a.s. if and only if the next three series also converge:*

(a) $\sum_{n=1}^{\infty} P\{|X_n| > A\}$,

(b) $\sum_{n=1}^{\infty} E(Y_n)$,

(c) $\sum_{n=1}^{\infty} \sigma^2(Y_n)$.

Proof. Suppose (a), (b), (c). To see the a.s. convergence of $\sum_{n=1}^{\infty} X_n$, we will prove the a.s. convergence of $\sum_{n=1}^{\infty}(Y_n - E(Y_n))$. Consider the set

$$N := \bigcap_{m=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \left\{ \left| \sum_{i=n_0+1}^{n} (Y_i - E(Y_i)) \right| \le \frac{1}{m} \right\},$$

i.e. for any $m \ge 1$, there is $n_0 \ge 1$ s.t. for all $n \ge n_0$, $\left|\sum_{i=n_0+1}^{n} (Y_i - E(Y_i))\right| \le \frac{1}{m}$. Consider $M_n := \left\{\max_{n_0 \le k \le n} \left|\sum_{i=n_0+1}^{n} (Y_i - E(Y_i))\right|\right\}$, then observe the inclusion:

$$\bigcap_{n=n_0}^{\infty} \left\{ M_n \le \frac{1}{m} \right\} \subseteq \bigcap_{n=n_0}^{\infty} \left\{ \left| \sum_{i=n_0+1}^{n} (Y_i - E(Y_i)) \right| \le \frac{1}{m} \right\}.$$

Then taking probabilities,

$$P\left\{\bigcap_{n=n_0}^{\infty}\left(M_n \le \frac{1}{m}\right)\right\} \le P\left\{\bigcap_{n=n_0}^{\infty}\left(\left|\sum_{i=n_0+1}^{n}(Y_i - E(Y_i))\right| \le \frac{1}{m}\right)\right\}.$$

By the First Kolmogorov's Inequality,

$$P\left\{M_n > \frac{1}{m}\right\} \le \frac{\sigma^2(\sum_{i=n_0+1}^{n}(Y_i - E(Y_i)))}{\frac{1}{m^2}}.$$

Therefore, considering the complement set,

$$\begin{aligned} P\left\{M_n \le \frac{1}{m}\right\} &\ge 1 - \sigma^2\left(\sum_{i=n_0+1}^{k}(Y_i - E(Y_i))\right)m^2 \\ &= 1 - m^2\sum_{i=n_0+1}^{k}\sigma^2(Y_i), \end{aligned} \tag{3.2.7}$$

for all $m \ge 1$, since Y_n are independent and the variance of a constant is zero. Taking limits and by (3.2.7), then

$$\begin{aligned} P(N) &= P\left\{\bigcap_{m=1}^{\infty}\bigcup_{n_0=1}^{\infty}\bigcap_{n=n_0}^{\infty}\left(\left|\sum_{i=n_0+1}^{n}(Y_i - E(Y_i))\right| \le \frac{1}{m}\right)\right\} \\ &= \lim_{m\to\infty}\lim_{n_0\to\infty} P\left\{\bigcap_{n=n_0}^{\infty}\left(\left|\sum_{i=n_0+1}^{n}(Y_i - E(Y_i))\right| \le \frac{1}{m}\right)\right\} \\ &\ge \lim_{m\to\infty}\lim_{n_0\to\infty}\left(1 - m^2\sum_{i=n_0+1}^{k}\sigma^2(Y_i)\right) = 1, \end{aligned}$$

where $\sum_{i=n_0+1}^{k}\sigma^2(Y_i) \to 0$, as $n_0 \to \infty$, since this series is convergent by *hypothesis (c)*.[5] As $P(B) \ge 1$, then $P(B) = 1$,. Finally, the series $\sum_{n=1}^{\infty}(Y_n - E(Y_n))$ converges a.s. Since $\sum_{n=1}^{\infty}E(Y_n)$ converges by *(b)*, then $\sum_{n=1}^{\infty}Y_n$ converges a.s. Since $\sum_{n=1}^{\infty}P\{|X_n| > A\}$ converges by *(a)* and by definition of Y_n, then $\sum_{n=1}^{\infty}P\{X_n \ne Y_n\}$ converges. By the First Borel-Cantelli's Lemma, $P\{\limsup_n\{X_n \ne Y_n\}\} = 0$, i.e. $\{X_n, n \ge 1\}, \{Y_n, n \ge 1\}$ differ in an infinite number of points. Since $\sum_{n=1}^{\infty}Y_n$ converges a.s., then finally $\sum_{n=1}^{\infty}X_n$ converges a.s. too.

Suppose that $\sum_{n=1}^{\infty}X_n$ converges a.s. First, we prove *(a)* by contradiction. Let $A > 0$, suppose $\sum_{n=1}^{\infty}P\{|X_n| > A\} = \infty$. By the Second Borel-Cantelli's Lemma, $P\{\limsup_n\{|X_n| > A\}\} = 1$. Then $\sum_{n=1}^{\infty}X_n$ does not converge, which contradicts the hypothesis. Therefore $\sum_{n=1}^{\infty}P\{|X_n| > A\} = \infty$ converges and *(a)* is correct.

[5]Note the limit is independent of m.

Now we prove *(c)* by contradiction. Following the previous argument, $\sum_{n=1}^{\infty} Y_n$ is convergent. By the Second Kolmogorov's Inequality,

$$P\left\{\max_{n_0\leq k\leq n}\left|\sum_{i=n_0+1}^{k} Y_i\right|\leq 1\right\}\leq \frac{(4+4A)^2}{\sum_{i=n_0}^{n}\sigma^2(Y_i)},$$

using that $|Y_i - E(Y_i)| < 2A$, by hypothesis and $\sigma^2\left(\sum_{i=n_0}^{n} Y_i\right) = \sum_{i=n_0}^{n}\sigma^2(Y_i)$, by independence of Y_n. Suppose $\sum_{i=n_0}^{n}\sigma^2(Y_i)$ does not converge, then the previous probability tends to 0, as $n\to\infty$, for all n_0. Then the series $\sum_{n=1}^{\infty} Y_n$ is not a Cauchy series, which is a contradiction with the convergence of $\sum_{i=n_0}^{n}\sigma^2(Y_i)$. Finally, $\sum_{n=1}^{\infty}\sigma^2(Y_n)$ converges and *(c)* is proved. Now we prove *(b)*. By $E(Y_i - E(Y_i)) = 0$, then $\sum_{n=1}^{\infty}\sigma^2(Y_i - E(Y_i)) = \sum_{n=1}^{\infty}\sigma^2(Y_i)$ which is convergent as proved. By definition of variance and convergence, $\sum_{n=1}^{\infty} Y_i - E(Y_i)$ is convergent a.s. Since $\sum_{n=1}^{\infty} Y_i$ is convergent, then finally $\sum_{n=1}^{\infty}\sigma^2(Y_n)$ converges and *(b)* is proved. [6] ∎

Another curious result for almost sure convergence is the following.

Proposition 3.2.4. *Let $\{X_n, n\geq 1\}$ be a sequence of independent and centered r.v, with finite second-order moment. If $\sum_{n=1}^{\infty}\sigma^2(X_n)<\infty$, then $\sum_{n=1}^{\infty} X_n$ is a.s. convergent.*

Proof. Applying the First Kolmogorov's Inequality to $X_{m+1}, ..., X_{m+k}$, then

$$\begin{aligned}P\{\sup_{j\geq 1}|S_{m+j}-S_m|>\epsilon\} &= \lim_{k\to\infty} P\{\sup_{1\leq j\leq k}|S_{m+j}-S_m|>\epsilon\}\\ &\leq \lim_{k\to\infty}\frac{1}{\epsilon^2}\sum_{j=1}^{k}\sigma^2(X_{m+j}) = \frac{1}{\epsilon^2}\sum_{j=1}^{\infty}\sigma^2(X_{m+j}),\end{aligned}$$

where $\sum_{j=1}^{\infty}\sigma^2(X_{m+j}) \underset{m\to\infty}{\longrightarrow} 0$. Finally, using basic properties of convergence in probability and almost sure convergence, S_n is almost sure convergent. ∎

3.3 Kronecker's Lemma

Kronecker's Lemma is a result of the relationship between the convergence of infinite sums and the convergence of sequences.

Lemma 3.3.1. ***(Kronecker's Lemma).*** *Consider $\{x_n, n\geq 1\}$, $\{a_n, n\geq 1\}$ two sequences of real numbers such that $0 < a_n \nearrow \infty$. If $\sum_{n=1}^{\infty}\frac{x_n}{a_n}$ converges, then*

$$\lim_{n\to\infty}\frac{1}{a_n}\sum_{j=1}^{n} x_j = 0.$$

[6] Notice that the independence of $\{X_n, n\geq 1\}, \{Y_n, n\geq 1\}$ is used throughout the proof.

Proof. Define the sequence $a_0 = b_0 = 0$, $b_n = \sum_{j=1}^{n} \frac{x_j}{a_j}$, for all $n \geq 1$. Therefore, $x_n = a_n (b_n - b_{n-1})$, for all $n \geq 1$. Then,

$$\frac{1}{a_n}\sum_{j=1}^{n} x_j = \frac{1}{a_n}\sum_{j=1}^{n} a_j(b_j - b_{j-1}) = b_n - \frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j).$$

Define $b_\infty = \lim_n b_n$. We want to prove that

$$\lim_{n\to\infty} \frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) = b_\infty.$$

Observe that $\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j) = 1$. Therefore, for all $n > m$,

$$\left|\frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) - b_\infty\right| = \left|\frac{1}{a_n}\sum_{j=0}^{n-1}(b_j - b_\infty)(a_{j+1} - a_j)\right|$$
$$\leq \frac{1}{a_n}\left|\sum_{j=0}^{m-1}(b_j - b_\infty)(a_{j+1} - a_j)\right| + \frac{1}{a_n}\left|\sum_{j=m}^{n-1}(b_j - b_\infty)(a_{j+1} - a_j)\right|.$$

Fix $\epsilon > 0$, consider m such that $|b_j - b_\infty| < \epsilon$, for all $j \geq m$. Since $a_{j+1} - a_j \geq 0$, then the second part of the previous sum is bounded by ϵ for this m. If ϵ, m, then

$$\limsup_{n\to\infty}\left|\frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) - b_\infty\right| \leq \epsilon,$$

using that $\frac{1}{a_n}\left|\sum_{j=0}^{m-1}(b_j - b_\infty)(a_{j+1} - a_j)\right| \underset{n\to\infty}{\longrightarrow} 0$. Taking $\epsilon = 0$, we are finally done. ∎

3.4 Strong Law of Large Numbers

To close this chapter, we will see three versions of the Strong Law of Large Numbers, using Kronecker's Lemma and other previous results.

Theorem 3.4.1. *Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables, centered and with finite second-order moment. Let $\{a_n, n \geq 1\}$ be a sequence of real numbers, such that $0 < a_n \nearrow \infty$. If $\sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{a_n^2} < \infty$, then*

$$\frac{S_n}{a_n} \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0. \tag{3.4.1}$$

Proof. Observe that $\sum_{n=1}^{\infty} \sigma^2\left(\frac{X_n}{a_n}\right) = \sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{a_n^2} < \infty$, therefore $\sum_{n=1}^{\infty} \frac{X_n}{a_n}$ is a.s. convergent, using *Proposition 3.2.4* applied to $\frac{X_n}{a_n}$. By Kronecker's Lemma *(Lemma 3.3.1)*, then $\lim_{n\to\infty} \frac{1}{a_n}\sum_{j=1}^{n} X_j = 0$, *a.s.* ∎

Theorem 3.4.2. *Let $\{X_n, n \geq 1\}$ be a sequence of uncorrelated random variables with second-order moment bounded by a constant C, independent of n. Then,*

$$\frac{S_n - E(S_n)}{n} \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof. Suppose that $\{X_n, n \geq 1\}$ is centered, then $\{S_n, n \geq 1\}$ is too. First, we want to prove that $\frac{S_{n^2}}{n^2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. Fix $\epsilon > 0$, by the Chebyshev's Inequality, then

$$P\left\{\left|\frac{S_{n^2}}{n^2}\right| > \epsilon\right\} \leq \frac{1}{n^4\epsilon^2}E[S_{n^2}^2] = \frac{1}{n^4\epsilon^2}Var(S_{n^2}) = \frac{1}{n^4\epsilon^2}\sum_{i=1}^{n^2} Var(X_i) \leq \frac{1}{n^4\epsilon}n^2C = \frac{C}{n^2\epsilon},$$

where $Var(S_{n^2}) = E(S_{n^2}^2)$, since S_{n^2} are centered and $Var(S_{n^2}) = \sum_{i=1}^{n^2} Var(X_i)$, as $\{X_n, n \geq 1\}$ are uncorrelated and centered. Taking series, then

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{S_{n^2}}{n^2}\right| > \epsilon\right\} \leq \sum_{n=1}^{\infty} \frac{C}{n^2\epsilon} < \infty.$$

By the First Borel-Cantelli's Lemma, $P\left\{\limsup_n \left\{\left|\frac{S_{n^2}}{n^2}\right| > \epsilon\right\}\right\} = 0$, $\forall\epsilon > 0$, i.e. $\left|\frac{S_{n^2}}{n^2}\right| \xrightarrow{\text{a.s.}} 0$ and then $\frac{S_{n^2}}{n^2} \xrightarrow{\text{a.s.}} 0$. Define $D_n := \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$, we want to prove that $\frac{D_n}{n^2} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. By the Chebyshev's Inequality, we have

$$P\left\{\frac{D_n}{n^2} > \epsilon\right\} \leq \frac{1}{n^4\epsilon^2}E\left[|D_n|^2\right] \leq \frac{1}{n^4\epsilon^2}(2n)^2C = \frac{4C}{n^2\epsilon^2},$$

where the next result has been used in the last inequality

$$E\left(|D_n|^2\right) \leq E\left[\sum_{k=n^2}^{(n+1)^2-1} |S_k - S_{n^2}|\right] = \sum_{k=n^2}^{n^2+2n} E\left[\left(\sum_{l=n^2+1}^{k} X_l\right)^2\right]$$
$$= \sum_{k=n^2}^{n^2+2n}\sum_{l=n^2+1}^{k} E\left(X_l^2\right) \leq \sum_{k=n^2}^{n^2+2n}\sum_{l=n^2+1}^{n^2+2n} C < (2n)^2C < \infty.$$

Then by the First Borel-Cantelli's Lemma, $P\left\{\limsup_n \left\{\frac{D_n}{n^2} > \epsilon\right\}\right\} = 0$, $\forall\epsilon > 0$, i.e. $\frac{D_n}{n^2} \xrightarrow{\text{a.s.}} 0$. Since $|S_k| = |S_k - S_{n^2} + S_{n^2}| \leq |S_k - S_{n^2}| + |S_{n^2}| \leq D_n + |S_{n^2}|$, notice there exists n, such that for $k \in \left[n^2, (n+1)^2\right)$,

$$\left|\frac{S_k}{k}\right| \leq \frac{|S_{n^2}| + D_n}{n^2}.$$

Finally, taking limits,

$$\lim_{n\to\infty} \frac{|S_k|}{k} \leq \lim_{n\to\infty} \frac{|S_{n^2}| + D_n}{n^2} = 0,$$

almost surely. Equivalently, $\frac{S_n - E(S_n)}{n} \xrightarrow[n\to\infty]{\text{a.s.}} 0$, as we wanted. ∎

We need one last previous result. The interpretation is that Y has finite first-order moment if and only if the series $\sum_{n=1}^{\infty} P\{|Y| \geq n\}$ is convergent.

Lemma 3.4.3. *Let Y be a random variable, then it satisfies*

$$\sum_{n=1}^{\infty} P\{|Y| \geq n\} \leq E(|Y|) \leq 1 + \sum_{n=1}^{\infty} P\{|Y| \geq n\}. \tag{3.4.2}$$

Proof. We will prove both inequalities. Firstly, observe that

$$\sum_{n=1}^{\infty} P\{|Y| \geq n\} = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} P\{k \leq |Y| < k+1\} = \sum_{k=1}^{\infty} kP\{k \leq Y < k+1\}$$
$$= \sum_{k=0}^{\infty} kP\{k \leq |Y| < k+1\} \leq \sum_{k=0}^{\infty} \int_{\{k \leq |Y| < k+1\}} |Y| \, dP = E(|Y|),$$

as we wanted. On the other hand, the second inequality is a consequence of

$$E(|Y|) = \sum_{k=0}^{\infty} \int_{\{k \leq |Y| < k+1\}} |Y| \, dP \leq \sum_{n=1}^{\infty} P\{|Y| \geq n\} + 1.$$

■

Finally, we can prove Kolmogorov's Strong Law of Large Numbers, where the only hypothesis is that $\{X_n, n \geq 1\}$ have finite first-order moment.[7]

Theorem 3.4.4. ***(Kolmogorov's Strong Law of Large Numbers).*** *Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. random variables. Then*

1) *If $E(|X_1|) < \infty$, then $\lim_{n\to\infty} \frac{S_n}{n} = E(X_1)$, a.s.*
2) *If $E(|X_1|) = \infty$, then $\limsup_n \frac{|S_n|}{n} = +\infty$, a.s.*

Proof. The idea of this proof is to truncate the X_n with zero values that are not in the interval $(-n, n)$ and then apply the *Theorem 3.4.1*. We will first prove *1)*. Define the sequence $Y_n = X_n \mathbb{1}_{\{|X_n| < n\}}$. By the previous *Lemma 3.4.3.*, since the X_n are i.i.d.,

$$\sum_{n=1}^{\infty} P\{X_n \neq Y_n\} \leq \sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \sum_{n=1}^{\infty} P\{|X_1| \geq n\} \leq E(|X_1|) < \infty,$$

By the First Borel-Cantelli's Lemma, $P\{\limsup_n \{X_n \neq Y_n\}\} = 0$. Define the previous set as $A := \liminf_n \{X_n = Y_n\}$. By definition, if $\omega \in A$, there exists a $n_0(\omega)$ s.t. for all $n \geq n_0$, $\omega \in \{X_n = Y_n\}$, i.e. $X_n(\omega) = Y_n(\omega)$. Then, on the set,

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = \frac{1}{n}\sum_{i=1}^{n} Y_i(\omega) + \frac{1}{n}\sum_{i=1}^{n_0(\omega)} (X_i - Y_i)(\omega).$$

[7] Since the hypothesis is weaker, this is a stronger result than the previous ones.

Therefore, it is enough to prove that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\text{a.s.}} E(X_1). \tag{3.4.3}$$

Equivalently, to prove (3.4.3) it is enough to show that

$$\frac{1}{n}\sum_{i=1}^{n} (Y_i - E(Y_i)) \xrightarrow[n\to\infty]{\text{a.s.}} 0, \tag{3.4.4}$$

since

$$E(Y_n) = E[X_n \mathbb{1}_{\{|X_n|<n\}}] = E[X_1 \mathbb{1}_{\{|X_1|<n\}}] \xrightarrow[n\to\infty]{} E(X_1),$$

implies that

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_i) \xrightarrow[n\to\infty]{} E(X_1).$$

By *Theorem 3.4.1* applied to the sequence $\{Y_n - E(Y_n), n \geq 1\}$,[8] (3.4.4) is true if the r.v. satisfy that $\sum_{n=1}^{\infty} \frac{\sigma^2(Y_n)}{n} < \infty$. Using the hypothesis on X_n and Y_n, notice that

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\sigma^2(Y_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} E\left[X_n^2 \mathbb{1}_{\{|X_n|<n\}}\right] = \sum_{n=1}^{\infty} \frac{1}{n^2} E\left[X_1^2 \mathbb{1}_{\{|X_1|<n\}}\right] \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E\left[X_1^2 \mathbb{1}_{\{k-1\leq|X_1|<k\}}\right] = \sum_{k=1}^{\infty} E\left[X_1^2 \mathbb{1}_{\{k-1\leq|X_1|<k\}}\right] \sum_{n=k}^{\infty} \frac{1}{n^2} \\
&\leq \sum_{k=1}^{\infty} kE\left[|X_1| \mathbb{1}_{\{k-1\leq|X_1|<k\}}\right] \frac{2}{k} \leq 2E\left(|X_1|\right) < \infty,
\end{aligned}$$

where in the second-to-last inequality it has been used the next bound

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \int_k^{\infty} \frac{1}{x^2}\, dx = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k}.$$

Now (3.4.4) is proved and therefore the result *1)*. Finally, we prove *2)*. For any constant $K > 0$,

$$\sum_{n=1}^{\infty} P\{|X_n| \geq Kn\} = \sum_{n=1}^{\infty} P\left\{\frac{|X_1|}{K} \geq n\right\} \geq \frac{E(|X_1|)}{K} - 1 = \infty.$$

Define the previous set $B_K := \limsup_n \{|X_n| \geq Kn\}$, for any K, then by the Second Borel-Cantelli's Lemma, $P(B_K) = 1$. Consider the set $B := \bigcap_{K=1}^{\infty} B_K$, then $P(B) = 1$. By definition of B, observe that

$$\frac{|S_n|}{n} + \frac{|S_{n-1}|}{n-1} > \frac{|S_n| + |S_{n-1}|}{n} \geq \frac{|S_n - S_{n-1}|}{n} = \frac{|X_n|}{n} \geq K,$$

for infinite n and for all $K \geq 1$, i.e. it is satisfied that $\frac{|S_n|}{n} \geq \frac{K}{2}$ for infinite n. Finally, $\limsup_n \frac{|S_n|}{n} = \infty$ in B, as we wanted to see. ∎

[8] $\{Y_n - E(Y_n), n \geq 1\}$ are independent, centered and have finite second-order moment.

Chapter 4

Weak convergence of probability measures

Let $\{\mu_n, n \geq 1\}$ be a sequence of probability measures in $\mathbb{R}$. The next question may arise: when is it possible to say that μ_n converges to a probability μ? The quick definition is $\lim_{n\to\infty} \mu_n(B) = \mu(B)$, for all $B \subset \mathcal{B}(\mathbb{R})$, but this is not sufficient in our studies. Consider a sequence $\{\mu_n, n \geq 1\}$ of discrete probabilities that converges to a continuous probability μ. This is not possible with the previous definition, for instance, if A is a countable set equal to the support of all μ_n, then $\mu_n(A) = 1$, for every n, but $\mu(A) = 0$, which is a contradiction with the statement.

The next mode of convergence differs conceptually from the previously studied ones since it is defined on a sequence of probability measures. It is the weakest type however, it is frequently used since it is used in the CLT. In this chapter, we will begin studying the Weak convergence of probabilities and some criteria for it, such as Paul-Lévy's Continuity Theorem. Finally, we will focus on convolutions of probability measures, which will be crucial for the next chapter. For further information regarding Weak convergence, we recommend *([4], Chapter 6)*. To review the concept of probability measures and measures check *([4], Sections 1.2 and 2.1)*.

4.1 Weak convergence of finite measures

The terms weak convergence, convergence in law and convergence in distribution are often used interchangeably. We will define these concepts and study their relation, as well as some similar results.

Definition 4.1.1. *Let $\{\mu_n, n \geq 1\}$, μ be probability measures in $\mathbb{R}$. This sequence* ***converges weakly*** *to μ if*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\ \mu_n(dx) = \int_{\mathbb{R}} f(x)\ \mu(dx), \tag{4.1.1}$$

for all $f \in \mathcal{C}_b(\mathbb{R})$.[1] *This is denoted* $\omega - lim_{n\to\infty}\mu_n = \mu$.

The limit of μ_n is unique, if exists. Note that if $f \equiv 1$, then $\mu_n(\mathbb{R}) \xrightarrow{n} \mu(\mathbb{R})$.

We will begin to study different **criterion of weak convergence**. The first one is known as **convergence in distribution** and is in terms of distribution functions.

Theorem 4.1.2. *A sequence of probability measures* $\{\mu_n, n \geq 1\}$ *converges weakly to a probability* μ *if and only if* $\lim_{n\to\infty} F_n(x) = F(x)$, *for every continuity point* $x \in \mathbb{R}$ *of* F, *where* F_n *and* F *are the distribution functions of* μ_n *and* μ, *respectively.*

Proof. Assume that $\omega - \lim_{n\to\infty} \mu_n = \mu$. Fix a continuity point x in F. Fix a real number $\epsilon > 0$, then consider the bounded and continuous functions

$$f_\epsilon^+(y) = \mathbb{1}_{(-\infty,x]}(y) + \left(1 - \frac{y-x}{\epsilon}\right)\mathbb{1}_{(x,x+y)}(y)$$

and

$$f_\epsilon^-(y) = \mathbb{1}_{(-\infty,x-\epsilon]}(y) + \frac{x-y}{\epsilon}\mathbb{1}_{(x-\epsilon,x)}(y).$$

Define $l := \liminf_{n\to\infty} F_n(x)$ and $L := \limsup_{n\to\infty} F_n(x)$. Hence, note that

$$\begin{aligned} F(x-\epsilon) &= \mu((-\infty, x-\epsilon]) \leq \int_{\mathbb{R}} f_\epsilon^- \, d\mu = \lim_{n\to\infty}\int_{\mathbb{R}} f_\epsilon^- \, d\mu_n \\ &\leq \liminf_n \mu_n((-\infty,x]) = l \leq L = \limsup_n \mu_n((-\infty,x]) \leq \lim_n \int_{\mathbb{R}} f_\epsilon^+ \, d\mu_n \\ &= \int_{\mathbb{R}} f_\epsilon^+ \, d\mu \leq \mu((-\infty, x+\epsilon]) = F(x+\epsilon), \end{aligned}$$

i.e. $F(x-\epsilon) \leq l \leq L \leq F(x+\epsilon)$, then taking $\epsilon \searrow 0$, $l = L = F(x)$. Equivalently, $\lim_{n\to\infty} F_n(x) = F(x)$, for every continuity point x of F, as we wanted.

Assume that $\lim_{n\to\infty} F_n(x) = F(x)$, for every continuity point x of F. Consider $f \in \mathcal{C}_b(\mathbb{R})$. Fix $\epsilon > 0$, there is a natural k such that $\mu\left(((-k,k]^c\right) < \epsilon$, since $\bigcap_{k=1}^\infty \left((-k,k]^c\right) = \emptyset$. Let $a,\ b$ be continuity points of F with $a \leq -k,\ b \geq k$, i.e. $[-k.k] \subseteq [a,b]$. Then $\mu\left((a,b]^c\right) < \epsilon$. Since f is uniformly continuous in $[a,b]$, there exists $\delta > 0$ such that: if $x,y \in [a,b]$ and $|x-y| \leq \delta$, then $|f(x) - f(y)| < \epsilon$. Decompose the interval $(a,b]$ in a finite number of $I_i = (a_i, b_i]$, $1 \leq i \leq r$, each with length lower or equal to δ and their endpoints are continuity points of F. Therefore, we will first decompose $(a,b]$ in intervals with lengths lower than $\frac{\delta}{2}$.

[1]Throughout this chapter, denote $\mathcal{C}_b(\mathbb{R})$ the set of bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$.

Then in each one, we will take a continuity point of F. Observe that

$$\begin{aligned}\left|\int_{\mathbb{R}} f\, d\mu_n - \int_{\mathbb{R}} f\, d\mu\right| &= \left|\int_{(a,b]^c} f\, d\mu_n + \int_{(a,b]} f\, d\mu_n - \int_{(a,b]^c} f\, d\mu - \int_{(a,b]} f\, d\mu\right| \\ &= \left|\int_{(a,b]^c} f\, d\mu_n + \sum_{i=1}^{r}\int_{I_i} f\, d\mu_n - \int_{(a,b]^c} f\, d\mu - \sum_{i=1}^{r}\int_{I_i} f\, d\mu\right| \\ &\leq \left|\int_{(a,b]^c} f\, d\mu_n - \int_{(a,b]^c} f\, d\mu\right| + \left|\sum_{i=1}^{r}\int_{I_i} f\, d\mu_n - \sum_{i=1}^{r}\int_{I_i} f\, d\mu\right| \\ &\leq \|f\|_\infty \left(\mu_n((a,b]^c) + \mu(a,b]^c)\right) + \sum_{i=1}^{r}\left|\int_{I_i} f\, d\mu_n - \int_{I_i} f\, d\mu\right|.\end{aligned} \tag{4.1.2}$$

By hypothesis,

$$\mu_n((a,b]^c) = F_n(a) + 1 - F_n(b) \underset{n\to\infty}{\longrightarrow} F(a) + 1 - F(b) = \mu((a,b]^c).$$

We want to bound the last term of (4.1.2). Since f is uniformly continuous, then

$$\begin{aligned}\left|\int_{I_i} f\, d\mu_n - \int_{I_i} f\, d\mu\right| &\leq \left|\int_{I_i}(f - f(a_i))\, d\mu_n\right| + |f(a_i)[\mu_n(I_i) - \mu(I_i)]| \\ &\quad + \left|\int_{I_i}(f - f(a_i))\, d\mu\right| \\ &\leq \epsilon(\mu_n(I_i) + \mu(I_i)) + |f(a_i)[\mu_n(I_i) - \mu(I_i)].\end{aligned} \tag{4.1.3}$$

Remember that $\lim_{n\to\infty} \mu_n(I_i) = \mu(I)$, for any $i \in \{1 \leq i \leq r\}$. Taking limits on (4.1.2) and (4.1.3) as $n \to \infty$ and using the previous results, then

$$\begin{aligned}\limsup_n \left|\int_{\mathbb{R}} f\, d\mu_n - \int_{\mathbb{R}} f\, d\mu\right| &\leq 2\|f\|_\infty \mu((a,b]^c)) + \sum_{i=1}^{r}(2\epsilon\mu(I_i) + 0) \\ &\leq \epsilon(2\|f\|_\infty + 1).\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, take ϵ close to 0 and finally, $\lim_n \int_{\mathbb{R}} f\, d\mu_n = \int_{\mathbb{R}} f\, d\mu$. ∎

4.2 Weak convergence of probabilities

In this section, we aim to prove two more criteria for weak convergence of probabilities. First, we will need to define the weak compactness and tightness of a family of probabilities and study some related results.

Definition 4.2.1. *Let $\mathcal{P}$ be a family of probability measures in $\mathbb{R}$. Then*

(a) *$\mathcal{P}$ is **relatively compact** if every sequence with elements in $\mathcal{P}$, has a subsequence that converges weakly.*

(b) *$\mathcal{P}$ is **tight** if for all $\epsilon > 0$, there exists $a > 0$ such that $\mu([-a,a]^c) < \epsilon$, for every $\mu \in \mathcal{P}$.*

Notice that if $\mathcal{P} := \{\mu_1, ..., \mu_r\}$ is finite, then $\mathcal{P}$ is tight, since

$$\sup_{\mu \in \mathcal{P}} \mu([-n,n]^c) \leq \sum_{i=1}^{r} \mu_i([-n,n]^c) \longrightarrow 0,$$

as $n \to \infty$.

We aim now to prove the equivalence of the two previous concepts. We will prove a more general case and consider functions with the image in $[0, M]$.

Theorem 4.2.2. ***(Helly-Bray's Theorem).*** *Let $\{F_n, n \geq 1\}$ be a sequence of functions $F_n : \mathbb{R} \longrightarrow [0, M]$, increasing and right-continuous. Then, there exists another function $F : \mathbb{R} \longrightarrow [0, M]$, increasing and right-continuous and a subsequence F_{n_k} such that*

$$\lim_{n} F_{n_k}(x) = F(x), \tag{4.2.1}$$

for all $x \in \mathbb{R}$ continuity point of F.

Proof. First of all, we define the sequence F_{n_k}. Fix the set $D := \{x_n, n \geq 1\}$ in $\mathbb{R}$. We will use Cantor's diagonal method to construct a subsequence $\{F_{n_k}(x_j), k \geq 1\}$ that converges to a limit y_j, for all $j \geq 1$. Since $0 \leq F_n(x_1) \leq M$, for all $n \geq 1$, there exists a subsequence $\{F_{1,n}(x_1)\}$ that converges to a limit $y_1 \in [0, M]$. Since $0 \leq F_{1,n}(x_2) \leq M$, for all $n \geq 1$, there exists a subsequence $\{F_{2,n}(x_2)\}$ that converges to a limit $y_2 \in [0, M]$. In general, since $0 \leq F_{m,n}(x_{m+1}) \leq M$, for all $n \geq 1$, there exists a subsequence $\{F_{m+1,n}(x_{m+1})\}$ that converges to a limit $y_{m+1} \in [0, M]$. Consider now the diagonal sequence $F_{n_k}(x) = F_{k,k}(x)$. Then, for all $x_j \in D$, $F_{n_k}(x_j) = F_{k,k}(x_j)$ is a partial sequence of $\{F_{j,k}(x_j), k \geq 1\}$ for $k \geq j$, satisfying $\lim_{n\to\infty} F_{n_k}(x_j) = y_j$.

For better clarity, define the function $F_D : D \longrightarrow [0, M]$, such that $F_D(x_j) := y_j$. F_D satisfies that $\lim_{k\to\infty} F_{n_k}(x) = F_D(x)$, for all $x \in D$. Notice that it is also increasing: if $x \leq y$, $F_{n_k}(x) \leq F_{n_k}(y)$, for every $k \geq 1$, then $F_D(x) \leq F_D(y)$, since F is increasing.

Secondly, we desire to define an increasing and right-continuous function as $F : \mathbb{R} \longrightarrow [0, M]$. Consider the function $F : \mathbb{R} \longrightarrow [0, M]$, defined by the infimum $F(x) := \inf\{F_D(y) : y \in D, y > x\}$. We see it satisfies the required properties:

- F is increasing: If $x_1 \leq x_2$, then $F(x_1) = \inf\{F_D(y) : y \in D, y > x_1\} \leq \inf\{F_D(y) : y \in D, y > x_2\} = F(x_2)$, by the definition of F_D.

- F is right-continuous: Consider the decreasing sequence $\{z_n, n \geq 1\}$ to x. Then $F(z_n) \searrow b \geq F(x)$. Suppose that $b > F(x)$. By the definition of $F(x)$, there exists a $y_0 \in D$, satisfying that $y_0 > x$ and $F_D(y_0) > b$. Then for n sufficiently big, $x \leq z_n < y_0$ and consequently, $F(z_n) \leq F_D(y_0) < b$, by the definition of $F(z_n)$. Then $b = \lim_{n\to\infty} F(z_n) \leq F_D(y_0) < b$, which is a contradiction. Therefore $b = F(x)$.

Finally, we want to show that $\lim_{n\to\infty} F_{n_k}(x) = F(x)$, for every continuity point x of F. Consider a $y \in D$ such that $y > x$, where x is a continuity point of F. On one hand,

$$F_D(x) = \limsup_k F_{n_k}(x) \leq \limsup_k F_{n_k}(y) = F_D(y)$$

and therefore,

$$\limsup_k F_{n_k}(x) \leq F(x). \tag{4.2.2}$$

On the other hand, if $x' < y < x$, $y \in D$, then

$$\liminf_k F_{n_k}(x) \geq \liminf_k F_{n_k}(y) = F_D(y) \geq F(x'),$$

by the definition of $F(x')$. Since this is satisfied for every $x' < x$ and F is continuous in x, then

$$\liminf_k F_{n_k}(x) \geq F(x') > F(x). \tag{4.2.3}$$

By (4.2.2) and (4.2.3), finally, $\lim_n F_{n_k}(x) = F(x)$, as we wanted. ∎

Observations 4.2.3. Two consequences are the following:

1) A similar result can be formulated for functions $F : \mathbb{R}^n \longrightarrow [0, M]$, right-continuous and with non-negative rectangular increments.

2) In particular, taking $M = 1$, the Helly-Bray's Theorem does not guarantee, in general, that F is a distribution function, since it may not satisfy the required conditions $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

At last, we can prove two criteria for weak convergence. Firstly, we will see a useful theorem, which guarantees the equivalence between weak compactness and tightness of probabilities.

Theorem 4.2.4. ***(Prokhorov's Theorem).*** *Let $\mathcal{P}$ be a family of probabilities in $\mathbb{R}$. Then, $\mathcal{P}$ is tight if and only if it is relatively compacted.*

Proof. Assume that $\mathcal{P}$ is tight. Consider a sequence $\{\mu_n, n \geq 1\} \subseteq \mathcal{P}$. Let F_n be the distribution function of μ_n, for every $n \geq 1$. By Helly-Bray's Theorem *(Theorem 4.2.2)*, there exists a subsequence $\{F_{n_k}, k \geq 1\}$ such that $\lim_k F_{n_k}(x) = F(x)$, for every continuity point x of F, where the function $F : \mathbb{R} \longrightarrow [0,1]$ is increasing and right-continuous. Now we only need to show that

$$\lim_{x \to -\infty} F(x) = 0 \; ; \; \lim_{x \to \infty} F(x) = 1. \tag{4.2.4}$$

Fix $\epsilon > 0$ and $a > 0$ such that $\mu([-a,a]^c) < \epsilon$, for all $n \geq 1$. Let $b \geq a$ and $c < -a$ be continuity points of F. Then

$$F_n(b) \geq \mu_n([-a,a]) > 1 - \epsilon$$

and

$$F_n(c) \leq \mu_n([-a,a]^c) < \epsilon.$$

Taking $n \to \infty$, $F(b) > 1 - \epsilon$ and $F(c) < \epsilon$ and taking ϵ close to 0, (4.2.4) is satisfied. Finally, $\mathcal{P}$ is relatively compact, as we wanted.

Assume that $\mathcal{P}$ is relatively compact. We will prove $\mathcal{P}$ is tight by contradiction. Suppose $\mathcal{P}$ is not tight. Then, there exists $\epsilon > 0$ such that for all $n \geq 1$, there exists $\mu_n \in \mathcal{P}$ satisfying that $\mu_n([-n,n]^c) \geq \epsilon$. By hypothesis, there exists a subsequence $\{\mu_{n_k}, k \geq 1\}$ such that $\omega - \lim_{k\to\infty} \mu_{n_k} = \mu$. Consider the function $f^m(x) := [|x| - (m-1)]^+ \wedge 1$. For every $n_k \geq m$,

$$\begin{aligned} 0 < \epsilon \leq \limsup_k \mu_{n_k}([-n_k, n_k]^c) &\leq \limsup_k \int_{\mathbb{R}} f^m \, d\mu_{n_k} = \int_{\mathbb{R}} f^m \, d\mu \\ &\leq \mu([-m+1, m-1]^c), \end{aligned}$$

and taking $m \to \infty$, we get a contradiction. Therefore, $\mathcal{P}$ is tight, as we desired. ∎

Finally, we can study another **criterion for weak convergence**.

Theorem 4.2.5. *Let $\{p_n, n \geq 1\}$ be a tight sequence of probabilities in $\mathbb{R}$, such that all convergent subsequences have the same limit p. Then, $\omega - \lim_{n\to\infty} p_n = p$.*

Proof. We will prove it by contradiction. Suppose that $\{p_n, n \geq 1\}$ does not weakly converge to p. Then, there exists a function $f \in \mathcal{C}_b(\mathbb{R})$ such that the set $\{\int_{\mathbb{R}} f \, dp_n, n \geq 1\}$ does not converge to $\int_{\mathbb{R}} f(x) \, dp(x)$, i.e. there exists $\epsilon > 0$ and a subsequence $\{p_{n_k}, k \geq 1\}$ such that

$$\left| \int_{\mathbb{R}} f \, dp_n - \int_{\mathbb{R}} f \, dp \right| \geq \epsilon, \tag{4.2.5}$$

for all $k \geq 1$. By Prokhorov's Theorem *(Theorem 4.2.4)*, there exists a subsequence $\{p_{n_{k_i}}, i \geq 1\}$ of the weakly convergent sequence $\{p_{n_k}, k \geq 1\}$ and with limit p, which is a contradiction with (4.2.5). Finally, $\omega - \lim_{n\to\infty} p_n = p$, as we wanted. ∎

At last, we can study the final criterion for weak convergence, which guarantees the weak convergence in terms of the characteristic function. This is exactly Paul-Lévy's Continuity Theorem and it will be crucial for the next chapter. We need a prior truncation inequality.

Proposition 4.2.6. ***(Truncation inequality).*** *Let p be a probability in $\mathbb{R}$ and φ be its respective characteristic function. Then, for every $a > 0$,*

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt \geq p\left\{x:\ |x| \geq \frac{2}{a}\right\}. \tag{4.2.6}$$

Proof. By Fubini's Theorem and using that sinus is an odd function,

$$\begin{aligned}\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt &= \frac{1}{a}\int_{-a}^{a}\left(\int_{\mathbb{R}}(1-e^{itx})\,p(dx)\right)\,dt\\ &= \frac{1}{a}\int_{\mathbb{R}}\left(\int_{-a}^{a}(1-cos(tx))\,dt\right)\,p(dx)\\ &= \int_{\mathbb{R}}2\left(1-\frac{sin(ax)}{ax}\right)\,p(dx)\\ &\geq 2\inf_{\{|y|\geq 2\}}\left(1-\frac{sin(y)}{y}\right)\int_{\{x:\ |ax|\geq 2\}}p(dx)\\ &\geq p\{x:|x|\geq\frac{2}{a}\},\end{aligned}$$

using in the last step that $\left|\frac{sin(y)}{y}\right| \leq 1$ and particularly $\left|\frac{sin(y)}{y}\right| \leq \frac{1}{2}$, if $|y| \geq 2$. ∎

Theorem 4.2.7. ***(Paul-Lévy's Continuity Theorem).*** *Let $\{p_n, n \geq 1\}$ be a sequence of probabilities in $\mathbb{R}$ and $\{\varphi_n, n \geq 1\}$ the sequence of its respective characteristic functions. Therefore,*

1) *If p_n converges weakly to a probability p as $n \to \infty$, then $\lim_{n\to\infty}\varphi_n(t) = \varphi(t)$, for every $t \in \mathbb{R}$, where φ is the characteristic function of p.*

2) *If $\lim_{n\to\infty}\varphi_n(t) = \varphi(t)$, where φ is the continuous function in $t = 0$, then φ is the characteristic function of p and it is satisfied that $\omega - \lim_{n\to\infty} p_n = p$.*

Proof. We will prove both statements. The first one 1) is trivial by the definition of weak convergence and characteristic functions, since $cos(tx)$ and $sin(tx)$ are bounded and continuous functions.

To prove the statement 2) we will need *Proposition 4.2.6* and the Dominated Convergence Theorem (which can be used since $|\varphi_n(t)| \leq 1$). Therefore,

$$p_n\left\{x:\ |x| \geq \frac{2}{a}\right\} \leq \frac{1}{a}\int_{-a}^{a}(1-\varphi_n(t))\,dt$$

and by hypothesis,

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi_n(t))\,dt \underset{n\to\infty}{\longrightarrow} \frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt.$$

On the other hand, φ is continuous in $t=0$ by hypothesis,

$$\lim_{a\to 0}\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt = 2(1-\varphi(0)) = 0.$$

Fix $\epsilon > 0$ and let $a > 0$ be such that $0 \leq \frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt \leq \frac{\epsilon}{2}$. Taking $n_0 \geq n$, then

$$p_n\left(\left[-\frac{2}{a},\frac{2}{a}\right]^{\mathsf{c}}\right) \leq p_n\left\{x:\ |x| \geq \frac{2}{a}\right\} \leq \frac{1}{a}\int_{-a}^{a}(1-\varphi_n(t))\,dt$$
$$\leq \frac{1}{a}\int_{-a}^{a}(1-\varphi(t))\,dt + \frac{\epsilon}{2} \leq \epsilon.$$

For $1 \leq n \leq n_0$, there is $a_n > 0$ such that $p_n([-a_n,a_n]^{\mathsf{c}}) \leq \epsilon$. Therefore, considering $b = \max\{a_1,...,a_{n_0},\frac{2}{a}\}$, it is satisfied that $\sup_n \mu_n([-b,b]^{\mathsf{c}}] \leq \epsilon$. We have proved that $\{\mu_n, n \geq 1\}$ is a tight sequence of probabilities in $\mathbb{R}$. By the criterion *Theorem 4.2.5*, to prove that $\omega - \lim_{n\to\infty} p_n = p$, it is sufficient to show that all convergent subsequences have the same limit p. Consider a subsequence $\{p_{n_i}, i \geq 1\}$ convergent to a probability named v. Using the first statement 1),

$$\varphi_v(t) = \lim_i \varphi_{n_i}(t) = \varphi(t),$$

implying that φ is the characteristic function of a probability denoted $p := v$ and also $\omega - \lim_{n\to\infty} p_n = p$, since all the convergent subsequences have a limit p. ∎

4.3 Convolutions of probability measures

This section aims to introduce a new concept: convolutions of probability measures. We will begin with convolutions of measures, followed by convolutions of probability measures and conclude with the distance between them. This will be important to prove different versions of the Central Limit Theorem *(Chapter 5)*.

4.3.1 Convolutions of measures

Throughout this section we will define $\mathcal{M} := \{$*set of finite measures in* $\mathbb{R}\}$. Consider $\mu_1, \mu_2 \in \mathcal{M}$ and $B \subset \mathcal{B}(\mathbb{R})$.

Definition 4.3.1. *Let μ_1, μ_2 be finite measures in $\mathbb{R}$. The **convolution of the measures** μ_1 and μ_2 is defined by*

$$(\mu_1 * \mu_2)(B) := \int_{\mathbb{R}} \mu_2(B-x)\mu_1(dx), \tag{4.3.1}$$

where $B - x := \{y : y + x \in B\}$ and $\mu_1(dx) = d\mu_1(x)$. This is denoted $\mu_1 * \mu_2$.

We will study some properties of a convolution of two measures.

Proposition 4.3.2. *The convolution $\mu_1 * \mu_2$ is a* ***measure****. Furthermore, the operation is commutative, associative and has a neutral element.*

Proof. First of all, we want to prove that the operation $\mu_1 * \mu_2$ is a measure. Define the map $T : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$, $T(x,y) := x + y$. Then the preimage of T is given by $T^{-1}(B) := \{(x,y) : x + y \in B\}$. The measure of the image is

$$\begin{aligned} v(B) :=& (\mu_1 * \mu_2)(T^{-1}(B)) = \int_{\mathbb{R}} \mathbb{1}_{\{T^{-1}(B)\}} \, d(\mu_1 * \mu_2)(x) \\ =& \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_{\{T^{-1}(B)\}} \, d\mu_2(x) d\mu_1(y) = \int_{\mathbb{R}} \left[\int_{\{x:\, x \in B-y\}} d\mu_2(x) \right] d\mu_1(y) \\ =& \int_{\mathbb{R}} \mu_2(B-y) \, d\mu_1(y) = (\mu_1 * \mu_2)(B) \end{aligned} \tag{4.3.2}$$

Therefore, $\mu_1 * \mu_2$ is indeed a measure. We will check now the desired properties. First, notice that $\mu_1 * \mu_2$ is commutative, i.e. $\mu_1 * \mu_2 = \mu_2 * \mu_1$. Note that if we interchange μ_1 and μ_2 in (4.3.2), we get $v(B) = (\mu_2 * \mu_1)(B) = (\mu_1 * \mu_2)(B)$, therefore the desired property is fulfilled. Secondly, we see that $\mu_1 * \mu_2$ is associative, i.e. $(\mu_1 * \mu_2) * \mu_3 = \mu * (\mu_2 * \mu_3)$. Consider the measures $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$. By the definition of convolution,

$$\begin{aligned} ((\mu_1 * \mu_2) * \mu_3)(B) &= \int_{\mathbb{R}} \mu_3(B-x) \, d(\mu_1 * \mu_2)(x) \\ &= \int_{\mathbb{R}\times\mathbb{R}} \mu_3(B-(x+y)) \, d\mu_1(x) d\mu_2(y), \end{aligned}$$

and

$$\begin{aligned} (\mu_1 * (\mu_2 * \mu_3))(B) &= \int_{\mathbb{R}} (\mu_2 * \mu_3)(B-x) \, d\mu_1(x) \\ &= \int_{\mathbb{R}\times\mathbb{R}} \mu_3(B-(x+y)) \, d\mu_2(y) d\mu_1(x), \end{aligned}$$

then $((\mu_1 * \mu_2) * \mu_3)(B) = (\mu_1 * (\mu_2 * \mu_3))(B)$, as we wanted. Thirdly, there exists a neutral element denoted by μ_e. Define the measure $\mu_e(x) := \mathbb{1}_{\{x=0\}} \in \mathcal{M}$ and consider $\mu \in \mathcal{M}$. Observe that $\mu(B) * \mu_e(x) = \mu(B)$ and also

$$\mu_e(x) * \mu(B) = \int_{\mathbb{R}} \mu(B-x) \, d\mu_e(x) = \int_{\mathbb{R}} \mu(B) = \mu(B).$$

∎

4.3.2 Convolutions of probability measures

In this subsection, we want to discover what happens if we consider probability measures instead of measures. We will define $\mathcal{P} := \{\textit{set of probability measures}\}$.

Consider $p_1,\ p_2 \in \mathcal{P}$. Applying the definition of **convolution of** measures to the **probabilities** p_1 and p_2, the convolution $p_1 * p_2$ is indeed a probability:

$$(p_1 * p_2)(\mathbb{R}) = \int_{\mathbb{R}} p_2(\mathbb{R} - x)\ dp_1(x) = \int_{\mathbb{R}} dp_1(x) = p_1(\mathbb{R}) = 1, \tag{4.3.3}$$

since $p_i(\mathbb{R}) = 1,\ i = 1, 2$, by definition of probability. This is denoted $p_1 * p_2$.

Now we wonder how to describe the distribution function, density and characteristic function of $p_1 * p_2$, knowing the respective functions of p_1 and p_2.

Definition 4.3.3. *Let F_1, F_2 be the distribution functions of p_1, p_2, respectively. Let F be the* ***distribution function of the convolution*** *$p_1 * p_2$. F is defined by*

$$F(x) := \int_{\mathbb{R}} F_2(x - y)\ dF_1(y). \tag{4.3.4}$$

F is the convolution of F_1 and F_2, i.e. $F = F_1 * F_2$.

The equation (4.3.4) is the result of

$$F(x) = (p_1 * p_2)((-\infty, x]) = \int_{\mathbb{R}} p_2((-\infty, x - y])\ dp_1(y) = \int_{\mathbb{R}} F_2(x - y)\ dF_1(y),$$

where by definition $F_2(x - y) = p_2((-\infty, x - y])$ and by notation $dF_1(y) = dp_1(y)$.

Definition 4.3.4. *Let p_1, p_2 be continuous probabilities and f_1, f_2 their density functions, respectively. Let f be the* ***density function of the convolution*** *$p_1 * p_2$. f is*

$$f(t) := \int_{\mathbb{R}} f_2(t - y) f_1(y)\ dy. \tag{4.3.5}$$

f is the convolution of f_1 and f_2, i.e. $f = f_1 * f_2$.

The equation (4.3.5) is a consequence of

$$\begin{aligned} F(x) &= \int_{\mathbb{R}} F_2(x - y)\ dF_1(y) = \int_{\mathbb{R}} \left[\int_{-\infty}^{x-y} f_2(z)\ dz \right] f_1(y)\ dy \\ &= \int_{\mathbb{R}} \left[\int_{-\infty}^{x} f_2(t - y)\ dt \right] f_1(y)\ dy = \int_{-\infty}^{x} \left[\int_{\mathbb{R}} f_2(t - y) f_1(y)\ dy \right]\ dt, \end{aligned}$$

taking $t = z + y$.

Definition 4.3.5. *Let φ_1, φ_2 be the characteristic functions of p_1, p_2, respectively. The* ***characteristic function*** *$\varphi_{p_1 * p_2}$* ***of the convolution*** *$p_1 * p_2$ is $\varphi_{p_1 * p_2} = \varphi_{p_1} \varphi_{p_2}$.*

The definition is due to

$$\begin{aligned} \varphi_{p_1 * p_2}(t) &= \int_{\mathbb{R}} e^{itx}\ d(p_1 * p_2)(x) = \int_{\mathbb{R}} e^{it(x+y)}\ dp_1(x) dp_2(y) \\ &= \int_{\mathbb{R}} e^{itx}\ dp_1(x) \int_{\mathbb{R}} e^{ity}\ dp_2(y) = \varphi_{p_1}(t) \varphi_{p_2}(t), \end{aligned}$$

using that $e^{it(x+y)} = e^{itx} e^{ity}$.

4.3.3 Distance of convolutions

We aim to define a distance and study its relation with weak convergence and convolutions of probabilities. For any $k \geq 1$, define $\mathcal{C}_b^k := \{f : \mathbb{R} \longrightarrow \mathbb{R} :$ *bounded, k-times differentiable and with continuous and bounded derivatives*$\}$.

Definition 4.3.6. *The* ***distance*** *between two measures $\mu,\ v \in \mathcal{M}$ is defined by*

$$d_k(\mu, v) := \sup \left\{ \left| \int_{\mathbb{R}} f(x)\ \mu(dx) - \int_{\mathbb{R}} f(x)\ v(dx) \right|,\ f \in \mathcal{C}_b^k,\ \sum_{i=0}^{k} \|f^{(i)}\|_\infty \leq 1 \right\}. \tag{4.3.6}$$

Observe the distance decreases as k increases, i.e. $d_1 \geq d_2 \geq d_3 \geq ....$

First, we state weak convergence in terms of the previous distance d_k.

Proposition 4.3.7. *Let $\{\mu_n, n \geq 1\}, \mu$ be probability measures. Then, it is satisfied that*

$$\mu_n \xrightarrow{\omega} \mu \overset{1)}{\iff} \exists k \geq 1,\ d_k(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0 \overset{2)}{\iff} \forall k \geq 1,\ d_k(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0. \tag{4.3.7}$$

Proof. We aim to prove both equivalences. We will start with 1). Suppose there is $k \geq 1$, $d_k(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0$. If F is the distribution function of μ, let x be a continuity point of F. Let $f_n(x), g_n(x) \in \mathcal{C}_b^k$ be two sequences of functions such that

$$\mathbb{1}_{(-\infty,x]} \leq f_n(x) \leq \mathbb{1}_{(-\infty,x+\frac{1}{n}]}$$

and

$$\mathbb{1}_{(-\infty,x-\frac{1}{n}]} \leq g_n(x) \leq \mathbb{1}_{(-\infty,x]}.$$

Notice that $d_k(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0$ implies that $\int_{\mathbb{R}} f(x)\ d\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\ d\mu(x)$ and $f \in \mathcal{C}_b^k$ satisfies that $\sum_{i=0}^{k} \|f^{(i)}\|_\infty \leq 1$, by definition. Observe that for all $n \geq 1$,

$$\begin{aligned}
F\left(x - \frac{1}{n}\right) &= \mu\left(\left(-\infty, x - \frac{1}{n}\right]\right) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x-\frac{1}{n}]}\ d\mu \leq \int_{\mathbb{R}} g_n\ d\mu \\
&= \lim_{m\to\infty} \int_{\mathbb{R}} g_n\ d\mu_m = \liminf_{m\to\infty} \int_{\mathbb{R}} g_n\ d\mu_m \\
&\leq \liminf_{m\to\infty} \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x]}\ d\mu_m \leq \limsup_{m\to\infty} \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x]}\ d\mu_m \\
&\leq \limsup_{m\to\infty} \int_{\mathbb{R}} f_n(x)\ d\mu_m = \lim_{m\to\infty} \int_{\mathbb{R}} f_n\ d\mu_m \\
&= \int_{\mathbb{R}} f_n\ d\mu \leq \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x+\frac{1}{n}]}\ d\mu = F\left(x + \frac{1}{n}\right).
\end{aligned}$$

Therefore, for every $n \geq 1$,

$$F\left(x - \frac{1}{n}\right) \leq \liminf_{m\to\infty} F_m(x) \leq \limsup_{m\to\infty} F_m(x) \leq F\left(x + \frac{1}{n}\right)$$

and taking limits as $n \to \infty$,

$$\liminf_{m\to\infty} F_m(x) = \limsup_{m\to\infty} F_m(x),$$

i.e. $\lim_{m\to\infty} F_m(x) = F(x)$, for every x continuity point of F and by *Theorem 4.1.2*, finally $\mu_n \xrightarrow{\omega} \mu$.

Suppose $\mu_n \xrightarrow{\omega} \mu$. We want to see that there exists a $k \geq 1$ satisfying $d_k(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0$. Since $d_1 \geq d_2 \geq ...$, it is enough to prove $d_1(\mu_n, \mu) \underset{n\to\infty}{\longrightarrow} 0$, i.e. take $k = 1$. Fix $\epsilon > 0$, then observe the disjoint union $\mathbb{R} = \bigcup_{n=1}^{\infty}(a_n, b_n]$, where a_n, b_n are the continuity points of F with $b_n - a_n < \epsilon$, for every $n \geq 1$.[2] Define $I_n := (a_n, b_n]$. Consider $f \in \mathcal{C}_b'$ such that $\|f\|_\infty + \|f'\|_\infty \leq 1$. Then

$$\begin{aligned}
&\left| \int_{\mathbb{R}} f(x)\ d\mu_n(x) - \int_{\mathbb{R}} f(x)\ d\mu(x) \right| \\
&\quad = \left| \int_{\mathbb{R}} f(x)\ d\mu_n(x) + \sum_{k=1}^{\infty} f(a_k)\mu_n(I_k) - \sum_{k=1}^{\infty} f(a_k)\mu_n(I_k) - \int_{\mathbb{R}} f(x)\ d\mu(x) \right| \\
&\quad = \left| \sum_{k=1}^{\infty} \int_{I_k} (f(x) - f(a_k))\ d\mu_n(x) + \sum_{k=1}^{\infty} \int_{I_k} (f(a_k) - f(x))\ d\mu(x) \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} f(a_k)\left[\mu_n(I_k) - \mu(I_k)\right] \right| \\
&\quad \leq \sum_{k=1}^{\infty} \int_{I_k} |f(x) - f(a_k)|\ d\mu_n(x) + \sum_{k=1}^{\infty} \int_{I_k} |f(a_k) - f(x)|\ d\mu(x) \\
&\qquad + \sum_{k=1}^{\infty} |f(a_k)|\,|\mu_n(I_k) - \mu(I_k)| \\
&\quad \leq \sum_{k=1}^{\infty} \int_{I_k} \epsilon\ d\mu_n(x) + \sum_{k=1}^{\infty} \int_{I_k} \epsilon\ d\mu(x) + \sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)| \\
&\quad \leq 2\epsilon + \sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)|
\end{aligned} \tag{4.3.8}$$

The second-to-last inequality is a consequence of the following. Taylor's expansion is $f(x) = f(a_k) + \mathcal{R}(\alpha)$, where $\mathcal{R}(\alpha) = f'(\alpha)(x - a_k)$ and $\alpha \in (a_k, x)$. Recall that $\|f'(\alpha)\|_\infty \leq 1$ and $|x - a_k| < \epsilon$, therefore

$$|f(x) - f(a_k)| = |\mathcal{R}(\alpha)| \leq \|f'(\alpha)\|_\infty |x - a_k| < \epsilon.$$

Also $\|f\|_\infty, \|f'\|_\infty < 1$, due to $\|f\|_\infty + \|f'\|_\infty \leq 1$. Taking the supremum on f in (4.3.8),

$$\sup_{f\in\mathcal{C}_b'} \left\{ \left| \int_{\mathbb{R}} f(x)\ d\mu_n(x) - \int_{\mathbb{R}} f(x)\ d\mu(x) \right| \right\} \leq 2\epsilon + \sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)|\,.$$

[2] Note that the discontinuity points of F_μ might be countable, finite or non-existent.

Since f satisfies $\|f\|_\infty + \|f'\|_\infty \le 1$, the left-side corresponds to $d_1(\mu_n, \mu)$. We need to see that the right side tends to 0. We want to bound $\sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)|$,[3]

$$\begin{aligned}\sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)| &= 2\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)]^+ - \sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)] \\ &= 2\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)]^+ ,\end{aligned} \tag{4.3.9}$$

since

$$\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)] = \sum_{k=1}^{\infty} \mu_n(I_k) - \sum_{k=1}^{\infty} \mu(I_k) = 1 - 1 = 0.$$

Define $g_n(x) := 2\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)]^+ \mathbb{1}_{(k,k+1]}(x)$, then it is fulfilled the equality $\sum_{k=1}^{\infty} |\mu_n(I_k) - \mu(I_k)| = \int_{\mathbb{R}} g_n(x)\, dx$, by (4.3.9). Observe that $g_n(x) \underset{n\to\infty}{\longrightarrow} 0$, for every $x \in \mathbb{R}$, since $\mu_n \overset{\omega}{\longrightarrow} \mu$. On the other hand,

$$|g_n(x)| \le \left| \sum_{k=1}^{\infty} 2\mu_n(I_k)\mathbb{1}_{(k,k+1]}(x) \right| =: g(x).$$

Lebesgue's Dominated Convergence Theorem can be applied to g since

$$\int_{\mathbb{R}} g(x)\, dx = \sum_{k=1}^{\infty} 2\mu_n(I_k) = 2 < \infty,$$

therefore

$$\lim_{n\to\infty} \int_{\mathbb{R}} g_n(x)\, dx = \int_{\mathbb{R}} g(x)\, dx = 0,$$

which implies that (4.3.9) tends to 0, by definition of g, i.e.

$$\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)] = 2\sum_{k=1}^{\infty} [\mu_n(I_k) - \mu(I_k)]^+ \underset{n\to\infty}{\longrightarrow} 0.$$

Finally,

$$\left\{ \left| \int_{\mathbb{R}} f(x)\, d\mu_n(x) - \int_{\mathbb{R}} f(x)\, d\mu(x) \right| \right\} \underset{n\to\infty}{\longrightarrow} 2\epsilon$$

and taking ϵ close to 0, we are done.

Now we want to prove 2). It is a one-line proof since the forward implication is a consequence of the property $d_1 \ge d_2 \ge \ldots$ and the reverse implication is trivial. Finally, we are done. ∎

The last result of this chapter gives a bound of the defined distance in terms of convolutions of probabilities.

[3] Recall that $|A| = 2A^+ - A$, for every set A.

Proposition 4.3.8. *Let $\mu_1, ..., \mu_n, v_1, ..., v_n \in \mathcal{P}$ be finite probability measures. Then,*

$$d_k(\mu, v) \leq \sum_{i=1}^{n} d_k(\mu_i, v_i), \tag{4.3.10}$$

where $\mu := \mu_1 * ... * \mu_n,\ v := v_1 * ... * v_n \in \mathcal{P}$.

Proof. We will do the proof by induction. First, we show that the statement holds for $n = 2$. Consider the probability measures $\mu_1, \mu_2, v_1, v_2 \in \mathcal{P}$. By the triangular inequality and definition of d_k,

$$\begin{aligned}
d_k(\mu_1 * \mu_2, v_1 * v_2) &= \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} f(x)\ d(\mu_1 * \mu_2)(x) - \int_{\mathbb{R}} f(x)\ d(v_1 * v_2)(x) \right| \right\} \\
&= \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} f(x+y)\ d\mu_1(x) d\mu_2(y) - \int_{\mathbb{R}} f(x+y)\ dv_1(x) dv_2(y) \right| \right\} \\
&\leq \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} f(x+y)\ d\mu_1(x) d\mu_2(y) - \int_{\mathbb{R}} f(x+y)\ d\mu_1(x) dv_2(y) \right| \right. \\
&\quad \left. + \left| \int_{\mathbb{R}} f(x+y)\ d\mu_1(x) dv_2(y) - \int_{\mathbb{R}} f(x+y)\ dv_1(x) dv_2(y) \right| \right\} \\
&= \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x+y)\ (d\mu_2(y) - dv_2(y)) \right)\ d\mu_1(x) \right| \right. \\
&\quad \left. + \left| \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x+y)\ (d\mu_1(x) - dv_1(x)) \right)\ dv_2(y) \right| \right\} \\
&\leq \int_{\mathbb{R}} \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} f(x+y)\ ((d\mu_2(y) - dv_2(y)))\ d\mu_1(x) \right| \right\} \\
&\quad + \int_{\mathbb{R}} \sup_{f \in \mathcal{C}_b^k} \left\{ \left| \int_{\mathbb{R}} f(x+y)\ ((d\mu_1(x) - dv_1(x)))\ dv_2(y) \right| \right\} \\
&= d_k(\mu_1, v_1) + d_k(\mu_2, v_2).
\end{aligned}$$

The induction step consists of proving the statement for $n+1$, if it is true for n. Consider the probability measures $\mu_1, ..., \mu_n, \mu_{n+1}, v_1, ..., v_n, v_{n+1} \in \mathcal{P}$. Denote $\mu := \mu_1 * ... * \mu_n,\ v := v_1 * ... * v_n \in \mathcal{P}$. By assumption, $d_k(\mu, v) \leq \sum_{i=1}^{n} d_k(\mu_i, v_i)$, therefore it is fulfilled that

$$\begin{aligned}
d_k(\mu * \mu_{n+1}, v * v_{n+1}) &\leq d_k(\mu, v) + d_k(\mu_{n+1}, v_{n+1}) \\
&\leq \sum_{i=1}^{n} d_k(\mu_i, v_i) + d_k(\mu_{n+1}, v_{n+1}) = \sum_{i=1}^{n+1} d_k(\mu_i, v_i),
\end{aligned}$$

using the base case in the first step. Finally, the statement is proved. ∎

Chapter 5

Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution of a normalized version of the sample mean converges to a standard normal distribution, as the sample size becomes larger. The convergence is weak and the assumption is that all samples are independent, identical in size, and have a finite second-order moment, regardless of the population's actual distribution.

The theorem has faced many changes from classical to modern probability theory. We will go through the principal variants of the CLT, such as their differences, interpretations and historical prominence. For in-depth learning of the normal distribution and its applicability, we suggest *([5], Chapter 4)* and *([6], Chapter 5)*.

5.1 De Moivre-Laplace's Theorem

We will start with the de Moivre-Laplace's, which is a particular case. Its first appearance was in the second edition of *The Doctrine of Chances* (1738), by Abraham de Moivre. The interpretation is that the normal distribution can approximate the normal distribution. Let $X_1, ..., X_n$ be independent r.v. Let $\mathcal{L}(X_1), ..., \mathcal{L}(X_n)$ be their respective distribution functions. Denote $S_n := \sum_{i=1}^n X_i$. Then

$$\mathcal{L}(S_n) = \mathcal{L}(X_1) * ... * \mathcal{L}(X_n), \tag{5.1.1}$$

due to the independence of $\{X_i, i \geq 1\}$. On the other hand, if $X_i \sim N(0, \sigma_i^2)$, for every $i \in \{1, ..., n\}$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$, then

$$N(0, \sigma^2) = N(0, \sigma_1^2) * ... * N(0, \sigma_n^2). \tag{5.1.2}$$

Proposition 5.1.1. *Let $X_1, ..., X_n$ be independent r.v, centered and bounded by C (i.e. $|X_i| \leq C$ a.s. for every $i \in \{1, ..., n\}$). Denote $\sigma_i^2 = E(X_i^2)$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. Then,*

$$d_3(\mathcal{L}(S_n), N(0, \sigma^2)) \leq C\sigma^2 \frac{1}{6}\left(1 + \sqrt{\frac{8}{\pi}}\right). \tag{5.1.3}$$

Proof. Using the equations (5.1.1) and (5.1.2) and *Proposition 4.3.8*, then

$$d_3(\mathcal{L}(S_n), N(0,\sigma^2)) \leq \sum_{i=1}^{n} d_3(\mathcal{L}(x_i), N(0,\sigma_i^2)). \tag{5.1.4}$$

Denote $Y_i \sim N(0,\sigma_i^2)$ for all $i \in \{1,...,n\}$ and $Y \sim N(0,\sigma^2)$. Consider $f \in \mathcal{C}_b^k$ such that $\|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty + \|f'''\|_\infty \leq 1$. Taylor's Development around $x = 0$ is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{x^3}{6}f'''(\alpha).$$

Denote $R(\alpha) := \frac{x^3}{6}f'''(\alpha)$ with $|R(\alpha)| \leq \|f'''(\alpha)\|_\infty \leq 1$ and α between 0 and x. Considering the random variables X_i ans Y_i and $\omega \in \Omega$, then

$$f(X_i(\omega)) = f(0) + f'(0)X_i(\omega) + \frac{f''(0)}{2}X_i^2(\omega) + R(T(\omega)),$$
$$f(Y_i(\omega)) = f(0) + f'(0)Y_i(\omega) + \frac{f''(0)}{2}Y_i^2(\omega) + R(T'(\omega)),$$

where $|R(T(\omega))|$, $|R(T'(\omega))| \leq 1$, and also $T(\omega)$ is between 0 and $X_i(\omega)$, and $T'(\omega)$ is between 0 and $Y_i(\omega)$. Taking expectations on $f(X_i)$ and $f(Y_i)$,

$$E[f(X_i)] = f(0) + \frac{\sigma_i^2}{2}f''(0) + \frac{E[X_i^3R(T(\omega))]}{6},$$
$$E[f(Y_i)] = f(0) + \frac{\sigma_i^2}{2}f''(0) + \frac{E[Y_i^3R(T'(\omega))]}{6}.$$

Therefore,

$$E[f(X_i) - f(Y_i)] = \frac{1}{6}E[X_i^3R(T(\omega)) - Y_i^3R(T'(\omega))]. \tag{5.1.5}$$

Applying the definition of d_3 and taking absolute values on (5.1.5), then

$$d_3(X_i, Y_i) \leq \frac{1}{6}(E[|X_i|^3|R(T(\omega))|] + E[|Y_i|^3|R(T'(\omega))|]) \leq \frac{1}{6}(E[|X_i|^3] + E[|Y_i|^3]), \tag{5.1.6}$$

using $|R(T(\omega))|, |R(T'(\omega))| \leq 1$. Notice that $E(|X_i|^3) \leq CE(|X_i|^2) = C\sigma_i^2$ and also

$$E(|Y_i|^3) = \frac{2}{\sqrt{2\pi\sigma_i^2}}\int_0^\infty y^3 \exp\left(-\frac{1}{2}\frac{y^2}{\sigma_i^2}\right) dy = \frac{4\sigma_i^2}{\sqrt{2\pi}\sigma_i}\sigma_i^2 = \sqrt{\frac{8}{\pi}}\sigma_i^3 \leq \sqrt{\frac{8}{\pi}}C\sigma_i^2,$$

since $\{X_i, i \in \{1,...,n\}\}$ are bounded a.s. by C and $Y_i \sim N(0,\sigma_i^2)$, by hypothesis. Plugging these into (5.1.6),

$$d_3(\mathcal{L}(X_i), N(0,\sigma_i^2)) \leq \frac{1}{6}\left(C\sigma_i^2 + \sqrt{\frac{8}{\pi}}C\sigma_i^2\right) = \frac{C}{6}\left(1 + \sqrt{\frac{8}{\pi}}\right)\sigma_i^2.$$

Finally, substituting this in (5.1.4),

$$d_3(\mathcal{L}(S_n), N(0,\sigma^2)) \leq \sum_{i=1}^{n} \left(\frac{C}{6} \left(1 + \sqrt{\frac{8}{\pi}} \right) \sigma_i^2 \right) = \frac{C}{6} \left(1 + \sqrt{\frac{8}{\pi}} \right) \sigma^2.$$

∎

Now we can prove the first Central Limit Theorem in history.

Theorem 5.1.2. ***(De Moivre-Laplace's Theorem).*** *Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables with $X_i \sim Bernoulli(p)$. Then,*

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{\omega} N(0,1). \tag{5.1.7}$$

Proof. Let $X_1, ..., X_n$ be independent r.v with $X_i \sim Bernoulli(p)$. Let $Y_i := \frac{X_i - p}{\sqrt{np(1-p)}}$, for every $i \in \{1, ..., n\}$. Note that Y_i are independent r.v, centered and bounded by

$$|Y_i| = \left| \frac{X_i - p}{\sqrt{np(1-p)}} \right| \leq \frac{\max(p, 1-p)}{\sqrt{np(1-p)}} =: C.$$

Denote $\sigma_i^2 = E(Y_i^2)$, then

$$\sigma_i^2 = Var\left(\frac{X_i - p}{\sqrt{np(1-p)}} \right) = \frac{Var(X_i)}{np(1-p)} = \frac{1}{n}.$$

Define $S_n := \sum_{i=1}^{n} \frac{X_i - p}{\sqrt{np(1-p)}} = \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}$, with $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} \frac{1}{n} = 1$. Finally, by *Proposition 5.1.1*, then

$$d_3(\mathcal{L}(S_n), N(0,1)) \leq \frac{1}{6} \left(1 + \frac{8}{\pi} \right) \frac{\max(p,q)}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{} 0.$$

∎

5.2 Lindeberg-Lévy's CLT

Now we will study the classical and most common CLT. It was independently developed by Paul Lévy and Harald Cramér in the 1920s. This theorem provides the conditions for which the distribution of a normalized version of the sample mean converges to a standard normal distribution, as the sample size increases.

Theorem 5.2.1. ***(Lindeberg-Lévy's CLT).*** *Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed r.v. with finite second-order moment. Let $m = E(X_i)$ and $\sigma^2 = Var(X_i)$, for every $i \in \{1, ..., n\}$. Denote $S_n := X_1 + ... + X_n$, then*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\omega} N(0,1). \tag{5.2.1}$$

Proof. Define the random variable $Y_n := \frac{S_n - nm}{\sigma\sqrt{n}}$, for $n \geq 1$. Observe that $E(Y_n) = 0$ and $\sigma^2(Y_n) = E(Y_n^2) = 1$. Compute the characteristic function of Y_n, if $t \in \mathbb{R}$, then

$$\begin{aligned} E(e^{itY_n}) &= E\left[\exp\left(it\frac{S_n - nm}{\sigma\sqrt{n}}\right)\right] = E\left[\exp i\left(\frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}}\right)t\right] \\ &= E\left[\exp i\left(\sum_{i=1}^n \frac{X_i - m}{\sigma\sqrt{n}}\right)t\right] = \left(E\left[\exp i\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)t\right]\right)^n = (\varphi_n(t))^n, \end{aligned}$$

where φ_n is the characteristic function of the r.v. $\frac{X_1 - m}{\sigma\sqrt{n}}$. Observe that $\varphi_n(0) = 1$. φ_n is two times differentiable with continuous derivatives, since $E(|X_1|^2) < \infty$. Then,

$$\varphi_n'(0) = iE\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)$$

and

$$\varphi_n''(t) = -\int_{\mathbb{R}} x^2 e^{itx}\left[P_0\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)^{-1}\right]dx = -E\left[\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)^2 \exp\left(it\frac{X_1 - m}{\sigma\sqrt{n}}\right)\right].$$

In particular at $t = 0$, $\varphi_n''(0) = -\frac{1}{n}$. Taylor's Expansion applied to the function $\varphi_n(t)$ around $t = 0$ is

$$\varphi_n(t) = 1 - \frac{t^2}{2n} + \frac{t^2}{2}[\varphi_n''(\theta t) - \varphi_n''(0)], \tag{5.2.2}$$

where $|\theta| \leq 1$. Therefore,

$$\begin{aligned} n[\varphi_n''(\theta t) - \varphi_n''(0)] &= -nE\left[\left(\frac{X_1 - m}{\sigma\sqrt{n}}\right)^2\left(\exp\left(it\theta\frac{X_1 - m}{\sigma\sqrt{n}}\right) - 1\right)\right] \\ &= -\frac{1}{\sigma^2}E\left[(X_1 - m)^2\left(\exp\left(it\theta\frac{X_1 - m}{\sigma\sqrt{n}}\right) - 1\right)\right]. \end{aligned} \tag{5.2.3}$$

Notice that

$$\left|(X_1 - m)^2\left(\exp\left(it\theta\frac{X_1 - m}{\sigma\sqrt{n}}\right) - 1\right)\right| \leq 2|X_1 - m|^2$$

and

$$\lim_{n\to\infty}(X_1 - m)^2\left(\exp\left(it\theta\frac{X_1 - m}{\sigma\sqrt{n}}\right) - 1\right) = 0.$$

Applying the Dominated Convergence Theorem to the last expression in (5.2.3),

$$n[\varphi_n''(\theta t) - \varphi_n''(0)] \underset{n\to\infty}{\longrightarrow} 0.$$

Plugging this into (5.2.2), then

$$\lim_{n\to\infty} (\varphi_n(t))^n = e^{-\frac{t^2}{2}}, \ \forall t \in \mathbb{R}.$$

By the Paul-Lévy's Continuity Theorem *(Theorem 4.2.7)*, the sequence $\{Y_n, n \geq 1\}$ converges weakly to a r.v. with distribution $N(0,1)$.[1]

∎

5.3 Lindeberg's Theorem

We will focus on a stronger version of the CLT. This was published by Jarl Waldemar Lindeberg in 1920. The Lindeberg's Condition is a sufficient condition for the CLT to hold. We will study the definition of a triangular family, following with the Lindeberg's CLT and the Feller's Theorem.

Definition 5.3.1. *A* ***triangular family*** *is a sequence $\{X_{n_j}, 1 \leq j \leq k_n, k_n \nearrow \infty\}$, where the rows are independent random variables.*

Theorem 5.3.2. ***(Lindeberg's CLT).*** *Let $\{X_{n_j}, 1 \leq j \leq k_n, k_n \nearrow \infty\}$ be a triangular family with centered random variables. If it is satisfied that*

(a) $\sum_{j=1}^{k_n} E(X_{n_j}^2) \longrightarrow \sigma^2$,

(b) ***(Lindeberg's Condition).*** $\sum_{j=1}^{k_n} E[X_{n_j}^2 \mathbb{1}_{\{|X_{n_j}|>\delta\}}] \underset{n\to\infty}{\longrightarrow} 0$, *for every* $\delta > 0$.

Then,

$$S_n := \sum_{j=1}^{k_n} X_{n_j} \xrightarrow[n\to\infty]{\omega} N(0,\sigma^2). \tag{5.3.1}$$

Proof. Define the r.v. $X_{n_j,\delta} := X_{n_j}\mathbb{1}_{\{|X_{n_j}|\leq\delta\}}$. Denote $S_{n,\delta} := \sum_{j=1}^{k_n} X_{n_j,\delta}$ and also $\sigma^2_{n,\delta} := \sum_{j=1}^{k_n} \sigma^2(X_{n_j,\delta})$. We want to see that the next terms tend to 0 as $n \to \infty$,

$$\begin{aligned} d_3(\mathcal{L}(S_n), N(0,\sigma^2)) \leq\ & d_3(\mathcal{L}(S_n), \mathcal{L}(S_{n,\delta} - E[S_{n,\delta}])) \\ & + d_3(\mathcal{L}(S_{n,\delta} - E[S_{n,\delta}]), N(0,\sigma^2_{n,\delta})) + d_3(N(0,\sigma^2_{n,\delta}), N(0,\sigma^2)). \end{aligned} \tag{5.3.2}$$

We will start with $d_3(N(0,\sigma^2_{n,\delta}), N(0,\sigma^2)) \underset{n\to\infty}{\longrightarrow} 0$. Note that

- $(X_{n_j} - X_{n_j,\delta})^2 = X_{n_j}^2 - X_{n_j,\delta}^2 = X_{n_j}^2 \mathbb{1}_{\{|X_{n_j}|>\delta\}}$,
- $\sigma^2(S_n) - \sigma^2(S_{n,\delta}) \geq 0$,

[1] Note that $e^{-\frac{t^2}{2}}$ is the characteristic function of a $N(0,1)$ variable *(Chapter 2)*.

- $E(X_{n_j}) = 0 = E[X_{n_j}\mathbb{1}_{\{|X_{n_j}|\leq\delta\}}] + E[X_{n_j}\mathbb{1}_{\{|X_{n_j}|>\delta\}}]$, by definition of X_{n_j}. Taking squares, $(E(X_{n_j,\delta}))^2 = (E[X_{n_j,\delta}\mathbb{1}_{\{|X_{n_j}|\leq\delta\}}])^2 = (E[X_{n_j,\delta}\mathbb{1}_{\{|X_{n_j}|>\delta\}}])^2$.

Hence, for every $\delta > 0$,

$$0 \leq \sigma^2(S_n) - \sigma^2(S_{n,\delta}) = \sum_{j=1}^{k_n} E(X_{n_j}^2) - \sum_{j=1}^{k_n} E[X_{n_j,\delta} - E(X_{n_j,\delta})]^2$$
$$= \sum_{j=1}^{k_n}\left(E(X_{n_j}^2) - E[X_{n_j,\delta} - E(X_{n_j,\delta})]^2\right) = \sum_{j=1}^{k_n}\left(E(X_{n_j}^2) - E(X_{n_j,\delta}^2) + [E(X_{n_j,\delta})]^2\right)$$
$$= \sum_{j=1}^{k_n}\left(E[X_{n_j}^2\mathbb{1}_{\{|X_{n_j}|>\delta\}}] + \left[E[X_{n_j}\mathbb{1}_{\{|X_{n_j}|>\delta\}}]\right]^2\right) \leq 2\sum_{j=1}^{k_n} E[X_{n_j}^2\mathbb{1}_{\{|X_{n_j}|>\delta\}}] \to 0,$$

by hypothesis *b)*, i.e. $\lim_{n\to\infty}(\sigma^2(S_n) - \sigma^2(S_{n,\delta})) = 0$. Since $\lim_{n\to\infty}\sigma_{n,\delta}^2 = \sigma^2$, $d_3(N(0,\sigma_{n,\delta}^2), N(0,\sigma^2)) \underset{n\to\infty}{\longrightarrow} 0$. Secondly, we see $d_3(\mathcal{L}(S_n), \mathcal{L}(S_{n,\delta} - E[S_{n,\delta}])) \underset{n\to\infty}{\longrightarrow} 0$. We will show that $S_n - (S_{n,\delta} - E[S_{n,\delta}]) \xrightarrow{\omega} 0$. Using that $E(S_n) = 0$,

$$E[(S_n - (S_{n,\delta} - E[S_{n,\delta}]))^2] = E[((S_n - S_{n,\delta}) - E[S_n - S_{n,\delta}])^2] = \sigma^2(S_n - S_{n,\delta})$$
$$= \sigma^2\left(\sum_{j=1}^{k_n}(X_{n_j} - X_{n_j,\delta})\right) = \sum_{j=1}^{k_n}\sigma^2(X_{n_j} - X_{n_j,\delta}) \leq \sum_{k=1}^{k_n} E[(X_{n_j} - X_{n_j,\delta})^2]$$
$$= \sum_{j=1}^{k_n} E[X_{n_j}^2\mathbb{1}_{\{|X_{n_j}|>\delta\}}] \underset{n\to\infty}{\longrightarrow} 0,$$

by *b)*. At last, we prove that $d_3(\mathcal{L}(S_{n,\delta} - E[S_{n,\delta}]), N(0,\sigma_{n,\delta}^2)) \underset{n\to\infty}{\longrightarrow} 0$. Observe that

- $S_{n,\delta} - E(S_{n,\delta}) = \sum_{j=1}^{k_n} X_{n_j,\delta} - E(X_{n_j,\delta})$,
- $|X_{n_j,\delta} - E(X_{n_j,\delta})| \leq |X_{n_j,\delta}| + |E(X_{n_j,\delta})| \leq 2\delta$,
- $\sigma^2(S_{n,\delta} - E(S_{n,\delta})) = \sigma_{n,\delta}^2$.

Therefore, by *Proposition 5.1.1*,

$$d_3(\mathcal{L}(S_{n,\delta} - E(S_{n,\delta})), N(0,\sigma_{n,\delta}^2)) \leq \frac{1}{6}\left(1 + \sqrt{\frac{8}{\pi}}\right)\sigma_{n,\delta}^2 2\delta.$$

Since $\lim_{n\to\infty}\sigma_{n,\delta}^2 = \sigma^2$, taking limits on $n \to \infty$, then

$$d_3(\mathcal{L}(S_{n,\delta} - E(S_{n,\delta})), N(0,\sigma_{n,\delta}^2)) \leq \frac{1}{6}\left(1 + \sqrt{\frac{8}{\pi}}\right)\sigma^2 2\delta \underset{\delta\to 0}{\longrightarrow} 0.$$ [2]

Plugging the three limits in (5.3.2), finally, $d_3(\mathcal{L}(S_n), N(0,\sigma^2)) \underset{n\to\infty}{\longrightarrow} 0$, as we wanted. ∎

[2] Note that in (5.3.3) we are taking the limit on $\delta \to 0$.

We will now study the reciprocal result: the Feller's Theorem. This provides a way to prove the Lindeberg's Condition, based on infinitesimal triangular families.

Definition 5.3.3. *A **triangular family** $\{X_{n_j}, 1 \leq j \leq k_n, k_n \nearrow \infty\}$ is **infinitesimal** if*

$$\lim_{n\to\infty} \max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\} = 0, \ \forall \epsilon > 0. \tag{5.3.3}$$

Corollary 5.3.4. *Let $\{X_{n_j}, 1 \leq j \leq k_n, k_n \nearrow \infty\}$ be a triangular family. If X_{n_j} satisfies the Lindeberg's Condition, then X_{n_j} is an infinitesimal triangular family.*

Proof. By hypothesis, $\sum_{j=1}^{k_n} E[X_{n_j}^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}] \underset{n\to\infty}{\longrightarrow} 0$, for every $\epsilon > 0$. Notice that

$$\sum_{j=1}^{k_n} E[X_{n_j}^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}] \geq \epsilon^2 \sum_{j=1}^{k_n} P\{|X_{n_j}| > \epsilon\} \geq \epsilon^2 \max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\} \geq 0.$$

Therefore,

$$\limsup_{n\to\infty} \sum_{j=1}^{k_n} E[X_{n_j}^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}] \geq \epsilon^2 \limsup_{n\to\infty} \max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\} \geq 0.$$

Putting everything together, $\max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\} \underset{n\to\infty}{\longrightarrow} 0$, as we wanted. ∎

Theorem 5.3.5. ***(Feller's Theorem).*** *Let $\{X_{n_j}, 1 \leq j \leq k_n, k_n \nearrow \infty\}$ be an infinitesimal triangular family with centered random variables such that:*

(a) $\sigma^2(S_n) = \sum_{j=1}^{k_n} E(X_{n_j}^2) \underset{n\to\infty}{\longrightarrow} \sigma^2$,

(b) $S_n = \sum_{j=1}^{k_n} X_{n_j} \overset{\omega}{\underset{n\to\infty}{\longrightarrow}} N(0, \sigma^2)$.

Then, X_{n_j} satisfies the Lindeberg's Condition.

Proof. The proof will consist of two parts. At the first part, we will show that

$$\sum_{j=1}^{k_n} E[|X_{n_j}| \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}] \underset{n\to\infty}{\longrightarrow} 0, \tag{5.3.4}$$

for every $\epsilon > 0$. Therefore, we will be able to see that for every $\epsilon > 0$,

$$\sum_{j=1}^{k_n} P\{|X_{n_j}| > \epsilon\} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.3.5}$$

At the second part, we will prove the Lindeberg's Condition. First, we see (5.3.4). Let μ be a measure in $\mathbb{R}$ such that $\frac{1}{|x|}$ is integrable. By the Jensen's Inequality,[3]

$$\int_{\mathbb{R}} \frac{1}{|x|} \, d\mu(x) \geq \frac{1}{\int_{\mathbb{R}} |x| \, d\mu(x)}. \tag{5.3.6}$$

[3] The Jensen's Inequality is reviewed in *Chapter 1*.

Applying (5.3.6) to the distribution of $|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}$, then

$$\int_{\mathbb{R}} \frac{1}{|x|} d\mathcal{L}\left(|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right) = E\left[\frac{1}{|X_{n_j}|^2} \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right] \geq E[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}]^{-1},$$

equivalently, $E\left[\frac{1}{|X_{n_j}|^2} \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]^{\frac{1}{2}} E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]^{\frac{1}{2}} \geq \sqrt{1} = 1$. Applying the Schwarz's Inequality in the first step, then

$$\begin{aligned} E\left[|X_{n_j}| \mathbb{1}^2_{\{|X_{n_j}|>\epsilon\}}\right] &\leq \left(E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]\right)^{\frac{1}{2}} \left(E\left[\mathbb{1}^2_{\{|X_{n_j}|>\epsilon\}}\right]\right)^{\frac{1}{2}} \\ &\leq \left(E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]\right)^{\frac{1}{2}} \left(P\{|X_{n_j}| > \epsilon\}\right)^{\frac{1}{2}} 1. \end{aligned}$$

Applying the previous inequalities,

$$E\left[|X_{n_j}| \mathbb{1}^2_{\{|X_{n_j}|>\epsilon\}}\right] \leq E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]^{\frac{1}{2}} \left(P\{|X_{n_j}| > \epsilon\}\right)^{\frac{1}{2}} E\left[\frac{1}{|X_{n_j}|^2} \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]^{\frac{1}{2}}$$

and using the inequality $\left(E\left[\frac{1}{|X_{n_j}|^2} \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]\right)^{\frac{1}{2}} \leq \left(\frac{1}{\epsilon^2} P\{|X_{n_j}| > \epsilon\}\right)^{\frac{1}{2}}$, then

$$\begin{aligned} E\left[|X_{n_j}| \mathbb{1}^2_{\{|X_{n_j}|>\epsilon\}}\right] &\leq \frac{1}{\epsilon} E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right] P\{|X_{n_j}| > \epsilon\} \\ &\leq \frac{1}{\epsilon} E(|X_{n_j}|^2) \max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\}. \end{aligned}$$

Taking sums on j and applying the hypothesis *(a)*, then

$$0 \leq \sum_{j=1}^{k_n} E[|X_{n_j}| \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}] \leq \frac{1}{\epsilon} \max_{1\leq j\leq k_n} P\{|X_{n_j}| > \epsilon\} \sum_{j=1}^{k_n} E(|X_{n_j}|^2) \underset{n\to\infty}{\longrightarrow} 0,$$

since also X_{n_j} is an infinitesimal triangular family. Finally, (5.3.4) is proved. We have also seen (5.3.5), by the definition of $E\left[|X_{n_j}| \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]$ and using (5.3.4). Now we can prove the second part. Fix $\delta > 0$, then define $X_{n_j,\delta} := X_{n_j} \mathbb{1}_{\{|X_{n_j}|\leq\delta\}}$. Then, denote $S_{n,\delta} = \sum_{j=1}^{k_n} X_{n_j,\delta}$ and $E(S_{n,\delta}) = \sum_{j=1}^{k_n} E(X_{n_j,\delta})$. On one hand, we will prove

$$S_{n,\delta} - E(S_{n,\delta}) \xrightarrow[n\to\infty]{\omega} N(0, \sigma^2). \tag{5.3.7}$$

On the other hand, define $\sigma_\delta^2 := \liminf_{n\to\infty} \sum_{j=1}^{k_n} E(X_{n_j}^2)$. Therefore, we will prove

$$S_{n,\delta} - E(S_{n,\delta}) \xrightarrow[n\to\infty]{\omega} N(0, \sigma_\delta^2). \tag{5.3.8}$$

Proving (5.3.7) and (5.3.8) and using the unicity of the limit in weak convergence, we will have $\sigma_\delta^2 = \sigma^2$. Since $|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\delta\}} = X_{n_j}^2 - X_{n_j,\delta}^2$,

$$\sum_{j=1}^{k_n} E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\delta\}}\right] = \sum_{j=1}^{k_n} E(|X_{n_j}|^2) - \sum_{j=1}^{k_n} E(|X_{n_j,\delta}|^2).$$

Then,

$$\begin{aligned} 0 &\leq \limsup_{n\to\infty} \sum_{j=1}^{k_n} E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}>\delta\}}\right] \\ &\leq \limsup_{n\to\infty} \sum_{j=1}^{k_n} E\left(|X_{n_j}|^2\right) - \limsup_{n\to\infty} \sum_{j=1}^{k_n} E\left(|X_{n_j,\delta}|^2\right) = \sigma^2 - \sigma_\delta^2 = 0, \end{aligned}$$

i.e. $\lim_{n\to\infty} \sum_{j=1}^{k_n} E\left[|X_{n_j}|^2 \mathbb{1}_{\{|X_{n_j}|>\delta\}}\right] = 0$. Finally, the Lindeberg's Condition is satisfied. If we prove (5.3.7) and (5.3.8), we will be done. Note that

$$\begin{aligned} P\{|S_{n,\delta} - S_n| > \epsilon\} \leq P\{S_{n,\delta} \neq S_n\} &\leq P\{|X_{n_j}| > \delta, \text{ for some } j \in \{1, ..., k_n\}\} \\ &\leq \sum_{j=1}^{k_n} P\{|X_{n_j}| > \delta\} \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

applying (5.3.5) in the limit and also

$$|E(S_{n,\delta})| = \left|\sum_{j=1}^{k_n} E\left[X_{n_j} \mathbb{1}_{\{|X_{n_j}|\leq\delta\}}\right]\right| = \left|\sum_{j=1}^{k_n} E\left[X_{n_j} \mathbb{1}_{\{|X_{n_j}|>\delta\}}\right]\right| \leq \sum_{j=1}^{k_n} E\left[X_{n_j} \mathbb{1}_{\{|X_{n_j}|>\delta\}}\right],$$

where the last term tends to 0 as $n \to \infty$ by (5.3.4). Therefore,

$$S_{n,\delta} - E(S_{n,\delta}) = S_{n,\delta} - S_n + S_n - E(S_{n,\delta}) \underset{n\to\infty}{\longrightarrow} 0 + N(0, \sigma^2) + 0 = N(0, \sigma^2).$$

We prove (5.3.8). If $\sigma_\delta^2 := \liminf_{n\to\infty} \sum_{j=1}^{k_n} E(X_{n_j,\delta}^2)$, then there exists a subsequence $\sum_{j=1}^{k_{n_i}} E(X_{n_j,\delta}^2) \underset{n\to\infty}{\longrightarrow} \sigma_\delta^2$. Construct $\{X_{n_j,\delta} - E(X_{n_j,\delta}),\ 1 \leq j \leq n,\ k_n \nearrow \infty\}$. We want to show it satisfies both hypotheses in the Lindeberg's Theorem:

(a) $\sigma^2(S_{n,\delta}) = \sum_{j=1}^{k_{n_i}} E(X_{n_j,\delta}^2) - [E(X_{n_j,\delta})]^2 \longrightarrow \sigma_\delta^2$,

(b) $\sum_{j=1}^{k_n} E\left[|X_{n_j,\delta} - E(X_{n_j,\delta})|^2 \mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right] \to 0$.

Then, $S_{n,\delta} - E(S_{n,\delta}) \underset{n\to\infty}{\longrightarrow} N(0, \sigma_\delta^2)$, as we wanted. First, we see a). Applying (5.3.4),

$$0 \leq \sum_{j=1}^{k_{n_i}} [E(X_{n_j,\delta})]^2 = \sum_{j=1}^{k_{n_i}} (E[X_{n_j} \mathbb{1}_{\{|X_{n_j}|>\delta\}}])^2 \leq \sum_{j=1}^{k_{n_i}} (E[|X_{n_j}| \mathbb{1}_{\{|X_{n_j}|>\delta\}}])^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Then,

$$\sum_{j=1}^{k_{n_i}}\left(E(X^2_{n_j,\delta})-[E(X_{n_j,\delta})]^2\right)=\sum_{j=1}^{k_{n_i}}E(X^2_{n_j,\delta})-\sum_{j=1}^{k_{n_i}}[E(X_{n_j,\delta})]^2\to\sigma_1^2+0=\sigma_1^2.$$

Now, we see b). Notice that

$$\{|X_{n_j,\delta}-E(X_{n_j,\delta})|>\epsilon\}\subset\{|X_{n_j,\delta}|>\frac{\epsilon}{2}\}\subset\{|X_{n_j}|>\frac{\epsilon}{2}\},$$

since $\sum_{j=1}^{k_n}|E(X_{n_j,\delta})|\to 0$, when $|E(X_{n_j,\delta})|<\frac{\epsilon}{2}$. Finally, using (5.3.5),

$$\begin{aligned}0&\le\sum_{j=1}^{k_{n_i}}E\left[|X_{n_j,\delta}-E(X_{n_j,\delta})|^2\mathbb{1}_{\{|X_{n_j,\delta}-E(X_{n_j,\delta})|>\epsilon\}}\right]\\&\le(2\delta)^2\sum_{j=1}^{k_{n_i}}P\left\{|X_{n_j,\delta}-E(X_{n_j,\delta})|>\epsilon\right\}\le(2\delta)^2\sum_{j=1}^{k_{n_i}}P\left\{|X_{n_j}|>\frac{\epsilon}{2}\right\}\underset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

■

5.4 Lyapunov's Theorem

In this section, we will study a generalized variation of the CLT, which is named after Aleksandr Lyapunov. It is also a sufficient condition for the CLT.

Theorem 5.4.1. ***(Lyapunov's Theorem).*** *Let $\{X_{n_j}, 1\le j\le k_n, k_n\nearrow\infty\}$ be a triangular family of centered random variables satisfying that*

(a) $\sum_{j=1}^{k_n}E(X^2_{n_j})\underset{n\to\infty}{\longrightarrow}\sigma^2$,

(b) ***(Lyapunov's Condition).*** *There is a $\delta>0$, such that $\sum_{j=1}^{k_n}E(|X_{n_j}|^{2+\delta})\underset{n\to\infty}{\longrightarrow}0$.*

Then,

$$S_n:=\sum_{j=1}^{k_n}X_{n_j}\overset{\omega}{\underset{n\to\infty}{\longrightarrow}}N(0,\sigma^2).\tag{5.4.1}$$

Proof. We will show that the Lyapunov's Condition implies the Lindeberg's Condition and we will be done. Notice that for every $\epsilon>0$, then

$$\sum_{j=1}^{k_n}E\left[|X_{n_j}|^{2+\delta}\right]\ge\sum_{j=1}^{k_n}E\left[|X_{n_j}|^{2+\delta}\mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]\ge\epsilon^2\sum_{j=1}^{k_n}E\left[|X_{n_j}|^2\mathbb{1}_{\{|X_{n_j}|>\epsilon\}}\right]\ge 0.$$

By hypothesis *b)* $\sum_{j=1}^{k_n}E(|X_{n_j}|^{2+\delta})\underset{n\to\infty}{\longrightarrow}0$ and ϵ is fixed, then $E[|X_{n_j}|^2\mathbb{1}_{\{|X_{n_j}|>\epsilon\}}]\to 0$, i.e. the Lindeberg's Condition is satisfied, as we wanted. ■

Observation 5.4.2. If the Lyapunov's CLT holds, then the Lindeberg's CLT also holds. However, the reverse is not always true. Consequently, the Lyapunov's CLT is stronger, although it is harder to check since it involves higher-order moments.

5.5 Multidimensional CLT

It might be interesting studying the multidimensional version of the classical CLT *(Subsection 5.2)*, since it is often used in Statistics. We will state and proof the multidimensional CLT.[4] The next result is a theorem-definition.[5]

Theorem 5.5.1. *Let $m \in \mathbb{R}^n$ and Λ be a $n \times n$ non-negative symmetric matrix. There is a probability in $\mathbb{R}^n$ named the **multivariate normal distribution** and with*

$$\varphi(t) := \exp\left(it^*m - \frac{1}{2}t^*\Lambda t\right), \tag{5.5.1}$$

for every $t \in \mathbb{R}^n$. This probability is denoted $N(m, \Lambda)$.

Proof. Let C be an orthogonal matrix s.t $C\Lambda C^* = D$, where D is a diagonal matrix. Let $\lambda_1, ..., \lambda_n$ be the elements of the principal diagonal of D. These are the eigenvalues of Λ and so non-negative, by hypothesis. Then, $\Lambda = C^*DC$. Let $\mathbf{Y} = (Y_1, ..., Y_n)$, where all the components are independent and have $N(0, \lambda_i)$ if $\lambda_i \neq 0$ or $Y_i \equiv 0$ if $\lambda_i = 0$, for every $i \in \{1, ..., n\}$. Notice that

$$\varphi_{\mathbf{Y}}(t) := E\left[\exp\left(i\sum_{j=1}^{n} t_j Y_j\right)\right] = \prod_{j=1}^{n} e^{-\frac{1}{2}tj^2\lambda_j} = e^{-\frac{1}{2}t^*Dt}.$$

Consider $\mathbf{X} := C^*\mathbf{Y} + m$. Consequently, the characteristic function of $\mathbf{X}$ is exactly

$$\varphi_{\mathbf{X}}(t) := e^{it^*m}\varphi_{\mathbf{Y}}(Ct) = e^{it^*m}e^{-\frac{1}{2}(Ct)^*DCt} = e^{it^*m-\frac{1}{2}t^*\Lambda t}.$$

∎

We see now the main properties of the Multidimensional Normal Distribution.

1) Let $\mathbf{X} \sim N(m, \Lambda)$ and C be orthogonal s.t $\Lambda = C^*DC$. Then, $\mathbf{Y} := C(\mathbf{X} - m)$ has independent components with $N(0, \lambda_i)$ if $\lambda_i \neq 0$ or equal to 0 if $\lambda_i = 0$.

   Proof. A consequence of *Proposition 5.5.1* and

   $$\varphi_{\mathbf{Y}}(t) := e^{it^*(-Cm)}\varphi_{\mathbf{X}}(C^*t) = \exp(-it^*Cm + it^*Cm - \frac{1}{2}t^*C\Lambda C^*t).$$

   ∎

2) Let $\mathbf{X} \sim N(m, \Lambda)$, where m, Λ are the vector of means and the matrix of variances and covariances of $\mathbf{X}$, respectively.

[4] The same notation will be used throughout this section.
[5] A theorem-definition combines the role of a definition and of a theorem at the same time.

Proof. Check that $E(\mathbf{X}) = C^*E(\mathbf{Y}) + m = m$ and also

$$E[(\mathbf{X}-m)(\mathbf{X}-m)^*] = E[C^*\mathbf{Y}\mathbf{Y}^*C] = C^*DC = \Lambda,$$

∎

3) Let $\mathbf{X} \sim N(m, \Lambda)$ and n-dimensional. Let A be a $r \times n$ matrix. Then, the random vector $\mathbf{AX}$ has distribution $N(Am, A\Lambda A^*)$.

Proof. For every $t \in \mathbb{R}^r$, it is a consequence of

$$\begin{aligned}\varphi_{\mathbf{AX}}(t) :=& \varphi_{\mathbf{X}}(A^*t) = \exp\left(it^*Am - \frac{1}{2}(A^*t)^*\Lambda(A^*t)\right)\\ =& \exp\left(it^*(Am) - \frac{1}{2}t^*(A\Lambda A^*)t\right).\end{aligned}$$

∎

In particular, every random vector $(X_{i_1}, ..., X_{i_m})$ with $m \le n$ has a normal distribution and also every linear combination $\sum_{i=1}^n a_i X_i$ does. Reciprocally, if $\mathbf{X} = (X_1, ..., X_n)$ is s.t every linear combination $\sum_{i=1}^n a_i X_i$ is normal, then $\mathbf{X}$ has a normal multidimensional distribution.

Proof. Since $E(e^{it^*\mathbf{X}}) = \varphi_{t^*\mathbf{X}}(1) = e^{-\frac{1}{2}\sigma^2(t^*\mathbf{X})+iE(t^*\mathbf{X})}$, then $E(t^*\mathbf{X}) = t^*E(\mathbf{X})$ and

$$\sigma^2(t^*\mathbf{X}) := E[[t^*(\mathbf{X}-E(\mathbf{X}))]^2] = E[(t^*(\mathbf{X}-E(\mathbf{X}))((\mathbf{X}-E(\mathbf{X}))^*t] = t^*\Lambda t.$$

Therefore, $\mathbf{X} \sim N(m, \Lambda)$, with $m := E(\mathbf{X})$ and $\Lambda := E[(\mathbf{X}-m)(\mathbf{X}-m)^*]$. ∎

4) Let $\mathbf{X} = (X_1, ..., X_n) \sim N(m, \Lambda)$. The independence of the r.v $X_1, ..., X_n$ is equivalent to the matrix Λ being diagonal, i.e $X_1, ..., X_n$ are uncorrelated.

Proof. If $X_1, ..., X_n$ are independent, then $\mathrm{Cov}(X_i, X_j) = 0$, for every $i \neq j$. Reciprocally, if Λ is a diagonal matrix and $\lambda_1, ..., \lambda_n$ denote the components of the diagonal, then

$$\varphi_{\mathbf{X}}(t) := e^{it^*m - \frac{1}{2}t^*\Lambda t} = \prod_{j=1}^n e^{it_j m_j - \frac{1}{2}\lambda_j t_j^2} = \prod_{j=1}^n \varphi_{X_j}(t_j).$$

∎

5) If Λ is an invertible (or non-singular) matrix (i.e. $det(\lambda) > 0$), then the **normal distribution** $N(m, \Lambda)$ is **non-degenerate**.

Proof. By hypothesis, $\lambda_i > 0$, for every $i \in \{1, ..., n\}$. Define $\varphi(y) := C^* y + m = x$. Since $\mathbf{X} := \varphi(\mathbf{Y})$, [6]

$$f_{\mathbf{X}}(\mathbf{X}) = f_{\mathbf{Y}}(\varphi^{-1}(\mathbf{X})) \left|J_\varphi(y)\right|^{-1} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{1}{2\lambda_i} y_i^2}$$
$$= [(2\pi)^n \det(\Lambda)]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-m)^* \Lambda^{-1}(x-m)\right),$$

using that $y^* D^{-1} y = (x-m)^* C^* (x-m) = (x-m)^* \Lambda^{-1}(x-m)$ and also $|J_\varphi(y)| = |detC^*| = 1, \quad \lambda_1 ... \lambda_n = \det(\Lambda)$. ∎

On the other hand, if Λ is a singular matrix, then $N(m, \Lambda)$ is **degenerate**.

Proof. Assume that the rang of Λ is $r < n$. Let $\lambda_1, ..., \lambda_r$ be the non-zero eigenvalues of Λ. Therefore, $Y_{r+1} = ... = Y_n = 0$ and the random vector $\mathbf{Z} = (Y_1, ..., Y_r)$ has an r-dimensional non-degenerate normal distribution. ∎

Finally, we can study the multidimensional CLT. The theorem states that sums converge to a multivariate normal distribution as the sample increases.

Theorem 5.5.2. ***(Multidimensional CLT).*** *Let $\{\boldsymbol{X}_n, n \geq 1\}$ be a sequence of random vectors k-dimensional, independent and identically distributed. Let $\boldsymbol{S}_n := \boldsymbol{X}_1 + ... + \boldsymbol{X}_n$. Suppose that the components of $\boldsymbol{X}_1$ have finite second-order moment. Denote $m := E(\boldsymbol{X}_1)$ and $\Lambda := E[(\boldsymbol{X}_1 - m)(\boldsymbol{X}_1 - m)^*]$. Then,*

$$\frac{\boldsymbol{S}_n - nm}{\sqrt{n}} \xrightarrow[n\to\infty]{\omega} N(0, \Lambda). \tag{5.5.2}$$

Proof. Define the random vector $\mathbf{Y}_n := \frac{S_n - nm}{\sqrt{n}}$. Fix $t \in \mathbb{R}^k$, then $\{t^* \mathbf{X}_n, n \geq 1\}$ is a sequence of independent and identically distributed random variables. Their mean is $E(t^* \mathbf{X}_n) = t^* m$ and their variance is

$$\sigma^2(t^* \mathbf{X}_1) = E[t^*(\mathbf{X}_1 - m)(\mathbf{X}_1 - m)^* t] = t^* \Lambda t.$$

Applying the Lévy-Lindeberg's CLT *(Theorem 5.2.1)*, then

$$t^* \mathbf{Y}_n = \frac{1}{\sqrt{n}} \left[\sum_{j=1}^{n} t^* \mathbf{X}_j - n t^* m\right] \xrightarrow[n\to\infty]{\omega} N(0, t^* \Lambda t).$$

Therefore,

$$\varphi_{\mathbf{Y}_n}(t) := E[e^{it^* \mathbf{Y}_n}] = \varphi_{t^* \mathbf{Y}_n}(1) \xrightarrow[n\to\infty]{\omega} \varphi_{N(0, t^* \Lambda t)}(1) = e^{-\frac{1}{2} t^* \Lambda t} = \varphi_{N(0,\Lambda)}(t),$$

for every $t \in \mathbb{R}^k$, i.e. the characteristic function of $\mathbf{Y}_n$ converges to the characteristic function of the normal distribution $N(0, \Lambda)$. Finally, by the multidimensional version of the Paul-Lévy's Continuity Theorem *(Theorem 4.2.7)*, $\mathbf{Y}_n \xrightarrow[n\to\infty]{\omega} N(0, \Lambda)$. ∎

[6] The transformation of random variable's formula is applied to $\mathbf{X}$.

5.6 Convergence of the compound Poisson distribution

In this section, we aim to consider sequences where the underlying distribution of the r.v. does not meet the hypothesis of the CLT. For instance, sequences with distributions that do not have finite variance, such as the Cauchy distribution. In particular, we will focus on the compound Poisson distribution.

Definition 5.6.1. *Let μ be a finite measure in $\mathbb{R}$. The* ***compound Poisson distribution*** *is given by*

$$Poiss(\mu)(B) := e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{\mu * ... * \mu(B)}{k!}, \tag{5.6.1}$$

for every $B \in \mathcal{B}(\mathbb{R})$. It is denoted by $Poiss(\mu)$ and the first term will be $\delta_{\{0\}}(B)$.

Proposition 5.6.2. *The compound Poisson distribution is a probability.*

Proof. First of all, note that $Poiss(\mu)(B) \geq 0$, for every $B \in \mathcal{B}(\mathbb{R})$. Secondly, observe that $Poiss(\mu)(\emptyset) = 0$ and also

$$Poiss(\mu)(\mathbb{R}) = e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{\mu * ... * \mu(\mathbb{R})}{k!} = e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{(\mu(\mathbb{R}))^k}{k!} = 1.$$

Finally, we show that the function is σ-additive. Take $A \in \mathcal{B}(\mathbb{R})$ with $A = \bigcup_{i=1}^{\infty} A_i$, where A_i are pairwise disjoint sets. By the additivity of convolution of measures,

$$\begin{aligned} Poiss(\mu)(A) &= e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{\mu * ... * \mu(A)}{k!} = e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \frac{\mu * ... * \mu(A_i)}{k!} \\ &= \sum_{i=0}^{\infty} e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{\mu * ... * \mu(A_i)}{k!} = \sum_{i=0}^{\infty} Poiss(\mu)(A_i). \end{aligned}$$

∎

We will now see that the sum of a sequence of r.v. with a Poisson distribution is exactly a compound Poisson distribution.

Proposition 5.6.3. *Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed r.v such that $X_i \sim v$, for every $i \in \{1, ..., n\}$. Let N be a random variable such that $N \sim Poiss(\lambda)$ and N is independent to X_i, for every $i \in \{1, ..., n\}$. Then,*

$$S := \sum_{i=1}^{N} X_i \sim Poiss(\mu), \tag{5.6.2}$$

where $\mu := \lambda v$ and v is a probability. In particular, if also $N \equiv 0$, then $S \equiv 0$.

Proof. Consider $B \in \mathcal{B}(\mathbb{R})$. Hence, using in the last equality $\mu(\mathbb{R}) = \lambda v(\mathbb{R}) = \lambda$ and $\lambda^k v * ... * v = (\lambda v) * ... * (\lambda v)$,

$$\begin{aligned} P\{S \in B\} &= P\left\{\sum_{i=1}^{N} X_i \in B\right\} = \sum_{k=0}^{\infty} P\left\{\sum_{i=1}^{N} X_i \in B,\ N = k\right\} \\ &= \sum_{k=0}^{\infty} P\left\{\sum_{i=1}^{N} X_i \in B\right\} P\{N = k\} = \sum_{k=0}^{\infty} P\left\{\sum_{i=1}^{k} X_i \in B\right\} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{v * ... * v(B)}{k!} \lambda^k = e^{-\mu(\mathbb{R})} \sum_{k=0}^{\infty} \frac{\mu * ... * \mu(B)}{k!}. \end{aligned}$$

∎

Now we may compute the characteristic function of the compound Poisson.

Proposition 5.6.4. *The characteristic function of the compound Poisson S is*

$$\varphi_S(t) := \exp\left(\int_{\mathbb{R}} (e^{itx} - 1)\, d\mu(x)\right). \tag{5.6.3}$$

Proof. By the definition of conditional expectation,[7] then

$$\begin{aligned} \varphi_{Poiss(\mu)}(t) &= E(e^{itS}) = E[E(e^{itS}|N)] = \sum_{k=0}^{\infty} P\{N = k\} E(e^{itS}|N = k) \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} E\left[e\left(it \sum_{i=1}^{k} X_i\right)\right] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} (\varphi_X(t))^k = \exp(-\lambda + \lambda\varphi_X(t)) \\ &= \exp\left(\int_{\mathbb{R}} (e^{itx} - 1)\right)\, d\mu(x), \end{aligned}$$

using that $\lambda\varphi_X(t) = \lambda \int_{\mathbb{R}} e^{itx}\, dx = \int_{\mathbb{R}} e^{itx}\, d\mu(x)$ in the last equality. ∎

We will see some basic properties of the compound Poisson before the final result.

Proposition 5.6.5. *Let $\{\mu_n, n \geq 1\}, \mu$ be probability measures. Then, it is satisfied that*

1) *If $\mu_n \xrightarrow[n\to\infty]{\omega} \mu$, then also $Poiss(\mu_n) \xrightarrow[n\to\infty]{\omega} Poiss(\mu)$.*

2) *$Poiss(\mu) = Poiss(\mu_1) * ... * Poiss(\mu_n)$, where $\mu = \sum_{i=1}^{n} \mu_i$.*

3) $\sup_{A \in \mathcal{B}(\mathbb{R})} |Poiss(\mu)(A) - \mu(A)| \leq (\mu(\mathbb{R} - 0))^2$.

[7] The concept of conditional expectation can be reviewed in *([4], Chapter 6)*.

Proof. We will begin proving 1). Since $\mu_n \xrightarrow[n\to\infty]{\omega} \mu$, then $\mu_n(\mathbb{R}) \longrightarrow \mu(\mathbb{R})$ and $\varphi_{\mu_n}(t) \longrightarrow \varphi_\mu(t)$, for every $t \in \mathbb{R}$. Hence,

$$\begin{aligned}\varphi_{Poiss(\mu_n)}(t) &= \exp\left(\int_{\mathbb{R}} (e^{itx}-1)\, d\mu_n(x)\right) = \exp(\varphi_{\mu_n}(t) - \mu_n(\mathbb{R}))\\ &\xrightarrow[n\to\infty]{} \exp(\varphi_\mu(t) - \mu(\mathbb{R})) = \varphi_{Poiss(\mu)}(t).\end{aligned}$$

Now we will show 2). Note that

$$\begin{aligned}\varphi_{Poiss(\mu)}(t) &= \exp\left(\int_{\mathbb{R}} (e^{itx}-1)\, d\mu(x)\right) = \exp\left(\int_{\mathbb{R}} (e^{itx}-1)\, d\sum_{i=1}^{\infty}\mu_i(x)\right)\\ &= \exp\left(\sum_{i=1}^{\infty}\int_{\mathbb{R}} (e^{itx}-1)\, d\mu_i(x)\right) = \prod_{i=1}^{n} \varphi_{Poiss\mu_i}(t).\end{aligned}$$

Let $A \in \mathcal{B}(\mathbb{R})$. First of all, we want to prove both inequalities to prove 3). Since this will be shown for every $A \in \mathcal{B}(\mathbb{R})$, then taking the supremum of A, $\sup_{A\in\mathcal{B}(\mathbb{R})} |Poiss(\mu)(A) - \mu(A)| \leq (\mu(\mathbb{R}-0))^2$. Observe the following. Let v be a measure s.t $v(A) = \mu(A) - \mu(\{0\})\delta_{\{0\}}(A)$. Note that $v = \mu_{|\mathbb{R}-\{0\}}$. Assume $v(\mathbb{R}) = \epsilon \geq 0$, then also $(\mu(\mathbb{R}-\{0\}))^2 = (v(\mathbb{R}))^2 = \epsilon^2$. Furthermore, note that $1 = \mu(\mathbb{R}) = v(\mathbb{R}) + \mu(\{0\})\delta_{\{0\}}(A) = \epsilon + \mu(\{0\})$, i.e. $\mu(\{0\}) = 1-\epsilon$. Consequently, $\mu(A) = v(A) + (1-\epsilon)\delta_{\{0\}}(A)$. We will first suppose the inequality $Poiss(\mu)(A) - \mu(A) \geq 0$. Applying the previous observation, then

$$\begin{aligned}0 &\leq Poiss(\mu)(A) - \mu(A) \leq Poiss(v)(A) - v(A) - (1-\epsilon)\delta_{\{0\}}(A)\\ &= e^{-v(\mathbb{R})}\left(\sum_{k=0}^{\infty} \frac{v * ... * v(A)}{k!}\right) - v(A) - (1-\epsilon)\delta_{\{0\}}(A)\\ &= e^{-\epsilon}\left(\delta_{\{0\}}(A) + v(A) + \sum_{k=2}^{\infty} \frac{v * ... * v(A)}{k!}\right) - v(A) - (1-\epsilon)\delta_{\{0\}}(A) \qquad (5.6.4)\\ &\leq (e^{-\epsilon} - 1 + \epsilon)\delta_{\{0\}}(A) + (e^{-\epsilon}-1)v(A) + e^{-\epsilon}\sum_{k=2}^{\infty}\frac{\epsilon^k}{k!}\\ &= (e^{-\epsilon} - 1 + \epsilon)\delta_{\{0\}}(A) + (e^{-\epsilon}-1)v(A) + e^{-\epsilon}\left(\sum_{k=1}^{\infty}\frac{\epsilon^k}{k!} - \epsilon\right).\end{aligned}$$

Let $f(x) = e^{-x} - 1 + x$ s.t. $f(0) = 0$ and $f'(x) = e^{-x} + 1 \geq 0$, then f is increasing and taking $x = \epsilon \geq 0$, $e^{-\epsilon} - 1 + \epsilon \geq 0$, i.e. $\epsilon \geq 1 - e^{-\epsilon}$. Applying in (5.6.4) that the first term is positive and suppressing the second term since $e^{-\epsilon} - 1 \leq 0$ and $v(A) \geq 0$, then

$$\begin{aligned}0 &\leq Poiss(v)(A) - v(A) - (1-\epsilon)\delta_{\{0\}}(A) \leq e^{-\epsilon} - 1 + \epsilon + e^{-\epsilon}(e^{\epsilon} - 1 + \epsilon)\\ &\leq e^{-\epsilon} - 1 + \epsilon + 1 - e^{-\epsilon} - \epsilon e^{-\epsilon} = \epsilon(1 - e^{-\epsilon}) \leq \epsilon^2,\end{aligned}$$

as we wanted. On the other hand, suppose $(\mu(A) - Poiss(\mu(A))) \leq 0$. Similarly,

$$
\begin{aligned}
0 &\leq v(A) + (1-\epsilon)\delta_{\{0\}}(A) - e^{-\epsilon}\left(\delta_{\{0\}}(A) + v(A) + \sum_{k=2}^{\infty}\frac{v * ... * v(A)}{k!}\right) \\
&\leq (1-e^{-\epsilon})v(A) + (1-\epsilon-e^{-\epsilon})\delta_{\{0\}}(A) - e^{-\epsilon}\sum_{k=2}^{\infty}\frac{v * ... * v(A)}{k!} \\
&\leq (1-e^{-\epsilon})v(A) \leq \epsilon v(\mathbb{R}) \leq \epsilon^2,
\end{aligned}
$$

applying in the second-to-last inequality that $(1-\epsilon-e^{-\epsilon}) \leq 0$ and the last term is also negative. Finally, we are done. ∎

The compound Poisson is used to approximate random variables of a sum of random variables that are almost always different to 0. At last, we are able to relate a serie of a compound Poisson to the distribution of a sum of r.v.

Proposition 5.6.6. *Let $X_1, ..., X_n$ be independent random variables. Define $S_n := \sum_{i=1}^n X_i$ and $\epsilon_k = P\{X_k \neq 0\}$. Then,*

$$
\sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss\left(\sum_{i=1}^n \mathcal{L}(X_i)\right)(A) - \mathcal{L}(S_n)(A)\right| \leq \sum_{i=1}^n \epsilon_i^2. \tag{5.6.5}
$$

Proof. We will do the prove by induction. First of all, applying the property 2) of *Proposition 5.6.5*, then

$$
\begin{aligned}
&\sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss\left(\sum_{i=1}^n \mathcal{L}(X_i)\right)(A) - \mathcal{L}(S_n)(A)\right| \\
&= \sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss(\mathcal{L}(X_1)) * ... * Poiss(\mathcal{L}(X_n))(A) - \mathcal{L}(X_1) * ... * \mathcal{L}(X_1)(A)\right|.
\end{aligned}
$$

We want to prove that

$$
\begin{aligned}
&\sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss(\mathcal{L}(X_1)) * ... * Poiss(\mathcal{L}(X_n))(A) - \mathcal{L}(X_1) * ... * \mathcal{L}(X_1)(A)\right| \\
&\leq \sum_{i=1}^n \sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss(\mathcal{L}(X_i))(A) - \mathcal{L}(X_i)(A))\right|.
\end{aligned} \tag{5.6.6}
$$

Now we show that the statement holds for $n = 2$. Note that[8]

$$
\begin{aligned}
Z_n &= \sup_{A\in\mathcal{B}(\mathbb{R})}\left|Poiss(\mathcal{L}(X_1) + \mathcal{L}(X_2))(A) - \mathcal{L}(X_1 + X_2)(A)\right| \\
&= \left|\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_A(x+y)\, dPoiss(\mathcal{L}(X_1))(x) dPoiss(\mathcal{L}(X_2))(y)\right. \\
&\quad \left. - \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_A(x+y)\, d(\mathcal{L}(X_1))(x) d(\mathcal{L}(X_2))(y)\right|.
\end{aligned}
$$

[8]Note that since we are taking the supremum in $A \in \mathcal{B}(\mathbb{R})$, it is the same to consider either the indicator $\mathbb{1}_{A-x}(y)$ or $\mathbb{1}_A(y)$.

Continue with

$$
\begin{aligned}
Z_n = &\left| \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{1}_A(x+y)(dPoiss\mathcal{L}(X_2)(y) - d\mathcal{L}(X_2)(y)) \right] dPoiss\mathcal{L}(X_1)(x) \right. \\
&\left. - \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{1}_A(x+y)(dPoiss\mathcal{L}(X_1)(x) - d\mathcal{L}(X_1)(x)) \right] dPoiss\mathcal{L}(X_2)(y) \right| \\
\leq &\int_{\mathbb{R}} \sup_{A\in\mathcal{B}(\mathbb{R})} \left\{ \int_{\mathbb{R}} \mathbb{1}_A(x+y) |dPoiss\mathcal{L}(X_2)(y) - d\mathcal{L}(X_2)(y)| \; dPoiss\mathcal{L}(X_1)(x) \right\} \\
&+ \int_{\mathbb{R}} \sup_{A\in\mathcal{B}(\mathbb{R})} \left\{ \int_{\mathbb{R}} \mathbb{1}_A(x+y) |d\mathcal{L}(X_1)(x) - dPoiss(\mathcal{L}(X_1))(x)| \right\} d\mathcal{L}(X_2)(y) \\
\leq &\int_{\mathbb{R}} \sup_{A\in\mathcal{B}(\mathbb{R})} \left\{ \int_{\mathbb{R}} \mathbb{1}_A(y) |dPoiss\mathcal{L}(X_2)(y) - d\mathcal{L}(X_2)(y)| \; dPoiss\mathcal{L}(X_1)(x) \right\} \\
&+ \int_{\mathbb{R}} \sup_{A\in\mathcal{B}(\mathbb{R})} \left\{ \int_{\mathbb{R}} \mathbb{1}_A(y) |d\mathcal{L}(X_1)(x) - dPoiss(\mathcal{L}(X_1))(x)| \right\} d\mathcal{L}(X_2)(y).
\end{aligned}
$$

Finally, integrating,

$$
\begin{aligned}
&\sup_{A\in\mathcal{B}(\mathbb{R})} |Poiss(\mathcal{L}(X_1) + \mathcal{L}(X_2))(A) - \mathcal{L}(X_1 + X_2)(A)| \\
&\leq \sup_{A\in\mathcal{B}(\mathbb{R})} |Poiss(\mathcal{L}(X_2))(A) - \mathcal{L}(X_2)(A)| + \sup_{A\in\mathcal{B}(\mathbb{R})} |\mathcal{L}(X_1)(A) - Poiss(\mathcal{L}(X_1))(A)|.
\end{aligned}
$$

The induction step consists of proving (5.6.6), which is done following the same arguments. Applying the property 3) from *Proposition 5.6.5*, then

$$
\begin{aligned}
&\sup_{A\in\mathcal{B}(\mathbb{R})} \left| Poiss\left(\sum_{i=1}^{n} \mathcal{L}(X_i) \right)(A) - \mathcal{L}(S_n)(A) \right| \\
&\leq \sum_{i=1}^{n} \sup_{A\in\mathcal{B}(\mathbb{R})} |Poiss(\mathcal{L}(X_i))(A) - \mathcal{L}(X_i)(A))| \leq \sum_{i=1}^{n} |\mathcal{L}(X_i)(\mathbb{R} - 0)|^2 \\
&= \sum_{i=1}^{n} (P\{X_i \neq 0\})^2 = \sum_{i=1}^{n} \epsilon_i^2.
\end{aligned}
$$

■

In conclusion, although the CLT is a core theorem in probability theory and many other fields, it is important to recognize the limitations and conditions under which it holds. For instance, r.v. with infinite variance and distributions that deviate from the normality. Hence, in this case, it might be more appropriate to use alternative methods and distributions. With this theme, a new extensive work could begin, but we will leave it for another time.

Bibliography

[1] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, (2013), 159–243.

[2] D. Nualart, M. Sanz, *Curs de probabilitats*, PPU (Promociones y Publicaciones Universitarias, S.A.), Edition 1, (1990).

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., Volume 1, Edition 3, (1950), 1–26, 174–196, 212–264.

[4] P. Billingsley, *Probability and Measure*, John Wiley & Sons, Inc., Edition 3, (1995), 67–92, 254–297, Chapter 5.

[5] D. D. Wackerly, W. Mendenhall III, R. L. Scheaffer, *Mathematical Statistics with Applications*, Thomson Brooks / Cole, Edition 7, (2008), 20–157, 178–185, 223–296, 346–390.

[6] M. H. DeGroot, M. J. Schervish, *Probability and Statistics*, Pearson Education, Inc., Edition 4, (2012), 1–275, 302–316, 347–376.

[7] M. Sanz i Solé, *Probabilitats*, Universitat de Barcelona, (1999).