MASTERS

MATHEMATICAL FINANCE

I.S.E.G. Lisbon School of Economics and Management

MASTERS FINAL WORK

DISSERTATION

Central Limit Theorem Variations

Author:

Daniel VITAL DE ALCÂNTARA

Supervision:

João LOPES DIAS

December 15, 2019

"You know that I write slowly. This is chiefly because I am never satisfied until I have said as much as possible in a few words, and writing briefly takes far more time than writing at length."

Carl Friedrich Gauss

UNIVERSIDADE DE LISBOA

Abstract

I.S.E.G. Lisbon School of Economics and Management
Department of Mathematics

Master

Central Limit Theorem Variations

by Daniel VITAL DE ALCÂNTARA

One of the most important theorems of Probability Theory is the Central Limit Theorem. It states that if X_n is a sequence of random variables then the normalized partial sums converge to a normal distribution. This result omits any rate of convergence. Furthermore the lack of assumptions makes us wonder if some generalizations are possible.

Particularly in this essay we will focus on two questions: **Does it exist a (universal) rate of convergence for the Central Limit Theorem? Furthermore in which circumstances can we apply the Central Limit Theorem?**

The Lévy Continuity Theorem states that convergence on distribution functions is equivalent to convergence on characteristic functions. Furthermore when we apply Taylor expansions to characteristic functions we get a polynomial with the moments as coefficients. For these reasons, on our case computing with characteristic functions is preferable.

By the Berry Essen Theorem we can in fact find the rate of convergence we are looking for. And by the Lindeberg Theorem and Lyapunov Condition we find that the Central Limit Theorem applies to sequences that are not identically distributed. Finally, using the Ergodic Theorem we will explain how stochastic processes are related to Ergodic Theory. With this we will show how this theorem can be used to find a result when the sequence is not independent.

Knowing the distribution of the sum of a sequence of random variables is very useful in applications. However we rarely have infinite random variables to sum. Therefore knowing the rate of convergence allows us to decide when it is acceptable to assume the variable is normally distributed. Furthermore knowing that the Central Limit Theorem holds even when some assumptions on the variables are relaxed may help us in many applications where we do not have i.i.d sequences.

Acknowledgements

Chiefly I would like to thank professor João Lopes Dias - my supervisor - for all the advice and guidance provided for this thesis. I would also like to thank my parents for all the support they have given me. Specially the opportunity to dedicate myself full time to this thesis. I will also take this opportunity to thank my friends for always believing in my success, and to all my professors that helped me achieve it. Finally I am grateful to Bárbara Cotrim Rodrigues for proof reading parts of the text.

Contents

Introduction

The Central Limit Theorem is one of the major results of Probability Theory. Many of the concepts in Probability theory drew inspiration from casinos, card games, dices and coin tosses. The Bernoulli distribution is a mathematical model for a coin toss, while the (discrete) Uniform distribution is a generalization of a roll of a dice. We will also begin our study of the Central Limit Theorem by drawing inspiration from simple examples.

Let us start by the following example: Most board games require two - fair six sided - dice to be played. A reasonable question is: why not build a twelve sided dice instead? A fair twelve sided dice is easy to understand. We are as likely to get a four as a seven or a one - or any other number for that matter. However with two six sided dice this process is different. You sum the observed values. On each dice you can observe one through six. Hence you have thirty six (6×6) possible outcomes. Since the dice are fair all these outcomes are equally likely. They are divided into eleven events - two through twelve. Each corresponding to a possible sum of both dices. Let us call one dice *dice A* and another *dice B*. A possible outcome is dice A yielding one and dice B yielding three. In this case we got the event four ($1 + 3$). Another possible outcome is if both dice yielded two. We would also get the event four ($2 + 2$). However if we got the event two, we would know the outcome i.e. what each dice yielded. Contrary to four - for which there are a few possibilities for each dice - if we get a two the only option is if both dice yielded one. This means that there is an higher probability to get a four than to get a two. Since there are more outcomes for which we observe a four. If this is not convincing enough: see example (1.2). Or - with a lot of determination - throw two six sided dice one hundred times and write down the sum of the faces observed.
Therefore if you throw a twelve sided dice the result is completely random. All numbers are equally likely to be observed. However if you throw two six sided dice you have some insight on the result. In some sense it is *less random* to have two six sided dice than one twelve sided dice. Given this a reasonable question is: what about an eighteen sided dice and three six sided dice? Are the three six sided dice even *less random* than the eighteen sided dice? We could go even further and wonder about one million six sided dice and a six million sided dice. How would they compare? Would the six sided dice be *much less random* than the six million sided dice? That is exactly the question answered by the Central Limit Theorem and we will explore it in this essay. This gives us the first intuition of what Central Limit Theorem is.

Let us try another example: Consider your favorite team based sport. In most of these sports it is quite hard to predict - with high certainty - which team is going to win a given match. This is one of the reasons why there are websites that allow people to bet on which team they think is going to win. This shows the unpredictability of such matches. This unpredictability is one of the reasons sports are appealing to so many people. We are never sure who is going to win. We try to analyze every decision in an attempt to predict the result. This contrasts with the predictability of

seasons. It is much easier to predict with high certainty which team is going to win a given season. Consider one of the most popular sports in the world: football. It would not come as a shock if I predicted that F.C. Barcelona or Real de Madrid will win the next Spanish football season. This is even true for very balanced leagues like the Premier League in England. From the twenty teams competing if I claim the next champion will be among the teams: Arsenal, Chelsea, Liverpool, Manchester United and Manchester City, no one would be surprised.
In the season of 1995/1996 of the Premier League the number of teams competing was reduced from twenty two to twenty and has been twenty ever since. That is twenty three years of twenty teams competing against each other. Each team plays with all other nineteen teams twice. Therefore there are thirty eight games to be played by each team.
Let me propose a game: Consider a team, for our example we will use Manchester United. Then you pick a season from the 1995/1996 to 2017/2018 and I have three chances to guess how many of the thirty eight games played the team did not lose (i.e. it won or drawn). How hard is it for me to guess it correctly? One would expect for me to fail quite often. If the number of games a team did not lose was (uniformly) random then one would expect for me to be right about 1 in each 38 times (approximately 2.6%) since I had three guesses - and assuming I was smart enough not to guess the same number twice - one would expect me to win about 3 in each 38 times (approximately 7.8%). Even if we said that we are sure that Manchester United is a decent team that loses less than half the games we would still be looking at me winning 3 in 19 times (approximately 16%).
Then you pick the season and whichever season you pick my guess would be that Manchester United did not lose 33, 32 or 31 games i.e Manchester United lost either 5, 6 or 7 games in that particular season. Would you care to guess how often I will be right? If we played enough times you would find that of all the seasons I would be correct in 14 of those 23 seasons (an astonishing 61%).

These examples suggest a certain type of long term predictability. Even if we do not know the result of a given experience if we repeat it enough times we get a good insight about the sum of all results. Nowadays the knowledge of the long term predictability of systems like these is known as the Central Limit Theorem, which is the main focus of this essay. We will explore when and how can we apply the Central Limit Theorem. The main questions I aim to answer in this thesis are: Is there a (universal) rate of convergence for the Central Limit Theorem? Furthermore in which circumstances can we apply the Central Limit Theorem?

In the first chapter the reader may find an introduction to Central Limit Theorem: the main definitions, some examples and a proof of the Theorem. A reader that has studied the Central Limit Theorem before may wish to skip this chapter. Chapters 2, 3 and 4 are independent and may be read in whichever order is preferred. Chapter 2 studies rates of convergence for the Central Limit Theorem. Chapter 3 studies if it is possible to prove a Central Limit Theorem when the sequence of variables is not identical. Finally Chapter 2 studies a proof of the Central Limit Theorem for non independent sequences (under some special circumstances).

Chapter 1

Classic Central Limit Theorem

On the first section of this chapter we will review some concepts that are important to understand the rest of this essay. Most of these concepts are basic measure theory properties and most readers may already be familiar with them. We will go through the properties in order to recall them, as such most proofs of this section are omitted, and references for the proofs will be provided when possible. On the second section we will state and prove the Central Limit theorem. Even though the proof is quite simple, it is included because this theorem is the motivation for all the questions we will answer throughout this essay.

1.1 Characteristic Function

Proposition 1.1.1. *Let X be a random variable. Then for every $t \in \mathbb{R}$ we have*

$$|\mathbb{E}e^{itX}| \leq 1.$$

Proof. $|\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = \mathbb{E}1 = 1.$ □

Remark 1.1. By the above proposition (1.1.1) we can conclude that $\mathbb{E}e^{itX}$ always exists.

Definition. Let X be a random variable. Define the function $\phi_X : \mathbb{R} \mapsto \mathbb{R}$ by

$$\phi_X(t) = \mathbb{E}e^{itX}$$

and call it *Characteristic function* of X. When there is no ambiguity we denote ϕ_X simply by ϕ.

Proposition 1.1.2. *Let X be a random variable. Then for every $t, b \in \mathbb{R}$ we have*

$$\phi_{bX}(t) = \phi_X(bt).$$

Proof. $\phi_{bX}(t) = \mathbb{E}e^{it(bX)} = \mathbb{E}e^{i(bt)X} = \phi_X(bt).$ □

Proposition 1.1.3. *Let X and Y be two independent random variables. Then for every $t \in \mathbb{R}$ we have*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t),$$

i.e. the characteristic function of the sum is the product of the characteristic functions.

Proof. $\phi_{X+Y}(t) = \mathbb{E}e^{it(X+Y)} = \mathbb{E}e^{itX}\mathbb{E}e^{itY} = \phi_X(t)\phi_Y(t)$ □

Definition. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be an i.i.d. sequence of random variables. Then we denote by S_n^X the random variable

$$S_n^X = \sum_{i=1}^{n} X_i$$

and call it the *partial sum of order n* of X. When there is no ambiguity we denote S_n^X simply by S_n.

Remark 1.2. Two random variables with the same distribution have the same characteristic function[1].

Proposition 1.1.4. *Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be an i.i.d. sequence of random variables. Then*

$$\phi_{S_n}(t) = \left(\phi_{X_1}(t)\right)^n.$$

Proof. The result follows from applying proposition (1.1.3) n times. □

Example 1.1. The characteristic function of a random variable X with normal distribution (with $\mu = 0$ and σ^2 as its variance) is

$$\phi_X(t) = e^{-\frac{\sigma^2 t^2}{2}}.$$

Solution.

$$\begin{aligned}\phi_X(t) = \mathbb{E}e^{itX} = \int_\Omega e^{itX} dP = \int_\mathbb{R} e^{itx} f_X(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} e^{itx} e^{-\frac{x^2}{2\sigma^2}} dx \\ = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \exp\left(-\frac{x^2 - 2itx\sigma^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \exp\left(-x\frac{x - 2it\sigma^2}{2\sigma^2}\right) dx\end{aligned}$$

Consider a new variable y such that $x = y + it\sigma^2$ then

$$\phi_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \exp\left(-\frac{(y + it\sigma^2)(y - it\sigma^2)}{2\sigma^2}\right) dy = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \exp\left(-\frac{y^2 - (it\sigma^2)^2}{2\sigma^2}\right) dy$$

Now we can separate the constant part. Notice that the remaining integral is equal to one since it is the integral of the function of probability density of a (centered) normal distribution over its support.

$$\phi_X(t) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy}_{=1} e^{-\frac{\sigma^2 t^2}{2}} = e^{-\frac{\sigma^2 t^2}{2}}.$$

We will from now on use this result in many proofs. □

1.2 Central limit theorem

Remark 1.3. The Central Limit theorem states that the sum of any i.i.d. sequence of random variables converges (in distribution) to the normal distribution, so one may be wondering why are we studying characteristic functions instead of convergence (in distribution) of random variables. The reason for this is the following Lévy

[1] This stems from the fact that the Fourier Transform is a bijection, for more on Fourier Transforms see the Appendices.

Continuity Theorem which states that convergence in distribution is equivalent to pointwise convergence of characteristic functions.

Definition. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence such that X_i is a random variable with distribution function F_i. Let Y be a random variable with distribution function F_Y. If for every $x \in \mathbb{R}$ where F_Y is continuous,

$$\lim_{n\to\infty} F_n(x) = F_Y(x),$$

we say that X_n converges in distribution to Y and denote it by $X \xrightarrow{\text{d}} Y$.

Theorem (Lévy Continuity Theorem). *Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence such that X_n is a random variable with distribution function F_n and characteristic function ϕ_n. Let Y be a random variable with distribution function F_Y and characteristic function ϕ_Y. Then for every $x \in \mathbb{R}$*

$$X \xrightarrow{d} Y \iff \lim_{n\to\infty} \phi_n(x) = \phi_Y(x).$$

Proof. see Varadhan, 2001 in the pages 39 and 40. □

Theorem (Central Limit Theorem). *Let X_i be a sequence of i.i.d. random variables with mean* 0 *and variance* 1. *Then,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \leq a\right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

i.e.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{d} Y \quad \textit{where} \quad Y \sim N(0,1).$$

Proof. Consider the Taylor expansion of order 2 of the characteristic function of X_n

$$\begin{aligned}\phi(t) &= \phi(0) + \phi'(0)t + \frac{\phi''(0)}{2}t^2 + O(t^3)\\ &= 1 + i\mathbb{E}(X)t + i^2\frac{\mathbb{E}(X^2)}{2}t^2 + O(t^3)\\ &= 1 + i0t - \frac{\sigma^2}{2}t^2 + O(t^3). \end{aligned} \tag{1.2.1}$$

Then by proposition (1.1.2) we have

$$\phi_{\frac{X_n}{\sqrt{n}}}(t) = 1 - \frac{\sigma^2}{2n}t^2 + O\left(\frac{t^3}{n^{3/2}}\right),$$

hence by proposition (1.1.4)

$$\phi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 - \frac{\sigma^2}{2n}t^2 + O\left(\frac{t^3}{n^{3/2}}\right)\right)^n.$$

Applying the limit on both sides[2]

$$\lim_{n\to\infty} \phi_{\frac{S_n}{\sqrt{n}}}(t) = e^{-\frac{\sigma^2 t^2}{2}} = \phi_Y(t).$$

[2]For more details review proposition (A.5.1.) with $\alpha = 1/2$

The result follows from the Lévy Continuity Theorem. □

Definition. Let X be a random variable such that $\mathbb{E}X = \mu \in \mathbb{R}$ and $\mathbb{E}(X-\mu)^2 = \sigma^2 \in \mathbb{R}_0^+$. We denote by $\widetilde{X}$ the random variable

$$\widetilde{X} = \frac{X-\mu}{\sigma}$$

and call it the *normalization* of X.

Remark 1.4. From the definition above it is easy to check[3] that for any random variable X, if $\widetilde{X}$ exists then $\mathbb{E}\widetilde{X} = 0$ and $Var(\widetilde{X}) = \mathbb{E}\widetilde{X}^2 = 1$.

Definition. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence of random variables and $Y \sim N(0,1)$. If for all $n \in \mathbb{N}$ exists $\widetilde{S}_n$ and

$$\frac{\widetilde{S}_n}{\sqrt{n}} \xrightarrow{\text{d}} Y,$$

we say that the *Central Limit Theorem holds on* X.

Remark 1.5. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence of random variables with mean μ and finite variance $\sigma^2 < \infty$. Let $Y \sim N(0,1)$ be a random variable with standard normal distribution. Note that,

$$\sqrt{n}\left(\frac{1}{n\sigma}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{\text{d}} Y \iff \frac{\widetilde{S}_n}{\sqrt{n}} \xrightarrow{\text{d}} Y.$$

Example 1.2. Consider $\Omega = \{1,2,3,4,5,6\}$ and a random variable $X : \Omega \mapsto \mathbb{R}$ such that X is the identity function. Consider a probability measure on Ω such that for all $x \in \{1,2,3,4,5,6\}$ we have

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) = 1/6,$$

i.e we have a mathematical model for a fair 6 sided dice. Here we have $\mu_X = 3.5$ and $\sigma_X^2 \approx 1.458$.

[n = 1] We rolled this (imaginary) dice 1 000 000 times and observed which face we got each time, and with this information we made the top left graph of figure 1.1. As expected we got each face approximately $\frac{1}{6} \approx 0.166667$

[n = 2] Then we rolled this (imaginary) dice 2 times and observed the sum of the faces we got each time (S_2), then we repeated this process one million times and with this information we made the top right graph of figure 1.1. Here the results are more interesting. We get more often the "middle numbers" (6, 7 and 8) then the "extreme numbers" (2,3,11 and 12). This makes sense since to get 2 there is only one way ($dice1 \implies 1$ and $dice2 \implies 1$) while there are several ways of getting 7 ($dice1 \implies 3$ and $dice2 \implies 4$; $dice1 \implies 4$ and $dice2 \implies 3$; $dice1 \implies 5$ and $dice2 \implies 2$; etc...).

[n = 10] Then we rolled this (imaginary) dice 10 times and observed the sum of the faces we got each time (S_{10}), then we repeated this process one million times and with this information we made the bottom left graph of figure 1.1.

[n = 30] Then we rolled this (imaginary) dice 30 times and observed the sum of the faces we got each time (S_{30}), then we repeated this process one million times and with this information we made the bottom right graph of figure 1.1. Here we can see

[3] $\mathbb{E}\widetilde{X} = \frac{1}{\sigma}(\mathbb{E}X - \mu) = 0$ and $Var(\widetilde{X}) = \mathbb{E}\widetilde{X}^2 - (\mathbb{E}\widetilde{X})^2 = \mathbb{E}\widetilde{X}^2 = \frac{1}{\sigma^2}(\mathbb{E}(X-\mu)^2) = \frac{Var(X)}{\sigma^2} = 1$.

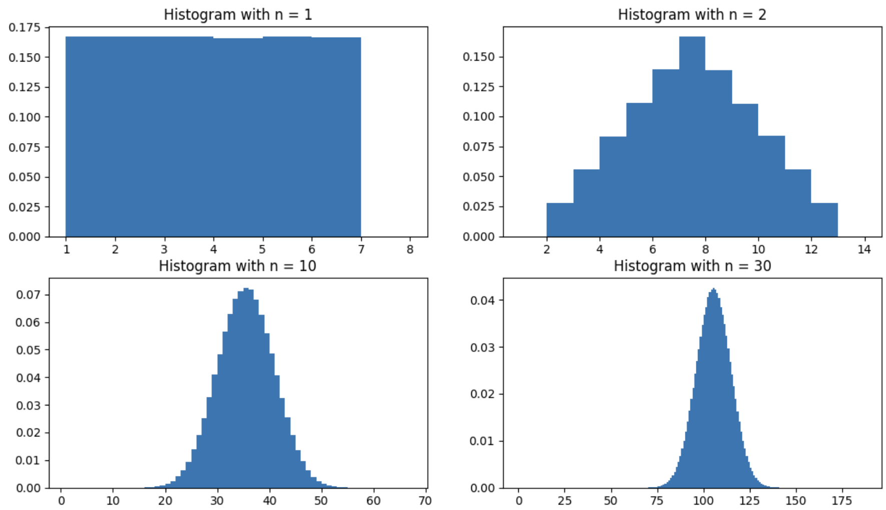

FIGURE 1.1: Central Limit Theorem example using uniform distribution

that the distribution is already quite similar to the normal. What the Central Limit Theorem tells us is that the further we go the closer we get to the normal distribution.

We have that $\mathbb{E}S_n = n\mu_X$ which can be easily seen in the figure. We also have that $Var(S_n) = n\sigma_X^2$

1.3 Centered Variables

Proposition 1.3.1. *Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be i.i.d. random variables. Then the Central Limit Theorem holds on X if and only if it holds on $\widetilde{X}$ i.e. let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be i.i.d. random variables such that for all natural n $\mathbb{E}X_n = 0$ and $\mu = \{\mu_n\}_{\forall n \in \mathbb{N}}$ be a sequence of real numbers, define for all natural n and $Y_n = X_n + \mu_n$. Then the central limit theorem holds on Y if and only if it holds on X.*

Proof. Let $Z \sim N(0,1)$. Notice that $\mathbb{E}S_n^Y = \sum_{i=1}^{n} \mu_i$ and $Var(S_n^Y) = \frac{1}{n\sigma^2}$. Then

$$\widetilde{S}_n^X \xrightarrow{\text{d}} Z \iff \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n}(Y_i - \mu_i)\right) \xrightarrow{\text{d}} Z.$$

To get the second equivalence above sum and subtract μ_i to X_i and substitute Y_i using its definition.

$$\frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n}\mu_i\right) \xrightarrow{\text{d}} Z \iff \frac{1}{\sqrt{n}\sigma}\left(S_n^Y - \sum_{i=1}^{n}\mu_i\right) \xrightarrow{\text{d}} Z,$$

which is equivalent to $\widetilde{S}_n^Y \xrightarrow{\text{d}} Z$. □

Remark 1.6. From the proposition above it becomes clear that considering sequences of random variables with any means or means equal to zero is equivalent when studying central limit convergence. Therefore in the name of simplicity and without any loss of generality all random variables considered in this essay will be assumed to have mean zero. Then, for all $X : \Omega \mapsto \mathbb{R}$ we assume that $\mathbb{E}X = 0$.

Chapter 2

Rates of Convergence

If we are working with a sum of i.i.d. random variables, the Central Limit Theorem tells us that the bigger the number of variables the smaller the error, and that the error tends to zero. However, in applications we are often summing a finite number of random variables and would like to know how big is the error we incur by assuming that the sum has normal distribution. This Chapter aims at answering this question. For this chapter the reader should be familiar with the *Fourier Inversion Formula* (section A.3), *Fubini Theorem* (section A.3), *Riemann-Lebesgue Lemma* (section A.2) and *Holder's inequality* (section A.4). All the claims of the theorems are provided in the appendices. References to the proofs are also provided. The main theorem for this chapter is called the Berry-Esseen Theorem.

2.1 The Integrability Problem

Remark 2.1. Given a sequence of i.i.d. random variables we would like to find a good upper bound for the error between the density function of $\widetilde{S}_n^X$ and the density of a standard normal distribution. An upper bound is an approximation and therefore a Taylor expansion may be an appropriate place to start. Since we want to put our constrains in the moments a good idea may be to study the characteristic function of $\widetilde{S}_n^X$. This can be derived from the characteristic functions of X_n. The Taylor expansions have as coefficients exactly the moments of X. We will start by trying to find a good bound for the difference between the characteristic functions to check if this idea has any chance of working.

Remark 2.2. Let X be a random variable such that $\mu = \mathbb{E}X$, $\sigma^2 = \mathbb{E}X^2$ and $\rho = \mathbb{E}X^3$ all exist (i.e. are finite). Then[1] $\rho \geq \sigma^3$.

Proposition 2.1.1. *Let a, b and C be real numbers such that $|a|$ and $|b|$ are less or equal to C. Then for any natural number n we have the inequality $|a^n - b^n| \leq n|a-b|C^{n-1}$.*

Proof. see proposition (A.4.3). □

Lemma 2.1. *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of normalized random variables such that $\mathbb{E}X_i^3 = \rho < \infty$. Let ϕ_n be the characteristic function of $\tilde{S}_n$ and $Y \sim N(0,1)$. Then exists a natural number N such that for all $0 < k < \sqrt{2}$, t in $]-{}^{k\sqrt{n}}/\rho, {}^{k\sqrt{n}}/\rho[$ and $n > N$ we have*

$$|\phi_n(t) - \phi_Y(t)| \leq \frac{\rho}{\sqrt{n}} \exp\left(-\frac{t^2}{4}\right)\left(\frac{t^3}{6} + \frac{|t|^4}{24}\right).$$

Proof. Let $T = \frac{k\sqrt{n}}{\rho}$. We will start by tying to apply the above proposition to $\phi({}^{t}/\sqrt{n})$ and $\exp\left(-{}^{t^2}/{2n}\right)$. Hence we must find an upper bound for both. By Taylor expansion

[1] see Holder's inequality and corollary (A.4).

of order 3, exists c in the interval $]0,t[$ such that

$$\phi_X(t) = 1 + i\mathbb{E}Xt - \mathbb{E}X^2\frac{t^2}{2} - i\mathbb{E}X^3\phi(c)\frac{t^3}{6}.$$

Taking the absolute value on both sides and noting that the characteristic function is always bounded (1.1.1) and that X has expected value 0 and variance 1 we have

$$|\phi(t)| \leq \left|1 - \frac{t^2}{2}\right| + \rho\frac{|t|^3}{6} \qquad \text{and} \qquad \left|\phi(t) - 1 + \frac{t^2}{2}\right| \leq \rho\frac{|t|^3}{6}. \tag{2.1.1}$$

Then, if we replace t by $t/\sqrt{n}$ on the inequality on the right (the left one will be used later)

$$\left|\phi\left(\frac{t}{\sqrt{n}}\right)\right| \leq \left|1 - \frac{t^2}{2n}\right| + \rho\frac{|t|^3}{6n^{\frac{3}{2}}} = \left|1 - \frac{t^2}{2n}\right| + \frac{t^2}{2n}\frac{\rho|t|}{3\sqrt{n}}.$$

By remark (2.2) we have $\rho \geq \sigma^3 = 1$. Hence[2] we have $T^2 \leq 2n$ which implies $t^2 \leq 2n$. Therefore we may remove the absolute value on the right hand side. Then

$$\left|\phi\left(\frac{t}{\sqrt{n}}\right)\right| \leq 1 - \frac{t^2}{2n} + \frac{t^2}{2n}\frac{1}{3}\frac{\rho|t|}{\sqrt{n}} \leq 1 - \frac{t^2}{2n} + \frac{t^2}{2n}\frac{1}{3}\frac{\rho T}{\sqrt{n}}.$$

Replacing T by its definition for the first step and knowing[3] that $1 - x < e^{-x}$ for the third step we get

$$\left|\phi\left(\frac{t}{\sqrt{n}}\right)\right| \leq 1 - \frac{t^2}{2n}\left(1 - \frac{k}{3}\right) = 1 - \frac{(3-k)t^2}{6n} \leq \exp\left(-\frac{(3-k)t^2}{6n}\right).$$

Hence we found an upper bound for $\phi(t/\sqrt{n})$. Now we need to check if the upper bound is valid for $\exp\left(-t^2/2n\right)$. Since k is positive we have that $-1/2 < -(3-k)/6$, thus

$$\phi_Y^n(t) = \exp\left(-\frac{t^2}{2n}\right) \leq \exp\left(-\frac{(3-k)t^2}{6n}\right).$$

Now we have a good candidate for our upper bound. Then by the above proposition

$$|\phi_n(t) - \phi_Y(t)| \leq n\left|\phi(\frac{t}{\sqrt{n}}) - \exp\left(-\frac{t^2}{2n}\right)\right|\exp\left(-\frac{(3-k)(n-1)t^2}{6n}\right).$$

For sufficiently large n we have $0 < k < \sqrt{2} \implies -\frac{(3-k)(n-1)}{6n} < -\frac{1}{4}$. Hence

$$|\phi_n(t) - \phi_Y(t)| \leq n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{t^2}{2n}\right)\right|\exp\left(-\frac{t^2}{4}\right). \tag{2.1.2}$$

Now sum and subtract $1 + \frac{t^2}{2n}$ and use the triangular inequality

$$n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - \exp(-\frac{t^2}{2n})\right| \leq n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - 1 + \frac{t^2}{2n}\right| + n\left|1 - \frac{t^2}{2n} - \exp\left(-\frac{t^2}{2n}\right)\right|.$$

Use inequality (2.1.1) on the first term and for the second term recall that $\frac{t^2}{2n} < 1$ and the fact that $|e^{-x} - 1 + x| \leq \frac{x^2}{2}$ for all x in the interval $]0,1[$ (by proposition (A.4.1).

[2] $T^2 = \frac{k^2 n}{\rho^2} \leq k^2 n \leq 2n$ since $k \leq \sqrt{2}$.

[3] for a proof see proposition (A.4).

Then

$$n\left|\phi(\frac{t}{\sqrt{n}}) - \exp\left(-\frac{t^2}{2n}\right)\right| \leq \rho\frac{|t|^3}{6\sqrt{n}} + \frac{t^4}{8n},$$

by substituting this last equation on (2.1.2)

$$|\phi_n(t) - \phi_Y(t)| \leq \exp\left(-\frac{t^2}{4}\right)\left(\frac{\rho}{\sqrt{n}}\frac{|t|^3}{6} + \frac{t^4}{8}\frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\right).$$

By construction[4] of T we have

$$|\phi_n(t) - \phi_Y(t)| \leq \exp\left(-\frac{t^2}{4}\right)\left(\frac{k|t|^3}{6T} + \frac{kt^4}{24T}\right).$$

The result follows from noting that $k/T = \rho/\sqrt{n}$. □

Lemma 2.2. *Let X and Y be centered random variables such that their characteristic functions are integrable. Then*

$$F_X(x) - F_Y(x) = \frac{1}{2\pi}\int -e^{-ity}\frac{\phi_X(t) - \phi_Y(t)}{it}dt$$

Proof. Since ϕ_X and ϕ_Y are integrable, we can use the Fourier Inversion Formula. Then

$$f_X(x) = \frac{1}{2\pi}\int e^{-itx}\phi_X(t)dt \qquad \text{and} \qquad f_Y(x) = \frac{1}{2\pi}\int e^{-itx}\phi_Y(t)dt.$$

Subtracting the above equations one by the other an integrating from a to x we get

$$\int_a^x f_X(x) - f_Y(x)dx = \frac{1}{2\pi}\int_a^x \int e^{-ity}(\phi_X(t) - \phi_Y(t))dtdy.$$

Note that F_X and F_Y are respectively the indefinite integrals of f_X and f_Y. Hence we can compute the integral on the left hand side. On the right hand side we have a bounded domain on y and everything is integrable on t thus we can apply Fubini Theorem to get

$$(F_X(x) - F_Y(x)) - (F_X(a) - F_Y(a)) = \frac{1}{2\pi}\int (e^{-itx} - e^{-ita})\frac{(\phi_X(t) - \phi_Y(t))}{it}dt.$$

The above integral exists because $\lim_{t\to 0}\frac{(\phi_X(t)-\phi_Y(t))}{it} \leq \left|\lim_{t\to 0}\frac{1-\phi_Y(t)}{t}\right| = |\phi_Y'(0)| \leq \mathbb{E}|Y| < \infty$. To complete the proof apply the limit as a tends to $-\infty$ to get the result from Riemann-Lebesgue Lemma. □

Remark 2.3. This lemma has a limitation: the characteristic functions must be integrable. Our aim is to compare the normal distribution - which has integrable characteristic function - to the normalized sum of other random variables. This random variables may not have integrable characteristic function. However all characteristic functions are bounded[5]. Therefore if we multiply our characteristic function - which we do not know if it is integrable - by one that is zero outside a closed interval this

[4] Since $T = k\sqrt{n}/\rho$ we have $\rho/\sqrt{n} = k/T$ (first term) and if $n \geq 10 \implies 1/\sqrt{n} < 1/3$ then $1/\sqrt{n} \times 1/\sqrt{n} < \rho k/T \times 1/3 \leq k/T \times 1/3$ (second term).

[5] see proposition 1.1.1.

multiplication will be integrable. Since the integral will be less or equal to $1 \times L$ where L is the lengh of the interval and 1 is the maximum value a characteristic function may have. Therefore we only need to find a characteristic function that is zero outside a closed interval.

2.2 Polya Distribution

Proposition 2.2.1. *Let θ be a strictly positive real. Then $(1-\cos\theta x)/\pi\theta x^2 \geq 0$.*

Proof. For the inequality just note that $\pi\theta x^2 \geq 0$ and the cosine is always smaller than one. □

Proposition 2.2.2. *Let θ be a strictly positive real. Then*

$$\int \frac{1-\cos\theta y}{\pi\theta y^2} dy = 1.$$

Proof. Consider a random variable X with triangular density[6] $f_X(y) = (1-|y|)^+$. Then $\phi_X(u) = 2\frac{1-\cos u}{u^2}$. Therefore by the Fourier Inversion Formula we have that

$$\frac{1}{2\pi}\int 2\frac{1-\cos u}{u^2} e^{-iuy} du = (1-|y|)^+ .$$

Replace $u = \theta x$ and $t = \theta y$

$$\int \theta \frac{1-\cos\theta x}{\pi\theta^2 x^2} e^{-ixt} dx = \left(1-\left|\frac{t}{\theta}\right|\right)^+ .$$

Note that the above formula is valid for every t. Simplify θ and set $t=0$ to get the result. □

Remark 2.4. By the previous proposition we have that $\mu(A) = \int_A \frac{1-\cos\theta x}{\pi\theta x^2} dx$ is a probability measure that has $f(x) = \frac{1-\cos\theta x}{\pi\theta x^2}$ as its probability density function.

Definition. Let X be a random variable with law

$$\mu(A) = \int_A \frac{1-\cos\theta x}{\pi\theta x^2} dx.$$

We call X a *random variable with Polya distribution* of parameter θ and write it by $X \sim Polya(\theta)$.

Remark 2.5. Let $X \sim Polya(\theta)$. From the proof of the above proposition (2.2.2) follows that the characteristic function of X is $\phi_X(t) = (1-|t/\theta|)^+$. Furthermore the characteristic function is zero outside the interval $]-\theta,\theta[$.

Remark 2.6. Let $X \sim Polya(\theta)$. Then $f_X(x) = \frac{1-\cos\theta x}{\pi\theta x^2}$. We have that that f_X is symmetric.

2.3 Berry-Esseen Theorem

Remark 2.7. Let X be any random variable and $P \sim Polya(T)$ a random variable with Polya distribution of parameter T. Since - just like $\phi_P(t)$ - is zero outside of the interval $[-T,T]$, we have that $\phi_X(t)\phi_P(t)$ is integrable.

[6] For more on the triangular distribution see Durrett, 2010, page 93, example 3.3.5.

Remark 2.8. We have that $\phi_X(t)\phi_P(t)$ is the characteristic function of $X + P$. Therefore we will work with the distribution function of $X + P$ which is computed as the convolution between F_X and F_P and denoted by $F_X * F_P$.

Remark 2.9. Therefore we will apply lemma (2.2) to the convoluted random variables. Hence we need to compare the difference between the convoluted densities and the original ones. That is the aim of the next lemma.

Lemma 2.3. *Let X and Y be random variables such that F_Y is continuous and differentiable and let F_P be the distribution of a Polya random variable with parameter T. Then*

$$|F_X(x) - F_Y(x)| \leq 2|F_{X*P}(x) - F_{Y*P}(x)| + \frac{24 \sup_x F_Y'(x)}{\pi T}.$$

Remark 2.10. Let $\lambda = \sup F_Y'(x)$, $\Delta(x) = F_X(x) - F_Y(x)$ and $\Delta^* = \Delta * F_P$. Furthermore let $\widehat{\lambda}$, $\widehat{\Delta}$ and $\widehat{\Delta}^*$ have analogous definitions but using $-X$ and $-Y$. Considering the formula[7] for the distribution function of the symmetric variable, we have for all $x \in \mathbb{R}$

$$\lambda = \widehat{\lambda}, \qquad \Delta(x^-) = -\widehat{\Delta}(-x), \qquad \Delta^*(x^-) = -\widehat{\Delta}^*(-x)$$

since it is valid for all x we can just substitute $y = x^-$ and $z = -x$ to conclude that the lemma is valid for X and Y if and only if it is valid for $-X$ and $-Y$.

Proof. If $\sup_x F_Y'(x) = \infty$ the proposition is trivially true. Otherwise consider the notation of the above remark (2.10) and let

$$\eta = \sup |\Delta(x)|, \qquad \eta^* = \sup |\Delta^*(x)|.$$

Since Δ is a difference of distribution functions, it decays to zero as x tends to $\pm\infty$. Therefore exists a closed neighborhood[8] of zero, A, such that the supremum on A is equal to η. Thus exists a $x_0 \in A$ such that $\eta = |\Delta(x_0^-)|$. If $\Delta(x_0^-) > 0$ then $F_X(x_0^-) > F_Y(x_0^-)$. Since F_X is right continuous and F_Y is continuous we have

$$F_X(x_0) \geq F_X(x_0^-) > F_Y(x_0^-) = F_Y(x_0).$$

Hence $\Delta(x_0) = \eta$. If $\Delta(x_0^-) < 0$, by the above remark (2.10) we consider $\widehat{\Delta}$ and we have $\widehat{\Delta}(x_0) = \eta$. Therefore without loss of generality exists x_0 such that $\Delta(x_0) = \eta$. Consider $s > 0$. Since F_X is increasing and by first order Taylor expansion[9] on F_Y we have

$$\Delta(x_0 + s) = F_X(x_0 + s) - F_Y(x_0 + s) \geq F_X(x_0) - (F_Y(x_0) + \lambda s) = \eta - \lambda s$$

Let $\delta = \frac{\eta}{2\lambda}$.

$$\Delta(x_0 + \delta - x) \geq \begin{cases} \dfrac{\eta}{2} + \lambda x & \text{if} \quad |x| \leq \delta \\ -\eta & \text{if} \quad \text{otherwise} \end{cases} \tag{2.3.1}$$

[7] $F_{-X}(-x) = \mathbb{P}(-X \leq -x) = \mathbb{P}(X \geq x) = 1 - \mathbb{P}(X < x) = 1 - F_X(x^-)$.

[8] A closed neighborhood of zero is such that $A = [a, b]$ for some $a < 0$ and $b > 0$.

[9] $F_Y(x_0 + s) = F_Y(x_0) + sF_Y'(\xi) \leq F_Y(x_0) + s\lambda$ where $\xi \in]x_0, x_0 + s[$.

Hence by definition of η^* (first step) and Δ^* (second step),

$$\eta^* \geq \Delta^*(x_0+\delta) = \int \Delta(x_0+\delta-x) f_P(x)dx$$
$$= \int_{-\delta}^{\delta} \Delta(x_0+\delta-x) f_P(x)dx + \int_{\mathbb{R}/[-\delta,\delta]} \Delta(x_0+\delta-x) f_P(x)dx.$$

Separate the second integral into two, corresponding to the two intervals. Then by inequality (2.3.1), we have that

$$\eta^* \geq \int_{-\delta}^{\delta} \left(\frac{\eta}{2}+\lambda x\right) f_P(x)dx + \int_{\delta}^{\infty} -\eta f_P(x)dx + \int_{-\delta}^{-\infty} -\eta f_P(x)dx.$$

By symmetry of f_P we have that

$$\eta^* \geq \underbrace{\int_{-\delta}^{\delta} \lambda x f_P(x)dx}_{=0} + \frac{\eta}{2}\left(1-2\int_{\delta}^{\infty} f_P(x)dx\right) - 2\int_{\delta}^{\infty} \eta f_P(x)dx. \tag{2.3.2}$$

Now note that

$$\int_{\delta}^{\infty} f_P(x)dx = \int_{\delta}^{\infty} \frac{1-\cos Tx}{T\pi x^2}dx \leq \int_{\delta}^{\infty} \frac{2}{T\pi x^2}dx = \frac{2}{T\pi\delta}.$$

Then replace the above equation in the inequality (2.3.2) to get

$$\eta^* \geq \frac{\eta}{2}\left(1-\frac{4}{T\pi\delta}\right) - 2\eta\frac{2}{T\pi\delta}.$$

Recall that $\delta = \frac{\eta}{2\lambda}$

$$\eta^* \geq \frac{\eta}{2}\left(1-\frac{8\lambda}{T\pi\eta}\right) - 2\eta\frac{4\lambda}{T\pi\eta} = \frac{\eta}{2} - \frac{4\lambda}{T\pi} - \frac{8\lambda}{T\pi} = \frac{\eta}{2} - \frac{12\lambda}{T\pi}.$$

Now just note that $\eta^* \geq \eta/2 - 12\lambda/\pi T \iff \eta \leq 2\eta^* + 24\lambda/\pi T$. □

Now we are ready to prove the theorem.

Theorem (Berry-Esseen). *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of centered random variables such that $\mathbb{E}X_i^2 = \sigma^2$ and $\mathbb{E}X_i^3 = \rho < \infty$. Let F_n be the distribution function of $\tilde{S}_n$. Then*

$$|F_n(x) - F(x)| \leq 3\frac{\rho}{\sigma^3\sqrt{n}}.$$

Remark 2.11. If all other conditions are met and the above theorem is valid for all random sequences such that $\sigma = 1$ then it is valid for all sequences. To see this let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence in the conditions of the above theorem and consider the sequence of random variables $\{Y_i\}_{i\in\mathbb{N}}$ such that for all natural number i we have $Y_i = \frac{X_i}{\sigma}$ then $\mathbb{E}Y_i^2 = \frac{\mathbb{E}X_i^2}{\sigma^2} = 1$ therefore we have the above theorem holds for Y, i.e.

$$|F_n(x) - F(x)| \leq 3\frac{\rho_Y}{\sqrt{n}}$$

Notice that F_n is the distribution of $\widetilde{S}_n^Y$ which is equal to $\widetilde{S}_n^X$ since it is a normalized random variable. We also have that $\rho_Y = \mathbb{E}Y_i^3 = \mathbb{E}\frac{X_i^3}{\sigma^3} = \frac{\rho}{\sigma^3}$ hence the result is valid for X.

Proof. By the remark above assume without loss of generality that $\sigma^2 = 1$. Let ϕ_n be the respective characteristic functions of X_n. Let P be a random variable with Polya's distribution of parameter T. We are in the conditions of lemma (2.3) and therefore

$$|F_X(x) - F_Y(x)| \leq 2|F_{X*P}(x) - F_{Y*P}(x)| + \frac{24 \sup_x F'_Y(x)}{\pi T}.$$

To find an upper bound for the second term just note that $f_Y(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ is the probability density function of Y (which has standard normal distribution) hence

$$\frac{24\lambda}{\pi T} \leq \frac{24 \sup_x F'_Y(x)}{\pi T} = \frac{24 \sup_x f_Y(x)}{\pi T} = \frac{24 f_Y(0)}{\pi T} = \frac{24}{\sqrt{2\pi}\pi T} \leq \frac{9.6}{\pi T}.$$

Considering the above for the second term and applying lemma (2.2) to the first term we get

$$|F_X(x) - F_Y(x)| \leq 2 \left| \frac{1}{2\pi} \int -e^{-ity} \frac{\phi_n(t)\phi_P(t) - \phi_Y(t)\phi_P(t)}{it} dt \right| + \frac{9.6}{\pi T}.$$

Now just recall that ϕ_P is zero outside the interval $]-T, T[$ and triangular inequality to get

$$|F_X(x) - F_Y(x)| \leq \frac{1}{2\pi} \int_{-T}^{T} \frac{|\phi_X(t) - \phi_Y(t)|}{|t|} dt + \frac{9.6}{\pi T}.$$

Apply lemma (2.1) to the term inside the integral (from now on $T = {}^{k\sqrt{n}}/_{\rho}$ with $0 < k \leq \sqrt{2}$) to get

$$|F_X(x) - F_Y(x)| \leq \int_{-T}^{T} \frac{\rho}{\sqrt{n}\pi} \exp\left(-\frac{t^2}{4}\right) \left(\frac{t^2}{6} + \frac{|t|^3}{24}\right) dt + \frac{9.6\rho}{k\pi\sqrt{n}}.$$

Using the distributive property and noting that the integral on the whole set is larger than the integral on $]-T, T[$ we get

$$\frac{\sqrt{n}}{\rho}|F_X(x) - F_Y(x)| \leq \frac{1}{3\sqrt{\pi}} \int \frac{1}{\sqrt{4\pi}} t^2 \exp\left(-\frac{t^2}{4}\right) dt + \frac{1}{24\pi} \int |t|^3 \exp\left(-\frac{t^2}{4}\right) dt + \frac{9.6}{k\pi}.$$

Then compute the integrals[10]. The first yields 2 and the second 16.

$$|F_X(x) - F_Y(x)| \leq \frac{\rho}{\sqrt{n}\pi} \left(\frac{2\sqrt{\pi}}{3} + \frac{16}{24} + \frac{9.6}{k}\right).$$

Now just pick the highest value of k to get the smallest bound ($k = \sqrt{2}$). Note that ${}^1/_\pi \left({}^{2\sqrt{\pi}}/_3 + {}^{16}/_{24} + {}^{9.6}/_{\sqrt{2}}\right) \approx 2.492690498 < 3$ then the result follows. □

Example 2.1. Consider a sequence of independent random variables $X = \{X_n\}_{\forall n \in \mathbb{N}}$ with exponential distribution of parameter $\lambda = 3$. Take several observations from X and draw an histogram of them. We will get something similar to the top left graph of figure (2.1).

Now we can normalize X and draw a cumulative graph of the histogram. This is what we have done and drew it in top middle graph of figure (2.1). For comparison we also drew on the same graph the cumulative probability distribution of a standard normal random variable Y. $|F_X(a) - F_Y(a)|$ is drawn on the top right graph of

[10] For more details see proposition (A.3.2) and proposition (A.3)

the figure of the (2.1). Now we need to compare the supremum of the function on

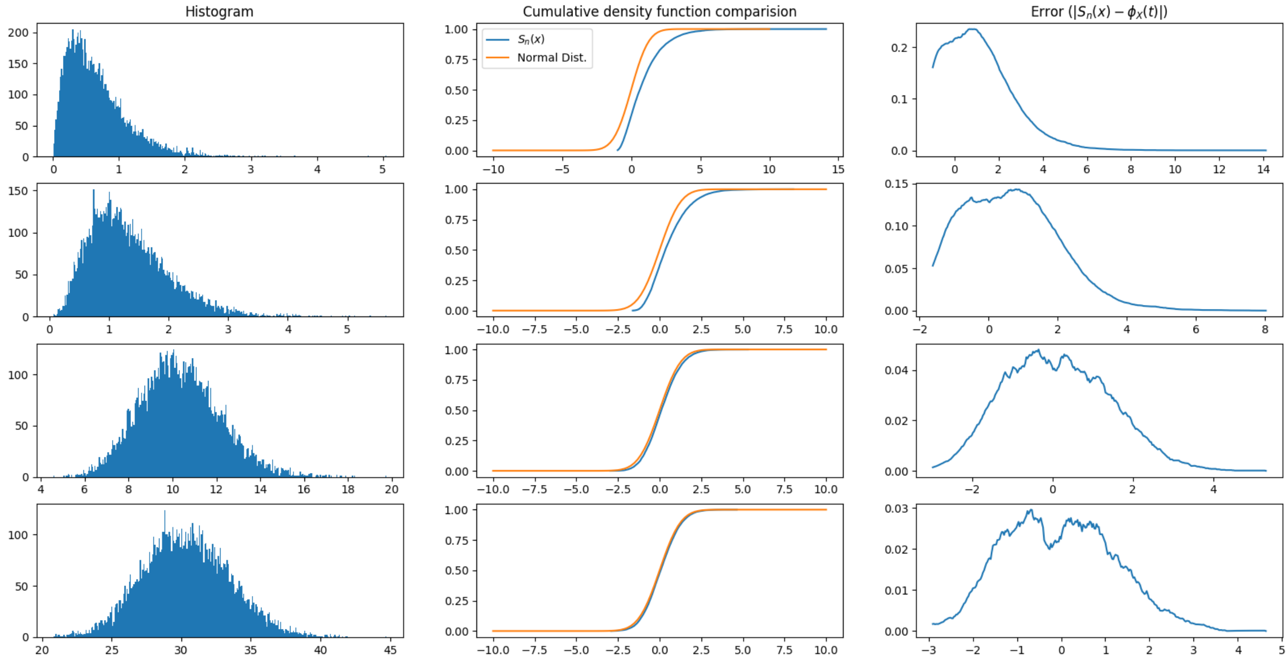

FIGURE 2.1: The Central Limit Theorem. The first row refers to $n = 1$, the second to $n = 3$, the third to $n = 30$ and the last to $n = 90$.

the right for different values of n, i.e instead of taking a sample of X (which is S_n^X for $n = 1$) we took a sample from S_n^X with growing values of n. This what we have done and drew the graphs in the subsequent rows of figure (2.1). Each row corresponds respectively to n equal to one, three, thirty and ninety. To see the behavior

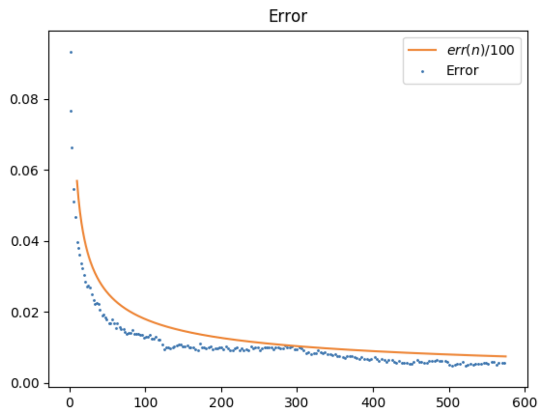

FIGURE 2.2: The error of the Central Limit Theorem as n becomes larger. The function $err(n) = 3\rho/\sigma^3\sqrt{n}$

of the error clearer: we plotted the errors for many values n. On figure (2.2) each error corresponds to the maximum of the function on the right of figure (2.1). For comparison we also drew on the same graph a multiple of the error predicted by the theorem, in this case the error is much smaller.

Chapter 3

Generalizations for nonidentical sequences

We proved the Central Limit Theorem and explored the rates of convergence. Now we will search for possible generalizations of the Central Limit Theorem. Consider a family of random variables that are not *i.i.d.*, can we still apply the Central Limit Theorem? In which circumstances can we do so? These are the questions we aim at answering with this chapter.

3.1 Lindeberg's Theorem

Notation. Consider a sequence $X = \{X_n\}_{\forall n \in \mathbb{N}}$ of random variables. In this chapter we are not assuming that the variables are identically distributed. Therefore the standard deviation of S_n is not equal to the standard deviation of any X_n. Hence for the rest of this chapter given any $n \in \mathbb{N}$ we will use the notation $s_n = \sqrt{\sum_{i=0}^{n} \sigma_i^2}$.

Remark 3.1. It is simple[1] to see that s_n^2 represents the variance of S_n.

Remark 3.2. We will start by taking another look into what it means for the Central Limit Theorem to hold. We hope this will give us some insight into how to prove it. Consider any $t \in \mathbb{R}$. Then

$$\lim_{n\to\infty} \left| \log \phi_{\widetilde{S}_n}(t) + \frac{t^2}{2} \right| = 0 \iff \lim_{n\to\infty} \phi_{\widetilde{S}_n}(t) = \exp\left(-\frac{t^2}{2}\right).$$

To see this apply the exponential function and the logarithm to $t^2/2$. Then use the rules of the logarithmitic function to get a single logarithm on the left hand side. Finally apply the exponential function to both sides and move the appropriate term to the right hand side.

Remark 3.3. Since we have a new expression to prove the Central Limit Theorem using the logarithm of $\phi_{\widetilde{S}_n}$ we may want to take a closer look at it. Consider any $t \in \mathbb{R}$, then it is easily seen [2] that

$$\log \phi_{\widetilde{S}_n}(t) = \log \phi_{\sum_{j=1}^{n} \frac{X_j}{s_n}}(t) = \log \prod_{j=1}^{n} \phi_{\frac{X_j}{s_n}}(t) = \sum_{j=0}^{n} \log \phi_{\frac{X_j}{s_n}}(t).$$

[1] $Var(S_n) = Var(\sum_{i=0}^{n} X_i) = \sum_{i=0}^{n} Var(X_i) = \sum_{i=0}^{n} \sigma_i^2$.

[2] To get the second equality one may wish to review proposition (1.1.3).

Remark 3.4. It may also be useful for us to note that - by the Taylor expansion of the logarithm - we have

$$\log x = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}(x-1)^n = (x-1) + O(x-1)^2.$$

Remark 3.5. We have that the previous two remarks (3.3 and 3.4) suggest that

$$\log \phi_{\tilde{S}_n} = \sum_{j=0}^{n} \log \phi_{\frac{X_j}{s_n}} \approx \sum_{j=0}^{n}\left(\phi_{\frac{X_j}{s_n}} - 1\right).$$

Remark 3.6. In the other hand recall that the characteristic function is an expected value of a exponential function[3]. Furthermore by third order Taylor expansion of the exponential function we have for all $j \leq n$

$$\phi_{\frac{X_j}{s_n}}(t) = 1 + \frac{it}{s_n}\mathbb{E}X_j - \frac{t^2}{2s_n}\mathbb{E}X_j^2 + O(t^3)$$

Subtract 1 to both sides and consider the sum of the above expression from X_1 to X_n. Recall that we are always considering $\mathbb{E}X = 0$. This suggests that

$$\sum_{j=0}^{n}\phi_{\frac{X_j}{s_n}}(t) - 1 \approx \frac{1}{s_n}\sum_{j=0}^{n}\mathbb{E}X_j^2\frac{t^2}{2} = -\frac{t^2}{2}.$$

Remark 3.7. Therefore our strategy will be to show that

$$\log \phi_{\tilde{S}_n}(t) \approx \sum_{j=0}^{n}\left(\phi_{\frac{X_j}{s_n}}(t) - 1\right) \approx -\frac{t^2}{2},$$

in the hopes that our "close enough ($\approx$)" will allow us to use remark (3.2) to finish the proof. We will show the first "close enough" in lemma (3.1), the second "close enough" in lemma (3.2) and finally in the proof of the theorem we will show that two "close enough's" is still "close enough".

Notation. For the rest of this chapter given a sequence $X = \{X_n\}_{\forall n \in \mathbb{N}}$ of random variables and for any $n \in \mathbb{N}$ we will use the notation

$$\psi_n(t) = \sum_{j=0}^{n}\left(\phi_{\frac{X_j}{s_n}}(t) - 1\right).$$

Remark 3.8. Since we will be using ψ_n for our prove, it may be useful to take a closer look at it. For the last step recall[4] that we are considering $\int x d\alpha_{X_n} = \mathbb{E}X = 0$.

$$\phi_{\frac{X_j}{s_n}}(t) - 1 = \phi_{X_j}\left(\frac{t}{s_n}\right) - 1 = \int e^{it\frac{x}{s_n}} - 1 d\alpha_{X_j} \leq \int \left| e^{it\frac{x}{s_n}} - 1 - it\frac{x}{s_n}\right| d\alpha_{X_j}.$$

We know[5] that for every $x \in \mathbb{R}$ exists C such that $|e^{ix} - 1 - ix| \leq Cx^2$. Hence

$$\phi_{\frac{X_j}{s_n}}(t) - 1 \leq Ct^2 \int \frac{x^2}{s_n^2} d\alpha_{X_j} = Ct^2\frac{\sigma_j^2}{s_n^2}$$

[3] see definition (1.1).
[4] see remark (1.6).
[5] see proposition (A.4.1).

In the previous equation consider the sum on j from 0 to n to get a bound on ψ_n. Then by definition of s_n

$$\psi_n(t) \leq \sum_{j=0}^{n} |\phi_{\frac{X_j}{s_n}}(t) - 1| \leq \frac{Ct^2}{s_n^2} \sum_{j=0}^{n} \sigma_j^2 = Ct^2.$$

Consider the limit as n grows arbitrarily.

$$\lim_{n\to\infty} \psi_n(t) \leq \sum_{j=0}^{\infty} |\phi_{\frac{X_j}{s_n}}(t) - 1| \leq Ct^2.$$

Remark 3.9. In the above series the general term $|\phi_{X_j/s_n}(t) - 1|$ depends on j and n. Then if we make both j and n diverge to infinity, we know that the general term will converge to zero - otherwise the series would not converge. However we have $j \leq n$. Hence it is possible to let n diverge to infinity while having fixed j. Even though at this point may not be completely clear why; we will compute the supremum of the term with fixed j and then let n diverge to infinity.

Remark 3.10. Consider the supremum in j of the general term considered in the above remark (3.9)

$$\sup_{1\leq j\leq n} \phi_{\frac{X_j}{s_n}}(t) - 1 \leq \sup_{1\leq j\leq n} Ct^2 \int \frac{x^2}{s_n^2} d\alpha_{X_j}.$$

For all positive real $\epsilon > 0$

$$\sup_{1\leq j\leq n} \phi_{\frac{X_j}{s_n}}(t) - 1 \leq Ct^2 \sup_{1\leq j\leq n} \left(\frac{1}{s_n^2} \int_{|x|\leq \epsilon s_n} x^2 d\alpha_{X_j} + \frac{1}{s_n^2} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_{X_j} \right).$$

Given the set on which we are integrating the first integral, if we replace x by ϵs_n we obtain a bigger term. This term does not depend on j, hence it can be removed from the supremum. For the second integral notice that since all terms are positive the supremum over a finite set is smaller than the sum over the same set.

$$\sup_{1\leq j\leq n} \phi_{\frac{X_j}{s_n}}(t) - 1 \leq Ct^2 \left(\frac{1}{s_n^2} \epsilon^2 s_n^2 + \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_{X_j} \right).$$

Simplifying s_n and taking the limit as n increases in both sides, we get

$$\lim_{n\to\infty} \sup_{1\leq j\leq n} \phi_{\frac{X_j}{s_n}}(t) - 1 \leq Ct^2\epsilon^2 + Ct^2 \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_{X_j}.$$

Since ϵ is arbitrary take the limit as ϵ decreases to zero to get a curious limit

$$\lim_{n\to\infty} \sup_{1\leq j\leq n} \phi_{\frac{X_j}{s_n}}(t) - 1 \leq Ct^2 \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_{X_j}.$$

Remark 3.11. We can not explicitly compute the above limit for all sequences of random variables. However we can argue that for many sequences the above limit is zero. For a simple example consider a sequence X_n of i.i.d. random variables. The variance of X_n is constant. We have that s_n is diverging to infinity and ϵ is fixed. Then we have that the sequence of sets $\{\omega : |X_n(\omega)| \geq \epsilon s_n\}$ are converging to the empty set. In this case the limit of the sum of the integrals is convergent while s_n

diverges. Therefore the above limit is zero. We will use the definition below to differentiate the sequences for which this limit is zero.

Definition. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence of independent random variables and $\{\alpha_n\}_{n\in\mathbb{N}}$ their respective laws. If for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_i(x) = 0,$$

we say the X satisfies the Lindeberg's condition, or is a *Lindeberg's family*.

Lemma 3.1. *Let $X = \{X_n\}_{\forall n\in\mathbb{N}}$ be a Lindeberg's family then for every $t \in \mathbb{R}$*

$$\lim_{n\to\infty} \left|\log \phi_{\widetilde{S}_n}(t) - \psi_n(t)\right| = 0.$$

Proof. Let us start by evaluating the absolute value. Consider t as above, then by manipulation of the characteristic function (see remark 3.3) and by the definition of ψ_n we have

$$\left|\log \phi_{\widetilde{S}_n} - \psi_n\right| = \left|\sum_{j=0}^{n} \log \phi_{\frac{X_j}{s_n}} - \sum_{j=0}^{n} \left(\phi_{\frac{X_j}{s_n}}(t) - 1\right)\right| \leq \sum_{j=0}^{n} \left|\log \phi_{\frac{X_j}{s_n}} - \left(\phi_{\frac{X_j}{s_n}}(t) - 1\right)\right|.$$

By Taylor expansion of the logarithmic function (see remark 3.4) and the definition of order[6] we have that exists a positive $M \in \mathbb{R}$ such that

$$\left|\log \phi_{\widetilde{S}_n} - \psi_n\right| \leq M \sum_{j=0}^{n} \left|\phi_{\frac{X_j}{s_n}}(t) - 1\right|^2 \leq M \left(\sup_{1\leq j\leq n} \left|\phi_{\frac{X_j}{s_n}}(t) - 1\right|\right) \sum_{j=0}^{n} \left|\phi_{\frac{X_j}{s_n}}(t) - 1\right|.$$

Notice that the sum above is bounded when n converges to infinity (see remark 3.8). On the other hand the supremum becomes zero since X is a Lindeberg's family (see remark 3.10). The claim follows. □

Lemma 3.2. *Let $X = \{X_n\}_{\forall n\in\mathbb{N}}$ be a Lindeberg's family and $\{\alpha_n\}_{n\in\mathbb{N}}$ their respective laws.Then for every $t \in \mathbb{R}$*

$$\lim_{n\to\infty} \left|\psi_n(t) + \frac{t^2}{2}\right| = 0.$$

Proof. Multiply and divide $t^2/2$ by s_n^2. Recall the definition of ψ and the definition of s_n^2 for the numerator.

$$\left|\psi_n + \frac{t^2}{2}\right| = \left|\sum_{j=0}^{n} \left(\phi_{\frac{X_j}{s_n}}(t) - 1\right) + \frac{t^2}{2} \sum_{j=0}^{n} \frac{\sigma_j^2}{s_n^2}\right| \leq \sum_{j=0}^{n} \left|\phi_{\frac{X_j}{s_n}}(t) - 1 + \frac{t^2\sigma_j^2}{2s_n^2}\right|$$

Recall[7] the definition of characteristic function, that we are considering $\int x d\alpha_{X_n} = \mathbb{E}X = 0$ and $\int x^2 d\alpha_{X_n} = \mathbb{E}X^2 = \sigma_j^2$, to get

$$\left|\psi_n + \frac{t^2}{2}\right| \leq \sum_{j=0}^{n} \int \left|\exp\left(it\frac{x}{s_n}\right) - 1 - it\frac{x}{s_n} + \frac{t^2x^2}{2s_n^2}\right| d\alpha_j.$$

[6]For more details on order see section A.5.
[7]see definition (1.1) and remark (1.6).

For all $\epsilon > 0$

$$\left|\psi_n + \frac{t^2}{2}\right| \leq \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} \left|\exp\left(it\frac{x}{s_n}\right) - 1 - it\frac{x}{s_n}\right| + \left|\frac{t^2x^2}{2s_n^2}\right| d\alpha_j +$$
$$+ \sum_{j=0}^{n} \int_{|x|<\epsilon s_n} \left|\exp\left(it\frac{x}{s_n}\right) - 1 - it\frac{x}{s_n} + \frac{t^2x^2}{2s_n^2}\right| d\alpha_j.$$

We know[8] that exists $C \in \mathbb{R}$ such that for all $y \in \mathbb{R}$ we have $|e^{iy} - 1 - iy| \leq Cy^2$ and $|e^{iy} - 1 - iy + \frac{y^2}{2}| \leq Cy^3$. Use this on the first and second term respectively to get

$$\left|\psi_n + \frac{t^2}{2}\right| \leq \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} C\frac{t^2x^2}{s_n^2} + \frac{t^2x^2}{2s_n^2} d\alpha_j + \sum_{j=0}^{n} \int_{|x|<\epsilon s_n} C\frac{t^3x^3}{s_n^3} d\alpha_j.$$

On the first term use the distributive property. On the second term, since we are integrating for $|x| < \epsilon s_n$ we have $|x|^3 < \epsilon s_n x^2$. Hence

$$\left|\psi_n + \frac{t^2}{2}\right| \leq \frac{(C + 1/2)t^2}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_j + \frac{Ct^3}{s_n^3} \sum_{j=0}^{n} \int_{|x|<\epsilon s_n} x^2 \epsilon s_n d\alpha_j.$$

On the second term simplify s_n. Then notice that integrating over Ω yields a greater integral than integrating for $|x| < \epsilon s_n$. Finally notice that the integral over Ω is - by definition - σ^2. Thus

$$\left|\psi_n + \frac{t^2}{2}\right| \leq \frac{(C + 1/2)t^2}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_j + Ct^3\epsilon \frac{1}{s_n^2} \sum_{j=0}^{n} \sigma_j^2.$$

The sum on the second term is equal to s_n^2. Then simplify s_n^2. If we take the limit as n increases arbitrarily on both sides we get

$$\lim_{n\to\infty} \left|\psi_n + \frac{t^2}{2}\right| \leq (C + 1/2)t^2 \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x|\geq \epsilon s_n} x^2 d\alpha_{X_j} + Ct^3\epsilon.$$

Since X is a Lindeberg's Family, the limit on the first term is zero. Since ϵ is arbitrary, take the limit as ϵ decreases to zero to get the result. □

Theorem (Lindeberg's theorem). *Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a Lindeberg's family. Then the Central Limit Theorem holds on X.*

Proof. Just notice that by triangular inequality

$$\left|\log \phi_{\widetilde{S}_n}(t) + \frac{t^2}{2}\right| \leq \left|\log \phi_{\widetilde{S}_n}(t) + \psi_n(t)\right| + \left|\psi_n(t) + \frac{t^2}{2}\right|.$$

Take the limit as n increases arbitrarily on both sides. Both terms on the right hand side become zero as a direct consequence of lemmas (3.1) and (3.2) respectively. Then

$$\lim_{n\to\infty} \left|\log \phi_{\widetilde{S}_n}(t) + \frac{t^2}{2}\right| = 0.$$

By remark (3.2) this implies that the Central Limit Theorem holds on X. □

[8] see propositions (A.4.1) and (A.4.2).

Example 3.1. Let $A \in \Omega$ such that $\mathbb{P}(A) = 1/2$. For all $n \in \mathbb{N}$ let

$$X_n = \begin{cases} a_n & \text{if } \omega \in A \\ -a_n & \text{if } \omega \in \Omega \backslash A \end{cases}.$$

Show that the Central Limit Theorem holds on X if

a) $a_n = 1/\sqrt{n}$ b) $a_n = n$

Solution (a). Compute σ_n^2

$$\sigma_n^2 = \left(\frac{1}{\sqrt{n}}\right)^2 \times \mathbb{P}(A) + \left(-\frac{1}{\sqrt{n}}\right)^2 \times \mathbb{P}(\Omega \backslash A) = \frac{1}{n} \times \frac{1}{2} + \frac{1}{n} \times \frac{1}{2} = \frac{1}{n}.$$

Therefore

$$s_n = \sqrt{\sum_{j=0}^{n} \sigma_j^2} = \sqrt{\sum_{j=0}^{n} \frac{1}{j}}$$

On the other hand, for all $\epsilon > 0$ we have

$$\int_{|x| \geq \epsilon s_n} x^2 d\alpha_j = \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(\left(\frac{1}{\sqrt{j}}\right)^2 \times \mathbb{P}(A) + \left(-\frac{1}{\sqrt{j}}\right)^2 \times \mathbb{P}(\Omega \backslash A) \right) = \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(\frac{1}{j}\right).$$

Therefore

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x| \geq \epsilon s_n} x^2 d\alpha_j = \lim_{n \to \infty} \frac{\sum_{j=0}^{n} \mathbb{1}_{(|a_j| \geq \epsilon s_n)} (1/j)}{\left(\sum_{j=0}^{n} 1/j\right)^2}.$$

In the above limit the denominator diverges to infinity. If we show that the numerator is bounded we have that the limit is zero. This will imply that the Central Limit Theorem holds on X. Note that

$$\lim_{n \to \infty} \frac{|a_n|}{s_n} = \lim_{n \to \infty} \frac{\frac{1}{\sqrt{n}}}{\sum_{j=0}^{n} \frac{1}{j}} = \lim_{n \to \infty} \frac{1}{\sqrt{n} \sum_{j=0}^{n} 1/j} = 0.$$

Hence, by definition of limit, for all $\epsilon > 0$ exists $N \in \mathbb{N}$ such that $n > N$ implies

$$\left| \frac{|a_n|}{s_n} - 0 \right| < \epsilon \iff |a_n| < \epsilon s_n.$$

Thus $\sum_{j=0}^{n} \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(\frac{1}{j}\right)$ has a finite number of terms and as a consequence is finite. □

Solution (b). By analogous computations we have

$$\sigma_n^2 = n^2, \qquad s_n = \sqrt{\sum_{j=0}^{n} j^2}, \qquad \int_{|x| \geq \epsilon s_n} x^2 d\alpha_j = \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(j^2\right).$$

Therefore

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{j=0}^{n} \int_{|x| \geq \epsilon s_n} x^2 d\alpha_j = \lim_{n \to \infty} \frac{\sum_{j=0}^{n} \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(j^2\right)}{\left(\sum_{j=0}^{n} j^2\right)^2}.$$

In the above limit the denominator diverges to infinity. If we show that the numerator is bounded we have that the limit is zero. This will imply that the Central Limit Theorem holds on X. Note that

$$\lim_{n\to\infty} \frac{|a_n|}{s_n} = \lim_{n\to\infty} \frac{n}{\sum_{j=0}^{n} j^2} < \lim_{n\to\infty} \frac{n}{n^2} = 0.$$

Hence, by definition of limit, for all $\epsilon > 0$ exists $N \in \mathbb{N}$ such that $n > N$ implies

$$\left| \frac{|a_n|}{s_n} - 0 \right| < \epsilon \iff |a_n| < \epsilon s_n.$$

Thus $\sum_{j=0}^{n} \mathbb{1}_{(|a_j| \geq \epsilon s_n)} \left(j^2\right)$ has a finite number of terms and as a consequence is finite.
□

3.2 Lyapunov's Condition

Definition. Let $X = \{X_n\}_{\forall n \in \mathbb{N}}$ be a sequence of independent random variables, $\{\alpha_n\}_{n\in\mathbb{N}}$ their respective laws. If exists a positive real δ such that

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=0}^{n} \int |x|^{2+\delta} d\alpha_i = 0,$$

we say the X satisfies the Lyapunov's condition, or is a *Lyapunov's family*.

Proposition 3.2.1 (Lyapunov's Condition)**.** *All Lyapunov's families are Lindeberg's families. Let $X = \{X_n\}_{\forall n\in\mathbb{N}}$ satisfy the Lyapunov's condition. Then X is a Lindeberg's family.*

Proof. Let $X = \{X_n\}_{\forall n\in\mathbb{N}}$ be a Lyapunov's family. Then exists $\delta > 0$ such that the Lyapunov's condition holds. Consider that δ. Let ϵ be any positive real. Recall that integrating over Ω yields a greater integral than integrating for $|x| \geq \epsilon s_n$. Then notice that for $|x| \geq \epsilon s_n$ we have $|x|^2|x|^\delta \geq |x|^2(\epsilon s_n)^\delta$ to get

$$\int |x|^{2+\delta} d\alpha_i \geq \int_{|x|\geq \epsilon s_n} |x|^2|x|^\delta d\alpha_i \geq (\epsilon s_n)^\delta \int_{|x|\geq\epsilon s_n} |x|^2 d\alpha_i.$$

Sum both sides from α_0 to α_n, divide by $s_n^{2+\delta}$ and take the limit as n increases arbitrarily. Notice that on the right hand side s_n^δ simplifies.

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=0}^{n} \int |x|^{2+\delta} d\alpha_i \geq \epsilon^\delta \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=0}^{n} \int_{|x|\geq\epsilon s_n} |x|^2 d\alpha_i.$$

Recall that ϵ and δ are positive. Then ϵ^δ is positive. Furthermore the integrals on the right hand side are always positive, as is s_n. Hence the entire term on the right hand side is positive. On the other hand, since X is a Lyapunov's family the left hand side is zero. Therefore

$$\epsilon^\delta \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=0}^{n} \int_{|x|\geq\epsilon s_n} |x|^2 d\alpha_i = 0.$$

Since ϵ^δ is strictly positive we must have that the limit is zero. Since ϵ is arbitrary the result follows. □

Remark 3.12. As a consequence of the previous proposition and the Lindeberg's theorem, we have that if $X = \{X_n\}_{\forall n \in \mathbb{N}}$ is a Lypunov's family then the Central Limit Theorem holds on X.

Example 3.2. Let $A \in \Omega$ such that $\mathbb{P}(A) = 1/2$. For all $n \in \mathbb{N}$ let

$$X_n = \begin{cases} a_n & \text{if } \omega \in A \\ -a_n & \text{if } \omega \in \Omega \backslash A \end{cases}.$$

Show that the Central Limit Theorem holds on X if $a_n = n$.

Solution. By the example above, we have

$$\sigma_n^2 = n^2, \qquad s_n = \sqrt{\sum_{j=0}^{n} j^2}.$$

Furthermore

$$\int |x|^{2+\delta} d\alpha_j = |j|^{2+\delta} \times \mathbb{P}(A) + |-j|^{2+\delta} \times \mathbb{P}(\Omega \backslash A) = j^{2+\delta} \times \frac{1}{2} + j^{2+\delta} \times \frac{1}{2} = j^{2+\delta}.$$

Therefore

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=0}^{n} \int |x|^{2+\delta} d\alpha_j = \lim_{n\to\infty} \frac{\sum_{j=0}^{n} j^{2+\delta}}{\left(\sqrt{\sum_{j=0}^{n} j^2}\right)^{2+\delta}} < \lim_{n\to\infty} \frac{\sum_{j=0}^{n} j^{2+\delta}}{\left(\sum_{j=0}^{n} j\right)^{2+\delta}}.$$

Consider $\delta = 3$. On the denominator separate the cross terms

$$\left(\sum_{j=0}^{n} j\right)^3 = \sum_{j=0}^{n} j^3 + \sum_{\substack{0 \le j_1, j_2, j_3 \le n \\ j_1 \ne j_2 \ne j_3}} j_1 j_2 j_3.$$

The sum on the left simplifies with the numerator while the sum on the right diverges. Therefore the limit is zero. □

Proposition 3.2.2. *There are Lindeberg's families that are not Lyapunov's families.*

Proof. Let a_n be a sequence of positive reals such that the sum of its squares is convergent i.e. exists a strictly positive real k such that $\sum_{n=0}^{\infty} a_n^2 = k$. This implies that a_n is an infinitesimal, i.e. $\lim_{n\to\infty} a_n = 0$. Let $A \in \Omega$ such that $\mathbb{P}(A) = 1/2$. For all $n \in \mathbb{N}$ let

$$X_n = \begin{cases} a_n & \text{if } \omega \in A \\ -a_n & \text{if } \omega \in \Omega \backslash A \end{cases}$$

Let us show that X is a Lindeberg's family. Consider ϵ any positive real. Then

$$s_n^2 = \sum_{i=0}^{n} \sigma_i^2 = \sum_{i=0}^{n} \left(\frac{1}{2} a_n^2 + \frac{1}{2}(-a_n)^2\right) = \sum_{i=0}^{n} a_n^2.$$

Also

$$\int_{|x| \ge \epsilon s_n} x^2 d\alpha_i = \begin{cases} \frac{1}{2} a_n^2 + \frac{1}{2}(-a_n)^2 = a_n^2 & \text{if } a_n \ge \epsilon s_n^2 \\ 0 & \text{if } a_n < \epsilon s_n^2 \end{cases}.$$

Note that $\lim_{n\to\infty} s_n^2 = k$ and $\lim_{n\to\infty} a_n = 0$ hence we have that exists a natural number N such that for every natural n bigger or equal than N the above integral is zero and for every n smaller than N the above integral is k^2 and N is the smallest natural with this property.

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=0}^{n}\int_{|x|\geq\epsilon s_n}x^2d\alpha_i=\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=0}^{N}\int_{|x|\geq\epsilon s_n}x^2d\alpha_i=\lim_{n\to\infty}\frac{1}{nk^2}Na_n^2=0,$$

and then X is a Lindeberg's family. Now let us show that X is not a Lyapunov's family, let δ be any real positive. Note that

$$\int|x|^{2+\delta}d\alpha_i=\frac{1}{2}|a_n|^{2+\delta}+\frac{1}{2}|-a_n|^{2+\delta}=a_n^{2+\delta}.$$

Now let us evaluate the definition of Lyapunov's family. Just note that $\lim_{n\to\infty}s_n^2=k$ and then

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{i=0}^{n}\int|x|^{2+\delta}d\alpha_i=\lim_{n\to\infty}\frac{1}{(s_n^2)^{\frac{2+\delta}{2}}}\sum_{i=0}^{n}a_n^{2+\delta}=k^{-\frac{2+\delta}{2}}\lim_{n\to\infty}\sum_{i=0}^{n}a_n^{2+\delta}>0.$$

The last limit is a sum of strictly positive terms and therefore it it larger than zero.

Remark 3.13. To complete the proof we just have to provide at least one sequence that satisfies the conditions to be used as a_n. Many exist, one such sequence is $a_n=\sqrt{\frac{1}{2^n}}$ and we have $k=\sum_{n=0}^{\infty}a_n^2=\sum_{n=0}^{\infty}\frac{1}{2^n}=2$.

□

Remark 3.14. From the the proposition above and the proof we can conclude that in general a Lyapunov's family is not a Lindeberg's family.

Chapter 4

Dependence of Variables

4.1 Ergodic Introduction

Remark 4.1. In this chapter the proof of the Kolmogorov Extension and Ergodic Theorems were omitted since the proofs are long and fall outside the scope of this essay. If the reader is interested in reading the proofs, references to them are provided.

Remark 4.2. In previous chapters we were considering sequences of independent random variables, sometimes identically distributed. In this chapter, unless specifically stated, no sequence of random variables is assumed to be independent.

Remark 4.3. Our aim is to prove the Central Limit Theorem for sequences of random variables that are not necessarily independent. To do this we will use Ergodic Theory. We will not digress too much into these topics but will start by presenting the relation between Ergodic Theory and stochastic processes. In this first part we will explain how the Ergodic theorem relates to stochastic processes.

Definition. Let $\mathscr{F}^n$ be the σ-algebra generated by the finite dimensional rectangles, i.e. sets of the form $\{(r_1, r_2, ..., r_n) : r_i \in (a_i, b_i], i = 1, ..., n\}$. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be measures on the space of n-tuples of real numbers $(\mathbb{R}^n, \mathscr{F}^n)$. If for every $n \in \mathbb{N}$ and $a_i, b_i \in \mathbb{R}$ we have

$$\mu_{n+1}((a_1, b_1] \times ... \times (a_n, b_n], \mathbb{R}) = \mu_n((a_1, b_1] \times ... \times (a_n, b_n])$$

We say that $\{\mu_n\}_{n\in\mathbb{N}}$ are *consistent*.

Remark 4.4. Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a stochastic process and $\{\alpha_n\}_{n\in\mathbb{N}}$ their respective laws. Consider the product measures[1] $\mu_n = \alpha_1 \times ... \times \alpha_n$ then

$$\mu_{n+1}((a_1, b_1] \times ... \times (a_n, b_n], \mathbb{R}) = \alpha_1((a_1, b_1]) \times ... \times \alpha_n((a_n, b_n]) \times \alpha_{n+1}(\mathbb{R})$$

Since α_{n+1} is a law, is in particular a probability measure. Therefore we have that $\alpha_{n+1}(\mathbb{R})$ is one. This implies that these laws are consistent .

Theorem (Kolmogorov Extension Theorem)**.** *If $\{\mu_n\}_{n\in\mathbb{N}}$ are consistent measures on the space of n-tuples of real numbers $(\mathbb{R}^n, \mathscr{F}^n)$ then exists a unique probability measure $\mathbb{P}$ on the set of real sequences $(\mathbb{R}^\mathbb{N}, \mathscr{F}^\mathbb{N})$ such that*

$$\mathbb{P}(\{s_n\}_{n\in\mathbb{N}} : 1 \leq i \leq n \quad s_i \in (a_i, b_i]) = \mu_n((a_1, b_1] \times ... \times (a_n, b_n])$$

Proof. see Durrett, 2010 Theorem A.3.1. page 366. □

[1]For more details on product measures see Capinski and Kopp, 2013 Chapter 6 page 159.

Remark 4.5. Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a stochastic process and $\{\alpha_n\}_{n\in\mathbb{N}}$ be their respective laws. Consider the family of functions $Y = \{Y_n\}_{n\in\mathbb{N}}$ such that for any natural n and for any real sequence $s = \{s_n\}_{n\in\mathbb{N}}$, we have

$$Y_n : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R} \quad \text{such that} \quad Y_n(s) = s_n.$$

The laws $\{\alpha_n\}_{n\in\mathbb{N}}$ are consistent by remark (4.4). Hence we can apply the Kolmogorov Extension Theorem to argue that exists a unique probability measure $\mathbb{Q}$ on the set of all real sequences such that

$$\mathbb{Q}(\{s_n\}_{n\in\mathbb{N}} : 1 \leq i \leq n \quad s_i \in (a_i, b_i]) = \mu_n((a_1, b_1] \times ... \times (a_n, b_n]).$$

Notice that the fact that $\mathbb{Q}$ is a probability measure implies that Y is a sequence of random variables. Furthermore by the above equation we have that X and Y have the same distribution (i.e. the same laws) - even though they have different domains. This means that - for the purpose of proving the Central Limit Theorem - working with X or working with Y is equivalent.

Remark 4.6. Given any real sequence we can define a shift operator T that drops the first term in any sequence of real numbers i.e.$T(a_1, a_2, ...) = (a_2, a_3, ...)$. Then it is easy[2] to see that

$$Y_n(s) = Y_1(T^{n-1}s).$$

Hence there exists a $\widehat{T}$ that gives us a similar expression for X. To check this, first note note that given any ω in Ω we can consider the sequence $(X_1(\omega), X_2(\omega), ...)$ which we will denote simply by $\widehat{X}(\omega)$. Hence we have that $\widehat{X} : \Omega \mapsto \mathbb{R}^{\mathbb{N}}$ is a well defined function. Furthermore for any real sequence s such that $\mathbb{Q}(s)$ is not zero we have that $\widehat{X}^{-1}(s)$ is a non empty set[3]. Then define $\widehat{T} : \Omega \mapsto \Omega$ such that $\widehat{T}(\omega)$ is any value in the set $\widehat{X}^{-1} \circ T \circ \widehat{X}(\omega)$.

Remark 4.7. Hence for any process $\{X_n\}_{n\in\mathbb{N}_0}$ exists $T : \Omega \mapsto \Omega$ such that $X_n(s) = X_0(T^n s)$. More importantly given a random variable $X : \Omega \mapsto \mathbb{R}$ and a function $T : \Omega \mapsto \Omega$, the stochastic process

$$X_n(\omega) = X(T^n\omega)$$

is well defined.

Definition. Let $\{X_n\}_{n\in\mathbb{Z}}$ be a sequence of random variables. Then by remark (4.7.) exists a function $T : \Omega \mapsto \Omega$ and a random variable X such that $X_n(s) = X(T^n s)$. We will call (X, T) a *Ergodic representation* of X.

Definition. Let $X = \{X_n\}_{n\in\mathbb{Z}}$ be a sequence of random variables. For all $a, k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$, if the random vectors $(X_a, X_{a+1}, ..., X_{a+b})$ and $(X_{a+n}, X_{a+1+n}, ..., X_{a+b+n})$ have the same distribution (i.e. the same law), we call X a *stationary stochastic process*.

Definition. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $T : \Omega \mapsto \Omega$ a function. If for every element A of $\mathscr{F}$ we have $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$, then we call T *measure-preserving*.

[2]Note that $Y_2(s) = s_2 = (Ts)_1 = Y_1(Ts)$ and then apply an induction argument to conclude for all naturals

[3]since $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\widehat{X}^{-1}(s)) = \mathbb{P}_{\widehat{X}}(s) = \mathbb{Q}(s)$ which is not is not zero.

Proposition 4.1.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables, (X,T) its Ergodic representation and $\sigma(X_1, X_2, ...)$ the sigma algebra generated*[4] *by the random variables $\{X_n\}_{n\in\mathbb{N}}$. Then $\{X_n\}_{n\in\mathbb{N}}$ is a stationary process if and only if T is measure preserving on $\sigma(X)$.*

Proof. Let A be any set in $\sigma(X_1, X_2, ...)$. Then by definition[5], exists a natural n and Borel set B such that $X_n^{-1}(B) = A$.
[$\Longrightarrow$]

$$\mathbb{P}(T^{-1}A) = \mathbb{P}(T^{-1}X_n^{-1}(B)) = \mathbb{P}(\{\omega : X_n(T\omega) \in B\})$$

Using that (X,T) is a Ergodic representation of the process, we get the expression below. Then use the fact that X is stationary and finally use the definition of B to get

$$\mathbb{P}(\{\omega : X_{n+1}(\omega) \in B\}) = \mathbb{P}(\{\omega : X_n(\omega) \in B\}) = \mathbb{P}(A).$$

[$\Longleftarrow$]

$$\mathbb{P}(X_n^{-1}B) = \mathbb{P}(\{\omega : X_n(\omega) \in B\}) = \mathbb{P}(\{\omega : X(T^n\omega) \in B\}).$$

Solving the composition of inverse functions and noting that T is measure preserving we get

$$\mathbb{P}((X \circ T^n)^{-1}B) = \mathbb{P}(T^{-n}(X^{-1}B)) = \mathbb{P}(X^{-1}B).$$

Hence for all n we have that X_n and X have the same distribution, then by definition the process is stationary. □

Theorem (Ergodic Theorem). *Let $f \in L_1(\mathbb{R})$ be a measurable integrable function, $T : \Omega \mapsto \Omega$ be measure preserving and $I = \sigma(\{A : TA = A\})$. Then*

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} f(T^k\omega)}{n} = \mathbb{E}(f|I) \quad a.\ s.$$

Proof. see Varadhan, 2001, Theorem 6.1., page 180. □

Corollary. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a stationary stochastic process, (X,T) one of its stochastic representations and $I = \sigma(X_1, X_2, ...)$. Then*

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} X_k(\omega)}{n} = \mathbb{E}(f|I) \quad a.\ s.$$

Proof. direct consequence of the Ergodic Theorem, proposition (4.1.1) and remark (4.7) □

4.2 Central Limit Theorem

Remark 4.8. Given a Stochastic process $X = \{X_n\}_{n\in\mathbb{N}}$ we defined

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[(X_1 + ... + X_n)^2\right]$$

If the random variables X_n are independent and identically distributed, the above formula simplifies to be just the variance of X_1. If the variables are not identically distributed but are independent, we can simplify the above formula to get what we called s_n in Chapter 3. If the variables are independent then the covariances are

[4]For more details on generated sigma algebras, see Capinski and Kopp, 2013, Subsection 3.5.2., page 67.
[5]Since $\sigma(Y) = \{Y^{-1}(A) : A \in \beta\}$, where β is the Borel sigma algebra.

zero, but the converse is not true. Therefore there are dependent variables with zero covariance. This seems a great place to start since it may give us a easier challenge while still improving our result.

Remark 4.9. Let $Y = \{Y_n\}_{n\in\mathbb{N}}$ be a martingale[6]. For every $n \in \mathbb{N}$ define $X_n = Y_{n+1} - Y_n$. Then for any $t, s \in \mathbb{N}$ such that $t > s$ we have that

$$\mathbb{E}[X_t X_s] = \mathbb{E}[Y_{t+1}Y_{s+1}] - \mathbb{E}[Y_{t+1}Y_s] - \mathbb{E}[Y_t Y_{s+1}] + \mathbb{E}[Y_t Y_s].$$

Since the expected value of the conditional expectation is just the expected value [7] we have that the above expression is equal to

$$\mathbb{E}[\mathbb{E}[Y_{t+1}Y_{s+1}|\mathscr{F}_{s+1}]] - \mathbb{E}[\mathbb{E}[Y_{t+1}Y_s|\mathscr{F}_s]] - \mathbb{E}[\mathbb{E}[Y_t Y_{s+1}|\mathscr{F}_{s+1}]] + \mathbb{E}[\mathbb{E}[Y_t Y_s|\mathscr{F}_s]].$$

In each term take the measurable variable with respect to the sigma algebra outside the conditional expectation. Then use the martingale property to simplify the conditional expectations. Notice that the first term becomes $\mathbb{E}[Y_{s+1}^2]$ and the fourth term its symmetric. A similar thing happens to the second and fourth term with $\mathbb{E}[Y_s^2]$. Then we have that the covariances are zero. It is also simple to verify[8] that - for any n - the expected value of X_n is zero.

Remark 4.10. By the above remark (4.9), if we consider martingale differences not only does the expected become zero but we get

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[X_1^2 + ... + X_n^2\right].$$

Remark 4.11. Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a sequence of martingale differences with finite variance σ^2. We say that the Central Limit Theorem holds[9] on X if for any $t \in \mathbb{R}$ after some algebraic manipulation we have

$$\lim_{n\to\infty} \phi_{\frac{S_n}{\sqrt{n}}}(t) = \exp\left(-\frac{\sigma^2 t^2}{2}\right) \iff \lim_{n\to\infty} \exp\left(\frac{\sigma^2 t^2}{2}\right)\mathbb{E}\left(\exp\left(it\frac{S_n}{\sqrt{n}}\right)\right) - 1 = 0.$$

Notation. For the rest of this chapter we will use the notation

$$\varphi(n, j, t) = \exp\left(\frac{\sigma^2 t^2 j}{2n}\right)\mathbb{E}\left(\exp\left(it\frac{S_j}{\sqrt{n}}\right)\right).$$

When there is no ambiguity in which n and t are being considered we will denote $\varphi(n, j, t)$ simply by φ_j.

Remark 4.12. Hence by the above remark (4.11) we have that if $\lim_{n\to\infty}\varphi(n, n, t) - 1 = 0$ the Central Limit Theorem holds on X.

Remark 4.13. For the second step below sum and subtract $\varphi(n, j, t)$ for each j (such that $1 \le j \le n$)

$$\varphi(n, n, t) - 1 = \varphi(n, n, t) - \varphi(n, 0, t) = \sum_{j=1}^{n}(\varphi(n, j, t) - \varphi(n, j-1, t)).$$

[6] For more details on martingales see Varadhan, 2001 Chapter 5 page 149. For our purposes knowing the definition will be enough.

[7] i.e. $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}]]$

[8] $\mathbb{E}X_n = \mathbb{E}Y_{n+1} - \mathbb{E}Y_n = 0$ since martingales have constant expected value. See proposition A.3.3.

[9] Using the characteristic function definition for the Central Limit Theorem.

Then if

$$\lim_{n\to\infty}\left|\sum_{j=1}^{n}(\varphi(n,j,t)-\varphi(n,j-1,t))\right|=0,$$

the Central Limit Theorem holds on X.

Remark 4.14. Hence it may prove helpful to us to analyze the expressions $\varphi_j - \varphi_{j-1}$ for each j. By definition we have

$$\varphi_j-\varphi_{j-1}=\exp\left(\frac{\sigma^2t^2j}{2n}\right)\mathbb{E}\left(\exp\left(it\frac{S_{j-1}+X_j}{\sqrt{n}}\right)\right)-\exp\left(\frac{\sigma^2t^2j}{2n}-\frac{\sigma^2t^2}{2n}\right)\mathbb{E}\left(\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\right).$$

Use the distributive property on the exponential function of the first term. The rest of the expression can be placed inside the expected value. Then use again the distributive property, this time on the term with S_{j-1} inside the expected value to get

$$\varphi_j-\varphi_{j-1}=\exp\left(\frac{\sigma^2t^2j}{2n}\right)\mathbb{E}\left[\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\left(\exp\left(it\frac{X_j}{\sqrt{n}}\right)-\exp\left(-\frac{\sigma^2t^2}{2n}\right)\right)\right].$$

Remark 4.15. Let us look closer to the term that we could not use the distributive property on. By Taylor expansion of the variable X_t and t in the first and second terms respectively, exists - for each t - a $C_t \in \mathbb{R}$ such that

$$\exp\left(it\frac{X_j}{\sqrt{n}}\right)-\exp\left(-\frac{\sigma^2t^2}{2n}\right)=\left(1+it\frac{X_j}{\sqrt{n}}-\frac{X_j^2t^2}{2n}+t^3\frac{X_j^3}{n^{\frac{3}{2}}}C_t\right)-\left(1-\frac{\sigma^2t^2}{2n}+\frac{\sigma^4t^3}{n^2}C_t\right).$$

The terms of order n^{-1} we will *pass* to the left. The terms of higher order we will leave on the right. The term with X_j we will also leave on the right since it has zero expected value. We get

$$\exp\left(it\frac{X_j}{\sqrt{n}}\right)-\exp\left(-\frac{\sigma^2t^2}{2n}\right)-\frac{(\sigma^2-\xi_j^2)t^2}{2n}=it\frac{X_j}{\sqrt{n}}+t^3\frac{X_j^3}{n^{\frac{3}{2}}}C_t+\frac{\sigma^4t^3}{n^2}C_t.$$

Remark 4.16. The above remark (4.15) gives us an idea of can we prove the Central Limit Theorem in this case. The expression computed in the above remark was taken from inside an expected value so if we apply the expected value to both sides and assume that orders higher than n^{-1} will converge to zero *fast enough* we get

$$\exp\left(it\frac{X_j}{\sqrt{n}}\right)-\exp\left(-\frac{\sigma^2t^2}{2n}\right)\approx\frac{(\sigma^2-X_j^2)t^2}{2n},$$

hence if we look at the sum we wish to show is zero we have

$$\sum_{j=1}^{n}\varphi_j-\varphi_{j-1}\approx\sum_{j=1}^{n}\exp\left(\frac{\sigma^2t^2j}{2n}\right)\mathbb{E}\left[\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\left(\frac{(\sigma^2-X_j^2)t^2}{2n}\right)\right]. \tag{4.2.1}$$

Now if we consider a variable u *close enough* to j such that we pass the sum inside we would get

$$\sum_{j=1}^{n}\varphi_j-\varphi_{j-1}\approx\exp\left(\frac{\sigma^2t^2u}{2n}\right)\mathbb{E}\left[\exp\left(it\frac{S_{u-1}}{\sqrt{n}}\right)\left(\sum_{j=1}^{n}\frac{(\sigma^2-X_j^2)t^2}{2n}\right)\right].$$

The limit of the above sum is shown to be zero by the Ergodic Theorem. Hence the entire sum is zero. Let us now reformulate this idea in mathematical language. Equation 4.2.1 will be shown by lemma 4.1, the choice of the appropriate variable u will be explored in lemma 4.2 and the use of the Ergodic Theorem will be used in the proof of the Central Limit Theorem.

Notation. For the rest of this chapter we will use the notation

$$\theta(n,j,t) = \exp\left(\frac{\sigma^2 t^2 j}{2n}\right) \mathbb{E}\left[\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\left(\frac{(\sigma^2 - X_j^2)t^2}{2n}\right)\right]$$

When there is no ambiguity in which n and t are being used we will denote $\theta(n,j,t)$ simply by θ_j.

Lemma 4.1. *Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a sequence of martingale differences with finite variance σ^2 and $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n|^3$ finite. Then*

$$\lim_{n\to\infty} \sum_{j=1}^{n} |(\varphi_j - \varphi_{j-1}) - \theta_j| = 0$$

Proof. By the considerations made in remark (4.14) and the definition of θ, we have that $(\varphi_j - \varphi_{j-1}) - \theta_j$ is equal to

$$\exp\left(\frac{\sigma^2 t^2 j}{2n}\right) \mathbb{E}\left[\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\left(\exp\left(it\frac{X_j}{\sqrt{n}}\right) - \exp\left(-\frac{\sigma^2 t^2}{2n}\right) - \frac{(\sigma^2 - X_j^2)t^2}{2n}\right)\right].$$

By the Taylor expansion computed in remark (4.15) the expression becomes

$$(\varphi_j - \varphi_{j-1}) - \theta_j = \exp\left(\frac{\sigma^2 t^2 j}{2n}\right) \mathbb{E}\left[\exp\left(it\frac{S_{j-1}}{\sqrt{n}}\right)\left(it\frac{X_j}{\sqrt{n}} + t^3\frac{X_j^3}{n^{\frac{3}{2}}}C_t + \frac{\sigma^4 t^3}{n^2}C_t\right)\right].$$

Apply the absolute value to both sides and use the properties of the expected value to get

$$|(\varphi_j - \varphi_{j-1}) - \theta_j| \leq \exp\left(\frac{\sigma^2 t^2 j}{2n}\right) \mathbb{E}\left|it\frac{X_j}{\sqrt{n}} + t^3\frac{X_j^3}{n^{\frac{3}{2}}}C_t + \frac{\sigma^4 t^3}{n^2}C_t\right|.$$

Note that the expected value of X is zero and that $j < n$ and you get the expression

$$|(\varphi_j - \varphi_{j-1}) - \theta_j| \leq \exp\left(\frac{\sigma^2 t^2}{2}\right)\left(|t|^3\frac{\mathbb{E}|X_j^3|}{n^{\frac{3}{2}}}|C_t| + \frac{\sigma^4|t|^3}{n^2}|C_t|\right).$$

The result follows from summing both sides in j between 1 and n and apply the limit as n goes to infinity. □

Remark 4.17. For any $m \in \mathbb{N}$ such that $m < n$ we have that exists $q, r \in \mathbb{N}$ such that $n = qm + r$. Therefore for any $m \in \mathbb{N}$ we may divide the interval $[0, n]$ into q intervals of size m and final interval of size r. We will denote each of the q intervals by $B_m^n(0), \ldots, B_m^n(q-1)$ and the final interval[10] by $B_m^n(q)$.

Notation. Consider $n, m \in \mathbb{N}$ fixed such that $m < n$ and the blocks $B_m^n(0), \ldots,$ $B_m^n(q-1)$. For any $j \in \mathbb{N}$ in the interval $[0, n]$ we have that exists a unique $k \in \mathbb{N}$ such that $j \in B_m^n(k)$. We will denote by b_j the smallest integer in $B_m^n(k)$.

[10]note that this final interval may be empty.

Notation. For any k from zero to q we can define a step function - block by block - such that for j in the block $B_m^n(k)$ we have

$$\widehat{\theta}^{(m)}(n,j,t) = \exp\left(\frac{\sigma^2 t^2 b_j}{2n}\right) \mathbb{E}\left[\exp\left(it\frac{S_{b_j}}{\sqrt{n}}\right)\left(\frac{(\sigma^2 - X_j^2)t^2}{2n}\right)\right].$$

When there is no ambiguity in which n, t and m are being considered we will denote $\widehat{\theta}(n,j,t)$ simply by $\widehat{\theta}_j$.

Remark 4.18. Notice that we still have the same problem with passing the sum inside when considering the interval $[0,n]$ since b_j depends of j. However within each block we have that b_j is constant. Therefore we can circumvent the problem. We hope that this transformation allows for $\widehat{\theta}$ *close enough* to θ.

Lemma 4.2. *Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a sequence of martingale differences with zero mean and finite variance σ^2. Then*

$$\lim_{n\to\infty}\sum_{j=1}^{n}|\widehat{\theta}_j - \theta_j| = 0.$$

Proof. We start by taking a closer look at $|\widehat{\theta}_j - \theta_j|$. Notice that using the distribution law and a simple algebraic manipulation we get

$$|\widehat{\theta}_j - \theta_j| \leq \mathbb{E}\left|\frac{(\sigma^2 - X_j^2)t^2}{2n}\left(\exp\left(\frac{\sigma^2 t^2 b_j}{2n}\right)\exp\left(it\frac{S_{b_j}}{\sqrt{n}}\right) - \exp\left(\frac{\sigma^2 t^2 j}{2n}\right)\exp\left(it\frac{S_j}{\sqrt{n}}\right)\right)\right|.$$

We have an inequality above because we passed the absolute value inside expected value. Using the distributive law[11] again we get that the above expected value is equal to

$$\mathbb{E}\left|\exp\left(\frac{\sigma^2 t^2 b_j}{2n}\right)\exp\left(it\frac{S_{b_j}}{\sqrt{n}}\right)\frac{(\sigma^2 - X_j^2)t^2}{2n}\left(1 - \exp\left(\frac{\sigma^2 t^2 (j-b_j)}{2n}\right)\exp\left(it\frac{S_j - S_{b_j}}{\sqrt{n}}\right)\right)\right|.$$

The second exponential in the expected value - by definition of (imaginary) exponential function - has absolute value equal to one. On the other hand by construction $b_k < n$. Then the first exponential in the expected value has absolute value that does not depend on n, i.e. for each t exists C_t such that

$$|\widehat{\theta}_j - \theta_j| \leq \mathbb{E}\left|C_t\frac{(\sigma^2 - X_j^2)t^2}{2n}\left(1 - \exp\left(\frac{\sigma^2 t^2 (j-b_j)}{2n}\right)\exp\left(it\frac{S_j - S_{b_j}}{\sqrt{n}}\right)\right)\right|.$$

Sum both sides on j from 1 to n. For the right side notice that the sum of a finite number of terms is smaller than the maximum of those terms multiplied by the number of times we are summing, to get

$$\sum_{j=1}^{n}|\widehat{\theta}_k - \theta_j| \leq n\sup_{1\leq j\leq n}\mathbb{E}\left|C_t\frac{(\sigma^2 - X_j^2)t^2}{2n}\left(1 - \exp\left(\frac{\sigma^2 t^2 (j-b_j)}{2n}\right)\exp\left(it\frac{S_j - S_{b_j}}{\sqrt{n}}\right)\right)\right|.$$

Simplify n. By an argument of stationarity conclude that $S_j - S_{b_j}$ is equal in distribution to S_{j-b_j} and X_j is equal in distribution to X_{j-b_j}. Furthermore $j - b_j$ with

[11] We have two terms - each with two exponentials - multiply and divide the second term by the first and then use the distribuitive property on the first term.

$0 \leq j \leq n$ is equivalent - by construction of b_j - to have j just in the first block. Hence

$$\sum_{j=1}^{n} |\widehat{\theta}_k - \theta_j| \leq \frac{t^2}{2} \sup_{1\leq j\leq m} \mathbb{E}\left[C_t|\sigma^2 - X_j^2| \left|1 - \exp\left(\frac{\sigma^2 t^2 j}{2n}\right)\exp\left(it\frac{S_j}{\sqrt{n}}\right)\right|\right].$$

Now take the limit as n grows to infinity on both sides. Using the Lebesgue Convergence Theorem on the right side to place the limit inside the expected value, get us the result. The use of the theorem is justified since

$$C_t|\sigma^2 - X_j^2| \left(1 + \left|\exp\left(\frac{\sigma^2 t^2 (j - b_k)}{2n}\right)\exp\left(it\frac{S_j - S_{b_k}}{\sqrt{n}}\right)\right|\right) \leq (1 + C_t)C_t|\sigma^2 - X_j^2|,$$

which is integrable (i.e. has finite expected value), does not depend on n and is larger than what is inside the expected value for all natural n. □

Theorem (Central Limit Theorem). *Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a martingale difference with zero mean and finite variance σ^2 and $\mathbb{E}|X_n|^3$ finite. Then the Central Limit Theorem holds on X.*

Proof. Sum and subtract θ and $\widehat{\theta}$. Now use triangular inequality twice to get

$$\lim_{n\to\infty}\left|\sum_{j=1}^{n}\varphi_j - \varphi_{j-1}\right| \leq \lim_{n\to\infty}\left|\sum_{j=1}^{n}\varphi_j - \varphi_{j-1} - \theta_j\right| + \lim_{n\to\infty}\sum_{j=1}^{n}\left|\theta_j - \widehat{\theta}_j\right| + \lim_{n\to\infty}\left|\sum_{j=1}^{n}\widehat{\theta}_j\right|.$$

The first and second term are zero as direct conclusions of lemma (4.1) and lemma (4.2) respectively. Hence if we show that the last term is zero we complete the proof. To do this notice that $\widehat{\theta}_j$ is defined in blocks. Therefore we may arrange the sum to reflect that (we are only using the commutative law of the sum). Recall that there are q blocks of size m and final block of size r.

$$\left|\sum_{j\in B_s}\widehat{\theta}_j\right| = \left|\sum_{k=0}^{q}\sum_{j\in B_m^n(k)}\exp\left(\frac{\sigma^2 t^2 b_k}{2n}\right)\mathbb{E}\left[\exp\left(it\frac{S_{b_k}}{\sqrt{n}}\right)\left(\frac{(\sigma^2 - X_{b_k+j}^2)t^2}{2n}\right)\right]\right|.$$

Everything that does not depend directly on j may be put outside the inner sum[12]

$$\left|\frac{1}{n}\sum_{k=0}^{q}\frac{t^2}{2}\exp\left(\frac{\sigma^2 t^2 b_k}{2n}\right)\mathbb{E}\left[\exp\left(it\frac{S_{b_k}}{\sqrt{n}}\right)\sum_{j\in B_m^n(k)}\sigma^2 - X_{b_k+j}^2\right]\right|.$$

Notice that by construction $b_k < n$ so the first exponential that appears has absolute value smaller than a constant C_t that only depends on t. Furthermore when passing the absolute value inside the expected value we get a bigger expression and the exponential inside the expected value has absolute value equal to one. Hence the above expression is less or equal than

$$\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq \sum_{k=0}^{q}\frac{C_t}{n}\mathbb{E}\left|\sum_{j\in B_m^n(k)}\sigma^2 - X_{b_k+j}^2\right|.$$

[12]Notice that this step would not be possible if we were working with θ instead of $\widehat{\theta}$ and is the reason for the change.

Recall that the last block is slightly different from the others. To address this we will separate the last block. Notice that we are working with stationary sequences hence starting the sum at b_k or at 1 shall make no difference in terms of expected value, therefore we will start at one.

$$\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq \sum_{k=0}^{q-1}\frac{C_t}{n}\mathbb{E}\left|\sum_{j=1}^{m}\sigma^2 - X_j^2\right| + \frac{C_t}{n}\mathbb{E}\left|\sum_{j=1}^{r}\sigma^2 - X_j^2\right|.$$

Notice that nothing depends on k anymore, hence we can express that sum as a simple multiplication. Furthermore we may multiply and divide by m and r in the normal block and last block respectively.

$$\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq C_t\frac{qm}{n}\mathbb{E}\left|\sum_{j=1}^{m}\frac{\sigma^2 - X_j^2}{m}\right| + C_t\frac{r}{n}\mathbb{E}\left|\sum_{j=1}^{r}\frac{\sigma^2 - X_j^2}{r}\right|.$$

In the above expression take the limit when n grows arbitrary.

Remark 4.19. Let n be any natural number, m be the smallest natural number bigger (not equal) than $n/2$. Then let r be such that $n = m \times 1 + r$. This means that r is greatest natural smaller or equal to $n/2$. In this case as n increases arbitrarily both m and r increase arbitrarily.

By the above remark we can conclude that we may consider m and r also growing arbitrarily, hence we have

$$\lim_{n\to\infty}\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq \lim_{n\to\infty}\lim_{m\to\infty} C_t\frac{qm}{n}\mathbb{E}\left|\sum_{j=1}^{m}\frac{\sigma^2 - X_j^2}{m}\right| + \lim_{n\to\infty}\lim_{r\to\infty} C_t\frac{r}{n}\mathbb{E}\left|\sum_{j=1}^{r}\frac{\sigma^2 - X_j^2}{r}\right|$$

Notice that the random variables $\sigma^2 - X_j^2$ are in the conditions of the Ergodic Theorem[13]. Hence we apply the Ergodic theorem to both sums inside the expected values and conclude that those sums converge to $\mathbb{E}[\sigma - X^2|I]$, where I is determined by the theorem[14]. Therefore we will represent the first sum as $\delta(m)$ and the second sum as $\delta(r)$. Finally let $\delta^*(m,r) = \max\{\delta(m), \delta(r)\}$. Since we are only looking for an upper bond we get

$$\lim_{n\to\infty}\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq \lim_{n\to\infty}\lim_{m\to\infty}\lim_{r\to\infty}\left(C_t\frac{qm}{n}\mathbb{E}\left|\delta^*(m,r)\right| + C_t\frac{r}{n}\mathbb{E}\left|\delta^*(m,r)\right|\right).$$

Now use the distributive law and recall that $n = mq + r$ to conclude that the fractions simplify to one. Therefore our expression no longer depends on n and as such the limit on n is dropped. Furthermore, by definition, the limit of δ^* when both variables are increasing arbitrarily is equal to the limit of $\delta(k)$ as k grows arbitrarily. Hence we have

$$\lim_{n\to\infty}\left|\sum_{j\in B_s}\widehat{\theta}_j\right| \leq C_t\lim_{k\to\infty}\mathbb{E}\left|\delta(k)\right| = C_t\mathbb{E}\left|\mathbb{E}[\sigma - X^2|I]\right| = C_t\mathbb{E}\left|\sigma - X^2\right| = 0$$

and the result follows. □

[13]They are stationary and integrable.

[14]The definition of I will not be relevant, so we will not worry to much about it. Just keep in mind that it is a well defined sigma algebra.

Appendix A

General results

A.1 Convolutions

Notation. Let X be random variable, μ_X be the the law of X, F_X its distribution function and f be a measurable function. Then we denote $\int f(x)\mu(dx)$ by $\int f(x)dF_X(x)$.

Definition. Let F be a function, Y be random variable and μ_Y be the law of Y. We call

$$\int F(y-x)d\mu_Y$$

the *convolution* between F and F_Y and denote it by $F * F_Y$.

Proposition A.1.1. *Let $F : \mathbb{R} \mapsto [0,1]$ be a function such that*

- *is increasing (i.e. $\forall x \leq y \quad F(x) \leq F(y)$)*
- *is continuous from the right at every point (i.e. $\forall a \in \mathbb{R} \quad \lim_{x\to a^+} F(x) = F(a)$)*
- *the limit of $F(x)$ as x increases is 1 and as decreases is zero (i.e. $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$).*

Then exists a random variable X such that $F_X = F$.

Proof. See Theorem 1.2.2., page 9 of Durrett, 2010. □

Proposition A.1.2. *If F and G are distributions then $F * G$ is a distribution, i.e. exists a random variable X such that $F_X = F * G$.*

Proof. Note that $\forall x \leq y \quad F(x) \leq F(y) \implies F * G(x) = \int F(x-z)dG(z) \leq \int F(y-z)dG(z) = \int F(y-z)dG(z) = F * G(y)$.

On the other hand, by the monotonic convergence theorem, $\forall a \in \mathbb{R}, \quad \lim_{x\to a^+} F * G(x) = \lim_{x\to a} \int F(x-z)dG(z) = 1 - \lim_{x\to a^+} \int 1 - F(x-z)dG(z) = 1 - \int 1 - F(a^+ - z)dG(z) = F * G(a)$.

Finally, by the monotonic convergence theorem $\lim_{x\to\infty} F * G(x) = \lim_{n\to\infty} \int F(x-z)dG(z) = \int \lim_{x\to\infty} F(x-z)dG(z) = \int 1dG(z) = 1$. Analogously one can prove $\lim_{x\to-\infty} F(x) = 0$

Therefore by proposition (A.1.1) the result follows. □

Notation. Given two random variables X and Y, we will denote by $X * Y$ the random variable which has distribution $F_X * F_Y$.

Proposition A.1.3. *Let X and Y be random variables, F_X and F_Y their respective distributions. The random variable with distribution $F * G$ is $Z = X + Y$.*

Proof. See Theorem 2.1.10., page 42 of Durrett, 2010. □

A.2 Fourier Transform

This section serves as an introduction to this chapter, many of the concepts discussed here are elemental and reader that are familiar with Fourier transforms can skip it. For those that find this chapther too sumarized in the book of Rudin, 1987 starting on page 178, one may find a more detailed introduction to Fourier transforms.

Definition. Let f be an integrable function. Then we define the *Fourier transform of* f as

$$\hat{f}(t) = \int f(x)e^{-itx}dx.$$

Remark A.1. Recall that for any non zero reals k, a and b we have

$$\int e^{ikx}dx = \int (cos(kx) + isin(kx))dx$$

Computing the integral

$$\frac{1}{k}sin(kx) - \frac{i}{k}cos(kx) = \frac{1}{ik}(isin(kx) + cos(kx)) = \frac{1}{ik}e^{ikx}$$

, then

$$\int (ax+b)e^{ikx}dx = (ax+b)\frac{1}{ik}e^{ikx} - \int a\frac{1}{ik}e^{ikx}dx = (ax+b)\frac{1}{ik}e^{ikx} + a\frac{1}{ik}\frac{1}{ik}e^{ikx} \tag{A.2.1}$$

$$= (ax+b)\frac{1}{ik}e^{ikx} - \frac{a}{k^2}e^{ikx}, \tag{A.2.2}$$

and,

$$sin(z) = \frac{e^{iz} - e^{-iz}}{2i}$$

. This will be necessary for the computations in the next example.

Example A.1. Let a and b be real numbers such that $a < b$, let h be a positive real such that $h < \frac{b-a}{2}$ and consider the function

$$f_{a,b,h}(x) = \begin{cases} 0 & \text{if} \quad x \leq a-h \\ \dfrac{x-a+h}{2h} & \text{if} \quad a-h \leq x \leq a+h \\ 1 & \text{if} \quad a+h \leq x \leq b-h \\ 1 - \dfrac{x-b+h}{2h} & \text{if} \quad b-h \leq x \leq b+h \\ 0 & \text{if} \quad b+h \leq x. \end{cases} \tag{A.2.3}$$

Show that

$$\hat{f}_{a,b,h}(t) = \frac{1}{\sqrt{2\pi}}\frac{e^{-iat} - e^{-ibt}}{it}\frac{\sin ht}{ht} \tag{A.2.4}$$

Solution:

$$\begin{aligned}
\sqrt{2\pi}\hat{f}_{a,b,h}(t) &= \int_{\mathbb{R}} f_{a,b,h}(x)e^{-itx}dx \\
&= \int_{a-h}^{a+h} \frac{x-a+h}{2h} e^{-itx}dx + \int_{a+h}^{b-h} e^{-itx}dx + \int_{b-h}^{b+h}\left(1-\frac{x-b+h}{2h}\right)e^{-itx}dx \\
&= \left[\frac{x-a+h}{2h}\frac{1}{it}e^{-itx} - \frac{1}{2h(-it)^2}e^{-itx}\right]_{a-h}^{a+h} + \left[\frac{1}{it}e^{-itx}\right]_{a+h}^{b-h} + \\
&\quad + \left[\left(1-\frac{x-b+h}{2h}\right)\frac{1}{it}e^{-itx} - \frac{1}{-2h(-it)^2}e^{itx}\right]_{b-h}^{b+h} \\
&= \left[\frac{1}{it}e^{-it(a+h)} - \frac{1}{2hi^2t^2}e^{-it(a+h)} + \frac{1}{2hi^2t^2}e^{-it(a-h)}\right] + \left[\frac{1}{it}e^{-it(b-h)} - \frac{1}{it}e^{-it(a+h)}\right] + \\
&\quad + \left[-\frac{1}{-2hi^2t^2}e^{-it(b+h)} - \frac{1}{it}e^{-it(b-h)} + \frac{1}{-2hi^2t^2}e^{-it(b-h)}\right] \\
&= \left[-\frac{1}{2hi^2t^2}e^{-it(a+h)} + \frac{1}{2hi^2t^2}e^{-it(a-h)}\right] + \left[\frac{1}{2hi^2t^2}e^{-it(b+h)} - \frac{1}{2hi^2t^2}e^{-it(b-h)}\right] \\
&= \frac{1}{2hi^2t^2}\left(-e^{-it(a+h)} + e^{-it(a-h)} + e^{-it(b+h)} - e^{-it(b-h)}\right) \\
&= \frac{1}{2hi^2t^2}\left(\left(e^{-ita}-e^{-itb}\right)e^{ith} - \left(e^{-ita}-e^{-itb}\right)e^{-ith}\right) \\
&= \frac{1}{hit^2}\left(e^{-ita}-e^{-itb}\right)\frac{\left(e^{ith}-e^{-ith}\right)}{2i} \\
&= \frac{e^{-iat}-e^{-ibt}}{it}\frac{\sin ht}{ht}.
\end{aligned}$$

Then we have $\hat{f}_{a,b,h}(t) = \frac{1}{\sqrt{2\pi}}\frac{e^{-iat}-e^{-ibt}}{it}\frac{\sin ht}{ht}$

Proposition A.2.1. *Let f be an $L^1(\mathbb{R})$ function. Then the Fourier transform $\hat{f}$ of f is bounded by a real constant C.*

Proof. Since f is in $L^1(\mathbb{R})$ we have that $\int_{\mathbb{R}} |f(x)|\,dx < \infty$.

$$\begin{aligned}
|\hat{f}(t)| &= \left|\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)e^{-itx}dx\right| \\
&\leq \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |f(x)|\left|e^{-itx}\right|dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |f(x)|\,dx < \infty.
\end{aligned}$$

□

Theorem (Riemann-Lebesgue)**.** *Let f be an L^1 function. Then*

$$\lim_{|t|\to\infty}\int_{\mathbb{R}} f(x)e^{itx}dx = 0$$

.

Example A.2. Let X and Y be random variables, μ and λ their respective probability measures and ϕ_μ and ϕ_λ their respective characteristic functions. Then we have

$$\lim_{b\to\infty}\frac{1}{2\pi}\int_{\mathbb{R}}(\phi_\mu(y)-\phi_\lambda(y))\frac{e^{-iby}}{iy}\frac{\sin hy}{hy}dy \tag{A.2.5}$$

Solution: We will compute the limit using the Riemann Lebesgue theorem (A.2), in order to justify its use we must ensure that $g(y)=\frac{(\phi_\mu(y)-\phi_\lambda(y))\sin hy}{ihy^2}$ is integrable.

$$\begin{aligned}\int_{\mathbb{R}}\left|\frac{(\phi_\mu(y)-\phi_\lambda(y))\sin hy}{ihy^2}\right|dy &= \int_{\mathbb{R}}\frac{|(\phi_\mu(y)-\phi_\lambda(y))||\sin hy|}{|h|y^2}dy\\ &\leq \int_{\mathbb{R}}\frac{2|\sin hy|}{|h|y^2}dy\\ &= 2\int_0^{+\infty}\frac{|\sin hy|}{|h|y^2}dy\\ &= 2\int_0^1\frac{|\sin hy|}{|h|y^2}dy+2\int_1^{+\infty}\frac{|\sin hy|}{|h|y^2}dy\end{aligned}$$

We have that

$$\int_1^{+\infty}\frac{|\sin hy|}{|h|y^2}dy \leq \int_1^{+\infty}\frac{1}{|h|y^2}dy<\infty$$

To show that g is integrable on the closed compact set $[0,1]$ we must only show that it is continuous on the same set. It is obviously continuous on $]0,1]$ so we just have to check that it is continuous on zero. To achieve that we must only check that $\lim_{y\to 0}|g(y)|<\infty$. First notice that by (A.4.1)

$$\begin{aligned}|\phi_\mu(t)-\phi_\lambda(t)| &= |\mathbb{E}(e^{itX})-\mathbb{E}(e^{itY})|\\ &= |\mathbb{E}(e^{itX}-1-itX)-\mathbb{E}(e^{itY}-1-itY)+\mathbb{E}(it(X-Y))|\\ &\leq \mathbb{E}|e^{itX}-1-itX|+\mathbb{E}|e^{itY}-1-itY|+\mathbb{E}|it(X-Y)|\\ &\leq \mathbb{E}(Ct^2)+\mathbb{E}(Ct^2)+t\mathbb{E}|(X-Y)|\leq Ct^2+Ct=Ct(t+1).\end{aligned}$$

Then

$$\lim_{y\to 0}|g(y)|=\lim_{y\to 0}\frac{|\phi_\mu(y)-\phi_\lambda(y)|}{y}\frac{\sin hy}{hy}\leq \lim_{y\to 0}\frac{Cy(y+1)}{y}\lim_{y\to 0}\frac{\sin hy}{hy}=C\times 1\leq\infty.$$

Then

$$\lim_{b\to\infty}\frac{1}{2\pi}\int_{\mathbb{R}}\phi(y)\frac{e^{-iby}}{iy}\frac{\sin hy}{hy}dy=\lim_{b\to\infty}\frac{1}{2\pi}\int_{\mathbb{R}}\frac{\phi(y)\sin hy}{ihy^2}e^{i(-b)y}dy=0.$$

A.3 measure theory

Proposition A.3.1. *Let $(\Omega,\mathscr{F},\mathbb{P})$ a probability space. Then if X is a random variable,*

$$\mathbb{P}\circ X^{-1}$$

is a probability measure.

Definition. For all X random variable we call the *law of* X to the measure $\mathbb{P} \circ X^{-1}$. We represent the law of X by α_X, where the subscript might get dropped (or changed) when there is no ambiguity.

Remark A.2. The normal distribution has bounded density.

Proposition A.3.2. *Let a be any non zero real number. Then*

$$\int \frac{1}{\sqrt{2\pi a}} x^2 \exp\left(-\frac{x^2}{2a}\right) dx = a.$$

Proof. Let X be a random variable with normal distribution (with $\mu = 0$ and $\sigma^2 = a$) then

$$\int \frac{1}{\sqrt{2\pi a}} x^2 \exp\left(-\frac{x^2}{2a}\right) dx = \int x^2 f_X(x) dx = \mathbb{E}X^2 = \sigma^2 = a.$$

□

Remark A.3.

$$\int_0^\infty x^3 \exp\left(-\frac{x^2}{4}\right) dx = -\int_0^\infty 2x^2 \frac{x}{2} \exp\left(-\frac{x^2}{4}\right) dx$$

Integrating by parts,

$$= -0 + \int_0^\infty 4x \exp\left(-\frac{x^2}{4}\right) dx = -8 \exp\left(-\frac{x^2}{4}\right)\Bigg|_0^\infty = 0 - (-8) = 8.$$

Theorem (Inversion Formula)**.** *Let* $f \in L^1$, $\widehat{f} \in L^1$ *and*

$$g(x) = \int_{\mathbb{R}} \widehat{f}(t) e^{itx} dm(t).$$

Then g *is continuous and* $f(x) = g(x)$ *a.e.*

Proof. see Rudin, 1987, Theorem 9.11., page 185. □

Theorem (Fatou)**.** *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be a sequence of non-negative measurable functions. Then*

$$\liminf_{n\to\infty} \int f_n(x) d\mu \leq \int \liminf_{n\to\infty} f_n(x) d\mu.$$

Proof. See Capinski and Kopp, 2013, Theorem 4.7., page 82. □

Theorem (Monotonic Convergence Theorem)**.** *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be a point wise increasing sequence of non-negative measurable functions. If* $\lim_{n\to\infty} f_n(x) : \mathbb{R} \mapsto \mathbb{R}$ *is measurable then*

$$\lim_{n\to\infty} \int f_n(x) d\mu = \int \lim_{n\to\infty} f_n(x) d\mu$$

Proof. By Fatou (A.3) (first inequality), by definition of lim inf (second inequality), lim sup (third inequality) and by the fact that f_n is increasing hence for any n we have $f_n \leq \lim_{n\to\infty} f_n(x)$ (fourth inequality),

$$\int \lim_{n\to\infty} f_n(x) d\mu \leq \liminf_{n\to\infty} \int f_n(x) d\mu \leq \lim_{n\to\infty} \int f_n(x) d\mu \leq \limsup_{n\to\infty} \int f_n(x) d\mu \leq \int \lim_{n\to\infty} f_n(x) d\mu$$

the Theorem follows. □

Theorem (Fubini Theorem). *Let $f \in L^1(\Omega_1, \Omega_2)$. Then the sections are integrable in appropriate spaces, the functions*

$$\omega_1 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) dP_2(\omega_2), \qquad \omega_2 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) dP_1(\omega_1)$$

are in $L^1(\Omega_1)$, $L^1(\Omega_2)$, respectively, and

$$\int_{\Omega_1 \times \Omega_2} f d(P_1 \times P_2) = \int_{\Omega_1} \int_{\Omega_2} f dP_2 dP_1 = \int_{\Omega_2} \int_{\Omega_1} f dP_1 dP_2.$$

Proof. See Capinski and Kopp, 2013, Theorem 6.10., page 171. □

Proposition A.3.3. *Martingales have constant expected value.*

Proof. Let X_n be a martingale then

$$\mathbb{E}X_n = \mathbb{E}\mathbb{E}[X_n | \mathscr{F}_0] = \mathbb{E}X_0.$$

□

A.4 Bounds

Remark A.4. Just note that by Taylor expansion of order 2 we have that exists c in the interval $]0, x[$ such that

$$e^{-x} = 1 - x + \underbrace{\frac{c^2}{2}}_{\geq 0}.$$

Hence

$$1 - x \leq e^{-x}.$$

Proposition A.4.1. *Let x be a real number in the interval $]0, 1[$. Then $|e^{-x} - 1 + x| \leq \frac{x^2}{2}$*

Proof. We have that $e^{-x} = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!}$. Then

$$|e^{-x} - 1 + x| \leq \left| \frac{x^2}{2} - \frac{x^3}{3!} + \frac{x^4}{4!} - ... \right| \leq \frac{x^2}{2}$$

The last step is because we have a series of decreasing alternating terms (they are decreasing because the series is convergent). □

Proposition A.4.2. *Given a positive integer n, exists a real number C such that for every real number x*

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

Proof. Consider the Taylor expansion of order $n+1$ of the exponencial function $e^x = \sum_{j=0}^{n} \frac{x^{n+1}}{j!} + \frac{x^j}{(n+1)!} e^{\theta}$, where θ is a specific real number in the interval $]0, x[$.

$$\begin{aligned}\left|e^{ix}-\sum_{j=0}^{n}\frac{(ix)^j}{j!}\right| &= \left|\sum_{j=0}^{n}\frac{(ix)^j}{j!}+\frac{(ix)^{n+1}}{(n+1)!}e^{i\theta}-\sum_{j=0}^{n}\frac{(ix)^j}{j!}\right| \\ &= \left|\frac{(ix)^{n+1}}{(n+1)!}e^{i\theta}\right| \\ &\leq \underbrace{\left|\frac{(i)^{n+1}}{(n+1)!}\right|}_{C}|x|^{n+1}\underbrace{\sup_{a\in\mathbb{R}}\left|e^{ia}\right|}_{\leq 1}.\end{aligned}$$

□

Corollary. *Exists a real number C such that for every real number x*

$$|e^{ix}-1-ix|\leq Cx^2. \tag{A.4.1}$$

Corollary. *Exists a real number C such that for every real number x*

$$\left|e^{ix}-1-ix+\frac{x^2}{2}\right|\leq Cx^3. \tag{A.4.2}$$

Proposition A.4.3. *Let a, b and C be real numbers such that $|a|$ and $|b|$ are less or equal to C. Then for any natural number n we have the inequality*

$$|a^n-b^n|\leq n|a-b|C^{n-1}.$$

Proof. For every natural m let $\alpha_m=a^{n-m}b^m$. Then we have that for any natural n

$$\begin{aligned}\alpha_0-\alpha_n &= \alpha_0-\alpha_1+\alpha_1-\alpha_2+...+\alpha_{n-2}-\alpha_{n-1}+\alpha_{n-1}-\alpha_n \\ &= \sum_{m=0}^{n-1}\alpha_{n-m}-\alpha_m.\end{aligned}$$

Now note that for every natural m we have

$$\alpha_{n-m}-\alpha_m=a^{n-m}b^m-a^{n-m-1}b^{m+1}=a^{n-m-1}b^m(a-b)\leq C^{n-m-1}C^m(a-b)=C^{n-1}(a-b)$$

Then

$$|a^n-b^n|=\left|\sum_{m=0}^{n-1}a^{n-m}b^m-a^{n-m-1}b^{m+1}\right|\leq\sum_{m=0}^{n-1}\left|C^{n-1}(a-b)\right|=n|a-b|C^{n-1}$$

□

Theorem (Holder)**.** *Let X, Y be random variables and p and q be positive real numbers such that $\mathbb{E}|X|^p<\infty$, $\mathbb{E}|Y|^q<\infty$ and $\frac{1}{p}+\frac{1}{q}=1$. Then*

$$\mathbb{E}|XY|\leq(\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|X|^q)^{\frac{1}{q}}.$$

Corollary. *Let a and b be real numbers such that $a < b$, and X be a random variable such that $\mathbb{E}|X|^b < \infty$. Then $\mathbb{E}|X|^a$ exists and*

$$\mathbb{E}|X|^a \leq (\mathbb{E}|X|^b)^{\frac{a}{b}}.$$

Proof. Just apply the Holder's Theorem (A.4) to X and 1_Ω. □

A.5 Order

Definition. Let A and B be any sets. Then denote

$$A^B = \{f : B \mapsto A\}.$$

Let $f, g \in \mathbb{R}^{\mathbb{R}}$. If

$$\exists x_0 \in \mathbb{R} \quad \exists M \in \mathbb{R}^+ : x > x_0 \implies |f(x)| \leq Mg(x)$$

we say that f *has order* g and we denote

$$O(f) = \{g \in \mathbb{R}^{\mathbb{R}} : g \text{ has order } f\}$$

Remark A.5. Many times, when we are interested that $\exists g \in O(f)$, this is the only property we are interested about of g. Therefore it is very common, and we shall do it too, to abuse the above notation $O(f)$ and identify O(f) as some particular function that belongs to O(f). The example below clarifies the notation.

Example A.3. Correct Notation:

$$\exists g \in O(x-1)^2 \quad \log x = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} (x-1)^n.$$

Separating the first term

$$(x-1) + \sum_{n=2}^{\infty} (-1)^{n+1} \frac{1}{n} (x-1)^n = (x-1) + g(x).$$

By abuse we will use

$$\log x = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} (x-1)^n = (x-1) + \sum_{n=2}^{\infty} (-1)^{n+1} \frac{1}{n} (x-1)^n = (x-1) + O(x-1)^2. \tag{A.5.1}$$

Proposition A.5.1. *Let $a \in \mathbb{R}$, $\alpha > 0$ and $f(n) = O\left(\frac{1}{n^{1+\alpha}}\right)$. Then*

$$\lim_{n \to \infty} \left(1 + \frac{a}{n} + f(n)\right)^n = e^a.$$

Proof. By the Newton Binomial (also known as Binomial Theorem), we have that

$$\left(1 + \frac{a}{n} + f(n)\right)^n = \sum_{k=0}^{n} C_k^n \left(\frac{a}{n} + f(n)\right)^k 1^{n-k} = \sum_{k=0}^{n} C_k^n \left(\frac{a}{n} + f(n)\right)^k.$$

Let $0 \le k \le n$ be a natural number. Then

$$C_k^n \left(\frac{a}{n} + f(n)\right)^k = \frac{n!}{k!(n-k)!} \left(\frac{a}{n} + f(n)\right)^k.$$

Now notice that $\frac{n!}{(n-k)!} = n(n-1)...(n-k+1)$ is a polynomial on n of degree k, and we will denote it by p_k. Then

$$C_k^n \left(\frac{a}{n} + f(n)\right)^k = \frac{p_k(n)}{k!} \frac{1}{n^k} (a + nf(n))^k,$$

Now we are ready to compute $\lim_{n\to\infty} \left(1 + \frac{a}{n} + f(n)\right)^n$. Note that

$$\lim_{n\to\infty} \frac{p_k(n)}{n^k} = 1$$

since both polynomials have the same degree and have maximum order coefficient equal to one and

$$\lim_{n\to\infty} (a + nf(n))^k = a^k,$$

because $f(n) = O\left(\frac{1}{n^{1+\alpha}}\right)$ and by definition of order $\lim_{n\to\infty} f(n)n = 0$. Since by Taylor expansion $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$, we have

$$\begin{aligned}\lim_{n\to\infty} \left(1 + \frac{a}{n} + f(n)\right)^n &= \lim_{n\to\infty} \sum_{k=0}^{n} C_k^n \left(\frac{a}{n} + f(n)\right)^k \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} \frac{p_k(n)}{n^k} \frac{(a + nf(n))^k}{k!} = \sum_{k=0}^{\infty} \frac{a^k}{k!} = e^a.\end{aligned}$$

□

Corollary. *We have that*

$$\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$$

Proof. Notice $f(n) = 0$ satisfies the order constraint of the above theorem for every α. □

Bibliography

Capinski, M. and E. Kopp (2013). *Measure, Integral and Probability*. Springer Undergraduate Mathematics Series. Springer London. ISBN: 9781447136316. URL: `https://books.google.pt/books?id=AorbBwAAQBAJ`.

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 9781139491136. URL: `https://books.google.pt/books?id=evbGTPhuvSoC`.

Rudin, W. (1987). *Real and Complex Analysis*. Mathematics series. McGraw-Hill. ISBN: 9780071002769. URL: `https://books.google.pt/books?id=NmW7QgAACAAJ`.

Varadhan, S.R.S. (2001). *Probability Theory*. Courant Lecture Notes. Courant Institute of Mathematical Sciences. ISBN: 9780821828526. URL: `https://books.google.pt/books?id=jdGFAwAAQBAJ`.