SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

A Note on a Central Limit Theorem of Ibragimov-Linnik

av

Hedwig Nora Nordlinder

2026 - No K20

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

A Note on a Central Limit Theorem of Ibragimov-Linnik

Hedwig Nora Nordlinder

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Alan Sola

2026

Abstract

In this thesis we study the generalisation of the central limit theorem from the case of independent and identically distributed random variables to stationary stochastic processes where independence is replaced by **mixing**. An overview of basic probability theory is given, as well as definitions central to the field and the results themselves.

Sammanfattning på svenska

I denna uppsats studerar vi generaliseringen av centrala gränsvärdessatsen från oberoende och likafördelade slumpvariabler till stationära stokastiska processer där oberoendeantagandet ersätts med ett **blandningsantagande**. En översikt över grundläggande sannolikhetsteori ges, samt definitioner centrala till fältet och resultaten själva.

Contents

1 Acknowledgements

I would like to first and foremost thank my supervisor Alan Sola, whose attention to detail and generosity with his time made this thesis possible. This help was especially valuable when navigating the details of the central proof reproduced herein. This also means a thank you to Elis Rislund, who recommended Alan Sola as an advisor and initially encouraged me to write this thesis, is in order, as is a thank you to Wilma Eklund for being with me throughout the writing process. Finally, I would like to humbly acknowledge the help of large language models for tasks such as assistance with proof-reading, searching for references and implementation of the limited numerical simulations performed.

2 Introduction

This thesis aims to serve as an exposition to the generalisation of perhaps the most well-known result in probability theory (the central limit theorem) to sequences of non-i.i.d random variables. Alongside a brief historical exposition an introduction to measure theoretic probability will also be given, to aid in self-contained-ness.

2.1 Historical background

The most well-known form of the central limit theorem says, informally: A re-normalised sum of independent and identically distributed random variables will tend towards the normal distribution as the number of summands increases. Or more formally:

Theorem 2.1. *(The Central Limit Theorem). Let* $\{X_k\}_{k=1}^{\infty}$ *be a sequence of independent and identically distributed random variables with common mean* μ *and finite common variance* σ^2*. Define* $S_n := \frac{1}{n}\sum_{k=1}^{n} X_k$*. Then it holds that the centred, re-normalised sequence of random variables* $\sqrt{n}(S_n - \mu)$ *converges in distribution to* $\mathcal{N}(0, \sigma^2)$*.*

This modern form of the central limit theorem is due primarily to Finnish mathematician Jarl Waldemar Lindeberg and French mathematician Paul Pierre Lévy. However, the first CLT-like theorem is due to de Moivre and Laplace, who proved the following special case:

Theorem 2.2. *(de Moivre-Laplace theorem) Let* $p \in (0, 1)$ *and let* k *be close to* np*.*

$$\lim_{n\to\infty} \frac{\frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}}{\binom{n}{k} p^k (1-p)^{n-k}} = 1.$$

The numerator in Theorem 2.2 is the probability density function of a normal distribution with mean np and variance $np(1-p)$, and the denominator is the probability mass function of a binomial distribution $\text{Bin}(n, p)$. Such a distribution has mean np and variance $np(1-p)$. The theorem of de Moivre and Laplace may therefore be re-stated as ”For large n and k close to np, we may approximate the probability $\mathbb{P}[X = k]$ for $X \sim \text{Bin}(n, p)$ by the density of a normal distribution with the same mean and variance as X”. Recall that the binomial distribution is the distribution of the number of successes one observes when performing n independent trials, each with a probability p of succeeding. We may therefore re-state Theorem 2.2 in the following way

Corollary 2.3. *(re-statement of the de Moivre-Laplace theorem) Let* $\{X_k\}_{k=1}^{\infty}$ *be independent and identically distributed random variables taking the value* 1 *with probability* p *and the value* 0 *with probability* $1-p$*. Define* $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$*. Then it holds that the centred, re-normalised sequence of random variables* $\sqrt{n}(S_n - p)$ *converges in distribution to* $\mathcal{N}(0, p(1-p))$*.*

With this re-statement the correspondence between the modern CLT (Theorem 2.1) and the de Moivre-Laplace theorem is evident: de Moivre and Laplace simply proved it for the special case of Bernoulli random variables. It should be noted, however, that "proved" is perhaps too generous: the de Moivre-Laplace theorem pre-dates the rigorous measure-theoretic formalisation of probability theory due to Kolmogorov, and would by modern standards not be considered a valid proof of the theorem. For example, it is not immediately clear why convergence of the ratios of densities implies convergence in distribution, nor is it clear why the ratio of the probability *mass* function of a *discrete* random variable with the probability *density* function of a *continuous* random variable is even a meaningful quantity.

3 Background

In this section the necessary background in stochastic processes and measure theory will be given. The reader is assumed to be familiar with basic mathematical analysis in order to follow it, but other than that it should be self-contained. However, not much attention will be given to intuition and examples, making this a rather terse introduction for the completely uninitiated.

3.1 Gentle Introduction to Probability Theory

The rigorous, measure-theoretic foundations of probability theory are due to Kolmogorov. We will attempt to give a brief introduction to measure theory here, focusing chiefly on results most relevant to the content of this thesis.

3.1.1 Measure theory

The need for measure theory in probability theory arises from the fact that no contradiction-free definition of measuring the probability of an *event* (here understood as a subset of $[0, 1]$) satisfying the following conditions exists:

1. $\mathbb{P}[(a, b)] = b - a$ for $0 \leq a \leq b \leq 1$
2. $\mathbb{P} : 2^{[0,1]} \to [0, 1]$
3. For countable collections of pairwise disjoint sets $(A_k)_{k=1}^{\infty}$ belonging to $2^{[0,1]}$ it holds that $\mathbb{P}\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mathbb{P}[A_k]$

This can be thought of as the failure of the existence of a *uniform* probability distribution on the interval $[0, 1]$. One may (if one assumes the *Axiom of Choice*[1]) construct sets that violate the conditions listed no matter what numerical value one assigns to them, and one must therefore restrict oneself to only speak of *measurable* sets. Before formally defining measurable sets we will use our intuition about what constitutes an *event* in order to motivate the need for certain properties of the family of measurable sets. If A and B are events that we can assign a probability to we would likely also be interested in making statements about the probability that both occur, that either one occurs or that one of them does **not** occur, that is, if $\mathbb{P}$ is

[1]If one does not assume this, the truth value of the statement is dependent on the logical model of the real numbers one uses. This is far outside the scope of this thesis, however.

defined for A and B we would like $\mathbb{P}$ to also be defined for $A \cup B, A \cap B$, A^c and B^c. This (and a little more) is what is afforded to us by a σ-algebra:

Definition 3.1. *(σ-Algebra) Let Ω be some set. A collection $\mathcal{F}$ of subsets of Ω is said to be a σ-algebra if the following conditions hold*

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ *(Closure under complements)*
3. *Let $(A_k)_{k\in\mathbb{N}}$ be a countable collection with $A_k \in \mathcal{F}$ for each $k \in \mathbb{N}$. Then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ (Closure under countable unions)*

Example 3.2. *One easily checks that the set of **all** subsets of a set Ω is a σ-algebra on Ω. Furthermore, the set $\{\emptyset, \Omega\}$ is a σ-algebra. For a non-trivial example, let A be a proper non-empty subset of Ω. Then the set $\{\emptyset, A, A^c, \Omega\}$ is a σ-algebra.*

Before continuing we note an important property of σ-algebras: The property of being one is preserved under intersections. More formally

Theorem 3.3. *Let $\mathcal{F}_1, \mathcal{F}_2$ be two σ-algebras on the space Ω. Then the set $\mathcal{F} := \mathcal{A}_1 \cap \mathcal{A}_2$ is again a σ-algebra on Ω.*

Proof. We verify the properties. First, since $\mathcal{F}_1, \mathcal{F}_2$ are both σ-algebras they contain $\emptyset, \Omega$ so their intersection does as well. Now, consider an arbitrary set $A \in \mathcal{F}$. It holds that $A \in \mathcal{F} \implies A \in \mathcal{F}_1, \mathcal{F}_2$ and since $\mathcal{F}_1, \mathcal{F}_2$ are σ-algebras $A^c \in \mathcal{F}_1, \mathcal{F}_2$ so $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2 = \mathcal{F}$. Lastly, if (A_k) is a countable collection of sets in $\mathcal{F}$ its elements must belong to both $\mathcal{F}_1, \mathcal{F}_2$. Since these are σ-algebras therefore $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_1, \mathcal{F}_2$ so $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ as desired. $\square$

Remark 3.4. *An identical proof goes through for arbitrary intersections.*

This property is very important since it allows us to speak of the *smallest* (in the sense of set inclusion) σ-algebra with some prescribed property, since we may simply take the intersection of all σ-algebras with this property in order to obtain one that is a subset of all. This is of critical importance when defining the Borel sigma algebra, which is essential to probability theory. Using the notion of a σ-algebra we may now proceed by introducing measurable spaces, measures and measure spaces.

Definition 3.5. *(Measurable Space) Let X be some set and let $\mathcal{A}$ be a σ-Algebra over X. Then the pair $(X, \mathcal{A})$ is said to be a measurable space.*

Definition 3.6. *(Measure) Let* $(X, \mathcal{A})$ *be a measurable space and let* $\mu : \mathcal{A} \to [0, \infty]$ *be a set function satisfying the following two properties:*

1. $\mu(\emptyset) = 0$
2. *Let* $(A_k)_{k \in \mathbb{N}}$ *be a countable collection of pairwise disjoint elements in* $\mathcal{A}$*. Then*

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$

Example 3.7. *For a somewhat trivial example of a measure, let* $\Omega = \mathbb{R}, \mathcal{A} = 2^{\mathbb{R}}$ *and define for some* $x \in \mathbb{R}$ *the function* $\delta_x : 2^{\mathbb{R}} \to [0, \infty]$ *as*

$$\delta_x(A) := \begin{cases} 1, & x \in A \\ 0, & \textit{otherwise.} \end{cases}$$

Since $x \notin \emptyset$ *it holds that* $\delta_x(\emptyset) = 0$*. If we let* $(A_k)_{k \in \mathbb{N}}$ *be a countable collection of pairwise disjoint subsets of* $\mathbb{R}$ *it either holds that* $x \notin A_k$ *for all* k *or that there exists exactly one* k *such that* $x \in A_k$ *(the possibility of more than one such* k *existing is precluded by disjointness) and so we have countable additivity.*

Definition 3.8. *(Measure space) Let* X *be some set,* $\mathcal{A}$ *be a* σ*-algebra over* X *and let* μ *be a measure defined on* $(X, \mathcal{A})$*. Then the triple* $(X, \mathcal{A}, \mu)$ *is said to be a measure space.*

And we now may give our first definition from probability theory: the definition of the probability space.

Definition 3.9. *(Probability space) The measure space* $(\Omega, \mathcal{F}, \mathbb{P})$ *where* Ω *is some set,* $\mathcal{F}$ *is a* σ*-algebra over* Ω *and* $\mathbb{P}$ *is a measure is said to be a probability space if* $\mathbb{P}(\Omega) = 1$*.*

The preceding discussion motivating the need for σ-algebras and measure theory did not make it immediately clear *which* σ-algebra one ought to choose. It is evident that one must not take the power set as the σ-algebra, despite it being a perfectly valid one, since this would yield *too many* measurable sets. We however would also not like to take our σ-algebra to be $\{\emptyset, \Omega\}$ since this yields *too few* measurable sets. In general, the heuristic is that one choses the smallest σ-algebra containing "enough" measurable sets. Since we later will use the formalism introduced thus far to define an integral which operates on a larger class of functions than the Riemann integral (the *measurable* functions) we must chose our σ-algebra to respect this. Before introducing this choice, we first define what it means for a function to be *measurable.*

Definition 3.10. *(Measurable function) Let* $(X, \mathcal{A})$*,* $(X', \mathcal{A}')$ *be two measurable spaces. A function* $f : X \to X'$ *is said to be measurable if it holds for all* $A \in \mathcal{A}'$ *that* $f^{-1}(A) \in \mathcal{A}$

One sees that given two measurable spaces one obtains a class of measurable functions. If we wish to develop a theory of integration superior to the one of Riemann we therefore would like our class of measurable functions to be richer than the class of Riemann integrable functions. While there are discontinuous Riemann integrable functions, a known result from analysis is that each continuous real-valued function is Riemann integrable on compact intervals. We recall that a function $f : X \to Y$ is said to be continuous if it holds for each open $U \subset Y$ that $f^{-1}(U)$ is open in X. Therefore for the purpose of integrating real-valued functions we would like our σ-algebra to contain at the very least all open sets of $\mathbb{R}$, since this would mean every continuous function is also measurable. This motivates the following choice of σ-algebra:

Definition 3.11. *(Borel σ-algebra) Let* (X, τ) *be a topological space and denote by* $\mathcal{B}(X)$ *the smallest*[2] *σ-algebra $\mathcal{A}$ satisfying* $\tau \subset \mathcal{A}$*.* $\mathcal{B}(X)$ *is called the Borel σ-algebra*

The observant reader may at this point object: The class of Riemann-integrable functions includes discontinuous functions, so a priori guaranteeing only the measurability of continuous functions does not give that all Riemann integrable functions are also Lebesgue integrable. This is however the case, but the proof is outside the scope of this thesis. With all of the theory we have now built up we are able to define the fundamental object of study of probability theory: Random variables [3]

Definition 3.12. *(Random variable) Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. The function* $X : \Omega \to \mathbb{R}$ *is said to be (Borel) measurable if the following holds:* $A \in \mathcal{B}(\mathbb{R}) \implies X^{-1}(A) \in \mathcal{F}$ *where we equip* $\mathbb{R}$ *with the Euclidean topology. We call Borel measurable functions random variables.*

[2]Smallest in the sense of set inclusion

[3]Which, despite their name, are neither random nor variables

3.1.2 Lebesgue Integration Theory

In this section we will develop the aforementioned theory of Lebesgue integration, as it is essential to probability theory. As a sketch, this section begins by defining the integral for simple functions where linearity and other desirable properties are sufficient to determine what the integral of such simple functions must be, and then extends this definition to more complicated functions by taking limits. We begin by defining a finite partition.

Definition 3.13. *(Finite partition) Let X be a set. A collection of sets $(A_i)_{i=1}^n$ is said to be a finite partition of X if the following two conditions hold*

1. *$A_i \cap A_j = \emptyset$ for when $i \neq j$ ((A_i) are pairwise disjoint)*

2. *$\bigcup_{i=1}^n A_i = X$*

With this in mind, we are now ready to define a simple function, which is the building block of the Lebesgue integral.

Definition 3.14. *(Simple function) Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(A_k)_{k=1}^n$ be a measurable finite partition of X. A function $f : X \to \mathbb{R}$ is said to be simple if it can be written in the following form*

$$f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{x \in A_i}$$

with the restriction that if $\mu(A_i) = \infty$ then $a_i = 0$

It is quite natural to see what the integral of a simple function "ought" to be. We wish for our integral to be a linear functional, so therefore we require that

$$\int_X \left(\sum_{i=1}^{n} a_i \mathbf{1}_{x \in A_i} \right) d\mu(x) = \sum_{i=1}^{n} a_i \int_X \mathbf{1}_{x \in A_i} d\mu(x)$$

and from here we use the other "desirable" property that the integral of the function that is constantly equal to 1 on some set ought to simply return the area or volume of this set, that is we wish for $\int_X \mathbf{1}_{x \in A} d\mu(x) = \mu(A)$. These two properties now uniquely define the Lebesgue integral for simple functions:

Definition 3.15. *(Lebesgue integral for simple functions) Let f be a simple function on some measure space $(X, \mathcal{A}, \mu)$. We define the Lebesgue integral of f as*

$$\int_X f d\mu = \int_X \left(\sum_{i=1}^{n} a_i \mathbf{1}_{x \in A_i} \right) d\mu := \sum_{i=1}^{n} a_i \mu(A_i)$$

where we adopt the convention that $0 \cdot \infty = 0$.

In order to use these simple functions to extend the definition of the integral further we must define what it means to take a limit of measurable functions. We introduce our first notion of convergence:

Definition 3.16. *(Convergence almost everywhere). Let (f_n) be a sequence of measurable functions on $(X, \mathcal{A}, \mu)$. The sequence (f_n) is said to converge almost everywhere to the function f if*

$$\mu \left(\{x \in X : \lim_{n \to \infty} f_n(x) \neq f(x)\} \right) = 0.$$

If so, we write $f_n \stackrel{a.e}{\to} f$.

Definition 3.17. *(Cauchy in L^1) Let (f_n) be a sequence of simple functions on $(X, \mathcal{A}, \mu)$. (f_n) is said to be Cauchy in L^1 if for all $\epsilon > 0$ there exists an N such*

$$\int_X |f_n - f_m| \, d\mu < \epsilon$$

for all $n, m > N$.

Definition 3.18. *(Lebesgue integral) Let f be some function and suppose there exists a sequence of simple functions f_n such that*

1. $f_n \stackrel{a.e}{\to} f$
2. (f_n) *is Cauchy in* L^1

then the Lebesgue integral of f is defined as

$$\int_X f d\mu := \lim_{n \to \infty} \int_X f_n d\mu.$$

Remark 3.19. *It can be shown that this is well defined, that is, independent of the choice of approximating sequence f_n*

3.1.3 Modes of stochastic convergence

Definition 3.20. *(Almost sure convergence) Let (X_n) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the sequence X_n converges almost surely to the random variable X if the following holds*

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

If so, we write $X_n \stackrel{a.s}{\to} X$.

Definition 3.21. *(Convergence in L^p) Let $p > 0$ and let (X_n) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the sequence X_n converges in L^p to the random variable X if*

$$\lim_{n\to\infty} \int_\Omega |X_n - X|^p \, d\mathbb{P} = 0.$$

If so, we write $X_n \stackrel{L^p}{\to} X$

Definition 3.22. *(Convergence in probability) Let (X_n) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X_n converges in probability to the random variable X if the following holds*

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0$$

for all $\epsilon > 0$. If so, we write $X_n \stackrel{\mathbb{P}}{\to} X$.

Definition 3.23. *(Convergence in distribution) Let (X_n) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the sequence X_n converges in distribution[4] to X if the following holds*

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

for all $x \in \mathbb{R}$ where $F(x) := \mathbb{P}(X \leq x)$ is continuous. If so, we write $X_n \stackrel{d}{\to} X$

Theorem 3.24. *(Hierarchy of stochastic convergence) We have the following relationship between the modes of stochastic convergence*

1. $X_n \stackrel{a.s}{\to} X \implies X_n \stackrel{\mathbb{P}}{\to} X$
2. $X_n \stackrel{L^p}{\to} X \implies X_n \stackrel{\mathbb{P}}{\to} X$ *for all* $p > 0$
3. $X_n \stackrel{\mathbb{P}}{\to} X \implies X_n \stackrel{d}{\to} X$

[4] There are many equivalent characterisations of convergence in distribution, which are unified under the Portmanteau lemma

3.1.4 Other Results from Probability and Measure Theory

Definition 3.25. *(Independence) Let X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X and Y are independent if it holds for all $A, B \in \mathcal{B}(\mathbb{R})$ that*

$$\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) = \mathbb{P}(X^{-1}(A))\mathbb{P}(Y^{-1}(B))$$

Definition 3.26. *(Absolute continuity) Let μ, ν be two measures on the measurable space $(X, \mathcal{A})$. Then ν is said to be absolutely continuous with respect to μ if $\mu(A) = 0 \implies \nu(A) = 0$ for $A \in \mathcal{A}$. If so, we write $\nu \ll \mu$.*

Definition 3.27. *(σ-finiteness) A measure μ is said to be σ-finite if there exists a countable collection of measurable sets $\{A_i\}_{i=1}^{\infty}$ such that $\mu(A_i) < \infty$ for all i and $\bigcup_{i=1}^{\infty} A_i = X$.*

Remark 3.28. *σ-finiteness guarantees every measurable set is either a set of finite measure or a countable union of sets of finite measure. All probability measures are σ-finite since $\mathbb{P}[\Omega] = 1 < \infty$.*

Theorem 3.29. *(Radon-Nikodym) Let ν, μ be two σ-finite measures on the measurable space $(X, \mathcal{A})$ and suppose $\nu \ll \mu$. Then there exists a non-negative $\mathcal{A}$-measurable function $f : X \to \mathbb{R}$ such that*

$$\nu(A) = \int_A f d\mu$$

holds for all $A \in \mathcal{A}$. We call f the Radon-Nikodym derivative of ν with respect to μ and denote it by $\frac{d\nu}{d\mu}$. The function f is unique μ-almost everywhere.

Definition 3.30. *(Pushforward measure). Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and define a new measure $\mathbb{P}_X$ on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by*

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)).$$

We call $\mathbb{P}_X$ the pushforward of $\mathbb{P}$ under X

Definition 3.31. *(Normal distribution [5]) Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X is $\mathcal{N}(\mu, \sigma^2)$-distributed (read as normally distributed with mean μ and variance σ^2) if*

$$\frac{d\mathbb{P}_X}{d\lambda} \overset{a.e}{=} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

[5] Technically, this defines a non-degenerate Normal distribution.

where $\frac{d\mathbb{P}_X}{d\lambda}$ *denotes the Radon-Nikodym derivative of the pushforward* $\mathbb{P}_X$ *with respect to the Lebesgue measure on* $\mathbb{R}$.

Definition 3.32. *(Expectation) Let* X *be a random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$. *We write*

$$\mathbb{E}[X] := \int_\Omega X d\mathbb{P}$$

and call this operator the expectation of X *when it exists.*

Theorem 3.33. *(Slutsky) Let* X_n, Y_n *be sequences of random variables satisfying* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{\mathbb{P}} c$. *Then* $X_n + Y_n \xrightarrow{d} X + c$

Theorem 3.34. *(L^1 is stronger than probability) Let* $X_n \xrightarrow{L^1} X$. *Then* $X_n \xrightarrow{\mathbb{P}} X$

Proof. We wish to show that if

$$\lim_{n\to\infty} \int |X_n - X| \, d\mathbb{P} = 0$$

then it holds for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| \geq \epsilon] = 0.$$

Let $A_n = \{\omega \in \Omega : |X_n - X| \geq \epsilon\}$ and write

$$\int |X_n - X| \, d\mathbb{P} = \int_{A_n^c} |X_n - X| \, d\mathbb{P} + \int_{A_n} |X_n - X| \, d\mathbb{P} \geq \epsilon \mathbb{P}[A_n].$$

But since $X_n \xrightarrow{L^1} X$ it must hold that $\epsilon\mathbb{P}[A_n] \to 0$ by domination, so $X_n \xrightarrow{\mathbb{P}} X$. □

Corollary 3.35. *(L^2 is stronger than probability)*

Proof. Follows from the Cauchy-Schwartz inequality and Theorem 3.34 □

Both the standard proof of the central limit theorem as well as the central proof under α-mixing given herein makes use of *characteristic functions*, which we will now introduce and give some basic properties of:

Definition 3.36. *(Characteristic function). Let* X *be a random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$. *The characteristic function of* X *is*

$$\varphi_X(t) := \mathbb{E} \exp\{itX\}.$$

Characteristic functions always exist for all t and share a close connection to the fourier transform of the probability density function of X (should it exist). Perhaps most important property, which is heavily used in this thesis is the following:

Theorem 3.37. *(Lévy Continuity Theorem) Let $\{X_n\}_{n=1}^{\infty}$ be random variables and consider the corresponding characteristic functions $\{\varphi_n\}_{n=1}^{\infty}$. If $\varphi_n \overset{pointwise}{\to} \varphi$ for some $\varphi(t)$ continuous at zero, then $X_n \overset{d}{\to} X$ where X is a random variable with characteristic function φ.*

It is often easier to show convergence for characteristic functions than for example convergence for the cumulative distribution function at continuity points, making them a strong tool for proving asymptotic results in probability theory.

3.1.5 Results from analysis

Definition 3.38. *(Slowly varying function) The function $f : (0, \infty) \to \mathbb{R}$ is said to be slowly varying if it holds for any $\alpha > 0$ that*

$$\lim_{x\to\infty} \frac{f(\alpha x)}{f(x)} = 1.$$

Example 3.39. *Trivially, constant functions are slowly varying. Since*

$$\lim_{x\to\infty} \frac{\log \alpha x}{\log x} = \lim_{x\to\infty} \frac{\log x + \log \alpha}{\log x} = 1$$

the logarithm function is also slowly varying.

3.2 Ergodic Theory and Stochastic Processes

Definition 3.40. *(Stochastic process) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let T be some index set. A stochastic process is a collection $(X_t)_{t\in T}$ of random variables indexed by T.*

Remark 3.41. *The values in T are often interpreted as time. If T is at most countable one says under this interpretation that (X_t) is a discrete-time stochastic process, and otherwise one calls (X_t) a continuous-time stochastic process.*

Definition 3.42. *(Stationarity) A stochastic process (X_t) is said to be stationary if it holds for every finite tuple of time points $(t_1, t_2, \ldots, t_n)$ and every valid shift $h > 0$*

that

$$(X_{t_1}, \dots X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h}).$$

Here, we of course require some additive structure on our time set.

Often times one studies stochastic processes defined on a *filtered* probability space. Informally, a filtration is a mathematical representation of what information one has access to at a certain point in time. Formally, a filtration is defined in the following way:

Definition 3.43. *(Filtration) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let T be some ordered index set. The collection of sub σ-algebras of $\mathcal{F}$ $(\mathcal{F}_t)_{t\in T}$ is said to be a filtration if the following holds*

$$t \leq s \implies \mathcal{F}_t \subseteq \mathcal{F}_s.$$

Remark 3.44. *Intuitively, the sub σ-algbras become finer and finer as time progresses, that is they contain more and more measurable sets (thought of as "information" being accessible to us)*

Definition 3.45. *(Filtered probability space) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{F}_t)_{t\in T}$ be a filtration of $\mathcal{F}$. Then $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in T})$ is said to be a filtered probability space.*

We will now begin to build up to the definition of α-mixing for a stochastic process. To understand this definition we must first understand how stochastic processes give rise to filtrations, and what it means for a σ-algebra to be generated by a random variable

Definition 3.46. *(Generated σ-algebra) Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\sigma'(X)$ be some sub σ-algebra of $\mathcal{F}$ such that X is $\sigma'(X)$ measurable. We call $\sigma(X)$ the σ-algebra generated by X if it is the smallest σ-algebra in which X is measurable, that is $\sigma(X) \subseteq \sigma'(X)$ for all $\sigma'(X)$.*

Remark 3.47. *This definition extends mutatis mutandis to σ-algebras generated by multiple random variables $X_1, X_2, \dots$*

Now, we are ready to see how stochastic processes give rise to filtrations, and how one may speak of canonical filtrations for some stochastic process

Definition 3.48. *(Filtration generated by a stochastic process) Let (X_t) be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration generated by (X_t) is defined as*

$$\mathcal{F}_t = \sigma(\{X_s : s \leq t\}).$$

We will also introduce some notation for the sub σ-algebra generated by a "slice in time" $a, b \in T$ with $a \leq b$, that is we will let $\mathcal{F}_a^b$ be defined as $\mathcal{F}_a^b := \sigma(\{X_s : a \leq s \leq b\})$ (Note that here we allow $b = \infty$ and $a = -\infty$). We are now ready to define the α-mixing coefficients of a stationary stochastic process, which will be central to this entire thesis.

Definition 3.49. *(α-mixing coefficients) Let (X_t) be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in T})$ where $(\mathcal{F}_t)_{t\in\mathbb{R}}$ is the filtration generated by (X_t). Define*

$$\alpha(s) := \sup_{t\in T} \sup_{A\in\mathcal{F}_t, B\in\mathcal{F}_{t+s}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

We call $\alpha(s)$ the α-mixing coefficients of the stochastic process (X_t).

Definition 3.50. *(α-mixing) The stochastic process (X_t) is said to be α-mixing if*

$$\lim_{s\to\infty} \alpha(s) = 0.$$

Remark 3.51. *Intuitively, a process is α-mixing if the dependence of X_{t+s} on X_t decays to 0 as $s \to \infty$ even when considering the worst case in terms of t. Put succinctly, dependence decays to zero as the relative time between two points in the process increases, no matter the absolute time of these points.*

Definition 3.52. *(AR(1)-process). Let $(\epsilon_i)_{i=1}^n$ be independent and identically distributed with $\operatorname{Var} \epsilon_i \equiv \sigma_\epsilon < \infty$ and $\mathbb{E}\epsilon_i \equiv 0$. We call this process the innovation process. Define $X_0 \sim \mathcal{D}$ for some distribution $\mathcal{D}$ with finite second moment and let*

$$X_i = \phi X_{i-1} + \epsilon_i$$

for $i \geq 1$. We call $(X_i)_{i=0}^{\infty}$ an AR(1)-process. When $|\phi| < 1$ the process has a stationary distribution.

4 Standard Limit Theorems

Theorem 4.1. *(Strong law of large numbers) Let X_k be a sequence of independent and identically distributed random variables with expectation $\mathbb{E}[X_k] \equiv \mu$. Let $S_n := \frac{1}{n}\sum_{k=1}^{n} X_k$. Then it holds that $S_n \overset{a.s}{\to} \mu$*

Theorem 4.2. *(Central Limit Theorem) Let X_k be a sequence of independent and identically distributed random variables with $\mathbb{E}[X_k] \equiv \mu$ and $\mathrm{Var}(X_k) \equiv \sigma^2$ and define $S_n = \frac{1}{\sqrt{n\sigma^2}}\sum_{k=1}^{n}(X_k - \mu)$. Then it holds that $S_n \overset{d}{\to} X$ where $X \sim \mathcal{N}(0,1)$*

The central purpose of the first part of this thesis will be to study under what conditions one may weaken the assumption that X_k are independent and identically distributed. One such condition is α-mixing, which we will now state

Theorem 4.3. *("Weak" Central Limit Theorem under α-mixing) [3]Let (X_n) be a stationary discrete-time stochastic process and denote by $\alpha(s)$ its α-mixing coefficients. Suppose that $\mathbb{E}[X_n] = 0$, $\mathbb{E}[X_n^{12}] < \infty$ and $\alpha(s) = \mathcal{O}(s^{-5})$. Define*

$$S_n := \sum_{k=1}^{n} X_k.$$

Then the limit

$$\sigma^2 := \lim_{n\to\infty} \frac{\mathbb{E}[S_n^2]}{n}$$

exists and

$$\frac{1}{\sqrt{n\sigma^2}} S_n \overset{d}{\to} S$$

where $S \sim \mathcal{N}(0,1)$.

Example 4.4. *As an example of a stochastic process satisfying the requirements of Theorem 4.3, consider a process defined as follows*

$$X_1 \sim \mathrm{Rademacher}\left(\frac{1}{2}\right)$$
$$X_n \sim \mathrm{Rademacher}\left(\frac{1}{2} + \frac{X_{n-1}}{4}\right),\ n \geq 2.$$

$\mathbb{E}X_n^{12} < \infty$ is immediate since each X_n is bounded. We now proceed to compute the α-mixing coefficients. We first note that this is a Markov chain on $\{-1, 1\}$ with stationary distribution $\mathrm{Rademacher}(\frac{1}{2})$*, so the process as defined is stationary. Thus,*

we may compute the α-mixing coefficients as

$$\alpha(k) = \sup_{A \in \mathcal{F}_1, B \in \mathcal{F}_{1+k}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

(that is, taking the supremum over the absolute time is redundant). By symmetry we may fix the event $A = \{1\}$ and thus we may further simplify the α-mixing coefficients to

$$\alpha(k) = \mathbb{P}(X_1 = 1) \sup_{R \in \{-1,1\}} |\mathbb{P}(X_k = R | X_1 = 1) - \mathbb{P}(X_k = R)| \, .$$

By stationarity we have $\mathbb{P}(X_k = R) = \frac{1}{2}$ and so the α-mixing coefficients are

$$\alpha(k) = \frac{1}{2} \sup_{R \in \{-1,1\}} \left| \mathbb{P}(X_k = R | X_1 = 1) - \frac{1}{2} \right| .$$

Intuitively, since $X_1 = 1$ increases the probability of subsequent 1s and decreases the probability of subsequent -1s it will hold that $\mathbb{P}(X_k = 1 | X_1 = 1) > \mathbb{P}(X_k = -1 | X_1 = 1)$ and so we have

$$\alpha(k) = \frac{1}{2} \sup_{R \in \{-1,1\}} \left| \mathbb{P}(X_k = 1 | X_1 = 1) - \frac{1}{2} \right| .$$

In order to compute $\mathbb{P}(X_k = 1 | X_1 = 1$ we use the fact that the process is a Markov chain on $\{-1, 1\}$ with transition matrix given by $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$. The k-step distribution given $X_1 = 1$ is therefore

$$\begin{pmatrix} \mathbb{P}(X_k = 1 | X_1 = 1)) \\ \mathbb{P}(X_k = -1 | X_1 = 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}^{k-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} .$$

Diagonalization yields

$$\begin{pmatrix} \mathbb{P}(X_k = 1 | X_1 = 1)) \\ \mathbb{P}(X_k = -1 | X_1 = 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}^{k-1} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} .$$

After straightforward but tedious computations we obtain

$$\mathbb{P}(X_k = 1 | X_1 = 1) = \frac{1}{2} + 2^{-k}$$

and so the α-mixing coefficients are

$$\frac{1}{2}\left|\frac{1}{2} + 2^{-k} - \frac{1}{2}\right| = 2^{-(k+1)}$$

so the process mixes at a geometric rate, faster than $\mathcal{O}(k^{-5})$ which is prescribed by Theorem 4.4. Thus, we have a central limit theorem for the process, which is graphically illustrated in figures 1, 2 and 3 which compare the convergence to the theoretical limiting distribution for the Rademacher process described with an i.i.d sequence of Rademacher($\frac{1}{2}$).

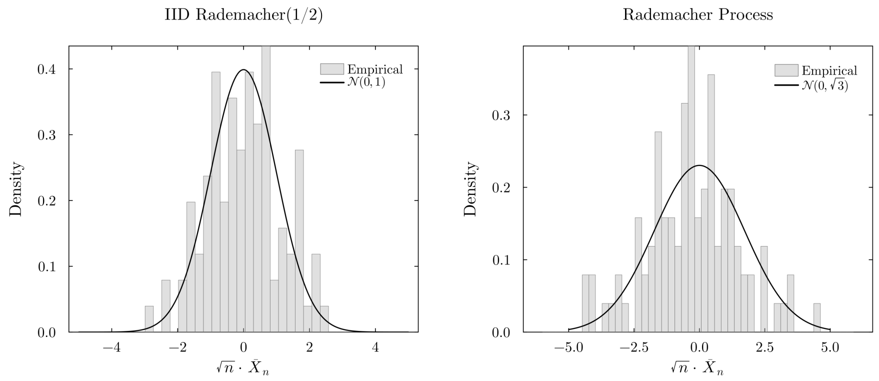

Figure 1: i.i.d vs Rademacher process, $n = 100$

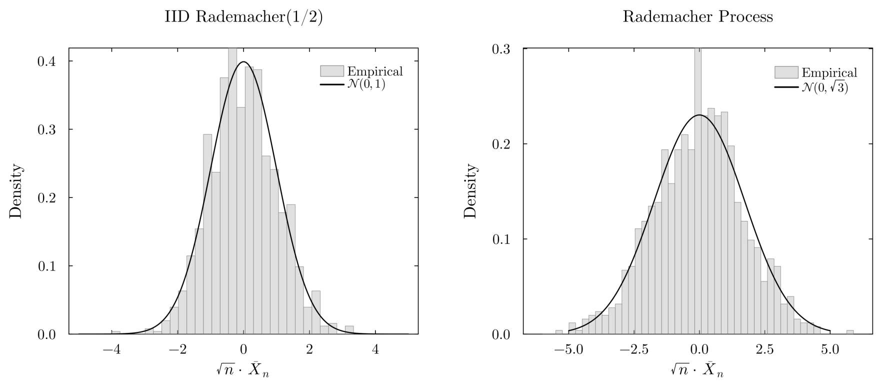

Figure 2: i.i.d vs Rademacher process, $n = 1000$

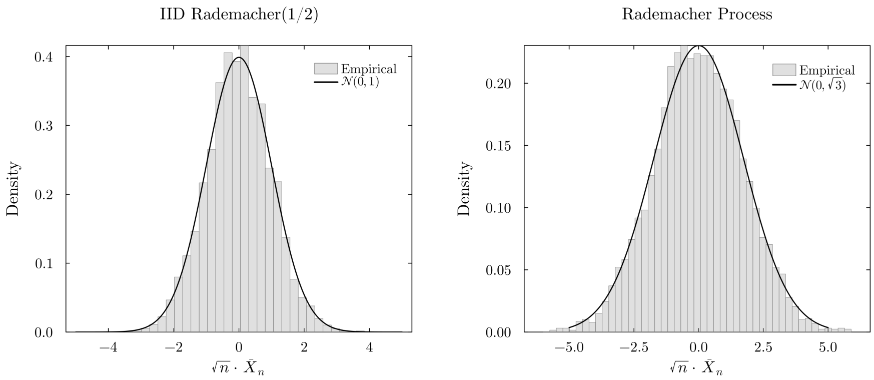

Figure 3: i.i.d vs Rademacher process, $n = 5000$

The use of the transition matrix in the example given above can be viewed as a primitive *spectral argument* for α-mixing. Arguing from the spectral properties of the transition operator is a standard route in the field. The aforementioned example is rather simple and we could have explicitly computed $\mathbb{P}(X_k = 1|X_1 = 1)$ by a combinatorial argument by marginalising over all paths from $X_1 = 1$ to $X_k = 1$, but it is in general not possible to identify the events A, B which constitue the supremum (should they even exist)

Theorem 4.5. *("Strong" Central Limit Theorem under α-mixing, [3]) Let (X_n) be a stationary discrete-time stochastic process and denote by $\alpha(s)$ its α-mixing coefficients. Suppose that $\mathbb{E}[X_n] = 0$ and that there exists $\delta > 0$ such that $\mathbb{E}\left[|X_n|^{2+\delta}\right] < \infty$ and*

$$\sum_{k=1}^{\infty} \alpha(k)^{\frac{\delta}{2(2+\delta)}} < \infty.$$

Then the conclusion of Theorem 4.3 still holds.

Proposition 4.6. *The conditions of Theorem 4.3 are stronger than the conditions of Theorem 4.5.*

Proof. Suppose the conditions of Theorem 4.3 hold and let $\delta = 10$. Then $\mathbb{E}\left[|X_n|^{2+\delta}\right] < \infty$ is immediate from the assumption $\mathbb{E}[X_n^{12}] < \infty$. We will also have for large k $\alpha(k) \le Ck^{-5}$ and hence $\alpha(k)^{\frac{5}{12}} \le C'k^{-\frac{25}{12}}$ so we will have

$$0 \le \sum_{k=1}^{\infty} \alpha(k)^{\frac{\delta}{2(2+\delta)}} \le C'' \sum_{k=1}^{\infty} k^{-2} = \frac{\pi^2}{6} < \infty.$$

□

Theorem 4.7. *The conditions of Theorem 4.5 are sufficient, but not necessary.*

Before giving the proof of Theorem 4.7 we introduce a random variable that has finite second moment but infinite $2+\delta$ moment for all $\delta > 0$. The construction is due to an answer on Math Stack Exchange [4]. For completeness we re-state the exact construction [6] and show that it indeed has the desired properties

Lemma 4.8. *(Existence of a random variable with infinite $2+\delta$-moments but finite second moment, [4]) Let X be a discrete random variable defined by*

$$\mathbb{P}(X = n) = \begin{cases} 0, & n = 0, 1 \\ \frac{c}{n^3 \log^2 n}, & n \geq 2 \end{cases}$$

where $c = \left(\sum_{n=2}^{\infty} \frac{1}{n^3 \log^2 n}\right)^{-1}$. Then $\mathbb{E}X^2 < \infty$ but $\mathbb{E}\,|X|^{2+\delta} = \infty$ for all $\delta > 0$.

Proof. We first recognise the random variable is well-defined since $\sum_{n=2}^{\infty} \frac{1}{n^3 \log^2 n} \leq \frac{1}{8 \log^2 2} + \sum_{n=3}^{\infty} \frac{1}{n^3}$ which is a p-series and thus convergent, so the normalising constant c is finite. Furthermore

$$\mathbb{E}X^2 = c\sum_{n=2}^{\infty} n^2 \frac{1}{n^3 \log^2 n} = c\sum_{n=2}^{\infty} \frac{1}{n \log^2 n} \leq \frac{c}{2\log^2 2} + c\int_2^{\infty} \frac{dx}{x \log^2 x} = \frac{c}{2\log^2 2} + \frac{c}{\log 2} < \infty$$

so the second (and therefore also first) moment of X is finite. Now, consider $\mathbb{E}\,|X|^{2+\delta}$ for some $\delta \in (0, 1)$. We first show that

$$f(x) = \frac{x^{\delta}}{x \log^2 x}$$

is monotonically decreasing for $x \geq 2$ and $\delta \in (0, 1)$:

$$\frac{f'(x)}{f(x)} = \frac{d}{dx} \log f(x) = \frac{d}{dx}\left(\delta \log x - \log(x) - \log\log^2 x\right) = \frac{\delta}{x} - \frac{1}{x} - \frac{2}{x \log x}$$

which is less than zero for $\delta \in (0, 1)$ so $\frac{f'(x)}{f(x)} < 0$ which in particular means $f'(x) < 0$ since $f(x) > 0$ for all $x > 0$. This allows us to use the following integral comparison

$$\mathbb{E}\,|X|^{2+\delta} = c\sum_{n=2}^{\infty} \frac{1}{n^{1-\delta} \log^2 n} \geq c\int_2^{\infty} \frac{x^{\delta} dx}{x \log^2 x} \underbrace{=}_{u=\log x} c\int_{\log 2}^{\infty} \frac{e^{u\delta}}{u^2} du = \infty$$

[6] We also correct a very minor indexing error.

showing that no p-moments for $p \in (2,3)$ exist. Finally, the extension to $q \in [3, \infty)$ follows from that $\mathbb{E}\,|X|^q > \mathbb{E}\,|X|^p$ for $p \in (2,3)$ since $X \geq 2$ a.s. $\square$

The proof of Theorem 4.7 now follows from taking an i.i.d sequence of random variables as defined in 4.8 (and centering), since i.i.d sequences have all α-mixing coefficients equal to zero. For less trivial examples, one may construct stationary $AR(1)$-processes with such an i.i.d sequence as its innovation process and appeal to the Markov Chain Central Limit Theorem.

5 A Detailed Discussion on a Theorem of Ibragimov-Linnik

Before proving the main theorem of interest of this thesis (the Ibragimov-Linnik Central Limit Theorem) we state a result of Lindeberg:

Theorem 5.1. *(Lindeberg condition) Let $(X_k)_{k=1}^{\infty}$ be a sequence of independent but not necessarily identically distributed random variables satisfying $\mathbb{E}X_k \equiv 0$ for all k. Define $\sigma_n^2 := \sum_{k=1}^n \operatorname{Var}(X_k)$. If the sequence satisfies the Lindeberg condition*

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{k=1}^n \int_{|z|>\epsilon\sigma_n} z^2 dF_k = 0$$

for all $\epsilon > 0$ then the re-normalised sequence $\frac{1}{\sigma_n}\sum_{k=1}^n X_k$ converges in distribution to the standard normal distribution $\mathcal{N}(0,1)$.

Remark 5.2. *The Lindeberg condition is sufficient, but not necessary. It is however necessary if the worst-case contribution to the total variance of any single X_k is of vanishing importance, that is if*

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{\operatorname{Var}(X_k)}{\sigma_n^2} = 0.$$

Finally, we need a result due to Potter in order to repair what is believed to be a minor error in the proof due to Ibragimov and Linnik given in [2]

Theorem 5.3. *(Potter bound, [1]) Let h be slowly varying and let $\delta > 0, A > 1$ be arbitrary. Then there exists an N such that for all $n_1, n_2 > N$ it holds that*

$$\frac{h(n_2)}{h(n_1)} \le A \max\left\{\left(\frac{n_2}{n_1}\right)^{\delta}, \left(\frac{n_1}{n_2}\right)^{\delta}\right\}$$

Theorem 5.4. *(Sufficient conditions for a central limit theorem, [2]) Let $(X_k)_{k=1}^{\infty}$ be a strongly mixing stationary stochastic process with $\mathbb{E}X_1 = 0$ and define $\sigma_n := \mathbb{E}S_n^2$. The sequence satisfies a central limit theorem if it holds that $\sigma_n^2 = nh(n)$ for some slowly varying function h and that there exists a pair of sequences $p_n, q_n \to \infty$ that satisfy*

1. *$q_n = o(p_n)$, $p_n = o(n)$*
2. *$\lim_{n\to\infty} n^{1-\beta} q_n^{1+\beta} p_n^{-2} = 0$ holds for all $\beta > 0$*

3. $\lim_{n\to\infty} np^{-1}\alpha(q_n) = 0$ *where* $\alpha(n)$ *are the alpha-mixing coefficients of the process*

We also require a Lindeberg-type condition in the sense that we require that

$$\bar{F}_l(z) := \mathrm{P}[X_1 + \cdots + X_l < z]$$

satisfies

$$\lim_{n\to\infty} \frac{n}{p_n \sigma_n^2} \int_{|z|>\epsilon\sigma_n} z^2 d\bar{F}_{p_n}(z) = 0$$

for all $\epsilon > 0$*.*

Proof. (Deep breath) Let $S_n := \sum_{k=1}^n X_k$. We will now define two auxiliary sequences of random variables ζ_i, η_i as follows (where we define $k := [\frac{n}{p_n+q_n}]$

$$\zeta_i := \sum_{ip_n+iq_n+1}^{(i+1)p_n+iq_n} X_j$$

$$\eta_i := \begin{cases} \sum_{(i+1)p_n+iq_n+1}^{(i+1)p_n+(i+1)q_n} X_j & 0 \le i \le k-1, \\ \sum_{kp_n+kq_n+1}^{n} X_j & i = k. \end{cases}$$

We then decompose S_n as

$$S_n = \underbrace{\sum_{i=0}^{k-1} \zeta_i}_{S_n'} + \underbrace{\sum_{i=0}^{k} \eta_i}_{S_n''} .$$

Remark: We note that the number of terms in the sum defining S_n'' is kq_n and the number of terms in the sum defining S_n' is $p_n(k-1)$. We have $\frac{kq_n}{p_n(k-1)} \sim \frac{q_n}{p_n} \to 0$ since q_n is of order $o(p_n)$. This means that as $n \to \infty$ a larger and larger share of the sequence X_k will "belong" to the term S_n', and the share of terms in the sequence S_n'' will vanish. The main idea of the proof technique is to show that S_n'' is of vanishing importance to the sum S_n. More specifically, what we wish to show is that the renormalised sum $S_n''/\sigma_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$, since we will then by Slutskys theorem (3.33) have that $\frac{1}{\sigma_n} S_n$ has the same distributional limit as $\frac{1}{\sigma_n} S_n'$, and thus it will be sufficient to show a central limit theorem for S_n', which is an easier task than showing it directly for S_n since we have "thinned" the sequence, making the terms of S_n' "nearly independent".

We now continue the proof. We introduce

$$p_n = \max\left\{\frac{n\alpha(n^{1/4})}{\lambda_n}, \frac{n^{3/4}}{\lambda_n}\right\}$$
$$q_n = n^{1/4}$$

where
$$\lambda_n = \max\left\{\alpha(n^{1/4})^{1/3}, \frac{1}{\log n}\right\}.$$

We will now show that $\frac{1}{\sigma_n}S''_n \xrightarrow{L^2} 0$. Note that

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{\sigma_n}S''_n\right|^2\right] &= \frac{1}{\sigma_n^2}\mathbb{E}\left[\left|\sum_{i=0}^{k}\eta_i\right|^2\right] = \frac{1}{\sigma_n^2}\mathbb{E}\left[\left|\sum_{i=0}^{k-1}\eta_i + \eta_k\right|^2\right] \\
&\leq \frac{1}{\sigma_n^2}\left(\mathbb{E}\left[\left|\sum_{i=0}^{k-1}\eta_i\right|^2\right] + 2\mathbb{E}\left[\eta_k\sum_{i=0}^{k-1}\eta_i\right] + \mathbb{E}[\eta_k^2]\right) \\
&\leq \frac{1}{\sigma_n^2}\left(k^2\sigma_{q_n}^2 + 2k\sqrt{\sigma^2_{n-k(p_n+q_n)}\sigma^2_{q_n}} + \sigma^2_{n-k(p_n+q_n)}\right)
\end{aligned}$$

Before continuing the author feels some motivation for the final step is necessary, since the proof is very convoluted. Recall that $\sigma_m^2 := \operatorname{Var}\sum_{k=n}^{n+m} X_k$ where this definition does not depend on n due to stationarity (that is, joint distributions are shift-invariant). The quantity η_i consists for each i of exactly q_n terms, hence when bounding the η_i sum via the Cauchy-Schwartz inequality we use the fact that we have k^2 terms all bounded by the common variance $\sigma^2_{q_n}$. Similarly, η_k consists of $n - k(p_n + q_n)$ terms and thus has variance $\sigma^2_{n-k(p_n+q_n)}$. We will now use the assumption that the variance is of the form $nh(n)$ where h is slowly varying. Before proceeding, for compactness we write $n - k(q_n + p_n) = r_n$

$$\mathbb{E}\left[\left|\frac{1}{\sigma_n}S''_n\right|^2\right] \underbrace{\leq}_{\text{By previous computations}} \frac{k^2 q_n h(q_n) + 2k\sqrt{q_n r_n h(q_n)h(r_n)} + r_n h(r_n)}{nh(n)}.$$

We first note that r_n is of order $p_n + q_n$. This can be realised by writing

$$r_n = n-(p_n+q_n)\left[\frac{n}{p_n+q_n}\right] = n-(p_n+q_n)\left(\frac{n}{p_n+q_n} - \left\{\frac{n}{p_n+q_n}\right\}\right) = (p_n+q_n)\left\{\frac{n}{p_n+q_n}\right\}$$

where $\{\cdot\}$ denotes the *fractional part*, which always belongs to $[0,1)$ so $r_n = \mathcal{O}(p_n +$

q_n). In order to continue, we also require knowledge about the following property of slowly varying functions:

Lemma 5.5. *(Lemma 18.2.4 in [2]) For sufficiently small c and sufficiently large n it holds for a slowly varying function h that*

$$\frac{h(cn)}{h(n)} \leq c^{-\frac{1}{2}}.$$

We omit the proof. We will now show that the following terms

$$\frac{k^2 q_n h(q_n)}{nh(n)} \tag{A_n}$$

$$\frac{2k\sqrt{q_n r_n h(q_n) h(r_n)}}{nh(n)} \tag{B_n}$$

$$\frac{r_n h(r_n)}{nh(n)} \tag{C_n}$$

all converge to zero. Noting first that $B_n = 2\sqrt{A_n C_n}$ we realise we must only show the result for A_n, C_n. For C_n we may apply Lemma 5.5 in the following way: Let $c = \frac{r_n}{n}$. Then $C_n = \frac{r_n}{n}\frac{h(cn)}{h(n)}$ and since $r_n = \mathcal{O}(p_n + q_n) = o(n)$ we may make c arbitrarily small for sufficiently large n so we may bound

$$C_n \leq \frac{r_n}{n}\sqrt{\frac{n}{r_n}} = \sqrt{\frac{r_n}{n}} \to 0 \text{ as } n \to \infty$$

and so we conclude [7] $C_n \to 0$. For the case of A_n we will first note the following:

$$\frac{k^2}{n} \sim \frac{1}{n}\frac{n^2}{p_n^2 + q_n^2} \sim \frac{n}{p_n^2}$$

which follows from the definition of k as well has the fact that $q_n = o(p_n)$. If we are being a bit more careful we write

$$\frac{k^2}{n} < K\frac{1}{n}\frac{n^2}{(p_n + q_n)^2} \leq K\frac{\frac{n}{p_n^2}}{1 + \frac{2q_n}{p_n} + \frac{q_n^2}{p_n^2}}$$

for some constant K and we realise that it holds by virtue of $q_n = o(p_n)$ that for

[7] It should be noted that it is the opinion of the author that the employed lemma is miss-stated in [2], however the conditions required for the believed corrected statement are satisfied. If the reader is not convinced, they may instead argue via the Potter bound.

any $\epsilon > 0$ it holds for sufficiently large n that $\frac{q_n}{p_n} < \epsilon$ so in fact

$$\frac{k^2}{n} \leq \frac{\frac{n}{p_n^2}}{1 + 2\epsilon + \epsilon^2} < \frac{n}{p_n^2}.$$

Substituting this into the definition of A_n yields that we wish to show

$$\frac{nq_n h(q_n)}{p_n^2 h(n)} \to 0.$$

We first apply Lemma 5.5 and obtain that we wish to show

$$\frac{nq_n}{p_n^2}\sqrt{\frac{n}{q_n}} \to 0 \iff \frac{n^{1+\frac{1}{2}} q_n^{\frac{1}{2}}}{p_n^2} \to 0.$$

We would now like some condition requiring that this is true. After toiling with this problem for weeks, the conclusion of the author is that this can not hold (despite this being heavily implied by the proof sketch in [2]). Consider the i.i.d case, in which all α-mixing coefficients are zero and thus the terms p_n, q_n are reducible to [8]

$$p_n = n^{3/4} \log n$$
$$q_n = n^{1/4}.$$

In this case we have

$$\frac{n^{1+\beta} n^{\frac{1-\beta}{4}}}{n^{3/2}(\log n)^2} = \frac{n^{\frac{5+3\beta}{4}}}{n^{3/2}(\log n)^2}.$$

which diverges for $\beta > \frac{1}{3}$. As a substitute for the believed to be erroneous condition (ii) we introduce the following:

Assumption 1. *It holds for some $\beta > 0$ that*

$$\frac{n^{1+\beta} q_n^{1-\beta}}{p_n^2} \to 0 \text{ as } n \to \infty.$$

Now apply the Potter bound (Theorem 5.3) to $\frac{nq_n h(q_n)}{p_n^2 h(n)}$:

$$\frac{nq_n h(q_n)}{p_n^2 h(n)} \leq A \frac{nq_n}{p_n^2} \frac{n^\delta}{q_n^\delta} = A \frac{n^{1+\delta} q_n^{1-\delta}}{p_n^2}$$

[8]Technically p_n, q_n may be chosen arbitrarily subject to the imposed constraints, but this is an explicit construction given in [2].

and by Assumption 1 there exists some $\delta > 0$ such that the expression in Equation 5 goes to zero, and we conclude that also A_n goes to zero. We have now shown that $\frac{1}{\sigma_n} S''_n \xrightarrow{\mathbb{P}} 0$ so what remains to be proven is the simpler central limit theorem for S'_n. To do this, we will show that the characteristic function of $\frac{1}{\sigma_n} S'_n$ converges pointwise to $\phi_n(t)^k$ where $\phi_n(t)$ is the characteristic function of $\frac{1}{\sigma_n}\zeta_0$. We begin by noting that the random variable

$$\exp\left\{\frac{it}{\sigma_n}\sum_{m=0}^{k-2}\zeta_m\right\}$$

is $\mathcal{F}_{-\infty}^{(k-1)p+(k-2)q}$-measurable since the largest index occurring in any ζ_m will be $(k-1)p_n + (k-2)q_n$ by definition. Furthermore it will hold that $\exp\{\frac{it}{\sigma_n}\zeta_{k-1}\}$ is $\mathcal{F}^{\infty}_{(k-1)p_n+(k-1)q_n+1}$-measurable. These conditions are required for the following statement:

Theorem 5.6. *(Theorem 17.2.1 in [2]) Let X, Y be almost surely bounded (by scalars C_1, C_2 respectively) random variables such that X is $\mathcal{F}^t_{-\infty}$-measurable and Y is $\mathcal{F}^{\infty}_{t+\tau}$-measurable where $\mathcal{F}^n_m$ is generated by some strongly mixing stationary process. Then it holds that*

$$|\mathrm{Cov}(X, Y)| \leq 4C_1C_2\alpha(\tau)$$

where $\alpha(\tau)$ is the strong mixing coefficient

We now finally use our strong mixing assumption in order to apply Theorem 5.6 (by taking $X = \exp\left\{\frac{it}{\sigma_n}\sum_{j=0}^{k-2}\zeta_j\right\}$ and $Y = \exp\{\frac{it}{\sigma_n}\zeta_{k-1}\}$) in order to obtain the bound [9]

$$\left|\mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\sum_{j=0}^{k-1}\zeta_j\right\}\right] - \mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\sum_{j=0}^{k-2}\zeta_j\right\}\right]\mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\zeta_{k-1}\right\}\right]\right| \leq K\alpha(q_n)$$

where K is some constant. It also holds for similar reasons that

$$\left|\mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\sum_{j=0}^{\ell}\zeta_j\right\}\right] - \mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\sum_{j=0}^{\ell-1}\zeta_j\right\}\right]\mathbb{E}\left[\exp\left\{\frac{it}{\sigma_n}\zeta_{\ell}\right\}\right]\right| \leq K\alpha(q_n)$$

for $\ell \leq k-2$. This allows us to proceed via a telescoping product argument in the following sense: We write $\Psi_{n,m}(t) := \mathbb{E}\exp\{\frac{it}{\sigma_n}\sum_{j=0}^{m-1}\zeta_j\}$ and $\varphi_n(t) := \mathbb{E}\exp\{\frac{it}{\sigma_n}\zeta_0\}$. By stationarity $\varphi_n(t)$ is also the characteristic function for each ζ_ℓ. We now apply

[9]Using of course that characteristic functions are bounded

the age-old analysis trick of adding and subtracting zero:

$$\underbrace{|\Psi_{n,m}(t) - \varphi_n(t)^m|}_{\heartsuit} = |\Psi_{n,m}(t) - \Psi_{n,m-1}(t)\varphi_n(t) + \Psi_{n,m-1}(t)\varphi_n(t) - \varphi_n(t)^m|$$

$$\leq |\Psi_{n,m}(t) - \Psi_{n,m-1}(t)\varphi_n(t)| + |\Psi_{n,m-1}(t)\varphi_n(t) - \varphi_n(t)^m|$$

$$= |\Psi_{n,m}(t) - \Psi_{n,m-1}(t)\varphi_n(t)| + |\varphi_n(t)| \left|\Psi_{n,m-1}(t) - \varphi_n(t)^{m-1}\right|.$$

The first term may by our previous argument be bounded above by $K\alpha(q_n)$, giving that

$$\heartsuit \leq K\alpha(q_n) + |\varphi_n(t)| \left|\Psi_{n,m-1}(t) - \varphi_n(t)^{m-1}\right|.$$

We then recursively apply an identical argument to the second term (and use that the absolute value of a characteristic function is bounded above by 1 uniformly in t) in order to conclude that

$$\left|\mathbb{E}[\exp\{\frac{it}{\sigma_n}S'_n\}] - \phi_n(t)^k\right| \leq Kk\alpha(q_n).$$

By assumption (iii) this tends to zero as $n \to \infty$, showing that the characteristic function of Z'_n converges pointwise to the characteristic function of a sum of k independent copies of $\frac{1}{\sigma_n}\zeta_0$. This result is precisely why we split $S_n = S'_n + S''_n$: The split allows us to collect q_n terms into S''_n as "padding" between the terms constituting S'_n. This padding then allows us to apply the covariance-mixing bound of Theorem 5.6 to make the difference in the characteristic functions of sums of independent copies of ζ_0 and the characteristic function of S'_n small. We now expand upon this argument: let $(\zeta^\star_{n,j})_{j=0}^{k-1}$ be an independent sequence satisfying $\zeta^\star_{n,j} \overset{d}{=} \frac{1}{\sigma_n}\zeta_0$. Now, let $T_n := \sum_{j=0}^{k-1} \zeta^\star_{n,j}$. By our previous argument the asymptotic distribution of T_n is the same as the asymptotic distribution of Z'_n, and so what remains to be shown is convergence to a normal distribution for T_n. Similar to the Lindeberg condition[10] of Theorem 5.1 it suffices to show that

$$\lim_{n\to\infty} \sum_{j=1}^{k} \int_{|z|>\epsilon} z^2 d\mathbb{P}[\zeta^\star_{n,j} < z] = 0.$$

[10] We say similar instead of exactly since we require a Lindeberg-condition for triangular arrays, but we omit this technical detail in the interest of the flow of the argument

First, note that $d\mathbb{P}[\zeta^{\star}_{n,j}] \equiv d\mathbb{P}[\frac{1}{\sigma_n}\zeta_0]$ for all j so our condition becomes

$$\lim_{n\to\infty} k \int_{|z|>\epsilon} z^2 d\mathbb{P}\left[\frac{1}{\sigma_n}\zeta_0 < z\right] = 0.$$

By change of variables we obtain that we wish to show

$$\lim_{n\to\infty} \frac{k}{\sigma_n^2} \int_{|z|>\sigma_n\epsilon} z^2 d\mathbb{P}\left[\zeta_0 < z\right] = 0.$$

Recall that $k := \left[\frac{n}{p_n+q_n}\right]$ so therefore $k \sim \frac{n}{p_n}$. Combining this with the fact ζ_0 is a sum of p_n consecutive elements in the original sequence we have

$$\lim_{n\to\infty} \frac{k}{\sigma_n^2} \int_{|z|>\sigma_n\epsilon} z^2 d\mathbb{P}\left[\zeta_0 < z\right] = \lim_{n\to\infty} \frac{n}{p_n\sigma_n^2} \int_{|z|>\sigma_n\epsilon} z^2 dF_{p_n}(z)$$

which we assumed converges to 0, and so by invoking the Lindeberg condition we have shown the sufficiency part of Theorem 5.4. In the interest of brevity and limitation of scope, we leave out necessity, which can be proven by showing the condition of Remark 5.2 (see [2]). □

References

[1] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1987.

[2] I.A. Ibragimov and Yu. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, 1971.

[3] Wikipedia. *Central limit theorem — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Central%20limit%20theorem&oldid=1353138943`. [Online; accessed 13-May-2026]. 2026.

[4] Yuval Filmus. *Answer to "finite variance but infinite higher moments"*. `https://math.stackexchange.com/questions/1085558/finite-variance-but-infinite-higher-moments`. Math Stack Exchange. 2014-12-30. (Visited on 12/23/2025).