Central Limit Theorem of Some Statistics Associated with Self-Normalized Subordinators

CENTRAL LIMIT THEOREM OF SOME STATISTICS ASSOCIATED WITH SELF-NORMALIZED SUBORDINATORS

BY

BRIGHT M. K. AGBEWU

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2019) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE: Central Limit Theorem of Some Statistics Associated with Self-Normalized Subordinators

AUTHOR: Bright M. K. Agbewu

B.Sc., (Mathematics & Statistics)

University of Ghana, Ghana

SUPERVISOR: Dr. Shui Feng

NUMBER OF PAGES: viii, 61

To my family and friends

Abstract

Consider a population of m-type individuals labelled by $\{1, 2, \ldots, m\}$. Let $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ denote the relative frequencies of all types with x_i denoting the relative frequency of type i for $1 \leq i \leq m$. For a random sample of size 2 from the population, the probability that the individuals of the sample are of the same type is given by

$$H = \sum_{i=1}^{m} x_i^2.$$

In this thesis, we focus on the case where $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ is a random vector. The quantity H appears in various fields of study. For instance, it is associated with the Shannon entropy in communication, the Herfindahl-Hirschman index in economics and known as the homozygosity in population genetics.

In [7], fluctuation theorems for the infinite dimensional case $\{\varphi_r(\mathbf{x}) : r \geq 2\}$ defined as

$$\varphi_r(\mathbf{x}) = \sum_{i=1}^{\infty} x_i^r$$

are considered. In this thesis we present, under a moment assumption, a Central

Limit Theorem (CLT) associated with H and present as examples the Gamma subordinator case, which is a well known result by Griffiths [10], and the generalized Gamma subordinator case.

Acknowledgements

I want to use this opportunity to thank my supervisor, Professor Shui Feng, for his continual patience and guidance throughout the research period. This thesis would not have been possible without his immense knowledge and expertise.

I would also like to express my gratitude to my committee members: Professor Fred M. Hoppe and Professor Noah Forman for their precious time and knowledge.

I thank all the professors in the Mathematics and Statistics department of the McMaster University who have helped me one way or the other.

I also want to thank my family and friends for being there for me throughout the research period. Their encouragement kept me going.

Contents

Chapter 1

Introduction

A subordinator is a very important concept in the study of stochastic processes. Subordinators are special cases of Lévy processes which are explored into detail in Section 2.1 of Chapter 2. The representation by subordinators on which this thesis is based is presented in Section 2.2 of Chapter 2. In this thesis, we focus on self-normalized subordinators which represent normalized random measures with independent increments [17]. These notions of subordinators and normalized random measures with independent increments are rigorously presented in Sections 2.2 and 2.3 of Chapter 2 respectively. The main focus of this thesis is to provide a Central Limit Theorem for

$$H = \sum_{i=1}^{m} X_i^2,$$

where $(X_1, X_2, \ldots, X_m)$ is a random vector with $\sum_{i=1}^{m} X_i = 1$ and $0 < X_i < 1$ for each i.

Our interest in this statistic stems from its importance in several fields of study.

In Section 2.4 of Chapter 2, we present some areas of study and their respective applications of H.

In Chapter 3, we review Central Limit Theorems where we take a look at different structures of random variables (independent and identically distributed, independent but not identically distributed, and dependent random variables) and how CLT can be established in each of these cases.

In Chapter 4, the main results of the thesis are presented. We establish a CLT for H under a moment condition. Two examples of this result are presented by considering the Gamma subordinator and the generalized Gamma subordinator.

Chapter 2

Subordinators and Related Concepts

In this chapter, we introduce subordinators in detail. We first introduce Lévy processes and present subordinators as a special case of such processes. Examples of subordinators including the Gamma subordinator, the generalized Gamma subordinator and the stable subordinator are presented. Self-normalized random measures are briefly discussed. In the final part of the chapter, the homozygosity and some diversity indices related to H are also presented. The definitions and concepts presented here are mainly based on [1].

2.1 Lévy Processes

Defintion 2.1.1 (Lévy Process). *A stochastic process* $X = \{X_t : t \geq 0\}$ *in* $\mathbb{R}^n$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *is said to be a Lévy process if it satisfies the* *following properties:*

1. $X_0 = 0$ *almost surely, i.e,* $\mathbb{P}(X_0 = 0) = 1$

2. *For any* $0 \le t_1 < t_2 < \ldots < t_n < \infty$, $X_{t_1}, X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ *are independent.*

3. *For any* $s < t$, $X_t - X_s$ *is equal in distribution to* X_{t-s}.

4. *The paths of* X *are right-continuos with left limits.*

Lévy processes may be thought of as an extension of the classical family of *random walks* (sums of independent and identically distributed random variables). They are analogues of random walks in continuous time [1].

2.1.1 Examples of Lévy Processes

Brownian Motion

A (standard) Brownian motion [20] (sometimes called a Wiener process) in $\mathbb{R}^d$ is a Lévy process $B = \{B(t), t \ge 0\}$ such that

1. $B(0) = 0$

2. $B(t) - B(s) \sim N(0, t - s)$ for $0 \le s \le t$

3. $B(t)$ has independent increments.

It is one of the best known Lévy processes.

Poisson Process

The Poisson process [21] with rate $\lambda > 0$ is a Lévy process $N = \{N(t), t \geq 0\}$ such that

1. $N(0) = 0$
2. $N(t)$ has independent increments
3. The number of events in any interval of length t is Poisson distributed with mean λt.

2.2 Subordinators

2.2.1 The Lévy-Khintchine Formula

Here, a formula first established by Paul Lévy and A. Ya. Khintchine in the 1930s is presented. This formula gives a characterisation of infinitely divisible random variables by giving a representation of their characteristic functions.

Before stating the formula, we first take a look at what *infinitely divisible random variables* are.

Defintion 2.2.1 (Infinite Divisibility). *A real-valued random variable X is said to be infinitely divisible, if for all $n \in \mathbb{N}$ there exist independent and identically distributed (i.i.d.) random variables $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that*

$$X \stackrel{d}{=} Y_1^{(n)} + \ldots + Y_n^{(n)}.$$

Remark. *[1] The distribution of a Lévy process has the property of infinite divisibility.*

Lévy and Khintchine show the exact form of the characteristic function of an infinitely divisible random variable is

$$-\log \mathbb{E}(e^{i\xi\cdot X}) = -il\cdot\xi + \frac{1}{2}\xi\cdot Q\xi + \int_{\mathbb{R}^d-\{0\}}(1 - e^{iy\cdot\xi} + i\xi\cdot y\mathbb{1}_{\{|y|<1\}})\Lambda(dy) \quad (2.1)$$

where $l \in \mathbb{R}^d$, $Q \in \mathbb{R}^{d\times d}$ is a positive semidefinite symmetric matrix, and Λ is the Lévy measure on $\mathbb{R}^d - \{0\}$ satisfying $\int_{\mathbb{R}^d-\{0\}} \min\{1, |y|^2\}\Lambda(dy) < \infty$.

Defintion 2.2.2 (**Subordinator**). *A process $\{\tau_s : s \geq 0\}$ is called a subordinator if it has stationary, independent, and non-negative increments with $\tau_0 = 0$.*

Remark. *[1] A subordinator can be defined as a non-decreasing (a.s.) $\mathbb{R}$-valued Lévy process.*

Defintion 2.2.3. *A subordinator $\{\tau_s : s \geq 0\}$ has no drift (pure-jump) if for all $\lambda \geq 0$, $s \geq 0$, the Laplace transform is given by*

$$\mathbb{E}(e^{-\lambda\tau_s}) = \exp\left\{-s\int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \quad (2.2)$$

where Λ is the Lévy measure on $[0, \infty)$.

The Lévy measure Λ is a measure on $\mathbb{R}$ that satisfies:

1. $\Lambda(0, \infty) = \infty$
2. $\int_0^\infty \min\{x, 1\}\Lambda(dx) < \infty$.

In the remainder of this thesis, we focus on subordinators with no drift.

Now, we give a representation of a subordinator that would be used throughout this thesis.

Let $\{\tau_s : s \geq 0\}$ be a pure-jump subordinator up to time m. Let

$$Y_1 = \tau_1 - \tau_0, Y_2 = \tau_2 - \tau_1, \ldots, Y_m = \tau_m - \tau_{m-1}.$$

We refer to $Y_1, Y_2, \ldots, Y_m$ as the increments of $\{\tau_s : s \geq 0\}$. By definition, $Y_1, Y_2, \ldots, Y_m$ are independent and identically distributed and

$$\tau_m = \sum_{i=1}^{m} Y_i < \infty.$$

Define

$$X_i = \frac{Y_i}{\tau_m}, \quad i = 1, 2, \ldots, m \tag{2.3}$$

Then the random vector $(X_1, X_2, \ldots, X_m)$ specifies a random discrete probability on the set of integers $\{1, \ldots, m\}$ [16].

2.2.2 Examples of Subordinators

Gamma Subordinator

The subordinator $\{\gamma_s : s \geq 0\}$ is called a gamma subordinator if its Lévy measure is

$$\Lambda(dx) = x^{-1} e^{-x} dx, \quad x > 0.$$

The marginals of the gamma subordinator have the gamma distribution. In particular, for the Lévy measure specified above, the increments of the gamma process have the gamma distribution with shape parameter 1. We see that

1. $\Lambda(0,\infty) = \int_0^\infty x^{-1}e^{-x}dx$ diverges to infinity since $1/x$ is non-integrable in any interval that includes 0 because it diverges, hence $x^{-1}e^{-x}$ diverges too.

2. $$\begin{aligned}\int_0^\infty \min\{x,1\}\Lambda(dx) &= \int_0^\infty \min\{x,1\}x^{-1}e^{-x}dx \\ &= \int_0^1 x\cdot x^{-1}e^{-x}dx + \int_1^\infty x^{-1}e^{-x}dx \\ &= 1-e^{-1}+\Gamma(0,1) < \infty\end{aligned}$$

   where

   $$\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$$

   is the incomplete gamma function.

Its Laplace transform becomes

$$\mathbb{E}(e^{-\lambda\gamma_s}) = \frac{1}{(1+\lambda)^s}.$$

Generalized Gamma Subordinator

The subordinator $\{\rho_s : s \geq 0\}$ is a generalized Gamma process with scale parameter one [4, 14] if its Lévy measure is

$$\Lambda(dx) = \Gamma(1-\alpha)^{-1}x^{-(1+\alpha)}e^{-x}dx, \quad x>0$$

where $\alpha \in (0,1)$.

We have that,

1.
$$\int_0^\infty \Gamma(1-\alpha)^{-1}x^{-(1+\alpha)}e^{-x}dx = \Gamma(1-\alpha)^{-1}\int_0^\infty x^{-(1+\alpha)}e^{-x}dx$$
$$= \infty$$

since $\int_0^\infty x^{-(1+\alpha)}dx$ diverges, so would $\int_0^\infty x^{-(1+\alpha)}e^{-x}dx$.

2.
$$\begin{aligned}
&\int_0^\infty \min\{1,x\}\Lambda(dx)\\
&\quad= \int_0^1 x\cdot\Gamma(1-\alpha)^{-1}x^{-(1+\alpha)}e^{-x}dx + \int_1^\infty \Gamma(1-\alpha)^{-1}x^{-(1+\alpha)}e^{-x}dx\\
&\quad= \Gamma(1-\alpha)^{-1}\int_0^1 x^{-\alpha}e^{-x}dx + \Gamma(1-\alpha)^{-1}\int_1^\infty x^{-(1+\alpha)}e^{-x}dx\\
&\quad= \Gamma(1-\alpha)^{-1}\left(\Gamma(1-\alpha)-\Gamma(1-\alpha,1)\right) + \frac{\Gamma(1-\alpha)^{-1}}{\alpha}\left(e^{-1}-\Gamma(1-\alpha,1)\right)\\
&\quad< \infty
\end{aligned}$$

Here the Laplace transform is,

$$\mathbb{E}(e^{-\lambda\rho_s}) = \exp\left\{-\frac{s}{\alpha}((\lambda+1)^\alpha - 1)\right\}.$$

Stable Subordinator

The subordinator $\{\rho_s : s \geq 0\}$ is a stable subordinator with index $\alpha \in (0, 1]$ if its Lévy measure is

$$\Lambda(dx) = c_\alpha x^{-(1+\alpha)} dx, \quad x > 0, \quad c_\alpha > 0.$$

Realize that,

1. $\Lambda(0, \infty) = \int_0^\infty c_\alpha x^{-(1+\alpha)} dx = \infty$

2. $$\begin{aligned} \int_0^\infty \min\{x, 1\} \Lambda(dx) &= \int_0^\infty \min\{x, 1\} c_\alpha x^{-(1+\alpha)} dx \\ &= c_\alpha \int_0^1 x^{-\alpha} dx + c_\alpha \int_1^\infty x^{-(1+\alpha)} dx \\ &= \frac{c_\alpha}{\alpha(1-\alpha)} < \infty. \end{aligned}$$

Its Laplace transform is given by

$$\mathbb{E}(e^{-\lambda \rho_s}) = \exp\{-s\Gamma(1-\alpha)\lambda^\alpha\}.$$

Note that a stable subordinator with index $\alpha \in (0, 1]$ is a subordinator with zero drift and Lévy measure as defined above and this is the case we work with in this thesis.

2.3 Self-Normailized Random Measures

Definitions of a completely random measure (CRM) and a normalized random measure with independent increments (NRMIs) are given below as introduced by [11].

Let $\mathbb{X}$ be a Polish space with the Borel σ-algebra $\mathscr{X}$. Let $\mathcal{M}_{\mathbb{X}}$ denote the space of finite measures on $(\mathbb{X}, \mathscr{X})$ endowed with the Borel σ-algebra $\mathscr{M}_{\mathbb{X}}$.

Defintion 2.3.1. *[25] Let $\tilde{\mu}$ be a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathcal{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ such that for any $A_1, \ldots, A_m$ in $\mathscr{X}$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_m)$ are mutually independent. The measurable map $\tilde{\mu}$ is called a completely random measure.*

Defintion 2.3.2. *[25] Let $\tilde{\mu}$ be a CRM such that $0 < \tilde{\mu}(\mathbb{X}) < +\infty$ almost surely. The random probability measure $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$ is called normalized random measure with independent increments.*

From the above definitions, we note that generally, an NRMIs is defined as a random probability measure obtained by normalizing a CRM with finite total mass and characterized by some Lévy intensity measure $v(ds, dx) = \rho(ds|x)\alpha(dx)$ [25].

We say v is homogeneous when $v(ds, dx) = \rho(ds)\alpha(dx)$. In this case, the corresponding CRM (NRMI) is referred to as homogeneous CRM (homogeneous NRMI) [25].

Some examples of homogeneous NRMIs are the Dirichlet process [8], normalized generalized Gamma process [14] and the normalized stable process [12].

From the above discussions, realize that (2.3) is a self-normalized random measure.

Remark. *[8] Normalizing a Gamma CRM yields a Dirichlet process.*

2.4 The Homozygosity and Diversity Indices

For any integer $m \geq 2$, the function

$$H = \sum_{i=1}^{m} x_i^2, \quad (x_i, i \text{ in } [1,n]) \text{ in } \left\{0 < x_1, \ldots, x_m < 1, \sum_{i=1}^{m} x_i = 1\right\}$$

is very important in various fields of study. For instance, in population genetics, the statistic represents the homozygosity. It is associated with the Shannon entropy in communication, the Herfindahl-Hirschman index in economics, and the Gini-Simpson's index in ecology.

2.4.1 Homozygosity

Homozygosity is a term used in genetics to refer to the situation where a person inherits the same alleles for a single trait. Diploids typically have two alleles for any given trait.

Consider a locus with $m \geq 2$ alleles. Suppose the frequency of allele i is $x_i > 0$ and let the alleles be placed in decreasing order of frequency so that $x_i \geq x_j$ if $i < j$. Then for diploids, the fraction of homozygotes expected under the assumption of Hardy-Weinberg proportions can be defined as [18]

$$H = \sum_{i=1}^{m} x_i^2$$

where

$$\sum_{i=1}^{m} x_i = 1.$$

Now, consider the more general setting where we have an infinite number of alleles at the locus in question. Differently put, take a population of individuals with types labelled by $\{1, 2, \ldots\}$. Suppose the frequency of type i for $i \geq 1$ is $x_i > 0$ so that $\mathbf{x} = (x_1, x_2, \ldots)$ represents the relative frequencies of all types. For $r \geq 2$, the probability that out of a random sample of size r, individuals in the sample are of the same type is given by

$$\varphi_r(\mathbf{x}) = \sum_{i=1}^{\infty} x_i^r \tag{2.4}$$

with

$$\sum_{i=1}^{\infty} x_i = 1.$$

If $\mathbf{x}$ follow the one-parameter Poisson-Dirichlet distribution, then $\varphi_2(\mathbf{x})$ is known as the *homozygosity* of the population in population genetics [7]. For details on the one-parameter Poisson-Dirichlet distribution and related concepts, see Feng [7]. In Feng [7], fluctuation theorems for $\{\varphi_n(\mathbf{x}) : n \geq 2\}$ are presented. This thesis focuses on the finite case H.

2.4.2 Herfindahl-Hirschman Index

The Herfindahl-Hirschman Index (HHI), named after economists Orris C. Herfindahl and Albert O. Hirschman, is a commonly accepted measure of market concentration. This index is considered as superior to other concentration measures such as the CR_4 and CR_8 concentration ratios. It has been in use by the U.S. Department of Justice since 1982 and is the measure of concentration used in governmental merger analysis [13].

Let x_i denote the market share of firm i in the market. Then the HHI is given by

$$H = \sum_{i=1}^{m} x_i^2$$

where m is the number of firms in the industry.

The values of H are in the interval $(0, 1]$ when the market shares are expressed as fractions and $(0, 10000]$ points when the market shares are expressed as whole percentages.

Smaller values of HHI indicate a very competitive market (unconcentrated markets) while larger values indicate a monopolistic market (highly concentrated markets). An increase in HHI usually means a decrease in competition and an increase of market power.

Usually, a threshold is set to determine which values of HHI indicate an unconcentrated market, a moderately concentrated market and a highly concentrated market. For instance, according to the Horizontal Merger Guidelines of the U.S. Department of Justice and Federal Trade Commission, markets are classified into three types:

- Unconcentrated Markets: HHI below 0.15
- Moderately Concentrated Markets: HHI between 0.15 and 0.25
- Highly Concentrated Markets: HHI above 0.25

2.4.3 Simpson's Index

The Simpson's index [22] is a measure of the concentration of individuals in an infinite population.

More formally, consider an infinite population with individuals belonging to one of m groups, and let $x_1, \ldots, x_m$ be the frequency of individuals in the various groups with $\sum_{i=1}^{m} x_i = 1$. Then the Simpson's index

$$H = \sum_{i=1}^{m} x_i^2$$

is a measure of the concentration of the classification.

The smaller the value of H, the more diverse or less concentrated the population is and the larger the value of H, the more concentrated or equivalently, less diverse the population is.

Chapter 3

Central Limit Theorems

The Central Limit Theorem (CLT) is one of the most important and widely used results in probability theory with many useful applications in most fields. In this chapter, we begin with the classical CLT where we consider independent and identically distributed (i.i.d.) random variables and then proceed to Feller-Lindeberg theorems where we present results regarding independent variables that are not necessarily identical. We conclude with Stein's method which handles the case of dependent random variables.

3.1 Classical CLT

Let $X_1, X_2, \ldots, X_m$ be i.i.d. random variables with finite common mean and variance. Central Limit Theorems associated with this basic structure of random variables are often referred to as the classical CLT. In the following theorem, we present the classical CLT.

Theorem 3.1.1 (Classical CLT). *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}(X_i) = \mu < \infty$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. If $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, then*

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1). \tag{3.1}$$

Proof.

$$\begin{aligned} T_n &= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \qquad \text{where } Y_i = \frac{X_i - \mu}{\sigma} \end{aligned}$$

Then we have that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with $\mathbb{E}(Y_i) = 0$ and $\mathrm{Var}(Y_i) = 1$. The characteristic function of T_n by definition is

$$\begin{aligned} \varphi_{T_n}(t) &= \mathbb{E}\left(e^{itT_n}\right) \\ &= \mathbb{E}\left(e^{it\sum_{i=1}^n Y_i/\sqrt{n}}\right) \\ &= \mathbb{E}\left(e^{itY_1/\sqrt{n}+\ldots+itY_n/\sqrt{n}}\right) \\ &= \mathbb{E}\left(e^{itY_1/\sqrt{n}} \cdot e^{itY_2/\sqrt{n}} \cdots e^{itY_n/\sqrt{n}}\right) \\ &= \mathbb{E}\left(e^{itY_1/\sqrt{n}}\right)\mathbb{E}\left(e^{itY_2/\sqrt{n}}\right) \cdots \mathbb{E}\left(e^{itY_n/\sqrt{n}}\right) \qquad \text{(indep. of } Y_i\text{'s)} \\ &= \left[\mathbb{E}\left(e^{itY_1/\sqrt{n}}\right)\right]^n \qquad (Y_i\text{'s are identically distributed)} \\ &= \left[\varphi_{Y_1/\sqrt{n}}(t)\right]^n \end{aligned}$$

Observe that

$$e^{itY_1/\sqrt{n}} = 1 + \frac{itY_1}{\sqrt{n}} - \frac{t^2Y_1^2}{2n} + \ldots$$

Thus

$$\varphi_{Y_1/\sqrt{n}}(t) = \mathbb{E}\left(e^{itY_1/\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

and

$$\varphi_{T_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n$$

so that

$$\varphi_{T_n}(t) \longrightarrow e^{-\frac{1}{2}t^2} \quad \text{as } n \to \infty$$

But $e^{-\frac{1}{2}t^2}$ is the characteristic function of the standard normal distribution and the proof is complete from the continuity theorem. For more on continuity theorem, see [9]. □

The two fundamental questions that arise from the classical CLT are:

1. Whether or not CLT still holds when random variables are independent but not necessarily identical.

2. Whether or not CLT still holds when random variables are not independent.

These fundamental questions are addressed by the works of Lindeberg and Feller and Stein.

Lindeberg and Feller showed that under some conditions, CLT still holds for independent random variables that are not necessarily identical. Charles Stein, on the other hand, developed a sophisticated approach for handling the case of dependent random variables.

Now, we present concepts and theorems associated with the works of Feller and Lindeberg in establishing CLT.

3.2 Feller-Lindeberg Theorems

Before presenting the main results of this section, we need the idea of *triangular array of independent random variables* which we present as follows.

Suppose $X_1, \ldots, X_n$ are independent random variables, even possibly identically distributed, but with their distributions depending on n. For instance, suppose random variables have the Poisson distribution with mean λ_n, where λ_n changes with n. Then it becomes necessary to have a way of representing these random variables, and this brings about the idea of triangular array of random variables. Definitions and results presented in this section are mainly based on [3].

Defintion 3.2.1. *For each $n \geq 1$, let $\{X_{n1}, \ldots, X_{nr_n}\}$ be a collection of random variables defined on a probability space $(\Omega_n, \mathcal{F}_n, P_n)$ such that $X_{n1}, \ldots, X_{nr_n}$ are independent. Then, $\{X_{n1}, \ldots, X_{nr_n}\}_{n\geq 1}$ is called a triangular array of independent random variables.*

Let $\{X_{n1}, \ldots, X_{nr_n}\}_{n\geq 1}$ be a triangular array of independent random variables. Then, the sums defined as

$$S_n = \sum_{j=1}^{r_n} X_{nj}, \quad n \geq 1 \tag{3.2}$$

are called the row sums.

We are now in a good position to state the *Lindeberg Condition* which plays a major role in establishing CLT for row sums.

Defintion 3.2.2. *Let* $\{X_{n1}, \ldots, X_{nr_n}\}_{n\geq 1}$ *be a triangular array of independent random variables with*

$$\mathbb{E}(X_{nj}) = 0, \quad \mathbb{E}(X_{nj}^2) = \sigma_{nj}^2 < \infty \quad \textit{for all} \quad 1 \leq j \leq r_n, \quad n \geq 1. \tag{3.3}$$

Then $\{X_{n1}, \ldots, X_{nr_n}\}_{n\geq 1}$ *is said to satisfy the Lindeberg condition if for every* $\epsilon > 0$,

$$\lim_{n\to\infty} s_n^{-2} \sum_{j=1}^{r_n} \mathbb{E}(X_{nj}^2 \cdot \mathbb{1}(|X_{nj}| > \epsilon s_n)) = 0, \tag{3.4}$$

where $s_n^2 = \sum_{j=1}^{r_n} \sigma_{nj}^2, \quad n \geq 1.$

In the following example, we show that (3.1) satisfies the Lindeberg condition.

Example 1. Let $X_1, \ldots, X_n$ be i.i.d. random variables such that $\mathbb{E}(X_i) = \mu < \infty$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define T_n as in (3.1). Writing T_n as the row sum of a triangular array of independent random variables, we have:

$$T_n = \sum_{j=1}^{n} X_{nj},$$

where $X_{nj} = (X_j - \mu)/\{\sigma\sqrt{n}\}$, $1 \leq j \leq n$, $n \geq 1$.

It is clear that $\mathbb{E}(X_{nj}) = 0$ and $\mathbb{E}(X_{nj}^2) = 1/n$ for all $1 \leq j \leq n$, $n \geq 1$. We are left to show that

$$\lim_{n\to\infty} s_n^{-2} \sum_{j=1}^{n} \mathbb{E}(X_{nj}^2 \cdot \mathbb{1}(|X_{nj}| > \epsilon s_n)) = 0,$$

where $s_n^2 = \sum_{j=1}^{n} \sigma_{nj}^2 = \sum_{j=1}^{n} \frac{1}{n} = 1.$

We have,

$$\begin{aligned}
s_n^{-2}\sum_{j=1}^{n}\mathbb{E}(X_{nj}^2\cdot\mathbb{1}(|X_{nj}|>\epsilon s_n)) &= \sum_{j=1}^{n}\mathbb{E}\left[\left(\frac{X_j-\mu}{\sigma\sqrt{n}}\right)^2\cdot\mathbb{1}\left(\left|\frac{X_j-\mu}{\sigma\sqrt{n}}\right|>\epsilon\right)\right] \\
&= \sum_{j=1}^{n}\frac{1}{\sigma^2 n}\mathbb{E}[(X_j-\mu)^2\cdot\mathbb{1}(|X_j-\mu|>\epsilon\sigma\sqrt{n})] \\
&= n\cdot\frac{1}{\sigma^2 n}\mathbb{E}[(X_1-\mu)^2\cdot\mathbb{1}(|X_1-\mu|>\epsilon\sigma\sqrt{n})] \\
&= \sigma^{-2}\mathbb{E}[(X_1-\mu)^2\cdot\mathbb{1}(|X_1-\mu|>\epsilon\sigma\sqrt{n})]
\end{aligned}$$

Let $Z_n = (X_1-\mu)^2\cdot\mathbb{1}(|X_1-\mu|>\epsilon\sigma\sqrt{n})$. We have that $|Z_n|\leq(X_1-\mu)^2$ and $\mathbb{E}(X_1-\mu)^2<\infty$. It is also clear that, the probability of Z_n approaches zero as $n\to\infty$ and hence we conclude that $Z_n \xrightarrow{P} 0$. Now, applying the Dominated Convergence Theorem, it is concluded that $\mathbb{E}(Z_n)\to 0$ and the Lindeberg condition is satisfied.

Next, the main result of this section is presented. This is the Lindeberg CLT for triangular array of independent random variables.

Theorem 3.2.1 (Lindeberg's CLT). *Let $\{X_{nj}: 1\leq j\leq r_n\}_{n\geq1}$ be a triangular array of independent random variables satisfying (3.3) and the Lindeberg's condition (3.4). Then,*

$$\frac{S_n}{s_n}\xrightarrow{d} N(0,1)$$

where $S_n=\sum_{j=1}^{r_n}X_{nj}$ and $s_n^2=\sum_{j=1}^{r_n}\sigma_{nj}^2$.

Remark. *Another way to establish the result of the classical CLT is to observe that, from Example 1, i.i.d. random variables satisfy the Lindeberg condition and thus the result follows from the Lindeberg CLT. Clearly, the Lindeberg CLT generalizes the classical CLT.*

Defintion 3.2.3 (Lyapounov's condition). *A triangular array $\{X_{nj} : 1 \leq j \leq r_n\}_{n\geq 1}$ of independent random variables satisfying (3.3) is said to satisfy Lyapounov's condition if there exists a $\delta \in (0, \infty)$ such that*

$$\lim_{n\to\infty} s_n^{-(2+\delta)} \sum_{j=1}^{r_n} \mathbb{E}\left(|X_{nj}|^{2+\delta}\right) = 0, \tag{3.5}$$

where $s_n^2 = \sum_{j=1}^{r_n} \sigma_{nj}^2$

Lyapounov's condition is a stronger condition than Lindeberg's condition and it is often easier to check.

Lemma 3.2.1. *Lyapounov's condition implies Lindeberg's condition.*

Proof. We need to show that

$$\lim_{n\to\infty} s_n^{-2} \sum_{j=1}^{r_n} \mathbb{E}(X_{nj}^2 \cdot \mathbb{1}(|X_{nj}| > \epsilon s_n)) = 0$$

whenever

$$\lim_{n\to\infty} s_n^{-(2+\delta)} \sum_{j=1}^{r_n} \mathbb{E}\left(|X_{nj}|^{2+\delta}\right) = 0.$$

Suppose $\lim_{n\to\infty} s_n^{-(2+\delta)} \sum_{j=1}^{r_n} \mathbb{E}\left(|X_{nj}|^{2+\delta}\right) = 0$. Let $\epsilon, \delta > 0$. Observe that for any random variable $|X_{nj}| > \epsilon s_n$, we have

$$X_{nj}^2 = \frac{|X_{nj}|^{2+\delta}}{|X_{nj}|^{\delta}} \leq \frac{|X_{nj}|^{2+\delta}}{(\epsilon s_n)^{\delta}}$$

Thus for any random variable X_{nj}, we have

$$\mathbb{E}(X_{nj}^2 \cdot \mathbb{1}(|X_{nj}| > \epsilon s_n)) \leq \frac{1}{\epsilon^\delta \cdot s_n^\delta} \mathbb{E}\left(|X_{nj}|^{2+\delta}\right)$$

This implies

$$\lim_{n\to\infty} s_n^{-2} \sum_{j=1}^{r_n} \mathbb{E}(X_{nj}^2 \cdot \mathbb{1}(|X_{nj}| > \epsilon s_n)) \leq \lim_{n\to\infty} \frac{1}{\epsilon^\delta} s_n^{-(2+\delta)} \sum_{j=1}^{r_n} \mathbb{E}\left(|X_{nj}|^{2+\delta}\right) = 0$$

Hence the Lindeberg condition is satisfied. □

Corollary 3.2.1.1 (Lyapounov's CLT). *Let $\{X_{nj} : 1 \leq j \leq r_n\}_{n\geq 1}$ be a triangular array of independent random variables satisfying (3.3) and Lyapounov's condition (3.5). Then*

$$\frac{S_n}{s_n} \xrightarrow{d} N(0,1)$$

where $S_n = \sum_{j=1}^{r_n} X_{nj}$ and $s_n^2 = \sum_{j=1}^{r_n} \sigma_{nj}^2$

The Lindeberg's condition and Lyapounov's condition presented are both sufficient conditions for the validity of the CLT. The necessary condition for the CLT is provided by Feller.

Theorem 3.2.2 (Feller's Theorem). *Let $\{X_{nj} : 1 \leq j \leq r_n\}_{n\geq 1}$ be a triangular array of independent random variables satisfying (3.3) such that*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \max_{1\leq j\leq r_n} \sigma_{nj}^2 = 0 \qquad (3.6)$$

where $s_n^2 = \sum_{j=1}^{r_n} \sigma_{nj}^2$. *If for* $S_n = \sum_{j=1}^{r_n} X_{nj}$,

$$\frac{S_n}{s_n} \xrightarrow{d} N(0,1),$$

then $\{X_{nj} : 1 \leq j \leq r_n\}_{n\geq 1}$ *satisfies the Lindeberg condition.*

3.3 Stein's Method

Stein's method is a sophisticated technique developed by Charles Stein to quantify the error in the approximation of a distribution by another distribution in various metrics. Initially, Stein's method was introduced in [24] to estimate the error in approximating the distribution of sum of dependent random variables of certain structure by the normal distribution. This method has however been extended beyond the normal distribution to other distributions such as the Poisson distribution. See, for instance, [2] for Poisson approximation using Stein's method. In this section, we focus on Stein's method for normal approximation.

Stein's method usually deals with bounding the distance between two probability distributions in a particular metric. To this effect, some useful probability metrics are introduced and an important relationship between the *Kolmogorov* metric and *Wasserstein* metric is established.

Definitions, concepts and results presented here are based mainly on [5, 19].

3.3.1 Metrics on Probability Measures

In this section, we review some commonly used metrics on probability measures in applying Stein's method. For a comprehensive review of probability metrics, see [23]. Let Ω denote a measurable space. Let μ and ν represent two probability measures on Ω. We are interested in the family of distance measures known as the *integral probability metrics* (IPMs) [15], defined as

$$d_{\mathscr{G}}(\mu, \nu) := \sup_{g \in \mathscr{G}} \left| \int_{\Omega} g(x) d\mu(x) - \int_{\Omega} g(x) d\nu(x) \right| \tag{3.7}$$

where $\mathscr{G}$ is a class of real-valued bounded measurable functions on Ω.

Given two random variables W and Z, with respective distribution functions $F_W(x)$ and $F_Z(x)$, the probability metric takes the form;

$$d_{\mathscr{G}}(W, Z) = \sup_{g \in \mathscr{G}} \left| \int g(x) dF_W(x) - \int g(x) dF_Z(x) \right| = \sup_{g \in \mathscr{G}} \Big| \mathbb{E}(g(W)) - \mathbb{E}(g(Z)) \Big|.$$

By taking special cases of $\mathscr{G}$ in (3.7), different useful probability metrics are obtained as follows:

1. By letting $\mathscr{G} = \{\mathbb{1}[\cdot \leq x] : x \in \mathbb{R}\}$ in (3.7), the *Kolmogorov* metric is obtained. This is denoted by d_K. In this metric, a sequence of distributions converging to a fixed distribution implies weak convergence.

2. By letting $\mathscr{G} = \{g : \mathbb{R} \longrightarrow \mathbb{R} : |g(x) - g(y)| \leq |x - y|\}$ in (3.7), i.e. a collection of 1-Lipschitz functions, in (3.7), the *Wasserstein* metric is obtained. This is denoted by d_W.

3. If $\mathscr{G} = \{\mathbb{1}[\cdot \in A] : A \in \text{Borel}(\mathbb{R})\}$ in (3.7), the *total variation* metric is obtained. This metric is denoted d_{TV}

In using Stein's method for normal approximation, the Wasserstein metric is frequently used. The following result shows one of the reasons why this choice of metric works in the approximation of a given distribution by the normal distribution.

Lemma 3.3.1. *If the random variable Z has Lebesgue density bounded by C, then for any random variable W,*

$$d_K(W, Z) \leq \sqrt{2C d_W(W, Z)}.$$

Proof. Let $\varepsilon > 0$. Define $g_x(w) = \mathbb{1}[w \leq x]$ and

$$g_{x,\varepsilon}(w) = \begin{cases} 1 & w \leq x, \\ -\frac{1}{\varepsilon}(w - (x + \varepsilon)) & x < w < x + \varepsilon, \\ 0 & w \geq x + \varepsilon. \end{cases}$$

Then we have

$$\begin{aligned} \mathbb{E}(g_x(W)) - \mathbb{E}(g_x(Z)) &= \mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x) \\ &= \mathbb{P}(W \leq x) - \mathbb{E}(g_{x,\varepsilon}(Z)) + \mathbb{E}(g_{x,\varepsilon}(Z)) - \mathbb{P}(Z \leq x) \\ &\leq \mathbb{E}(g_{x,\varepsilon}(W)) - \mathbb{E}(g_{x,\varepsilon}(Z)) + \mathbb{E}(g_{x,\varepsilon}(Z)) - \mathbb{P}(Z \leq x) \end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}(g_{x,\varepsilon}(Z)) - \mathbb{P}(Z \le x) &= \mathbb{P}(Z \le x) - \frac{1}{\varepsilon}\int_x^{x+\varepsilon}(z-(x+\varepsilon))f_Z(z)dz - \mathbb{P}(Z \le x) \\
&= -\frac{1}{\varepsilon}\int_x^{x+\varepsilon}(z-(x+\varepsilon))f_Z(z)dz \\
&= -\frac{1}{\varepsilon}\int_{-\varepsilon}^{0} uf(u)du \quad \text{where } u = z-(x+\varepsilon) \\
&\le -\frac{C}{\varepsilon}\int_{-\varepsilon}^{0} udu \\
&= \frac{C\varepsilon}{2}
\end{aligned}$$

Also, $\mathbb{E}(g_{x,\varepsilon}(W)) - \mathbb{E}(g_{x,\varepsilon}(Z)) \le \frac{1}{\varepsilon}d_W(W,Z)$ since $g_{x,\varepsilon}(w)$ is $\frac{1}{\varepsilon}-$Lipschitz by definition.

Thus we have

$$\mathbb{P}(W \le x) - \mathbb{P}(Z \le x) \le \frac{1}{\varepsilon}d_W(W,Z) + \frac{C\varepsilon}{2}$$

Taking $\varepsilon = \sqrt{2d_W(W,Z)/C}$ yields

$$\mathbb{P}(W \le x) - \mathbb{P}(Z \le x) \le \sqrt{2Cd_W(W,Z)}.$$

Using a similar argument and defining

$$g^*_{x,\varepsilon}(w) = \begin{cases} 1 & w \le x-\varepsilon, \\ -\frac{1}{\varepsilon}(w-x) & x-\varepsilon < w < x, \\ 0 & w \ge x. \end{cases}$$

yields the same upper bound for $\mathbb{P}(Z \leq x) - \mathbb{P}(W \leq x)$. Therefore,

$$d_K(W, Z) \leq \sqrt{2C d_W(W, Z)}.$$

□

Lemma 3.3.1 implies that, an upper bound for the Wasserstein metric gives an upper bound for the Kolmogorov metric. Therefore, a convergence in the Wasserstein metric implies weak convergence. This is one of the reasons why working with the Wasserstein metric is not a bad idea. Another reason why the Wasserstein metric is often used in the normal approximation as opposed to directly working with the Kolmogorov metric would become clear after the following results.

3.3.2 Basics of Stein's Method

The following results point out Stein's idea for normal approximation.

Defintion 3.3.1. *A function $f : [a, b] \to \mathbb{R}$ is said to be absolutely continuous on $[a, b]$ if, $\forall \epsilon > 0$, $\exists \delta > 0$ such that*

$$\sum_{j=1}^{n} |f(y_j) - f(x_j)| < \epsilon,$$

whenever $\{(x_j, y_j) : j = 1, \ldots, n\}$ is a finite collection of mutually disjoint subintervals of $[a, b]$ with $\sum_{j=1}^{n} |y_j - x_j| < \delta$.

Lemma 3.3.2 (Stein's Identity). *If $W \sim N(0,1)$, then*

$$\mathbb{E}(f'(W)) = \mathbb{E}(Wf(W)) \tag{3.8}$$

for all absolute continuous functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$. Conversely, if (3.8) holds for all bounded, continuous and piecewise continuously differentiable functions f with $\mathbb{E}|f'(Z)| < \infty$, then $W \sim N(0,1)$.

The functional operator defined by

$$\mathscr{D}(f(x)) = f'(x) - xf(x) \tag{3.9}$$

is referred to as the characterizing operator of the standard normal distribution. Stein's method for normal approximation depends heavily on this characterizing operator of the standard normal distribution.

The proof of **Lemma 3.3.2** is later presented in the section.

Now, consider testing how close the distributions of Z and W are by taking the difference between their expectations $\mathbb{E}(g(Z))$ and $\mathbb{E}(g(W))$ over some set of functions g. Then intuitively, we expect $\mathbb{E}(g(W)) - \mathbb{E}(g(Z))$ to be close to zero if the distributions of Z and W are close. Assume one of the random variables, say Z, has a characterizing operator. In particular, take $Z \sim N(0,1)$, then from (3.8) and (3.9), we know that

$$\mathbb{E}(\mathscr{D}(f(Z))) = 0.$$

In this case, if the distribution of W is close to the distribution of Z, then we expect

$\mathbb{E}(\mathscr{D}(f(W)))$ to be close to zero. Putting these two differences together gives rise to the Stein's equation

$$f'(w) - wf(w) = g(w) - \mathbb{E}(g(Z)). \tag{3.10}$$

Now, realize that, if for every $g \in \mathscr{G}$, there exists $f \in \mathscr{F}$ such that for all w (3.10) is satisfied, then we have

$$d_{\mathscr{G}}(W, Z) = \sup_{g \in \mathscr{G}} \Big| \mathbb{E}(g(W)) - \mathbb{E}(g(Z)) \Big| \leq \sup_{f \in \mathscr{F}} \Big| \mathbb{E}(f'(W)) - \mathbb{E}(Wf(W)) \Big| \tag{3.11}$$

which is obvious by taking expectations on both sides of (3.10).

It is important to note that the solution f of Stein's equation always exists when g is 1-Lipschitz.

The following result gives some boundary conditions on the solution of Stein's equation (3.10).

Lemma 3.3.3. *Let f be the solution of the differential equation*

$$f'(w) - wf(w) = g(w) - \mathbb{E}(g(Z))$$

1. *If g is bounded, then*

$$\|f\|_\infty \leq \sqrt{\frac{\pi}{2}} \|g(\cdot) - \mathbb{E}(g(Z))\|_\infty, \quad \textit{and} \quad \|f'\|_\infty \leq 2\|g(\cdot) - \mathbb{E}(g(Z))\|_\infty.$$

2. If g is absolute continuous, then

$$\|f\|_\infty \leq 2\|g'\|_\infty, \quad \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}}\|g'\|_\infty \quad \textit{and} \quad \|f''\|_\infty \leq 2\|g'\|_\infty.$$

Observe from **Lemma 3.3.3** and (3.11) that, if g is 1-Lipschitz and if we define $\mathscr{F} = \{f : \|f\|_\infty \leq 1, \quad \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \|f''\|_\infty \leq 2\}$, then (3.11) becomes

$$d_W(W,Z) = \sup_{g\in\mathscr{G}}\Big|\mathbb{E}(g(W)) - \mathbb{E}(g(Z))\Big| \leq \sup_{f\in\mathscr{F}}\Big|\mathbb{E}(f'(W) - Wf(W))\Big|.$$

Now, to show that $W \xrightarrow{d} N(0,1)$, we have to show that the upper bound of

$$\sup_{f\in\mathscr{F}}\Big|\mathbb{E}(f'(W) - Wf(W))\Big|$$

is approximately zero so that $d_W(W,Z) \to 0$ and from **Lemma 3.3.1**, $d_K(W,Z) \to 0$. As noted earlier, convergence in the Kolmogorov metric implies weak convergence and so we can conclude $W \xrightarrow{d} N(0,1)$ whenever $d_K(W,Z) \to 0$.

In Section 3.3.3, various approaches are presented in bounding $|\mathbb{E}(f'(W)-Wf(W))|$ using the structure of W.

It should now be clear from **Lemma 3.3.3** why it is often better to work in the Wasserstein metric. Specifically, from item 2 of **Lemma 3.3.3**, we see that the solution of Stein's equation (3.10) f is bounded with two bounded derivatives in the Wasserstein metric. In contrast, in the Kolmogorov metric, we see from item 1 of **Lemma 3.3.3** that, f is bounded with one bounded derivative but is not twice differentiable.

The following result is needed to prove **Lemma 3.3.2** (Stein's identity).

Lemma 3.3.4. *If $\Phi(x)$ is the cumulative distribution function of Z, then the unique bounded solution f_x of the differential equation*

$$f_x'(w) - wf_x(w) = \mathbb{1}[w \leq x] - \Phi(x)$$

is given by

$$\begin{aligned} f_x(w) &= e^{w^2/2} \int_w^\infty e^{-t^2/2}(\Phi(x) - \mathbb{1}[t \leq x])dt \\ &= -e^{w^2/2} \int_{-\infty}^w e^{-t^2/2}(\Phi(x) - \mathbb{1}[t \leq x])dt. \end{aligned}$$

At this point, we have all the necessary results to prove **Lemma 3.3.2**.

Proof of Lemma 3.3.2. Suppose f is an absolute continuous function with $\mathbb{E}|f'(Z)| < \infty$. Assume W has a standard normal distribution. Then

$$\begin{aligned} \mathbb{E}(f'(W)) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(t)e^{-t^2/2}dt \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty f'(t)e^{-t^2/2}dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(t)e^{-t^2/2}dt \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty f'(t) \left(\int_t^\infty xe^{-x^2/2}dx \right) dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(t) \left(\int_{-\infty}^t xe^{-x^2/2}dx \right) dt \end{aligned}$$

By Fubini's theorem, we can exchange the order of integral to get

$$\begin{aligned}
\mathbb{E}(f'(W)) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \left(\int_0^x f'(t)dt \right) xe^{-x^2/2}dx + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \left(\int_x^0 f'(t)dt \right) (-x)e^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty (f(x) - f(0))\, xe^{-x^2/2}dx + \frac{1}{2\pi} \int_{-\infty}^0 (f(x) - f(0))\, xe^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (f(x) - f(0)) xe^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} xf(x)e^{-x^2/2}dx - \frac{1}{\sqrt{2\pi}} f(0) \int_{\mathbb{R}} xe^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} xf(x)e^{-x^2/2}dx \\
&= \mathbb{E}(Wf(W))
\end{aligned}$$

Conversely, assume $\mathbb{E}(f'(W) - Wf(W)) = 0$ for all bounded, continuous, and piecewise continuously differentiable functions f with $\mathbb{E}|f'(Z)| < \infty$. In particular, we see from **Lemma 3.3.4** that, the unique bounded solution f_x of the differential equation

$$f_x' - wf_x(w) = \mathbb{1}[w \leq x] - \Phi(x)$$

is continuous and piecewise continuously differentiable. Thus we have

$$0 = \mathbb{E}(f_x'(W) - Wf_x(W)) = \mathbb{E}(\mathbb{1}[W \leq x] - \Phi(x)) = \mathbb{P}(W \leq x) - \Phi(x).$$

Thus W has a standard normal distribution. $\square$

In the following theorem, we present a summary of our discussion of Stein's method to this point.

Theorem 3.3.1. *If W is a random variable and Z has the standard normal distribution, and we define the family of functions* $\mathcal{F} = \left\{ f : \|f\|_\infty \leqslant 1, \|f'\|_\infty \leqslant \sqrt{2/\pi}, \|f''\|_\infty \leqslant 2 \right\}$, *then*

$$d_{\mathrm{W}}(W, Z) \leqslant \sup_{f \in \mathcal{F}} \Big| \mathbb{E}(f'(W) - Wf(W)) \Big|.$$

3.3.3 Upper Bound of the Wasserstein Metric

Depending on the structure of W, there are various methods employed in bounding the Wasserstein metric. Some of these techniques include the dependency graph approach, method of exchangeable pairs, size-bias coupling among others. In this section, we explore the dependency graph and method of exchangeable pairs in detail.

Before diving into details on dependency graphs and method of exchangeable pairs, we present an example that shows how the classical CLT is established in the Wasserstein metric.

Sum of Independent Random Variables

Theorem 3.3.2. *Let $E_1, \ldots, E_m$ be independent random variables with $\mathbb{E}|E_i|^3 < \infty$, $\mathbb{E}(E_i) = 0$, and $\mathbb{E}(E_i^2) = 1$. If $W = (\sum_{i=1}^m E_i)/\sqrt{m}$ and Z has the standard normal distribution, then*

$$d_W(W, Z) \leq \frac{3}{m^{\frac{3}{2}}} \sum_{i=1}^m \mathbb{E}|E_i|^3$$

and $W \xrightarrow{d} N(0,1)$.

Proof. Take any $f \in C'$ with f' absolutely continuous, and satisfying $\|f\|_\infty \leqslant 1, \|f''\|_\infty \leqslant 2, \|f'\|_\infty \leqslant \sqrt{2/\pi}$.

Let $W_i = \frac{\sum_{j \neq i} E_j}{\sqrt{m}} = W - \frac{E_i}{\sqrt{m}}$ so that W_i and E_i are independent.

Now, $\mathbb{E}(Wf(W)) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \mathbb{E}(E_i f(W))$.

Note that $\mathbb{E}(E_i f(W_i)) = \mathbb{E}(E_i)\mathbb{E}(f(W_i)) = 0$ since E_i and W_i are independent and $\mathbb{E}(E_i) = 0$.

From this, we have

$$\begin{aligned}\mathbb{E}(E_i f(W)) &= \mathbb{E}(E_i(f(W) - f(W_i))) \\ &= \mathbb{E}(E_i(f(W) - f(W_i) - (W - W_i)f'(W_i))) + \mathbb{E}(E_i(W - W_i)f'(W_i))\end{aligned}$$

Note by Taylors expansion that

$$|f(b) - f(a) - (b-a)f'(a)| \leq \frac{1}{2}(b-a)^2 |f''|_\infty$$

So we have

$$\begin{aligned}|\mathbb{E}[E_i(f(W) - f(W_i) - (W - W_i)f'(W_i))]| &\leq \frac{1}{2}|f''|_\infty \mathbb{E}\left|E_i \cdot \frac{E_i^2}{m}\right| \\ &\leq \frac{1}{m}\mathbb{E}|E_i|^3\end{aligned}$$

and thus

$$\left|\mathbb{E}(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} E_i(f(W) - f(W_i) - (W - W_i)f'(W_i)))\right| \leq \frac{1}{m^{\frac{3}{2}}} \sum_{i=1}^{m} \mathbb{E}|E_i|^3$$

which yields

$$\left|\mathbb{E}(Wf(W)) - \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}(E_i^2)\mathbb{E}(f'(W_i))\right| \leq \frac{1}{m^{\frac{3}{2}}} \sum_{i=1}^{m} \mathbb{E}|E_i|^3$$

and we have

$$\left|\mathbb{E}(Wf(W)) - \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}(f'(W_i))\right| \leq \frac{1}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|^3$$

Also, note that

$$\mathbb{E}(E_i(W - W_i)f'(W_i)) = \mathbb{E}\left[\frac{E_i^2}{\sqrt{m}}f'(W_i)\right] = \frac{1}{\sqrt{m}}\mathbb{E}(f'(W_i))$$

This implies

$$\begin{aligned}\left|\mathbb{E}(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}E_i(W - W_i)f'(W_i)) - \mathbb{E}(f'(W))\right| &= \left|\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}(f'(W_i)) - \mathbb{E}(f'(W))\right| \\ &\leq \frac{|f''|_\infty}{m}\sum_{i=1}^{m}\mathbb{E}|W - W_i| \\ &\leq \frac{2}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|\end{aligned}$$

Combining the results above, we get

$$\begin{aligned}|\mathbb{E}(Wf(W)) - \mathbb{E}(f'(W))| &\leq \frac{1}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|^3 + \frac{2}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i| \\ &\leq \frac{3}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|^3 \quad \text{since } \mathbb{E}|E_i| \leq (\mathbb{E}|E_i|^3)^{\frac{1}{3}} \leq \mathbb{E}|E_i|^3\end{aligned}$$

Therefore

$$d_W(W, Z) \leq \frac{3}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|^3$$

But

$$\frac{3}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathbb{E}|E_i|^3 \to 0 \quad \text{as } m \to \infty \text{ since } \mathbb{E}|E_i|^3 < \infty.$$

Hence we conclude that $W \xrightarrow{d} N(0,1)$. $\square$

Dependency Graph

Here, the result obtained under sums of independent random variables (Theorem 3.3.2) is extended to sums of random variables with local dependence.

Defintion 3.3.2. *A dependency graph for random variables $\{X_i\}_{i\in V}$ is any graph G with vertex set V such that if A, B are two disjoint subsets of V so that there are no edges between vertices in A and B, then $\{X_i\}_{i\in A}$ and $\{X_i\}_{i\in B}$ are mutually independent.*

Defintion 3.3.3 (Dependency Neighborhood)**.** *A collection of random variables $\{X_1,\ldots,X_n\}$ is said to have a dependency neighborhoods $N_i \subseteq \{1,\ldots,n\}$, $i = 1,\ldots,n$, if X_i is independent of $\{X_j\}_{j\notin N_i}$.*

Next, we present a result that gives an upper bound in the Wasserstein metric between sums of locally dependent random variables and the standard normal random variable.

Theorem 3.3.3. *[19] Let $X_1,\ldots,X_n$ be random variables with $\mathbb{E}|X_i|^4 < \infty$, $\mathbb{E}(X_i) = 0$ and $\mathrm{Var}(\sum_{i=1} X_i) = \sigma^2$. Define $W = \sum_{i=1} X_i/\sigma$. Let the collection $(X_1,\ldots,X_n)$ have dependency graph with* max *degree $D-1$. Then for $Z \sim N(0,1)$,*

$$d_W(W,Z) \leq \frac{4}{\sqrt{\pi}\sigma^2}\sqrt{D^3\sum_{i=1}^{n}\mathbb{E}|X_i|^4} + \frac{D^2}{\sigma^3}\sum_{i=1}^{n}\mathbb{E}|X_i|^3.$$

Exchangeable Pairs

We start with the definitions of an *exchangeable pair* and and a *Stein pair.*

Defintion 3.3.4. *The ordered pair* (W, W') *of random variables is called an exchangeable pair if* $(W, W') =_d (W', W)$*, where* $=_d$ *indicates equality in distribution.*

Defintion 3.3.5. *If* (W, W') *is an exchangeable pair and satisfies the relation*

$$\mathbb{E}(W' - W|W) = -\lambda W \tag{3.12}$$

for some $\lambda \in (0, 1)$*, then we call* (W, W') *a* $\lambda-$*Stein pair.*

Note from the definitions of an exchangeable pair and Stein pair that $\mathbb{E}(W) = \mathbb{E}(W') = 0$ and $\mathbb{E}(W^2) = \mathbb{E}(W'^2)$. The following result establishes an upper bound of the Wasserstein distance between an arbitrary random variable W, not necessarily a sum, and the standard random variable Z.

Theorem 3.3.4. *Suppose* (W, W') *is a* $\lambda-$*Stein pair with* $\mathbb{E}(W^2) = 1$*. If* $Z \sim N(0, 1)$*, then*

$$d_W(W, Z) \leq \sqrt{\frac{2}{\pi}\,\mathrm{Var}\left(\mathbb{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)\right)} + \frac{1}{3\lambda}\mathbb{E}|W' - W|^3$$

See [19] for proof.

Exchangeable Pair by Substitution

Generally, an exchangeable pair is constructed using the following procedure:
Let $\{X_1, \ldots, X_n\}$ be a collection of random variables and define $W = g(X_1, \ldots, X_n)$.

Let $\{X_1', \ldots, X_n'\}$ be an independent copy of $\{X_1, \ldots, X_n\}$. Let I be chosen uniformly at random from $\{1, \ldots, n\}$, independent of $\{X_i, X_i', i = 1, \ldots, n\}$. Define $W' = g(X_1, \ldots, X_{I-1}, X_I', X_{I+1}, \ldots, X_n)$. That is, to obtain W', replace X_I by X_I' in the definition of W while the other variables remain the same. Then (W, W') is an exchangeable pair.

Note that even though this procedure always yields an exchangeable pair, the relation (3.12) is not automatically satisfied and must be verified.

Chapter 4

Main Results

In this chapter, we present the main result of this thesis. We provide a condition under which pure jump subordinators have a finite fourth moment. With this condition, we establish a CLT associated with H. In the last section, we provide two examples of the main result. Among these examples is the result obtained by Griffiths [10], which involves the Gamma subordinator.

4.1 Central Limit Theorem Associated With H

In this section, we state with proof a CLT associated with

$$H = \sum_{i=1}^{m} X_i^2$$

where

$$X_i = \frac{Y_i}{\sum_{j=1}^{m} Y_j}, \quad i = 1, \dots, m$$

with $Y_1, \ldots, Y_m$ denoting the increments of $\{\tau_s : s \geq 0\}$ up to time m.

Lemma 4.1.1. *Let $\{\tau_s : s \geq 0\}$ be a subordinator with no drift. If $\int_0^\infty x^4 \Lambda(dx) < \infty$, then $\mathbb{E}(\tau_s^4) < \infty$.*

Proof. Recall that a subordinator $\{\tau_s : s \geq 0\}$ has no drift (pure-jump) if for any $\lambda \geq 0$, $s \geq 0$, the Laplace transform is given by

$$\mathbb{E}(e^{-\lambda \tau_s}) = \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \tag{4.1}$$

where Λ is the Lévy measure on $[0, \infty)$.

From this definition, the fourth moment of a subordinator with no drift can be derived by differentiating (4.1) four times and setting $\lambda = 0$. We have

$$\begin{aligned}
\mathbb{E}(e^{-\lambda \tau_s}) &= \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \\
\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}(e^{-\lambda \tau_s}) &= \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \cdot \left[-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right] \\
\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathbb{E}(e^{-\lambda \tau_s}) &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\exp\left\{-s \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \cdot \left(-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right)\right] \\
&= \left[-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right]^2 + \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx)\right\} \cdot \left[s \int_0^\infty (x^2 e^{-\lambda x})\Lambda(dx)\right] \\
\frac{\mathrm{d}^3}{\mathrm{d}\lambda^3}\mathbb{E}(e^{-\lambda \tau_s}) &= 2\left(s \int_0^\infty (x^2 e^{-\lambda x})\Lambda(dx)\right)\left(-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx\right) \\
&\quad + \left[s \int_0^\infty (x^2 e^{-\lambda x})\Lambda(dx)\right] \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x)}\Lambda(dx))\right\} \cdot \left[-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right] \\
&\quad + \exp\left\{-s \int_0^\infty (1 - e^{-\lambda x)}\Lambda(dx))\right\} \cdot \left[-s \int_0^\infty (x^3 e^{-\lambda x})\Lambda(dx)\right] \\
\frac{\mathrm{d}^4}{\mathrm{d}\lambda^4}\mathbb{E}(e^{-\lambda \tau_s}) &= 2\left[\left(-s \int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right)\left(-s \int_0^\infty (x^3 e^{-\lambda x})\Lambda(dx)\right) + \left(s \int_0^\infty (x^2 e^{-\lambda x})\Lambda(dx)\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+\left[s\int_0^\infty (x^2e^{-\lambda x})\Lambda(dx)\right]\exp\left\{-s\int_0^\infty (1-e^{-\lambda x})\Lambda(dx))\right\}\cdot\left[-s\int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right]\\
&+\exp\left\{-s\int_0^\infty (1-e^{-\lambda x})\Lambda(dx))\right\}\cdot\left[-s\int_0^\infty (x^3e^{-\lambda x})\Lambda(dx)\right]\cdot\left[-s\int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right]\\
&+\left[s\int_0^\infty (x^2e^{-\lambda x})\Lambda(dx)\right]\cdot\left[s\int_0^\infty (x^2e^{-\lambda x})\Lambda(dx)\right]\cdot\exp\left\{-s\int_0^\infty (1-e^{-\lambda x})\Lambda(dx))\right\}\\
&+\exp\left\{-s\int_0^\infty (1-e^{-\lambda x})\Lambda(dx))\right\}\cdot\left[-s\int_0^\infty (xe^{-\lambda x})\Lambda(dx)\right]\cdot\left[-s\int_0^\infty (x^3e^{-\lambda x})\Lambda(dx)\right]\\
&+\left[s\int_0^\infty (x^4e^{-\lambda x})\Lambda(dx)\right]\cdot\exp\left\{-s\int_0^\infty (1-e^{-\lambda x})\Lambda(dx))\right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}^4}{\mathrm{d}\lambda^4}\mathbb{E}(e^{-\lambda\tau_s})\Big|_{\lambda=0} &= \mathbb{E}(\tau_s^4)\\
&= s\int_0^\infty x^4\Lambda(dx) + 4s^2\left(\int_0^\infty x^3\Lambda(dx)\right)\left(\int_0^\infty x\Lambda(dx)\right) + 3s^2\left(\int_0^\infty x^2\Lambda(dx)\right)^2\\
&+ s^3\left(\int_0^\infty x^2\Lambda(dx)\right)\left(\int_0^\infty x\Lambda(dx)\right)^2
\end{aligned}
$$

From the assumptions on Λ, in particular, from $\int_0^\infty \min\{x,1\}\Lambda(dx) < \infty$, we have that $\int_0^\infty x^4\Lambda(dx) < \infty$ implies $\int_0^\infty x^k\Lambda(dx) < \infty$ for $k = 1, 2, 3$. Clearly, $\mathbb{E}(\tau_s^4) < \infty$ whenever $\int_0^\infty x^4\Lambda(dx) < \infty$. This completes the proof. □

Note that since $\int_0^\infty \min\{x,1\}\Lambda(dx) < \infty$, $\mathbb{E}(\tau_s^4) < \infty$ implies that up to the fourth moment of the increments of τ_s exist. The following theorem is the main result of this thesis.

Theorem 4.1.1. *Let $\{\tau_s : s \geq 0\}$ be a subordinator with no drift. If $\int_0^\infty x^4\Lambda(dx) < \infty$, then there are some $a, \delta \in (0,\infty)$ such that as $m \to \infty$,*

$$\sqrt{\frac{m}{a}}\left(m\sum_{i=1}^m X_i^2 - \delta\right) \xrightarrow{d} N(0,1).$$

Proof. Assume $\int_0^\infty x^4 \Lambda(dx) < \infty$.

Using the representation $X_i = \frac{Y_i}{\sum_{j=1}^m Y_j}$ for $i = 1, 2, \ldots, m$, we have

$$\begin{aligned} m \sum_{i=1}^m X_i^2 &= \frac{m^2}{(\sum_{i=1}^m Y_i)^2} \cdot \frac{1}{m} \sum_{i=1}^m Y_i^2 \\ &= \left(\frac{m^2}{(\sum_{j=1}^m Y_j)^2} - \frac{1}{b^2} + \frac{1}{b^2} \right) \cdot \frac{1}{m} \sum_{i=1}^m Y_i^2 \qquad \text{where } b = \mathbb{E}(Y_i) \neq 0 \\ &= \left(\frac{m^2}{(\sum_{j=1}^m Y_j)^2} - \frac{1}{b^2} \right) \cdot \frac{1}{m} \sum_{i=1}^m Y_i^2 + \frac{1}{b^2 m} \sum_{i=1}^m Y_i^2 \end{aligned}$$

Hence,

$$m \sum_{i=1}^m X_i^2 - \delta = \left(\frac{m^2}{(\sum_{j=1}^m Y_j)^2} - \frac{1}{b^2} \right) \cdot \frac{1}{m} \sum_{i=1}^m Y_i^2 + \frac{1}{b^2 m} \sum_{i=1}^m Y_i^2 - \delta$$

where $\delta = \frac{c}{b^2}$, $c = \mathbb{E}(Y_i^2)$.

Notice that

$$\begin{aligned} \left(\frac{m^2}{(\sum_{j=1}^m Y_j)^2} - \frac{1}{b^2} \right) &= \left(\frac{m}{\sum_{j=1}^m Y_j} - \frac{1}{b} \right) \left(\frac{m}{\sum_{j=1}^m Y_j} + \frac{1}{b} \right) \\ &= \frac{bm - \sum_{j=1}^m Y_j}{b \sum_{j=1}^m Y_j} \left(\frac{m}{\sum_{j=1}^m Y_j} + \frac{1}{b} \right) \\ &= -\frac{1}{\frac{b}{m} \sum_{j=1}^m Y_j} \left(\frac{m}{\sum_{j=1}^m Y_j} + \frac{1}{b} \right) \left(\frac{1}{m} \sum_{j=1}^m (Y_j - b) \right) \end{aligned}$$

This implies

$$m \sum_{i=1}^m X_i^2 - \delta = -\frac{\frac{1}{m} \sum_{j=1}^m Y_j^2}{b \cdot \frac{1}{m} \sum_{j=1}^m Y_j} \cdot \left(\frac{1}{\frac{1}{m} \sum_{j=1}^m Y_j} + \frac{1}{b} \right) \frac{1}{m} \sum_{j=1}^m (Y_j - b) + \frac{1}{b^2} \cdot \frac{1}{m} \sum_{i=1}^m (Y_i^2 - c)$$

By the Weak Law of Large Numbers (WLLN),

$$-\frac{\frac{1}{m}\sum_{j=1}^{m} Y_j^2}{b \cdot \frac{1}{m}\sum_{j=1}^{m} Y_j} \cdot \left(\frac{1}{\frac{1}{m}\sum_{j=1}^{m} Y_j} + \frac{1}{b} \right) = A \xrightarrow{\mathbb{P}} -\frac{2c}{b^3} = \varepsilon_1$$

and let $\varepsilon_2 = \frac{1}{b^2}$.

By Slutsky's theorem, we have that

$$\begin{aligned} \sqrt{\frac{m}{a}} \left(m \sum_{i=1}^{m} X_i^2 - \delta \right) &\approx \varepsilon_1 \cdot \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \left(\frac{Y_j - b}{\sqrt{a}} \right) + \varepsilon_2 \cdot \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{Y_i^2 - c}{\sqrt{a}} \right) \\ &= \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{\varepsilon_2 Y_i^2 + \varepsilon_1 Y_i - \varepsilon_2 c - \varepsilon_1 b}{\sqrt{a}} \right) \end{aligned}$$

where $a = \text{Var}(\varepsilon_2 Y_i^2 + \varepsilon_1 Y_i)$.

Set

$$E_i = \frac{\varepsilon_2 Y_i^2 + \varepsilon_1 Y_i - \varepsilon_2 c - \varepsilon_1 b}{\sqrt{a}}$$

Then E_i's are iid with $\mathbb{E}(E_i) = 0$ and $\text{Var}(E_i) = 1$ and we can apply Feller-Lindeberg CLT to

$$W = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} E_i$$

to conclude that $W \xrightarrow{d} N(0, 1)$ as $m \to \infty$. □

If in particular we have that $\int_0^\infty x^6 \Lambda(dx) < \infty$, then we can apply Stein's method for Sum of Independent random variables (Theorem 3.3.2) to get

$$d_W(W, Z) \leq \frac{3}{m^{\frac{3}{2}}} \sum_{i=1}^{m} \mathbb{E}|E_i|^3$$

where $Z \sim N(0,1)$.

As $m \to \infty$, we have $d_W(W,Z) \to 0$ and we can conclude that $W \xrightarrow{d} N(0,1)$. Realize that unlike the Feller-Lindeberg CLT, applying Stein's method provides an additional information in the form of an upper bound of the error between W and Z. That is, the error between W and Z can be controlled.

4.1.1 Alternate Proof

We give an alternative proof for Theorem 4.1.1 by applying Stein's method directly.

Proof. Let

$$W = \sqrt{\frac{m}{a}} \left(m \sum_{i=1}^{m} X_i^2 - \delta \right)$$

We define $a, b, c, \delta, \varepsilon_1, \varepsilon_2$ as in the previous proof. As before, we can write W as

$$W = \sqrt{\frac{m}{a}} \left(m \sum_{i=1}^{m} X_i^2 - \delta \right) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{\varepsilon_2 Y_i^2 + AY_i - \varepsilon_2 c - Ab}{\sqrt{a}} \right)$$

Let

$$U = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{\varepsilon_2 Y_i^2 + \varepsilon_1 Y_i - \varepsilon_2 c - \varepsilon_1 b}{\sqrt{a}} \right)$$

Then realize that U is the sum of independent random variables with $\mathbb{E}(U) = 0$ and $\mathrm{Var}(U) = 1$, and by Stein's method for sum of independent random variables, we have that U is approximately standard normal. This means, as $m \to \infty$,

$$\mathbb{E}(Uf(U) - f'(U)) \approx 0$$

where f is any twice differentiable function satisfying $\|f\|_\infty \leq 1$, $\|f'\|_\infty \leq \sqrt{\frac{2}{\pi}}$ and $\|f''\|_\infty \leq 2$.

Note that

$$\begin{aligned}|\mathbb{E}(Wf(W)) - \mathbb{E}(Uf(U))| &\leq |E(Wf(W))| + |\mathbb{E}(Uf(U))| \\ &\leq |\mathbb{E}(W)| + |\mathbb{E}(U)| \quad \text{since } \|f\|_\infty \leq 1 \\ &\to 0 \quad \text{as } m \to \infty \text{ since } \mathbb{E}(W) \to 0 \text{ as } m \to \infty.\end{aligned}$$

By Taylor's theorem,

$$f'(W) = f'(U) + r$$

where $r = f''(\tilde{U})(W - U)$, $\tilde{U}$ is between W and U.

Thus

$$\begin{aligned}|\mathbb{E}(f'(W) - f'(U))| &= |\mathbb{E}(f''(\tilde{U})(W - U))| \\ &\leq 2|\mathbb{E}(W - U)| \quad \text{since } \|f''\|_\infty \leq 2 \\ &= 2|\mathbb{E}(W) - \mathbb{E}(U)| \\ &\to 0 \quad \text{as } m \to \infty.\end{aligned}$$

So we have

$$\begin{aligned}|\mathbb{E}(Wf(W) - f'(W)) - \mathbb{E}(Uf(U) - f'(U))| &\leq |\mathbb{E}(Wf(W)) - \mathbb{E}(Uf(U))| + |\mathbb{E}(f'(W) - f'(U))| \\ &\to 0 \quad \text{as } m \to \infty.\end{aligned}$$

And we get

$$\mathbb{E}(Wf(W) - f'(W)) \approx \mathbb{E}(Uf(U) - f'(U)) \approx 0$$

This implies that

$$\mathbb{E}(Wf(W)) \approx \mathbb{E}(f'(W))$$

and we conclude by Stein's method that $W \xrightarrow{d} N(0,1)$. $\square$

In the remainder of this section, we give examples of Theorem 4.1.1 by considering the Gamma subordinator and the Generalized Gamma subordinator.

4.2 Examples with some subordinators

4.2.1 Gamma Subordinators

Recall from Chapter 2 that a subordinator $\{\tau_s : s \geq 0\}$ is a Gamma subordinator if its Lévy measure is

$$\Lambda(dx) = x^{-1}e^{-x}dx, \quad x > 0.$$

As an illustration of Theorem 4.1.1, consider $(X_1, X_2, \ldots, X_m) \sim \text{Dir}(1, 1, \ldots, 1)$. Then each X_i is such that

$$X_i = \frac{Y_i}{\sum_{i=1}^{m} Y_i}$$

where $Y_i \sim \text{Exp}(1)$ for $i = 1, 2, \ldots, m$.

Moreover, $\int_0^\infty x^4 x^{-1} e^{-x} dx = 3! = 6 < \infty$.

Using this representation of X_i, we have

$$\begin{aligned}\sum_{i=1}^{m} X_i^2 &= \sum_{i=1}^{m} \left(\frac{Y_i}{\sum_{j=1}^{m} Y_j} \right)^2 \\ &= \sum_{i=1}^{m} \left(\frac{Y_i}{\frac{m}{m}\sum_{j=1}^{m} Y_j} \right)^2 \\ &= \frac{1}{\left(\frac{1}{m}\sum_{j=1}^{m} Y_j\right)^2} \cdot \frac{1}{m^2} \sum_{i=1}^{m} Y_i^2\end{aligned}$$

Hence

$$\begin{aligned}m\sum_{i=1}^{m} X_i^2 &= \frac{m^2}{(\sum_{i=1}^{m} Y_i)^2} \cdot \frac{1}{m}\sum_{i=1}^{m} Y_i^2 \\ &= \left(\frac{m^2}{(\sum_{i=1}^{m} Y_i)^2} - 1 + 1 \right) \cdot \frac{1}{m}\sum_{i=1}^{m} Y_i^2 \\ &= \left(\frac{m^2}{(\sum_{i=1}^{m} Y_i)^2} - 1 \right) \cdot \frac{1}{m}\sum_{i=1}^{m} Y_i^2 + \frac{1}{m}\sum_{i=1}^{m} Y_i^2\end{aligned}$$

And we have,

$$\sqrt{\frac{m}{4}}\left(m\sum_{i=1}^{m} X_i^2 - 2 \right) = \sqrt{\frac{m}{4}}\left(\frac{m^2}{(\sum_{i=1}^{m} Y_i)^2} - 1 \right)\frac{1}{m}\sum_{i=1}^{m} Y_i^2 + \sqrt{\frac{m}{4}}\left(\frac{1}{m}\sum_{i=1}^{m}(Y_i^2 - 2) \right)$$

Note that

$$\begin{aligned}\frac{m^2}{(\sum_{i=1}^{m} Y_i)^2} - 1 &= \left(\frac{m}{\sum_{i=1}^{m} Y_i} - 1 \right)\left(\frac{m}{\sum_{i=1}^{m} Y_i} + 1 \right) \\ &= -\left(1 + \frac{1}{\frac{1}{m}\sum_{i=1}^{m} Y_i} \right)\left(\frac{\sum_{i=1}^{m} Y_i - m}{\sum_{i=1}^{m} Y_i} \right) \\ &= \frac{-\left(1 + \frac{1}{\frac{1}{m}\sum_{i=1}^{m} Y_i} \right)}{\frac{1}{m}\sum_{i=1}^{m} Y_i}\left(\frac{1}{m}\sum_{i=1}^{m}(Y_i - 1) \right)\end{aligned}$$

So that

$$\sqrt{\frac{m}{4}}\left(m\sum_{i=1}^{m}X_i^2-2\right)=\frac{-\left(1+\frac{1}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\right)}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\cdot\frac{1}{m}\sum_{i=1}^{m}Y_i^2\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{Y_i-1}{\sqrt{4}}\right)\right)$$
$$+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{Y_i^2-2}{\sqrt{4}}\right)$$

Consider

$$\frac{-\left(1+\frac{1}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\right)}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\cdot\frac{1}{m}\sum_{i=1}^{m}Y_i^2$$

and note that $\frac{1}{m}\sum_{i=1}^{m}Y_i\xrightarrow{\text{P}}1$ and $\frac{1}{m}\sum_{i=1}^{m}Y_i^2\xrightarrow{\text{P}}2$.

Hence

$$\frac{-\left(1+\frac{1}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\right)}{\frac{1}{m}\sum_{i=1}^{m}Y_i}\cdot\frac{1}{m}\sum_{i=1}^{m}Y_i^2\xrightarrow{\text{P}}-4$$

Thus by Slutsky's theorem,

$$\sqrt{\frac{m}{4}}\left(m\sum_{i=1}^{m}X_i^2-2\right)\approx\frac{-4}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{Y_i-1}{\sqrt{4}}\right)+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{Y_i^2-2}{\sqrt{4}}\right)$$
$$=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{Y_i^2-4Y_i-(-2)}{\sqrt{4}}\right)$$

Let $E_i=\frac{Y_i^2-4Y_i-(-2)}{\sqrt{4}}$, then $E_1,E_2,\ldots,E_m$ are iid (since they are functions of Y_i's) with $\mathbb{E}(E_i)=0$ and $\text{Var}(E_i)=1$. Apply Theorem 3.3.2 and conclude that

$$\sqrt{\frac{m}{4}}\left(m\sum_{i=1}^{m}X_i^2-2\right)\xrightarrow{d}N(0,1).$$

Computation of Variance

It is observed from above that a scale factor of $\sqrt{\frac{m}{4}}$ is used in obtaining a standard normal distribution. The 4 represents the variance of $\sqrt{m}\,(m\sum_{i=1}^{m} X_i^2 - 2)$ and this can be derived in at least two ways. Two methods are presented here.

Method I (Slutsky's Approach)

It is observed that by applying Slutsky's theorem, we have that

$$\begin{aligned}\sqrt{m}\left(m\sum_{i=1}^{m} X_i^2 - 2\right) &\approx \frac{-4}{\sqrt{m}}\sum_{i=1}^{m}(Y_i - 1) + \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(Y_i^2 - 2\right) \\ &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(Y_i^2 - 4Y_i + 2\right)\end{aligned}$$

With this, the variance of $\sqrt{m}\,(m\sum_{i=1}^{m} X_i^2 - 2)$ can be approximated by computing the variance of $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(Y_i^2 - 4Y_i + 2)$.

Now,

$$\begin{aligned}\text{Var}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(Y_i^2 - 4Y_i + 2\right)\right) &= \frac{1}{m}\cdot m\,\text{Var}(Y_i^2 - 4Y_i) \\ &= \text{Var}(Y_i^2 - 4Y_i) \\ &= \text{Var}(Y^2) + 16\,\text{Var}(Y) - 2(4)\,\text{Cov}(Y^2, Y)\end{aligned}$$

Here, $Y \sim \text{Exp}(1)$ hence $f_Y(y) = e^{-y}, \quad y > 0$

Computing the first four moments of Y, we have

$$\mathbb{E}(Y) = \int_0^\infty ye^{-y}dy = \Gamma(2) = 1$$

$$\mathbb{E}(Y^2) = \int_0^\infty y^2e^{-y}dy = \Gamma(3) = 2$$

$$\mathbb{E}(Y^3) = \int_0^\infty y^3e^{-y}dy = \Gamma(4) = 6$$

$$\mathbb{E}(Y^4) = \int_0^\infty y^4e^{-y}dy = \Gamma(5) = 24$$

Thus

$$\text{Var}(Y^2) = \mathbb{E}(Y^4) - \mathbb{E}(Y^2)^2 = 24 - 2^2 = 20$$

$$\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = 2 - 1 = 1$$

and

$$\text{Cov}(Y^2, Y) = \mathbb{E}(Y^3) - \mathbb{E}(Y^2)\mathbb{E}(Y) = 6 - 2(1) = 4$$

Putting the results together, we have

$$\text{Var}\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(Y_i^2 - 4Y_i + 2\right)\right) = 20 + 16(1) - 2(4)(4) = 4$$

Method II (Direct Calculation)

Using the fact that $(X_1, X_2, \ldots, X_m) \sim \text{Dir}(1, 1, \ldots, 1)$, the variance of $\sqrt{m}\left(m \sum_{i=1}^{m} X_i^2 - 2\right)$ can be calculated directly as follows:

$$\text{Var}\left(\sqrt{m}\left(m\sum_{i=1}^{m} X_i^2 - 2\right)\right) = m^3 \,\text{Var}\left(\sum_{i=1}^{m} X_i^2\right)$$

Note that

$$\text{Var}\left(\sum_{i=1}^{m} X_i^2\right) = \sum_{i=1}^{m} \text{Var}\left(X_i^2\right) + \sum_{i\neq j} \text{Cov}\left(X_i^2, X_j^2\right)$$

For $(X_1, \ldots, X_m) \sim \text{Dir}(\alpha_1, \ldots, \alpha_m)$ we have,

$$\begin{aligned}
\mathbb{E}(X_j^4) &= \int \cdots \int x_j^4 \cdot \frac{\Gamma\left(\sum_{i=1}^{m} \alpha_i\right)}{\prod_{i=1}^{m} \Gamma(\alpha_i)} \prod_{i=1}^{m} x_i^{\alpha_i - 1} dx_1 \cdots dx_m \\
&= \frac{\Gamma\left(\sum_{i=1}^{m} \alpha_i\right)}{\prod_{i=1}^{m} \Gamma(\alpha_i)} \int \cdots \int \prod_{\substack{i=1 \\ i\neq j}}^{m} x_i^{\alpha_i - 1} \cdot x_j^{\alpha_j + 4 - 1} dx_1 \cdots dx_m \\
&= \frac{\Gamma\left(\sum_{i=1}^{m} \alpha_i\right)}{\prod_{i=1}^{m} \Gamma(\alpha_i)} \cdot \frac{\prod_{\substack{i=1 \\ i\neq j}}^{m} \Gamma(\alpha_i)\Gamma(\alpha_j + 4)}{\Gamma\left(\sum_{i=1}^{m} \alpha_i + 4\right)} \\
&= \frac{(\alpha_j + 3)(\alpha_j + 2)(\alpha_j + 1)\alpha_j}{\left(\sum_{i=1}^{m} \alpha_i + 3\right)\left(\sum_{i=1}^{m} \alpha_i + 2\right)\left(\sum_{i=1}^{m} \alpha_i + 1\right)\left(\sum_{i=1}^{m} \alpha_i\right)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}(X_j^2) &= \frac{\Gamma\left(\sum_{i=1}^{m} \alpha_i\right)}{\prod_{i=1}^{m} \Gamma(\alpha_i)} \cdot \frac{\prod_{\substack{i=1 \\ i\neq j}}^{m} \Gamma(\alpha_i)\Gamma(\alpha_j + 2)}{\Gamma\left(\sum_{i=1}^{m} \alpha_i + 2\right)} \\
&= \frac{(\alpha_j + 1)\alpha_j}{\left(\sum_{i=1}^{m} \alpha_i + 1\right)\left(\sum_{i=1}^{m} \alpha_i\right)}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(X_j^2 X_k^2) &= \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} \cdot \frac{\prod_{\substack{i=1\\ i\neq j\\ i\neq k}}^m \Gamma(\alpha_i)\Gamma(\alpha_j+2)\Gamma(\alpha_k+2)}{\Gamma\left(\sum_{i=1}^m \alpha_i + 4\right)} \\
&= \frac{(\alpha_j+1)(\alpha_j)(\alpha_k+1)\alpha_k}{\left(\sum_{i=1}^m \alpha_i + 3\right)\left(\sum_{i=1}^m \alpha_i + 2\right)\left(\sum_{i=1}^m \alpha_i + 1\right)\left(\sum_{i=1}^m \alpha_i\right)}
\end{aligned}
$$

Since $\alpha_1 = \ldots = \alpha_m = 1$ in our case, we have

$$\mathbb{E}(X_j^4) = \frac{24}{m(m+1)(m+2)(m+3)}$$

$$\mathbb{E}(X_j^2) = \frac{2}{m(m+1)}$$

and

$$\mathbb{E}(X_j^2 X_k^2) = \frac{4}{m(m+1)(m+2)(m+3)}$$

From which we have

$$
\begin{aligned}
\mathrm{Var}(X_i^2) &= \mathbb{E}(X_i^4) - \mathbb{E}(X_i^2)^2 \\
&= \frac{20m^2 + 4m - 24}{m^2(m+1)^2(m+2)(m+3)} \geq 0 \quad \text{since } m \geq 1
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(X_j^2, X_k^2) &= \mathbb{E}(X_j^2 X_k^2) - \mathbb{E}(X_j^2)\mathbb{E}(X_k^2) \\
&= \frac{-16m - 24}{m^2(m+1)^2(m+2)(m+3)}
\end{aligned}
$$

Now,

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{m} X_i^2\right) &= \frac{(20m^2+4m-24)m}{m^2(m+1)^2(m+2)(m+3)} + \frac{(-16m-24)m(m-1)}{m^2(m+1)^2(m+2)(m+3)} \\ &= \frac{4m^2-4m}{m(m+1)^2(m+2)(m+3)}\end{aligned}$$

Finally, we have

$$\begin{aligned}\operatorname{Var}\left(\sqrt{m}\left(m\sum_{i=1}^{m} X_i^2 - 2\right)\right) &= m^3 \operatorname{Var}\left(\sum_{i=1}^{m} X_i^2\right) \\ &= \frac{4m^5-4m^4}{m(m+1)^2(m+2)(m+3)} \\ &= \frac{4-\frac{4}{m}}{1(1+\frac{1}{m})^2(1+\frac{2}{m})(1+\frac{3}{m})} \to 4 \quad \text{as } m \to \infty\end{aligned}$$

4.2.2 Generalized Gamma Subordinators

From Chapter 2, we know that a subordinator $\{\rho_s : s \geq 0\}$ is a generalized Gamma subordinator with scale parameter one if its Lévy measure is given by

$$\Lambda(dx) = \Gamma(1-\alpha)^{-1} x^{-(1+\alpha)} e^{-x} dx, \quad x > 0.$$

Now,

$$\begin{aligned}\int_0^\infty x^4 \Lambda(dx) &= \int_0^\infty x^4 \Gamma(1-\alpha)^{-1} x^{-(1+\alpha)} e^{-x} dx \\ &= \frac{\Gamma(4-\alpha)}{\Gamma(1-\alpha)} \\ &< \infty\end{aligned}$$

and we can conclude from Theorem 4.1.1 that

$$\sqrt{\frac{m}{a}}\left(m\sum_{i=1}^{m}X_i^2-\delta\right)\xrightarrow{d}N(0,1)$$

for $a,\delta\in(0,\infty)$ as $m\to\infty$.

Chapter 5

Conclusion and Future Directions

5.1 Summary

In this thesis, we presented the importance of the statistic $H = \sum_{i=1}^{m} X_i^2$ in various fields of study and established a Central Limit Theorem associated with H under a moment condition. We assumed $X_i, i = 1, 2, \ldots, m$ is a self-normalized random measure so that it has the representation $X_i = Y_i / \sum_{j=1}^{m} Y_j$, where $Y_i, i = 1, 2, \ldots, m$ are the increments of the subordinator $\{\tau_s : s \geq 0\}$ up to time m. We showed that our result obtained includes as a special case the result obtained by Griffiths [10] and also extends to the generalized Gamma subordinator, which is a new result.

5.2 Future Directions

In our result, we assumed the fourth moment of the subordinator exists. As a result, subordinators such as the stable subordinators are not included in our result. It may be of interest to investigate whether or not this moment assumption can be relaxed so that our result extends to more classes of subordinators.

In Feng [7], fluctuation theorems are presented for $\{\varphi_r(\mathbf{x}) : r \geq 2\}$ (2.4). It may be of interest to investigate whether Stein's method could be applied to this infinite dimensional case.

Apart from CLT, there are other limiting distributions such as the Large Deviation Principle (LDP). For instance, in [6], large deviation principles are established for the Fleming-Viot processes with neutral mutation and selection. One may be interested in establishing LDP for H.

Bibliography

[1] Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

[2] Arratia, R., Goldstein, L., Gordon, L., et al. (1989). Two moments suffice for poisson approximations: the chen-stein method. *The Annals of Probability*, 17(1):9–25.

[3] Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory*. Springer Science & Business Media.

[4] Brix, A. (1999). Generalized gamma measures and shot-noise cox processes. *Advances in Applied Probability*, 31(4):929–953.

[5] Chen, L. H., Goldstein, L., and Shao, Q.-M. (2010). *Normal approximation by Stein's method*. Springer Science & Business Media.

[6] Dawson, D. A. and Feng, S. (1998). Large deviations for the fleming–viot process with neutral mutation and selection. *Stochastic processes and their applications*, 77(2):207–232.

[7] Feng, S. (2010). *The Poisson-Dirichlet distribution and related topics: models and asymptotic behaviors*. Springer Science & Business Media.

[8] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230.

[9] Fristedt, B. E. and Gray, L. F. (2013). *A modern approach to probability theory*. Springer Science & Business Media.

[10] Griffiths, R. (1979). On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology*, 15(1):140–158.

[11] Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

[12] Kingman, J. F. (1975). Random discrete distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(1):1–15.

[13] Laine, C. R. (1995). The herfindahl-hirschman index: a concentration measure taking the consumer's point of view. *The Antitrust Bulletin*, 40(2):423–432.

[14] Lijoi, A., Mena, R. H., and Prünster, I. (2007). Controlling the reinforcement in bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):715–740.

[15] Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443.

[16] Perman, M. (1993). Order statistics for jumps of normalised subordinators. *Stochastic Processes and their Applications*, 46(2):267–281.

[17] Regazzini, E., Lijoi, A., Prünster, I., et al. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31(2):560–585.

[18] Rosenberg, N. A. and Jakobsson, M. (2008). The relationship between homozygosity and the frequency of the most frequent allele. *Genetics*, 179(4):2027–2036.

[19] Ross, N. (2011). Fundamentals of stein's method. *Probability Surveys*, 8:210–293.

[20] Ross, S. (1996a). *Stochastic processes*. Wiley series in probability and statistics: Probability and Statistics. Wiley.

[21] Ross, S. (1996b). *Stochastic processes*. Wiley series in probability and statistics: Probability and Statistics. Wiley.

[22] Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(4148):688.

[23] Sriperumbudur, B. K., Fukumizu, K., Gretton, A., Schölkopf, B., and Lanckriet, G. R. (2009). On integral probability metrics,\phi-divergences and binary classification. *arXiv preprint arXiv:0901.2698*.

[24] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California.

[25] Trippa, L. and Favaro, S. (2012). A class of normalized random measures with an exact predictive sampling scheme. *Scandinavian Journal of Statistics*, 39(3):444–460.