ON THE CENTRAL LIMIT THEOREM OF RANDOM SUM OF INDEPENDENT RANDOM VARIABLES USING STEINS METHOD

ON THE CENTRAL LIMIT THEOREM OF RANDOM SUM OF INDEPENDENT RANDOM VARIABLES USING STEIN'S METHOD

By YICHEN HUANG, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2024)

Hamilton, Ontario, Canada (Mathematics and Statistics)

TITLE:	ON THE CENTRAL LIMIT THEOREM OF RANDOM SUM OF INDEPENDENT RANDOM VARIABLES USING STEIN'S METHOD
AUTHOR:	Yichen Huang B.Sc. (Mathematics and Statistics), McMaster University, Hamilton, Canada
SUPERVISOR:	Shui Feng
NUMBER OF PAGES:	xi, 41

Abstract

Stein's method provides a powerful tool for quantifying the distance between a random variable and a normal distribution without relying on the characteristic function as in the traditional method. The traditional characteristic function method is often challenging when dealing with complicated random variables. Stein's method was first proposed to solve the problem of normal approximation that the characteristic function cannot solve. This method provides a powerful tool to give an upper bound between any random variable and the normal distribution, which can been seen as a speed of convergence to the normal distribution. The core idea of Stein's method is to construct a differential equation for the target random variable and analyze its solution. In this thesis, I use Stein's method to quantify the error bound of the standardized random sum of independent random variables with respect to the approximation of the normal distribution to establish the conditions that need to be met for the Central Limit Theorem to hold. Additionally, the error term is bounded using the Wasserstein distance, which demonstrates Stein's method's effectiveness in controlling approximation errors by bounding the expectation of the Stein's Identity. In addition, the results of Charles Stein's first paper, published in 1972, which describes the Stein method in detail, are also given in the paper. Stein's paper gives error bound for the sum of dependent variable sequences and the standard normal

under certain conditions.

Acknowledgements

I would first like to express my highest respect and gratitude to my supervisor, Dr. Shui Feng. Before I started my graduate studies, Dr. Feng gave me the opportunity to learn theoretical statistics. During my graduate studies, Dr. Feng tried to teach me how to think about some mathematical ideas and methods. Despite my frequent simple questions, my supervisor always took the time to patiently explain concepts in detail. I am very grateful to Dr. Feng for his great help during my one-and-a-half years of study.

I would also like to thank Dr. Paguyo and Dr. Forman. Dr. Paguyo provided many ideas and helped me when I was in trouble. He is a lecturer who is very willing to help students. Dr. Forman's engaging teaching style has helped clarify complicated theorems, and his enthusiasm for statistics is inspiring.

In addition, I would like to thank my parents for their care and support throughout my life. They always give me advice but still support my final decisions. I also wish my grandparents a lifetime of happiness and well-being.

Finally, I would like to express my heartfelt gratitude to my girlfriend, Yi Wan. She has been a constant source of encouragement in my studies and has also supported me in my daily life. I am incredibly fortunate to have her by my side.

Table of Contents

Notation, Definitions, and Abbreviations

Notation

- Ω: The sample space in a probability space.
- $\mathcal{F}$: The sigma-algebra in a probability space.
- $\mathbb{P}$: The probability measure in the given probability space.
- $\mathbb{P}(A)$: The probability of event A occurring.
- $\mathbb{E}[X]$: The expected value (mean) of a random variable X.
- $\mathrm{Var}(X)$: The variance of a random variable X.
- $d_W(X, Y)$: The Wasserstein distance between two probability distributions X and Y.
- $d_K(X, Y)$: The Kolmogorov distance between two probability distributions X and Y.
- Lip: Lipschitz functions.

- $\Phi(x)$: The cumulative distribution function (CDF) of the standard normal distribution.
- $\mathcal{N}(0,1)$: The standard normal distribution.
- $X_n \xrightarrow{d} X$: Convergence in distribution.
- **RV**: Random Variable
- **LD1**: Local Dependence conditions 1

Definitions

Classical Central Limit Theorem (CLT): It states that the standardized sum of independent and identically distributed random variables with finite variance converges in distribution to a standard normal distribution as the number of variables approaches infinity.

Wasserstein Distance: A probability metric of the difference between two probability distributions, often used to compare the distribution of sums of random variables to a normal distribution in Stein's method.

Kolmogorov Distance: A probability metric used to quantify the maximal difference between the cumulative distribution functions of two probability distributions.

Total Variance Distance: A probability metric defined as the maximum difference in probabilities they assign to the same event. We use the total variation metric for approximation by discrete distributions.

Weakly Dependent Random Variables: A sequence of random variables where the dependence between two variables weakens when the difference between their indices is greater.

Convergence in distribution: a sequence of random variables X_n is said to converge in distribution to a random variable X, if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for every continuity point x of the cumulative distribution function F_X of X.

Absolutely continuous functions: A function $f : [a, b] \to \mathbb{R}$ is absolutely continuous if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that, for any collection of non-overlapping subintervals $[a_1, b_1], [a_2, b_2], \ldots, [a_i, b_i] \subset [a, b]$, if

$$\sum_{j=1}^{i} (b_j - a_j) < \delta,$$

then we have

$$\sum_{j=1}^{i} |f(b_j) - f(a_j)| < \epsilon.$$

Chapter 1

Introduction

Statistical analysis becomes easier if, given a collection of samples, we can tell approximately what distribution these samples follow. The Central Limit Theorem plays an important role in probability by giving an approximate distribution. This fundamental probability theory describes the asymptotic behavior of the sum of properly scaled random variables. It states that under the conditions that every random variable has the same finite expectation and variance, then as the number of random variables grows, their standardized sum converges in distribution to a normal distribution, regardless of the original distribution of random variables. Specifically, for a sequence of i.i.d random variable X_i with finite mean μ and variance σ^2, the standardized sum

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)$$

converges in distribution to the standard normal distribution $N(0, 1)$ as $n \to \infty$. One of the classical methods of proving the central limit theorem is utilizing the characteristic function. Given a random variable X, its characteristic function is defined as:

$$\varphi_X(t) = \mathbb{E}[e^{itX}],$$

where $t \in \mathbb{R}$, i is an imaginary unit, and $\mathbb{E}$ denotes the expectation. It is known that the distribution of a random variable is uniquely determined by its characteristic function. Namely, different distributions correspond to different characteristic functions. The idea of using characteristic functions to prove the central limit theorem is first to compute the characteristic functions of the sum of independently and identically distributed random variables and then use the convergence theorem in complex analysis to prove that this characteristic function converges to the characteristic function of the normal distribution.

The Stein method introduced by Charles Stein in 1972 is another idea for proving the central limit theorem, which provides a more general framework for proving the central limit theorem. In other words, the core idea of Stein's method is to transform the problem of approximating the limit between different distributions into a problem of solving differential equations by constructing Stein's differential equation. Considering the construction is different, the Stein Method can handle more complicated variables, such as non-independent and non-identical random variables. In addition, the Stein Method allows us to find a convergence rate when dealing with normal distribution approximation problems, i.e., how fast the target distribution will converge to the normal distribution. In addition to its powerful application in approximating normal distributions, the Stein method has been successfully extended to other essential probability distributions. These extensions include the approximation of Poisson and exponential and geometric distributions. In the Poisson approximation, using size-bias coupling, Stein's method provides an efficient error bound for Poisson approximations, which is widely used in situations with dependence structures or local dependence. Stein's method uses equilibrium coupling to approximate exponential distributions. The approximation of geometric distributions mainly involves discrete equilibrium coupling, which is suitable for handling the approximation requirements of discrete random variables. These applications demonstrate Stein's method's versatility and flexibility, making it a powerful tool in many statistics and probability theory scenarios. One can find how Stein's method uses these different coupling techniques to control distributional approximations in the paper [10].

In this thesis, I will use the Stein Method to show that a random sum of independent random variables will converge to a normal distribution under some conditions. A **random sum** [6] is a sum of random variables such that the number of terms is itself a random variable. It is typically referred to as:

$$S_\xi = X_1 + X_2 + \cdots + X_\xi,$$

where ξ is a non-negative random variable representing the number of terms in the sum, and $X_1, X_2, \ldots$ is a sequence of independent random variables representing each term in the sum. If the random sum ξ is deterministic, then the problem degenerates into the original problem of summing random variables. In particular, X_i can be independent of ξ, or there can be some dependency. In our case, we set X_i and ξ as independent. In the study of the Central Limit Theorem (CLT) for random sums, many key results have extended the applicability of the classical CLT. One fundamental theorem proves a standardized random sum still converges to the standard normal distribution $N(0, 1)$. Specifically, let $X_1, X_2, \ldots$ be an independent and identically distributed(i.i.d.) sequence of random variables with mean zero and finite variance σ^2, and let the random index $\tau(t)$ satisfy $\frac{\tau(t)}{t} \xrightarrow{p} \theta$ (where $0 < \theta < \infty$). Then, the theorem states that the standardized random sums

$$\frac{S_{\tau(t)}}{\sigma\sqrt{\tau(t)}} \quad \text{and} \quad \frac{S_{\tau(t)}}{\sigma\sqrt{\theta t}}$$

both converge in distribution to the standard normal distribution $N(0,1)$ as $t \to \infty$. This result generalizes the classical CLT to cases with random indices. One can find a more detailed proof and related theorems in [7]. Random sums have many practical applications, particularly in finance, insurance, queueing theory, and bio-statistics. [6] In insurance actuarial science, total claim models are often modelled as random sums. Suppose ξ represents the number of claims occurring in a given period, and X_i represents the amount of each claim. Then, the total claim amount S_ξ can be expressed as a random sum. In queueing theory, random sums can describe waiting times or service times. Considering a service system, ξ represents the number of customers arriving during a service period, and X_i represents the service time for each customer. The total service time S_ξ can be expressed as a random sum. In bio-statistics, especially in gene mutation and DNA sequence analysis, random sums are used to model the total number of mutation events. Suppose ξ denotes the number of mutations occurring in a gene sequence and X_i denotes the length of each mutation. The total length affected S_ξ is a random sum. The model can be used to assess the effect of mutations on functional regions of the genome.

Chapter 2

Normal Approximation

2.1 Classical Central Limit Theorem

The history of the CLT goes back to Laplace in the late 18th century, who used it as a tool to solve other mathematical problems [8]. He dealt with a sequence of games, each with two possible outcomes, "win" and "lose". Laplace first studied the limit problem for sums of binomial distributions, and a generalization of this problem gave birth to the earliest CLT, the study of sums of a set of independent and identically distributed Bernoulli variables. After this, Laplace generalized the CLT. He found that the standardized sum of any independent identically distributed random variables converges to the standard normal distribution. Lyapunov [11] generalized the CLT in 1901, relaxing the assumption of identical distributions by allowing random variables to be distributed differently but requiring that the third-order moments of each random variable must not be too large to affect the convergence. This work set the stage for later developments by Lindeberg in 1922 and Feller, who formulated the Lindeberg–Feller CLT [4]. The Lindeberg–Feller CLT expands the applicability of the theorem by removing the identically distributed assumption and introducing the Lindeberg condition. This condition ensures that the influence of any individual term in the sum does not dominate as the number of terms grows. As long as these extreme-valued increments do not have a dominant effect when normalized by the cumulative variance, the central limit theorem still holds. In this chapter, we first discuss the classical CLT, which concerns a set of i.i.d. random variables.

Given a standard normally distributed random variable X, recall that the probability density function(PDF) of X is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R},$$

and its cumulative distribution function(CDF) is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Let Y_i be another sequence of i.i.d. random variables. To demonstrate that Y_i converges in distribution to X, one of the approaches is to directly compute the characteristic function of Y_i and observe whether it has the form of $\varphi_Z(t) = e^{-\frac{t^2}{2}}$ as $n \to \infty$, the characteristic function of an $N(0,1)$ variable. This approach also proves the classical CLT.

Theorem 2.1.1 [2] Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with finite mean μ and variance σ^2. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ denotes a standard normal variable.

Proof: The characteristic function of X_i is given by $\varphi_X(t) = \mathbb{E}[e^{itX_i}]$. Since the random variables $X_1, X_2, \ldots, X_n$ are independent, the characteristic function of the sum of these random variables is

$$\varphi_{Z_n}(t) = \left(\varphi_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Using the Taylor expansion of $\varphi_X(t)$ around $t = 0$ and Substituting this expansion into the expression for $\varphi_{Z_n}(t)$, we must get

$$\varphi_{Z_n}(t) = \left(1 + i\frac{t\mu}{\sigma\sqrt{n}} - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

As $n \to \infty$, the term $i\frac{t\mu}{\sigma\sqrt{n}}$ tends to 0, and the expression simplifies to

$$\varphi_{Z_n}(t) \to e^{-t^2/2},$$

which is the characteristic function of the standard normal distribution $N(0,1)$. Recall that the characteristic function is precisely the Fourier transform of the PDF. From it, one can derive the CDF or the PDF. The invertibility of the Fourier transform also guarantees the uniqueness of the characteristic functions.

2.2 Probability Metrics

We often need to measure the difference between two probability distributions in probability theory. We will focus on the following three metrics: Total Variation

Distance, Kolmogorov Distance, and Wasserstein Distance. The biggest difference between these three probability metrics can be derived from the definition of 'distance', specifically in choosing the set of test functions. For any two probability measures μ and ν, We define the probability metric in the following form:

$$d_{\mathcal{H}}(\mu, \nu) = \sup_{g \in \mathcal{H}} \left| \int g(x) d\mu(x) - \int g(x) d\nu(x) \right|,$$

where $\mathcal{H}$ is some family of test functions. The Kolmogorov Distance measures how large the difference between two different distributions is, so the test functions are exactly the indicator functions $\mathcal{H} = \{\mathbf{1}[\cdot \leq x] : x \in \mathbb{R}\}$, which tells us how large the difference between two different distributions is within a range of differences between the two distributions. The total variance distance measures the maximum probability difference between the two distributions over the set of all possible events, so its set of test functions is defined as $\mathcal{H} = \{\mathbf{1}[\cdot \in A] : A \in \text{Borel}(\mathbb{R})\}$. Instead, we will focus on the Wasserstein distance in this paper, specifically the Wasserstein distance, which maximizes the difference in the expectation of the two distributions for any Lipschitz functions. So its test function set is defined as $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leq |x - y|\}$. Then for two random variables X and Y, the probability metric will be in the form of:

$$d_{\mathcal{H}}(X, Y) = \sup_{g \in \mathcal{H}} \left| \int g(x) dF_X - \int g(y) dF_Y \right| = \sup_{g \in \mathcal{H}} \left| \mathbb{E}[g(X)] - \mathbb{E}[g(Y)] \right|.$$

Definition 2.2.1 [13] A function $f : \mathbb{R} \to \mathbb{R}$ is called a Lipschitz function if there exists a constant $C \geq 0$ such that, for all $x, y \in \mathbb{R}$, the following condition is satisfied:

$$|f(x) - f(y)| \leq C|x - y|.$$

For simplicity, we choose $C = 1$.

Given the test functions to be a set of Lipschitz functions, the Wasserstein distance is given by:

$$d_W(X, Y) = \sup_{g \in Lip} |\mathbb{E}[g(X)] - \mathbb{E}[g(Y)]|.$$

We also have theorems about probability metrics to expand the relationship between those metrics.

Theorem 2.2.2 [1] Suppose X, Z are two random variables, and If Z is a standard normal random variable then,

$$d_K(X, Z) \leq 2\sqrt{\frac{1}{\sqrt{2\pi}} d_W(X, Z)}.$$

Proof: The core idea is to construct a Lipschitz function $g_\epsilon(x)$ which satisfies the assumption. Set

$$g_\epsilon(x) = \begin{cases} 1 & \text{if } x \leq z, \\ 0 & \text{if } z + \epsilon \leq x, \\ \text{linear} & \text{if } z \leq x \leq z + \epsilon. \end{cases}$$

Then $|g_\epsilon(x)'| = \frac{1}{\epsilon}$ for some constants ϵ.

First, recall that $d_K(X, Z) = \sup_z |\mathbb{P}(X \leq z) - \mathbb{P}(Z \leq z)|$. By direct calculation, we have,

$$\begin{aligned}
\mathbb{P}(X \leq z) - \mathbb{P}(Z \leq z) &\leq \mathbb{P}(X \leq z) + \mathbb{P}(z \leq X \leq z+\epsilon) - \mathbb{P}(z \leq Z \leq z+\epsilon) \\
&\quad + \mathbb{P}(z \leq Z \leq z+\epsilon) - \mathbb{P}(Z \leq z) + \mathbb{P}(Z \leq z) - \mathbb{P}(Z \leq z) \\
&= [\mathbb{P}(X \leq z) + \mathbb{P}(z \leq X \leq z+\epsilon)] - [\mathbb{P}(z \leq Z \leq z+\epsilon) + \mathbb{P}(Z \leq z)] \\
&\quad + \mathbb{P}(z \leq Z \leq z+\epsilon) - \mathbb{P}(Z \leq z) + \mathbb{P}(Z \leq z) \\
&\leq \mathbb{E}[g_\epsilon(X)] - \mathbb{E}[g_\epsilon(Z)] + \mathbb{P}(z \leq Z \leq z+\epsilon) \\
&\leq \frac{d_W(X,Z)}{\epsilon} + \mathbb{P}(z \leq Z \leq z+\epsilon) \\
&\leq \frac{d_W(X,Z)}{\epsilon} + \int_z^{z+\epsilon} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&\leq \frac{d_W(X,Z)}{\epsilon} + \frac{1}{\sqrt{2\pi}} \int_z^{z+\epsilon} dx \\
&\leq \frac{d_W(X,Z)}{\epsilon} + \frac{\epsilon}{\sqrt{2\pi}},
\end{aligned}$$

and we set $f(\epsilon) = \frac{d_W(X,Z)}{\epsilon} + \frac{\epsilon}{\sqrt{2\pi}}$. Then,

$$f'(\epsilon) = -\frac{d_W(X,Z)}{\epsilon^2} + \frac{1}{\sqrt{2\pi}}.$$

We need to find the least ϵ that makes $f'(\epsilon) = 0$. Thus, we have

$$\epsilon = \sqrt{\frac{d_W(X,Z)}{\frac{1}{\sqrt{2\pi}}}}.$$

Combining all results, we have

$$\mathbb{P}(X \leq z) - \mathbb{P}(Z \leq z) \leq 2\sqrt{\frac{d_W(X,Z)}{\sqrt{2\pi}}} = 2\sqrt{\frac{1}{\sqrt{2\pi}} d_W(X,Z)}.$$

For any two random variables X and Y, we expect them to have approximately the same distribution if $d_W(X, Y) = 0$.

2.3 Stein's Equation

The characteristic function and Stein's method greatly differ in comparing the distributions between two random variables. Specifically, the main idea of the Stein method is to convert the traditional problem of comparing the difference between the distributions of two random variables into a problem of analyzing the solution of a differential equation by constructing a differential equation, which is a Stein equation. This chapter will discuss how to construct a Stein equation for normal approximation. We begin with Stein's characterizing operator first.

Stein's Lemma: [10] Define the operator A acting on a function f as follows:

$$Af(x) = f'(x) - xf(x).$$

This operator A has the following properties: If Z is a standard normally distributed random variable, then for any absolutely continuous function f with $\mathbb{E}|f'(Z)| < \infty$, we have

$$\mathbb{E}[Af(Z)] = 0. \tag{2.3.10}$$

Conversely, if for some random variable W, the condition

$$\mathbb{E}[Af(W)] = 0. \tag{2.3.11}$$

holds for all absolutely continuous functions f with $\|f'\| < \infty$, then W follows the standard normal distribution. Stein's Lemma indicates that a random variable X is a standard normal variable if and only if $\mathbb{E}[\mathcal{A}f(X)] = 0$. We will first prove (2.3.10).

Theorem 2.3.1 [3] Let $Z \sim N(0, 1)$. Then for any bounded function f with continuous first-order derivatives f', we have

$$\mathbb{E}\left[f'(Z) - Zf(Z)\right] = 0. \tag{2.3.12}$$

The proof requires a direct calculation of $\mathbb{E}[f'(Z)]$ and $\mathbb{E}[Zf(Z)]$ using the definition of expectation. Applying the integral by part would give a straightforward result. Let

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

It is clear that $g'(z) = -zg(z)$. By direct calculation, we have:

$$\begin{aligned}
\mathbb{E}[f'(Z) - Zf(Z)] &= \int_{-\infty}^{\infty} \left(f'(z) - zf(z)\right) g(z)dz \\
&= \int_{-\infty}^{\infty} f'(z)g(z)dz - \int_{-\infty}^{\infty} zf(z)g(z)dz \\
&= 0.
\end{aligned}$$

Theorem 2.3.1 proves that if a random variable is a standard normal random variable, then it satisfies $E[f'(Z) - Zf(Z)] = 0$. The interesting fact is that Stein first observed the identity(2.3.12) and then found the solution of $f(x)$ to the differential equation in the form of $f'(x) - xf(x) = h(x)$. By direct calculation, giving,

$$f(x) = e^{\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} h(x)\, dx.$$

Then for any bounded function $g(x)$, we can verify that the function $f(x)$ defined in the form of

$$f(x) := e^{x^2/2} \int_{-\infty}^{x} \left(g(t) - \mathbb{E}[g(Z)]\right) e^{-t^2/2}\, dt \tag{2.3.13}$$

satisfies the differential equation $f'(x) - xf(x) = g(x) - \mathbb{E}[g(Z)]$. If we take expectations on both sides, then yields:

$$\mathbb{E}[f'(x) - xf(x)] = \mathbb{E}g(x) - \mathbb{E}[g(Z)]. \tag{2.3.14}$$

We call the equation $f'(x) - xf(x) = g(x) - \mathbb{E}[g(Z)]$ the Stein equation. One point is that if we take the upper bound of the right-hand side of equation 2.3.14, it is equivalent to the Wasserstein distance mentioned in section 2.2. Thus, the Stein equation also provides an upper bound for the normal approximation of the Wasserstein distance mentioned above. Considering a random variable Y, we must have the following,

$$\sup_{f\in\mathcal{F}} \left|\mathbb{E}[f'(Y) - Yf(Y)]\right| = \sup_{g\in\mathcal{H}} \left|\mathbb{E}g(Y) - \mathbb{E}g(Z)\right|. \tag{2.3.15}$$

where $\mathcal{H}$ is a set of any Lipschitz functions.

Equation 2.3.15 tells us that the Wasserstein distance between two random variables can be bounded by calculating the expectation on the left side of the equation. Note that the expectation on the left side of the equation depends only on variable Y, so it is generally easier to find an upper bound for the expectation on the left side than on the right.

Theorem 2.3.2 [3] Let Y be a random variable, $\mathcal{F}$ be a large class of real-valued functions. If for all $f \in \mathcal{F}$,

$$\mathbb{E}[f'(Y) - Yf(Y)] = 0,$$

then $Y \sim N(0,1)$. Where f is a function in a function set

$$\mathcal{F} = \left\{ f : \|f\|_\infty \leq 1, \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}}, \|f''\|_\infty \leq 2 \right\}. \tag{2.3.21}$$

Proof: We first notice that if $P(Y \leq z) - \Phi(z) = 0$, then Y is a standard normal random variable. We have

$$P(Y \leq z) - \Phi(z) = \mathbb{E}\left[1_{(-\infty,z]}(Y) - \Phi(z)\right].$$

Now we set the Stein equation as:

$$f'(Y) - Yf(Y) = 1_{(-\infty,z]}(Y) - \Phi(z).$$

Recall that (2.3.12) gives us a solution for any bounded function $g(x)$, which is,

$$f(y) = e^{y^2/2} \int_{-\infty}^{y} \left[1_{(-\infty,z]}(t) - \Phi(z)\right] e^{-t^2/2}\, dt$$

$$= -e^{y^2/2} \int_{y}^{\infty} \left[1_{(-\infty,z]}(t) - \Phi(z)\right] e^{-t^2/2}\, dt.$$

Now, evaluating the integral, we get,

$$f(y) = \begin{cases} \sqrt{2\pi} e^{y^2/2} \Phi(y) \left[1 - \Phi(z)\right], & y \leq z. \\ \sqrt{2\pi} e^{y^2/2} \Phi(z) \left[1 - \Phi(y)\right], & y > z. \end{cases}$$

The function $f(y)$ can be verified that it is a bounded continuous differentiable function in the function set $\mathcal{F}$. Suppose (2.3.12) holds for any $f \in \mathcal{F}$, thus it holds for $f(y)$. Therefore, we must have:

$$\mathbb{E}[f'(Y) - Yf(Y)] = \mathbb{E}[1_{(-\infty,z]}(Y) - \Phi(z)] = 0.$$

Thus, we conclude that Y is a standard normal random variable.

Our framework so far consists of the probability metric, focusing on the Wasserstein distance, Stein's Lemma, Stein identity and Stein equation. The Stein identity leads to the most central part of the Stein method, the Stein equation. We focus on the Wasserstein distance because the Stein equation gives an upper bound on the Wasserstein distance under the Lipschitz condition, and the Wasserstein distance can bound the Kolmogorov distance. In the context of Normal Approximation, we also give a general solution to the Stein equation and prove that any random variable is standard normal distributed if and only if the Stein identity holds. The remaining key element we didn't deal with is the property of the solution to Stein equation. Specifically, Theorem 2.3.2 states that the solution f should satisfy (2.3.21), and the following few theorems explain where (2.3.21) comes from.

2.4 Property of the Solution to the Stein Equation

Theorem 2.4.1 [14] For any bounded function $g(x)$, if g is Lipschitz, the another form of solution f satisfies the Stein equation is

$$f(x) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbb{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] dt, \quad Z \sim N(0,1).$$

The proof requires the use of Stein's Identity where $f(Z) = g(\sqrt{t}x + \sqrt{1-t}Z)$. Then substitute it into $f'(x)$ and $xf(x)$.

First note that

$$f'(x) = -\int_0^1 \frac{\sqrt{t}}{2\sqrt{t(1-t)}} \mathbb{E}\left[Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt.$$

Then set $g(Z) = g(\sqrt{t}x + \sqrt{1-t}Z)$, and by Stein's identity we must have

$$\mathbb{E}\left[Zg(Z)\right] = \mathbb{E}\left[g'(Z)\right].$$

This implies

$$\mathbb{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] = \mathbb{E}\left[\sqrt{1-t}g'(\sqrt{t}x + \sqrt{1-t}Z)\right].$$

By direct calculation, we have,

$$\begin{aligned} f'(x) - xf(x) &= -\int_0^1 \frac{\sqrt{t}}{2\sqrt{t(1-t)}}\mathbb{E}[Zg'(\sqrt{t}x + \sqrt{1-t}Z)]dt \\ &+ \int_0^1 \frac{x}{2\sqrt{t(1-t)}}\mathbb{E}[Zg(\sqrt{t}x + \sqrt{1-t}Z)]dt \\ &= -\int_0^1 \mathbb{E}\left[\frac{1}{\sqrt{1-t}}Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right]dt \\ &+ \int_0^1 \mathbb{E}\left[\frac{x}{\sqrt{2t(1-t)}}g'(\sqrt{t}x + \sqrt{1-t}Z)\right]dt \\ &= \int_0^1 \mathbb{E}\left[\left(\frac{x}{2\sqrt{t}} - \frac{Z}{2\sqrt{1-t}}\right)g'(\sqrt{t}x + \sqrt{1-t}Z)\right]dt. \end{aligned}$$

Since $Z \sim N(0,1)$ has finite expectation and $\left(\frac{x}{2\sqrt{t}} - \frac{Z}{2\sqrt{1-t}}\right)$ is a linear combination of Z, also we notice that g is a Lipschitz function with $|g'(x)| \leq C$ for some constant C, by Fubini's Theorem we have,

$$= \mathbb{E}\left[\int_0^1 dg(\sqrt{t}x + \sqrt{1-t}Z)\right] = \mathbb{E}\left[g(\sqrt{t}x + \sqrt{1-t}Z)\right]\Bigg|_0^1 = g(x) - \mathbb{E}[g(Z)].$$

Theorem 2.4.2 [14] If g is a Lipschitz function, then the following inequalities hold:

$$|f|_\infty \leq |g'|_\infty, \tag{2.4.2}$$

$$|f'|_\infty \leq \sqrt{\frac{\pi}{2}}|g'|_\infty, \tag{2.4.3}$$

$$|f''|_\infty \leq 2|g'|_\infty. \tag{2.4.4}$$

Equations (2.4.2) and (2.4.3) are simple to check by directly calculating using another solution form f. Equation (2.4.4) is tricky to prove. We need to take the second derivative of the Stein equation with respect to x and then rewrite the second derivative as $f''(x) = g'(x) + f(x) + x\left[xf(x) + g(x) - \mathbb{E}g(Z)\right] = g'(x) + x\left[g(x) - \mathbb{E}g(Z)\right] + (1+x^2)f(x)$. Write $g(x)-\mathbb{E}g(Z)$ and $f(x)$ as the form of integrals, then we can bound them respectively. The theorem assumes that $g(x)$ is a Lipschitz function, which is also consistent with the definition of the Wasserstein distance. Specifically, if $g(x)$ is 1-Lipschitz, then$|g'|_\infty = 1$. Also, one can find a $f(x)$ to solve the Stein equation for any 1-Lipschitz $g(x)$. based on this relationship, giving,

$$d_W(W, Z) = \sup_{g\in\mathcal{H}} \left|\mathbb{E}\left[g(W)\right] - \mathbb{E}\left[g(Z)\right]\right| \leq \sup_{f\in\mathcal{F}} \left|\mathbb{E}\left[f'(W) - Wf(W)\right]\right|.$$

All these theorems complete the whole framework of our general setup. That is, we found and verified the conditions of the solution $f(x)$, which is in the $\mathcal{F}$ set.

2.5 Classical Central Limit Theorem with Stein's Method

As mentioned, if the Wasserstein distance is approximately equal to 0, the two random variables have approximately the same distribution. However, it is difficult to calculate precisely how much the two distributions differ directly. Hence, the Stein method gives us a new way to estimate the distance between two distributions. This section will discuss a specific case where the Stein method provides an upper bound on the Wasserstein distance for a set of independently and identically distributed random variables to analyze the difference between the two distributions. In addition, the Stein method will also give a rate of convergence for the approximation.

Theorem 2.5.1

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, and let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables defined on a common probability space such that:

- $\mathbb{E}[X_i] = 0$, i.e., the expectation of X_i is 0.
- $\mathbb{V}ar(X_i) = 1$, i.e., the variance of X_i is 1.
- A sequence $\{X_i\}_{i=1}^n$ such that for any n, the third-moment condition holds:

$$\lim_{n\to\infty} \frac{1}{n^{3/2}} \sum_{k=1}^{n} \mathbb{E}[|X_k|^3] = 0.$$

Define the partial sum and normalized random variables as follows:

$$S_n = X_1 + X_2 + \cdots + X_n, \quad W_n = \frac{S_n}{\sqrt{n}}.$$

Then, as $n \to \infty$, the distribution of W_n converges to the standard normal distribution, i.e.,

$$W_n \xrightarrow{d} Z \sim N(0,1), \quad n \to \infty.$$

Here, $\xrightarrow{d}$ denotes convergence in distribution.

Proof: The main idea is to control the upper bound of Wasserstein distance $d_W(W, Z)$. Recall that $d_W(W, Z) \leq \sup_{f\in\mathcal{F}} |\mathbb{E}\left[f'(W) - Wf(W)\right]|$. We begin by calculating the expectation of the left-hand side of the Stein equation. For any $i = 1, \ldots, n$, set

$$W_n^i = W_n - \frac{X_i}{\sqrt{n}}.$$

We first notice that,

$$\mathbb{E}\left[f'(W_n) - W_n f(W_n)\right] = \mathbb{E}\left[f'(W_n) - \frac{1}{n}\sum_{i=1}^{n} f'(W_n^i) + \frac{1}{n}\sum_{i=1}^{n} f'(W_n^i) - W_n f(W_n)\right],$$

where we set

$$\mathrm{I} = \mathbb{E}\left[W_n f'(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i^2 f'(W_n^i)\right],$$

and

$$\mathrm{II} = \mathbb{E}\left[f'(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[f'(W_n^i)\right].$$

For any $f \in \mathcal{H}$, it follows from direct calculation that

$$\begin{aligned}
\mathbb{E}[W_n f(W_n)] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left[X_i f(W_n)\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left[X_i\big(f(W_n) - f(W_n^i)\big) + X_i f(W_n^i)\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left[X_i\big(f(W_n) - f(W_n^i)\big)\right],
\end{aligned}$$

using the independence of X_i and $f(W_n^i)$ in the last line.

Noting that

$$f(W_n) - f(W_n^i) = f(W_n) - f(W_n^i) - f'(W_n^i)(W_n - W_n^i) + f'(W_n^i)(W_n - W_n^i)$$

$$= f(W_n) - f(W_n^i) - f'(W_n^i)(W_n - W_n^i) + \frac{X_i}{\sqrt{n}} f'(W_n^i).$$

Thus

$$\mathrm{I} = \mathbb{E}\left[W_n f(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i^2 f'(W_n^i)\right]$$

$$= \mathbb{E}\left[W_n f(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[f'(W_n^i)\right]$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left[X_i(f(W_n) - f(W_n^i) - f'(W_n^i)(W_n - W_n^i))\right].$$

By Taylor's expansion, we obtain

$$\begin{aligned}
|I| &= \left|\mathbb{E}\left[W_n f(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[f'(W_n^i)\right]\right| \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left|X_i\left(f(W_n) - f(W_n^i) - f'(W_n^i)(W_n - W_n^i)\right)\right| \\
&\leq \frac{|f''|_\infty}{2} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{E}\left[|X_i|(W_n - W_n^i)^2\right] \\
&\leq \frac{|f''|_\infty}{2} \cdot \frac{1}{n^{3/2}}\sum_{i=1}^{n} \mathbb{E}\left[|X_i|^3\right].
\end{aligned}$$

Thus, we have:

$$|I| \leq \frac{1}{n^{3/2}}\sum_{i=1}^{n} \mathbb{E}\left[|X_i|^3\right].$$

On the other hand, set

$$|II| = \left|\mathbb{E}\left[f'(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[f'(W_n^i)\right]\right| = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left|f'(W_n) - f'(W_n^i)\right|.$$

Applying Taylor's expansion, we obtain

$$
\begin{aligned}
|II| &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[|f''|_\infty |W_n - W_n^i|\right] \\
&\leq \frac{|f''|_\infty}{n}\sum_{i=1}^{n}\mathbb{E}\left[|W_n - W_n^i|\right] \\
&\leq \frac{2}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[|X_i|^3\right].
\end{aligned}
$$

By Hölder's inequality, it follows that

$$
\mathbb{E}\left[X_i^2\right] = 1 \leq \left(\mathbb{E}\left[|X_i|^3\right]\right)^{2/3},
$$

$$
\mathbb{E}\left[|X_i|\right] \leq \left(\mathbb{E}\left[|X_i|^3\right]\right)^{1/3} \leq \left(\mathbb{E}\left[|X_i|^3\right]\right).
$$

Thus, we have:

$$
|II| \leq \frac{2}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left[|X_i|^3\right].
$$

Finally, we have

$$
\begin{aligned}
|\mathbb{E}\left[f'(W_n) - W_n f(W_n)\right]| &= \left|\mathbb{E}\left[f'(W_n)\right] - \frac{1}{n}\sum_{i=1}^{n} f'(W_n^i) + \frac{1}{n}\sum_{i=1}^{n} f'(W_n^i) - W_n f(W_n)\right| \\
&\leq |I| + |II| \\
&\leq \frac{3}{n^{3/2}}\sum_{k=1}^{n}\mathbb{E}\left[|X_k|^3\right] \to 0,
\end{aligned}
$$

which also indicates

$$
d_K(W_n, Z) \leq d_W(W_n, Z) \leq \frac{3}{n^{3/2}}\sum_{k=1}^{n}\mathbb{E}\left[|X_k|^3\right] \to 0.
$$

Therefore, we conclude that W_n is a standard normal random variable. Set $C = \max\{\mathbb{E}\left[|X_k|^3\right]\}$ for any k from 1 to n. Using this, we have,

$$d_K(W_n, Z) \leq d_W(W_n, Z) \leq \frac{3}{n^{3/2}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^3\right]$$

$$\leq \frac{3}{n^{3/2}} \cdot nC = \frac{3C}{n^{1/2}} \approx O\left(\frac{1}{\sqrt{n}}\right).$$

Thus, we observe that the variable W_n converges to the standard normal at the speed of approximately $O\left(\frac{1}{\sqrt{n}}\right)$.

2.6 A Bound for the Error in Normal Approximation of Dependent Random Variables

In this section, we will present key results from Stein's original work [12] on the normal approximation for sums of dependent random variables. The classical CLT only applies to i.i.d. random variables. The Berry-Esseen theorem provides an upper bound on the error in the independent and identically-distributed case as $O\left(n^{-1/2}\right)$. However, this result no longer holds in the dependent case, so Stein wanted to find a new tool to deal with these complex dependence structures.

In the dependent random variable case, Phillip [9] proved that when the random variables X_i are bounded and exhibit exponentially decaying dependence, the error is of the order $n^{-1/4}$. Stein's work further derives an error bound under more general conditions. His paper proves that the error can reach $O\left(n^{-1/2}\left(\log n\right)^2\right)$ in the case of weakly dependent random variables and, under certain assumptions, even $O\left(n^{-1/2}\right)$.

Stein's reasoning goes like this: to transform the problem of normal approximation into one of analyzing the difference in expectations by constructing a differential equation associated with a normal distribution. The Stein method also provides a quantitative analysis of the upper bound on the error. For a stationary and m-dependence sequence of random variables $X_1, X_2, \ldots, X_n$, the Stein method can show that the upper bound on the error of the distribution of its standardized sum $S_n = \frac{1}{\sqrt{\operatorname{Var} \sum_{i=1}^n X_i}} \sum_{i=1}^n X_i$ approximating the normal distribution is $An^{-1/2}$, where A is a constant depending on the distribution of the sequence $X_1, X_2, X_3, \ldots.$

We first recall the definition of a stationary sequence and a m-dependence sequence.

Definition 2.6.1 [12] A sequence $X_1, X_2, \ldots$ of random variables is said to be stationary if, for every pair t, j of natural numbers, the sequence $X_{t+1}, \ldots, X_{t+j}$ has the same distribution as $X_1, \ldots, X_j$.

Definition 2.6.2 [12] A sequence $X_1, X_2, \ldots$ of random variables is said to be m-dependent, where m is a nonnegative integer, if for any two subsets $A, B \subset \{1, 2, \ldots\}$ for which

$$\inf_{i \in A, j \in B} |i - j| \geq m + 1,$$

the sets of random variables $\{X_i\}_{i \in A}$ and $\{X_j\}_{j \in B}$ are independent. The intuition of this definition is if two sets of random variables are separated by more than m positions in the sequence, then these sets of variables are independent of each other. We consider a stationary m-dependent sequence of random variables $X_1, X_2, \ldots$ also satisfying the following conditions:

$$\mathbb{E}[X_i] = 0, \quad \text{for all } i, \tag{2.6.1}$$

$$\mathbb{E}[X_i^2] = 1, \quad \text{for all } i, \tag{2.6.2}$$

$$\beta = \mathbb{E}[X_i^8] < \infty, \tag{2.6.3}$$

$$0 < C = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) < \infty. \tag{2.6.4}$$

Then we must have the following corollary.

Corollary 2.6.3 [12] If $X_1, X_2, \dots$ is a stationary m-dependent sequence of random variables satisfying (2.6.1), (2.6.2),(2.6.3) and (2.6.4), then there exists a constant A, where A is a constant depending on the distribution of the sequence $X_1, X_2, X_3, ...,$

$$\left| P\left(\frac{1}{\sqrt{nC}} \sum_{i=1}^{n} X_i \le a\right) - \Phi(a) \right| \le A n^{-1/2},$$

where $\Phi(a)$ is the cumulative distribution function of the standard normal distribution.

In the original Stein paper, the dependent sequence was treated relatively strictly, requiring that the sequence not only satisfy stationarity but also be m-dependent. This setting requires each variable depends only on a fixed number of neighbors and statistical properties such as expectation and variance remain consistent across the sequence. Next the following theorems generalize Stein's method to a wider range of dependence structures.

Theorem 2.6.4 [1] Let $\{\xi_1, \xi_2, \dots\}$ be a sequence of random variables $\xi_i, i \in \mathbb{Z}$. Let J be a finite index set with cardinality n, and let $\{\xi_i, i \in J\}$ be a random field with zero means and finite variances, such that $\mathrm{Var}(W) = 1$ for $W = \sum_{i \in J} \xi_i$. For each $A \subset J$, define $\xi_A = \{\xi_i, i \in A\}$ and $A^c = \{j \in J : j \notin A\}$. The following one assumption, defining two different conditions of local dependence:

- LD1: for any $i \in J$, one can find subsets $A_i \subset B_i \subset J$ ξ_i is independent of $\xi_{A_i^c}$, and ξ_{A_i} is independent of $\xi_{B_i^c}$.

We set

$$\eta_i = \sum_{j \in A_i} \xi_j \quad \text{and} \quad \tau_i = \sum_{j \in B_i} \xi_j.$$

It is clear, for independent random variables ξ_i, we can take $A_i = B_i = \{i\}$. Then the difference δ is given by:

$$\delta = 2 \sum_{i \in J} \left(\mathbb{E}[\xi_i \eta_i \tau_i] + \mathbb{E}[\xi_i \eta_i] \mathbb{E}[\tau_i] \right) + \sum_{i \in J} \mathbb{E}[\xi_i^2 \eta_i],$$

where δ is defined as:

$$d_W(\mathcal{L}(W), \mathcal{N}(0,1)) := \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \delta.$$

Next we introduce a new dependence structure called Dependency Neighborhoods.

Definition 2.6.5 [10] A set of random variables $(X_1, \ldots, X_n)$ has dependency neighborhoods $N_i \subset \{1, \ldots, n\}$, $i = 1, \ldots, n$, if $i \in N_i$ and X_i is independent of $\{X_j\}_{j \notin N_i}$. We first note that dependency neighborhoods is a more flexible condition. The intuition behind neighborhoods dependency is to limit the dependencies of each variable to a manageable subset. This setting is useful since each random variable X_i can define a unique subset that every other random variable in this subset is dependent to X_i.

Theorem 2.6.6 [10] Let $X_1, \ldots, X_n$ be random variables such that $\mathbb{E}[X_i^4] < \infty$, $\mathbb{E}[X_i] = 0$, $\sigma^2 = \mathrm{Var}\left(\sum_i X_i\right)$, and define $W = \sum_i X_i/\sigma$. Let the collection $(X_1, \ldots, X_n)$ have dependency neighborhoods N_i, $i = 1, \ldots, n$, and also define $D := \max_{1 \le i \le n} |N_i|$. Then for Z a standard normal random variable,

$$d_W(W, Z) \le \frac{D^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}\left|X_i\right|^3 + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}\left[X_i^4\right]}.$$

These theorems give an error bound of wasserstein distance under the different structures of dependent sequence. The restrictions on the dependency structure have been relaxed from the stationary m-dependent sequence to the LD2 condition and then to the neighborhood dependence, enabling the Stein method to be applied to a wider range of scenarios and complex dependencies. A detailed discussion of these structures and proofs under these conditions can be found in further paper [12] [1] [10].

Chapter 3

Stein's Method For Random Index CLT

In the previous chapter, we discussed the classical CLT for i.i.d. random variables. We study the difference between proving the central limit theorem using characteristic functions and Stein's method, and also discuss the work of Charles Stein in 1972, which provides an error bound for weakly dependent random variables. However, in many practical applications, random variables are not identically distributed, which brings a challenge for the application of classical CLT.

In this chapter, we discuss the case of independent but not necessarily identical random variables. Specifically, we'll focus on randdom sums of independent random variables. This is common in a variety of fields, such as insurance and finance, where the number of sums is often determined by a random mechanism. The aim of this chapter is to analyze conditions under which the distribution of a normalized random sum, W_{Z_n}, converges to a standard normal distribution using the Stein's method. We derive two key conditions that must be met for this convergence to occur. This chapter will deduce these conditions in detail and explore their implications. Compared to the classical CLT discussed in the previous chapter, the conditions proposed in this chapter are broader and can accommodate more cases.

Theorem 3.1

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{X_i : \Omega \to \mathbb{R}\}_{i=1}^{\infty}$ be a sequence of independent random variables defined on a common space such that $\mathbb{E}[X_i] = 0$ and $\sigma_i^2 = \mathrm{Var}(X_i) < \infty$ for each i. Let $Z_n : \Omega \to \mathbb{Z}^+$ be a positive integer-valued random variable, independent of $\{X_i\}$, defined on the same probability space. Define the partial sum S_{Z_n}:

$$S_{Z_n} = \sum_{i=1}^{Z_n} X_i.$$

Let

$$\sigma_{S_{Z_n}}^2 = \mathrm{Var}(S_{Z_n}).$$

Let

$$W_{Z_n} = \frac{S_{Z_n}}{\sigma_{S_{Z_n}}}$$

be the standardized sum. If the following conditions hold:

$$\lim_{n\to\infty} \frac{3\sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[|X_i|^3\right]}{\left(\sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \mathbb{E}\left[X_i^2\right]\right)^{3/2}} = 0, \tag{3.0.1}$$

then $W_{Z_n} \Rightarrow Z \sim \mathcal{N}(0,1)$ as $n \to \infty$ and the wasserstein distance is given by:

$$d_W(W_{Z_n}, \mathcal{N}(0,1)) \leq \frac{3}{\sigma_{S_{Z_n}}^3} \left[\sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[|X_i|^3\right]\right].$$

Proof: Before the proof we first recall that the wasserstein distance is bounded by $\mathbb{E}\left[f'(W_{Z_n}) - W_{Z_n}f(W_{Z_n})\right]$ and $\frac{3\sum_{i=1}^{Z_n}\mathbb{E}[|X_i|^3]}{\left(\sum_{i=1}^{Z_n}\mathbb{E}[X_i^2]\right)^{3/2}}$ is the upper bound of $\mathbb{E}[f'(W_{Z_n}) - W_{Z_n}f(W_{Z_n})]$. We need to make its limit (3.0.1) equal to 0 to satisfy the Stein's Lemma.

By direct calculation, we have,

$$\begin{aligned}
\mathbb{E}[S_{Z_n}] &= \mathbb{E}\left(\mathbb{E}[S_{Z_n} \mid Z_n = n]\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n}\mathbb{E}[X_i \mid Z_n = n]\right) \\
&= \mathbb{E}\left[0\right] \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\sigma^2_{S_{Z_n}} &= \mathbb{E}\left[S^2_{Z_n}\right] - \left(\mathbb{E}[S_{Z_n}]\right)^2 \\
&= \mathbb{E}\left[S^2_{Z_n}\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{Z_n} X_i\right)^2\right] \\
&= \sum_{k=1}^{\infty} P(Z_n = k)\cdot\mathbb{E}\left[\left(\sum_{i=1}^{k} X_i\right)^2\right] \\
&= \sum_{k=1}^{\infty} P(Z_n = k)\cdot\mathbb{E}\left[\sum_{j=1}^{k}\sum_{i=1}^{k} X_jX_i\right] \\
&= \sum_{k=1}^{\infty} P(Z_n = k)\cdot\sum_{j=1}^{k}\sum_{i=1}^{k}\mathbb{E}[X_jX_i]
\end{aligned}$$

$$= \begin{cases} 0, & \text{if } j \neq i, \\ \sum_{k=1}^{\infty} P(Z_n = k)\cdot\sum_{i=1}^{k}\sigma_i^2, & \text{if } j = i. \end{cases}$$

For any $i = 1, \ldots, Z_n$, set

$$W_{Z_n}^i = W_{Z_n} - \frac{X_i}{\sqrt{\sigma_{S_{Z_n}}}}, \text{ which means } W_{Z_n}^i \text{ is independent of } X_i.$$

Recall that W_{Z_n} is standard normal if and only if $E[f'(W_{Z_n}) - W_{Z_n} f(W_{Z_n})] \to 0$. Rewrite it, giving,

$$\begin{aligned}
&E\left[f'(W_{Z_n}) - W_{Z_n} f(W_{Z_n})\right] \\
&= E\left[f'(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right] \\
&- \left(E\left[W_{Z_n} \cdot f(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right] \right).
\end{aligned}$$

Then set

$$I = E\left[W_{Z_n} \cdot f(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right].$$

$$II = E\left[f'(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right].$$

Notice that

$$\left|E\left[f'(W_{Z_n}) - W_{Z_n} f(W_{Z_n})\right]\right| = |I - II| \leq |I| + |II|.$$

Also note that,

$$\begin{aligned}\mathbb{E}\left[W_{Z_n} f(W_{Z_n})\right] &= \mathbb{E}\left[\frac{S_{Z_n}}{\sigma_{S_{Z_n}}} \cdot f(W_{Z_n})\right] \\ &= \frac{1}{\sigma_{S_{Z_n}}} \cdot \mathbb{E}\left[S_{Z_n} \cdot f(W_{Z_n})\right] \\ &= \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \mathbb{E}\left[X_i f(W_k)\right] \\ &= \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \mathbb{E}\left[X_i (f(W_k) - f(W_k^i))\right].\end{aligned}$$

Since we have,

$$f(W_k) - f(W_k^i) = f(W_k) - f(W_k^i) - f'(W_k^i)(W_k - W_k^i) + f'(W_k^i)(W_k - W_k^i)$$

$$= f(W_k) - f(W_k^i) - f'(W_k^i)(W_k - W_k^i) + \frac{X_i}{\sigma_{S_{Z_n}}} f'(W_k^i).$$

Thus we can conclude that,

$$\begin{aligned}I &= E\left[W_{Z_n} \cdot f(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right] \\ &= E\left[W_{Z_n} \cdot f(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2\right] \cdot E\left[f'(W_k^i)\right] \\ &= \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\Bigg[X_i (f(W_k) - f(W_k^i) - f'(W_k^i)(W_k - W_k^i) \\ &\quad + \frac{X_i}{\sigma_{S_{Z_n}}} f'(W_k^i)) - \frac{X_i^2}{(\sigma_{S_{Z_n}})} f'(W_k^i)\Bigg] \\ &= \frac{1}{(\sigma_{S_{Z_n}})} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\Bigg[X_i \left(f(W_k) - f(W_k^i) - f'(W_k^i)(W_k - W_k^i)\right)\Bigg].\end{aligned}$$

Then expand I at W_k^i using Taylor expansion we get:

$$f(W_k) = f(W_k^i) + f'(W_k^i)(W_k - W_k^i) + \frac{f''(W_k^i)}{2}(W_k - W_k^i)^2 + O\left((W_k - W_k^i)^3\right).$$

So,

$$\begin{aligned}
|I| &= \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[X_i\left(f(W_k) - f(W_k^i) - f'(W_k^i)(W_k - W_k^i)\right)\right] \\
&\leq \frac{|f''|_\infty}{2} \cdot \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[\left|X_i(W_k - W_k^i)^2\right|\right] \\
&\leq \frac{|f''|_\infty}{2} \cdot \frac{1}{\sigma_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \frac{1}{(\sigma_{S_k})^2} \cdot \sum_{i=1}^{k} E\left[|X_i|^3\right] \\
&\leq \frac{|f''|_\infty}{2} \cdot \frac{1}{(\sigma_{S_{Z_n}})^3} \cdot \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[|X_i|^3\right].
\end{aligned}$$

Similarly, we have,

$$\begin{aligned}
II &= E\left[f'(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2 \cdot f'(W_k^i)\right] \\
&= E\left[f'(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} E\left[X_i^2\right] \cdot E\left[f'(W_k^i)\right] \\
&= \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} \sigma_i^2 \cdot E\left[f'(W_{Z_n})\right] - \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} \sigma_i^2 \cdot E\left[f'(W_k^i)\right] \\
&= \frac{1}{(\sigma_{S_{Z_n}})^2} \sum_{k=1}^{\infty} P(Z_n = k) \sum_{i=1}^{k} \sigma_i^2 \cdot E\left[f'(W_k) - f'(W_k^i)\right].
\end{aligned}$$

Then expand Π at W_k^i using Taylor expansion we get:

$$f'(W_k) = f'(W_k^i) + f''(W_k^i)(W_k - W_k^i) + O\left((W_k - W_k^i)^n\right).$$

So we conclude that,

$$
\begin{aligned}
|II| &= \frac{1}{\sigma^2_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \sigma_i^2 \left| E\left[f'(W_k) - f'(W_k^i) \right] \right| \\
&\leq \frac{1}{\sigma^2_{S_{Z_n}}} \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \sigma_i^2 E\left[\left| f'(W_k) - f'(W_k^i) \right| \right] \\
&\leq \frac{|f''|_\infty}{\sigma^2_{S_{Z_n}}} \cdot \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \sigma_i^2 \cdot E\left[\left| W_k - W_k^i \right| \right] \\
&\leq \frac{2}{\sigma^3_{S_{Z_n}}} \cdot \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[X_i^2 \right] \cdot E\left[|X_i| \right],
\end{aligned}
$$

where we use the Theorem 2.4.4 in last inequality.

By Holder's inequality, it follows that:

$$
E\left[X_i^2\right] \leq \left(E\left[|X_i|^3\right]\right)^{\frac{2}{3}} \quad \text{and} \quad E\left[|X_i|\right] \leq \left(E\left[|X_i|^3\right]\right)^{\frac{1}{3}}.
$$

It implies that:

$$
|II| \leq \frac{2}{\sigma^3_{S_{Z_n}}} \cdot \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[|X_i|^3 \right].
$$

Finally, we put all together and obtain that,

$$
\begin{aligned}
&\left| E\left[f'(W_{Z_n}) - W_{Z_n} f(W_{Z_n}) \right] \right| \\
&\leq |I| + |II| \\
&= \frac{3 \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} E\left[|X_i|^3 \right]}{\left(\sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \mathbb{E}\left[X_i^2 \right] \right)^{3/2}}.
\end{aligned}
\tag{3.0.3}
$$

If equation 3.0.3 goes to 0 as n goes to infinity, then by Theorem 2.3.2 we must have W_{Zn} approaches to standard normal.

Also, by the constraints of central limit theorem we obtain that:

$$\begin{aligned}\mathrm{Var}(S_{Z_n}) &= \sum_{k=1}^{\infty} P(Z_n = k) \cdot \sum_{i=1}^{k} \sigma_i^2 \\ &\leq \sum_{k=1}^{\infty} P(Z_n = k) \cdot k \cdot \sigma_{max_i}^2 \\ &= \sigma_{max_i}^2 \cdot \mathbb{E}[Z_n] \quad \Rightarrow \quad \mathbb{E}[Z_n] < \infty.\end{aligned}$$

This completes the proof of Theorem 3.1. Notice that if we are given the distribution of Zn then we are able to find the convergence speed of W_{Zn} to Standard normal.

3.1 Concrete Example

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider two sequences of independent Bernoulli random variables $\{X_i\}_{i=1}^n$ and $\{Y_k\}_{k=1}^n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The sequences X_i and Y_k follow these distributions:

$$\mathbb{P}(X_i = 1) = \frac{1}{i}, \quad \mathbb{P}(X_i = 0) = 1 - \frac{1}{i}.$$

For the X_i's, and similarly for Y_k:

$$\mathbb{P}(Y_k = 1) = \frac{1}{k}, \quad \mathbb{P}(Y_k = 0) = 1 - \frac{1}{k}.$$

The random variable Z_n is defined as the sum of the Y_i's, that is,

$$Z_n = Y_1 + Y_2 + \cdots + Y_n.$$

Now, define a new sum S_{Z_n} as:

$$S_{Z_n} = X_2 + X_3 + \cdots + X_{Z_n},$$

where each X_i follows the Bernoulli distribution as stated above. Our task is to verify the standardized sum converges to standard normal, specifically we have $W_{Z_n} = \frac{S_{Z_n} - \mathbb{E}[S_{Z_n}]}{\sqrt{\mathrm{Var}(S_{Z_n})}} \xrightarrow{d} N(0,1)$.
By direct calculation we have,

$$\mathbb{E}[X_i] = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = \frac{1}{i}.$$

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}\,[X_i])^2 = \frac{1}{i}.$$

Thus, the variance is:

$$\mathrm{Var}(X_i) = \frac{1}{i} - \frac{1}{i^2}.$$

$$\sum_{i=1}^{k} \mathrm{Var}(X_i) = \sum_{i=1}^{k} \left(\frac{1}{i} - \frac{1}{i^2} \right) = \sum_{i=1}^{k} \frac{1}{i} - \sum_{i=1}^{k} \frac{1}{i^2}.$$

For large k, the harmonic series $\sum_{i=1}^{k} \frac{1}{i}$ grows asymptotically as $\log k$, while the sum $\sum_{i=1}^{k} \frac{1}{i^2}$ converges to a constant $\zeta(2) = \frac{\pi^2}{6}$. Thus, for large k:

$$\sum_{i=1}^{k} \mathrm{Var}(X_i) \approx \log k - \frac{\pi^2}{6}.$$

By direct calculation we have,

$$\mathbb{E}\left[(X_i - \mathbb{E}[X_i])^3\right] = \mathbb{E}[X_i^3] - 3\mathbb{E}[X_i^2]\mathbb{E}[X_i] + 2(\mathbb{E}[X_i])^3.$$

Using the fact that

$$\mathbb{E}[X_i^3] = \mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \frac{1}{i}.$$

Thus, the third moment is:

$$\mathbb{E}\left[(X_i - \mathbb{E}[X_i])^3\right] = \frac{1}{i} - \frac{3}{i^2} + \frac{2}{i^3}.$$

and we have,

$$\sum_{i=1}^{k} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])^3\right] = \sum_{i=1}^{k} \left(\frac{1}{i} - \frac{3}{i^2} + \frac{2}{i^3}\right).$$

For large k:

- The first sum $\sum_{i=1}^{k} \frac{1}{i} \approx \log k$.

- The terms $\sum_{i=1}^{k} \frac{3}{i^2}$ and $\sum_{i=1}^{k} \frac{2}{i^3}$ converge to constants.

Therefore, for large k:

$$\sum_{i=1}^{k} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])^3\right] \approx \log k.$$

Now we substitute the asymptotic approximations of the second and third moments into the limit expression, for the numerator,

$$\sum_{k=1}^{\infty} \mathbb{P}(Z_n = k) \sum_{i=1}^{k} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])^3\right] \approx \sum_{k=1}^{\infty} \mathbb{P}(Z_n = k) \log k = \mathbb{E}[\log Z_n].$$

For the denominator:

$$\begin{aligned}\left(\sum_{k=1}^{\infty}\mathbb{P}(Z_n=k)\sum_{i=1}^{k}\mathbb{E}\left[(X_i-\mathbb{E}[X_i])^2\right]\right)^{3/2} &\approx \left(\sum_{k=1}^{\infty}\mathbb{P}(Z_n=k)\log k\right)^{3/2}\\ &= \left(\mathbb{E}\left[\log Z_n\right]\right)^{3/2}.\end{aligned}$$

Thus, the limit of equation (3.0.1) becomes:

$$\lim_{n\to\infty} 3\cdot\frac{\mathbb{E}[\log Z_n]}{(\mathbb{E}[\log Z_n])^{3/2}} = \lim_{n\to\infty}\frac{3}{(\mathbb{E}[\log Z_n])^{1/2}}. \tag{3.1.1}$$

By large deviation principles [5] we have that,

$$\mathbb{E}\left[\log Z_n\right] \text{ has the same scale as } \log\log n.$$

Thus equation 3.1.1 becomes

$$\lim_{n\to\infty}\frac{3}{(\mathbb{E}[\log Z_n])^{1/2}} = \lim_{n\to\infty}\frac{3}{(\log\log n)^{1/2}} = 0.$$

Also, we notice that for any fixed n,

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\sum_{k=1}^{n} Y_k\right] \approx \log n \text{ is finite.}$$

Thus, we conclude that condition 3.0.1 and 3.0.2 are both satisfied. Therefore, we must have the CLT hold in this case, which is $W_{Z_n} = \frac{S_{Z_n}-\mathbb{E}[S_{Z_n}]}{\sqrt{\mathrm{Var}(S_{Z_n})}} \xrightarrow{d} N(0,1)$.

3.2 Future Work

My future work aims to extend the application of Stein's method to prove the convergence of random sums of weakly dependent random variables to a normal distribution after standardization. This may involve identifying some appropriate conditions under which the sum of sequence of weakly dependent random variables still approximates the standard normal. One specific case is to investigate random variables that exhibit $m-dependence$, as introduced in locally dependent sequences by Stein's paper. Furthermore, another goal is to give an exact error bound for the difference between the sum of dependent sequences and the standard normal distribution. I plan to investigate it within the framework of both stationary and non-stationary sequences. This future work has the potential to solve more complicated questions.

Bibliography

[1] A. D. Barbour and L. H. Y. Chen. *An introduction to Stein's method*, volume 4. World Scientific, 2005.

[2] P. Billingsley. *Probability and Measure*. Wiley, 1995.

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

[4] W. Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

[5] S. Feng and F. M. Hoppe. Large deviation principles for some random combinatorial structures in population genetics and Brownian motion. *The Annals of Applied Probability*, 8(4):975–994, 1998.

[6] B. V. Gnedenko and V. Y. Korolev. *Random summation: limit theorems and applications*. CRC press, 2020.

[7] A. Gut. Anscombe's theorem 60 years later. *Sequential Analysis*, 31(3):368–396, 2012.

[8] P. S. Laplace. *Théorie analytique des probabilités*. Courcier, 1820.

[9] W. Philipp. The remainder in the central limit theorem for mixing stochastic processes. *The Annals of Mathematical Statistics*, 40(2):601–609, 1969.

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

[11] Y. Sinai. *Russian mathematicians in the 20th century*. World scientific, 2003.

[12] C. Stein. 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*, volume 6, pages 583–603. University of California Press, 1972.

[13] E. M. Stein and R. Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2009.

[14] X. Zhang. Fundamentals of Stein's method and its application in proving central limit theorem. *University of Toronto Project*, 2016.