Stein's Method: A Direct Approach for Proving the Central Limit Theorem

By Xuelin Zhu
Email: xuelin@umich.edu

December 25, 2022

Abstract

In 1972, Charles Stein published a proof of the Central Limit Theorem that utilizes a novel approach. The main objective of this project is to understand the fundamental principles of Stein's method, demonstrate its use in establishing the Central Limit Theorem, examine the techniques employed in this method, and evaluate its applications and potential.

Keywords: Stein's equation, Stein's identity, zero bias distribution, normal approximation

1 Introduction

1.1 The Central Limit Theorem

Convergence is a fundamental concept in both probability and statistics, and the Central Limit Theorem (CLT) is a powerful result in this area. Lindeberg-Feller-Lévy Theorem provides a comprehensive version of the CLT stating that for an independently but not necessarily identically distributed random variable sequence $\{X_1, \ldots, X_n\}$ with $EX_i = 0$ and $\sum_{i=1}^n \mathrm{Var}X_i \to \sigma^2$,

$$\text{(Lindeberg's Condition)} \quad \forall \epsilon > 0: \quad \lim_{n\to\infty} \sum_{i=1}^n E\big[X_i^2 I(|X_i| > \epsilon)\big] = 0$$

is necessary and sufficient for both the CLT and Feller's condition to hold, i.e.

$$\text{(CLT)} \qquad S_n := \sum_{i=1}^n X_i \Rightarrow N(0, \sigma^2),$$

$$\text{(Feller's condition)} \qquad \lim_{n\to\infty} \max_{1\le i\le n} EX_i^2 = 0.$$

Typically in a classroom setting, the proof of this powerful result relies on the characteristic function method, which involves the use of Fourier transforms. This proof may not be intuitive from a probabilistic perspective as it involves the manipulation of complex functions in the transform domain. While an alternative method involving the moment-generating function may also be used, this approach is not always feasible as the existence of the moment-generating function may not be guaranteed.

In 1972, Charles Stein[1] introduced an innovative method for proving the Central Limit Theorem, referred to as Stein's method. This method does not require the use of Fourier transforms, instead relying on a direct and intuitive characterization of the distribution. Stein's method has numerous extensions and applications, making it a particularly noteworthy topic for study.

In the following sections, we will delve into the fundamental concepts of Stein's method, provide necessary definitions, demonstrate the application of Stein's method in proving the Central Limit Theorem, present a brief overview of other Stein identities, and discuss the potential of this method.

1.2 Conventions and preliminaries

In the following sections, unless otherwise specified:

- Z is a random variable with the standard normal distribution $N(0,1)$;
- W is the random variable for which we aim to prove asymptotic normality;
- For any given function $h : \mathbb{R} \mapsto \mathbb{R}$, the notation $||\cdot||$ refers to the supremum norm, i.e.

$$||h|| = \sup_{x\in\mathbb{R}} h(x) \; ; \tag{1}$$

- For any given function, we define

$$Nh := Eh(Z).$$

We will also make use of several definitions and properties related to continuity, which are listed here for reference. These concepts are standard in analysis and will not be proven in this report.

Definition 1. *A function $f : \mathbb{R} \mapsto \mathbb{R}$ is Lipschitz continuous if there exists a constant K s.t.*

$$\forall x, y \in \mathbb{R} : \quad |f(x) - f(y)| \leq K|x - y|.$$

A function $f : \mathbb{R} \mapsto \mathbb{R}$ is absolutely continuous on $[a, b]$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that for any finite disjoint intervals $[a_i, b_i] \subset [a, b]$, $i = 1, \dots, N$, if

$$\sum_{i=1}^{N}(b_i - a_i) < \delta \quad \Rightarrow \quad \sum_{i=1}^{N}|f(b_i) - f(a_i)| < \epsilon.$$

Proposition 1. *Properties of Lipschitz and absolute continuity:*

(a) If f is absolutely continuous, then f' exists almost everywhere and

$$f(b) - f(a) = \int_a^b f'(x)dx.$$

(b) If g is integrable, then

$$f(x) = \int_a^x g(t)dt$$

is absolutely continuous.

(c) f absolutely continuous if and only if there exists an integrable function $g \in L_1$ such that

$$f(x) = f(a) + \int_a^x g(t)dt.$$

(d) f is Lipschitz continuous if and only if f is absolutely continuous with a bounded derivative, i.e. the supremum norm $||f'|| < \infty$.

Proposition 2. *If f and g are both non-decreasing functions of x, then*

$$Ef(X)Eg(X) \leq E\big[f(X)g(X)\big].$$

2 Fundamentals of Stein's method

In this section, we will first provide an overview of the general idea of Stein's method for proving convergence in distribution, followed by introducing some fundamental definitions that are essential to this method.

2.1 The general idea of Stein's method

In order to establish the convergence in distribution of a random variable W to a standard normal distribution $Z \sim N(0,1)$, Stein's method proposes considering a class of functions $\mathcal{L}$ such that

$$\forall h \in \mathcal{L}: \quad Eh(W) = Eh(Z) \quad \text{or equivalently} \quad Eh(W) - Eh(Z) = 0\,. \tag{2}$$

If this condition holds for a sufficiently large class of functions, we can infer that W and Z have the same distribution. One example of such a class of functions is $\mathcal{L} = \{h_t(W) = e^{itW} : \ t \in \mathbb{R}\}$, which is utilized in the characteristic function method. However, proving this condition (2) directly can be challenging. Stein's method instead considers a fixed function $h \in \mathcal{L}$ and the following differential equation:

$$f'(w) - wf(w) = h(w) - Eh(Z).$$

If this equation has a corresponding solution f_h, then plugging $w = W$ and taking expectations of both sides yields

$$E\big[f_h'(W) - Wf_h(W)\big] = Eh(W) - Eh(Z). \tag{3}$$

Therefore, showing that

$$E\big[f_h'(W) - Wf_h(W)\big] = 0,$$

is equivalent to demonstrating that $Eh(W) = Eh(Z)$, which is the central idea behind Stein's method. We will utilize the concept of the zero-bias distribution in order to complete the proof of the CLT using this approach.

2.2 Stein's equation

Definition 2. *For a given real-valued measurable function h with $E|h(Z)| < \infty$, we call*

$$f'(w) - wf(w) = h(w) - Nh \tag{4}$$

the Stein's equation for h.

The introduction of Stein's equation, as described in equation (3), is motivated by the desire to replace $Eh(W) - Eh(Z)$ with $E[f'(W) - Wf(W)]$. It is important that a solution to this equation exists in order for the method to be effective. Fortunately, the following lemma guarantees the existence of such a solution.

Lemma 3. *For a given real-valued measurable function h with $E|h(Z)| < \infty$, the unique bounded solution f_h to Stein's equation (4) for f is given by*

$$f_h(w) = e^{w^2/2} \int_{-\infty}^{w} \big(h(x) - Nh\big) e^{-x^2/2} dx \tag{5}$$

Proof. To solve Stein's equation (4), we multiply both sides by the integrating factor $e^{-w^2/2}$ and get

$$e^{-w^2/2} f'(w) - we^{-w^2/2} f(w) = e^{-w^2/2}\big(h(w) - Nh\big). \tag{6}$$

By the chain rule, the left hand side (LHS) of (6) equals

$$\frac{d}{dw}\Big(e^{-w^2/2} f(w)\Big) = e^{-w^2/2} f'(w) - we^{-w^2/2} f(w),$$

therefore equation (6) means

$$\frac{d}{dw}\Big(e^{-w^2/2} f(w)\Big) = e^{-w^2/2}\big(h(w) - Nh\big).$$

Integrating both sides from $-\infty$ to w, we get

$$e^{-w^2/2} f_h(w) - \lim_{x \to -\infty} e^{-x^2/2} f_h(x) = \int_{-\infty}^{w} e^{-x^2/2}\big(h(x) - Nh\big) dx + c,$$

where $c \in \mathbb{R}$ is a fixed constant. We shall see later when we add a continuous condition on h, its solution f_h satisfies $e^{-x^2/2} f_h(x) \to 0$ as $x \to -\infty$. Then the last equality becomes

$$\begin{aligned} e^{-w^2/2} f_h(w) &= \int_{-\infty}^{w} e^{-x^2/2}\big(h(x) - Nh\big) dx + c \\ \iff \quad f_h(w) &= e^{-w^2/2} \int_{-\infty}^{w} e^{-x^2/2}\big(h(x) - Nh\big) dx + ce^{-w^2/2}. \end{aligned} \tag{7}$$

Obviously, right hand side (RHS) of equation (7) is bounded only when $c = 0$. □

The solution to Stein's equation, as given in equation (5), possesses several useful properties. First, if the function h is integrable, the absolute continuity of the solution f_h follows from Proposition 1(b). Additionally, for a fixed constant $z \in \mathbb{R}$, setting $h(w) = I(w \le z)$ results in the bounded solution given by

$$f_z(w) = e^{w^2/2} \int_{-\infty}^{w} \big[I(x \le z) - \Phi(z)\big] e^{-x^2/2} dx, \tag{8}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Further properties of this solution are presented in the subsequent lemma.

Lemma 4. *For a given function $h : \mathbb{R} \mapsto \mathbb{R}$, let f_h be the solution to Stein's equation for h. If h is bounded, then*

$$\|f_h\| \le \sqrt{\pi/2}\|h(\cdot) - Nh\| \quad \text{and} \quad \|f_h'\| \le 2\|h(\cdot) - Nh\|. \tag{9}$$

If h is absolutely continuous, then

$$\|f_h\| \le 2\,\|h'\|, \quad \|f_h'\| \le \sqrt{2/\pi}\,\|h'\| \quad \text{and} \quad \|f_h''\| \le 2\,\|h'\|. \tag{10}$$

The proof of Lemma 4 is technically involved and lengthy, but it is not directly relevant to the central ideas of this report. Therefore, we will omit this proof and refer interested readers to the appendix of Chapter 2 of the *Normal Approximation by Stein's method*[2] for further details.

2.3 Characterization of Normal Distribution

In 1972, Charles Stein[1] published a sufficient and necessary condition for a distribution to be normal. This condition can be used to characterize normal distributions.

Lemma 5. $W \sim N(0, \sigma^2)$ *if and only if*

$$\sigma^2 Ef'(W) = E\big[Wf(W)\big] \tag{11}$$

for all absolutely continuous functions f *such that the above expectations exist.*

Proof. We will prove the following lemma for the standard normal case, with the understanding that a similar proof can be constructed for the general case. We first prove $\Rightarrow$ direction. Suppose $W \sim N(0,1)$ and let f be an absolutely continuous function such that $E|f'(W)| < \infty$ and $E\big|Wf(W)\big| < \infty$. It follows that

$$\begin{aligned} Ef'(W) &= \int_{-\infty}^{\infty} f'(w) \times \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw \\ &= \int_{-\infty}^{0} f'(w) \times \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw + \int_{0}^{\infty} f'(w) \times \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw \end{aligned} \tag{12}$$

Note that

$$e^{-w^2/2} = \int_{-\infty}^{w} -xe^{-x^2/2} dx \quad \text{and} \quad e^{-w^2/2} = \int_{w}^{\infty} xe^{-x^2/2} dx.$$

Plugging the two identities into equation (12), it follows that

$$\begin{aligned} Ef'(W) &= \int_{-\infty}^{0} f'(w) \times \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} -xe^{-x^2/2} dx\right) dw + \int_{0}^{\infty} f'(w) \times \left(\frac{1}{\sqrt{2\pi}} \int_{w}^{\infty} xe^{-x^2/2} dx\right) dw \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \int\limits_{w=-\infty}^{w=0} \int\limits_{x=-\infty}^{x=w} -xf'(w)e^{-x^2/2} dx dw + \int\limits_{w=0}^{w=\infty} \int\limits_{x=w}^{x=\infty} xf'(w)e^{-x^2/2} dx dw \right\} \\ \text{(By Fubini)} \quad &= \frac{1}{\sqrt{2\pi}} \left\{ \int\limits_{x=-\infty}^{x=0} \int\limits_{w=x}^{w=0} -xf'(w)e^{-x^2/2} dw dx + \int\limits_{x=0}^{x=\infty} \int\limits_{w=0}^{w=x} xf'(w)e^{-x^2/2} dw dx \right\} \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{0} \left(\int_{x}^{0} f'(w) dw\right) (-x) e^{-x^2/2} dx + \int_{0}^{\infty} \left(\int_{0}^{x} f'(w) dw\right) xe^{-x^2/2} dx \right\} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \big[f(x) - f(0)\big] xe^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} xf(x) \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx - f(0) \int_{-\infty}^{\infty} x \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= E\big[Wf(W)\big] - f(0) \cdot EW \\ \text{(By } EW = 0\text{)} \quad &= E\big[Wf(W)\big]. \end{aligned}$$

Now we prove $\Leftarrow$ direction. Suppose for all absolutely continuous function f, the identity

$$Ef'(W) = E\big[Wf(W)\big]$$

holds. Fixing $z \in \mathbb{R}$, consider the special case $h(w) = I(w \leq z)$. The solution f_z is given in equation (8) and f_z is absolutely continuous. So by condition, the following identity holds for f_z'

$$\forall z \in \mathbb{R}: \quad Ef_z'(W) = E\big[Wf_z(W)\big] \quad \Rightarrow \quad E\big[f_z'(W) - Wf_z(W)\big] = 0. \tag{13}$$

Since f_z' is the solution to Stein's equation (4), it follows that

$$f_z'(W) - Wf_z(W) = I(W \leq z) - \Phi(z). \tag{14}$$

By equation (13), (14), and the arbitrary choice of $z \in \mathbb{R}$, we conclude that

$$\forall z \in \mathbb{R}: \quad E\big[I(W \leq z) - \Phi(z)\big] = 0,$$

which is followed by

$$\forall z \in \mathbb{R}: \quad P(W \leq z) - \Phi(z) = 0,$$

i.e., the c.d.f. of W is the same as that of the standard normal, in other words, $W \sim N(0,1)$. $\square$

2.4 Zero bias distribution

The characterization given in equation (11) holds only when the random variable W is normal. However, we can generalize this idea by introducing the concept of the zero-bias distribution for a random variable with zero mean and finite variance. This allows us to extend the results of this characterization beyond the normal distribution.

Definition 3. *Suppose W is any random variable with $EW = 0$ and $\mathrm{Var}W = \sigma^2$. We say another random variable W^* has the W-zero bias distribution if*

$$\sigma^2 Ef'(W^*) = E\big[Wf(W)\big] \tag{15}$$

for all absolutely continuous functions f such that the above expectations exist.

For illustrative purposes, we provide an example of the zero-bias distribution for a Bernoulli distribution.

Example. Suppose $\xi \sim Ber(p)$ and let $W = \xi - p$. Then for any absolutely continuous function f such that the following expectations exist:

$$\begin{aligned} E\big[Wf(W)\big] &= E\big[(\xi - p)f(\xi - p)\big] \\ &= (0-p)f(0-p)P(\xi = 0) + (1-p)f(1-p)P(\xi = 1) \\ &= -p(1-p)f(-p) + p(1-p)f(1-p) \\ &= \sigma^2\big[f(1-p) - f(-p)\big]. \end{aligned} \tag{16}$$

Since f is absolutely continuous, by Proposition 1(a), f' exists almost everywhere with

$$f(1-p) - f(p) = \int_{-p}^{1-p} f'(u)du.$$

Using this identity, equation (16) indicates that

$$E\big[Wf(W)\big] = \sigma^2 \int_{-p}^{1-p} f'(u)du = \sigma^2 Ef'(U),$$

where $U \sim Uni[-p, 1-p]$. According to the definition (15) of zero bias, we conclude $W^* \stackrel{d}{=} U$.

The definition of the zero-bias distribution given in equation (15) is well-defined, as demonstrated by the following lemma which shows that this distribution exists uniquely.

Proposition 6. *Let W be a random variable with $EW = 0$ and $\mathrm{Var}W = \sigma^2 < \infty$. Then the W-zero bias distribution exists and is unique. Moreover, the distribution of W^* is absolutely continuous with the density*

$$p^*(x) = E\big[WI(W > x)\big]/\sigma^2 \quad \textit{or equivalently} \quad p^*(x) = -E\big[WI(W \le x)\big]/\sigma^2, \tag{17}$$

and the c.d.f.

$$G^*(x) = E\big[W(W-x)I(W \le x)\big]/\sigma^2. \tag{18}$$

Proof. We will complete the proof for $\sigma = 1$, and the general case could be done similarly. We first notice that the two expressions in (17) are the same since for all $x \in \mathbb{R}$,

$$EW = 0 \quad \Rightarrow \quad E\big[WI(W > x) + WI(W \le x)\big] = 0 \quad \Rightarrow \quad E\big[WI(W > x)\big] = -E\big[WI(W \le x)\big].$$

With the observations below:

$$\begin{cases} E\big[WI(W > x)\big] \ge 0 & \text{when } x > 0, \\ -E\big[WI(W \le x)\big] \ge 0 & \text{when } x \le 0, \end{cases}$$

we claim that

$$\forall x \in \mathbb{R}: \quad E\big[WI(W > x)\big] = -E\big[WI(W \le x)\big] \ge 0,$$

i.e., the two expressions in (17) are equal and non-negative on the whole $\mathbb{R}$. We denote the zero-bias distribution as $p^*(x)$, and it remains to show that:

(a) $p^*(\cdot)$ integrates to 1, thus forming a density;

(b) p^* satisfies the definition of the zero-bias distribution given in equation (15), implying that W^* follows the W-zero bias distribution;

(c) the zero-bias distribution is unique;

(d) the c.d.f. of the zero-bias distribution is given by equation (18).

Suppose f is an absolutely continuous function on $\mathbb{R}$. By Proposition 1(c), there exists an integrable function $g \in L_1$ such that

$$f(x) = \int_0^x g(u)du + c, \tag{19}$$

where $c \in \mathbb{R}$ is a constant and $f' = g$ almost everywhere on $\mathbb{R}$. Then it follows that

$$
\begin{aligned}
\int_0^\infty f'(u)E\big[WI(W>u)\big]du &= \int\limits_{u=0}^{u=\infty} g(u)\left(\int\limits_{\{W>u\}} W\,dP\right)du \\
\text{(By Fubini)} \quad &= \int\limits_{\{0<W<\infty\}} \left(\int\limits_{u=0}^{u=W} Wg(u)du\right)dP \\
&= \int\limits_{\{0<W<\infty\}} W\left(\int_0^W g(u)du\right)dP \\
\Big(\text{By absolute continuity (19)}\Big) \quad &= \int\limits_{\{0<W<\infty\}} W\big[f(W)-c\big]dP \\
&= E\big[Wf(W)I(W>0)\big] - c\cdot E\big[WI(W>0)\big].
\end{aligned} \tag{20}
$$

Similarly, we can conclude that

$$
\int_{-\infty}^0 f'(u)E\big[WI(W>u)\big]du = E\big[Wf(W)I(W\le 0)\big] - c\cdot E\big[WI(W\le 0)\big]. \tag{21}
$$

Summing up (20) and (21), for any absolutely continuous function f, we have

$$
\begin{aligned}
\int_{-\infty}^\infty f'(u)E\big[WI(W>u)\big]du &= \int_{-\infty}^0 f'(u)E\big[WI(W>u)\big]du + \int_0^\infty f'(u)E\big[WI(W>u)\big]du \\
&= E\Big\{Wf(W)\big[I(W>0)+I(W\le 0)\big]\Big\} - c\cdot E\Big\{W\big[I(W>0)+I(W\le 0)\big]\Big\} \\
&= E\big[Wf(W)\big] - c\cdot EW \\
\text{(Since } EW=0) \quad &= E\big[Wf(W)\big]
\end{aligned} \tag{22}
$$

Now we let $f(x) = x$ particularly, then equation (22) means

$$
\int_{-\infty}^\infty p^*(u)du = \int_{-\infty}^\infty E\big[WI(W>u)\big]du = E(W^2) = \sigma^2 = 1,
$$

i.e. (a) is proved and p^* is a density. To prove (b), suppose W^* has the distribution with density p^*. Return to equation (22), for any absolutely continuous function f,

$$
Ef'(W^*) = \int_{-\infty}^\infty f'(u)p^*(u)du = \int_{-\infty}^\infty f'(u)E\big[WI(W>u)\big]du = E\big[Wf(W)\big].
$$

Then (b) is done. To prove (c), suppose there are two zero bias distributions, say W_1^* and W_2^*. Then by the definition of zero bias, for all absolutely continuous functions f:

$$
Ef'(W_1^*) = E\big[Wf(W)\big] = Ef'(W_2^*).
$$

This implies $W_1^* \overset{d}{=} W_2^*$. Now it remains to calculate its c.d.f. to get (d) and this can be checked by

$$\begin{aligned}
\int_{-\infty}^{x} p^*(u)du &= \int_{-\infty}^{x} -E\big[WI(W \le u)\big]du \\
&= -\int\limits_{u=-\infty}^{u=x} \left(\int\limits_{\{W \le u\}} WdP \right) du \\
\text{(By Fubini)} \quad &= -\int\limits_{\{W \le x\}} \left(\int\limits_{u=W}^{u=x} Wdu \right) dP \\
&= -E\big[W(x-W)I(W \le x)\big] = E\big[W(W-x)I(W \le x)\big].
\end{aligned}$$

□

As a reminder, our ultimate goal is to prove the Central Limit Theorem. In order to use the zero-bias distribution as a tool in this proof, we need to find a random variable with the zero-bias distribution of the sum of independent random variables. This random variable can be constructed as follows.

Lemma 7. *Suppose $\{X_i,\ i = 1, \dots, n\}$ are independent random variables with*

$$EX_i = 0, \quad \mathrm{Var}X_i = \sigma_i^2, \quad \text{and} \quad \sum_{i=1}^{n} \sigma_i^2 = 1.$$

Set

$$W = \sum_{i=1}^{n} X_i.$$

Let X_i^* have the X_i-zero bias distribution with $\{X_i^*\ i = 1, \dots, n\}$ mutually independent and X_i^* independent of X_j for all $i \ne j$. Further, let I be a random index, independent of X_i's and X_i^*'s, with distribution

$$P(I = i) = \sigma_i^2. \tag{23}$$

Then

$$W^* \overset{d}{=} W - X_I + X_I^*,, \tag{24}$$

i.e., $W - X_I + X_I^*$ has the W-zero bias distribution.

Proof. To avoid confusion, we use $\mathbf{1}(\cdot)$ as the indicator function and I as the random variable defined in (23) in this proof. Since the index is random, we can write

$$X_I = \sum_{i=1}^{n} \mathbf{1}(I = i)X_i \quad \text{and} \quad X_I^* = \sum_{i=1}^{n} \mathbf{1}(I = i)X_i^*.$$

All we need to prove is W^* defined in equation (24) satisfies the definition (15) of zero bias. Suppose f is an absolutely continuous function. It follows that

$$E\big[Wf(W)\big] = E\big[\sum_{i=1}^{n} X_i f(W)\big] = \sum_{i=1}^{n} E\big[X_i f(W)\big]. \tag{25}$$

Since $f(\cdot)$ is absolutely continuous, the function $g(X_i^*) = f(W - X_i + X_i^*)$ is also absolutely continuous with respect to X_i^*, with the derivative

$$g'(X_i^*) = f'(W - X_i + X_i^*).$$

Since X_i^* follows X_i-zero bias distribution and $g(\cdot)$ is absolutely continuous, it follows that

$$\sigma_i^2 E g'(X_i^*) = E\big[X_i g(X_i)\big] \quad \Longleftrightarrow \quad \sigma_i^2 E f'(W - X_i + X_i^*) = E\big[X_i f(W - X_i + X_i)\big] = E\big[X_i f(W)\big]$$

Plugging this identity into equation (25), it follows that

$$\begin{aligned} E\big[W f(W)\big] &= \sum_{i=1}^n \sigma_i^2 E\big[f'(W - X_i + X_i^*)\big] \\ \text{(By the distribution (23) of } I \text{)} \quad &= \sum_{i=1}^n P(I = i) E\big[f'(W - X_i + X_i^*)\big] \\ \text{(Conditional expectation of discrete r.v. } I \text{)} \quad &= E\big[f'(W - X_I + X_I^*)\big]. \end{aligned} \tag{26}$$

Since f is an arbitrary absolutely continuous function, equation (26) concludes our result. $\square$

If X_i and X_i^* are defined on the same probability space, then we can do subtraction. Define

$$\Delta = W^* - W \quad \text{or equivalently} \quad \Delta = X_I^* - X_I. \tag{27}$$

Using this notation, (26) means

$$E\big[W f(W)\big] = E f'(W + \Delta). \tag{28}$$

for all absolutely continuous functions f.

3 Proof of the Central Limit Theorem

In this section, we will demonstrate the application of Stein's method in establishing the Central Limit Theorem. Before moving on to the details, it is helpful to review the definitions and concepts that have been introduced thus far and to consider their relevance to the proof of the CLT. Specifically, for any measurable function $h : \mathbb{R} \to \mathbb{R}$, we have defined the solution to Stein's equation (29) as f, which satisfies

$$f'(w) - w f(w) = h(w) - Nh, \tag{29}$$

where $Nh := Eh(Z)$ and Z follows the standard normal distribution. Substituting $w = W$ into this equation and taking expectations on both sides yields

$$E\big[f'(W) - W f(W)\big] = Eh(W) - Nh. \tag{30}$$

By using the identity given in equation (28), we can rewrite (30) as

$$Eh(W) - Nh = E\big[f'(W) - f'(W + \Delta)\big]. \tag{31}$$

If the RHS of this equation (31) is small, it may suggest that W and Z are approximately equal. This idea will be elaborated in this section.

3.1 Continuity and Berry Esseen bounds

It is not possible to control $E\big[f'(W) - f'(W+\Delta)\big]$ for all measurable functions h. However, for functions with Lipschitz continuity, this approach can be effective. For convenience, we denote

$$\beta_2 := \sum_{i=1}^{n} E\Big\{X_i^2 I(|X_i| > 1)\Big\} \quad \text{and} \quad \beta_3 := \sum_{i=1}^{n} E\Big\{|X_i|^3 I(|X_i| \le 1)\Big\}.$$

Theorem 8. *If $W = \sum_{i=1}^n X_i$ is the sum of independent random variables with*

$$EX_i = 0, \quad \mathrm{Var} X_i = \sigma_i^2, \quad \textit{and} \quad \sum_{i=1}^{n} \sigma_i^2 = 1,$$

then for any Lipschitz continuous function h

$$|Eh(W) - Nh| \le ||h'||(4\beta_2 + 3\beta_3). \tag{32}$$

Proof. Let f be the solution to the Stein's equation (29) for h. By triangle inequality and the mean value theorem:

$$|f'(W) - f'(W+\Delta)| \le \Big(2 \cdot ||f'||\Big) \wedge \Big(||f''|| \cdot |\Delta|\Big), \tag{33}$$

where $a \wedge b := \min(a, b)$. Recall the properties (10) of the Stein's equation solutions, we have

$$||f'|| \le \sqrt{2/\pi}||h'|| \quad \text{and} \quad ||f''|| \le 2||h'||.$$

Plugging them in equation (33), we have

$$|f'(W) - f'(W+\Delta)| \le \Big(2\sqrt{2/\pi} \cdot ||h'||\Big) \wedge \Big(2||h'|| \cdot |\Delta|\Big) \le 2||h'|| \cdot \Big(1 \wedge |\Delta|\Big), \tag{34}$$

Using equation (31), it follows that

$$\begin{aligned}
|Eh(W) - Nh| &= \Big|E\big[f'(W) - f'(W+\Delta)\big]\Big| \\
\Big(\text{By equation (34)}\Big) \quad &\le \Big|E\Big\{2||h'|| \cdot \Big(1 \wedge |\Delta|\Big)\Big\}\Big| \\
&\le 2||h'|| \cdot E\Big|1 \wedge |\Delta|\Big| \\
\Big(\text{Plugging in } \Delta \text{ of (27)}\Big) \quad &= 2||h'|| \cdot E\Big|1 \wedge |X_I^* - X_I|\Big| \\
(\text{By triangle inequality}) \quad &\le 2||h'|| \cdot E\Big\{1 \wedge \Big(|X_I^*| + |X_I|\Big)\Big\} \\
&\le 2||h'|| \cdot E\Big\{\Big(1 \wedge |X_I^*|\Big) + \Big(1 \wedge |X_I|\Big)\Big\}.
\end{aligned} \tag{35}$$

We will bound the two terms on the RHS of equation (35) one by one. Let $sgn(x)$ be the sign of $x \in \mathbb{R}$, and set

$$f(x) = xI(|x| > 1) + \frac{1}{2}x^2 sgn(x) I(|x| \le 1) \quad \Rightarrow \quad f'(x) = 1 \wedge |x|.$$

One can check that f is absolutely continuous. Then using definition (15) of zero bias and plugging in the f defined above, we get for all $i \in \{1, 2, \ldots, n\}$

$$\sigma_i^2 E f'(X_i^*) = E\big[X_i f(X_i)\big] \quad \Rightarrow \quad \sigma_i^2 E\Big(1 \wedge |X_i^*|\Big) = E\Big\{X_i^2 I(|X_i| > 1) + \frac{1}{2}|X_i|^3 I(|X_i| \le 1)\Big\} \tag{36}$$

Recall that I is distributed with $P(I=i)=\sigma_i^2$. It then follows that

$$
\begin{aligned}
E\Big(1\wedge|X_I^*|\Big) &= \sum_{i=1}^n E\Big(1\wedge|X_i^*|\Big)\cdot P(I=i) = \sum_{i=1}^n \sigma_i^2 E\Big(1\wedge|X_i^*|\Big) \\
\text{(By equation (36))} \quad &= \sum_{i=1}^n E\Big\{X_i^2 I(|X_i|>1) + \frac{1}{2}|X_i|^3 I(|X_i|\le 1)\Big\} \\
&= \beta_2 + \frac{1}{2}\beta_3 \qquad (37)
\end{aligned}
$$

For the other term in equation (35), by $P(I=i)=\sigma_i^2 = EX_i^2$, it follows that

$$
\begin{aligned}
E\Big(1\wedge|X_I|\Big) &= \sum_{i=1}^n E\big(1\wedge|X_i|\big)\cdot P(I=i) \\
&= \sum_{i=1}^n E\big(1\wedge|X_i|\big)\cdot EX_i^2 \\
\text{(By Proposition 2)} \quad &\le \sum_{i=1}^n E\Big\{\Big(1\wedge|X_i|\Big)X_i^2\Big\},
\end{aligned}
$$

where the last inequality holds for the fact that $1\wedge|x|$ and $|x|^2$ are non-decreasing w.r.t. $|x|$. Note that

$$
\Big(1\wedge|X_i|\Big)X_i^2 = X_i^2 I(|X_i|>1) + |X_i|^3 I(|X_i|<1).
$$

So, it follows that

$$
E\Big(1\wedge|X_I|\Big) \le \sum_{i=1}^n E\Big\{X_i^2 I(|X_i|>1) + |X_i|^3 I(|X_i|\le 1)\Big\} = \beta_2+\beta_3. \qquad (38)
$$

By equation (35), (37) and (38), we get

$$
|Eh(W)-Nh| \le ||h'||(4\beta_2+3\beta_3).
$$

□

Theorem 8 provides an upper bound for the difference between $Eh(W)$ and $Eh(Z)$. The next theorem states that this bound on expectations can be used to obtain an upper Berry-Esseen bound, which refers to an upper bound for the error in the normal approximation of a random variable W in the form

$$
\sup_{z\in\mathbb{R}} |P(W\le z)-\Phi(z)| \le c.
$$

Theorem 9. *For any random variable W and the standard normal Z, if there exists a $\delta>0$ such that, for any Lipschitz continuous function h,*

$$
|Eh(W)-Nh| \le \delta||h'||, \qquad (39)
$$

then

$$
\sup_{z\in\mathbb{R}} |P(W\le z)-\Phi(z)| \le 2\delta^{1/2}. \qquad (40)
$$

Proof. Note that $P(W\le z)$ and $\Phi(z)$ are two probabilities, which means $0\le P(\cdot)\le 1$ and so is $\Phi(\cdot)$. Then the LHS of (40) is no greater than 1. If $\delta>1/4$, equation (40) holds naturally. So, we can assume $\delta\le 1/4$.

Let $\alpha = \delta^{1/2}(2\pi)^{1/4}$, fix $z \in \mathbb{R}$ and consider the following function:

$$h_\alpha(w) = \begin{cases} 1 & \text{if } w \leq z, \\ 1 + (z-w)/\alpha & \text{if } z < w \leq z + \alpha, \\ 0 & \text{if } w > z + \alpha. \end{cases}$$

One can check that h_α is Lipschitz continuous with $||h'|| = 1/\alpha$, and $I(W \leq z) \leq h_\alpha(W) \leq I(W \leq z + \alpha)$, and $I(Z \leq z) \leq h_\alpha(Z) \leq I(Z \leq z + \alpha)$. Then it follows that

$$\begin{aligned}
P(W \leq z) - \Phi(z) &= EI(W \leq z) - P(Z \leq z) \\
&\leq Eh_\alpha(W) - P(Z \leq z) \\
&= Eh_\alpha(W) - Nh_\alpha + Nh_\alpha - P(Z \leq z) \\
\text{(By definition of } Nh) \quad &= Eh_\alpha(W) - Nh_\alpha + Eh_\alpha(Z) - P(Z \leq z) \\
\Big(\text{By condition (39) and } h_\alpha(Z) \leq I(Z \leq z + \alpha)\Big) \quad &\leq \delta ||h'_\alpha|| + P(Z \leq z + \alpha) - P(Z \leq z) \\
\text{(By } ||h'_\alpha|| = 1/\alpha) \quad &= \frac{\delta}{\alpha} + P(z \leq Z \leq z + \alpha) \\
\text{(By the density of standard normal)} \quad &\leq \frac{\delta}{\alpha} + \frac{\alpha}{\sqrt{2\pi}} \\
\text{(Plugging in } \alpha = \delta^{1/2}(2\pi)^{1/4}) \quad &= \delta^{1/2}(2\pi)^{-1/4} + \delta^{1/2}(2\pi)^{-1/4} \\
&= 2\delta^{1/2}(2\pi)^{-1/4} \leq 2\delta^{1/2}.
\end{aligned}$$

If we take

$$h_\alpha(w) = \begin{cases} 1 & \text{if } w \leq z - \alpha, \\ 1 + (z-w)/\alpha & \text{if } z - \alpha < w \leq z, \\ 0 & \text{if } w > z, \end{cases}$$

by a similar argument, we get

$$P(W \leq z) - \Phi(z) \geq -2\delta^{1/2},$$

which concludes that for all $z \in \mathbb{R}$,

$$|P(W \leq z) - \Phi(z)| \leq 2\delta^{1/2}.$$

□

3.2 The sufficiency of Lindeberg's condition

Now we will demonstrate the Lindeberg-Feller-Lévy Central Limit Theorem, specifically the implication

$$\text{(Lindeberg Condition)} \quad \Rightarrow \quad \text{(CLT)}.$$

This result follows as a special case of Theorems 8 and 9. We will not prove the implication

$$\text{(Lindeberg Condition)} \quad \Rightarrow \quad \text{(Feller's Condition)}$$

as it is provided in the lecture notes.

Proposition 10. *Suppose $\{X_1, \ldots, X_n\}$ are independent random variables with*

$$EX_i = 0, \quad \mathrm{Var} X_i = \sigma_i^2, \quad \textit{and} \quad \sum_{i=1}^n \sigma_i^2 = 1.$$

Set $W = \sum_{i=1}^n X_i$. If Lindeberg's condition holds, i.e., for all $\epsilon > 0$,

$$\sum_{i=1}^n E\Big\{X_i^2 I\big(|X_i| > \epsilon\big)\Big\} \to 0 \quad \textit{as } n \to \infty, \tag{41}$$

then

$$W \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty. \tag{42}$$

Proof. We keep using the notation of β_2 and β_3,

$$\beta_2 := \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > 1)\Big\} \quad \text{and} \quad \beta_3 := \sum_{i=1}^n E\Big\{|X_i|^3 I(|X_i| \le 1)\Big\}.$$

Now fix $\epsilon \in (0,1)$, and it follows that

$$\begin{aligned}
\beta_2 + \beta_3 &= \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > 1)\Big\} + \sum_{i=1}^n E\Big\{|X_i|^3 I(|X_i| \le 1)\Big\} \\
\text{(By choosing } 0 < \epsilon < 1) \quad &\le \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > 1)\Big\} + \sum_{i=1}^n E\Big\{\epsilon \cdot X_i^2 I(|X_i| \le \epsilon) + 1 \cdot X_i^2 I(\epsilon < |X_i| \le 1)\Big\} \\
&= \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > \epsilon)\Big\} + \epsilon \cdot \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| \le \epsilon)\Big\} \\
&\le \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > \epsilon)\Big\} + \epsilon \cdot \sum_{i=1}^n EX_i^2 \\
\Big(\text{By } \sum_{i=1}^n \sigma_i^2 = 1\Big) \quad &= \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > \epsilon)\Big\} + \epsilon.
\end{aligned} \tag{43}$$

If Lindeberg's condition (41) holds, taking limits of $n \to \infty$ on both sides of (43), it follows that:

$$\forall \epsilon \in (0,1): \quad \lim_{n\to\infty} \big(\beta_2 + \beta_3\big) \le \epsilon \quad \Rightarrow \quad \beta_2 + \beta_3 \to 0.$$

Recall that (32) in Theorem 8 implies for any Lipschitz continuous function h,

$$|Eh(W) - Nh| \le ||h'||(4\beta_2 + 3\beta_3) \le 4||h'||(\beta_2 + \beta_3),$$

and (40) in Theorem 9 further implies

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le 8(\beta_2 + \beta_3) \to 0 \quad \text{as } n \to 0.$$

This implies the c.d.f. of W converges weakly to that of Z, the standard normal distribution. $\square$

Proposition 10 demonstrates that, when $\sum_{i=1}^n \mathrm{Var} X_i = 1$, Lindeberg's condition implies the CLT. By making a simple change of variable, it can be shown that this result also holds when $\sum_{i=1}^n \mathrm{Var} X_i = \sigma^2 < \infty$.

3.3 The necessity of Lindeberg's condition

If we assume that Feller's condition holds, i.e.

$$\lim_{n\to\infty} \max_{1\le i\le n} EX_i^2 = 0,$$

we can also use Stein's method to prove that the CLT implies Lindeberg's condition. To do so, we need a lower Berry-Esseen bound, which refers to a lower bound for the error in the normal approximation of a random variable W in the form

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \ge c.$$

This bound is guaranteed in the next theorem.

Theorem 11. *Suppose $\{X_1, \dots, X_n\}$ are independent random variables with*

$$EX_i = 0, \quad \mathrm{Var} X_i = \sigma_i^2, \quad \textit{and} \quad \sum_{i=1}^n \sigma_i^2 = 1 \,.$$

Set $W = \sum_{i=1}^n X_i$. Then there exists an absolute constant C such that for all $\epsilon > 0$,

$$\Big(1 - e^{-\epsilon^2/4}\Big) \sum_{i=1}^n E\Big\{X_i^2 I(|X_i| > \epsilon)\Big\} \le C\Big(\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| + \sum_{i=1}^n \sigma_i^4\Big). \tag{44}$$

Proof. For any given measurable function h, denote f as the solution to Stein's equation of h. By the same argument of (30),

$$E\big[f'(W) - Wf(W)\big] = Eh(W) - Nh. \tag{45}$$

The idea of this proof is to choose a proper h and use some inequalities to derive the lower bound.

We first restrict the choice of h to be bounded and absolutely continuous. By Proposition 1(a), h' exists almost everywhere. And we further restrict h with

$$\int_{-\infty}^{\infty} |h'(w)| dw < \infty.$$

With this condition, the following identity holds

$$\begin{aligned}
\int_{-\infty}^{\infty} h'(w)\big[P(W \le w) - \Phi(w)\big] dw &= \int_{-\infty}^{\infty} \big[P(W \le w) - \Phi(w)\big] dh(w) \\
\text{(Integration by parts)} \quad &= h(w)\big[P(W \le w) - \Phi(w)\big]\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(w) d\big\{P(W \le w) - \Phi(w)\big\} \\
&= 0 \cdot \lim_{w\to\infty} h(w) - 0 \cdot \lim_{w\to-\infty} h(w) - \int_{-\infty}^{\infty} h(w) dF_Z + \int_{-\infty}^{\infty} h(w) dF_W \\
(\because h \text{ is bounded}) \quad &= Eh(W) - Nh,
\end{aligned}$$

and the following bound follows

$$\left|\int_{-\infty}^{\infty} h'(w)\big[P(W \le w) - \Phi(w)\big] dw\right| \le \sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \times \int_{-\infty}^{\infty} |h'(w)| dw.$$

For convenience, we denote

$$\delta := \sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)|,$$

then we have

$$|Eh(W) - Nh| \le \delta \int_{-\infty}^{\infty} |h'(w)| dw. \tag{46}$$

The bound (46) is not enough. Now we are going to bound the LHS of equation (45). Let $W^{(i)} = W - X_i$, and $W^{(i)}$ is independent of X_i. Then we have

$$\begin{aligned} E\big[Wf(W)\big] &= \sum_{i=1}^{n} E\big[X_i f(W)\big] \\ (\because X_i \perp\!\!\!\perp W^{(i)} \text{ and } EX_i = 0) \quad &= \sum_{i=1}^{n} E\big[X_i f(W) - X_i f(W^{(i)})\big] \\ &= \sum_{i=1}^{n} E\big[X_i f(W) - X_i f(W^{(i)}) - X_i^2 f'(W^{(i)})\big] + \sum_{i=1}^{n} E\big[X_i^2 f'(W^{(i)})\big] \end{aligned} \tag{47}$$

and with $\sum_{i=1}^{n} \sigma_i^2 = 1$, we also have

$$\begin{aligned} Ef'(W) = 1 \times Ef'(W) &= \sum_{i=1}^{n} \sigma_i^2 \times Ef'(W) \\ &= \sum_{i=1}^{n} \Big\{\sigma_i^2 E\big[f'(W) - f'(W^{(i)})\big]\Big\} + \sum_{i=1}^{n} \sigma_i^2 Ef'(W^{(i)}) \\ (\because X_i \perp\!\!\!\perp W^{(i)} \text{ and } EX_i^2 = \sigma_i^2) \quad &= \sum_{i=1}^{n} \Big\{\sigma_i^2 Ef'(W) - E\big[X_i^2 f'(W^{(i)})\big]\Big\} + \sum_{i=1}^{n} E\big[X_i^2 f'(W^{(i)})\big]. \end{aligned} \tag{48}$$

Equation (48) subtracting (47), we get

$$E\big[f'(W) - Wf(W)\big] = \sum_{i=1}^{n} \Big\{\sigma_i^2 Ef'(W) + X_i f(W^{(i)}) - X_i f(W)\Big\}. \tag{49}$$

For convenience, denote

$$g(w, y) = -\frac{1}{y}\Big[f(w + y) - f(w) - yf'(w)\Big], \tag{50}$$

then

$$E\Big[X_i^2 g(W^{(i)}, X_i)\Big] = E\Big\{X_i\big[f(W^{(i)}) + X_i f'(W^{(i)}) - f(W)\big]\Big\}. \tag{51}$$

Equation (49) subtracting the summation of (51), we get

$$\begin{aligned} E\big[f'(W) - Wf(W)\big] - \sum_{i=1}^{n} E\big[X_i^2 g(W^{(i)}, X_i)\big] &= \sum_{i=1}^{n} \Big\{\sigma_i^2 Ef'(W) - E\big[X_i^2 f'(W^{(i)})\big]\Big\} \\ (\because X_i \perp\!\!\!\perp W^{(i)} \text{ and } EX_i^2 = \sigma_i^2) \quad &= \sum_{i=1}^{n} \Big\{\sigma_i^2 E\big[f'(W) - f'(W^{(i)})\big]\Big\}. \end{aligned} \tag{52}$$

Now recall that $W = W^{(i)} + X_i$, so we can use Taylor expansion and get:

$$f'(W) - f'(W^{(i)}) = f'(W^{(i)} + X_i) - f'(W^{(i)}) = f''(W^{(i)})X_i + \frac{f'''(\xi_i)}{2}X_i^2,$$

where ξ_i is a random number between $W^{(i)}$ and $W^{(i)} + X_i$. Taking expectation on both sides, it follows that

$$(\because EX_i = 0) \qquad E\Big[f'(W) - f'(W^{(i)})\Big] = \frac{f'''(\xi_i)}{2}EX_i^2.$$

Using this identity, we can get a bound for (52),

$$\begin{aligned}
\left| E\big[f'(W) - Wf(W)\big] - \sum_{i=1}^n E\big[X_i^2 g(W^{(i)}, X_i)\big] \right| &\leq \sum_{i=1}^n \left\{ \sigma_i^2 \cdot \left| E\Big[f'(W) - f'(W^{(i)})\Big] \right| \right\} \\
&\leq \sum_{i=1}^n \sigma_i^2 \cdot \left| \frac{f'''(\xi_i)}{2}EX_i^2 \right| \\
&\leq \frac{1}{2}||f'''|| \sum_{i=1}^n \sigma_i^4. \qquad (53)
\end{aligned}$$

For convenience, we further denote

$$R_1 := \sum_{i=1}^n E\big[X_i^2 g(W^{(i)}, X_i)\big] \quad \text{and} \quad R := \sum_{i=1}^n E\big[X_i^2 g(Z, X_i)\big]. \qquad (54)$$

Note that with equation (46) and (53), we get the bound:

$$\sup_{z\in\mathbb{R}} |P(W \leq z) - \Phi(z)| := \delta \geq \frac{R_1 - ||f'''|| \sum_{i=1}^n \sigma_i^4/2}{\int_{-\infty}^{\infty} |h'(w)|dw}, \qquad (55)$$

where f is the solution to Stein's equation for h, so all elements in the RHS of (55) depends only on the choice of h. And since the inequality (55) holds for all bounded, absolutely continuous function h with

$$E\big[f'(W) - Wf(W)\big] = Eh(W) - Nh,$$

by the genius choice of

$$h(w) = (1 - 2w^2)e^{-w^2/2},$$

the lower bound could be derived. However, this is a non-trivial task, as it involves finding a lower bound for the quantity R_1 by considering the relationship between R and R_1 and finding a lower bound for R.

One can check that using this h, the following facts follow:

Fact 1: the solution to Stein's equation for h is $f(w) = we^{-w^2/2}$, and $f'(w) = (1 - w^2)e^{-w^2/2}$.

Fact 2: $c_1 := \int_{-\infty}^{\infty} |h'(w)|dw \leq 7$, $\quad c_2 := \int_{-\infty}^{\infty} |f''(w)|dw \leq 4$, $\quad$ and $\quad c_3 := ||f'''|| = 3$.

Fact 3: using this h and f, the g defined in (50) is non-decreasing.

And using this h and the definition (50) of function g, we have

$$\begin{aligned}
Eg(Z,y) &= \int_{-\infty}^{\infty} g(Z,y)d\Phi \\
\text{(Plugging } g\text{)} \quad &= \int_{-\infty}^{\infty} -\frac{1}{y}\Big[f(Z+y)-f(Z)-yf'(Z)\Big]d\Phi \\
\text{(Plugging } f\text{)} \quad &= -\frac{1}{y}\int_{-\infty}^{\infty}\Big[(z+y)e^{-(z+y)^2/2} - ze^{-z^2/2} - y(1-z^2)e^{-z^2/2}\Big]\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \\
\text{(By calculation)} \quad &= \frac{1}{2\sqrt{2}}\big(1-e^{-y^2/4}\big). \qquad (56)
\end{aligned}$$

By the simplified expression (56), we can conclude that for a fixed $\epsilon > 0$,

$$\text{whenever } |y| \geq \epsilon: \quad Eg(Z,y) \geq \frac{1}{2\sqrt{2}}\big(1-e^{-\epsilon^2/4}\big), \qquad (57)$$

so with the definition (54) of R, it follows that

$$\begin{aligned}
R &:= \sum_{i=1}^{n} E\big[X_i^2 g(Z,X_i)\big] \\
&\geq \sum_{i=1}^{n} E\big[X_i^2 g(Z,X_i) I^2(|X_i|\geq\epsilon)\big] \\
\text{(By Proposition 2)} \quad &\geq \sum_{i=1}^{n}\Big\{E\big[X_i^2 I(|X_i|\geq\epsilon)\big]\times E\big[g(Z,X_i)I(|X_i|\geq\epsilon)\big]\Big\} \\
\Big(\text{By equation (57)}\Big) \quad &\geq \frac{1}{2\sqrt{2}}\big(1-e^{-\epsilon^2/4}\big)\sum_{i=1}^{n} E\big[X_i^2 I(|X_i|\geq\epsilon)\big]. \qquad (58)
\end{aligned}$$

So R has a lower bound now and it remains to find the relation between R and R_1. Let W' be a random variable that has the same distribution as W, but is independent of all X_i's, and define

$$R_2 := \sum_{i=1}^{n} E\big[X_i^2 g(W',X_i)\big].$$

Recall the definition (54) of R_1, and it follows that

$$\begin{aligned}
R_1 &= \sum_{i=1}^{n} E\big[X_i^2 g(W^{(i)},X_i)\big] \\
\text{(By definition (50) of } g\text{)} \quad &= -\sum_{i=1}^{n} E\Big\{X_i^2 \times \frac{1}{X_i}\Big[f(W)-f(W^{(i)})-X_i f'(W^{(i)})\Big]\Big\}. \qquad (59)
\end{aligned}$$

Note that we can write the RHS of (59) in an integral form:

$$\begin{aligned}
\int_0^1 f'(W^{(i)}+tX_i) - f'(W^{(i)})dt &= \frac{1}{X_i}f(W^{(i)}+tX_i)\Big|_{t=0}^{t=1} - f'(W^{(i)})\Big|_{t=0}^{t=1} \\
&= \frac{1}{X_i}\Big[f(W)-f(W^{(i)})-X_i f'(W^{(i)})\Big].
\end{aligned}$$

Therefore equation (59) becomes

$$R_1 = -\sum_{i=1}^{n} E\Big\{X_i^2 \int_0^1 \Big[f'(W^{(i)} + tX_i) - f'(W^{(i)})\Big]dt\Big\}. \tag{60}$$

Similarly we can write

$$R_2 = \sum_{i=1}^{n} E\big[X_i^2 g(W', X_i)\big] = -\sum_{i=1}^{n} E\Big\{X_i^2 \int_0^1 \Big[f'(W' + tX_i) - f'(W')\Big]dt\Big\}. \tag{61}$$

By the expression (60) and (61), algebraic simplification leads to

$$\begin{aligned} R_1 = R_2 + \sum_{i=1}^{n} & E\Big\{X_i^2 \int_0^1 \Big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\Big]dt\Big\} \\ & - \sum_{i=1}^{n} E\Big\{X_i^2 \int_0^1 \Big[f'(W') - f'(W^{(i)})\Big]dt\Big\}. \end{aligned} \tag{62}$$

By Taylor expansion, for any $\theta \in \mathbb{R}$, there exists a random number ξ_i between $W^{(i)} + \theta$ and $W + \theta$ such that

$$\begin{aligned} & f'(W + \theta) - f'(W^{(i)} + \theta) = X_i f''(W^{(i)} + \theta) + \frac{X_i^2}{2} f'''(\xi_i), \\ (\because EX_i = 0) \quad \Rightarrow \quad & \Big|E\big[f'(W + \theta) - f'(W^{(i)} + \theta)\big]\Big| \le \frac{||f'''||}{2}\sigma_i^2 = \frac{1}{2}c_3\sigma_i^2. \end{aligned}$$

Recall that, by our choice, W and W' have the same distribution, so it follows that

$$\Big|E\big[f'(W' + \theta) - f'(W^{(i)} + \theta)\big]\Big| = \Big|E\big[f'(W + \theta) - f'(W^{(i)} + \theta)\big]\Big| \le \frac{1}{2}c_3\sigma_i^2$$

Letting $\theta = 0$ and $\theta = tX_i$ in particular, the inequality indicates that

$$-\frac{1}{2}c_3\sigma_i^2 \le E\Big[f'(W') - f'(W^{(i)})\Big] \le \frac{1}{2}c_3\sigma_i^2, \tag{63}$$

$$-\frac{1}{2}c_3\sigma_i^2 \le E\Big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\Big] \le \frac{1}{2}c_3\sigma_i^2. \tag{64}$$

Therefore, we have

$$\begin{aligned} E\left\{X_i^2 \int_0^1 \Big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\Big]dt\right\} &= E\left\{\int_0^1 X_i^2\Big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\Big]dt\right\} \\ \text{(By Fubini)} \quad &= \int_0^1 E\left\{X_i^2\big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\big]\right\}dt \\ \text{(By Proposition 2)} \quad &\ge \int_0^1 EX_i^2 \cdot E\big[f'(W' + tX_i) - f'(W^{(i)} + tX_i)\big]dt \\ \Big(\text{By inequality (64)}\Big) \quad &\ge \int_0^1 \sigma_i^2 \cdot \big(-\frac{1}{2}c_3\sigma_i^2\big)dt \\ &= -\frac{1}{2}c_3\sigma_i^4. \end{aligned} \tag{65}$$

On the other hand, by the independence of X_i and $W', W^{(i)}$, it follows that

$$
\begin{aligned}
E\Big\{X_i^2 \int_0^1 \Big[f'(W') - f'(W^{(i)})\Big] dt\Big\} &= EX_i^2 \cdot E\Big\{\int_0^1 \Big[f'(W') - f'(W^{(i)})\Big] dt\Big\} \\
\text{(By Fubini)} \quad &= EX_i^2 \cdot \int_0^1 E\big[f'(W') - f'(W^{(i)})\big] dt \\
\Big(\text{By inequality (63)}\Big) \quad &\geq \sigma_i^2 \int_0^1 \frac{1}{2} c_3 \sigma_i^2 dt \\
&= \frac{1}{2} c_3 \sigma_i^4.
\end{aligned} \tag{66}
$$

Using inequality (65), (66) and the identity (62), we get the bound:

$$
R_1 \geq R_2 - c_3 \sum_{i=1}^n \sigma_i^4. \tag{67}
$$

Similarly, we can also write R in an integral form:

$$
R = \sum_{i=1}^n E\big[X_i^2 g(Z, X_i)\big] = \sum_{i=1}^n E\Big\{X_i^2 \int_0^1 \Big[f'(Z + tX_i) - f'(Z)\Big] dt\Big\},
$$

and by the integral form (61), algebraic simplification leads to

$$
\begin{aligned}
R_2 = R + \sum_{i=1}^n E\Big\{X_i^2 \int_0^1 \Big[f'(Z + tX_i) - f'(W' + tX_i)\Big] dt\Big\} \\
- \sum_{i=1}^n E\Big\{X_i^2 \int_0^1 \Big[f'(Z) - f'(W')\Big] dt\Big\}.
\end{aligned} \tag{68}
$$

And note that for any $\theta > 0$,

$$
\begin{aligned}
|Ef'(W' + \theta) - Ef'(Z + \theta)| &= \Big| \int_{-\infty}^{\infty} f'(w + \theta) dF_{W'} - \int_{-\infty}^{\infty} f'(w + \theta) dF_Z \Big| \\
\text{(Integration by parts)} \quad &= \Big| \int_{-\infty}^{\infty} \Big[F_{W'}(w) - F_Z(w)\Big] df'(w + \theta) \Big| \\
\text{(Change of variables)} \quad &= \Big| \int_{-\infty}^{\infty} f''(w) \Big(P(W' \leq w - \theta) - \Phi(w - \theta)\Big) dw \Big| \leq c_2 \delta.
\end{aligned} \tag{69}
$$

Therefore, by equation (68) and (69), we get the bound

$$
R_2 \geq R - 2c_2 \delta. \tag{70}
$$

Now recall all the bounds (46), (53), (67) and (70), and we will find

$$
c_1 \delta \geq R - \frac{3}{2} c_3 \sum_{i=1}^n \sigma_i^4 - 2c_2 \delta.
$$

Finally, we plug the bound (58) of R in the last inequality and collect terms to get

$$\begin{aligned}\frac{1}{2\sqrt{2}}(1-e^{-\epsilon^2/4})\sum_{i=1}^n E\Big[X_i^2 I(|X_i|>\epsilon)\Big] &\le \delta(c_1+2c_2)+\frac{3}{2}c_3\sum_{i=1}^n \sigma_i^4 \\ (\text{Plug in } \delta) \quad &= (c_1+2c_2)\cdot\sup_{z\in\mathbb{R}}|P(W\le z)-\Phi(z)|+\frac{3}{2}c_3\sum_{i=1}^n\sigma_i^4.\end{aligned}$$

Note that c_1, c_2 and c_3 are universal constants since they are not related to the choice of ϵ, which concludes this theorem. □

The necessity of Lindeberg's condition can be derived as a special case of Theorem 11.

Proposition 12. *Suppose $\{X_1,\dots,X_n\}$ are independent random variables with*

$$EX_i=0,\quad \mathrm{Var}X_i=\sigma_i^2,\quad \textit{and}\quad \sum_{i=1}^n\sigma_i^2=1.$$

Set $W=\sum_{i=1}^n X_i$. If Feller's condition is satisfied, i.e.,

$$\max_{1\le i\le n} EX_i^2\to 0,$$

then CLT implies Lindeberg's condition.

Proof. By Feller's condition, we have

$$\sum_{i=1}^n\sigma_i^4\le\sum_{i=1}^n\Big\{\sigma_i^2\times\max_{1\le j\le n}\sigma_j^2\Big\}=\max_{1\le j\le n}\sigma_j^2\times\sum_{i=1}^n\sigma_i^2=\max_{1\le j\le n}\sigma_j^2\to 0, \tag{71}$$

and by the CLT, we have

$$\sup_{z\in\mathbb{R}}|P(W\le z)-\Phi(z)|\to 0. \tag{72}$$

Using the limits (71), (72), by Theorem 11:

$$\Big(1-e^{-\epsilon^2/4}\Big)\sum_{i=1}^n E\Big\{X_i^2 I(|X_i|>\epsilon)\Big\}\le C\Big(\sup_{z\in\mathbb{R}}|P(W\le z)-\Phi(z)|+\sum_{i=1}^n\sigma_i^4\Big)\to 0,$$

for all $\epsilon>0$. Therefore, Lindeberg's condition follows. □

4 Summaries and discussions

In general, to prove convergence in distribution $W\Rightarrow Z$, by Stein's idea, it is equivalent to demonstrate that, for a sufficiently large class of functions $\mathcal{L}$,

$$\forall h\in\mathcal{L}:\quad Eh(W)\to Eh(Z).$$

With the introduction of Stein's equation, this can be achieved by showing that

$$E\big[f'(W)-Wf(W)\big]\to 0, \tag{73}$$

where f is the solution to Stein's equation for h. One method for bounding the LHS of equation (73) is to use the concept of zero bias distributions, as was done in the proof of the Central Limit Theorem. There are also other approaches for bounding the LHS of equation (73), such as the K function approach, exchange pairs, and size bias, which are also effective but not covered in this discussion due to limited space and time.

The application of Stein's method is not limited to independent random variables, but also extends to dependent random variables and multivariate normal distributions. In addition, Stein's method has also been utilized in the study of approximations of distributions that are not necessarily normal. The versatility of Stein's method continues to be a subject of ongoing research.

We would like to express our gratitude to Professor Tailen Hsing for providing us with the opportunity to learn about Stein's method and for his insightful lectures on the foundations of theoretical probability. We would also like to thank Professor Qi-Man Shao for his assistance in answering our questions and offering valuable guidance. We look forward to further exploring this subject in the future.

References

[1] Charles 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.

[2] Louis HY Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*, volume 2. Springer, 2011.