FUNDAMENTALS OF STEIN'S METHOD AND ITS APPLICATION IN PROVING CENTRAL LIMIT THEOREM

Xinyi(Cindy) Zhang

Department of Mathematics, University of Toronto

E-mail: xyi.zhang@mail.utotonto.ca.

November 22, 2016

Contents

Introduction

Stein's method is a sophisticated approach for proving generalized central limit theorem, pioneered in the 1970s by Charles Stein, one of the leading statisticians of the 20th century. In the ordinary central limit theorem, if $X_1, X_2, \cdots, X_n$ are independent and identically distributed random variables, then the simple average of these random variables follows the standard normal distribution (Gaussian distribution)

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \sim N(0,1),$$

where $\mathbb{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$.

The usual method to prove central limit theorem when random variables $X_1, X_2, \cdots, X_n$ are independent and identically distributed is to demonstrate convergence in distribution

$$\mathrm{P}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right) \longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \text{as } n \to \infty,$$

where the probability on the left-hand side can be computed by Fourier transform. The Fourier transform will decompose as a product because of independence. But what if these random variables X_i's are not independent? The technique mentioned above to prove central limit theorem is no longer useful. As a result, Stein came up with a new approach to prove that a random variable W approximately follows Gaussian distribution by bounding the Wassertein distance between two random variables, W and Z, where Z follows standard normal distribution, even if the condition of independence is violated. In this paper, we will introduce four techniques to bound the Wassertein distance between W and Z, including dependency graph, method of exchangeable pairs, size-bias coupling and zero-bias coupling, based on which we can prove ordinary central limit thorem, Hoeffding combinatorial central limit theorem, Lindeberg-Feller central limit theorem and other normal approximations in the Wasserstein metric.

1 Fundamentals of Stein's Method

In this section, we will introduce an overview of Stein's method, including the motivation and significance of this approach, some basic idea of probability metrics, relationship between Kolmogorov-Smirnov distance and Wasserstein distance, as well as how Stein's idea derived from these nice properties in the probability metric.

1.1 Overview

Given a collection of samples, it will be easier for us to make statistical analysis if we can obtain the approximate distribution (for instance, normal distribution, poisson distribution, exponential distribution, etc.) of the random variable that we are interested in. The ordinary central limit thorem can help us obtain the approximate distribution given that a sequence of such random variables are independent.

However, in real applications, random variables are not always independent and identically distributed (independence is an indispensable condition for the ordinary central limit theorem). Thus we need a more sophisticated approach to obtain the limiting distribution, and Stein's method helps us to deal with such problems on the basis of the following idea:

Let W be a random variable and Z be a standard Gaussian random variable, i.e. $Z \sim N(0,1)$. The density function of Z is

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

and the cumulative distribution function is

$$\mathbf{P}(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

To show that W is approximately Gaussian distributed, we need to prove that $\mathbf{P}$ $(W \leq x) \approx \mathbf{P}$ $(Z \leq x)$.
Recall the fact:

$$\mathbf{P}(W \leq x) = \mathbb{E}g(W),$$

where

$$g(u) = \begin{cases} 1 & \text{if } u \leq x, \\ 0 & \text{if } u > x. \end{cases}$$

The most intuitive way to prove convergence in distribution is that if we can show

$$\mathbb{E}g(W) \approx \mathbb{E}g(Z), \tag{1}$$

then we can conclude that random variable W approximately follows standard Gaussian distribution.

In Stein's method, we can obtain this approximation of distribution by bounding the Wassertein distance between W and Z when g is 1-*Lipschitz* from the perspective of probability metric.

Now we will start the introduction of Stein's method in more details, including why getting the upper bound of Wassertein distance works, some important properties of the solution to Stein's equation for both general g and specific g, some useful techniques to bound Wassertein distance, like dependency graph, method of exchangeable pairs, size-bias coupling, zero-bias coupling, and application of Stein's method in proving generalized central limit theorem using these techniques.

1.2 Probability metrics

Definition 1 *For two probability measures μ and ν, the probability metric is:*

$$d_{\mathscr{G}}(\mu, \nu) = \sup_{g \in \mathscr{G}} \left| \int g(x) d\mu(x) - \int g(x) d\nu(x) \right|, \tag{2}$$

where $\mathscr{G}$ is some family of test functions.

As for two random variables W and Z, the probability metric has the form

$$d_{\mathscr{G}}(W,Z) = \sup_{g\in\mathscr{G}}\left|\int g(x)dF_W(x) - \int g(x)dF_Z(x)\right| = \sup_{g\in\mathscr{G}}\left|\mathbb{E}g(W) - \mathbb{E}g(Z)\right|, \tag{3}$$

where $F_W(x)$ and $F_Z(x)$ are distribution functions of random variables W and Z respectively.

Kolmogorov-Smirnov distance and Wasserstein distance for random variables W and Z can be defined as follows for special g:

1. If $g = \mathbf{1}_{[\cdot\leq x]}$ is an indicator function, then the Kolmogorov-Smirnov distance can be defined as follows
$$\mathrm{Kolm}(W,Z) = \sup_{x\in\mathbb{R}}|\mathbf{P}(W\leq x) - \mathbf{P}(Z\leq x)|.$$
2. If $\mathscr{G}$ is a collection of 1-*Lipschitz* functions, the Wasserstein distance can be defined as follows
$$\mathrm{Wass}(W,Z) = \sup_{g\in\mathscr{G}}|\mathbb{E}g(W) - \mathbb{E}g(Z)|.$$

Lemma 1 *Suppose W, Z are two random variables, and Z has a density with respect to Lebesgue measure bounded by a constant C. Then* $\mathrm{Kolm}(W,Z)\leq 2\sqrt{C\mathrm{Wass}(W,Z)}$.

Proof. Fix $\varepsilon > 0$. Define $g^1_{x,\varepsilon}(w)$ to be 1 when $w\leq x$, 0 when $w\geq x+\varepsilon$, and linear between.

$$\begin{aligned}\mathbf{P}(W\leq x) - \mathbf{P}(Z\leq x) &= \mathbf{P}(W\leq x) - \mathbb{E}[g^1_{x,\varepsilon}(Z)] + \mathbb{E}[g^1_{x,\varepsilon}(Z)] - \mathbf{P}(Z\leq x)\\ &\leq \mathbb{E}[g^1_{x,\varepsilon}(W)] - \mathbb{E}[g^1_{x,\varepsilon}(Z)] + \mathbb{E}[g^1_{x,\varepsilon}(Z)] - \mathbf{P}(Z\leq x)\\ &\leq \frac{1}{\varepsilon}\mathrm{Wass}(W,Z) + C\varepsilon,\end{aligned}$$

since $g^1_{x,\varepsilon}(w)$ is $\frac{1}{\varepsilon}$-*Lipschitz*, and the density of Z is bounded by C.
If we take $\varepsilon = \sqrt{\frac{\mathrm{Wass}(W,Z)}{C}}$, then we have

$$\mathbf{P}(W\leq x) - \mathbf{P}(Z\leq x)\leq 2\sqrt{C\mathrm{Wass}(W,Z)}.$$

Define $g^2_{x,\varepsilon}(w)$ to be 1 when $w\leq x-\varepsilon$, 0 when $w\geq x$, and linear between. Then we can get the same upper bound for $\mathbf{P}(Z\leq x) - \mathbf{P}(W\leq x)$.
Therefore,

$$\mathrm{Kolm}(W,Z) = \sup_{x\in\mathbb{R}}|\mathbf{P}(W\leq x) - \mathbf{P}(Z\leq x)|\leq 2\sqrt{C\mathrm{Wass}(W,Z)}. \tag{4}$$

In particular, if $Z\sim N(0,1)$, then $C = \frac{1}{\sqrt{2\pi}}$. This finishes the proof. □

1.3 Stein's idea

From Gaussian integration by parts, we know that if X is a random variable with mean 0 and $\mathbb{E}(x^2) = \sigma^2$, then the density function of X is

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

Note that $x\varphi(x) = -\sigma^2\varphi'(x)$. Then given a continuously differential function $f : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}xF(x)$ satisfies

$$\begin{aligned}\mathbb{E}xF(x) &= \int_{-\infty}^{+\infty} xF(x)\varphi(x)dx = -\sigma^2 F(x)\varphi(x)\Big|_{-\infty}^{+\infty} + \sigma^2 \int_{-\infty}^{\infty} F'(x)\varphi(x)dx \\ &= \sigma^2 \mathbb{E}F'(x),\end{aligned}$$

when the limits $\lim_{x\to\infty} F(x)\varphi(x) = 0$ and integrals on both sides are finite.

Therefore, for standard Gaussian distribution, we have $\mathbb{E}xF(x) = \mathbb{E}F'(x)$, as a result of the special case with $\sigma = 1$. Based on this result, we have the following nice property for standard normal distribution, which is also known as Stein's Identity.

Lemma 2 *(**Stein's Identity**) If $Z \sim N(0,1)$, then*

$$\mathbb{E}f'(Z) = \mathbb{E}Zf(Z), \tag{5}$$

for all absolute continuous functions $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$.
Conversely, if $\mathbb{E}[f'(Z)] = \mathbb{E}[Zf(Z)]$ for all bounded, continuous and piecewise continuously differentiable functions f with $\mathbb{E}|f'(Z)| < \infty$, then Z has a standard normal distribution.

We will later show the proof of this lemma in the next section. Stein's method is deduced from the above idea. If a random variables W satisfies (5) approximately, i.e. $\mathbb{E}f'(W) \approx \mathbb{E}Wf(W)$, then we can conclude that W is approximately a standard Gaussian random variable.
Recall the probability metric of two random variables:

$$d_{\mathscr{G}}(W,Z) = \sup_{g\in\mathscr{G}} |\mathbb{E}g(W) - \mathbb{E}g(Z)|,$$

where $\mathscr{G}$ is some given class of functions.

If we can find another class of function $f \in \mathscr{F}$ such that

$$\sup_{g\in\mathscr{G}} \big|\mathbb{E}g(W) - \mathbb{E}g(Z)\big| \le \sup_{f\in\mathscr{F}} \big|\mathbb{E}[f'(W) - Wf(W)]\big|, \tag{6}$$

then we can bound $d_{\mathscr{G}}(W,Z)$ by bounding $|\mathbb{E}[f'(W) - Wf(W)]|$. On the basis of this relationship, Stein's method can be characterized by the following equation

$$f'(w) - wf(w) = g(w) - \mathbb{E}g(Z), \tag{7}$$

which is also known as Stein's equation.

If $\mathscr{F}$ is a class of functions such that for every $g \in \mathscr{G}$, there exists $f \in \mathscr{F}$ such that (7) holds, for $Z \sim N(0,1)$, then (6) holds, which is obvious by taking expectation on both sides of (7)

$$\mathbb{E}[f'(W) - Wf(W)] = \mathbb{E}g(W) - \mathbb{E}g(Z). \tag{8}$$

Noticeably, there are several boundary conditions for the solution of equation (7), which will be explained in details in **section 2**.

Lemma 3 *Given a function $g : \mathbb{R} \to \mathbb{R}$ that is bounded, then there exists absolutely continuous f solving (3) for all x, satisfying*

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

And if g is Lipschitz, then

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

(Proof of **Lemma 3** will be given in the next section.)

Particularly, if g is 1-*Lipschitz*, the probability metric becomes Wasserstein distance, and if $\mathscr{F}$ is defined as a family of functions satisfying

$$\mathscr{F} = \{\forall f \in \mathscr{F},\ |f|_\infty \leq 1,\ f'|_\infty \leq \sqrt{\frac{2}{\pi}}, \text{ and } |f''|_\infty \leq 2\},$$

then we have

$$\text{Wass}(W,Z) = \sup_{g \in \mathscr{G}} \big|\mathbb{E}g(W) - \mathbb{E}g(Z)\big| \leq \sup_{f \in \mathscr{F}} \big|\mathbb{E}[f'(W) - Wf(W)]\big|. \tag{9}$$

From **Lemma 1** we have already known that $\text{Kolm}(W,Z) \leq 2\sqrt{C\text{Wass}(W,Z)}$, hence if we can show that the upper bound of $\text{Wass}(W,Z)$ approximates 0, then we can conclude that
$|\mathbf{P}(W \leq x) - \mathbf{P}(Z \leq x)| \to 0$. Finally, by definition of convergence in distribution, we get

$$W \dot{\sim} N(0,1).$$

In **section 3**, we will see some applications of Stein's method to show normal approximation by bounding Wasserstein distance.

2 Detailed Illustration of Stein's method

In this section, we will give the proof of Stein's identity and study the solution of Stein's equation

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

for both general g and special g.

2.1 Stein's Identity (proof)

To prove **Lemma 2 (Stein's Identity)**, we first need the following result.

Lemma 4 *For fixed $z \in \mathbb{R}$ and $\Phi(z) = \boldsymbol{P}(Z \leq z)$, the cumulative distribution function of Z, the unique bounded solution $f(w)$ of the equation*

$$f'(w) - wf(w) = 1_{(w \leq z)} - \Phi(z)$$

is given by

$$f(w) = \begin{cases} \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(w)(1-\Phi(z)), & w \leq z \\ \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(z)(1-\Phi(w)), & w > z \end{cases} \tag{10}$$

Proof. Multiply both sides of the equation by $e^{-\frac{w^2}{2}}$:

$$e^{-\frac{w^2}{2}}[f'(w) - wf(w)] = e^{-\frac{w^2}{2}}[\mathbf{1}_{(w \leq z)} - \Phi(z)],$$

and the above equation yields

$$(f(w)e^{\frac{-w^2}{2}})' = e^{-\frac{w^2}{2}}[\mathbf{1}_{(w \leq z)} - \Phi(z)].$$

Integration yields

$$f(w)e^{\frac{-w^2}{2}} = \int_{-\infty}^{w} [\mathbf{1}_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx, \tag{11}$$

and since f is unique, there is no constant term in (11). Then multiply both sides by $e^{\frac{w^2}{2}}$, we have

$$\begin{aligned} f(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^{w} [\mathbf{1}_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx \\ &= -e^{\frac{w^2}{2}} \int_{w}^{\infty} [\mathbf{1}_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx \end{aligned} \quad [1]$$

which is equivalent to (10), since

[1] Two forms of solution come from the fact that $\int_{-\infty}^{\infty} [\mathbf{1}_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx = 0$.

a. $w \leq z$

$$
\begin{aligned}
f(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^{w} [\mathbf{1}_{(x\leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx \\
&= e^{\frac{w^2}{2}} (1-\Phi(z)) \sqrt{2\pi} \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(w)(1-\Phi(z))
\end{aligned}
$$

b. $w > z$

$$
\begin{aligned}
f(w) &= -e^{\frac{w^2}{2}} \int_{w}^{\infty} [\mathbf{1}_{(x\leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx \\
&= e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{x^2}{2}} \Phi(z) dx \\
&= \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(z) \int_{w}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(z)(1-\Phi(w)).
\end{aligned}
$$

□

Based on the result in **Lemma 4**, now we can show the proof of Stein's identity.

Proof of Lemma 2 (Stein's Identity) . "$\Longrightarrow$" Suppose that $Z \sim N(0,1)$, and $\mathbb{E}|f'(Z)| < \infty$, we can then write $\mathbb{E}f'(Z)$ as an integral and exchange the order of integral to get the final result using Fubini theorem [2].

$$
\begin{aligned}
\mathbb{E}f'(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(z) e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(z) e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} f'(z) e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(z) \int_{-\infty}^{z} (-x) e^{-\frac{x^2}{2}} dx dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} f'(z) \int_{z}^{+\infty} x e^{-\frac{x^2}{2}} dx dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} (\int_{0}^{x} f'(z) dz)(-x e^{-\frac{x^2}{2}}) dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} (\int_{0}^{x} f'(z) dz) x e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (f(x) - f(0)) x e^{-\frac{x^2}{2}} dx \\
&= \mathbb{E}Zf(Z) - \frac{1}{\sqrt{2\pi}} f(0) \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx \\
&= \mathbb{E}Zf(Z),
\end{aligned}
$$

which implies that if $Z \sim N(0,1)$, then we have $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$.

[2] Suppose A and B are complete measure space. Suppose $f(x,y)$ is $A\times B$ measurable. If $\iint_{A\times B} |f(x,y)| d(x,y) < \infty$, then $\int_A (\int_B f(x,y) dy) dx = \int_B (\int_A f(x,y) dx) dy = \iint_{A\times B} f(x,y) d(x,y)$.

"$\Longleftarrow$" Suppose that $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$.
Recall Stein's equation:

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

where $Z \sim N(0,1)$.
Taking $g(w) = 1_{[w \leq x]}$, the solution implied in **Lemma 4** satisfies the conditions of **Lemma 2 (Stein's Identity)**, thus we have

$$0 = \mathbb{E}[f'(W) - Wf(W)] = \mathbb{E}[\mathbf{1}_{(w \leq x)}] - \Phi(z) = \mathbf{P}(W \leq x) - \mathbf{P}(Z \leq x).$$

Therefore, $W \sim N(0,1)$. This finish the proof. □

2.2 Solution to Stein's equation

Lemma 5 *For a given real valued measurable function g with* $\mathbb{E}|g(Z)| < \infty$, *and* $Z \sim N(0,1)$, *the Stein's equation for g is:*

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

and the unique, bounded solution to this differential equation is

$$f(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{\frac{-x^2}{2}} \big[g(x) - \mathbb{E}g(Z)\big] dx = -e^{\frac{-w^2}{2}} \int_{w}^{+\infty} e^{\frac{-x^2}{2}} \big[g(x) - \mathbb{E}g(Z)\big] dx. \tag{12}$$

Proof. Mutilplying both sides of Stein's equation by $e^{-\frac{w^2}{2}}$ yields

$$\big(f(w)e^{\frac{-w^2}{2}}\big)' = e^{\frac{-w^2}{2}}\big(g(w) - \mathbb{E}g(Z)\big).$$

Integration yields

$$f(w)e^{\frac{-w^2}{2}} = \int_{-\infty}^{w} e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx + Ce^{\frac{w^2}{2}}$$
$$f(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx + C.$$

In order to get a unique and bounded solution $f(w)$, C must be 0. So the final solution for Stein's equation is

$$\begin{aligned} f(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx \\ &= -e^{\frac{w^2}{2}} \int_{w}^{+\infty} e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx. \end{aligned}$$

We have two forms of solution as a result of the following fact:

$$\begin{aligned}\int_{-\infty}^{+\infty} e^{\frac{-x^2}{2}}[g(x)-\mathbb{E}g(Z)]dx &= \int_{-\infty}^{+\infty} e^{\frac{-x^2}{2}} g(x)dx - \mathbb{E}g(Z)\int_{-\infty}^{+\infty} e^{\frac{-x^2}{2}}dx \\ &= \sqrt{2\pi}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} g(x)dx - \sqrt{2\pi}\mathbb{E}g(Z) \\ &= \sqrt{2\pi}\mathbb{E}g(X) - \sqrt{2\pi}\mathbb{E}g(Z) \\ &= 0.\end{aligned}$$

This finishes the proof. □

Additionally, we have another form of solution,

$$f(w) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}}\mathbb{E}[Zg(\sqrt{t}w+\sqrt{1-t}Z)]dt, \quad Z\sim N(0,1), \tag{13}$$

if g is *Lipschitz*.

Proof. We need to check that (13) satisfies $f'(w)-wf(w)=g(w)-\mathbb{E}g(Z)$.
First, let's take first derivative on $f(w)$ with respect to w.

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

By Stein's identity $\mathbb{E}Zf(Z)=\mathbb{E}f'(Z)$, we have

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

Combining the above calculations, we obtain that

$$\begin{aligned}f'(w)-wf(w) &= -\int_0^1 \mathbb{E}\big[\frac{Z}{2\sqrt{1-t}}g'(\sqrt{t}w+\sqrt{1-t}Z)\big]dt + \int_0^1 \mathbb{E}\big[\frac{w}{2\sqrt{t(1-t)}}\cdot Z\cdot g(\sqrt{t}w+\sqrt{1-t}Z)\big]dt \\ &= -\int_0^1 \mathbb{E}\big[\frac{Z}{2\sqrt{1-t}}g'(\sqrt{t}w+\sqrt{1-t}Z)\big]dt + \\ &\quad \int_0^1 \mathbb{E}\big[\frac{w}{2\sqrt{t(1-t)}}\cdot\sqrt{1-t}\cdot\mathbb{E}\big[g'(\sqrt{t}w+\sqrt{1-t}Z)\big]dt \\ &= \int_0^1 \mathbb{E}(\frac{-Z}{2\sqrt{1-t}}+\frac{w}{2\sqrt{t}})g'(\sqrt{t}w+\sqrt{1-t}Z)dt \\ &= \mathbb{E}\int_0^1 \frac{dg(\sqrt{t}w+\sqrt{1-t}Z)}{dt}dt \\ &= \mathbb{E}\Big[g(\sqrt{t}w+\sqrt{1-t}Z)\Big|_0^1\Big] \\ &= g(x)-\mathbb{E}g(Z).\end{aligned}$$

This finishes the proof. □

2.3 Boundary conditions for the solution of Stein's equation

In this section, we will show the proof of **Lemma 3**, the boundary conditions for the solution of Stein's equation.

Lemma 6 *If $g:\mathbb{R}\to\mathbb{R}$ is bounded, then* $|f|_\infty \le \sqrt{\frac{\pi}{2}}\Big|g(w)-\mathbb{E}g(Z)\Big|_\infty$

Proof. We need to consider two cases, $w>0$ and $w\le 0$.

a. $w>0$
Recall the solution of Stein's equation for general g

$$f(w) = -e^{\frac{w^2}{2}}\int_w^{+\infty} e^{\frac{-x^2}{2}}\left[g(x)-\mathbb{E}g(Z)\right]dx.$$

The supremum norm of f satisfies

$$|f|_\infty \le |g(w)-\mathbb{E}g(Z)|_\infty \cdot e^{\frac{w^2}{2}}\int_w^{+\infty} e^{\frac{-x^2}{2}}dx.$$

Let $h(w)=e^{\frac{w^2}{2}}\int_w^{+\infty}e^{\frac{-x^2}{2}}dx$, then $h'(w)=-1+we^{\frac{w^2}{2}}\int_w^{+\infty}e^{\frac{-x^2}{2}}dx$. Define

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

By Mill's ratio inequality:

$$\frac{x}{1+x^2}<r(x)<\frac{1}{x},\quad \forall x>0,$$

then we have $we^{\frac{w^2}{2}}\int_w^\infty e^{\frac{-x^2}{2}}dx<1$ and $h'(w)<0,$which implies that $h(w)$ is decreasing on $[0,+\infty]$, and thus

$$h(w)\le h(0)=\sqrt{\frac{\pi}{2}}.$$

Therefore,

$$|f|_\infty \le \sqrt{\frac{\pi}{2}}\Big|g(w)-\mathbb{E}g(Z)\Big|_\infty,$$

when $w>0$.

b. $w\le 0$
In this case, we use another form of the soluiton

$$e^{\frac{w^2}{2}}\int_{-\infty}^{w} e^{\frac{-x^2}{2}}[g(x)-\mathbb{E}g(Z)]dx,$$

and the same method as $w > 0$ as well as the symmetry of standard Gaussian distribution to get the upper bound of $|f|_\infty$.

Therefore,

$$|f|_\infty \le \sqrt{\frac{\pi}{2}}\Big|g(w) - \mathbb{E}(g(Z))\Big|_\infty, \quad \forall w.$$

This finishes the proof. □

Lemma 7 *If $g : \mathbb{R} \to \mathbb{R}$ is bounded, then $|f'|_\infty \le 2|g(w) - \mathbb{E}g(Z)|_\infty$.*

Proof. We only consider the case when $w > 0$, and the upper bound can be attained using the same manner by using another form of the solution when $w < 0$.
By taking first derivative with respect to w, we have

$$\begin{aligned} f'(w) &= wf(w) + g(w) - \mathbb{E}g(Z) \\ &= g(w) - \mathbb{E}g(Z) + w(-e^{\frac{w^2}{2}}) \int_w^{+\infty} e^{\frac{-x^2}{2}} [g(w) - \mathbb{E}g(Z)]dx, \end{aligned}$$

and hence

$$|f'(w)|_\infty \le |g(w) - \mathbb{E}g(Z)|_\infty (1 + we^{\frac{w^2}{2}} \int_w^{+\infty} e^{\frac{-x^2}{2}} dx).$$

By Mill's ratio inequality $r(x) < \frac{1}{x}$, we can get

$$we^{\frac{w^2}{2}} \int_w^{+\infty} e^{\frac{-x^2}{2}} dx < 1, \text{when } w > 0.$$

Therefore,

$$|f'|_\infty \le 2|g(w) - \mathbb{E}g(Z)|_\infty.$$

□

Lemma 8 *If g is Lipschitz, but not necessarily bounded, then $|f|_\infty \le |g'|_\infty$.*

Proof. If g is *Lipschitz*, then we have another form of solution that have been proved in the previous subsection

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

Suppose that

$$f(Z) = g(\sqrt{t}w + \sqrt{1-t}Z),$$

and based on Stein's identity that for any absolutely continuous function f

$$\mathbb{E}Zf(Z) = \mathbb{E}f'(Z),$$

then we can obtain that

$$\mathbb{E}[Zg(\sqrt{t}w + \sqrt{1-t}Z)] = \sqrt{1-t}\mathbb{E}[g'(\sqrt{t}w + \sqrt{1-t}Z)]. \tag{14}$$

Plugging in (14) into the solution, we can get

$$f(w) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}} \cdot \sqrt{1-t}\mathbb{E}[g'(\sqrt{t}w + \sqrt{1-t}Z)]dt,$$

and thus

$$|f|_\infty \le |g'|_\infty \int_0^1 \frac{1}{2\sqrt{t}} dt = |g'|_\infty.$$

This finishes the proof. □

Lemma 9 *If g is Lipschitz, but not necessarily bounded, then* $|f'|_\infty \le \sqrt{\frac{\pi}{2}}|g'|_\infty$.

Proof. If g is *Lipschitz*, then we have

$$f(w) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbb{E}[Zg(\sqrt{t}w + \sqrt{1-t}Z)]dt,$$

and taking the first derivative with respect to w, we obtain that

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

So the supremum norm of $f'(w)$ satisfies

$$|f'|_\infty \le |g'|_\infty \cdot \mathbb{E}|Z| \le 2|g'|_\infty \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \sqrt{\frac{\pi}{2}}|g'|_\infty.$$

□

Lemma 10 *If g is Lipschitz, but not necessarily bounded, then* $|f''|_\infty \le 2|g'|_\infty$.

Proof. By taking second derivative on Stein's equation with respect to w, we get

$$\begin{aligned} f''(w) &= g'(w) + f(w) + wf'(w) \\ &= g'(w) + f(w) + w\big[wf(w) + g(w) - \mathbb{E}g(Z)\big]. \end{aligned}$$

Rewrite $g(w)-\mathbb{E}g(Z)$ as an integral:

$$\begin{aligned}
g(w)-\mathbb{E}g(Z) &= g(w)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} g(y)e^{-\frac{y^2}{2}}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(g(w)-g(y))e^{-\frac{y^2}{2}}dy \\
&= \frac{1}{\sqrt{2\pi}}[\int_{-\infty}^{w}(g(w)-g(y))e^{-\frac{y^2}{2}}dy+\int_{w}^{+\infty}(g(w)-g(y))e^{-\frac{y^2}{2}}dy] \\
&= \frac{1}{\sqrt{2\pi}}[\int_{-\infty}^{w}e^{-\frac{y^2}{2}}\int_{y}^{w}g'(z)dzdy-\int_{w}^{+\infty}1\int_{w}^{y}e^{-\frac{y^2}{2}}g'(z)dzdy] \\
&= \frac{1}{\sqrt{2\pi}}[\int_{-\infty}^{w}g'(z)\int_{-\infty}^{z}e^{-\frac{y^2}{2}}dydz-\int_{w}^{+\infty}g'(z)\int_{z}^{+\infty}e^{-\frac{y^2}{2}}dydz] \\
&= \int_{-\infty}^{w}g'(z)\Phi(z)dz-\int_{w}^{+\infty}g'(z)(1-\Phi(z))dz
\end{aligned}$$

Next, using $\Phi(z)=\mathbf{P}(Z\le z)$ and $\bar{\Phi}(z)=\mathbf{P}(Z>z)$, we can rewrite $f(w)$ as follows

$$\begin{aligned}
f(w) &= e^{\frac{w^2}{2}}\int_{-\infty}^{w}e^{\frac{-y^2}{2}}[g(y)-\mathbb{E}g(z)]dy \\
&= e^{\frac{w^2}{2}}\int_{-\infty}^{w}e^{\frac{-y^2}{2}}\big(\int_{-\infty}^{y}g'(z)\Phi(z)dz-\int_{y}^{\infty}g'(z)\bar{\Phi}(z)dz\big)dy \\
&= e^{\frac{w^2}{2}}\int_{-\infty}^{w}g'(z)\Phi(z)\int_{z}^{w}e^{\frac{-y^2}{2}}dydz-e^{\frac{w^2}{2}}\big[\int_{-\infty}^{w}g'(z)\bar{\Phi}(z)\int_{-\infty}^{z}e^{\frac{-y^2}{2}}dydz+ \\
&\quad \int_{w}^{+\infty}g'(z)\bar{\Phi}(z)\int_{-\infty}^{w}e^{\frac{-y^2}{2}}dydz\big]. \qquad (15)
\end{aligned}$$

And Since

$$\begin{aligned}
\sqrt{2\pi}e^{\frac{w^2}{2}}\int_{-\infty}^{w}g'(z)\Phi(z)\int_{z}^{w}\frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dydz &= \sqrt{2\pi}e^{\frac{w^2}{2}}\int_{-\infty}^{w}g'(z)\Phi(z)(\Phi(w)-\Phi(z))dz \\
&= \sqrt{2\pi}e^{\frac{w^2}{2}}\int_{-\infty}^{w}g'(z)\Phi(z)(\bar{\Phi}(z)-\bar{\Phi}(w))dz,
\end{aligned}$$

we can rewrite (15) as follows

$$\begin{aligned}
(15) &= \sqrt{2\pi}e^{\frac{w^2}{2}}\int_{-\infty}^{w}g'(z)\Phi(z)(\bar{\Phi}(z)-\bar{\Phi}(w))dz-\sqrt{2\pi}e^{\frac{w^2}{2}}\big[\int_{-\infty}^{w}g'(z)\bar{\Phi}(z)\Phi(z)dz+\int_{w}^{\infty}g'(z)\bar{\Phi}(z)\Phi(w)dz\big] \\
&= \sqrt{2\pi}e^{\frac{w^2}{2}}\big[\int_{-\infty}^{w}g'(z)\Phi(z)(\bar{\Phi}(z)-\bar{\Phi}(w))dz-\int_{-\infty}^{w}g'(z)\bar{\Phi}(z)\Phi(z)dz-\int_{w}^{\infty}g'(z)\bar{\Phi}(z)\Phi(w)dz\big] \\
&= -\sqrt{2\pi}e^{\frac{w^2}{2}}\big[\int_{-\infty}^{w}g'(z)\Phi(z)\bar{\Phi}(w)dz+\int_{w}^{\infty}g'(z)\bar{\Phi}(z)\Phi(w)dz\big] \\
&= -\sqrt{2\pi}e^{\frac{w^2}{2}}\big[\bar{\Phi}(w)\int_{-\infty}^{w}g'(z)\Phi(z)dz+\Phi(w)\int_{w}^{\infty}g'(z)\bar{\Phi}(z)dz\big].
\end{aligned}$$

Combining all these calculations together, we get

$$\begin{aligned}
f''(w) &= g'(w) + f(w) + w\big[wf(w) + g(w) - \mathbb{E}g(z)\big] \\
&= g'(w) + w\big[g(w) - \mathbb{E}g(z)\big] + (1+w^2)f(w) \\
&= g'(w) + w\Big[\int_{-\infty}^{w} g'(z)\Phi(z)dz - \int_{w}^{+\infty} g'(z)\bar{\Phi}(z)dz\Big] - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\Big[\bar{\Phi}(w)\int_{-\infty}^{w} g'(z)\Phi(z)dz \\
&\quad + \Phi(w)\int_{w}^{\infty} g'(z)\bar{\Phi}(z)dz\Big] \\
&= g'(w) + \big[w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\bar{\Phi}(w)\big]\int_{-\infty}^{w} g'(z)\Phi(z)dz + \\
&\quad \big[-w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\Phi(w)\big]\int_{w}^{+\infty} g'(z)\bar{\Phi}(z)dz.
\end{aligned}$$

Hence the supremum norm of $f''(w)$ satisfies

$$|f''|_{\infty} \le |g'|_{\infty}\Big[1 + |w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\bar{\Phi}(w)|\int_{-\infty}^{w}\Phi(z)dz + \big|-w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\Phi(w)\big|\int_{w}^{\infty}\bar{\Phi}(z)dz\Big]. \tag{16}$$

Recall Mill's ratio inequality ($x > 0$)

$$\frac{x}{1+x^2} \le e^{\frac{x^2}{2}}\int_{x}^{\infty} e^{-\frac{t^2}{2}}dt < \frac{1}{x},$$

which is equivalent to

$$\frac{xe^{\frac{-x^2}{2}}}{\sqrt{2\pi}(1+x^2)} \le 1 - \Phi(x) \le \frac{e^{\frac{-x^2}{2}}}{x\sqrt{2\pi}}. \tag{17}$$

It's easy to see that (17) implies

$$-x + \sqrt{2\pi}(1+x^2)e^{\frac{x^2}{2}}(1 - \Phi(x)) \ge 0.$$

Additionally, since $x + \sqrt{2\pi}(1+x^2)e^{\frac{x^2}{2}}\Phi(x) > 0$, based on the fact that $x > 0$, then we can remove the absolute value sign. Moreover, by simple integration, we can obtain that

$$\int_{-\infty}^{w}\Phi(z)dz = w\Phi(w) + \frac{1}{\sqrt{2\pi}}e^{\frac{-w^2}{2}}, \quad [3] \tag{18}$$

and

$$\int_{w}^{\infty}(1 - \Phi(z))dz = -w(1 - \Phi(w)) + \frac{1}{\sqrt{2\pi}}e^{\frac{-w^2}{2}}. \quad [4] \tag{19}$$

[3] $\int_{-\infty}^{w}\Phi(z)dz = \int_{-\infty}^{w}(z)'\Phi(z)dz$, and then using integration by parts, we can get the above result.

[4] $\int_{w}^{\infty}(1-\Phi(z))dz = \int_{w}^{\infty}(z)'(1-\Phi(z))dz$, and then using integration by parts, we can get the above result.

Finally, by substituion, we get

$$\left[-w+\sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\bar{\Phi}(w)\right]\int_{-\infty}^{w}\Phi(z)dz+\left[w+\sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\Phi(w)\right]\int_{w}^{\infty}\bar{\Phi}(z)dz=1.$$

Therefore,

$$|f''|_\infty \le 2|g'|_\infty.$$

□

Lemma 11 *A less constrained bound: $|f''|_\infty \le 4|g'|_\infty$, when g is Lipschitz.*

Proof. Recall

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

and

$$f''(w)-wf'(w)=g'(w)+f(w). \tag{20}$$

Let $h(w)=g'(w)+f(w)$, then we have $\mathbb{E}h(Z)=\mathbb{E}[g'(Z)+f(Z)]$. Since $Z\sim N(0,1)$, and f' is also absolutely continuous, then by Stein's Identity we have

$$\mathbb{E}f''(Z)=\mathbb{E}Zf'(Z).$$

Hence,

$$\mathbb{E}h(Z)=\mathbb{E}[g'(Z)+f(Z)]=\mathbb{E}[f''(Z)-Zf'(Z)]=0,$$

and (20) can be rewritten as

$$f''(w)-wf'(w)=h(w)+\mathbb{E}[h(Z)], \tag{21}$$

where $h:=g'+f$. Therefore, f' is the solution for the new Stein's equation (21), and thus also satisfies the boundary condition that we have proved before:

$$\begin{aligned}|f''|_\infty &\le 2\big|g'+f-\mathbb{E}[g'(Z)+f(Z)]\big|_\infty = 2|g'+f|_\infty\\ &\le 2(|g'|_\infty+|f|_\infty)\\ &\le 2(|g'|_\infty+|g'|_\infty)=4|g'|_\infty\end{aligned}$$

This finishes the proof. □

2.4 Example-Ordinary Central Limit Theorem in the Wasserstein metric

Let $X_1,X_2,\ldots,X_n$ be independent random variables with $\mathbb{E}(X_i)=0$, $\mathbf{E}(X_i^2)=1$ and $\mathbb{E}|X_i^3|<\infty$. Let $W=\frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}$, and then $W\sim N(0,1)$.

Proof. Take any $f\in C^1$withf' absolutely continuous, and satisfying $|f|\le 1,\ |f'|\le\sqrt{\frac{2}{\pi}},\ |f''|\le 2$.

Let $W_i = W - \frac{X_i}{\sqrt{n}}$, which implies that $W_i \perp X_i$ ("$\perp$" is the sign for independence). Note that

$$\begin{aligned}\mathbb{E}\big(X_i f(W)\big) &= \mathbb{E}[X_i(f(W) - f(W_i)) + X_i f(W_i)] = \mathbb{E}[X_i(f(W) - f(W_i))] \\ &= \mathbb{E}\big[X_i(f(W) - f(W_i) - (W - W_i)f'(W_i))\big] + \mathbb{E}\big[X_i(W - W_i)f'(W_i)\big].\end{aligned}$$

Note that $f''(x) = \frac{f'(x+h) - f'(x)}{h}$, and thus we can obtain

$$\mathbb{E}[Wf(W)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}[X_i f(W)],$$

and

$$|f(W) - f(W_i) - (W - W_i)f'(W_i)| \le \frac{1}{2}(W - W_i)^2 |f''|_\infty,$$

thus

$$\mathbb{E}\big[X_i(f(W) - f(W_i) - (W - W_i)f'(W_i))\big] \le \frac{1}{2}|f''|_\infty \cdot \mathbb{E}\Big|X_i \frac{X_i^2}{n}\Big| \le \frac{1}{n}\mathbb{E}|X_i|^3.$$

Again,

$$\mathbb{E}\big[X_i(W - W_i)f'(W_i)\big] = \frac{1}{\sqrt{n}}\mathbb{E}[X_i^2] \cdot \mathbf{E}[f'(W_i)] = \frac{1}{\sqrt{n}}\mathbb{E}[f'(W_i)],$$

since W_i is independent of X_i. Based on the above calculations, we can get

$$\big|\mathbb{E}Wf(W) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[f'(W_i)]\big| \le \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \mathbb{E}|X_i|^3.$$

Finally, note that

$$\begin{aligned}\big|\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[f'(W_i)] - \mathbb{E}[f'(W)]\big| &\le \frac{|f''|_\infty}{n} \sum_{i=1}^{n} \mathbb{E}|W - W_i| \\ &\le \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \mathbb{E}|X_i|.\end{aligned}$$

Combining all these together, we obtain that

$$\big|\mathbb{E}[f'(W) - Wf(W)]\big| \le \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \mathbb{E}|X_i|^3 + \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \mathbb{E}|X_i|.$$

Since $\mathbb{E}(X_i^2) = 1$, we can say that $\mathbb{E}|X_i|^3 \ge 1$, and $\mathbb{E}|X_i| \le \big(\mathbb{E}|X_i|^3\big)^{\frac{1}{3}} \le \mathbb{E}|X_i|^3$. Therefore,

$$\text{Wass}(W, Z) \le \frac{3}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \mathbb{E}|X_i|^3,$$

which converges to 0, since $\mathbf{E}|X_i|^3 < \infty$. Finally, we can conclude that $W \dot{\sim} N(0,1)$. □

3 Appication of Stein's method

Based on the idea from Stein's method, we know that if we can figure out that Wasserstein distance between random variables W and Z approximates 0, where Z is standard Gaussian, then we can conclude that W approximately follows standard normal distribution. Unlike the proof of ordinary central limit theorem in the Wasserstein metric, it not quite easy to compute the upper bound of $\text{Wass}(W,Z)$ directly in most cases. In this section, we will introduce some useful techniques that are mostly used to obtain the upper bound of Wasserstein distance $\text{Wass}(W,Z)$, including dependency graph, method of exchangeable pairs, size-bias coupling and zero-bias coupling.

3.1 Dependency Graph

We firstly will introduce the definitions of dependency graph and dependency neighborhoods.

Definition 2 *$\{X_i\}_{i\in V}$ are random variables. A dependency graph for $\{X_i\}_{i\in V}$ is any graph G with vertex set V such that if S,T are two disjoint subsets of V so that there is no edge of G between any vertex in S to any vertex in T, then $\{X_i\}_{i\in S}$ and $\{X_i\}_{i\in T}$ are mutually independent.*

Definition 3 ***(Dependency neighborhoods)*** *We say that a collection of random variables $\{X_1, X_2, \cdots, X_n\}$ has dependency neighborhoods $N_i \subseteq \{1, \cdots, n\}$, $i = 1,\cdots,n$, if for $i \in N_i$, X_i is independent of $\{X_j\}_{j\notin N_i}$.*

Lemma 12 *Given a graph G, let $D = 1 +$ maximum degree[5] of G. Let $W = \sum_{i\in V}^{n} X_i$, then*

$$\text{Var}(W) \le D \sum_{i\in V} \text{Var}(X_i). \tag{22}$$

Proof. $Var(W)$ can be expressed as the following:

$$\text{Var}(W) = \text{Var}(\sum_{i\in V}^{n} X_i) = \sum_{i,j}[\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)].$$

We denote the neighborhood of i by N_i, and if $j \in N_i$, then $j \sim i$.
WLOG, we assume that $\mathbb{E}(X_i) = 0$, then

$$\text{Var}(W) = \sum_{i,j}\mathbb{E}(X_iX_j) = \sum_{i,j\sim i}\mathbb{E}(X_iX_j).$$

Note that $X_iX_j \le \frac{X_i^2+X_j^2}{2}$, thus $\text{Var}(W)$ satisfies the following inequality

$$\begin{aligned}\text{Var}(W) &\le \sum_{i,j\sim i}\mathbb{E}(\frac{X_i^2+X_j^2}{2})\\ &\le \sum_{i,j\sim i}\frac{\text{Var}(X_i)+\text{Var}(X_j)}{2}\\ &\le D\sum_{i\in V}\text{Var}(X_i),\end{aligned}$$

This finishes the proof. □

3.1.1 Application of Dependency Graph

In this section, we will introduce an example about how to use **Lemma 12** to compute the upper bound of Wasserstein distance. This example also illustrates that a sum of locally dependent random variables will be approximately normal.

Example Suppose that $\mathbb{E}(X_i)=0,\ \text{Var}(\sum_i X_i)=\sigma^2,\ W=\frac{\sum_i X_i}{\sigma}$ and $Z\sim N(0,1)$, and then we have

$$Wass(W,Z) \le \frac{2}{\sigma^2\sqrt{\pi}}\sqrt{D^3\sum_i \mathbb{E}|X_i|^4}+\frac{D^3}{\sigma^3}\sum_i \mathbb{E}|X_i|^3.$$

Proof. Let $W_i=\frac{1}{\sigma}\sum_{j\notin N_i}X_j$. Obviously, W_i is independent of X_i, and $W-W_i=\frac{1}{\sigma}\sum_{j\in N_i}X_j$. Take any f such that $|f|\le 1,\ |f'|\le\sqrt{\frac{2}{\pi}},\ |f''|_\infty\le 2$.
Let's first look at $\mathbb{E}[Wf(W)]$. By assumption, $\mathbb{E}(X_i)=0$, and thus we have

$$\begin{aligned}\mathbb{E}[Wf(W)] &= \frac{1}{\sigma}\sum_i \mathbb{E}[X_if(W)]\\ &= \frac{1}{\sigma}\sum_i \mathbb{E}[X_if(W)]-\frac{1}{\sigma}\sum_i \mathbb{E}[X_if(W_i)]\\ &= \frac{1}{\sigma}\sum_i \mathbb{E}\big[\big(X_i(f(W)-f(W_i)\big)\big]\\ &= (I)\ +\ (II),\end{aligned}$$

where $(I)=\frac{1}{\sigma}\sum_i \mathbb{E}[X_i(f(W)-f(W_i)-(W-W_i)f'(W))]$ and $(II)=\frac{1}{\sigma}\sum_i \mathbb{E}[X_i(W-W_i)f'(W)]$.
Note that $W-W_i=\frac{1}{\sigma}\sum_{j\in N_i}X_j$. As a result of Taylor expansion and conditions satisfies by f, we

can get

$$
\begin{aligned}
(I) &\leq \frac{1}{\sigma}\sum_i \mathbb{E}\big|X_i(W-W_i)^2 \cdot \frac{f''}{2}\big| \\
&\leq \frac{1}{2\sigma}|f''|\sum_i \mathbf{E}\big|X_i(W-W_i)^2\big| \\
&\leq \frac{1}{\sigma^3}\sum_i \mathbb{E}|X_i(\sum_{j\in N_i} X_j)^2|.
\end{aligned}
$$

Plugging in $W - W_i = \frac{1}{\sigma}\sum_{j\in N_i} X_j$ to equation (II), we have

$$
\begin{aligned}
(II) &= \frac{1}{\sigma}\sum_i \mathbb{E}\big[X_i(W-W_i)f'(W)\big] \\
&= \frac{1}{\sigma}\sum \mathbb{E}[X_i(\frac{1}{\sigma}\sum_{j\in Ni} X_j)f'(W)] \\
&= \mathbb{E}\big[f'(W)\big(\frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j)\big)\big].
\end{aligned}
$$

Let $T = \frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j)$, then the expectation of T satisfies

$$
\begin{aligned}
\mathbb{E}(T) &= \frac{1}{\sigma^2}\sum \mathbb{E}[X_i \cdot \sigma(W-W_i)] \\
&= \frac{1}{\sigma}\sum \mathbb{E}[X_i W] = \frac{1}{\sigma}\mathbb{E}[W\sum_i X_i] \\
&= \mathbb{E}(W^2) = 1.
\end{aligned}
$$

Next, we can bound the absolute value of $(II) - f'(W)$ by

$$
\begin{aligned}
|(II) - f'(W)| &= \big|\mathbb{E}[f'(W)(T-1)]\big| \\
&\leq |f'|_\infty \cdot \mathbb{E}|T-1| = \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}|T-1| \cdot \mathbb{E}|T-1|} \\
&\leq \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}(T-1)^2} = \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)},
\end{aligned}
$$

Note that $|\mathbb{E}Wf(W) - \mathbb{E}f'(W)| \leq (I) + |(II) - f'(W)|$. Combining the above results, we can obtain that

$$
|\mathbb{E}Wf(W) - \mathbb{E}f'(W)| \leq \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)} + \frac{1}{\sigma^3}\sum \mathbb{E}\big|X_i(\sum_{j\in N_i} X_j)^2\big|
$$

and $\frac{1}{\sigma^3}\sum\mathbb{E}\left|X_i(\sum_{j\in N_i}X_j)^2\right|$ can be bounded by

$$\begin{aligned}\frac{1}{\sigma^3}\sum\mathbb{E}|X_i(\sum_{j\in N_i}X_j)^2| &\leq \frac{1}{\sigma^3}\sum_i\sum_{j,k\in N_i}\mathbb{E}|X_iX_jX_k| \\ &\leq \frac{1}{3\sigma^3}\sum_i\sum_{j,k\in N_i}(\mathbb{E}|X_i|^3+\mathbb{E}|X_j|^3+\mathbb{E}|X_k|^3) \\ &\leq \frac{D^2}{\sigma^3}\sum_i\mathbb{E}|X_i|^3,\end{aligned}$$

where the last two inequalities come from A-G inequality: $\frac{X_1+\cdots+X_n}{n}\geq\sqrt[n]{X_1X_2\cdots X_n}$ [6].
To get the upper bound of $\mathrm{Var}(T)$, let's first compute the upper bound of $\mathrm{Var}(\sum X_i(\sum_{j\in N_i}X_j))$.

$$\mathrm{Var}(\sum X_i(\sum_{j\in N_i}X_j)) = \mathrm{Var}(\sum_{i,j\in N_i}X_iX_j),$$

where

$$\mathrm{Var}(X_iX_j)\leq\mathbb{E}(X_i^2X_j^2)\leq\frac{\mathbb{E}(X_i^4)+\mathbb{E}(X_j^4)}{2}.$$

Finally, using the result in **Lemma 5**, we can obtain that

$$\mathrm{Var}(\sum_{i,j\in N_i}X_iX_j)\leq 2D^2\sum_{i\sim j}\mathrm{Var}(X_iX_j)\leq 2D^2\times D\sum_i\mathbb{E}(X_i^4),$$

because of the following reason: X_iX_j is independent of X_kX_l if $k,l\notin\{N_i\cup N_j\}$. $|N_i\cup N_j|\leq 2D$ and since $N_i,N_j\subseteq V$, each vertex in $N_i\cup N_j$ has at most D neighbors, which implies that the maximum degree of the new dependency graph is $2D^2$ and $\{X_iX_j,\ i\in V,\ j\in N_i\}$ is a collection with a dependency graph of maxmimum degree $2D^2$.
Thus,

$$\mathrm{Var}(T)\leq\frac{1}{\sigma^4}\cdot 2D^3\sum_i\mathbb{E}X_i^4.$$

Combining all these results together, we get

$$Wass(W,Z)\leq\frac{2}{\sigma^2\sqrt{\pi}}\sqrt{D^3\sum_i\mathbb{E}|X_i|^4}+\frac{D^3}{\sigma^3}\sum_i\mathbb{E}|X_i|^3.$$

□

[6] $X_iX_jX_k=\sqrt[3]{X_i^3X_j^3X_k^3}\leq\frac{X_i^3+X_j^3+X_k^3}{3}$

3.2 Method of Exchangeable Pairs

Definition 4 *(W,W') is an exchangeable pair of random variables if (W,W') and (W,W') have the same distribution, i.e. $(W,W') \stackrel{d}{=} (W',W)$.*

The following lemma gives an upper bound of the Wasserstein distance between random variables W and Z using the method of exchangeable pair.

Lemma 13 *Suppose (W,W') is an exchangeable pair and there is a constant $\lambda \in (0,1)$ such that*

$$\mathbb{E}(W'-W|W) = -\lambda W. \tag{23}$$

Also assume that $\mathbb{E}(W^2)=1$, then we have the following inequality:

$$\begin{aligned} Wass(W,Z) &= \sup\big|\mathbb{E}f'(w) - \mathbb{E}(Wf(W))\big| \\ &\leq \sqrt{\frac{2}{\pi}\text{Var}[\mathbb{E}(\frac{1}{2\lambda}(W'-W)^2|W)]} + \frac{1}{3\lambda}\mathbb{E}|W'-W|^3, \end{aligned}$$

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

Based on the definition of exchangeable pair, we have $\mathbb{E}(W) = \mathbb{E}(W')$ and $\mathbb{E}(W^2) = \mathbb{E}(W'^2) = 1$. Additionally, from equation (23), we have the following conclusions:

1. $\mathbb{E}(W) = 0$, since
$$\mathbb{E}[\mathbb{E}(W-W'|W)] = \mathbb{E}(W-W') = \mathbb{E}(\lambda W) = 0.$$

2. $\mathbb{E}(W'-W)^2 = 2\lambda$, since
$$\begin{aligned} \mathbb{E}(W'-W)^2 &= \mathbb{E}[W^2 + W'^2 - 2WW'] \\ &= \mathbb{E}[2W^2 - 2WW'] \\ &= \mathbb{E}[2W(W-W')] \\ &= \mathbb{E}(\mathbb{E}[2W(W-W')|W]) \\ &= \mathbb{E}(2W\mathbb{E}[(W-W')|W]) \\ &= \mathbb{E}\big(2W\mathbb{E}(\lambda W)\big) = 2\lambda\mathbb{E}(W^2) = 2\lambda \end{aligned}$$

Next, we will prove **Lemma 13**.
Proof. Take any twice differentiable function f such that

$$|f| \leq 1, \quad |f'| \leq \sqrt{\frac{2}{\pi}}, \quad |f''|_\infty \leq 2.$$

Let $F'(x) = f(x)$, and by Taylor expansion, we have

$$0 = \mathbb{E}(F(W') - F(W)) = \mathbb{E}[(W'-W)f(W)] + \frac{1}{2}\mathbb{E}[(W'-W)^2 f'(W)] + r, \tag{24}$$

where $r = \frac{1}{6}\mathbb{E}\big[(W'-W)^3 f''(\widetilde{W})\big]$, $\widetilde{W}$ is between W and W', and thus $|r| \le \frac{1}{6}|f''|_\infty \cdot \mathbb{E}|W'-W|^3 \le \frac{1}{3}\mathbb{E}|W'-W|^3$. Since there exists $\lambda \in (0,1)$ such that

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

multiplying both sides by $f(W)$ and taking expectation, then we have

$$\mathbb{E}\big[(W'-W)f(W)\big] = -\lambda \mathbb{E}[Wf(W)].$$

Based on the relationship implied by equation (24), we can get

$$\begin{aligned}
-\lambda \mathbb{E}[Wf(W)] &= \mathbb{E}\big[(W'-W)f(W)\big] \\
&= -\frac{1}{2}\mathbb{E}\big[(W'-W)^2 f'(W)\big] - r \\
&= -\mathbb{E}\big[\frac{1}{2}\mathbb{E}((W'-W)^2 f'(W)|W)\big] - r \\
&= -\mathbb{E}\big[\frac{1}{2}f'(W)\mathbb{E}((W'-W)^2|W)\big] - r.
\end{aligned}$$

Next we can bound $\mathbb{E}\big[f'(W) - Wf(W)\big]$ using the above computations,

$$\begin{aligned}
|\mathbb{E}f'(W) - \mathbb{E}[Wf(W)]| &\le |\mathbb{E}[f'(W)(\mathbb{E}(\frac{1}{2\lambda}(W'-W)^2|W) - 1)]| + \frac{1}{3\lambda}\mathbb{E}|W'-W|^3 \\
&\le \sqrt{\frac{2}{\pi}}\mathbb{E}|\mathbb{E}(\frac{1}{2\lambda}(W'-W)^2|W) - 1| + \frac{1}{3\lambda}\mathbb{E}|W'-W|^3.
\end{aligned}$$

Let $Y = \mathbb{E}(\frac{1}{2\lambda}(W'-W)^2|W)]$, then we have $\mathbb{E}(Y) = 1$, and

$$\mathbb{E}\big[Y - \mathbb{E}(Y)\big] = \sqrt{\mathbb{E}^2\big[Y - \mathbb{E}(Y)\big]} \le \sqrt{\mathbb{E}\big[(Y - \mathbb{E}(Y))^2\big]} = \sqrt{\mathrm{Var}(Y)},$$

where the last inequality comes from Jessen's inequality. Therefore,

$$Wass(W,Z) \le \sqrt{\frac{2}{\pi}\mathrm{Var}[\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)]} + \frac{1}{3\lambda}\mathbb{E}|W'-W|^3.$$

□

Remark: this inequality also holds if W and W' have the same distribution, but not necessarily exchangeable.

3.2.1 Construction of an Exchangeable Pair

General idea for construction: Let $X_1', X_2', \cdots, X_n'$ be an independent copy of $X_1, X_2, \cdots, X_n$, and choose index I uniformly at random from $\{1,2,\cdots,n\}$. i.e. $(X_1', X_2', \cdots, X_n')$ is independent of $(X_1, X_2, \cdots, X_n)$.

Example

$X_1, X_2, \cdots, X_n$ is a sequence of random variables, let $W = \sum_i^n X_i$ and

$$W' = \sum_{j \neq I}^{n} X_j + X_I' = W + X_I' - X_I,$$

then (W, W') is an exchangeable pair.

Proof. In order to prove that $(W, W') \overset{\text{d}}{=} (W', W)$, we need to show that

$$\mathbf{P}(W \in A,\ W' \in B) = \mathbf{P}(W' \in A, W \in B).$$

Let $S_{-i} = X_1 + \cdots + X_{i-1} + X_{i+1} + \cdots + X_n$. Since $\mathbf{P}(I = i) = \frac{1}{n}$, we have

$$\mathbf{P}(W \in A,\ W' \in B) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}(W \in A, W_i' \in B).$$

Since X_i' is an independent copy of X_i, we have $X_i' \perp X_i,\ X_i' \perp S_{-i},\ X_i \perp S_{-i}$, where "$\perp$" denotes independence, and then we can conclude that $(S_{-i},\ X_i,\ X_i') \overset{\text{d}}{=} (S_{-i},\ X_i',\ X_i)$.

Note that $W = S_{-i} + X_i$, $W' = S_{-i} + X_i'$, thus we obtain that

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^{n} \mathbf{P}(W \in A, W_i' \in B) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}(W_i' \in A, W \in B) \\ &= \mathbf{P}(W' \in A, W \in B).\end{aligned}$$

Therefore, W and W' is an exchangeable pair. □

3.2.2 Application of exchangeable pairs in proving Central Limit Theorem

In this part, we will prove the Ordinary Central Limit Theorem for weighted sum of independent random variables and Hoeffding Central limit theorem, in which random variables are not independent, using the method of exchangeable pairs.

3.2.2.1 Weighted sum of independent random variables

Let $X_1, X_2, \cdots, X_n$ be independent random variables with mean 0, variance 1 and $\mathbb{E}|X_i|^4 \leq \infty$.
Let

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

then W approximately follows standard Gaussian distribution.

Proof. Let $X_1', X_2', \cdots, X_n'$ be the independent copy of $X_1, X_2, \cdots, X_n$, and choose index I uniformly and randomly from $\{1, 2, \cdots, n\}$. In addition, X_i' is independent of X_i, and X_i, X_i' have the same distribution.

Since (W, W') is an exchangeable pair, then we need to check if there exists $\lambda \in (0,1)$ such that

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

Let $W' = \frac{1}{\sqrt{n}}\sum_{j\neq I}^{n} X_j + \frac{X_I'}{\sqrt{n}} = W + \frac{X_I' - X_I}{\sqrt{n}}$, then we have $W' - W = \frac{X_I' - X_I}{\sqrt{n}}$, and we can compute $\mathbb{E}(W' - W|W)$ as follows

$$\begin{aligned}
\mathbb{E}(W' - W|W) &= \frac{1}{\sqrt{n}}\mathbb{E}(X_I' - X_I|W) \\
&= \frac{1}{\sqrt{n}} \times \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i' - X_i|W) \\
&= \frac{1}{n}\mathbb{E}(\sum_{i=1}^{n} \frac{1}{\sqrt{n}}(X_i' - X_i|W)) \\
&= \frac{1}{n}\mathbb{E}(\sum_{i=1}^{n} \frac{1}{\sqrt{n}}X_i') - \frac{1}{n}\mathbb{E}(W|W) = -\frac{1}{n}W.
\end{aligned}$$

Hence, in our example, $\lambda = \frac{1}{n}$, and we have

$$Wass(W,Z) \leq \sqrt{\frac{2}{\pi}\text{Var}[\mathbb{E}(\frac{1}{2\lambda}(W - W')^2|W)]} + \frac{1}{3\lambda}\mathbb{E}|W - W'|^3.$$

Let's first compute the upper bound of $\frac{1}{3\lambda}\mathbf{E}|W - W'|^3$. By replacing $W' - W$ by $\frac{X_I' - X_I}{\sqrt{n}}$, we have

$$\begin{aligned}
\frac{1}{3\lambda}\mathbb{E}|W - W'|^3 &= \frac{n}{3}\mathbb{E}|\frac{X_I' - X_I}{\sqrt{n}}|^3 \\
&= \frac{n}{3n^{\frac{3}{2}}}\mathbb{E}|X_I' - X_I|^3 = \frac{1}{3n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathbb{E}|X_i' - X_i|^3 \\
&\leq \frac{8}{3n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathbb{E}|X_i|^3
\end{aligned}$$

Next we wiil compute the upper bound of $\text{Var}[\mathbb{E}(\frac{1}{2\lambda}(W - W')^2|W)]$. Since

$$\mathbb{E}(\frac{1}{2\lambda}(W - W')^2|W) = \frac{1}{2}\mathbb{E}[(X_I' - X_I)^2|W] = \frac{1}{2n}[\sum_{i}\mathbb{E}(X_i' - X_i)^2|W],$$

where $\mathbb{E}(X_i'-X_i)^2|W]$ can be expressed as follows

$$\begin{aligned}\mathbb{E}[(X_i'-X_i)^2|W] &= \mathbb{E}(X_i'^2|W)+\mathbb{E}(X_i^2|W)+2\mathbb{E}(X_i'X_i|W)\\ &= 1+\mathbb{E}(X_i^2|W)+0\\ &= 1+\mathbb{E}(X_i^2|W),\end{aligned}$$

thus we obtain that

$$\begin{aligned}\mathrm{Var}[\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)] &= \mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}(1+\mathbb{E}(X_i^2|W))]\\ &= \mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(X_i^2|W)].\end{aligned}$$

However, $\mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(X_i^2|W)]$ is not easy to compute. Thus we need the following property to compute the upper bound of $\mathrm{Var}[\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)]$.

Remark

$$\mathrm{Var}\big(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)\big)\le \mathrm{Var}\big(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|\mathscr{F})\big)$$

for any σ-field $\mathscr{F}$ that is larger than the σ-field generated by W. Thus we can compute the upper bound of $\mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|\mathscr{F}))$ instead. Let's first prove why this inequality holds.

Proof. Recall that $\mathrm{Var}(X)=\mathbb{E}(X^2)-\mathbb{E}^2(X)$, thus we have

$$\begin{aligned}\mathrm{Var}(\mathbb{E}((W-W')^2|W)) &= \mathbb{E}\{\mathbb{E}[(W-W')^2|W]\}^2-\mathbb{E}^2(\mathbb{E}[(W-W')^2|W])\\ &= \mathbb{E}\{\mathbb{E}[(W-W')^2|W]\}^2-\mathbb{E}^2[(W-W')^2]\\ \mathrm{Var}(\mathbb{E}((W-W')^2|\mathscr{F})) &= \mathbb{E}\{\mathbb{E}[(W-W')^2|\mathscr{F}]\}^2-\mathbb{E}^2[(W-W')^2]\end{aligned}$$

Hence, we only need to compare $\mathbb{E}\{\mathbb{E}[(W-W')^2|W]\}^2$ and $\mathbb{E}\{\mathbb{E}[(W-W')^2|\mathscr{F}]\}^2$ to get the relationship between $\mathrm{Var}(\mathbb{E}((W-W')^2|W))$ and $\mathrm{Var}(\mathbb{E}((W-W')^2|\mathscr{F}))$. Let $X=(W-W')^2$, then we will compare $\mathbb{E}[\mathbb{E}(X|W)]^2$ and $\mathbb{E}[\mathbb{E}(X|\mathscr{F})]^2$.
Take $W=\sum_{i=1}^{n}X_i$ and $\mathscr{F}=(X_1,\cdots,X_n)$, then we have

$$\mathbb{E}(X|W)=\mathbb{E}(X|\mathscr{F})=\mathbb{E}[\mathbb{E}(X|\mathscr{F})|W],$$

because of the fact that $\sigma(W)\subseteq\mathscr{F}$, W contains less information than $\mathscr{F}$, and $\mathscr{F}$ includes more randomness. Thus,

$$\mathbb{E}[\mathbb{E}(X|W)]^2=\mathbb{E}[\mathbb{E}(\mathbb{E}(X|\mathscr{F})|W)]^2.$$

Let $Y=\mathbb{E}(X|\mathscr{F})$, $\mathbb{E}^2(Y|W)\le\mathbb{E}[Y^2|W]$, which comes from Jensen's inequality, and then we can get

$$\mathbb{E}[\mathbb{E}(X|W)]^2\le\mathbb{E}[\mathbb{E}(Y^2|W)]=\mathbb{E}(Y^2)=\mathbb{E}[\mathbb{E}(X|\mathscr{F})]^2.$$

Therefore,

$$\mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)) \leq \mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|\mathscr{F}))$$

□

Given the property mentioned above, then we can compute the upper bound of the variance as follows

$$\begin{aligned}
\mathrm{Var}[\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)] &= \mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(X_i^2|W)] \\
&\leq \mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(X_i^2|(X_1,\cdots,X_n))] \\
&= \mathrm{Var}[\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(X_i^2|X_i)] \\
&= \frac{1}{4n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i^2) = \frac{1}{4n^2}\sum_{i=1}^{n}[\mathbb{E}(X_i^4)-\mathbb{E}^2(X_i^2)] \\
&\leq \frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{E}(X_i^4)
\end{aligned}$$

Therefore, if $\mathbb{E}(X_i^4) < \infty$,

$$Wass(W,Z) \leq \sqrt{\frac{1}{2n^2\pi}\sum_{i=1}^{n}\mathbb{E}(X_i^4)} + \frac{8}{3n^{\frac{3}{2}}}\mathbb{E}|X_i|^3 \ \rightarrow\ 0,\ \text{as}\ n\ \rightarrow\ \infty$$

and $W \dot{\sim} N(0,1)$. □

3.2.2.2 Hoeffding Combinatorial Central Limit Theorem

Suppose $(a_{i,j})_{i,j=1}^{n}$ is an array of numbers. Let π be a uniform random permutation of $\{1,2,\cdots,n\}$. Let $W = \sum_{i=1}^{n} a_{i\pi(i)}$. Then we have Wasserstein distance bounded by

$$\mathrm{Wass}(W,Z) \leq L\sqrt{\frac{1}{n}\sum_{i,j}a_{ij}^4} + \frac{\tilde{L}}{n}\sum_{i,j}|a_{ij}|^3,$$

where L, $\tilde{L}$ are some constants.

Proof. WLOG, we can assume that

$$\sum_{i=1}^{n}a_{ij} = 0, \quad \sum_{j=1}^{n}a_{ij} = 0, \quad \frac{1}{n-1}\sum_{i,j=1}^{n}a_{ij}^2 = 1,$$

based on which we can show that

$$\frac{W - \mathbb{E}(W)}{\sqrt{\text{Var}(W)}} \sim N(0,1).$$

at the end of our proof.

Let's first prove why this assumption does not compromise generality. Let $\sigma = \sqrt{\text{Var}(W)}$ and $\mu = \mathbb{E}(W)$. Define

$$a_{i.} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}, \quad a_{.j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}, \quad a_{..} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij},$$

and

$$\tilde{a}_{ij} = \frac{a_{ij} - a_{i.} - a_{.j} + a_{..}}{\sigma}.$$

Since $\mathbb{E}(W) = \mu$, we can express μ in terms of $a_{..}$ based on the definition of $a_{i.}$, $a_{.j}$, $a_{..}$ and $\tilde{a}_{ij}$

$$\mu = \mathbb{E}(W) = \mathbb{E}(\sum_{i=1}^{n} a_{i\pi(i)}) = \frac{1}{n}\sum_{i,j} a_{ij} = \frac{1}{n}\cdot n^2 a_{..} = na_{..},$$

and we can also find the explicit form of $\tilde{a}_{i.}$ and $\tilde{a}_{.j}$

$$\begin{aligned}
\tilde{a}_{i.} &= \frac{1}{n}\sum_{j=1}^{n}\tilde{a}_{ij} = \frac{1}{n}\sum_{j=1}^{n}(\frac{a_{ij} - a_{i.} - a_{.j} + a_{..}}{\sigma}) \\
&= \frac{1}{\sigma}\cdot\frac{1}{n}\sum_{j=1}^{n}(a_{ij} - a_{i.} - a_{.j} + a_{..}) \\
&= \frac{1}{\sigma n}(\sum_{j=1}^{n} a_{ij} - \sum_{j=1}^{n} a_{i.} - \sum_{j=1}^{n} a_{.j} + \sum_{j=1}^{n} a_{..}) \\
&= \frac{1}{\sigma n}(\sum_{j=1}^{n} a_{ij} - \sum_{j=1}^{n}\frac{1}{n}\sum_{j=1}^{n} a_{ij} - \frac{1}{n}\sum_{i,j} a_{ij} + \frac{1}{n}\sum_{i,j} a_{ij}) \\
&= \frac{1}{\sigma n}(\sum_{j=1}^{n} a_{ij} - \sum_{j=1}^{n} a_{ij} - \frac{1}{n}\sum_{i,j} a_{ij} + \frac{1}{n}\sum_{i,j} a_{ij})) \\
&= 0.
\end{aligned}$$

Similarly, we can also get $\tilde{a}_{.j} = 0$. Next, rewrite $\widetilde{W}$ and we obtain that

$$\begin{aligned}\widetilde{W} &= \frac{1}{\sigma}\sum_{i=1}^{n}\tilde{a}_{i\pi(i)} = \frac{1}{\sigma}\sum_{i=1}^{n}(a_{i\pi(i)} - a_{i.} - a_{.\pi(i)} + a_{..})\\ &= \frac{1}{\sigma}(\sum_{i=1}^{n}a_{i\pi(i)} - \sum_{i=1}^{n}a_{i.} - \sum_{i=1}^{n}a_{.\pi(i)} + na_{..})\\ &= \frac{1}{\sigma}(\sum_{i=1}^{n}a_{i\pi(i)} - na_{..})^{7}\\ &= \frac{1}{\sigma}(W-\mu).\end{aligned}$$

Hence,

$$\mathbf{E}(\widetilde{W}) = \frac{1}{\sigma}(\mathbf{E}(W)-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0.$$

Note that

$$\frac{W-\mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} = \frac{\widetilde{W}-\mathbb{E}(\widetilde{W})}{\sqrt{\mathrm{Var}(\widetilde{W})}},$$

as a result of the following fact:

$$\begin{aligned}\frac{\widetilde{W}-\mathbb{E}(\widetilde{W})}{\sqrt{\mathrm{Var}(\widetilde{W})}} &= \frac{\frac{1}{\sigma}(\sum_{i=1}^{n}a_{i\pi(i)} - na_{..}) - \frac{1}{\sigma}\mathbb{E}(\sum_{i=1}^{n}a_{i\pi(i)} - na_{..})}{\sqrt{\frac{1}{\sigma^2}\mathrm{Var}(\sum_{i=1}^{n}a_{i\pi(i)} - na_{..})}}\\ &= \frac{\big(W-\mathbb{E}(W)\big) - \big(na_{..} - \mathrm{E}(na_{..})\big)}{\sqrt{\mathrm{Var}(\sum_{i=1}^{n}a_{i\pi(i)} - na_{..})}}\\ &= \frac{W-\mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}},\end{aligned}$$

since $na_{..}$ is constant.
Thus, if we can show that $\frac{\widetilde{W}-\mathbb{E}(\widetilde{W})}{\sqrt{\mathrm{Var}(\widetilde{W})}} \sim N(0,1)$, then we can conclude that $\frac{W-\mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} \sim N(0,1)$.

Now, let's continue proving why the assumption made at the beginning does not compromise generality. By simple computations, we obtain that $\mathrm{Var}(\widetilde{W})$ satisfies

$$\mathrm{Var}(\widetilde{W}) = \mathrm{Var}(\frac{W-\mu}{\sigma}) = \frac{1}{\sigma^2}\mathrm{Var}(W) = \frac{\sigma^2}{\sigma^2} = 1,$$

and both $\mathbb{E}[\tilde{a}_{i\pi(i)}]$ and $\mathbf{E}[\tilde{a}_{j\pi(j)}]$ equal to 0

$$\mathbb{E}[\tilde{a}_{i\pi(i)}] = \frac{1}{n}\sum_{j=1}^{n}\tilde{a}_{ij} = \tilde{a}_{i.} = 0, \quad \mathbf{E}[\tilde{a}_{j\pi(j)}] = 0.$$

[7] $\sum_{i=1}^{n}a_{i\pi(i)} = \sum_{i=1}^{n}\frac{1}{n}\sum_{k=1}^{n}a_{k\pi(i)} = \frac{1}{n}\sum_{j,k}a_{kj} = \sum_{i,j}a_{ij} = na_{..}$.

Rewrite $\mathrm{Var}(\widetilde{W})$

$$\begin{aligned}\mathrm{Var}(\widetilde{W}) &= \mathrm{Var}[\sum_{i=1}^{n}\tilde{a}_{i\pi(i)}]\\ &= \sum_{i=1}^{n}\mathrm{Var}(\tilde{a}_{i\pi(i)}) + \sum_{i\neq j}\mathrm{cov}(\tilde{a}_{i\pi(i)},\tilde{a}_{j\pi(j)})\\ &= \sum_{i=1}^{n}\left[\mathbb{E}(\tilde{a}^2_{i\pi(i)}) - \mathbb{E}^2(\tilde{a}_{i\pi(i)})\right] + \sum_{i\neq j}\left[\mathbb{E}(\tilde{a}_{i\pi(i)}\tilde{a}_{j\pi(j)}) - \mathbb{E}(\tilde{a}_{i\pi(i)})\mathbb{E}(\tilde{a}_{j\pi(j)})\right]\\ &= \sum_{i=1}^{n}\mathbb{E}(\tilde{a}^2_{i\pi(i)}) + \sum_{i\neq j}\mathbb{E}(\tilde{a}_{i\pi(i)}\tilde{a}_{j\pi(j)})\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{a}^2_{ij} + \frac{1}{n(n-1)}\sum_{i\neq j}\sum_{k\neq l}\tilde{a}_{ik}\tilde{a}_{jl}.\end{aligned}$$

And since $\tilde{a}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\tilde{a}_{ij} = 0$, we have $\sum_{j=1}^{n}\tilde{a}_{ij} = 0$ and similarly, $\sum_{i=1}^{n}\tilde{a}_{ij} = 0$. Then $\sum_{i\neq j}\sum_{k\neq l}\tilde{a}_{ik}\tilde{a}_{jl}$ can be expressed as follows

$$\begin{aligned}\sum_{i\neq j}\sum_{k\neq l}\tilde{a}_{ik}\tilde{a}_{jl} &= \sum_{i}\sum_{j\neq i}\sum_{k}\sum_{l\neq k}\tilde{a}_{ik}\tilde{a}_{jl}\\ &= \sum_{i}\sum_{k}\tilde{a}_{ik}\sum_{j\neq i}(\sum_{l\neq k}\tilde{a}_{jl})\\ &= \sum_{i}\sum_{k}\tilde{a}_{ik}\sum_{j\neq i}(-\tilde{a}_{jk})\\ &= \sum_{i,k}\tilde{a}_{ik}\cdot\tilde{a}_{ik} = \sum_{i,k}\tilde{a}^2_{ik}.\end{aligned}$$

Thus, $\mathrm{Var}(\widetilde{W})$ finally has the form

$$\begin{aligned}1 = \mathrm{Var}(\widetilde{W}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{a}^2_{ij} + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{a}^2_{ij} = \frac{1}{n-1}\sum_{i,j}\tilde{a}^2_{ij}\\ &= \frac{1}{n-1}\sum_{i,j}(\frac{a_{ij}-a_{i.}-a_{.j}+a_{..}}{\sigma})^2\end{aligned}$$

Therefore,

$$\mathrm{Var}(W) = \sigma^2 = \frac{1}{n-1}\sum_{i,j}(a_{ij}-a_{i.}-a_{.j}+a_{..})^2 = \frac{1}{n-1}(\sum_{i,j}a^2_{i,j} - n\sum_{i=1}^{n}a^2_{i.} - n\sum_{j=1}^{n}a^2_{.j} + n^2a^2_{..}).$$

On the basis of the relationship that we have analyzed between W and $\widetilde{W}$, we can see that it's reasonable to assume that

$$\sum_{i=1}^{n} a_{i,j} = 0, \quad \sum_{j=1}^{n} a_{i,j} = 0, \quad \frac{1}{n-1}\sum_{i,j}^{n} a_{i,j}^2 = 1,$$

where $\sum_{j=1}^{n} a_{i,j} = \tilde{a}_{i.}$, $\sum_{i=1}^{n} a_{i,j} = \tilde{a}_{.j}$, and $\frac{1}{n-1}\sum_{i,j}^{n} a_{i,j}^2$ is equivalent to

$$\frac{1}{n-1}\cdot\frac{1}{\sigma^2}\sum_{i,j}(a_{ij} - a_{i.} - a_{.j} + a_{..})^2 = \mathrm{Var}(\widetilde{W}) = 1,$$

and then we have $\mathbb{E}(W) = 0$, $\mathrm{Var}(W) = 1$.

Before computing the upper bound of Wasserstein distance, we first need to construct an exchangeable pair (π, π'), which will be applied to construct exchangeable pair (W, W').

Theorem 1 *Let $\pi'(I) = \pi(J)$, $\pi'(J) = \pi(I)$ and $\pi'(k) = \pi(k)$ if $k \neq I, J$, where I, J are selected uniformly at random on $\{1, 2, \cdots, n\}$, where π' is denoted by $\pi \circ (I, J)$. Then (π, π') is an exchangeable pair.*

Proof. We know that π is permutation from $\{1, 2, \cdots, n\}$, and $\pi(I) = \frac{1}{n}\sum_{i=1}^{n}\pi(i)$, then we have

$$\begin{aligned}\mathbf{P}(\pi \in A, \pi' \in B) &= \mathbf{P}(\pi \in A, \pi \circ (I, J) \in B)\\ &= \frac{1}{n(n-1)}\sum_{i,j}\mathbf{P}(\pi \in A, \pi \circ (i, j) \in B).\end{aligned}$$

Let $\widetilde{\pi} = \pi \circ (i, j)$, which is still a uniform permutation of indices. and thus, $\pi = \widetilde{\pi} \circ (i, j)$. So we finally get

$$\mathbf{P}(\pi \in A, \pi \circ (i, j) \in B) = \mathbf{P}(\widetilde{\pi} \circ (i, j) \in A, \widetilde{\pi} \in B) = \mathbf{P}(\pi \in B, \pi \circ (i, j) \in A),$$

where the last equality comes from the fact that π and $\widetilde{\pi}$ have the same distribution.
Therefore, $(\pi, \pi') \stackrel{\mathrm{d}}{=} (\pi', \pi)$. □

Now we can construct an exchageable pair for W. Let $W' = \sum_{i=1}^{n} a_{i\pi(i)}$. Then (W, W') is an exchangeable pair. And obviously,

$$\begin{aligned}W' - W &= a_{I\pi'(I)} + a_{J\pi'(J)} - a_{I\pi(I)} - a_{J\pi(J)}\\ &= a_{I\pi(J)} + a_{J\pi(I)} - a_{I\pi(I)} - a_{J\pi(J)}\end{aligned}$$

Recall that if W and W' is an exchangeable pair, then $\mathrm{E}(W' - W|W) = -\lambda W$. To obtain λ, we firstly compute $\mathbb{E}(W' - W|\pi)$, and later we will see that $\mathrm{E}(W' - W|W) = \mathrm{E}(W' - W|\pi)$.

$$\mathbb{E}(W' - W|\pi) = \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}(a_{i\pi(j)} + a_{j\pi(i)} - a_{i\pi(i)} - a_{j\pi(j)}).$$

Note that

$$\begin{aligned}\frac{1}{n(n-1)}\sum_{i\neq j}a_{i\pi(i)} &= \frac{1}{n(n-1)}\sum_{i}\sum_{j\neq i}a_{i\pi(i)}\\ &= \frac{1}{n(n-1)}\cdot(n-1)\sum_{i}a_{i\pi(i)}\\ &= \frac{W}{n}.\end{aligned}$$

Similarly,

$$\frac{1}{n(n-1)}\sum_{i\neq j}a_{j\pi(j)} = \frac{W}{n}.$$

Additionally, by assumption, we have $\sum_{i\neq j}a_{i\pi(j)} + a_{i\pi(i)} = 0$ by fixing i, then we can obtain that

$$\frac{1}{n(n-1)}\sum_{i\neq j}a_{i\pi(j)} = \frac{1}{n(n-1)}\sum_{i}\sum_{i\neq j}a_{i\pi(j)} = -\frac{1}{n(n-1)}\sum_{i}a_{i\pi(i)} = -\frac{1}{n(n-1)}W.$$

Similarly, $\frac{1}{n(n-1)}\sum_{i\neq j}a_{j\pi(i)} = -\frac{1}{n(n-1)}W$.
Combining all these together, we get the following equation

$$\mathbb{E}(W'-W|\pi) = -\frac{2}{n-1}W, \text{ which only depends on } W. \tag{25}$$

So we have

$$\mathbb{E}(W'-W|W) = \mathbb{E}(W'-W|\pi) = -\frac{2}{n-1}W,$$

and hence in our example, $\lambda = \frac{2}{n-1}$.

Next we will bound the Wasserstein distance by computing the upper bound of $\mathrm{Var}[\mathbb{E}(\frac{1}{2\lambda}(W'-W)^2|W)]$ and $\mathbb{E}|W'-W|^3$ separately in order to complete the proof of Hoeffding combinatorial CLT.

In the previous part we have proved that $\mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)) \leq \mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|\mathscr{F}))$, if W is measurable in σ-field $\mathscr{F}$, and $\mathscr{F}$ is larger, hence we can bound $\mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W))$ by bounding $\mathrm{Var}(\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|\pi))$.

Let's first bound $\mathbb{E}|W'-W|^3$. Note that

$$\mathbb{E}|W'-W|^3 = \mathbb{E}[\mathbb{E}|W'-W|^3|\pi],$$

where $\mathbb{E}\big(|W'-W|^3|\pi\big)$ satisfies

$$\begin{aligned}\mathbb{E}(|W'-W|^3|\pi) &= \frac{1}{n(n-1)}\sum_{i,j\neq i}|a_{i\pi(i)}+a_{j\pi(j)}-a_{i\pi(j)}-a_{j\pi(i)}|^3\\ &\leq \frac{16}{n(n-1)}\sum_{i,j\neq i}(|a_{i\pi(i)}|^3+|a_{j\pi(j)}|^3+|a_{i\pi(j)}|^3+|a_{j\pi(i)}|^3),\end{aligned}$$

where the last inequality is based on the fact that $(a+b-c-d)^3 \leq 16(a^3+b^3+c^3+d^3)$. The we can get

$$\begin{aligned}\frac{1}{3\lambda}\mathbb{E}|W'-W|^3 &\leq \frac{1}{3\lambda}\mathbb{E}[\frac{16}{n(n-1)}\sum_{i,j\neq i}(|a_{i\pi(i)}|^3+|a_{j\pi(j)}|^3+|a_{i\pi(j)}|^3+|a_{j\pi(i)}|^3)]\\ &\leq \frac{n-1}{6}\times\frac{16}{n(n-1)}\sum_i\sum_{i\neq j}4\times\mathbb{E}|a_{i\pi(i)}|^3\\ &\leq \frac{n-1}{6}\times\frac{64}{n(n-1)}\times(n-1)\times\frac{1}{n}\sum_{i,j}|a_{ij}|^3\\ &\leq \frac{n-1}{6}\times\frac{64}{n^2}\sum_{i,j}|a_{ij}|^3.\end{aligned}$$

Note that based on our assumption at the beginning we have $a_{ij}=O(\frac{1}{\sqrt{n}})$, then we have

$$\frac{1}{3\lambda}\mathbb{E}|W'-W|^3 \leq \frac{C}{\sqrt{n}},$$

where C is a constant.
Now we will prove the concentration for the conditional variance in the upper bound of $\text{Wass}(W,Z)$. Let's first look at $\mathbb{E}((W'-W)^2|\pi)$.

$$\mathbb{E}((W'-W)^2|\pi) = \frac{1}{n(n-1)}\sum_{i\neq j}(a_{i\pi(i)}+a_{j\pi(j)}-a_{i\pi(j)}-a_{j\pi(i)})^2. \tag{26}$$

Let $A_{ij}=a_{i\pi(i)}+a_{j\pi(j)}-a_{i\pi(j)}-a_{j\pi(i)}$. Then the above equation (26) becomes

$$\mathbb{E}((W'-W)^2|\pi) = \frac{1}{n(n-1)}\sum_{i\neq j}A_{ij}^2 = X.$$

Based on the fact that $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ and

$$\begin{aligned}\text{Var}(\sum_i X_i) &= \sum_i \text{Var}(X_i) + \sum_{i\neq j} \text{Cov}(X_i, X_j) \\ &= \sum_i \text{Var}(X_i) + \sum_{i\neq j} [\mathbb{E}(X_i X_j - \mathbf{E}(X_i)\mathbf{E}(X_j)] \\ &= \sum_{i,j} [\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)],\end{aligned}$$

we have

$$\begin{aligned}\text{Var}(X) &= \frac{1}{n^2(n-1)^2} \text{Var}(\sum_{i\neq j} A_{ij}^2) \\ &= \frac{1}{n^2(n-1)^2} \sum_{i\neq j}\sum_{k\neq l} [\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)],\end{aligned}$$

where

$$\begin{aligned}\sum_{i\neq j}\sum_{k\neq l}\left[\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)\right] = &\sum_{i\neq j, k\neq l}{}^{8} \left[\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)\right] \\ &+ \sum_{i\neq j\neq k\neq l} \left[\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)\right].\end{aligned}$$

Let $\text{I} = \sum_{i\neq j, k\neq l}\left[\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)\right]$, then we have

$$\begin{aligned}\text{I} &= \sum_{i\neq j, k\neq l} [\mathbb{E}(A_{ij}^2 A_{kl}^2) - \mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)] \leq \sum_{i\neq j, k\neq l} \mathbb{E}(A_{ij}^2 A_{kl}^2) \\ &\leq 4 \sum_{i\neq j, l} \mathbb{E}(A_{ij}^2 A_{il}^2) \leq 4 \sum_{i\neq j, l} \frac{\mathbb{E}(A_{ij}^4) + \mathbb{E}(A_{il}^4)}{2} \quad \text{(using the fact that } xy \leq \frac{x^2+y^2}{2}) \\ &\leq 2 \sum_{i,j,l} [\mathbb{E}(A_{ij}^4) + \mathbb{E}(A_{il}^4)] \\ &\leq 2[n \sum_{i,j} \mathbb{E}(A_{ij}^4) + n \sum_{i,l} \mathbb{E}(A_{il}^4)] = 4n \sum_{i,j} \mathbb{E}(A_{ij}^4).\end{aligned}$$

Plugging in the original form

$$\mathbb{E}(A_{ij}^4) = \mathbb{E}[(a_{i\pi(i)} + a_{j\pi(j)} - a_{i\pi(j)} - a_{j\pi(i)})^4],$$

and using the fact that

$$(a+b+c+d)^4 \leq 8(a^4+b^4+c^4+d^4),$$

we can obtain that

$$\begin{aligned}\mathbb{E}(A_{ij}^4) &\leq 8[\mathbb{E}(a_{i\pi(i)}^4)+\mathbb{E}(a_{j\pi(j)}^4)+\mathbb{E}(a_{i\pi(j)}^4)+\mathbb{E}(a_{j\pi(i)}^4)] \\ &= \frac{8}{n}[2\sum_l a_{il}^4 + 2\sum_l a_{jl}^4].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{I} &\leq 4n\sum_{i,j}\mathbb{E}(A_{ij}^4) \\ &\leq 4n\sum_{i,j}\frac{16}{n}(\sum_l a_{il}^4+\sum_l a_{jl}^4) \\ &\leq 64\sum_{i,j}(\sum_l a_{il}^4+\sum_l a_{jl}^4) = 64(n\sum_{i,l}a_{il}^4 + n\sum_{j,l}a_{jl}^4) = 64\times 2n\sum_{i,j}a_{ij}^4.\end{aligned}$$

Now we move on to compute II, where $\mathrm{II} = \sum_{i\neq j\neq k\neq l}[\mathbb{E}(A_{ij}^2A_{kl}^2)-\mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)]$.

a. Let's first look at $\mathbb{E}(A_{ij}^2A_{kl}^2)$.

$$A_{ij}^2A_{kl}^2 = (a_{i\pi(i)}+a_{j\pi(j)}-a_{i\pi(j)}-a_{j\pi(i)})^2(a_{k\pi(k)}+a_{l\pi(l)}-a_{k\pi(l)}-a_{l\pi(k)})^2.$$

and

$$(\pi(i),\pi(j),\pi(k),\pi(l)) = (i_1,i_2,i_3,i_4), \quad \text{with probability } \frac{1}{n(n-1)(n-2)(n-3)},$$

where $\{i_s\}_{s=1}^4 \in \{1,2,\cdots,n\}$. Thus,

$$\begin{aligned}\mathbb{E}(A_{ij}^2A_{kl}^2) = &\frac{1}{n(n-1)(n-2)(n-3)}\times \\ &\sum_{i_1\neq i_2\neq i_3\neq i_4}(a_{i,i_1}+a_{j,i_2}-a_{i,i_2}-a_{j,i_1})^2(a_{k,i_3}+a_{l,i_4}-a_{k,i_4}-a_{l,i_3})^2.\end{aligned}$$

b. Next we compute $\mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2)$. Let $A_1 = (a_{i,i_1}+a_{j,i_2}-a_{i,i_2}-a_{j,i_1})^2$ and $A_2 = (a_{k,i_3}+a_{l,i_4}-$

$a_{k,i_4} - a_{l,i_3})^2$. Then we have

$$
\begin{aligned}
\mathbb{E}(A_{ij}^2)\mathbb{E}(A_{kl}^2) &= \frac{1}{n(n-1)} \sum_{i_1 \neq i_2} (a_{i,i_1} + a_{j,i_2} - a_{i,i_2} - a_{j,i_1})^2 \times \\
&\quad \frac{1}{n(n-1)} \sum_{i_3 \neq i_4} (a_{k,i_3} + a_{l,i_4} - a_{k,i_4} - a_{l,i_3})^2 \\
&= \frac{1}{n^2(n-1)^2} \sum_{i_1 \neq i_2, i_3 \neq i_4} (a_{i,i_1} + a_{j,i_2} - a_{i,i_2} - a_{j,i_1})^2 (a_{k,i_3} + a_{l,i_4} - a_{k,i_4} - a_{l,i_3})^2 \\
&= \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} A_1 A_2 + \\
&\quad [\frac{1}{n^2(n-1)^2} - \frac{1}{n(n-1)(n-2)(n-3)}] \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} A_1 A_2 + \\
&\quad \frac{1}{n^2(n-1)^2} \sum_{i_1 \neq i_2, i_3 \neq i_4} A_1 A_2.
\end{aligned}
$$

As a result, II can be rewritten as follows

$$
\mathrm{II} = - \sum_{i \neq j \neq k \neq l} \big[(\frac{1}{n^2(n-1)^2} - \frac{1}{n(n-1)(n-2)(n-3)}) \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} A_1 A_2 + \frac{1}{n^2(n-1)^2} \sum_{i_1 \neq i_2, i_3 \neq i_4} A_1 A_2 \big].
$$

Obviously, bounding $|\mathrm{II}|$ is enough. As for $A_1 A_2$, we have the following inequality

$$
\begin{aligned}
A_1 A_2 &\leq \frac{(a_{i,i_1} + a_{j,i_2} - a_{i,i_2} - a_{j,i_1})^4 + (a_{k,i_3} + a_{l,i_4} - a_{k,i_4} - a_{l,i_3})^4}{2} \\
&\leq 4[(a_{i,i_1}^2 + a_{j,i_2}^2 + a_{i,i_2}^2 + a_{j,i_1}^2) + (a_{k,i_3}^2 + a_{l,i_4}^2 + a_{k,i_4}^2 + a_{l,i_3}^2)].
\end{aligned}
$$

Let $(*) = (\frac{1}{n^2(n-1)^2} - \frac{1}{n(n-1)(n-2)(n-3)}) \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} A_1 A_2$, and $(**) = \frac{1}{n^2(n-1)^2} \sum_{i_1 \neq i_2, i_3 \neq i_4} A_1 A_2$. Then when n is large enough, sum of $(*)$ and sum of $(**)$ satisfy

$$
\begin{aligned}
\sum_{i \neq j \neq k \neq l} (*) &\sim \frac{1}{n^5} \sum_{i \neq j \neq k \neq l} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} A_1 A_2 \\
&\leq \frac{C_1}{n^5} \times n^6 \sum_{i,j} a_{ij}^4 = C_1 n a_{ij}^4.
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{i \neq j \neq k \neq l} (**) &\sim \frac{1}{n^4} \sum_{i \neq j \neq k \neq l} \sum_{i_1 \neq i_2, i_3 \neq i_4} A_1 A_2 \\
&\leq \frac{C_2}{n^4} (n^3 \times n^2 \sum_{i,j} a_{ij}^4) = C_2 n \sum_{i,j} a_{ij}^4,
\end{aligned}
$$

where "$\sim$" means approximation.

Thus,

$$|\text{II}| \leq \tilde{C}n\sum_{i,j} a_{ij}^4. \quad (C'\text{s are constants}),$$

and

$$\text{Var}(X) \leq \frac{1}{n^2(n-1)^2} \times Ln\sum_{i,j} a_{ij}^4. \quad (L \text{ is constant) and } \frac{4}{\lambda^2}\text{Var}(X) \leq \frac{L}{n}\sum_{i,j} a_{ij}^4.$$

Therefore,

$$\begin{aligned}\text{Wass}(W,Z) &\leq \sqrt{\frac{2}{\pi}\text{Var}[\mathbb{E}(\frac{1}{2\lambda}(W-W')^2|W)]} + \frac{1}{3\lambda}\mathbb{E}|W-W'|^3 \\ &\leq L\sqrt{\frac{1}{n}\sum_{i,j} a_{ij}^4} + \frac{\tilde{L}}{n}\sum_{i,j}|a_{ij}|^3.\end{aligned}$$

Based on our assumption at the beginnig, we have $a_{ij} \sim O(\frac{1}{n})$, and thus $\text{Wass}(W,Z)$ is bounded by $\frac{C}{\sqrt{n}}$, where C is some constant. Finally, we can conclude that $W \dot{\sim} N(0,1)$, as $n \to \infty$. This finishes the proof of Hoeffding combinatorial central limit theorem. □

3.3 Size-bias Coupling

3.3.1 Some basic idea about size-bias coupling

Definition 5 *For a random variable $\mathbb{E}(X) > 0$ with $\mathbb{E}(X) = \mu < \infty$, we say that the random variable X^s has the size-bias distribution with respect to X if for all f such that $\mathbb{E}|Xf(X)| < \infty$, we have*

$$\mathbb{E}[Xf(X)] = \mu\mathbb{E}[f(X^s)].$$

Theorem 2 *Let $X \geq 0$ be a random variable such that $\mathbb{E}(X) = \mu < \infty$, and $\text{Var}(X) = \sigma^2$. Let X^s be defined on the same space as X and have the size-bias distribution with respect to X. If $W = \frac{X-\mu}{\sigma}$ and $Z \sim N(0,1)$, then*

$$\text{Wass}(X,Z) \leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\text{Var}\big(\mathbb{E}(X^s - X|X)\big)} + \frac{\mu}{\sigma^3}\mathbb{E}[(X^s-X)^2].$$

Proof. Take any twice differentiable function f satisfying $|f|_\infty \leq 1,\ |f'|_\infty \leq \sqrt{\frac{2}{\pi}}$, and $|f''|_\infty \leq 2$. $\mathbb{E}[Wf(W)]$ can be expressed as follows:

$$\begin{aligned}\mathbb{E}[Wf(W)] = \mathbb{E}\Big[\frac{X-\mu}{\sigma}f(\frac{X-\mu}{\sigma})\Big] &= \mathbb{E}\Big[\frac{X}{\sigma}f(\frac{X-\mu}{\sigma}) - \frac{\mu}{\sigma}f(\frac{X-\mu}{\sigma})\Big] \\ &= \mathbb{E}\Big[\frac{\mu}{\sigma}\big(f(\frac{X^s-\mu}{\sigma}) - f(\frac{X-\mu}{\sigma})\big)\Big],\end{aligned}$$

where the last equality is from the definition of size-bias distribution.

Then a Taylor expansion yields

$$\mathbb{E}[Wf(W)] = \frac{\mu}{\sigma}\mathbb{E}\Big[\frac{X^s-\mu}{\sigma}f'\Big(\frac{X-\mu}{\sigma}\Big) + \frac{(X^s-X)^2}{2\sigma^2}f''\Big(\frac{X^*-\mu}{\sigma}\Big)\Big],$$

for some X^* lies between X and X^s. Then using definition of W in terms of X we obtain that

$$\begin{aligned}|\mathbb{E}[f'(W) - Wf(W)]| &= \Big|\mathbb{E}[f'(W) - \frac{\mu}{\sigma^2}(X^s-X)f'(W)] - \mathbb{E}[\frac{\mu}{2\sigma^3}(X^s-X)^2 f''(\frac{X^*-\mu}{\sigma})]\Big| \\ &\leq \Big|\mathbb{E}[f'(W)(1-\frac{\mu}{\sigma^2}(X^s-X))]\Big| + \frac{\mu}{2\sigma^3}\Big|\mathbb{E}[(X^s-X)^2 f''(\frac{X^*-\mu}{\sigma})]\Big|.\end{aligned}$$

Since $|f'|_\infty \leq \sqrt{\frac{2}{\pi}}$, and by taking conditional expectation, we have

$$\begin{aligned}\Big|\mathbb{E}[f'(W)(1-\frac{\mu}{\sigma^2}(X^s-X))]\Big| \leq \sqrt{\frac{2}{\pi}}\Big|\mathbb{E}\big(1-\frac{\mu}{\sigma^2}(X^s-X)\big)\Big| &= \sqrt{\frac{2}{\pi}}\Big|\mathbb{E}[\mathbb{E}(1-\frac{\mu}{\sigma^2}(X^s-X)|X)]\Big| \\ &\leq \sqrt{\frac{2}{\pi}}\mathbb{E}\Big|1-\frac{\mu}{\sigma^2}\mathbb{E}(X^s-X|X)\Big|,\end{aligned}$$

Note that $\sigma^2+\mu^2 = \mathbf{E}(X^2) = \mathbf{E}(X\cdot X) = \mu\mathbf{E}(X^s)$, which implies that $\mathbf{E}(X^s-X) = \frac{\sigma^2}{\mu}$. Then we get

$$\begin{aligned}\sqrt{\frac{2}{\pi}}\mathbb{E}\Big|1-\frac{\mu}{\sigma^2}\mathbb{E}(X^s-X|X)\Big| &= \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\mathbb{E}\Big|\frac{\sigma^2}{\mu} - \mathbb{E}(X^s-X|X)\Big| \\ &= \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\mathbb{E}\Big|\mathbb{E}[\mathbb{E}(X^s-X|X)] - \mathbb{E}(X^s-X|X)\Big|.\end{aligned}$$

Now by Cauchy-Schwarz inequality ($|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$, which implies $\mathbb{E}^2(X) \leq \mathbb{E}(X^2)$), we have

$$\begin{aligned}\frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\mathbb{E}\Big|\mathbb{E}[\mathbb{E}(X^s-X|X)] - \mathbb{E}(X^s-X|X)\Big| &\leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}\Big[\mathbb{E}\big(\mathbb{E}(X^s-X|X)\big) - \mathbb{E}(X^s-X|X)\Big]^2} \\ &= \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}\big(\mathbb{E}(X^s-X|X)\big)}\end{aligned}$$

and since $|f''| \leq 2$, we can get

$$\frac{\mu}{2\sigma^3}\Big|\mathbb{E}[(X^s-X)^2 f''(\frac{X^*-\mu}{\sigma})]\Big| \leq \frac{\mu}{2\sigma^3}\cdot 2\mathbb{E}|(X^s-X)^2| = \frac{\mu}{\sigma^3}\mathbb{E}[(X^s-X)^2].$$

Combining the above results, we finally bound the Wasserstein distance by

$$\mathrm{Wass}(X,Z) \leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}\big(\mathbb{E}(X^s-X|X)\big)} + \frac{\mu}{\sigma^3}\mathbb{E}[(X^s-X)^2].$$

This finishes the proof. □

3.3.2 Construction of size-bias coupling

Considering the fact that there is an explicit form of the upper bound of Wasserstein distance $\mathrm{Wass}(W,Z)$ if we can construct a size-bias coupling, we need some approaches for this construction. In this section, we will introduce three methods about constructing a random variable X^s that has the size-bias distribution with respect to X, given some specific conditions.

Method 1
Let $X=\sum_{i=1}^{n} X_i$, where $X_i \geq 0$ and $\mathbb{E}(X_i)=\mu_i$. The following steps show the construction of size-bias of X.

1. For each $i=1,2,\cdots,n$, let X_i^s has the size-bias distribution of X_i and independent of X_j, X_j^s, $j \neq i$. Given $X_i^s=x$, define the vector $(X_j^{(i)})_{j=i}$ to have the same distribution of $(X_j)_{j\neq i}$ conditional on $X_i=x$.
2. Choose a random summand X_I , where the index I is chosen proportional to μ_i, and independent of all else. Let $\mathbb{E}(X)=\mu$, $\mathbf{P}(I=i)=\frac{\mu_i}{\mu}$.
3. Define $X^s=\sum_{j\neq I} X_j^I + X_I^s$.

If X^s is constructed by the above 3 steps, then X^s has size-bias distribution of X.

Method 2
Let $X_1,X_2,\cdots,X_n$ be non-negative independent random variables with $\mathbb{E}(X_i)=\mu_i$, and for each $i=1,2,\cdots,n$, let X_i^s have the size-bias distribution of X_i, and X_i^s is independent of X_j and X_j^s, $i\neq j$. If $X=\sum_{i=1}^{n} X_i$, $\mathbb{E}(X)=\mu$, and I is chosen independent from all else with. If $\mathbf{P}(I=i)=\frac{\mu_i}{\mu}$, then $X^s=X-X_I+X_I^s$ has the size-bias distribution of X.

Proof. Because of the independence stated in this method, the conditioning in the construction has no effect.

Method 3
Let $X_1,X_2,\cdots,X_n$ be zero-one random variables and $\mathbf{P}(X_i=1)=p_i$. For each $i=1,2,\cdots,n$, let $(X_j^{(i)})_{j\neq i}$ have the distribution of $(X_j)_{i\neq j}$ conditional on $X_i=1$. If $X=\sum_{i=1}^{n} X_i$, $\mathbb{E}(X)=\mu$, and I is chosen independent from all else with $\mathbf{P}(I=i)=\frac{p_i}{\mu}$, then $X^s=\sum_{j\neq I} X_j^{(I)}+1$ has the size-bias distribution of X.
Proof. Based on the notations in method 3, we have $\mu=\sum_{i=1}^{n} p_i$ and $X^s=\sum_{j\neq I} X_j^{(I)}+1$, $(X_j^{(i)})_{j\neq i}=(X_j)_{j\neq i}|X_i=1$. Then we can easily get

$$\mathbb{E}[f(X^s)]=\sum_{i=1}^{n}\frac{p_i}{\mu}\mathbb{E}[f(1+\sum_{j\neq i} X_j^{(i)})].$$

Additionally, we can also get the following relation by simple calculations:

$$\mathbb{E}[X_i f(\sum_{j=1}^{n} X_j)] = p_i \mathbb{E}[f(1+\sum_{j\neq i} X_j)|X_i = 1] = p_i \mathbb{E}[f(1+\sum_{j\neq i} X_j^{(i)})].$$

Finally, we obtain that

$$\begin{aligned}\mathbb{E}[f(X^s)] &= \sum_{i=1}^{n} \frac{1}{\mu} \mathbb{E}\big[X_i f(\sum_{j=1}^{n} X_j)\big] = \frac{1}{\mu}\mathbb{E}\big[(\sum_{i=1}^{n} X_i) f(\sum_{j=1}^{n} X_j)\big] \\ &= \frac{1}{\mu}\mathbb{E}[X f(X)].\end{aligned}$$

This finishes the proof. □

3.3.3 Application of size-bias coupling

A very important application of size-bias coupling is to prove the concentration of measure for the number of isolated vertives in the Erdös-Rényi random graph. A full version of proof can be found in Ghosh's paper [4].

From the mehtods introduced above, we can see that a necessary component is that W is nonnegative. However, this is not quite natural, especially when the distribution of W is symmetric around zero, since W should be closer to a standard Gaussian random variable Z. This shortcoming motivates the introduction of zero-bias coupling.

3.4 Zero-bias coupling

3.4.1 Basic idea about zero-bias coupling

Definition 6 *For a random variable W with $\mathbb{E}(W) = 0$ and $\mathrm{Var}(W) = \sigma^2 < \infty$, we say that random variable W^z has the zero-bias distribution with respect to W if for all absolutely continuous f such that $\mathbb{E}[Wf(W)] < \infty$, we have*

$$\mathbb{E}[Wf(W)] = \sigma^2 \mathbb{E}[f'(W^z)].$$

Theorem 3 *If W is a random variable with $\mathbb{E}(W) = 0$ and $\mathrm{Var}(W) = 1$, and let W^z defined on the same space as W and have the zero-bias distribution with respect to W. If $Z \sim N(0,1)$, then*

$$\mathrm{Wass}(W,Z) \leq 2\mathbb{E}|W^z - W|.$$

Proof. Take function f satisfying $|f|_\infty \leq 1$, $|f'|_\infty \leq \sqrt{\frac{2}{\pi}}$, and $|f''|_\infty \leq 2$. Obviously,

$$\mathrm{Wass}(W,Z) \leq \sup\big|\mathbb{E}[f'(W) - Wf(W)]\big| = \sup\big|\mathbb{E}[f'(W) - f'(W^z)]\big|.$$

Then by Taylor expansion, we have

$$\text{Wass}(W,Z) \leq |f''|_\infty \mathbb{E}|W - W^z| = 2\mathbb{E}|W^z - W|.$$

□

Proposition 1 *Let W be a random variable with $\mathbb{E}(W) = 0$, and* $\text{Var}(W) = \sigma^2 < \infty$.

1. *There is a unique probability distribution for W^z satisfying*

$$\mathbb{E}[Wf(W)] = \sigma^2 \mathbb{E}[f'(W^z)],$$

for absolutely continuous f such that $\mathbb{E}[Wf(W)] < \infty$.

2. *The distribution of W^z is absolutely continuous with respect to Lebesgue measure with density*

$$f_z(w) = \frac{1}{\sigma^2}\mathbb{E}[W \cdot 1_{(W>w)}] = -\frac{1}{\sigma^2}\mathbb{E}[W \cdot 1_{(W\leq w)}].$$

Proposition 2 *If W is a random variable with $\mathbb{E}(W) = 0$, and* $\text{Var}(W) = \sigma^2 < \infty$, *then $(aW)^z$ has the same distribution as aW^z.*

3.4.2 Construction of zero-bias coupling

In general, the construction of zero-bias coupling is difficult to achieve. So we consider a simpler and special case that in which the random variable W is the sum of independent random variables.

Theorem 4 *Let $X_1, X_2, \cdots, X_n$ be independent random variables with $\mathbb{E}(X_i) = 0$,* $\text{Var}(X_i) = \sigma_i^2$, *and $\sum_{i=1}^n \sigma_i^2 = 1$. Define $W = \sum_{i=1}^n X_i$, then we have the following method to construct zero-bias W^z for W.*

1. *For each $i = 1, 2, \cdots, n$, let X_i^z have the zero-bias distribution of X_i, independent of X_j and X_j^z, $j \neq i$.*

2. *Choose a random summand X_I, where the index I satisfies $\boldsymbol{P}(I = i) = \sigma_i^2$, and is independent from others.*

3. *Define $W^z = \sum_{j\neq I} X_j + X_I^z = W - X_I + X_I^z$.*

Then W^z has the zero-bias distribution of W.

Proof. Note that $\mathbb{E}[X_i f(X_i)] = \sigma_i^2 \mathbb{E}[f(X_i^z)]$ and $W - X_i$ is independent of X_i. Then we have

$$\begin{aligned}
\mathbb{E}[Wf(W)] &= \sum_{i=1}^n \mathbb{E}[X_i f(W)] = \sum_{i=1}^n \mathbb{E}\big(X_i f(W - X_i + X_i)\big) \\
&= \sum_{i=1}^n \sigma_i^2 \mathbb{E}[f'(W - X_i + X_i^z)].
\end{aligned}$$

Additionally, since $\mathbb{E}[f'(W - X_I + X_I^z)] = \sum_{i=1}^n \sigma_i^2 f'(W - X_i + X_i^z)$, and then we obtain that

$$\mathbb{E}[Wf(W)] = \mathbb{E}[f'(W - X_I + X_I^z)] = \mathbb{E}[f'(W^z)] \tag{27}$$

Therefore, we can conclude that W^z has the zero-bias distribution of W by definition. □

3.4.3 Application of zero-bias coupling

Zero-bias coupling are generally applied to prove Lindeberg-Feller Central Limit Theorem.

First, let's introduce the triangular array of random variables.

$$\begin{array}{llllll} X_{11} & X_{12} & X_{13} & \cdots X_{1n_1} & & \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n_2} & \\ X_{31} & X_{32} & X_{33} & \cdots & & X_{3n_3} \\ \cdots & & & & & \\ X_{n1} & X_{n2} & X_{n3} & \cdots & & X_{nn_n} \\ \cdots, & & & & & \end{array}$$

where X_{ij}'s are random variables and satisfy the following properties

1. For each i, the n_i random variables $X_{i1}, \cdots, X_{in_i}$ in the i-th row are mutually independent.
2. $\mathbb{E}(X_{ij}) = 0$, for all i, j.
3. $\sum_j \mathbb{E}(X_{ij}^2) = 1$, for all i.

Let $(X_{i,n})_{1 \le i \le n, n \ge 1}$ be the triangular array of random variables defined as above such that $\mathrm{Var}(X_{i,n}) = \sigma_{i,n}^2 < \infty$. Let $W_n = \sum_{i=1}^n X_{i,n}$, and then $\mathrm{Var}(W_n) = 1$. A sufficient condition for W_n to satisfy central limit theorem is the following *Lindeberg* condition:

$$\sum_{i=1}^n \mathbb{E}\left[X_{i,n}^2 1_{(|X_{i,n}| > \varepsilon)}\right] \to 0, \text{ as } n \to \infty,$$

for all $\varepsilon > 0$.

Theorem 5 *Let $(X_{i,n})_{1 \le i \le n, n \ge 1}$ be the triangular array of random variables defined as above and let $X_{I_n,n}$ be a random variable independent of $X_{i,n}$, with $\boldsymbol{P}(I_n = i) = \sigma_{i,n}^2$. For each i, let $X_{i,n}^z$ have the zero-bias dsitribution of $X_{i,n}$ independent of all else. Then Lindeberg condition*

$$\sum_{i=1}^n \mathbb{E}\left[X_{i,n}^2 1_{(|X_{i,n}| > \varepsilon)}\right] \to 0, \text{ as } n \to \infty,$$

holds for all $\varepsilon > 0$ if and only if

$$X_{I_n,n}^z \xrightarrow{\boldsymbol{p}} 0, \text{ as } n \to \infty.$$

Proof. Let $f'(x) = 1_{(|x|\geq \varepsilon)}$ and $f(0) \to 0$, for some fixed $\varepsilon > 0$. Note that

$$f'(x) = \begin{cases} 1 & |x| \geq \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

then we have $f(x) = |x| \cdot 1_{(|x|\geq\varepsilon)} - \varepsilon$ and $xf(x) = \big(x^2 - \varepsilon|x|\big)1_{(|x|\geq\varepsilon)}$. And $\mathbf{P}(|X^z_{I_n,n}| \geq \varepsilon)$ can be bounded as follows

$$\begin{aligned}
\mathbf{P}(|X^z_{I_n,n}| \geq \varepsilon) &= \sum_{i=1}^{n} \sigma^2_{i,n}\mathbf{P}(|X^z_{i,n}| \geq \varepsilon) \\
&= \sum_{i=1}^{n} \sigma^2_{i,n}\mathbb{E}\big[1_{(|X^z_{i,n}|\geq\varepsilon)}\big] = \sum_{i=1}^{n} \sigma^2_{i,n}\mathbb{E}[f'(X^z_{i,n})] \\
&= \sum_{i=1}^{n} \mathbb{E}\big[X^z_{i,n} f(X^z_{i,n})\big] \\
&= \sum_{i=1}^{n} \mathbb{E}\big[\big((X^z_{i,n})^2 - \varepsilon|X^z_{i,n}|\big)1_{(|X^z_{i,n}|\geq\varepsilon)}\big] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\big[(X^z_{i,n})^2 1_{(|X^z_{i,n}|\geq\varepsilon)}\big],
\end{aligned}$$

which indicates that $0 \leq \mathbf{P}(|X^z_{I_n,n}| \geq \varepsilon) \leq \sum_{i=1}^{n} \mathbb{E}\big[(X^z_{i,n})^2 1_{(|X^z_{i,n}|\geq\varepsilon)}\big]$. Therefore, $X^z_{I_n,n} \xrightarrow{\mathbf{p}} 0$, as $n \to \infty$, which is equivalent to *Lindeberg* condition. □

Theorem 6 *If $X^z_{I_n,n} \xrightarrow{p} 0$, as $n \to \infty$, then W_n satisfies CLT.*

Proof. To prove that W_n satisfies a CLT, we need to show that $\text{Wass}(W_n, Z) \to 0$. Based on the proof of **Theorem 3**, we have already known that $\text{Wass}(W_n, Z) \leq \sup\big|\mathbb{E}[f'(W_n) - f'(W^z_n)]\big|$. Additionally, note that $\mathbb{E}(x) = \int_0^\infty \mathbf{P}(X \geq t)dt$ [9], then we have

$$\begin{aligned}
\big|\mathbb{E}[f'(W_n) - f'(W^z_n)]\big| \leq \mathbb{E}\big|f'(W_n) - f'(W^z_n)\big| &= \int_0^\infty \mathbf{P}\big(|f'(W_n) - f'(W^z_n)| \geq t\big)dt \\
&\leq \int_0^{2|f'|_\infty} \mathbf{P}\big(|f'(W_n) - f'(W^z_n)| \geq t\big)dt \;^{10} \\
&\leq \int_0^{2|f'|_\infty} \mathbf{P}\big(|f''|_\infty|W_n - W^z_n| \geq t\big)dt = \int_0^{2|f'|_\infty} \mathbf{P}\Big(|W_n - W^z_n| \geq \frac{t}{|f''|_\infty}\Big)dt,
\end{aligned}$$

where the last inequality is from Taylor expansion.

Next, we will show $|W_n - W^z_n| \xrightarrow{\mathbf{p}} 0$, since if $|W_n - W^z_n| \xrightarrow{\mathbf{p}} 0$, then $\mathbf{P}\big(|W_n - W^z_n| \geq \frac{t}{|f''|_\infty}\big) \to 0$ by definition of convergence in probability.

Note that $|W_n - W^z_n| = |X^z_{I_n,n} - X_{I_n,n}|$ and $X^z_{I_n,n} \xrightarrow{\mathbf{p}} 0$, hence in order to prove $|W_n - W^z_n| \xrightarrow{\mathbf{p}} 0$, it's

[9] This is because $X = \int_0^\infty 1_{(x\geq t)}dt$, thus, $\mathbb{E}(x) = \int_0^\infty \mathbf{P}(X \geq t) \cdot 1dt$.

[10] $t \leq |f'(W_n) - f'(W^z_n)| \leq 2|f'|_\infty$

enough to show $X_{I_n,n} \xrightarrow{\mathbf{p}} 0$. Let $m_n = \max_i\{\sigma_{i,n}^2\}$, then by Chebyshev's inequality, we have

$$\begin{aligned}\mathbf{P}\big(|X_{I_n,n}| \geq \varepsilon\big) &\leq \frac{\mathrm{Var}(X_{I_n,n})}{\varepsilon^2} = \frac{\sum_{i=1}^n \sigma_{i,n}^2 \cdot \mathrm{Var}(X_{i_n,n})}{\varepsilon^2} = \frac{\sum_{i=1}^n \sigma_{i,n}^4}{\varepsilon^2} \\ &\leq \frac{m_n}{\varepsilon^2}\sum_{i=1}^n \sigma_{i,n}^2 = \frac{m_n}{\varepsilon^2},\end{aligned}$$

where the last equality comes from the fact that $\mathrm{Var}(W_n) = 1$.

In the final step, what left to show is $m_n \to 0$. $\forall \delta > 0$, we have

$$\sigma_{i,n}^2 = \mathbb{E}(X_{i,n}^2) = \mathbb{E}\big[X_{i,n}^2 1_{(|X_{i,n}| \leq \delta)}\big] + \mathbb{E}\big[X_{i,n}^2 1_{(|X_{i,n}| > \delta)}\big].$$

And since $\mathbb{E}\big[X_{i,n}^2 1_{(|X_{i,n}| \leq \delta)}\big] = X_{i,n}^2 \mathbf{P}(|X_{i,n}| \leq \delta) \leq \delta^2$, we can get

$$\sigma_{i,n}^2 \leq \delta^2 + \mathbb{E}\big[X_{i,n}^2 1_{(|X_{i,n}| > \delta)}\big].$$

According to **Theorem 5**, $X_{I_n,n}^z \xrightarrow{\mathbf{p}} 0$, as $n \to \infty$ is equivalent to *Lindeberg* condition, thus,

$$\mathbb{E}[X_{i,n}^2 1_{(|X_{i,n}| > \delta)}] \to 0, \text{as } n \to \infty.$$

Then we can conclude that $\forall \delta > 0$, $0 \leq m_n \leq \delta^2$, as $n \to \infty$. Therefore, $m_n \to 0$, as $n \to \infty$. This finishes the proof. □

4 References

[1] Louis H.Y. Chen, Larry Goldstein and Qi-Man Shao: *Normal Approximation by Stein's Method* (2011).

[2] David Rosenberg: *Stein's method and applications* (2007).

[3] Nathan Ross: *Fundamentals of Stein's method* (2011).

[4] Subhankar Ghosh and Larry Goldstein: *Concentration of measure for the number of isolated vertices in the Erdös-Rényi random graph by size bias couplings*.