

STEIN'S THEOREM, INTEGRATION BY PARTS AND THEIR APPLICATIONS

Rafael Aznar

Department of Mathematics, University of Toronto

Email: rafael.aznar@mail.utotonto.ca.

December 7, 2018

Contents

1	FUNDAMENTALS OF STEIN'S METHOD	3
1.1	Overview	3
1.2	Probability metrics	4
1.3	Integration by Parts	6
2	STEIN'S METHOD	10
2.1	Introduction to Stein's Method	10
2.2	Solution to Stein's equation	11
2.3	Boundary conditions for the solution of Stein's equation	14
2.4	Example-Ordinary Central Limit Theorem in the Wasserstein metric	22
3	APPLICATION OF STEIN'S METHOD	23
3.1	Dependency Graph	23
3.1.1	Application of Dependency Graph	25
3.2	Method of Exchangeable Pairs	27
3.2.1	Construction of an Exchangeable Pair	29
3.2.2	Application of exchangeable pairs in proving Central Limit Theorem	30
4	REFERENCES	34

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

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

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

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 Wasserstein 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 also study two applications of Stein's method: dependency graphs and exchangeable pairs. These applications provide us with bounds for the Wasserstein distance between our variables, which in turn allow us to approximate W and Z .

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 theorem 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

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

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

One way to do this is to approximate the supremum norm in the difference of the cumulative distribution functions of W, Z . In other words, we wish to bound the quantity

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x)| \quad (1)$$

Recall the fact:

$$\mathbb{P}(W \leq x) = \mathbb{E}\chi_{[\leq x]}(W),$$

where χ_E is the standard indicator function on the set E , namely

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

Stein's method consists of using the *Wasserstein* and *Kolmogorov* probability metrics in order to bound the difference (1). To do this, we introduce the notion of a metric in a probability space. These metrics will give us a notion how close two random variables are in distribution. In particular, we will have a means of establishing when a random variable converges to a Gaussian distribution, and how fast.

To do this, we now introduce the notion of a metric generated by a class of functions. In the next section, we will construct the metrics that we will need to study convergence to Gaussian distribution.

1.2 Probability metrics

Definition 1 For two probability measures μ and ν , the probability metric generated by a class \mathcal{G} of test functions is:

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

Definition 2 Let W, Z be random variables with probability distribution functions $F_W(x)$ and $F_Z(x)$, respectively. We define the probability metric generated by a class \mathcal{G} of test functions, between W and Z as the distance detailed in (2), applied to $F_W(x)$ and $F_Z(x)$:

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

We now define the Kolmogorov and Wasserstein metrics for random variables by using Def. 2 for appropriate choices of \mathcal{G} :

Definition 3 We obtain the Kolmogorov metric by setting our class of functions $\mathcal{G} = \{\chi_{[\leq x]} : x \in \mathbb{R}\}$ so that

$$\text{Kolm}(W, Z) := \sup_{x \in \mathbb{R}} |\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x)|.$$

Notice that the Kolmogorov distance is precisely the quantity which we set out to bound. Next we introduce the Wasserstein metric, which gives us a more straightforward approach to show convergence in distribution.

Definition 4 The Wasserstein metric is obtained by setting \mathcal{G} to be the class of all 1-Lipschitz functions, to get

$$\text{Wass}(W, Z) := \sup_{g \in \mathcal{G}} |\mathbb{E}g(W) - \mathbb{E}g(Z)|.$$

Lemma 1 Let W, Z be random variables, such that Z has a density function bounded by some constant C . Then

$$\text{Kolm}(W, Z) \leq 2\sqrt{C\text{Wass}(W, Z)}$$

Proof. Fix $x \in \mathbb{R}$, $\varepsilon > 0$. Let $g_{x,\varepsilon}^1(t)$ a linear transition for $\chi_{[\leq x]}$ between x and $x + \varepsilon$. Explicitly:

$$g_{x,\varepsilon}^1(t) = \begin{cases} 1 & \text{if } t \leq x, \\ \frac{1}{\varepsilon}(x + \varepsilon - t) & \text{if } x \leq t \leq x + \varepsilon \\ 0 & \text{if } t > x + \varepsilon. \end{cases}$$

Notice that by construction, this function is $\frac{1}{\varepsilon}$ -Lipschitz, and therefore $\varepsilon g_{x,\varepsilon}^1(t)$ is 1-Lipschitz. This means that

$$\text{Wass}(W, Z) \geq |\mathbb{E}(\varepsilon g_{x,\varepsilon}^1(W)) - \mathbb{E}(\varepsilon g_{x,\varepsilon}^1(Z))| = \varepsilon |\mathbb{E}(g_{x,\varepsilon}^1(W)) - \mathbb{E}(g_{x,\varepsilon}^1(Z))|$$

We now consider the difference $\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x)$.

$$\begin{aligned} \mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x) &= \mathbb{P}(W \leq x) - \mathbb{E}[g_{x,\varepsilon}^1(Z)] + \mathbb{E}[g_{x,\varepsilon}^1(Z)] - \mathbb{P}(Z \leq x) \\ &\leq \mathbb{E}(g_{x,\varepsilon}^1(W)) - \mathbb{E}[g_{x,\varepsilon}^1(Z)] + \mathbb{E}[g_{x,\varepsilon}^1(Z)] - \mathbb{P}(Z \leq x) \\ &\leq \frac{1}{\varepsilon} \text{Wass}(W, Z) + \int [g_{x,\varepsilon}^1(t) - \chi_{[\leq x]}(t)] dF_Z(t) \end{aligned}$$

And now notice that $g_{x,\varepsilon}^1$ and $\chi_{[\leq x]}(t)$ agree outside of the interval $[x, x + \varepsilon]$. So to estimate this last integral, we need only consider it within this bound. This together with the bound on $F_Z(t)$ gives us:

$$\begin{aligned} \int [g_{x,\varepsilon}^1(t) - \chi_{[\leq x]}(t)] dF_Z(t) &\leq \int_x^{x+\varepsilon} C \frac{1}{\varepsilon} (x + \varepsilon - t) dt \\ &= -\frac{1}{2\varepsilon} C (x + \varepsilon - t)^2 \Big|_x^{x+\varepsilon} = \frac{C\varepsilon}{2} < C\varepsilon \end{aligned}$$

Putting it all together we get that

$$\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x) \leq \frac{1}{\varepsilon} \text{Wass}(W, Z) + C\varepsilon$$

Now, considering the expression above as a function of ε , it is a straightforward computation that the minimizer of this expression is $\varepsilon = \sqrt{C^{-1} \text{Wass}(W, Z)}$. Plugging in we have:

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

Next, fix $\varepsilon > 0$ once more, and define $g_{x,\varepsilon}^2(w)$ analogously to $g_{x,\varepsilon}^1(w)$, but this time, performing our linear transition on the interval $[x - \varepsilon, x]$. Performing the same calculation and minimizing with respect to ε once more gives us the same upper bound for $\mathbf{P}(Z \leq x) - \mathbf{P}(W \leq x)$.

Finally, taking a supremum on the previous equation we obtain:

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

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

Lemma 1 gives us a sufficient condition for convergence which is remarkably easier to work with than the Kolmogorov metric, which is our first and most natural notion of convergence in probability. For the remainder of this paper, the Wasserstein metric will be our principal method for proving convergence to Gaussian distribution.

1.3 Integration by Parts

Now that we have a relation between the Kolmogorov distance and the Wasserstein distance given by **Lemma 1**, we may shift our attention to studying the expectation of suitable classes of functions to obtain a bound for Wasserstein metrics. With this in mind, we show an important result for normal distributions.

Theorem 1 *Integration by Parts:* If $W \sim N(0, 1)$, then for all absolutely continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ with $\mathbb{E}|f'(W)| < \infty$, we have

$$\mathbb{E}f'(W) = \mathbb{E}Wf(W) \quad (5)$$

Conversely, if $\mathbb{E}[f'(W)] = \mathbb{E}[Wf(W)]$ for all bounded, continuous and piece-wise continuously differentiable functions f with $\mathbb{E}|f'(W)| < \infty$, then $W \sim N(0, 1)$.

We will need an intermediate result before we prove **Theorem 1**. If we wish to say that $W \sim N(0, 1)$, then it will suffice show that $\mathbb{P}(W \leq x) = \mathbb{P}(Z \leq x)$ for all $x \in \mathbb{R}$. Therefore, it is natural to look for a function f such that $\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x) = \mathbb{E}[f'(W) - Wf(W)]$. Because of this we will show the following lemma.

Lemma 2 Let Z be a standard Gaussian random variable with cumulative distribution function $\Phi(x) = \mathbb{P}(Z \leq x)$, and let $z \in \mathbb{R}$. Then the differential equation

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

has a unique bounded solution given by

$$f(w) = \begin{cases} \sqrt{2\pi}e^{\frac{w^2}{2}}\Phi(w)(1 - \Phi(z)), & w \leq z \\ \sqrt{2\pi}e^{\frac{w^2}{2}}\Phi(z)(1 - \Phi(w)), & w > z \end{cases} \quad (6)$$

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

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

and notice that

$$\begin{aligned} e^{-\frac{w^2}{2}}[f'(w) - wf(w)] &= (f(w)e^{-\frac{w^2}{2}})' \\ \implies (f(w)e^{-\frac{w^2}{2}})' &= e^{-\frac{w^2}{2}}[\chi_{(w \leq z)} - \Phi(z)]. \end{aligned}$$

And now we may integrate with respect to w on the interval $[-\infty, w]$, which gives us a formula for the general solution:

$$f(w)e^{-\frac{w^2}{2}} + C = \int_{-\infty}^w [\chi_{(x \leq z)} - \Phi(z)]e^{-\frac{x^2}{2}} dx, \quad (7)$$

Next, we use the fact that Z is Gaussian to obtain that for all $w \geq z$:

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

In particular, using (7) and letting $w \rightarrow \infty$:

$$\lim_{w \rightarrow \infty} f(w) e^{-\frac{w^2}{2}} + C = \lim_{w \rightarrow \infty} \sqrt{2\pi} \Phi(z) [1 - \Phi(w)] = 0$$

So that only $C = 0$ gives us a bounded solution f . This solution will have the form

$$f(w) = e^{-\frac{w^2}{2}} \int_{-\infty}^w [\chi_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx$$

We now claim that this solution is equivalent to (6). Indeed, we know that

$$\begin{aligned}
&\int_{-\infty}^{\infty} [\chi_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx = 0 \\
\implies \int_{-\infty}^w [\chi_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx &= - \int_w^{\infty} [\chi_{(x \leq z)} - \Phi(z)] e^{-\frac{x^2}{2}} dx
\end{aligned}$$

Therefore, we consider what happens when $w \leq z$ and when $z > w$:

Case 1: if $w \leq z$

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

Case 2: if $w > z$

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

□

We are now ready to prove Theorem 1.

Proof of Theorem 1. “ \implies ” Suppose that $Z \sim N(0, 1)$. First, write $\mathbb{E}f'(Z)$ as an integral:

$$\begin{aligned}
\mathbb{E}f'(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(z) e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(z) e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f'(z) e^{-\frac{z^2}{2}} dz
\end{aligned}$$

Next, we may integrate by parts to find that: $e^{-\frac{z^2}{2}} = \int_{-\infty}^z (-x) e^{-\frac{x^2}{2}} dx = \int_z^\infty x e^{-\frac{x^2}{2}} dx$

So substituting we obtain

$$\mathbb{E}f'(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(z) \int_{-\infty}^z (-x) e^{-\frac{x^2}{2}} dx dz + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f'(z) \int_z^{+\infty} x e^{-\frac{x^2}{2}} dx dz$$

And using Fubini's theorem:

$$\begin{aligned}
\mathbb{E}f'(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \left(\int_x^0 f'(z) dz \right) (-x e^{-\frac{x^2}{2}}) dx + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \left(\int_0^x f'(z) dz \right) x e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (f(x) - f(0)) x e^{-\frac{x^2}{2}} dx \\
&= \mathbb{E}Zf(Z) - \frac{1}{\sqrt{2\pi}} f(0) \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx \\
&= \mathbb{E}Zf(Z),
\end{aligned}$$

Where we solve this last integral by noticing that $x e^{-\frac{x^2}{2}}$ is an odd function.

“ \Leftarrow ” Suppose that $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$ for all bounded, continuous and piece-wise continuously differentiable functions f with $\mathbb{E}|f'(Z)| < \infty$.

Now for $x \in \mathbb{R}$, take f to be the unique bounded solution of the differential solution in **Lemma 2**.

By the formula (6), we have that f is piecewise continuously differentiable, so we have

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

Therefore, $W \sim N(0, 1)$. This finishes the proof. \square

2 Stein's method

2.1 Introduction to Stein's Method

Equipped with **Theorem 1**, we will now construct a new method of bounding how far (in the Wasserstein sense) a random variable is from Gaussian distribution.

We will find a class of functions $f \in \mathcal{F}$ such that

$$\text{Wass}(W, Z) = \sup_{g \in \mathcal{G}} |\mathbb{E}g(W) - \mathbb{E}g(Z)| \leq \sup_{f \in \mathcal{F}} |\mathbb{E}[f'(W) - Wf(W)]|, \quad (8)$$

then we can bound the Wasserstein distance by bounding $|\mathbb{E}[f'(W) - Wf(W)]|$. In order to do this, for all $g \in \mathcal{G}$, we would like \mathcal{F} to contain a function f such that

$$f'(w) - wf(w) = g(w) - \mathbb{E}g(Z) \quad (9)$$

(as usual, Z a random variable s.t. $Z \sim N(0, 1)$) This is known as Stein's equation.

If this condition holds, then taking expectations and then evaluating the supremum across \mathcal{F} will give us the inequality (8).

$$\mathbb{E}[f'(W) - Wf(W)] = \mathbb{E}g(W) - \mathbb{E}g(Z). \quad (10)$$

Noticeably, the solution to Stein's equation has several boundary conditions, detailed as follows.

Lemma 3 *Given a function $g : \mathbb{R} \rightarrow \mathbb{R}$ that is bounded, then there exists absolutely continuous solution f to Stein's equation, satisfying*

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

And if g is Lipschitz, then

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

The proof this Lemma is long, and will be solved in section 2.3.

Particularly, if \mathcal{F} is defined as a family of functions satisfying

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

then we have that \mathcal{F} contains all solutions to Stein's equation with g 1-Lipschitz. Therefore,

$$\text{Wass}(W, Z) = \sup_{g \in \mathcal{G}} |\mathbb{E}g(W) - \mathbb{E}g(Z)| \leq \sup_{f \in \mathcal{F}} |\mathbb{E}[f'(W) - Wf(W)]|. \quad (11)$$

From **Lemma 1** we know 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 $|\mathbf{P}(W \leq x) - \mathbf{P}(Z \leq x)| \rightarrow 0$. Finally, by definition of convergence in distribution, we get

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

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

In this section, we will study the solution of Stein's equation

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

for both general g and special g .

2.2 Solution to Stein's equation

Lemma 4 *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

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

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

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

Integrating on w gives us

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

In order to get a unique and bounded solution $f(w)$, we take limits as $w \rightarrow \infty$, and notice that to achieve boundedness, we must have:

$$\begin{aligned}
-C &= \lim_{w \rightarrow \infty} \int_{-\infty}^w e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx \\
&= \lim_{w \rightarrow \infty} \left[\int_{-\infty}^w e^{\frac{-x^2}{2}} g(x) dx - \int_{-\infty}^w e^{\frac{-x^2}{2}} \mathbb{E}g(Z) dx \right] \\
&= \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} g(x) dx - \mathbb{E}g(Z) \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx \\
&= \sqrt{2\pi} \mathbb{E}g(Z) - \sqrt{2\pi} \mathbb{E}g(Z) = 0
\end{aligned}$$

Therefore, C must be 0. So the final solution for Stein's equation is

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

Furthermore, a consequence of the fact that

$$\int_{-\infty}^w e^{\frac{-x^2}{2}} [g(x) - \mathbb{E}g(Z)] dx = 0$$

which was just showed, is that we get our desired expression:

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

This finishes the proof. □

Claim: additionally, if g is *Lipschitz*, we have an equivalent form of the solution:

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

This form will be helpful in proving our boundary conditions.

Proof. We need to check that (13) satisfies $f'(w) - wf(w) = g(w) - \mathbb{E}g(Z)$.

First, let's take first derivative of $f(w)$ with respect to w . Since expectation is a linear operator, our derivative will be

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

Next, we would like to use integration by parts on f . In order to do this, we must first check that

for a normal Gaussian random variable Y we have $\mathbb{E}|f'(Y)| < \infty$. In fact,

$$\begin{aligned}
\mathbb{E}f'(Y) &= -\mathbb{E}\left[\int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)|Y] dt\right] \\
&= -\int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E}\left[\mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)|Y]\right] dt \\
&= -\int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)] dt \\
&\leq -\int_0^1 \frac{1}{2\sqrt{1-t}} \inf_{t \in [0,1]} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)] dt
\end{aligned}$$

Notice that this can be done because $\mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)]$ is continuous in t , so it attains a minimum on $[0, 1]$. Next:

$$\begin{aligned}
\mathbb{E}f'(Y) &\leq -\inf_{t \in [0,1]} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)] \int_0^1 \frac{1}{2\sqrt{1-t}} dt \\
&= -\inf_{t \in [0,1]} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)] < \infty
\end{aligned}$$

Similarly, we may take the supremum instead to find the bound

$$\begin{aligned}
-\infty &< -\sup_{t \in [0,1]} \mathbb{E}[Zg'(\sqrt{t}Y + \sqrt{1-t}Z)] \leq \mathbb{E}f'(Y) \\
&\implies \mathbb{E}|f'(Y)| < \infty
\end{aligned}$$

As wanted. Then using integration by parts, $\mathbb{E}Zf(Z) = \mathbb{E}f'(Z)$, 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

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

This finishes the proof. \square

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. To do this, we will employ the following useful inequality.

Lemma 5 (Mill's Ratio Inequality) For all $x \in \mathbb{R}$:

$$\frac{x}{1+x^2} < e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt < \frac{1}{x} \quad (14)$$

Proof: This argument is due to R.D. Gordon ¹. For simplicity, define $R(x) = e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt$. First note that $R(x)$ is positive for all x . Next, using the Fund. Theorem of Calculus and uniform continuity of the normal distribution:

$$\begin{aligned}
R'(x) &= x e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt + e^{\frac{x^2}{2}} \left[e^{-\frac{x^2}{2}} \right] \Big|_x^\infty \\
&= xR(x) - 1
\end{aligned}$$

This means that, if at any point x_0 , $R(x)$ is increasing, then $R'(x) = xR(x) - 1$ will also be increasing. This means that if $R(x)$ is increasing at any point, it will have unbounded growth as $x \rightarrow \infty$.

¹R.D. Gordon, "Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument" *Annals of Math. Stat.*, Vol. 12 (1941), pp. 364-366

However,

$$\begin{aligned}\lim_{x \rightarrow \infty} xR(x) &= \lim_{x \rightarrow \infty} x e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt \\ &\leq e^{\frac{x^2}{2}} \int_x^\infty t e^{-\frac{t^2}{2}} dt = e^{\frac{x^2}{2}} \left[-e^{-\frac{t^2}{2}} \right]_x^\infty = 1\end{aligned}$$

Therefore, R cannot have unbounded growth, so R is monotonically decreasing. This means that $R'(x) < 0$ for all real x . Therefore,

$$R(x) = e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt \leq \frac{1}{x} \quad \forall x \in \mathbb{R}.$$

To show the other inequality, take the second and third derivatives of R using our formula for the first derivative:

$$\begin{aligned}R''(x) &= xR'(x) + R(x) = xR'(x) + \frac{1}{x}(R'(x) + 1) \\ &= \frac{1+x^2}{x}R'(x) + \frac{1}{x}, \\ R'''(x) &= \left(1 + \frac{2}{1+x^2}\right)xR''(x) - \frac{2}{1+x^2}\end{aligned}$$

This means that $R''(x)$ is decreasing whenever it is negative. Therefore, if at any point x_0 we have $R''(x_0) < 0$, then we get that R'' is monotonically decreasing on $[x_0, \infty]$, and furthermore, $R'''(x) \leq xR''(x)$ for $x > x_0$, so that $\lim_{x \rightarrow \infty} R''(x) = -\infty$. However, this contradicts the positivity of R . Therefore, we have $R''(x) > 0$ for all x .

From this we may conclude that

$$\begin{aligned}0 < R''(x) &= \frac{1+x^2}{x}R'(x) + \frac{1}{x} \\ &= \frac{1+x^2}{x}(xR(x) - 1) + \frac{1}{x} \\ &= (1+x^2)R(x) - x \implies \frac{x}{1+x^2} < R(x),\end{aligned}$$

which is precisely our inequality. □

Now we proceed to prove the boundary conditions detailed in the previous section.

Lemma 6 *If $g : \mathbb{R} \rightarrow \mathbb{R}$ is bounded, then the bounded solution f to Stein's equation satisfies*

$$(i) \quad |f|_\infty \leq \sqrt{\frac{\pi}{2}} \left| g(w) - \mathbb{E}g(Z) \right|_\infty$$

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

Proof. (i) We consider the cases $w > 0$ and $w \leq 0$ separately.

Case 1 $w > 0$

Recall the solution of Stein's equation for general g

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

Then the supremum norm of f satisfies

$$\begin{aligned} \sup_{w \in [0, \infty]} |f| &\leq |g(w) - \mathbb{E}g(Z)|_{\infty} \cdot e^{\frac{w^2}{2}} \int_w^{+\infty} e^{-\frac{x^2}{2}} dx \\ &= |g(w) - \mathbb{E}g(Z)|_{\infty} \cdot R(w) \end{aligned}$$

And we have shown that $R(x)$ is monotonically decreasing on $[0, \infty]$, so

$$\begin{aligned} \sup_{w \in [0, \infty]} |f| &\leq |g(w) - \mathbb{E}g(Z)|_{\infty} R(0) \\ &= \sqrt{\frac{\pi}{2}} |g(w) - \mathbb{E}g(Z)|_{\infty} \end{aligned}$$

when $w > 0$.

b. $w \leq 0$

In this case, we use another form of the solution

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

Next, using the fact that the Gaussian distribution is an even function, we may apply the same method as above to obtain the same bound for $\sup_{w \in [-\infty, 0]} |f|$.

Therefore,

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

(ii): Similarly to the previous part, we use the fact that the Gaussian distribution function is even. Therefore, we need only evaluate the region where $w > 0$.

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

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

By Mill's ratio inequality $R(x) < \frac{1}{x}$, we conclude

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

Therefore,

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

□

Lemma 7 *If g is Lipschitz, but not necessarily bounded, then the bounded solution f to Stein's equation satisfies*

$$(i) \quad |f|_\infty \leq |g'|_\infty$$

$$(ii) \quad |f'|_\infty \leq \sqrt{\frac{\pi}{2}} |g'|_\infty$$

$$(iii) \quad |f''|_\infty \leq 2|g'|_\infty.$$

Proof. (i): If g is Lipschitz, then we have another form of solution that has been proved in the previous subsection

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

Define the function h by

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

and using integration by parts we know that for the absolutely continuous function h

$$\mathbb{E}Zh(Z) = \mathbb{E}h'(Z),$$

then we can obtain that

$$\mathbb{E}[Zg(\sqrt{t}w + \sqrt{1-t}Z)] = \sqrt{1-t} \mathbb{E}[g'(\sqrt{t}w + \sqrt{1-t}Z)]. \quad (15)$$

Using this equality in our exppression for f :

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

and thus

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

(ii): Once again 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

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

(iii): This time we will use the standard form of the solution f . By differentiating both sides of Stein's equation with respect to w , we get

$$\begin{aligned} f''(w) &= g'(w) + f(w) + wf'(w) \\ &= g'(w) + f(w) + w[wf(w) + g(w) - \mathbb{E}g(Z)] \\ &= g'(w) + (1+w^2)f(w) + (g(w) - \mathbb{E}g(Z)) \end{aligned}$$

First we write an equivalent expression to $g(w) - \mathbb{E}g(Z)$. To do this we write it as an integral:

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

Next, use Fubini's theorem to change the order of integration

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

Next we rewrite $f(w)$ by using $\Phi(z) = \mathbf{P}(Z \leq z)$ and $\bar{\Phi}(z) = \mathbf{P}(Z > z) = 1 - \Phi(z)$, and the

equality we just obtained as follows:

$$\begin{aligned} f(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^w e^{\frac{-y^2}{2}} [g(y) - \mathbb{E}g(z)] dy \\ &= e^{\frac{w^2}{2}} \int_{-\infty}^w e^{\frac{-y^2}{2}} \left(\int_{-\infty}^y g'(z) \Phi(z) dz - \int_y^{\infty} g'(z) \bar{\Phi}(z) dz \right) dy \end{aligned}$$

So once again using Fubini's, and splitting up intervals of integration in our last integral, we have

$$\begin{aligned} f(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^w g'(z) \Phi(z) \int_z^w e^{\frac{-y^2}{2}} dy dz - e^{\frac{w^2}{2}} \left[\int_{-\infty}^w g'(z) \bar{\Phi}(z) \int_{-\infty}^z e^{\frac{-y^2}{2}} dy dz + \right. \\ &\quad \left. \int_w^{+\infty} g'(z) \bar{\Phi}(z) \int_{-\infty}^w e^{\frac{-y^2}{2}} dy dz \right]. \end{aligned} \quad (16)$$

And Since

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

we can rewrite (16) as follows

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

Combining all these calculations, we get

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

Then, taking the supremum of $f''(w)$ on $[, \infty]$:

$$\begin{aligned}
&\sup_{w \in [0,1]} |f''| \leq \\
&|g'|_{\infty} \sup_{w \in [0,1]} \left[1 + |w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\bar{\Phi}(w)|\int_{-\infty}^w \Phi(z)dz + |-w - \sqrt{2\pi}(1+w^2)e^{\frac{w^2}{2}}\Phi(w)|\int_w^{\infty} \bar{\Phi}(z)dz\right].
\end{aligned} \tag{17}$$

Now we wish to compute the sign of the terms in absolute value. To do this, recall that by Mill's ratio inequality, for $x > 0$:

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

but notice that $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1 - \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ so our inequality is equivalent to

$$\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}(1+x^2)} \leq 1 - \Phi(x) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}. \tag{18}$$

Then, using the left-hand side of the inequality we conclude that

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

so the first quantity in absolute values in (17) is nonnegative.

Additionally, if $x > 0$, we get $x + \sqrt{2\pi}(1+x^2)e^{\frac{x^2}{2}}\Phi(x) > 0$, so in (17), the second quantity in absolute values is negative. Moreover, using standard integration by parts, we notice that

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

and

$$\int_w^\infty (1 - \Phi(z))dz = -w(1 - \Phi(w)) + \frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}. \quad (20)$$

Finally, by substituting, we get that (17) equals

$$|g'|_\infty [-w + \sqrt{2\pi}(1 + w^2)e^{\frac{w^2}{2}}\bar{\Phi}(w)] \int_{-\infty}^w \Phi(z)dz + [w + \sqrt{2\pi}(1 + w^2)e^{\frac{w^2}{2}}\Phi(w)] \int_w^\infty \bar{\Phi}(z)dz = |g'|_\infty$$

Finally, just as before, the same bound applies to the on the interval $[-\infty, 0]$ by evenness of the Gaussian distribution function. Therefore,

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

□

Lemma 8 *A less constrained bound: $|f''|_\infty \leq 4|g'|_\infty$, when g is Lipschitz.*

Proof. Recall

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

and

$$f''(w) - wf'(w) = g'(w) + f(w). \quad (21)$$

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

$$f''(w) - wf'(w) = h(w) + \mathbb{E}[h(Z)], \quad (22)$$

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:

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

This finishes the proof.

□

2.4 Example-Ordinary Central Limit Theorem in the Wasserstein metric

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

Proof.

Take any $f \in C^1$ with f' absolutely continuous, and satisfying $|f| \leq 1$, $|f'| \leq \sqrt{\frac{2}{\pi}}$, $|f''| \leq 2$. Let $W_i = W - \frac{X_i}{\sqrt{n}}$, which implies that $W_i \perp X_i$ (where “ \perp ” denotes independence). We will show that

$$|\mathbb{E}[f'(W) - Wf(W)]| \rightarrow 0 \text{ as } n \rightarrow \infty$$

First we split $\mathbb{E}Wf(W)$ into the sum $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}[X_i f(W)]$

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

We will approximate both of these terms separately. Now note that, using mean value theorem, $\forall x < y \in \mathbb{R}$, $\exists \theta_1 \in [x, y]$ such that

$$(x - y)f'(\theta_1) = f(x) - f(y),$$

Then, using MVT once more, $\exists \theta_2 \in [\theta_1, y]$ so that

$$\begin{aligned} (\theta_1 - y)f''(\theta_2) &= f'(\theta_1) - f'(y) \\ \implies |f''(\theta_2)| &\geq \left| \frac{f'(\theta_1) - f'(y)}{x - y} \right| = \left| \frac{f(x) - f(y) - f'(y)(x - y)}{(x - y)^2} \right| \\ \implies |f''|_\infty &\geq \left| \frac{f(x) - f(y) - f'(y)(x - y)}{(x - y)^2} \right| \forall x, y \in \mathbb{R} \end{aligned}$$

and now we may compute:

$$|f(W) - f(W_i) - (W - W_i)f'(W_i)| \leq (W - W_i)^2 |f''|_\infty,$$

thus

$$\mathbb{E}[X_i(f(W) - f(W_i) - (W - W_i)f'(W_i))] \leq |f''|_\infty \cdot \mathbb{E}\left|X_i \frac{X_i^2}{n}\right| \leq \frac{2}{n} \mathbb{E}|X_i|^3.$$

And next,

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

since W_i is independent of X_i . Putting both of our approximations together, we obtain

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

Finally, note that

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

Combining these results and using the triangle inequality, we obtain that

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

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

$$\left| \mathbb{E}[f'(W) - Wf(W)] \right| \leq \frac{4}{n^{\frac{3}{2}}} \sum_{i=1}^n \mathbb{E}|X_i|^3$$

So in particular, by Stein's method,

$$\text{Wass}(W, Z) \leq \frac{4}{n^{\frac{3}{2}}} \sum_{i=1}^n \mathbb{E}|X_i|^3$$

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

□

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 and the method of exchangeable pairs.

3.1 Dependency Graph

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

Definition 5 $\{X_i\}_{i \in V}$ 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.

In particular, any two variables which are not independent must be neighbors in G . Here, by a neighbor of a vertex X , we mean a vertex Y of G such that Y, X are connected by an edge. Similarly, the neighbourhood of a vertex X is the set of all neighbours Y of X .

Notice however, that we do not require the converse of this statement. In other words, if two vertices in G are neighbours, then we may not conclude that they are dependent. We will call this neighbourhood the standard neighbourhood to differentiate it from the following concept.

Definition 6 (Dependency neighborhoods) We say that a collection of random variables $\{X_1, X_2, \dots, X_n\}$ has dependency neighborhoods $N_i \subseteq \{1, \dots, n\}$, $i = 1, \dots, n$, if for all i , X_i is independent of $\{X_j\}_{j \notin N_i}$.

Lemma 9 Given a graph G , let $D = 1 + \text{maximum degree}^2$ of G . Let $W = \sum_{i \in V} X_i$, then

$$\text{Var}(W) \leq D \sum_{i \in V} \text{Var}(X_i). \quad (23)$$

Proof. $\text{Var}(W)$ can be expressed as the following:

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

We denote the neighborhood of i by N_i , and when $j \in N_i$, we say $j \sim i$.

WLOG, we assume that $\mathbb{E}(X_i) = 0$, then

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

This is because whenever $j \not\sim i$, X_i and X_j are independent, so we may express the expectation of their product as the product of their expectations, which are zero.

Now note that $X_i X_j \leq \frac{X_i^2 + X_j^2}{2}$ by the arithmetic mean - geometric mean inequality, thus $\text{Var}(W)$ satisfies the following inequality

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

²In graph theory, the degree of a vertex of a graph is the number of edges incident to the vertex, with loops (edges which have the same vertex as both of its endpoints) counted twice.

This finishes the proof. \square

3.1.1 Application of Dependency Graph

In this section, we will introduce an example about how to use **Lemma 9** to compute the upper bound of Wasserstein distance using Stein's method. 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

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

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

Let's first look at $\mathbb{E}[Wf(W)]$. By assumption, $\mathbb{E}(X_i) = 0$, and thus we have

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

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 may see that

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

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

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

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

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

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

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

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

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

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

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

where the last two inequalities come from the fact that each variable X_i 's expectation is summed

once for each neighbor of X_i , so at most D times. As well as the AM-GM inequality: $\frac{X_1 + \dots + X_n}{n} \geq \sqrt[n]{X_1 X_2 \dots X_n}$.

To get the upper bound of $\text{Var}(T)$, let's first compute the upper bound of $\text{Var}(\sum X_i (\sum_{j \in N_i} X_j))$.

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

Let us also keep track of the fact that

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

Finally, the collection of variables $\{X_i X_j : 1 \leq i, j \leq n\}$ generates a new dependency graph. Now, $X_i X_j$ is independent of $X_k X_l$ if $k, l \notin \{N_i \cup N_j\}$. Then, since $|N_i \cup N_j| \leq 2D$ and since $N_i, N_j \subseteq V$, each vertex in $N_i \cup N_j$ has at most D neighbors. This implies that the new dependency graph has maximum degree at most $2D^2$. Now, using the result in **Lemma 9**, we obtain that

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

Thus,

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

Combining all these results together, we get

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

□

Notably, if D is fixed, i.e., if each variable has at most D other variables which are not independent to it, then W approaches normal Gaussian distribution as $n \rightarrow \infty$.

3.2 Method of Exchangeable Pairs

Definition 7 (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 10 Suppose (W, W') is an exchangeable pair and there is a constant $\lambda \in (0, 1)$ such that

$$\mathbb{E}(W' - W | W) = -\lambda W. \tag{24}$$

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

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

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

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

2. $\mathbb{E}(W' - W)^2 = 2\lambda$, since

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

Next, we will prove **Lemma 10**.

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

$$0 = \mathbb{E}(F(W') - F(W)) = \mathbb{E}[(W' - W)f(W)] + \frac{1}{2}\mathbb{E}[(W' - W)^2 f'(W)] + r, \quad (25)$$

where $r = \frac{1}{6}\mathbb{E}[(W' - W)^3 f''(\tilde{W})]$, \tilde{W} is between W and W' , and thus $|r| \leq \frac{1}{6}|f''|_\infty \cdot \mathbb{E}|W' - W|^3 \leq \frac{1}{3}\mathbb{E}|W' - W|^3$. 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

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

Based on the relationship implied by equation (25), we proceed by saying that

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

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

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

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

$$\mathbb{E}[Y - \mathbb{E}(Y)] = \sqrt{\mathbb{E}^2[Y - \mathbb{E}(Y)]} \leq \sqrt{\mathbb{E}[(Y - \mathbb{E}(Y))^2]} = \sqrt{\text{Var}(Y)},$$

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

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

□

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

The general idea for the construction of an exchangeable pairs to take the sum of a finite collection of random variables. Then, from this sum, we choose one of the summands uniformly at random, and replace it with an independent copy of a variable with identical distribution to the original summand. Our claim is that these two sums will be exchangeable pairs. Formally, let X'_1, X'_2, \dots, X'_n be an independent copy of X_1, X_2, \dots, X_n , and choose index I uniformly at random from $\{1, 2, \dots, n\}$. i.e. $(X'_1, X'_2, \dots, X'_n)$ is independent of (X_1, X_2, \dots, X_n) .

Example

X_1, X_2, \dots, 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 $(W, W') \stackrel{d}{=} (W', W)$, we need to show that for all measurable sets A, B

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

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

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

Where $W'_i = \sum_{j=1}^n X_j - X_i + X'_i$. 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 $(S_{-i}, X_i, X'_i) \stackrel{d}{=} (S_{-i}, X'_i, X_i)$.

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. This is a much stronger result than our traditional central limit theorem, as now we may take random variables with any distribution. In many applications, identical distributions are too much to ask for. However, it is still common to deal with variables which are independent of one another, and hence the value of the following result.

3.2.2.1 Weighted sum of independent random variables

Let X_1, X_2, \dots, 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, \dots, X'_n be the independent copy of X_1, X_2, \dots, X_n , and choose index I uniformly and randomly from $\{1, 2, \dots, n\}$. 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

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

Where our last equality stems from the fact that each X'_i is independent of all X_j , and therefore is independent of $\frac{1}{\sqrt{n}} \sum_i X_i = W$. Hence, in our example, $\lambda = \frac{1}{n}$, and we have by Lemma 10 that

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

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

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

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

$$\mathbb{E}\left(\frac{1}{2\lambda} (W - W')^2 | W\right) = \frac{1}{2} \mathbb{E}[(X'_I - X_I)^2 | W] = \frac{1}{2n} \left[\sum_i^n \mathbb{E}(X'_i - X_i)^2 | W \right],$$

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

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

thus we obtain that

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

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

Remark

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

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

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

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

Hence, we only need to compare $\mathbb{E}\{\mathbb{E}[(W - W')^2|W]\}^2$ and $\mathbb{E}\{\mathbb{E}[(W - W')^2|\mathcal{F}]\}^2$ to get the relationship between $\text{Var}(\mathbb{E}((W - W')^2|W))$ and $\text{Var}(\mathbb{E}((W - W')^2|\mathcal{F}))$. Let $X = (W - W')^2$, then we will compare $\mathbb{E}[\mathbb{E}(X|W)]^2$ and $\mathbb{E}[\mathbb{E}(X|\mathcal{F})]^2$.

Take $W = \sum_{i=1}^n X_i$ and $\mathcal{F} = (X_1, \dots, X_n)$, then we have

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

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

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

Let $Y = \mathbb{E}(X|\mathcal{F})$, $\mathbb{E}^2(Y|W) \leq \mathbb{E}[Y^2|W]$, which comes from Jensen's inequality, and then we continue as follows

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

Therefore,

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

□

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

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

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

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

and $W \overset{\cdot}{\sim} N(0, 1)$.

□

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 Erdős-Rényi random graph by size bias couplings*.
- [5] R.D. Gordon: *Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument*. Annals of Mathematical Statistics, Vol 12 (1941), pp 364-366.