MATH30011 Project (Semester One)

Stein's Method and Its Application to Random Graphs

Author: Wenlin Chen (10407337)
Supervisor: Dr Robert Gaunt

Department of Mathematics, University of Manchester

January 12, 2020

Abstract

Stein's method is a powerful and elegant method which can be used to obtain distributional approximation results and their respective error bounds with respect to integral probability metrics. The method can be easily adapted to many distributions and is applicable to dependence structures. In this thesis, we study the basics of Stein's method and use them to obtain approximation results for normal and Poisson distributions. Error bounds and rates of convergence are obtained for both the independent and identically distributed case and the weakly dependent case in each of the normal and Poisson approximations. We also give an original example of subgraph counts in Bernoulli random graphs, of which normal and Poisson approximations are obtained using the results from Stein's method. The validity of such approximations is discussed and rates of convergence are compared.

Contents

1 Introduction **3**
- 1.1 Motivation . . . 3
- 1.2 Historical Background of Stein's Method . . . 4
- 1.3 Summary of the Thesis . . . 4
- 1.4 Contributions of the Thesis . . . 5

2 Probability Metrics **6**
- 2.1 Integral Probability Metrics . . . 6
- 2.2 Kolmogorov, Wasserstein and Total Variation Metrics . . . 6

3 Normal Approximation **9**
- 3.1 Stein's Method for Normal Approximation . . . 9
- 3.2 Bounding the Error . . . 12
- 3.3 Sum of Independent Random Variables . . . 13
- 3.4 Extension to Local Dependence . . . 15
- 3.5 Example: Complete Graphs K_4 in Random Graphs . . . 20

4 Poisson Approximation **24**
- 4.1 Stein's Method for Poisson Approximation . . . 24
- 4.2 Bounding the Error . . . 27
- 4.3 Law of Small Numbers . . . 27
- 4.4 Extension to Local Dependence . . . 29
- 4.5 Revisiting the Example of Complete Graphs K_4 Counts . . . 31

5 Conclusion **34**

Chapter 1

Introduction

1.1 Motivation

An important concern in probability theory is to prove distributional limit theorems and to obtain explicit rates of convergence for them. In general, obtaining error bounds is much harder than showing convergence, since it requires many extra efforts, as discussed on page 211 of [18].

We consider the central limit theorem as a motivating example. The central limit theorem is a fundamental theorem in probability and statistics. It states that a sum of independent and identically distributed random variables that follow any distribution with a non-zero and finite variance is approximately distributed as a normal distribution, provided that the sample size is sufficiently large. However, the central limit theorem does not tell us the speed at which it converges. In other words, it does not indicate how large the sample size is would be sufficient for a given problem.

The Berry-Esséen theorem (Berry [3] and Esséen [10]) quantifies the error in the central limit theorem. It states as follows:

Theorem 1.1 (The Berry-Esséen theorem). *Let $X, X_1, X_2, \cdots$ be independent and identically distributed random variables with mean μ and non-zero and finite variance σ^2 $(\sigma > 0)$. Let the random variable Z have the standard normal distribution and*

$$W_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}.$$

Suppose that $\mathbb{E}|X|^3 < +\infty$, then

$$\sup_{z\in\mathbb{R}} |\mathbb{P}(W_n \leqslant z) - \mathbb{P}(Z \leqslant z)| \leqslant \frac{C\mathbb{E}|X|^3}{\sigma^3\sqrt{n}},$$

where $C \leqslant 0.4748$ [11].

However, there are several disadvantages and limitations of such a result, as discussed on pages 211 and 212 of [18]:

1. The proof of the Berry-Esséen theorem typically employs characteristic function analysis (Fourier analysis), which results in difficulties in applying the proof method for dependent random variables and to non-normal approximations.

2. It would require many added efforts to extract rates of convergence if we prove distributional limit theorems using traditional methods such as Fourier analysis and method of moments.

Stein's method is well suited to address those problems. It is a powerful technique that can be used to obtain distributional approximation results and their respective error bounds. As we will show in this thesis, the method overcomes the drawbacks of the Berry-Esséen theorem mentioned above:

1. Stein's method is an elegant method that shows distributional approximations with automatic rates of convergence.

2. The basic method for normal approximation can be adapted to many other distributions (e.g. Poisson, exponential, geometric, variance-gamma, etc.).

3. It is very natural to extend the basic method to sums of non-identically distributed random variables with dependence structures.

(These are discussed on page 3 of [11] and page 212 of [18].)

1.2 Historical Background of Stein's Method

Stein's method was first published in the paper [21] authored by Charles Stein in 1972. The method was initially invented to quantify the errors in normal approximation for sums of dependent random variables of a certain structure [18].

However, due to its nature of flexibility, Stein's method has gone far beyond its intended purpose. In 1975, Louis Hsiao Yun Chen adapted Stein's method to approximating sums of dependent random variables by Poisson distribution in the paper [5]. Since then, the method has been applied to many other distributions such as exponential distribution [4], geometric distribution [17], variance-gamma distribution [12], etc. Recently, Stein's method has been adapted to more general distributions with applications to measuring sample quality [13], Bayesian inference [15], and so on.

1.3 Summary of the Thesis

In this thesis, we will study the basics of Stein's method, use them to obtain approximation results with their respective error bounds for normal and Poisson approximations, and apply Stein's method to a subgraph counting problem in Bernoulli random graphs. We will follow the expositions of [11] and [18]. Readers are expected to be familiar with undergraduate level real analysis and probability theory.

We begin Chapter 2 by defining integral probability metrics. We give three examples of such probability metrics, which plays an essential role in Stein's method. We also establish a useful relation between two of these metrics, the Kolmogorov and Wasserstein metrics.

In Chapter 3, we study Stein's method for normal approximation. We develop the essential ingredients of Stein's method for normal approximation and discuss the strategy for bounding the distance between probability distributions using Stein's method. Error bounds for normal approximation are obtained for both the independent and identically distributed case and the locally dependent case. We finish the chapter by giving an original example of subgraph counts in Bernoulli random graphs, of which the normal approximation is obtained.

In Chapter 4, we study Stein's method for Poisson approximation. We follow the same framework and strategy as those for Stein's method for normal approximation. We also revisit the subgraph counting problem and obtain a Poisson approximation for it, which partially complements the normal approximation for this problem.

In Chapter 5, we summarize the thesis and point out some extra work that could be done if time permits.

1.4 Contributions of the Thesis

We highlight contributions made in this thesis in order of significance.

In Section 3.5 and 4.5, we give an original example of counting the number of complete graphs K_4 in a Bernoulli random graph. The validity of such approximations is discussed and rates of convergence are compared. The approaches we take to deal with this type of problem are also slightly different from those in [18] (Section 3.2.1 and Section 4.3.2).

We give our own detailed proofs of the first half of Lemma 4.2 and Lemma 4.5, as the proofs of them are not given in [18].

Thorough proofs of Proposition 2.3, Lemma 3.2, Theorem 3.7, Theorem 3.12, Corollary 3.14, Theorem 4.6, and Theorem 4.8 are given in this thesis, whereas significant details are missing in the proofs of them in [11] and [18].

Chapter 2

Probability Metrics

2.1 Integral Probability Metrics

A probability metric is a measure of distance between two probability distributions. In this section, we define a family of probability metrics, integral probability metrics [16], which plays an essential role in Stein's method.

Definition 2.1. *Let X and Y be random variables with distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, respectively. An **integral probability metric** is defined by*

$$d_{\mathcal{H}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X) - \mathbb{E}h(Y)|,$$

for some class of test functions $\mathcal{H}$.

From here onwards, for random variables X and Y with their respective distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, we will abuse notation to write $d_{\mathcal{H}}(X, Y)$ (the distance between random variables X and Y) in place of $d_{\mathcal{H}}(\mathcal{L}(X), \mathcal{L}(Y))$.

2.2 Kolmogorov, Wasserstein and Total Variation Metrics

Definition 2.2. *For a set $A \subseteq \Omega$, we define the **indicator function** $\mathbb{I}_A : \Omega \to \{0, 1\}$ as*

$$\mathbb{I}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}, \quad \textit{for } x \in \Omega.$$

We now introduce three useful examples of integral probability metrics and some of their properties and relations. These will be frequently employed in Stein's method.

1. ***Kolmogorov*** **metric**

$$\begin{aligned} d_K(X, Y) &= \sup_{h \in \mathcal{H}_K} |\mathbb{E}h(X) - \mathbb{E}h(Y)| \\ &= \sup_{z \in \mathbb{R}} |\mathbb{P}(X \leqslant z) - \mathbb{P}(Y \leqslant z)|, \end{aligned}$$

   where $\mathcal{H}_K = \{\mathbb{I}_{(-\infty, z]} : z \in \mathbb{R}\}$. The Kolmogorov metric is the maximum distance between the cumulative distribution functions of the random variables X and Y. A sequence of random variables converging to a fixed random variable in this metric implies convergence in distribution (or, equivalently, weak convergence).

2. ***Wasserstein*** **metric**

$$d_W(X,Y) = \sup_{h\in\mathcal{H}_W} |\mathbb{E}h(X) - \mathbb{E}h(Y)|,$$

where $\mathcal{H}_W = \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leqslant |x-y|, \text{ for any } x, y \in \mathbb{R}\}$ is a class of Lipschitz functions with Lipschitz constants $L_h \leqslant 1$. The Wasserstein metric is usually used for approximation by continuous distributions [18].

3. ***Total variation*** **metric**

$$d_{TV}(X,Y) = \sup_{h\in\mathcal{H}_{TV}} |\mathbb{E}h(X) - \mathbb{E}h(Y)|,$$

where $\mathcal{H}_{TV} = \{\mathbb{I}_B : B \in \text{Borel}(\mathbb{R})\}$. The total variation metric is usually used for approximation by discrete distributions [18].

Proposition 2.3. *Suppose Y is a continuous random variable with probability density function p_Y bounded by a constant C. Then, for any random variable X,*

$$d_K(X,Y) \leqslant \sqrt{2C d_W(X,Y)}.$$

Proof. Consider test functions $h_z = \mathbb{I}_{(-\infty,z]} \in \mathcal{H}_K$, $z \in \mathbb{R}$. For $\varepsilon > 0$, we define

$$h_{z,\varepsilon}(x) = \begin{cases} 1, & x \leqslant z \\ 1 - \frac{x-z}{\varepsilon}, & z < x \leqslant z+\varepsilon \\ 0, & x > z+\varepsilon \end{cases} \quad \text{(see Figure 2.1).}$$

Then, for all $z \in \mathbb{R}$, we have

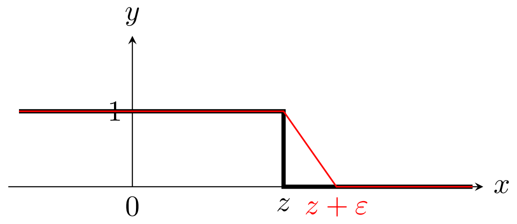
$$\begin{aligned} \mathbb{E}h_z(X) - \mathbb{E}h_z(Y) &= \mathbb{E}h_z(X) - \mathbb{E}h_{z,\varepsilon}(Y) + \mathbb{E}h_{z,\varepsilon}(Y) - \mathbb{E}h_z(Y) \\ &\leqslant \mathbb{E}h_{z,\varepsilon}(X) - \mathbb{E}h_{z,\varepsilon}(Y) + \int_z^{z+\varepsilon} \left(1 - \frac{x-z}{\varepsilon}\right) p_Y(x)dx \\ &\leqslant (\mathbb{E}[\varepsilon h_{z,\varepsilon}(X)] - \mathbb{E}[\varepsilon h_{z,\varepsilon}(Y)])/\varepsilon + C\varepsilon/2 \\ &\leqslant d_W(X,Y)/\varepsilon + C\varepsilon/2, \end{aligned}$$

where the final inequality follows since $\varepsilon h_{z,\varepsilon} \in \mathcal{H}_W$. Taking $\varepsilon = \sqrt{2d_W(X,Y)/C}$, we obtain half of the desired inequality

$$\mathbb{E}h_z(X) - \mathbb{E}h_z(Y) \leqslant \sqrt{2C d_W(X,Y)}, \quad \text{for all } z \in \mathbb{R}.$$

Similarly, for $\varepsilon > 0$, we define

$$h_{z,-\varepsilon}(x) = \begin{cases} 1, & x \leqslant z-\varepsilon \\ -\frac{x-z}{\varepsilon}, & z-\varepsilon < x \leqslant z \\ 0, & x > z \end{cases} \quad \text{(see Figure 2.1).}$$



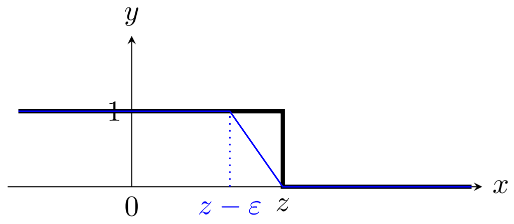


Figure 2.1: Graphs of functions $y = h_z(x)$ (the black lines in both figures), $y = h_{z,\varepsilon}(x)$ (the red line in the left figure), and $y = h_{z,-\varepsilon}(x)$ (the blue line in the right figure).

Then, for all $z \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}h_z(Y) - \mathbb{E}h_z(X) &= \mathbb{E}h_z(Y) - \mathbb{E}h_{z,-\varepsilon}(Y) + \mathbb{E}h_{z,-\varepsilon}(Y) - \mathbb{E}h_z(X) \\
&\leqslant \int_{z-\varepsilon}^{z} \left(1 + \frac{x-z}{\varepsilon}\right) p_Y(x)dx + \mathbb{E}h_{z,-\varepsilon}(Y) - \mathbb{E}h_{z,-\varepsilon}(X) \\
&\leqslant C\varepsilon/2 + (\mathbb{E}[\varepsilon h_{z,-\varepsilon}(Y)] - \mathbb{E}[\varepsilon h_{z,-\varepsilon}(X)])/\varepsilon \\
&\leqslant C\varepsilon/2 + d_W(X,Y)/\varepsilon,
\end{aligned}$$

where the final inequality follows since $\varepsilon h_{z,-\varepsilon} \in \mathcal{H}_W$. Taking $\varepsilon = \sqrt{2d_W(X,Y)/C}$, we obtain the other half of the desired inequality

$$\mathbb{E}h_z(Y) - \mathbb{E}h_z(X) \leqslant \sqrt{2Cd_W(X,Y)}, \quad \text{for all } z \in \mathbb{R}.$$

Hence, we have shown that

$$|\mathbb{E}h_z(X) - \mathbb{E}h_z(Y)| \leqslant \sqrt{2Cd_W(X,Y)}, \quad \text{for all } z \in \mathbb{R}.$$

Therefore,

$$d_K(X,Y) = \sup_{z\in\mathbb{R}} |\mathbb{E}h_z(X) - \mathbb{E}h_z(Y)| \leqslant \sqrt{2Cd_W(X,Y)}.$$

□

Remark 2.4. Proposition 2.3 implies that a bound between a given distribution and the target distribution in the Wasserstein metric immediately yields a bound between them in the Kolmogorov metric. This nice property reduces the problem of bounding the distance between probability distributions in the Kolmogorov metric to the problem of bounding that in the Wasserstein metric. However, such a bound in the Kolmogorov metric is often of suboptimal order. As we will see in Section 3.3, the rate of convergence of the central limit theorem with respect to the Kolmogorov metric obtained using Proposition 2.3 is slower than that would arise from Theorem 1.1 (the Berry-Esséen theorem).

Chapter 3

Normal Approximation

3.1 Stein's Method for Normal Approximation

The key idea behind Stein's method for distributional approximation is to show distributional convergence using a **characterizing operator** $\mathcal{A}$ [18].

Definition 3.1. *A function $f : \mathbb{R} \to \mathbb{R}$ is said to be **absolutely continuous** if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\sum_{i=1}^{n} |f(x_i) - f(x_i')| < \varepsilon$$

for every finite collection $\{(x_i, x_i')\}_{i=1}^{n}$ of non-overlapping intervals with

$$\sum_{i=1}^{n} |x_i - x_i'| < \delta$$

[19, 22].

We first obtain a characterizing operator for the standard normal distribution.

Lemma 3.2 (**Stein's lemma**). *Let $f : \mathbb{R} \to \mathbb{R}$ be a function, and let W be a random variable. We define the functional operator $\mathcal{A}_\Phi$ by*

$$\mathcal{A}_\Phi f(x) = f'(x) - xf(x).$$

Then, W has the standard normal distribution if and only if, for all f such that f is absolutely continuous and $\mathbb{E}|f'(Z)| < +\infty$ for $Z \sim \mathcal{N}(0, 1)$,

$$\mathbb{E}\mathcal{A}_\Phi f(W) = 0 \quad (i.e.\ \mathbb{E}[f'(W) - Wf(W)] = 0).$$

The differential operator $\mathcal{A}_\Phi$ is referred to as a characterizing operator of the standard normal distribution [18].

Before proving Stein's lemma, we discuss the heuristic of Stein's method. The distance between any random variable W and $Z \sim \mathcal{N}(0, 1)$ can be measured by

$$\mathbb{E}h(W) - Nh \tag{3.1}$$

over a class of test functions $\mathcal{H}$, where $Nh = \mathbb{E}h(Z)$. For normal approximation, our aim is to bound (3.1) for any $h \in \mathcal{H}$. By Stein's lemma, if $\mathbb{E}\mathcal{A}_\Phi f(W)$ is close to zero for many functions f, then we would expect W to be close to Z in distribution [11]. This intuition motivates the so-called **Stein equation**:

$$f'(x) - xf(x) = h(x) - Nh. \tag{3.2}$$

Definition 3.3. *Let $f : D_f \subseteq \mathbb{R} \to \mathbb{R}$ be a function. We define the **supremum norm** of the function f by*

$$\|f\| = \sup_{x \in D_f} |f(x)|.$$

Lemma 3.4. *The unique bounded solution of the Stein equation* (3.2) *is given by*

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} [h(t) - Nh] e^{-t^2/2} dt. \tag{3.3}$$

1. *If h is bounded, then*

$$\|f_h\| \leqslant \sqrt{\frac{\pi}{2}} \|h - Nh\| \quad \text{and} \quad \|f_h'\| \leqslant 2\|h - Nh\|.$$

2. *If h is absolutely continuous, then*

$$\|f_h''\| \leqslant 2\|h'\|.$$

Proof. Multiplying both sides of the Stein equation by an integrating factor $I = e^{-x^2/2}$ and integrating both sides over $(-\infty, x]$ gives the general solution:

$$f_{h,C}(x) = e^{x^2/2} \int_{-\infty}^{x} [h(t) - Nh] e^{-t^2/2} dt + Ce^{x^2/2}, \quad C \in \mathbb{R} \text{ is a constant.}$$

Therefore, (3.3) is a solution to the Stein equation.

We refer to [6] (Lemma 2.4) for the proof of the bounds for f_h and its derivatives. The proof consists of elementary calculations, which is technical and irrelevant to the idea of Stein's method. Therefore, we omit it in this thesis.

Now, we show that (3.3) is the unique bounded solution. Suppose that u and v are two bounded solutions to the Stein equation, then u and v satisfy

$$u'(x) - xu(x) = h(x) - Nh \quad \text{and} \quad v'(x) - xv(x) = h(x) - Nh.$$

We define $w = u - v$, and thus w satisfies

$$w'(x) - xw(x) = 0,$$

to which the general solution is $w(x) = C_w e^{x^2/2}$. Since u and v are bounded, w is bounded. Therefore, $C_w = 0$ and $w = 0$, which implies $u = v$. Hence, we have shown uniqueness. □

Proof of Lemma 3.2 (Stein's lemma)

Proof. ("$\Longrightarrow$") We prove this part assuming that $\lim_{x\to\pm\infty} f(x)e^{-x^2/2} = 0$. Suppose that W has the standard normal distribution, then by integration by parts, for any absolutely continuous function f with $\mathbb{E}|f'(W)| < +\infty$,

$$\begin{aligned}
\mathbb{E}f'(W) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} f'(x)e^{-x^2/2}dx \\
&= \left[\frac{1}{\sqrt{2\pi}} f(x)e^{-x^2/2}\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} xf(x)e^{-x^2/2}dx \\
&= \mathbb{E}Wf(W).
\end{aligned}$$

We refer to [18] (Lemma 2.1) for the proof of this part without the extra assumption, where Fubini's theorem is employed.

("$\Longleftarrow$") Consider the Stein equation (3.2) with test functions $h_z = \mathbb{I}_{(-\infty,z]} \in \mathcal{H}_K, z \in \mathbb{R}$:

$$f'(x) - xf(x) = \mathbb{I}_{(-\infty,z]}(x) - \Phi(z), \tag{3.4}$$

where Φ is the cumulative distribution function of the standard normal distribution. By Lemma 3.4, the unique bounded solution of (3.4) is given by

$$\begin{aligned}
f_z(x) &= e^{x^2/2} \int_{-\infty}^{x} [\mathbb{I}_{(-\infty,z]}(t) - \Phi(z)]e^{-t^2/2}dt \\
&= \begin{cases} e^{x^2/2} \int_{-\infty}^{x} (1-\Phi(z))e^{-t^2/2}dt, & x \leqslant z \\ e^{x^2/2} \int_{-\infty}^{z} (1-\Phi(z))e^{-t^2/2}dt + e^{x^2/2} \int_{z}^{x} (-\Phi(z))e^{-t^2/2}dt, & x > z \end{cases} \\
&= \begin{cases} \sqrt{2\pi}e^{x^2/2}\Phi(x)(1-\Phi(z)), & x \leqslant z \\ \sqrt{2\pi}e^{x^2/2}\Phi(z)(1-\Phi(x)), & x > z \end{cases}.
\end{aligned}$$

Since $h_z(x) = \mathbb{I}_{(-\infty,z]}(x) \in \{0,1\}$ is bounded and $\|h_z - Nh_z\| = \|h_z - \Phi(z)\| \leqslant 1$, f_z' is bounded by

$$\|f_z'\| \leqslant 2\|h_z - Nh_z\| \leqslant 2.$$

Therefore, f_z is Lipschitz with Lipschitz constant $L_{f_z} \leqslant 2$. Let $\varepsilon > 0$ be given, and choose $\delta = \varepsilon/2$. Assume that the finite collection $\{(x_i, x_i')\}_{i=1}^{n}$ of non-overlapping intervals satisfies

$$\sum_{i=1}^{n} |x_i - x_i'| < \delta = \frac{\varepsilon}{2}.$$

Then, for such intervals $\{(x_i, x_i')\}_{i=1}^{n}$, we have

$$\sum_{i=1}^{n} |f_z(x_i) - f_z(x_i')| \leqslant \sum_{i=1}^{n} L_{f_z}|x_i - x_i'| \leqslant 2\sum_{i=1}^{n} |x_i - x_i'| < 2\delta = 2\left(\frac{\varepsilon}{2}\right) = \varepsilon.$$

Hence, f_z is absolutely continuous. If we denote the probability density function of $Z \sim \mathcal{N}(0,1)$ by ϕ, then

$$\mathbb{E}|f_z'(Z)| = \int_{-\infty}^{+\infty} |f_z'(x)|\phi(x)dx \leqslant \|f_z'\| \int_{-\infty}^{+\infty} \phi(x)dx \leqslant 2 < +\infty.$$

Now, suppose that W is a random variable such that $\mathbb{E}[f'(W) - Wf(W)] = 0$ for all functions $f : \mathbb{R} \to \mathbb{R}$ such that f is absolutely continuous and $\mathbb{E}|f'(Z)| < +\infty$ for $Z \sim \mathcal{N}(0,1)$. We have verified that the function f_z satisfying (3.4) is such a function. If we evaluate both sides of the equation

$$f_z'(x) - xf_z(x) = \mathbb{I}_{(-\infty,z]}(x) - \Phi(z)$$

at the random variable W and take expectations in both sides, we have

$$\mathbb{E}[f_z'(W) - Wf_z(W)] = \mathbb{E}[\mathbb{I}_{(-\infty,z]}(W) - \Phi(z)], \quad \text{for all } z \in \mathbb{R}.$$

Since

$$\mathbb{E}[f_z'(W) - Wf_z(W)] = 0 \quad \text{and} \quad \mathbb{E}[\mathbb{I}_{(-\infty,z]}(W) - \Phi(z)] = \mathbb{P}(W \leqslant z) - \Phi(z),$$

we have

$$\mathbb{P}(W \leqslant z) = \Phi(z), \quad \text{for all } z \in \mathbb{R},$$

which implies that W has the standard normal distribution. □

Corollary 3.5. *Suppose that f_z is the unique bounded solution of* (3.4). *Then, for any random variable W, we have*

$$|\mathbb{P}(W \leqslant z) - \Phi(z)| = |\mathbb{E}[f_z'(W) - Wf_z(W)]|,$$

where Φ is the cumulative distribution function of the standard normal distribution.

3.2 Bounding the Error

Our strategy for bounding the distance between any random variable W and the standard normal random variable $Z \sim \mathcal{N}(0,1)$ now becomes clear.

Let f_h be the unique bounded solution of the Stein equation (3.2). We evaluate both sides of the equation

$$f_h'(x) - xf_h(x) = h(x) - Nh$$

at a random variable W and take expectations in both sides to obtain

$$\mathbb{E}[f_h'(W) - Wf_h(W)] = \mathbb{E}h(W) - Nh.$$

Since the right-hand side of the equation above is the quantity that we want to bound, we have now reduced the problem to bounding

$$\mathbb{E}[f_h'(W) - Wf_h(W)]$$

for a given function $h \in \mathcal{H}$ [11]. We can bound it using the structure of the random variable W and the properties of the function f_h [18].

Proposition 3.6. *Let the random variable Z have the standard normal distribution. Then*

$$d_{\mathcal{H}}(W,Z) = \sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - Nh| = \sup_{h\in\mathcal{H}} |\mathbb{E}[f_h'(W) - Wf_h(W)]|,$$

for any random variable W, where $\mathcal{H}$ is some class of test functions, and f_h is the unique bounded solution of the Stein equation (3.2)*.*

3.3 Sum of Independent Random Variables

We now show two error bounds for the normal approximation of sums of n independent and identically distributed random variables for finite n using Stein's method. We focus mainly on obtaining an error bound in the Wasserstein metric, since the test functions $h \in \mathcal{H}_W$ for the Wasserstein metric are Lipschitz with $\|h'\| \leqslant 1$, which allows us to bound the second derivative of the unique bounded solution f_h of the Stein equation (3.2) by $\|f_h''\| \leqslant 2\|h'\| \leqslant 2$, by Lemma 3.4. Once we obtain an error bound in the Wasserstein metric, we will be able to immediately obtain an error bound in the Kolmogorov metric using Proposition 2.3.

We note that the proof employs only basic mathematical techniques such as Taylor's theorem, triangle inequality, and properties of expectation in probability theory, which shows why Stein's method is a powerful and elegant method.

Theorem 3.7. *Suppose that $X, X_1, X_2, \cdots, X_n$ are independent and identically distributed random variables with $\mathbb{E}(X) = 0$, $Var(X) = 1$, and $\mathbb{E}|X^3| < +\infty$. Suppose that the random variable Z has the standard normal distribution and $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, then*

$$d_W(W,Z) \leqslant \frac{1}{\sqrt{n}}(2 + \mathbb{E}|X^3|)$$

and

$$d_K(W,Z) = \sup_{z\in\mathbb{R}} |\mathbb{P}(W \leqslant z) - \Phi(z)| \leqslant \left(\frac{2}{n\pi}\right)^{1/4} \sqrt{2 + \mathbb{E}|X^3|}. \tag{3.5}$$

Proof. Let f_h be the unique bounded solution of the Stein equation (3.2). Define $W_i = W - \frac{X_i}{\sqrt{n}}$, and thus W_i and X_i are independent. By Taylor's theorem with Lagrange's form of remainder, for $x, x_0 \in \mathbb{R}$, we have

$$f_h(x) = f_h(x_0) + f_h'(x_0)(x - x_0) + \frac{f_h''(c_1)}{2}(x - x_0)^2, \tag{3.6}$$

where $c_1 = x_0 + \theta_1(x - x_0)$ for some $\theta_1 \in (0,1)$, and

$$f_h'(x) = f_h'(x_0) + f_h''(c_2)(x - x_0), \tag{3.7}$$

where $c_2 = x_0 + \theta_2(x - x_0)$ for some $\theta_2 \in (0,1)$. For each $i = 1, \cdots, n$, we obtain the following equations by replacing x and x_0 by the random variables W and W_i, respectively, in equations (3.6) and (3.7):

$$f_h(W) = f_h(W_i) + f_h'(W_i)\frac{X_i}{\sqrt{n}} + f_h''(C_{1i})\frac{X_i^2}{2n}, \tag{3.8}$$

where the random variable $C_{1i} = W_i + \theta_{1i}\frac{X_i}{\sqrt{n}}$ for some $\theta_{1i} \in (0,1)$, and

$$f_h'(W) = f_h'(W_i) + f_h''(C_{2i})\frac{X_i}{\sqrt{n}}, \tag{3.9}$$

where the random variable $C_{2i} = W_i + \theta_{2i}\frac{X_i}{\sqrt{n}}$ for some $\theta_{2i} \in (0,1)$. Then, we have

$$\begin{aligned}
\mathbb{E}[Wf_h(W)] &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i f_h(W)] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i f_h(W_i)] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i^2 f_h'(W_i)] + \frac{1}{2n\sqrt{n}}\sum_{i=1}^n \mathbb{E}[X_i^3 f_h''(C_{1i})] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}(X_i)\mathbb{E}f_h(W_i) + \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i^2)\mathbb{E}f_h'(W_i) + \frac{1}{2\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^3 f_h''(C_{1i})\right] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}f_h'(W_i) + \frac{1}{2\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^3 f_h''(C_{1i})\right] \\
&= \mathbb{E}f_h'(W) - \frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i f_h''(C_{2i})\right] + \frac{1}{2\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^3 f_h''(C_{1i})\right],
\end{aligned}$$

where the second equality follows by equation (3.8), the third equality follows since X_i and W_i are independent for each i, the fourth equality follows since $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^2) = \mathrm{Var}(X_i) + [\mathbb{E}(X_i)]^2 = 1$, and the final equality follows by equation (3.9). Therefore,

$$\begin{aligned}
|\mathbb{E}[f_h'(W) - Wf_h(W)]| &\leqslant \left|\frac{1}{\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i f_h''(C_{2i})\right]\right| + \left|\frac{1}{2\sqrt{n}}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^3 f_h''(C_{1i})\right]\right| \\
&\leqslant \frac{1}{\sqrt{n}}\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n X_i f_h''(C_{2i})\right| + \frac{1}{2\sqrt{n}}\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n X_i^3 f_h''(C_{1i})\right| \\
&\leqslant \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i f_h''(C_{2i})|\right) + \frac{1}{2\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i^3 f_h''(C_{1i})|\right) \\
&\leqslant \frac{\|f_h''\|}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i|\right) + \frac{\|f_h''\|}{2\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i^3|\right) \\
&= \frac{\|f_h''\|}{2\sqrt{n}}(2\mathbb{E}|X| + \mathbb{E}|X^3|) \\
&\leqslant \frac{\|f_h''\|}{2\sqrt{n}}(2 + \mathbb{E}|X^3|),
\end{aligned}$$

where the equality in the penultimate line follows since $X_1, X_2, \cdots, X_n$ are identically distributed as X, and the final inequality follows since $\mathbb{E}|X| \leqslant \sqrt{\mathbb{E}(X^2)} = 1$, by the Cauchy-Schwarz inequality $|\mathbb{E}(Y_1Y_2)| \leqslant \sqrt{\mathbb{E}(Y_1^2)}\sqrt{\mathbb{E}(Y_2^2)}$ with random variables $Y_1 = |X|$ and $Y_2 = 1$.

By Proposition 3.6, we have

$$\begin{aligned} d_W(W,Z) &= \sup_{h\in\mathcal{H}_W} |\mathbb{E}[f_h'(W) - Wf_h(W)]| \\ &\leqslant \sup_{h\in\mathcal{H}_W} \frac{\|f_h''\|}{2\sqrt{n}}(2+\mathbb{E}|X^3|) \\ &\leqslant \sup_{h\in\mathcal{H}_W} \frac{\|h'\|}{\sqrt{n}}(2+\mathbb{E}|X^3|) \\ &\leqslant \frac{1}{\sqrt{n}}(2+\mathbb{E}|X^3|), \end{aligned}$$

where the inequality in the penultimate line follows since $\|f_h''\| \leqslant 2\|h'\|$ by Lemma 3.4, and the final inequality follows since any $h \in \mathcal{H}_W$ is Lipschitz with $\|h'\| \leqslant 1$. Since the probability density function of the random variable $Z \sim \mathcal{N}(0,1)$ is bounded by

$$|\phi(x)| = \left|\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right| \leqslant \frac{1}{\sqrt{2\pi}}, \quad \text{for all } x \in \mathbb{R},$$

then, by Proposition 2.3, we have

$$d_K(W,Z) \leqslant \left(\frac{2}{\pi}\right)^{1/4}\sqrt{d_W(W,Z)} \leqslant \left(\frac{2}{n\pi}\right)^{1/4}\sqrt{2+\mathbb{E}|X^3|}.$$

□

Definition 3.8 (**Big-O notation**). *Let $\mathbb{N}$ denote the set of all natural numbers $\{0,1,2,\cdots\}$ and $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$. Let $f,g : \mathbb{N} \to \mathbb{R}^+$ be two functions. Then, we say "$f(n) = \mathcal{O}(g(n))$" to mean that "there exist positive constants c and n_0 such that, for all $n > n_0$, $f(n) \leqslant cg(n)$".*

Remark 3.9. The error bounds given in Theorem 3.7 indicate how good the approximation by the central limit theorem is for finite n. We note that the rate of convergence of the central limit theorem arising from (3.5) is $\mathcal{O}(n^{-1/4})$, which is slower than $\mathcal{O}(n^{-1/2})$ that would arise from Theorem 1.1 (the Berry-Esséen theorem). This is because the bound in the Kolmogorov metric obtained using Proposition 2.3 is often of suboptimal order.

3.4 Extension to Local Dependence

In this section, we generalize Theorem 3.7 to sums of random variables with local dependence. We first define the dependency neighbourhood of a random variable.

Definition 3.10. *Let $X_1, X_2, \cdots, X_n$ be random variables. For each $i = 1,\cdots,n$, we define the* ***dependency neighbourhood*** *of the random variable X_i as the set $D_i \subseteq \{1,\cdots,n\}$ such that X_i is independent of $\{X_j : j \notin D_i\}$.*

Remark 3.11. Dependency neighbourhoods are also referred to as dependency graphs, since their structure can be represented in the form of an undirected graph $G(V,E)$ with vertex set $V = \{1,\cdots,n\}$ and edge set $E = \{(i,j) : i \neq j \text{ and } j \in D_i\}$ (i.e. vertex i is connected to vertex $j \neq i$ if $j \in D_i$) [18].

We show an error bound for normal approximations of sums of n locally dependent random variables for finite n using the argument analogous to that in the proof of Theorem 3.7.

Theorem 3.12. *Let $X_1, X_2, \cdots, X_n$ be random variables with mean zero and finite variance. Let $W = \sum_{i=1}^{n} X_i$ and assume $Var(W) = 1$. Let D_i be the dependency neighbourhood of X_i. Suppose that the random variable Z has the standard normal distribution, then*

$$d_W(W,Z) \leqslant \sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i}\mathbb{E}|X_iX_jX_k| + 2\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_j\setminus D_i}\mathbb{E}|X_iX_jX_k| + 2\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i\cup D_j}|\mathbb{E}X_iX_j|\mathbb{E}|X_k|.$$

Proof. Let f_h be the unique bounded solution to the Stein equation. Define $W_i = W - \sum_{j\in D_i} X_j$ and $W_{i,j} = W - \sum_{k\in D_i\cup D_j} X_k = W_i - \sum_{k\in D_j\setminus D_i} X_k$. Since $i \in D_i$ for all i, then W_i and X_i are independent, and $W_{i,j}$ is independent of X_i and X_j.

For each $i = 1, \cdots, n$, we replace x and x_0 by the random variables W and W_i, respectively, in equation (3.6) to obtain the equation

$$f_h(W) = f_h(W_i) + f_h'(W_i)\sum_{j\in D_i} X_j + \frac{1}{2}f_h''(C_{1i})\left(\sum_{j\in D_i} X_j\right)^2, \tag{3.10}$$

where the random variable $C_{1i} = W_i + \theta_{1i}\sum_{j\in D_i} X_j$ for some $\theta_{1i} \in (0,1)$. For each $i = 1, \cdots, n$ and $j = 1, \cdots, n$, we replace x and x_0 by the random variables W_i and $W_{i,j}$, respectively, in equation (3.7) to obtain the equation

$$f_h'(W_i) = f_h'(W_{i,j}) + f_h''(C_{2ij})\sum_{k\in D_j\setminus D_i} X_k, \tag{3.11}$$

where the random variable $C_{2ij} = W_{i,j} + \theta_{2ij}\sum_{k\in D_j\setminus D_i} X_k$ for some $\theta_{2ij} \in (0,1)$. Similarly, we replace x and x_0 by the random variables W and $W_{i,j}$, respectively, in equation (3.7) to obtain the equation

$$f_h'(W) = f_h'(W_{i,j}) + f_h''(C_{3ij})\sum_{k\in D_i\cup D_j} X_k, \tag{3.12}$$

where the random variable $C_{3ij} = W_{i,j} + \theta_{3ij}\sum_{k\in D_i\cup D_j} X_k$ for some $\theta_{3ij} \in (0,1)$. Using these equations, we can obtain

$$\begin{aligned}
\mathbb{E}[Wf_h(W)] &= \sum_{i=1}^{n} \mathbb{E}[X_i f_h(W)] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i f_h(W_i)] + \sum_{i=1}^{n} \mathbb{E}\left[X_i f_h'(W_i) \sum_{j\in D_i} X_j\right] + R_1 \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i)\mathbb{E}f_h(W_i) + \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}[X_iX_j f_h'(W_i)] + R_1 \\
&= \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}[X_iX_j f_h'(W_i)] + R_1 \\
&= \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}[X_iX_j f_h'(W_{i,j})] + R_1 + R_2 \\
&= \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}(X_iX_j)\mathbb{E}f_h'(W_{i,j}) + R_1 + R_2 \\
&= \left[\sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}(X_iX_j)\right]\mathbb{E}f_h'(W) + R_1 + R_2 + R_3 \\
&= \mathbb{E}f_h'(W) + R_1 + R_2 + R_3,
\end{aligned}$$

where the second equality follows by equation (3.10), the third equality follows since X_i and W_i are independent for each i, the fourth equality follows by the assumption that $\mathbb{E}(X_i) = 0$, the fifth equality follows by equation (3.11), the sixth equality follows since $W_{i,j}$ is independent of X_i and X_j, the penultimate equality follows by equation (3.12), the final equality follows since

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}(X_iX_j) &= \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}(X_iX_j) + \sum_{i=1}^{n}\sum_{j\notin D_i} \mathbb{E}(X_iX_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}(X_iX_j) \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\
&= \mathbb{E}(W^2) \\
&= \mathrm{Var}(W) + [\mathbb{E}(W)]^2 \\
&= 1 + \left[\sum_{i=1}^{n} \mathbb{E}(X_i)\right]^2 \\
&= 1,
\end{aligned}$$

and the remainders are

$$\begin{aligned}
R_1 &= \sum_{i=1}^{n} \mathbb{E}\left[\frac{f_h''(C_{1i})}{2} X_i \left(\sum_{j\in D_i} X_j\right)^2\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\frac{f_h''(C_{1i})}{2} X_i \sum_{j\in D_i}\sum_{k\in D_i} X_j X_k\right] \\
&= \sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i} \mathbb{E}\left[\frac{f_h''(C_{1i})}{2} X_i X_j X_k\right],
\end{aligned}$$

$$\begin{aligned}
R_2 &= \sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}\left[X_i X_j f_h''(C_{2ij}) \sum_{k\in D_j\setminus D_i} X_k\right] \\
&= \sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_j\setminus D_i} \mathbb{E}[f_h''(C_{2ij}) X_i X_j X_k],
\end{aligned}$$

$$\begin{aligned}
R_3 &= -\sum_{i=1}^{n}\sum_{j\in D_i} \mathbb{E}(X_i X_j)\mathbb{E}\left[f_h''(C_{3ij}) \sum_{k\in D_i\cup D_j} X_k\right] \\
&= -\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i\cup D_j} \mathbb{E}(X_i X_j)\mathbb{E}\left[f_h''(C_{3ij}) X_k\right].
\end{aligned}$$

Therefore, since $\|f_h''\| \leqslant 2\|h'\|$ and $\|h'\| \leqslant 1$ for any $h \in \mathcal{H}_W$, we have

$$\begin{aligned}
d_W(W,Z) &= \sup_{h\in\mathcal{H}_W} |\mathbb{E}[f_h'(W) - W f_h(W)]| \\
&\leqslant \sup_{h\in\mathcal{H}_W} (|R_1| + |R_2| + |R_3|) \\
&\leqslant \sup_{h\in\mathcal{H}_W} \Bigg[\frac{\|f_h''\|}{2}\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i} \mathbb{E}|X_i X_j X_k| + \|f_h''\|\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_j\setminus D_i} \mathbb{E}|X_i X_j X_k| \\
&\qquad + \|f_h''\|\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i\cup D_j} |\mathbb{E}X_i X_j|\mathbb{E}|X_k|\Bigg] \\
&\leqslant \sup_{h\in\mathcal{H}_W} \Bigg[\|h'\|\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i} \mathbb{E}|X_i X_j X_k| + 2\|h'\|\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_j\setminus D_i} \mathbb{E}|X_i X_j X_k| \\
&\qquad + 2\|h'\|\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i\cup D_j} |\mathbb{E}X_i X_j|\mathbb{E}|X_k|\Bigg] \\
&\leqslant \sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i} \mathbb{E}|X_i X_j X_k| + 2\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_j\setminus D_i} \mathbb{E}|X_i X_j X_k| \\
&\quad + 2\sum_{i=1}^{n}\sum_{j\in D_i}\sum_{k\in D_i\cup D_j} |\mathbb{E}X_i X_j|\mathbb{E}|X_k|.
\end{aligned}$$

$\square$

Remark 3.13. Consider the independent and identically distributed case discussed in the previous section, where $D_i = \{i\}$ for each $i = 1, \cdots, n$. Put $X_i = \frac{1}{\sqrt{n}} Y_i$, where $Y_1, \cdots, Y_n$ are independent and identically distributed random variables with mean zero and unit variance. So $\mathrm{Var}(W) = 1$. Since $D_j \backslash D_i = \emptyset$ for $j \in D_i = \{i\}$, we have

$$\begin{aligned} d_W(W, Z) &\leqslant \frac{1}{n^{3/2}} \sum_{i=1}^{n} [\mathbb{E}|Y_i^3| + 2|\mathbb{E}(Y_i^2)|\mathbb{E}|Y_i|] \\ &\leqslant \frac{1}{\sqrt{n}} (\mathbb{E}|Y^3| + 2), \end{aligned}$$

where Y is a random variable with the same distribution as Y_i, and the final inequality follows since $\mathbb{E}(Y_i^2) = 1$ and $\mathbb{E}|Y_i| \leqslant 1$, as we showed in the proof of Theorem 3.7. This result is exactly the same as that in Theorem 3.7.

We can simplify the error bound given in Theorem 3.12 as follows.

Corollary 3.14. *Let $D = \max_{1 \leqslant i \leqslant n} |D_i|$, where $|D_i|$ is the number of elements in the set D_i. If W, Z and X_i are as defined in Theorem 3.12, then*

$$d_W(W, Z) \leqslant 7D^2 \sum_{i=1}^{n} \mathbb{E}|X_i|^3.$$

Proof. By the arithmetic-geometric mean inequality, we have

$$\mathbb{E}|X_i X_j X_k| \leqslant \mathbb{E}\left(\frac{|X_i|^3 + |X_j|^3 + |X_k|^3}{3}\right) = \frac{1}{3}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right).$$

Therefore, we have

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j \in D_i} \sum_{k \in D_i} \mathbb{E}|X_i X_j X_k| &\leqslant \sum_{i=1}^{n} \sum_{j \in D_i} \sum_{k \in D_i} \frac{1}{3}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right) \\ &= \sum_{i=1}^{n} \sum_{j \in D_i} \sum_{k \in D_i} \mathbb{E}|X_i|^3 \\ &\leqslant D^2 \sum_{i=1}^{n} \mathbb{E}|X_i|^3, \end{aligned}$$

where the final inequality follows since $|D_i| \leqslant D$.

Similarly, since $|D_j \backslash D_i| \leqslant |D_j| \leqslant D$, we have

$$2 \sum_{i=1}^{n} \sum_{j \in D_i} \sum_{k \in D_j \backslash D_i} \mathbb{E}|X_i X_j X_k| \leqslant 2D^2 \sum_{i=1}^{n} \mathbb{E}|X_i|^3.$$

We find that

$$\begin{aligned} |\mathbb{E}X_i X_j|\mathbb{E}|X_k| &\leqslant \mathbb{E}|X_i X_j|\mathbb{E}|X_k| \\ &\leqslant (\mathbb{E}|X_i|^2)^{1/2}(\mathbb{E}|X_j|^2)^{1/2}\mathbb{E}|X_k| \\ &\leqslant (\mathbb{E}|X_i|^3)^{1/3}(\mathbb{E}|X_j|^3)^{1/3}(\mathbb{E}|X_k|^3)^{1/3} \\ &\leqslant \frac{1}{3}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right), \end{aligned}$$

where the second inequality follows by Cauchy-Schwarz inequality, the third inequality follows since $\mathbb{E}|X|^r \leqslant (\mathbb{E}|X|^s)^{r/s}$, for $0 < r < s$, by Hölder's inequality, and the final inequality follows by the arithmetic-geometric mean inequality. Then, since $|D_i \cup D_j| \leqslant |D_i| + |D_j| \leqslant 2D$, we have

$$\begin{aligned}
2\sum_{i=1}^n \sum_{j\in D_i} \sum_{k\in D_i\cup D_j} |\mathbb{E}X_iX_j|\mathbb{E}|X_k| &\leqslant 2\sum_{i=1}^n \sum_{j\in D_i} \sum_{k\in D_i\cup D_j} \frac{1}{3}\left(\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3\right) \\
&= 2\sum_{i=1}^n \sum_{j\in D_i} \sum_{k\in D_i\cup D_j} \mathbb{E}|X_i|^3 \\
&\leqslant 4D^2 \sum_{i=1}^n \mathbb{E}|X_i|^3.
\end{aligned}$$

Hence, following Theorem 3.12, we have

$$\begin{aligned}
d_W(W,Z) &\leqslant D^2\sum_{i=1}^n \mathbb{E}|X_i|^3 + 2D^2\sum_{i=1}^n \mathbb{E}|X_i|^3 + 4D^2\sum_{i=1}^n \mathbb{E}|X_i|^3 \\
&= 7D^2\sum_{i=1}^n \mathbb{E}|X_i|^3.
\end{aligned}$$

□

3.5 Example: Complete Graphs K_4 in Random Graphs

A nice application of Stein's method is to approximate the distribution of the number of copies of a certain type of graph occurring in a Bernoulli random graph.

In this section, we give an original example, in which we apply Stein's method to obtain a normal approximation of the distribution of the number of complete graphs K_4 in a Bernoulli random graph.

Definition 3.15. *A **Bernoulli random graph** (or a random graph generated by the **Erdős–Rényi model** [8]) $G = \mathcal{G}(n,p)$ is an undirected graph G with n vertices, where each pair of vertices is connected with probability p independent from every other pair of vertices. $p \in (0,1)$ is called edge probability.*

Definition 3.16. *A **complete graph** on n vertices, denoted by K_n, is an undirected graph on n vertices whose edge set includes every possible edge [14].*

For $n \geqslant 4$, let $G = \mathcal{G}(n,p)$ be a Bernoulli random graph. There are $N = \binom{n}{4}$ places where a complete graph K_4 can occur in G. We index each combination of four vertices in G by a distinct number $i \in \{1, \cdots, N\}$, in some arbitrary but fixed order, and refer to it as "the i-th set of four vertices"[1].

[1] A more precise notation may be helpful in other applications of Stein's method to random graph theory (e.g. the notation used in the paper [2]), but this slightly imprecise notation is absolutely fine for the examples considered in this thesis.

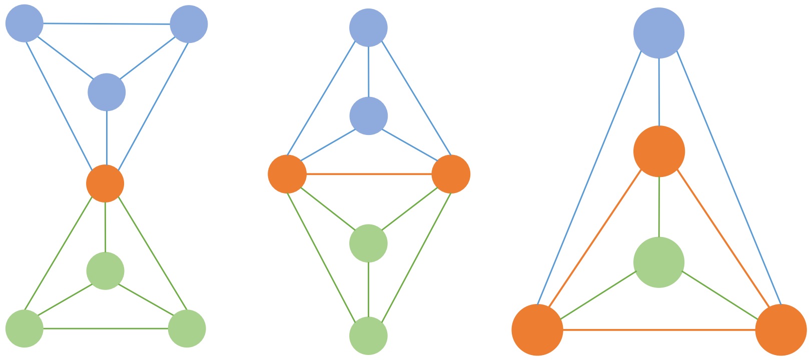


Figure 3.1: In the left diagram, the two complete graphs K_4 share one vertex and zero edge, and thus they are independent of each other. In the middle diagram, the two complete graphs K_4 share two vertices and one edge. In the right diagram, the two complete graphs K_4 share three vertices and three edges.

There are $\binom{4}{2} = 6$ edges in a complete graph K_4. For each $i = 1, \cdots, N$, we define $Y_i \sim Bernoulli(p^6)$ to be the indicator random variable that a complete graph K_4 is formed at the i-th set of four vertices. Then, $\mathbb{E}(Y_i) = p^6$ and $\text{Var}(Y_i) = p^6(1 - p^6)$.

We define $T = \sum_{i=1}^{N} Y_i$, which denotes the number of complete graphs K_4 occurring in G. Then, the expectation of T is given by

$$\mathbb{E}(T) = \mathbb{E}\left(\sum_{i=1}^{N} Y_i\right) = \sum_{i=1}^{N} \mathbb{E}(Y_i) = Np^6 = \binom{n}{4}p^6.$$

Now, we calculate the variance σ^2 of the random variable T. Let A_i be the set that contains indices of sets of four vertices which share exactly two vertices with the i-th set, and let B_i be the set that contains indices of sets of four vertices which share exactly three vertices with the i-th set. Figure 3.1 shows, from left to right, diagrams of two complete graphs K_4 which share exactly one vertex, two vertices and three vertices.

Then, for each $i = 1, \cdots, N$, we have

$$|A_i| = \binom{4}{2}\binom{n-4}{2} = 3(n-4)(n-3) \quad \text{and} \quad |B_i| = \binom{4}{3}\binom{n-4}{1} = 4(n-4).$$

We note that A_i and B_i are disjoint, and $i \notin A_i, B_i$. For any $j \neq i$, Y_j and Y_i are independent if and only if the collection of edges between the four vertices in the j-th set is disjoint from those between the four vertices in the i-th set. Therefore, for each $j \neq i$, there are three possible cases:

- if $j \notin A_i$ and $j \notin B_i$, then

$$\text{Cov}(Y_i, Y_j) = \mathbb{E}(Y_iY_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j) = \mathbb{E}(Y_i)\mathbb{E}(Y_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j) = 0;$$

- if $j \in A_i$, then
$$\begin{aligned}\mathrm{Cov}(Y_i, Y_j) &= \mathbb{E}(Y_iY_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j)\\ &= \mathbb{E}(Y_iY_j|Y_i = 1)\mathbb{P}(Y_i = 1) + \mathbb{E}(Y_iY_j|Y_i = 0)\mathbb{P}(Y_i = 0) - p^{12}\\ &= \mathbb{E}(Y_j|Y_i = 1)p^6 - p^{12}\\ &= p^5p^6 - p^{12}\\ &= p^{11}(1 - p),\end{aligned}$$
where the second equality follows by the law of total expectation;
- if $j \in B_i$, then
$$\begin{aligned}\mathrm{Cov}(Y_i, Y_j) &= \mathbb{E}(Y_iY_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j)\\ &= \mathbb{E}(Y_iY_j|Y_i = 1)\mathbb{P}(Y_i = 1) + \mathbb{E}(Y_iY_j|Y_i = 0)\mathbb{P}(Y_i = 0) - p^{12}\\ &= \mathbb{E}(Y_j|Y_i = 1)p^6 - p^{12}\\ &= p^3p^6 - p^{12}\\ &= p^9(1 - p^3),\end{aligned}$$
where the second equality follows by the law of total expectation.

Then, we have
$$\begin{aligned}\sigma^2 &= \mathrm{Var}(T)\\ &= \mathrm{Var}\left(\sum_{i=1}^N Y_i\right)\\ &= \sum_{i=1}^N \mathrm{Var}(Y_i) + \sum_{i=1}^N \sum_{j\neq i} \mathrm{Cov}(Y_i, Y_j)\\ &= Np^6(1 - p^6) + \sum_{i=1}^N \left[|A_i|p^{11}(1 - p) + |B_i|p^9(1 - p^3)\right]\\ &= Np^6(1 - p^6) + N\left[3(n - 4)(n - 3)p^{11}(1 - p) + 4(n - 4)p^9(1 - p^3)\right]\\ &= \binom{n}{4}p^6[1 - p^6 + 3(n - 4)(n - 3)p^5(1 - p) + 4(n - 4)p^3(1 - p^3)].\end{aligned}$$
Now, let W denote the standardized number of complete graphs K_4:
$$W = \frac{T - \mathbb{E}(T)}{\sqrt{\mathrm{Var}(T)}} = \frac{\sum_{i=1}^N Y_i - Np^6}{\sigma} = \sum_{i=1}^N \left(\frac{Y_i - p^6}{\sigma}\right) = \sum_{i=1}^N X_i,$$
where $X_i = (Y_i - p^6)/\sigma$. By Corollary 3.14, we have
$$d_W(W, Z) \leqslant 7D^2 \sum_{i=1}^N \mathbb{E}|X_i|^3,$$
where
$$\begin{aligned}D &= |A_1 \cup B_1 \cup \{1\}|\\ &= |A_1| + |B_1| + 1\\ &= 3(n - 4)(n - 3) + 4(n - 4) + 1\\ &= 3n^2 - 21n + 37\\ &< 3n^2, \quad \text{by assumption } n \geqslant 4,\end{aligned}$$

and, for $i = 1, \cdots, n$,

$$\begin{aligned}
\mathbb{E}|X_i|^3 &= \frac{1}{\sigma^3}\mathbb{E}|Y_i - p^6|^3 \\
&= \frac{1}{\sigma^3}\left[|0 - p^6|^3\mathbb{P}(Y_i = 0) + |1 - p^6|^3\mathbb{P}(Y_i = 1)\right] \\
&= \frac{1}{\sigma^3}[p^{18}(1 - p^6) + (1 - p^6)^3 p^6] \\
&= \frac{p^6}{\sigma^3}[p^{12}(1 - p^6) + (1 - p^6)^3] \\
&< \frac{p^6}{\sigma^3}[1 \times 1 + 1] \\
&= \frac{2p^6}{\sigma^3}.
\end{aligned}$$

Putting them together gives

$$\begin{aligned}
d_W(W, Z) &\leqslant 7D^2 \sum_{i=1}^{N} \mathbb{E}|X_i|^3 \\
&< 7 \times (3n^2)^2 \times \binom{n}{4} \times \frac{2p^6}{\sigma^3} \\
&= 7 \times 9 \times n^4 \times \frac{n(n-1)(n-2)(n-3)}{4 \times 3 \times 2 \times 1} \times \frac{2p^6}{\sigma^3} \\
&< \frac{21 n^8 p^6}{4\sigma^3},
\end{aligned}$$

where

$$\sigma^2 = \binom{n}{4} p^6[1 - p^6 + 3(n-4)(n-3)p^5(1-p) + 4(n-4)p^3(1-p^3)].$$

If p does not depend on n, then $\sigma^2 = \mathcal{O}(n^6)$ for large n, and thus $\sigma^3 = \mathcal{O}(n^9)$ for large n. Hence,

$$d_W(W, Z) = \mathcal{O}(n^{-1}).$$

This tells us that, for large n, the distribution of the standardized random variable W (the standardized number of complete graphs K_4 in a Bernoulli random graph $\mathcal{G}(n, p)$) converges to the standard normal distribution at the rate $\mathcal{O}(n^{-1})$.

If p does depend on n, then $\sigma^2 = \mathcal{O}(n^6 p^{11})$ for large n, and thus $\sigma^3 = \mathcal{O}(n^9 p^{33/2})$. Note that we use the approximation $1 - p \approx 1$ ($p \ll 1$) here, because real-world networks are often large and sparse, as discussed in the paper [7]. Hence

$$d_W(W, Z) = \mathcal{O}\left(\frac{1}{p^{21/2} n}\right).$$

This tells us that the distribution of the standardized random variable W converges to the standard normal distribution if $p = p(n) \gg n^{-2/21}$. For small p, a Poisson approximation may be better, which will be obtained by extending Stein's method to the Poisson distribution in Chapter 4.

Chapter 4

Poisson Approximation

4.1 Stein's Method for Poisson Approximation

One good thing about Stein's method is that it can be easily extended to many different distributions other than the normal distribution. In this chapter, we extend Stein's method to the Poisson distribution, one of the most important discrete probability distributions. The general setup and the framework of Stein's method for Poisson approximation are similar to those of Stein's method for normal approximation.

Lemma 4.1 (**Stein's lemma for Poisson approximation**). *Let $\mathbb{N}$ denote the set of all natural numbers $\{0, 1, 2, \cdots\}$. Let $f : \mathbb{N} \to \mathbb{R}$ be a function, and let W be a non-negative integer-valued random variable. For $\lambda > 0$, we define the functional operator $\mathcal{A}_P$ by*

$$\mathcal{A}_P f(k) = \lambda f(k+1) - k f(k).$$

Then, W has the Poisson distribution with mean λ if and only if, for all bounded f,

$$\mathbb{E}\mathcal{A}_P f(W) = 0 \quad (i.e.\ \mathbb{E}[\lambda f(W+1) - W f(W)] = 0).$$

The difference operator $\mathcal{A}_P$ is referred to as a characterizing operator of the Poisson distribution [18].

The heuristic behind Stein's method for Poisson approximation is that we can measure the distance between any random variable W and $Z \sim Poisson(\lambda)$ by

$$\mathbb{E}h(W) - \mathbb{E}h(Z)$$

over a class of test functions $\mathcal{H}$. For Poisson approximation, we take $h_B = \mathbb{I}_B \in \mathcal{H}_{TV}$ (we restrict $B \subseteq \mathbb{N}$, since we only consider non-negative integer-valued random variables), which gives

$$\mathbb{E}\mathbb{I}_B(W) - \mathcal{P}_\lambda(B), \tag{4.1}$$

where $\mathcal{P}_\lambda$ denotes the probability with respect to the Poisson distribution with mean λ. Our aim is to bound (4.1) for any $\mathbb{I}_B \in \mathcal{H}_{TV}$ ($B \subseteq \mathbb{N}$). By Stein's lemma for Poisson approximation, if $\mathbb{E}\mathcal{A}_P f(W)$ is close to zero for many functions f, then we would expect W to be close to Z in distribution. This intuition motivates the **Stein equation for Poisson approximation**:

$$\lambda f(k+1) - k f(k) = \mathbb{I}_B(k) - \mathcal{P}_\lambda(B). \tag{4.2}$$

Lemma 4.2. *Let $B \subseteq \mathbb{N}$. The unique solution of* (4.2) *with the boundary condition $f(0) = 0$ is given by*

$$f_B(k) = \lambda^{-k} e^{\lambda} (k-1)! [\mathcal{P}_\lambda(B \cap \mathbb{Z}_k) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\mathbb{Z}_k)], \tag{4.3}$$

where $\mathbb{Z}_k = \{0, 1, \cdots, k-1\}$ $(k \in \mathbb{N}\backslash\{0\})$ and $\mathbb{Z}_0 = \emptyset$. Moreover,

$$\|f_B\| \leqslant \min\left\{1, \lambda^{-1/2}\right\} \quad \text{and} \quad \|\Delta f_B\| \leqslant \frac{1-e^{-\lambda}}{\lambda} \leqslant \min\left\{1, \lambda^{-1}\right\},$$

where $\Delta f_B(k) = f_B(k+1) - f_B(k)$.

Proof. We first verify that (4.3) is a solution to (4.2). Since $\mathbb{Z}_0 = \emptyset$, we have $f_B(0) = 0$. Since

$$\begin{aligned}
\lambda f_B(k+1) &= \lambda \lambda^{-(k+1)} e^{\lambda} k! [\mathcal{P}_\lambda(B \cap \mathbb{Z}_{k+1}) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\mathbb{Z}_{k+1})] \\
&= \lambda^{-k} e^{\lambda} k! [\mathcal{P}_\lambda(B \cap (\mathbb{Z}_k \cup \{k\})) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\mathbb{Z}_k \cup \{k\})] \\
&= \lambda^{-k} e^{\lambda} k! [\mathcal{P}_\lambda((B \cap \mathbb{Z}_k) \cup (B \cap \{k\})) - \mathcal{P}_\lambda(B)(\mathcal{P}_\lambda(\mathbb{Z}_k) + \mathcal{P}_\lambda(\{k\}))] \\
&= \lambda^{-k} e^{\lambda} k! [\mathcal{P}_\lambda(B \cap \mathbb{Z}_k) + \mathcal{P}_\lambda(B \cap \{k\}) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\mathbb{Z}_k) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\{k\})]
\end{aligned}$$

and

$$k f_B(k) = \lambda^{-k} e^{\lambda} k! [\mathcal{P}_\lambda(B \cap \mathbb{Z}_k) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\mathbb{Z}_k)],$$

we have

$$\begin{aligned}
\lambda f_B(k+1) - k f_B(k) &= \lambda^{-k} e^{\lambda} k! [\mathcal{P}_\lambda(B \cap \{k\}) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\{k\})] \\
&= \lambda^{-k} e^{\lambda} k! [\mathbb{I}_B(k)\mathcal{P}_\lambda(\{k\}) - \mathcal{P}_\lambda(B)\mathcal{P}_\lambda(\{k\})] \\
&= \lambda^{-k} e^{\lambda} k! \mathcal{P}_\lambda(\{k\}) [\mathbb{I}_B(k) - \mathcal{P}_\lambda(B)] \\
&= \lambda^{-k} e^{\lambda} k! \frac{\lambda^k}{k!} e^{-\lambda} [\mathbb{I}_B(k) - \mathcal{P}_\lambda(B)] \\
&= \mathbb{I}_B(k) - \mathcal{P}_\lambda(B).
\end{aligned}$$

Now, we show that (4.3) is the unique solution. Suppose that u and v are two solutions to (4.2) with $u(0) = v(0) = 0$, then u and v satisfy

$$\lambda u(k+1) - k u(k) = \mathbb{I}_B(k) - \mathcal{P}_\lambda(B) \quad \text{and} \quad \lambda v(k+1) - k v(k) = \mathbb{I}_B(k) - \mathcal{P}_\lambda(B).$$

We define $w = u - v$, and thus w satisfies

$$\lambda w(k+1) - k w(k) = 0 \quad \text{and} \quad w(0) = u(0) - v(0) = 0.$$

We prove that $w(k) = 0$ for all $k \in \mathbb{N}$ by induction. The base case is $w(0) = 0$, and the inductive hypothesis is that $w(n) = 0$ $(n \in \mathbb{N})$. Then, in the inductive step, we have

$$\lambda w(n+1) - n w(n) = \lambda w(n+1) = 0.$$

Since $\lambda > 0$, we have $w(n+1) = 0$. Therefore, $w(k) = 0$ for all $k \in \mathbb{N}$, which implies $u(k) = v(k)$ for all $k \in \mathbb{N}$. Hence, we have shown uniqueness.

We refer to [18] (Lemma 4.4) for the proof of the bounds for f_B and Δf_B. The proof consists of elementary calculations, which is technical and irrelevant to the idea of Stein's method. Therefore, we omit it in this thesis. □

Proof of Lemma 4.1 (Stein's lemma for Poisson approximation)

Proof. ("$\Longrightarrow$") Suppose that W has the Poisson distribution with mean λ, then, for any bounded function f, we have

$$\begin{aligned}\lambda \mathbb{E} f(W+1) &= \lambda \sum_{k=0}^{\infty} f(k+1) \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=0}^{\infty} (k+1) f(k+1) \frac{\lambda^{k+1}}{(k+1)!} e^{-\lambda} \\ &= \sum_{n=1}^{\infty} n f(n) \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} n f(n) \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \mathbb{E}[W f(W)].\end{aligned}$$

("$\Longleftarrow$") Suppose that W is a non-negative integer-valued random variable such that $\mathbb{E}[\lambda f(W+1) - W f(W)] = 0$ for all bounded functions $f : \mathbb{N} \to \mathbb{R}$. Taking $B = \{n\}$ ($n \in \mathbb{N}$), by Lemma 4.2, the function $f_{\{n\}}$ satisfying (4.2) with the boundary condition $f(0) = 0$ is a bounded function. If we evaluate both sides of the equation

$$\lambda f_{\{n\}}(k+1) - k f_{\{n\}}(k) = \mathbb{I}_{\{n\}}(k) - \mathcal{P}_\lambda(\{n\})$$

at the random variable W and take expectations in both sides, we have

$$\mathbb{E}[\lambda f_{\{n\}}(W+1) - W f_{\{n\}}(W)] = \mathbb{E}[\mathbb{I}_{\{n\}}(W) - \mathcal{P}_\lambda(\{n\})], \quad \text{for all } n \in \mathbb{N}.$$

Since

$$\mathbb{E}[\lambda f_{\{n\}}(W+1) - W f_{\{n\}}(W)] = 0 \quad \text{and} \quad \mathbb{E}[\mathbb{I}_{\{n\}}(W) - \mathcal{P}_\lambda(\{n\})] = \mathbb{P}(W = n) - \frac{\lambda^n}{n!} e^{-\lambda},$$

we have

$$\mathbb{P}(W = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \text{for all } n \in \mathbb{N},$$

which implies that W has the Poisson distribution with mean λ. □

Corollary 4.3. *Suppose that f_B is the unique solution of* (4.2) *with the boundary condition $f(0) = 0$. Then, we have*

$$|\mathbb{P}(W \in B) - \mathcal{P}_\lambda(B)| = |\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]|, \tag{4.4}$$

for any non-negative integer-valued random variable W.

We note that the left-hand side of (4.4) is equal to the quantity

$$|\mathbb{E} h_B(W) - \mathbb{E} h_B(Z)| \quad (h_B = \mathbb{I}_B \in \mathcal{H}_{TV} \text{ and } Z \sim \mathit{Poisson}(\lambda))$$

that we want to bound.

4.2 Bounding the Error

Following Corollary 4.3, we can reduce the problem of bounding the distance between any random variable W and the Possion random variable $Z \sim Poisson(\lambda)$ to bounding

$$|\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]|.$$

Proposition 4.4. *Let the random variable Z have the Poisson distribution with mean λ. Then, for any non-negative integer-valued random variable W,*

$$d_{TV}(W,Z) = \sup_{h_B \in \mathcal{H}_{TV}} |\mathbb{E}h_B(W) - \mathbb{E}h_B(Z)| = \sup_{B \subseteq \mathbb{N}} |\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]|,$$

where f_B is the unique solution of (4.2) with the boundary condition $f(0) = 0$.

Analogous to the strategy used for normal approximation, we will find the structure in the random variable W of interest and use it to break down the term $Wf(W)$ in order to compare it with $\lambda f(W+1)$ for appropriate f [18].

4.3 Law of Small Numbers

It is well known that a binomial distribution $Binomial(n,p)$ (or the distribution of a sum of n independent and identically distributed Bernoulli random variables with success probability p) can be approximated well by a Poisson distribution $Poisson(\lambda)$ when $\lambda = np$ and p is extremely small, which is where the term *law of small numbers* comes from. We first prove the convergence of such an approximation in distribution.

Lemma 4.5. *For fixed $\lambda > 0$, we define random variables $W_n \sim Binomial(n, \lambda/n)$ and $Z \sim Poisson(\lambda)$. Then,*

$$\lim_{n\to\infty} \mathbb{P}(W_n = k) = \mathbb{P}(Z = k), \quad \text{for any } k \in \mathbb{N}.$$

Proof. For any $k \in \mathbb{N}$, we have

$$\begin{aligned}
\mathbb{P}(W_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda^k}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\left[\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\right]\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n}.
\end{aligned}$$

Since

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) = 1, \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1,$$

and

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda},$$

we have

$$\lim_{n\to\infty} \mathbb{P}(W_n = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \mathbb{P}(Z = k), \quad \text{for any } k \in \mathbb{N}.$$

□

We now generalize Lemma 4.5 and give a corresponding error bound in the total variation metric using Stein's method for Poisson approximation. We will show that sums of n independent Bernoulli random variables with different success probabilities can also be approximated well by Poisson distributions.

Theorem 4.6. *Let $X_1, \cdots, X_n$ be independent Bernoulli random variables with success probabilities $\mathbb{P}(X_i = 1) = p_i$. Let $W = \sum_{i=1}^n X_i$ and $\lambda = \mathbb{E}(W) = \sum_{i=1}^n p_i$. Suppose that the random variable Z has the Poisson distribution with mean λ, then*

$$d_{TV}(W, Z) \leqslant \min\{1, \lambda^{-1}\} \sum_{i=1}^n p_i^2 \leqslant \min\{1, \lambda\} \max_{1\leqslant i\leqslant n} p_i.$$

Proof. Let f_B be the unique solution of (4.2) with the boundary condition $f(0) = 0$. Define $W_i = W - X_i$, and thus W_i and X_i are independent. Then

$$\begin{aligned}
\mathbb{E}[W f_B(W)] &= \sum_{i=1}^n \mathbb{E}[X_i f_B(W)] \\
&= \sum_{i=1}^n \mathbb{E}[X_i f_B(W)|X_i = 1]\mathbb{P}(X_i = 1) + \mathbb{E}[X_i f_B(W)|X_i = 0]\mathbb{P}(X_i = 0) \\
&= \sum_{i=1}^n p_i \mathbb{E}[X_i f_B(W)|X_i = 1] \\
&= \sum_{i=1}^n p_i \mathbb{E}[X_i f_B(W_i + X_i)|X_i = 1] \\
&= \sum_{i=1}^n p_i \mathbb{E}[f_B(W_i + 1)|X_i = 1] \\
&= \sum_{i=1}^n p_i \mathbb{E}[f_B(W_i + 1)],
\end{aligned}$$

where the second equality follows by law of total expectation, and the final equality follows since W_i and X_i are independent. Therefore,

$$\begin{aligned}
|\mathbb{E}[\lambda f_B(W + 1) - W f_B(W)]| &= \left| \sum_{i=1}^n p_i \mathbb{E}[f_B(W + 1) - f_B(W_i + 1)] \right| \\
&\leqslant \sum_{i=1}^n p_i \mathbb{E}|f_B(W + 1) - f_B(W_i + 1)|.
\end{aligned}$$

But

$$\begin{aligned}
|f_B(W + 1) - f_B(W_i + 1)| &= |f_B(W_i + X_i + 1) - f_B(W_i + 1)| \\
&= |f_B(W_i + 2) - f_B(W_i + 1)|X_i \\
&= |\Delta f_B(W_i + 1)|X_i \\
&\leqslant \|\Delta f_B\| X_i,
\end{aligned}$$

where the first equality follows since $W = W_i + X_i$, and the second equality follows since $X_i \in \{0,1\}$. Then, we have

$$\begin{aligned}
|\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]| &\leqslant \|\Delta f_B\| \sum_{i=1}^{n} p_i \mathbb{E}(X_i) \\
&\leqslant \min\{1, \lambda^{-1}\} \sum_{i=1}^{n} p_i \mathbb{E}(X_i) \\
&= \min\{1, \lambda^{-1}\} \sum_{i=1}^{n} p_i^2 \\
&\leqslant \min\{1, \lambda^{-1}\} \left(\sum_{i=1}^{n} p_i \right) \max_{1 \leqslant j \leqslant n} p_j \\
&= \min\{1, \lambda^{-1}\} \lambda \max_{1 \leqslant j \leqslant n} p_j \\
&= \min\{1, \lambda\} \max_{1 \leqslant j \leqslant n} p_j, \quad \text{for any } B \subseteq \mathbb{N},
\end{aligned}$$

where the second inequality follows by Lemma 4.2.

Hence, by Proposition 4.4, we have

$$d_{TV}(W, Z) = \sup_{B \subseteq \mathbb{N}} |\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]| \leqslant \min\{1, \lambda^{-1}\} \sum_{i=1}^{n} p_i^2 \leqslant \min\{1, \lambda\} \max_{1 \leqslant j \leqslant n} p_j.$$

□

Remark 4.7. If X_i are independent and identically distributed Bernoulli random variables with success probability p, then $W_n = \sum_{i=1}^{n} X_i \sim$ *Binomial*(n, p) and $\lambda = \mathbb{E}(W_n) = np$. Hence, by Theorem 4.6, an error bound for Poisson approximation of the binomial distribution (Lemma 4.5, law of small numbers) in the total variation metric is given by

$$d_{TV}(W_n, Z) \leqslant \min\{1, \lambda\} p = \frac{\min\{1, \lambda\}\lambda}{n},$$

of which the rate of convergence is $\mathcal{O}(n^{-1})$.

4.4 Extension to Local Dependence

We go a step further and generalize Theorem 4.6 to sums of locally dependent Bernoulli random variables with different success probabilities.

Theorem 4.8. *Let $X_1, \cdots, X_n$ be Bernoulli random variables with success probabilities $\mathbb{P}(X_i = 1) = p_i$. Let $W = \sum_{i=1}^{n} X_i$, $\lambda = \mathbb{E}(W) = \sum_{i=1}^{n} p_i$, and $p_{ij} = \mathbb{E}(X_i X_j)$. For $i = 1, \cdots, n$, let D_i be the dependency neighbourhood of X_i. Suppose that the random variable Z has the Poisson distribution with mean λ, then*

$$d_{TV}(W, Z) \leqslant \min\{1, \lambda^{-1}\} \left(\sum_{i=1}^{n} \sum_{j \in D_i} p_i p_j + \sum_{i=1}^{n} \sum_{j \in D_i \setminus \{i\}} p_{ij} \right).$$

Proof. Let f_B be the unique solution of (4.2) with the boundary condition $f(0) = 0$. Define $W_i = W - X_i$. Define $V_i = \sum_{j \notin D_i} X_j$, and thus V_i and X_i are independent. Since $X_i f(W) = X_i f(W_i + 1)$, we have

$$\begin{aligned}
&\mathbb{E}[\lambda f_B(W+1) - W f_B(W)] \\
&= \mathbb{E}[\lambda f_B(W+1) - \lambda f_B(W_i+1) + \lambda f_B(W_i+1) - W f_B(W)] \\
&= \lambda \mathbb{E}[f_B(W+1) - f_B(W_i+1)] + \mathbb{E}[\lambda f_B(W_i+1) - \sum_{i=1}^{n} X_i f_B(W_i+1)] \\
&= \sum_{i=1}^{n} p_i \mathbb{E}[f_B(W+1) - f_B(W_i+1)] + \sum_{i=1}^{n} \mathbb{E}[(p_i - X_i) f_B(W_i+1)].
\end{aligned}$$

We find that

$$\sum_{i=1}^{n} \mathbb{E}[(p_i - X_i) f_B(W_i+1)] = \sum_{i=1}^{n} \mathbb{E}[(p_i - X_i)(f_B(W_i+1) - f_B(V_i+1))],$$

since V_i and X_i being independent implies that

$$\mathbb{E}[(p_i - X_i) f_B(V_i+1)] = \mathbb{E}(p_i - X_i) \mathbb{E} f_B(V_i+1) = 0.$$

We found, in the proof of Theorem 4.6, that

$$|f_B(W+1) - f_B(W_i+1)| \leqslant \|\Delta f_B\| X_i.$$

We deal with the other term $|f_B(W_i+1) - f_B(V_i+1)|$ using a similar treatment. Suppose $D_i \backslash \{i\} = \{l_1^{(i)}, l_2^{(i)}, \cdots, l_{n_i}^{(i)}\}$ ($n_i = 0$ if $D_i \backslash \{i\} = \emptyset$). The reason why we can enumerate the elements in $D_i \backslash \{i\}$ is that there are only finitely many elements in it. Then, we have

$$\begin{aligned}
&|f_B(W_i+1) - f_B(V_i+1)| \\
&= \left| f_B\left(V_i + \sum_{j \in D_i \backslash \{i\}} X_j + 1\right) - f_B(V_i+1) \right| \\
&= \left| f_B\left(V_i + \sum_{j=1}^{n_i} X_{l_j^{(i)}} + 1\right) - f_B(V_i+1) \right| \\
&= \left| \sum_{k=1}^{n_i} \left[f_B\left(V_i + \sum_{j=1}^{n_i-k+1} X_{l_j^{(i)}} + 1\right) - f_B\left(V_i + \sum_{j=1}^{n_i-k} X_{l_j^{(i)}} + 1\right)\right] \right| \\
&= \left| \sum_{k=1}^{n_i} \left[f_B\left(V_i + \sum_{j=1}^{n_i-k} X_{l_j^{(i)}} + 2\right) - f_B\left(V_i + \sum_{j=1}^{n_i-k} X_{l_j^{(i)}} + 1\right)\right] X_{l_{n_i-k+1}^{(i)}} \right| \\
&\leqslant \sum_{k=1}^{n_i} \left| \Delta f_B\left(V_i + \sum_{j=1}^{n_i-k} X_{l_j^{(i)}} + 1\right) \right| X_{l_{n_i-k+1}^{(i)}} \\
&\leqslant \|\Delta f_B\| \sum_{k=1}^{n_i} X_{l_{n_i-k+1}^{(i)}} \\
&= \|\Delta f_B\| \sum_{j \in D_i \backslash \{i\}} X_j,
\end{aligned}$$

where $\sum_{j=1}^{0}$ is defined to be empty sum, and the first equality follows since

$$W_i = W - X_i = V_i + \sum_{j \in D_i} X_j - X_i = V_i + \sum_{j \in D_i \setminus \{i\}} X_j.$$

Therefore,

$$\begin{aligned}
&|\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]| \\
\leqslant &\left| \sum_{i=1}^{n} p_i \mathbb{E}[f_B(W+1) - f_B(W_i+1)] \right| + \left| \sum_{i=1}^{n} \mathbb{E}[(p_i - X_i)(f_B(W_i+1) - f_B(V_i+1))] \right| \\
\leqslant &\|\Delta f_B\| \sum_{i=1}^{n} p_i \mathbb{E}(X_i) + \|\Delta f_B\| \sum_{i=1}^{n} \mathbb{E}\left[|p_i - X_i| \sum_{j \in D_i \setminus \{i\}} X_j \right] \\
\leqslant &\|\Delta f_B\| \left(\sum_{i=1}^{n} p_i \mathbb{E}(X_i) + \sum_{i=1}^{n} \mathbb{E}\left[(p_i + X_i) \sum_{j \in D_i \setminus \{i\}} X_j \right] \right) \\
\leqslant &\min\{1, \lambda^{-1}\} \left(\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} \sum_{j \in D_i \setminus \{i\}} (p_i p_j + p_{ij}) \right) \\
= &\min\{1, \lambda^{-1}\} \left(\sum_{i=1}^{n} \sum_{j \in D_i} p_i p_j + \sum_{i=1}^{n} \sum_{j \in D_i \setminus \{i\}} p_{ij} \right), \quad \text{for any } B \subseteq \mathbb{N},
\end{aligned}$$

where the third inequality follows since $|p_i - X_i| \leqslant p_i + X_i$, by triangle inequality.

Hence, by Proposition 4.4, we have

$$\begin{aligned}
d_{TV}(W, Z) &= \sup_{B \subseteq \mathbb{N}} |\mathbb{E}[\lambda f_B(W+1) - W f_B(W)]| \\
&\leqslant \min\{1, \lambda^{-1}\} \left(\sum_{i=1}^{n} \sum_{j \in D_i} p_i p_j + \sum_{i=1}^{n} \sum_{j \in D_i \setminus \{i\}} p_{ij} \right).
\end{aligned}$$

□

Remark 4.9. Consider the independent case discussed in the previous section, where $D_i = \{i\}$ for each $i = 1, \cdots, n$. Since $D_i \setminus \{i\} = \emptyset$, we have

$$d_{TV}(W, Z) \leqslant \min\{1, \lambda^{-1}\} \sum_{i=1}^{n} p_i^2,$$

which is exactly the same as the result in Theorem 4.6.

4.5 Revisiting the Example of Complete Graphs K_4 Counts

In Section 3.5, we approximated the distribution of the number of complete graphs K_4 in a Bernoulli random graph using Stein's method for normal approximation. We saw that such an approximation failed when p was small and depended on the number of vertices

n in the graph. However, this is often the case in real-world problems, because most real-world networks are large and sparse, as discussed in the paper [7]. In fact, the number of edges in a real-world graph is often at most of order n [7], whereas there are $\binom{n}{2} \approx n^2/2$ possible places in total where edges could be formed.

In this section, we apply Stein's method for Poisson approximation to this problem, which allows us to obtain a valid approximation for small p which depends on n.

We recall that there are $N = \binom{n}{4} = \mathcal{O}(n^4)$ places a complete graph K_4 can occur in a Bernoulli random graph $G = \mathcal{G}(n,p)$. The random variable $T = \sum_{i=1}^{N} Y_i$ denotes the number of complete graphs K_4 occurring in G, where $Y_i \sim Bernoulli(p^6)$ are the indicator random variables as defined in Section 3.5. Then, $\mathbb{E}(T) = Np^6$.

Let A_i and B_i be the sets as defined in Section 3.5. Then, the dependency neighbourhoods are $D_i = A_i \cup B_i \cup \{i\}$. We obtained, in Section 3.5, that, for each i, A_i and B_i are disjoint, $i \notin A_i, B_i$, and

$$|A_i| = 3(n-4)(n-3) \quad \text{and} \quad |B_i| = 4(n-4).$$

Also, by the calculations in Section 3.5, we have:

- if $j \in A_i$, then $\mathbb{E}(Y_iY_j) = p^{11}$;
- if $j \in B_i$, then $\mathbb{E}(Y_iY_j) = p^9$.

Now, let Z be a random variable that has the Poisson distribution with mean $\lambda = Np^6$. By Theorem 4.8, we have

$$\begin{aligned}
d_{TV}(T,Z) &\leqslant \min\{1,\lambda^{-1}\}\left(\sum_{i=1}^{N}\sum_{j\in D_i} p_ip_j + \sum_{i=1}^{N}\sum_{j\in D_i\setminus\{i\}} p_{ij}\right) \\
&= \min\{1,\lambda^{-1}\}\sum_{i=1}^{N}\left(\sum_{j\in D_i}\mathbb{E}(Y_i)\mathbb{E}(Y_j) + \left(\sum_{j\in A_i}\mathbb{E}(Y_iY_j) + \sum_{j\in B_i}\mathbb{E}(Y_iY_j)\right)\right) \\
&= \min\{1,\lambda^{-1}\}\sum_{i=1}^{N}\left(\sum_{j\in D_i} p^{12} + \sum_{j\in A_i} p^{11} + \sum_{j\in B_i} p^9\right) \\
&= \min\{1,\lambda^{-1}\}\sum_{i=1}^{N}[(|A_i|+|B_i|+1)p^{12} + |A_i|p^{11} + |B_i|p^9] \\
&= \min\{1,\lambda^{-1}\}p^6\sum_{i=1}^{N}[p^6 + |A_i|p^5(1+p) + |B_i|p^3(1+p^3)] \\
&= \min\{1,\lambda^{-1}\}p^6\sum_{i=1}^{N}[p^6 + 3(n-4)(n-3)p^5(1+p) + 4(n-4)p^3(1+p^3)] \\
&= \min\{1,\lambda^{-1}\}Np^6[p^6 + 3(n-4)(n-3)p^5(1+p) + 4(n-4)p^3(1+p^3)] \\
&= \min\{1,\lambda^{-1}\}\lambda[p^6 + 3(n-4)(n-3)p^5(1+p) + 4(n-4)p^3(1+p^3)] \\
&= \min\{1,\lambda\}[p^6 + 3(n-4)(n-3)p^5(1+p) + 4(n-4)p^3(1+p^3)].
\end{aligned}$$

We consider the case where p is small and depends on n. Since $\lambda = \mathcal{O}(n^4p^6)$, we can view the error bound as an asymptotic result with $p = \mathcal{O}(n^{-2/3})$, which gives

$$d_{TV}(T,Z) = \mathcal{O}(n^{-1}).$$

	Normal approximation	Poisson approximation
p is independent of n	$\mathcal{O}(n^{-1})$	Not applicable
$p = p(n)$ depends on n	$\mathcal{O}(p^{-21/2}n^{-1})$ (valid if $p \gg n^{-2/21}$)	$\mathcal{O}(n^{-1})$ (valid if $p \leqslant n^{-2/3}$)

Table 4.1: Normal and Poisson approximations for the K_4 counting problem in Bernoulli random graphs $\mathcal{G}(n, p)$.

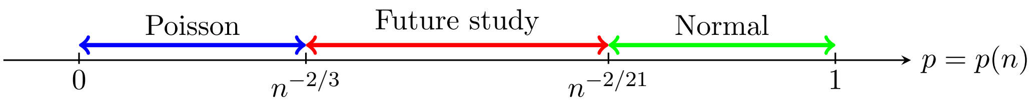


Figure 4.1: Validity of normal and Poisson approximations of K_4 counts in Bernoulli random graph when p depends on n.

This tells us that the distribution of the random variable T (the number of complete graphs K_4 in a Bernoulli random graph $\mathcal{G}(n, p)$) converges to the Poisson distribution with mean Np^6 at the rate $\mathcal{O}(n^{-1})$.

In Table 4.1, we summarize the validity and rates of convergence of normal and Poisson approximations for the K_4 counting problem in Bernoulli random graphs. A normal approximation is valid when the edge probability p does not depend on the number of vertices n. The rate of convergence of such an approximation is $\mathcal{O}(n^{-1})$. When p depends on n, a normal approximation is valid if $p \gg n^{-2/21}$ and a Poisson approximation is valid if $p \leqslant n^{-2/3}$ (see Figure 4.1), of which the rates of convergence are $\mathcal{O}(p^{-21/2}n^{-1})$ and $\mathcal{O}(n^{-1})$, respectively.

Chapter 5

Conclusion

Stein's method provides a powerful framework for showing distributional convergence with automatic error bounds with respect to integral probability metrics. In this thesis, we developed the essential ingredients of Stein's method (Stein's lemma and the Stein equation) for normal and Poisson approximations. Using the Stein equation, we converted the distance between the random variable of interest and the target random variable to the form of the expectation of some functional of the random variable of interest, which was then bounded using the properties of the functional and the structure of the random variable of interest. We applied this strategy to obtain error bounds for the central limit theorem and law of small numbers. We also generalized such results to sums of non-identically distributed random variables with dependence structures.

Counting K_4 in Bernoulli random graphs. We applied Stein's method to obtain normal and Poisson approximations of the distribution of the number of complete graphs K_4 in a Bernoulli random graph $\mathcal{G}(n,p)$. We saw that a normal approximation was valid when p did not depend on n. However, p usually depends on n in real-world problems, in which case a normal approximation is valid if $p \gg n^{-2/21}$ and a Poisson approximation is valid if $p \leqslant n^{-2/3}$ (see Figure 4.1). If we had more time, we could have improved our approximation results to eliminate the gap between $n^{-2/3}$ and $n^{-2/21}$.

Subgraph counts in Bernoulli random graphs. If time permits, we could generalize the results above to obtain Poisson approximations for counting the number of any type of strictly balanced subgraphs of K_n in Bernoulli random graphs. Furthermore, we could adapt Stein's method to the compound Poisson distribution, which would allow us to count the number of copies of a balanced graph in Bernoulli random graphs. These are discussed in the paper [20]. The concept of balanced graphs originated in the paper [9]. An example of a balanced but not strictly balanced graph is the whisker graph (a graph on four vertices which consists of a triangle with an edge attached to one of its three vertices).

Beyond the Erdős–Rényi model. The Erdős–Rényi model is a beautiful mathematical subject in itself. However, such a model is too simple and ideal for real-world problems, as discussed in [1] (Section 3.5). One drawback of the model is that it fails to capture the degree distribution of real-world networks, since most vertices in a Bernoulli random graph have comparable degrees, while there are many outliers (highly popular/unpopular individuals) observed in real-world networks. If time permits, we could apply Stein's method to obtain approximation results of subgraph counts in more realistic random graph models.

Bibliography

[1] A.L. Barabási, *Network science*, Cambridge university press, 2016.

[2] A.D. Barbour, *Poisson convergence and random graphs*, Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 92, No. 2, Cambridge University Press, 1982.

[3] A.C. Berry, *The Accuracy of the Gaussian Approximation to the Sum of Independent Variates*, Transactions of the American Mathematical Society, Vol. 49, No. 1, pp. 122-136, 1941.

[4] S. Chatterjee, J. Fulman, A. Rollin, *Exponential approximation by Stein's method and spectral graph theory*, ALEA Lat. Am. J. Probab. Math. Stat., 8:197–223, 2011.

[5] L.H.Y. Chen, *Poisson approximation for dependent trials*, The Annals of Probability, Vol. 3, No. 3, pp. 534-545, 1975.

[6] L.H.Y. Chen, L. Goldstein, Q-M. Shao, *Normal approximation by Stein's method*, Probability and its Applications (New York), Springer, Heidelberg, 2011.

[7] F. Chung, *Graph theory in the information age*, Notices of the AMS, Vol. 57, No. 6, pp. 726-732, 2010.

[8] P. Erdős, A. Rényi, *On Random Graphs. I*, Publicationes Mathematicae. 6, pp. 290–297, 1959.

[9] P. Erdős, A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci 5, No. 1, pp. 17-60, 1960.

[10] C-G Esseen, *Fourier analysis of distribution functions, a mathematical study of the Laplace-Gaussian law*, Acta Math. 77, 1945.

[11] R. Gaunt, *An introduction to Stein's method*, lecture notes, 2014.

[12] R. Gaunt, *Variance-Gamma approximation via Stein's method*, Electron. J. Probab. 19 (2014), no. 38, pp. 1–33, 2014.

[13] J. Gorham, L. Mackey, *Measuring Sample Quality with Stein's Method*, Advances in Neural Information Processing Systems (NIPS) 28, pp. 226-234, 2015.

[14] D. Jungnickel, *Graphs, Networks and Algorithms (second ed.)*, Springer, Berlin, Heidelberg, 2005.

[15] Q. Liu, D. Wang, *Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm*, Advances in Neural Information Processing Systems (NIPS) 29, pp. 2378-2386, 2016.

[16] A. Müller, *Integral Probability Metrics and Their Generating Classes of Functions*, Advances in Applied Probability, Vol. 29, No. 2 (Jun., 1997), pp. 429-443, 1997.

[17] E. Peköz, A. Röllin, N. Ross, *Total variation error bounds for geometric approximation*, Bernoulli, Vol. 19, No. 2 (2013), 610-632, 2013.

[18] N. Ross, *Fundamentals of Stein's method*, Probab. Surv 8, pp. 210-293, 2011.

[19] H.L. Royden, *Real Analysis (third ed.)*, Collier Macmillan, ISBN 0024041513, 1988.

[20] D. Stark, *Compound Poisson approximations of subgraph counts in random graphs*, Random Struct. Algorithms, Vol. 18 (Jan. 2001), pp. 39-60, 2001.

[21] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 583–602, Berkeley, Calif., 1972.

[22] T. Tao, *An Introduction to Measure Theory*, American Mathematical Society, ISBN 0821869191, 2011.