

## Stein's method for normal approximation

Louis H. Y. Chen and Qi-Man Shao

*Institute for Mathematical Sciences, National University of Singapore*

*3 Prince George's Park, Singapore 118402*

*E-mail: lhychen@ims.nus.edu.sg*

*and*

*Department of Statistics and Applied Probability*

*National University of Singapore*

*6 Science Drive 2, Singapore 117543;*

*Department of Mathematics, University of Oregon*

*Eugene, OR 97403, USA*

*E-mail: qmshao@darkwing.uoregon.edu*

Stein's method originated in 1972 in a paper in the Proceedings of the Sixth Berkeley Symposium. In that paper, he introduced the method in order to determine the accuracy of the normal approximation to the distribution of a sum of dependent random variables satisfying a mixing condition. Since then, many developments have taken place, both in extending the method beyond normal approximation and in applying the method to problems in other areas. In these lecture notes, we focus on univariate normal approximation, with our main emphasis on his approach exploiting an *a priori* estimate of the concentration function. We begin with a general explanation of Stein's method as applied to the normal distribution. We then go on to consider expectations of smooth functions, first for sums of independent and locally dependent random variables, and then in the more general setting of exchangeable pairs. The later sections are devoted to the use of concentration inequalities, in obtaining both uniform and non-uniform Berry–Esseen bounds for independent random variables. A number of applications are also discussed.

**Contents**

1	Introduction	2
2	Fundamentals of Stein’s method	8
2.1	Characterization	8
2.2	Properties of the solutions	10
2.3	Construction of the Stein identities	11
3	Normal approximation for smooth functions	13
3.1	Independent random variables	14
3.2	Locally dependent random variables	17
3.3	Exchangeable pairs	19
4	Uniform Berry–Esseen bounds: the bounded case	23
4.1	Independent random variables	23
4.2	Binary expansion of a random integer	28
5	Uniform Berry–Esseen bounds: the independent case	31
5.1	The concentration inequality approach	31
5.2	Proving the Berry–Esseen theorem	34
5.3	A lower bound	35
6	Non-uniform Berry–Esseen bounds: the independent case	39
6.1	A non-uniform concentration inequality	39
6.2	The final result	44
7	Uniform and non-uniform bounds under local dependence	48
8	Appendix	53
	References	58

**1. Introduction**

Stein’s method is a way of deriving explicit estimates of the accuracy of the approximation of one probability distribution by another. This is accomplished by comparing expectations, as indicated in the title of Stein’s (1986) monograph, *Approximate computation of expectations*. An upper bound is computed for the difference between the expectations of any one of a (large) family of test functions under the two distributions, each family of test functions determining an associated metric. Any such bound in turn implies a corresponding upper bound for the distance between the two distributions, measured with respect to the associated metric.

Thus, if the family of test functions consists of the indicators of all (measurable) subsets, then accuracy is expressed in terms of the total variation

distance  $d_{TV}$  between the two distributions:

$$d_{TV}(P, Q) := \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right| = \sup_A |P(A) - Q(A)|,$$

where  $\mathcal{H} = \{\mathbf{1}_A; A \text{ measurable}\}$ . If the distributions are on  $\mathbb{R}$ , and the test functions are the indicators of all half-lines, then accuracy is expressed using Kolmogorov distance, as is customary in Berry–Esseen theorems:

$$d_K(P, Q) := \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right| = \sup_{z \in \mathbb{R}} |P(-\infty, z] - Q(-\infty, z]|,$$

where  $\mathcal{H} = \{\mathbf{1}_{(-\infty, z]}; z \in \mathbb{R}\}$ . If the test functions consist of all uniformly Lipschitz functions  $h$  with constant bounded by 1, then the Wasserstein distance results:

$$d_W(P, Q) := \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right|,$$

where  $\mathcal{H} = \{h: \mathbb{R} \rightarrow \mathbb{R}; \|h'\| \leq 1\} =: \text{Lip}(1)$ , and where, for a function  $g: \mathbb{R} \rightarrow \mathbb{R}$ ,  $\|g\|$  denotes  $\sup_{x \in \mathbb{R}} |g(x)|$ . If the test functions are uniformly bounded and uniformly Lipschitz, bounded Wasserstein distances result:

$$d_{BW(k)}(P, Q) := \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right|,$$

where  $\mathcal{H} = \{h: \mathbb{R} \rightarrow \mathbb{R}; \|h\| \leq 1, \|h'\| \leq k\}$ . Stein's method applies to all of these distances, and to many more.

For normal approximation on  $\mathbb{R}$ , Stein began with the observation that

$$\mathbb{E}\{f'(Z) - Zf(Z)\} = 0 \tag{1.1}$$

for *any* bounded function  $f$  with bounded derivative, if  $Z$  has the standard normal distribution  $\mathcal{N}(0, 1)$ . This can be verified by partial integration:

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

However, the same partial integration can also be used to solve the differential equation

$$f'(x) - xf(x) = g(x), \quad \lim_{x \downarrow -\infty} f(x) e^{-x^2/2} = 0, \tag{1.2}$$

for  $g$  any bounded function, giving

$$\begin{aligned} \int_{-\infty}^y g(x)e^{-x^2/2} dx &= \int_{-\infty}^y \{f'(x) - xf(x)\}e^{-x^2/2} dx \\ &= \int_{-\infty}^y \frac{d}{dx}\{f(x)e^{-x^2/2}\} dx \\ &= f(y)e^{-y^2/2}, \end{aligned}$$

and hence

$$f(y) = e^{y^2/2} \int_{-\infty}^y g(x)e^{-x^2/2} dx.$$

Note that this  $f$  actually satisfies  $\lim_{y \downarrow -\infty} f(y) = 0$ , because

$$\Phi(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx \sim (y\sqrt{2\pi})^{-1} e^{-y^2/2}$$

as  $y \downarrow -\infty$ , and that  $f$  is bounded (and then, by the same argument, has  $\lim_{y \uparrow \infty} f(y) = 0$ ) if and only if  $\int_{-\infty}^{\infty} g(x)e^{-x^2/2} dx = 0$ , or, equivalently, if  $\mathbb{E}g(Z) = 0$ .

Hence, taking any bounded function  $h$ , we observe that the function  $f_h$ , defined by

$$f_h(x) := e^{x^2/2} \int_{-\infty}^x \{h(t) - \mathbb{E}h(Z)\}e^{-t^2/2} dt, \quad (1.3)$$

satisfies (1.2) for  $g(x) = h(x) - \mathbb{E}h(Z)$ ; substituting any random variable  $W$  for  $x$  in (1.2) and taking expectations then yields

$$\mathbb{E}\{f'_h(W) - Wf_h(W)\} = \mathbb{E}h(W) - \mathbb{E}h(Z). \quad (1.4)$$

Thus the characterization (1.1) of the standard normal distribution also delivers an upper bound for normal approximation with respect to any of the distances introduced above: for any class  $\mathcal{H}$  of (bounded) test functions  $h$ ,

$$\sup_{h \in \mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| = \sup_{h \in \mathcal{H}} |\mathbb{E}\{f'_h(W) - Wf_h(W)\}|. \quad (1.5)$$

The curious fact is that the supremum on the right hand side of (1.5), which contains only the random variable  $W$ , is frequently much simpler to bound than that on the left hand side. The differential characterization (1.1) of the normal distribution is reflected in the fact that the quantity  $\mathbb{E}\{f'(W) - Wf(W)\}$  is often relatively easily shown to be small, when the structure of  $W$  is such as to make normal approximation seem plausible; for instance, when  $W$  is a sum of individually small and only weakly dependent

random variables. This is illustrated in the following elementary argument for the classical central limit theorem.

Suppose that  $X_1, X_2, \dots, X_n$  are independent and identically distributed random variables, with mean zero and unit variance, and such that  $\mathbb{E}|X_1|^3 < \infty$ ; let  $W = n^{-1/2} \sum_{j=1}^n X_j$ . We evaluate the quantity  $\mathbb{E}\{f'(W) - Wf(W)\}$  for a smooth and very well-behaved function  $f$ . Since  $W$  is a sum of identically distributed components, we have

$$\mathbb{E}\{Wf(W)\} = n\mathbb{E}\{n^{-1/2}X_1f(W)\},$$

and  $W = n^{-1/2}X_1 + W_1$ , where  $W_1 = n^{-1/2} \sum_{j=2}^n X_j$  is independent of  $X_1$ . Hence, by Taylor's expansion,

$$\begin{aligned} \mathbb{E}\{Wf(W)\} &= n^{1/2}\mathbb{E}\{X_1f(W_1 + n^{-1/2}X_1)\} \\ &= n^{1/2}\mathbb{E}\{X_1(f(W_1) + n^{-1/2}X_1f'(W_1))\} + \eta_1, \end{aligned} \quad (1.6)$$

where  $|\eta_1| \leq \frac{1}{2}n^{-1/2}\mathbb{E}|X_1|^3\|f''\|$ . On the other hand, again by Taylor's expansion,

$$\mathbb{E}\{f'(W) - \mathbb{E}f'(W_1 + n^{-1/2}X_1)\} = \mathbb{E}\{f'(W) - \mathbb{E}f'(W_1)\} + \eta_2, \quad (1.7)$$

where  $|\eta_2| \leq n^{-1/2}\mathbb{E}|X_1|\|f''\|$ . But now, since  $\mathbb{E}X_1 = 0$  and  $\mathbb{E}X_1^2 = 1$ , and since  $X_1$  and  $W_1$  are independent, it follows that

$$|\mathbb{E}\{f'(W) - Wf(W)\}| \leq n^{-1/2}\{1 + \frac{1}{2}\mathbb{E}|X_1|^3\}\|f''\|,$$

for any  $f$  with bounded second derivative. Hence, if  $\mathcal{H}$  is any class of test functions and

$$C_{\mathcal{H}} := \sup_{h \in \mathcal{H}} \|f_h''\|,$$

then it follows that

$$\sup_{h \in \mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq C_{\mathcal{H}} n^{-1/2} \{1 + \frac{1}{2}\mathbb{E}|X_1|^3\}. \quad (1.8)$$

The inequality (1.8) thus establishes an accuracy of order  $O(n^{-1/2})$  for the normal approximation of  $W$ , with respect to the distance  $d_{\mathcal{H}}$  on probability measures on  $\mathbb{R}$  defined by

$$d_{\mathcal{H}}(P, Q) = \sup_{h \in \mathcal{H}} \left| \int h dP - \int h dQ \right|,$$

provided only that  $C_{\mathcal{H}} < \infty$ . For example, as observed by Erickson (1974), the class of test functions  $h$  for the Wasserstein distance  $d_W$  is just the Lipschitz functions  $\text{Lip}(1)$  with constant no greater than 1, and, for this class,  $C_{\mathcal{H}} = 2$ : see (2.13) of Lemma 2.3. The distinction between the bounds

for different distances is then reflected in the differences between the values of  $C_{\mathcal{H}}$ . Computing good estimates of  $C_{\mathcal{H}}$  involves studying the properties of the solutions  $f_h$  to the Stein equation given in (1.3) for  $h \in \mathcal{H}$ . Fortunately, for normal approximation, a lot is known.

It sadly turns out that this simple argument is not enough to prove the Berry–Esseen theorem, which states that

$$\sup_{z \in \mathbf{R}} |\mathbb{P}[W \leq z] - \Phi(z)| \leq Cn^{-1/2} \mathbb{E}|X_1|^3,$$

for a universal constant  $C$ . This is because, for  $\mathcal{H}$  the set of indicators of half-lines,  $C_{\mathcal{H}} = \infty$ . However, it is possible to refine the argument, by making less crude estimates of  $|\eta_1|$  and  $|\eta_2|$ . Broadly speaking, if  $h = \mathbf{1}_{(-\infty, z]}$ , and writing  $f_z$  for the corresponding solution  $f_h$ , then  $f'_z$  is smooth, except for a jump discontinuity at  $z$ , as can be seen from (1.2). Hence, taking  $|\eta_2|$  for illustration,

$$\begin{aligned} & |f'_z(W_1 + n^{-1/2}X_1) - f'_z(W_1)| \\ & \leq n^{-1/2}|X_1| M(z, f_z)(|W_1| + n^{-1/2}|X_1| + 1), \end{aligned} \quad (1.9)$$

where  $M(z, f) := \sup_{x \neq z} \{|f''(x)|/(|x| + 1)\}$ , provided that  $z$  does not lie between  $W_1 + n^{-1/2}X_1$  and  $W_1$ . If it does, the difference is still bounded by  $2\|f'_z\|$ . Hence, since both  $\sup_{z \in \mathbf{R}} \|f'_z\|$  and  $\sup_{z \in \mathbf{R}} M(z, f_z)$  are finite, explicit bounds of order  $O(n^{-1/2})$  can still be deduced for the Berry–Esseen theorem, provided that it can be shown that

$$\sup_{z \in \mathbf{R}} \mathbb{P}(W_1 \in [z, z + h]) \leq C(h + n^{-1/2} \mathbb{E}|X_1|^3) \quad (1.10)$$

for some constant  $C$  (the corresponding estimates for  $|\eta_1|$  are a little more involved, but the argument can be carried through analogously). Inequalities of this form are precisely the concentration inequalities that play a central part in these notes, and the above discussion illustrates their importance.

If the random variables  $X_i$  are dependent, these simple arguments are no longer valid. However, they can be adapted to work effectively, if less cleanly, for many weak dependence structures. For instance, if individual summands have little influence on  $W$ , as is typical when a normal approximation is expected to be good, then it may be possible to decompose  $W$  in the form

$$W = n^{-1/2}(X_i + U_i) + W_i,$$

where  $W_i$  is ‘almost’ independent of  $X_i$ , and  $|U_i|$  is not too ‘large’, enabling Taylor expansions analogous to (1.6) and (1.7) to be attempted.

However, especially if Kolmogorov distance is to be considered, it pays to be rather less brutal, making direct use of (1.4), and rewriting the errors in exact, integral form; this will appear time and again in what follows. Better bounds can be achieved as a result, and concentration arguments also become feasible: note that finding sharp bounds for a probability such as  $\mathbb{P}(a \leq W_i \leq a + n^{-1/2}(X_i + U_i))$  is not an easy task when  $W_i$ ,  $X_i$  and  $U_i$  are dependent.

In these lecture notes, we follow Stein's original theme, by presenting his ideas in the context of normal approximation. We shall focus on the approach using concentration inequalities. This approach dates back to Stein's lectures around 1970 (see Ho & Chen (1978)). It was later extended by Chen (1986) to dependent and non-identically distributed random variables with arbitrary index set. A proof of the Berry–Esseen theorem for independent and non-identically distributed random variables using the concentration inequality approach is given in Section 2 of Chen (1998). The approach has further been shown to be effective in obtaining not only uniform but also non-uniform Berry–Esseen bounds for the accuracy of normal approximation for independent random variables (Chen & Shao, 2001) and for locally dependent random fields as well (Chen & Shao, 2004a). As an extension, a randomized concentration inequality was used to establish uniform and non-uniform Berry–Esseen bounds for non-linear statistics in Chen & Shao (2004b). In view of its current successes, the technique seems well worth further exploration.

We also briefly discuss the exchangeable pairs coupling, and use it to obtain a Berry–Esseen bound for the number of 1's in the binary expansion of a random integer. The use of exchangeable pairs was discussed in Stein's monograph (Stein, 1986), and is now widely known (see, for example, Diaconis, Holmes & Reinert (2004)). Other proofs of the Berry–Esseen theorem which will not be included in these lectures are the inductive argument of Barbour & Hall (1984), Stroock (1993) and Bolthausen (1984), the size bias coupling used by Goldstein & Rinott (1996) and the zero bias transformation introduced by Goldstein & Reinert (1997). The inductive argument works well for uniform bounds when the conditional distribution of the sum under consideration, given any one of the variables in the sum, has a form similar to that of the corresponding unconditional distribution of the sum. The size biasing method works well for many combinatorial problems, such as counting the number of vertices in a random graph; theoretically speaking, the zero bias transformation approach works for arbitrary random variables  $W$  with mean zero and finite variance. More on these approaches

can be found in Sections 4.5, 5.2 and 8 of Chapter 4 of this volume. We also refer to Goldstein & Reinert (1997) for some of the important features of the zero bias transformation and to Goldstein (2005) for Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences using zero and size biasing approaches. Short surveys of normal approximation by Stein’s method were given in Rinott & Rotar (2000) and in Chen (1998). Stein’s method has also been used to establish bounds on the approximation error in the multivariate central limit theorem, in particular by Götze (1991) and by Rinott & Rotar (1996). A collection of expository lectures on a variety of aspects of Stein’s method and its applications has recently been edited by Diaconis & Holmes (2004).

These notes are organized as follows. We begin with a more detailed account of Stein’s method than the sketch provided in this section. In Section 3, we consider the expectation of smooth functions under independence and local dependence, and also in the setting of an exchangeable pair having the linear regression property. Some applications are given. Section 4 discusses sums of uniformly bounded random variables. Here, it is shown that Berry–Esseen bounds for distribution functions can be obtained without the use of concentration inequalities. Both the independent case and the problem of counting the number of ones in the binary expansion of a random integer are studied, where in the latter case the use of an exchangeable pair is demonstrated. In Sections 5 and 6, concentration inequalities are invoked to obtain both uniform and non-uniform Berry–Esseen bounds for sums for independent random variables. The last section is devoted to obtaining a uniform Berry–Esseen bound under local dependence.

## 2. Fundamentals of Stein’s method

In this section, we follow Stein (1986), and give a more detailed account of the basic method sketched in the introduction.

### 2.1. Characterization

Let  $Z$  be a standard normally distributed random variable, and let  $\mathcal{C}_{bd}$  be the set of continuous and piecewise continuously differentiable functions  $f: \mathcal{R} \rightarrow \mathcal{R}$  with  $\mathbb{E}|f'(Z)| < \infty$ . Stein’s method rests on the following characterization, which is a slight strengthening of (1.1).



**Lemma 2.1:** *Let  $W$  be a real valued random variable. Then  $W$  has a standard normal distribution if and only if*

$$\mathbb{E}f'(W) = \mathbb{E}\{Wf(W)\}, \quad (2.1)$$

for all  $f \in \mathcal{C}_{bd}$ .

**Proof:** *Necessity.* If  $W$  has a standard normal distribution, then for  $f \in \mathcal{C}_{bd}$

$$\begin{aligned} \mathbb{E}f'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w)e^{-w^2/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(w) \left( \int_{-\infty}^w (-x)e^{-x^2/2} dx \right) dw \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f'(w) \left( \int_w^{\infty} xe^{-x^2/2} dx \right) dw. \end{aligned}$$

By Fubini's theorem, it thus follows that

$$\begin{aligned} \mathbb{E}f'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \left( \int_x^0 f'(w) dw \right) (-x)e^{-x^2/2} dx \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left( \int_0^x f'(w) dw \right) xe^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [f(x) - f(0)]xe^{-x^2/2} dx \\ &= \mathbb{E}Wf(W). \end{aligned}$$

*Sufficiency.* For fixed  $z \in \mathbb{R}$ , let  $f(w) := f_z(w)$  denote the solution of the equation

$$f'(w) - wf(w) = \mathbf{1}_{(-\infty, z]}(w) - \Phi(z). \quad (2.2)$$

Multiplying by  $-e^{-w^2/2}$  on both sides of (2.2) yields

$$\left( e^{-w^2/2} f(w) \right)' = -e^{-w^2/2} (\mathbf{1}_{(-\infty, z]}(w) - \Phi(z)).$$

Thus  $f_z$  is given by

$$\begin{aligned} f_z(w) &= e^{w^2/2} \int_{-\infty}^w [\mathbf{1}_{(-\infty, z]}(x) - \Phi(z)]e^{-x^2/2} dx \\ &= -e^{w^2/2} \int_w^{\infty} [\mathbf{1}_{(-\infty, z]}(x) - \Phi(z)]e^{-x^2/2} dx \\ &= \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)[1 - \Phi(z)] & \text{if } w \leq z, \\ \sqrt{2\pi}e^{w^2/2}\Phi(z)[1 - \Phi(w)] & \text{if } w \geq z. \end{cases} \end{aligned} \quad (2.3)$$

By Lemma 2.2 below,  $f_z$  is a bounded continuous and piecewise continuously differentiable function. Suppose that (2.1) holds for all  $f \in \mathcal{C}_{bd}$ . Then it holds for  $f_z$ , and hence, by (2.2)

$$0 = \mathbb{E}[f'_z(W) - Wf_z(W)] = \mathbb{E}[\mathbf{1}_{(-\infty, z]}(w) - \Phi(z)] = P(W \leq z) - \Phi(z).$$

Thus  $W$  has a standard normal distribution.  $\blacksquare$

Equation (2.2) is a particular case of the more general Stein equation

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z), \quad (2.4)$$

which is to be solved for  $f$ , given a real valued measurable function  $h$  with  $\mathbb{E}|h(Z)| < \infty$ : c.f. (1.2). As for (2.3), the solution  $f = f_h$  is given by

$$\begin{aligned} f_h(w) &= e^{w^2/2} \int_{-\infty}^w [h(x) - \mathbb{E}h(Z)] e^{-x^2/2} dx \\ &= -e^{w^2/2} \int_w^{\infty} [h(x) - \mathbb{E}h(Z)] e^{-x^2/2} dx. \end{aligned} \quad (2.5)$$

## 2.2. Properties of the solutions

We now list some basic properties of the solutions (2.3) and (2.5) to the Stein equations (2.2) and (2.4). The reasons why we need them have already been indicated in (1.8) and (1.9), where estimates of  $\sup_{h \in \mathcal{H}} \|f''_h\|$ ,  $\sup_{z \in \mathbf{R}} \|f'_z\|$  and  $\sup_{z \in \mathbf{R}} \sup_{x \neq z} |f'_z(x)|$  were required, in order to determine our error bounds for the various approximations. For the more detailed arguments to come, further properties are also needed. We defer the proofs to the appendix, since they only involve careful real analysis.

We begin by considering the solution  $f_z$  to (2.2), given in (2.3).

**Lemma 2.2:** *The function  $f_z$  defined by (2.3) is such that*

$$wf_z(w) \text{ is an increasing function of } w. \quad (2.6)$$

Moreover, for all real  $w, u$  and  $v$ ,

$$|wf_z(w)| \leq 1, \quad |wf_z(w) - uf_z(u)| \leq 1 \quad (2.7)$$

$$|f'_z(w)| \leq 1, \quad |f'_z(w) - f'_z(v)| \leq 1 \quad (2.8)$$

$$0 < f_z(w) \leq \min(\sqrt{2\pi}/4, 1/|z|) \quad (2.9)$$

and

$$|(w+u)f_z(w+u) - (w+v)f_z(w+v)| \leq (|w| + \sqrt{2\pi}/4)(|u| + |v|). \quad (2.10)$$

We mostly use (2.8) and (2.9) for our approximations. If one does not care about the constants, one can easily obtain

$$|f'_z(w)| \leq 2 \quad \text{and} \quad 0 < f_z(w) \leq \sqrt{\pi/2}$$

by using the well-known inequality

$$1 - \Phi(w) \leq \min\left(\frac{1}{2}, \frac{1}{w\sqrt{2\pi}}\right) e^{-w^2/2}, \quad w > 0.$$

Next, we consider the solution  $f_h$  to the Stein equation (2.4), as given in (2.5), for any bounded absolutely continuous function  $h$ .

**Lemma 2.3:** *For any absolutely continuous function  $h: \mathbb{R} \rightarrow \mathbb{R}$ , the solution  $f_h$  given in (2.5) satisfies*

$$\|f_h\| \leq \min(\sqrt{\pi/2}\|h(\cdot) - \mathbb{E}h(Z)\|, 2\|h'\|), \quad (2.11)$$

$$\|f'_h\| \leq \min(2\|h(\cdot) - \mathbb{E}h(Z)\|, 4\|h'\|) \quad (2.12)$$

and

$$\|f''_h\| \leq 2\|h'\|. \quad (2.13)$$

### 2.3. Construction of the Stein identities

In this section, we return to the elementary use of Stein's method which led to the simple bound (1.8), but express the remainders  $\eta_1$  and  $\eta_2$  in a form better suited to deriving more advanced results. We now take  $\xi_1, \xi_2, \dots, \xi_n$  to be independent random variables satisfying  $\mathbb{E}\xi_i = 0$ ,  $1 \leq i \leq n$ , and such that  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . Thus  $\xi_i$  corresponds to the normalized random variable  $n^{-1/2}X_i$  of the introduction; here, however, we no longer require the  $\xi_i$  to be identically distributed. Much as before, we set

$$W = \sum_{i=1}^n \xi_i \quad \text{and} \quad W^{(i)} = W - \xi_i, \quad (2.14)$$

and define

$$K_i(t) = \mathbb{E}\{\xi_i(I_{\{0 \leq t \leq \xi_i\}} - I_{\{\xi_i \leq t < 0\}})\}. \quad (2.15)$$

It is easy to check that  $K_i(t) \geq 0$  for all real  $t$ , that

$$\int_{-\infty}^{\infty} K_i(t) dt = \mathbb{E}\xi_i^2 \quad \text{and that} \quad \int_{-\infty}^{\infty} |t|K_i(t) dt = \frac{1}{2}\mathbb{E}|\xi_i|^3. \quad (2.16)$$

Let  $h$  be a measurable function with  $\mathbb{E}|h(Z)| < \infty$ , and let  $f = f_h$  be the corresponding solution of the Stein equation (2.4). Our goal is to estimate

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}\{f'(W) - Wf(W)\}.$$

The following argument is fundamental to the approach, and many of the tricks reappear repeatedly in what follows.

Since  $\xi_i$  and  $W^{(i)}$  are independent for each  $1 \leq i \leq n$ , we have

$$\begin{aligned} \mathbb{E}\{Wf(W)\} &= \sum_{i=1}^n \mathbb{E}\{\xi_i f(W)\} \\ &= \sum_{i=1}^n \mathbb{E}\{\xi_i [f(W) - f(W^{(i)})]\}, \end{aligned}$$

where the last equality follows because  $\mathbb{E}\xi_i = 0$ . Writing the final difference in integral form, we thus have

$$\begin{aligned} \mathbb{E}\{Wf(W)\} &= \sum_{i=1}^n \mathbb{E}\left\{\xi_i \int_0^{\xi_i} f'(W^{(i)} + t) dt\right\} \\ &= \sum_{i=1}^n \mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W^{(i)} + t) \xi_i (I_{\{0 \leq t \leq \xi_i\}} - I_{\{\xi_i \leq t < 0\}}) dt\right\} \\ &= \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{f'(W^{(i)} + t)\} K_i(t) dt, \end{aligned} \quad (2.17)$$

from the definition of  $K_i$  and again using independence. However, because

$$\sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t) dt = \sum_{i=1}^n \mathbb{E}\xi_i^2 = 1,$$

it follows that

$$\mathbb{E}f'(W) = \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{f'(W)\} K_i(t) dt. \quad (2.18)$$

Thus, by (2.17) and (2.18),

$$\mathbb{E}\{f'(W) - Wf(W)\} = \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{f'(W) - f'(W^{(i)} + t)\} K_i(t) dt. \quad (2.19)$$

Equations (2.17) and (2.19) play a key role in proving good normal approximations. Note in particular that (2.19) is an *equality*, replacing the clumsier bounds for  $|\eta_1|$  and  $|\eta_2|$  which led to (1.8). This more careful treatment of the errors pays big dividends later on. Note also that (2.17) and (2.19) hold for all bounded absolute continuous  $f$ .

### 3. Normal approximation for smooth functions

Our goal in this section is to estimate  $\mathbb{E}h(W) - \mathbb{E}h(Z)$  for various classes of random variables  $W$ , when  $h$  is a smooth function satisfying

$$\|h'\| := \sup_x |h'(x)| < \infty. \quad (3.1)$$

Such estimates lead naturally to bounds on the accuracy of the standard normal approximation to  $W$  in terms of the Wasserstein distance  $d_W$ , and hence suffice to prove the central limit theorem. The next theorem highlights this, and also shows that Wasserstein bounds imply bounds, albeit rather weaker, with respect to the Kolmogorov distance  $d_K$ .

**Theorem 3.1:** *Assume that there exists a  $\delta$  such that, for any uniformly Lipschitz function  $h$ ,*

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \delta \|h'\|. \quad (3.2)$$

Then

$$d_W(\mathcal{L}(W), \mathcal{N}(0, 1)) := \sup_{h \in \text{Lip}(1)} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \delta; \quad (3.3)$$

$$d_K(\mathcal{L}(W), \mathcal{N}(0, 1)) := \sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 2\delta^{1/2}. \quad (3.4)$$

**Proof:** The first statement is immediate from the definition of  $d_W$ . For the second, we can assume that  $\delta \leq 1/4$ , since otherwise (3.4) is trivial. Let  $\alpha = \delta^{1/2}(2\pi)^{1/4}$ , and define for fixed  $z$

$$h_\alpha(w) = \begin{cases} 1 & \text{if } w \leq z, \\ 0 & \text{if } w \geq z + \alpha, \\ \text{linear} & \text{if } z \leq w \leq z + \alpha. \end{cases}$$

Then  $\|h'\| = 1/\alpha$ , and hence, by (3.2),

$$\begin{aligned} \mathbb{P}(W \leq z) - \Phi(z) &\leq \mathbb{E}h_\alpha(W) - \mathbb{E}h_\alpha(Z) + \mathbb{E}h_\alpha(Z) - \Phi(z) \\ &\leq \frac{\delta}{\alpha} + \mathbb{P}\{z \leq Z \leq z + \alpha\} \\ &\leq \frac{\delta}{\alpha} + \frac{\alpha}{\sqrt{2\pi}}, \end{aligned}$$

and hence

$$\mathbb{P}(W \leq z) - \Phi(z) \leq 2(2\pi)^{-1/4} \delta^{1/2} \leq 2\delta^{1/2}. \quad (3.5)$$

Similarly, we have

$$\mathbb{P}(W \leq z) - \Phi(z) \geq -2\delta^{1/2}, \quad (3.6)$$

proving (3.4). ■

In the next three sections, we show that (3.2) is satisfied with suitably small  $\delta$ , when  $W$  is a sum of (i) independent random variables, or (ii) locally dependent random variables. We also show that (3.2) is satisfied when (iii)  $W$  is such that an exchangeable pair  $(W, W')$  can be constructed having the linear regression property:

$$\mathbb{E}\{W' | W\} = (1 - \lambda)W \quad \text{for some } 0 < \lambda < 1. \quad (3.7)$$

### 3.1. Independent random variables

In this section, we use (2.19) to prove (3.2) for  $W$  a sum of independent random variables with zero means and finite third moments, extending (1.8) to non-identically distributed random variables.

**Theorem 3.2:** *Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables satisfying  $\mathbb{E}\xi_i = 0$  and  $\mathbb{E}|\xi_i|^3 < \infty$  for each  $1 \leq i \leq n$ , and such that  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . Then Theorem 3.1 can be applied with*

$$\delta = 3 \sum_{i=1}^n \mathbb{E}|\xi_i|^3. \quad (3.8)$$

In particular, we have

$$\left| \mathbb{E}|W| - \sqrt{\frac{2}{\pi}} \right| \leq 3 \sum_{i=1}^n \mathbb{E}|\xi_i|^3.$$

**Proof:** It follows from Lemma 2.3 that  $\|f_h''\| \leq 2\|h'\|$ . Therefore, by (2.19) and the mean value theorem,

$$\begin{aligned} |\mathbb{E}\{f_h'(W) - W f_h(W)\}| &\leq \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}|f_h'(W) - f_h'(W^{(i)} + t)| K_i(t) dt \\ &\leq 2\|h'\| \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}(|t| + |\xi_i|) K_i(t) dt. \end{aligned}$$

Using (2.16), it thus follows that

$$\begin{aligned} |\mathbb{E}\{f'_h(W) - W f_h(W)\}| &\leq 2\|h'\| \sum_{i=1}^n (\mathbb{E}|\xi_i|^3/2 + \mathbb{E}|\xi_i|\mathbb{E}\xi_i^2) \\ &\leq 3\|h'\| \sum_{i=1}^n \mathbb{E}|\xi_i|^3. \quad \blacksquare \end{aligned}$$

It is actually not necessary to assume the existence of finite third moments in Theorem 3.2. The quantity  $\delta$  can be computed in terms of the elements appearing in the statement of the Lindeberg central limit theorem.

**Theorem 3.3:** *Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables satisfying  $\mathbb{E}\xi_i = 0$  for each  $1 \leq i \leq n$  and such that  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . Then Theorem 3.1 can be applied with*

$$\delta = 4(4\beta_2 + 3\beta_3), \quad (3.9)$$

where

$$\beta_2 = \sum_{i=1}^n \mathbb{E}\xi_i^2 I_{\{|\xi_i| > 1\}} \quad \text{and} \quad \beta_3 = \sum_{i=1}^n \mathbb{E}|\xi_i|^3 I_{\{|\xi_i| \leq 1\}}. \quad (3.10)$$

**Proof:** We use (2.12) and (2.13) to show that

$$\begin{aligned} |f'_h(W) - f'_h(W^{(i)} + t)| &\leq \|h'\| \min(8, 2(|t| + |\xi_i|)) \\ &\leq 8\|h'\| (|t| \wedge 1 + |\xi_i| \wedge 1), \end{aligned}$$

where  $a \wedge b$  denotes  $\min(a, b)$ . Substituting this bound into (2.19), we obtain

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq 8\|h'\| \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}(|t| \wedge 1 + |\xi_i| \wedge 1) K_i(t) dt. \quad (3.11)$$

Now

$$\int_{-\infty}^{\infty} (|t| \wedge 1) \{ \mathbf{1}_{[0,x]}(t) - \mathbf{1}_{[-x,0]}(t) \} dt = \begin{cases} \frac{1}{2}|x|t^2 + |x|(|x| - 1) & \text{if } |x| > 1; \\ \frac{1}{2}|x|^3, & \text{if } |x| \leq 1, \end{cases}$$

so that

$$\begin{aligned} &\int_{-\infty}^{\infty} \mathbb{E}(|t| \wedge 1 + |\xi_i| \wedge 1) K_i(t) dt \\ &= \mathbb{E}\{|\xi_i|(|\xi_i| - 1) I_{\{|\xi_i| > 1\}}\} + \frac{1}{2} \mathbb{E}\{|\xi_i|(|\xi_i| \wedge 1)^2\} + \mathbb{E}\{\xi_i^2 \mathbb{E}(|\xi_i| \wedge 1)\} \\ &= \mathbb{E}\{\xi_i^2 I_{\{|\xi_i| > 1\}}\} - \frac{1}{2} \mathbb{E}\{|\xi_i| I_{\{|\xi_i| > 1\}}\} \\ &\quad + \frac{1}{2} \mathbb{E}\{|\xi_i|^3 I_{\{|\xi_i| \leq 1\}}\} + \mathbb{E}\{\xi_i^2 \mathbb{E}(|\xi_i| \wedge 1)\}. \end{aligned}$$

Adding over  $i$ , it follows that

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq 8\|h'\| \left\{ \beta_2 + \frac{1}{2}\beta_3 + \sum_{i=1}^n \mathbb{E}\xi_i^2 \mathbb{E}(|\xi_i| \wedge 1) \right\}. \quad (3.12)$$

However, since both  $x^2$  and  $(x \wedge 1)$  are increasing functions of  $x \geq 0$ , it follows that, for any random variable  $\xi$ ,

$$\mathbb{E}\xi^2 \mathbb{E}(|\xi| \wedge 1) \leq \mathbb{E}\{\xi^2(|\xi| \wedge 1)\} = \mathbb{E}|\xi|^3 I_{\{|\xi| \leq 1\}} + \mathbb{E}\xi^2 I_{\{|\xi| > 1\}}, \quad (3.13)$$

so that the final sum is no greater than  $\beta_3 + \beta_2$ , completing the bound. ■

Theorem 3.1 and Theorem 3.3 together yield the Lindeberg central limit theorem, as follows. Let  $X_1, X_2, \dots, X_n$  be independent random variables with  $\mathbb{E}X_i = 0$  and  $\mathbb{E}X_i^2 < \infty$  for each  $1 \leq i \leq n$ . Put

$$S_n = \sum_{i=1}^n X_i \quad \text{and} \quad B_n^2 = \sum_{i=1}^n \mathbb{E}X_i^2.$$

To apply Theorems 3.1 and 3.3, let

$$\xi_i = X_i/B_n \quad \text{and} \quad W = S_n/B_n. \quad (3.14)$$

Define  $\beta_2$  and  $\beta_3$  as in (3.10), and observe that, for any  $0 < \varepsilon < 1$ ,

$$\begin{aligned} \beta_2 + \beta_3 &= \frac{1}{B_n^2} \sum_{i=1}^n \mathbb{E}\{X_i^2 I_{\{|X_i| > B_n\}}\} + \frac{1}{B_n^3} \sum_{i=1}^n \mathbb{E}\{|X_i|^3 I_{\{|X_i| \leq B_n\}}\} \\ &\leq \frac{1}{B_n^2} \sum_{i=1}^n \mathbb{E}\{X_i^2 I_{\{|X_i| > B_n\}}\} + \frac{1}{B_n^3} \sum_{i=1}^n B_n \mathbb{E}\{X_i^2 I_{\{\varepsilon B_n \leq |X_i| \leq B_n\}}\} \\ &\quad + \frac{1}{B_n^3} \sum_{i=1}^n \varepsilon B_n \mathbb{E}\{X_i^2 I_{\{|X_i| < \varepsilon B_n\}}\} \\ &\leq \frac{1}{B_n^2} \sum_{i=1}^n \mathbb{E}\{X_i^2 I_{\{|X_i| > \varepsilon B_n\}}\} + \varepsilon. \end{aligned} \quad (3.15)$$

If the Lindeberg condition holds, that is if

$$\frac{1}{B_n^2} \sum_{i=1}^n \mathbb{E}\{X_i^2 I_{\{|X_i| > \varepsilon B_n\}}\} \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.16)$$

for all  $\varepsilon > 0$ , then (3.15) implies  $\beta_2 + \beta_3 \rightarrow 0$  as  $n \rightarrow \infty$ , since  $\varepsilon$  is arbitrary. Hence, by Theorems 3.1 and 3.3,

$$\sup_z |\mathbb{P}(S_n/B_n \leq z) - \Phi(z)| \leq 8(\beta_2 + \beta_3)^{1/2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This proves the Lindeberg central limit theorem.



### 3.2. Locally dependent random variables

An  $m$ -dependent sequence of random variables  $\xi_i$ ,  $i \in \mathbb{Z}$ , is one with the property that, for each  $i$ , the sets of random variables  $\{\xi_j, j \leq i\}$  and  $\{\xi_j, j > i + m\}$  are independent. As a special case, sequences of independent random variables are 0-dependent. Local dependence generalizes the notion of  $m$ -dependence to random variables with arbitrary index set. It is applicable, for instance, to random variables indexed by the vertices of a graph, and such that the collections  $\{\xi_i, i \in I\}$  and  $\{\xi_j, j \in J\}$  are independent whenever  $I \cap J = \emptyset$  and the graph contains no edges  $\{i, j\}$  with  $i \in I$  and  $j \in J$ .

Let  $\mathcal{J}$  be a finite index set of cardinality  $n$ , and let  $\{\xi_i, i \in \mathcal{J}\}$  be a random field with zero means and finite variances. Define  $W = \sum_{i \in \mathcal{J}} \xi_i$ , and assume that  $\text{Var}(W) = 1$ . For  $A \subset \mathcal{J}$ , let  $\xi_A$  denote  $\{\xi_i, i \in A\}$  and  $A^c = \{j \in \mathcal{J} : j \notin A\}$ . We introduce the following two assumptions, defining different strengths of local dependence.

- (LD1) For each  $i \in \mathcal{J}$  there exists  $A_i \subset \mathcal{J}$  such that  $\xi_i$  and  $\xi_{A_i^c}$  are independent.
- (LD2) For each  $i \in \mathcal{J}$  there exist  $A_i \subset B_i \subset \mathcal{J}$  such that  $\xi_i$  is independent of  $\xi_{A_i^c}$  and  $\xi_{A_i}$  is independent of  $\xi_{B_i^c}$ .

We then define  $\eta_i = \sum_{j \in A_i} \xi_j$  and  $\tau_i = \sum_{j \in B_i} \xi_j$ . Note that, for independent random variables  $\xi_i$ , we can take  $A_i = B_i = \{i\}$ , for which then  $\eta_i = \tau_i = \xi_i$ .

**Theorem 3.4:** *Theorem 3.1 can be applied with*

$$\delta = 4\mathbb{E} \left| \sum_{i \in \mathcal{J}} \{\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)\} \right| + \sum_{i \in \mathcal{J}} \mathbb{E} |\xi_i \eta_i^2| \quad (3.17)$$

under (LD1), and with

$$\delta = 2 \sum_{i \in \mathcal{J}} (\mathbb{E} |\xi_i \eta_i \tau_i| + |\mathbb{E}(\xi_i \eta_i)| \mathbb{E} |\tau_i|) + \sum_{i \in \mathcal{J}} \mathbb{E} |\xi_i \eta_i^2| \quad (3.18)$$

under (LD2).

**Remark:** For independent random variables, the value of  $\delta$  in (3.18) is  $5 \sum_{i \in \mathcal{J}} \mathbb{E} |\xi_i|^3$ , somewhat larger than the direct bound given in Theorem 3.2.

**Proof:** We first derive Stein identities similar to (2.17) and (2.19). Let  $f = f_h$  be the solution of the Stein equation (2.4). Then

$$\begin{aligned}\mathbb{E}\{Wf(W)\} &= \sum_{i \in \mathcal{J}} \mathbb{E}\xi_i f(W) \\ &= \sum_{i \in \mathcal{J}} \mathbb{E}\xi_i [f(W) - f(W - \eta_i)],\end{aligned}$$

by the independence of  $\xi_i$  and  $W - \eta_i$ . Hence

$$\begin{aligned}\mathbb{E}\{Wf(W)\} &= \sum_{i \in \mathcal{J}} \mathbb{E}\{\xi_i [f(W) - f(W - \eta_i) - \eta_i f'(W)]\} \\ &\quad + \mathbb{E}\left\{\left(\sum_{i \in \mathcal{J}} \xi_i \eta_i\right) f'(W)\right\}.\end{aligned}\quad (3.19)$$

Now, because  $\mathbb{E}\xi_i = 0$  for all  $i$ , and from (LD1), it follows that

$$1 = \mathbb{E}W^2 = \sum_{i \in \mathcal{J}} \sum_{j \in \mathcal{J}} \mathbb{E}\{\xi_i \xi_j\} = \sum_{i \in \mathcal{J}} \mathbb{E}\{\xi_i \eta_i\},$$

giving

$$\begin{aligned}\mathbb{E}\{f'(W) - Wf(W)\} &= -\mathbb{E}\left(\sum_{i \in \mathcal{J}} \{\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)\} f'(W)\right) \\ &\quad - \sum_{i \in \mathcal{J}} \mathbb{E}\{\xi_i [f(W) - f(W - \eta_i) - \eta_i f'(W)]\}.\end{aligned}\quad (3.20)$$

By (2.12) and (2.13),  $\|f'\| \leq 4\|h'\|$  and  $\|f''\| \leq 2\|h'\|$ . Therefore it follows from (3.20) and the Taylor expansion that

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \|h'\| \left\{ 4\mathbb{E}\left|\sum_{i \in \mathcal{J}} \{\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)\}\right| + \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i \eta_i^2| \right\}.$$

This proves (3.17).

When (LD2) is satisfied,  $f'(W - \tau_i)$  and  $\xi_i \eta_i$  are independent for each  $i \in \mathcal{J}$ . Hence, using (3.20), we can write

$$\begin{aligned}|\mathbb{E}h(W) - \mathbb{E}h(Z)| &\leq \left| \mathbb{E} \sum_{i \in \mathcal{J}} \{\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)\} (f'(W) - f'(W - \tau_i)) \right| + \|h'\| \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i \eta_i^2| \\ &\leq \|h'\| \left\{ 2 \sum_{i \in \mathcal{J}} (\mathbb{E}|\xi_i \eta_i \tau_i| + |\mathbb{E}(\xi_i \eta_i)| \mathbb{E}|\tau_i|) + \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i \eta_i^2| \right\},\end{aligned}$$

as desired. ■

Here are two examples of locally dependent random fields. We refer to Baldi & Rinott (1989), Rinott (1994), Baldi, Rinott & Stein (1989), Dembo & Rinott (1996), and Chen & Shao (2004) for more details.

**Example 1.** *Graphical dependence.*

Consider a set of random variables  $\{X_i, i \in \mathcal{V}\}$  indexed by the vertices of a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ .  $\mathcal{G}$  is said to be a dependency graph if, for any pair of disjoint sets  $\Gamma_1$  and  $\Gamma_2$  in  $\mathcal{V}$  such that no edge in  $\mathcal{E}$  has one endpoint in  $\Gamma_1$  and the other in  $\Gamma_2$ , the sets of random variables  $\{X_i, i \in \Gamma_1\}$  and  $\{X_i, i \in \Gamma_2\}$  are independent. Let  $D$  denote the maximal degree of  $G$ ; that is, the maximal number of edges incident to a single vertex. Let

$$A_i = \{i\} \cup \{j \in \mathcal{V} : \text{there is an edge connecting } j \text{ and } i\}$$

and  $B_i = \bigcup_{j \in A_i} A_j$ . Then  $\{X_i, i \in \mathcal{V}\}$  satisfies (LD2). Hence (3.18) holds.

**Example 2.** *The number of local maxima on a graph.*

Consider a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  (which is not necessarily a dependency graph) and independent and identically distributed continuous random variables  $\{Y_i, i \in \mathcal{V}\}$ . For  $i \in \mathcal{V}$  define the 0-1 indicator variable

$$X_i = \begin{cases} 1 & \text{if } Y_i > Y_j \text{ for all } j \in N_i \\ 0 & \text{otherwise} \end{cases}$$

where  $N_i = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}$ , so that  $X_i = 1$  indicates that  $Y_i$  is a local maximum. Let  $W = \sum_{i \in \mathcal{V}} X_i$  be the number of local maxima. Let

$$A_i = \{i\} \cup \left\{ \bigcup_{j \in N_i} N_j \right\} \quad \text{and} \quad B_i = \bigcup_{j \in A_i} A_j.$$

Then  $\{X_i, i \in \mathcal{V}\}$  satisfies (LD2), and (3.18) holds.

One can refer to Dembo & Rinott (1996) and Rinott & Rotar (1996) for more examples and results under local dependence.

### 3.3. Exchangeable pairs

Let  $W$  be a random variable that is not necessarily the partial sum of independent random variables. Suppose that  $W$  is approximately normal, and that we want to find how accurate the approximation is. Another basic approach to Stein's method is to introduce a second random variable  $W'$  on the same probability space, in such a way that  $(W, W')$  is an exchangeable pair; that is, such that  $(W, W')$  and  $(W', W)$  have the same distribution.

The approach makes essential use of the elementary fact that, if  $(W, W')$  is an exchangeable pair, then

$$\mathbb{E}g(W, W') = 0 \quad (3.21)$$

for all *antisymmetric* measurable functions  $g(x, y)$  such that the expected value exists.

A key identity is the following lemma (see Stein 1986).

**Lemma 3.5:** *Let  $(W, W')$  be an exchangeable pair of real random variables with finite variance, having the linear regression property*

$$\mathbb{E}(W' | W) = (1 - \lambda)W \quad (3.22)$$

for some  $0 < \lambda < 1$ . Then

$$\mathbb{E}W = 0 \quad \text{and} \quad \mathbb{E}(W' - W)^2 = 2\lambda\mathbb{E}W^2, \quad (3.23)$$

and, for every piecewise continuous function  $f$  satisfying the growth condition  $|f(w)| \leq C(1 + |w|)$ , we have

$$\mathbb{E}\{Wf(W)\} = \frac{1}{2\lambda}\mathbb{E}\{(W - W')(f(W) - f(W'))\}. \quad (3.24)$$

**Proof:** The proof exploits (3.21) with  $g(x, y) = (x - y)(f(y) + f(x))$ , for which  $\mathbb{E}g(W, W')$  exists, because of the assumption on  $f$ . Then (3.21) gives

$$\begin{aligned} 0 &= \mathbb{E}\{(W - W')(f(W') + f(W))\} \\ &= \mathbb{E}\{(W - W')(f(W') - f(W))\} + 2\mathbb{E}\{f(W)(W - W')\} \quad (3.25) \\ &= \mathbb{E}\{(W - W')(f(W') - f(W))\} + 2\mathbb{E}\{f(W)\mathbb{E}(W - W' | W)\} \\ &= \mathbb{E}\{(W - W')(f(W') - f(W))\} + 2\lambda\mathbb{E}\{Wf(W)\}, \end{aligned}$$

this last by (3.22), and this is just (3.24). ■

Using this lemma, we can deduce the following theorem.

**Theorem 3.6:** *If  $(W, W')$  is exchangeable and (3.22) holds, it follows that Theorem 3.1 can be applied with*

$$\delta = 4\mathbb{E}\left|1 - \frac{1}{2\lambda}\mathbb{E}((W' - W)^2 | W)\right| + \frac{1}{2\lambda}\mathbb{E}|W - W'|^3. \quad (3.26)$$

**Remark:** In the first term, note that

$$\mathbb{E}\left\{1 - \frac{1}{2\lambda}\mathbb{E}((W' - W)^2 | W)\right\} = 1 - \frac{1}{2\lambda}\mathbb{E}(W' - W)^2 = 1 - \mathbb{E}W^2,$$

so that the bound is unlikely to be useful unless  $\mathbb{E}W^2$  is close to 1.

**Proof:** Let  $f = f_h$  be the solution (2.5) to the Stein equation, and define

$$\widehat{K}(t) = (W - W')(I_{\{-(W-W') \leq t \leq 0\}} - I_{\{0 < t \leq -(W-W')\}}) \geq 0,$$

noting that

$$\int_{-\infty}^{\infty} \widehat{K}(t) dt = (W - W')^2. \quad (3.27)$$

By (3.24),

$$\begin{aligned} \mathbb{E}\{Wf(W)\} &= \frac{1}{2\lambda} \mathbb{E} \left\{ \int_{-(W-W')}^0 f'(W+t)(W-W') dt \right\} \\ &= \frac{1}{2\lambda} \mathbb{E} \left\{ \int_{-\infty}^{\infty} f'(W+t)\widehat{K}(t) dt \right\} \end{aligned}$$

and

$$\mathbb{E}f'(W) = \mathbb{E} \left\{ f'(W) \left( 1 - \frac{1}{2\lambda}(W - W')^2 \right) \right\} + \frac{1}{2\lambda} \mathbb{E} \left\{ \int_{-\infty}^{\infty} f'(W)\widehat{K}(t) dt \right\},$$

from (3.27). Putting the two together gives

$$\begin{aligned} |\mathbb{E}h(W) - \mathbb{E}h(Z)| &= |\mathbb{E}\{f'(W) - Wf(W)\}| \\ &= \left| \mathbb{E}f'(W) \left( 1 - \frac{1}{2\lambda}(W - W')^2 \right) + \frac{1}{2\lambda} \mathbb{E} \int_{-\infty}^{\infty} (f'(W) - f'(W+t))\widehat{K}(t) dt \right| \\ &\leq \left| \mathbb{E} \left\{ f'(W) \left( 1 - \frac{1}{2\lambda} \mathbb{E}((W - W')^2 | W) \right) \right\} \right| \\ &\quad + \frac{1}{2\lambda} \mathbb{E} \left| \int_{-\infty}^{\infty} (f'(W) - f'(W+t))\widehat{K}(t) dt \right|, \end{aligned}$$

and now the bounds in Lemma 2.3 give

$$\begin{aligned} &|\mathbb{E}h(W) - \mathbb{E}h(Z)| \\ &\leq \|h'\| \left\{ 4\mathbb{E} \left| 1 - \frac{1}{2\lambda} \mathbb{E}((W - W')^2 | W) \right| + \frac{1}{\lambda} \mathbb{E} \int_{-\infty}^{\infty} |t|\widehat{K}(t) dt \right\} \\ &= \|h'\| \left\{ 4\mathbb{E} \left| 1 - \frac{1}{2\lambda} \mathbb{E}((W' - W)^2 | W) \right| + \frac{1}{2\lambda} \mathbb{E}|W - W'|^3 \right\}, \end{aligned}$$

from (3.27), as desired.  $\blacksquare$

We use the following example to show how to apply the bound in the above theorem. Let  $\xi_i$  be independent random variables with zero means and  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ , and put  $W = \sum_{i=1}^n \xi_i$ . Let  $\{\xi_i^*, 1 \leq i \leq n\}$  be an independent copy of  $\{\xi_i, 1 \leq i \leq n\}$ , and let  $I$  have uniform distribution

on  $\{1, 2, \dots, n\}$ , independent of  $\{\xi_i\}$  and  $\{\xi_i^*\}$ . Define  $W' = W - \xi_I + \xi_I^*$ . Then  $(W, W')$  is an exchangeable pair, and

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

so that (3.22) is satisfied with  $\lambda = 1/n$ . Direct calculation also gives

$$\mathbb{E}|W - W'|^3 = \frac{1}{n} \sum_{i=1}^n \mathbb{E}|\xi_i - \xi_i^*|^3 \leq (8/n) \sum_{i=1}^n \mathbb{E}|\xi_i|^3$$

and

$$\mathbb{E}((W - W')^2 | W) = \frac{1}{n} \left(1 + \sum_{i=1}^n \mathbb{E}(\xi_i^2 | W)\right); \quad (3.28)$$

from the latter, we have

$$\begin{aligned} \mathbb{E} \left| 1 - \frac{1}{2\lambda} \mathbb{E}((W' - W)^2 | W) \right| &= \frac{1}{2} \mathbb{E} \left| 1 - \mathbb{E} \left( \sum_{i=1}^n \xi_i^2 | W \right) \right| \\ &\leq \frac{1}{2} \mathbb{E} \left| \sum_{i=1}^n (\xi_i^2 - \mathbb{E}\xi_i^2) \right|. \end{aligned}$$

Thus, if the  $\xi_i$  have finite fourth moments, Theorem 3.1 can be applied with

$$\delta = 2 \sqrt{\sum_{i=1}^n \text{Var}(\xi_i^2) + 4 \sum_{i=1}^n \mathbb{E}|\xi_i|^3} \leq 2 \sqrt{\sum_{i=1}^n \mathbb{E}|\xi_i|^4 + 4 \sum_{i=1}^n \mathbb{E}|\xi_i|^3}.$$

In particular, if  $\xi_i = n^{-1/2} X_i$ , where the random variables  $X_i$  are independent and identically distributed random variables with finite fourth moments, then the bound is of the correct order  $O(n^{-1/2})$ .

On the other hand, if we keep the original form

$$\left| \mathbb{E} \left\{ f'(W) \left( 1 - \frac{1}{2\lambda} \mathbb{E}((W - W')^2 | W) \right) \right\} \right|$$

which gave rise to the first term on the right hand side of (3.26), then we only need to assume finite third moments. To see this, by (3.28).

$$\begin{aligned} \left| \mathbb{E} \left\{ f'(W) \left( 1 - \frac{1}{2\lambda} \mathbb{E}((W - W')^2 | W) \right) \right\} \right| &= \frac{1}{2} \left| \mathbb{E} \left\{ f'(W) \sum_{i=1}^n (\mathbb{E}\xi_i^2 - \xi_i^2) \right\} \right| \\ &= \frac{1}{2} \left| \sum_{i=1}^n \mathbb{E} \left\{ (f'(W) - f'(W - \xi_i)) (\mathbb{E}\xi_i^2 - \xi_i^2) \right\} \right|, \end{aligned}$$

because  $W - \xi_i$  and  $\xi_i$  are independent. This is now bounded using Lemma 2.3 by

$$\|h'\| \sum_{i=1}^n \mathbb{E}|\xi_i(\mathbb{E}\xi_i^2 - \xi_i^2)| \leq 2\|h'\| \sum_{i=1}^n \mathbb{E}|\xi_i|^3,$$

and Theorem 3.1 can be applied with

$$\delta = 6 \sum_{i=1}^n \mathbb{E}|\xi_i|^3.$$

We therefore end this section with the following alternative to (3.26): if  $(W, W')$  is exchangeable and (3.22) holds, then

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \left| \mathbb{E}f'_h(W)(1 - \frac{1}{2\lambda}(W - W')^2) \right| + \frac{1}{2\lambda}\|h'\|\mathbb{E}|W - W'|^3. \quad (3.29)$$

#### 4. Uniform Berry–Esseen bounds: the bounded case

In the previous section, the sharpest bounds in Theorem 3.1 were of order  $O(\delta)$ , and were obtained for the Wasserstein distance  $d_W$ . Those for the Kolmogorov distance  $d_K$  were only of the larger order  $O(\delta^{1/2})$ . Here, we turn to deriving bounds for  $d_K$  which are of comparable order to those for  $d_W$ . We begin with the simplest case of independent summands. The method of proof motivates the formulation of a rather general theorem, which is then applied in a dependent setting, that of the number of 1's in the binary expansion of a randomly chosen integer, using an exchangeable pair.

##### 4.1. Independent random variables

Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables with zero means and with  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . We use the notation of Section 2.3:

$$W = \sum_{i=1}^n \xi_i, \quad W^{(i)} = W - \xi_i \quad \text{and} \quad K_i(t) = \mathbb{E}\{\xi_i(I_{\{0 \leq t \leq \xi_i\}} - I_{\{\xi_i \leq t < 0\}})\}.$$

Let  $f_z$  be the solution of the Stein equation (2.1). For bounded  $\xi_i$ , we are ready to apply (2.17) to obtain the following Berry–Esseen bound.

**Theorem 4.1:** *If  $|\xi_i| \leq \delta_0$  for  $1 \leq i \leq n$ , then*

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 3.3 \delta_0. \quad (4.1)$$

**Proof:** Write  $f = f_z$ . It follows from (2.17) and because  $f$  satisfies the Stein equation (2.1) that

$$\begin{aligned}\mathbb{E}\{Wf(W)\} &= \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{f'(W^{(i)} + t)\} K_i(t) dt \\ &= \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{(W^{(i)} + t)f(W^{(i)} + t) + I_{\{W^{(i)} + t \leq z\}} - \Phi(z)\} K_i(t) dt.\end{aligned}$$

Reorganizing this, and recalling that

$$\sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t) dt = \sum_{i=1}^n \mathbb{E}\xi_i^2 = 1,$$

we have

$$\begin{aligned}& \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{P}(W^{(i)} + t \leq z) K_i(t) dt - \Phi(z) \\ &= \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{Wf(W) - (W^{(i)} + t)f(W^{(i)} + t)\} K_i(t) dt.\end{aligned}\quad (4.2)$$

Now, by (2.10),

$$\begin{aligned}& \sum_{i=1}^n \mathbb{E} \int_{-\infty}^{\infty} |Wf(W) - (W^{(i)} + t)f(W^{(i)} + t)| K_i(t) dt \\ &\leq \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E} \left\{ (|W^{(i)}| + \sqrt{2\pi}/4)(|\xi_i| + |t|) \right\} K_i(t) dt \\ &\leq (1 + \sqrt{2\pi}/4) \sum_{i=1}^n \int_{-\infty}^{\infty} (\mathbb{E}|\xi_i| + |t|) K_i(t),\end{aligned}$$

since  $\mathbb{E}\{W^{(i)}\}^2 \leq 1$  and  $\xi_i$  and  $W^{(i)}$  are independent. Hence, recalling (2.16), we have

$$\begin{aligned}& \left| \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{P}(W^{(i)} + t \leq z) K_i(t) dt - \Phi(z) \right| \\ &\leq (1 + \sqrt{2\pi}/4) \sum_{i=1}^n \left\{ \mathbb{E}|\xi_i| \mathbb{E}\xi_i^2 + \frac{1}{2} \mathbb{E}|\xi_i|^3 \right\} \\ &\leq \frac{3}{2} (1 + \sqrt{2\pi}/4) \sum_{i=1}^n \mathbb{E}|\xi_i|^3.\end{aligned}\quad (4.3)$$



Hence we would be finished if  $\mathbb{P}(W^{(i)} + t \leq z)$  could be replaced by  $\mathbb{P}(W \leq z)$ , since we have  $\sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t) dt = 1$ . Clearly, in view of the fact that  $\mathbb{P}(W \leq z) = \mathbb{P}(W^{(i)} \leq z - \xi_i)$ , we have

$$\begin{aligned} & |\mathbb{P}(W^{(i)} + t \leq z) - \mathbb{P}(W \leq z)| \\ & \leq \mathbb{P}(z - \max\{\xi_i, t\} \leq W^{(i)} \leq z - \min\{\xi_i, t\}), \end{aligned} \quad (4.4)$$

and the difference should be small if both  $|t|$  and  $|\xi_i|$  are.

Since the  $|\xi_i|$  are uniformly bounded by  $\delta_0$ , proving that the difference is small is particularly simple. First, we note that  $|\xi_i| \leq \delta_0$  implies also that  $K_i(t) = 0$  for  $|t| > \delta_0$ , so that we only need to consider (4.4) with both  $|t|$  and  $|\xi_i|$  bounded by  $\delta_0$ . But then

$$\mathbb{P}(W^{(i)} + t \leq z) = \mathbb{P}(W - \xi_i + t \leq z) \begin{cases} \geq \mathbb{P}(W \leq z - 2\delta_0) \\ \leq \mathbb{P}(W \leq z + 2\delta_0). \end{cases} \quad (4.5)$$

Taking  $z + 2\delta_0$  for  $z$  in the first inequality in (4.5) and substituting it into (4.3) gives

$$\begin{aligned} \mathbb{P}(W \leq z) - \Phi(z) & \leq \Phi(z + 2\delta_0) - \Phi(z) + \frac{3}{2}(1 + \sqrt{2\pi}/4) \sum_{i=1}^n \mathbb{E}|\xi_i|^3 \\ & \leq \frac{2\delta_0}{\sqrt{2\pi}} + \frac{3}{2}(1 + \sqrt{2\pi}/4)\delta_0 \leq 3.3\delta_0, \end{aligned} \quad (4.6)$$

and the corresponding lower bound follows from the second inequality in (4.5) and (4.3) with  $z - 2\delta_0$  for  $z$ , completing the proof. ■

One can see from the above approach that the key ingredient of the proof is to rewrite  $\mathbb{E}\{Wf(W)\}$  in terms of a functional of  $f'$ . We formulate this in abstract form as follows.

**Theorem 4.2:** *Let  $W$  be a real valued random variable having  $\mathbb{E}|W| \leq 1$ , and let  $f_z$  be the solution of the Stein equation (2.1). Suppose that there exist random variables  $R_1, R_2$  and  $M(t) \geq 0$ , and constants  $\delta_0, \delta_1$  and  $\delta_2$  that do not depend on  $z$ , such that*

$$\int_{|t| \leq \delta_0} M(t) dt = 1, \quad (4.7)$$

$$|R_1| \leq \delta_1, \quad |\mathbb{E}(R_2)| \leq \delta_2 \quad (4.8)$$

and

$$\mathbb{E}\{W f_z(W)\} = \mathbb{E} \left\{ \int_{|t| \leq \delta_0} f'_z(W + R_1 + t) M(t) dt \right\} + \mathbb{E}(R_2). \quad (4.9)$$

Then it follows that

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 2.1(\delta_0 + \delta_1) + \delta_2. \quad (4.10)$$

**Remark:** Note that, although the function  $M$  is random, its integral always takes the fixed value 1.

**Proof:** Since  $f_z$  satisfies the Stein equation (2.1), we have

$$\begin{aligned} & \mathbb{E} \left\{ \int_{|t| \leq \delta_0} f'_z(W + R_1 + t) M(t) dt \right\} \\ &= \mathbb{E} \left\{ \int_{|t| \leq \delta_0} (I_{\{W + R_1 + t \leq z\}} - \Phi(z)) M(t) dt \right\} \\ & \quad + \mathbb{E} \left\{ \int_{|t| \leq \delta_0} (W + R_1 + t) f_z(W + R_1 + t) M(t) dt \right\} \\ &\leq \mathbb{E} \left\{ \int_{|t| \leq \delta_0} (I_{\{W \leq z + \delta_0 + \delta_1\}} - \Phi(z)) M(t) dt \right\} \\ & \quad + \mathbb{E} \left\{ \int_{|t| \leq \delta_0} (W + R_1 + t) f_z(W + R_1 + t) M(t) dt \right\}, \end{aligned}$$

where the last line follows because  $-R_1 - t \leq \delta_0 + \delta_1$ . Hence, and since  $\int_{|t| \leq \delta_0} M(t) dt = 1$ , we find that

$$\begin{aligned} & \mathbb{E} \left\{ \int_{|t| \leq \delta_0} f'_z(W + R_1 + t) M(t) dt \right\} \\ &\leq \mathbb{P}(W \leq z + \delta_0 + \delta_1) - \Phi(z) \\ & \quad + \int_{|t| \leq \delta_0} \mathbb{E}\{(W + R_1 + t) f_z(W + R_1 + t) M(t)\} dt \\ &\leq \mathbb{P}(W \leq z + \delta_0 + \delta_1) - \Phi(z + \delta_0 + \delta_1) \\ & \quad + \frac{(\delta_0 + \delta_1)}{\sqrt{2\pi}} + \int_{|t| \leq \delta_0} \mathbb{E}\{(W + R_1 + t) f_z(W + R_1 + t) M(t)\} dt. \end{aligned}$$

Thus by (4.9), (4.8) and (4.7)

$$\begin{aligned}
& \mathbb{P}(W \leq z + \delta_0 + \delta_1) - \Phi(z + \delta_0 + \delta_1) \\
& \geq -\frac{(\delta_0 + \delta_1)}{\sqrt{2\pi}} - \mathbb{E}R_2 \\
& \quad + \int_{|t| \leq \delta_0} \mathbb{E}\{[Wf_z(W) - (W + R_1 + t)f_z(W + R_1 + t)]M(t)\} dt \\
& \geq -\frac{(\delta_0 + \delta_1)}{\sqrt{2\pi}} - \delta_2 - \int_{|t| \leq \delta_0} \mathbb{E}\{(|W| + \sqrt{2\pi}/4)(|R_1| + |t|)M(t)\} dt,
\end{aligned}$$

this last by (2.10), and so

$$\begin{aligned}
& \mathbb{P}(W \leq z + \delta_0 + \delta_1) - \Phi(z + \delta_0 + \delta_1) \\
& \geq -\frac{(\delta_0 + \delta_1)}{\sqrt{2\pi}} - \delta_2 - \mathbb{E} \left\{ \int_{|t| \leq \delta_0} (|W| + \sqrt{2\pi}/4)(\delta_0 + \delta_1)M(t) dt \right\} \\
& = -\frac{(\delta_0 + \delta_1)}{\sqrt{2\pi}} - \delta_2 - (\mathbb{E}|W| + \sqrt{2\pi}/4)(\delta_0 + \delta_1) \\
& \geq -2.1(\delta_0 + \delta_1) - \delta_2. \tag{4.11}
\end{aligned}$$

A similar argument gives

$$\mathbb{P}(W \leq z - \delta_0 - \delta_1) - \Phi(z - \delta_0 - \delta_1) \leq 2.1(\delta_0 + \delta_1) + \delta_2, \tag{4.12}$$

and this proves (4.10).  $\blacksquare$

To see why Theorem 4.1 is a special case of Theorem 4.2, let  $I$  be independent of  $\{\xi_i, 1 \leq i \leq n\}$  with  $\mathbb{P}(I = i) = \mathbb{E}\xi_i^2$  for  $i = 1, 2, \dots, n$ . Then we can rewrite (2.17) as

$$\begin{aligned}
\mathbb{E}Wf(W) &= \mathbb{E} \int_{|t| \leq \delta_0} f'(W^{(I)} + t)\tilde{K}_I(t) dt \\
&= \mathbb{E} \int_{|t| \leq \delta_0} f'(W + W^{(I)} - W + t)\tilde{K}_I(t) dt,
\end{aligned}$$

where  $\tilde{K}_i(t) = K_i(t)/\mathbb{E}\xi_i^2$ . Set  $R_1 = W^{(I)} - W$ ,  $R_2 = 0$  and  $M(t) = \tilde{K}_I(t)$  in Theorem 4.2. It is easy to see that conditions (4.7) – (4.9) are satisfied with  $\delta_1 = \delta_0$  and  $\delta_2 = 0$ . Hence (4.10) holds with a bound of  $4.2\delta_0$ .

In the next section, we illustrate how to use Theorem 4.2 to get a Berry–Esseen bound for the number of ones in the binary expansion of a random integer, using an exchangeable pair approach.

#### 4.2. Binary expansion of a random integer

Let  $n \geq 2$  be a natural number and  $X$  be a random variable uniformly distributed over the set  $\{0, 1, \dots, n-1\}$ . Let  $k$  be such that  $2^{k-1} < n \leq 2^k$ . Write the binary expansion of  $X$  as

$$X = \sum_{i=1}^k X_i 2^{k-i}$$

and let  $S = X_1 + \dots + X_k$  be the number of ones in the binary expansion of  $X$ . When  $n = 2^k$ , the distribution of  $S$  is the binomial distribution for  $k$  trials with probability  $1/2$ , and hence can be approximated by a normal distribution. We shall show that the normal approximation is good for *any* large  $n$ .

**Theorem 4.3:** Let  $k$  be such that  $2^{k-1} < n \leq 2^k$ , and set  $W = \frac{S-(k/2)}{\sqrt{k/4}}$ . Then

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 6.2k^{-1/2}. \quad (4.13)$$

**Proof:** We use the exchangeable pair approach. Let  $I$  be a random variable uniformly distributed over the set  $\{1, 2, \dots, k\}$  and independent of  $X$ , and let the random variable  $X'$  be defined by

$$X' = \sum_{i=1}^k X'_i 2^{k-i},$$

where

$$X'_i = \begin{cases} X_i & \text{if } i \neq I \\ 1 - X_I & \text{if } i = I \text{ and } X + 2^{k-I} < n \\ 0 & \text{if } X_I = 0 \text{ and } X + 2^{k-I} \geq n. \end{cases}$$

Also let  $S' = \sum_{i=1}^k X'_i$  and  $W' = \frac{S'-(k/2)}{\sqrt{k/4}}$ . The ordered pair  $(X, X')$  of random variables is exchangeable, and thus the pairs  $(S, S')$  and  $(W, W')$  are both exchangeable. Hence, for any function  $f$  on  $\{0, 1, \dots, k\}$ , we have

$$\begin{aligned} 0 &= \mathbb{E}\{(S' - S)(f(S) + f(S'))\} \\ &= 2\mathbb{E}\{(S' - S)f(S)\} + \mathbb{E}\{(S' - S)(f(S') - f(S))\} \\ &= 2\mathbb{E}\{f(S)\mathbb{E}(S' - S|X)\} + \mathbb{E}\{\mathbb{E}((S' - S)(f(S') - f(S))|X)\}. \end{aligned} \quad (4.14)$$

Observe that

$$\begin{aligned}
\mathbb{E}(S' - S|X) &= \mathbb{P}(S' - S = 1|X) - \mathbb{P}(S' - S = -1|X) \\
&= \mathbb{P}(X_I = 0, X'_I = 1|X) - \mathbb{P}(X_I = 1|X) \\
&= \mathbb{P}(X_I = 0|X) - \mathbb{P}(X_I = 0, X'_I = 0|X) - \mathbb{P}(X_I = 1|X) \\
&= \mathbb{E}(1 - X_I|X) - \mathbb{P}(X_I = 0, X'_I = 0|X) - \mathbb{P}(X_I = 1|X) \\
&= \frac{1}{k} \sum_{i=1}^k (1 - X_i) - \frac{1}{k} \sum_{i=1}^k I_{\{X_i=0, X+2^{k-i} \geq n\}} - \frac{1}{k} \sum_{i=1}^k X_i \\
&= 1 - \frac{2S}{k} - \frac{Q}{k}, \tag{4.15}
\end{aligned}$$

where  $Q = \sum_{i=1}^k I_{\{X_i=0, X+2^{k-i} \geq n\}}$ . Now rewrite (4.14), with a re-definition of the (arbitrary) function  $f$ , as

$$k^{1/2} \mathbb{E}\{f(W) \mathbb{E}(S - S'|X)\} = \frac{1}{2} k^{1/2} \mathbb{E}\{\mathbb{E}((f(W') - f(W))(S' - S)|X)\}. \tag{4.16}$$

The left hand side of (4.16) is

$$\begin{aligned}
k^{1/2} \mathbb{E}\left\{f(W) \left(-1 + \frac{2S}{k} + \frac{Q}{k}\right)\right\} &= \mathbb{E}\left\{f(W) \left(\frac{S - k/2}{k^{1/2}/2} + \frac{Q}{k^{1/2}}\right)\right\} \\
&= \mathbb{E}\{W f(W)\} + k^{-1/2} \mathbb{E}Q f(W), \tag{4.17}
\end{aligned}$$

and the right hand side of (4.16) can be written as

$$\begin{aligned}
&\frac{1}{2} k^{1/2} \mathbb{E}\left\{\mathbb{E}\{(f(W') - f(W)) I_{\{S' - S = 1\}}|X\} \right. \\
&\quad \left. - \mathbb{E}\{(f(W') - f(W)) I_{\{S' - S = -1\}}|X\}\right\} \\
&= \frac{1}{2} k^{1/2} \mathbb{E}\left\{\mathbb{E}\{((f(W + 2k^{-1/2}) - f(W)) I_{\{S' - S = 1\}}|X\} \right. \\
&\quad \left. - \mathbb{E}\{((f(W - 2k^{-1/2}) - f(W)) I_{\{S' - S = -1\}}|X\}\right\} \\
&= \frac{1}{2} k^{1/2} \mathbb{E}\left\{(f(W + 2k^{-1/2}) - f(W)) \mathbb{P}(S' - S = 1|X) \right. \\
&\quad \left. - (f(W - 2k^{-1/2}) - f(W)) \mathbb{P}(S' - S = -1|X)\right\} \\
&= \frac{1}{2} k^{1/2} \mathbb{E}\left\{(f(W + 2k^{-1/2}) - f(W)) \left(1 - \frac{S}{k} - \frac{Q}{k}\right) \right. \\
&\quad \left. - (f(W - 2k^{-1/2}) - f(W)) \frac{S}{k}\right\}.
\end{aligned}$$

The latter expression can in turn be written as

$$\begin{aligned}
& \mathbb{E}\left\{(f(W + 2k^{-1/2}) - f(W))\left(\frac{k - S}{2k^{1/2}}\right)\right\} \\
& \quad - \mathbb{E}\left\{(f(W - 2k^{-1/2}) - f(W))\left(\frac{S}{2k^{1/2}}\right)\right\} \\
& \quad - \mathbb{E}\left\{(f(W + 2k^{-1/2}) - f(W))\frac{Q}{2k^{1/2}}\right\} \\
& = \mathbb{E}\int_{|t| \leq 2k^{-1/2}} f'(W + t)M(t) dt - \mathbb{E}\left\{(f(W + 2k^{-1/2}) - f(W))\frac{Q}{2k^{1/2}}\right\},
\end{aligned} \tag{4.18}$$

where

$$M(t) = \begin{cases} \frac{k-S}{2k^{1/2}} & \text{for } 0 \leq t \leq \frac{2}{k^{1/2}} \\ \frac{S}{2k^{1/2}} & \text{for } -\frac{2}{k^{1/2}} \leq t < 0. \end{cases}$$

Note that  $M(t) \geq 0$  and  $\int_{|t| \leq 2k^{-1/2}} M(t) dt = 1$ . So, taking  $f = f_z$  as given in (2.2), we have

$$\mathbb{E}\{Wf_z(W)\} = \int_{|t| \leq 2k^{-1/2}} \mathbb{E}\{f'_z(W + t)M(t)\} dt + \mathbb{E}(R_2), \tag{4.19}$$

of the form (4.9) with  $R_1 = 0$  and  $\delta_0 = 2k^{-1/2}$ , where

$$R_2 = -\left\{(f(W + 2k^{-1/2}) - f(W))\frac{Q}{2k^{1/2}}\right\} - k^{-1/2}Qf(W).$$

Then, by (2.9), it follows that

$$|\mathbb{E}(R_2)| \leq \left(\frac{\sqrt{2\pi}}{8} + \frac{\sqrt{2\pi}}{4}\right)\frac{\mathbb{E}Q}{\sqrt{k}} \leq 2k^{-1/2},$$

since  $\mathbb{E}Q \leq 2$ , as shown below. It thus follows from Theorem 4.2 that

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 2.1(2k^{-1/2}) + 2k^{-1/2} = 6.2k^{-1/2},$$

as desired.

To show that  $\mathbb{E}Q \leq 2$ , simply observe that, from its definition,

$$\begin{aligned}
\mathbb{E}Q &= \sum_{i=1}^k \mathbb{P}(X_i = 0, X + 2^{k-i} \geq n) \leq \sum_{i=1}^k \mathbb{P}(X \geq n - 2^{k-i}) \\
&= \sum_{i=1}^k \frac{2^{k-i}}{n} = \frac{2^k - 1}{n} \leq \frac{2^k - 1}{2^{k-1}} \leq 2.
\end{aligned}$$

This completes the proof. ■

The binary expansion of a random integer has previously studied by a number of authors. Diaconis (1977) and Stein (1986) also proved that the distribution of  $S$ , the number of ones in the binary expansion, is only order  $O(k^{-1/2})$  away from the  $B(k, 1/2)$ , whereas Barbour & Chen (1992) further proved that, if a mixture of the  $B(k-1, 1/2)$  and  $B(k, 1/2)$  is used as an approximation, the error can be reduced to  $O(k^{-1})$ .

### 5. Uniform Berry–Esseen bounds: the independent case

Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables with zero means and  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . Let

$$\gamma = \sum_{i=1}^n \mathbb{E}|\xi_i|^3 \quad (5.1)$$

and  $W = \sum_{i=1}^n \xi_i$ . If  $\mathbb{E}|\xi_i|^3 < \infty$ , then we have the uniform Berry–Esseen inequality

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq C_0 \gamma \quad (5.2)$$

and the non-uniform Berry–Esseen inequality

$$\forall z \in \mathbb{R}^1, \quad |\mathbb{P}(W \leq z) - \Phi(z)| \leq C_1(1 + |z|)^{-3} \gamma, \quad (5.3)$$

where both  $C_0$  and  $C_1$  are absolute constants. One can take  $C_0 = 0.7975$  [van Beeck (1972)] and  $C_1 = 31.935$  [Paditz (1989)]. We shall use the concentration inequality approach to give a direct proof of (5.2) in this section and (5.3) in next section, albeit with different constants. We refer to Chen & Shao (2001) for more details.

#### 5.1. The concentration inequality approach

As indicated in (1.10) and by (4.4) in the proof of Theorem 4.1, a key step in proving the Berry–Esseen bound is to have a good bound for the probability  $\mathbb{P}(a \leq W^{(i)} \leq b)$ , where  $W^{(i)} = \sum_{j \neq i} \xi_j$ . In this section, we prove such a concentration inequality. The idea is once again to use the Stein identity. More precisely, we use the fact that  $\mathbb{E}f'(W)$  is close to  $\mathbb{E}\{Wf(W)\}$  when  $W$  is close to the normal. If  $f'$  is taken to be the indicator  $\mathbf{1}_{[a,b]}$ , then  $\mathbb{E}f'(W) = \mathbb{P}(a \leq W \leq b)$ ; on the other hand, choosing  $f(\frac{1}{2}(b-a)) = 0$ , it follows that  $\|f\| = \frac{1}{2}(b-a)$ , so that

$$|\mathbb{E}\{Wf(W)\}| \leq \frac{1}{2}(b-a)\mathbb{E}|W| \leq \frac{1}{2}(b-a)$$

if  $\mathbb{E}W^2 = 1$ . Thus, if  $\mathbb{E}f'(W)$  and  $\mathbb{E}\{Wf(W)\}$  are close, it follows that  $\mathbb{P}(a \leq W \leq b)$  is close to  $\frac{1}{2}(b-a)$ . The proof of the inequality below makes this heuristic precise.

**Proposition 5.1:** *We have*

$$\mathbb{P}(a \leq W^{(i)} \leq b) \leq \sqrt{2}(b-a) + (1 + \sqrt{2})\gamma \quad (5.4)$$

for all real  $a < b$  and for every  $1 \leq i \leq n$ .

**Proof:** Define  $\delta = \gamma/2$  and take

$$f(w) = \begin{cases} -\frac{1}{2}(b-a) - \delta & \text{if } w < a - \delta, \\ w - \frac{1}{2}(b+a) & \text{if } a - \delta \leq w \leq b + \delta, \\ \frac{1}{2}(b-a) + \delta & \text{for } w > b + \delta, \end{cases} \quad (5.5)$$

so that  $f' = \mathbf{1}_{[a-\delta, b+\delta]}$  and  $\|f\| = \frac{1}{2}(b-a) + \delta$ . Set

$$\begin{aligned} \widehat{M}_j(t) &= \xi_j(I_{\{-\xi_j \leq t \leq 0\}} - I_{\{0 < t \leq -\xi_j\}}) \geq 0, \\ \widehat{M}(t) &= \sum_{1 \leq j \leq n} \widehat{M}_j(t), \quad M(t) = \mathbb{E}\widehat{M}(t). \end{aligned} \quad (5.6)$$

Since  $\xi_j$  and  $W^{(i)} - \xi_j$  are independent for  $j \neq i$ ,  $\xi_i$  is independent of  $W^{(i)}$  and  $\mathbb{E}\xi_j = 0$  for all  $j$ , we have

$$\begin{aligned} &\mathbb{E}\{W^{(i)}f(W^{(i)})\} - \mathbb{E}\{\xi_i f(W^{(i)} - \xi_i)\} \\ &= \sum_{j=1}^n \mathbb{E}\{\xi_j [f(W^{(i)}) - f(W^{(i)} - \xi_j)]\} = \sum_{j=1}^n \mathbb{E}\left\{\xi_j \int_{-\xi_j}^0 f'(W^{(i)} + t) dt\right\}. \end{aligned}$$

Hence, using (5.6), we have

$$\begin{aligned} &\mathbb{E}\{W^{(i)}f(W^{(i)})\} - \mathbb{E}\{\xi_i f(W^{(i)} - \xi_i)\} \\ &= \sum_{j=1}^n \mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W^{(i)} + t) \widehat{M}_j(t) dt\right\} = \mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W^{(i)} + t) \widehat{M}(t) dt\right\}. \end{aligned}$$

At this point, we have reached a precise replacement for the statement ' $\mathbb{E}f'(W)$  is close to  $\mathbb{E}\{Wf(W)\}$ ' of the heuristic, with  $W^{(i)}$  for  $W$ . To exploit it, we first note that

$$\mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W^{(i)} + t) \widehat{M}(t) dt\right\} \geq \mathbb{E}\left\{\int_{|t| \leq \delta} f'(W^{(i)} + t) \widehat{M}(t) dt\right\},$$



because  $f'(t) \geq 0$  and  $\widehat{M}(t) \geq 0$ . But now the definition of  $f$  implies that

$$\begin{aligned} \mathbb{E} \left\{ \int_{|t| \leq \delta} f'(W^{(i)} + t) \widehat{M}(t) dt \right\} &\geq \mathbb{E} \left\{ I_{\{a \leq W^{(i)} \leq b\}} \int_{|t| \leq \delta} \widehat{M}(t) dt \right\} \\ &= \mathbb{E} \left\{ I_{\{a \leq W^{(i)} \leq b\}} \sum_{j=1}^n |\xi_j| \min(\delta, |\xi_j|) \right\} \geq H_{1,1} - H_{1,2}, \end{aligned} \quad (5.7)$$

where

$$H_{1,1} = \mathbb{P}(a \leq W^{(i)} \leq b) \sum_{j=1}^n \mathbb{E} |\xi_j| \min(\delta, |\xi_j|)$$

and

$$H_{1,2} = \mathbb{E} \left| \sum_{j=1}^n \{ |\xi_j| \min(\delta, |\xi_j|) - \mathbb{E} |\xi_j| \min(\delta, |\xi_j|) \} \right|.$$

A direct calculation yields

$$\min(x, y) \geq x - x^2/(4y) \quad (5.8)$$

for  $x > 0$  and  $y > 0$ , implying that

$$\sum_{j=1}^n \mathbb{E} |\xi_j| \min(\delta, |\xi_j|) \geq \sum_{j=1}^n \left\{ \mathbb{E} \xi_j^2 - \frac{\mathbb{E} |\xi_j|^3}{4\delta} \right\} = \frac{1}{2}, \quad (5.9)$$

by choice of  $\delta$ , and hence that

$$H_{1,1} \geq \frac{1}{2} \mathbb{P}(a \leq W^{(i)} \leq b). \quad (5.10)$$

By the Hölder inequality,

$$\begin{aligned} H_{1,2} &\leq \left( \text{Var} \left\{ \sum_{j=1}^n |\xi_j| \min(\delta, |\xi_j|) \right\} \right)^{1/2} \\ &\leq \left( \sum_{j=1}^n \mathbb{E} \xi_j^2 \min(\delta, |\xi_j|)^2 \right)^{1/2} \\ &\leq \delta \left( \sum_{j=1}^n \mathbb{E} \xi_j^2 \right)^{1/2} = \delta. \end{aligned} \quad (5.11)$$

Hence

$$\mathbb{E} \left\{ \int_{-\infty}^{\infty} f'(W^{(i)} + t) \widehat{M}(t) dt \right\} \geq \frac{1}{2} \mathbb{P}(a \leq W^{(i)} \leq b) - \delta, \quad (5.12)$$

to be compared with the equation ‘ $\mathbb{E}f'(W) = \mathbb{P}(a \leq W \leq b)$ ’ of the heuristic.

On the other hand, recalling that  $\|f\| \leq \frac{1}{2}(b-a) + \delta$ , we have

$$\begin{aligned}
& \mathbb{E}\{W^{(i)}f(W^{(i)})\} - \mathbb{E}\{\xi_i f(W^{(i)} - \xi_i)\} \\
& \leq \left\{ \frac{1}{2}(b-a) + \delta \right\} (\mathbb{E}|W^{(i)}| + \mathbb{E}|\xi_i|) \\
& \leq \frac{1}{\sqrt{2}} \left( (\mathbb{E}|W^{(i)}|^2) + (\mathbb{E}|\xi_i|^2) \right)^{1/2} (b-a+2\delta) \\
& \leq \frac{1}{\sqrt{2}} \left( \mathbb{E}|W^{(i)}|^2 + \mathbb{E}|\xi_i|^2 \right)^{1/2} (b-a+2\delta) \\
& = \frac{1}{\sqrt{2}} (b-a+2\delta), \tag{5.13}
\end{aligned}$$

the inequality to be compared with ‘ $|\mathbb{E}\{Wf(W)\}| \leq \frac{1}{2}(b-a)$ ’ from the heuristic. Combining (5.12) and (5.13) thus gives

$$\mathbb{P}(a \leq W^{(i)} \leq b) \leq \sqrt{2}(b-a) + 2(1+\sqrt{2})\delta = \sqrt{2}(b-a) + (1+\sqrt{2})\gamma$$

as desired. ■

## 5.2. Proving the Berry–Esseen theorem

We are now ready to prove the classical Berry–Esseen theorem, with a constant of 7.

**Theorem 5.2:** *We have*

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 7\gamma. \tag{5.14}$$

**Proof:** As discussed in the previous section, we need to find a way to replace  $\mathbb{P}(W^{(i)} + t \leq z)$  by  $\mathbb{P}(W \leq z)$  in (4.3). However, it follows from (5.4) that

$$\begin{aligned}
& \left| \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{P}(W^{(i)} + t \leq z) K_i(t) dt - \mathbb{P}(W \leq z) \right| \\
& \leq \sum_{i=1}^n \int_{-\infty}^{\infty} |\mathbb{P}(W^{(i)} + t \leq z) - \mathbb{P}(W \leq z)| K_i(t) dt \\
& = \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E}\{\mathbb{P}(z - \max(t, \xi_i) \leq W^{(i)} \leq z - \min(t, \xi_i) \mid \xi_i)\} K_i(t) dt.
\end{aligned}$$

Proposition 5.1 thus gives the bound

$$\begin{aligned}
& \left| \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{P}(W^{(i)} + t \leq z) K_i(t) dt - \mathbb{P}(W \leq z) \right| \\
& \leq \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbb{E} \{ \sqrt{2}(|t| + |\xi_i|) + (1 + \sqrt{2})\gamma \} K_i(t) dt \\
& = (1 + \sqrt{2})\gamma + \sqrt{2} \sum_{i=1}^n \left( \frac{1}{2} \mathbb{E} |\xi_i|^3 + \mathbb{E} |\xi_i| \mathbb{E} \xi_i^2 \right) \\
& \leq (1 + 2.5\sqrt{2})\gamma.
\end{aligned} \tag{5.15}$$

Now by (4.2)

$$|\mathbb{P}(W \leq z) - \Phi(z)| \leq (1 + 2.5\sqrt{2} + 1.5(1 + \sqrt{2\pi}/4))\gamma \leq 7\gamma,$$

which is (4.1). ■

We remark that following the above lines of proof, one can prove that

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 7 \sum_{i=1}^n (\mathbb{E} \xi_i^2 I_{\{|\xi_i| > 1\}} + \mathbb{E} |\xi_i|^3 I_{\{|\xi_i| \leq 1\}}),$$

dispensing with the third moment assumption. We leave the proof to the reader. With a more refined concentration inequality, the constant 7 can be reduced to 4.1 (see Chen & Shao (2001)).

### 5.3. A lower bound

Let  $X_i$ ,  $i \geq 1$ , be independent random variables with zero means and finite variances, and define  $B_n = \sum_{i=1}^n \text{Var} X_i$ . It is known that if the Feller condition

$$\max_{1 \leq i \leq n} \mathbb{E} X_i^2 / B_n^2 \rightarrow 0, \tag{5.16}$$

is satisfied, then the Lindeberg condition is necessary for the central limit theorem. Barbour & Hall (1984) used Stein's method to provide not only a nice proof of the necessity, but also a lower bound for the Kolmogorov distance between the distribution of  $W$  and the normal, which is as close as can be expected to being a multiple of one of Lindeberg's sums  $B_n^{-2} \sum_{i=1}^n \mathbb{E} \{ X_i^2 I_{\{B_n^{-1} |X_i| > \varepsilon\}} \}$ . Note that no lower bound could simply be a multiple of a Lindeberg sum, since a sum of  $n$  identically distributed *normal* random variables is itself normally distributed, but the corresponding Lindeberg sum is not zero. The next theorem and its proof are due to Barbour & Hall (1984).

**Theorem 5.3:** Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables which have zero means and finite variances  $\mathbb{E}\xi_i^2 = \sigma_i^2$ ,  $1 \leq i \leq n$ , and satisfy  $\sum_{i=1}^n \sigma_i^2 = 1$ . Then there exists an absolute constant  $C$  such that for all  $\varepsilon > 0$

$$(1 - e^{-\varepsilon^2/4}) \sum_{i=1}^n \mathbb{E}\{\xi_i^2 I_{\{|\xi_i| > \varepsilon\}}\} \leq C \left( \sup_z |\mathbb{P}(W \leq z) - \Phi(z)| + \sum_{i=1}^n \sigma_i^4 \right). \quad (5.17)$$

**Remark:** For the sequence of independent random variables  $\{X_i, i \geq 1\}$ , we take  $\xi_i = X_i/B_n$ . Clearly, Feller's negligibility condition (5.16) implies that  $\sum_{i=1}^n \sigma_i^4 \leq \max_{1 \leq i \leq n} \sigma_i^2 \rightarrow 0$  as  $n \rightarrow \infty$ . Therefore, if  $S_n/B_n$  is asymptotically normal, then

$$\sum_{i=1}^n \mathbb{E}\{\xi_i^2 I_{\{|\xi_i| > \varepsilon\}}\} \rightarrow 0$$

as  $n \rightarrow \infty$  for every  $\varepsilon > 0$ , and the Lindeberg condition is satisfied.

**Proof:** Once again, the argument starts with the Stein identity

$$\mathbb{E}\{f'_h(W) - W f_h(W)\} = \mathbb{E}h(W) - \mathbb{E}h(Z),$$

for  $h$  yet to be chosen. Integrating by parts, the right hand side is bounded by

$$\begin{aligned} |\mathbb{E}h(W) - \mathbb{E}h(Z)| &= \left| \int_{-\infty}^{\infty} h'(w) \{\mathbb{P}(W \leq w) - \Phi(w)\} dw \right| \\ &\leq \Delta \int_{-\infty}^{\infty} |h'(w)| dw, \end{aligned} \quad (5.18)$$

where  $\Delta = \sup_z |\mathbb{P}(W \leq z) - \Phi(z)|$ . For the left hand side, in the usual way, because  $\xi_i$  and  $W^{(i)}$  are independent and  $\mathbb{E}\xi_i = 0$ , we have

$$\begin{aligned} \mathbb{E}\{W f_h(W)\} &= \sum_{i=1}^n \mathbb{E}\{\xi_i^2 f'_h(W^{(i)})\} \\ &\quad + \sum_{i=1}^n \mathbb{E}\{\xi_i (f_h(W^{(i)} + \xi_i) - f_h(W^{(i)})) - \xi_i f'_h(W^{(i)})\}, \end{aligned}$$

and, because  $\sum_{i=1}^n \sigma_i^2 = 1$ ,

$$\begin{aligned} \mathbb{E}f'_h(W) &= \sum_{i=1}^n \sigma_i^2 \mathbb{E}f'_h(W^{(i)} + \xi_i) \\ &= \sum_{i=1}^n \sigma_i^2 \mathbb{E}f'_h(W^{(i)}) + \sum_{i=1}^n \sigma_i^2 \mathbb{E}\{f'_h(W) - f'_h(W^{(i)})\}, \end{aligned}$$

with the last term easily bounded by  $\frac{1}{2}\|f_h'''\| \sum_{i=1}^n \sigma_i^4$ . Hence

$$\left| \mathbb{E}\{f_h'(W) - W f_h(W)\} - \sum_{i=1}^n \mathbb{E}\{\xi_i^2 g(W^{(i)}, \xi_i)\} \right| \leq \frac{1}{2}\|f_h'''\| \sum_{i=1}^n \sigma_i^4, \quad (5.19)$$

where

$$g(w, y) = g_h(w, y) = -y^{-1}\{f_h(w+y) - f_h(w) - y f_h'(w)\}.$$

Intuitively, if  $W$  is close to being normally distributed, then

$$R_1 := \sum_{i=1}^n \mathbb{E}\{\xi_i^2 g(W^{(i)}, \xi_i)\} \quad \text{and} \quad R := \sum_{i=1}^n \mathbb{E}\{\xi_i^2 g(Z, \xi_i)\},$$

for  $Z \sim \mathcal{N}(0, 1)$  independent of the  $\xi_i$ 's, should be close to one another.

Taking (5.18) and (5.19) together, it follows that we have a lower bound for  $\Delta$ , provided that a function  $h$  can be found with  $\int_{-\infty}^{\infty} |h'(w)| dw < \infty$  for which  $\mathbb{E}g_h(Z, y)$  is of constant sign, and provided also that  $\|f_h'''\| < \infty$ . In practice, it is easier to look for a suitable  $f$ , and then define  $h$  by setting  $h(w) = f'(w) - w f(w)$ . The function  $g$  is zero for any linear function  $f$ , and, for even functions  $f$ , it follows that  $\mathbb{E}g(Z, y)$  is antisymmetric in  $y$  about zero, but odd functions are possibilities; for instance,  $f(x) = x^3$  has  $\mathbb{E}g(Z, y) = -y^2$ , of constant sign. Unfortunately, this  $f$  fails to have finite  $\int_{-\infty}^{\infty} |h'(w)| dw$ , but the choice  $f(w) = w e^{-w^2/2}$  is a good one; it behaves much like the sum of a linear and a cubic function for those values of  $w$  where  $\mathcal{N}(0, 1)$  puts most of its mass, yet dies away to zero fast when  $|w|$  is large. Making the computations,

$$\begin{aligned} \mathbb{E}g(Z, y) &= -\frac{1}{y\sqrt{2\pi}} \int_{-\infty}^{\infty} \left\{ (w+y)e^{-(w+y)^2/2} \right. \\ &\quad \left. - w e^{-w^2/2} - y e^{-w^2/2}(1-w^2) \right\} e^{-w^2/2} dw \\ &= -y^{-1} \left\{ \frac{y}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(w^2 + (w+y)^2)\right\} dw - \frac{y}{2}\sqrt{\frac{\pi}{2}} \right\} \\ &= \frac{1}{2\sqrt{2}}(1 - e^{-y^2/4}) \geq 0, \end{aligned} \quad (5.20)$$

and  $\mathbb{E}g(Z, y) \geq \frac{1}{2\sqrt{2}}(1 - e^{-\varepsilon^2/4})$  whenever  $|y| \geq \varepsilon$ . Hence, for this choice of  $f$ , we have

$$R \geq \frac{1}{2\sqrt{2}}(1 - e^{-\varepsilon^2/4}) \sum_{i=1}^n \mathbb{E}\{\xi_i^2 I_{\{|\xi_i| > \varepsilon\}}\} \quad (5.21)$$

for any  $\varepsilon > 0$ . It thus remains to show that  $R$  and  $R_1$  are close enough, after which (5.18), (5.19) and (5.21) complete the proof.

To show this, note first that, taking this choice of  $f$ , and setting  $h(w) = f'(w) - wf(w)$ , we have

$$c_1 := \int_{-\infty}^{\infty} |h'(w)| dw \leq 5; \quad c_2 := \int_{-\infty}^{\infty} |f''(w)| dw \leq 4; \quad c_3 := \sup_w |f'''(w)| = 3.$$

Now define an intermediate  $R_2$  between  $R_1$  and  $R$ , by

$$R_2 := \sum_{i=1}^n \mathbb{E}\{\xi_i^2 g(W^*, \xi_i)\},$$

where  $W^*$  has the same distribution as  $W$ , but is independent of the  $\xi_i$ 's. Then

$$\begin{aligned} R_1 &= - \sum_{i=1}^n \mathbb{E}\left\{\xi_i^2 \int_0^1 [f'(W^{(i)} + t\xi_i) - f'(W^{(i)})] dt\right\} \\ &= R_2 + \sum_{i=1}^n \mathbb{E}\left\{\xi_i^2 \int_0^1 [f'(W^* + t\xi_i) - f'(W^{(i)} + t\xi_i)] dt\right\} \\ &\quad - \sum_{i=1}^n \mathbb{E}\left\{\xi_i^2 \int_0^1 [f'(W^*) - f'(W^{(i)})] dt\right\}. \end{aligned} \quad (5.22)$$

Now, for any  $\theta$ , and because  $\xi_i$  and  $W^{(i)}$  are independent, with  $\mathbb{E}\xi_i = 0$ ,

$$\begin{aligned} &|\mathbb{E}(f'(W^* + \theta) - f'(W^{(i)} + \theta))| \\ &= |\mathbb{E}(f'(W^{(i)} + \xi_i + \theta) - f'(W^{(i)} + \theta))| \\ &= |\mathbb{E}(f'(W^{(i)} + \xi_i + \theta) - f'(W^{(i)} + \theta) - \xi_i f''(W^{(i)} + \theta))| \\ &\leq \frac{1}{2} c_3 \sigma_i^2, \end{aligned}$$

by Taylor's theorem. Hence, from (5.22) and because  $\sum_{i=1}^n \sigma_i^2 = 1$ ,

$$R_1 \geq R_2 - c_3 \sum_{i=1}^n \sigma_i^4. \quad (5.23)$$

Similarly,

$$\begin{aligned} R_2 &= R + \sum_{i=1}^n \mathbb{E}\left\{\xi_i^2 \int_0^1 [f'(Z) + t\xi_i] - f'(W^* + t\xi_i)] dt\right\} \\ &\quad - \sum_{i=1}^n \mathbb{E}\left\{\xi_i^2 \int_0^1 [f'(Z) - f'(W^*)] dt\right\}, \end{aligned}$$

and, for any  $\theta$ ,

$$\begin{aligned} & |\mathbb{E}f'(W^* + \theta) - \mathbb{E}f'(Z + \theta)| \\ &= \left| \int_{-\infty}^{\infty} f''(w) (\mathbb{P}(W^* \leq w - \theta) - \Phi(w - \theta)) dw \right| \leq c_2 \Delta, \end{aligned}$$

so that

$$R_2 \geq R - 2c_2 \Delta. \quad (5.24)$$

Combining (5.18) and (5.19) with (5.23) and (5.24), it follows that

$$c_1 \Delta \geq R_1 - \frac{1}{2} c_3 \sum_{i=1}^n \sigma_i^4 \geq R - \frac{3}{2} c_3 \sum_{i=1}^n \sigma_i^4 - 2c_2 \Delta.$$

In view of (5.21), collecting terms, it follows that

$$\Delta(c_1 + 2c_2) + \frac{3}{2} c_3 \sum_{i=1}^n \sigma_i^4 \geq \frac{1}{2\sqrt{2}} (1 - e^{-\varepsilon^2/4}) \sum_{i=1}^n \mathbb{E}\{\xi_i^2 I_{\{|\xi_i| > \varepsilon\}}\} \quad (5.25)$$

for any  $\varepsilon > 0$ . This proves (5.17), with  $C \leq 40$ .  $\blacksquare$

## 6. Non-uniform Berry–Esseen bounds: the independent case

In this section, we prove a non-uniform Berry–Esseen bound similar to (5.3), following the proof in Chen & Shao (2001). To do this, as in the previous section, we first need a concentration inequality; it must itself now be non-uniform.

### 6.1. A non-uniform concentration inequality

Let  $\xi_1, \xi_2, \dots, \xi_n$  be independent random variables satisfying  $\mathbb{E}\xi_i = 0$  for every  $1 \leq i \leq n$  and  $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$ . Let

$$\bar{\xi}_i = \xi_i I_{\{\xi_i \leq 1\}}, \quad \bar{W} = \sum_{i=1}^n \bar{\xi}_i, \quad \bar{W}^{(i)} = \bar{W} - \bar{\xi}_i.$$

**Proposition 6.1:** *We have*

$$\mathbb{P}(a \leq \bar{W}^{(i)} \leq b) \leq e^{-a/2} (5(b-a) + 7\gamma) \quad (6.1)$$

for all real  $b > a$  and for every  $1 \leq i \leq n$ , where  $\gamma = \sum_{i=1}^n \mathbb{E}|\xi_i|^3$ .

To prove it, we first need to have the following Bennett–Hoeffding inequality.

**Lemma 6.2:** *Let  $\eta_1, \eta_2, \dots, \eta_n$  be independent random variables satisfying  $\mathbb{E}\eta_i \leq 0$ ,  $\eta_i \leq \alpha$  for  $1 \leq i \leq n$ , and  $\sum_{i=1}^n \mathbb{E}\eta_i^2 \leq B_n^2$ . Put  $S_n = \sum_{i=1}^n \eta_i$ . Then*

$$\mathbb{E}e^{tS_n} \leq \exp\left(\alpha^{-2}(e^{t\alpha} - 1 - t\alpha)B_n^2\right) \quad (6.2)$$

for  $t > 0$ ,

$$\mathbb{P}(S_n \geq x) \leq \exp\left(-\frac{B_n^2}{\alpha^2}\left[\left(1 + \frac{\alpha x}{B_n^2}\right)\ln\left(1 + \frac{\alpha x}{B_n^2}\right) - \frac{\alpha x}{B_n^2}\right]\right) \quad (6.3)$$

and

$$\mathbb{P}(S_n \geq x) \leq \exp\left(-\frac{x^2}{2(B_n^2 + \alpha x)}\right) \quad (6.4)$$

for  $x > 0$ .

**Proof:** It is easy to see that  $(e^s - 1 - s)/s^2$  is an increasing function of  $s \in \mathbb{R}$ , from which it follows that

$$e^{ts} \leq 1 + ts + (ts)^2(e^{t\alpha} - 1 - t\alpha)/(t\alpha)^2 \quad (6.5)$$

for  $s \leq \alpha$ , if  $t > 0$ . Using the properties of the  $\eta_i$ 's, we thus have

$$\begin{aligned} \mathbb{E}e^{tS_n} &= \prod_{i=1}^n \mathbb{E}e^{t\eta_i} \\ &\leq \prod_{i=1}^n \left(1 + t\mathbb{E}\eta_i + \alpha^{-2}(e^{t\alpha} - 1 - t\alpha)\mathbb{E}\eta_i^2\right) \\ &\leq \prod_{i=1}^n \left(1 + \alpha^{-2}(e^{t\alpha} - 1 - t\alpha)\mathbb{E}\eta_i^2\right) \\ &\leq \exp\left(\alpha^{-2}(e^{t\alpha} - 1 - t\alpha)B_n^2\right). \end{aligned}$$

This proves (6.2).

To prove inequality (6.3), let

$$t = \frac{1}{\alpha} \ln\left(1 + \frac{\alpha x}{B_n^2}\right).$$



Then, by (6.2),

$$\begin{aligned} \mathbb{P}(S_n \geq x) &\leq e^{-tx} \mathbb{E} e^{tS_n} \\ &\leq \exp\left(-tx + \alpha^{-2}(e^{t\alpha} - 1 - t\alpha)B_n^2\right) \\ &= \exp\left(-\frac{B_n^2}{\alpha^2} \left[ \left(1 + \frac{\alpha x}{B_n^2}\right) \ln\left(1 + \frac{\alpha x}{B_n^2}\right) - \frac{\alpha x}{B_n^2} \right]\right). \end{aligned}$$

In view of the fact that

$$(1+s)\ln(1+s) - s \geq \frac{s^2}{2(1+s)}$$

for  $s > 0$ , (6.4) follows from (6.3).  $\blacksquare$

**Proof of Proposition 6.1.** It follows from (6.2) with  $\alpha = 1$  and  $B_n^2 = 1$  that

$$\begin{aligned} \mathbb{P}(a \leq \bar{W}^{(i)} \leq b) &\leq e^{-a/2} \mathbb{E} e^{\bar{W}^{(i)}/2} \\ &\leq e^{-a/2} \exp(e^{1/2} - \frac{3}{2}) \leq 1.19 e^{-a/2}. \end{aligned}$$

Thus, (6.1) holds if  $7\gamma \geq 1.19$ .

We now assume that  $\gamma \leq 1.19/7 = 0.17$ . Much as in the proof of Proposition 5.1, define  $\delta = \gamma/2$  ( $\leq 0.085$ ), and set

$$f(w) = \begin{cases} 0 & \text{if } w < a - \delta, \\ e^{w/2}(w - a + \delta) & \text{if } a - \delta \leq w \leq b + \delta, \\ e^{w/2}(b - a + 2\delta) & \text{if } w > b + \delta. \end{cases} \quad (6.6)$$

Put

$$\bar{M}_i(t) = \xi_i(I_{\{-\bar{\xi}_i \leq t \leq 0\}} - I_{\{0 < t \leq -\bar{\xi}_i\}}), \quad \bar{M}^{(i)}(t) = \sum_{j \neq i} \bar{M}_j(t).$$

Clearly,  $\bar{M}^{(i)}(t) \geq 0$ ,  $f'(w) \geq 0$  and  $f'(w) \geq e^{w/2}$  for  $a - \delta \leq w \leq b + \delta$ .

Arguing much as in the derivation of (5.7), we obtain

$$\begin{aligned} \mathbb{E}\{W^{(i)} f(\bar{W}^{(i)})\} &= \sum_{j \neq i} \mathbb{E} \left\{ \xi_j [f(\bar{W}^{(i)}) - f(\bar{W}^{(i)} - \bar{\xi}_j)] \right\} \\ &= \sum_{j \neq i} \mathbb{E} \left\{ \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t) \bar{M}_j(t) dt \right\} \\ &= \mathbb{E} \left\{ \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t) \bar{M}^{(i)}(t) dt \right\}. \end{aligned}$$

This expression is now bounded below, giving

$$\begin{aligned}
\mathbb{E}\{W^{(i)}f(\bar{W}^{(i)})\} &\geq \mathbb{E}\left\{I_{\{a \leq \bar{W}^{(i)} \leq b\}} \int_{|t| \leq \delta} f'(\bar{W}^{(i)} + t) \bar{M}^{(i)}(t) dt\right\} \\
&\geq \mathbb{E}\left\{e^{(\bar{W}^{(i)} - \delta)/2} I_{\{a \leq \bar{W}^{(i)} \leq b\}} \int_{|t| \leq \delta} \bar{M}^{(i)}(t) dt\right\} \\
&\geq \mathbb{E}\left\{e^{(\bar{W}^{(i)} - \delta)/2} I_{\{a \leq \bar{W}^{(i)} \leq b\}} \sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|)\right\} \\
&\geq e^{-\delta/2} (H_{2,1} - H_{2,2}),
\end{aligned} \tag{6.7}$$

where

$$\begin{aligned}
H_{2,1} &= \mathbb{E}\left\{e^{\bar{W}^{(i)}/2} I_{\{a \leq \bar{W}^{(i)} \leq b\}} \sum_{j \neq i} \mathbb{E}\{|\xi_j| \min(\delta, |\bar{\xi}_j|)\}\right\}, \\
H_{2,2} &= \mathbb{E}\left\{e^{\bar{W}^{(i)}/2} \left| \sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|) - \mathbb{E}|\xi_j| \min(\delta, |\bar{\xi}_j|) \right|\right\}.
\end{aligned}$$

Noting that  $\delta \leq .085$  and that  $\gamma \leq .17$ , and following the proof of (5.9), we have

$$\begin{aligned}
\sum_{j \neq i} \mathbb{E}\{|\xi_j| \min(\delta, |\bar{\xi}_j|)\} &= \sum_{j \neq i} \mathbb{E}(|\xi_j| [\min(\delta, |\xi_j|) - \delta I_{\{\xi_j > 1\}}]) \\
&\geq -\delta \mathbb{E}|\xi_i| + \sum_{j=1}^n \mathbb{E}\{|\xi_j| \min(\delta, |\xi_j|)\} - \delta \gamma \\
&\geq -\delta \gamma^{1/3} + 0.5 - \delta \gamma \\
&\geq -0.085(0.17)^{1/3} + 0.5 - 0.085(0.17) \geq 0.43.
\end{aligned} \tag{6.8}$$

Hence

$$H_{2,1} \geq 0.43 e^{a/2} \mathbb{P}(a \leq \bar{W}^{(i)} \leq b). \tag{6.9}$$

By the Bennett inequality (6.2) again, we have

$$\mathbb{E}e^{\bar{W}^{(i)}} \leq \exp(e - 2)$$

and hence, as for (5.11),

$$\begin{aligned}
H_{2,2} &\leq \left(\mathbb{E}e^{\bar{W}^{(i)}}\right)^{1/2} \left(\text{Var}\left(\sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|)\right)\right)^{1/2} \\
&\leq \exp\left(\frac{e}{2} - 1\right) \delta \leq 1.44 \delta.
\end{aligned} \tag{6.10}$$

As to the left hand side of (6.7), we have

$$\begin{aligned} \mathbb{E}\{W^{(i)}f(\overline{W}^{(i)})\} &\leq (b-a+2\delta)\mathbb{E}\left\{|W^{(i)}|e^{\overline{W}^{(i)}/2}\right\} \\ &\leq (b-a+2\delta)(\mathbb{E}|W^{(i)}|^2)^{1/2}\left(\mathbb{E}e^{\overline{W}^{(i)}}\right)^{1/2} \\ &\leq (b-a+2\delta)\exp(e-2) \leq 2.06(b-a+2\delta). \end{aligned}$$

Combining the above inequalities yields

$$\begin{aligned} \mathbb{P}(a \leq \overline{W}^{(i)} \leq b) &\leq \frac{e^{-a/2}}{0.43} \left( e^{\delta/2} 2.06(b-a+2\delta) + 1.44\delta \right) \\ &\leq \frac{e^{-a/2}}{0.43} \left( e^{0.425} 2.06(b-a+2\delta) + 1.44\delta \right) \\ &\leq e^{-a/2} (5(b-a) + 13.4\delta) \\ &\leq e^{-a/2} (5(b-a) + 7\gamma). \end{aligned}$$

This proves (6.1). ■

Before proving the main theorem, we need the following moment inequality.

**Lemma 6.3:** *Let  $2 < p \leq 3$ , and let  $\{\eta_i, 1 \leq i \leq n\}$  be independent random variables with  $\mathbb{E}\eta_i = 0$  and  $\mathbb{E}|\eta_i|^p < \infty$ . Put  $S_n = \sum_{i=1}^n \eta_i$  and  $B_n^2 = \sum_{i=1}^n \mathbb{E}\eta_i^2$ . Then*

$$\mathbb{E}|S_n|^p \leq (p-1)B_n^p + \sum_{i=1}^n \mathbb{E}|\eta_i|^p \quad (6.11)$$

**Remark:** Moment inequalities of this kind were first proved by Rosenthal (1970), in a more general martingale setting.

**Proof:** Let  $S_n^{(i)} = S_n - \eta_i$ . Then

$$\begin{aligned} \mathbb{E}|S_n|^p &= \sum_{i=1}^n \mathbb{E}\eta_i S_n |S_n|^{p-2} \\ &= \sum_{i=1}^n \mathbb{E}\eta_i (S_n |S_n|^{p-2} - S_n^{(i)} |S_n|^{p-2}) \\ &\quad + \sum_{i=1}^n \mathbb{E}\eta_i (S_n^{(i)} |S_n|^{p-2} - S_n^{(i)} |S_n^{(i)}|^{p-2}), \end{aligned}$$

once again because  $\eta_i$  and  $S_n^{(i)}$  are independent, and  $\mathbb{E}\eta_i = 0$ . Thus we have

$$\begin{aligned} \mathbb{E}|S_n|^p &\leq \sum_{i=1}^n \mathbb{E}\eta_i^2 |S_n|^{p-2} + \sum_{i=1}^n \mathbb{E}|\eta_i| |S_n^{(i)}| \{(|S_n^{(i)}| + |\eta_i|)^{p-2} - |S_n^{(i)}|^{p-2}\} \\ &\leq \sum_{i=1}^n \mathbb{E}\eta_i^2 (|\eta_i|^{p-2} + |S_n^{(i)}|^{p-2}) \\ &\quad + \sum_{i=1}^n \mathbb{E}|\eta_i| |S_n^{(i)}|^{p-1} \{(1 + |\eta_i|/|S_n^{(i)}|)^{p-2} - 1\}. \end{aligned}$$

Since  $(1+x)^{p-2} - 1 \leq (p-2)x$  in  $x \geq 0$ , we thus have

$$\begin{aligned} \mathbb{E}|S_n|^p &\leq \sum_{i=1}^n \mathbb{E}|\eta_i|^p + \sum_{i=1}^n \mathbb{E}\eta_i^2 \mathbb{E}|S_n^{(i)}|^{p-2} \\ &\quad + \sum_{i=1}^n \mathbb{E}|\eta_i| |S_n^{(i)}|^{p-1} (p-2) |\eta_i|/|S_n^{(i)}| \\ &= \sum_{i=1}^n \mathbb{E}|\eta_i|^p + (p-1) \sum_{i=1}^n \mathbb{E}\eta_i^2 \mathbb{E}|S_n^{(i)}|^{p-2}, \end{aligned}$$

and Hölder's inequality now gives

$$\begin{aligned} \mathbb{E}|S_n|^p &\leq \sum_{i=1}^n \mathbb{E}|\eta_i|^p + (p-1) \sum_{i=1}^n \mathbb{E}\eta_i^2 (\mathbb{E}|S_n^{(i)}|^2)^{(p-2)/2} \\ &\leq \sum_{i=1}^n \mathbb{E}|\eta_i|^p + (p-1) B_n^p, \end{aligned}$$

as desired. ■

## 6.2. The final result

We are now ready to prove the non-uniform Berry–Esseen inequality.

**Theorem 6.4:** *There exists an absolute constant  $C$  such that for every real number  $z$ ,*

$$|\mathbb{P}(W \leq z) - \Phi(z)| \leq \frac{C\gamma}{1 + |z|^3}. \quad (6.12)$$

**Proof:** Without loss of generality, assume  $z \geq 0$ . By (6.11),

$$\mathbb{P}(W \geq z) \leq \frac{1 + \mathbb{E}|W|^3}{1 + z^3} \leq \frac{1 + 2 + \gamma}{1 + z^3}.$$

Thus (6.12) holds if  $\gamma \geq 1$ , and we can now assume  $\gamma < 1$ . Let

$$\bar{\xi}_i = \xi_i I_{\{\xi_i \leq 1\}}, \quad \bar{W} = \sum_{i=1}^n \bar{\xi}_i, \quad \bar{W}^{(i)} = \bar{W} - \bar{\xi}_i.$$

Observing that

$$\begin{aligned} \{W \geq z\} &= \{W \geq z, \max_{1 \leq i \leq n} \xi_i > 1\} \cup \{W \geq z, \max_{1 \leq i \leq n} \xi_i \leq 1\} \\ &\subset \{W \geq z, \max_{1 \leq i \leq n} \xi_i > 1\} \cup \{\bar{W} \geq z\}, \end{aligned}$$

we have

$$\mathbb{P}(W > z) \leq \mathbb{P}(\bar{W} > z) + \mathbb{P}(W > z, \max_{1 \leq i \leq n} \xi_i > 1), \quad (6.13)$$

and, since clearly  $W \geq \bar{W}$ ,

$$\mathbb{P}(\bar{W} > z) \leq \mathbb{P}(W > z). \quad (6.14)$$

Note that

$$\begin{aligned} \mathbb{P}(W > z, \max_{1 \leq i \leq n} \xi_i > 1) &\leq \sum_{i=1}^n \mathbb{P}(W > z, \xi_i > 1) \\ &\leq \sum_{i=1}^n \mathbb{P}(\xi_i > \max(1, z/2)) + \sum_{i=1}^n \mathbb{P}(W^{(i)} > z/2, \xi_i > 1) \\ &= \sum_{i=1}^n \mathbb{P}(\xi_i > \max(1, z/2)) + \sum_{i=1}^n \mathbb{P}(W^{(i)} > z/2) \mathbb{P}(\xi_i > 1) \\ &\leq \frac{\gamma}{\max(1, z/2)^3} + \sum_{i=1}^n \frac{(1 + \mathbb{E}|W^{(i)}|^3)}{1 + (z/2)^3} \mathbb{E}|\xi_i|^3 \leq \frac{C\gamma}{1 + z^3}; \end{aligned}$$

here, and in what follows,  $C$  denotes a generic absolute constant, whose value may be different at each appearance. Thus, to prove (6.12), it suffices to show that

$$|\mathbb{P}(\bar{W} \leq z) - \Phi(z)| \leq Ce^{-z/2}\gamma. \quad (6.15)$$

Let  $f_z$  be the solution to the Stein equation (2.2), and define

$$\bar{K}_i(t) = \mathbb{E}\{\bar{\xi}_i(I_{\{0 \leq t \leq \bar{\xi}_i\}} - I_{\{\bar{\xi}_i \leq t < 0\}})\}.$$

We follow the proof of (2.17), noting that  $\bar{\xi}_i \leq 1$ , and that  $\mathbb{E}\bar{\xi}_i$  need no longer in general be equal to zero, but may be negative. This gives

$$\mathbb{E}\{\bar{W}f_z(\bar{W})\} = \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}f'_z(\bar{W}^{(i)} + t)\bar{K}_i(t) dt + \sum_{i=1}^n \mathbb{E}\bar{\xi}_i \mathbb{E}f_z(\bar{W}^{(i)}).$$

From

$$\sum_{i=1}^n \int_{-\infty}^1 \bar{K}_i(t) dt = \sum_{i=1}^n \mathbb{E} \bar{\xi}_i^2 = 1 - \sum_{i=1}^n \mathbb{E} \xi_i^2 I_{\{\xi_i > 1\}},$$

we obtain that

$$\begin{aligned} \mathbb{P}(\bar{W} \leq z) - \Phi(z) &= \mathbb{E} f'_z(\bar{W}) - \mathbb{E}\{\bar{W} f_z(\bar{W})\} \\ &= \sum_{i=1}^n \mathbb{E}\{\xi_i^2 I_{\{\xi_i > 1\}}\} \mathbb{E} f'_z(\bar{W}) \\ &\quad + \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}[f'_z(\bar{W}^{(i)} + \bar{\xi}_i) - f'_z(\bar{W}^{(i)} + t)] \bar{K}_i(t) dt \\ &\quad + \sum_{i=1}^n \mathbb{E}\{\xi_i I_{\{\xi_i > 1\}}\} \mathbb{E} f_z(\bar{W}^{(i)}) \\ &:= R_1 + R_2 + R_3. \end{aligned} \tag{6.16}$$

By (8.3), (2.8) and (6.2),

$$\begin{aligned} \mathbb{E}|f'_z(\bar{W})| &= \mathbb{E}\{|f'_z(\bar{W})| I_{\{\bar{W} \leq z/2\}}\} + \mathbb{E}\{|f'_z(\bar{W})| I_{\{\bar{W} > z/2\}}\} \\ &\leq (1 + \sqrt{2\pi}(z/2)e^{z^2/8})(1 - \Phi(z)) + \mathbb{P}(\bar{W} > z/2) \\ &\leq (1 + \sqrt{2\pi}(z/2)e^{z^2/8})(1 - \Phi(z)) + e^{-z^2/2} \mathbb{E} e^{\bar{W}} \\ &\leq C e^{-z^2/2}, \end{aligned}$$

and hence

$$|R_1| \leq C \gamma e^{-z^2/2}. \tag{6.17}$$

Similarly, we have  $\mathbb{E} f_z(\bar{W}^{(i)}) \leq C e^{-z^2/2}$  and

$$|R_3| \leq C \gamma e^{-z^2/2}. \tag{6.18}$$

To estimate  $R_2$ , write

$$R_2 = R_{2,1} + R_{2,2},$$

where

$$\begin{aligned} R_{2,1} &= \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}[I_{\{\bar{W}^{(i)} + \bar{\xi}_i \leq z\}} - I_{\{\bar{W}^{(i)} + t \leq z\}}] \bar{K}_i(t) dt, \\ R_{2,2} &= \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}[(\bar{W}^{(i)} + \bar{\xi}_i) f_z(\bar{W}^{(i)} + \bar{\xi}_i) \\ &\quad - (\bar{W}^{(i)} + t) f_z(\bar{W}^{(i)} + t)] \bar{K}_i(t) dt. \end{aligned}$$

By Proposition 6.1,

$$\begin{aligned}
R_{2,1} &\leq \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}\{I_{\{\bar{\xi}_i \leq t\}} \mathbb{P}(z-t < \bar{W}^{(i)} \leq z - \bar{\xi}_i \mid \bar{\xi}_i)\} \bar{K}_i(t) dt \\
&\leq C \sum_{i=1}^n \int_{-\infty}^1 e^{-(z-t)/2} \mathbb{E}(|\bar{\xi}_i| + |t| + \gamma) \bar{K}_i(t) dt \\
&\leq C e^{-z/2} \gamma.
\end{aligned} \tag{6.19}$$

From Lemma 6.5, proved below, it follows that

$$\begin{aligned}
R_{2,2} &\leq \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}\{I_{\{t \leq \bar{\xi}_i\}} [\mathbb{E}(\{\bar{W}^{(i)} + \bar{\xi}_i\} f_z(\bar{W}^{(i)} + \bar{\xi}_i) \mid \bar{\xi}_i) \\
&\quad - \mathbb{E}(\bar{W}^{(i)} + t) f_z(\bar{W}^{(i)} + t)]\} \bar{K}_i(t) dt \\
&\leq C e^{-z/2} \sum_{i=1}^n \int_{-\infty}^1 \mathbb{E}(|\bar{\xi}_i| + |t|) \bar{K}_i(t) dt \\
&\leq C e^{-z/2} \gamma.
\end{aligned} \tag{6.20}$$

Therefore

$$R_2 \leq C e^{-z/2} \gamma. \tag{6.21}$$

Similarly, we have

$$R_2 \geq -C e^{-z/2} \gamma. \tag{6.22}$$

This proves the theorem.  $\blacksquare$

It remains to prove the following lemma.

**Lemma 6.5:** For  $s < t \leq 1$ , we have

$$\begin{aligned}
&\mathbb{E}\{(\bar{W}^{(i)} + t) f_z(\bar{W}^{(i)} + t)\} - \mathbb{E}\{(\bar{W}^{(i)} + s) f_z(\bar{W}^{(i)} + s)\} \\
&\leq C e^{-z/2} (|s| + |t|).
\end{aligned} \tag{6.23}$$

**Proof:** Let  $g(w) = (w f_z(w))'$ . Then

$$\mathbb{E}\{(\bar{W}^{(i)} + t) f_z(\bar{W}^{(i)} + t)\} - \mathbb{E}\{(\bar{W}^{(i)} + s) f_z(\bar{W}^{(i)} + s)\} = \int_s^t \mathbb{E} g(\bar{W}^{(i)} + u) du.$$

From the definition of  $g$  and  $f_z$ , we get

$$g(w) = \begin{cases} \left( \sqrt{2\pi}(1+w^2)e^{w^2/2}(1-\Phi(w)) - w \right) \Phi(z), & w \geq z \\ \left( \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w) + w \right) (1-\Phi(z)), & w < z. \end{cases}$$

By (2.6),  $g(w) \geq 0$  for all real  $w$ . A direct calculation shows that

$$\sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w) + w \leq 2 \quad \text{for } w \leq 0. \quad (6.24)$$

Thus, we have

$$g(w) \leq \begin{cases} 4(1+z^2)e^{z^2/8}(1-\Phi(z)) & \text{if } w \leq z/2 \\ 4(1+z^2)e^{z^2/2}(1-\Phi(z)) & \text{if } z/2 < w \leq z, \end{cases}$$

and this latter bound holds also for  $w > z$ , as can be seen by applying (6.24) with  $-w$  for  $w$  to the formula for  $g$  in  $w \geq z$ . Hence, for any  $u \in [s, t]$ , we have

$$\begin{aligned} & \mathbb{E}g(\overline{W}^{(i)} + u) \\ &= \mathbb{E}\left\{g(\overline{W}^{(i)} + u)I_{\{\overline{W}^{(i)} + u \leq z/2\}}\right\} + \mathbb{E}\left\{g(\overline{W}^{(i)} + u)I_{\{\overline{W}^{(i)} + u > z/2\}}\right\} \\ &\leq 4(1+z^2)e^{z^2/8}(1-\Phi(z)) + 4(1+z^2)e^{z^2/2}(1-\Phi(z))\mathbb{P}(\overline{W}^{(i)} + u > z/2) \\ &\leq Ce^{-z/2} + C(1+z)e^{-z+2u}\mathbb{E}e^{2\overline{W}^{(i)}}. \end{aligned}$$

But  $u \leq t \leq 1$ , and so

$$\mathbb{E}g(\overline{W}^{(i)} + u) \leq Ce^{-z/2} + C(1+z)e^{-z}\mathbb{E}e^{2\overline{W}^{(i)}} \leq Ce^{-z/2},$$

by (6.2). This gives

$$\mathbb{E}\{(\overline{W}^{(i)} + t)f_z(\overline{W}^{(i)} + t)\} - \mathbb{E}\{(\overline{W}^{(i)} + s)f_z(\overline{W}^{(i)} + s)\} \leq Ce^{-z/2}(|s| + |t|),$$

proving (6.23). ■

## 7. Uniform and non-uniform bounds under local dependence

In this section, we extend the discussion of normal approximation under local dependence using Stein's method, which was begun in Section 3.2. Our aim is to establish optimal uniform and non-uniform Berry–Esseen bounds under local dependence.

Throughout this section, let  $\mathcal{J}$  be an index set of cardinality  $n$  and let  $\{\xi_i, i \in \mathcal{J}\}$  be a random field with zero means and finite variances. Define  $W = \sum_{i \in \mathcal{J}} \xi_i$  and assume that  $\text{Var}(W) = 1$ . For  $A \subset \mathcal{J}$ , let  $\xi_A$  denote  $\{\xi_i, i \in A\}$ ,  $A^c = \{j \in \mathcal{J} : j \notin A\}$  and  $|A|$  the cardinality of  $A$ . We introduce four dependence assumptions, the first two of which appeared in Section 3.2

(LD1) For each  $i \in \mathcal{J}$  there exists  $A_i \subset \mathcal{J}$  such that  $\xi_i$  and  $\xi_{A_i^c}$  are independent.



(LD2) For each  $i \in \mathcal{J}$  there exist  $A_i \subset B_i \subset \mathcal{J}$  such that  $\xi_i$  is independent of  $\xi_{A_i^c}$  and  $\xi_{A_i}$  is independent of  $\xi_{B_i^c}$ .

(LD3) For each  $i \in \mathcal{J}$  there exist  $A_i \subset B_i \subset C_i \subset \mathcal{J}$  such that  $\xi_i$  is independent of  $\xi_{A_i^c}$ ,  $\xi_{A_i}$  is independent of  $\xi_{B_i^c}$ , and  $\xi_{B_i}$  is independent of  $\xi_{C_i^c}$ .

(LD4\*) For each  $i \in \mathcal{J}$  there exist  $A_i \subset B_i \subset B_i^* \subset C_i^* \subset D_i^* \subset \mathcal{J}$  such that  $\xi_i$  is independent of  $\xi_{A_i^c}$ ,  $\xi_{A_i}$  is independent of  $\xi_{B_i^c}$ , and then  $\xi_{A_i}$  is independent of  $\{\xi_{A_j}, j \in B_i^{*c}\}$ ,  $\{\xi_{A_l}, l \in B_i^*\}$  is independent of  $\{\xi_{A_j}, j \in C_i^{*c}\}$ , and  $\{\xi_{A_l}, l \in C_i^*\}$  is independent of  $\{\xi_{A_j}, j \in D_i^{*c}\}$ .

It is clear that (LD4\*) implies (LD3), (LD3) yields (LD2) and (LD1) is the weakest assumption. Roughly speaking, (LD4\*) is a version of (LD3) for  $\{\xi_{A_i}, i \in \mathcal{J}\}$ . On the other hand, (LD1) in many cases actually implies (LD2), (LD3) and (LD4\*) and  $B_i, C_i, B_i^*, C_i^*$  and  $D_i^*$  could be chosen as:  $B_i = \cup_{j \in A_i} A_j$ ,  $C_i = \cup_{j \in B_i} A_j$ ,  $B_i^* = \cup_{j \in A_i} B_j$ ,  $C_i^* = \cup_{j \in B_i^*} B_j$  and  $D_i^* = \cup_{j \in C_i^*} B_j$ .

We first present a general uniform Berry–Esseen bound under assumption (LD2).

**Theorem 7.1:** Let  $N(B_i) = \{j \in \mathcal{J} : B_j \cap B_i \neq \emptyset\}$  and  $2 < p \leq 4$ . Assume that (LD2) is satisfied with  $|N(B_i)| \leq \kappa$ . Then

$$\begin{aligned} & \sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \\ & \leq (13 + 11\kappa) \sum_{i \in \mathcal{J}} (\mathbb{E}|\xi_i|^{3\wedge p} + \mathbb{E}|Y_i|^{3\wedge p}) + 2.5 \left( \kappa \sum_{i \in \mathcal{J}} (\mathbb{E}|\xi_i|^p + \mathbb{E}|Y_i|^p) \right)^{1/2}, \end{aligned} \quad (7.1)$$

where  $Y_i = \sum_{j \in A_i} \xi_j$ . In particular, if  $\mathbb{E}|\xi_i|^p + \mathbb{E}|Y_i|^p \leq \theta^p$  for some  $\theta > 0$  and for each  $i \in \mathcal{J}$ , then

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq (13 + 11\kappa) n \theta^{3\wedge p} + 2.5 \theta^{p/2} \sqrt{\kappa n}, \quad (7.2)$$

where  $n = |\mathcal{J}|$ .

Note that in many cases  $\kappa$  is bounded and  $\theta$  is of order of  $n^{-1/2}$ . In those cases,  $\kappa n \theta^{3\wedge p} + \theta^{p/2} \sqrt{\kappa n} = O(n^{-(p-2)/4})$ , which is of the best possible order of  $n^{-1/2}$  when  $p = 4$ . However, the cost is the existence of fourth moments. To reduce the assumption on moments, we need the stronger condition (LD3).

**Theorem 7.2:** Let  $2 < p \leq 3$ . Assume that (LD3) is satisfied with  $|N(C_i)| \leq \kappa$ , where  $N(C_i) = \{j \in \mathcal{J} : C_i B_j \neq \emptyset\}$ . Then

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 75\kappa^{p-1} \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i|^p. \quad (7.3)$$

We now present a general non-uniform bound for locally dependent random fields  $\{\xi_i, i \in \mathcal{J}\}$  under (LD4\*).

**Theorem 7.3:** Assume that  $\mathbb{E}|\xi_i|^p < \infty$  for  $2 < p \leq 3$  and that (LD4\*) is satisfied. Let  $\kappa = \max_{i \in \mathcal{J}} \max(|D_i^*|, |\{j : i \in D_j^*\}|)$ . Then

$$|\mathbb{P}(W \leq z) - \Phi(z)| \leq C\kappa^p(1 + |z|)^{-p} \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i|^p, \quad (7.4)$$

where  $C$  is an absolute constant.

The above results can immediately be applied to  $m$ -dependent random fields. Let  $d \geq 1$  and  $Z^d$  denote the  $d$ -dimensional space of positive integers. The distance between two points  $i = (i_1, \dots, i_d)$  and  $j = (j_1, \dots, j_d)$  in  $Z^d$  is defined by  $|i - j| = \max_{1 \leq l \leq d} |i_l - j_l|$  and the distance between two subsets  $A$  and  $B$  of  $Z^d$  is defined by  $\rho(A, B) = \inf\{|i - j| : i \in A, j \in B\}$ . For a given subset  $\mathcal{J}$  of  $Z^d$ , a set of random variables  $\{\xi_i, i \in \mathcal{J}\}$  is said to be an  $m$ -dependent random field if  $\{\xi_i, i \in A\}$  and  $\{\xi_j, j \in B\}$  are independent whenever  $\rho(A, B) > m$ , for any subsets  $A$  and  $B$  of  $\mathcal{J}$ . Thus, choosing

$$\begin{aligned} A_i &= \{j : |j - i| \leq m\} \cap \mathcal{J}, & B_i &= \{j : |j - i| \leq 2m\} \cap \mathcal{J}, \\ C_i &= \{j : |j - i| \leq 3m\} \cap \mathcal{J}, & B_i^* &= \{j : |j - i| \leq 3m\} \cap \mathcal{J}, \\ C_i^* &= \{j : |j - i| \leq 4m\} \cap \mathcal{J} & \text{and } D_i^* &= \{j : |j - i| \leq 5m\} \cap \mathcal{J} \end{aligned}$$

in Theorems 7.2 and 7.3 yields a uniform and a non-uniform bound.

**Theorem 7.4:** Let  $\{\xi_i, i \in \mathcal{J}\}$  be an  $m$ -dependent random fields with zero means and finite  $\mathbb{E}|\xi_i|^p < \infty$  for  $2 < p \leq 3$ . Then

$$\sup_z |\mathbb{P}(W \leq z) - \Phi(z)| \leq 75(10m + 1)^{(p-1)d} \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i|^p \quad (7.5)$$

and

$$|\mathbb{P}(W \leq z) - \Phi(z)| \leq C(1 + |z|)^{-p} 11^{pd} (m + 1)^{(p-1)d} \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i|^p, \quad (7.6)$$

where  $C$  is an absolute constant.

The main idea of the proof is similar to that in Sections 3 and 4, first deriving a Stein identity and then uniform and non-uniform concentration inequalities. We outline some main steps in the proof and refer to Chen & Shao (2004a) for details.

Define

$$\begin{aligned}\widehat{K}_i(t) &= \xi_i \{I(-Y_i \leq t < 0) - I(0 \leq t \leq -Y_i)\}, \quad K_i(t) = \mathbb{E}\widehat{K}_i(t), \\ \widehat{K}(t) &= \sum_{i \in \mathcal{J}} \widehat{K}_i(t), \quad K(t) = \mathbb{E}\widehat{K}(t) = \sum_{i \in \mathcal{J}} K_i(t).\end{aligned}\quad (7.7)$$

We first derive a Stein identity for  $W$ . Let  $f$  be a bounded absolutely continuous function. Then

$$\begin{aligned}\mathbb{E}\{Wf(W)\} &= \sum_{i \in \mathcal{J}} \mathbb{E}\{\xi_i(f(W) - f(W - Y_i))\} = \sum_{i \in \mathcal{J}} \mathbb{E}\left\{\xi_i \int_{-Y_i}^0 f'(W+t) dt\right\} \\ &= \sum_{i \in \mathcal{J}} \mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W+t) \widehat{K}_i(t) dt\right\} = \mathbb{E}\left\{\int_{-\infty}^{\infty} f'(W+t) \widehat{K}(t) dt\right\},\end{aligned}\quad (7.8)$$

and hence, by the fact that  $\int_{-\infty}^{\infty} K(t) dt = \mathbb{E}W^2 = 1$ ,

$$\begin{aligned}\mathbb{E}\{f'(W) - Wf(W)\} &= \mathbb{E} \int_{-\infty}^{\infty} f'(W)K(t) dt - \mathbb{E} \int_{-\infty}^{\infty} f'(W+t)\widehat{K}(t) dt \\ &= \mathbb{E} \int_{-\infty}^{\infty} (f'(W) - f'(W+t))K(t) dt \\ &\quad + \mathbb{E}f'(W) \int_{-\infty}^{\infty} (K(t) - \widehat{K}(t)) dt \\ &\quad + \mathbb{E} \int_{-\infty}^{\infty} (f'(W+t) - f'(W))(\widehat{K}(t) - K(t)) dt \\ &:= R_1 + R_2 + R_3.\end{aligned}$$

Now let  $f = f_z$  be the Stein solution (2.3). Then

$$\begin{aligned}|R_1| &\leq \mathbb{E} \int_{-\infty}^{\infty} (|W| + 1)|tK(t)| dt + \left| \mathbb{E} \int_{-\infty}^{\infty} (I_{\{W \leq z\}} - I_{\{W+t \leq z\}})K(t) dt \right| \\ &\leq \frac{1}{2} \sum_{i=1}^n \mathbb{E}(|W| + 1)|\xi_i|Y_i^2 \\ &\quad + \int_{-\infty}^{\infty} \mathbb{P}(z - \max(t, 0) \leq W \leq z - \min(t, 0))K(t) dt \\ &:= R_{1,1} + R_{1,2}.\end{aligned}$$

Estimating  $R_{1,1}$  is not so difficult, while  $R_{1,2}$  can be estimated by using a concentration inequality given below.

Observe that

$$R_2 = \mathbb{E} \left\{ f'(W) \sum_{i=1}^n (\xi_i Y_i - \mathbb{E}(\xi_i Y_i)) \right\},$$

which can also be estimated easily. The main difficulty arises from estimating  $R_3$ . The reader may refer to Chen & Shao (2004a) for details.

To conclude this section, we give the simplest concentration inequality in the paper of Chen & Shao (2004a), and provide a detailed proof to illustrate the difficulty for dependent variables.

**Proposition 7.5:** *Assume (LD1). Then for any real numbers  $a < b$ ,*

$$\mathbb{P}(a \leq W \leq b) \leq 0.625(b - a) + 4r_1 + 4r_2, \quad (7.9)$$

where  $r_1 = \sum_{i \in \mathcal{J}} \mathbb{E}|\xi_i Y_i|^2$  and  $r_2 = \int_{-\infty}^{\infty} \text{Var}(\widehat{K}(t)) dt$ .

**Proof:** Let  $\alpha = r_1$  and define

$$f(w) = \begin{cases} -(b - a + \alpha)/2 & \text{for } w \leq a - \alpha \\ \frac{1}{2\alpha}(w - a + \alpha)^2 - (b - a + \alpha)/2 & \text{for } a - \alpha < w \leq a \\ w - (a + b)/2 & \text{for } a < w \leq b \\ -\frac{1}{2\alpha}(w - b - \alpha)^2 + (b - a + \alpha)/2 & \text{for } b < w \leq b + \alpha \\ (b - a + \alpha)/2 & \text{for } w > b + \alpha. \end{cases} \quad (7.10)$$

Then  $f'$  is a continuous function given by

$$f'(w) = \begin{cases} 1, & \text{for } a \leq w \leq b \\ 0, & \text{for } w \leq a - \alpha \text{ or } w \geq b + \alpha, \\ \text{linear,} & \text{for } a - \alpha \leq w \leq a \text{ or } b \leq w \leq b + \alpha. \end{cases}$$

Clearly  $|f(w)| \leq (b - a + \alpha)/2$ . With this  $f$ , and with  $\widehat{K}(t)$  and  $K(t)$  as defined in (7.7), we have, by (7.8),

$$\begin{aligned} (b - a + \alpha)/2 &\geq \mathbb{E}Wf(W) = \mathbb{E} \int_{-\infty}^{\infty} f'(W + t)\widehat{K}(t) dt \\ &:= \mathbb{E}f'(W) \int_{-\infty}^{\infty} K(t) dt + \mathbb{E} \int_{-\infty}^{\infty} (f'(W + t) - f'(W))K(t) dt \\ &\quad + \mathbb{E} \int_{-\infty}^{\infty} f'(W + t)(\widehat{K}(t) - K(t)) dt \\ &:= H_1 + H_2 + H_3. \end{aligned} \quad (7.11)$$

Clearly,

$$H_1 = \mathbb{E}f'(W) \geq \mathbb{P}(a \leq W \leq b). \quad (7.12)$$

By the Cauchy inequality,

$$\begin{aligned} |H_3| &\leq (1/8)\mathbb{E} \int_{-\infty}^{\infty} [f'(W+t)]^2 dt + 2\mathbb{E} \int_{-\infty}^{\infty} (\widehat{K}(t) - K(t))^2 dt \\ &\leq (b-a+2\alpha)/8 + 2r_2. \end{aligned} \quad (7.13)$$

To bound  $H_2$ , let

$$L(\alpha) = \sup_{x \in \mathbb{R}} \mathbb{P}(x \leq W \leq x + \alpha).$$

Then, by writing

$$\begin{aligned} H_2 &= \mathbb{E} \int_0^{\infty} \int_0^t f''(W+s) ds K(t) dt - \mathbb{E} \int_{-\infty}^0 \int_t^0 f''(W+s) ds K(t) dt \\ &= \alpha^{-1} \int_0^{\infty} \int_0^t \{\mathbb{P}(a-\alpha \leq W+s \leq a) - \mathbb{P}(b \leq W+s \leq b+\alpha)\} ds K(t) dt \\ &\quad - \alpha^{-1} \int_{-\infty}^0 \int_t^0 \{\mathbb{P}(a-\alpha \leq W+s \leq a) - \mathbb{P}(b \leq W+s \leq b+\alpha)\} ds K(t) dt, \end{aligned}$$

we have

$$\begin{aligned} |H_2| &\leq \alpha^{-1} \int_0^{\infty} \int_0^t L(\alpha) ds |K(t)| dt + \alpha^{-1} \int_{-\infty}^0 \int_t^0 L(\alpha) ds |K(t)| dt \\ &= \alpha^{-1} L(\alpha) \int_{-\infty}^{\infty} |tK(t)| dt \leq \frac{1}{2} \alpha^{-1} r_1 L(\alpha) = \frac{1}{2} L(\alpha). \end{aligned} \quad (7.14)$$

It follows from (7.11) – (7.14) that

$$\mathbb{P}(a \leq W \leq b) \leq 0.625(b-a) + 0.75\alpha + 2r_2 + 0.5L(\alpha). \quad (7.15)$$

Substituting  $a = x$  and  $b = x + \alpha$  in (7.15), we obtain

$$L(\alpha) \leq 1.375\alpha + 2r_2 + 0.5L(\alpha)$$

and hence

$$L(\alpha) \leq 2.75\alpha + 4r_2. \quad (7.16)$$

Finally combining (7.15) and (7.16), we obtain (7.9).  $\blacksquare$

## 8. Appendix

Here we give detailed proofs of the basic properties of the solutions to the Stein equations (2.2) and (2.4), given in Lemmas 2.2 and 2.3. The proofs of Lemma 2.2 and part of Lemma 2.3 follow Stein (1986), and part of the proof of Lemma 2.3 is due to Stroock (1993).

**Proof of Lemma 2.2:** Since  $f_z(w) = f_{-z}(-w)$ , we need only consider the case  $z \geq 0$ . Note that for  $w > 0$

$$\int_w^\infty e^{-x^2/2} dx \leq \int_w^\infty \frac{x}{w} e^{-x^2/2} dx = \frac{e^{-w^2/2}}{w},$$

which also yields

$$(1 + w^2) \int_w^\infty e^{-x^2/2} dx \geq w e^{-w^2/2},$$

by comparing the derivatives of the two functions. Thus

$$\frac{w e^{-w^2/2}}{(1 + w^2)\sqrt{2\pi}} \leq 1 - \Phi(w) \leq \frac{e^{-w^2/2}}{w\sqrt{2\pi}}. \quad (8.1)$$

It follows from (2.3) that

$$(w f_z(w))' = \begin{cases} \sqrt{2\pi}[1 - \Phi(z)] \left( (1 + w^2)e^{w^2/2}\Phi(w) + \frac{w}{\sqrt{2\pi}} \right) & \text{if } w < z; \\ \sqrt{2\pi}\Phi(z) \left( (1 + w^2)e^{w^2/2}(1 - \Phi(w)) - \frac{w}{\sqrt{2\pi}} \right) & \text{if } w > z, \\ \geq 0, & \end{cases}$$

by (8.1). This proves (2.6).

In view of the fact that

$$\lim_{w \rightarrow -\infty} w f_z(w) = \Phi(z) - 1 \text{ and } \lim_{w \rightarrow \infty} w f_z(w) = \Phi(z), \quad (8.2)$$

(2.7) follows by (2.6).

By (2.2), we have

$$\begin{aligned} f'_z(w) &= w f_z(w) + I_{\{w \leq z\}} - \Phi(z) \\ &= \begin{cases} w f_z(w) + 1 - \Phi(z) & \text{for } w < z, \\ w f_z(w) - \Phi(z) & \text{for } w > z; \end{cases} \\ &= \begin{cases} (\sqrt{2\pi} w e^{w^2/2} \Phi(w) + 1)(1 - \Phi(z)) & \text{for } w < z, \\ (\sqrt{2\pi} w e^{w^2/2} (1 - \Phi(w)) - 1)\Phi(z) & \text{for } w > z. \end{cases} \end{aligned} \quad (8.3)$$

Since  $w f_z(w)$  is an increasing function of  $w$ , by (8.1) and (8.2),

$$0 < f'_z(w) \leq z f_z(z) + 1 - \Phi(z) < 1 \text{ for } w < z \quad (8.4)$$

and

$$-1 < z f_z(z) - \Phi(z) \leq f'_z(w) < 0 \text{ for } w > z. \quad (8.5)$$

Hence, for any  $w$  and  $v$ ,

$$|f'_z(w) - f'_z(v)| \leq \max(1, z f_z(z) + 1 - \Phi(z) - (z f_z(z) - \Phi(z))) = 1.$$

This proves (2.8).

Observe that, by (8.4) and (8.5),  $f_z$  attains its maximum at  $z$ . Thus

$$0 < f_z(w) \leq f_z(z) = \sqrt{2\pi}e^{z^2/2}\Phi(z)(1 - \Phi(z)). \quad (8.6)$$

By (8.1),  $f_z(z) \leq 1/z$ . To finish the proof of (2.9), let

$$g(z) = \Phi(z)(1 - \Phi(z)) - e^{-z^2/2}/4 \quad \text{and} \quad g_1(z) = \frac{1}{\sqrt{2\pi}} + \frac{z}{4} - \frac{2\Phi(z)}{\sqrt{2\pi}}.$$

Observe that  $g'(z) = e^{-z^2/2}g_1(z)$  and that

$$g_1'(z) = \frac{1}{4} - \frac{1}{\pi}e^{-z^2} \begin{cases} < 0 & \text{if } 0 \leq z < z_0, \\ = 0 & \text{if } z = z_0, \\ > 0 & \text{if } z > z_0, \end{cases}$$

where  $z_0 = (2 \ln(4/\pi))^{1/2}$ . Thus,  $g_1(z)$  is decreasing on  $[0, z_0]$  and increasing on  $(z_0, \infty)$ . Since  $g_1(0) = 0$  and  $g_1(\infty) = \infty$ , there exists  $z_1 > 0$  such that  $g_1(z) < 0$  for  $0 < z < z_1$  and  $g_1(z) > 0$  for  $z > z_1$ . Therefore,  $g(z)$  attains its maximum at either  $z = 0$  or  $z = \infty$ , that is

$$g(z) \leq \max(g(0), g(\infty)) = 0,$$

which is equivalent to  $f_z(z) \leq \sqrt{2\pi}/4$ . This completes the proof of (2.9).

The last inequality (2.10) is a consequence of (2.8) and (2.9) by rewriting

$$\begin{aligned} (w+u)f_z(w+u) - (w+v)f_z(w+v) \\ = w(f_z(w+u) - f_z(w+v)) + uf_z(w+u) - vf_z(w+v) \end{aligned}$$

and using the Taylor expansion. ■

**Proof of Lemma 2.3:** We define  $\tilde{h}(w) = h(w) - \mathbf{E}h(Z)$ , and then put  $c_0 = \sup_w |\tilde{h}(w)|$ ,  $c_1 = \sup_w |h'(w)|$ . Since  $\tilde{h}$  and  $f_h$  are unchanged when  $h$  is replaced by  $h - h(0)$ , we may assume that  $h(0) = 0$ . Therefore  $|h(t)| \leq c_1|t|$  and  $|\mathbf{E}h(Z)| \leq c_1\mathbf{E}|Z| = c_1\sqrt{2/\pi}$ .

First we verify (2.11). From the definition (2.5) of  $f_h$ , it follows that

$$\begin{aligned} |f_h(w)| &\leq \begin{cases} e^{w^2/2} \int_{-\infty}^w |\tilde{h}(x)|e^{-x^2/2} dx & \text{if } w \leq 0, \\ e^{w^2/2} \int_w^{\infty} |\tilde{h}(x)|e^{-x^2/2} dx & \text{if } w \geq 0 \end{cases} \\ &\leq e^{w^2/2} \min\left(c_0 \int_{|w|}^{\infty} e^{-x^2/2} dx, c_1 \int_{|w|}^{\infty} (|x| + \sqrt{2/\pi})e^{-x^2/2} dx\right) \\ &\leq \min(\sqrt{\pi/2}, 2c_1), \end{aligned}$$

where in the last inequality we used the fact that

$$e^{w^2/2} \int_{|w|}^{\infty} e^{-x^2/2} dx \leq \sqrt{\pi/2}.$$

Next we prove (2.12). By (2.4), for  $w \geq 0$ ,

$$\begin{aligned} |f'_h(w)| &\leq |h(w) - \mathbb{E}h(Z)| + we^{w^2/2} \int_w^{\infty} |h(x) - \mathbb{E}h(Z)| e^{-x^2/2} dx \\ &\leq |h(w) - \mathbb{E}h(Z)| + c_0 we^{w^2/2} \int_w^{\infty} e^{-x^2/2} dx \leq 2c_0, \end{aligned}$$

by (8.1). It follows from (2.5) again that

$$f''(w) - wf'(w) - f(w) = h'(w),$$

or equivalently that

$$(e^{-w^2/2} f'(w))' = e^{-w^2/2} (f(w) + h'(w)).$$

Therefore

$$f'(w) = -e^{w^2/2} \int_w^{\infty} (f(x) + h'(x)) e^{-x^2/2} dx$$

and, by (2.11),

$$|f'(w)| \leq 3c_1 e^{w^2/2} \int_w^{\infty} e^{-x^2/2} dx \leq 3c_1 \sqrt{\pi/2} \leq 4c_1.$$

Thus we have

$$\sup_{w \geq 0} |f'(w)| \leq \min(2c_0, 4c_1).$$

Similarly, the above bound holds for  $\sup_{w \leq 0} |f'(w)|$ . This proves (2.12).

Now we prove (2.13). Differentiating (2.4) gives

$$\begin{aligned} f''_h(w) &= wf'_h(w) + f_h(w) + h'(w) \\ &= (1 + w^2)f_h(w) + w(h(w) - \mathbb{E}h(Z)) + h'(w). \end{aligned} \quad (8.7)$$

From

$$\begin{aligned} h(x) - \mathbb{E}h(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [h(x) - h(s)] e^{-s^2/2} ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \int_s^x h'(x) dt e^{-s^2/2} ds - \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \int_x^s h'(t) dt e^{-s^2/2} ds \\ &= \int_{-\infty}^x h'(t) \Phi(t) dt - \int_x^{\infty} h'(t) (1 - \Phi(t)) dt, \end{aligned} \quad (8.8)$$



it follows that

$$\begin{aligned}
f_h(w) &= e^{w^2/2} \int_{-\infty}^w [h(x) - \mathbf{E}h(Z)] e^{-x^2/2} dx \\
&= e^{w^2/2} \int_{-\infty}^w \left( \int_{-\infty}^x h'(t) \Phi(t) dt - \int_x^{\infty} h'(t) (1 - \Phi(t)) dt \right) e^{-x^2/2} dx \\
&= -\sqrt{2\pi} e^{w^2/2} (1 - \Phi(w)) \int_{-\infty}^w h'(t) \Phi(t) dt \\
&\quad - \sqrt{2\pi} e^{w^2/2} \Phi(w) \int_w^{\infty} h'(t) [1 - \Phi(t)] dt. \tag{8.9}
\end{aligned}$$

From (8.7) – (8.9) and (8.1) we now obtain

$$\begin{aligned}
|f_h''(w)| &\leq |h'(w)| + |(1 + w^2)f_h(w) + w(h(w) - \mathbf{E}h(Z))| \\
&\leq |h'(w)| + \left| \left( w - \sqrt{2\pi}(1 + w^2)e^{w^2/2}(1 - \Phi(w)) \right) \int_{-\infty}^w h'(t) \Phi(t) dt \right| \\
&\quad + \left| \left( -w - \sqrt{2\pi}(1 + w^2)e^{w^2/2}\Phi(w) \right) \int_w^{\infty} h'(t) (1 - \Phi(t)) dt \right| \\
&\leq |h'(w)| + c_1 \left( -w + \sqrt{2\pi}(1 + w^2)e^{w^2/2}(1 - \Phi(w)) \right) \int_{-\infty}^w \Phi(t) dt \\
&\quad + c_1 \left( w + \sqrt{2\pi}(1 + w^2)e^{w^2/2}\Phi(w) \right) \int_w^{\infty} (1 - \Phi(t)) dt. \tag{8.10}
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
|f_h''(w)| &\leq |h'(w)| \\
&\quad + c_1 \left( -w + \sqrt{2\pi}(1 + w^2)e^{w^2/2}(1 - \Phi(w)) \right) \left( w\Phi(w) + \frac{e^{-w^2/2}}{\sqrt{2\pi}} \right) \\
&\quad + c_1 \left( w + \sqrt{2\pi}(1 + w^2)e^{w^2/2}\Phi(w) \right) \left( -w(1 - \Phi(w)) + \frac{e^{-w^2/2}}{\sqrt{2\pi}} \right) \\
&= |h'(w)| + c_1 \leq 2c_1, \tag{8.11}
\end{aligned}$$

as desired. ■

**Acknowledgement:** We are grateful to Andrew Barbour for his careful editing of the paper and, in particular, for substantially improving the exposition in the Introduction.

## References

1. P. BALDI & Y. RINOTT (1989) On normal approximation of distributions in terms of dependency graph. *Ann. Probab.* **17**, 1646–1650.
2. P. BALDI, Y. RINOTT & C. STEIN (1989) A normal approximation for the number of local maxima of a random function on a graph. In: *Probability, statistics, and mathematics. Papers in honor of Samuel Karlin*, Eds: T. W. Anderson, K. B. Athreya, D. L. Iglehart, pp. 59–81. Academic Press, Boston.
3. A. D. BARBOUR & L. H. Y. CHEN (1992) On the binary expansion of a random integer. *Statist. Probab. Lett.* **14**, 235–241.
4. A. D. BARBOUR & P. HALL (1984) Stein’s method and the Berry–Esseen theorem. *Austral. J. Statist.* **26**, 8–15.
5. P. VAN BEECK (1972) An application of Fourier methods to the problem of sharpening the Berry–Esseen inequality. *Z. Wahrsch. verw. Geb.* **23**, 187–196.
6. A. BIKELIS (1966) Estimates of the remainder in the central limit theorem. *Litovsk. Mat. Sb.* **6**, 323–346 (in Russian).
7. E. BOLTHAUSEN (1984) An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. verw. Geb.* **66**, 379–386.
8. E. BOLTHAUSEN & F. GÖTZE (1993) The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.
9. L. H. Y. CHEN (1986) The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. IMA Preprint Series #243, Univ. Minnesota.
10. L. H. Y. CHEN (1998) Stein’s method: some perspectives with applications. In: *Probability towards 2000*, Eds: L. Accardi & C. C. Heyde, Lecture Notes in Statistics **128**, pp. 97–122. Springer-Verlag, Berlin.
11. L. H. Y. CHEN (2000) Non-uniform bounds in probability approximations using Stein’s method. In: *Probability and statistical models with applications: a volume in honor of Theophilos Cacoullos*, Eds: N. Balakrishnan, Ch. A. Charalambides & M. V. Koutras, pp. 3–14. Chapman & Hall/CRC Press, Boca Raton, Florida.
12. L. H. Y. CHEN & Q. M. SHAO (2001) A non-uniform Berry–Esseen bound via Stein’s method. *Probab. Theory Rel. Fields* **120**, 236–254.
13. L. H. Y. CHEN & Q. M. SHAO (2004a) Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028.
14. L. H. Y. CHEN & Q. M. SHAO (2004b) Normal approximation for non-linear statistics using a concentration inequality approach. Preprint.
15. A. DEMBO & Y. RINOTT (1996) Some examples of normal approximations by Stein’s method. In: *Random Discrete Structures*, Eds: D. Aldous & R. Pemantle, pp. 25–44. Springer, New York.
16. P. DIACONIS (1977) The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5**, 72–81.
17. P. DIACONIS & S. HOLMES (2004) *Stein’s method: expository lectures and applications*. IMS Lecture Notes, Vol. **46**, Hayward, CA.

18. P. DIACONIS, S. HOLMES & G. REINERT (2004) Use of exchangeable pairs in the analysis of simulations. In: *Stein's method: expository lectures and applications*, Eds: P. Diaconis & S. Holmes, pp. 1–26. IMS Lecture Notes Vol. **46**, Hayward, CA.
19. R. V. ERICKSON (1974)  $L_1$  bounds for asymptotic normality of  $m$ -dependent sums using Stein's technique. *Ann. Probab.* **2**, 522–529.
20. C. G. ESSEEN (1945) Fourier analysis of distribution functions: a mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1–125.
21. W. FELLER (1968) On the Berry-Esseen theorem. *Z. Wahrsch. verw. Geb.* **10**, 261–268.
22. L. GOLDSTEIN (2005) Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** (to appear).
23. L. GOLDSTEIN & G. REINERT (1997) Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.
24. L. GOLDSTEIN & Y. RINOTT (1996) Multivariate normal approximations by Stein's method and size bias couplings. *Adv. Appl. Prob.* **33**, 1–17.
25. F. GÖTZE (1991) On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.
26. S. T. HO & L. H. Y. CHEN (1978) An  $L_p$  bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231–249.
27. S. V. NAGAEV (1965) Some limit theorems for large deviations. *Theory Probab. Appl.* **10**, 214–235.
28. L. PADITZ (1989) On the analytical structure of the constant in the nonuniform version of the Esseen inequality. *Statistics* **20**, 453–464.
29. V. V. PETROV (1995) *Limit theorems of probability theory: sequences of independent random variables*. Oxford Studies in Probability 4, Clarendon Press, Oxford.
30. Y. RINOTT (1994) On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.* **55**, 135–143.
31. Y. RINOTT & V. ROTAR (1996) A multivariate CLT for local dependence with  $n^{-1/2} \log n$  rate and applications to multivariate graph related statistics. *J. Multiv. Anal.* **56**, 333–350.
32. Y. RINOTT & V. ROTAR (2000) Normal approximations by Stein's method. *Decis. Econ. Finance* **23**, 15–29.
33. H. P. ROSENTHAL (1970) On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel J. Math.* **8**, 273–303.
34. C. STEIN (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2**, 583–602. Univ. California Press, Berkeley, CA.
35. C. STEIN (1986) *Approximation computation of expectations*. IMS Lecture Notes Vol. **7**, Hayward, CA.
36. D. W. STROOCK (1993) *Probability theory: an analytic view*. Cambridge Univ. Press, Cambridge, U.K.