Tutorial Notes for the Workshop on Stein's Method and Applications

Stein's Method and Normal Approximation

Louis H.Y. Chen

National University of Singapore

Qi-Man Shao

National University of Singapore and University of Oregon

1 Introduction

Let $X_1, X_2, \cdots, X_n$ be independent random variables with zero means and finite variances. Put

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

It is well-known that if the Lindeberg condition

$$\forall\, \varepsilon > 0,\ \frac{1}{B_n^2} \sum_{i=1}^{n} EX_i^2 I_{\{|X_i| > \varepsilon B_n\}} \to 0 \text{ as } n \to \infty \tag{1.1}$$

is satisfied, then

$$\frac{S_n}{B_n} \xrightarrow{d.} N(0,1).$$

Furthermore, if $E|X_i|^3 < \infty$, then we have the uniform Berry-Esseen inequality

$$\sup_z |P\Big(\frac{S_n}{B_n} \le z\Big) - \Phi(z)| \le C_0 B_n^{-3} \sum_{i=1}^{n} E|X_i|^3 \tag{1.2}$$

and the non-uniform Berry-Esseen inequality

$$\forall\, z \in R^1,\ |P\Big(\frac{S_n}{B_n} \le z\Big) - \Phi(z)| \le C_1 (1+|z|)^{-3} B_n^{-3} \sum_{i=1}^{n} E|X_i|^3, \tag{1.3}$$

where $\Phi(z)$ is the standard normal distribution function, and both C_0 and C_1 are absolute constants. One can take $C_0 = 0.7975$ [van Beeck (1972)] and $C_1 = 114.7$ for independent random variables [Paditz (1977)] and $C_1 = 30.54$ for i.i.d. random variables [Michel (1988)] . The standard proof of Berry-Esseen inequalities is based on the method of characteristic function or the Fourier transform, which works well for independent random variables although it is already very complicated. A totally new method of normal approximation was introduced by Stein in 1972. Stein's method is striking. It works well not only for independent random variables but also for dependent variables. Stein's ideas can be also applied to many other probability approximations, notably to Poisson, Poisson process, compound Poisson and binomial approximations.

In this tutorial, we shall give an overview of the use of the Stein method for normal approximation. We start with basic results on the Stein equations and their solutions and then prove several classical limit theorems to illustrate the beauty of the Stein method. The focus will be on the ideas behind different approaches such as the concentration inequality approach, induction approach and exchangeable pair approach. We shall present a totally self-contained proof for (1.2) and (1.3) via Stein's method.

2 Stein's method

2.1 The Stein equation

Let Z be a standard normally distributed random variable and let $\mathcal{C}_{bd}$ be the set of continuous and piecewise continuously differential functions $f: R \to R$ with $E|f'(Z)| < \infty$. Stein's method rests on the following observation.

Lemma 2.1 *Let W be a real valued random variable. Then W has a standard normal distribution it is necessary and sufficient that for all $f \in \mathcal{C}_{bd}$*

$$Ef'(W) = EWf(W). \tag{2.1}$$

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

$$\begin{aligned}
Ef'(W) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w) e^{-w^2/2} dw \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(w) \Big(\int_{-\infty}^{w} (-x) e^{-x^2/2} dx \Big) dw \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(w) \Big(\int_{w}^{\infty} x e^{-x^2/2} dx \Big) dw \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \Big(\int_{x}^{0} f'(w) dw \Big) (-x) e^{-x^2/2} dx \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \Big(\int_{0}^{x} f'(w) dw \Big) x e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [f(x) - f(0)] x e^{-x^2/2} dx \\
&= EWf(W).
\end{aligned}$$

Sufficiency. For fixed $z \in R^1$, let $f(w) := f_z(w)$ be the solution of the following equation

$$f'(w) - wf(w) = I_{\{w \le z\}} - \Phi(z). \tag{2.2}$$

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

$$\left(e^{-w^2/2} f(w) \right)' = -e^{-w^2/2} \left(I_{\{w \le z\}} - \Phi(z) \right)$$

Thus,

$$f_z(w) = e^{w^2/2} \int_{-\infty}^{w} [I_{\{x \le z\}} - \Phi(z)] e^{-x^2/2} dx$$

$$
\begin{aligned}
&= -e^{w^2/2}\int_w^\infty [I_{\{x\le z\}} - \Phi(z)]e^{-x^2/2}dx \\
&= \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)[1-\Phi(z)] & \text{if } w \le z, \\ \sqrt{2\pi}e^{w^2/2}\Phi(z)[1-\Phi(w)] & \text{if } w \ge z. \end{cases} \qquad (2.3)
\end{aligned}
$$

By Lemma 2.2 below, the solution f_z above 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. By (2.2)

$$
0 = E[f_z'(W) - Wf_z(W)] = E[I_{\{W\le z\}} - \Phi(z)] = P(W \le z) - \Phi(z).
$$

Thus, W has a standard normal distribution. □

Equation (2.2) is called the Stein equation. In general, for a real valued measurable function h with $E|h(Z)| < \infty$, the Stein equation refers to

$$
f'(w) - wf(w) = h(w) - Eh(Z). \qquad (2.4)
$$

Clearly, if $h(w) = I_{\{w\le z\}}$, (2.4) reduces to (2.2). Similar to (2.3), the solution $f = f_h$ is given by

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

2.2 Properties of solutions to the Stein equations

In this subsection we study basic properties of solutions to the Stein equations (2.3) and (2.5). First, we consider the solution f_z to (2.3).

Lemma 2.2 *For the function f_z defined by* (2.3) *we have*

$$
wf_z(w) \text{ is an increasing function of } w, \qquad (2.6)
$$

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

$$
|f_z'(w)| \le 1, \ |f_z'(w) - f_z'(v)| \le 1 \qquad (2.8)
$$

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

and

$$
|(w+u)f_z(w+u) - (w+v)f_z(w+v)| \le (|w| + \sqrt{2\pi}/4)(|u| + |v|) \qquad (2.10)
$$

for all real w, u, and v.

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

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

which also yields

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

by comparing the derivatives of the two functions. Thus

$$\frac{we^{-w^2/2}}{(1+w^2)\sqrt{2\pi}} \leq 1 - \Phi(w) \leq \frac{e^{-w^2/2}}{w\sqrt{2\pi}}. \tag{2.11}$$

It follows from (2.3) that

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

by (2.11). This proves (2.6).

In view of the fact that

$$\lim_{w\to-\infty} wf_z(w) = \Phi(z) - 1 \text{ and } \lim_{w\to\infty} wf_z(w) = \Phi(z), \tag{2.12}$$

(2.7) follows by (2.6).

By (2.2), we have

$$\begin{aligned} f_z'(w) &= wf_z(w) + I_{\{w\leq z\}} - \Phi(z) \\ &= \begin{cases} wf_z(w) + 1 - \Phi(z) & \text{for } w < z, \\ wf_z(w) - \Phi(z) & \text{for } w > z. \end{cases} \\ &= \begin{cases} (\sqrt{2\pi}we^{w^2/2}\Phi(w) + 1)(1-\Phi(z)) & \text{for } w < z, \\ (\sqrt{2\pi}we^{w^2/2}(1-\Phi(w)) - 1)\Phi(z) & \text{for } w > z. \end{cases} \end{aligned} \tag{2.13}$$

Since $wf_z(w)$ is an increasing function of w, by (2.11) and (2.12)

$$0 < f_z'(w) \leq zf_z(z) + 1 - \Phi(z) < 1 \quad \text{for} \quad w < z \tag{2.14}$$

and

$$-1 < zf_z(z) - \Phi(z) \leq f_z'(w) < 0 \quad \text{for} \quad w > z\ . \tag{2.15}$$

Hence for any w and v,

$$|f_z'(w) - f_z'(v)| \le \max(1, zf_z(z) + 1 - \Phi(z) - (zf_z(z) - \Phi(z)) = 1.$$

This proves (2.8).

Observe that by (2.14) and (2.15), f_z attains its maximum at z. Thus

$$0 < f_z(w) \le f_z(z) = \sqrt{2\pi} e^{z^2/2}\Phi(z)(1 - \Phi(z)). \tag{2.16}$$

By (2.11), $f_z(z) \le 1/z$. To finish the proof of (2.9), let

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

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

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

where $z_0 = (2\ln(4/\pi))^{1/2}$. Thus, $g_1(z)$ is decreasing on $[0, z_0)$ and increasing on (z_0, ∞). 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 maximum at either $z = 0$ or $z = \infty$, that is

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

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

The last inequality (2.10) is a consequence of (2.8) and (2.9) by rewriting $(w + u)f_z(w + u)$-$(w + v)f_z(w + v) = w(f_z(w + u) - f_z(w + v)) + uf_z(w + u) - vf_z(w + v)$ and using the Taylor expansion. □

Next, we discuss the solution f_h for bounded absolutely continuous function h.

Lemma 2.3 *For absolutely continuous function h: $R \to R$*

$$\sup_w |f_h(w)| \le \min\left(\sqrt{\pi/2}\sup_w |h(w) - Eh(Z)|, 2\sup_w |h'(w)|\right), \tag{2.17}$$

$$\sup_w |f_h'(w)| \le \min\left(2\sup_w |h(w) - Eh(Z)|, 4\sup_w |h'(w)|\right) \tag{2.18}$$

and

$$\sup_w |f_h''(w)| \le 2\sup_w |h'(w)|. \tag{2.19}$$

Proof. Let $\tilde{h}(w) = h(w) - Eh(Z)$ and 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 $|h(t)| \le c_1|t|$ and $|Eh(Z)| \le c_1 E|Z| = c_1\sqrt{2/\pi}$.

First we verify (2.17). From the definition (2.5) of f_h, it follows that

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

where in the last inequality we used the fact that

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

Next we prove (2.18). By (2.4), for $w \ge 0$

$$\begin{aligned} |f_h'(w)| &\le |h(w) - Eh(Z)| + we^{w^2/2}\int_{w}^{\infty} |h(x) - Eh(Z)|e^{-x^2/2}dx \\ &\le |h(w) - Eh(Z)| + c_0 we^{w^2/2}\int_{w}^{\infty} e^{-x^2/2}dx \le 2c_0 \end{aligned}$$

by (2.11). It follows from (2.5) again that

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

or equivalently

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

Therefore

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

and by (2.17)

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

Thus we have

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

Similarly, the above bound holds for $\sup_{w\le 0}|f'(w)|$. This proves (2.18).

Now we prove (2.19). Differentiating (2.4) gives

$$\begin{aligned} f_h''(w) &= wf_h'(w) + f_h(w) + h'(w) \\ &= (1+w^2)f_h(w) + w(h(w) - Eh(Z)) + h'(w). \end{aligned} \tag{2.20}$$

From

$$\begin{aligned} & h(x) - Eh(Z) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [h(x) - h(s)] e^{-s^2/2} ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \int_{s}^{x} h'(x) dt e^{-s^2/2} ds - \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \int_{x}^{s} h'(t) dt e^{-s^2/2} ds \\ &= \int_{-\infty}^{x} h'(t)\Phi(t) dt - \int_{x}^{\infty} h'(t)(1 - \Phi(t)) dt, \end{aligned} \tag{2.21}$$

it follows that

$$\begin{aligned} f_h(w) &= e^{w^2/2} \int_{-\infty}^{w} [h(x) - Eh(Z)] e^{-x^2/2} dx \\ &= e^{w^2/2} \int_{-\infty}^{w} \Big(\int_{-\infty}^{x} h'(t)\Phi(t) dt - \int_{x}^{\infty} h'(t)(1 - \Phi(t)) dt \Big) e^{-x^2/2} dx \\ &= -\sqrt{2\pi} e^{w^2/2} (1 - \Phi(w)) \int_{-\infty}^{w} h'(t)\Phi(t) dt \\ & \quad -\sqrt{2\pi} e^{w^2/2} \Phi(w) \int_{w}^{\infty} h'(t)[1 - \Phi(t)] dt. \end{aligned} \tag{2.22}$$

From (2.20) - (2.22) and (2.11) we obtain

$$\begin{aligned} |f_h''(w)| &\le |h'(w)| + |(1+w^2)f_h(w) + w(h(w) - Eh(Z))| \\ &\le |h'(w)| + |\Big(w - \sqrt{2\pi}(1+w^2)e^{w^2/2}(1 - \Phi(w))\Big) \int_{-\infty}^{w} h'(t)\Phi(t) dt| \\ & \quad + |\Big(- w - \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big) \int_{w}^{\infty} h'(t)(1 - \Phi(t)) dt| \\ &\le |h'(w)| + c_1\Big(- w + \sqrt{2\pi}(1+w^2)e^{w^2/2}(1 - \Phi(w))\Big) \int_{-\infty}^{w} \Phi(t) dt \\ & \quad + c_1\Big(w + \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big) \int_{w}^{\infty} (1 - \Phi(t)) dt \\ &= |h'(w)| + c_1\Big(- w + \sqrt{2\pi}(1+w^2)e^{w^2/2}(1 - \Phi(w))\Big)\Big(w\Phi(w) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}\Big) \\ & \quad + c_1\Big(w + \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big)\Big(- w(1 - \Phi(w)) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}\Big) \\ &= |h'(w)| + c_1 \le 2c_1, \end{aligned} \tag{2.23}$$

as desired. □

2.3 The main idea of the Stein approach

The Stein equation(2.4) is the starting point for normal approximations. To illustrate the main idea of this approach, let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$ for each $1 \le i \le n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Put

$$W = \sum_{i=1}^n \xi_i, \ W^{(i)} = W - \xi_i \tag{2.24}$$

and

$$K_i(t) = E\xi_i(I_{\{0\le t\le \xi_i\}} - I_{\{\xi_i \le t<0\}}). \tag{2.25}$$

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

$$\int_{-\infty}^{\infty} K_i(t)dt = E\xi_i^2 \text{ and } \int_{-\infty}^{\infty} |t|K_i(t)dt = E|\xi_i|^3/2. \tag{2.26}$$

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

$$Eh(W) - Eh(Z) = Ef'(W) - EWf(W).$$

Since ξ_i and $W^{(i)}$ are independent and $E\xi_i = 0$ for each $1 \le i \le n$, we have

$$\begin{aligned} EWf(W) &= \sum_{i=1}^n E\xi_i f(W) \\ &= \sum_{i=1}^n E\xi_i[f(W) - f(W^{(i)})] \\ &= \sum_{i=1}^n E\xi_i \int_0^{\xi_i} f'(W^{(i)} + t)dt \\ &= \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W^{(i)} + t)\xi_i(I_{\{0\le t\le\xi_i\}} - I_{\{\xi_i\le t<0\}})dt \\ &= \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W^{(i)} + t)K_i(t)dt. \end{aligned} \tag{2.27}$$

From

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

it follows that

$$Ef'(W) = \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W)K_i(t)dt. \tag{2.28}$$

Thus, by (2.27) and (2.28)

$$Ef'(W) - EWf(W) = \sum_{i=1}^{n} E \int_{-\infty}^{\infty} [f'(W) - f'(W^{(i)} + t)] K_i(t) dt. \tag{2.29}$$

Equations (2.27) and (2.29) play a key role in proving a Berry-Esseen type inequality. We remark that it holds for all bounded absolute continuous f. Let

$$\gamma = \sum_{i=1}^{n} E|\xi_i|^3. \tag{2.30}$$

2.4 Expectation of smooth functions

Equation (2.29) is ready to drive a Berry-Esseen type bound for smooth function h.

Theorem 2.1 *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$, $E|\xi_i|^3 < \infty$ for each $1 \le i \le n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Then for any absolutely continuous function h satisfying $\sup_x |h'(x)| \le c_1$*

$$|Eh(W) - Eh(Z)| \le 2c_1 \sum_{i=1}^{n} E|\xi_i|^3. \tag{2.31}$$

In particular, we have

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

Proof. It follows from (2.19) that $|f_h''| \le 2c_1$. Therefore, by (2.29) and the mean value theorem

$$\begin{aligned} |Ef_h'(W) - EWf_h(W)| &\le \sum_{i=1}^{n} E \int_{-\infty}^{\infty} |f_h'(W) - f_h'(W^{(i)} + t)| K_i(t) dt \\ &\le 2c_1 \sum_{i=1}^{n} E \int_{-\infty}^{\infty} (|t| + |\xi_i|) K_i(t) dt \\ &= 2c_1 \sum_{i=1}^{n} (E|\xi_i|^3/2 + E|\xi_i| E\xi_i^2) \\ &\le 3c_1 \sum_{i=1}^{n} E|\xi_i|^3. \end{aligned} \tag{2.32}$$

We note that it is not necessary to assume the existence of finite third moments in Theorem 2.1.

Theorem 2.2 *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$ for each $1 \le i \le n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Let h be absolutely continuous with $|h'| \le c_1$. Then*

$$|Eh(W) - Eh(Z)| \le 4c_1(4\beta_2 + 3\beta_3), \tag{2.33}$$

where

$$\beta_2 = \sum_{i=1}^n E\xi_i^2 I_{\{|\xi_i|>1\}} \text{ and } \beta_3 = \sum_{i=1}^n E|\xi_i|^3 I_{\{|\xi_i|\le 1\}}. \tag{2.34}$$

Proof. Observing from (2.18) and (2.19) that

$$|f_h'(W) - f_h'(W^{(i)} + t)| \le \min(8c_1, 2c_1(|t| + |\xi_i|)) \le 8c_1(|t| \wedge 1 + |\xi_i| \wedge 1),$$

where $a \wedge b$ denotes $\min(a, b)$, we have by (2.29)

$$\begin{aligned}
&|Eh(W) - Eh(Z)| \\
\le\ & 8c_1 \sum_{i=1}^n E \int_{-\infty}^{\infty} (|t| \wedge 1 + |\xi_i| \wedge 1) K_i(t) dt \\
=\ & 8c_1 \sum_{i=1}^n \Big(E|\xi_i|(|\xi_i| - 1) I_{\{|\xi_i|>1\}} + \frac{1}{2} E|\xi_i|(|\xi_i| \wedge 1)^2 + E\xi_i^2 E(|\xi_i| \wedge 1) \Big) \\
=\ & 8c_1 \sum_{i=1}^n \Big(E\xi_i^2 I_{\{|\xi_i|>1\}} - \frac{1}{2} E|\xi_i| I_{\{|\xi_i|>1\}} + \frac{1}{2} E|\xi_i|^3 I_{\{|\xi_i|\le 1\}} + E\xi_i^2 E(|\xi_i| \wedge 1) \Big) \\
\le\ & 8c_1 \Big\{ \beta_2 + \frac{1}{2}\beta_3 + \sum_{i=1}^n E\xi_i^2 E(|\xi_i| \wedge 1) \Big\}.
\end{aligned} \tag{2.35}$$

We need the following fact: for any random variable ξ

$$E\xi^2 E(|\xi| \wedge 1) \le E|\xi|^3 I_{\{|\xi|\le 1\}} + E\xi^2 I_{\{|\xi|>1\}}. \tag{2.36}$$

To see this, let η be an independent copy of ξ. It is easy to verify that

$$\xi^2(|\eta| \wedge 1) + \eta^2(|\xi| \wedge 1) \le |\xi|^3 I_{\{|\xi|\le 1\}} + |\eta|^3 I_{\{|\eta|\le 1\}} + |\xi|^2 I_{\{|\eta|>1\}} + |\eta|^2 I_{\{|\eta|>1\}}.$$

Taking expectation on both sides yields (2.36).

Now (2.33) follows from (2.35) - (2.36). □

Remark 2.1 *If h is bounded by c_0, then the bound in* (2.33) *can be replaced by* $\max(c_1, 4(c_0 \wedge c_1))(4\beta_2 + 3\beta_3)$.

2.5 The Lindeberg central limit theorem

Since an indicator function is not continuous, unfortunately, Theorem (2.2) does not give a sharp Berry-Esseen bound directly. However, one can use a bounded absolutely continuous function to approximate the indicator function and then apply Theorem 2.2 to obtain a weak version of the Berry-Esseen bound which is good enough to recover the Lindebderg central limit theorem.

Theorem 2.3 *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$ for each $1 \le i \le n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Then*

$$\sup_z |P(W \le z) - \Phi(z)| \le 2.2(4\beta_2 + 3\beta_3)^{1/2}, \tag{2.37}$$

where β_2 and β_3 are defined in (2.34).

Proof. We can assume that $(4\beta_2 + 3\beta_3)^{1/2} \le 1/2$. Otherwise, (2.37) is trivial. Let $\alpha = 0.5(4\beta_2 + 3\beta_3)^{1/2}$, and define for fixed z

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

It is easy to see that $|h| \le 1$, $|h'| \le 1/\alpha$. By Remark 2.1 and the assumption $\alpha \ge 4$, we have

$$|Eh_\alpha(W) - Eh_\alpha(Z)| \le \max(4, 1/\alpha)(4\beta_2 + 3\beta_3) \le (4\beta_2 + 3\beta_3)/\alpha \tag{2.38}$$

and hence

$$\begin{aligned} P(W \le z) - \Phi(z) &\le Eh_\alpha(W) - Eh_\alpha(Z) + Eh_\alpha(Z) - \Phi(Z) \\ &\le (4\beta_2 + 3\beta_3)/\alpha + EI_{\{z \le Z \le z+\alpha\}} \\ &\le (4\beta_2 + 3\beta_3)/\alpha + \frac{\alpha}{\sqrt{2\pi}} \le 2.2(4\beta_2 + 3\beta_3)^{1/2}. \end{aligned} \tag{2.39}$$

Similarly, we have

$$P(W \le z) - \Phi(z) \ge -2.2(4\beta_2 + 3\beta_3)^{1/2}. \tag{2.40}$$

This proves (2.37), by (2.39) and (2.40). □

Although Theorem 2.3 does not give a sharp Berry-Esseen bound, it does provide a self-contained proof for the central limit theorem under Lindeberg's condition.

Let $X_1, X_2, \cdots, X_n$ be independent random variables with $EX_i = 0$ and $EX_i^2 < \infty$ for each $1 \le i \le n$. Put

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

To apply Theorem 2.3, let

$$\xi_i = X_i/B_n \text{ and } W = S_n/B_n. \tag{2.41}$$

Observe that for any $0 < \varepsilon < 1$

$$\begin{aligned}
\beta_2 + \beta_3 &= \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i|>B_n\}} + \frac{1}{B_n^3}\sum_{i=1}^n E|X_i|^3 I_{\{|X_i|\le B_n\}} \\
&\le \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i|>B_n\}} + \frac{1}{B_n^3}\sum_{i=1}^n B_n EX_i^2 I_{\{\varepsilon B_n \le |X_i|\le B_n\}} \\
&\quad + \frac{1}{B_n^3}\sum_{i=1}^n \varepsilon B_n EX_i^2 I_{\{|X_i|<\varepsilon B_n\}} \\
&\le \varepsilon + \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i|>\varepsilon B_n\}}.
\end{aligned} \tag{2.42}$$

If Lindeberg's condition (1.1) is satisfied, then (2.42) implies $\beta_2 + \beta_3 \to 0$ as $n \to \infty$ since ε is arbitrary. This shows

$$\sup_z |P(S_n/B_n \le z) - \Phi(z)| \to 0 \text{ as } n \to \infty$$

by Theorem 2.3.

2.6 Converse to the Lindeberg-Feller theorem

Let $X_1, X_2, \cdots, X_n$ be independent random variables with $EX_i = 0$ and $EX_i^2 < \infty$ for each $1 \le i \le n$. The notation is as in Section 2.5. It is known that if the Feller condition is satisfied

$$\max_{1\le i\le n} EX_i^2/B_n^2 \to 0, \tag{2.43}$$

then Lindebderg's condition is necessary for the central limit theorem. Stein's method can provide a nice proof for the necessity.

Theorem 2.4 *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$ for each $1 \le i \le n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Then there exists an absolute constant C such that for all $\varepsilon > 0$*

$$(1 - e^{-\varepsilon^2/4})\sum_{i=1}^n E\xi_i^2 I_{\{|\xi_i|>\varepsilon\}} \le C\Big(\sup_z |P(W \le z) - \Phi(z)| + \sum_{i=1}^n (E\xi_i^2)^2\Big). \tag{2.44}$$

Now for the sequence of independent random variables $\{X_i, i \geq 1\}$, recall $\xi_i = X_i/B_n$. Clearly, Feller's condition (2.43) implies that $\sum_{i=1}^n (E\xi_i^2)^2 \leq \max_{1\leq i\leq n} E\xi_i^2 \to 0$ as $n \to \infty$. Therefore, if S_n/B_n is asymptotically normal, then

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

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

Proof of Theorem 2.4. Let $f(w) = we^{-w^2/2}$ and $h(w) = f'(w) - wf(w)$. Put

$$c_1 = \int_{-\infty}^{\infty} |h'(w)|dw,\ c_2 = \sup_w |f'''(w)|,\ c_3 = \int_{-\infty}^{\infty} |f''(w)|dw,\ \Delta = \sup_z P(W \leq z) - \Phi(z)|.$$

Numerical computation gives $c_1 \leq 5, c_2 = 3, c_3 \leq 4$.

Since $Eh(Z) = 0$ by (2.1), we have

$$|Eh(W)| = |Eh(W) - Eh(Z)| = |\int_{-\infty}^{\infty} h'(w)\{P(W \leq w) - \Phi(w)\}dw| \leq c_1\Delta. \tag{2.45}$$

Furthermore with $\sigma_i^2 = E\xi_i^2$

$$\begin{aligned} Eh(W) &= \sum_{i=1}^n E(\sigma_i^2 f'(W^{(i)} + \xi_i) - \xi_i f(W^{(i)} + \xi_i)) \\ &= \sum_{i=1}^n \sigma_i^2 E(f'(W^{(i)} + \xi_i) - f'(W^{(i)}) - \xi_i f''(W^{(i)})) \\ &\quad - \sum_{i=1}^n E(\xi_i\{f(W^{(i)} + \xi_i) - f(W^{(i)}) - \xi_i f'(W^{(i)})\}) \\ &\geq -0.5c_2 \sum_{i=1}^n \sigma_i^4 + R_1, \end{aligned} \tag{2.46}$$

where

$$R_1 = -\sum_{i=1}^n E(\xi_i\{f(W^{(i)} + \xi_i) - f(W^{(i)}) - \xi_i f'(W^{(i)})\}).$$

Let W^*, Z and $\{\xi_i, 1 \leq i \leq n\}$ be independent, where W^* and W have the same distribution and Z has the standard normal distribution. Put

$$\begin{aligned} R_2 &= \sum_{i=1}^n E(\xi_i\{f(W^* + \xi_i) - f(W^*) - \xi_i f'(W^*)\}), \\ R_3 &= \sum_{i=1}^n E(\xi_i\{f(Z + \xi_i) - f(Z) - \xi_i f'(Z)\}). \end{aligned}$$

We shall prove that R_1 can be approximated by R_2 and eventually by R_3. Note that

$$\begin{aligned} R_1 &= \sum_{i=1}^n E\Big\{\xi_i^2 \int_0^1 [f'(W^{(i)}+t\xi_i) - f'(W^{(i)})]dt\Big\} \\ &= R_2 + \sum_{i=1}^n E\Big\{\xi_i^2 \int_0^1 [f'(W^*+t\xi_i) - f'(W^{(i)}+t\xi_i)]dt\Big\} \\ &\quad - \sum_{i=1}^n E\Big\{\xi_i^2 \int_0^1 [f'(W^*) - f'(W^{(i)})]dt\Big\}. \end{aligned} \tag{2.47}$$

For any constant θ,

$$\begin{aligned} &|E(f'(W^*+\theta) - f'(W^{(i)}+\theta))| \\ &= |E(f'(W^{(i)}+\xi_i+\theta) - f'(W^{(i)}+\theta))| \\ &= |E(f'(W^{(i)}+\xi_i+\theta) - f'(W^{(i)}+\theta) - \xi_i f''(W^{(i)}+\theta))| \\ &\le 0.5c_2\sigma_i^2. \end{aligned}$$

So by (2.47),

$$R_1 \ge R_2 - 0.5c_2 \sum_{i=1}^n \sigma_i^4. \tag{2.48}$$

Similarly,

$$\begin{aligned} R_2 &= R_3 + \sum_{i=1}^n E\Big\{\xi_i^2 \int_0^1 [f'(Z)+t\xi_i) - f'(W^*+t\xi_i)]dt\Big\} \\ &\quad - \sum_{i=1}^n E\Big\{\xi_i^2 \int_0^1 [f'(Z) - f'(W^*)]dt\Big\} \end{aligned}$$

and for constant θ

$$|Ef'(W^*+\theta) - Ef'(Z+\theta)| = |\int_{-\infty}^{\infty} f''(w)(P(W^* \le w-\theta) - \Phi(w-\theta))dw| \le c_3\Delta.$$

Combining the above estimates with (2.45) - (2.48) yields

$$R_3 \le (c_1+2c_3)\Delta + 1.5c_2 \sum_{i=1}^n \sigma_i^4. \tag{2.49}$$

Observing that

$$g(y) := -y^{-1}E(f(Z+y) - f(Z) - yf'(Z)) = 2^{-1.5}(1-e^{-y^2/4}),$$

we have

$$\begin{aligned} R_3 &= \sum_{i=1}^n E\xi_i^2 g(\xi_i) \\ &\geq 2^{-1.5}(1-e^{-\varepsilon^2/4}) \sum_{i=1}^n E\xi_i^2 I_{\{|\xi_i|>\varepsilon\}} \end{aligned} \tag{2.50}$$

for every $\varepsilon > 0$. This proves (2.44). $\square$

3 Uniform Berry-Esseen Bounds

Throughout this section we assume that $\xi_1, \xi_2, \cdots, \xi_n$ are independent random variables with zero means and finite second moment. We also assume $\sum_{i=1}^n E\xi_i^2 = 1$. Use the notation in the previous section,

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

Let f_z be the solution of the Stein equation (2.1). Our goal is to use Stein's method to prove the uniform Berry-Esseen inequality

$$\sup_z |P(W \leq z) - \Phi(z)| \leq C \sum_{i=1}^n E|\xi_i|^3.$$

3.1 Bounded random variables

For bounded ξ_i, we are ready to apply (2.27) to obtain the following Berry-Esseen type bound.

Theorem 3.1 *If $|\xi_i| \leq \delta_0$ for $1 \leq i \leq n$, then*

$$\sup_z |P(W \leq z) - \Phi(z)| \leq 3.3\delta_0 \tag{3.1}$$

Proof. Write $f = f_z$. It follows from (2.27) that

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

and

$$\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)}+t \le z)K_i(t)dt - \Phi(z) = \sum_{i=1}^n E\int_{-\infty}^{\infty} \{Wf(W)-(W^{(i)}+t)f(W^{(i)}+t)\}K_i(t)dt. \quad (3.2)$$

By (2.10),

$$\begin{aligned}
&\sum_{i=1}^n E\int_{-\infty}^{\infty} |Wf(W)-(W^{(i)}+t)f(W^{(i)}+t)|K_i(t)dt \\
&\le \sum_{i=1}^n \int_{-\infty}^{\infty} E(|W^{(i)}|+\sqrt{2\pi}/4)(|\xi_i|+|t|)K_i(t)dt \\
&\le (1+\sqrt{2\pi}/4)\sum_{i=1}^n \int_{-\infty}^{\infty} (E|\xi_i|+|t|)K_i(t) \\
&= (1+\sqrt{2\pi}/4)\sum_{i=1}^n \{E|\xi_i|E\xi_i^2 + 0.5E|\xi_i|^3\} \\
&\le 1.5(1+\sqrt{2\pi}/4)\sum_{i=1}^n E|\xi_i|^3. \quad (3.3)
\end{aligned}$$

Noting that the assumption $|\xi_i| \le \delta_0$ implies $K_i(t)=0$ for $|t|>\delta_0$, we have

$$\begin{aligned}
&\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)}+t \le z)K_i(t)dt \\
&= \sum_{i=1}^n \int_{|t|\le\delta_0} P(W-\xi_i+t \le z)K_i(t)dt \\
&\ge \sum_{i=1}^n \int_{|t|\le\delta_0} P(W \le z-2\delta_0)K_i(t)dt \\
&= P(W \le z-2\delta_0).
\end{aligned}$$

Combining with (3.2) and (3.3) gives

$$\begin{aligned}
&P(W \le z-2\delta_0) - \Phi(z-2\delta_0) \\
&\le \Phi(z)-\Phi(z-2\delta_0) + 1.5(1+\sqrt{2\pi}/4)\sum_{i=1}^n E|\xi_i|^3 \\
&\le \frac{2\delta_0}{\sqrt{2\pi}} + 1.5(1+\sqrt{2\pi}/4)\delta_0 \le 3.3\delta_0 \quad (3.4)
\end{aligned}$$

Similarly, we have

$$\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)}+t \le z)K_i(t)dt \le P(W \le z+2\delta_0).$$

and

$$P(W \le z + 2\delta_0) - \Phi(z + 2\delta_0) \ge -3.3\delta_0 \tag{3.5}$$

This proves (3.1) by (3.4) and (3.5). $\Box$

One can see from the above proof that the boundness of ξ_i is used only in the approximation of $\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)} + t \le z) K_i(t) dt$ by $P(W \le z)$. On the other hand, it is intuitively appealing that $P(W^{(i)} + t \le z)$ should be close to $P(W \le z)$. We shall present two different approaches in the next two subsections.

3.2 The inductive approach

Assume that $E|\xi_i|^3 < \infty$. We shall prove

$$\sup_z |P(W \le z) - \Phi(z)| \le C \sum_{i=1}^n E|\xi_i|^3 \tag{3.6}$$

by induction, where C can be taken 76.

Let $\gamma = \sum_{i=1}^n E|\xi_i|^3$, $\tau_i^2 = EW^{(i)2}$ and $\tau = \min_{1 \le i \le n} \tau_i$. Since (3.6) is trivial if $\gamma > 1/76$, we can assume $\gamma < 1/76$ which in turn implies that $\tau^2 \ge (1 - \gamma^{2/3}) \ge 0.9$.

If $n = 1$, $E|\xi_1|^3 \ge (E\xi_i^2)^{2/3} = 1$. (3.6) is true. Suppose inductively that (3.6) has been established whenever W consists of fewer than n summands. Then, in particular

$$\begin{aligned} P(a < W^{(i)} \le b) &= \Phi(b/\tau_i) - \Phi(a/\tau_i) + P(W^{(i)} \le b) - \Phi(b/\tau_i) - \{P(W^{(i)} \le a) - \Phi(a/\tau_i)\} \\ &\le 2C\tau_i^{-3} \sum_{j \ne i} E|\xi_j|^3 + (2\pi)^{-1/2} \tau_i^{-1} (b - a) \\ &\le 4C\gamma + (b - a) \end{aligned} \tag{3.7}$$

for $a < b$.

For $\delta = 16\gamma$, we have

$$\begin{aligned} &\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)} + t \le z) K_i(t) dt \\ &= P(W \le z - 2\delta) + \sum_{i=1}^n \int_{-\infty}^{\infty} \{P(W^{(i)} + t \le z) - P(W^{(i)} + \xi_i \le z - 2\delta)\} K_i(t) dt \\ &\ge P(W \le z - 2\delta) - \sum_{i=1}^n \int_{-\infty}^{\infty} E\Big\{ \int_{t \ge 2\delta + \xi_i} I_{\{z - t \le W^{(i)} \le z - 2\delta - \xi_i\}} K_i(t) \Big\} dt \end{aligned}$$

$$
\begin{aligned}
&= P(W \le z - 2\delta) - \sum_{i=1}^n \int_{-\infty}^{\infty} E\Big\{ \int_{t \ge 2\delta + \xi_i} P(z - t \le W^{(i)} \le z - 2\delta - \xi_i) K_i(t) \Big\} dt \\
&\ge P(W \le z - 2\delta) - \sum_{i=1}^n \int_{-\infty}^{\infty} E\Big\{ \int_{t \ge 2\delta + \xi_i} (4C\gamma + t - 2\delta - \xi_i) K_i(t) dt \Big\} \qquad \text{[by (3.7)]} \\
&\ge P(W \le z - 2\delta) - 2\delta - 1.5\gamma - 4C\gamma \sum_{i=1}^n E\Big\{ \int_{t \ge 2\delta + \xi_i} K_i(t) dt \Big\} \\
&\ge P(W \le z - 2\delta) - 2\delta - 1.5\gamma \\
&\quad -4C\gamma \sum_{i=1}^n E\Big\{ I_{\{\xi_i \le -\delta\}} \int_{-\infty}^{\infty} K_i(t) dt + \int_{t \ge \delta} K_i(t) dt \Big\} \\
&\ge P(W \le z - 2\delta) - 34\gamma - 4C\gamma \sum_{i=1}^n (P(\xi_i < -\delta) E\xi_i^2 + E\xi_i^2 I_{\{\xi_i > \delta\}}) \\
&\ge P(W \le z - 2\delta) - 34\gamma - 4C\gamma \sum_{i=1}^n E|\xi_i|^3/\delta \\
&= P(W \le z - 2\delta) - 34\gamma - C\gamma/2. \qquad (3.8)
\end{aligned}
$$

Thus, by (3.2) and (3.3)

$$P(W \le z - 2\delta) - \Phi(z - 2\delta) \le 38\gamma + C\gamma/2 = C\gamma$$

if we take $C = 76$. Similarly, we have

$$P(W \le z + 2\delta) - \Phi(z + 2\delta) \ge -38\gamma - C\gamma/2 = -C\gamma.$$

This completes the proof of (3.6). □

3.3 The concentration inequality approach

The proofs in previous two subsections suggest that the key step in proving the Berry-Esseen bound is the concentration inequality (3.7). In this subsection, we give a direct proof for (3.7) and hence the Berry-Esseen inequality. Let $\gamma = \sum_{i=1}^n E|\xi_i|^3$.

Proposition 3.1 *We have*

$$P(a \le W^{(i)} \le b) \le \sqrt{2}(b - a) + (1 + \sqrt{2})\gamma \qquad (3.9)$$

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

Proof. Define $\delta = \gamma/2$ and

$$f(w) = \begin{cases} -\frac{1}{2}(b-a) - \delta & \text{if } \; w < a - \delta, \\ w - \frac{1}{2}(b+a) & \text{if } \; a - \delta \le w \le b + \delta, \\ \frac{1}{2}(b-a) + \delta & \text{for } \; w > b + \delta \end{cases} \tag{3.10}$$

Let

$$\begin{aligned} \hat{M}_j(t) &= \xi_j (I_{\{-\xi_j \le t \le 0\}} - I_{\{0 < t \le -\xi_j\}}), \\ \hat{M}(t) &= \sum_{1 \le j \le n} \hat{M}_j(t), \;\; M(t) = E\hat{M}(t). \end{aligned}$$

Since ξ_j and $W^{(i)} - \xi_j$ are independent for $j \neq i$, $E\xi_j = 0$, $\hat{M}(t) \geq 0$ and $f'(t) \geq 0$, we have

$$\begin{aligned} & EW^{(i)} f(W^{(i)}) - E\xi_i f(W^{(i)} - \xi_i) \\ &= \sum_{j=1}^n E\xi_j [f(W^{(i)}) - f(W^{(i)} - \xi_j)] \\ &= \sum_{j=1}^n E\xi_j \int_{-\xi_j}^0 f'(W^{(i)} + t) dt \\ &= \sum_{j=1}^n E\xi_j [f(W^{(i)}) - f(W^{(i)} - \xi_j)] \\ &= \sum_{j=1}^n E \int_{-\infty}^{\infty} f'(W^{(i)} + t) \hat{M}_j(t) dt \\ &= E \int_{-\infty}^{\infty} f'(W^{(i)} + t) \hat{M}(t) dt \\ &\geq E \int_{|t| \le \delta} f'(W^{(i)} + t) \hat{M}(t) dt \\ &\geq EI_{\{a \le W^{(i)} \le b\}} \int_{|t| \le \delta} \hat{M}(t) dt \\ &= EI_{\{a \le W^{(i)} \le b\}} \sum_{j=1}^n |\xi_j| \min(\delta, |\xi_j|) \\ &\geq H_{1,1} - H_{1,2}, \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} H_{1,1} &= P(a \le W^{(i)} \le b) \sum_{j=1}^n E|\xi_j| \min(\delta, |\xi_j|), \\ H_{1,2} &= E|\sum_{j=1}^n |\xi_j| \min(\delta, |\xi_j|) - E|\xi_j| \min(\delta, |\xi_j|)|. \end{aligned}$$

A direct calculation yields

$$\min(x,y) \geq x - x^2/(4y) \tag{3.12}$$

for $x > 0$ and $y > 0$. Then

$$\sum_{j=1}^n E|\xi_j|\min(\delta,|\xi_j|) \geq \sum_{j=1}^n \Big\{E\xi_j^2 - \frac{E|\xi_j|^3}{4\delta}\Big\} = \frac{1}{2} \tag{3.13}$$

and hence

$$H_{1,1} \geq .5P(a \leq W^{(i)} \leq b). \tag{3.14}$$

By the Hölder inequality,

$$\begin{aligned}
H_{1,2} &\leq \Big(\mathrm{Var}(\sum_{j=1}^n |\xi_j|\min(\delta,|\xi_j|))\Big)^{1/2} \\
&\leq \Big(\sum_{j=1}^n E\xi_j^2 \min(\delta,|\xi_j|)^2\Big)^{1/2} \\
&\leq \delta\Big(\sum_{j=1}^n E\xi_j^2\Big)^{1/2} = \delta. \qquad (3.15)
\end{aligned}$$

Combining (3.14) and (3.15) with (3.11) and observing that

$$|f| \leq .5(b-a)+\delta,$$

we have

$$\begin{aligned}
P(a \leq W^{(i)} \leq b) &\leq 2\delta + (E|W^{(i)}| + E|\xi_i|)(b-a+2\delta) \\
&\leq 2\delta + \sqrt{2}\left((E|W^{(i)}|)^2 + (E|\xi_i|)^2\right)^{1/2}(b-a+2\delta) \\
&\leq 2\delta + \sqrt{2}\left(E|W^{(i)}|^2 + E|\xi_i|^2\right)^{1/2}(b-a+2\delta) \\
&= 2\delta + \sqrt{2}\,(EW^2)^{1/2}(b-a+2\delta) \\
&= \sqrt{2}(b-a) + 2(1+\sqrt{2})\delta \\
&= \sqrt{2}(b-a) + (1+\sqrt{2})\gamma
\end{aligned}$$

as desired. □

We are now ready to prove

Theorem 3.2 *We have*

$$\sup_z |P(W \leq z) - \Phi(z)| \leq 7\gamma. \tag{3.16}$$

Proof. It follows from (3.9) that

$$
\begin{aligned}
&|\sum_{i=1}^n \int_{-\infty}^{\infty} P(W^{(i)} + t \le z) K_i(t) dt - P(W \le z)| \\
\le\ & \sum_{i=1}^n \int_{-\infty}^{\infty} |P(W^{(i)} + t \le z) - P(W \le z)| K_i(t) dt \\
=\ & \sum_{i=1}^n \int_{-\infty}^{\infty} E\{P(z - \max(t, \xi_i) \le W^{(i)} \le z - \min(t, \xi_i) \mid \xi_i)\} K_i(t) dt \\
\le\ & \sum_{i=1}^n \int_{-\infty}^{\infty} E\{\sqrt{2}(|t| + |\xi_i|) + (1+\sqrt{2})\gamma\} K_i(t) dt \\
=\ & (1+\sqrt{2})\gamma + \sqrt{2} \sum_{i=1}^n (0.5 E|\xi_i|^3 + E|\xi_i| E\xi_i^2) \\
\le\ & (1 + 2.5\sqrt{2})\gamma. \qquad (3.17)
\end{aligned}
$$

Now by (3.2)

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

which is (3.1). □

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

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

We leave the proof to the reader. A refined concentration inequality can lead to reduce the constant 7 to 4.1.

3.4 A randomized concentration inequality

In this subsection we present a randomized concentration inequality, which is useful to establish the Berry-Esseen inequality for functions of independent random variables, in particular for non-linear statistics. Let Δ_1 and Δ_2 be real-valued Borel measurable functions of $(\xi_i, 1 \le i \le n)$.

Theorem 3.3 *We have*

$$
\begin{aligned}
P(\Delta_1 \le W \le \Delta_2) \le\ & E|W(\Delta_2 - \Delta_1)| + 2\gamma \\
& + \sum_{i=1}^n \{E|\xi_i(\Delta_1 - \Delta_{1,i})| + E|\xi_i(\Delta_2 - \Delta_{2,i})|\}, \qquad (3.18)
\end{aligned}
$$

where $\Delta_{1,i}$ and $\Delta_{2,i}$ are Borel measurable functions of $(\xi_j, 1 \le j \le n,\ j \ne i)$, and γ is defined as in (2.30).

Proof. We follow the proof of Proposition 3.1. Define $\delta = 0.5\gamma$ and

$$f_{\Delta_1,\Delta_2}(w) = \begin{cases} -(\Delta_2 - \Delta_1)/2 - \delta & \text{for } w \le \Delta_1 - \delta \\ w - \frac{1}{2}(\Delta_1 + \Delta_2) & \text{for } \Delta_1 - \delta \le w \le \Delta_2 + \delta \\ (\Delta_2 - \Delta_1)/2 + \delta & \text{for } w > \Delta_2 + \delta. \end{cases}$$

Let

$$\hat{M}_i(t) = \xi_i\{I(-\xi_i \le t \le 0) - I(0 < t \le -\xi_i)\}, \ \hat{M}(t) = \sum_{i=1}^n \hat{M}_i(t).$$

Since ξ_i and $f_{\Delta_{1,i},\Delta_{2,i}}(W - \xi_i)$ are independent for $1 \le i \le n$ and $E\xi_i = 0$, we have

$$\begin{aligned} EWf_{\Delta_1,\Delta_2}(W) &= \sum_{i=1}^n E\xi_i[f_{\Delta_1,\Delta_2}(W) - f_{\Delta_1,\Delta_2}(W - \xi_i))] \\ &\quad + \sum_{i=1}^n E\xi_i[f_{\Delta_1,\Delta_2}(W - \xi_i) - f_{\Delta_{1,i},\Delta_{2,i}}(W - \xi_i))\} \\ &:= H_1 + H_2. \end{aligned} \tag{3.19}$$

Using the fact that $\hat{M}(t) \ge 0$ and $f'_{\Delta_1,\Delta_2}(w) \ge 0$, we have

$$\begin{aligned} H_1 &= \sum_{i=1}^n E\Big\{\xi_i \int_{-\xi_i}^0 f'_{\Delta_1,\Delta_2}(W+t)dt\Big\} \\ &= \sum_{i=1}^n E\Big\{\int_{-\infty}^{\infty} f'_{\Delta_1,\Delta_2}(W+t)\hat{M}_i(t)dt\Big\} \\ &= E\Big\{\int_{-\infty}^{\infty} f'_{\Delta_1,\Delta_2}(W+t)\hat{M}(t)dt\Big\} \\ &\ge E\Big\{\int_{|t|\le\delta} f'_{\Delta_1,\Delta_2}(W+t)\hat{M}(t)dt\Big\} \\ &\ge E\Big\{I_{\{\Delta_1 \le W \le \Delta_2\}} \int_{|t|\le\delta} \hat{M}(t)dt\Big\} \\ &= E\Big\{I_{\{\Delta_1 \le W \le \Delta_2\}} \sum_{i=1}^n |\xi_i| \min(\delta, |\xi_i|)\Big\}. \end{aligned} \tag{3.20}$$

From the proof of (3.14) and (3.15) one can see that

$$H_1 \ge .5P(\Delta_1 \le W \le \Delta_2) - \delta. \tag{3.21}$$

As to H_2, it is easy to see that

$$|f_{\Delta_1,\Delta_2}(w) - f_{\Delta_{1,i},\Delta_{2,i}}(w)| \le |\Delta_1 - \Delta_{1,i}|/2 + |\Delta_2 - \Delta_{2,i}|/2.$$

Hence

$$|H_2| \le (1/2) \sum_{i=1}^{n} \{E|\xi_i(\Delta_1 - \Delta_{1,i})| + E|\xi_i(\Delta_2 - \Delta_{2,i})|\}. \tag{3.22}$$

It follows from the definition of f_{Δ_1,Δ_2} that

$$|f_{\Delta_1,\Delta_2}(w)| \le (1/2)(\Delta_2 - \Delta_1) + \delta.$$

Hence, by (3.19), (3.21), and (3.22)

$$\begin{aligned}
&P(\Delta_1 \le W \le \Delta_2) \\
&\le \; 2EWf_{\Delta_1,\Delta_2}(W) + 2\delta + \sum_{i=1}^{n} \{E|\xi_i(\Delta_1 - \Delta_{1,i})| + E|\xi_i(\Delta_2 - \Delta_{2,i})|\} \\
&\le \; E|W(\Delta_2 - \Delta_1)| + 2\delta E|W| + 2\delta + \sum_{i=1}^{n} \{E|\xi_i(\Delta_1 - \Delta_{1,i})| + E|\xi_i(\Delta_2 - \Delta_{2,i})|\} \\
&\le \; E|W(\Delta_2 - \Delta_1)| + 2\gamma + \sum_{i=1}^{n} \{E|\xi_i(\Delta_1 - \Delta_{1,i})| + E|\xi_i(\Delta_2 - \Delta_{2,i})|\}.
\end{aligned}$$

□

It follows easily from Theorems 3.3 and 3.2 that

Theorem 3.4 *Let $\Delta = \Delta(\xi_1, \cdots, \xi_n) : R^n \longrightarrow R^1$ be a Borel measurable function. Then we have Then we have*

$$\sup_z |P(W + \Delta \le z) - \Phi(z)| \le 9\gamma + E|W\Delta| + \sum_{i=1}^{n} E|\xi_i(\Delta - \Delta_i)|, \tag{3.23}$$

where Δ_i is a measurable function of $(\xi_j, 1 \le j \le n, j \ne i)$.

Theorem 3.4 provides a general result on Berry-Esseen type bounds for many non-linear statistics. To see the usefulness of the above general result, let's consider the U-statistic. Let $X_1, X_2, \cdots, X_n$ be a sequence of independent identically distributed random variables, and let $h(x, y)$ be a real-valued Borel measurable symmetric function, i.e., $h(x, y) = h(y, x)$. Define the U-statistic with the kernel h by

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

Theorem 3.5 *Assume that* $Eh(X_1, X_2) = 0$ *and* $\sigma^2 = Eh^2(X_1, X_2) < \infty$. *Let* $g(x) = Eh(x, X_2)$ *and* $\sigma_1^2 = Eg^2(X_1)$. *If* $\sigma_1 > 0$, *then*

$$\sup_z |P(\frac{\sqrt{n}U_n}{2\sigma_1} \le z) - \Phi(z)| \le \frac{2\sigma}{(n-1)^{1/2}\sigma_1} + \frac{9E|g(X_1)|^3}{n^{1/2}\sigma_1^3}. \tag{3.24}$$

Proof. Let

$$\begin{aligned} W &= \frac{1}{\sqrt{n}\sigma_1}\sum_{i=1}^n g(X_i), \\ \Delta &= \frac{\sqrt{n}}{n(n-1)\sigma_1}\sum_{1\le i<j\le n}\{h(X_i, X_j) - g(X_i) - g(X_j)\}, \\ \Delta_l &= \frac{\sqrt{n}}{n(n-1)\sigma_1}\sum_{1\le i<j\le n, i\ne l, j\ne l}\{h(X_i, X_j) - g(X_i) - g(X_j)\}. \end{aligned}$$

It is easy to see that

$$\frac{\sqrt{n}U_n}{2\sigma_1} = W + \Delta$$

and that Δ_l is a measurable function of $(X_j, 1 \le j \le n, j \ne l)$. By Theorem 3.4, it suffices to show that

$$E\Delta^2 \le \frac{\sigma^2}{2(n-1)\sigma_1^2} \tag{3.25}$$

and

$$E|\Delta - \Delta_l|^2 \le \frac{\sigma^2}{n(n-1)\sigma_1^2}. \tag{3.26}$$

It is known that $\{\sum_{i=1}^{j-1}(h(X_i, X_j) - g(X_i) - g(X_j)), 2 \le j \le n\}$ is a martingale difference sequence. Hence

$$\begin{aligned} E\Delta^2 &= \frac{1}{n(n-1)^2\sigma_1^2}\sum_{j=2}^n E\Big(\sum_{i=1}^{j-1}\{h(X_i, X_j) - g(X_i) - g(X_j)\}\Big)^2 \\ &= \frac{1}{n(n-1)^2\sigma_1^2}\sum_{j=2}^n E\Big(E\Big\{\Big(\sum_{i=1}^{j-1}\{h(X_i, X_j) - g(X_i) - g(X_j)\}\Big)^2 \mid X_j\Big\}\Big) \\ &= \frac{1}{n(n-1)^2\sigma_1^2}\sum_{j=2}^n (j-1)E\Big\{E\Big((h(X_1, X_2) - g(X_1) - g(X_2))^2 \mid X_j\Big)\Big\} \\ &= \frac{1}{2(n-1)\sigma_1^2}\Big\{Eh^2(X_1, X_2) - 2Eg^2(X_1)\Big\} \\ &\le \frac{\sigma^2}{2(n-1)\sigma_1^2}. \end{aligned}$$

This proves (3.25).

As to (3.26), note that $\Delta - \Delta_l$, $1 \le l \le n$ are identically distributed. Thus,

$$\begin{aligned} E|\Delta - \Delta_l|^2 &= E|\Delta - \Delta_1|^2 \\ &= \frac{1}{n(n-1)^2\sigma_1^2} E\Big(\sum_{j=2}^n \{h(X_1, X_j) - g(X_1) - g(X_j)\}\Big)^2 \\ &= \frac{1}{n(n-1)\sigma_1^2}\Big\{Eh^2(X_1, X_2) - 2Eg^2(X_1)\Big\} \\ &\le \frac{\sigma^2}{n(n-1)\sigma_1^2}. \end{aligned}$$

This is (3.26). □

4 Non-uniform Berry-Esseen Bounds

We shall prove the non-uniform Berry-Esseen bound in the normal approximation in this section. To do this, we first need to have a non-uniform concentration inequality.

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

$$\bar\xi_i = \xi_i I_{\{\xi_i \le 1\}},\ \bar W = \sum_{i=1}^n \bar\xi_i,\ \bar W^{(i)} = \bar W - \bar\xi_i.$$

Proposition 4.1 *We have*

$$P(a \le \bar W^{(i)} \le b) \le e^{-a/2}(5(b-a) + 7\gamma) \tag{4.1}$$

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

We first need to have the following Bennett-Hoeffding inequality.

Lemma 4.1 *Let $\eta_1, \eta_2, \cdots, \eta_n$ be independent random variables satisfying $E\eta_i \le 0$, $\eta_i \le a$ for $1 \le i \le n$, and $\sum_{i=1}^n E\eta_i^2 \le B_n^2$. Put $S_n = \sum_{i=1}^n \eta_i$. Then*

$$Ee^{tS_n} \le \exp\Big(a^{-2}(e^{ta} - 1 - ta)B_n^2\Big) \tag{4.2}$$

for $t > 0$,

$$P(S_n \ge x) \le \exp\Big(- \frac{B_n^2}{a^2}\Big[(1 + \frac{ax}{B_n^2})\ln(1 + \frac{ax}{B_n^2}) - \frac{ax}{B_n^2}\Big]\Big) \tag{4.3}$$

and

$$P(S_n \geq x) \leq \exp\Big(- \frac{x^2}{2(B_n^2 + ax)}\Big) \tag{4.4}$$

for $x > 0$.

Proof. It is easy to see that $(e^s - 1 - s)/s^2$ is an increasing function of s. We have

$$e^{ts} \leq 1 + ts + (ts)^2 (e^{ta} - 1 - ta)/(ta)^2 \tag{4.5}$$

for $s \leq a$. We have

$$\begin{aligned} Ee^{tS_n} &= \prod_{i=1}^n Ee^{t\eta_i} \\ &\leq \prod_{i=1}^n \big(1 + tE\eta_i + a^{-2}(e^{ta} - 1 - ta)E\eta_i^2\big) \\ &\leq \prod_{i=1}^n \big(1 + a^{-2}(e^{ta} - 1 - ta)E\eta_i^2\big) \\ &\leq \exp\Big(a^{-2}(e^{ta} - 1 - ta)B_n^2\Big). \end{aligned}$$

This proves (4.2).

To prove (4.3), let

$$t = \frac{1}{a} \ln\Big((1 + \frac{1x}{B_n}\Big).$$

Then, by (4.2)

$$\begin{aligned} P(S_n \geq x) &\leq e^{-tx} Ee^{tS_n} \\ &\leq \exp\Big(- tx + a^{-2}(e^{ta} - 1 - ta)B_n^2\Big) \\ &= \exp\Big(- \frac{B_n^2}{a^2}\Big[(1 + \frac{ax}{B_n^2}) \ln(1 + \frac{ax}{B_n^2}) - \frac{ax}{B_n^2}\Big]\Big). \end{aligned}$$

In view of the fact that

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

for $s > 0$, (4.4) follows from (4.3). $\square$

Proof of Proposition 4.1. It follows from (4.2) that

$$P(a \leq W^{(i)} \leq b) \leq e^{-a/2} Ee^{W^{(i)}/2} \leq e^{-a/2} \exp(e^{0.5} - 1.5) \leq 1.19e^{-a/2}.$$

Thus, (4.1) holds if $7\gamma \geq 1.19$.

We now assume $\gamma \le 0.17$. Similarly to the proof of Proposition 3.1, define $\delta = \gamma/2(\le 0.085)$ and

$$f(w) = \begin{cases} 0 & \text{if } w < a - \delta, \\ e^{w/2}(w - a + \delta) & \text{if } a - \delta \le w \le b + \delta, \\ e^{w/2}(b - a + 2\delta) & \text{if } w > b + \delta \end{cases} \tag{4.6}$$

Put

$$\bar{M}_i(t) = \xi_i(I_{\{-\bar{\xi}_i \le t \le 0\}} - I_{\{0 < t \le -\bar{\xi}_i\}}), \ \bar{M}(t) = \sum_{i=1}^n \bar{M}_i(t).$$

Clearly, $\bar{M}(t) \ge 0$, $f'(w) \ge 0$ and $f'(w) \ge e^{w/2}$ for $a - \delta \le w \le b + \delta$. Analogous to (3.11),

$$\begin{aligned} & EW^{(i)} f(\bar{W}^{(i)}) \\ = \ & \sum_{j \ne i} E\xi_j [f(\bar{W}^{(i)}) - f(W^{(i)} - \bar{\xi}_j)] \\ = \ & \sum_{j \ne i}^n E \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t)\bar{M}_i(t) dt \\ = \ & E \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t)\bar{M}^{(i)}(t) dt \\ \ge \ & EI_{\{a \le \bar{W}^{(i)} \le b\}} \int_{|t| \le \delta} f'(\bar{W}^{(i)} + t)\bar{M}^{(i)}(t) dt \\ \ge \ & Ee^{\bar{W}^{(i)} - \delta} I_{\{a \le \bar{W}^{(i)} \le b\}} \int_{|t| \le \delta} \bar{M}^{(i)}(t) dt \\ \ge \ & Ee^{\bar{W}^{(i)} - \delta} I_{\{a \le \bar{W}^{(i)} \le b\}} \sum_{j \ne i} |\xi_j| \min(\delta, |\bar{\xi}_j|) \\ \ge \ & e^{-\delta/2}(H_{2,1} - H_{2,2}), \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} H_{2,1} &= Ee^{\bar{W}^{(i)}/2} I_{\{a \le \bar{W}^{(i)} \le b\}} \sum_{j \ne i} E|\xi_j| \min(\delta, |\bar{\xi}_j|), \\ H_{2,2} &= Ee^{\bar{W}^{(i)}/2} |\sum_{j \ne i} |\xi_j| \min(\delta, |\bar{\xi}_j|) - E|\xi_j| \min(\delta, |\bar{\xi}_j|)|. \end{aligned}$$

Noting that $\delta \le .085$ and $\gamma \le .17$ and following the proof of (3.13), we have

$$\begin{aligned} \sum_{j \ne i} E|\xi_j| \min(\delta, |\bar{\xi}_j|) &= \sum_{j \ne i} E|\xi_j| \min(\delta, |\xi_j|) \\ &\ge -\delta E|\xi_i| + \sum_{j=1}^n E|\xi_j| \min(\delta, |\xi_j|) \\ &\ge 0.5 - \delta\gamma^{1/3} \ge 0.5 - 0.085(0.17)^{1/3} \ge 0.45. \end{aligned} \tag{4.8}$$

Hence

$$H_{2,1} \geq .45 e^{a/2} P(a \leq \bar{W}^{(i)} \leq b) \tag{4.9}$$

By the Bennett inequality (4.2) again, we have

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

and hence

$$\begin{aligned} H_{2,2} &\leq (E e^{\bar{W}^{(i)}})^{1/2} \Big(\mathrm{Var}(\sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|))\Big)^{1/2} \\ &\leq \exp(.5e - 1)\delta \leq 1.44\delta. \end{aligned} \tag{4.10}$$

As to the left hand side of (4.6), we have

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

Combining the above inequalities yields

$$\begin{aligned} P(a \leq \bar{W}^{(i)} \leq b) &\leq \frac{e^{-a/2}}{.45}\Big(e^{\delta/2} 2.06(b - a + 2\delta) + 1.44\delta\Big) \\ &\leq \frac{e^{-a/2}}{.45}\Big(e^{.0425} 2.06(b - a + 2\delta) + 1.44\delta\Big) \\ &\leq e^{-a/2}(4.8(b-a) + 12.76\delta) \\ &\leq e^{-a/2}(5(b-a) + 7\gamma). \end{aligned}$$

This proves (4.1). $\square$

We also need the following moment inequality.

Lemma 4.2 *Let $2 < p \leq 3$, and $\{\eta_i, 1 \leq i \leq n\}$ be independent random variables with $E\eta_i = 0$ and $E|\eta_i|^p < \infty$. Put $S_n = \sum_{i=1}^n \eta_i$ and $B_n^2 = \sum_{i=1}^n E\eta_i^2$. Then*

$$E|S_n|^p \leq (p-1)B_n^p + \sum_{i=1}^n E|\eta_i|^p \tag{4.11}$$

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

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

as desired. □

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

Theorem 4.1 *There exists an absolute constant C such that for every real number z,*

$$
|P(W \le z) - \Phi(z)| \le \frac{C\gamma}{1 + |z|^3}. \tag{4.12}
$$

Proof. Without loss of generality, assume $z \ge 0$. By (4.11),

$$
P(W \ge z) \le \frac{1 + E|W|^3}{1 + z^3}.
$$

So (4.12) holds if $\gamma \ge 1$, and we can assume $\gamma \le 1$. Let

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

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

$$P(W > z) \leq P(\bar{W} > z) + P(W > z, \max_{1\leq i\leq n} \xi_i > 1) \tag{4.13}$$

and similarly,

$$P(\bar{W} > z) \leq P(W > z) + P(\bar{W} > z, \max_{1\leq i\leq n} \xi_i > 1). \tag{4.14}$$

Note that

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

here and in the sequel, C denotes an absolute constant but whose value may be different at each appearance. Similarly,

$$\begin{aligned}&P(\bar{W} > z, \max_{1\leq i\leq n} \xi_i > 1) \\ &\leq \sum_{i=1}^n P(\bar{W} > z, \xi_i > 1) \\ &= \sum_{i=1}^n P(\bar{W}^{(i)} > z - \xi_i I_{\{\xi_i \leq 1\}}, \xi_i > 1) \\ &= \sum_{i=1}^n P(\bar{W}^{(i)} > z, \xi_i > 1) \\ &= \sum_{i=1}^n P(\bar{W}^{(i)} > z) P(\xi_i > 1)\end{aligned}$$

$$
\begin{aligned}
&\le e^{-z/2}\sum_{i=1}^n Ee^{\bar W^{(i)}/2}P(\xi_i>1)\\
&\le 2e^{-z/2}\gamma \le \frac{C\gamma}{1+z^3}
\end{aligned}
$$

by (4.2). Thus, to prove (4.1), it suffices to show that

$$
|P(\bar W\le z)-\Phi(z)|\le Ce^{-z/2}\gamma. \tag{4.15}
$$

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

$$
\bar K_i(t)=E\bar\xi_i(I_{\{0\le t\le\bar\xi_i\}}-I_{\{\bar\xi_i\le t<0\}}).
$$

Following the proof of (2.27) and noting that $\bar\xi_i\le 1$, we have

$$
E\bar W f_z(\bar W)=\sum_{i=1}^n E\int_{-\infty}^1 f_z'(\bar W^{(i)}+t)\bar K_i(t)dt+\sum_{i=1}^n E\bar\xi_i Ef_z(\bar W^{(i)}).
$$

From

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

we obtain that

$$
\begin{aligned}
&P(\bar W\le z)-\Phi(z)\\
&= Ef_z'(\bar W)-E\bar W f_z(\bar W)\\
&= \sum_{i=1}^n E\xi_i^2 I_{\{\xi_i>1\}}Ef_z'(\bar W)\\
&\quad+\sum_{i=1}^n E\int_{-\infty}^1[f_z'(\bar W^{(i)}+\bar\xi_i)-f_z'(\bar W^{(i)}+t)]\bar K_i(t)dt\\
&\quad+\sum_{i=1}^n E\xi_i I_{\{\xi_i>1\}}Ef_z(\bar W^{(i)})\\
&:= R_1+R_2+R_3.
\end{aligned} \tag{4.16}
$$

By (2.13), (2.8) and (4.2),

$$
\begin{aligned}
E|f_z'(\bar W)| &= E|f_z'(\bar W)|I_{\{\bar W\le z/2\}}+E|f_z'(\bar W)|I_{\{\bar W>z/2\}}\\
&\le (1+\sqrt{2\pi}(z/2)e^{z^2/8})(1-\Phi(z))+P(\bar W>z/2)\\
&\le (1+\sqrt{2\pi}(z/2)e^{z^2/8})(1-\Phi(z))+e^{-z/2}Ee^{\bar W}\\
&\le Ce^{-z/2}
\end{aligned}
$$

Hence

$$|R_1| \le C\gamma e^{-z/2}. \tag{4.17}$$

Similarly, we have $Ef_z(\bar{W}^{(i)}) \le Ce^{-z/2}$ and

$$|R_3| \le C\gamma e^{-z/2}. \tag{4.18}$$

To estimate R_2, write

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

where

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

By Proposition 4.1,

$$\begin{aligned} R_{2,1} &\le \sum_{i=1}^n E\int_{-\infty}^1 I_{\{\bar{\xi}_i\le t\}}P(z-t<\bar{W}^{(i)}\le z-\bar{\xi}_i \mid \bar{\xi}_i)\bar{K}_i(t)dt \\ &\le C\sum_{i=1}^n E\int_{-\infty}^1 e^{-(z-t)/2}(|\bar{\xi}_i|+|t|+\gamma)\bar{K}_i(t)dt \\ &\le Ce^{-z/2}\gamma. \end{aligned} \tag{4.19}$$

From Lemma 4.3 below it follows that

$$\begin{aligned} R_{2,2} &\le \sum_{i=1}^n E\int_{-\infty}^1 I_{\{t\le\bar{\xi}_i\}}[E(\{\bar{W}^{(i)}+\bar{\xi}_i\}f_z(\bar{W}^{(i)}+\bar{\xi}_i) \mid \bar{\xi}_i) - E(\bar{W}^{(i)}+t)f_z(\bar{W}^{(i)}+t)]\bar{K}_i(t)dt \\ &\le Ce^{-z/2}\sum_{i=1}^n E\int_{-\infty}^1 (|\bar{\xi}_i|+|t|)\bar{K}_i(t)dt \\ &\le Ce^{-z/2}\gamma. \end{aligned} \tag{4.20}$$

Therefore

$$R_2 \le Ce^{-z/2}\gamma. \tag{4.21}$$

Similarly, we have

$$R_2 \ge -Ce^{-z/2}\gamma. \tag{4.22}$$

This proves the theorem. □

We remain to prove the following lemma.

Lemma 4.3 *For $s < t \leq 1$ we have*

$$\begin{aligned} & E(\bar{W}^{(i)} + t) f_z(\bar{W}^{(i)} + t) - E(\bar{W}^{(i)} + s) f_z(\bar{W}^{(i)} + s) \\ \leq \;& Ce^{-z/2}(|s| + |t|) \end{aligned} \tag{4.23}$$

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

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

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

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

Thus, we have

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

Hence, by (4.2)

$$\begin{aligned} Eg(W^{(i)} + u) &= Eg(W^{(i)} + u) I_{\{W^{(i)} + u \leq z/2\}} + Eg(W^{(i)} + u) I_{\{W^{(i)} + u > z/2\}} \\ &\leq 4(1 + z^2) e^{z^2/8}(1 - \Phi(z)) + 4(1 + z^2) e^{z^2/2}(1 - \Phi(z)) P(W^{(i)} + u > z/2) \\ &\leq Ce^{-z/2} + C(1 + z) e^{-z + 2u} Ee^{2W^{(i)}} \\ &\leq Ce^{-z/2} + C(1 + z) e^{-z} Ee^{2W^{(i)}} \qquad \text{since } u \leq 1 \\ &\leq Ce^{-z/2}, \end{aligned}$$

which gives

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

This proves (4.23). □

5 Uniform and Non-uniform Bounds under Local Dependence

In this section we discuss normal approximation under local dependence using Stein's method. Our aim is to establish optimal uniform and non-uniform Berry-Esseen bounds under local dependence. Local dependence is more general than m-dependence for sequences of random variables. It applies to random variables with arbitrary index set, such as those indexed by the vertices of a graph with dependence defined in terms of common edges.

Throughout this section let $\mathcal{J}$ be an index set and $\{\xi_i, i \in \mathcal{J}\}$ be a random field with zero means and finite variances, and let n be the cardinality of $\mathcal{J}$. Define $W = \sum_{i\in\mathcal{J}} \xi_i$ and assume that $\mathrm{Var}(W) = 1$.

For $A \subset \mathcal{J}$, let ξ_A denote $\{\xi_i, i \in A\}$, $A^c = \{j \in \mathcal{J} : j \notin A\}$, and $|A|$ the cardinality of A.

We first introduce dependence assumptions.

(LD1) For each $i \in \mathcal{J}$ there exists $A_i \subset \mathcal{J}$ such that ξ_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 ξ_i is independent of $\xi_{A_i^c}$ and ξ_{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 ξ_i is independent of $\xi_{A_i^c}$, ξ_{A_i} is independent of $\xi_{B_i^c}$, and ξ_{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 ξ_i is independent of $\xi_{A_i^c}$, ξ_{A_i} is independent of $\xi_{B_i^c}$, ξ_{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 5.1 *Let $N(B_i) = \{j \in \mathcal{J} : B_j B_i \neq \emptyset\}$ and $2 < p \leq 4$. Assume that $(LD2)$ is satisfied with $|N(B_i)| \leq \kappa$. Then*

$$\sup_z |P(W \leq z) - \Phi(z)| \leq (13 + 11\kappa) \sum_{i\in\mathcal{J}} (E|\xi_i|^{3\wedge p} + E|Y_i|^{3\wedge p}) + 2.5\Big(\kappa \sum_{i\in\mathcal{J}} (E|\xi_i|^p + E|Y_i|^p)\Big)^{1/2}\Big\}$$

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

$$\sup_z |P(W \le z) - \Phi(z)| \le (13 + 11\kappa)\, n\, \theta^{3 \wedge p} + 2.5\theta^{p/2}\sqrt{\kappa n}, \tag{5.1}$$

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

Note that in many cases κ is bounded and θ 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 5.2 *Suppose that* $(LD3)$ *is satisfied. Let* $2 < p \le 3$. *Assume that* $(LD3)$ *is satisfied with* $|N(C_i)| \le \kappa$, *where* $N(C_i) = \{j \in \mathcal{J} : \ C_i B_j \ne \emptyset\}$. *Then*

$$\sup_z |P(W \le z) - \Phi(z)| \le 75\kappa^{p-1} \sum_{i \in \mathcal{J}} E|\xi_i|^p. \tag{5.2}$$

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

Theorem 5.3 *Assume that* $E|\xi_i|^p < \infty$ *for* $2 < p \le 3$ *and that* $(LD4^*)$ *is satisfied. Let* $\kappa = \max_{i \in \mathcal{J}} \max(|D_i^*|, |\{j : i \in D_j^*\}|)$. *Then*

$$|P(W \le z) - \Phi(z)| \le C\kappa^p (1 + |z|)^{-p} \sum_{i \in \mathcal{J}} E|\xi_i|^p, \tag{5.3}$$

where C *is an absolute constant.*

The above results can immediately be applied to m-dependent random fields. Let $d \ge 1$ and Z^d denote the d-dimensional space of positive integers. The distance between two points $i = (i_1, \cdots, i_d)$ and $j = (j_1, \cdots, j_d)$ in Z^d is defined by $|i - j| = \max_{1 \le l \le d} |i_l - j_l|$ 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 $A_i = \{j : \ |j - i| \le m\} \cap \mathcal{J}$, $B_i = \{j : \ |j - i| \le 2m\} \cap \mathcal{J}$, $C_i = \{j : \ |j - i| \le 3m\} \cap \mathcal{J}$, $B_i^* = \{j : |j - i| \le 3m\} \cap \mathcal{J}$, $C_i^* = \{j : \ |j - i| \le 4m\} \cap \mathcal{J}$, and $D_i^* = \{j : \ |j - i| \le 5m\} \cap \mathcal{J}$ in Theorems 5.2 and 5.3 yields a uniform and a non-uniform bound.

Theorem 5.4 *Let $\{\xi_i, i \in \mathcal{J}\}$ be an m-dependent random fields with zero means and finite $E|\xi_i|^p < \infty$ for $2 < p \le 3$. Then*

$$\sup_z |P(W \le z) - \Phi(z)| \le 75(10m+1)^{(p-1)d} \sum_{i\in\mathcal{J}} E|\xi_i|^p \tag{5.4}$$

and

$$|P(W \le z) - \Phi(z)| \le C(1+|z|)^{-p} 11^{pd}(m+1)^{(p-1)d} \sum_{i\in\mathcal{J}} E|\xi_i|^p, \tag{5.5}$$

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 and Shao (2002) for details.

Define

$$\begin{aligned} \hat{K}_i(t) &= \xi_i\{I(-Y_i \le t < 0) - I(0 \le t \le -Y_i)\},\ K_i(t) = E\hat{K}_i(t), \\ \hat{K}(t) &= \textstyle\sum_{i\in\mathcal{J}} \hat{K}_i(t),\ K(t) = E\hat{K}(t) = \sum_{i\in\mathcal{J}} K_i(t). \end{aligned} \tag{5.6}$$

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

$$\begin{aligned} E\{Wf(W)\} &= \sum_{i\in\mathcal{J}} E\{\xi_i(f(W) - f(W - Y_i))\} \\ &= \sum_{i\in\mathcal{J}} E\Big\{\xi_i \int_{-Y_i}^0 f'(W+t)dt\Big\} \\ &= \sum_{i\in\mathcal{J}} E\Big\{\int_{-\infty}^{\infty} f'(W+t)\hat{K}_i(t)dt\Big\} \\ &= E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \end{aligned} \tag{5.7}$$

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

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

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

$$\begin{aligned}
|R_1| &\leq E\int_{-\infty}^{\infty}(|W|+1)|t||K(t)|dt \\
&\quad +|E\int_{-\infty}^{\infty}(I_{\{W\leq z\}} - I_{\{W+t\leq z\}})K(t)dt| \\
&\leq 0.5\sum_{i=1}^{n} E(|W|+1)|\xi_i|Y_i^2 + \int_{-\infty}^{\infty} P(z-\max(t,0)\leq W\leq z-\min(t,0))K(t)dt \\
&:= R_{1,1}+R_{1,2}
\end{aligned}$$

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

Observe that

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

which can also be estimated easily. The main difficulty arises from estimating R_3. The reader may refer to Chen and Shao (2002) for details.

At the end this section, we give the simplest non-uniform concentration inequality in the paper Chen and Shao (2002) and provide a detailed proof to illustrate the difficulty for dependent variables.

Proposition 5.1 *Assume (LD1). Then for any real numbers $a < b$,*

$$P(a\leq W\leq b)\leq 0.625(b-a)+4r_1+4r_2, \tag{5.8}$$

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

Proof. Let $\alpha = r_1$ and define

$$f(w) = \begin{cases} -(b-a+\alpha)/2 & \text{for } w\leq a-\alpha \\ \frac{1}{2\alpha}(w-a+\alpha)^2-(b-a+\alpha)/2 & \text{for } a-\alpha < w\leq a \\ w-(a+b)/2 & \text{for } a < w\leq b \\ -\frac{1}{2\alpha}(w-b-\alpha)^2+(b-a+\alpha)/2 & \text{for } b < w\leq b+\alpha \\ (b-a+\alpha)/2 & \text{for } w > b+\alpha \end{cases} \tag{5.9}$$

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 $|f(w)| \le (b-a+\alpha)/2$. With this f, Y_i, and $\hat{K}(t)$ and $K(t)$ as defined in (5.6), we have by (5.7)

$$\begin{aligned}
(b-a+\alpha)/2 &\ge EWf(W) = E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \\
&:= Ef'(W)\int_{-\infty}^{\infty} K(t)dt + E\int_{-\infty}^{\infty}(f'(W+t)-f'(W))K(t)dt \\
&\quad +E\int_{-\infty}^{\infty} f'(W+t)(\hat{K}(t)-K(t))dt \\
&:= H_1+H_2+H_3. \qquad (5.10)
\end{aligned}$$

Clearly,

$$H_1 = Ef'(W) \ge P(a \le W \le b). \qquad (5.11)$$

By the Cauchy inequality,

$$\begin{aligned}
|H_3| &\le (1/8)E\int_{-\infty}^{\infty}[f'(W+t)]^2dt + 2E\int_{-\infty}^{\infty}(\hat{K}(t)-K(t))^2dt \\
&\le (b-a+2\alpha)/8 + 2r_2. \qquad (5.12)
\end{aligned}$$

To bound H_2, let

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

Then by writing

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

we have

$$\begin{aligned}
|H_2| &\le \alpha^{-1}\int_0^{\infty}\int_0^t L(\alpha)ds|K(t)|dt + \alpha^{-1}\int_{-\infty}^0\int_t^0 L(\alpha)ds|K(t)|dt \qquad (5.13) \\
&= \alpha^{-1}L(\alpha)\int_{-\infty}^{\infty}|tK(t)|dt \le 0.5\alpha^{-1}r_1L(\alpha) = 0.5L(\alpha).
\end{aligned}$$

It follows from (5.10) - (5.13) that

$$P(a \le W \le b) \le 0.625(b-a) + 0.75\alpha + 2r_2 + 0.5L(\alpha). \qquad (5.14)$$

Substituting $a = x$ and $b = x + \alpha$ in (5.14), we obtain

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

and hence

$$L(\alpha) \le 2.75\alpha + 4r_2. \tag{5.15}$$

Finally combining (5.14) and (5.15), we obtain (5.8). □

6 Exchangeable Pair Approach

Let W be a random variable which is not necessary the partial sum of independent random variables. Suppose that W is approximately normal, we want to get the rate of convergence. Another basic approach of Stein's method is via introducing an exchangeable pair $(W, \hat{W})$. That is, $(W, \hat{W})$ and $(\hat{W}, W)$ have the same distribution. The approach is based on the fact that for all antisymmetric measurable function $g(x, y)$

$$Eg(W, \hat{W}) = 0 \tag{6.1}$$

provided the expected value exists.

A key identity is the following lemma.

Lemma 6.1 *Let $(W, \hat{W})$ be an exchangeable pair of real random variables such that*

$$E(\hat{W}|W) = (1-\lambda)W, \ E(\hat{W} - W)^2 = 2\lambda, \tag{6.2}$$

where $0 < \lambda < 1$. Then for every piecewise continuous function f satisfying $|f(w)| \le C(1 + |w|)$, we have

$$EWf(W) = \frac{1}{2\lambda}E(W - \hat{W})(f(W) - f(\hat{W})). \tag{6.3}$$

Proof. By (6.1),

$$\begin{aligned} 0 &= E(W - \hat{W})(f(\hat{W}) + f(W)) \\ &= E(W - \hat{W})(f(\hat{W}) - f(W)) + 2Ef(W)(W - \hat{W}) \\ &= E(W - \hat{W})(f(\hat{W}) - f(W)) + 2E\{f(W)E(W - \hat{W} \mid W)\} \\ &= E(W - \hat{W})(f(\hat{W}) - f(W)) + 2\lambda EWf(W) \qquad \text{[by (6.2)]}, \end{aligned}$$

which gives (6.3). □

Now we prove

Theorem 6.1 *Let h be absolute continuous with bounded h'. Then under the condition of Lemma 6.1*

$$|Eh(W) - Eh(Z)| \le 2\sup_x |h(x) - Eh(Z)|E|1 - \frac{1}{2\lambda}E((\hat{W} - W)^2 \mid W)| + \frac{1}{4\lambda}\sup_x |h'(x)|E|W - \hat{W}|^3. \tag{6.4}$$

Proof. Let $f = f_h$ be the Stein solution in (2.5) and define

$$\hat{K}(t) = (W - \hat{W})(I_{\{-(W-\hat{W}) \le t \le 0\}} - I_{\{0 < t \le -(W-\hat{W})\}}).$$

By (6.3),

$$EWf(W) = \frac{1}{2\lambda}E\int_{-(W-\hat{W})}^{0} f'(W+t)(W-\hat{W})dt = \frac{1}{2\lambda}E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt$$

and

$$Ef'(W) = Ef'(W)(1 - \frac{1}{2\lambda}(W-\hat{W})^2) + \frac{1}{2\lambda}E\int_{-\infty}^{\infty} f'(W)\hat{K}(t)dt.$$

Therefore

$$\begin{aligned}
|Eh(W) - Eh(Z)| &= |Ef'(W) - EWf(W)| \\
&= |Ef'(W)(1 - \frac{1}{2\lambda}(W-\hat{W})^2) + \frac{1}{2\lambda}E\int_{-\infty}^{\infty}(f'(W) - f'(W+t))\hat{K}(t)dt| \\
&\le |E\Big\{f'(W)(1 - \frac{1}{2\lambda}E((W-\hat{W})^2 \mid W))\Big\}| \\
&\quad + \frac{1}{2\lambda}E|\int_{-\infty}^{\infty}(f'(W) - f'(W+t))\hat{K}(t)dt| \\
&\le 2\sup_x |h(x) - Eh(Z)|E|(1 - \frac{1}{2\lambda}E((W-\hat{W})^2 \mid W)| \\
&\quad + \frac{1}{2\lambda}\sup_x |h'(x)|E|\int_{-\infty}^{\infty}|t|\hat{K}(t)dt| \qquad \text{[by (2.17) and (2.19)]} \\
&= 2\sup_x |h(x) - Eh(Z)|E|1 - \frac{1}{2\lambda}E((\hat{W}-W)^2 \mid W)| + \frac{1}{4\lambda}\sup_x |h'(x)|E|W-\hat{W}|^3
\end{aligned}$$

as desired. □

We end this section with the following example to show how to estimate the bound in the above theorem. Let ξ_i be independent random variables with zero means and $\sum_{i=1}^n E\xi_i^2 = 1$, and put $W = \sum_{i=1}^n \xi_i$. Let $\{\eta_i^*, 1 \le i \le n\}$ be an independent copy of $\{\xi_i, 1 \le i \le n\}$, and I have uniform distribution on $\{1, 2, \cdots, n\}$. Assume that I, $\{\xi_i\}$, $\{\xi_i^*\}$ are independent. Define $\hat{W} = W - X_I + X_I^*$. Then $(W, \hat{W})$ is an exchangeable pair satisfying

$$E(\hat{W}|W) = (1 - \frac{1}{n})W, \quad E(W - \hat{W})^2 = \frac{2}{n}.$$

That means (6.2) is satisfied with $\lambda = 1/n$. Directly calculation also gives

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

and

$$E((W - \hat{W})^2 \mid W) = \frac{1}{n}(1 + \sum_{i=1}^n E(\xi_i^2 \mid W)).$$

Thus,

$$\begin{aligned} & E|1 - \frac{1}{2\lambda}E((\hat{W} - W)^2 \mid W)| \\ = \ & (1/2)E|1 - E(\sum_{i=1}^n \xi_i^2 \mid W)| \\ \le \ & (1/2)E|\sum_{i=1}^n (\xi_i^2 - E\xi_i^2)|. \end{aligned}$$

So the bound is sharp if the fourth moment of ξ_i exists.

References

[1] Barbour, A. D. and Hall, P. (1984). Stein's method and the Berry-Esseen theorem. *Austral. J. Statist.* **26**, 8-15.

[2] Bikelis, A. (1966). Estimates of the remainder in the central limit theorem. *Litovsk. Mat. Sb.* **6(3),** 323-46 (in Russian).

[3] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379-386.

[4] Chen, L.H.Y., Shao, Q.M. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Th. Related Fields* **120**, 236-254.

[5] Chen, L.H.Y., Shao, Q.M. (2002). Normal approximation under local dependence.

[6] Chen, L.H.Y., Shao, Q.M. (2003). Uniform and non-uniform bounds in normal approximation for non-linear statistics.

[7] Esseen, C. G. (1945). Fourier analysis of distribution functions: a mathematical study of the Laplace-Gaussian law. *Acta Math.* **77,** 1-125.

[8] Feller, W. (1968). On the Berry-Esseen theorem. *Z. Wahrsch. Verw. Gebiete* **10**, 261-268.

[9] Michel, R. (1981). On the constant in the non-uniform version of the Berry-Esseen Theorem. *Z. Wahrsch. Verw. Gebiete* **55,** 109-117.

[10] Nagaev, S. V. (1965). Some limit theorems for large deviations. *Theory Probab. Appl.* **10,** 214-235.

[11] Paditz, L. (1977). Über die Annäherung der Verteilungsfunktionen von Summen unabhängiger Zufallsgröben gegen unberrenzt teilbare Verteilungsfunktionen unter besonderer berchtung der Verteilungsfunktion der standarddisierten Normalverteilung. Dissertation, A.TU Dresden.

[12] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Varaibles.* Oxford Studies in Probability 4, Clarendon Press, Oxford.

[13] Stein, C. (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. Stat. Prob.* **2,** 583-602, Univ. California Press. Berkeley, Calif.

[14] Stein, C. (1986). *Approximation Computation of Expectations.* Lecture Notes 7, Inst. Math. Statist., Hayward, Calif.

[15] Stroock, D. W. (1993). *Probability Theory: an analytic view.* Cambridge Univ. Press, Cambridge, U.K.