Chapter 3
Berry–Esseen Bounds for Independent Random Variables

In this chapter we illustrate some of the main ideas of the Stein method by proving the classical Lindeberg central limit theorem and the Berry–Esseen inequality for sums of independent random variables. We begin with Lipschitz functions, which suffice to prove the Lindeberg theorem. We then prove the Berry–Esseen inequality by developing a concentration inequality. Throughout this chapter we assume that $W = \xi_1 + \cdots + \xi_n$ where $\xi_1, \ldots, \xi_n$ are independent random variables satisfying

$$E\xi_i = 0, \quad 1 \le i \le n \quad \text{and} \quad \sum_{i=1}^n \operatorname{Var}(\xi_i) = 1. \tag{3.1}$$

Though we focus on the independent case, the ideas developed here provide a basis for handling more general situations, see, for instance, Theorem 3.5 and its consequence, Theorem 5.2.

Recall that the supremum, L^∞, or Kolmogorov distance between two distribution functions F and G is given by

$$\|F - G\|_\infty = \sup_{z\in\mathbb{R}} \bigl|F(z) - G(z)\bigr|.$$

The main goal of this chapter is to prove the Berry–Esseen inequality, first shown by Berry (1941), and Esseen (1942), which gives a uniform bound between F, the distribution function of W, and Φ, that of the standard normal Z, of the form

$$\|F - \Phi\|_\infty \le C \sum_{i=1}^n E|\xi_i|^3 \tag{3.2}$$

where C is an absolute constant. The upper bound on the smallest possible value of C has decreased from Esseen's original estimate of 7.59 to its current value of 0.4785 by Tyurin (2010). After proving the Lindeberg and Berry–Esseen theorems, the latter using both the concentration inequality and inductive approaches, we end the chapter with a lower bound on $\|F - \Phi\|_\infty$.

L.H.Y. Chen et al., *Normal Approximation by Stein's Method*,
Probability and Its Applications,
DOI 10.1007/978-3-642-15007-4_3, © Springer-Verlag Berlin Heidelberg 2011

3.1 Normal Approximation with Lipschitz Functions

We recall that $h:\mathbb{R}\to\mathbb{R}$ is a Lipschitz continuous function if there exists a constant K such that

$$\big|h(x)-h(y)\big|\le K|x-y| \quad \text{for all } x,y\in\mathbb{R}.$$

Equivalently, h is Lipschitz continuous if and only if h is absolutely continuous with $\|h'\|<\infty$.

Theorem 3.1 *Let $W=\sum_{i=1}^n \xi_i$ be the sum of mean zero independent random variables ξ_i, $1\le i\le n$ with $\sum_{i=1}^n \mathrm{Var}(\xi_i)=1$, and h a Lipschitz continuous function. If $E|\xi_i|^3<\infty$ for $i=1,\dots,n$, then*

$$\big|Eh(W)-Nh\big|\le 3\|h'\|\gamma, \tag{3.3}$$

where

$$\gamma=\sum_{i=1}^n E|\xi_i|^3. \tag{3.4}$$

Proof By Lemma 2.8, (2.77) holds with $\Delta=\xi_I^*-\xi_I$, where ξ_i^* has the ξ_i-zero bias distribution and is independent of ξ_j, $j\neq i$, and I is a random index with distribution (2.60), independent of all other variables. Invoking (2.79) of Proposition 2.4,

$$\begin{aligned}
\big|Eh(W)-Nh\big| &\le 2\|h'\|E\big|\xi_I^*-\xi_I\big| \\
&= 2\|h'\|\sum_{i=1}^n E\big|\xi_i^*-\xi_i\big|E\xi_i^2 \\
&\le 2\|h'\|\sum_{i=1}^n \big(E\big|\xi_i^*\big|E\xi_i^2+E|\xi_i|E\xi_i^2\big) \\
&= 2\|h'\|\sum_{i=1}^n \left(\frac{1}{2}E|\xi_i|^3+E|\xi_i|E\xi_i^2\right) \\
&\le 3\|h'\|\sum_{i=1}^n E|\xi_i|^3,
\end{aligned}$$

where we have invoked (2.57) to obtain the second equality, followed by Hölder's inequality. □

The constant 3 is improved to 1 in Corollary 4.2.

The following theorem shows that one can bound $|Eh(W)-Nh|$ in terms of sums of the truncated second and third moments

$$\beta_2=\sum_{i=1}^n E\xi_i^2\mathbf{1}_{\{|\xi_i|>1\}} \quad\text{and}\quad \beta_3=\sum_{i=1}^n E|\xi_i|^3\mathbf{1}_{\{|\xi_i|\le 1\}}, \tag{3.5}$$

without the need to assume the existence of third moments as in Theorem 3.1.

Theorem 3.2 *If $W = \sum_{i=1}^n \xi_i$ is the sum of mean zero independent random variables ξ_i, $1 \le i \le n$ with $\sum_{i=1}^n \mathrm{Var}(\xi_i) = 1$, then for any Lipschitz function h*

$$\left|Eh(W) - Nh\right| \le \|h'\|(4\beta_2 + 3\beta_3). \tag{3.6}$$

Proof We adopt the same notation as in the proof of Theorem 3.1. The key observation is that we can follow the proof of (2.79) in Proposition 2.4, but instead of applying $|f_h'(W) - f_h'(W+\Delta)| \le 2\|h'\||\Delta|$, we instead use

$$\left|f_h'(W) - f_h'(W+\Delta)\right| \le \min\bigl(2\left\|f_h'\right\|, \left\|f_h''\right\||\Delta|\bigr) \le 2\|h'\|\bigl(1 \wedge |\Delta|\bigr), \tag{3.7}$$

which holds by (2.13), where $a \wedge b$ denotes $\min(a,b)$. Hence

$$\begin{aligned}\left|Eh(W) - Nh\right| &\le 2\|h'\|E\bigl(1 \wedge \left|\xi_I^* - \xi_I\right|\bigr)\\ &\le 2\|h'\|E\bigl(1 \wedge \bigl(\left|\xi_I^*\right| + |\xi_I|\bigr)\bigr)\\ &\le 2\|h'\|\bigl(E\bigl(1 \wedge \left|\xi_I^*\right|\bigr) + E\bigl(1 \wedge |\xi_I|\bigr)\bigr).\end{aligned} \tag{3.8}$$

Letting $\mathrm{sign}(x)$ be $+1$ for $x > 0$, -1 for $x < 0$ and 0 for $x = 0$, setting

$$f(x) = x\mathbf{1}_{|x|>1} + \frac{1}{2}x^2\,\mathrm{sign}(x)\mathbf{1}_{|x|\le 1} \quad \text{we have} \quad f'(x) = 1 \wedge |x|.$$

Hence, (2.60) and (2.51) now yield

$$\begin{aligned}E\bigl(1 \wedge \left|\xi_I^*\right|\bigr) &= \sum_{i=1}^n E\bigl(1 \wedge \left|\xi_i^*\right|\bigr)E\xi_i^2\\ &= \sum_{i=1}^n E\left(\xi_i^2\mathbf{1}_{|\xi_i|>1} + \frac{1}{2}|\xi_i|^3\mathbf{1}_{|\xi_i|\le 1}\right) = \beta_2 + \frac{1}{2}\beta_3.\end{aligned} \tag{3.9}$$

We recall the fact that if g and h are increasing functions, then $Eg(\xi)Eh(\xi) \le Eg(\xi)h(\xi)$. Now, regarding the second term in (3.8), since both $1 \wedge |x|$ and x^2 are increasing functions of $|x|$, again applying (2.60),

$$\begin{aligned}E\bigl(1 \wedge |\xi_I|\bigr) &= \sum_{i=1}^n E\bigl(1 \wedge |\xi_i|\bigr)E\xi_i^2 \le E\sum_{i=1}^n\bigl(1 \wedge |\xi_i|\bigr)\xi_i^2\\ &\le E\sum_{i=1}^n \xi_i^2\mathbf{1}_{\{|\xi_i|>1\}} + |\xi_i|^3\mathbf{1}_{\{|\xi_i|\le 1\}} = \beta_2 + \beta_3.\end{aligned} \tag{3.10}$$

Substituting the bounds (3.9) and (3.10) into (3.8) now gives the result. □

One cannot derive a sharp Berry–Esseen bound for W using the smooth function bounds (3.3) or (3.6). Nevertheless, as noted by Erickson (1974), these smooth function bounds imply a weak L^∞ bound, as highlighted in the following theorem.

Theorem 3.3 *Assume that there exists a δ such that, for any Lipschitz function h,*

$$\left|Eh(W) - Nh\right| \le \delta\|h'\|. \tag{3.11}$$

Then

$$\sup_{z\in\mathbb{R}}\bigl|P(W\le z)-\Phi(z)\bigr|\le 2\delta^{1/2}. \tag{3.12}$$

Proposition 2.4 shows that (3.11) is satisfied under conditions (2.76) or (2.77). Though the resulting bound δ has the optimal rate in many applications, see, for example, (3.3) and (3.6), the rate of a Berry–Esseen bound of the type (3.12) may not be optimal.

Proof We can assume that $\delta \le 1/4$, since otherwise (3.12) is trivial. Let $\alpha = \delta^{1/2}(2\pi)^{1/4}$, and for some fixed $z\in\mathbb{R}$ define

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

Then h is Lipschitz continuous with $\|h'\| = 1/\alpha$, and hence, by (3.11),

$$\begin{aligned}P(W\le z)-\Phi(z)&\le Eh_\alpha(W)-Nh_\alpha+Nh_\alpha-\Phi(z)\\&\le\frac{\delta}{\alpha}+P(z\le Z\le z+\alpha)\\&\le\frac{\delta}{\alpha}+\frac{\alpha}{\sqrt{2\pi}}.\end{aligned}$$

Therefore

$$P(W\le z)-\Phi(z)\le 2(2\pi)^{-1/4}\delta^{1/2}\le 2\delta^{1/2}.$$

Similarly, we have

$$P(W\le z)-\Phi(z)\ge -2\delta^{1/2},$$

proving (3.12). □

3.2 The Lindeberg Central Limit Theorem

Let $\xi_1,\dots,\xi_n$ be independent random variables satisfying $E\xi_i=0$, $1\le i\le n$ and $\sum_{i=1}^n \mathrm{Var}(\xi_i)=1$, and let $W=\sum_{i=1}^n\xi_i$. The classical Lindeberg central limit theorem states that

$$\sup_{z\in\mathbb{R}}\bigl|P(W\le z)-\Phi(z)\bigr|\to 0\quad\text{as } n\to\infty$$

if the Lindeberg condition is satisfied, that is, if for all $\varepsilon>0$

$$\sum_{i=1}^n E\xi_i^2\mathbf{1}_{\{|\xi_i|>\varepsilon\}}\to 0\quad\text{as } n\to\infty. \tag{3.13}$$

With β_2 and β_3 as in (3.5), observe that for any $0 < \varepsilon < 1$,

$$\begin{aligned}\beta_2 + \beta_3 &= \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i|>1\}} + \sum_{i=1}^n E|\xi_i|^3 \mathbf{1}_{\{|\xi_i|\le 1\}} \\ &\le \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i|>1\}} + \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{\varepsilon<|\xi_i|\le 1\}} + \sum_{i=1}^n \varepsilon E\xi_i^2 \mathbf{1}_{\{|\xi_i|\le \varepsilon\}} \\ &\le \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i|>\varepsilon\}} + \varepsilon. \end{aligned} \tag{3.14}$$

Hence, if the Lindeberg condition (3.13) holds, then (3.14) implies $\beta_2 + \beta_3 \to 0$ as $n \to \infty$, since ε is arbitrary. Therefore, by Theorems 3.2 and 3.3,

$$\sup_{z\in\mathbb{R}} \bigl| P(W \le z) - \Phi(z) \bigr| \le 4(\beta_2 + \beta_3)^{1/2} \to 0 \quad \text{as } n \to \infty, \tag{3.15}$$

thus proving the Lindeberg central limit theorem.

In Sect. 3.5 we prove the partial converse, that if $\max_{1\le i\le n} E\xi_i^2 \to 0$, then the Lindeberg condition (3.13) is necessary for normal convergence.

3.3 Berry–Esseen Inequality: The Bounded Case

In the previous section, the smooth function bounds in Theorem 3.3 (see also Proposition 2.4) are of order $O(\delta)$, while the L^∞ bounds are only of the larger order $O(\delta^{1/2})$. Here, we turn to deriving L^∞ bounds which are of comparable order to those of the smooth function bounds. We will use the notation introduced in Sect. 2.3.1,

$$\begin{aligned} W &= \sum_{i=1}^n \xi_i, \qquad W^{(i)} = W - \xi_i, \\ \text{and} \quad K_i(t) &= E\xi_i \bigl(\mathbf{1}_{\{0\le t\le \xi_i\}} - \mathbf{1}_{\{\xi_i \le t<0\}}\bigr). \end{aligned} \tag{3.16}$$

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

Theorem 3.4 *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with zero means satisfying $\sum_{i=1}^n \mathrm{Var}(\xi_i) = 1$, and $W^{(i)}$ and $K_i(t)$ as in (3.16). Then*

$$\left| \sum_{i=1}^n \int_{-\infty}^{\infty} P\bigl(W^{(i)} + t \le z\bigr) K_i(t)\,dt - \Phi(z) \right| \le 2.44\gamma \tag{3.17}$$

where γ is given in (3.4).

If in addition $|\xi_i| \le \delta_0$ for $1 \le i \le n$, then

$$\sup_{z\in\mathbb{R}} \bigl| P(W \le z) - \Phi(z) \bigr| \le 3.3\delta_0. \tag{3.18}$$

Before starting the proof, we note that by (2.62), we may write (3.17) as

$$\left|P(W^* \le z) - \Phi(z)\right| \le 2.44\gamma. \tag{3.19}$$

Proof For $z \in \mathbb{R}$, let $f = f_z$ be the solution of the Stein equation (2.2). From (2.27) and (2.2),

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

Reorganizing this equality, using $\sum_{i=1}^{n} \int_{-\infty}^{\infty} K_i(t)\,dt = 1$ from (2.28), and recalling $K_i(t)$ is real yields

$$\begin{aligned}
&\sum_{i=1}^{n} \int_{-\infty}^{\infty} P(W^{(i)} + t \le z) K_i(t)\,dt - \Phi(z) \\
&\quad = \sum_{i=1}^{n} \int_{-\infty}^{\infty} E\{Wf(W) - (W^{(i)} + t) f(W^{(i)} + t)\} K_i(t)\,dt.
\end{aligned} \tag{3.20}$$

Now, by (2.10), we may bound the absolute value of (3.20) by

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

since $E(W^{(i)})^2 \le 1$ and ξ_i and $W^{(i)}$ are independent. Hence, recalling (2.25), we have

$$\begin{aligned}
&\left|\sum_{i=1}^{n} \int_{-\infty}^{\infty} P(W^{(i)} + t \le z) K_i(t)\,dt - \Phi(z)\right| \\
&\quad \le (1 + \sqrt{2\pi}/4) \sum_{i=1}^{n} \left\{E|\xi_i| E\xi_i^2 + \frac{1}{2} E|\xi_i|^3\right\} \\
&\quad \le \frac{3}{2}(1 + \sqrt{2\pi}/4)\gamma \le 2.44\gamma
\end{aligned}$$

proving (3.17).

The proof would be finished if $P(W^{(i)} + t \le z)$ could be replaced by $P(W \le z)$, since $\sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t)\,dt = 1$. Note that $|\xi_i| \le \delta_0$ implies that $K_i(t) = 0$ for $|t| > \delta_0$, and when both $|t|$ and $|\xi_i|$ are bounded by δ_0 then

$$P\bigl(W^{(i)} + t \le z\bigr) = P(W - \xi_i + t \le z) \ge P(W \le z - 2\delta_0). \tag{3.21}$$

Replacing z by $z + 2\delta_0$ in (3.17) and (3.21) we obtain

$$\begin{aligned}
2.44\gamma &\ge \sum_{i=1}^n \int_{-\infty}^{\infty} P\bigl(W^{(i)} + t \le z + 2\delta_0\bigr) K_i(t)\,dt - \Phi(z + 2\delta_0) \\
&\ge \sum_{i=1}^n \int_{-\infty}^{\infty} P(W \le z) K_i(t)\,dt - \Phi(z + 2\delta_0) \\
&\ge P(W \le z) - \Phi(z) - \frac{2\delta_0}{\sqrt{2\pi}},
\end{aligned}$$

where we have applied (2.28) followed by an elementary inequality. Next, as $|\xi_i| \le \delta_0$ for all $i = 1, \ldots, n$,

$$\gamma = \sum_{i=1}^n E|\xi_i|^3 \le \delta_0 \sum_{i=1}^n E|\xi_i|^2 = \delta_0,$$

from which we now obtain

$$P(W \le z) - \Phi(z) \le 2.44\gamma + \frac{2\delta_0}{\sqrt{2\pi}} \le 3.3\delta_0. \tag{3.22}$$

The proof is completed by proving the corresponding lower bound using similar reasoning. □

The key ingredient in the proof of Theorem 3.4 is to rewrite $E[Wf(W)]$ in terms of a functional of f'. We now formulate a result along these same lines, taking as our basis the Stein identity (2.76).

Theorem 3.5 *For W any random variable, suppose that for every $z \in \mathbb{R}$ there exist a random variable R_1 and random function $\hat{K}(t) \ge 0$, $t \in \mathbb{R}$, and constants δ_0 and δ_1 not depending on z, such that $|ER_1| \le \delta_1$ and*

$$EWf_z(W) = E\int_{|t|\le\delta_0} f_z'(W + t)\hat{K}(t)\,dt + ER_1, \tag{3.23}$$

where f_z is the solution of the Stein equation (2.2). Then

$$\sup_{z\in\mathbb{R}} \bigl|P(W \le z) - \Phi(z)\bigr| \le \delta_0\bigl(1.1 + E\bigl[|W|\hat{K}_1\bigr]\bigr) + 2.7E|1 - \hat{K}_1| + \delta_1, \tag{3.24}$$

where $\hat{K}_1 = E(\int_{|t|\le\delta_0} \hat{K}(t)\,dt \,|\, W)$.

Proof We can assume $\delta_0 \le 1$ because (3.24) is trivial otherwise. Using that f_z satisfies the Stein equation (2.2), and the nonnegativity of $\hat{K}(t)$, we have

$$\begin{aligned}
&E\int_{|t|\le\delta_0} f_z'(W+t)\hat{K}(t)\,dt\\
&\quad= E\int_{|t|\le\delta_0}\big(\mathbf{1}_{\{W+t\le z\}}-\Phi(z)\big)\hat{K}(t)\,dt + E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt\\
&\quad\le E\int_{|t|\le\delta_0}\big(\mathbf{1}_{\{W\le z+\delta_0\}}-\Phi(z)\big)\hat{K}(t)\,dt + E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt\\
&\quad= E\big(\mathbf{1}_{\{W\le z+\delta_0\}}-\Phi(z)\big)\hat{K}_1 + E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt,
\end{aligned}$$

where the inequality holds because $-t\le\delta_0$.

Now, writing $\hat{K}_1 = 1-(1-\hat{K}_1)$, we find that

$$\begin{aligned}
&E\int_{|t|\le\delta_0} f_z'(W+t)\hat{K}(t)\,dt\\
&\quad\le P(W\le z+\delta_0)-\Phi(z)+E|1-\hat{K}_1| + E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt\\
&\quad\le P(W\le z+\delta_0)-\Phi(z+\delta_0)+\frac{\delta_0}{\sqrt{2\pi}}\\
&\qquad + E|1-\hat{K}_1| + E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt.
\end{aligned}$$

Thus, rearranging and using (3.23) to obtain the first equality,

$$\begin{aligned}
&P(W\le z+\delta_0)-\Phi(z+\delta_0)\\
&\quad\ge -\frac{\delta_0}{\sqrt{2\pi}} - E|1-\hat{K}_1| + E\int_{|t|\le\delta_0} f_z'(W+t)\hat{K}(t)\,dt\\
&\qquad - E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt\\
&\quad= -\frac{\delta_0}{\sqrt{2\pi}} - E|1-\hat{K}_1| + EWf_z(W) - ER_1\\
&\qquad - E\int_{|t|\le\delta_0}(W+t)f_z(W+t)\hat{K}(t)\,dt\\
&\quad= -\frac{\delta_0}{\sqrt{2\pi}} - E|1-\hat{K}_1| + E\big[Wf_z(W)(1-\hat{K}_1)\big] - ER_1\\
&\qquad + E\int_{|t|\le\delta_0}\big\{Wf_z(W)-(W+t)f_z(W+t)\big\}\hat{K}(t)\,dt\\
&\quad\ge -\frac{\delta_0}{\sqrt{2\pi}} - 2E|1-\hat{K}_1| - \delta_1 - \int_{|t|\le\delta_0} E\big(|W|+\sqrt{2\pi}/4\big)|t|\hat{K}(t)\,dt,
\end{aligned}$$

this last by (2.7), the hypotheses $|ER_1|\le\delta_1$, and (2.10). Hence,

$$\begin{aligned}
&P(W\le z+\delta_0)-\Phi(z+\delta_0)\\
&\quad\ge -\frac{\delta_0}{\sqrt{2\pi}} - 2E|1-\hat{K}_1| - \delta_1 - E\int_{|t|\le\delta_0}\big(|W|+0.7\big)\delta_0\hat{K}(t)\,dt
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{\delta_0}{\sqrt{2\pi}} - 2E|1-\hat{K}_1| - \delta_1 - \delta_0 E\big(|W|+0.7\big)\hat{K}_1 \\
&\geq -\frac{\delta_0}{\sqrt{2\pi}} - 2E|1-\hat{K}_1| - \delta_1 - \delta_0\big\{E\big[|W|\hat{K}_1\big] + 0.7 + 0.7E|1-\hat{K}_1|\big\} \\
&\geq -\delta_0\big(1.1 + E\big[|W|\hat{K}_1\big]\big) - 2.7E|1-\hat{K}_1| - \delta_1, \qquad (3.25)
\end{aligned}
$$

recalling that $\delta_0 \leq 1$. A similar argument gives

$$
\begin{aligned}
&P(W \leq z - \delta_0) - \Phi(z-\delta_0) \\
&\quad \leq \delta_0\big(1.1 + E\big[|W|\hat{K}_1\big]\big) + 2.7E|1-\hat{K}_1| + \delta_1, \qquad (3.26)
\end{aligned}
$$

completing the proof of (3.24). □

In Chap. 5 we illustrate how to use Theorem 3.5 to obtain Berry–Esseen bounds in various applications.

3.4 The Berry–Esseen Inequality for Unbounded Variables

Theorem 3.4 demonstrates the Berry–Esseen inequality when W is a sum of uniformly bounded, mean zero, independent random variables $\xi_1, \ldots, \xi_n$ with variances summing to one. Here we drop the boundedness restriction and prove, using two different methods, that there exists a universal constant C such that

$$
\sup_{z\in\mathbb{R}}\big|P(W \leq z) - \Phi(z)\big| \leq C\gamma \quad \text{where } \gamma = \sum_{i=1}^{n} E|\xi_i|^3. \qquad (3.27)
$$

Tyurin (2010) has shown that C can be taken 0.4785. Both of our two approaches, using concentration inequalities in Sect. 3.4.1, and an inductive method in Sect. 3.4.2, lead to somewhat larger constants, but as the sequel shows, these approaches generalize to many cases where the independence condition can be dropped.

3.4.1 *The Concentration Inequality Approach*

Noting that (3.17) in Theorem 3.4 holds without the uniform boundedness restriction, with $W^{(i)}$ as in (3.16) we see that one can prove the Berry–Esseen inequality more generally by showing that

$$
P\big(W^{(i)} + t \leq z\big) \quad \text{is close to} \quad P(W \leq z) = P\big(W^{(i)} + \xi_i \leq z\big),
$$

which it suffices to have a good bound for $P(a \leq W^{(i)} \leq b)$. Intuitively, the distribution of $W^{(i)}$ is close to the standard normal, and hence we should be able to bound $P(a \leq W^{(i)} \leq b)$ using some multiple of $b-a$. This heuristic is made precise by the concentration inequality

Lemma 3.1 *For all real $a < b$, and for every $1 \le i \le n$,*

$$P\big(a \le W^{(i)} \le b\big) \le \sqrt{2}(b-a) + 2(\sqrt{2}+1)\gamma \tag{3.28}$$

where γ is as in (3.27).

We remark that Chen (1998) was the first to apply the concentration inequality approach to independent but non-identically distributed variables. Postponing the proof of (3.28) to the end of this section, we demonstrate the following Berry–Esseen bound with a constant of 9.4.

Theorem 3.6 *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with zero means, satisfying $\sum_{i=1}^n \operatorname{Var}(\xi_i) = 1$. Then $W = \sum_{i=1}^n \xi_i$ satisfies*

$$\sup_{z\in\mathbb{R}} \big|P(W \le z) - \Phi(z)\big| \le 9.4\gamma \quad \textit{where } \gamma = \sum_{i=1}^n E|\xi_i|^3. \tag{3.29}$$

Proof With $W^{(i)}$ and $K_i(t)$ as in (3.16), by (2.25) and (3.28) we have

$$\begin{aligned}
&\left|\sum_{i=1}^n \int_{-\infty}^{\infty} P\big(W^{(i)} + t \le z\big) K_i(t)\,dt - P(W \le z)\right| \\
&\quad = \left|\sum_{i=1}^n \int_{-\infty}^{\infty} \big(P\big(W^{(i)} + t \le z\big) - P(W \le z)\big) K_i(t)\,dt\right| \\
&\quad \le \sum_{i=1}^n \int_{-\infty}^{\infty} \big|P\big(W^{(i)} + t \le z\big) - P(W \le z)\big| K_i(t)\,dt \\
&\quad = \sum_{i=1}^n \int_{-\infty}^{\infty} \big|P\big(W^{(i)} + t \le z\big) - P\big(W^{(i)} + \xi_i \le z\big)\big| K_i(t)\,dt \\
&\quad = \sum_{i=1}^n \int_{-\infty}^{\infty} E\big\{P\big(z - t \vee \xi_i \le W^{(i)} \le z - t \wedge \xi_i \,|\, \xi_i\big)\big\} K_i(t)\,dt \\
&\quad \le \sum_{i=1}^n \int_{-\infty}^{\infty} E\big\{\sqrt{2}\big(|t| + |\xi_i|\big) + 2(\sqrt{2}+1)\gamma\big\} K_i(t)\,dt \\
&\quad = \sqrt{2} \sum_{i=1}^n \left(\frac{1}{2} E|\xi_i|^3 + E|\xi_i| E\xi_i^2\right) + 2(\sqrt{2}+1)\gamma \\
&\quad \le (3.5\sqrt{2} + 2)\gamma \le 6.95\gamma,
\end{aligned} \tag{3.30}$$

where we have again applied (2.25). Invoking (3.17) now yields the claim. □

As in Theorem 3.2, one can dispense with the third moment assumption in Theorem 3.6 and replace γ in (3.29) by $\beta_2 + \beta_3$, defined in (3.5); we leave the details to the reader. Additionally, with a more refined concentration inequality, the constant can be reduced further, resulting in

$$\sup_{z\in\mathbb{R}}\left|P(W\le z)-\Phi(z)\right|\le 4.1(\beta_2+\beta_3), \tag{3.31}$$

see Chen and Shao (2001).

We now prove the concentration inequality (3.28). The idea is to use the fact that if f' equals the indicator $\mathbf{1}_{[a,b]}$ of some interval, then $Ef'(W)=P(a\le W\le b)$. This fixes f up to a constant, and choosing $f((a+b)/2)=0$ the norm $\|f\|=(b-a)/2$ takes on its minimal value, yielding the smallest factor in the right hand side of the inequality

$$\left|EWf(W)\right|\le\frac{1}{2}(b-a)E|W|\le\frac{1}{2}(b-a),$$

which holds whenever $EW^2\le 1$.

Proof of Lemma 3.1 Define $\delta=\gamma$ and take

$$f(w)=\begin{cases}-\frac{1}{2}(b-a)-\delta & \text{if } w<a-\delta,\\ w-\frac{1}{2}(b+a) & \text{if } a-\delta\le w\le b+\delta,\\ \frac{1}{2}(b-a)+\delta & \text{for } w>b+\delta,\end{cases} \tag{3.32}$$

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

$$\hat{K}_j(t)=\xi_j(\mathbf{1}_{\{-\xi_j\le t\le 0\}}-\mathbf{1}_{\{0<t\le-\xi_j\}})\quad\text{and}\quad\hat{K}(t)=\sum_{j=1}^n\hat{K}_j(t). \tag{3.33}$$

Since ξ_j and $W^{(i)}-\xi_j$ are independent for $j\ne i$, ξ_i is independent of $W^{(i)}$, and $E\xi_j=0$ for all j, similarly to (2.27), we have

$$\begin{aligned}&EW^{(i)}f\big(W^{(i)}\big)-E\xi_i f\big(W^{(i)}-\xi_i\big)\\&=\sum_{j=1}^n E\xi_j\big[f\big(W^{(i)}\big)-f\big(W^{(i)}-\xi_j\big)\big]\\&=\sum_{j=1}^n E\xi_j\int_{-\xi_j}^0 f'\big(W^{(i)}+t\big)\,dt\\&=\sum_{j=1}^n E\int_{-\infty}^{\infty}f'\big(W^{(i)}+t\big)\hat{K}_j(t)\,dt\\&=E\int_{-\infty}^{\infty}f'\big(W^{(i)}+t\big)\hat{K}(t)\,dt.\end{aligned} \tag{3.34}$$

Noting that $f'(t)\ge 0$ and $\hat{K}(t)\ge 0$, we have by the definition of f

$$E\int_{-\infty}^{\infty} f'\big(W^{(i)}+t\big)\hat{K}(t)\,dt \geq E\int_{|t|\leq\delta} f'\big(W^{(i)}+t\big)\hat{K}(t)\,dt$$
$$\geq E\mathbf{1}_{\{a\leq W^{(i)}\leq b\}}\int_{|t|\leq\delta}\hat{K}(t)\,dt.$$

Letting $K(t)=E\hat{K}(t)$, we may write this last expression as

$$E\mathbf{1}_{\{a\leq W^{(i)}\leq b\}}\int_{|t|\leq\delta}\big[\hat{K}(t)-K(t)\big]\,dt + P\big(a\leq W^{(i)}\leq b\big)\int_{|t|\leq\delta}K(t)\,dt. \tag{3.35}$$

As in (2.28) and (2.25), respectively, the function $K(t)$ is a density and $E|T|=\gamma/2$ for T so distributed. Hence, for the integral in the second term of (3.35), recalling $\delta=\gamma$,

$$\int_{|t|\leq\delta}K(t)\,dt = P\big(|T|\leq\delta\big) = 1-P\big(|T|>\delta\big)\geq 1-\frac{\gamma}{2\delta}=1/2.$$

For the first term of (3.35), applying the Cauchy–Schwarz inequality and integrating yields the bound

$$\begin{aligned}\mathrm{Var}\bigg(\int_{|t|\leq\delta}\hat{K}(t)\,dt\bigg)^{1/2} &\leq \Bigg(\mathrm{Var}\Bigg\{\sum_{j=1}^{n}|\xi_j|\min\big(\delta,|\xi_j|\big)\Bigg\}\Bigg)^{1/2}\\ &\leq \Bigg(\sum_{j=1}^{n}E\xi_j^2\min\big(\delta,|\xi_j|\big)^2\Bigg)^{1/2}\\ &\leq \delta\Bigg(\sum_{j=1}^{n}E\xi_j^2\Bigg)^{1/2}=\delta.\end{aligned}$$

Hence, from (3.34) and (3.35) we obtain

$$EW^{(i)}f\big(W^{(i)}\big)-E\xi_i f\big(W^{(i)}-\xi_i\big)\geq\frac{1}{2}P\big(a\leq W^{(i)}\leq b\big)-\delta. \tag{3.36}$$

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

$$\begin{aligned}&EW^{(i)}f\big(W^{(i)}\big)-E\xi_i f\big(W^{(i)}-\xi_i\big)\\ &\quad\leq\bigg(\frac{1}{2}(b-a)+\delta\bigg)\big(E\big|W^{(i)}\big|+E|\xi_i|\big)\\ &\quad\leq\frac{1}{\sqrt{2}}\Big(\big(E\big|W^{(i)}\big|\big)^2+\big(E|\xi_i|\big)^2\Big)^{1/2}(b-a+2\delta)\\ &\quad\leq\frac{1}{\sqrt{2}}\big(E\big|W^{(i)}\big|^2+E|\xi_i|^2\big)^{1/2}(b-a+2\delta)\\ &\quad=\frac{1}{\sqrt{2}}(b-a+2\delta).\end{aligned} \tag{3.37}$$

Combining (3.36) and (3.37) thus gives

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

as desired. □

By reasoning as above, and as in the proofs of Theorem 8.1 and Propositions 10.1 and 10.2, one can prove the following stronger concentration inequality.

Proposition 3.1 *If W is the sum of the independent mean zero random variables $\xi_1, \ldots, \xi_n$, then for all real $a < b$*

$$P(a \le W \le b) \le b - a + 2(\beta_2 + \beta_3) \tag{3.38}$$

where β_2 and β_3 are defined in (3.5). In addition, if $W^{(i)} = W - \xi_i$, then

$$P\big(a \le W^{(i)} \le b\big) \le \sqrt{2}(b-a) + (\sqrt{2}+1)(\beta_2 + \beta_3) \tag{3.39}$$

for every $1 \le i \le n$.

We leave the proof to the reader. Clearly, $\beta_2 + \beta_3 \le \gamma$, so Proposition 3.1 not only relaxes the moment assumption required by (3.28) but improves the constant as well.

3.4.2 An Inductive Approach

In this section we prove the following Berry–Esseen inequality by induction.

Theorem 3.7 *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with zero means, satisfying $\sum_{i=1}^n \mathrm{Var}(\xi_i) = 1$. Then $W = \sum_{i=1}^n \xi_i$ satisfies*

$$\sup_{z \in \mathbb{R}} \big|P(W \le z) - \Phi(z)\big| \le 10\gamma \quad \text{where } \gamma = \sum_{i=1}^n E|\xi_i|^3. \tag{3.40}$$

Though the constant produced is not optimal, the inductive approach is quite useful in more general settings when the removal of some variables leaves a structure similar to the original one; see Theorem 6.2 in Sect. 6.1.1 for one example involving dependence where the inductive method succeeds, and references to other such examples. Use of induction in the independent case appears in the text of Stroock (2000).

Proof Without loss of generality we may assume $E\xi_i^2 \ne 0$ for all $i = 1, \ldots, n$. Let

$$\tau_i^2 = E\big(W^{(i)}\big)^2 \quad \text{and} \quad \tau = \min_{1 \le i \le n} \tau_i.$$

Since (3.40) is trivial if $\gamma \ge 1/10$, we can assume $\gamma < 1/10$. Since

$$1 = EW^2 = E\big(\big(W^{(i)}\big)^2 + \xi_i\big)^2 = E\big(W^{(i)}\big)^2 + E\xi_i^2 \le E\big(W^{(i)}\big)^2 + \big(E|\xi_i|^3\big)^{2/3},$$

we have

$$\tau^2 \geq 1 - \gamma^{2/3} \geq 0.7845. \tag{3.41}$$

When $n = 1$, since $\gamma = E|\xi_1|^3 \geq (E\xi_1^2)^{3/2} = 1$, inequality (3.40) is trivially true. Now take $n \geq 2$ and assume that (3.40) has been established for a sum composed of fewer than n summands. Then for all $i = 1, \ldots, n$ and $a < b$, with $C = 10$ we have

$$\begin{aligned}
&P\big(a < W^{(i)} \leq b\big) \\
&\quad = \Phi(b/\tau_i) - \Phi(a/\tau_i) + P\big(W^{(i)} \leq b\big) - \Phi(b/\tau_i) \\
&\qquad - \big\{P\big(W^{(i)} \leq a\big) - \Phi(a/\tau_i)\big\} \\
&\quad \leq \frac{2C}{\tau_i^3} \sum_{j \neq i} E|\xi_j|^3 + \frac{b-a}{\sqrt{2\pi}\,\tau_i} \\
&\quad \leq 2.88C\gamma + (b-a)/2,
\end{aligned} \tag{3.42}$$

using (3.41) twice in the final inequality.

Let ξ_i^* have the ξ_i-zero bias distribution and be independent of ξ_j, $j \neq i$, and let I be a random index, independent of all other variables, with distribution (2.60). Then, by Lemma 2.8, letting $\delta = 2\gamma$, we have

$$\begin{aligned}
&P(W^* \leq z) - P(W \leq z - 2\delta) \\
&\quad = P\big(W^{(I)} + \xi_I^* \leq z\big) - P\big(W^{(I)} + \xi_I \leq z - 2\delta\big) \\
&\quad \geq -EP\big(z - \xi_I^* \leq W^{(I)} \leq z - \xi_I - 2\delta | \xi_I, \xi_I^*\big)\mathbf{1}\big(\xi_I^* \geq \xi_I + 2\delta\big) \\
&\quad \geq -E\big(2.88C\gamma + \big(\xi_I^* - \xi_I\big)/2 - \delta\big)\mathbf{1}\big(\xi_I^* \geq \xi_I + 2\delta\big) \\
&\quad \geq -2.88C\gamma P\big(\xi_I^* - \xi_I \geq 2\delta\big) - E\big(\xi_I^* - \xi_I\big)\mathbf{1}\big(\xi_I^* \geq \xi_I + 2\delta\big)/2 - \delta,
\end{aligned}$$

where we have invoked (3.42) to obtain the second inequality. By Theorem 4.3, ξ_i and ξ_i^* may be coupled so that

$$E\big|\xi_i^* - \xi_i\big| \leq \frac{E|\xi_i|^3}{2E\xi_i^2} \quad \text{so, by (2.60),} \quad E\big|\xi_I^* - \xi_I\big| \leq \gamma/2.$$

But now

$$P\big(\xi_I^* - \xi_I \geq 2\delta\big) \leq \gamma/(4\delta) \quad \text{and} \quad E\big(\xi_I^* - \xi_I\big)\mathbf{1}\big(\xi_I^* \geq \xi_I + 2\delta\big) \leq \gamma/2.$$

Hence, recalling $\delta = 2\gamma$,

$$P(W^* \leq z) - P(W \leq z - 2\delta) \geq -2.88C\gamma/8 - \gamma/4 - 2\gamma = -5.85\gamma.$$

Thus, by (3.19),

$$\begin{aligned}
P(W \leq z - 2\delta) - \Phi(z - 2\delta) &\leq P(W^* \leq z) - \Phi(z - 2\delta) + 5.85\gamma \\
&\leq 2.44\gamma + \frac{4\gamma}{\sqrt{2\pi}} + 5.85\gamma < 10\gamma.
\end{aligned}$$

Similarly, we may obtain

$$P(W \leq z + 2\delta) - \Phi(z + 2\delta) \geq -10\gamma,$$

thus completing the proof. □

3.5 A Lower Berry–Esseen Bound

Again, let $\xi_1, \ldots, \xi_n$ be independent random variables with zero means satisfying $\sum_{i=1}^n \mathrm{Var}(\xi_i) = 1$. Feller (1935) and Lévy (1935) proved independently (see LeCam 1986) that if the Feller–Lévy condition

$$\max_{1 \le i \le n} E\xi_i^2 \to 0, \tag{3.43}$$

is satisfied, then the Lindeberg condition (3.13) is necessary for the central limit theorem. The theorem below is due to Hall and Barbour (1984) who used Stein's method to provide not only a nice proof of the necessity, but also a lower bound for the L^∞ distance between the distribution of W and the normal.

Theorem 3.8 *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with zero means and finite variances $E\xi_i^2 = \sigma_i^2$, $1 \le i \le n$, satisfying $\sum_{i=1}^n \sigma_i^2 = 1$, and let $W = \sum_{i=1}^n \xi_i$. Then there exists an absolute constant C such that for all $\varepsilon > 0$,*

$$\begin{aligned}
&\bigl(1 - e^{-\varepsilon^2/4}\bigr) \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i| > \varepsilon\}} \\
&\quad \le C\left(\sup_{z \in \mathbb{R}} \bigl|P(W \le z) - \Phi(z)\bigr| + \sum_{i=1}^n \sigma_i^4\right).
\end{aligned} \tag{3.44}$$

Clearly, the Feller–Lévy condition (3.43) implies that $\sum_{i=1}^n \sigma_i^4 \le \max_{1 \le i \le n} \sigma_i^2 \to 0$ as $n \to \infty$. Therefore, if W is asymptotically normal,

$$\sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i| > \varepsilon\}} \to 0$$

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

Proof Once again, the argument starts with the Stein equation

$$E\bigl\{f_h'(W) - W f_h(W)\bigr\} = Eh(W) - Nh, \tag{3.45}$$

for a function h yet to be chosen. Taking h absolutely continuous with $\int_{-\infty}^{\infty} |h'(w)| dw < \infty$, we may integrate by parts and obtain the bound

$$\begin{aligned}
\bigl|Eh(W) - Nh\bigr| &= \left|\int_{-\infty}^{\infty} h'(w)\bigl\{P(W \le w) - \Phi(w)\bigr\} dw\right| \\
&\le \delta \int_{-\infty}^{\infty} \bigl|h'(w)\bigr| dw,
\end{aligned} \tag{3.46}$$

where $\delta = \sup_{z \in \mathbb{R}} |P(W \le z) - \Phi(z)|$.

For the left hand side of (3.45), in the usual way, because ξ_i and $W^{(i)} = W - \xi_i$ are independent, and $E\xi_i = 0$, we have

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

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

$$
Ef_h'(W) = \sum_{i=1}^n \sigma_i^2 Ef_h'\big(W^{(i)}\big) + \sum_{i=1}^n \sigma_i^2 E\big\{f_h'(W) - f_h'\big(W^{(i)}\big)\big\},
$$

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

$$
\left| E\big\{f_h'(W) - Wf_h(W)\big\} - \sum_{i=1}^n E\xi_i^2 g\big(W^{(i)}, \xi_i\big)\right| \le \frac{1}{2}\|f_h'''\|\sum_{i=1}^n \sigma_i^4, \tag{3.47}
$$

where

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

Intuitively, if the distribution of W is close to that of the standard normal Z, taken to be independent of the ξ_i's, then

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

should be close to one another.

Taking (3.46) and (3.47) together, we will be able to compute a lower bound for δ, if we can produce an absolutely continuous function h satisfying $\int_{-\infty}^{\infty}|h'(w)|\,dw < \infty$ for which $Eg_h(Z,y)$ is of constant sign, provided also that $\|f_h'''\| < \infty$. In practice, it is easier to look for a suitable f, and then define $h(w) = f'(w) - wf(w)$. The function g is zero for any linear function f, and when f is an even function then $Eg(Z,y)$ is odd. Choosing f to be the odd function $f(y) = y^3$ yields $Eg(Z,y) = -y^2$, of constant sign. Unfortunately, this f fails to yield an h satisfying $\int_{-\infty}^{\infty}|h'(w)|\,dw < \infty$.

A good choice is $f(w) = we^{-w^2/2}$, which behaves much like the sum of a linear and a cubic function for those values of w where Z puts most of its mass, yet decays to zero quickly when $|w|$ is large. Making the computations, we have

$$
\begin{aligned}
Eg(Z,y) &= -\frac{y^{-1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\big\{(w+y)e^{-(w+y)^2/2} \\
&\qquad - we^{-w^2/2} - ye^{-w^2/2}\big(1-w^2\big)\big\}e^{-w^2/2}\,dw \\
&= \frac{1}{2\sqrt{2}}\big(1 - e^{-y^2/4}\big),
\end{aligned} \tag{3.48}
$$

a nonnegative function which satisfies

$$Eg(Z, y) \geq \frac{1}{2\sqrt{2}}\bigl(1 - e^{-\varepsilon^2/4}\bigr) \quad \text{whenever } |y| \geq \varepsilon$$

for all $\varepsilon > 0$. Hence, for this choice of f we have

$$R \geq \frac{1}{2\sqrt{2}}\bigl(1 - e^{-\varepsilon^2/4}\bigr) \sum_{i=1}^{n} E\xi_i^2 \mathbf{1}_{\{|\xi_i| > \varepsilon\}}. \tag{3.49}$$

It thus remains to show that R and R_1 are close enough, after which (3.1), (3.47) and (3.49) complete the proof.

For this step, note that for $f(w) = we^{-w^2/2}$ and $h(w) = f'(w) - wf(w)$ we have

$$c_1 := \int_{-\infty}^{\infty} \bigl|h'(w)\bigr|\,dw \leq 7,$$
$$c_2 := \int_{-\infty}^{\infty} \bigl|f''(w)\bigr|\,dw \leq 4; \quad \text{and} \quad c_3 := \sup_w \bigl|f'''(w)\bigr| = 3.$$

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

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

where W' has the same distribution as W, but is independent of the ξ_i's. Then

$$\begin{aligned}
R_1 &= -\sum_{i=1}^{n} E\biggl\{\xi_i^2 \int_0^1 \bigl[f'\bigl(W^{(i)} + t\xi_i\bigr) - f'\bigl(W^{(i)}\bigr)\bigr]\,dt\biggr\} \\
&= R_2 + \sum_{i=1}^{n} E\biggl\{\xi_i^2 \int_0^1 \bigl[f'\bigl(W' + t\xi_i\bigr) - f'\bigl(W^{(i)} + t\xi_i\bigr)\bigr]\,dt\biggr\} \\
&\quad - \sum_{i=1}^{n} E\biggl\{\xi_i^2 \int_0^1 \bigl[f'\bigl(W'\bigr) - f'\bigl(W^{(i)}\bigr)\bigr]\,dt\biggr\}.
\end{aligned} \tag{3.50}$$

Now, for any θ, using that W and W' have the same distribution, that ξ_i and $W^{(i)}$ are independent, and that $E\xi_i = 0$,

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

by Taylor's theorem. Hence, from (3.50),

$$R_1 \geq R_2 - c_3 \sum_{i=1}^{n} \sigma_i^4. \tag{3.51}$$

Similarly,

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

and, for any θ, as $\int_{-\infty}^{\infty} |f''(w)| dw = c_2 < \infty$,

$$\left|Ef'(W' + \theta) - Ef'(Z + \theta)\right|$$
$$= \left|\int_{-\infty}^{\infty} f''(w)\big(P(W' \le w - \theta) - \Phi(w - \theta)\big)\, dw\right| \le c_2\delta,$$

so that

$$R_2 \ge R - 2c_2\delta. \tag{3.52}$$

Combining (3.46) and (3.47) with (3.51) and (3.52), it follows that

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

In view of (3.49), collecting terms, it follows that

$$\delta(c_1 + 2c_2) + \frac{3}{2}c_3 \sum_{i=1}^n \sigma_i^4 \ge \frac{1}{2\sqrt{2}}\left(1 - e^{-\varepsilon^2/4}\right) \sum_{i=1}^n E\xi_i^2 \mathbf{1}_{\{|\xi_i| > \varepsilon\}} \tag{3.53}$$

for any $\varepsilon > 0$. This proves (3.44), with $C \le 43$. □