# Stein's Method and Self-normalized Limit Theory

Qi-Man Shao

*Hong Kong University of Science and Technology* and *University of Oregon*

E-mail: *maqmshao@ust.hk*

**Abstract.** This lecture series consists of two parts. The first part is on Stein's method. In contrast to the traditional method of characteristic function, Stein in 1972 introduced a totally new method to determine the accuracy of the normal approximation. The method works well not only for independent random variables but also for dependent variables. Stein's ideas have been extended beyond normal approximation and applied to problems in other areas. In these lectures, we focus on univariate normal approximation with emphasis on the main ideas behind the method. The second part is about the self-normalized limit theorems. The normalizing constants in classical limit theorems are usually sequences of real numbers. Moment conditions or other related assumptions are necessary and sufficient for many classical limit theorems. However, the situation becomes very different when the normalizing constants are sequences of random variables. A self-normalized large deviation holds without any moment conditions. A self-normalized law of the iterated logarithm remains valid for all distributions in the domain of attraction of a normal or stable law. This reveals that the self-normalization preserves much better properties than deterministic normalization does. We shall briefly review recent development on the self-normalized limit theorems and show how Stein's method can be used to recover some of the results.

2005 @ All rights reserved

**Part I. Stein's Method** (based on Chen and Shao (2005))

# 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 the first part of this lecture series, 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. The second part will focus on the self-normalized limit theorems. In contrast to the classical limit theorems, the self-normalized limit theorems remain valid under much less moment conditions than those in the classical limit theorems. We shall briefly review recent developments in this area and especially, use Stein's method to recover the Cramér type self-normalized moderate deviations.

# 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 differentiable 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,

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

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

$$f'(w) - wf(w) = h(w) - Eh(Z). \tag{2.4}$$

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

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

## 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*

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

$$|wf_z(w)| \le 1, \quad |wf_z(w) - uf_z(u)| \le 1 \tag{2.7}$$

$$|f_z'(w)| \le 1, \;\; |f_z'(w) - f_z'(v)| \le 1 \tag{2.8}$$

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

*and*

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

*for all real* $w$, $u$, *and* $v$.

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

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

$$\sup_w |f_h(w)| \le \min\left(\sqrt{\pi/2}\sup_w |h(w) - Eh(Z)|, 2\sup_w |h'(w)|\right), \tag{2.11}$$

$$\sup_w |f_h'(w)| \le \min\left(2\sup_w |h(w) - Eh(Z)|, 4\sup_w |h'(w)|\right) \tag{2.12}$$

*and*

$$\sup_w |f_h''(w)| \le 2\sup_w |h'(w)|. \tag{2.13}$$

Proofs of Lemmas 2.2 and 2.3 are given in the Appendix.

## 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

$$W = \sum_{i=1}^n \xi_i, \;\; W^{(i)} = W - \xi_i \tag{2.14}$$

and

$$K_i(t) = E\xi_i(I_{\{0\le t\le \xi_i\}} - I_{\{\xi_i \le t < 0\}}). \tag{2.15}$$

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

$$\int_{-\infty}^{\infty} K_i(t)dt = E\xi_i^2 \text{ and } \int_{-\infty}^{\infty} |t|K_i(t)dt = E|\xi_i|^3/2. \tag{2.16}$$

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

$$\begin{aligned} EWf(W) &= \sum_{i=1}^n E\xi_i f(W) \\ &= \sum_{i=1}^n E\xi_i [f(W) - f(W^{(i)})] \\ &= \sum_{i=1}^n E\xi_i \int_0^{\xi_i} f'(W^{(i)} + t)dt \\ &= \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W^{(i)} + t)\xi_i (I_{\{0\le t\le \xi_i\}} - I_{\{\xi_i \le t < 0\}})dt \\ &= \sum_{i=1}^n E\int_{-\infty}^{\infty} f'(W^{(i)} + t)K_i(t)dt. \end{aligned} \tag{2.17}$$

From

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

it follows that

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

Thus, by (2.17) and (2.18)

$$Ef'(W) - EWf(W) = \sum_{i=1}^n E\int_{-\infty}^{\infty} [f'(W) - f'(W^{(i)} + t)]K_i(t)dt. \tag{2.19}$$

Equations (2.17) and (2.19) play a key role in proving a Berry-Esseen type inequality. We remark that it holds for all bounded absolute continuous $f$. Let

$$\gamma = \sum_{i=1}^n E|\xi_i|^3. \tag{2.20}$$

## 2.4 Expectation of smooth functions

Equation (2.19) 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* $\sup_x |h'(x)| \le c_1$

$$|Eh(W) - Eh(Z)| \le 3c_1 \sum_{i=1}^n E|\xi_i|^3. \tag{2.21}$$

*In particular, we have*

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

**Proof.** It follows from (2.13) that $|f_h''| \le 2c_1$. Therefore, by (2.19) and the mean value theorem

$$\begin{aligned} |Ef_h'(W) - EWf_h(W)| &\le \sum_{i=1}^{n} E \int_{-\infty}^{\infty} |f_h'(W) - f_h'(W^{(i)} + t)| K_i(t) dt \\ &\le 2c_1 \sum_{i=1}^{n} E \int_{-\infty}^{\infty} (|t| + |\xi_i|) K_i(t) dt \\ &= 2c_1 \sum_{i=1}^{n} (E|\xi_i|^3/2 + E|\xi_i| E\xi_i^2) \\ &\le 3c_1 \sum_{i=1}^{n} E|\xi_i|^3. \end{aligned} \tag{2.22}$$

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*

$$|Eh(W) - Eh(Z)| \le 4c_1(4\beta_2 + 3\beta_3), \tag{2.23}$$

*where*

$$\beta_2 = \sum_{i=1}^{n} E\xi_i^2 I_{\{|\xi_i|>1\}} \text{ and } \beta_3 = \sum_{i=1}^{n} E|\xi_i|^3 I_{\{|\xi_i|\le 1\}}. \tag{2.24}$$

**Proof.** Observing from (2.12) and (2.13) 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 (2.19)

$$\begin{aligned} &|Eh(W) - Eh(Z)| \\ &\le 8c_1 \sum_{i=1}^{n} E \int_{-\infty}^{\infty} (|t| \wedge 1 + |\xi_i| \wedge 1) K_i(t) dt \\ &= 8c_1 \sum_{i=1}^{n} \Big( E|\xi_i|(|\xi_i| - 1) I_{\{|\xi_i|>1\}} + \frac{1}{2} E|\xi_i|(|\xi_i| \wedge 1)^2 + E\xi_i^2 E(|\xi_i| \wedge 1) \Big) \\ &= 8c_1 \sum_{i=1}^{n} \Big( E\xi_i^2 I_{\{|\xi_i|>1\}} - \frac{1}{2} E|\xi_i| I_{\{|\xi_i|>1\}} + \frac{1}{2} E|\xi_i|^3 I_{\{|\xi_i|\le 1\}} + E\xi_i^2 E(|\xi_i| \wedge 1) \Big) \\ &\le 8c_1 \Big\{ \beta_2 + \frac{1}{2}\beta_3 + \sum_{i=1}^{n} E\xi_i^2 E(|\xi_i| \wedge 1) \Big\}. \end{aligned} \tag{2.25}$$

We need the following fact: for any random variable $\xi$

$$E\xi^2 E(|\xi| \wedge 1) \leq E|\xi|^3 I_{\{|\xi|\leq 1\}} + E\xi^2 I_{\{|\xi|>1\}}. \tag{2.26}$$

To see this, let $\eta$ be an independent copy of $\xi$. It is easy to verify that

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

Taking expectation on both sides yields (2.26).

Now (2.23) follows from (2.25) - (2.26). □

**Remark 2.1** *If $h$ is bounded by $c_0$, then the bound in* (2.23) *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 $1 \leq i \leq n$ and $\sum_{i=1}^n E\xi_i^2 = 1$. Then*

$$\sup_z |P(W \leq z) - \Phi(z)| \leq 2.2(4\beta_2 + 3\beta_3)^{1/2}, \tag{2.27}$$

*where $\beta_2$ and $\beta_3$ are defined in* (2.24).

**Proof.** We can assume that $(4\beta_2 + 3\beta_3)^{1/2} \leq 1/2$. Otherwise, (2.27) is trivial. Let $\alpha = 0.5(4\beta_2 + 3\beta_3)^{1/2}$, and define for fixed $z$

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

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

$$|Eh_\alpha(W) - Eh_\alpha(Z)| \leq \max(4, 1/\alpha)(4\beta_2 + 3\beta_3) \leq (4\beta_2 + 3\beta_3)/\alpha \tag{2.28}$$

and hence

$$\begin{aligned} P(W \le z) - \Phi(z) &\le Eh_\alpha(W) - Eh_\alpha(Z) + Eh_\alpha(Z) - \Phi(Z) \\ &\le (4\beta_2 + 3\beta_3)/\alpha + EI_{\{z \le Z \le z+\alpha\}} \\ &\le (4\beta_2 + 3\beta_3)/\alpha + \frac{\alpha}{\sqrt{2\pi}} \le 2.2(4\beta_2 + 3\beta_3)^{1/2}. \end{aligned} \tag{2.29}$$

Similarly, we have

$$P(W \le z) - \Phi(z) \ge -2.2(4\beta_2 + 3\beta_3)^{1/2}. \tag{2.30}$$

This proves (2.27), by (2.29) and (2.30). □

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

$$\xi_i = X_i/B_n \text{ and } W = S_n/B_n. \tag{2.31}$$

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

$$\begin{aligned} \beta_2 + \beta_3 &= \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i| > B_n\}} + \frac{1}{B_n^3}\sum_{i=1}^n E|X_i|^3 I_{\{|X_i| \le B_n\}} \\ &\le \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i| > B_n\}} + \frac{1}{B_n^3}\sum_{i=1}^n B_n EX_i^2 I_{\{\varepsilon B_n \le |X_i| \le B_n\}} \\ &\quad + \frac{1}{B_n^3}\sum_{i=1}^n \varepsilon B_n EX_i^2 I_{\{|X_i| < \varepsilon B_n\}} \\ &\le \varepsilon + \frac{1}{B_n^2}\sum_{i=1}^n EX_i^2 I_{\{|X_i| > \varepsilon B_n\}}. \end{aligned} \tag{2.32}$$

If Lindeberg's condition (1.1) is satisfied, then (2.32) implies $\beta_2 + \beta_3 \to 0$ as $n \to \infty$ since $\varepsilon$ 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

$$\max_{1\le i\le n} EX_i^2/B_n^2 \to 0, \tag{2.33}$$

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 $\varepsilon > 0$*

$$(1 - e^{-\varepsilon^2/4}) \sum_{i=1}^n E\xi_i^2 I_{\{|\xi_i| > \varepsilon\}} \le C\Big(\sup_z |P(W \le z) - \Phi(z)| + \sum_{i=1}^n (E\xi_i^2)^2\Big). \tag{2.34}$$

Now for the sequence of independent random variables $\{X_i, i \ge 1\}$, recall $\xi_i = X_i/B_n$. Clearly, Feller's condition (2.33) implies that $\sum_{i=1}^n (E\xi_i^2)^2 \le \max_{1\le i\le n} E\xi_i^2 \to 0$ as $n \to \infty$. Therefore, if $S_n/B_n$ is asymptotically normal, then

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

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

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

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

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

$$|Eh(W)| = |Eh(W) - Eh(Z)| = |\int_{-\infty}^{\infty} h'(w)\{P(W \le w) - \Phi(w)\} dw| \le c_1 \Delta. \tag{2.35}$$

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

$$\begin{aligned} Eh(W) &= \sum_{i=1}^n E(\sigma_i^2 f'(W^{(i)} + \xi_i) - \xi_i f(W^{(i)} + \xi_i)) \\ &= \sum_{i=1}^n \sigma_i^2 E(f'(W^{(i)} + \xi_i) - f'(W^{(i)}) - \xi_i f''(W^{(i)})) \\ &\quad - \sum_{i=1}^n E(\xi_i \{f(W^{(i)} + \xi_i) - f(W^{(i)}) - \xi_i f'(W^{(i)})\}) \\ &\ge -0.5 c_2 \sum_{i=1}^n \sigma_i^4 + R_1, \end{aligned} \tag{2.36}$$

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 $\{\xi_i, 1 \le i \le n\}$ 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

$$\begin{aligned} R_1 &= \sum_{i=1}^{n} E\Big\{\xi_i^2 \int_0^1 [f'(W^{(i)}+t\xi_i) - f'(W^{(i)})]dt\Big\} \\ &= R_2 + \sum_{i=1}^{n} E\Big\{\xi_i^2 \int_0^1 [f'(W^*+t\xi_i) - f'(W^{(i)}+t\xi_i)]dt\Big\} \\ &\quad -\sum_{i=1}^{n} E\Big\{\xi_i^2 \int_0^1 [f'(W^*) - f'(W^{(i)})]dt\Big\}. \end{aligned} \tag{2.37}$$

For any constant $\theta$,

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

So by (2.37),

$$R_1 \ge R_2 - c_2 \sum_{i=1}^{n} \sigma_i^4. \tag{2.38}$$

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 $\theta$

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

Combining the above estimates with (2.35) - (2.38) yields

$$R_3 \le (c_1 + 2c_3)\Delta + 1.5c_2 \sum_{i=1}^n \sigma_i^4. \tag{2.39}$$

Observing that

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

we have

$$\begin{aligned} R_3 &= \sum_{i=1}^n E\xi_i^2 g(\xi_i) \\ &\ge 2^{-1.5}(1 - e^{-\varepsilon^2/4}) \sum_{i=1}^n E\xi_i^2 I_{\{|\xi_i| > \varepsilon\}} \end{aligned} \tag{2.40}$$

for every $\varepsilon > 0$. This proves (2.34). $\square$

# 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,

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

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

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

## 3.1 Bounded random variables

For bounded $\xi_i$, we are ready to apply (2.17) to obtain the following Berry-Esseen type bound.

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

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

**Proof.** Write $f = f_z$. It follows from (2.17) that

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

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),

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

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

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

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). □

One can see from the above proof that the boundness of $\xi_i$ 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 (11.2) 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 $n=1$, $E|\xi_1|^3\ge(E\xi_i^2)^{2/3}=1$. (11.2) is true. Suppose inductively that (11.2) 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

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

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

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 (11.2). □

### 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

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

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

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

where

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

A direct calculation yields

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

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

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

and hence

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

By the Hölder inequality,

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

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

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

$$
\begin{aligned}
&\leq \sum_{i=1}^{n} \int_{-\infty}^{\infty} E\{\sqrt{2}(|t| + |\xi_i|) + (1+\sqrt{2})\gamma\} K_i(t)dt \\
&= (1+\sqrt{2})\gamma + \sqrt{2}\sum_{i=1}^{n}(0.5E|\xi_i|^3 + E|\xi_i|E\xi_i^2) \\
&\leq (1+2.5\sqrt{2})\gamma. \qquad (3.17)
\end{aligned}
$$

Now by (3.2)

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

which is (3.1). □

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

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

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 $\Delta_1$ and $\Delta_2$ be real-valued Borel measurable functions of $(\xi_i, 1 \leq i \leq n)$.

**Theorem 3.3** *We have*

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

*where $\Delta_{1,i}$ and $\Delta_{2,i}$ are Borel measurable functions of $(\xi_j, 1 \leq j \leq n,\ j \neq i)$, and $\gamma$ is defined as in* (2.20).

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

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

Let

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

Since $\xi_i$ 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 $f_{\Delta_1,\Delta_2}$ that

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

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

$$P(\Delta_1 \le W \le \Delta_2)$$

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

□

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 $\Delta_i$ 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 $\Delta_l$ 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*

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

*for $x > 0$.*

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

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

for $s \le a$. We have

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

This proves (4.2).

To prove (4.3), let

$$t = \frac{1}{a}\ln\Big((1+\frac{ax}{B_n^2}\Big).$$

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). □

**Lemma 4.2** *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent* **non-negative** *random variables with $E\xi_j^2 < \infty$ for each $1 \leq j \leq n$. Put $S = \sum_{j=1}^n \xi_j$, $\mu = \sum_{j=1}^n E\xi_j$ and $\sigma^2 = \sum_{j=1}^n E\xi_j^2$. Then for any $0 < x \leq \mu$*

$$P(S \leq x) \leq \exp\left(-\frac{(\mu - x)^2}{2\sigma^2}\right). \tag{4.6}$$

**Proof.** Noting that

$$e^{-a} \leq 1 - a + a^2/2 \quad \text{for } a \geq 0,$$

We have for any $t \geq 0$ and $x \leq \mu$

$$\begin{aligned} P(S \leq x) &\leq e^{tx}Ee^{-tS} = e^{tx}\prod_{j=1}^n Ee^{-t\xi_j} \\ &\leq e^{tx}\prod_{j=1}^n E(1 - t\xi_j + t^2\xi_j^2/2) \\ &\leq \exp\left(-t(\mu - x) + t^2\sigma^2/2\right). \end{aligned}$$

Letting $t = (\mu - x)/\sigma^2$ yields (4.6). ∎

**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 $\gamma \leq 0.17$. Similarly to the proof of Proposition 3.1, define $\delta = \gamma/2(\leq 0.085)$ and

$$f(w) = \begin{cases} 0 & \text{if } w < a - \delta, \\ e^{w/2}(w - a + \delta) & \text{if } a - \delta \leq w \leq b + \delta, \\ e^{w/2}(b - a + 2\delta) & \text{if } w > b + \delta \end{cases} \tag{4.7}$$

Put

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

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

$$\begin{aligned} & EW^{(i)} f(\bar{W}^{(i)}) \\ = & \sum_{j \neq i} E\xi_j [f(\bar{W}^{(i)}) - f(W^{(i)} - \bar{\xi}_j)] \\ = & \sum_{j \neq i}^{n} E \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t)\bar{M}_i(t)dt \\ = & E \int_{-\infty}^{\infty} f'(\bar{W}^{(i)} + t)\bar{M}^{(i)}(t)dt \\ \geq & EI_{\{a \leq \bar{W}^{(i)} \leq b\}} \int_{|t| \leq \delta} f'(\bar{W}^{(i)} + t)\bar{M}^{(i)}(t)dt \\ \geq & Ee^{(\bar{W}^{(i)} - \delta)/2} I_{\{a \leq \bar{W}^{(i)} \leq b\}} \int_{|t| \leq \delta} \bar{M}^{(i)}(t)dt \\ \geq & Ee^{(\bar{W}^{(i)} - \delta)/2} I_{\{a \leq \bar{W}^{(i)} \leq b\}} \sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|) \\ \geq & e^{-\delta/2}(H_{2,1} - H_{2,2}), \end{aligned} \tag{4.8}$$

where

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

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

$$\begin{aligned} \sum_{j \neq i} E|\xi_j| \min(\delta, |\bar{\xi}_j|) &= \sum_{j \neq i} E(|\xi_j|(\min(\delta, |\xi_j|) - \delta I_{\{\xi_j > 1\}})) \\ &\geq -\delta E|\xi_i| - \delta\gamma + \sum_{j=1}^{n} E|\xi_j| \min(\delta, |\xi_j|) \\ &\geq 0.5 - \delta\gamma^{1/3} - \delta\gamma \\ &\geq 0.5 - 0.085(0.17)^{1/3} - 0.085(0.017) \geq 0.43. \end{aligned} \tag{4.9}$$

Hence

$$H_{2,1} \geq .45 e^{a/2} P(a \leq \bar{W}^{(i)} \leq b) \tag{4.10}$$

By the Bennett inequality (4.2) again, we have

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

and hence

$$\begin{aligned} H_{2,2} &\leq (E e^{\bar{W}^{(i)}})^{1/2} \Big( \mathrm{Var}(\sum_{j \neq i} |\xi_j| \min(\delta, |\bar{\xi}_j|)) \Big)^{1/2} \\ &\leq \exp(.5e-1)\delta \leq 1.44\delta. \end{aligned} \tag{4.11}$$

As to the left hand side of (4.7), we have

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

Combining the above inequalities yields

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

This proves (4.1). $\square$

We also need the following moment inequality.

**Lemma 4.3** *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*

$$E|S_n|^p \leq (p-1)B_n^p + \sum_{i=1}^n E|\eta_i|^p \tag{4.12}$$

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

$$E|S_n|^p = \sum_{i=1}^n E\eta_i S_n |S_n|^{p-2}$$

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

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 $z$,*

$$|P(W \le z) - \Phi(z)| \le \frac{C\gamma}{1 + |z|^3}. \tag{4.13}$$

**Proof.** Without loss of generality, assume $z \ge 0$. By (4.12),

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

So (4.13) 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

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

we have

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

and similarly,

$$P(\bar{W} > z) \le P(W > z) + P(\bar{W} > z, \max_{1\le i\le n} \xi_i > 1). \tag{4.15}$$

Note that

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

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

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

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

$$|P(\bar{W} \le z) - \Phi(z)| \le C e^{-z/2}\gamma. \tag{4.16}$$

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 (2.17) 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

$$\begin{aligned} & P(\bar{W}\le z) - \Phi(z) \\ = \ & Ef_z'(\bar{W}) - E\bar{W}f_z(\bar{W}) \\ = \ & \sum_{i=1}^n E\xi_i^2 I_{\{\xi_i>1\}} Ef_z'(\bar{W}) \\ & + \sum_{i=1}^n E\int_{-\infty}^1 [f_z'(\bar{W}^{(i)}+\bar{\xi}_i) - f_z'(\bar{W}^{(i)}+t)]\bar{K}_i(t)dt \\ & + \sum_{i=1}^n E\xi_i I_{\{\xi_i>1\}} Ef_z(\bar{W}^{(i)}) \\ := \ & R_1 + R_2 + R_3. \end{aligned} \tag{4.17}$$

By $(A1.3)$, (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

$$|R_1| \le C\gamma e^{-z/2}. \tag{4.18}$$

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

$$|R_3| \le C\gamma e^{-z/2}. \tag{4.19}$$

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,

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

From Lemma 4.4 below it follows that

$$\begin{aligned} R_{2,2} &\le \sum_{i=1}^n E \int_{-\infty}^1 I_{\{t \le \bar{\xi}_i\}} [E(\{\bar{W}^{(i)}+\bar{\xi}_i\} f_z(\bar{W}^{(i)}+\bar{\xi}_i) \mid \bar{\xi}_i) - E(\bar{W}^{(i)}+t) f_z(\bar{W}^{(i)}+t)] \bar{K}_i(t) dt \\ &\le C e^{-z/2} \sum_{i=1}^n E \int_{-\infty}^1 (|\bar{\xi}_i| + |t|) \bar{K}_i(t) dt \\ &\le C e^{-z/2} \gamma. \end{aligned} \tag{4.21}$$

Therefore

$$R_2 \le C e^{-z/2} \gamma. \tag{4.22}$$

Similarly, we have

$$R_2 \ge -C e^{-z/2} \gamma. \tag{4.23}$$

This proves the theorem. □

We remain to prove the following lemma.

**Lemma 4.4** *For $s < t \le 1$ we have*

$$\begin{aligned} & E(\bar{W}^{(i)}+t) f_z(\bar{W}^{(i)}+t) - E(\bar{W}^{(i)}+s) f_z(\bar{W}^{(i)}+s) \\ & \quad \le C e^{-z/2} (|s| + |t|) \end{aligned} \tag{4.24}$$

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

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

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

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

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

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

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 (4.24). □

# 5 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 $g(x, y)$

$$Eg(W, \hat{W}) = 0 \tag{5.1}$$

provided the expected value exists.

A key identity is the following lemma.

**Lemma 5.1** *Let* $(W, \hat{W})$ *be an exchangeable pair of real random variables such that*

$$E(\hat{W}|W) = (1-\lambda)W, \ E(\hat{W}-W)^2 = 2\lambda, \tag{5.2}$$

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

$$EWf(W) = \frac{1}{2\lambda} E(W-\hat{W})(f(W)-f(\hat{W})). \tag{5.3}$$

**Proof.** By (5.1),

$$\begin{aligned} 0 &= E(W-\hat{W})(f(\hat{W})+f(W)) \\ &= E(W-\hat{W})(f(\hat{W})-f(W)) + 2Ef(W)(W-\hat{W}) \\ &= E(W-\hat{W})(f(\hat{W})-f(W)) + 2E\{f(W)E(W-\hat{W} \mid W)\} \\ &= E(W-\hat{W})(f(\hat{W})-f(W)) + 2\lambda EWf(W) \qquad \text{[by (5.2)]}, \end{aligned}$$

which gives (5.3). □

Now we prove

**Theorem 5.1** *Let* $h$ *be absolute continuous with bounded* $h'$. *Then under the condition of Lemma 5.1*

$$|Eh(W)-Eh(Z)| \le 2\sup_x |h(x)-Eh(Z)| E|1-\frac{1}{2\lambda}E((\hat{W}-W)^2 \mid W)| + \frac{1}{4\lambda}\sup_x |h'(x)| E|W-\hat{W}|^3. \tag{5.4}$$

**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 (5.3),

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

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

$$\begin{aligned} |Eh(W)-Eh(Z)| &= |Ef'(W)-EWf(W)| \\ &= |Ef'(W)(1-\frac{1}{2\lambda}(W-\hat{W})^2) + \frac{1}{2\lambda}E\int_{-\infty}^{\infty}(f'(W)-f'(W+t))\hat{K}(t)dt| \end{aligned}$$

$$
\begin{aligned}
&\le |E\Big\{f'(W)(1-\frac{1}{2\lambda}E((W-\hat W)^2 \mid W))\Big\}| \\
&\quad +\frac{1}{2\lambda}E|\int_{-\infty}^{\infty}(f'(W)-f'(W+t))\hat K(t)dt| \\
&\le 2\sup_x |h(x)-Eh(Z)|E|(1-\frac{1}{2\lambda}E((W-\hat W)^2 \mid W)| \\
&\quad +\frac{1}{2\lambda}\sup_x |h'(x)|E|\int_{-\infty}^{\infty}|t|\hat K(t)dt| \qquad \text{[by (2.11) and (2.13)]} \\
&= 2\sup_x |h(x)-Eh(Z)|E|1-\frac{1}{2\lambda}E((\hat W-W)^2 \mid W)| + \frac{1}{4\lambda}\sup_x |h'(x)|E|W-\hat W|^3
\end{aligned}
$$

as desired. □

We end this section with the following example to show how to estimate the bound in the above theorem. Let $\xi_i$ 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 (5.2) 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,

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

So the bound is sharp if the fourth moment of $\xi_i$ exists.

## 6 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 $\xi_A$ 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 $\xi_i$ and $\xi_{A_i^c}$ are independent.

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

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

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

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

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

**Theorem 6.1** *Let $N(B_i) = \{j \in \mathcal{J} : B_j B_i \neq \emptyset\}$ and $2 < p \le 4$. Assume that $(LD2)$ is satisfied with $|N(B_i)| \le \kappa$. Then*

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

*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*

$$\sup_z |P(W \le z) - \Phi(z)| \le (13 + 11\kappa)\, n\, \theta^{3\wedge p} + 2.5\theta^{p/2}\sqrt{\kappa n}, \tag{6.1}$$

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

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

**Theorem 6.2** *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*

$$\sup_z |P(W \le z) - \Phi(z)| \le 75\kappa^{p-1} \sum_{i\in\mathcal{J}} E|\xi_i|^p. \tag{6.2}$$

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

**Theorem 6.3** *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*

$$|P(W \le z) - \Phi(z)| \le C\kappa^p (1 + |z|)^{-p} \sum_{i\in\mathcal{J}} E|\xi_i|^p, \tag{6.3}$$

*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 12.3 and 6.3 yields a uniform and a non-uniform bound.

**Theorem 6.4** *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*

$$\sup_z |P(W \le z) - \Phi(z)| \le 75(10m + 1)^{(p-1)d} \sum_{i\in\mathcal{J}} E|\xi_i|^p \tag{6.4}$$

*and*

$$|P(W \le z) - \Phi(z)| \le C(1+|z|)^{-p} 11^{pd} (m+1)^{(p-1)d} \sum_{i\in\mathcal{J}} E|\xi_i|^p, \tag{6.5}$$

*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

$$\begin{array}{rcl} \hat{K}_i(t) & = & \xi_i\{I(-Y_i \le t < 0) - I(0 \le t \le -Y_i)\},\ K_i(t) = E\hat{K}_i(t), \\ \hat{K}(t) & = & \sum_{i\in\mathcal{J}} \hat{K}_i(t),\ K(t) = E\hat{K}(t) = \sum_{i\in\mathcal{J}} K_i(t). \end{array} \tag{6.6}$$

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

$$\begin{aligned} E\{Wf(W)\} &= \sum_{i\in\mathcal{J}} E\{\xi_i(f(W) - f(W-Y_i))\} \\ &= \sum_{i\in\mathcal{J}} E\Big\{\xi_i \int_{-Y_i}^0 f'(W+t)dt\Big\} \\ &= \sum_{i\in\mathcal{J}} E\Big\{\int_{-\infty}^{\infty} f'(W+t)\hat{K}_i(t)dt\Big\} \\ &= E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \end{aligned} \tag{6.7}$$

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

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

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 (2004) for details.

At the end this section, we give the simplest non-uniform concentration inequality in the paper Chen and Shao (2004) and provide a detailed proof to illustrate the difficulty for dependent variables.

**Proposition 6.1** *Assume (LD1). Then for any real numbers $a < b$,*

$$P(a \le W \le b) \le 0.625(b-a) + 4r_1 + 4r_2, \tag{6.8}$$

*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

$$f(w) = \begin{cases} -(b-a+\alpha)/2 & \text{for } w \le a-\alpha \\ \frac{1}{2\alpha}(w-a+\alpha)^2 - (b-a+\alpha)/2 & \text{for } a-\alpha < w \le a \\ w-(a+b)/2 & \text{for } a < w \le b \\ -\frac{1}{2\alpha}(w-b-\alpha)^2 + (b-a+\alpha)/2 & \text{for } b < w \le b+\alpha \\ (b-a+\alpha)/2 & \text{for } w > b+\alpha \end{cases} \tag{6.9}$$

Then $f'$ is a continuous function given by

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

Clearly $|f(w)| \le (b-a+\alpha)/2$. With this $f$, $Y_i$, and $\hat{K}(t)$ and $K(t)$ as defined in (6.6), we have by (6.7)

$$\begin{aligned} (b-a+\alpha)/2 \ &\ge\ EWf(W) = E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \\ &:=\ Ef'(W)\int_{-\infty}^{\infty} K(t)dt + E\int_{-\infty}^{\infty}(f'(W+t)-f'(W))K(t)dt \\ &\quad\ +E\int_{-\infty}^{\infty} f'(W+t)(\hat{K}(t)-K(t))dt \\ &:=\ H_1 + H_2 + H_3. \end{aligned} \tag{6.10}$$

Clearly,

$$H_1 = Ef'(W) \ge P(a \le W \le b). \tag{6.11}$$

By the Cauchy inequality,

$$\begin{aligned} |H_3| &\le (1/8)E\int_{-\infty}^{\infty}[f'(W+t)]^2dt + 2E\int_{-\infty}^{\infty}(\hat{K}(t)-K(t))^2dt \\ &\le (b-a+2\alpha)/8 + 2r_2. \end{aligned} \tag{6.12}$$

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

$$\begin{aligned} |H_2| &\le \alpha^{-1}\int_0^{\infty}\int_0^t L(\alpha)ds|K(t)|dt + \alpha^{-1}\int_{-\infty}^0\int_t^0 L(\alpha)ds|K(t)|dt \\ &= \alpha^{-1}L(\alpha)\int_{-\infty}^{\infty}|tK(t)|dt \le 0.5\alpha^{-1}r_1L(\alpha) = 0.5L(\alpha). \end{aligned} \tag{6.13}$$

It follows from (6.10) - (6.13) that

$$P(a\le W\le b) \le 0.625(b-a) + 0.75\alpha + 2r_2 + 0.5L(\alpha). \tag{6.14}$$

Substituting $a = x$ and $b = x+\alpha$ in (6.14), we obtain

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

and hence

$$L(\alpha) \le 2.75\alpha + 4r_2. \tag{6.15}$$

Finally combining (6.14) and (6.15), we obtain (6.8). □

# 7 Appendix 1

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

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

which also yields

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

by comparing the derivatives of the two functions. Thus

$$\frac{we^{-w^2/2}}{(1+w^2)\sqrt{2\pi}} \le 1-\Phi(w) \le \frac{e^{-w^2/2}}{w\sqrt{2\pi}}. \tag{A1.1}$$

It follows from (2.3) that

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

by $(A1.1)$. This proves (2.6).

In view of the fact that

$$\lim_{w\to-\infty} wf_z(w) = \Phi(z)-1 \text{ and } \lim_{w\to\infty} wf_z(w) = \Phi(z), \tag{A1.2}$$

(2.7) follows by (2.6).

By (2.2), we have

$$\begin{aligned} f_z'(w) &= wf_z(w) + I_{\{w\le z\}} - \Phi(z) \\ &= \begin{cases} wf_z(w)+1-\Phi(z) & \text{for } w<z, \\ wf_z(w)-\Phi(z) & \text{for } w>z. \end{cases} \\ &= \begin{cases} (\sqrt{2\pi}we^{w^2/2}\Phi(w)+1)(1-\Phi(z)) & \text{for } w<z, \\ (\sqrt{2\pi}we^{w^2/2}(1-\Phi(w))-1)\Phi(z) & \text{for } w>z. \end{cases} \end{aligned} \tag{A1.3}$$

Since $wf_z(w)$ is an increasing function of $w$, by $(A1.1)$ and $(A1.2)$

$$0 < f_z'(w) \le zf_z(z)+1-\Phi(z) < 1 \ \text{ for } \ w<z \tag{A1.4}$$

and

$$-1 < zf_z(z)-\Phi(z) \le f_z'(w) < 0 \ \text{ for } \ w>z\ . \tag{A1.5}$$

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 $(A1.4)$ and $(A1.5)$, $f_z$ attains its maximum at $z$. Thus

$$0 < f_z(w) \le f_z(z) = \sqrt{2\pi} e^{z^2/2} \Phi(z)(1 - \Phi(z)). \tag{A1.6}$$

By $(A1.1)$, $f_z(z) \le 1/z$. 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 $(z_0, \infty)$. Since $g_1(0) = 0$ and $g_1(\infty) = \infty$, there exists $z_1 > 0$ such that $g_1(z) < 0$ for $0 < z < z_1$ and $g_1(z) > 0$ for $z > z_1$. Therefore, $g(z)$ attains 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. □

**Proof of Lemma 2.3**. 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 (2.11). 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 (2.12). By (2.4), for $w \geq 0$

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

by $(A1.1)$. 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 (2.11)

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

Thus we have

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

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

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

$$\begin{aligned} f_h''(w) &= wf_h'(w) + f_h(w) + h'(w) \\ &= (1 + w^2) f_h(w) + w(h(w) - Eh(Z)) + h'(w). \end{aligned} \tag{A1.7}$$

From

$$\begin{aligned} & h(x) - Eh(Z) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty [h(x) - h(s)] e^{-s^2/2} ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \int_s^x h'(x) dt e^{-s^2/2} ds - \frac{1}{\sqrt{2\pi}} \int_x^\infty \int_x^s h'(t) dt e^{-s^2/2} ds \\ &= \int_{-\infty}^x h'(t) \Phi(t) dt - \int_x^\infty h'(t)(1 - \Phi(t)) dt, \end{aligned} \tag{A1.8}$$

it follows that

$$f_h(w) = e^{w^2/2} \int_{-\infty}^w [h(x) - Eh(Z)] e^{-x^2/2} dx$$

$$
\begin{aligned}
&= e^{w^2/2}\int_{-\infty}^{w}\Big(\int_{-\infty}^{x} h'(t)\Phi(t)dt - \int_{x}^{\infty} h'(t)(1-\Phi(t))dt\Big)e^{-x^2/2}dx \\
&= -\sqrt{2\pi}e^{w^2/2}(1-\Phi(w))\int_{-\infty}^{w} h'(t)\Phi(t)dt \\
&\quad -\sqrt{2\pi}e^{w^2/2}\Phi(w)\int_{w}^{\infty} h'(t)[1-\Phi(t)]dt. \qquad \text{(A1.9)}
\end{aligned}
$$

From $(A1.7)$ - $(A1.9)$ and $(A1.1)$ we obtain

$$
\begin{aligned}
|f_h''(w)| &\leq |h'(w)| + |(1+w^2)f_h(w) + w(h(w) - Eh(Z))| \\
&\leq |h'(w)| + |\Big(w - \sqrt{2\pi}(1+w^2)e^{w^2/2}(1-\Phi(w))\Big)\int_{-\infty}^{w} h'(t)\Phi(t)dt| \\
&\quad +|\Big(-w - \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big)\int_{w}^{\infty} h'(t)(1-\Phi(t))dt| \\
&\leq |h'(w)| + c_1\Big(-w + \sqrt{2\pi}(1+w^2)e^{w^2/2}(1-\Phi(w))\Big)\int_{-\infty}^{w}\Phi(t)dt \\
&\quad +c_1\Big(w + \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big)\int_{w}^{\infty}(1-\Phi(t))dt \\
&= |h'(w)| + c_1\Big(-w + \sqrt{2\pi}(1+w^2)e^{w^2/2}(1-\Phi(w))\Big)\Big(w\Phi(w) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}\Big) \\
&\quad +c_1\Big(w + \sqrt{2\pi}(1+w^2)e^{w^2/2}\Phi(w)\Big)\Big(-w(1-\Phi(w)) + \frac{e^{-w^2/2}}{\sqrt{2\pi}}\Big) \\
&= |h'(w)| + c_1 \leq 2c_1, \qquad \text{(A1.10)}
\end{aligned}
$$

as desired. □

# References


[1] P. Baldi and Y. Rinott (1989). On normal approximation of distributions in terms of dependency graph. *Ann. Probab.* **17**, 1646-1650.

[2] P. Baldi, Y. Rinott and C. Stein (1989). A normal approximation for the number of local maxima of a random function on a graph. In: Anderson, T.W., Athreya, K.B., Iglehart, D.L. (edts): Probability, statistics, and mathematics. Papers in honor of Samuel Karlin. Academic Press, Boston, pp. 59-81.

[3] A.D. Barbour and L.H.Y. Chen (1992). On the binary expansion of a random integer. *Statist. Probab. Lett.* **14**, 235-241.

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

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

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

[7] E. Bolthausen and F. Götze (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.

[8] L.H.Y. Chen (1986). The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. IMA Preprint Series #243, Univ. Minnesota.

[9] L.H.Y. Chen (1998). Stein's method: some perspectives with applications. *Probability Towards 2000* (L. Accardi and C. C. Heyde, eds.), Lecture Notes in Statistics No. 128, Springer Verlag, 97-122.

[10] L.H.Y. Chen (2000). Non-uniform bounds in probability approximations using Stein's method. In: *Probability and Statistical Models with Applications: A Volume in Honor of Theophilos Cacoullos* (N. Balakrishnan, Ch. A. Charalambides and M.V. Koutras eds). Chapman & Hall/CRC Press, Boca Raton, Florida, 3-14.

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

[12] L.H.Y. Chen and Q.M. Shao (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, 1985-2028.

[13] L.H.Y. Chen and Q.M. Shao (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method* (editors: A.D. Barbour and L.H.Y. Chen), pp. 1-59. Lecture Notes Series, IMS, NUS, Vol. **4**, Singapore.

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

[15] A. Dembo and Y. Rinott (1996). Some examples of normal approximations by Stein's method. In: *Random Discrete Structures.* (Aldous, D. and Pemantle, R. eds). Springer, New York, pp. 25-44.

[16] P. Diaconis (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5**, 72-81.

[17] P. Diaconis and S. Holmes (2004). *Stein's method: expository lectures and applications.* IMS Leture Notes-Monograph, Vol. **46**.

[18] P. Diaconis, S. Holmes, G. Reinert. (2004). Use of exchangeable pairs in the analysis of simulations. In: *Stein's method: expository lectures and applications* (Editors: P. Diaconis and S. Holmes). IMS Leture Notes-Monograph, Vol. **46**, pp.1-26.

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

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

[21] L. Goldstein (2003). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. Preprint.

[22] L. Goldstein and G. Reinert (1997) Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.

[23] L. Goldstein and Y. Rinott (1996). Multivariate normal approximations by Stein's method and size bias couplings. *A. Appl. Probab.* **33**, 1-17.

[24] F. Götze (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

[25] S.T. Ho and L.H.Y. Chen (1978). An $L_p$ bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231-249.

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

[27] L. Paditz (1989). On the analytical structure of the constant in the nonuniform version of the Esseen inequality. *Statistics* **20**, 453-464.

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

[29] Y. Rinott (1994). On normal approximation rates for certain sums of dependent random variables. *J. Computational and Applied Mathematics* **55**, 135-143.

[30] Y. Rinott and V. Rotar (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivate graph related statistics. *J. Multivariate Anal.* **56**, 333-350.

[31] Y. Rinott and V. Rotar (2000). Normal approximations by Stein's method. *Decis. Econ. Finance* **23**, 15-29.

[32] C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **2,** 583-602, Univ. California Press. Berkeley, Calif.

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

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

[35] P. van Beeck (1972). An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Z. Wahrs. verw. Gebiete* **23**, 187-196.

# Part II. Self-normalized limit theorems

## 8 Introduction

Let $X$, $X_1, X_2, \cdots$ be independent and identically distributed random variables. Put

$$S_n = \sum_{i=1}^{n} X_i \text{ and } V_n^2 = \sum_{i=1}^{n} X_i^2. \tag{8.1}$$

It is well-known that moment conditions or other related conditions are necessary and sufficient for many classical limit theorems. For example, the strong law of large numbers holds if and only if the mean of $X$ is finite; the central limit theorem holds if and only if $EX^2 I(|X| \le x)$ is slowing varying as $x \to \infty$; and a necessary and sufficient condition for the large deviation is that the moment generating function of $X$ is finite in a neighborhood of zero. On the other hand, limit theorems for self-normalized sums $S_n/V_n$ put a totally new countenance upon the classical limit theorems. In contrast to the well-known Hartman-Wintner law of the iterated logarithm and its converse by Strassen (1966), Griffin and Kuelbs (1989) obtained a self-normalized law of the iterated logarithm for all distributions in the domain of attraction of a normal or stable law. Shao (1997) showed that no moment conditions are needed for a self-normalized large deviation result $P(S_n/V_n \ge x\sqrt{n})$ and that the tail probability of $S_n/V_n$ is Gaussian like when $X_1$ is in the domain of attraction of the normal law and sub-Gaussian like when $X$ is in the domain of attraction of a stable law, while Giné, Götze and Mason (1997) proved that the tails of $S_n/V_n$ are uniformly sub-Gaussian when the sequence is stochastically bounded. Shao (1999) established a Cramér type result for self-normalized sums only under a finite third moment condition. These results strongly show that self-normalized partial sums preserve desirable properties much better than non-randomized partial sums. Self-normalization is also commonly used in statistics. Many statistical inferences require the use of classical limit theorems. However, these classical results often involve some unknown parameters, one needs to first estimate the unknown parameters and then substitute the estimators into the classical limit theorems. This commonly used practice is exactly the self-normalization. A typical case is the Student t-statistic. The close relationship between the Student t-statistic $T_n$ and the self-normalized sum $S_n/V_n$ can be seen below:

$$T_n := \frac{\bar{X}}{s/\sqrt{n}} = \frac{S_n}{V_n}\Big(\frac{n-1}{n-(S_n/V_n)^2}\Big)^{1/2} \tag{8.2}$$

and

$$\{T_n \ge t\} = \left\{\frac{S_n}{V_n} \ge t\Big(\frac{n}{n+t^2-1}\Big)^{1/2}\right\}, \tag{8.3}$$

where $\bar{X}$ is the sample mean and $s$ is the sample standard deviation. In the second part of lecture series, we shall first give a brief survey on recent developments in the direction of self-normalized limit theory and then focus on how to use Stein's method to prove the Berry-Esseen inequality and Cramér type large deviation for self-normalized sums. The following topics will be reviewed:

1. Self-normalized large deviations
2. Self-normalized saddlepoint approximations
3. Self-normalized moderate deviations
4. Cramér type large deviations for independent random variables
5. Self-normalized laws of the iterated logarithm
6. Limiting distributions of self-normalized sums
7. Weak invariance principle for self-normalized partial sum processes
8. Exponential inequalities for self-normalized processes
9. Applications to statistics

# 9 Self-normalized large deviations

Let $X, X_1, X_2, \cdots$ be a sequence of independent and identically distributed (iid) random variables. The classical Chernoff large deviation [8] states that if

**A**) $Ee^{t_0 X} < \infty$ for some $t_0 > 0$,

then for every $x > EX$,

$$\lim_{n\to\infty} P\left(\frac{S_n}{n} \geq x\right)^{1/n} = \rho(x), \tag{9.4}$$

where $\rho(x) = \inf_{t\geq 0} e^{-tx} Ee^{tX}$.

Roughly speaking, this type of large deviation shows that the convergence rate in the law of large numbers is exponential if the moment generating function is finite in a right neighbourhood of zero. The latter is also necessary for an exponential scale. Essentially built on condition **A**), the area of large deviations in finite dimensional spaces, as well as in abstract spaces, has been well developed, and various applications in statistics (c.f., e.g., Bahadur (1971)), engineering, statistical mechanics and applied probability have been found in recent years. We refer to de Acosta (1988), Stroock (1984), Donsker and Varadhan (1987) and the book by Dembo and Zeitouni (1992), and references therein, for more details.

On the other hand, the following result shows that a self-normalized large deviation remains valid for arbitrary random variables without any moment conditions.

**Theorem 9.1** *[Shao (1997)] Assume that either $E(X) \geq 0$ or $E(X^2) = \infty$. Then*

$$\lim_{n\to\infty} P\left(\frac{S_n}{V_n\, n^{1/2}} \geq x\right)^{1/n} = \sup_{c\geq 0} \inf_{t\geq 0} e^{-tx^2} E e^{t(2cX - c^2X^2)} \tag{9.5}$$

*for $x > E(X)/(E(X^2))^{1/2}$, where $E(X)/(E(X^2))^{1/2} = 0$ if $E(X^2) = \infty$.*

**Remark 9.1** *Note that for any random variable $X$, $E(X^2)$ is either finite or infinite. When $E(X^2)$ is finite, of course, $E(|X|)$ is finite. It is reasonable to assume $E(X) \geq 0$. In fact, when $E(X) \leq 0$,* (9.5) *remains true for $x > 0$.*

A key observation of the proof of Theorem 9.1 is the following identity

$$\forall \;\; x \geq 0, y \geq 0, \quad x^{1/2}\, y^{1/2} = (1/2) \inf_{c>0} (x/c + yc). \tag{9.6}$$

Thus, one can write

$$x\sqrt{n} V_n = (1/2) \inf_{c>0} (V_n^2/c + cx^2 n)$$

and

$$\begin{aligned} P(S_n/V_n \geq x\sqrt{n}) &= P\Big(S_n \geq (1/2) \inf_{c>0} (x^2 n/c + V_n^2 c)\Big) \\ &= P\Big(\sup_{c>0} \sum_{i=1}^{n} (2cX_i - c^2 X_i^2) \geq x^2 n\Big). \end{aligned} \tag{9.7}$$

Now for each fixed $c > 0$, $\{2cX_i - c^2X_i^2, i \geq 1\}$ is a sequence of i.i.d. random variables with finite moment generating function in a right neighborhood of zero. Applying the Chnorff large deviation (9.4) yields

$$\liminf_{n\to\infty} P(S_n/V_n \geq x\sqrt{n})^{1/n} \geq \sup_{c\geq 0} \inf_{t\geq 0} e^{-tx^2} E(e^{t(2cX - c^2X^2)}).$$

As to the upper bound, intuitively, the order of magnitude for the probability of the union of events should be close to the maximum of the probability of the individual event. The detailed proof is much more involved.

Theorem 9.1 has been extended to high dimension by Dembo and Shao (1998a) and to self-normalized empirical processes by Bercu, Gassiat and Rio (2002).

## 10 Self-normalized saddlepoint approximations

Let $X, X_1, X_2, \cdots$ be i.i.d random variables and denote $\bar{X}$ the sample mean of $\{X_i, 1 \le i \le n\}$. Large deviation result provides an exponential rate of convergence for tail probability. However, a more fine-tuned approximation can be offered by saddlepoint approximations. Daniels (1954) showed that the density function of $\bar{X}$ satisfies

$$f_{\bar{X}}(x) = e^{-n(\tau x - \kappa(\tau))}\Big(\frac{n}{2\pi\kappa''(\tau)}\Big)^{1/2}(1 + O(n^{-1})),$$

where $\tau$ is the saddlepoint satisfying $\kappa'(\tau) = x$. Lugannani and Rice (1980) obtained the tail probability of $\bar{X}$:

$$P(\bar{X} \ge x) = 1 - \Phi(\sqrt{n}\hat{w}) + \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}}\Big(\frac{1}{\hat{u}} - \frac{1}{\hat{w}} + O(n^{-1})\Big),$$

where $\kappa'(\tau) = x$, $\hat{w} = \{2[\tau\kappa'(\tau) - \kappa(\tau)]\}^{1/2}\text{sign}\{\tau\}$, $\hat{u} = \tau[\kappa''(\tau)]^{1/2}$, $\Phi$ and $\phi$ denote the standard normal distribution function and density function, respectively. So, the error incurred by the saddlepoint approximation is $O(n^{-1})$ as against the more usual $O(n^{-1/2})$ associated with the normal approximation. Another desirable feature of saddlepoint approximation is that the approximation is quite satisfactory even when the sample size $n$ is small. The book by Jensen (1995) gives a detailed account of saddlepoint approximations and related techniques. However, a finite moment generating function is an essential requirement for saddlepoint expansions. Daniels and Young (1991) derived saddlepoint approximations for the tail probability of the Student t-statistic under the assumption that the moment generating function of $X^2$ exists. Note that Theorem 9.1 holds without any moment assumption. It is natural to ask whether the saddle point approximation is still valid without any moment condition for the t statistic or equivalently, for the self-normalized sum $S_n/V_n$. Jing, Shao and Zhou (2004) recently give an affirmative answer to this question. Let $K(s,t) = \ln Ee^{sX + tX^2}$,

$$K_{11}(s,t) = \frac{\partial^2 K(s,t)}{\partial s^2}, K_{12}(s,t) = \frac{\partial^2 K(s,t)}{\partial s \partial t}, K_{22}(s,t) = \frac{\partial^2 K(s,t)}{\partial t^2}.$$

For $0 < x < 1$, let $\hat{t}_0$ and $a_0$ be solutions $t$ and $a$ to the equations

$$\frac{EXe^{t(-2aX/x^2 + X^2)}}{Ee^{t(-2aX/x^2 + X^2)}} = a, \quad \frac{EX^2e^{t(-2aX/x^2 + X^2)}}{Ee^{t(-2aX/x^2 + X^2)}} = \frac{a^2}{x^2}$$

It is proved in [36] that $\hat{t}_0 < 0$. Put $\hat{s}_0 = -2a_0\hat{t}_0/x^2$ and define

$$\begin{aligned} \Lambda_0(x) &= \hat{s}_0 a_0 + \hat{t}_0 a_0^2/x^2 - K(\hat{s}_0, \hat{t}_0), \\ \Lambda_1(x) &= 2\hat{t}_0/x^2 + (1, 2a_0/x^2)\Delta^{-1}(1, 2a_0/x^2)', \end{aligned}$$

where

$$\Delta = \begin{pmatrix} K_{11}(\hat{s}_0, \hat{t}_0) & K_{12}(\hat{s}_0, \hat{t}_0) \\ K_{12}(\hat{s}_0, \hat{t}_0) & K_{22}(\hat{s}_0, \hat{t}_0) \end{pmatrix}.$$

**Theorem 10.1** [*Jing, Shao and Zhou* (2004)]. *Assume $EX = 0$ or $EX^2 = \infty$ and that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |Ee^{isX+itX^2}|^r dsdt < \infty$ for some $r > 1$. Then for $0 < x < 1$*

$$P(S_n/V_n \geq x\sqrt{n}) = 1 - \Phi(\sqrt{n}w) - \frac{\phi(\sqrt{n}w)}{\sqrt{n}}\Big(\frac{1}{w} - \frac{1}{v} + O(n^{-1})\Big),$$

*where $w = \sqrt{2\Lambda_0(x)}$, and $v = (-\hat{t}_0/2)^{1/2} x^{3/2} a_0^{-1} (\det\Delta)^{1/2}\Lambda_1(x)^{1/2}$.*

## 11 Self-normalized moderate deviations

Let $X, X_1, X_2, \cdots$ be i.i.d random variables and let $\{x_n,\, n \geq 1\}$ be a sequence of positive numbers with $x_n \to \infty$ as $n \to \infty$. It is known that

$$\lim_{n\to\infty} x_n^{-2} \ln P\left(\frac{|S_n|}{\sqrt{n}} \geq x_n\right) = -\frac{1}{2}$$

holds for any sequence $\{x_n\}$ with $x_n \to \infty$ and $x_n = o(\sqrt{n})$ if and only if $EX = 0,\ EX^2 = 1$ and $Ee^{t_0|X|} < \infty$ for some $t_0 > 0$. The sufficient part follows from the Cramér large deviation (c.f. Petrov (1975) ). Following Shao (1989), the necessary part can be proved by studying the increments of $S_n$. The next result shows again that the situation is quite different in the case of self-normalized limit theorems. It tells us that the main term of the asymptotic probability of $P(S_n \geq x_n V_{n,2})$ is distribution free as long as $X$ is in the domain of attraction of a normal law and $x_n = o(\sqrt{n})$.

**Theorem 11.1** *[Shao (1997)] Let $\{x_n,\, n \geq 1\}$ be a sequence of positive numbers with $x_n \to \infty$ and $x_n = o(\sqrt{n})$ as $n \to \infty$. If $EX = 0$ and $EX^2 I\{|X| \leq x\}$ is slowly varying as $x \to \infty$, then*

$$\lim_{n\to\infty} x_n^{-2} \ln P\left(\frac{S_n}{V_{n,2}} \geq x_n\right) = -\frac{1}{2}.$$

Similar results to that of Theorem 11.1 remain valid when $X$ is in the domain of attraction of a stable law.

**Theorem 11.2** *[Shao (1997)] Assume that there exist $0 < \alpha < 2,\ c_1 \geq 0, c_2 \geq 0,\ c_1 + c_2 > 0$ and a slowly varying function $h(x)$ such that*

$$P(X \geq x) = \frac{c_1 + o(1)}{x^\alpha} h(x) \quad \textit{and} \quad P(X \leq -x) = \frac{c_2 + o(1)}{x^\alpha} h(x) \quad \textit{as} \quad x \to \infty.$$

*Moreover, assume that $EX = 0$ if $1 < \alpha < 2$, $X$ is symmetric if $\alpha = 1$ and that $c_1 > 0$ if $0 < \alpha < 1$. Let $\{x_n, n \ge 1\}$ be a sequence of positive numbers with $x_n \to \infty$ and $x_n = o(n^{1/2})$ as $n \to \infty$. Then, we have*

$$\lim_{n\to\infty} x_n^{-2} \ln P\left(\frac{S_n}{V_n} \ge x_n\right) = -\beta(\alpha, c_1, c_2),$$

*where $\beta(\alpha, c_1, c_2)$ is the solution of $\Gamma(\beta, \alpha, c_1, c_2) = 0$ and*

$$\Gamma(\beta,\alpha,c_1,c_2) = \begin{cases} c_1 \int_0^\infty \frac{1+2x-e^{2x-x^2/\beta}}{x^{\alpha+1}} dx + c_2 \int_0^\infty \frac{1-2x-e^{-2x-x^2/\beta}}{x^{\alpha+1}} dx & \text{if } 1 < \alpha < 2, \\ c_1 \int_0^\infty \frac{2-e^{2x-x^2/\beta}-e^{-2x-x^2/\beta}}{x^2} dx & \text{if } \alpha = 1, \\ c_1 \int_0^\infty \frac{1-e^{2x-x^2/\beta}}{x^{\alpha+1}} dx + c_2 \int_0^\infty \frac{1-e^{-2x-x^2/\beta}}{x^{\alpha+1}} dx & \text{if } 0 < \alpha < 1. \end{cases}$$

*In particular, if $X$ is symmetric, then*

$$\lim_{n\to\infty} x_n^{-2} \ln P\left(\frac{S_n}{V_n} \ge x_n\right) = -\beta(\alpha),$$

*where $\beta(\alpha)$ is the solution of* $\displaystyle\int_0^\infty \frac{2-e^{2x-x^2/\beta}-e^{-2x-x^2/\beta}}{x^{\alpha+1}} dx = 0.$

## 12 Cramér type large deviations for independent random variables

Let $X_1, X_2, \cdots$ be a sequence of independent random variables with $EX_i = 0$ and $0 < EX_i^2 < \infty$ for $i \ge 1$. Set

$$S_n = \sum_{i=1}^n X_i, \qquad B_n^2 = \sum_{i=1}^n EX_i^2, \qquad V_n^2 = \sum_{i=1}^n X_i^2.$$

It is well-known that the central limit theorem holds if the Lindeberg condition is satisfied. There are mainly two approaches for estimating the error of the normal approximation. One approach is to study the absolute error in the central limit theorem via Berry-Esseen bounds or Edgeworth expansions. Another approach is to estimate the relative error of $P(S_n \ge xB_n)$ to $1 - \Phi(x)$. One of the typical results in this direction is the so-called Cramér large deviation. If $X_1, X_2, \cdots$ are a sequence of i.i.d. random variables with zero means and a finite moment-generating function $Ee^{tX_1} < \infty$ for $t$ in a neighborhood of zero, then for $x \ge 0$ and $x = o(n^{1/2})$

$$\frac{P(S_n \ge xB_n)}{1-\Phi(x)} = \exp\left\{x^2 \lambda(\frac{x}{\sqrt{n}})\right\} \left[1 + O(\frac{1+x}{\sqrt{n}})\right]$$

where $\lambda(t)$ is the so-called Cramér's series (see [Petrov (1975), Chapter VIII] for details). In particular, if $Ee^{t|X_1|^{1/2}} < \infty$ for some $t > 0$, then

$$\frac{P(S_n \ge xB_n)}{1-\Phi(x)} \to 1 \tag{12.1}$$

holds uniformly for $x \in (0, o(n^{1/6}))$. Note that the moment condition $Ee^{t|X_1|^{1/2}} < \infty$ is necessary. Similar results are also available for independent but not necessarily identically distributed random variables under a finite moment-generating function condition.

Shao (1999) established a (12.1) type result for self-normalized sums only under a finite third-moment condition. More precisely, he showed that if $E|X_1|^{2+\delta} < \infty$ for $0 < \delta \le 1$, then

$$\frac{P(S_n \ge xV_n)}{1-\Phi(x)} \to 1 \tag{12.2}$$

holds uniformly for $x \in (0, o(n^{\delta/(2(2+\delta))}))$. Recently, several papers have focused on the self-normalized limit theorems for independent but not necessarily identically distributed random variables. Bentkus, Bloznelis and Götze (1996) obtained the following Berry-Esseen bound:

$$\begin{aligned} & |P(S_n/V_n \ge x) - (1-\Phi(x))| \\ \le \quad & A\Big(B_n^{-2}\sum_{i=1}^n EX_i^2 I_{\{|X_i|>B_n\}} + B_n^{-3}\sum_{i=1}^n E|X_i|^3 I_{\{|X_i|\le B_n\}}\Big), \end{aligned}$$

where $A$ is an absolute constant. Assuming only finite third moments, Wang and Jing (1999) derived exponential non-uniform Berry-Esseen bounds. Chistyakov and Götze (2003) refined Wang and Jing' results and obtained the following result among others: If $X_1, X_2, \cdots$ are *symmetric* independent random variables with finite third moments, then

$$P(S_n/V_n \ge x) = (1-\Phi(x))\Big(1 + O(1)(1+x)^3 B_n^{-3}\sum_{i=1}^n E|X_i|^3\Big) \tag{12.3}$$

for $0 \le x \le B_n/(\sum_{i=1}^n E|X_i|^3)^{1/3}$, where $O(1)$ is bounded by an absolute constant.

Result (12.3) is useful because it provides not only the relative error but also a Berry-Esseen rate of convergence. It should be noted that if $X_i$ is symmetric, then

$$S_n/V_n \text{ and } \Big(\sum_{i=1}^n \varepsilon_i X_i\Big)/V_n$$

have the same distribution, where $\{\varepsilon_i, i \ge 1\}$ is a Rademacher sequence that is independent of $\{X_i, i \ge 1\}$. Hence given $\{X_i, 1 \le i \le n\}$, the problem reduces to estimate the tail probability of the partial sum of independent random variables $\varepsilon_i X_i/V_n, 1 \le i \le n$. So, the assumption of symmetry not only takes away the main difficulty in proving a self-normalized limit theorem but also limits its potential applications. Jing, Shao and Wang (2003) recently obtained a Cramér-type large deviation for general independent random variables. In particular, they show that (12.3) remains valid for non-symmetric independent random variables. Let

$$\Delta_{n,x} = \frac{(1+x)^2}{B_n^2} \sum_{i=1}^{n} EX_i^2 I\{|X_i| > B_n/(1+x)\} + \frac{(1+x)^3}{B_n^3} \sum_{i=1}^{n} E|X_i|^3 I\{|X_i| \le B_n/(1+x)\}$$

for $x \ge 0$.

**Theorem 12.1** [*Jing, Shao and Wang (2003)*]. *There is an absolute constant $A$ $(> 1)$ such that*

$$\frac{P(S_n \ge xV_n)}{1-\Phi(x)} = e^{O(1)\Delta_{n,x}}$$

*for all $x \ge 0$ satisfying*

$$x^2 \max_{1\le i\le n} EX_i^2 \le B_n^2$$

*and*

$$\Delta_{n,x} \le (1+x)^2/A,$$

*where $|O(1)| \le A$.*

Theorem 18.1 provides a very general framework. The following result is a direct consequence of the above general theorem.

**Theorem 12.2** [*Jing, Shao and Wang (2003)*]. *Let $\{a_n, n \ge 1\}$ be a sequence of positive numbers. Assume that*

$$a_n^2 \le B_n^2 / \max_{1\le i\le n} EX_i^2 \tag{12.4}$$

*and*

$$\forall\, \varepsilon > 0, \quad B_n^{-2} \sum_{i=1}^{n} EX_i^2 I\{|X_i| > \varepsilon B_n/(1+a_n)\} \to 0 \quad \text{as } n \to \infty. \tag{12.5}$$

*Then*

$$\frac{\ln P(S_n/V_n \ge x)}{\ln(1-\Phi(x))} \to 1 \tag{12.6}$$

*holds uniformly for $x \in (0, a_n)$.*

When the $X_i$'s have a finite $(2+\delta)$th moment for $0 < \delta \le 1$, we obtain (12.3) without assuming any symmetric condition.

**Theorem 12.3** *[Jing, Shao and Wang (2003)]. Let* $0 < \delta \le 1$ *and set*

$$L_{n,\delta} = \sum_{i=1}^n E|X_i|^{2+\delta}, \ d_{n,\delta} = B_n / L_{n,\delta}^{1/(2+\delta)}.$$

*Then,*

$$\frac{P(S_n/V_n \ge x)}{1-\Phi(x)} = 1 + O(1)\Big(\frac{1+x}{d_{n,\delta}}\Big)^{2+\delta} \tag{12.7}$$

*for* $0 \le x \le d_{n,\delta}$*, where* $O(1)$ *is bounded by an absolute constant. In particular, if* $d_{n,\delta} \to \infty$ *as* $n \to \infty$*, we have*

$$\frac{P(S_n \ge xV_n)}{1-\Phi(x)} \to 1 \tag{12.8}$$

*uniformly in* $0 \le x \le o(d_{n,\delta})$*.*

By the fact that $1 - \Phi(x) \le 2e^{-x^2/2}/(1+x)$ for $x \ge 0$, it follows from (12.7) that the following exponential non-uniform Berry-Esseen bound

$$|P(S_n/V_n \ge x) - (1-\Phi(x))| \le A(1+x)^{1+\delta} e^{-x^2/2} / d_{n,\delta}^{2+\delta} \tag{12.9}$$

holds for $0 \le x \le d_{n,\delta}$.

The original proof of Theorem 18.1 is quite complicated. We shall use Stein's method to give a new proof.

Theorem 18.1 has been successfully applied to study the studentized bootstrap and the self-normalized law of the iterated logarithm. The proof of Theorem 18.1 is quite complicated. Shao (2004) refined Theorem 12.2 which only requires condition (12.5).

**Theorem 12.4** *[Shao (2004)] Let* $x_n$ *be a sequence of real numbers such that* $x_n \to \infty$ *and* $x_n = o(B_n)$*. Assume*

$$\forall \ \varepsilon > 0, \ B_n^{-2} \sum_{i=1}^n E X_i^2 I\{|X_i| > \varepsilon B_n / x_n\} \to 0. \tag{12.10}$$

*Then we have*

$$\ln P(S_n/V_n \ge x_n) \sim -x_n^2/2. \tag{12.11}$$

## 13 Self-normalized laws of the iterated logarithm

Let $X, X_1, X_2, \cdots$ be i.i.d. random variables. It is well-known that the Hartman-Wintner law of the iterated logarithm holds if and only if the second moment of $X$ is finite. In contrast to this classical result, Griffin and Kuelbs (1989) established a self-normalized law of the iterated logarithm for all distributions in the domain of attraction of a normal or stable law:

**Theorem 13.1** *[Griffin and Kuelbs (1989)]*

*(a) If* $EX = 0$ *and* $EX^2 I\{|X| \le x\}$ *is slowly varying as* $x \to \infty$*, then*

$$\limsup_{n\to\infty} \frac{S_n}{V_n(2\log\log n)^{1/2}} = 1 \quad a.s.$$

*(b) Under the conditions of Theorem 11.2, there is a positive constant* $C$ *such that*

$$\limsup_{n\to\infty} \frac{S_n}{V_n(2\log\log n)^{1/2}} = C \quad a.s.$$

By applying the self-normalized moderate deviation result of Theorem 11.2 and the subsequence method, the precise constant $C$ in Griffin and Kuelbs's LIL can be determined.

**Theorem 13.2** *[Shao (1997)] Under the conditions of Theorem 11.2, we have*

$$\limsup_{n\to\infty} \frac{S_n}{(\log\log n)^{1/2} V_n} = (\beta(\alpha, c_1, c_2))^{1/2} \quad a.s.$$

For non-identically distributed independent random variables, as a direct consequence of Theorem 12.4, we have

**Theorem 13.3** [*Shao*(2004)] *Let* $X_1, X_2, \cdots$ *be independent random variables with* $E(X_i) = 0$ *and* $E(X_i^2) < \infty$*. Assume that* $B_n^2 := \sum_{i=1}^n EX_i^2 \to \infty$ *as* $n \to \infty$ *and that*

$$\forall\ \varepsilon > 0, B_n^{-2} \sum_{i=1}^n EX_i^2 I\{|X_i| > \varepsilon B_n/(\log\log B_n)^{1/2}\} \to 0,$$

*then*

$$\limsup_{n\to\infty} \frac{S_n}{V_n(2\log\log B_n)^{1/2}} = 1 \quad a.s.$$

## 14 Limiting distributions of self-normalized sums

Let $X, X_1, X_2, \cdots$ be i.i.d random variables. It is well-known that $S_n/b_n - a_n$ has a non-degenerate limiting distribution for some suitably chosen real constants $a_n$ and $b_n > 0$ if and only if $X$ is in the domain of attraction of a stable law with index $\alpha$ $(0 < \alpha \le 2)$. When $\alpha = 2$, it is equivalent to $EX^2 I(|X| \le x)$ is slowly varying as $x \to \infty$, while for $0 < \alpha < 2$, it is equivalent to

$$P(X \ge x) = \frac{(c_1 + o(1))h(x)}{x^\alpha},\ P(X \le -x) = \frac{(c_2 + o(1))h(x)}{x^\alpha}$$

as $x \to \infty$, where $c_1 \geq 0, c_2 \geq 0, c_1 + c_2 > 0$ and $h(x)$ is slowly varying at $\infty$. It is also known that the normalizing constants $a_n$ and $b_n$ are complicatedly determined by the slowly varying function $l$. On the other hand, self-normalization shows again its neatness. Logan, Mallows, Rice and Shepp (LMRS)(1973) proved that all limit laws of $S_n/V_{n,2}$ for $X$ in the domain of attraction of a stable law with index $\alpha \in (0,\ 2)$ have a subgaussian tail which depends in a complicated way on the parameter $\alpha$, and posed a conjecture which was solved in [47].

**Theorem 14.1** *[Logan, Mallows, Rice and Shepp (1973) and Shao (1997)]*

*Under the conditions of Theorem 11.2, the limiting density function $p(x)$ of $S_n/V_{n,2}$ exists and satisfies as $x \to \infty$*

$$p(x) \sim \frac{1}{\alpha}\left(\frac{2}{\pi}\right)^{1/2} \sqrt{2\beta(\alpha, c_1, c_2)} e^{-x^2 \beta(\alpha, c_1, c_2)}.$$

Logan, Mallows, Rice and Shepp (1973) also stated that $S_n/V_n$ is asymptotically normal if [and perhaps only if] $X$ is in the domain of attraction of the normal law and that it seems worthy of conjecture that the only possible nontrivial limiting distributions of $S_n/V_n$ are those obtained when $X$ follows a stable law. The first conjecture of LMRS was proved by Giné, Götze and Mason (1997) and the second conjecture was recently confirmed by Chistyakov and Götze (2004).

**Theorem 14.2** *[Gine, Götze and Mason (1997)]*

$$S_n/V_{n,2} \xrightarrow{d.} N(0,1)$$

*if and only if $EX = 0$ and $EX^2 I\{|X| \leq x\}$ is slowly varying.*

**Theorem 14.3** [*Chistyakov and Götze* (2004)]. *$S_n/V_n$ converges weakly to a random variable $Z$ such that $P(|Z| = 1) < 1$ if and only if*

*(i) $X$ is in the domain of attraction of a stable law with index $\alpha \in (0, 2]$;*

*(ii) $EX = 0$ if $1 < \alpha \leq 2$;*

*(iii) if $\alpha = 1$, then $X$ is in the domain of attraction of the Cauchy law and Feller's condition holds, that is, $\lim_{n\to\infty} nE\sin(X/a_n)$ exists and is finite, where $a_n = \inf\{x > 0 : nx^{-2}(1 + EX^2 I\{|X| \leq x\}) \leq 1\}$.*

Chistyakov and Götze (2004) also proved that $S_n/V_n$ converges weakly to $Z$ with $P(|Z|=1)$ if and only if $P(|X|>x)$ is a slowly varying function at $+\infty$.

For the asymptotic distribution of self-normalized censored sums and trimmed sums, we refer to Hahn, Kuelbs and Weiner (1990) and Griffin and Pruitt (1991).

# 15 Weak invariance principle for self-normalized partial sum processes

Let $X, X_1, X_2, \cdots$ be i.i.d. random variables. As a generalization of the central limit theorem, the classical weak invariance principle states that on an appropriate probability space,

$$\sup_{0\le t\le 1} |\frac{1}{\sqrt{n}\sigma} S_{[nt]} - \frac{1}{\sqrt{n}} W(nt)| = o(1) \text{ in probability}$$

if and only if $EX=0$ and $\mathrm{Var}(X)=\sigma^2$, where $\{W(t), t\ge 0\}$ is a standard Wiener process. The weak invariance principle is a stronger version of Donsker's classical functional central limit theorem. In view of the self-normalized central limit theorem, Csörgő, Szyszkowicz and Wang (2003) proved a self-normalized version of the weak invariance principle.

**Theorem 15.1** [*Csörgő, Szyszkowicz and Wang* (2003)]. *As $n\to\infty$, the following statements are equivalent:*

*(a) $EX=0$ and $X$ is in the domain of attraction of the normal law;*

*(b) $S_{[nt]}/V_n \to W(t)$ weakly on $(D[0,1],\rho)$, where $\rho$ is the sup-norm metric for functions in $D[0,1]$, and $\{W(t), 0\le t\le 1\}$ is a standard Wiener process;*

*(c) On an appropriate probability space, we can construct a standard Wiener process $\{W(t), t\ge 0\}$ such that*

$$\sup_{0\le t\le 1} |S_{[nt]}/V_n - W(nt)/\sqrt{n}| = o(1) \quad \textit{in probability.}$$

The following are immediate corollaries of the above weak invariance principle. If $EX=0$ and $X$ is in the domain of attraction of the normal law, then

$$P(\max_{1\le i\le n} S_i/V_n \le x) \to P(\sup_{0\le t\le 1} W(t)\le x),$$

$$P(\max_{1\le i\le n} |S_i|/V_n \le x) \to P(\sup_{0\le t\le 1} |W(t)|\le x),$$

$$P(n^{-1}\sum_{i=1}^{n} S_i^2/V_n^2 \le x) \to P\Big(\int_0^1 W^2(t)dt \le x\Big),$$

$$P(n^{-1}\sum_{i=1}^{n} |S_i|/V_n \le x) \to P\Big(\int_0^1 |W(t)|dt \le x\Big).$$

When $X$ is in the domain of attraction of a stable law, Hu and Shao (2005) obtained a similar invariance principle.

Let $N_1, N_2$ be two independent poisson random measures with intensity $L$ on $\sigma$-finite measure space $(R^2, \mathcal{B}^2, L)$, where $\mathcal{B}^2$ is the Borel $\sigma$-algebra of $R^2$ and $L$ is the Lebesgue measure on $R^2$. For $i = 1, 2$ put

$$N_i(t, y) := N_i([0, t] \times [0, y]) \quad \text{for} \quad t \ge 0,\ y \ge 0$$

and define the following processes:

$$\Delta_{\alpha,i}(t) = \begin{cases} \alpha^{-1}\int_0^\infty N_i(t,y)y^{-1-1/\alpha}dy & \text{if } 0 < \alpha < 1 \\ \int_e^\infty (N_i(t,y) - ty)y^{-2}dy + \int_0^e N_i(t,y)y^{-2}dy & \text{if } \alpha = 1 \\ \alpha^{-1}\int_0^\infty (N_i(t,y) - ty)y^{-1-1/\alpha}dy & \text{if } 1 < \alpha \le 2 \end{cases}$$

**Theorem 15.2** *[Hu and Shao (2005)] Assume that there exists* $\alpha \in (0, 2)$, $c_1 \ge 0, c_2 \ge 0, c_1 + c_2 > 0$ *and a slowly varying function* $h(x)$ *such that*

$$P(X \ge x) = \frac{c_1 + o(1)}{x^\alpha}h(x) \quad \textit{and } P(X \le -x) = \frac{c_2 + o(1)}{x^\alpha}h(x)$$

*as* $x \to \infty$ *Moreover, assume that* $EX = 0$ *if* $1 < \alpha < 2$ *and* $X$ *is symmetric if* $\alpha = 1$. *Write*

$$\Delta_\alpha(t) = -\omega_1\Delta_{\alpha,1}(t) + \omega_2\Delta_{\alpha,2}(t),$$

$$\widetilde{\Delta}_\alpha(t) = \omega_1^2\Delta_{\alpha/2,1}(t) + \omega_2^2\Delta_{\alpha/2,2}(t), \quad 0 \le t \le 1,$$

*where* $w_1 = \left(\frac{c_2}{c_1+c_2}\right)^{\frac{1}{\alpha}}$, $w_2 = \left(\frac{c_1}{c_1+c_2}\right)^{\frac{1}{\alpha}}$, *then in* $D[0, 1]$,

$$S_{[nt]}/V_n \Rightarrow \Delta_\alpha(t)/\sqrt{\widetilde{\Delta}_\alpha(1)}.$$

Theorem 15.2 can be applied to obtain the limiting distribution of a unit-root test statistic when the data is in the domain of attraction of a stable law.

## 16 Exponential inequalities for self-normalized processes

We have focused on limit theorems for self-normalized sums so far. Several papers have recently discussed moment inequalities for self-normalized processes. de la Pena, Klass and Lai (2004) established very interesting exponential inequalities for general self-normalized processes.

**Theorem 16.1** *[de la Pena, Klass and Lai (2004)]. Let $S$ and $V$ be two variables with $V \geq 0$ such that*

$$E \exp\{\lambda S - \frac{\lambda^2}{2} V^2\} \leq 1 \tag{16.1}$$

*for all $\lambda \in \mathbf{R}$. Then for all $y > 0$,*

$$E \frac{y}{\sqrt{V^2 + y^2}} \exp\{\frac{S^2}{2(V^2 + y^2)}\} \leq 1. \tag{16.2}$$

*Consequently, if $EV > 0$, then*

$$E \exp\Big(\frac{S^2}{4(V^2 + (EV)^2)}\Big) \leq \sqrt{2}$$

*and*

$$E \ \exp\Big(\frac{xS}{\sqrt{V^2 + (EV)^2}}\Big) \leq \sqrt{2} \ \exp(x^2)$$

*for $x > 0$. Moreover, for all $p > 0$,*

$$E\Big(\frac{|S|}{\sqrt{V^2 + (EV)^2}}\Big)^p \leq 2^{p+\frac{1}{2}} p\, \Gamma(p/2).$$

Condition (16.1) is satisfied for a large classes of random variables $(S, V)$. Important examples include: (i) $S = W_T$, $V = \sqrt{T}$, where $W_t$ is a standard Brownian motion, and $T$ is a stopping time with $T < \infty$ a.s.; (ii) $S = M_t$, $V = \sqrt{\langle M_t \rangle}$, where $M_t$ is a continuous square-integrable martingale with $M_0 = 0$; (iii) $S = \sum_{i=1}^n d_i$, $V = \sqrt{\sum_{i=1}^n d_i^2}$, where $\{d_i\}$ is a sequence of variables adapted to an increasing sequence of $\sigma$-fields $\{F_i\}$ and the $d_i$'s are conditionally symmetric. They also developed maximal inequalities and iterated logarithm bounds for self-normalized martingales.

# 17 Application to statistics

The idea of self-normalization is by no means new. Self-normalized statistics have been used, out of necessity, for quite some time by now. The celebrated "Student t-statistic" $t_n$, a classical object familiar to virtually everyone who does data analysis, is closely related to $S_n/V_{n,2}$ via the following identities

$$t_n = \frac{S_n}{V_{n,2}} \left(\frac{n-1}{n - (S_n/V_{n,2})^2}\right)^{1/2}$$

and

$$\{t_n \geq t\} = \left\{ \frac{S_n}{V_{n,2}} \geq t \left(\frac{n}{n + t^2 - 1}\right)^{1/2} \right\}.$$

Thus, appropriate properties of self-normalized partial sums can be easily transformed to the t-statistic. For example, under the conditions of Theorem 11.1,

$$\lim_{n\to\infty} x_n^{-2} \ln P(t_n \geq x_n) = -\frac{1}{2}$$

for any $x_n \to \infty$ with $x_n = o(\sqrt{n})$.

The above result enables one to evaluate the exact Bahadur slope of all score tests for any univariate model. He and Shao (1996) derived the exact Bahadur slopes of studentized score tests and showed that under mild conditions the likelihood score is locally optimal in Bahadur efficiency. The self-normalized technique was successfully applied in Chen and Shao (1997) to study the performance of Monte Carlo methods for estimating ratios of normalizing constants.

Cao (2005) recently studied the moderate deviations for two-sample t-statistics.

Let $X_1, \cdots, X_{n_1}$ be a random sample from a population with mean $\mu_1$ and variance $\sigma_1^2$, and $Y_1, \cdots, Y_{n_2}$ be a random sample from another population with mean $\mu_2$ and variance $\sigma_2^2$. Define the two sample t-statistic

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}},$$

where

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2, s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$

**Theorem 17.1** [*Cao*(2005)] *Let $n = n_1 + n_2$. Assume that there are $0 < c_1 \leq c_2 < \infty$ such that $c_1 \leq n_1/n_2 \leq c_2$. Then for any $x := x(n_1, n_2)$ satisfying $x \to \infty$, $x = o(n^{1/2})$*

$$\ln P(T \geq x) \sim -x^2/2$$

*as $n_1, n_2 \to \infty$. If in addition, $E|X_1|^3 < \infty$ and $E|Y_1|^3 < \infty$, then*

$$\frac{P(T \geq x)}{1 - \Phi(x)} = 1 + O(1)(1+x)^3 n^{-1/2} d^3$$

*for $0 \leq x \leq n^{1/6}/d$, where $d^3 = (E|X_1 - \mu_1|^3 + E|Y_1 - \mu_2|^3)/(\sigma_1^2 + \sigma_2^2)^{3/2}$ and $O(1)$ is a finite constant depending only on $c_1$ and $c_2$. In particular,*

$$\frac{P(T \geq x)}{1 - \Phi(x)} \to 1$$

*uniformly in $x \in (0, o(n^{1/6}))$.*

## 18 An explicit Berry-Esseen bound for Student's t-statistic via Stein's method

Let $X_1$, $X_2$, ..., $X_n$ be a sequence of independent random variables with $EX_i = 0$ and $EX_i^2 < \infty$. Put

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

Define the Student t-statistic by

$$T_n = \frac{S_n}{\sqrt{n}\ s},$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - S_n/n)^2$. The study of Berry-Esseen bound for Student's statistic has a long history (see, e.g., Bentkus and Götze (1996) and references therein) and the first optimal result is due to Bentkus and Götze (1996) for i.i.d case, which is extended to the non-i.i.d case by Bentkus, Bloznelis and Götze (1996). In particular, they show that

$$\sup_z |P(T_n \le z) - \Phi(z)| \le C_1 B_n^{-2} \sum_{i=1}^n EX_i^2 I_{\{|X_i| > B_n/2\}} + C_2 B_n^{-3} \sum_{i=1}^n E|X_i|^3 I_{\{|X_i| \le B_n/2\}}, \tag{18.1}$$

where $C_1$ and $C_2$ are absolute constants. The bound coincides the classical Berry-Esseen bound for the standardized mean $S_n/B_n$ up to an absolute constant. Their proof is based on the characteristic function approach. Her we give a direct proof of (18.1) with explicit values of $C_1$ and $C_2$ via Stein's method. As we mentioned before, the Student t-statistic is closely related to the self-normalized sum $S_n/V_n$ via the following identity

$$\{T_n \ge x\} = \Big\{\frac{S_n}{V_n} \ge x\Big(\frac{n}{n+x^2-1}\Big)^{1/2}\Big\}.$$

Hence we state our main result in terms of the self-normalized sum.

**Theorem 18.1** *We have*

$$\sup_z |P(S_n/V_n \le z) - \Phi(z)| \le 10.2\beta_2 + 25\beta_3, \tag{18.2}$$

*where*

$$\beta_2 = B_n^{-2} \sum_{i=1}^n EX_i^2 I_{\{|X_i| > 0.5B_n\}} \text{ and } \beta_3 = B_n^{-3} \sum_{i=1}^n E|X_i|^3 I_{\{|X_i| \le 0.5B_n\}}. \tag{18.3}$$

*In particular, we have*

$$\sup_z |P(S_n/V_n \le z) - \Phi(z)| \le 25 B_n^{-p} \sum_{i=1}^n E|X_i|^p \tag{18.4}$$

*for* $2 < p \le 3$.

**Remark.** For i.i.d case, Nagaev (2002) also obtained an explicit bound similar to (18.4) but with a bigger value of the constant.

**Proof.** We assume that $\{X_i, 1 \le i \le n\}$ are independent with $EX_i = 0$ and finite second moments. Without loss of generality, assume $B_n = 1$ and $z \ge 0$. The proof is carried out via (i) truncation and (ii) applying the Stein's method to the truncated variables. The main difficulty arises from the variable $V_n$ in the denominator which can not be simply replaced by $(E(V_n^2))^{1/2}$.

(18.2) is obviously true if $\beta_2 > 0.1$ or $\beta_3 > 0.04$. So we can assume that

$$\beta_2 \le 0.1 \ \text{ and } \ \beta_3 \le 0.04. \tag{18.5}$$

Recall a Hoeffding (1963) type inequality for non-negative independent random variables (see (4.6))

*Let* $\{Y_i, 1 \le i \le n\}$ *be independent non-negative random variables with* $\mu = \sum_{i=1}^n EY_i$ *and* $\sigma^2 = \sum_{i=1}^n EY_i^2 < \infty$*. Then for* $0 < x < \mu$

$$P(\sum_{i=1}^n Y_i \le x) \le \exp\Big(-\frac{(\mu - x)^2}{2\sigma^2}\Big). \tag{18.6}$$

A direct corollary of (18.6) is: for $c^2 < 1 - \beta_2$

$$\begin{aligned} P(V_n \le c) &\le P(\sum_{i=1}^n X_i^2 I_{\{|X_i| \le 0.5\}} \le c^2) \\ &\le \exp\Big(-\frac{(\sum_{i=1}^n EX_i^2 I_{\{|X_i|\le 0.5\}} - c^2)^2}{2\sum_{i=1}^n EX_i^4 I_{\{|X_i|\le 0.5\}}}\Big) \\ &\le \exp\Big(-\frac{(1-\beta_2 - c^2)^2}{\beta_3}\Big). \end{aligned} \tag{18.7}$$

Hence for $z > 0$

$$\begin{aligned} P(S_n/V_n \ge z) &\le P(V_n \le 0.7) + P(S_n \ge 0.7z) \\ &\le \exp\Big(-\frac{(1-0.1-0.49)^2}{0.04}\Big) + \frac{1}{1+(0.7z)^2} \\ &\le 0.015 + \frac{1}{1+(0.7z)^2} \end{aligned}$$

and

$$\begin{aligned} &\sup_{z\ge 0}(P(S_n/V_n \ge z) - (1-\Phi(z))) \\ &\le 0.015 + \sup_{z \ge 0}\Big(\frac{1}{1+(0.7z)^2} - (1-\Phi(z))\Big) \\ &\le 0.015 + 0.585 = 0.6. \end{aligned} \tag{18.8}$$

Thus, (18.2) holds if $\beta_2 > 0.6/10 = 0.06$ or $\beta_3 > 0.6/25 = 0.024$. Therefore we can further assume

$$\beta_2 \le 0.06 \text{ and } \beta_3 \le 0.024. \tag{18.9}$$

Let

$$\begin{aligned}
&\bar{X}_i = X_i I_{\{|X_i| \le 0.5\}},\ \bar{S}_n = \textstyle\sum_{i=1}^n \bar{X}_i,\\
&\bar{V}_n^2 = \textstyle\sum_{i=1}^n \bar{X}_i^2,\ V_n^* = \max(\bar{V}_n, 0.8),\\
&\xi_i = \bar{X}_i/V_n^*,\ W = \textstyle\sum_{i=1}^n \xi_i = \bar{S}_n/V_n^*,\\
&\kappa_i(t) = \xi_i I_{\{-\xi_i \le t \le 0\}} - \xi_i I_{\{0 < t \le -\xi_i\}},\quad \kappa(t) = \textstyle\sum_{i=1}^n \kappa_i(t).
\end{aligned} \tag{18.10}$$

Note that

$$\begin{aligned}
&|P(S_n/V_n \le z) - P(W \le z)|\\
&\le\ P(\max_{1\le i\le n} |X_i| > 0.5) + P(\bar{V}_n < 0.8)\\
&\le\ 4\beta_2 + \exp\Big(-\frac{(1-0.06-0.64)^2}{\beta_3}\Big) \qquad \text{[by (18.7)]}\\
&\le\ 4\beta_2 + \beta_3 \sup_{0<x<0.024} \frac{1}{x}\exp\Big(-\frac{0.09}{x}\Big)\\
&\le\ 4\beta_2 + 0.98\beta_3.
\end{aligned} \tag{18.11}$$

Now we use the Stein method to bound $P(W \le z) - \Phi(z)$. Observe that for any absolute continuous function $f$

$$\begin{aligned}
&Wf(W) - \sum_{i=1}^n \xi_i f(W - \xi_i) \qquad (18.12)\\
&= \sum_{i=1}^n \xi_i (f(W) - f(W-\xi_i)) = \sum_{i=1}^n \xi_i \int_{-\xi_i}^0 f'(W+t)dt\\
&= \sum_{i=1}^n \xi_i \int_{-1}^1 f'(W+t)[I_{\{-\xi_i \le t \le 0\}} - I_{\{0 < t \le -\xi_i\}}]dt\\
&= \sum_{i=1}^n \int_{-1}^1 f'(W+t)m_i(t)dt = \int_{-1}^1 f'(W+t)m(t)dt.
\end{aligned}$$

Noting that

$$\int_{-1}^1 m(t)dt = \sum_{i=1}^n \xi_i^2 = \bar{V}_n^2/V_n^{*2} = 1 - (1 - \bar{V}_n^2/0.64)I_{\{\bar{V}_n < 0.8\}},$$

we have

$$\begin{aligned}
&f'(W) - Wf(W) \qquad (18.13)\\
&= f'(W)(1 - \bar{V}_n^2/0.64)I_{\{\bar{V}_n<0.8\}} - \sum_{i=1}^n \xi_i f(W-\xi_i) + \int_{-1}^1 [f'(W) - f'(W+t)]m(t)dt.
\end{aligned}$$

Now we let $f := f_z$ be the solution of the following Stein equation:

$$f'(w) - wf(w) = I(w \le z) - \Phi(z). \tag{18.14}$$

Then

$$P(W \le z) - \Phi(z) = R_1 - R_2 + R_3, \tag{18.15}$$

where

$$R_1 = Ef_z'(W)(1 - \bar{V}_n^2/0.64)I_{\{\bar{V}_n < 0.8\}}, \tag{18.16}$$

$$R_2 = \sum_{i=1}^n E\xi_i f_z(W - \xi_i), \tag{18.17}$$

$$R_3 = \int_{-1}^1 E\{(f_z'(W) - f_z'(W+t))m(t)\}dt. \tag{18.18}$$

Recall

$$0 \le f_z(w) \le \sqrt{2\pi}/4 < 0.627, \tag{18.19}$$

$$|f_z'(w)| \le 1, \tag{18.20}$$

$$|wf_z(w) - (w+t)f_z(w+t)| \le (|w| + 0.627)|t|. \tag{18.21}$$

It follows from (18.20), (18.7) and the proof of (18.11) that

$$|R_1| \le P(\bar{V}_n \le 0.8) \le 0.98\beta_3. \tag{18.22}$$

Now by Lemmas 18.1 and 18.2 below,

$$|R_2| \le 1.568\beta_2 + 1.85\beta_3 \text{ and } |R_3| \le 4.59\beta_2 + 21.92\beta_3.$$

This proves (18.2) by combining the inequalities above. □

We now present two lemmas used in the main proof.

**Lemma 18.1** *Under* (18.9)*, we have*

$$|R_2| \le 1.568\beta_2 + 1.85\beta_3. \tag{18.23}$$

**Proof.** Note that $\xi_i$ and $W - \xi_i$ are almost independent and the dependence is mainly because of the denominator $V_n^*$ in $\xi_i$ and $W - \xi_i$. To eliminate the dependence, we introduce

$$\bar{V}_{(i)} = \Big(\sum_{1 \le j \le n, j \ne i} \bar{X}_j^2\Big)^{1/2},\ V_{(i)}^* = \max(\bar{V}_{(i)}, 0.8),\ W_{(i)}^* = \frac{\bar{S}_n - \bar{X}_i}{V_{(i)}^*}. \tag{18.24}$$

Since $E(X_i) = 0$,

$$|E(\bar{S}_n - \bar{X}_i)| = |\sum_{j \neq i} E(X_j I_{\{|X_j|>0.5\}}| \leq 2\beta_2.$$

Thus, under (18.9), we have

$$\begin{aligned} E|\bar{S}_n - \bar{X}_i| &\leq (E(\bar{S}_n - \bar{X}_i)^2)^{1/2} \\ &\leq \Big(\sum_{j=1}^n E\bar{X}_j^2 + (E(\bar{S}_n - \bar{X}_i))^2\Big)^{1/2} \\ &\leq (1 + 4\beta_2^2)^{1/2} \leq 1.008. \end{aligned} \tag{18.25}$$

Noting that by (18.19) and (18.20)

$$|\frac{d}{dx}(axf_z(bx))| = |af_z(bx) + abxf_z'(bx)| \leq 0.627|a| + |ab|/0.8$$

for $|x| \leq 1/0.8$ and for constants $a$ and $b$, and

$$\begin{aligned} 0 &\leq \frac{1}{V_{(i)}^*} - \frac{1}{V_n^*} = \frac{V_n^{*2} - V_{(i)}^{*2}}{V_n^* V_{(i)}^* (V_n^* + V_{(i)}^*)} \\ &\leq \frac{\bar{X}_i^2}{V_n^* V_{(i)}^* (V_n^* + V_{(i)}^*)} \leq \frac{\bar{X}_i^2}{2(0.8)^3} \leq 0.98\bar{X}_i^2, \end{aligned} \tag{18.26}$$

we have

$$\begin{aligned} |R_2| &\leq \sum_{i=1}^n |E\Big\{\frac{\bar{X}_i}{V_n^*} f_z\Big(\frac{\bar{S}_n - \bar{X}_i}{V_n^*}\Big)\Big\} - E\Big\{\frac{\bar{X}_i}{V_{(i)}^*} f_z\Big(\frac{\bar{S}_n - \bar{X}_i}{V_{(i)}^*}\Big)\Big\}| \\ &\quad + \sum_{i=1}^n |E\Big\{\frac{\bar{X}_i}{V_{(i)}^*} f_z\Big(\frac{\bar{S}_n - \bar{X}_i}{V_{(i)}^*}\Big)\Big\}| \\ &\leq \sum_{i=1}^n E\Big\{(0.627|\bar{X}_i| + \frac{|\bar{X}_i(\bar{S}_n - \bar{X}_i)|}{0.8})0.98\bar{X}_i^2\Big\} \\ &\quad + \sum_{i=1}^n |E\bar{X}_i| \; |E\Big\{\frac{1}{V_{(i)}^*} f_z\Big(\frac{\bar{S}_n - \bar{X}_i}{V_{(i)}^*}\Big)\Big\}| \\ &\leq 0.615 \sum_{i=1}^n E|\bar{X}_i|^3 + (0.98/0.8) \sum_{i=1}^n E|\bar{X}_i|^3 E|\bar{S}_n - \bar{X}_i| \\ &\quad + 0.627(0.8)^{-1} \sum_{i=1}^n |E\bar{X}_i| \qquad \text{[by (18.19)]} \\ &\leq (0.615 + (1.008)(0.98)/0.8)\beta_3 + 0.627/0.4\beta_2 \qquad \text{[by (18.25)]} \\ &\leq 1.568\beta_2 + 1.85\beta_3, \end{aligned} \tag{18.27}$$

as desired. □

Next lemma provides an estimate for $R_3$.

**Lemma 18.2** *Under* (18.9), *we have*

$$|R_3| \le 4.59\beta_2 + 21.06\beta_3. \tag{18.28}$$

**Proof.** First we prove

$$R_3 \le 4.59\beta_2 + 21.92\beta_3. \tag{18.29}$$

The proof of $-R_3 \le 4.59\beta_2 + 21.92\beta_3$ is similar and hence will be omitted.

By (18.14) and (18.21),

$$\begin{aligned} R_3 &\le E\int_{-1}^{1} m(t)(|W| + 0.627)|t|dt + E\int_{-1}^{1} m(t)(I_{\{W\le z\}} - I_{\{W+t\le z\}})dt \\ &\le 0.5E\Big((|W| + 0.627)\sum_{i=1}^{n}|\xi_i|^3\Big) + \sum_{i=1}^{n} E\int_0^1 m(t)I_{\{z-t<W\le z\}}dt \\ &\le 0.5(0.8)^{-3}\sum_{i=1}^{n} E\Big(|\bar{X}_i|^3(|\bar{S}_n|/0.8 + 0.627)\Big) + \sum_{i=1}^{n} E(\xi_i^2 I_{\{z+\xi_i<W\le z,\xi_i\le 0\}}) \\ &\le 0.98\sum_{i=1}^{n} E\Big(|\bar{X}_i|^3(0.627 + |\bar{S}_n|/0.8)\Big) + \sum_{i=1}^{n} E(\xi_i^2 I_{\{z-|\bar{X}_i|/0.8<W\le z\}}) \\ &:= R_{3,1} + R_{3,2}. \end{aligned} \tag{18.30}$$

By (18.25),

$$\begin{aligned} R_{3,1} &\le 0.98\sum_{i=1}^{n}\Big\{(0.627 + 0.5/0.8)E|\bar{X}_i|^3 + E|\bar{X}_i|^3 E|\bar{S}_n - \bar{X}_i|/0.8\Big\} \\ &\le 0.98(0.627 + 1/1.6 + 1.008/0.8)\beta_3 \le 2.47\beta_3. \end{aligned} \tag{18.31}$$

To bound $R_{3,2}$, let

$$\delta = 0.75\sum_{i=1}^{n}|\xi_i|^3, \ \ \eta_i = |\bar{X}_i|/0.8$$

and define

$$h_{i,\delta}(x) = \begin{cases} -\delta - \eta_i/2 & \text{for } x \le z - \eta_i - \delta \\ x - (2z - \eta_i)/2 & \text{for } z - \eta_i - \delta < x < z + \delta \\ \delta + \eta_i/2 & \text{for } x \ge z + \delta \end{cases} \tag{18.32}$$

Then, by (18.12)

$$\begin{aligned} &W\,h_{i,\delta}(W) - \sum_{j=1}^{n}\xi_j h_{i,\delta}(W - \xi_j) \\ &= \int_{-1}^{1} m(t)h'_{i,\delta}(W+t)dt \ge \int_{|t|\le\delta} I_{\{z-\eta_i\le W\le z\}}m(t)dt \\ &= I_{\{z-\eta_i\le W\le z\}}\sum_{j=1}^{n}|\xi_j|\min(\delta, |\xi_j|) \end{aligned}$$

$$
\begin{aligned}
&\geq I_{\{z-\eta_i \leq W \leq z\}} \sum_{j=1}^{n} |\xi_j|(|\xi_j| - \xi_j^2/(4\delta)) \\
&= I_{\{z-\eta_i \leq W \leq z\}} (\bar{V}_n^2/\max(\bar{V}_n^2, 0.8^2) - 1/3) \\
&\geq (2/3) I_{\{z-\eta_i \leq W \leq z\}} - I_{\{\bar{V}_n < 0.8\}}, \qquad (18.33)
\end{aligned}
$$

where in the last but second inequality, we used the fact that:

$$
\min(a, b) \geq a - a^2/(4b) \ \text{ for any } a \geq 0, b > 0.
$$

Hence

$$
\begin{aligned}
&(2/3) \sum_{i=1}^{n} E\xi_i^2 I_{\{z-\eta_i \leq W \leq z\}} \\
&\leq EI_{\{\bar{V}_n < 0.8\}} \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} E\xi_i^2 W h_{i,\delta}(W) - \sum_{j=1}^{n} \sum_{i=1}^{n} E\xi_i^2 \xi_j h_{i,\delta}(W - \xi_j) \\
&\leq P(\bar{V}_n < 0.8) + R_{3,3} + R_{3,4} + R_{3,5} \\
&\leq 0.98\beta_3 + R_{3,3} + R_{3,4} + R_{3,5} \qquad (18.34)
\end{aligned}
$$

by (18.22), where

$$
\begin{aligned}
R_{3,3} &= \sum_{i=1}^{n} E\Big(\xi_i^2 |W| (\delta + \eta_i/2)\Big), \\
R_{3,4} &= -\sum_{i=1}^{n} E\xi_i^3 h_{i,\delta}(W - \xi_i), \\
R_{3,5} &= -\sum_{i=1}^{n} \sum_{j \neq i} E\xi_i^2 \xi_j h_{i,\delta}(W - \xi_j).
\end{aligned}
$$

From (18.25) we obtain that

$$
\begin{aligned}
|R_{3,3}| &\leq E(|W|\delta) + 0.5 \sum_{i=1}^{n} E(\xi_i^2 \eta_i |W|) \\
&\leq (0.5 + 0.75) 0.8^{-4} \sum_{i=1}^{n} E|\bar{X}_i^3 \bar{S}_n| \\
&\leq 1.25(0.8)^{-4}(0.5 + 1.008)\beta_3 \leq 4.603\beta_3 \qquad (18.35)
\end{aligned}
$$

and

$$
\begin{aligned}
|R_{3,4}| &\leq \sum_{i=1}^{n} E|\xi_i^3|(\delta + 0.5\eta_i) \\
&\leq 0.75 E\Big(\sum_{i=1}^{n} |\xi_i|^3\Big)^2 + 0.5(0.8)^{-4} \sum_{i=1}^{n} E|\bar{X}_i|^4
\end{aligned}
$$

$$
\begin{aligned}
&\leq (0.75/1.6)E\Big(\sum_{i=1}^{n}|\xi_i|^3\Big) + 0.6104\sum_{i=1}^{n}E|\bar{X}_i|^3 \\
&\leq 1.53\beta_3, \qquad (18.36)
\end{aligned}
$$

where we used the fact that $\sum_{i=1}^n |\xi_i|^3 \leq (0.5/0.8)\sum_{i=1}^n \xi_i^2 \leq 1/1.6$.

To bound $R_{3,5}$, we define $V_{(j)}^*$ as in (18.24) and $h_{i,\delta_j}$ as in (18.32) with

$$
\delta_j = 0.75 \sum_{1\leq l\leq n, l\neq j}^{n} |\bar{X}_l|^3/V_{(j)}^{*3}.
$$

Observe that

$$
\begin{aligned}
0 \quad &\leq \quad \frac{1}{V_{(j)}^{*3}} - \frac{1}{V_n^{*3}} = \frac{V_n^{*3} - V_{(j)}^3}{V_n^{*3}V_{(j)}^{*3}} \\
&= \quad \frac{1.5\int_{V_{(j)}^{*2}}^{V_n^{*2}} \sqrt{t}\,dt}{V_n^{*3}V_{(j)}^{*3}} \leq \frac{1.5|\bar{X}_j|^2}{V_n^{*2}V_{(j)}^{*3}}.
\end{aligned}
$$

Clearly for any $x$

$$
\begin{aligned}
|h_{i,\delta}(x) - h_{i,\delta_j}(x)| \quad &\leq \quad |\delta - \delta_j| \qquad (18.37) \\
&\leq \quad \frac{0.75|\bar{X}_j|^3}{V_n^{*3}} + 0.75\sum_{l\neq j}|\bar{X}_l|^3(1/V_{(j)}^{*3} - 1/V_n^{*3}) \\
&\leq \quad \frac{0.75|\bar{X}_j|^3}{V_n^{*3}} + \frac{1.5(0.75)|\bar{X}_j|^2}{V_n^{*2}V_{(j)}^{*3}}\sum_{l\neq j}|\bar{X}_l|^3 \\
&\leq \quad \frac{0.375|\bar{X}_j|^2}{0.8^3} + \frac{1.125|\bar{X}_j|^2}{V_n^{*2}0.8^3}\sum_{l\neq j}0.5\bar{X}_l^2 \\
&\leq \quad 1.832\bar{X}_j^2.
\end{aligned}
$$

Write $R_{3,5} = R_{3,6} + R_{3,7}$, where

$$
\begin{aligned}
R_{3,6} \quad &= \quad \sum_{i=1}^{n}\sum_{j\neq i} E\{\xi_i^2\xi_j(h_{i,\delta_j}(W-\xi_j) - h_{i,\delta}(W-\xi_j))\}, \\
R_{3,7} \quad &= \quad -\sum_{i=1}^{n}\sum_{j\neq i} E\xi_i^2\xi_j h_{i,\delta_j}(W-\xi_j).
\end{aligned}
$$

By (18.37),

$$
\begin{aligned}
|R_{3,6}| \quad &\leq \quad 1.832\sum_{i=1}^{n}\sum_{j\neq i}E\{\xi_i^2|\xi_j|\bar{X}_j^2\} \\
&\leq \quad 1.832\sum_{j=1}^{n}E\{|\xi_j|\bar{X}_j^2\} \\
&\leq \quad (1.832/0.8)\beta_3 = 2.29\beta_3. \qquad (18.38)
\end{aligned}
$$

Notice that

$$|h_{i,\delta_j}| \le \delta_i + \eta_i/2 \le 0.75(0.5/0.8) + 0.5/1.6 = 0.782.$$

Since $\bar{X}_j$ and $\{V^*_{(j)}, \bar{S}_n - \bar{X}_j, \bar{X}_i\}$ are independent for $j \neq i$, we have

$$\begin{aligned}
|R_{3,7}| \le\ & |\sum_{i=1}^{n}\sum_{j\neq i}|E(\bar{X}_j)\ E\bar{X}_i^2 V_{(j)}^{-3} h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_{(j)})\Big\}| \\
& + \sum_{i=1}^{n}\sum_{j\neq i}|E\Big(\bar{X}_i^2\bar{X}_j(1/V_n^*)^3 h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_n^*) - \bar{X}_i^2\bar{X}_j(1/V_{(j)}^*)^3 h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_{(j)}^*)\Big)| \\
\le\ & 0.782(0.8)^{-3}\sum_{i=1}^{n}\sum_{j\neq i}E|X_j|I_{\{|X_j|>0.5\}}\ E\bar{X}_i^2 \\
& + \sum_{i=1}^{n}\sum_{j\neq i}|E\Big(\bar{X}_i^2\bar{X}_j(V_n^{*-3} - V_{(j)}^{*-3})h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_{(j)}^*)\Big)| \\
& + \sum_{i=1}^{n}\sum_{j\neq i}|E\Big(\bar{X}_i^2\bar{X}_j V_n^{*-3}(h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_n^*) - h_{i,\delta_j}((\bar{S}_n - \bar{X}_j)/V_{(j)}^*))\Big)| \\
\le\ & 3.06\beta_2 + 0.782\sum_{i=1}^{n}\sum_{j\neq i}E\Big(\frac{1.5\bar{X}_i^2|\bar{X}_j|^3}{V_n^{*2}V_{(j)}^3}\Big) \\
& + \sum_{i=1}^{n}\sum_{j\neq i}E\Big(\frac{\bar{X}_i^2|\bar{X}_j||\bar{S}_n - \bar{X}_j|}{V_n^{*3}}(\frac{1}{V_{(j)}^*} - \frac{1}{V_n^*})\Big) \\
\le\ & 3.06\beta_2 + \frac{0.782(1.5)}{0.8^3}\beta_3 \\
& + \sum_{i=1}^{n}\sum_{j\neq i}E\Big(\frac{\bar{X}_i^2|\bar{X}_j|^3|\bar{S}_n - \bar{X}_j|}{V_n^{*3}}\Big) \qquad \text{[by (18.26)]} \\
\le\ & 3.06\beta_2 + 2.292\beta_3 + \frac{1}{0.8}\sum_{j=1}^{n}E|\bar{X}_j|^3 E|\bar{S}_n - \bar{X}_j| \qquad \text{[by (18.25)]} \\
\le\ & 3.06\beta_2 + 3.56\beta_3. \qquad (18.39)
\end{aligned}$$

Now combining (18.34), (18.35), (18.36), (18.38) and (18.39) yields

$$R_{3,2} \le 1.5(0.98\beta_3 + 4.603\beta_3 + 1.53\beta_3 + 2.29\beta_3 + 3.06\beta_2 + 3.56\beta_3) \le 4.59\beta_2 + 19.45\beta_3. \qquad (18.40)$$

This proves (18.29) by (18.30), (18.31) and (18.40). □

# References


[1] A. de Acosta, Large deviations for vector-valued functionals of a Markov chain: lower bounds. Ann. Probab. 16 (1988), 925–960.

[2] R.R. Bahadur, Some limit theorems in statistics. Regional Conference Series in Appl. Math. No. 4, SIAM, Philadelphia, 1971.

[3] V. Bentkus, M. Bloznelis and F. Götze. A Berry-Esseen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.* **9** (1996), 765–796.

[4] V. Bentkus and F. Götze, The Berry-Esséen bound for Student's statistic. Ann. Probab. 24 (1996), 491-503.

[5] B. Bercu, E. Gassiat and E. Rio, Concentration inequalities, large and moderate deviations for self-normalized empirical processes. *Ann. Probab.* **30** (2002), 1576–1604.

[6] H.Y. Cao, Moderate deviations for two sample t-statistics. Preprint.

[7] M.H. Chen and Q.M. Shao, On Monte Carlo methods for estimating ratios of normalizing constants. Ann. Statist. 25 (1997), 1563-1594.

[8] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist. 23 (1952), 493-507.

[9] G.P. Chistyakov and F. Götze, Moderate deviations for self-normalized sums. *Theory Probab. Appl.* **47** (2003), 415–428.

[10] G.P. Chistyakov and F. Götze, Limit distributions of Studentized means. *Ann. Probab.* **32** (2004), 28-77.

[11] H. Cramér, Sur un nouveaux théoréme limite de la théorie des probabilités. Actualités Sci. Indust No. 736 (1938), 5-23, Hermann, Paris.

[12] S. Csörgő, E. Haeusler and D.M. Mason, The asymptotic distribution of trimmed sums. Ann. Probab. 16 (1988), 672-699.

[13] S. Csörgő, E. Haeusler and D.E. Mason, The quantile-transform approach to the asymptotic distribution of modulus trimmed sums. Sums, Trimmed Sums and Extremes (Hahn, M. G., Mason, D.M., Weiner, D.C., eds) 337-354, Progr. Probab. 23 (1991), Birkhäuser, Boston.

[14] M. Csörgő and L. Horváth, Asymptotic representations of self-normalized sums. Probab. Math. Statist. 9 (1988), 15–24.

[15] M. Csörgő and L. Horváth, Weighted Approximations in Probability and Statistics. Wiley & Sons, New York, 1993.

[16] M. Csörgő, B. Szyszkowicz and Q. Wang, Donsker's theorem for self-normalized partial sums processes. *Ann. Probab.* **31** (2003a), 1228-1240.

[17] M. Csörgő, B. Szyszkowicz and Q. Wang, Darling-Erdős theorems for self-normalized sums. *Ann. Probab.* **31** (2003b), 676-692.

[18] H.E. Daniels, Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650 (1954).

[19] H.E. Daniels and G.A. Young, Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169-179 (1991).

[20] D.A. Darling, The influence of the maximum term in the addition of independent random variables. Trans. Amer. Math. Soc. 73 (1952), 95–107.

[21] A. Dembo and Q.M. Shao, Self-normalized large deviations in vector spaces. In: *Progress in Probability* (Eberlein, Hahn, Talagrand, eds) Vol. **43** (1998a), 27-32

[22] A. Dembo and Q.M. Shao, Self-normalized moderate deviations and LILs *Stochastic Process. Appl.* **75** (1998b), 51-65.

[23] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications. Jones and Bartlett, Boston, 1992.

[24] M.D. Donsker and S.R.S Varadhan, Large deviations for noninteracting particle systems. J. Stat. Physics 46 (1987), 1195–1232.

[25] E. Gine,, F. Götze and D.M. Mason, When is the Student t-statistic asymptotically standard normal? Ann. Probab. 25 (1997), 1514-1531.

[26] P.S.Griffin and J. Kuelbs, Self-normalized laws of the iterated logarithm. Ann. Probab. 17 (1989), 1571–1601.

[27] P.S. Griffin and D.M. Mason, On the asymptotic normality of self-normalized sums. Math. Proc. camb. Phil. Soc. 109 (1991), 597-610.

[28] P.S. Griffin and W.E. Pruitt, Asymptotic normality and subsequential limits of trimmed sums. Ann. Probab. 17 (1989), 1186-1219.

[29] P.S. Griffin and W.E. Pruitt, Weak convergence of trimmed sums. Sums, Trimmed Sums and Extremes (Hahn, M. G., Mason, D.M., Weiner, D.C., eds) 55-80, Progr. Probab. 23 (1991), Birkhäuser, Boston.

[30] M.G. Hahn, J. Kuelbs and D.C. Weiner, The asymptotic joint distribution of self-normalized censored sums and sums of squares. Ann. Probab. 18 (1990), 1284-1341.

[31] M.G. Hahn, J. Kuelbs and D.C. Weiner, Asymptotic behavior of partial sums: a more robust approach via trimming and self-normalization. Sums, Trimmed Sums and Extremes (Hahn, M. G., Mason, D.M., Weiner, D.C., eds) 1-54, Progr. Probab. 23 (1991), Birkhäuser, Boston.

[32] X. He and Q.M. Shao, Bahadur efficiency and robustness of studentized score tests. Ann. Inst. Statist. Math. 48 (1996), 295-314.

[33] Z.S. Hu and Q.M. Shao (2005). Self-normalized weak invariance principle in the domain of a stable law (in preparation)

[34] J.L. Jensen, *Saddlepoint Approximations.* Oxford University Press, New York, 1995.

[35] B.Y. Jing, Q.M. Shao and Q.Y. Wang, Self-normalized Cramér type large deviations for independent random variables. *Ann. Probab.* (with B. Y. Jing and Q.Y. Wang) **31** (2003), 2167-2215.

[36] B.Y. Jing, Q.M. Shao and W. Zhou, Saddlepoint approximation for Student's t-statistic with no moment conditions. *Ann. Statist.* **32** (2004), 2679-2711.

[37] R. LePage, M. Woodroofe and J. Zinn, Convergence to a stable distribution via order statistics. Ann. Probab. 9 (1981), 624-632.

[38] Yu. V. Linnik, Limit theorems for sums of independent random variables, taking account of large deviations. Theory Probab. Appl. 7 (1962), 121-134.

[39] B.F.Logan, C.L. Mallows, S.O.Rice, and L.A. Shepp, Limit distributions of self-normalized sums. Ann. Probab. 1 (1973), 788–809.

[40] R. Lugannani and S. Rice, Saddlepoint approximations for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12**, 475-490 (1980).

[41] R. Maller, Asymptotic normality of trimmed means in higher dimensions. Ann. Probab. 16 (1988), 1608-1622.

[42] R. Maller, A review of some asymptotic properties of trimmed sums of multivariate data. Sums, Trimmed Sums and Extremes (Hahn, M. G., Mason, D.M., Weiner, D.C., eds) 179-214, Progr. Probab. 23 (1991), Birkhäuser, Boston.

[43] S.V. Nagaev, The Berry-Esseen bound for self-normalized sums. *Siberian Adv. Math.* **12** (2002), 79-125.

[44] V. de la Pena, M. Klass and T.Z. Lai, Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.* **32** (2004), 1902–1933.

[45] V.V. Petrov, Sums of Independent Random Variables. Springer-Verlag, New York, 1975.

[46] Q.M. Shao, On a problem of Csörgő and Révész. Ann. Probab. 17 (1989), 809-812.

[47] Q.M. Shao, Self-normalized large deviations. Ann. Probab. 25 (1997), 285-328.

[48] Q.M. Shao, Cramér-type large deviation for Student's $t$ statistic. *J. Theorect. Probab.* **12**, 387-398 (1999).

[49] Q.M. Shao, An explicit Berry-Esseen bound for Student's t-statistic via Stein's method. In: *Stein's Method and Applications* (Editors: A.D. Barbour, L.H.Y. Chen). Lecture Notes Series, IMS, NUS, Vol. **5** (2005), pp.143-155.

[50] V. Strassen, A converse to the law of the iterated logarithm. Z. Wahrsch. verw. Gebiete 4 (1966), 265–268.

[51] D.W. Stroock, An Introduction to the Theory of Large Deviations. Springer-Verlag, Berlin, 1984.

[52] Q. Wang and B.Y. Jing, An exponential non-uniform Berry-Esseen bound for self-normalized sums. *Ann. Probab.* **27**, 2068-2088 (1999).