Title
Stein's method for nonlinear statistics: A brief survey and recent progress

Authors
Shao, Qi-Man
Zhang, Kan
Zhou, Wen-Xin

Publication Date
2016

DOI
10.1016/j.jspi.2015.06.008

Peer reviewed

Journal of Statistical Planning and Inference 168 (2016) 68–89

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Stein's method for nonlinear statistics: A brief survey and recent progress

Qi-Man Shao [a,*], Kan Zhang [b], Wen-Xin Zhou [c]

[a] Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
[b] Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong
[c] School of Mathematics and Statistics, University of Melbourne, Parkville, VIC, 3010, Australia

ARTICLE INFO

Article history:
Received 11 May 2014
Accepted 19 June 2015
Available online 7 September 2015

Keywords:
Berry–Esseen bound
Cramér-type moderate deviation
Stein's method
Nonlinear statistics
U-statistics
Studentized statistics

ABSTRACT

Stein's method is a powerful tool for proving central limit theorems along with explicit error bounds in probability theory, where uniform and non-uniform Berry–Esseen bounds spark general interest. Nonlinear statistics, typified by Hoeffding's class of *U*-statistics, *L*-statistics, random sums and functions of nonlinear statistics, are building blocks in various statistical inference problems. However, because the standardized statistics often involve unknown nuisance parameters, the Studentized analogues are most commonly used in practice. This paper begins with a brief review of some standard techniques in Stein's method, and their applications in deriving Berry–Esseen bounds and Cramér moderate deviations for nonlinear statistics, and then using the concentration inequality approach, establishes Berry–Esseen bounds for Studentized nonlinear statistics in a general framework. As direct applications, sharp Berry–Esseen bounds for Studentized *U*- and *L*- statistics are obtained.

1. Introduction

The classical approach to the central limit theorem and the accuracy of approximations for independent random variables rely heavily on Fourier transform methods. However, the use of Fourier methods is highly limited without an independence structure, which makes it far less possible to capture the explicit bounds for the accuracy of approximations. In 1972, Charles Stein introduced a novel technique, now known as Stein's method, for normal approximation. The method works for both independent and dependent random variables. The method also provides bounds of approximation accuracy. Extensive applications of Stein's method to obtain uniform and non-uniform Berry–Esseen-type bounds for independent and dependent random variables can be found in, for example, Diaconis (1977), Baldi et al. (1989), Barbour (1990), Dembo and Rinott (1996), Goldstein and Reinert (1997), Chen and Shao (2001, 2004, 2007), Chatterjee (2008), Nourdin and Peccati (2009) and Chen and Fang (2011). In addition to the traditional study of Berry–Esseen bounds, new developments to Stein's method have triggered a series of research on Cramér-type moderate deviations, which address the relative error of two tail probabilities. See, for example, Raič (2007), Chen et al. (2013) and Shao and Zhou (in press), among others. Various extensions of Stein's idea have been applied to many other probability approximations, most notably to Poisson, Poisson process, compound Poisson, binomial approximations and more recently to multivariate, combinatorial and discretized normal approximations. Stein's method has also found diverse applications in a wide range of fields, see for example,

* Corresponding author.
E-mail address: qmshao@cuhk.edu.hk (Q.-M. Shao).

http://dx.doi.org/10.1016/j.jspi.2015.06.008
0378-3758/

Arratia et al. (1990), Barbour et al. (1992) and Chen (1993). Expositions of Stein's method and its applications in normal and other distributional approximations can be found in Diaconis and Holmes (2004), Barbour and Chen (2005). We also refer to Chen et al. (2011) a thorough coverage of the method's fundamentals and recent developments in both theory and applications.

The paper is organized as follows. In the next section, we give a brief review on recent developments on Stein's method. In Section 3, we present the main results in this paper, the Berry–Esseen bounds and Cramér type moderate deviations for Studentized nonlinear statistics. Applications to Studentized U-statistics and L-statistics are discussed in Section 4. The proofs of the main results are in Section 5, while other technical proofs are postponed to Appendix.

Notation. For the convenience, we summarize here some of the standard notations used throughout this paper. For any real-valued random variable X, let $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ for $p \geq 1$. For any real number a and b, set $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For any event E, denote by $I(E)$ the corresponding indicator function. Letters C, c will denote absolute constants whose values may change from line to line.

2. A brief survey on Stein's method

In this section, we briefly revisit Stein's method and summarize several latest results in this area. Due to the limit of space, we only focus on some of the fundamental techniques that are required for proving the results in this paper. We refer to Ross (2011) and Chatterjee (in press), respectively, for an elaborate and an advanced survey of Stein's method from its origin to recent developments.

2.1. Stein's equation

Let Z be a standard normal $N(0, 1)$ random variable and let $\mathcal{C}_{bd}$ be the class of bounded, continuous and piecewise differentiable functions $f : \mathbb{R} \mapsto \mathbb{R}$ satisfying $\mathbb{E}|f'(Z)| < \infty$. Stein's method for normal approximation rests on the following characterization.

Lemma 2.1. *Let W be a real-valued random variable. Then W follows a standard normal distribution if and only if*

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

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

The proof of necessity is essentially a direct consequence of integration by parts. For the sufficiency, let $f(w) := f_x(w)$ be the solution to the equation

$$f'(w) - wf(w) = I(w \leq x) - \Phi(x), \tag{2.2}$$

where x is a fixed number and $\Phi(x)$ is the standard normal distribution function. Indeed, Stein (1972) showed that a bounded solution always exists and can be written as

$$f_x(w) = \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)\{1-\Phi(x)\} & \text{if } w \leq x, \\ \sqrt{2\pi}e^{w^2/2}\Phi(x)\{1-\Phi(w)\} & \text{if } w \geq x. \end{cases} \tag{2.3}$$

Clearly, $f_x \in \mathcal{C}_{bd}$ and has the following properties (Chen et al., 2011): for any real w, u and v,

$$0 < f_x(w) \leq \min\big(\sqrt{2\pi}/4, 1/|x|\big), \tag{2.4}$$

and

$$|f_x'(w)| \leq 1, \qquad |f_x'(w) - f_x'(v)| \leq 1. \tag{2.5}$$

In fact, Eq. (2.2) is a particular case of the more general Stein equation

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z), \tag{2.6}$$

where h is a given real-valued measurable function with $\mathbb{E}|h(Z)| < \infty$. Similar to (2.3), the solution $f = f_h$ is given by

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

If h is bounded, then

$$\|f_h\| \leq \sqrt{\pi/2}\,\|h - \mathbb{E}h(Z)\| \leq 2\|h\|, \tag{2.7}$$

and

$$\|f_h'\| \le 2\|h - \mathbb{E}h(Z)\| \le 4\|h\|, \tag{2.8}$$

where $\|\cdot\|$ denotes the sup-norm. If h is absolutely continuous, then

$$\|f_h\| \le 2\|h'\|, \qquad \|f_h'\| \le \|h'\|, \qquad \|f_h''\| \le 2\|h'\|. \tag{2.9}$$

2.2. Normal approximation for smooth functions and Berry–Esseen bounds

Let $T := T_n$ be the random variable of interest. In many applications, it is useful to have a good estimate on $\mathbb{E}h(T) - \mathbb{E}h(Z)$. In view of (2.6), this is actually equivalent to estimate $\mathbb{E}f_h'(T) - \mathbb{E}Tf_h(T)$, which in some cases proves to be much easier to deal with. When T is a standardized sum of independent or locally dependent random variables, Stein's method has been successfully applied to prove both uniform and non-uniform Berry–Esseen bounds (Chen and Shao, 2001, 2004). A key idea of Stein's method is to rewrite $E\{Tf(T)\}$ as close as possible to $Ef'(T)$. Following Chen et al. (2011), we say that T satisfies a general framework of Stein's identity if there exist some random function $\hat{K}(u)$ and a "negligible" random variable R, such that

$$\mathbb{E}\{Tf(T)\} = \mathbb{E}\int_{-\infty}^{\infty} f'(T+u)\hat{K}(u)\,du + \mathbb{E}\{Rf(T)\} \tag{2.10}$$

for all absolutely continuous functions f whenever all the above expectations exist. The following theorem provides the normal approximation for smooth functions.

Theorem 2.1. *Let h be absolutely continuous with $\|h'\| < \infty$ and $\mathcal{F}$ be any σ-algebra containing $\sigma(T)$. Then, as long as (2.10) is satisfied,*

$$|\mathbb{E}h(T) - \mathbb{E}h(Z)| \le \|h'\|(\mathbb{E}|1-\hat{K}_1| + 2\mathbb{E}\hat{K}_2 + 2\mathbb{E}|R|), \tag{2.11}$$

where

$$\hat{K}_1 = E\left\{\int_{-\infty}^{\infty} \hat{K}(u)\,du \,\middle|\, \mathcal{F}\right\} \quad \text{and} \quad \hat{K}_2 = \int_{-\infty}^{\infty} |u\hat{K}(u)|\,du.$$

Chen et al. (2011) provided four different approaches to construct $\hat{K}$ in (2.10). In particular, for the exchangeable pairs approach (Stein, 1986), one constructs T' so that (T, T') is exchangeable. Suppose that there exist a constant $\alpha \in (0, 1]$ and a random variable R, such that

$$\mathbb{E}(T - T' \mid T) = \alpha(T - R). \tag{2.12}$$

Then for all f,

$$\mathbb{E}[(T-T')\{f(T)+f(T')\}] = 0,$$

provided that the expectation exists. It follows that

$$\begin{aligned}\mathbb{E}\{Tf(T)\} &= (2\alpha)^{-1}\mathbb{E}[(T-T')\{f(T')-f(T)\}] + \mathbb{E}\{Rf(T)\} \\ &= \mathbb{E}\int_{-\infty}^{\infty} f'(T+u)\hat{K}(u)\,du + \mathbb{E}\{Rf(T)\}\end{aligned} \tag{2.13}$$

for all absolutely continuous functions f whenever all expectations exist, where $\hat{K}(u) = (2\alpha)^{-1}\Delta\{I(-\Delta < u < 0) - I(0 \le u < -\Delta)\}$ and $\Delta = T - T'$. Therefore, with

$$\hat{K}_1 = (2\alpha)^{-1}\mathbb{E}(\Delta^2|\mathcal{F}) \quad \text{and} \quad \hat{K}_2 = (4\alpha)^{-1}|\Delta|^3. \tag{2.14}$$

Theorem 2.1 leads to

Theorem 2.2. *Let h be absolutely continuous with $\|h'\| < \infty$ and $\mathcal{F}$ any σ-algebra containing $\sigma(T)$, and let (T, T') be an exchangeable pair satisfying (2.12). Then*

$$|\mathbb{E}h(T) - \mathbb{E}h(Z)| \le \|h'\| \times \{\mathbb{E}|1-\hat{K}_1| + (2\alpha)^{-1}\mathbb{E}|\Delta|^3 + 2\mathbb{E}|R|\}. \tag{2.15}$$

From the L_1 bound one can derive a Berry–Esseen bound, as highlighted below:

$$\sup_{x\in\mathbb{R}} |P(T \le z) - \Phi(z)| \le 2\left\{\sup_{\|h'\|\le 1} |\mathbb{E}h(T) - \mathbb{E}h(Z)|\right\}^{1/2}. \tag{2.16}$$

Unfortunately, the upper bound in (2.16) is usually not sharp. When $\hat{K}(u)$ in (2.10) has a bounded support, Chen et al. (2011) established the following Berry–Esseen bound under the framework of (2.10).

Theorem 2.3. *Let T be any random variable and let f_x be the solution of the Stein equation (2.2) for $x \in \mathbb{R}$. Suppose that there exist random variables R_1 and $\hat{K}(u) \ge 0$, $u \in \mathbb{R}$, and constants δ and δ_1 independent of x, such that $|\mathbb{E}R_1| \le \delta_1$ and*

$$\mathbb{E}\{Tf_x(T)\} = \mathbb{E}\int_{|u|\le\delta} f_x'(T+u)\hat{K}(u)\,du + \mathbb{E}R_1. \tag{2.17}$$

Then

$$\sup_{x\in\mathbb{R}} |P(T \le x) - \Phi(x)| \le \delta(1.1 + \mathbb{E}|T\hat{K}_1|) + 2.7\mathbb{E}|1-\hat{K}_1| + \delta_1, \tag{2.18}$$

where $\hat{K}_1 = E\{\int_{|u|\le\delta_0} \hat{K}(u)\,du|T\}$. In particular, if T, T' are zero mean, unit variance exchangeable random variables satisfying (2.12) for some $\alpha \in (0,1]$ and some random variable R, and if $|\Delta| \le \delta$ for some constant δ, then

$$\sup_{x\in\mathbb{R}} |P(T \le x) - \Phi(x)| \le \delta(1.1 + \mathbb{E}|T\hat{K}_1|) + 2.7\mathbb{E}|1-\hat{K}_1| + \mathbb{E}|R|, \tag{2.19}$$

where $\hat{K}_1 = (2\alpha)^{-1}\mathbb{E}\big(\Delta^2|T)$ and $\Delta = T - T'$.

It is also noteworthy that Stein's exchangeable pairs method has been further developed to solve problems in measure concentration, in particular for deriving concentration inequalities with explicit constants for functions of dependent random variables. We refer to Chatterjee (2005, 2007) for this line of research. More recently, Mackey et al. (2014) extend Chatterjee's argument to the matrix setting and establish exponential matrix concentration inequalities.

2.3. Randomized concentration inequalities

The concentration inequality approach is a powerful technique for normal approximation by Stein's method. Chen and Shao (2001, 2004, 2007) obtain optimal uniform and non-uniform Berry–Essen bounds for independent random variables, dependent random variables under local dependence and a class of nonlinear statistics by developing uniform and non-uniform concentration inequalities. More recently, Shao (2010) developed the following exponential-type concentration inequality, making it possible to obtain Berry–Esseen bounds for Studentized statistics, including the self-normalized sums as a prototypical example (see Section 2).

Let $\xi_1, \dots, \xi_n$ be independent random variables with zero means and finite second moments. Let $W = \sum_{i=1}^n \xi_i$, and Δ_1, Δ_2 be measurable functions of $\{\xi_i\}_{i=1}^n$.

Theorem 2.4 (Shao, 2010). *Assume that there exist $c_1 > c_2 > 0$, $\delta > 0$ such that*

$$\sum_{i=1}^n \mathbb{E}\xi_i^2 \le c_1 \quad \text{and} \quad \sum_{i=1}^n \mathbb{E}\{|\xi_i| \min(\delta, |\xi_i|/2)\} \ge c_2. \tag{2.20}$$

Then, for any $\lambda \ge 0$, the following inequality

$$\begin{aligned}\mathbb{E}\{e^{\lambda W} I(\Delta_1 \le W \le \Delta_2)\} \le{}& (\mathbb{E}e^{2\lambda W})^{1/2}\exp\{-c_2^2/(16c_1\delta^2)\}\\ &+ 2c_2^{-1}e^{\lambda\delta}\Big[\mathbb{E}\{e^{\lambda W}|W_n|(|\Delta_2-\Delta_1|+2\delta)\}\\ &+ 2\sum_{i=1}^n \mathbb{E}\{e^{\lambda W^{(i)}}|\xi_i|(|\Delta_1-\Delta_1^{(i)}|+|\Delta_2-\Delta_2^{(i)}|)\}\Big]\end{aligned} \tag{2.21}$$

holds for all measurable functions $\Delta_1^{(i)}, \Delta_2^{(i)}$ as long as ξ_i and $(W^{(i)}, \Delta_1^{(i)}, \Delta_2^{(i)})$ are independent, where $W^{(i)} = W - \xi_i$.

The proof relies on the particular construction of an exponential-type testing function that is different from those in Chen and Shao (2001, 2004, 2007). We refer to Shao (2010) for full details.

All the previous results are essentially about estimating the absolute distributional error; that is, $|P(T \le x) - \Phi(x)|$. In some applications, it is more important to measure the relative error of $P(T \ge x)$ to $1 - \Phi(x)$, through the Chernoff large deviation or the Cramér-type moderate deviation. This line of work usually requires different techniques, including exponential inequalities and the method of conjugated distributions, also known as the measure transform approach which can be traced back to Harald Cramér in 1938. For this purpose, in Shao and Zhou (in press), a new randomized concentration inequality is developed to establish Cramér-type moderate deviations for general Studentized nonlinear statistics.

Using the previous notation, we assume that

$$\mathbb{E}\xi_i = 0 \quad \text{for } i = 1, 2, \ldots, n, \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E}\xi_i^2 = 1. \tag{2.22}$$

Moreover, write

$$\beta_2 = \sum_{i=1}^{n} \mathbb{E}\{\xi_i^2 I(|\xi_i| > 1)\}, \qquad \beta_3 = \sum_{i=1}^{n} \mathbb{E}\{|\xi_i|^3 I(|\xi_i| \le 1)\}. \tag{2.23}$$

Theorem 2.5 (*Shao and Zhou, in press*). *For each $1 \le i \le n$, let $\Delta_1^{(i)}$ and $\Delta_2^{(i)}$ be random variables such that ξ_i and $(\Delta_1^{(i)}, \Delta_2^{(i)}, W - \xi_i)$ are independent. Then*

$$P(\Delta_1 \le W \le \Delta_2) \le 17(\beta_2 + \beta_3) + 5\mathbb{E}|\Delta_2 - \Delta_1| + 2\sum_{i=1}^{n}[\mathbb{E}|\xi_i\{\Delta_1 - \Delta_1^{(i)}\}| + \mathbb{E}|\xi_i\{\Delta_2 - \Delta_2^{(i)}\}|]. \tag{2.24}$$

We remark that a similar result was obtained by Chen and Shao (2007) with $\mathbb{E}|W(\Delta_2 - \Delta_1)|$ instead of $\mathbb{E}|\Delta_2 - \Delta_1|$ in (2.24). However, using the term $\mathbb{E}|W(\Delta_2 - \Delta_1)|$ does not yield the sharp bound when Theorem 3.2 is applied to Studentized U-statistics. This is exactly why we need to develop the concentration inequality (2.24).

2.4. Stein's method for standardized statistics

Nonlinear statistics are building blocks in various statistical inference problems. It is known that many of them can be written as a linear statistic plus a negligible term. Typical examples include U-statistics, multi-sample U-statistics, L-statistics, random sums and functions of nonlinear statistics. More precisely, let $X_1, X_2, \ldots, X_n$ be independent random variables and let $T := T_n(X_1, \ldots, X_n)$ be a general sampling statistic of interest that can be decomposed as a standardized partial sum plus a remainder; that is,

$$T = W + D, \qquad W = \sum_{i=1}^{n} g_i(X_i), \tag{2.25}$$

where $D := D_n(X_1, \ldots, X_n) = T - W$ and $g_i := g_{n,i}$ are Borel measurable functions. Assume that

$$\mathbb{E}g_i(X_i) = 0 \quad \text{for } i = 1, \ldots, n \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E}g_i^2(X_i) = 1. \tag{2.26}$$

It is clear that if $D \xrightarrow{P} 0$, then the central limit theorem holds:

$$\sup_{x\in\mathbb{R}} |P(T \le x) - \Phi(x)| \to 0$$

provided the Lindeberg condition is satisfied, i.e. for any $\varepsilon > 0$, $\sum_{i=1}^{n} \mathbb{E}\{\xi_i^2 I(|\xi_i| > \varepsilon)\} \to 0$, where $\xi_i = g_i(X_i)$.

The rate of convergence to normality has been intensively studied in various situations. The classical Berry–Esseen (B–E) bound for sample means, i.e. $D \equiv 0$, is well-known. Bounds for other commonly used statistics are also available in the literature. For example, for U-statistics, B–E bounds were established under different sets of assumptions and in increasing generality by Bickel (1974), Chan and Wierman (1977), Callaert and Janssen (1978), Helmers and van Zwet (1982) and Ghosh (1985). For L-statistics, we refer to Bjerve (1977), Helmers (1977) and Helmers et al. (1990), among others. Results for R-statistics were provided by Hajek (1968).

It should be noted that each of the aforementioned references on B–E bounds for U-, L- and R-statistics focused on the individual structure of each statistic itself. A first unifying approach was proposed by Van Zwet (1984) for general symmetric statistics. Friedrich (1989) removed the symmetry assumption and relaxed the moment conditions. In the following, we briefly review the contributions of Chen and Shao (2007) in which both the uniform and non-uniform B–E bounds are established for general nonlinear statistics in the form of (2.25). A direct and unifying treatment is provided and the bounds are the best possible for many known statistics.

For $i = 1, \ldots, n$, we put $\xi_i = g_i(X_i)$ and assume that (2.26) is satisfied throughout the following. Let β_2, β_3 be as in (2.23), and let $\delta > 0$ satisfy

$$\sum_{i=1}^{n} \mathbb{E}\{|\xi_i| \min(\delta, |\xi_i|)\} \ge 1/2. \tag{2.27}$$

The following uniform and non-uniform Berry–Esseen theorems are Theorems 2.1 and 2.2 in Chen and Shao (2007), respectively.

Theorem 2.6. *For $i = 1, \ldots, n$, let $D^{(i)}$ be a random variable such that X_i and $(D^{(i)}, W^{(i)} = W - \xi_i)$ are independent. Then*

$$\sup_{x\in\mathbb{R}} |P(T \le x) - P(W \le x)| \le 4\delta + \mathbb{E}|WD| + \sum_{i=1}^n \mathbb{E}|\xi_i\{D - D^{(i)}\}| \tag{2.28}$$

for δ satisfying (2.27). In particular, we have

$$\sup_{x\in\mathbb{R}} |P(T \le x) - \Phi(x)| \le 6.1(\beta_2 + \beta_3) + \mathbb{E}|WD| + \sum_{i=1}^n \mathbb{E}|\xi_i\{D - D^{(i)}\}|. \tag{2.29}$$

Theorem 2.7. *Under the assumptions of Theorem 2.6 we have, for every $x \in \mathbb{R}$,*

$$|P(T \le x) - P(W \le x)| \le \gamma_x + e^{-|x|/3}\tau, \tag{2.30}$$

where

$$\gamma_x = P\{|D| > (1 + |x|)/3\} + \sum_{i=1}^n P\{|\xi_i| > (1 + |x|)/3\} + \sum_{i=1}^n P\{|W^{(i)}| > (|x| - 2)/3\}P(|\xi_i| > 1),$$

$$\tau = 22\delta + 8.5\|D\|_2 + 3.6\sum_{i=1}^n \|\xi_i\|_2 \times \|D - D^{(i)}\|_2.$$

In particular, if $\|\xi_i\|_p < \infty$ for some $2 < p \le 3$, then

$$|P(T \le x) - \Phi(x)| \le P\{|D| > (1 + |x|)/3\} + C(1 + |x|)^{-p}\left\{\|D\|_2 + \sum_{i=1}^n \|\xi_i\|_2 \times \|D - D^{(i)}\|_2 + \sum_{i=1}^n \|\xi_i\|_p\right\}. \tag{2.31}$$

Remark 2.1. For readers' convenience, we list several choices of δ such that condition (2.27) is satisfied:

- If $\beta_2 + \beta_3 \le \frac{1}{2}$, then (2.27) is satisfied with $\delta = \frac{1}{2}(\beta_2 + \beta_3)$;
- Let $\delta > 0$ be such that $\sum_{i=1}^n \mathbb{E}\{\xi_i^2 I(|\xi_i| > \delta)\} \le \frac{1}{2}$, then (2.27) holds;
- Assume $\mathbb{E}|\xi_i|^p < \infty$ for $p > 2$. Then (2.27) is satisfied with

$$\delta = \left\{\frac{2(p-2)^{p-2}}{(p-1)^{p-1}}\sum_{i=1}^n \mathbb{E}|\xi_i|^p\right\}^{1/(p-2)}.$$

Remark 2.2. The choices of $D^{(i)}$ and g_i are rather flexible. The most common approach in the literature is to choose $D^{(i)} = D_n(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and $g_i(x) = \mathbb{E}(T|X_i = x)$, which is also the main choice in this paper.

The main idea of the proof for bounds (2.28) and (2.30) comes from the simple observation that, for every $x \in \mathbb{R}$,

$$-P(x - |D| \le W \le x) \le P(T \le x) - P(W \le x) \le P(x \le W \le x + |D|).$$

This motivates us to develop uniform and non-uniform randomized concentration inequalities via Stein's method.

Theorems 2.6 and 2.7 provide a very general framework. We refer to Chen and Shao (2007, Sect. 3) for further applications to U-, multi-sample U-, L-statistics, random sums and functions of nonlinear statistics. A more specific statistical implication arises in the study of certain properties of the Pitman asymptotic relative efficiency (ARE) between Pearson's, Kendall's and Spearman's correlation coefficients. It is well known that the standard expression for the Pitman ARE is applicable when the distribution of the corresponding test statistics is close to normality uniformly over a neighborhood of the null set of distributions (Noether, 1955). Such uniform closeness can usually be provided by B–E bounds. In fact, Kendall's and Spearman's correlation coefficients are instances of U-statistics and the Pearson statistic can be regarded as a function of sums of independent random vectors. We refer to Pinelis and Molzon (2009) for further extensions of Theorems 2.6 and 2.7 and B–E bounds for smooth nonlinear functions of sums of independent random vectors.

3. Stein's method for Studentized statistics

Since the standardized statistics often involve certain unknown nuisance parameters, the Studentized analogues are most commonly used in practice. We are thus motivated to consider the following Studentized (self-normalized) counterpart of T,

$$T_{SN} = \frac{W + D_{1n}}{(1 + D_{2n})^{1/2}}. \tag{3.1}$$

Here, both D_{1n} and D_{2n} are measurable functions of $\{X_1, \ldots, X_n\}$. Examples satisfying (3.1) include Student's t-statistic, Studentized U- and L-statistics and Studentized functions of the sample mean. See, for example, Wang et al. (2000).

Without loss of generality, we assume that the Studentized statistic T_{SN} of interest converges in distribution to the standard normal. As mentioned in the beginning, there are two ways to measure the error of the normal approximation. One is to study the absolute error in the central limit theorem via the Berry–Esseen bounds or Edgeworth expansions (Petrov, 1975). The other way is to evaluate the relative error of the tail probability of T_{SN} against the tail probability of its limiting distribution, i.e. $1-\Phi(\cdot)$. One of the typical results in this direction is the Cramér-type moderate deviation. In this section, we study these two types of problems in general schemes. Applications to particular statistics are provided in Section 4.

3.1. Berry–Esseen bounds

Motivated by practical considerations, in this section we study the B–E bounds of optimal order on the closeness of normality for a class of Studentized statistics, say, T_{SN} given in (3.1), under minimal or near minimal conditions. The B–E bounds for Studentized statistics, including the Studentized U- and L-statistics, have been studied by various authors. We refer to Wang et al. (2000) for a general treatment.

First, let $\{X_i, \xi_i = g_i(X_i)\}_{i=1}^n$, T_{SN}, W, D_{1n} and D_{2n} be defined as above, and write

$$W_{\mathrm{b}} = W_{n,\mathrm{b}} = \sum_{i=1}^n \xi_{\mathrm{b},i}, \quad \xi_{\mathrm{b},i} = \xi_i I(|\xi_i| \le 1),\ i = 1,\dots,n. \tag{3.2}$$

For $j = 1, 2$ and $i = 1,\dots,n$, put $W^{(i)} = W - \xi_i$ and let $D_{jn}^{(i)}$ be a random variable such that X_i and $D_{jn}^{(i)}$ are independent. Moreover, for $j = 1, 2$, define truncated versions of D_{jn} and $D_{jn}^{(i)}$ as

$$\bar{D}_{jn} = D_{jn}(\xi_{\mathrm{b},1},\dots,\xi_{\mathrm{b},n}) I(|D_{jn}| \le 1/2), \qquad \bar{D}_{jn}^{(i)} = D_{jn}^{(i)}(\xi_{\mathrm{b},1},\dots,\xi_{\mathrm{b},n}) I\{|D_{jn}^{(i)}| \le 1/2\}. \tag{3.3}$$

We now present the following main result.

Theorem 3.1. *Let $2 \le p \le 3$ and $1/p + 1/q = 1$. There exists a positive absolute constant C such that*

$$\begin{aligned}\sup_{x\in\mathbb{R}} |P(T_{SN} \le x) - \Phi(x)| \le C\Big[&\beta_2 + \beta_3 + \|\bar{D}_{1n}\|_2 + \mathbb{E}\{\bar{D}_{2n}^2(1+e^{W_{\mathrm{b}}})\} + P(|D_{2n}| > 1/2) \\ &+ \sum_{j=1}^2 \sum_{i=1}^n \|\xi_i\|_p \times \|\bar{D}_{jn} - \bar{D}_{jn}^{(i)}\|_q + \sup_{x\in\mathbb{R}} |x\mathbb{E}\{\bar{D}_{2n} f_x(W_{\mathrm{b}})\}|\Big],\end{aligned} \tag{3.4}$$

where f_x denotes the unique solution of Stein equation (2.2) and β_2, β_3 are defined as in (2.23).

The proof of Theorem 3.1 is postponed to Section 5.

3.2. Cramér-type moderate deviations

The Cramér-type moderate deviation theory addresses the problem of estimating the relative error of the tail probability of T against that of the limiting distribution; that is, $P(T \ge x)/P(Z \ge x),\ x \ge 0$. The most interesting problem is to find the largest possible a_n ($a_n \to \infty$) so that

$$\frac{P(T \ge x)}{P(Z \ge x)} = 1 + o_n(1)$$

holds uniformly for $0 \le x \le a_n$. The problem is important for statistical hypothesis testing. Assume that the p-value of the test is $P(T \ge x)$. Because the exact p-value is usually unknown, it is a common practice to use the limiting tail probability $P(Z \ge x)$ to estimate the p-value. The Cramér-type moderate deviation quantifies the accuracy of the estimated p-value. The moderate deviation results have been successfully applied to multiple hypothesis tests based on t-statistics in Fan et al. (2007), Clarke and Hall (2009) and Delaigle et al. (2011). Regarding feature selection in classification and square-root Lasso for recovery of sparse signals, see Fan and Lv (2010) and Belloni et al. (2011) and the references therein. The moderate deviation result for Hotelling's T^2 statistic (Liu and Shao, 2013) is a key ingredient in controlling the false discovery rate for multiple tests on the equality of mean vectors. In summary, moderate deviation probabilities have played an important role in high dimensional data analysis.

The moderate deviation for the Student t-statistic is now well-understood. Let $X_1, X_2, \dots$ be i.i.d. non-degenerate real-valued random variables with mean μ and variance σ^2. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ be the sample mean and the sample variance, respectively. Student's t-statistic is defined by $T_n = \frac{\sqrt{n}}{s_n}(\bar{X}_n - \mu)$. Assume without loss of generality that $\mu = 0$. In contrast to the moderate deviation for the z-statistic $\sqrt{n}\bar{X}_n/\sigma$, which requires a finite moment generating function of $\sqrt{|X_1|}$, Shao (1999) established a Cramér-type moderate deviation theorem under a finite third moment. More precisely, he showed that if $\mathbb{E}|X_1|^3 < \infty$, then

$$\frac{P(T_n \ge x)}{1-\Phi(x)} = 1 + o(1) \quad \text{holds uniformly for } 0 \le x \le o(n^{1/6}). \tag{3.5}$$

Note that t-statistic is closely related to the self-normalized sum S_n/V_n via the following identity

$$T_n = \frac{S_n}{V_n}\left\{\frac{n-1}{n-(S_n/V_n)^2}\right\}^{1/2}, \tag{3.6}$$

where $S_n = \sum_{i=1}^n X_i$, $V_n^2 = \sum_{i=1}^n X_i^2$. Result (3.5) is equivalent to the moderate deviation for the self-normalized sum

$$\frac{P(S_n \ge xV_n)}{1-\Phi(x)} = 1 + o(1) \quad \text{holds uniformly for } 0 \le x \le o(n^{1/6}). \tag{3.7}$$

Result (3.7) has been further extended to independent (not necessarily identically distributed) random variables by Jing et al. (2003) under a Lindeberg type condition. In particular, for independent random variables with $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^3 < \infty$, the general result in Jing et al. (2003) gives

$$\frac{P(S_n/V_n \ge x)}{1-\Phi(x)} = 1 + O(1)(1+x)^3 \frac{\sum_{i=1}^n \mathbb{E}|X_i|^3}{\left(\sum_{i=1}^n \mathbb{E}X_i^2\right)^{3/2}} \tag{3.8}$$

for $0 \le x \le (\sum_{i=1}^n \mathbb{E}X_i^2)^{1/2}/(\sum_{i=1}^n \mathbb{E}|X_i|^3)^{1/3}$.

The past two decades have also witnessed significant progress on the development of self-normalized limit theory. For a systematic presentation of general self-normalized limit theory and its statistical application, we refer to de la Peña et al. (2009).

In this subsection, we present the main results of Shao and Zhou (in press), which extend (3.8) to a general class of Studentized nonlinear statistics in the form

$$\widehat{T}_n = \frac{W_n + D_{1n}}{Q_n(1+D_{2n})^{1/2}}, \tag{3.9}$$

where as before, $\{\xi_i := g_i(X_i)\}_{i=1}^n$ satisfies (2.26),

$$W_n = \sum_{i=1}^n \xi_i, \qquad Q_n^2 = \sum_{i=1}^n \xi_i^2$$

and $D_{jn} = D_{jn}(X_1, \ldots, X_n)$, $j = 1, 2$ are measurable functions of $\{X_i\}$. Note that the expression of $\widehat{T}_n$ is slightly different from that of T_{SN} given in (3.1). When $D_{1n} = D_{2n} = 0$, $\widehat{T}_n$ reduces to the self-normalized sum W_n/Q_n.

It is noteworthy that the proof in Jing et al. (2003) is lengthy and complicated and their method can hardly be adopted for general Studentized ratios. Enlightened by the work of Chen and Shao (2007), Shao and Zhou (in press) developed a new randomized concentration inequality (Theorem 2.5) to establish a general Cramér-type moderate deviation theorem. The proof was more transparent and direct.

For $1 \le i \le n$ and $x \ge 1$, let

$$\delta_{i,x} = x^2\mathbb{E}\xi_i^2 I\{x|\xi_i| > 1\} + x^3\mathbb{E}|\xi_i|^3 I\{x|\xi_i| \le 1\} \tag{3.10}$$

and

$$L_{n,x} = \sum_{i=1}^n \delta_{i,x}, \qquad I_{n,x} = \mathbb{E}\exp(xW_n - x^2Q_n^2/2) = \prod_{i=1}^n \mathbb{E}\exp(x\xi_i - x^2\xi_i^2/2). \tag{3.11}$$

Let $D_{1n}^{(i)}$ and $D_{2n}^{(i)}$, for each $1 \le i \le n$, be arbitrary measurable functions of $\{X_j\}_{1\le j\le n,\, j\ne i}$. For $x > 0$, set

$$R_{n,x} = I_{n,x}^{-1} \times \left(\mathbb{E}\left\{(x|D_{1n}| + x^2|D_{2n}|)e^{\sum_{j=1}^n (x\xi_j - x^2\xi_j^2/2)}\right\}\right.$$
$$\left. + \sum_{i=1}^n \mathbb{E}\left[(x|\xi_i| \wedge 1)\{|D_{1n} - D_{1n}^{(i)}| + x|D_{2n} - D_{2n}^{(i)}|\}e^{\sum_{j\ne i}^n (x\xi_j - x^2\xi_j^2/2)}\right]\right). \tag{3.12}$$

The following results are Theorems 2.1 and 2.2 in Shao and Zhou (in press).

Theorem 3.2. *Let $\widehat{T}_n$ be as in (3.9). Then there exists positive absolute constants C_1–C_4 and c_1 such that*

$$\exp\{-C_2 L_{n,x}\}\{1-\Phi(x)\}(1-C_1 R_{n,x}) \le P(\widehat{T}_n \ge x) \tag{3.13}$$

and

$$P(\widehat{T}_n \ge x) \le \{1-\Phi(x)\}(1+C_3 R_{n,x})\exp\{C_4 L_{n,x}\} + P(x|D_{1n}| > Q_n/4) + P(x^2|D_{2n}| > 1/4) \tag{3.14}$$

for all $x \ge 1$ satisfying

$$\max_{1\le i\le n} \delta_{i,x} \le 1 \tag{3.15}$$

and

$$L_{n,x} \le c_1 x^2. \tag{3.16}$$

Theorem 3.2 provides the upper and lower bounds of relative errors for $x \ge 1$. To cover the case of $0 \le x \le 1$, we present a rough estimate of the absolute error in the next theorem.

Theorem 3.3. *There exists a positive absolute constant C such that*

$$|P(\widehat{T}_n \le x) - \Phi(x)| \le C(L_{n,1+x} + \widetilde{R}_{n,x}) \tag{3.17}$$

for $x \ge 0$, where

$$\widetilde{R}_{n,x} = \mathbb{E}(|D_{1n}| + x|D_{2n}|) + \sum_{i=1}^{n} \mathbb{E}[|\xi_i\{D_{1n} - D_{1n}^{(i)}\}| + x|\xi_i\{D_{2n} - D_{2n}^{(i)}\}|]. \tag{3.18}$$

4. Applications

In this section, we apply the main results presented in Section 3.1 to several well-known examples; namely, the Studentized U- and L-statistics. The B–E bounds for these statistics have been studied by Wang et al. (2000), also in a general scheme, particularly for Studentized U-statistics where the optimal B–E bound is of order $O(n^{-1/2})$ when $\mathbb{E}|h(X_1,X_2)|^3 < \infty$. Using the explicit bound given in Theorem 3.1, we extend this result to more general heavy-tailed cases; that is, $\mathbb{E}|h(X_1,X_2)|^{2+\delta} < \infty$ for some $0 < \delta \le 1$.

4.1. U-statistics

Let $X_1, X_2, \dots, X_n$ be a sequence of independent identically distributed (i.i.d.) random variables and let $h(x,y)$ be a real-valued Borel measurable function, symmetric in its arguments with $\theta = \mathbb{E}h(X_1,X_2)$. The U-statistic of degree 2 for the estimation of θ with kernel $h(x,y)$ is

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

The U-statistic is a basic statistic and its asymptotic properties have been extensively studied in the literature. Write $g(x) = \mathbb{E}h(x,X_1)$ and assume $\sigma_g^2 = \mathrm{Var}\{g(X_1)\} > 0$, then the standardized (non-degenerate) U-statistic is given by

$$Z_n = \frac{\sqrt{n}}{2\sigma_g}(U_n - \theta). \tag{4.1}$$

A systematic presentation of U-statistics theory was given by Korolyuk and Borovskikh (1994). Studies on the uniform B–E bounds for U-statistics arose successively in Filippova (1962), Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977), Callaert and Janssen (1978), Serfling (1980), Van Zwet (1984) and Friedrich (1989). In particular, Friedrich (1989) obtained the order $O(n^{-1/2})$ when $\mathbb{E}|h(X_1,X_2)|^{5/3} < \infty$. We refer the readers to Bentkus et al. (1994) and Jing and Zhou (2005) for discussions on the necessity of this moment condition. For the non-uniform B–E bound, see Zhao and Chen (1983) and Wang (2002).

Applying Theorems 2.6 and 2.7 to the U-statistics with the kernel of degree 2 yields the following results. We refer to Theorem 3.1 of Chen and Shao (2007) for general results on U-statistics with degree $2 \le m < \frac{1}{2}n$.

Theorem 4.1. *Assume that $\theta = \mathbb{E}h(X_1,X_2) = 0$, $\sigma = \|h(X_1,X_2)\|_2 < \infty$ and $\sigma_g = \|g(X_1)\|_2 > 0$. Then,*

$$\sup_{x\in\mathbb{R}} \left| P(Z_n \le x) - P\left\{ \frac{1}{\sigma_g\sqrt{n}} \sum_{i=1}^{n} g(X_i) \le x \right\} \right| \le \frac{(1+\sqrt{2})\sigma}{\sigma_g\sqrt{2(n-1)}} + \frac{c_0}{\sqrt{n}} \tag{4.2}$$

for Z_n as in (4.1), where c_0 is a constant such that $\mathbb{E}g^2(X_1)I\{|g(X_1)| > c_0\sigma_g\} \le \frac{1}{2}\sigma_g^2$. If $\sigma_p := \|g(X_1)\|_p < \infty$ for some $2 < p \le 3$, then

$$\sup_{x\in\mathbb{R}} |P(Z_n \le x) - \Phi(x)| \le \frac{(1+\sqrt{2})\sigma}{\sigma_g\sqrt{2(n-1)}} + \frac{6.1\sigma_p^p}{\sigma_g^p n^{p/2-1}}, \tag{4.3}$$

and for every $x \in \mathbb{R}$,

$$|P(Z_n \le x) - \Phi(x)| \le \frac{18\sigma^2}{\sigma_g^p(1+|x|)^2(n-1)} + \frac{13.5\sqrt{2}e^{-|x|/3}\sigma}{\sigma_g\sqrt{n-1}} + \frac{C\sigma_p^p}{\sigma_g^p(1+|x|)^p n^{p/2-1}}. \tag{4.4}$$

Moreover, if $\|h(X_1, X_2)\|_p < \infty$ for some $2 < p \le 3$, then for every $x \in \mathbb{R}$,

$$|P(Z_n \le x) - \Phi(x)| \le C\left\{\frac{\|h(X_1,X_2)\|_p^p}{\sigma_g^p(1+|x|)^p\sqrt{n}} + \frac{\sigma_p^p}{\sigma_g^p(1+|x|)^p n^{p/2-1}}\right\}. \tag{4.5}$$

In the case where $\sigma_g^2 = \mathrm{Var}\{g(X_1)\}$ is unknown, we consider the Studentized U-statistic (Arvesen, 1969) defined as

$$T_n = \frac{\sqrt{n}}{2r_n}(U_n - \theta), \tag{4.6}$$

where r_n^2 is the Jackknife estimator of σ_g^2,

$$r_n^2 = \frac{(n-1)}{(n-2)^2}\sum_{i=1}^n (q_i - U_n)^2 \quad \text{with } q_i = \frac{1}{n-1}\sum_{j=1, j\ne i}^n h(X_i, X_j).$$

In contrast to the standardized U-statistics, few optimal limit theorems are available for Studentized U-statistics in the literature. A uniform B–E bound was proved in Wang et al. (2000) when $\mathbb{E}|h(X_1, X_2)|^3 < \infty$. However, a finite third moment of $h(X_1, X_2)$ may not be an optimal condition.

As a direct but non-trivial consequence of Theorem 3.1, we establish the following uniform bound for Studentized U-statistic T_n.

Theorem 4.2. *Assume that $\theta = \mathbb{E}h(X_1, X_2) = 0$, $\|h(X_1, X_2)\|_2 < \infty$ and for some $2 < p \le 3$, $\sigma_p := \|g(X_1)\|_p < \infty$. Set $g(x) = \mathbb{E}h(x, X_1)$ for $x \in \mathbb{R}$ and $\sigma_g = \|g(X_1)\|_2 > 0$. Then there exists an absolute constant $C > 0$ such that, for all $n \ge 3$,*

$$\sup_{x\in\mathbb{R}} |P(T_n \le x) - \Phi(x)| \le C\left\{\frac{\|h(X_1,X_2)\|_2}{\sigma_g\sqrt{n}} + \frac{\sigma_p^p}{\sigma_g^p n^{p/2-1}}\right\}. \tag{4.7}$$

Remark 4.1. The uniform upper bound given in (4.7) is more general than the one provided in Wang et al. (2000). In particular, when $p = 3$, the right-hand side of (4.7) is of order

$$n^{-1/2}\{\sigma_g^{-1}\|h(X_1,X_2)\|_2 + \sigma_g^{-3}\|g(X_1)\|_3^3\},$$

which is in line with (4.3) and is slightly different from that in Wang et al. (2000), i.e. $n^{-1/2}\sigma_g^{-3}\|h(X_1, X_2)\|_3^3$.

Remark 4.2. It remains an open question whether the moment conditions in Theorem 4.2 can be further weakened to $\|g(X_1)\|_3 < \infty$, $\|h(X_1, X_2)\|_{\frac{5}{3}} < \infty$ and $\sigma_g^2 > 0$, while our result requires $\|h(X_1, X_2)\|_2 < \infty$ and provides the explicit rate of convergence.

4.2. L-statistics

Let $X_1, \ldots, X_n$ be i.i.d. real random variables with distribution F. Denote by F_n the empirical distribution; that is, $F_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$. Let $J(t)$ be a Borel function on $[0, 1]$. An *L-functional* is defined as

$$T(G) = \int_{-\infty}^{\infty} xJ\{G(x)\}\,dG(x), \quad G \in \mathcal{F}_0,$$

where $\mathcal{F}_0$ contains all cumulative distribution functions on $\mathbb{R}$ for which T is well-defined. The statistic $T(F_n)$ is called an L-statistic (or L-estimator) of $T(F)$. The variance of $T(F)$ can be written as

$$\sigma^2 := \sigma^2(J, F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J\{F(s)\}J\{F(t)\}F(s\wedge t)\{1 - F(s\vee t)\}\,ds\,dt.$$

Clearly, a natural estimate of σ^2 is given by $\hat{\sigma}^2 := \sigma^2(J, F_n)$.

Uniform B–E bounds for the L-statistic with J satisfying some smoothness conditions were first studied by Helmers (1977) and Helmers et al. (1990). Applying Theorems 2.6 and 2.7 yields the following uniform and non-uniform bounds for standardized L-statistics. Put

$$g(x) = \int_{-\infty}^{\infty} \{I(x \le s) - F(s)\} J\{F(s)\}\, ds.$$

The following result is Theorem 3.3 in Chen and Shao (2007).

Theorem 4.3. *Let $n \ge 4$. Assume that $\mathbb{E}X_1^2 < \infty$ and $\|g(X_1)\|_p < \infty$ for some $2 < p \le 3$. If the weight function $J(t)$ is Lipschitz of order 1 on $[0, 1]$; that is, there is a constant c_0 such that $|J(t) - J(s)| \le c_0|t - s|$ for all $0 \le s, t \le 1$. Then,*

$$\sup_{x\in\mathbb{R}} |P[\sqrt{n}\sigma^{-1}\{T(F_n) - T(F)\} \le x] - \Phi(x)| \le \frac{(1+\sqrt{2})c_0\|X_1\|_2}{\sigma\sqrt{n}} + \frac{6.1\|g(X_1)\|_p^p}{\sigma^p n^{p/2-1}}. \tag{4.8}$$

Moreover, for every $x \in \mathbb{R}$,

$$|P[\sqrt{n}\sigma^{-1}\{T(F_n) - T(F)\} \le x] - \Phi(x)| \le \frac{9c_0^2\|X_1\|_2^2}{\sigma^2(1+|x|)^2 n} + \frac{C}{(1+|x|)^p}\left\{\frac{c_0\|X_1\|_2}{\sigma\sqrt{n}} + \frac{\|g(X_1)\|_p^p}{\sigma^p n^{p/2-1}}\right\}. \tag{4.9}$$

As for Studentized L-statistics, a B–E bound was given by Helmers (1982) under a stronger moment condition $\mathbb{E}|X_1|^{4.5} < \infty$. This condition was finally relaxed to $\mathbb{E}|X_1|^3 < \infty$ by Wang et al. (2000). Applying our Theorem 3.1 also leads to a uniform bound for Studentized L-statistics whenever the third moment of $|X_1|$ is finite.

Theorem 4.4. *Let $n \ge 4$. Assume that $\mathbb{E}|X_1|^3 < \infty$ and $\sigma^2 > 0$. If the weight function $J(t)$ satisfies that $J^{(2)}(t)$ is bounded on $t \in [0, 1]$, then there exists a positive constant $C(J)$ such that*

$$\sup_{x\in\mathbb{R}} |P[\sqrt{n}\hat{\sigma}^{-1}\{T(F_n) - T(F)\} \le x] - \Phi(x)| \le C(J)\frac{\mathbb{E}|X_1|^3}{\sigma^3\sqrt{n}}. \tag{4.10}$$

Here, the constant $C(J)$ depends on J only through the quantities $c_j = \max_{1\le t\le 1} |J^{(j)}(t)|$ for $j = 0, 1, 2$.

5. Proof of Theorem 3.1

We start with some preliminary observations. Note that D_{2n} is close to 0, the key ingredient of the proof is first transforming $(1 + D_{2n})^{1/2}$ to $1 + \frac{1}{2}D_{2n}$ plus a small term, in the spirit of Taylor's expansion, and then applying the concentration inequality (5.14).

Without loss of generality, we assume $x \ge 0$. Otherwise we consider $-T_{SN}$ instead. Recall that

$$T_{SN} = \frac{W_n + D_{1n}}{(1 + D_{2n})^{1/2}}.$$

Applying the elementary inequalities

$$1 + s/2 - s^2/2 \le (1+s)^{1/2} \le 1 + s/2, \quad s \ge -1$$

to $s = D_{2n}$ and set $\Delta_{n,x} = D_{1n} - \frac{1}{2}xD_{2n}$, we have

$$\{T_{SN} \ge x\} \subseteq \{W_n + \Delta_{n,x} \ge x\} \bigcup \left\{x + x(D_{2n} - D_{2n}^2)/2 \le W_n + D_{1n} \le x + xD_{2n}/2\right\} \tag{5.1}$$

and

$$\{T_{SN} \ge x\} \supseteq \{W_n + \Delta_{n,x} \ge x\}. \tag{5.2}$$

The remainder of the proof will be formulated into two parts by deriving upper bounds for

$$|P(W_n + \Delta_{n,x} \le x) - \Phi(x)| \tag{5.3}$$

and

$$P_0 := P\{x + x(D_{2n} - D_{2n}^2)/2 \le W_n + D_{1n} \le x + xD_{2n}/2\} \tag{5.4}$$

separately, where the second part requires a randomized concentration inequality. In fact, using the standard truncation argument, we only need to estimate (5.3) and (5.4) with W_n, D_{1n} and D_{2n} replaced with W_b, $\bar{D}_{1n}$ and $\bar{D}_{2n}$ and with $|\bar{D}_{1n}| \vee |\bar{D}_{2n}| \le \frac{1}{2}$, $|\Delta_{n,x}| = |\bar{D}_{1n} - \frac{1}{2}x\bar{D}_{2n}| \le \frac{1}{2} + \frac{1}{4}x$. The difference is bounded by

$$\beta_2 + P(|D_{1n}| > 1/2) + P(|D_{2n}| > 1/2).$$

5.1. First part of the proof

In this subsection we derive an upper bound on $P(W_b+\Delta_{n,x}\le x)-\Phi(x)$ for $x\ge 0$ fixed. For brevity of notation, we use $f:=f_x$ to denote the solution of (2.2) when there is no ambiguity. For each i, set $W_b^{(i)}=W_b-\xi_{b,i}$ and define

$$k_{b,i}(t)=\mathbb{E}[\xi_{b,i}\{I(0\le t\le \xi_{b,i})-I(\xi_{b,i}\le t\le 0)\}]. \tag{5.5}$$

It is straightforward to verify that $k_{b,i}(t)\ge 0$ for all $t\in\mathbb{R}$ and

$$\int_{-\infty}^{\infty}k_{b,i}(t)\,dt=\mathbb{E}\xi_{b,i}^2,\qquad \int_{-\infty}^{\infty}|t|k_{b,i}(t)\,dt=\frac{1}{2}\mathbb{E}|\xi_{b,i}|^3. \tag{5.6}$$

By Stein's equation (2.2) we can rewrite $P(W_b+\Delta_{n,x}\le x)-\Phi(x)$ as

$$\begin{aligned}
&\mathbb{E}f'(W_b+\Delta_{n,x})-\mathbb{E}(W_b+\Delta_{n,x})f(W_b+\Delta_{n,x})\\
&\quad=\mathbb{E}f'(W_b+\Delta_{n,x})-\mathbb{E}\{W_bf(W_b+\Delta_{n,x})\}\\
&\qquad-\mathbb{E}[\Delta_{n,x}\{f(W_b+\Delta_{n,x})-f(W_b)\}]-\mathbb{E}\{\Delta_{n,x}f(W_b)\}\\
&\quad=\left[\sum_{i=1}^{n}\mathbb{E}\int_{-1}^{1}\{f'(W_b+\Delta_{n,x})-f'(W_b^{(i)}+\Delta_{n,x}^{(i)}+t)\}k_{b,i}(t)dt\right] &(5.7)\\
&\qquad+\left\{\beta_2\mathbb{E}f'(W_b+\Delta_{n,x})-\sum_{i=1}^{n}\mathbb{E}\xi_{b,i}f(W_b^{(i)}+\Delta_{n,x}^{(i)})-\mathbb{E}\Delta_{n,x}f(W_b)\right\}\\
&\qquad+\left[-\sum_{i=1}^{n}\mathbb{E}\xi_{b,i}\{f(W_b+\Delta_{n,x})-f(W_b+\Delta_{n,x}^{(i)})\}\right]\\
&\qquad+\left\{-\mathbb{E}\Delta_{n,x}\times\int_0^{\Delta_{n,x}}f'(W_b+t)dt\right\}\\
&\quad:=R_1+R_2+R_3+R_4. &(5.8)
\end{aligned}$$

The estimates of R_1,R_2,R_3 and R_4 are presented in the following proposition. The proof is postponed to Section 5.3.

Proposition 5.1. *Let $2\le p\le 3$ and $1/p+1/q=1$. There exists a universal constant $C>0$ such that*

$$R_1\le C\left\{\beta_2+\beta_3+\sum_{j=1}^{2}\sum_{i=1}^{n}\|\xi_{b,i}\|_p\times\|\bar D_{jn}-\bar D_{jn}^{(i)}\|_q\right\}, \tag{5.9}$$

$$R_2\le 2\beta_2+\|\bar D_{1n}\|_2+x|\mathbb{E}\{\bar D_{2n}f_x(W_b)\}|, \tag{5.10}$$

$$R_3\le C\sum_{j=1}^{2}\sum_{i=1}^{n}\|\xi_{b,i}\|_p\times\|\bar D_{jn}-\bar D_{jn}^{(i)}\|_q \tag{5.11}$$

and

$$R_4\le C\left\{\|\bar D_{1n}\|_2^2+\|\bar D_{2n}\|_2^2+\mathbb{E}(\bar D_{2n}^2e^{W_b})\right\}. \tag{5.12}$$

Putting (5.9)–(5.12) together, we conclude that

$$\begin{aligned}
|P(W_n+\Delta_{n,x}\le x)-\Phi(x)|\le C\Bigg[&\beta_2+\beta_3+\|\bar D_{1n}\|_2+\|\bar D_{2n}\|_2^2+\mathbb{E}(\bar D_{2n}^2e^{W_b})\\
&+\sum_{j=1}^{2}\sum_{i=1}^{n}\|\xi_{b,i}\|_p\times\|\bar D_{jn}-\bar D_{jn}^{(i)}\|_q+|x\mathbb{E}\{\bar D_{2n}f_x(W_b)\}|\Bigg].
\end{aligned} \tag{5.13}$$

5.2. Second part of the proof

In this subsection, we estimate the probability P_0 given in (5.4). The main technical tool is a randomized concentration inequality (Theorem 2.4) developed by Shao (2010). To this end, put $\widetilde\Delta_{n,x}=\Delta_{n,x}+\frac{1}{2}x\bar D_{2n}^2$ so the target probability (5.4) reads

$$P_0=P(x-\widetilde\Delta_{n,x}\le W_b\le x-\Delta_{n,x}).$$

Set $\delta:=\frac{1}{2}(\beta_2+\beta_3)$. We will give the upper bound of P_0 by using the following lemma, which is Theorem 2.4 with a slight modification.

Lemma 5.1. *If* $\beta_2 + \beta_3 \le \frac{1}{2}$, *then*

$$\begin{aligned}
&\mathbb{E}\{e^{W_b/2} I(x - \widetilde{\Delta}_{n,x} \le W_b \le x - \Delta_{n,x})\} \\
&\le (\mathbb{E}e^{W_b})^{1/2} e^{-1/(32\delta^2)} + 4e^{\delta/2}\Big(\mathbb{E}\{e^{W_b/2}|W_b|(|\widetilde{\Delta}_{n,x} - \Delta_{n,x}| + 2\delta)\} \\
&\quad + 2\sum_{i=1}^{n} \mathbb{E}[e^{W_b^{(i)}/2}|\xi_{b,i}|\{|\widetilde{\Delta}_{n,x} - \widetilde{\Delta}_{n,x}^{(i)}| + |\Delta_{n,x} - \Delta_{n,x}^{(i)}|\}]\Big).
\end{aligned} \tag{5.14}$$

Because $|\widetilde{\Delta}_{n,x}| \le \frac{1}{2}(1+x)$, the left-hand side of (5.14) is bounded from below by

$$e^{(x-1)/4} P(x - \widetilde{\Delta}_{n,x} \le W_b \le x - \Delta_{n,x}).$$

For the right-hand side of (5.14), by Lemma 5.4 and the assumption that $\delta \le \frac{1}{4}$, we have

$$\mathbb{E}e^{W_b} \le \exp(e-2) \quad \text{and} \quad e^{-1/(32\delta^2)} \le 4e^{-1/2}\delta \le e^{-1/2}.$$

Moreover, by definition,

$$\mathbb{E}\{e^{W_b/2}|W_b(\widetilde{\Delta}_{n,x} - \Delta_{n,x})|\} \le \frac{x}{2}\mathbb{E}(\bar{D}_{2n}^2|W_b|e^{W_b/2}) \le x\mathbb{E}\bar{D}_{2n}^2(1+e^{W_b}) \tag{5.15}$$

and

$$\begin{aligned}
&\mathbb{E}[e^{W_b^{(i)}/2}|\xi_{b,i}|\{|\widetilde{\Delta}_{n,x} - \widetilde{\Delta}_{n,x}^{(i)}| + |\Delta_{n,x} - \Delta_{n,x}^{(i)}|\}] \\
&\le C\mathbb{E}[e^{W_b^{(i)}/2}|\xi_{b,i}|\{|\bar{D}_{1n} - \bar{D}_{1n}^{(i)}| + x|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}|\}] \\
&\le C\|e^{W_b^{(i)}/2}\xi_{b,i}\|_p\{\|\bar{D}_{1n} - \bar{D}_{1n}^{(i)}\|_q + x\|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}\|_q\} \\
&\le C\|\xi_{b,i}\|_p\{\|\bar{D}_{1n} - \bar{D}_{1n}^{(i)}\|_q + x\|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}\|_q\}.
\end{aligned}$$

Together with (5.15) and Lemma 5.1, this implies

$$P_0 \le C\Big[\beta_2 + \beta_3 + \mathbb{E}\{\bar{D}_{2n}^2(1+e^{W_b})\} + \sum_{j=1}^{2}\sum_{i=1}^{n}\|\xi_{b,i}\|_p \times \|\bar{D}_{jn} - \bar{D}_{jn}^{(i)}\|_q\Big]. \tag{5.16}$$

Combining (5.1), (5.2), (5.13) and (5.16) yields (3.4). □

5.3. *Proof of Proposition 5.1*

We begin by collecting a few useful technical lemmas. The first one follows directly from Theorem 2.4, which serves as a key tool to bound $|R_1|$. Recall that $W_b = \sum_{i=1}^n \xi_i I(|\xi_i| \le 1)$.

Lemma 5.2. *Let* $1/p + 1/q = 1$ *and* $2 \le p \le 3$. *Then for every* $a \in \mathbb{R}$,

$$P(D_{1n} \le W_b \le D_{2n}, D_{1n} \ge a) \le Ce^{-a/2}\Big\{\|D_{2n} - D_{1n}\|_2 + \beta_2 + \beta_3 + \sum_{j=1}^{2}\sum_{i=1}^{n}\|\xi_{b,i}\|_p \times \|D_{jn} - D_{jn}^{(i)}\|_q\Big\},$$

where $D_{1n}^{(i)}$ *and* $D_{2n}^{(i)}$ *are arbitrary measurable functions of* $\{\xi_j\}_{j\neq i}$.

To bound the remaining terms, we need the following properties of functions f_x and f_x'. The proof is based on direct computations from the explicit formula of f_x and thus is omitted.

Lemma 5.3. *For arbitrary* $x \ge 1$, *let* f_x *be the solution of* (2.2) *whose explicit form is given in* (2.3). *Then we have*

$$f_x(w) \le \begin{cases} 2.1e^{-x}, & w \le x-1, \\ 1, & w > x-1, \end{cases} \tag{5.17}$$

$$0 \le f_x'(w) \le \begin{cases} e^{1/2-x}, & w \le x-1, \\ 1, & x-1 < w \le x, \\ (1+x^2)^{-1}, & w > x. \end{cases} \tag{5.18}$$

Moreover, let $g(w) = \{wf_x(w)\}'$. *Then* $g(w) \geq 0$ *for all* $w \in \mathbb{R}$,

$$g(w) \leq \begin{cases} 2\{1-\Phi(x)\}, & w \leq 0, \\ 2(1+w^3)^{-1}, & w \geq x, \end{cases}$$

and $g(w)$ *is increasing for* $0 \leq w < x$ *with*

$$g(x-1) \leq xe^{1/2-x} \quad \text{and} \quad g(x-) \leq x\{2-\Phi(x)\} + 2(1+x^3)^{-1} \leq 2(1+x).$$

Lemma 5.4. *For* $a > 0$,

$$\mathbb{E}e^{aW_{\mathrm{b}}} \leq \exp(e^a - 1 - a),$$

and for $x > 1$ *and* $t \in \mathbb{R}$,

$$\mathbb{E}f_x(W_{\mathrm{b}}+t) \leq 2.1e^{-x} + 5.6\exp(t-x), \qquad \mathbb{E}f_x'(W_{\mathrm{b}}+t) \leq e^{1/2-x} + 5.6\exp(t-x).$$

Proof of Proposition 5.1. For R_1, define functions $G(w) = wf(w)$ and $g(w) = G'(w) \geq 0$ for $w \in \mathbb{R}$, such that the integrand in R_1 can be written as

$$\begin{aligned} &G(W_{\mathrm{b}} + \Delta_{n,x}) - G(W_{\mathrm{b}}^{(i)} + \Delta_{n,x}^{(i)} + t) + I(W_{\mathrm{b}} + \Delta_{n,x} \leq x) - I(W_{\mathrm{b}}^{(i)} + \Delta_{n,x}^{(i)} + t \leq x) \\ &\quad = \int_{t+\Delta_{n,x}^{(i)}}^{\xi_{\mathrm{b},i}+\Delta_{n,x}} g(W_{\mathrm{b}}^{(i)} + u)\,du + I(W_{\mathrm{b}} + \Delta_{n,x} \leq x) - I(W_{\mathrm{b}}^{(i)} + \Delta_{n,x}^{(i)} + t \leq x). \end{aligned}$$

Accordingly, we have $R_1 = R_{11} + R_{12}$, where

$$\begin{aligned} R_{11} &:= \sum_{i=1}^n \int_{-1}^1 \mathbb{E}\left\{\int_{t+\Delta_{n,x}^{(i)}}^{\xi_{\mathrm{b},i}+\Delta_{n,x}} g(W_{\mathrm{b}}^{(i)} + u)\,du\right\} k_{\mathrm{b},i}(t)\,dt, \\ R_{12} &:= \sum_{i=1}^n \int_{-1}^1 \{P(W_{\mathrm{b}} + \Delta_{n,x} \leq x) - P(W_{\mathrm{b}}^{(i)} + \Delta_{n,x}^{(i)} + t \leq x)\} k_{\mathrm{b},i}(t)\,dt. \end{aligned}$$

Put $\eta_1 = t + \Delta_{n,x}^{(i)}$, $\eta_2 = \xi_{\mathrm{b},i} + \Delta_{n,x}$, then

$$\begin{aligned} |R_{11}| &\leq \sum_{i=1}^n \int_{-1}^1 \mathbb{E}\int g(W_{\mathrm{b}}^{(i)} + u) I(\eta_1 < u < \eta_2)\,du\,k_{\mathrm{b},i}(t)\,dt \\ &\quad + \sum_{i=1}^n \int_{-1}^1 \mathbb{E}\int g(W_{\mathrm{b}}^{(i)} + u) I(\eta_2 < u < \eta_1)\,du\,k_{\mathrm{b},i}(t)\,dt. \end{aligned}$$

When $x \geq 1$, using Lemma 5.2, Lemma 5.3, Hölder's inequality and the following identity

$$1 \equiv I\{W_{\mathrm{b}}^{(i)} + u \leq x-1\} + I\{W_{\mathrm{b}}^{(i)} + u > x-1, u \leq 3x/4\} + I\{W_{\mathrm{b}}^{(i)} + u > x-1, u > 3x/4\}$$

gives

$$\begin{aligned} &\mathbb{E}\int g(W_{\mathrm{b}}^{(i)} + u) I(\eta_1 < u < \eta_2)\,du \\ &\quad \leq xe^{1/2-x}\|\eta_2 - \eta_1\|_1 + 2(1+x)P(W_{\mathrm{b}}^{(i)} + 1 > x/4)^{1/p} \times \|\eta_2 - \eta_1\|_q \\ &\qquad + 2(1+x)P(\eta_2 > 3x/4)^{1/p} \times \|\eta_2 - \eta_1\|_q \\ &\quad \leq xe^{1/2-x}\|\eta_2 - \eta_1\|_1 + 2(1+x)[\{e^{1-x/4}\mathbb{E}e^{W_{\mathrm{b}}^{(i)}}\}^{1/p} + (e^{-3x/4}\mathbb{E}e^{\xi_{\mathrm{b},i}+\Delta_{n,x}})^{1/p}] \times \|\eta_2 - \eta_1\|_q. \end{aligned}$$

Together with Lemma 5.4 and the fact that $|\Delta_{n,x}| \vee |\Delta_{n,x}^{(i)}| \leq \frac{1}{2}(1+x)$, this implies that

$$\mathbb{E}\int g(W_{\mathrm{b}}^{(i)} + u) I(\eta_1 < u < \eta_2)\,du \leq Cxe^{-x/(4p)}\{\|\xi_{\mathrm{b},i}\|_q + \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q + |t|\}.$$

A similar argument holds for the case with indicator function $I(\eta_2 < u < \eta_1)$, which leads to the same bound. Plugging the above calculations into the expression of R_{11}, we get, for $x \geq 1$,

$$|R_{11}| \leq Cxe^{-x/6}\left\{\beta_2 + \beta_3 + \sum_{i=1}^n \|\xi_{\mathrm{b},i}\|_p \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q\right\}. \tag{5.19}$$

In contrast, when $0 \le x \le 1$, the function $g(w) = \{wf_x(w)\}'$ satisfies that $0 \le g(w) \le 2$ for all w and $\|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q \le \|\bar{D}_{1n} - \bar{D}_{1n}^{(i)}\|_q + \frac{1}{2}\|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}\|_q$. It follows that

$$|R_{11}| \le \beta_2 + 3\beta_3 + \sum_{i=1}^{n} \|\xi_{b,i}\|_p \times \left\{2\|\bar{D}_{1n} - \bar{D}_{1n}^{(i)}\|_q + \|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}\|_q\right\}. \tag{5.20}$$

For R_{12}, the integrand is bounded by

$$P\{x - \Delta_{n,x} \le W_b \le x - \Delta_{n,x}^{(i)} + \xi_{b,i} - t\} + P\{x - \Delta_{n,x}^{(i)} + \xi_{b,i} - t \le W_b \le x - \Delta_{n,x}\}.$$

Note that, for $|t| \le 1$,

$$x - \Delta_{n,x} \ge x/2 - 1/2, \qquad x - \Delta_{n,x}^{(i)} + \xi_{b,i} - t \ge x/2 - 5/2.$$

By Lemma 5.2, we see that the sum of the above two probabilities is bounded by some multiple of

$$e^{-x/4}\left\{|t| + \|\xi_{b,i}\|_q + \beta_2 + \beta_3 + \sum_{i=1}^{n} \|\xi_{b,i}\|_p \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q\right\},$$

which further yields, for $x \ge 0$,

$$|R_{12}| \le Ce^{-x/4}\left\{\beta_2 + \beta_3 + \sum_{i=1}^{n} \|\xi_{b,i}\|_p \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q\right\}. \tag{5.21}$$

Putting (5.19)–(5.21) together proves (5.9).

The bound (5.10) for R_2 follows directly by independence and the facts that $0 \le f \le 1$ and $|f'| \le 1$. Turning to R_3, when $0 \le x \le 1$,

$$|f(W_b + \Delta_{n,x}) - f(W_b + \Delta_{n,x}^{(i)})| \le \sum_{j=1}^{2} |\bar{D}_{jn} - \bar{D}_{jn}^{(i)}|$$

and when $x \ge 1$, it follows from (5.18) that

$$\begin{aligned} |f(W_b + \Delta_{n,x}) - f(W_b + \Delta_{n,x}^{(i)})| &= |f(W_b + \Delta_{n,x}) - f(W_b + \Delta_{n,x}^{(i)})|\{I(W_b \le x/2 - 3/2) + I(W_b > x/2 - 3/2)\} \\ &\le |\Delta_{n,x} - \Delta_{n,x}^{(i)}|\{e^{1/2-x} + I(W_b > x/2 - 3/2)\}. \end{aligned}$$

Furthermore, this implies

$$\begin{aligned} |\mathbb{E}\xi_{b,i}\{f(W_b + \Delta_{n,x}) - f(W_b + \Delta_{n,x}^{(i)})\}| &\le e^{1/2-x}\mathbb{E}|\xi_{b,i}\{\Delta_{n,x} - \Delta_{n,x}^{(i)}\}| + \mathbb{E}[|\xi_{b,i}\{\Delta_{n,x} - \Delta_{n,x}^{(i)}\}|I(W_b > x/2 - 3/2)] \\ &\le e^{1/2-x}\|\xi_{b,i}\|_p \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q \\ &\quad + \{e^{3/2-x/2}\mathbb{E}(|\xi_{b,i}|^p e^{\xi_{b,i}}) \times \mathbb{E}e^{W_b^{(i)}}\}^{1/p} \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q \\ &\le (e^{1/2-x} + e^{e/p+1/2p-x/2p})\|\xi_{b,i}\|_p \times \|\Delta_{n,x} - \Delta_{n,x}^{(i)}\|_q. \end{aligned}$$

This proves (5.11).

For the last term R_4, it follows from Lemma 5.3 that

$$\begin{aligned} \left|\mathbb{E}\Delta_{n,x}\int_0^{\Delta_{n,x}} f'(W_b + t)dt\right| &\le \{(1+x^2)^{-1} + e^{1/2-x}\}\mathbb{E}\Delta_{n,x}^2 + \mathbb{E}\Delta_{n,x}^2 I\{(x-3)/2 \le W_b \le (1+3x)/2\} \\ &\le 4\|\bar{D}_{1n}\|_2^2 + \|\bar{D}_{2n}\|_2^2 + 3x^2e^{-x/2}\mathbb{E}(\bar{D}_{2n}^2 e^{W_b}), \end{aligned}$$

whereas when $0 \le x \le 1$, it is straightforward that

$$\left|\mathbb{E}\Delta_{n,x}\int_0^{\Delta_{n,x}} f'(W_b + t)dt\right| \le \mathbb{E}\Delta_{n,x}^2 \le 2\|\bar{D}_{1n}\|_2^2 + \frac{1}{2}\|\bar{D}_{2n}\|_2^2.$$

The proof of Proposition 5.1 is now complete. □

Acknowledgments

The research of Qi-Man Shao was partially supported by Hong Kong RGC GRF 603710 and 403513. The research of Wen-Xin Zhou was supported in part by a grant from the Australian Research Council.

Appendix

Proof of Theorem 4.2. Observe that, if we put $\tilde{h} = \sigma_g^{-1}h$ and $\tilde{g} = \sigma_g^{-1}g$, then $\tilde{g}(x) = \mathbb{E}\tilde{h}(x, X_1)$ and $\tilde{g}(X_1), \ldots, \tilde{g}(X_n)$ are i.i.d. random variables with zero mean and unit variance. By the scaling invariance property of Studentized U-statistic, we can replace h and g with $\tilde{h}$ and $\tilde{g}$, respectively, which does not change the definition of T_n. For brevity of notation, we still use h and g but assume without loss of generality that $\sigma_g^2 = 1$.

Again, by the scaling invariance of T_n and the fact that

$$|P(T_n \le x) - \Phi(x)| = |P(-T_n \le -x) - \Phi(-x)|, \quad \forall x \in \mathbb{R},$$

we only need to prove that the upper bound in (4.7) holds for $\sup_{x\ge 0} |P(T_n \le x) - \Phi(x)|$. To begin with, write $\psi(x, y) = h(x, y) - g(x) - g(y)$, $\xi_i = n^{-1/2}g(X_i)$ for $i = 1, \ldots, n$ and set

$$S_n = \sum_{i=1}^n \xi_i, \qquad V_n^2 = \sum_{i=1}^n \xi_i^2, \qquad \Psi_n = \frac{1}{n-1}\sum_{i<j}\frac{\psi(X_i, X_j)}{\sqrt{n}},$$
$$\Lambda_n^2 = \sum_{i=1}^n \Psi_{n,i}^2, \qquad \Psi_{n,i} = \sum_{j=1, j\ne i}^n \frac{\psi(X_i, X_j)}{\sqrt{n}}.$$

By Hoeffding's decomposition, $\frac{\sqrt{n}}{2}U_n = S_n + \Psi_n$. Moreover, observe that $\sum_{i=1}^n (q_i - U_n)^2 = \sum_{i=1}^n q_i^2 - nU_n^2$, then we can rewrite T_n as

$$T_n = \frac{\sqrt{n}U_n}{2r_n} = T_n^* \Big/ \left\{1 - \frac{4(n-1)}{(n-2)^2}T_n^{*2}\right\}^{1/2}, \tag{A.1}$$

where

$$T_n^* = \frac{\sqrt{n}U_n}{2s_n}, \qquad s_n^2 = \frac{n-1}{(n-2)^2}\sum_{i=1}^n q_i^2. \tag{A.2}$$

By the one-to-one relationship between T_n and T_n^* as in (A.1), we have

$$\{T_n \ge x\} = \{T_n^* \ge x/\{1 + 4x^2(n-1)/(n-2)^2\}^{1/2}\}.$$

Therefore, we only need to prove the result for T_n^*, instead of T_n. To see this, note that

$$q_i = \frac{1}{n-1}\sum_{j=1, j\ne i}^n h(X_i, X_j) = \frac{\sqrt{n}}{n-1}\{(n-2)\xi_i + S_n + \Psi_{n,i}\},$$

which further leads to

$$\frac{(n-2)^2(n-1)}{n}s_n^2 = (n-2)^2V_n^2 + \Lambda_n^2 + (3n-4)S_n^2 + 2(n-2)\sum_{i=1}^n \xi_i\Psi_{n,i} + 2S_n\sum_{i=1}^n \Psi_{n,i}.$$

By the Cauchy–Schwarz inequality, the last term in the above identity is bounded by $2\sqrt{n}|S_n|\Lambda_n$. Then we can write

$$s_n^2 = \frac{n}{n-1}(V_n^2 + \delta_n), \tag{A.3}$$

where $\delta_n := \delta_n(\xi_1, \ldots, \xi_n, X_1, \ldots, X_n)$ is such that $\delta_n = \delta_{1n} + \delta_{2n}$ with

$$|\delta_{1n}| \le \frac{4(n-1)}{(n-2)^2}S_n^2 + \frac{2}{(n-2)^2}\Lambda_n^2 \quad \text{and} \quad \delta_{2n} = \frac{2}{n-2}\sum_{i=1}^n \xi_i\Psi_{n,i}. \tag{A.4}$$

Combining (A.2) and (A.3) gives

$$T_n^* = \frac{S_n + \Psi_n}{d_n\sqrt{1 + (V_n^2 - 1) + \delta_n}}, \tag{A.5}$$

where $d_n = \sqrt{\frac{n}{n-1}}$. By (A.4), we have

$$\mathbb{E}|\delta_{1n}| \le \frac{4(n-1)}{(n-2)^2} + \frac{2(n-1)}{(n-2)^2}\mathbb{E}\psi^2(X_1, X_2) = \frac{2(n-1)}{(n-2)^2}\mathbb{E}h^2(X_1, X_2).$$

Therefore, for $n \geq 3$,

$$P(|\delta_{1n}| > n^{-1/2}) \leq \frac{2\sqrt{n}(n-1)}{(n-2)^2}\mathbb{E}h^2(X_1, X_2).$$

Consequently, by (A.2) and the fact $|d_n - 1| = O(n^{-1})$, it suffices to bound

$$\sup_{x\geq 0}|P(T_n^{**} \leq x) - \Phi(x)|,$$

where in line with the notation in Theorem 3.1,

$$T_n^{**} = \frac{W + D_{1n}}{(1 + D_{2n})^{1/2}} \tag{A.6}$$

with

$$W = S_n, \qquad D_{1n} = \Psi_n, \qquad D_{2n} = (V_n^2 - 1) + \frac{1}{2\sqrt{n}} + \frac{2}{n-2}\sum_{i=1}^{n}\xi_i\Psi_{n,i}.$$

For $1 \leq \ell \leq n$, taking $D_{1n}^{(\ell)}$ and $D_{2n}^{(\ell)}$ in a natural way (cf. Chen and Shao, 2007) such that X_ℓ and $(D_{1n}^{(\ell)}, D_{2n}^{(\ell)})$ are independent and

$$D_{1n} - D_{1n}^{(\ell)} = \frac{1}{\sqrt{n}(n-1)}\sum_{i\neq\ell}\psi(X_i, X_\ell),$$

$$D_{2n} - D_{2n}^{(\ell)} = \xi_\ell^2 - \mathbb{E}\xi_\ell^2 + \frac{2}{\sqrt{n}(n-2)}\left\{\xi_\ell\sum_{j\neq\ell}\psi(X_j, X_\ell) + \sum_{i\neq\ell}\xi_i\psi(X_i, X_\ell)\right\}.$$

Moreover, set $\xi_{b,i} = \xi_i I(|\xi_i| \leq 1)$ for $i = 1, \ldots, n$, $W_b = \sum_{i=1}^n \xi_{b,i}$ and

$$\bar{D}_{2n} = D_{2n}(X_1, \ldots, X_n, \xi_{b,1}, \ldots, \xi_{b,n})I(|D_{2n}| \leq 1/2)$$

with

$$D_{2n}(X_1, \ldots, X_n, \xi_{b,1}, \ldots, \xi_{b,n}) = \sum_{i=1}^{n}\xi_{b,i}^2 - 1 + \frac{1}{2\sqrt{n}} + \frac{2}{n-2}\sum_{i=1}^{n}\xi_{b,i}\Psi_{n,i}.$$

Now with the above preparations, the final conclusion (4.7) follows from Theorem 3.1 and Lemma A.1, whose proof is postponed to the end of the Appendix.

Lemma A.1. *We have, for all $n \geq 3$,*

$$\|D_{1n}\|_2^2 \leq \frac{\|h(X_1, X_2)\|_2^2}{n-1},$$

$$\|\bar{D}_{2n}\|_2^2 \leq \frac{2\sigma_p^p}{n^{p/2-1}} + C\frac{\|h(X_1, X_2)\|_2^2}{n},$$

$$P(|D_{2n}| \geq 1/2) \leq C\left\{\frac{\sigma_p^p}{n^{p/2-1}} + \frac{\|h(X_1, X_2)\|_2^2}{n}\right\},$$

and for any $i = 1, \ldots, n$,

$$\|D_{1n} - D_{1n}^{(i)}\|_2 \leq \frac{\|h(X_1, X_2)\|_2}{\sqrt{n}(n-1)},$$

$$\|\bar{D}_{2n} - \bar{D}_{2n}^{(i)}\|_2 \leq \|\xi_{b,i}^2\|_2 + C\frac{\|h(X_1, X_2)\|_2}{n}.$$

Moreover,

$$\mathbb{E}(\bar{D}_{2n}^2 e^{W_b}) \leq C\left\{\frac{\sigma_p^p}{n^{p/2-1}} + \frac{\|h(X_1, X_2)\|_2}{\sqrt{n}}\right\}, \tag{A.7}$$

$$\sup_{x\geq 0}|x\mathbb{E}\{\bar{D}_{2n}f_x(W_b)\}| \leq C\left\{\frac{\sigma_p^p}{n^{p/2-1}} + \frac{\|h(X_1, X_2)\|_2}{\sqrt{n}}\right\}. \tag{A.8}$$

This completes the proof of Theorem 4.2. □

Proof of Theorem 4.4. Let $\psi(t) = \int_0^t J(s)\,ds$ for $0 \le t \le 1$. As in Serfling (1980, p. 265), we have

$$T(F_n) - T(F) = -\int_{-\infty}^{\infty} \{\psi(F_n(x)) - \psi(F(x))\}\,dx.$$

For brevity, we introduce the following notation

$$J_0(x) = J(F(x)), \qquad J_n(x) = J(F_n(x)),$$
$$\eta_i(x) = I_i(x) - F(x), \quad I_i(x) = I(X_i \le x), \qquad \bar{I}_i(x) = 1 - I_i(x),$$
$$g_i(X_i) = -\sigma n^{-1/2} \int \eta_i(x) J_0(x)\,dx, \qquad K(s,t) = F(s \wedge t)\{1 - F(s \vee t)\}$$

and write

$$\sigma^{-1}\sqrt{n}\{T(F_n) - T(F)\} = W_n + \Delta_n, \quad W_n = \sum_{i=1}^n g_i(X_i),$$

where

$$\Delta_n = \frac{-\sqrt{n}}{\sigma} \int [\psi(F_n(x)) - \psi(F(x)) - \{F_n(x) - F(x)\}J_0(x)]\,dx.$$

For each i, let $F_{n,i}(x) = \mathbb{E}\{F_n(x) \mid X_j, 1 \le j \le n, j \ne i\}$ and

$$\Delta_n^{(i)} = \sigma^{-1}\sqrt{n} \int [\psi(F_{n,i}(x)) - \psi(F(x)) - \{F_{n,i}(x) - F(x)\}J_0(x)]\,dx,$$

so that $\Delta_n^{(i)}$ and X_i are independent. It was already shown in Chen and Shao (2007) that

$$\sigma^2 \mathbb{E}\Delta_n^2 \le c_1^2 n^{-1} \mathbb{E}X_1^2, \qquad \sigma^2 \mathbb{E}|\Delta_n - \Delta_n^{(i)}|^2 \le 2c_1^2 n^{-2} \mathbb{E}X_1^2, \tag{A.9}$$

where $c_1 = \max_{0 \le t \le 1} |J'(t)|$. For the variance part, following Wang et al. (2000), we have $\hat{\sigma}^2/\sigma^2 = 1 + Q_n + R_n$, such that

$$\frac{\sqrt{n}}{\hat{\sigma}}\{T(F_n) - T(F)\} = \frac{W_n + \Delta_n}{(1 + Q_n + R_n)^{1/2}}, \tag{A.10}$$

where

$$Q_n = n^{-3} \sum_{i \ne j \ne k} \gamma_{i,j,k}, \quad \gamma_{i,j,k} = \xi_{i,j} + \varphi_{i,j,k},$$
$$\xi_{i,j} = \sigma^{-2} \iint J_0(s) J_0(t) \{I_i(s \wedge t)\bar{I}_j(s \vee t) - K(s,t)\}\,ds\,dt,$$
$$\varphi_{i,j,k} = \sigma^{-2} \iint J_0'(s) J_0'(t) \eta_i(t) I_j(s \wedge t) \bar{I}_k(s \vee t)\,ds\,dt,$$

and where $R_n = R_{1n} + R_{2n} + R_{3n}$ with

$$R_{1n} = 2\sigma^{-2} \iint [J_n(s) - J_0(s) - J_0'(s)\{F_n(s) - F(s)\}] \times J_0(s) K(s,t)\,ds\,dt,$$
$$R_{2n} = \sigma^{-2} \iint \{J_n(s) - J_0(s)\}\{J_n(t) - J_0(t)\} K(s,t)\,ds\,dt,$$
$$R_{3n} = \frac{1}{n^3} \sum_{j \ne k} (\xi_{j,k} + \varphi_{j,j,k} + \varphi_{k,j,k}) - \frac{1}{n\sigma^2} \iint F(s \wedge t)\{1 - F(s \vee t)\}\,ds\,dt.$$

Recall that $a_j = \max_{0 \le t \le 1} |J^{(j)}(t)|$, $j = 0, 1, 2$. Similar to the proof of Lemma A in Serfling (1980, p. 288), we can show that

$$|g_i(X_i)| \le c_0 \sigma^{-1} n^{-1/2} (|X_i| + \mathbb{E}|X_1|), \tag{A.11}$$
$$|\xi_{i,j}| \le C c_0^2 \sigma^{-2} (X_i^2 + X_j^2 + \mathbb{E}X_1^2), \tag{A.12}$$
$$|\varphi_{i,j,k}| \le C c_1^2 \sigma^{-2} (X_j^2 + X_k^2), \tag{A.13}$$

leading to

$$\mathbb{E}|R_{1n}| \le Cc_0c_2\sigma^{-2}n^{-1}\left(\int [F(t)\{1-F(t)\}]^{1/2}\,dt\right)^2,$$
$$\mathbb{E}|R_{2n}| \le Cc_1\sigma^{-2}n^{-1}\left(\int [F(t)\{1-F(t)\}]^{1/2}\,dt\right)^2,$$
$$\mathbb{E}|R_{3n}| \le C\max(1, c_0^2, c_1^2)\sigma^{-2}n^{-1}\mathbb{E}X_1^2,$$

where $\int [F(t)\{1-F(t)\}]^{1/2}\,dt \le 4\sigma^{-1/2}(\mathbb{E}|X_1|^3)^{1/2}$. Subsequently, by the Markov inequality we have

$$P(|R_n| \ge \sqrt{n}) \le C\max(1, c_0^2, c_1^2, c_2^2)n^{-1/2}(\sigma^{-2}\mathbb{E}X_1^2 + \sigma^{-3}\mathbb{E}|X_1|^3). \tag{A.14}$$

For $Q_n = n^{-3}\sum_{i\ne j\ne k}\gamma_{i,j,k}$, observe that

$$\begin{aligned}\sum_{i\ne j\ne k}\gamma_{i,j,k} &= \frac{1}{6}\sum_{i\ne j\ne k}(\gamma_{i,j,k}+\gamma_{i,k,j}+\gamma_{j,i,k}+\gamma_{j,k,i}+\gamma_{k,i,j}+\gamma_{k,j,i})\\ &:= \frac{1}{6}\sum_{i\ne j\ne k}\Gamma(X_i,X_j,X_k) = \sum_{i<j<k}\Gamma(X_i,X_j,X_k).\end{aligned}$$

In particular, $\binom{n}{3}^{-1}\sum_{i<j<k}\Gamma(X_i,X_j,X_k)$ is a U-statistic with $\mathbb{E}\Gamma(X_1,X_2,X_3)=0$. Set $\gamma_1(x) = \mathbb{E}\{\Gamma(X_1,X_2,X_3)|X_1=x\}$ and

$$\gamma_2(x_1,x_2) = \mathbb{E}\{\Gamma(X_1,X_2,X_3)|X_1=x_1,X_2=x_2\} - \sum_{j=1}^{2}\gamma_1(x_j),$$

$$\gamma_3(x_1,x_2,x_3) = \Gamma(x_1,x_2,x_3) - \sum_{j=1}^{3}\gamma_1(x_j) - \sum_{1\le i<j\le 3}\gamma_2(x_i,x_j).$$

Hoeffding's decomposition gives $Q_n = n^{-3}\sum_{i\ne j\ne k}\gamma_{i,j,k} = Q_{1n}+Q_{2n}$, where

$$Q_{1n} = n^{-3}\binom{n}{3}\frac{3}{n}\sum_{i=1}^{n}\gamma_1(X_i), \tag{A.15}$$

$$Q_{2n} = n^{-3}\binom{n}{3}\left\{\frac{3}{\binom{n}{2}}\sum_{i<j}\gamma_2(X_i,X_j) + \frac{1}{\binom{n}{3}}\sum_{i<j<k}\gamma_3(X_i,X_j,X_k)\right\}. \tag{A.16}$$

By Lemma 4 in Korolyuk and Borovskikh (1988) and inequalities (A.11)–(A.13), we have

$$\mathbb{E}|Q_{1n}|^{3/2} \le C(c_0^3+c_1^3)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3,$$
$$\mathbb{E}|Q_{2n}|^{3/2} \le C(c_0^3+c_1^3)n^{-1}\sigma^{-3}\mathbb{E}|X_1|^3.$$

Next, we apply (3.4) in Theorem 3.1 to the Studentized statistic given in (A.10). To this end, let $\bar{X}_i = X_iI(|X_i|\le\sigma\sqrt{n})$ and define the event

$$\mathcal{E}_n = \left\{\max_{1\le i\le n}|X_i| \le \sigma\sqrt{n}\right\}.$$

By the Markov inequality, $P(\mathcal{E}_n^c) \le n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3$. Further, in view of (A.14), we have

$$\begin{aligned}&\left|P\left\{\frac{W_n+\Delta_n}{(1+Q_n+R_n)^{1/2}} \ge x\right\} - P\left\{\frac{W+\Delta}{(1+\frac{1}{2}n^{-1/2}+\bar{Q}_n)^{1/2}} \ge x\right\}\right|\\ &\quad \le C\left(1\vee\max_{0\le k\le 2}c_k^3\right)n^{-1/2}(\sigma^{-3}\mathbb{E}|X_1|^3+\sigma^{-2}\mathbb{E}X_1^2),\end{aligned} \tag{A.17}$$

where $\bar{Q}_n = \bar{Q}_{1n}I(|\bar{Q}_{1n}|\le\frac{1}{8}) + Q_{2n}I(|Q_{2n}|\le\frac{1}{8})$, where

$$\bar{Q}_{1n} = n^{-3}\binom{n}{3}\frac{3}{n}\sum_{i=1}^{n}\gamma_{b,i}, \quad \gamma_{b,i} := \gamma_1(X_i)(|X_i|\le\sigma\sqrt{n}). \tag{A.18}$$

By the above calculations, put $D_{2n} = \frac{1}{2}n^{-1/2} + \bar{Q}_n$, so that $|D_{2n}| \le \frac{1}{2}$ whenever $n \ge 4$ and

$$\mathbb{E}|D_{2n}|^{3/2} \le C(c_0^3 + c_1^3)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3. \tag{A.19}$$

Moreover, by Lemma 5.4,

$$\mathbb{E}(D_{2n}^2 e^{W_b}) \le C\Big\{n^{-1} + (\mathbb{E}|Q_{2n}|^{3/2})^{1/2} + \mathbb{E}(\bar{Q}_{1n}^2 e^{W_b})\Big\}. \tag{A.20}$$

Recall that $W_b = \sum_{i=1}^n \xi_{b,i} = \sum_{i=1}^n g_i(X_i)I\{|g_i(X_i)| \le 1\}$. For $\gamma_{b,i}$ as in (A.18), it follows from (A.12) and (A.13) that

$$|\mathbb{E}\gamma_{b,i}| = |\mathbb{E}\gamma_1(X_i)I\{|X_i| > \sigma\sqrt{n}\}| \le C(c_0^2 + c_1^2)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3,$$

leading to

$$\begin{aligned}\mathbb{E}\Big(\sum_{i=1}^n \gamma_{b,i}\Big)^2 e^{W_b} &= \sum_{i=1}^n \mathbb{E}(\gamma_{b,i}^2 e^{\xi_{b,i}}) \times \mathbb{E}\{e^{W_b^{(i)}}\} + \sum_{i\ne j}\mathbb{E}(\gamma_{b,i}e^{\xi_{b,i}})\mathbb{E}(\gamma_{b,j}e^{\xi_{b,j}})\mathbb{E}\{e^{W_b^{(i,j)}}\}\\ &\le C(c_0^2 + c_1^2)^2\Big\{n\sigma^{-4}(\mathbb{E}X_1^2)^2 + \sqrt{n}\sigma^{-3}\mathbb{E}|X_1|^3\Big\} + C(1+c_0)^2(c_0^2+c_1^2)^2 n(\sigma^{-3}\mathbb{E}|X_1|^3)^2.\end{aligned}$$

Substituting the above calculations into (A.20) yields

$$\mathbb{E}(D_{2n}^2 e^{W_b}) \le C(J)\Big\{n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3 + n^{-1}(\sigma^{-3}\mathbb{E}|X_1|^3)^2\Big\}. \tag{A.21}$$

Here, and in what follows, $C(J)$ is a constant depending only on $c_j, j = 0, 1, 2$.

Next, we estimate $x\mathbb{E}\{D_{2n}f_x(W_b)\}$ for all $x \ge 0$. By (2.5) and Lemma 5.3, we have

$$\sup_{x\ge 0}\{x\mathbb{E}f_x(W_b)\} \le C$$

and

$$\sup_{x\ge 0}|x\mathbb{E}\{\bar{Q}_{2n}f_x(W_b)\}| \le 8^{-1/4}(\mathbb{E}|Q_{2n}|^{3/2})^{1/2} \times \sup_{x\ge 0} x\{\mathbb{E}f_x^2(W_b)\}^{1/2} \le C(J)n^{-1/2}(\sigma^{-3}\mathbb{E}|X_1|^3)^{1/2}.$$

By independence,

$$\begin{aligned}\mathbb{E}\{\bar{Q}_{1n}f_x(W_b)\} &= \frac{(n-1)(n-2)}{2n^3}\sum_{i=1}^n \mathbb{E}[\gamma_{b,i}\{f_x(W_b) - f_x(W_b^{(i)})\}]\\ &= \frac{(n-1)(n-2)}{2n^3}\sum_{i=1}^n \mathbb{E}\gamma_{b,i}\int_0^{\xi_{b,i}} f_x'(W_b^{(i)} + t)\,dt.\end{aligned}$$

This, together with (A.12) and (A.13) yields

$$\begin{aligned}\sup_{x\ge 0} x|\mathbb{E}\{\bar{Q}_{1n}f_x(W_b)\}| &\le Cn^{-1}\sum_{i=1}^n \mathbb{E}|\gamma_{b,i}\xi_{b,i}| \times \sup_{x\ge 0}\max_{|t|\le 1}\mathbb{E}|xf_x'(W_b^{(i)} + t)|\\ &\le C(J)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3.\end{aligned}$$

Putting pieces together, we conclude that

$$\sup_{x\ge 0}|x\mathbb{E}\{D_{2n}f_x(W_b)\}| \le C(J)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3. \tag{A.22}$$

For each $1 \le i \le n$, $|D_{2n} - D_{2n}^{(i)}|$ can be bounded by

$$\frac{(n-1)(n-2)}{2n^3}|\gamma_{b,i}| + \Big|\frac{n-2}{n^3}\sum_{j(\ne i)}\gamma_2(X_i, X_j) + \frac{1}{n^3}\sum_{j<k(\ne i)}\gamma_3(X_i, X_j, X_k)\Big|.$$

By similar arguments that lead to (A.15) and (A.16), we can show that

$$\mathbb{E}|D_{2n} - D_{2n}^{(i)}|^{3/2} \le C(c_0^3 + c_1^3)n^{-3/2}\sigma^{-3}\mathbb{E}|X_1|^3$$

and therefore,

$$\sum_{i=1}^n \|\xi_{b,i}\|_3 \times \|D_{2n} - D_{2n}^{(i)}\|_{\frac{3}{2}} \le C(J)n^{-1/2}\sigma^{-3}\mathbb{E}|X_1|^3. \tag{A.23}$$

Assembling (A.9), (A.10), (A.17), (A.19), (A.21), (A.22) and (A.23) proves (4.10) for all $x \geq 0$. Applying this result to $-T$ with weight function $-J$ covers the case of $x \leq 0$ and hence completes the proof of Theorem 4.4. □

Proof of Lemma A.1. We only prove the last two inequalities, i.e. (A.7) and (A.8), as the others can be obtained by routine computations. Observe that

$$\begin{aligned} D_{2n}(\xi_{\mathrm{b},1},\ldots,\xi_{\mathrm{b},n}) &= \sum_{i=1}^n(\xi_{\mathrm{b},i}^2-\mathbb{E}\xi_{\mathrm{b},i}^2)+\frac{2}{n-2}\sum_{i=1}^n\xi_{\mathrm{b},i}\Psi_{n,i}+\frac{1}{2\sqrt{n}}-\beta_2 \\ &:= \Pi_1+\Pi_2+\Pi_3, \end{aligned}$$

where $\Pi_1=\sum_{i=1}^n\eta_i=\sum_{i=1}^n(\xi_{\mathrm{b},i}^2-\mathbb{E}\xi_{\mathrm{b},i}^2)$, $\Pi_2=\frac{2}{n-2}\sum_{i=1}^n\xi_{\mathrm{b},i}\Psi_{n,i}$ and $\Pi_3=\frac{1}{2\sqrt{n}}-\beta_2$. Clearly, $|D_{2n}(\xi_{\mathrm{b},1},\ldots,\xi_{\mathrm{b},n})|\wedge 1\leq |\Pi_1|\wedge 1+|\Pi_2|\wedge 1+|\Pi_3|\wedge 1$.

Observe that conditional on X_i, $\Psi_{n,i}$ is a sum of independent random variables with zero means. By Hölder's inequality we have

$$\mathbb{E}\{(|\Pi_2|\wedge 1)^2e^{W_{\mathrm{b}}}\}\leq(\mathbb{E}\Pi_2^2)^{1/2}(\mathbb{E}e^{2W_{\mathrm{b}}})^{1/2}\leq Cn^{-1/2}\|h(X_1,X_2)\|_2$$

and it is straightforward to verify that $\mathbb{E}\{(|\Pi_3|\wedge 1)^2e^{W_{\mathrm{b}}}\}\leq C(n^{-1/2}+\beta_2)$. For Π_1, a direct calculation gives

$$\begin{aligned} \mathbb{E}(\Pi_1^2e^{W_{\mathrm{b}}}) &= \sum_{i=1}^n\mathbb{E}(\eta_i^2e^{\xi_{\mathrm{b},i}})\mathbb{E}\{e^{W_{\mathrm{b}}^{(i)}}\}+\sum_{i\neq j}\mathbb{E}(\eta_ie^{\xi_{\mathrm{b},i}})\mathbb{E}(\eta_je^{\xi_{\mathrm{b},j}})\mathbb{E}\{e^{W_{\mathrm{b}}^{(i,j)}}\} \\ &\leq C\sum_{i=1}^n\mathbb{E}\xi_{\mathrm{b},i}^4+C\left(\sum_{i=1}^n\mathbb{E}|\xi_{\mathrm{b},i}|^3\right)^2\leq C(\beta_p+\beta_p^2), \end{aligned}$$

where $\beta_p:=\sum_{i=1}^n\mathbb{E}|\xi_i|^p\leq\sigma_p^pn^{1-p/2}$. Putting the above calculations together implies (A.7).

Finally, we prove (A.8). For $0\leq x\leq 1$, the boundedness of f_x implies

$$|x\mathbb{E}\{(\Pi_2+\Pi_3)f_x(W_{\mathrm{b}})\}|\leq\mathbb{E}(|\Pi_2|+|\Pi_3|)\leq C\{n^{-1/2}\|h(X_1,X_2)\|_2+\beta_2\}.$$

When $x\geq 1$, it follows from (5.17) that

$$\begin{aligned} |x\mathbb{E}\{(\Pi_2+\Pi_3)f_x(W_{\mathrm{b}})\}| &\leq Cxe^{-x}\Big[\mathbb{E}(|\Pi_2|+|\Pi_3|)+\mathbb{E}\{(|\Pi_2|+|\Pi_3|)e^{W_{\mathrm{b}}}\}\Big] \\ &\leq C\{n^{-1/2}\|h(X_1,X_2)\|_2+\beta_2\}. \end{aligned}$$

As for Π_1, note that

$$\mathbb{E}\{\Pi_1f_x(W_{\mathrm{b}})\}=\sum_{i=1}^n\mathbb{E}[\eta_i\{f_x(W_{\mathrm{b}})-f_x(W_{\mathrm{b}}^{(i)})\}]=\sum_{i=1}^n\mathbb{E}\eta_i\int_0^{\xi_{\mathrm{b},i}}f_x'(W_{\mathrm{b}}^{(i)}+t)\,dt.$$

This, together with Lemma 5.4, yields

$$\sup_{x\geq 0}|x\mathbb{E}\{\Pi_1f_x(W_{\mathrm{b}})\}|\leq C\sum_{i=1}^n\mathbb{E}|\xi_{\mathrm{b},i}|^3\leq C\beta_3.$$

Consequently, (A.8) follows from the above calculations and the fact that $\beta_2+\beta_3\leq\sigma_p^pn^{1-p/2}$. □

References

Arratia, R., Goldstein, L., Gordon, L., 1990. Poisson approximation and the Chen–Stein method. Statist. Sci. 5, 403–434.
Arvesen, J.N., 1969. Jackknifing U-statistics. Ann. Math. Statist. 40, 2076–2100.
Baldi, P., Rinott, Y., Stein, C., 1989. A normal approximation for the number of local maxima of a random function on a graph. In: Probability, Statistics and Mathematics. Academic Press, Boston, MA, pp. 59–81.
Barbour, A.D., 1990. Stein's method for diffusion approximations. Probab. Theory Related Fields 84, 297–322.
Barbour, A.D., Chen, L.H.Y., 2005. An Introduction to Stein Method. In: Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 4. Singapore Univ. Press and World Scientific, Singapore.
Barbour, A.D., Holst, L., Janson, S., 1992. Poisson Approximation. Oxford University Press, Oxford.
Belloni, A., Chernozhukov, V., Wang, L., 2011. Square-root lasso: Pivotal recovery of sparse signals via conic programming. Biometrika 98, 791–806.
Bentkus, V., Götze, F., Zitikis, M., 1994. Lower estimates of the convergence rate for U-statistics. Ann. Probab. 22, 1707–1714.
Bickel, P.J., 1974. Edgeworth expansions in nonparametric statistics. Ann. Statist. 2, 1–20.
Bjerve, X., 1977. Error bounds for linear combinations of order statistics. Ann. Statist. 5, 237–255.
Callaert, H., Janssen, P., 1978. The Berry–Esseen theorem for U-statistics. Ann. Statist. 6, 417–421.
Chan, Y.K., Wierman, J., 1977. On the Berry–Esseen theorem for U-statistics. Ann. Probab. 5, 136–139.
Chatterjee, S., 2005. Concentration Inequalities with Exchangeable Pairs (Ph.D. thesis), Department of Statistics, Stanford University, Available on arXiv:math/0507526.
Chatterjee, S., 2007. Stein's method for concentration inequalities. Probab. Theory Related Fields 138, 305–321.
Chatterjee, S., 2008. A new method of normal approximation. Ann. Probab. 36, 1584–1610.

Chatterjee, S., 2014. A short survey of Stein's method. In: Proceedings of ICM 2014 (in press) Available at: arXiv:1404.1392.
Chen, L.H.Y., 1993. Extending the Poisson approximation. Science 262, 379–380.
Chen, L.H.Y., Fang, X., 2011. Multivariate normal approximation by Stein's method: The concentration inequality approach. arXiv:1111.4073.
Chen, L.H.Y., Fang, X., Shao, Q.-M., 2013. From Stein identities to moderate deviations. Ann. Probab. 41, 262–293.
Chen, L.H.Y., Goldstein, L., Shao, Q.-M., 2011. Normal Approximation by Stein's Method. Springer, Berlin.
Chen, L.H.Y., Shao, Q.-M., 2001. A non-uniform Berry–Esseen bound via Stein's method. Probab. Theory Related Fields 120, 236–254.
Chen, L.H.Y., Shao, Q.-M., 2004. Normal approximation under local dependence. Ann. Probab. 32, 1985–2028.
Chen, L.H.Y., Shao, Q.-M., 2007. Normal approximation for nonlinear statistics using a concentration inequality approach. Bernoulli 13, 581–599.
Clarke, S., Hall, P., 2009. Robustness of multiple testing procedures against dependence. Ann. Statist. 37, 332–358.
Delaigle, A., Hall, P., Jin, J., 2011. Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 283–301.
de la Peña, V.H., Lai, T.L., Shao, Q.-M., 2009. Self-Normalized Processes: Theory and Statistical Applications. Springer, Berlin.
Dembo, A., Rinott, Y., 1996. Some examples of normal approximations by Stein's method. In: Random Discrete Structures. In: The IMA Volumes in Mathematics and its Applications, vol. 76. Springer, New York, pp. 25–44.
Diaconis, P., 1977. The distribution of leading digits and uniform distribution mod 1. Ann. Probab. 5, 72–81.
Diaconis, P., Holmes, S., 2004. Stein's Method: Expository Lectures and Applications. In: Institute of Mathematical Statistics Lecture Notes, vol. 46. IMS, Hayward, CA.
Fan, J., Hall, P., Yao, Q., 2007. To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? J. Amer. Statist. Assoc. 102, 1282–1288.
Fan, J., Lv, J., 2010. A selective overview of variable selection in high dimensional feature space. Statist. Sinica 20, 101–148.
Filippova, A.A., 1962. Mises's theorem of the asymptotic behavior of functionals of empirical distributions functions and its statistical application. Theory Probab. Appl. 7, 24–57.
Friedrich, K.O., 1989. A Berry–Esseen bound for functions of independent random variables. Ann. Statist. 17, 170–183.
Ghosh, M., 1985. Berry–Esseen bounds for functionals of U-statistic. Sankhyā 47, 255–270.
Goldstein, L., Reinert, G., 1997. Stein's method and the zero bias transformation with application to simple random sampling. Ann. Appl. Probab. 7, 935–952.
Grams, W.F., Serfling, R.J., 1973. Convergence rates for U-statistics and related statistics. Ann. Statist. 1, 153–160.
Hajek, J., 1968. Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39, 325–346.
Helmers, R., 1977. The order of the normal approximation for linear combinations of order statistics with smooth weight functions. Ann. Probab. 5, 940–953.
Helmers, R., 1982. Edgeworth expansions for linear combinations of order statistics. In: Math. Centre Tracts, vol. 105. Math. Centrum, Amsterdam.
Helmers, R., Janssen, P., Serfling, R., 1990. Berry–Esseen and bootstrap results for generalized L-statistics. Scand. J. Stat. 17, 65–77.
Helmers, R., van Zwet, W.R., 1982. The Berry–Esseen bound for U-statistics. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics, III. vol. 1. Academic Press, New York, pp. 497–512.
Jing, B.-Y., Shao, Q.-M., Wang, Q., 2003. Self-normalized Cramér-type large deviation for independent random variables. Ann. Probab. 31, 2167–2215.
Jing, B.-Y., Zhou, W., 2005. A note on Edgeworth expansions for U-statistics under minimal conditions. Lith. Math. J. 45, 353–358.
Korolyuk, V.S., Borovskikh, Yu.V., 1988. Asymptotic theory of U-statistics. Ukrainian Math. J. 40, 142–154.
Korolyuk, V.S., Borovskikh, Yu.V., 1994. Theory of U-Statistics. Kluwer Academic Publishers, Dordrecht.
Liu, W.D., Shao, Q.-M., 2013. A Cramér moderate deviation theorem for Hotelling's T^2-statistic with applications to global tests. Ann. Statist. 41, 296–322.
Mackey, L., Jordan, M.I., Chen, R.Y., Farrell, B., Tropp, J.A., 2014. Matrix concentration inequalities via the method of exchangeable pairs. Ann. Probab. 42, 906–945.
Noether, G.E., 1955. On a theorem of Pitman. Ann. Math. Statist. 26, 64–68.
Nourdin, I., Peccati, G., 2009. Stein's method on Weiner Chaos. Probab. Theory Related Fields 145, 75–118.
Petrov, V.V., 1975. Sums of Independent Random Variables. Springer, New York.
Pinelis, I., Molzon, R., 2009. Berry–Esseen bounds for general nonlinear statistics, with applications to Pearson's and non-central student's and Hotelling's. Available at: arXiv:0906.0177.
Raič, M., 2007. CLT-related large deviation bounds based on Stein's method. Adv. Appl. Probab. 39, 731–752.
Ross, N., 2011. Fundamentals of Stein's method. Probab. Surv. 8, 210–293.
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.
Shao, Q.-M., 1999. Cramér type large deviation for student's t statistic. J. Theoret. Probab. 12, 387–398.
Shao, Q.-M., 2010. Stein's method, self-normalized limit theory and applications. In: Proceedings of the International Congress of Mathematicians. Volume IV. Hindustan Book Agency, New Delhi, pp. 2325–2350.
Shao, Q.-M., Zhou, W.-X., 2014. Cramér type moderate deviation theorems for self-normalized processes. Bernoulli (in press) Available at: arXiv:1405.1218.
Stein, C., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability Theory. Univ. California Press, Berkeley, CA, pp. 583–602.
Stein, C., 1986. Approximation Computation of Expectations. In: Institute of Mathematical Statistics Lecture Notes–Monograph Series, vol. 7. IMS, Hayward, CA.
van Zwet, W.R., 1984. A Berry–Esseen bound for symmetric statistics. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 66, 425–440.
Wang, Q., 2002. Non-uniform Berry–Esseen bound for U-statistics. Statist. Sinica 12, 1157–1169.
Wang, Q., Jing, B.-Y., Zhao, L., 2000. The Berry–Esseen bound for Studentized statistics. Ann. Probab. 28, 511–535.
Zhao, L.C., Chen, X.R., 1983. Non-uniform convergence rate for distributions of U-statistics. Sci. Sin. Ser. A 26, 795–810.