UC San Diego
UC San Diego Previously Published Works

Title
Cramér type moderate deviation theorems for self-normalized processes

Permalink
https://escholarship.org/uc/item/46d3n94p

Journal
Bernoulli, 22(4)

ISSN
1350-7265

Authors
Shao, Qi-Man
Zhou, Wen-Xin

Publication Date
2016

DOI
10.3150/15-bej719

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Bernoulli **22**(4), 2016, 2029–2079
DOI: 10.3150/15-BEJ719

Cramér type moderate deviation theorems for self-normalized processes

QI-MAN SHAO[1] and WEN-XIN ZHOU[2,3]

[1]*Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.*
E-mail: qmshao@cuhk.edu.hk
[2]*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA. E-mail: wenxinz@princeton.edu*
[3]*School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia*

Cramér type moderate deviation theorems quantify the accuracy of the relative error of the normal approximation and provide theoretical justifications for many commonly used methods in statistics. In this paper, we develop a new randomized concentration inequality and establish a Cramér type moderate deviation theorem for general self-normalized processes which include many well-known Studentized nonlinear statistics. In particular, a sharp moderate deviation theorem under optimal moment conditions is established for Studentized U-statistics.

Keywords: moderate deviation; nonlinear statistics; relative error; self-normalized processes; Studentized statistics; U-statistics

1. Introduction

Let T_n be a sequence of random variables and assume that T_n converges to Z in distribution. The problem we are interested in is to calculate the tail probability of T_n, $\mathbb{P}(T_n \geq x)$, where x may also depend on n and can go to infinity. Because the true tail probability of T_n is typically unknown, it is common practice to use the tail probability of Z to estimate that of T_n. A natural question is how accurate the approximation is? There are two major approaches for measuring the approximation error. One approach is to study the absolute error via Berry–Esseen type bounds or Edgeworth expansions. The other is to estimate the relative error of the tail probability of T_n against the tail probability of the limiting distribution, that is,

$$\frac{\mathbb{P}(T_n \geq x)}{\mathbb{P}(Z \geq x)}, \qquad x \geq 0.$$

A typical result in this direction is the so-called *Cramér type moderate deviation*. The focus of this paper is to find the largest possible a_n ($a_n \to \infty$) so that

$$\frac{\mathbb{P}(T_n \geq x)}{\mathbb{P}(Z \geq x)} = 1 + o(1)$$

holds uniformly for $0 \leq x \leq a_n$.

1350-7265 © 2016 ISI/BS

The moderate deviation, and other noteworthy limiting properties for self-normalized sums are now well-understood. More specifically, let $X_1, X_2, \dots, X_n$ be independent and identically distributed (i.i.d.) non-degenerate real-valued random variables with zero means, and let

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

be, respectively, the partial sum and the partial quadratic sum. The corresponding self-normalized sum is defined as S_n/V_n. The study of the asymptotic behavior of self-normalized sums has a long history. Here, we refer to [27] for weak convergence and to [20,21] for the law of the iterated logarithms when X_1 is in the domain of attraction of a normal or stable law. [4] derived the optimal Berry–Esseen bound, and [18] proved that S_n/V_n is asymptotically normal if and only if X_1 belongs to the domain of attraction of a normal law. Under the same necessary and sufficient conditions, [13] proved a self-normalized analogue of the weak invariance principle. It should be noted that all of these limiting properties also hold for the standardized sums. However, in contrast to the large deviation asymptotics for the standardized sums, which require a finite moment generating function of X_1, [30] proved a self-normalized large deviation for S_n/V_n without any moment assumptions. Moreover, [31] established a self-normalized Cramér type moderate deviation theorem under a finite third moment, that is, if $\mathbb{E}|X_1|^3 < \infty$, then

$$\frac{\mathbb{P}(S_n/V_n \ge x)}{1-\Phi(x)} \to 1 \qquad \text{holds uniformly for } 0 \le x \le o\big(n^{1/6}\big), \tag{1.1}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Result (1.1) was further extended to independent (not necessarily identically distributed) random variables by [23] 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 [23] gives

$$\frac{\mathbb{P}(S_n/V_n \ge x)}{1-\Phi(x)} = 1 + O(1)(1+x)^3 \frac{\sum_{i=1}^n \mathbb{E}|X_i|^3}{(\sum_{i=1}^n \mathbb{E}X_i^2)^{3/2}} \tag{1.2}$$

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

Over the past two decades, there has been significant progress in the development of the self-normalized limit theory. For a systematic presentation of the general self-normalized limit theory and its statistical applications, we refer to [14].

The main purpose of this paper is to extend (1.2) to more general self-normalized processes, including many commonly used Studentized statistics, in particular, Student's t-statistic and Studentized U-statistics. Notice that the proof in [23] is lengthy and complicated, and their method is difficult to adopt for general self-normalized processes. The proof in this paper is based on a new randomized concentration inequality and the method of conjugated distributions (also known as the change of measure method), which opens a new approach to studying self-normalized limit theorems.

The rest of this paper is organized as follows. The general result is presented in Section 2. To illustrate the sharpness of the general result, a result similar to (1.1) and (1.2) is obtained for Studentized U-statistics in Section 3. Applications to other Studentized statistics will be discussed in our future work. To establish the general Cramér type moderation theorem, a novel randomized concentration inequality is proved in Section 4. The proofs of the main results and key technical lemmas are given in Sections 5 and 6. Other technical proofs are provided in the Appendix.

2. Moderate deviations for self-normalized processes

Our research on self-normalized processes is motivated by Studentized nonlinear statistics. Nonlinear statistics are the building blocks in various statistical inference problems. It is known that many of these statistics can be written as a partial sum plus a negligible term. Typical examples include U-statistics, multi-sample U-statistics, L-statistics, random sums and functions of nonlinear statistics. We refer to [12] for a unified approach to uniform and non-uniform Berry–Esseen bounds for standardized nonlinear statistics.

Assume that the nonlinear process of interest can be decomposed as a standardized partial sum of independent random variables plus a remainder, that is,

$$\frac{1}{\sigma}\left(\sum_{i=1}^{n}\xi_i + D_{1n}\right),$$

where $\xi_1, \dots, \xi_n$ are independent random variables satisfying

$$\mathbb{E}\xi_i = 0 \qquad \text{for } i = 1, \dots, n \quad \text{and} \quad \sum_{i=1}^{n}\mathbb{E}\xi_i^2 = 1, \tag{2.1}$$

and where $D_{1n} = D_{1n}(\xi_1, \dots, \xi_n)$ is a measurable function of $\{\xi_i\}_{i=1}^n$. Because σ is typically unknown, a self-normalized process

$$T_n = \frac{1}{\widehat{\sigma}}\left(\sum_{i=1}^{n}\xi_i + D_{1n}\right)$$

is more commonly used in practice, where $\widehat{\sigma}$ is an estimator of σ. Assume that $\widehat{\sigma}$ can be written as

$$\widehat{\sigma} = \left\{\left(\sum_{i=1}^{n}\xi_i^2\right)(1 + D_{2n})\right\}^{1/2},$$

where D_{2n} is a measurable function of $\{\xi_i\}_{i=1}^n$. Without loss of generality and for the sake of convenience, we assume $\sigma = 1$. Therefore, under the assumptions in (2.1), we can rewrite the self-normalized process T_n as

$$T_n = \frac{W_n + D_{1n}}{V_n(1 + D_{2n})^{1/2}}, \tag{2.2}$$

where

$$W_n = \sum_{i=1}^{n}\xi_i, \qquad V_n = \left(\sum_{i=1}^{n}\xi_i^2\right)^{1/2}.$$

Essentially, this formulation (2.2) states that, for a nonlinear process that be can written as a linear process plus a negligible remainder, it is natural to expect that the corresponding normalizing term is dominated by a quadratic process. To ensure that T_n is well-defined, it is assumed implicitly in (2.2) that the random variable D_{2n} satisfies $1 + D_{2n} > 0$. Examples satisfying (2.2) include the t-statistic, Studentized U- and L-statistics. See [38] and the references therein for more details.

In this section, we establish a general Cramér type moderate deviation theorem for a self-normalized process T_n in the form of (2.2). We start by introducing some of the basic notation that is frequently used throughout this paper. For $x \geq 1$, write

$$L_{n,x} = \sum_{i=1}^{n} \delta_{i,x}, \qquad I_{n,x} = \mathbb{E}\exp\bigl(xW_n - x^2V_n^2/2\bigr) = \prod_{i=1}^{n} \mathbb{E}\exp\bigl(\xi_{i,x} - \xi_{i,x}^2/2\bigr), \tag{2.3}$$

where $\delta_{i,x} = \mathbb{E}\xi_{i,x}^2 I(|\xi_{i,x}| > 1) + \mathbb{E}|\xi_{i,x}|^3 I(|\xi_{i,x}| \leq 1)$ with $\xi_{i,x} := x\xi_i$. For $i = 1, \ldots, n$, let $D_{1n}^{(i)}$ and $D_{2n}^{(i)}$ be arbitrary measurable functions of $\{\xi_j\}_{j=1, j\neq i}^n$, such that $\{D_{1n}^{(i)}, D_{2n}^{(i)}\}$ and ξ_i are independent. Moreover, define

$$\begin{aligned} R_{n,x} = I_{n,x}^{-1} \times \Biggl(& \mathbb{E}\bigl\{\bigl(x|D_{1n}| + x^2|D_{2n}|\bigr)e^{\sum_{j=1}^n (\xi_{j,x} - \xi_{j,x}^2/2)}\bigr\} \\ & + \sum_{i=1}^{n} \mathbb{E}\bigl[\min\bigl(|\xi_{i,x}|, 1\bigr)\bigl\{\bigl|D_{1n} - D_{1n}^{(i)}\bigr| + x\bigl|D_{2n} - D_{2n}^{(i)}\bigr|\bigr\}e^{\sum_{j\neq i}(\xi_{j,x} - \xi_{j,x}^2/2)}\bigr]\Biggr). \end{aligned} \tag{2.4}$$

Here, and in the sequel, we use $\sum_{j\neq i} = \sum_{j=1, j\neq i}^n$ for brevity.

Now we are ready to present the main results.

Theorem 2.1. *Let T_n be defined in (2.2) under condition (2.1). Then there exist positive absolute constants C_1–C_4 and c_1 such that*

$$\mathbb{P}(T_n \geq x) \geq \bigl\{1 - \Phi(x)\bigr\} \exp\{-C_1 L_{n,x}\}(1 - C_2 R_{n,x}) \tag{2.5}$$

and

$$\begin{aligned} \mathbb{P}(T_n \geq x) \leq {} & \bigl\{1 - \Phi(x)\bigr\} \exp\{C_3 L_{n,x}\}(1 + C_4 R_{n,x}) \\ & + \mathbb{P}\bigl(x|D_{1n}| > V_n/4\bigr) + \mathbb{P}\bigl(x^2|D_{2n}| > 1/4\bigr) \end{aligned} \tag{2.6}$$

for all $x \geq 1$ satisfying

$$\max_{1\leq i\leq n} \delta_{i,x} \leq 1 \tag{2.7}$$

and

$$L_{n,x} \leq c_1 x^2. \tag{2.8}$$

Remark 2.1. The quantity $L_{n,x}$ in (2.3) is essentially the same as the factor $\Delta_{n,x}$ in [23], which is the leading term that describes the accuracy of the relative normal approximation error. To deal with the self-normalized nonlinear process T_n, first we need to "linearize" it in a proper way, although at the cost of introducing some complex perturbation terms. The linearized term is $xW_n - x^2V_n^2/2$, and its exponential moment is denoted by $I_{n,x}$ as in (2.3). A randomized concentration inequality is therefore developed (see Section 4) to cope with these random perturbations which lead to the quantity $R_{n,x}$ given in (2.4). Similar quantities also appear in the Berry–Esseen bounds for nonlinear statistics. See, for example, Theorems 2.1 and 2.2 in [12].

Theorem 2.1 provides the upper and lower bounds of the relative errors for $x \geq 1$. To cover the case of $0 \leq x \leq 1$, we present a rough estimate of the absolute error in the next theorem, and refer to [32] for the general Berry–Esseen bounds for self-normalized processes.

Theorem 2.2. *There exists an absolute constant $C > 1$ such that for all $x \geq 0$,*

$$\left|\mathbb{P}(T_n \leq x) - \Phi(x)\right| \leq C\breve{R}_{n,x}, \tag{2.9}$$

where

$$\begin{aligned}\breve{R}_{n,x} &:= L_{n,1+x} + \mathbb{E}|D_{1n}| + x\mathbb{E}|D_{2n}| \\ &\quad + \sum_{i=1}^{n} \mathbb{E}\big[\xi_i I\big\{|\xi_i| \leq 1/(1+x)\big\}\big\{\big|D_{1n} - D_{1n}^{(i)}\big| + x\big|D_{2n} - D_{2n}^{(i)}\big|\big\}\big]\end{aligned} \tag{2.10}$$

for $L_{n,1+x}$ as in (2.3).

The proof of Theorem 2.2 is deferred to the Appendix. In particular, when $0 \leq x \leq 1$, the quantity $L_{n,1+x}$ satisfies

$$\begin{aligned}L_{n,1+x} &= (1+x)^2\sum_{i=1}^{n}\mathbb{E}\xi_i^2 I\big\{|\xi_i| > 1/(1+x)\big\} + (1+x)^3\sum_{i=1}^{n}\mathbb{E}|\xi_i|^3 I\big\{|\xi_i| \leq 1/(1+x)\big\} \\ &\leq (1+x)^2\sum_{i=1}^{n}\mathbb{E}\xi_i^2 I\big(|\xi_i| > 1/2\big) + (1+x)^3\sum_{i=1}^{n}\mathbb{E}|\xi_i|^3 I\big(|\xi_i| \leq 1\big) \\ &\leq (1+x)^2\sum_{i=1}^{n}\mathbb{E}\xi_i^2 I\big(|\xi_i| > 1\big) + (1+x)^2\sum_{i=1}^{n}\mathbb{E}\xi_i^2 I\big(1/2 < |\xi_i| \leq 1\big) \\ &\quad + (1+x)^3\sum_{i=1}^{n}\mathbb{E}|\xi_i|^3 I\big(|\xi_i| \leq 1\big),\end{aligned}$$

which can be further bounded, up to a constant, by

$$\sum_{i=1}^{n}\mathbb{E}\xi_i^2 I\big(|\xi_i| > 1\big) + \sum_{i=1}^{n}\mathbb{E}|\xi_i|^3 I\big(|\xi_i| \leq 1\big).$$

Remark 2.2. 1. When $D_{1n} = D_{2n} = 0$, T_n reduces to the self-normalized sum of independent random variables, and thus Theorems 2.1 and 2.2 together immediately imply the main result in [23]. The proof therein, however, is lengthy and fairly complicated, especially the proof of Proposition 5.4, and can hardly be applied to prove the general result of Theorem 2.1. The proof of our Theorem 2.1 is shorter and more transparent.

2. D_{1n} and D_{2n} in the definitions of $R_{n,x}$ and $\breve{R}_{n,x}$ can be replaced by any non-negative random variables D_{3n} and D_{4n}, respectively, provided that $|D_{1n}| \le D_{3n}$, $|D_{2n}| \le D_{4n}$.

3. Condition (2.1) implies that ξ_i actually depends on both n and i; that is, ξ_i denotes ξ_{ni}, which is an array of independent random variables.

3. Studentized U-statistics

As a prototypical example of the self-normalized processes given in (2.2), we are particularly interested in Studentized U-statistics. In this section, we apply Theorems 2.1 and 2.2 to Studentized U-statistics and obtain a sharp Cramér moderate deviation under optimal moment conditions.

Let $X_1, X_2, \dots, X_n$ be a sequence of i.i.d. random variables and let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric Borel measurable function of m variables, where $2 \le m < n/2$ is fixed. The Hoeffding's U-statistic with a kernel h of degree m is defined as (Hoeffding [22])

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \dots, X_{i_m}),$$

which is an unbiased estimate of $\theta = \mathbb{E}h(X_1, \dots, X_m)$. Let

$$h_1(x) = \mathbb{E}\{h(X_1, X_2, \dots, X_m)|X_1 = x\}, \qquad x \in \mathbb{R}$$

and

$$\sigma^2 = \operatorname{Var}\{h_1(X_1)\}, \qquad \sigma_h^2 = \operatorname{Var}\{h(X_1, X_2, \dots, X_m)\}. \tag{3.1}$$

Assume $0 < \sigma^2 < \infty$, then the standardized non-degenerate U-statistic is given by

$$Z_n = \frac{\sqrt{n}}{m\sigma}(U_n - \theta).$$

The U-statistic is a basic statistic and its asymptotic properties have been extensively studied in the literature. We refer to [25] for a systematic presentation of the theory of U-statistics. For uniform Berry–Esseen bounds, see [1,2,5,8,9,16,17,19,29,35,39] and [12]. We refer to [15,24] and [6,7] for large and moderate deviation asymptotics.

Because σ is usually unknown, we are interested in the following Studentized U-statistic (Arvensen [3]), which is widely used in practice:

$$T_n = \frac{\sqrt{n}}{ms_1}(U_n - \theta),$$

where s_1^2 denotes the leave-one-out Jackknife estimator of σ^2 given by

$$
\begin{aligned}
s_1^2 &= \frac{(n-1)}{(n-m)^2} \sum_{i=1}^{n} (q_i - U_n)^2 \qquad \text{with} \\
q_i &= \frac{1}{\binom{n-1}{m-1}} \sum_{\substack{1 \le \ell_1 < \cdots < \ell_{m-1} \le n \\ \ell_j \ne i, j=1,\ldots,m-1}} h(X_i, X_{\ell_1}, \ldots, X_{\ell_{m-1}}).
\end{aligned} \tag{3.2}
$$

In contrast to the standardized U-statistics, few optimal limit theorems are available for Studentized U-statistics in the literature. A uniform Berry–Esseen bound for Studentized U-statistics was proved in [38] for $m = 2$ and $\mathbb{E}|h(X_1, X_2)|^3 < \infty$. However, a finite third moment of $h(X_1, X_2)$ may not be an optimal condition. Partial results on Cramér type moderate deviation were obtained in [36,37] and [26].

As a direct but non-trivial consequence of Theorems 2.1 and 2.2, we establish the following sharp Cramér type moderate deviation theorem for the Studentized U-statistic T_n.

Theorem 3.1. *Assume that $\sigma_p := (\mathbb{E}|h_1(X_1) - \theta|^p)^{1/p} < \infty$ for some $2 < p \le 3$. Suppose that there are constants $c_0 \ge 1$ and $\tau \ge 0$ such that*

$$
\big\{h(x_1, \ldots, x_m) - \theta\big\}^2 \le c_0 \left[\tau\sigma^2 + \sum_{i=1}^{m} \big\{h_1(x_i) - \theta\big\}^2\right]. \tag{3.3}
$$

Then there exist positive constants C_1 and c_1 independent of n such that

$$
\frac{\mathbb{P}(T_n \ge x)}{1 - \Phi(x)} = 1 + O(1)\left\{(\sigma_p/\sigma)^p \frac{(1+x)^p}{n^{p/2-1}} + (\sqrt{a_m} + \sigma_h/\sigma)\frac{(1+x)^3}{\sqrt{n}}\right\} \tag{3.4}
$$

holds uniformly for

$$
0 \le x \le c_1 \min\big\{(\sigma/\sigma_p) n^{1/2-1/p}, (n/a_m)^{1/6}\big\},
$$

where $|O(1)| \le C_1$ and $a_m = \max\{c_0\tau, c_0 + m\}$. In particular,

$$
\frac{\mathbb{P}(T_n \ge x)}{1 - \Phi(x)} \to 1 \tag{3.5}
$$

holds uniformly in $x \in [0, o(n^{1/2-1/p}))$.

It is easy to verify that condition (3.3) is satisfied for the t-statistic ($h(x_1, x_2) = (x_1 + x_2)/2$ with $c_0 = 2$ and $\tau = 0$), sample variance ($h(x_1, x_2) = (x_1 - x_2)^2/2$, $c_0 = 10$, $\tau = \theta^2/\sigma^2$), Gini's mean difference ($h(x_1, x_2) = |x_1 - x_2|$, $c_0 = 8$, $\tau = \theta^2/\sigma^2$) and one-sample Wilcoxon's statistic ($h(x_1, x_2) = I(x_1 + x_2 \le 0)$, $c_0 = 1$, $\tau = 1/\sigma^2$). Although it may be interesting to investigate whether condition (3.3) can be weakened, it seems that it is impossible to remove condition (3.3) completely. We also note that result (3.5) was earlier proved in [26] for $m = 2$. However, the approach used therein can hardly be extended to the case $m \ge 3$.

4. A randomized concentration inequality

To prove Theorem 2.1, we first develop a randomized concentration inequality via Stein's method. Stein's method (Stein [34]) is a powerful tool in the normal and non-normal approximation of both independent and dependent variables, and the concentration inequality is a useful approach in Stein's method. We refer to [10] for systematic coverage of the method and recent developments in both theory and applications and to [12] for uniform and non-uniform Berry–Esseen bounds for nonlinear statistics using the concentration inequality approach.

Let $\xi_1, \dots, \xi_n$ be independent random variables such that

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

Let

$$W = \sum_{i=1}^n \xi_i, \qquad V^2 = \sum_{i=1}^n \xi_i^2 \tag{4.1}$$

and let $\Delta_1 = \Delta_1(\xi_1, \dots, \xi_n)$ and $\Delta_2 = \Delta_2(\xi_1, \dots, \xi_n)$ be two measurable functions of $\xi_1, \dots, \xi_n$. Moreover, set

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

Theorem 4.1. *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*

$$\mathbb{P}(\Delta_1 \le W \le \Delta_2) \le 17(\beta_2 + \beta_3) + 5\mathbb{E}|\Delta_2 - \Delta_1| + 2\sum_{i=1}^n \sum_{j=1}^2 \mathbb{E}\bigl|\xi_i\bigl\{\Delta_j - \Delta_j^{(i)}\bigr\}\bigr|. \tag{4.2}$$

We note that a similar result was obtained by [12] with $\mathbb{E}|W(\Delta_2 - \Delta_1)|$ instead of $\mathbb{E}|\Delta_2 - \Delta_1|$ in (4.2). However, using the term $\mathbb{E}|W(\Delta_2 - \Delta_1)|$ will not yield the sharp bound in (3.4) when Theorem 2.1 is applied to Studentized U-statistics. This provides our main motivation for developing the new concentration inequality (4.2).

Proof of Theorem 4.1. Assume without loss of generality that $\Delta_1 \le \Delta_2$. The proof is based on Stein's method. For every $x \in \mathbb{R}$, let $f_x(w)$ be the solution to Stein's equation

$$f_x'(w) - wf_x(w) = I(w \le x) - \Phi(x), \tag{4.3}$$

which is given by

$$f_x(w) = \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)\bigl\{1 - \Phi(x)\bigr\}, & w \le x, \\ \sqrt{2\pi}e^{w^2/2}\Phi(x)\bigl\{1 - \Phi(w)\bigr\}, & w > x. \end{cases} \tag{4.4}$$

Set $f_{x,y} = f_x - f_y$ for any $x, y \in \mathbb{R}$, $\delta = (\beta_2 + \beta_3)/2$ and

$$\Delta_{1,\delta} = \Delta_1 - \delta, \qquad \Delta_{2,\delta} = \Delta_2 + \delta, \qquad \Delta_{1,\delta}^{(i)} = \Delta_1^{(i)} - \delta, \qquad \Delta_{2,\delta}^{(i)} = \Delta_2^{(i)} + \delta.$$

Noting that ξ_i and $(\Delta_1^{(i)}, \Delta_2^{(i)}, W^{(i)} = W - \xi_i)$ are independent and $\mathbb{E}\xi_i = 0$ for $i = 1, \ldots, n$, we have

$$\begin{aligned}
\mathbb{E}\{W f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\} &= \sum_{i=1}^n \mathbb{E}\{\xi_i f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\} \\
&= \sum_{i=1}^n \mathbb{E}\big[\xi_i \{f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W) - f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W^{(i)})\}\big] \\
&\quad + \sum_{i=1}^n \mathbb{E}\big[\xi_i \{f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W^{(i)}) - f_{\Delta_{2,\delta}^{(i)},\Delta_{1,\delta}^{(i)}}(W^{(i)})\}\big] \\
&:= H_1 + H_2.
\end{aligned} \tag{4.5}$$

By (4.4),

$$\frac{\partial}{\partial x} f_x(w) = \begin{cases} -e^{(w^2-x^2)/2}\Phi(w), & w \le x, \\ e^{(w^2-x^2)/2}\{1 - \Phi(w)\}, & w > x. \end{cases}$$

Clearly, $\sup_{x,w} |\frac{\partial}{\partial x} f_x(w)| \le 1$ and it follows that

$$|H_2| \le \sum_{i=1}^n \sum_{j=1}^2 \mathbb{E}\big|\xi_i \{\Delta_j - \Delta_j^{(i)}\}\big|. \tag{4.6}$$

As for H_1, let $\hat{k}_i(t) = \xi_i\{I(-\xi_i \le t \le 0) - I(0 < t \le -\xi_i)\}$ satisfying $\hat{k}_i(t) \ge 0$ and $\int_{\mathbb{R}} \hat{k}_i(t)\,dt = \xi_i^2$. Observe by (4.3) that

$$\begin{aligned}
&\xi_i \{f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W) - f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W^{(i)})\} \\
&\quad = \xi_i \int_{-\xi_i}^0 f'_{\Delta_{2,\delta},\Delta_{1,\delta}}(W + t)\,dt \\
&\quad = \int_{\mathbb{R}} f'_{\Delta_{2,\delta},\Delta_{1,\delta}}(W + t)\hat{k}_i(t)\,dt \\
&\quad = \int_{\mathbb{R}} (W + t) f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W + t)\hat{k}_i(t)\,dt \\
&\qquad + \xi_i^2 \{\Phi(\Delta_{1,\delta}) - \Phi(\Delta_{2,\delta})\} + \int_{\mathbb{R}} I(\Delta_{1,\delta} \le W + t \le \Delta_{2,\delta})\hat{k}_i(t)\,dt.
\end{aligned}$$

Adding up over $1 \le i \le n$ gives

$$\begin{aligned} H_1 &= \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} (W+t) f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W+t)\hat{k}_i(t)\,dt + \mathbb{E}\big[V^2\big\{\Phi(\Delta_{1,\delta}) - \Phi(\Delta_{2,\delta})\big\}\big] \\ &\quad + \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} I(\Delta_{1,\delta} \le W+t \le \Delta_{2,\delta})\hat{k}_i(t)\,dt \\ &:= H_{11} + H_{12} + H_{13} \end{aligned} \tag{4.7}$$

for V^2 given in (4.1). Following the proof of (10.59)–(10.61) in [10] (or see (5.6)–(5.8) in [12]), we have

$$H_{13} \ge (1/2)\mathbb{P}(\Delta_1 \le W \le \Delta_2) - \delta, \tag{4.8}$$

where $\delta = (\beta_2 + \beta_3)/2$. Assume that $\delta \le 1/8$. Otherwise, (4.2) is trivial. To finish the proof of (4.2), in view of (4.5), (4.6), (4.7) and (4.8), it suffices to show that

$$|H_{12}| \le 0.6\mathbb{E}|\Delta_2 - \Delta_1| + \beta_2 + 0.5\beta_3 \tag{4.9}$$

and

$$\mathbb{E}\big\{W f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\big\} - H_{11} \le 1.75\mathbb{E}|\Delta_2 - \Delta_1| + 7\beta_2 + 6\beta_3. \tag{4.10}$$

Next we prove (4.9) and (4.10), starting with (4.9).

Proof of (4.9). Recall that $\Delta_1 \le \Delta_2$ and $\sum_{i=1}^n \mathbb{E}\xi_i^2 = 1$. Let $\bar{\xi}_i = \xi_i I(|\xi_i| \le 1)$, we have

$$\begin{aligned} |H_{12}| &= \mathbb{E}\big[V^2\big\{\Phi(\Delta_2) - \Phi(\Delta_1)\big\}\big] \\ &\le \sum_{i=1}^n \mathbb{E}\xi_i^2 I\big(|\xi_i| > 1\big) + \mathbb{E}\left[\big\{\Phi(\Delta_2) - \Phi(\Delta_1)\big\}\sum_{i=1}^n \xi_i^2 I\big(|\xi_i| \le 1\big)\right] \\ &= \beta_2 + \mathbb{E}\big[\big\{\Phi(\Delta_2) - \Phi(\Delta_1)\big\}\big]\sum_{i=1}^n \mathbb{E}\bar{\xi}_i^2 + \mathbb{E}\left[\big\{\Phi(\Delta_2) - \Phi(\Delta_1)\big\}\sum_{i=1}^n \big(\bar{\xi}_i^2 - \mathbb{E}\bar{\xi}_i^2\big)\right] \\ &\le \beta_2 + \frac{1}{\sqrt{2\pi}}\mathbb{E}(\Delta_2 - \Delta_1) + \mathbb{E}\left\{\min\left(1, \frac{\Delta_2 - \Delta_1}{\sqrt{2\pi}}\right)\left|\sum_{i=1}^n \big(\bar{\xi}_i^2 - \mathbb{E}\bar{\xi}_i^2\big)\right|\right\} \\ &\le \beta_2 + \frac{1}{\sqrt{2\pi}}\mathbb{E}(\Delta_2 - \Delta_1) + \frac{1}{2}\mathbb{E}\min\left(1, \frac{\Delta_2 - \Delta_1}{\sqrt{2\pi}}\right)^2 + \frac{1}{2}\mathbb{E}\left\{\sum_{i=1}^n \big(\bar{\xi}_i^2 - \mathbb{E}\bar{\xi}_i^2\big)\right\}^2 \\ &\le \beta_2 + \frac{1}{\sqrt{2\pi}}\mathbb{E}(\Delta_2 - \Delta_1) + \frac{1}{2\sqrt{2\pi}}\mathbb{E}(\Delta_2 - \Delta_1) + \frac{1}{2}\beta_3 \\ &\le 0.6\mathbb{E}(\Delta_2 - \Delta_1) + \beta_2 + 0.5\beta_3, \end{aligned}$$

as desired. □

Proof of (4.10). Observe that

$$\begin{aligned}
&\mathbb{E}\{Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\} - H_{11} \\
&\quad = \mathbb{E}\{Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)(1-V^2)\} \\
&\qquad + \sum_{i=1}^{n} \mathbb{E}\int \{Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W) - (W+t)f_{\Delta_{2,\delta},\Delta_{1,\delta}}(W+t)\}\hat{k}_i(t)\,dt \\
&\quad := H_{31} + H_{32}.
\end{aligned} \tag{4.11}$$

Recall that $\sup_{x,w} |\frac{\partial}{\partial x} f_x(w)| \le 1$. This, together with the following basic properties of $f_x(w)$ (see, e.g., Lemma 2.3 in [10])

$$|wf_x(w)| \le 1, \qquad |f_x(w)| \le 1, \tag{4.12}$$

$$|wf_x(w) - (w+t)f_x(w+t)| \le \min\{1, (|w| + \sqrt{2\pi}/4)|t|\} \tag{4.13}$$

and $|f_{x,y}(w)| \le |x-y|$, yields

$$\begin{aligned}
H_{31} &= \mathbb{E}\left[Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\sum_{i=1}^{n}\{\mathbb{E}\xi_i^2 I(|\xi_i|>1) - \xi_i^2 I(|\xi_i|>1)\}\right] \\
&\quad + \mathbb{E}\left\{Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)\sum_{i=1}^{n}(\mathbb{E}\bar{\xi}_i^2 - \bar{\xi}_i^2)\right\} \\
&\le 2\beta_2 + 2\mathbb{E}\left\{I(\Delta_2 - \Delta_1 > 1)\left|\sum_{i=1}^{n}(\mathbb{E}\bar{\xi}_i^2 - \bar{\xi}_i^2)\right|\right\} \\
&\quad + \mathbb{E}\left\{Wf_{\Delta_{2,\delta},\Delta_{1,\delta}}(W)I(\Delta_2 - \Delta_1 \le 1)\sum_{i=1}^{n}(\mathbb{E}\bar{\xi}_i^2 - \bar{\xi}_i^2)\right\} \\
&\le 2\beta_2 + \mathbb{E}(\Delta_2 - \Delta_1) + \beta_3 \\
&\quad + \mathbb{E}\left\{|W|(2\delta + \Delta_2 - \Delta_1)I(\Delta_2 - \Delta_1 \le 1)\sum_{i=1}^{n}(\mathbb{E}\bar{\xi}_i^2 - \bar{\xi}_i^2)\right\} \\
&\le 2\beta_2 + \mathbb{E}(\Delta_2 - \Delta_1) + \beta_3 + 0.5\mathbb{E}\{(2\delta + \Delta_2 - \Delta_1)^2 I(\Delta_2 - \Delta_1 \le 1)\} \\
&\quad + 0.5\mathbb{E}\left[W^2\left\{\sum_{i=1}^{n}(\mathbb{E}\bar{\xi}_i^2 - \bar{\xi}_i^2)\right\}^2\right] \\
&\le 2\beta_2 + \mathbb{E}(\Delta_2 - \Delta_1) + \beta_3 + 2\delta^2 + 0.75\mathbb{E}(\Delta_2 - \Delta_1) + 2\beta_3 \\
&\le 2.125\beta_2 + 3.125\beta_3 + 1.75\mathbb{E}(\Delta_2 - \Delta_1),
\end{aligned} \tag{4.14}$$

where we used the facts that $\delta \le 1/8$,

$$\mathbb{E}\left\{\sum_{i=1}^n (\bar\xi_i^2 - \mathbb{E}\bar\xi_i^2)\right\}^2 \le \beta_3 \quad \text{and} \quad \mathbb{E}\left\{W \sum_{i=1}^n (\mathbb{E}\bar\xi_i^2 - \bar\xi_i^2)\right\}^2 \le 4\beta_3.$$

To see this, set $U = \sum_{i=1}^n \eta_i$ with $\eta_i = \bar\xi_i^2 - \mathbb{E}\bar\xi_i^2$, then by standard calculations,

$$\mathbb{E}U^2 = \sum_{i=1}^n \mathbb{E}\eta_i^2 \le \sum_{i=1}^n \mathbb{E}\bar\xi_i^4 \le \sum_{i=1}^n \mathbb{E}|\bar\xi_i|^3 = \beta_3$$

and

$$\mathbb{E}(W^2U^2) = \sum_{i,j,k,\ell} \mathbb{E}(\xi_i\xi_j\eta_k\eta_\ell) = \sum_{i=1}^n \mathbb{E}(\xi_i^2\eta_i^2) + \sum_{i\ne j} \mathbb{E}\xi_i^2\mathbb{E}\eta_j^2 + 2\sum_{i\ne j}\mathbb{E}\xi_i\eta_i\mathbb{E}\xi_j\eta_j \le 4\beta_3.$$

As for H_{32}, by (4.13)

$$\begin{aligned} H_{32} &\le \sum_{i=1}^n \mathbb{E}\int_{\mathbb{R}} 2\min\{1, (|W| + \sqrt{2\pi}/4)|t|\}\hat k_i(t)\,dt \\ &\le 2\sum_{i=1}^n \mathbb{E}\int_{|t|>1} \hat k_i(t)\,dt + 2\sum_{i=1}^n \mathbb{E}\int_{|t|\le 1} (|W| + \sqrt{2\pi}/4)|t|\hat k_i(t)\,dt \\ &\le 2\beta_2 + \mathbb{E}\left\{(|W| + \sqrt{2\pi}/4)\sum_{i=1}^n |\xi_i|\min(1, \xi_i^2)\right\} \\ &\le 2\beta_2 + \mathbb{E}\left[(|W| + \sqrt{2\pi}/4)\left\{\sum_{i=1}^n |\xi_i| I(|\xi_i| > 1) + \sum_{i=1}^n |\bar\xi_i|^3\right\}\right] \\ &\le 2\beta_2 + (2 + \sqrt{2\pi}/4)(\beta_2 + \beta_3) \\ &\le 4.7\beta_2 + 2.7\beta_3, \end{aligned} \tag{4.15}$$

where we used the inequalities

$$\mathbb{E}\{|W| \cdot |\xi_i| I(|\xi_i| > 1)\} \le \mathbb{E}|W^{(i)}| \cdot \mathbb{E}|\xi_i| I(|\xi_i| > 1) + \mathbb{E}\xi_i^2 I(|\xi_i| > 1) \le 2\mathbb{E}\xi_i^2 I(|\xi_i| > 1)$$

and $\mathbb{E}(|W| \cdot |\bar\xi_i|^3) \le \mathbb{E}|W^{(i)}| \cdot \mathbb{E}|\bar\xi_i|^3 + \mathbb{E}\bar\xi_i^4 \le 2\mathbb{E}|\bar\xi_i|^3$. Combining (4.11), (4.14) and (4.15) yields (4.10). □

□

5. Proof of Theorem 2.1

5.1. Main idea of the proof

Observe that V_n is close to 1 and $1 + D_{2n} > 0$. Remember that we are interested in a particular type of nonlinear process that can be written as a linear process plus a negligible remainder. Intuitively, the leading term of the normalizing factor should be a quadratic process, say V_n^2. The key idea of the proof is to first transform $V_n(1 + D_{2n})^{1/2}$ to $(V_n^2 + 1)/2 + D_{2n}$ plus a small term and then apply the method of conjugated distributions and the randomized concentration inequality (4.2). It follows from the elementary inequalities

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

that $(1 + D_{2n})^{1/2} \ge 1 + \min(D_{2n}, 0)$, which leads to

$$\begin{aligned} V_n(1 + D_{2n})^{1/2} &\ge V_n + V_n \min(D_{2n}, 0) \\ &\ge 1 + \big(V_n^2 - 1\big)/2 - \big(V_n^2 - 1\big)^2/2 + V_n \min(D_{2n}, 0) \\ &\ge V_n^2/2 + 1/2 - \big(V_n^2 - 1\big)^2/2 + \big\{1 + \big(V_n^2 - 1\big)/2\big\} \min(D_{2n}, 0) \\ &\ge V_n^2/2 + 1/2 - \big(V_n^2 - 1\big)^2 + \min(D_{2n}, 0). \end{aligned} \tag{5.1}$$

Using the inequality $2ab \le a^2 + b^2$ yields the reverse inequality

$$V_n(1 + D_{2n})^{1/2} \le (1 + D_{2n})/2 + V_n^2/2 = V_n^2/2 + 1/2 + D_{2n}/2.$$

Consequently, for any $x > 0$,

$$\begin{aligned} \{T_n \ge x\} &\subseteq \big\{W_n + D_{1n} \ge x\big(V_n^2/2 + 1/2 - \big(V_n^2 - 1\big)^2 + D_{2n} \wedge 0\big)\big\} \\ &= \big[xW_n - x^2V_n^2/2 \ge x^2/2 - x\big\{x\big(V_n^2 - 1\big)^2 + D_{1n} + xD_{2n} \wedge 0\big\}\big] \end{aligned} \tag{5.2}$$

and

$$\{T_n \ge x\} \supseteq \big\{xW_n - x^2V_n^2/2 \ge x^2/2 + x(xD_{2n}/2 - D_{1n})\big\}. \tag{5.3}$$

Proof of (2.6). By (5.2), we have for $x \ge 1$,

$$\begin{aligned} &\mathbb{P}(T_n \ge x) \\ &\quad \le \mathbb{P}\big\{W_n \ge xV_n(1 + D_{2n} \wedge 0) - D_{1n}, |D_{1n}| \le V_n/4x, |D_{2n}| \le 1/4x^2\big\} \\ &\qquad + \mathbb{P}\big(|D_{1n}|/V_n > 1/4x\big) + \mathbb{P}\big(|D_{2n}| > 1/4x^2\big) \\ &\quad \le \mathbb{P}\big(xW_n - x^2V_n^2/2 \ge x^2/2 - x\Delta_{1n}\big) + \mathbb{P}\big\{W_n \ge (x - 1/2x)V_n, \big|V_n^2 - 1\big| > 1/2x\big\} \\ &\qquad + \mathbb{P}\big(|D_{1n}|/V_n > 1/4x\big) + \mathbb{P}\big(|D_{2n}| > 1/4x^2\big), \end{aligned} \tag{5.4}$$

where

$$\Delta_{1n} = \min\{x(V_n^2 - 1)^2 + |D_{1n}| + xD_{2n} \wedge 0, 1/x\}. \tag{5.5}$$

Consequently, (2.6) follows from the next two propositions. We postpone the proofs to Section 5.2. □

Proposition 5.1. *There exist positive absolute constants C_1, C_2 such that*

$$\mathbb{P}(xW_n - x^2V_n^2/2 \geq x^2/2 - x\Delta_{1n}) \leq \{1 - \Phi(x)\}\exp(C_1L_{n,x})(1 + C_2R_{n,x}) \tag{5.6}$$

holds for $x \geq 1$ satisfying (2.7) and (2.8).

Proposition 5.2. *There exist positive absolute constants C_3, C_4 such that*

$$\mathbb{P}(W_n/V_n \geq x - 1/2x, |V_n^2 - 1| > 1/2x) \leq C_3\{1 - \Phi(x)\}\exp(C_4L_{n,x})L_{n,x} \tag{5.7}$$

holds for all $x \geq 1$.

Proof of (2.5). By (5.3),

$$\mathbb{P}(T_n \geq x) \geq \mathbb{P}(xW_n - x^2V_n^2/2 \geq x^2/2 + x\Delta_{2n}), \tag{5.8}$$

where $\Delta_{2n} = xD_{2n}/2 - D_{1n}$. Then (2.5) follows directly from the following proposition.

Proposition 5.3. *There exist positive absolute constants C_5, C_6 such that*

$$\mathbb{P}(xW_n - x^2V_n^2/2 \geq x^2/2 + x\Delta_{2n}) \geq \{1 - \Phi(x)\}\exp(-C_5L_{n,x})(1 - C_6R_{n,x}) \tag{5.9}$$

for $x \geq 1$ satisfying (2.7) and (2.8).

The proof of Theorem 2.1 is then complete. □

5.2. Proof of Propositions 5.1, 5.2 and 5.3

For two sequences of real numbers a_n and b_n, we write $a_n \lesssim b_n$ if there is a universal constant C such that $a_n \leq Cb_n$ holds for all n. Throughout this section, $C, C_1, C_2, \ldots$ denote positive constants that are independent of n. We start with some preliminary lemmas. The first two lemmas are Lemmas 5.1 and 5.2 in [23]. Let X be a random variable such that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, and set

$$\delta_1 = \mathbb{E}X^2I(|X| > 1) + \mathbb{E}|X|^3I(|X| \leq 1).$$

Lemma 5.1. *For $0 \leq \lambda \leq 4$ and $0.25 \leq \theta \leq 4$, we have*

$$\mathbb{E}e^{\lambda X - \theta X^2} = 1 + (\lambda^2/2 - \theta)\mathbb{E}X^2 + O(1)\delta_1, \tag{5.10}$$

where $O(1)$ is bounded by an absolute constant.

Lemma 5.2. *Let $Y = X - X^2/2$. Then for $0.25 \le \lambda \le 4$, we have*

$$\mathbb{E}e^{\lambda Y} = 1 + (\lambda^2/2 - \lambda/2)\mathbb{E}X^2 + O(1)\delta_1,$$
$$\mathbb{E}(Ye^{\lambda Y}) = (\lambda - 1/2)\mathbb{E}X^2 + O(1)\delta_1,$$
$$\mathbb{E}(Y^2 e^{\lambda Y}) = \mathbb{E}X^2 + O(1)\delta_1,$$
$$\mathbb{E}(|Y|^3 e^{\lambda Y}) = O(1)\delta_1 \quad \text{and} \quad \{\mathbb{E}(Ye^{\lambda Y})\}^2 = O(1)\delta_1,$$

where the $O(1)$'s are bounded by an absolute constant. In particular, when $\lambda = 1$, we have

$$e^{-5.5\delta_1} \le \mathbb{E}e^Y \le e^{2.65\delta_1}. \tag{5.11}$$

Lemma 5.3. *Let $Y = X - X^2/2$, $Z = X^2 - \mathbb{E}X^2$ and write*

$$\delta_{11} = \mathbb{E}X^2 I(|X| > 1), \qquad \delta_{12} = \mathbb{E}|X|^3 I(|X| \le 1).$$

Then

$$|\mathbb{E}(Ze^Y)| \le 4.2\delta_{11} + 1.5\delta_{12}, \tag{5.12}$$
$$\mathbb{E}(Z^2 e^Y) \le 4\delta_{11} + 2\delta_{12} + 2\delta_{11}^2, \tag{5.13}$$
$$\mathbb{E}(|YZ|e^Y) \le 2\delta_{11} + \delta_{12}, \tag{5.14}$$
$$\mathbb{E}(|Y|Z^2 e^Y) \le 3.1\delta_{11} + \delta_{12} + \delta_{11}^2. \tag{5.15}$$

Proof. See the Appendix. □

The next lemma provides an estimate of $I_{n,x}$ given in (2.3).

Lemma 5.4. *Let ξ_i be independent random variables satisfying (2.1) and let $L_{n,x}$ be defined as in (2.3). Then there exists an absolute positive constant C such that*

$$I_{n,x} = \exp\{O(1)L_{n,x}\} \tag{5.16}$$

for all $x \ge 1$, where $|O(1)| \le C$.

Proof. Applying (5.11) in Lemma 5.1 to $X = x\xi_i$ and $Y = X - X^2/2$ yields (5.16) with $|O(1)| \le 5.5$. □

Our proof is based on the following method of conjugated distributions or the change of measure technique (Petrov [28]), which can be traced back to Harald Cramér in 1938. Let ξ_i be independent random variables and $g(x)$ be a measurable function satisfying $\mathbb{E}e^{g(\xi_i)} < \infty$. Let $\hat{\xi}_i$ be independent random variables with the distribution functions given by

$$\mathbb{P}(\hat{\xi}_i \le y) = \frac{1}{\mathbb{E}e^{g(\xi_i)}}\mathbb{E}\{e^{g(\xi_i)} I(\xi_i \le y)\}.$$

Then, for any measurable function $f:\mathbb{R}^n\to\mathbb{R}$ and any Borel measurable set C,

$$\mathbb{P}\big\{f(\xi_1,\dots,\xi_n)\in C\big\}=\prod_{i=1}^n \mathbb{E}e^{g(\xi_i)}\times\mathbb{E}\big[e^{-\sum_{i=1}^n g(\hat{\xi}_i)}I\big\{f(\hat{\xi}_1,\dots,\hat{\xi}_n)\in C\big\}\big].$$

See, for example, [23] and [33] for the applications of the change of measure method in deriving moderate deviations.

Proof of Proposition 5.1. Let $Y_i=g(\xi_i)=\xi_{i,x}-\xi_{i,x}^2/2$ with $\xi_{i,x}=x\xi_i$, and let $\hat{\xi}_1,\dots,\hat{\xi}_n$ be independent random variables with $\hat{\xi}_i$ having the distribution function

$$V_i(y)=\mathbb{E}\big\{e^{Y_i}I(\xi_i\le y)\big\}/\mathbb{E}e^{Y_i},\qquad y\in\mathbb{R}.$$

Put $\widehat{Y}_i=g(\hat{\xi}_i)=x\hat{\xi}_i-x^2\hat{\xi}_i^2/2$ and recall that $xW_n-x^2V_n^2/2=\sum_{i=1}^n Y_i:=S_Y$. Then using the method of conjugated distributions gives

$$\begin{aligned}
&\mathbb{P}\big(xW_n-x^2V_n^2/2\ge x^2/2-x\Delta_{1n}\big)\\
&\quad=\mathbb{P}\left\{\sum_{i=1}^n g(\xi_i)\ge x^2-x\Delta_{1n}(\xi_1,\dots,\xi_n)\right\}\\
&\quad=\prod_{i=1}^n\mathbb{E}e^{Y_i}\times\mathbb{E}\big\{e^{-\widehat{S}_Y}I\big(\widehat{S}_Y\ge x^2/2-x\widehat{\Delta}_{1n}\big)\big\}\\
&\quad:=I_{n,x}\times H_n,
\end{aligned}\tag{5.17}$$

where $\widehat{S}_Y=\sum_{i=1}^n\widehat{Y}_i$, $H_n=\mathbb{E}\{e^{-\widehat{S}_Y}I(\widehat{S}_Y\ge x^2/2-x\widehat{\Delta}_{1n})\}$ and $\widehat{\Delta}_{1n}=\Delta_{1n}(\hat{\xi}_1,\dots,\hat{\xi}_n)$.

Set

$$m_n=\sum_{i=1}^n\mathbb{E}\widehat{Y}_i,\qquad \sigma_n^2=\sum_{i=1}^n\operatorname{Var}(\widehat{Y}_i)\quad\text{and}\quad \upsilon_n=\sum_{i=1}^n\mathbb{E}|\widehat{Y}_i|^3.$$

Then it follows from the definition of $\hat{\xi}_i$ that

$$\begin{aligned}
\mathbb{E}\widehat{Y}_i&=\mathbb{E}\big(Y_ie^{Y_i}\big)/\mathbb{E}e^{Y_i},\\
\operatorname{Var}(\widehat{Y}_i)&=\mathbb{E}\big(Y_i^2e^{Y_i}\big)/\mathbb{E}e^{Y_i}-(\mathbb{E}\widehat{Y}_i)^2,\\
\mathbb{E}|\widehat{Y}_i|^3&=\mathbb{E}\big(|Y_i|^3e^{Y_i}\big)/\mathbb{E}e^{Y_i}.
\end{aligned}$$

Applying Lemma 5.3 with $X=x\xi_i$ and $\lambda=1$ yields

$$\begin{aligned}
\mathbb{E}e^{Y_i}&=e^{O(1)\delta_{i,x}},\qquad \mathbb{E}\big(Y_ie^{Y_i}\big)=\big(x^2/2\big)\mathbb{E}\xi_i^2+O(1)\delta_{i,x},\\
\mathbb{E}\big(Y_i^2e^{Y_i}\big)&=x^2\mathbb{E}\xi_i^2+O(1)\delta_{i,x},\qquad \mathbb{E}\big(|Y_i|^3e^{Y_i}\big)=O(1)\delta_{i,x}
\end{aligned}\tag{5.18}$$

and $\{\mathbb{E}(Y_i e^{Y_i})\}^2 = O(1)\delta_{i,x}$. In view of (5.11) and (2.7), using a similar argument as in the proof of (7.11)–(7.13) in [23] gives

$$m_n = \sum_{i=1}^n \mathbb{E}\big(Y_i e^{Y_i}\big)/\mathbb{E}e^{Y_i} = x^2/2 + O(1)L_{n,x}, \tag{5.19}$$

$$\sigma_n^2 = \sum_{i=1}^n \big\{\mathbb{E}\big(Y_i^2 e^{Y_i}\big)/\mathbb{E}e^{Y_i} - (\mathbb{E}\widehat{Y}_i)^2\big\} = x^2 + O(1)L_{n,x}, \tag{5.20}$$

$$v_n = \sum_{i=1}^n \mathbb{E}\big(|Y_i|^3 e^{Y_i}\big)/\mathbb{E}e^{Y_i} = O(1)L_{n,x}, \tag{5.21}$$

where all of the $O(1)$'s appeared above are bounded by an absolute constant, say C_1. Taking into account the condition (2.8), we have $\sigma_n^2 \ge x^2/2$, provided the constant c_1 in (2.8) is sufficiently large, say, no larger than $(4C_1)^{-1}$.

Define the standardized sum $\widehat{W} := \widehat{W}_n = (\widehat{S}_Y - m_n)/\sigma_n$, and let

$$\varepsilon_n = \sigma_n^{-1}\big(x^2/2 - m_n\big), \qquad r_n = \varepsilon_n + \sigma_n.$$

By (5.19)–(5.21) and (2.8) with $c_1 \le (4C_1)^{-1}$,

$$|\varepsilon_n| \le \sqrt{2}C_1 x^{-1} L_{n,x}, \qquad v_n\sigma_n^{-3} \le \sqrt{8}C_1 x^{-3} L_{n,x}, \tag{5.22}$$

$$|r_n - x| \le |\varepsilon_n| + \big|\sigma_n^2 - x^2\big|/(\sigma_n + x) \le 2C_1 x^{-1} L_{n,x} \le x/2, \tag{5.23}$$

which leads to

$$H_n \le \mathbb{E}\big\{\exp(-\sigma_n \widehat{W} - m_n) I(\widehat{W} - \varepsilon_n \ge -x\widehat{\Delta}_{1n}/\sigma_n)\big\} \le H_{1n} + H_{2n} \tag{5.24}$$

with $H_{1n} = \mathbb{E}\{\exp(-\sigma_n \widehat{W} - m_n) I(\widehat{W} \ge \varepsilon_n)\}$ and

$$H_{2n} = \mathbb{E}\big\{\exp(-\sigma_n \widehat{W} - m_n) I(-x\widehat{\Delta}_{1n}/\sigma_n \le \widehat{W} - \varepsilon_n < 0)\big\}.$$

Denote by G_n the distribution function of $\widehat{W}$, then H_{1n} reads as

$$\begin{aligned} H_{1n} &= \int_{\varepsilon_n}^\infty e^{-\sigma_n t - m_n}\, dG_n(t) \\ &= e^{-x^2/2} \int_0^\infty e^{-\sigma_n s}\, dG_n(s + \varepsilon_n) \\ &= e^{-x^2/2} \bigg\{ \int_0^\infty e^{-\sigma_n s}\, d\big\{G_n(s+\varepsilon_n) - \Phi(s+\varepsilon_n)\big\} + \int_0^\infty e^{-\sigma_n s}\, d\Phi(s+\varepsilon_n) \bigg\} \\ &:= e^{-x^2/2}(J_{1n} + J_{2n}). \end{aligned} \tag{5.25}$$

Using integration by parts for the Lebesgue–Stieltjes integral, the Berry–Esseen inequality, (5.22) and the following upper and lower tail inequalities for the standard normal distribution

$$\frac{t}{1+t^2}e^{-t^2/2} \le \int_t^\infty e^{-u^2/2}\,du \le \frac{1}{t}e^{-t^2/2} \qquad \text{for } t>0, \tag{5.26}$$

we have

$$|J_{1n}| \le 2\sup_{t\in\mathbb{R}}\big|G_n(t)-\Phi(t)\big| \le 4\upsilon_n\sigma_n^{-3} \le C_2 e^{x^2/2}\big\{1-\Phi(x)\big\}x^{-2}L_{n,x}.$$

For J_{2n}, by the change of variables we have

$$J_{2n} = \frac{e^{-\varepsilon_n^2/2}}{\sqrt{2\pi}}\int_0^\infty \exp\big\{-(\sigma_n+\varepsilon_n)t-t^2/2\big\}\,dt = \frac{e^{-\varepsilon_n^2/2}}{\sqrt{2\pi}}\Psi(r_n),$$

where

$$\Psi(x) = \frac{1-\Phi(x)}{\Phi'(x)} = e^{x^2/2}\int_x^\infty e^{-t^2/2}\,dt.$$

By (5.26),

$$\Psi(s) \ge \frac{s}{1+s^2} \quad \text{and} \quad 0 < -\Psi'(s) = 1 - se^{s^2/2}\int_s^\infty e^{-t^2/2}\,dt \le \frac{1}{1+s^2} \qquad \text{for } s\ge 0.$$

In view of (5.23), $x/2 \le r_n \le 3x/2$. Consequently, $|\Psi(r_n)-\Psi(x)| \le 4|r_n-x|/(4+x^2)$, which further implies that

$$J_{2n} \le \frac{1}{\sqrt{2\pi}}\left\{\Psi(x)+\frac{4}{4+x^2}|r_n-x|\right\} \le e^{x^2/2}\big\{1-\Phi(x)\big\}\big(1+C_3x^{-2}L_{n,x}\big).$$

By (5.25) and the above upper bounds for J_{1n} and J_{2n},

$$H_{1n} \le \big\{1-\Phi(x)\big\}\big(1+C_4x^{-2}L_{n,x}\big). \tag{5.27}$$

As for H_{2n}, note that $x\widehat{\Delta}_{1n} \le 1$ by (5.5). Therefore,

$$H_{2n} \le e^{1-x^2/2} \times \mathbb{P}(\varepsilon_n - x\widehat{\Delta}_{1n}/\sigma_n \le \widehat{W} < \varepsilon_n). \tag{5.28}$$

Applying inequality (4.2) to the standardized sum $\widehat{W}$ gives

$$\begin{aligned}
&\mathbb{P}(\varepsilon_n - x\widehat{\Delta}_{1n}/\sigma_n \le \widehat{W} \le \varepsilon_n)\\
&\quad\le 17\upsilon_n\sigma_n^{-3} + 5x\sigma_n^{-1}\mathbb{E}|\widehat{\Delta}_{1n}| + 2x\sigma_n^{-2}\sum_{i=1}^n \mathbb{E}\big|\widehat{Y}_i\big\{\widehat{\Delta}_{1n}-\widehat{\Delta}_{1n}^{(i)}\big\}\big|,
\end{aligned} \tag{5.29}$$

where $\widehat{\Delta}_{1n}^{(i)}$ can be any random variable that is independent of $\hat{\xi}_i$. By (5.22), it is readily known that $\upsilon_n\sigma_n^{-3} \le \sqrt{8}C_1x^{-3}L_{n,x}$. For the other two terms, recall that the distribution function of $\hat{\xi}_i$

is given by $V_i(y) = \mathbb{E}\{e^{Y_i} I(\xi_i \le y)\}/\mathbb{E}e^{Y_i}$ with $Y_i = g(\xi_i)$. Then

$$\begin{aligned}\mathbb{E}|\widehat{\Delta}_{1n}| &= \int \cdots \int \Delta_{1n}(x_1, \dots, x_n)\, dV_1(x_1) \cdots dV_n(x_n) \\ &= I_{n,x}^{-1} \int \cdots \int \Delta_{1n}(x_1, \dots, x_n) \prod_{i=1}^{n} \{e^{g(x_i)}\, dF_{\xi_i}(x_i)\} \\ &= I_{n,x}^{-1} \times \mathbb{E}\big(|\Delta_{1n}| e^{\sum_{i=1}^{n} Y_i}\big).\end{aligned} \tag{5.30}$$

It can be similarly obtained that for each $i = 1, \dots, n$,

$$\mathbb{E}\big|\widehat{Y}_i\big\{\widehat{\Delta}_{1n} - \widehat{\Delta}_{1n}^{(i)}\big\}\big| = I_{n,x}^{-1} \times \mathbb{E}\big[\big|Y_i\big\{\Delta_{1n} - \Delta_{1n}^{(i)}\big\}\big| e^{\sum_{j=1}^{n} Y_j}\big]. \tag{5.31}$$

Assembling (5.28)–(5.31), we obtain from (5.26) that

$$\begin{aligned}H_{2n} &\le C_5\big\{1 - \Phi(x)\big\}\Bigg(x^{-2} L_{n,x} + I_{n,x}^{-1} \times x\mathbb{E}\big(|\Delta_{1n}| e^{\sum_{j=1}^{n} Y_j}\big) \\ &\quad + I_{n,x}^{-1} \sum_{i=1}^{n} \mathbb{E}\big[\big|Y_i\big\{\Delta_{1n} - \Delta_{1n}^{(i)}\big\}\big| e^{\sum_{j=1}^{n} Y_j}\big]\Bigg) \\ &\le C_5\big\{1 - \Phi(x)\big\}\Bigg[x^{-2} L_{n,x} + I_{n,x}^{-1} \times x\mathbb{E}\big(|\Delta_{1n}| e^{\sum_{j=1}^{n} Y_j}\big) \\ &\quad + 2I_{n,x}^{-1} \sum_{i=1}^{n} \mathbb{E}\big\{\min\big(|\xi_{i,x}|, 1\big)\big|\Delta_{1n} - \Delta_{1n}^{(i)}\big| e^{\sum_{j\ne i}^{n} Y_j}\big\}\Bigg],\end{aligned}$$

where the last step follows from the inequality $|t - t^2/2| e^{t - t^2/2} \le 2\min(1, |t|)$ for $t \in \mathbb{R}$.

Recall that $\Delta_{1n} \le x(V_n^2 - 1)^2 + |D_{1n}| + x|D_{2n}|$. To finish the proof of (5.6), we only need to consider the contribution from $x(V_n^2 - 1)^2$. For notational convenience, let $Z_i = \xi_i^2 - \mathbb{E}\xi_i^2$ for $1 \le i \le n$, such that $V_n^2 - 1 = \sum_{i=1}^{n} Z_i$ and

$$\big(V_n^2 - 1\big)^2 - \big\{\big(V_n^2 - 1\big)^2\big\}^{(i)} = Z_i^2 + 2Z_i \cdot \sum_{j \ne i} Z_j.$$

By Lemma 5.5, (5.28) and (5.29),

$$H_{2n} \le C_6\big\{1 - \Phi(x)\big\}\big\{R_{n,x} + x^{-2} L_{n,x}(1 + L_{n,x}) e^{C_7 \max_i \delta_{i,x}}\big\}. \tag{5.32}$$

Together, (5.17), (5.24), (5.27), (5.32) and Lemma 5.4 prove (5.6). □

Lemma 5.5. *For $x \ge 1$, we have*

$$\mathbb{E}\big\{\big(V_n^2 - 1\big)^2 e^{\sum_{j=1}^{n} Y_j}\big\} \lesssim I_{n,x} \times x^{-4} L_{n,x}(1 + L_{n,x}) \tag{5.33}$$

and

$$\sum_{i=1}^n \mathbb{E}\left\{\left|Y_i\left(Z_i^2 + 2Z_i \sum_{j\neq i} Z_j\right)\right| e^{\sum_{j=1}^n Y_j}\right\} \lesssim I_{n,x} \times x^{-4} L_{n,x}(1+L_{n,x}). \tag{5.34}$$

Proof. Recall that $V_n^2 - 1 = \sum_{i=1}^n Z_i$. By independence,

$$\begin{aligned}
&\mathbb{E}\left\{\left(\sum_{i=1}^n Z_i\right)^2 e^{\sum_{j=1}^n Y_j}\right\} \\
&= \sum_{i=1}^n \mathbb{E}\big(Z_i^2 e^{Y_i}\big)\mathbb{E}e^{\sum_{j\neq i} Y_j} + \sum_{i\neq j} \mathbb{E}\big(Z_i e^{Y_i}\big)\cdot \mathbb{E}\big(Z_j e^{Y_j}\big)\cdot \mathbb{E}e^{\sum_{k=1,k\neq i,j}^n Y_k} \qquad (5.35)\\
&= I_{n,x}\left\{\sum_{i=1}^n \mathbb{E}\big(Z_i^2 e^{Y_i}\big)/\mathbb{E}e^{Y_i} + \sum_{i\neq j} \mathbb{E}\big(Z_i e^{Y_i}\big)\cdot \mathbb{E}\big(Z_j e^{Y_j}\big)/\big(\mathbb{E}e^{Y_i}\mathbb{E}e^{Y_j}\big)\right\}.
\end{aligned}$$

It follows from Lemma 5.3 that $|\mathbb{E}(Z_i e^{Y_i})| \lesssim x^{-2}\delta_{i,x}$ and $\mathbb{E}(Z_i^2 e^{Y_i}) \lesssim x^{-4}(\delta_{i,x} + \delta_{i,x}^2)$. Substituting these into (5.35) proves (5.33) in view of (5.11).

Again, applying Lemma 5.3 gives us

$$\mathbb{E}\big(|Z_i Y_i| e^{Y_i}\big) \lesssim x^{-2}\delta_{i,x} \quad \text{and} \quad \mathbb{E}\big(Z_i^2 |Y_i| e^{Y_i}\big) \lesssim x^{-4}\big(\delta_{i,x} + \delta_{i,x}^2\big),$$

which together with Hölder's inequality imply

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\left\{\left|Y_i\left(Z_i^2 + 2Z_i \sum_{j\neq i} Z_j\right)\right| e^{\sum_{j=1}^n Y_j}\right\} \\
&\quad \lesssim I_{n,x} \times x^{-4} L_{n,x}(1+L_{n,x}) \\
&\qquad + 2\sum_{i=1}^n \mathbb{E}\big(|Z_i Y_i| e^{Y_i}\big)\left\{\mathbb{E}\left(\sum_{j\neq i} Z_j\right)^2 e^{\sum_{j\neq i} Y_j}\right\}^{1/2} \cdot \big(\mathbb{E}e^{\sum_{j\neq i} Y_j}\big)^{1/2} \\
&\quad \lesssim I_{n,x} \times x^{-4} L_{n,x}(1+L_{n,x}),
\end{aligned}$$

where we use (5.33) in the last step. This completes the proof of (5.34). □

Proof of Proposition 5.2. This proof is similar to the argument used in [31]. First, consider the following decomposition:

$$\begin{aligned}
&\mathbb{P}\big(W_n/V_n \geq x - 1/2x, \big|V_n^2 - 1\big| > 1/2x\big) \\
&\quad \leq \mathbb{P}\big\{W_n/V_n \geq x - 1/2x, (1+1/2x)^{1/2} < V_n \leq 4\big\} \\
&\qquad + \mathbb{P}\big\{W_n/V_n \geq x - 1/2x, V_n < (1-1/2x)^{1/2}\big\} \qquad (5.36)
\end{aligned}$$

$$
\begin{aligned}
&+ \mathbb{P}(W_n/V_n \geq x - 1/2x, V_n > 4) \\
&:= \sum_{\nu=1}^{3} \mathbb{P}\big\{(W_n, V_n) \in \mathcal{E}_\nu\big\},
\end{aligned}
$$

where $\mathcal{E}_\nu \subseteq \mathbb{R} \times \mathbb{R}^+$, $1 \leq \nu \leq 3$ are given by

$$
\begin{aligned}
\mathcal{E}_1 &= \big\{(u, v) \in \mathbb{R} \times \mathbb{R}^+ : u/v \geq x - 1/2x, \sqrt{1 + 1/2x} < v \leq 4\big\}, \\
\mathcal{E}_2 &= \big\{(u, v) \in \mathbb{R} \times \mathbb{R}^+ : u/v \geq x - 1/2x, v < \sqrt{1 - 1/2x}\big\}, \\
\mathcal{E}_3 &= \big\{(u, v) \in \mathbb{R} \times \mathbb{R}^+ : u/v \geq x - 1/2x, v > 4\big\}.
\end{aligned}
$$

To bound the probability $\mathbb{P}\{(W_n, V_n) \in \mathcal{E}_1\}$, put $t_1 = x\sqrt{1 + 1/2x}$ and $\lambda_1 = t_1(x - 1/2x)/8$. By Markov's inequality,

$$
\mathbb{P}\big\{(W_n, V_n) \in \mathcal{E}_1\big\} \leq x^2 e^{-\inf_{(u,v)\in\mathcal{E}_1}(t_1 u - \lambda_1 v^2)} \mathbb{E}\big\{\big(V_n^2 - 1\big)^2 e^{t_1 W_n - \lambda_1 V_n^2}\big\},
$$

where it can be easily verified that

$$
\inf_{(u,v)\in\mathcal{E}_1} \big(t_1 u - \lambda_1 v^2\big) = x^2 + x/2 - \lambda_1(1 + 1/x) - 1/2 - 1/4x.
$$

However, recall that $V_n^2 - 1 = \sum_{i=1}^n Z_i$ with $Z_i = \xi_i^2 - \mathbb{E}\xi_i^2$, it follows from the independence and (5.10) that

$$
\begin{aligned}
&\mathbb{E}\big\{\big(V_n^2 - 1\big)^2 e^{t_1 W_n - \lambda_1 V_n^2}\big\} \\
&\quad = \sum_{i=1}^{n} \mathbb{E}\big(Z_i^2 e^{t_1 \xi_i - \lambda_1 \xi_i^2}\big) \times \prod_{j \neq i} \mathbb{E}\big(e^{t_1 \xi_j - \lambda_1 \xi_j^2}\big) \\
&\qquad + \sum_{i \neq j} \mathbb{E}\big(Z_i e^{t_1 \xi_i - \lambda_1 \xi_i^2}\big) \mathbb{E}\big(Z_j e^{t_1 \xi_j - \lambda_1 \xi_j^2}\big) \times \prod_{k \neq i, j} \mathbb{E}\big(e^{t_1 \xi_k - \lambda_1 \xi_k^2}\big) \\
&\quad \lesssim x^{-4} L_{n,x}(1 + L_{n,x}) \exp\big(t_1^2/2 - \lambda_1 + C L_{n,x}\big),
\end{aligned}
\tag{5.37}
$$

where we use the fact $t_1^2/2 - \lambda_1 > 0$. Consequently,

$$
\begin{aligned}
&\mathbb{P}\big\{(W_n, V_n) \in \mathcal{E}_1\big\} / \big\{1 - \Phi(x)\big\} \\
&\lesssim x^{-2} L_{n,x}(1 + L_{n,x}) \exp(-3x/8 + C L_{n,x}) \lesssim L_{n,x} \exp(-3x/8 + C L_{n,x}).
\end{aligned}
\tag{5.38}
$$

Likewise, we can bound the probability $\mathbb{P}\{(W_n, V_n) \in \mathcal{E}_2\}$ by using (t_2, λ_2) instead of (t_1, λ_1), given by

$$
t_2 = x\sqrt{1 - 1/2x}, \qquad \lambda_2 = 2x^2 - 1.
$$

Note that $\inf_{(u,v)\in\mathcal{E}_2}(t_2u-\lambda_2v^2)=x^2-x/2-1/2+1/4x-\lambda_2(1-1/2x)$. Together with (5.37), this yields

$$\begin{aligned}&\mathbb{P}\{(W_n,V_n)\in\mathcal{E}_2\}/\{1-\Phi(x)\}\\&\quad\lesssim x^{-2}L_{n,x}(1+L_{n,x})\exp(-3x/4+CL_{n,x})\lesssim L_{n,x}\exp(-3x/4+CL_{n,x}).\end{aligned}\tag{5.39}$$

For the last term $\mathbb{P}\{(W_n,V_n)\in\mathcal{E}_3\}$, we use a truncation technique and the probability estimation of binomial distribution. Let $\widehat{W}_n=\sum_{i=1}^n\xi_iI(x\xi_i\le a_0)$, where a_0 is an absolute constant to be determined (see (5.43)). Observe that

$$\begin{aligned}\mathbb{P}\{(W_n,V_n)\in\mathcal{E}_3\}&\le\mathbb{P}\Bigg(\widehat{W}_n\ge 2x-1/x,\sum_{i=1}^n\xi_i^2I\big(x|\xi_i|\le 1\big)\ge 3\Bigg)\\&\quad+\mathbb{P}\Bigg(\widehat{W}_n\ge 2x-1/x,\sum_{i=1}^n\xi_i^2I\big(x|\xi_i|>1\big)\ge 13\Bigg)\\&\quad+\mathbb{P}\Bigg(\sum_{i=1}^n\xi_iI\{x\xi_i>a_0\}\ge(x-1/2x)V_n/2\Bigg)\\&:=J_{3n}+J_{4n}+J_{5n}.\end{aligned}$$

Let

$$\bar{V}_n^2=\sum_{i=1}^n\bar{\xi}_i^2\qquad\text{with }\bar{\xi}_i=\xi_iI\big(x|\xi_i|\le 1\big),1\le i\le n,$$

such that

$$\begin{aligned}J_{3n}&=\mathbb{P}\big(\widehat{W}_n\ge 2x-1/x,\bar{V}_n^2\ge 3\big)\le(\sqrt{e}/4)e^{-x^2}\mathbb{E}\big\{\big(\bar{V}_n^2-1\big)^2e^{x\widehat{W}_n/2}\big\}\\&\le e^{-x^2}\Bigg(\mathbb{E}\Bigg[\Bigg\{\sum_{i=1}^n\big(\bar{\xi}_i^2-\mathbb{E}\bar{\xi}_i^2\big)\Bigg\}^2e^{x\widehat{W}_n/2}\Bigg]+x^{-4}L_{n,x}^2\mathbb{E}e^{x\widehat{W}_n/2}\Bigg).\end{aligned}$$

Noting that $\mathbb{E}\{\xi_iI(x\xi_i\ge a_0)\}=-\mathbb{E}\{\xi_iI(x\xi_i>a_0)\}\le 0$ for every i, and

$$e^s\le 1+s+s^2/2+|s|^3e^{\max(s,0)}/6\qquad\text{for all }s,$$

we obtain

$$\begin{aligned}\mathbb{E}e^{x\widehat{W}_n/2}&\le\prod_{i=1}^n\Bigg[1+\frac{x^2}{8}\mathbb{E}\xi_i^2+\frac{e^{a_0/2}x^3}{48}\mathbb{E}\big\{|\xi_i|^3I\big(|x\xi_i|\le a_0\big)\big\}\Bigg]\\&\le\prod_{i=1}^n\Bigg\{1+\frac{x^2}{8}\mathbb{E}\xi_i^2+\frac{e^{a_0/2}x^3}{48}\mathbb{E}|\xi_i|^3I\big(x|\xi_i|\le 1\big)+\frac{a_0e^{a_0/2}x^2}{48}\mathbb{E}\xi_i^2I\big(x|\xi_i|>1\big)\Bigg\}\\&\le\exp\big\{x^2/8+O(1)L_{n,x}\big\}.\end{aligned}\tag{5.40}$$

Similar to the proof of (5.37), it follows that

$$J_{3n} \lesssim x^{-4} L_{n,x}(1+L_{n,x}) \exp\{-7x^2/8 + O(1)L_{n,x}\}. \tag{5.41}$$

To bound J_{4n}, let $\widehat{W}_n^{(i)} = \widehat{W}_n - \xi_i I(x\xi_i \le a_0)$, then applying (5.40) gives, for any i,

$$\mathbb{E}e^{x\widehat{W}_n^{(i)}/2} \le \exp\{x^2/8 + O(1)L_{n,x}\}.$$

Subsequently,

$$\begin{aligned} J_{4n} &\le (\sqrt{e}/13)e^{-x^2} \sum_{i=1}^n \mathbb{E}\{\xi_i^2 e^{(x/2)\xi_i I(x\xi_i \le a_0)} I(x|\xi_i| > 1)\} \times \mathbb{E}e^{x\widehat{W}_n^{(i)}/2} \\ &\le \big(\sqrt{e^{1+a_0}}/13\big) x^{-2} L_{n,x} \exp\{-7x^2/8 + O(1)L_{n,x}\}. \end{aligned} \tag{5.42}$$

Finally, we study J_{5n}. By Cauchy's inequality,

$$\begin{aligned} J_{5n} &\le \mathbb{P}\left\{\sum_{i=1}^n I\big(|x\xi_i| > a_0\big) \ge (x-1/2x)^2/4\right\} \\ &\le \frac{4e^{-(x-1/2x)^2}}{(x-1/2x)^2} \sum_{i=1}^n \mathbb{E}\{e^{4I(|x\xi_i|>a_0)} I\big(|x\xi_i| > a_0\big)\} \times \prod_{j\ne i} \mathbb{E}e^{4I(|x\xi_j|>a_0)} \\ &\lesssim x^{-2}e^{-x^2} \sum_{i=1}^n e^4 \mathbb{P}\big(|x\xi_i| > a_0\big) \times \prod_{j\ne i}\{1 + e^4\mathbb{P}\big(|x\xi_j| > a_0\big)\} \\ &\lesssim a_0^{-2} \exp\{\big(e^4 a_0^{-2} - 1\big)x^2\} \sum_{i=1}^n \mathbb{E}\xi_i^2 I\big(x|\xi_i| > 1\big) \\ &\lesssim x^{-2} L_{n,x} \exp\big(-x^2/2 - x^2/22\big) \end{aligned} \tag{5.43}$$

by letting $a_0 = 11$.

Adding up (5.41)–(5.43), we get

$$\mathbb{P}\{(W_n, V_n) \in \mathcal{E}_3\} \lesssim \{1 - \Phi(x)\} L_{n,x} \exp(CL_{n,x}).$$

This, together with (5.38) and (5.39) yields (5.7). □

Proof of Proposition 5.3. Retain the notation in the proof of Proposition 5.1, and recall that $\Delta_{2n} = xD_{2n}/2 - D_{1n}$, $\widehat{W} = \sum_{i=1}^n \widehat{Y}_i$. Analogous to (5.17) and (5.24), we see that

$$\begin{aligned} &\mathbb{P}\big(xW_n - x^2V_n^2/2 \ge x^2/2 + x\Delta_{2n}\big) \\ &\quad = I_{n,x}\mathbb{E}\{e^{-\widehat{W}} I\big(\widehat{W} \ge x^2/2 + x\widehat{\Delta}_{2n}\big)\} \end{aligned} \tag{5.44}$$

$$\begin{aligned}
&\geq I_{n,x}\big[\mathbb{E}\big\{\exp(-\sigma_n\widehat{W}-m_n)I(\widehat{W}\geq\varepsilon_n)\big\}\\
&\quad-\mathbb{E}\big\{\exp(-\sigma_n\widehat{W}-m_n)I(\varepsilon_n\leq\widehat{W}<\varepsilon_n+x\widehat{\Delta}_{2n}/\sigma_n)\big\}\big]\\
&\geq I_{n,x}\left\{\int_{\varepsilon_n}^{\infty}e^{-\sigma_n t-m_n}\,dG_n(t)-e^{-x^2/2}\mathbb{P}(\varepsilon_n\leq\widehat{W}<\varepsilon_n+x\widehat{\Delta}_{2n}/\sigma_n)\right\}\\
&:=I_{n,x}\big(H_{1n}-H'_{2n}\big),
\end{aligned}$$

for H_{1n} given in (5.24), and where $\varepsilon_n=\sigma_n^{-1}(x^2/2-m_n)$,

$$\widehat{\Delta}_{2n}=\Delta_{2n}(\hat{\xi}_1,\ldots,\hat{\xi}_n),\qquad H'_{2n}=e^{-x^2/2}\mathbb{P}(\varepsilon_n\leq\widehat{W}<\varepsilon_n+x\widehat{\Delta}_{2n}/\sigma_n).$$

Following the proof of (5.27), it can be similarly obtained that

$$H_{1n}\geq\big\{1-\Phi(x)\big\}\big(1-Cx^{-2}L_{n,x}\big). \tag{5.45}$$

Replacing $\widehat{\Delta}_{1n}$ with $\widehat{\Delta}_{2n}$ in (5.28) and using the same argument that leads to (5.32) implies

$$H'_{2n}\leq C\big\{1-\Phi(x)\big\}R_{n,x}. \tag{5.46}$$

Substituting (5.16), (5.45) and (5.46) into (5.44) proves (5.9). □

6. Proof of Theorem 3.1

Throughout this section, we use $C, C_1, C_2, \ldots$ and $c, c_1, c_2, \ldots$ to denote positive constants that are independent of n.

6.1. Outline of the proof

Put $\tilde{h}=(h-\theta)/\sigma$ and $\tilde{h}_1=(h_1-\theta)/\sigma$, such that $\tilde{h}_1(x)=\mathbb{E}\{\tilde{h}(X_1,X_2,\ldots,X_m)|X_1=x\}$ and $\tilde{h}_1(X_1),\ldots,\tilde{h}_1(X_n)$ are i.i.d. random variables with zero means and unit variances. Using this notation, condition (3.3) can be written as

$$\tilde{h}^2(x_1,\ldots,x_m)\leq c_0\left\{\tau+\sum_{i=1}^m\tilde{h}_1^2(x_i)\right\}. \tag{6.1}$$

By the scale-invariance property of Studentized U-statistics, we can replace, respectively, h and h_1 with $\tilde{h}$ and $\tilde{h}_1$, which does not change the definition of T_n. For ease of exposition, we still use h and h_1 but assume without loss of generality that $\mathbb{E}h_{1i}=0$ and $\mathbb{E}h_{1i}^2=1$, where $h_{1i}:=h_1(X_i)$ for $i=1,\ldots,n$.

For s_1^2 given in (3.2), observe that

$$\frac{(n-m)^2}{(n-1)}s_1^2=\sum_{i=1}^n(q_i-U_n)^2=\sum_{i=1}^n q_i^2-nU_n^2.$$

Define

$$T_n^* = \frac{\sqrt{n}}{ms_1^*} U_n, \qquad s_1^{*2} = \frac{(n-1)}{(n-m)^2} \sum_{i=1}^n q_i^2, \tag{6.2}$$

then by the definition of T_n,

$$T_n = T_n^* \Big/ \left(1 - \frac{m^2(n-1)}{(n-m)^2} T_n^{*2}\right)^{1/2},$$

such that for any $x \geq 0$,

$$\{T_n \geq x\} = \bigl\{T_n^* \geq x/\bigl(1 + x^2 m^2 (n-1)/(n-m)^2\bigr)^{1/2}\bigr\}. \tag{6.3}$$

Therefore, we only need to focus on T_n^*, instead of T_n.

To reformulate $T_n^* = \sqrt{n} U_n/(ms_1^*)$ in the form of (2.2), set

$$W_n = \sum_{i=1}^n \xi_i, \qquad V_n^2 = \sum_{i=1}^n \xi_i^2, \tag{6.4}$$

where $\xi_i = n^{-1/2} h_{1i}$ for $1 \leq i \leq n$. Moreover, put

$$r(x_1, \dots, x_m) = h(x_1, \dots, x_m) - \sum_{i=1}^m h_1(x_i). \tag{6.5}$$

For U_n, using Hoeffding's decomposition gives $\sqrt{n} U_n / m = W_n + D_{1n}$, where

$$D_{1n} = \frac{\sqrt{n}}{m\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} r(X_{i_1}, \dots, X_{i_m}). \tag{6.6}$$

However, a direct calculation shows that $s_1^2 = V_n^2(1 + D_{2n})$, where

$$\begin{aligned} (n-1) D_{2n} = 1 + V_n^{-2} \Biggl\{ & \frac{1}{\binom{n-2}{m-1}^2} \Lambda_n^2 + \frac{(m-1)\{(m+1)n - 2m\}n}{(n-m)^2} W_n^2 \\ & + \frac{2\sqrt{n}}{\binom{n-2}{m-1}} \sum_{i=1}^n \xi_i \psi_i + \frac{2m(m-1)n}{(n-m)^2} W_n D_{1n} \Biggr\}, \end{aligned} \tag{6.7}$$

$$\Lambda_n^2 = \sum_{i=1}^n \psi_i^2, \qquad \psi_i = \sum_{\substack{1 \leq \ell_1 < \cdots < \ell_{m-1} \leq n \\ \ell_j \neq i, j = 1, \dots, m-1}} r(X_i, X_{\ell_1}, \dots, X_{\ell_{m-1}}). \tag{6.8}$$

In particular, (6.7) generalizes (2.5) in [26] for $m = 2$. Combining the above decompositions of U_n and s_1^2, we obtain

$$T_n^* = \frac{W_n + D_{1n}}{V_n(1 + D_{2n})^{1/2}}. \tag{6.9}$$

To prove (3.4), by (6.3), it is sufficient to show that there exists a constant $C > 1$ independent of n such that

$$\mathbb{P}\big(T_n^* \ge x\big) \le \big\{1 - \Phi(x)\big\}e^{CL_{n,1+x}}\left\{1 + C(\sqrt{a_m} + \sigma_h)\frac{(1+x)^3}{\sqrt{n}}\right\} \tag{6.10}$$

and

$$\mathbb{P}\big(T_n^* \ge x\big) \ge \big\{1 - \Phi(x)\big\}e^{-CL_{n,1+x}}\left\{1 - C(\sqrt{a_m} + \sigma_h)\frac{(1+x)^3}{n^{1/2}}\right\} \tag{6.11}$$

hold uniformly for

$$0 \le x \le C^{-1}\min\big\{(\sigma/\sigma_p)n^{1/2-1/p}, (n/a_m)^{1/6}\big\}, \tag{6.12}$$

where $L_{n,x} = n\mathbb{E}\xi_{1,x}^2 I(|\xi_{1,x}| > 1) + n\mathbb{E}|\xi_{1,x}|^3 I(|\xi_{1,x}| \le 1)$ with $\xi_{i,x} = x\xi_i$ for $x \ge 1$.

The main strategy of proving (6.10) and (6.11) is to first partition the probability space into two parts, say $\mathcal{G}_{n,x}$ and its complement $\mathcal{G}_{n,x}^c$ such that $\mathbb{P}(\mathcal{G}_{n,x}^c)$ is sufficiently small, then find a tight upper bound for the tail probability of $|D_{2n}|$ on $\mathcal{G}_{n,x}$, and finally apply Theorem 2.1.

First, by Lemma 3.3 of [26], $\mathbb{P}(V_n^2 \le \sigma^2/2) \le \exp\{-n/(32a^2)\}$ for all $n \ge 1$, where $a > 0$ is such that $\mathbb{E}h_{1i}^2 I(|h_{1i}| \ge a\sigma) \le \sigma^2/4$. In particular, we take

$$a = 4^{1/(p-2)}(\sigma_p/\sigma)^{p/(p-2)} \le (2\sigma_p/\sigma)^{p/(p-2)}.$$

Then it follows from the inequality that $\sup_{2<p\le 3}\sup_{s\ge 0}(s^{p/2-1}e^{-s}) \le 1$ and (5.26) that (recall that $\sigma^2 = 1$)

$$\mathbb{P}\big(V_n^2 \le 1/2\big) \le C_1\big\{1 - \Phi(x)\big\}(\sigma_p/\sigma)^p(1+x)n^{1-p/2} \tag{6.13}$$

for all $0 \le x \le c_1(\sigma/\sigma_1)n^{p/2-1}$. We can therefore regard $\{V_n^2\}_{n\ge 1}$ as a sequence of positive random variables that are uniformly bounded away from zero. For W_n/V_n, applying Lemma 6.4 in [23] implies that for every $t > 0$,

$$\mathbb{P}\big\{|W_n| \ge t(4 + V_n)\big\} \le 4\exp\big(-t^2/2\big). \tag{6.14}$$

In view of (6.13) and (6.14), define the subset

$$\mathcal{G}_{n,x} = \big\{|W_n| \le \sqrt{x}n^{1/4}(4 + V_n), V_n^2 \ge 1/2\big\}, \tag{6.15}$$

such that

$$\mathbb{P}\big(\mathcal{G}_{n,x}^c\big) \le C_2\big\{1 - \Phi(x)\big\}(\sigma_p/\sigma)^p(1+x)n^{1-p/2} \tag{6.16}$$

holds uniformly for

$$0 \le x \le c_2 \min\{(\sigma/\sigma_1)n^{p/2-1}, \sqrt{n}\}. \tag{6.17}$$

Next, we restrict our attention to the subset $\mathcal{G}_{n,x}$. Recall the definition of D_{2n} in (6.7). For any $\varepsilon > 0$, we have

$$\left|\sum_{i=1}^{n} \xi_i \psi_i\right| \le (4\varepsilon)^{-1} V_n^2 + \varepsilon \Lambda_n^2. \tag{6.18}$$

In particular, taking $\varepsilon = \sigma/(xn^{m-1}\sigma_h)$ for σ_h^2 as in (6.18) yields

$$\begin{aligned} |D_{2n}| \le C_3\{&\sigma_h x n^{-1/2} + (\sigma_h x)^{-1} n^{3/2-2m} V_n^{-2} \Lambda_n^2 \\ &+ n^{-1}(W_n/V_n)^2 + n^{-1} V_n^{-2} |W_n||D_{1n}|\}. \end{aligned} \tag{6.19}$$

In addition to the subset $\mathcal{G}_{n,x}$ given in (6.15), put

$$\mathcal{E}_{n,x} = \mathcal{G}_{n,x} \cap \{|D_{1n}|/V_n \le 1/4x\}. \tag{6.20}$$

Together, (6.19) and (6.20) imply that

$$|D_{2n}| \le C_4\{\sigma_h x n^{-1/2} + (\sigma_h x)^{-1} n^{3/2-2m} \Lambda_n^2\} := D_{3n} \tag{6.21}$$

holds on $\mathcal{E}_{n,x}$ for all $1 \le x \le \sqrt{n}$.

Proof of (6.10). By (2.6), Remark 2.2, (6.9), (6.19) and condition (6.17), we have

$$\begin{aligned} \mathbb{P}(T_n^* \ge x) \le {} & \{1 - \Phi(x)\} e^{C_5 L_{n,x}} (1 + C_6 R_{n,x}) \\ & + \mathbb{P}(|D_{1n}|/V_n \ge 1/4x, \mathcal{G}_{n,x}) + \mathbb{P}(|D_{2n}| \ge 1/4x^2, \mathcal{E}_{n,x}) + \mathbb{P}(\mathcal{G}_{n,x}^c) \end{aligned} \tag{6.22}$$

for all $x \ge 1$ satisfying (6.17) and

$$L_{n,x} \le c_3 x^2, \tag{6.23}$$

where $R_{n,x}$ is given in (2.4) but with D_{2n} replaced by D_{3n}. In particular, for $2 < p \le 3$, we have $L_{n,x} \le (\sigma_p/\sigma)^p x^p n^{1-p/2}$, and thus the constraint (6.23) is satisfied whenever

$$1 \le x \le (c_3^{1/p}/2)(\sigma/\sigma_p)^{1/p} n^{1/2-1/p}. \tag{6.24}$$

However, for $0 \le x \le 1$, it follows from (2.9) that

$$\mathbb{P}(T_n^* \ge x) \le \mathbb{P}(\mathcal{G}_{n,x}^c) + \{1 - \Phi(x)\}(1 + C_7 \breve{R}_{n,x}),$$

for $\breve{R}_{n,x}$ as in (2.10) with D_{2n} replaced with D_{3n}.

In view of (6.16) and (6.22), (6.10) follows directly from the following two propositions. □

Proposition 6.1. *Under condition* (3.3), *there exists a positive constant* C *independent of* n *such that*

$$\mathbb{P}\big(|D_{1n}|/V_n \ge 1/4x, \mathcal{G}_{n,x}\big) + \mathbb{P}\big(|D_{2n}| \ge 1/4x^2, \mathcal{E}_{n,x}\big) \le C\sqrt{a_m}\big\{1-\Phi(x)\big\}x^2 n^{-1/2}, \tag{6.25}$$

holds for all $x \ge 1$ *satisfying* (6.12), *where* $a_m = \max\{c_0\tau, c_0 + m\}$, $\mathcal{G}_{n,x}$ *and* $\mathcal{E}_{n,x}$ *are given in* (6.15) *and* (6.20), *respectively.*

Proposition 6.2. *There is a positive constant* C *independent of* n *such that*

$$R_{n,x} \le C\sigma_h x^3 n^{-1/2} \tag{6.26}$$

for all $x \ge 1$ *and*

$$\breve{R}_{n,x} \le C\sigma_h n^{-1/2} \tag{6.27}$$

for $0 \le x \le 1$, *where* σ_h *is given in* (3.1).

Proof of (6.11). Observe that

$$\begin{aligned}\mathbb{P}\big(T_n^* \ge x\big) &\ge \mathbb{P}\big\{W_n + D_{1n} \ge xV_n(1+D_{2n})^{1/2}, \mathcal{G}_{n,x}\big\} \\ &\ge \mathbb{P}\big\{W_n + D_{1n} \ge xV_n(1+D_{3n})^{1/2}\big\} - \mathbb{P}\big(\mathcal{G}_{n,x}^c\big).\end{aligned}$$

Then (6.11) follows from (2.5), Remark 2.2, (6.16) and Proposition 6.2. Finally, assembling (6.17) and (6.24) yields (6.12) and completes the proof of Theorem 3.1. □

6.2. Proof of Propositions 6.1 and 6.2

We begin with a technical lemma, the proof of which is presented in the Appendix.

Lemma 6.1. *There exist an absolute constant* C *and constants* B_1–B_4 *independent of* n, *such that for all* $y \ge 0$,

$$\mathbb{P}\big\{\Lambda_n^2 \ge a_m y\big(B_1 + B_2V_n^2\big)n^{2m-2}\big\} \le Ce^{-y/4} \tag{6.28}$$

and

$$\mathbb{P}\left\{\frac{|\sum_{1\le i_1<\cdots<i_m\le n} r(X_{i_1},\dots,X_{i_m})|}{\sqrt{a_m}(B_3 + B_4V_n^2)^{1/2}n^{m-1}} \ge y\right\} \le Ce^{-y/4}, \tag{6.29}$$

where $a_m = \max\{c_0\tau, c_0+m\}$, *and* V_n^2 *and* Λ_n^2 *are given in* (6.4) *and* (6.8), *respectively.*

The above lemma generalizes and improves Lemma 3.4 of [26] where $m = 2$ and the bound was of the order $ne^{-y/8}$ instead of $e^{-y/4}$. Lemma C.2 in the Appendix makes it possible to eliminate the factor n.

Proof of Proposition 6.1. By (6.19) and the definition of $\mathcal{E}_{n,x}$ in (6.20), we get

$$\mathbb{P}\big(|D_{2n}| \geq 1/4x^2, \mathcal{E}_{n,x}\big) \leq \mathbb{P}\big(\Lambda_n^2 \geq c_4 V_n^2 x^{-4} n^{2m-1}, \mathcal{G}_{n,x}\big),$$

provided that $1 \leq x \leq c_5 n^{1/4}$. Because $V_n^2 \geq 1/2$ on $\mathcal{G}_{n,x}$, it is easy to see that

$$V_n^2 \geq (2B_1 + B_2)^{-1}\big(B_1 + B_2 V_n^2\big)$$

for B_1 and B_2 as in Lemma 6.1. Therefore, taking

$$y = \frac{c_4}{2B_1 + B_2} \cdot \frac{n}{a_m x^4}$$

in (6.28) leads to

$$\mathbb{P}\big(|D_{2n}| > 1/4x^2, \mathcal{E}_{n,x}\big) \leq C \exp\big\{-c_6 n/\big(a_m x^4\big)\big\}. \tag{6.30}$$

Using (6.29), it can be similarly shown that

$$\mathbb{P}\big(|D_{1n}|/V_n > 1/4x, \mathcal{G}_{n,x}\big) \leq C \exp\big\{-c_7 n^{1/2}/\big(a_m^{1/2} x\big)\big\}. \tag{6.31}$$

Together, (6.30), (6.31) and (5.26) imply (6.25) as long as

$$1 \leq x \leq c_8 (n/a_m)^{1/6}. \tag{6.32}$$

□

Proof of Proposition 6.2. For $x \geq 0$ and $1 \leq i \leq n$, put $Y_i = x\xi_i - x^2\xi_i^2/2$, and let

$$L_k := \mathbb{E}\big(r_{1,\dots,k} e^{Y_1 + \dots + Y_k}\big), \qquad \tilde{L}_k := \mathbb{E}\big(r_{1,\dots,k} e^{Y_2 + \dots + Y_k} | X_1\big)$$

for $2 \leq k \leq m$, where $r_{1,\dots,k} := \mathbb{E}\{r(X_1, \dots, X_m)|X_1, \dots, X_k\}$ for $r(X_1, \dots, X_m)$ as in (6.5). In particular, put $r_{1,\dots,m} := r(X_1, \dots, X_m)$ and note that $\mathbb{E}r_{1,\dots,m}^2 \leq \sigma_h^2$. The following lemma provides the upper bounds for L_m and $\tilde{L}_m$.

Lemma 6.2. *For any $0 \leq x \leq \sqrt{n}/2$, we have*

$$|L_m| \leq C\sigma_h x^2 n^{-1}, \tag{6.33}$$

$$|\tilde{L}_m| \leq C\big\{E\big(r_{1,\dots,m}^2 | X_1\big)\big\}^{1/2} x n^{-1/2}. \tag{6.34}$$

We postpone the proof of Lemma 6.2 to the end of this section. Recall the definition of D_{1n} in (6.6). Using Hölder's inequality, we estimate

$$\mathbb{E}\Big\{\Big(\sum r_{i_1,\dots,i_m}\Big)^2 e^{\sum_{j=1}^n Y_j}\Big\} = \sum\sum \mathbb{E}\big(r_{i_1,\dots,i_m} r_{j_1,\dots,j_m} e^{\sum_{j=1}^n Y_j}\big).$$

Put

$$\mathcal{C} = \{(i_1, j_1, \ldots, i_m, j_m) : 1 \le i_1 \le \cdots \le i_m \le n, 1 \le j_1 < \cdot < j_m \le n\}$$
$$= \bigcup_{k=0}^{m} \{(i_1, j_1, \ldots, i_m, j_m) \in \mathcal{C} : |\{i_1, \ldots, i_m\} \cap \{j_1, \ldots, j_m\}| = k\} := \bigcup_{k=0}^{m} \mathcal{C}_k.$$

By (5.11),

$$\begin{aligned}
&\mathbb{E}\Big\{\Big(\sum r_{i_1,\ldots,i_m}\Big)^2 e^{\sum_{j=1}^n Y_j}\Big\} \\
&\quad = \sum_{k=0}^{m} \sum_{(i_1,j_1,\ldots,i_m,j_m)\in\mathcal{C}_k} \mathbb{E}\big(r_{i_1,\ldots,i_m} r_{j_1,\ldots,j_m} e^{\sum_{j=1}^n Y_j}\big) \\
&\quad = \sum_{k=0}^{m} \binom{n}{m}\binom{n-k}{m-k} \mathbb{E}\big(r_{1,\ldots,m} r_{1,\ldots,k,m+1,\ldots,2m-k} e^{\sum_{j=1}^{2m-k} Y_j}\big) \cdot \big(\mathbb{E}e^{Y_1}\big)^{n-2m+k} \\
&\quad = \binom{n}{m}^2 \big(\mathbb{E}e^{Y_1}\big)^{-2m} I_{n,x} L_m^2 + \binom{n}{m}\binom{n-1}{m-1} \big(\mathbb{E}e^{Y_1}\big)^{1-2m} I_{n,x} \mathbb{E}\big(\tilde{L}_m^2 e^{Y_1}\big) \\
&\qquad + \sum_{k=2}^{m} \binom{n}{m}\binom{n-k}{m-k} \big(\mathbb{E}e^{Y_1}\big)^{k-2m} I_{n,x} \mathbb{E}\big(r_{1,\ldots,m} r_{1,\ldots,k,m+1,\ldots,2m-k} e^{\sum_{j=1}^{2m-k} Y_j}\big) \\
&\quad \le C I_{n,x} n^{2m} \big(L_m^2 + n^{-1}\mathbb{E}\tilde{L}_m^2 + \sigma_h^2 n^{-2}\big),
\end{aligned}$$

which together with Lemma 6.2 yields for $x \ge 1$,

$$\mathbb{E}\Big\{\Big(\sum r_{i_1,\ldots,i_m}\Big)^2 e^{\sum_{j=1}^n Y_j}\Big\} \le C\sigma_h^2 I_{n,x} x^4 n^{2m-2}.$$

This, together with (6.6) gives

$$\mathbb{E}\big(|D_{1n}| e^{\sum_{j=1}^n Y_j}\big) \le C\sigma_h I_{n,x} x^2 n^{-1/2}. \tag{6.35}$$

Recall that $\psi_i = \sum_{1 \le \ell_1 \le \cdots \le \ell_{m-1} (\ne i) \le n} r(X_i, X_{\ell_1}, \ldots, X_{\ell_{m-1}})$. Then it can be similarly derived that

$$\mathbb{E}\big(\psi_i^2 e^{\sum_{j=1}^n Y_j}\big) \le C\sigma_h^2 I_{n,x} x^2 n^{2m-3}. \tag{6.36}$$

Together with (6.21), this yields

$$\mathbb{E}\big(D_{3n} e^{\sum_{j=1}^n Y_j}\big) \le C\sigma_h I_{n,x} x n^{-1/2}. \tag{6.37}$$

Next, for each $1 \le i \le n$, let $D_{1n}^{(i)}$ and $D_{3n}^{(i)}$ be obtained from D_{1n} and D_{3n}, respectively, by throwing away the summands that depend on X_i. Then, by (6.6) and (6.21), we have

$$\left|D_{1n} - D_{1n}^{(i)}\right| \le \frac{\sqrt{n}}{m\binom{n}{m}}|\psi_i|$$

and

$$\begin{aligned}
x\left|D_{3n} - D_{3n}^{(i)}\right| \\
&\le C\sigma_h^{-1} n^{-2m+3/2}\Biggl\{\psi_i^2 + \sum_{j\neq i}\Biggl(\sum_{1\le j_1<\cdots<j_{m-2}(\neq i,j)\le n} r_{i,j,j_1,\dots,j_{m-2}}\Biggr)^2 \\
&\quad + 2\sum_{j\neq i}\Biggl|\Biggl(\sum_{1\le j_1<\cdots<j_{m-2}(\neq i,j)\le n} r_{i,j,j_1,\dots,j_{m-2}}\Biggr)\Biggl(\sum_{1\le j_1<\cdots<j_{m-1}(\neq j)\le n} r_{j,j_1,\dots,j_{m-1}}\Biggr)\Biggr|\Biggr\}.
\end{aligned}$$

Using a conditional analogue of the argument that leads to (6.36) implies

$$\mathbb{E}\bigl(\psi_i^2 e^{\sum_{j\neq i} Y_j}|X_i\bigr) \le C I_{n,x} x^2 n^{2m-3} \times \mathbb{E}\bigl(r_{1,\dots,m}^2|X_i\bigr), \tag{6.38}$$

as a consequence of which (recall that $\xi_{i,x} = x\xi_i$)

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\bigl\{\min\bigl(|\xi_{i,x}|,1\bigr)\bigl|D_{1n} - D_{1n}^{(i)}\bigr|e^{\sum_{j\neq i}^n Y_j}\bigr\} \\
&\quad\le Cn^{-m+1/2}\sum_{i=1}^n \mathbb{E}\bigl[\min\bigl(|\xi_{i,x}|,1\bigr)\bigl\{\mathbb{E}\bigl(\psi_i^2 e^{\sum_{j\neq i} Y_j}|X_i\bigr)\bigr\}^{1/2}\bigl\{\mathbb{E}\bigl(e^{\sum_{j\neq i} Y_j}\bigr)\bigr\}^{1/2}\bigr] \\
&\quad\le C I_{n,x} x^2 n^{-1}\sum_{i=1}^n \bigl(\mathbb{E}\xi_i^2\bigr)^{1/2}\bigl(Er_{1,\dots,m}^2\bigr)^{1/2} \\
&\quad\le C\sigma_h I_{n,x} x^2 n^{-1/2}.
\end{aligned} \tag{6.39}$$

For the contributions from $|D_{3n} - D_{3n}^{(i)}|$, we have

$$\begin{aligned}
\mathbb{E}\bigl\{\min\bigl(|\xi_{i,x}|,1\bigr)\psi_i^2 e^{\sum_{j\neq i} Y_j}\bigr\} &= \mathbb{E}\bigl\{\min\bigl(|\xi_{i,x}|,1\bigr) \times \mathbb{E}\bigl(\psi_i^2 e^{\sum_{j\neq i} Y_j}|X_i\bigr)\bigr\} \\
&\le C I_{n,x} x^2 n^{2m-3} \times \mathbb{E}\bigl\{\min\bigl(|\xi_{i,x}|,1\bigr) r_{1,\dots,m}^2\bigr\},
\end{aligned}$$

and for each pair (i, j) such that $1 \le i \ne j \le n$,

$$\begin{aligned}
&\mathbb{E}\Bigl\{\min\bigl(|\xi_{i,x}|,1\bigr)\Bigl|\Bigl(\sum \psi_{i,j,j_1,\dots,j_{m-2}}\Bigr)\Bigl(\sum \psi_{j,j_1,\dots,j_{m-1}}\Bigr)\Bigr|e^{\sum_{k\neq i} Y_k}\Bigr\} \\
&\quad\le \mathbb{E}\Bigl[\min\bigl(|\xi_{i,x}|,1\bigr)\mathbb{E}\Bigl\{\Bigl(\sum \psi_{i,j,j_1,\dots,j_{m-2}}\Bigr)^2 e^{\sum_{k\neq i} Y_k}\Bigl|X_i\Bigr\}^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{E}\Big\{\Big(\sum \psi_{j,j_1,\ldots,j_{m-1}}\Big)^2 e^{\sum_{k\neq i} Y_k}\Big\}^{1/2}\Big] \\
&\le C I_{n,x} x^2 n^{2m-7/2} \times \mathbb{E}|\xi_i r_{1,\ldots,m}| \times \big(\mathbb{E} r_{1,\ldots,m}^2\big)^{1/2} \\
&\le C\sigma_h^2 I_{n,x} x^2 n^{2m-4},
\end{aligned}
$$

where we used (6.36) in the second step. Similarly, it can be proved that

$$
\begin{aligned}
&\mathbb{E}\Big\{\min\big(|\xi_{i,x}|,1\big)\Big(\sum r_{i,j,j_1,\ldots,j_{m-2}}\Big)^2 e^{\sum_{k\neq i} Y_k}\Big\} \\
&\quad = \mathbb{E}\Big[\min\big(|\xi_{i,x}|,1\big)\mathbb{E}\Big\{\Big(\sum r_{i,j,j_1,\ldots,j_{m-2}}\Big)^2 e^{\sum_{k\neq i} Y_k}\big|X_i\Big\}\Big] \le C\sigma_h^2 I_{n,x} n^{2m-4}.
\end{aligned}
$$

Adding up the above calculations, we get

$$
\sum_{i=1}^n \mathbb{E}\big\{x \min\big(|\xi_{i,x}|,1\big)\big|D_{3n} - D_{3n}^{(i)}\big| e^{\sum_{j\neq i} Y_j}\big\} \le C\sigma_h I_{n,x} x^2 n^{-1/2}.
$$

This, together with (6.35), (6.37) and (6.39) implies (6.26).

Finally, we consider the case of $0 \le x \le 1$. By Hölder's inequality,

$$
\mathbb{E}|D_{1n}| \le C n^{1/2} \binom{n}{m}^{-1} \Big\{\mathbb{E}\Big(\sum r_{i_1,\ldots,i_m}\Big)^2\Big\}^{1/2} \le C\sigma_h n^{-1/2} \tag{6.40}
$$

and

$$
\mathbb{E} D_{3n} \le C\big(\sigma_h n^{-1/2} + \sigma_h^{-1} n^{-2m+3/2} \mathbb{E}\Lambda_n^2\big) \le C\sigma_h n^{-1/2}. \tag{6.41}
$$

Moreover, for any pair (i, j) such that $1 \le i \neq j \le n$,

$$
\mathbb{E}\psi_i^2 \le C\sigma_h^2 n^{2m-3}, \qquad \mathbb{E}\Big(\sum \psi_{i,j,j_1,\ldots,j_{m-2}}\Big)^2 \le C\sigma_h^2 n^{2m-4}
$$

and

$$
\begin{aligned}
&\mathbb{E}\Big\{\Big|\Big(\sum r_{i,j,\ell_1,\ldots,\ell_{m-2}}\Big)\Big(\sum r_{j,j_1,\ldots,j_{m-1}}\Big)\Big|\big|X_i\Big\} \\
&\le \Big[\mathbb{E}\Big\{\Big(\sum r_{i,j,\ell_1,\ldots,\ell_{m-2}}\Big)^2\big|X_i\Big\}\Big]^{1/2} \times \Big\{\mathbb{E}\Big(\sum \psi_{j,j_1,\ldots,j_{m-1}}\Big)^2\Big\}^{1/2} \\
&\le C\sigma_h n^{2m-7/2} \times \big\{\mathbb{E}\big(r_{1,\ldots,m}^2|X_i\big)\big\}^{1/2}.
\end{aligned}
$$

Combining the above calculations, we obtain

$$
\sum_{i=1}^n \mathbb{E}\big|\xi_i\big(D_{1n} - D_{1n}^{(i)}\big)\big| \le C n^{-m+1/2} \sum_{i=1}^n \big(\mathbb{E}\xi_i^2\big)^{1/2}\big(\mathbb{E}\psi_i^2\big)^{1/2} \le C\sigma_h n^{-1/2} \tag{6.42}
$$

and

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\big|x\xi_i I\big\{|\xi_i| \le 1/(1+x)\big\}\big(D_{3n} - D_{3n}^{(i)}\big)\big| \\
&\quad \le C\sigma_h^{-1} n^{-2m+3/2} \Bigg[\sum_{i=1}^{n} \mathbb{E}\psi_i^2 + \sum_{i\neq j} \mathbb{E}\Big(\sum \psi_{i,j,j_1,\ldots,j_{m-2}}\Big)^2 \\
&\qquad + 2\sum_{i\neq j} \mathbb{E}\Big\{|\xi_i| \times \Big|\Big(\sum r_{i,j,\ell_1,\ldots,\ell_{m-2}}\Big)\Big(\sum r_{j,j_1,\ldots,j_{m-1}}\Big)\Big|\Big\}\Bigg] \\
&\quad \le C\sigma_h n^{-1/2}.
\end{aligned} \tag{6.43}$$

Assembling (6.40)–(6.43) proves (6.27) and completes the proof of Proposition 6.2. □

Proof of Lemma 6.2. We prove (6.33) by the method of induction, and (6.34) follows a similar argument. First, for $m = 2$, observe that

$$L_2 = \mathbb{E}\big(r_{1,2}e^{Y_1+Y_2}\big) = \mathbb{E}\big\{r_{1,2}\big(e^{Y_1} - 1\big)\big(e^{Y_2} - 1\big)\big\}.$$

Using the inequality

$$\big|e^{t-t^2/2} - 1\big| \le 2|t| \qquad \text{for all } t \in \mathbb{R}, \tag{6.44}$$

we have (recall that $\xi_i = n^{-1/2}h_{1i}$)

$$|L_2| \le 4x^2 n^{-1}\mathbb{E}|r_{1,2}h_{11}h_{12}| \le 4\sigma_h x^2 n^{-1}.$$

Similarly, noting that $\tilde{L}_2 = \mathbb{E}\{r_{1,2}(e^{Y_2} - 1)|X_1\}$, we get

$$|\tilde{L}_2| \le 2\big\{\mathbb{E}\big(r_{1,2}^2|X_1\big)\big\}^{1/2} x n^{-1/2},$$

as desired.

For the general case where $m > 2$, we derive

$$\begin{aligned}
&\mathbb{E}\big(r_{1,\ldots,m}e^{Y_1+\cdots+Y_m}\big) \\
&\quad = \mathbb{E}\big\{r_{1,\ldots,m}\big(e^{Y_1} - 1\big)\cdots\big(e^{Y_m} - 1\big)\big\} + \sum_{1\le i_1<\cdots<i_{m-1}\le m} \mathbb{E}\big(r_{1,\ldots,m}e^{Y_{i_1}+\cdots+Y_{i_{m-1}}}\big) \\
&\qquad - \sum_{1\le i_1<\cdots<i_{m-2}\le m} \mathbb{E}\big(r_{1,\ldots,m}e^{Y_{i_1}+\cdots+Y_{i_{m-2}}}\big) + \cdots + (-1)^{m-1} \sum_{1\le i_1<i_2\le m} \mathbb{E}\big(r_{1,\ldots,m}e^{Y_{i_1}+Y_{i_2}}\big) \\
&\quad = \mathbb{E}\big\{r_{1,\ldots,m}\big(e^{Y_1} - 1\big)\cdots\big(e^{Y_m} - 1\big)\big\} + mL_{m-1} \\
&\qquad - \binom{m}{m-2} L_{m-2} + \cdots + (-1)^{m-1}\binom{m}{2} L_2,
\end{aligned}$$

where for each k-tuple $(i_1, \dots, i_k)$ $(2 \le k \le m-1)$ satisfying $1 \le i_1 < \cdots < i_k \le m$,

$$\begin{aligned}\mathbb{E}\big(r_{1,\dots,m} e^{Y_{i_1}+\cdots+Y_{i_k}}\big) &= \mathbb{E}\big[e^{Y_{i_1}+\cdots+Y_{i_k}} \mathbb{E}\big\{r(X_1,\dots,X_m)|X_{i_1},\dots,X_{i_k}\big\}\big] \\ &= \mathbb{E}\big(r_{i_1,\dots,i_k} e^{Y_{i_1}+\cdots+Y_{i_k}}\big) = L_k,\end{aligned}$$

by definition. Using inequality (6.44) again gives

$$\big|\mathbb{E}\big\{r_{1,\dots,m}\big(e^{Y_1}-1\big)\cdots\big(e^{Y_m}-1\big)\big\}\big| \le 2^m x^m n^{-m/2} \mathbb{E}|r_{1,\dots,m} h_{11}\cdots h_{1m}| \le \sigma_h (2x)^m n^{-m/2},$$

completing the proof of (6.33) by induction and under the condition that $x \le \sqrt{n}/2$. □

Appendix A: Proof of Theorem 2.2

The main idea of the proof is to first truncate ξ_i at a suitable level, and then apply the randomized concentration inequality to the truncated variables.

For $x \ge 0$ and $i = 1, \dots, n$, define $Y_i = x\xi_i - x^2\xi_i^2/2$, and

$$\bar{\xi}_i = \xi_i I\big\{|\xi_i| \le 1/(1+x)\big\}, \qquad \bar{Y}_i = Y_i I\big\{|\xi_i| \le 1/(1+x)\big\}.$$

Moreover, put $S_Y = \sum_{i=1}^n Y_i$ and $S_{\bar{Y}} = \sum_{i=1}^n \bar{Y}_i$.

We first consider the case of $x > 0$. Proceeding as in (5.2) and (5.3), we have

$$\mathbb{P}\big(S_Y \ge x^2/2 + x\Delta_{2n}\big) \le \mathbb{P}(T_n \ge x) \le \mathbb{P}\big(S_Y \ge x^2/2 - x\Delta_{1n}\big), \tag{A.1}$$

where $\Delta_{1n} = x(V_n^2 - 1)^2 + |D_{1n}| + xD_{2n} \wedge 0$ and $\Delta_{2n} = xD_{2n}/2 - D_{1n}$. Replacing the ξ_i^2's with their truncated versions, we put $\Delta_{3n} = x(\sum_{i=1}^n \bar{\xi}_i^2 - 1)^2 + |D_{1n}| + xD_{2n} \wedge 0$, such that

$$\begin{aligned}&\big|\mathbb{P}\big(S_Y \ge x^2/2 - x\Delta_{1n}\big) - \mathbb{P}\big(S_{\bar{Y}} \ge x^2/2 - x\Delta_{3n}\big)\big| \\ &\quad \le \mathbb{P}\Big\{\max_{1\le i\le n} |\xi_i| > 1/(1+x)\Big\} \le (1+x)^2 \sum_{i=1}^n \mathbb{E}\xi_i^2 I\big\{|\xi_i| > 1/(1+x)\big\},\end{aligned} \tag{A.2}$$

and the same bound holds for $|\mathbb{P}(S_Y \ge x^2/2 + x\Delta_{2n}) - \mathbb{P}(S_{\bar{Y}} \ge x^2/2 + x\Delta_{2n})|$.

It suffices to estimate the probabilities of the truncated random variables. Consider the following decomposition:

$$\mathbb{P}\big(S_{\bar{Y}} \ge x^2/2 - x\Delta_{3n}\big) \le \mathbb{P}\big(S_{\bar{Y}} \ge x^2/2\big) + \mathbb{P}\big(x^2/2 - x\Delta_{3n} \le S_{\bar{Y}} < x^2/2\big), \tag{A.3}$$

where $S_{\bar{Y}} = \sum_{i=1}^n \bar{Y}_i$ denotes the sum of the truncated random variables. Write $\bar{m}_n = \sum_{i=1}^n \mathbb{E}\bar{Y}_i$, $\bar{\sigma}_n^2 = \sum_{i=1}^n \mathrm{Var}(\bar{Y}_i)$ and $\bar{v}_n = \sum_{i=1}^n \mathbb{E}|\bar{Y}_i|^3$. By a similar calculation to that leading to (5.18),

$$\begin{aligned}\mathbb{E}\bar{Y}_i &= -\big(x^2/2\big)\mathbb{E}\xi_i^2 + O(1)\big(x+x^2\big)\mathbb{E}\xi_i^2 I\big\{|\xi_i| > 1/(1+x)\big\}, \\ \mathbb{E}\bar{Y}_i^2 &= x^2\mathbb{E}\xi_i^2 + O(1)\big[x^2\mathbb{E}\xi_i^2 I\big\{|\xi_i| > 1/(1+x)\big\} + x^3\mathbb{E}|\bar{\xi}_i|^3\big], \\ \mathbb{E}|\bar{Y}_i|^3 &= O(1)x^3\mathbb{E}|\bar{\xi}_i|^3\end{aligned}$$

and

$$\operatorname{Var}(\bar{Y}_i) = x^2\mathbb{E}\xi_i^2 + O(1)\big[x^2\mathbb{E}\xi_i^2 I\{|\xi_i| > 1/(1+x)\} + x^3\mathbb{E}|\bar{\xi}_i|^3\big],$$

where $|O(1)| \le C_1$ for some absolute constant C_1. Combining these calculations, we have

$$\begin{aligned} \bar{m}_n &= -x^2/2 + O(1)\big(x + x^2\big)\sum_{i=1}^n \mathbb{E}\xi_i^2 I\{|\xi_i| > 1/(1+x)\}, \\ \bar{\sigma}_n^2 &= x^2 + O(1)x^2\sum_{i=1}^n \big[\mathbb{E}\xi_i^2 I\{|\xi_i| > 1/(1+x)\} + x\mathbb{E}|\bar{\xi}_i|^3\big] \ge x^2/2, \end{aligned} \tag{A.4}$$

where the last inequality holds as long as $(1+x)^{-2}L_{n,1+x} \le (2C_1)^{-1}$. Otherwise, if this constraint is violated, then (2.9) is always true provided that $C > 2C_1$.

Applying the Berry–Esseen inequality to the first addend in (A.3) gives

$$\begin{aligned} \mathbb{P}\big(S_{\bar{Y}} \ge x^2/2\big) &= 1 - \Phi(\bar{\varepsilon}_n) + O(1)\bar{v}_n\bar{\sigma}_n^{-3} \\ &= 1 - \Phi(x) + O(1)(1+x)^{-1}L_{n,1+x}, \end{aligned} \tag{A.5}$$

where $\bar{\varepsilon}_n := \bar{\sigma}_n^{-1}(x^2/2 - \bar{m}_n) = x + O(1)(1+x)^{-1}L_{n,1+x}$ by (A.4).

For the second addend in (A.3), applying the concentration inequality (4.2) to $\bar{W}_n = \bar{\sigma}_n^{-1}(S_{\bar{Y}} - \bar{m}_n)$ and noting that $|\bar{Y}_i| \le 3x|\bar{\xi}_i|/2$, we obtain

$$\begin{aligned} &\mathbb{P}\big(x^2/2 - x|\Delta_{3n}| \le S_{\bar{Y}} < x^2/2\big) \\ &\quad= \mathbb{P}(\bar{\varepsilon}_n - x\Delta_{3n}/\bar{\sigma}_n \le \bar{W}_n \le \bar{\varepsilon}_n) \\ &\quad\le 17\bar{\sigma}_n^{-3}\sum_{i=1}^n \mathbb{E}|\bar{Y}_i|^3 + 5x\bar{\sigma}_n^{-1}\mathbb{E}|\Delta_{3n}| + 2x\bar{\sigma}_n^{-2}\sum_{i=1}^n \mathbb{E}\big|\bar{Y}_i\big\{\Delta_{3n} - \Delta_{3n}^{(i)}\big\}\big| \\ &\quad\le C\Bigg[\sum_{i=1}^n \mathbb{E}|\bar{\xi}_i|^3 + \mathbb{E}|\Delta_{3n}| + \sum_{i=1}^n \mathbb{E}\big|\bar{\xi}_i\big\{\Delta_{3n} - \Delta_{3n}^{(i)}\big\}\big|\Bigg], \end{aligned} \tag{A.6}$$

where $\Delta_{3n} = x(\sum_{i=1}^n \bar{\xi}_i^2 - 1)^2 + |D_{1n}| + x|D_{2n}|$. For $i = 1, \dots, n$, put

$$\begin{aligned} d_i &= \Bigg(\sum_{i=1}^n \bar{\xi}_i^2 - 1\Bigg)^2 - \Bigg(\sum_{j\ne i} \bar{\xi}_j^2 - 1\Bigg)^2 \\ &= \bar{\xi}_i^2\Bigg[\bar{\xi}_i^2 + 2\sum_{j\ne i}\big(\bar{\xi}_j^2 - \mathbb{E}\bar{\xi}_j^2\big) - 2\mathbb{E}\bar{\xi}_i^2 - 2\sum_{i=1}^n \mathbb{E}\xi_i^2 I\big\{|\bar{\xi}_i| > 1/(1+x)\big\}\Bigg]. \end{aligned}$$

Direct calculation shows that

$$\mathbb{E}\left(\sum_{i=1}^{n}\bar{\xi}_i^2 - 1\right)^2 \le C(1+x)^{-4}\big(L_{n,1+x} + L_{n,1+x}^2\big),$$
$$\sum_{i=1}^{n}\mathbb{E}|\bar{\xi}_i d_i| \le C(1+x)^{-5}\big(L_{n,1+x} + L_{n,1+x}^2\big).$$

Substituting this into (A.6), we get

$$\begin{aligned}
&\mathbb{P}\big(x^2/2 - x|\Delta_{3n}| \le S_{\bar{Y}} < x^2/2\big) \\
&\quad \le C\Bigg[(1+x)^{-2}L_{n,1+x} + \mathbb{E}|D_{1n}| + x\mathbb{E}|D_{2n}| \\
&\qquad + \sum_{i=1}^{n}\mathbb{E}\big\{|\bar{\xi}_i|\big(\big|D_{1n} - D_{1n}^{(i)}\big| + x\big|D_{2n} - D_{2n}^{(i)}\big|\big)\big\}\Bigg].
\end{aligned}$$

This, together with (A.1), (A.2), (A.3) and (A.5) implies

$$P(T_n \le x) \le \Phi(x) + C\breve{R}_{n,x}$$

for all $x > 0$, where $\breve{R}_{n,x}$ is given in (2.10). A lower bound can be similarly obtained by noting that $\mathbb{P}(S_{\bar{Y}} \ge x^2/2 + x\Delta_{2n}) \ge \mathbb{P}(S_{\bar{Y}} \ge x^2/2) - \mathbb{P}(x^2/2 \le S_{\bar{Y}} < x^2/2 + x\Delta_{2n})$.

We next consider the case of $x = 0$. It is straightforward that

$$\begin{aligned}
&\big|P(T_n \le 0) - \Phi(0)\big| \\
&\quad = \big|\mathbb{P}(W_n + D_{1n} \le 0) - \Phi(0)\big| \le \big|\mathbb{P}(W_n \le 0) - \Phi(0)\big| + \mathbb{P}\big(-|D_{1n}| \le W_n \le |D_{1n}|\big).
\end{aligned}$$

A uniform Berry–Esseen bound (see, e.g., [11]) gives $|P(W_n \le 0) - \Phi(0)| \le 4.1L_{n,1}$. As before, we can use the truncation technique and the concentration inequality (4.2) to upper bound the probability $\mathbb{P}(-|D_{1n}| \le W_n \le |D_{1n}|)$. The rest of the proof is almost identical to that for the case of $x > 0$ and is therefore omitted.

Appendix B: Proof of Lemma 5.3

Recall that $Z = X^2 - \mathbb{E}X^2$ and $Y = X - X^2/2$. Using the inequality $|e^s - 1| \le |s|e^{s\vee 0}$ implies

$$\begin{aligned}
\mathbb{E}\big\{Ze^Y I\big(|X| \le 1\big)\big\} &= \mathbb{E}\big[Z\big\{1 + O(1)|Y|e^{Y\vee 0}\big\}I\big(|X| \le 1\big)\big] \\
&= \mathbb{E}\big\{ZI\big(|X| > 1\big)\big\} + O(1)\mathbb{E}\big\{|Z| \cdot |Y|e^{Y\vee 0}I\big(|X| \le 1\big)\big\},
\end{aligned}$$

where $|O(1)| \le 1$. Because $|Y|e^{Y\vee 0}I(|X| \le 1) \le 1.5|X|I(|X| \le 1)$, we have

$$\mathbb{E}\big\{|Z| \times |Y|e^{Y\vee 0}I\big(|X| \le 1\big)\big\} \le 1.5\mathbb{E}\big\{|X|^3 I\big(|X| \le 1\big)\big\}. \tag{B.1}$$

Note that if both f and g are increasing functions, then $\mathbb{E}f(X)\mathbb{E}g(X) \le \mathbb{E}\{f(X)g(X)\}$. In particular, we have $\mathbb{E}X^2 \times \mathbb{P}(|X|>1) \le \mathbb{E}\{|X|^2 I(|X|>1)\}$, which further implies

$$\mathbb{E}\{|Z|e^Y I(|X|>1)\} \le \sqrt{e}\mathbb{E}\{X^2 I(|X|>1)\}.$$

Together with (B.1), this yields (5.12).

For (5.13), it is straightforward that

$$\begin{aligned}
\mathbb{E}(Z^2e^Y) &= \mathbb{E}\{Z^2e^Y I(|X|\le 1)\} + \mathbb{E}\{Z^2e^Y I(|X|>1)\} \\
&\le \sqrt{e}\big[\mathbb{E}\{X^4 I(|X|\le 1)\} + (\mathbb{E}X^2)^2\mathbb{P}(|X|\le 1) - 2\mathbb{E}X^2 \times \mathbb{E}\{X^2 I(|X|\le 1)\}\big] \\
&\quad + \mathbb{E}\{X^4 e^{X-X^2/2} I(|X|>1)\} + \sqrt{e}(\mathbb{E}X^2)^2 \times \mathbb{P}(|X|>1) \\
&\le \sqrt{e}\mathbb{E}\{X^4 I(|X|\le 1)\} + 4\mathbb{E}\{X^2 I(|X|>1)\} \\
&\quad + \sqrt{e}(\mathbb{E}X^2)^2 - 2\sqrt{e}\mathbb{E}X^2 \times \mathbb{E}\{X^2 I(|X|\le 1)\} \\
&\le \sqrt{e}\mathbb{E}\{X^4 I(|X|\le 1)\} + 4\mathbb{E}\{X^2 I(|X|>1)\} \\
&\quad + \sqrt{e}\mathbb{E}X^2 \times \mathbb{E}\{X^2 I(|X|>1)\} - \sqrt{e}\mathbb{E}X^2 \times \mathbb{E}\{X^2 I(|X|\le 1)\} \\
&\le \sqrt{e}\mathbb{E}\{|X|^3 I(|X|\le 1)\} + 4\mathbb{E}\{X^2 I(|X|>1)\} + \sqrt{e}\{\mathbb{E}X^2 I(|X|>1)\}^2,
\end{aligned}$$

where in the third inequality we use the inequality $\sup_{|x|>1}\{x^2\exp(x-x^2/2)\} \le 4$.

Moreover, noting that

$$\sup_{|x|\le 1}\{(1-x/2)\exp(x-x^2/2)\} \le 1 \quad \text{and} \quad \sup_{x\in\mathbb{R}}\{|x-x^2/2|\exp(x-x^2/2)\} \le \sqrt{e}/2,$$

we obtain

$$\begin{aligned}
\mathbb{E}(|YZ|e^Y) &= \mathbb{E}\{|YZ|e^Y I(|X|\le 1)\} + \mathbb{E}\{|YZ|e^Y I(|X|>1)\} \\
&\le \mathbb{E}\{|X^2-\mathbb{E}X^2| \times |X| I(|X|\le 1)\} + \frac{\sqrt{e}}{2}\mathbb{E}\{X^2 I(|X|>1)\} \\
&\le 2\mathbb{E}\{X^2 I(|X|>1)\} + \mathbb{E}\{|X|^3 I(|X|\le 1)\},
\end{aligned}$$

which proves (5.14).

Finally, for (5.15), it follows from the inequality $\sup_{|x|>1}\{|x^3-x^4/2|\exp(x-x^2/2)\} < 3.1$ that

$$\begin{aligned}
&\mathbb{E}(|Y|Z^2e^Y) \\
&\quad = \mathbb{E}\{Z^2|Y|e^Y I(|X|\le 1)\} + \mathbb{E}\{Z^2|Y|e^Y I(|X|>1)\} \\
&\quad \le \frac{\sqrt{e}}{2}\mathbb{E}\{Z^2 I(|X|\le 1)\} + \max\Big[3.1\mathbb{E}\{X^2 I(|X|>1)\}, \frac{\sqrt{e}}{2}(\mathbb{E}X^2)^2 P(|X|>1)\Big]
\end{aligned}$$

$$\le \frac{\sqrt{e}}{2}\mathbb{E}\{|X|^3 I(|X| \le 1)\}$$
$$+ \max\left[3.1\mathbb{E}\{X^2 I(|X| > 1)\}, \frac{\sqrt{e}}{2}\mathbb{E}\{X^2 I(|X| > 1)\} + \frac{\sqrt{e}}{2}\{\mathbb{E}X^2 I(|X| > 1)\}^2\right],$$

as desired.

Appendix C: Proof of Lemma 6.1

We start with two technical lemmas. The first follows [26].

Lemma C.1. *Let $\{\xi_i, \mathcal{F}_i, i \ge 1\}$ be a sequence of martingale differences with $\mathbb{E}\xi_i^2 < \infty$, and put*

$$D_n^2 = \sum_{i=1}^n \{\xi_i^2 + 2\mathbb{E}(\xi_i^2|\mathcal{F}_{i-1}) + 3\mathbb{E}\xi_i^2\}.$$

Then we have

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \ge x D_n\right) \le \sqrt{2}\exp(-x^2/8) \tag{C.1}$$

for all $x > 0$. In particular, if $\{\xi_i, i \ge 1\}$ is a sequence of independent random variables with zero means and finite variances, write

$$S_n = \sum_{i=1}^n \xi_i, \qquad V_n^2 = \sum_{i=1}^n \xi_i^2 \quad \text{and} \quad B_n^2 = \sum_{i=1}^n \mathbb{E}\xi_i^2,$$

such that $D_n^2 = V_n^2 + 5B_n^2$. Then for any $x \ge 0$,

$$\mathbb{P}(|S_n| \ge x D_n) \le \sqrt{2}\exp(-x^2/8) \tag{C.2}$$

and

$$\mathbb{E}[S_n^2 I\{|S_n| \ge x(V_n + 4B_n)\}] \le 23 B_n^2 \exp(-x^2/4). \tag{C.3}$$

The following result may be of independent interest.

Lemma C.2. *Let $\{\xi_i, i \ge 1\}$ and $\{\eta_i, i \ge 1\}$ be two sequences of arbitrary random variables. Assume that the η_i's are non-negative, and that for any $u > 0$,*

$$\mathbb{E}\{\xi_i I(\xi_i \ge u\eta_i)\} \le c_i e^{-cu}, \tag{C.4}$$

where $\{c, c_i, i \ge 1\}$ *are positive constants. Then, for any* $u > 0$, $v > 0$ *and* $n \ge 1$,

$$\mathbb{P}\left\{\sum_{i=1}^n \xi_i \ge u\left(v + \sum_{i=1}^n \eta_i\right)\right\} \le \frac{e^{-cu}}{cu^2 v}\sum_{i=1}^n c_i. \tag{C.5}$$

Proof. For any $u > 0$ and $v > 0$, applying Markov's and Jensen's inequalities gives

$$\begin{aligned}
\text{L.H.S. of (C.5)} &\le \mathbb{P}\left\{\sum_{i=1}^n (\xi_i - u\eta_i) \ge uv\right\} \\
&\le \frac{1}{uv}\mathbb{E}\left\{\sum_{i=1}^n (\xi_i - u\eta_i)\right\}_+ \\
&\le \frac{1}{uv}\sum_{i=1}^n \mathbb{E}(\xi_i - u\eta_i)_+,
\end{aligned} \tag{C.6}$$

where $x_+ = \max(0, x)$ for all $x \in \mathbb{R}$. For each $1 \le i \le n$ fixed, it follows from (C.4) that

$$\begin{aligned}
\mathbb{E}(\xi_i - u\eta_i)_+ &= \mathbb{E}\int_{u\eta_i}^\infty I(\xi_i \ge s)\,ds \\
&= \int_1^\infty u\mathbb{E}\{\eta_i I(\xi_i \ge tu\eta_i)\}\,dt \\
&\le \int_1^\infty t^{-1}\mathbb{E}\{\xi_i I(\xi_i \ge tu\eta_i)\}\,dt \\
&\le c_i \int_1^\infty t^{-1}\exp(-cut)\,dt \le \frac{e^{-cu}}{cu}c_i,
\end{aligned}$$

which completes the proof of (C.5) by (C.6). □

To prove Lemma 6.1, we use an inductive approach by formulating the proof into three steps. Here, C and $B_1, B_2, \ldots$ denote positive constants that are independent of n. Recalling (6.1), it is easy to verify that

$$r^2(x_1, \ldots, x_m) \le 2a_m\{1 + h_1^2(x_1) + \cdots + h_1^2(x_m)\}, \tag{C.7}$$

where $a_m = \max\{c_0\tau, c_0 + m\}$. In line with (6.4), let $W_n = n^{-1/2}\sum_{i=1}^n h_{1i}$ and $V_n^2 = n^{-1}\sum_{i=1}^n h_{1i}^2$. Here, and in the sequel, we write

$$h_{1i} = h_1(X_i), \qquad h_{j,i_1,\ldots,i_j} = \mathbb{E}\{h(X_1, \ldots, X_m)|X_{i_1}, \ldots, X_{i_j}\}, \qquad 2 \le j \le m,$$

for ease of exposition. The conclusion is obvious when $0 \le y \le 2$, therefore we assume $y \ge 2$ without loss of generality.

Step 1. Let $m = 2$, then (C.7) reduces to

$$r^2(x_1, x_2) \le 2a_2\big\{1 + h_1^2(x_1) + h_1^2(x_2)\big\}, \tag{C.8}$$

where $a_2 = \max\{c_0\tau, c_0 + 2\}$. We follow the lines of the proof of Lemma 3.4 in [26] with the help of Lemma C.2.

Retaining the notation in Section 6 for $m = 2$, we have

$$\Lambda_n^2 = \sum_{i=1}^n \psi_i^2, \qquad \psi_i = \sum_{j=1, j\neq i}^n r_{i,j} = \sum_{j=1, j\neq i}^n r(X_i, X_j), \qquad 1 \le i \le n.$$

Conditional on X_i, note that ψ_i is a sum of independent random variables with zero means. To apply inequality (C.3), put

$$t_i = v_i + 4b_i, \qquad v_i^2 = \sum_{j\neq i} r_{i,j}^2, \qquad b_i^2 = \sum_{j\neq i} \mathbb{E}\big(r_{i,j}^2|X_i\big)$$

for $1 \le i \le n$. By (C.3), $\mathbb{E}\{\psi_i^2 I(\psi_i^2 \ge y t_i^2)|X_i\} \le 23 b_i^2 e^{-y/4}$. Taking expectations on both sides yields

$$\mathbb{E}\big\{\psi_i^2 I\big(\psi_i^2 \ge y t_i^2\big)\big\} \le 23(n-1)e^{-y/4}\mathbb{E}\big(r_{1,2}^2\big).$$

Applying Lemma C.2 with $\xi_i = \psi_i^2$, $\eta_i = t_i$, $u = y$ and $v = a_2 n(n-1)$ gives

$$\mathbb{P}\left\{\Lambda_n^2 \ge y\left(\sum_{i=1}^n t_i^2 + a_2 n(n-1)\right)\right\} \le C\big(a_2 y^2\big)^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2}^2\big). \tag{C.9}$$

Direct calculation based on (C.8) shows

$$\sum_{i=1}^n v_i^2 \le a_2(n-1)n\big(2 + 4V_n^2\big), \qquad \sum_{i=1}^n b_i^2 \le a_2(n-1)n\big(4 + 2V_n^2\big),$$

which further implies

$$\sum_{i=1}^n t_i^2 + a_2 n(n-1) \le 17\sum_{i=1}^n \big(v_i^2 + b_i^2\big) + a_2 n(n-1) \le a_2(n-1)n\big(103 + 102V_n^2\big).$$

Substituting this into (C.9) with $y \ge 2$ proves (6.28).

As for (6.29), let $\mathcal{F}_j = \sigma\{X_i : i \le j\}$ and write

$$\sum_{1\le i<j\le n} r_{i,j} = \sum_{j=2}^n \sum_{i=1}^{j-1} r_{i,j} = \sum_{j=2}^n R_j, \qquad R_j = \sum_{i=1}^{j-1} r_{i,j}, \qquad 2 \le j \le n.$$

Note that $\{R_j, \mathcal{F}_j, j \geq 2\}$ is a martingale difference sequence. Then using the sub-Gaussian inequality (C.1) for self-normalized martingales yields

$$\mathbb{P}\left\{\left|\sum_{1\leq i<j\leq n} r_{i,j}\right| > \sqrt{2y}\left(Q_n^2 + 2\widehat{Q}_n^2 + 3\sum_{j=2}^n \mathbb{E}R_j^2\right)^{1/2}\right\} \leq \sqrt{2}e^{-y/4}, \tag{C.10}$$

where

$$Q_n^2 = \sum_{j=2}^n R_j^2, \qquad \widehat{Q}_n^2 = \sum_{j=2}^n \mathbb{E}\big(R_j^2|\mathcal{F}_{j-1}\big).$$

Observe that Q_n^2 and Λ_n^2 have same structure, thus it can be similarly proved that

$$\mathbb{P}\big\{Q_n^2 \geq a_2 y n^2\big(102V_n^2 + 103\big)\big\} \leq C a_2^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2}^2\big). \tag{C.11}$$

For $\widehat{Q}_n^2$, write

$$\hat{t}_j = u_j + 4d_j \qquad \text{where } u_j^2 = \sum_{i=1}^{j-1} r_{i,j}^2, \qquad d_j^2 = \sum_{i=1}^{j-1} \mathbb{E}\big(r_{i,j}^2|X_j\big), \qquad 2 \leq j \leq n, \tag{C.12}$$

then it follows from a conditional analogue of (C.3) that

$$\mathbb{E}\big\{R_j^2 I\big(R_j^2 \geq y\hat{t}_j^2\big)|X_j\big\} \leq 23 d_j^2 e^{-y/4}. \tag{C.13}$$

Therefore, for $y \geq 2$,

$$\begin{aligned} &\mathbb{P}\left[\widehat{Q}_n^2 > y\left\{\sum_{j=2}^n \mathbb{E}\big(\hat{t}_j^2|\mathcal{F}_{j-1}\big) + a_2 n(n-1)\right\}\right] \\ &\quad \leq \mathbb{P}\left[\frac{\sum_{j=2}^n \mathbb{E}\{R_j^2 I(R_j^2 \leq y\hat{t}_j^2)|\mathcal{F}_{j-1}\}}{\sum_{j=2}^n \mathbb{E}(\hat{t}_j^2|\mathcal{F}_{j-1})} > y\right] \\ &\qquad + \mathbb{P}\left[\sum_{j=2}^n \mathbb{E}\big\{R_j^2 I\big(R_j^2 > y\hat{t}_j^2\big)|\mathcal{F}_{j-1}\big\} \geq y a_2 n(n-1)\right] \\ &\quad \leq \frac{1}{a_2 y n(n-1)} \sum_{j=2}^n \mathbb{E}\big\{R_j^2 I\big(R_j^2 > y\hat{t}_j^2\big)\big\} \leq C a_2^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2}^2\big), \end{aligned} \tag{C.14}$$

where in the last step we used (C.13).

For d_j^2 and u_j^2 given in (C.12), we have

$$\mathbb{E}\big(u_j^2|\mathcal{F}_{j-1}\big) = \sum_{i=1}^{j-1} \mathbb{E}\big(r_{i,j}^2|X_i\big) \leq 4a_2(j-1) + 2a_2 \sum_{i=1}^{j-1} h_{1i}^2,$$

$$\mathbb{E}\big(d_j^2|\mathcal{F}_{j-1}\big)=\sum_{i=1}^{j-1} r_{i,j}^2 \le 2a_2(j-1)+2a_2\sum_{i=1}^{j-1}\big(h_{1i}^2+h_{1j}^2\big),$$

leading to

$$\sum_{j=2}^{n}\mathbb{E}\big(\hat{t}_j^2|\mathcal{F}_{j-1}\big)\le 17\sum_{j=2}^{n}\big\{\mathbb{E}\big(u_j^2|\mathcal{F}_{j-1}\big)+\mathbb{E}\big(d_j^2|\mathcal{F}_{j-1}\big)\big\}\le a_2(n-1)n\big(104+136V_n^2\big).$$

Substituting this into (C.14) yields

$$\mathbb{P}\big\{\widehat{Q}_n^2 > a_2 y n^2\big(136V_n^2+104\big)\big\}\le Ca_2^{-1}e^{-y/4}\mathbb{E}\big(r_{1,2}^2\big). \tag{C.15}$$

Together, (C.10), (C.11), (C.15) and the identity $\sum_{j=2}^{n}\mathbb{E}R_j^2=\frac{1}{2}n(n-1)\mathbb{E}(r_{1,2}^2)$ prove (6.29).
Step 2. Assume $m=3$. By (C.7),

$$r^2(x_1,x_2,x_3)\le 2a_3\big\{1+h_1^2(x_1)+h_1^2(x_2)+h_1^2(x_3)\big\} \tag{C.16}$$

and for $r_2(x_1,x_2)=E\{r(X_1,X_2,X_3)|X_1=x_1,X_2=x_2\}$,

$$r_2^2(x_1,x_2)\le 2a_3\big\{2+h_1^2(x_1)+h_1^2(x_2)\big\}. \tag{C.17}$$

Again, starting from $\Lambda_n^2=\sum_{i=1}^{n}\psi_i^2$ with

$$\begin{aligned}\psi_i &= \sum_{\substack{1\le j<k\le n\\ j,k\ne i}} r(X_i,X_j,X_k) := \sum_{\substack{1\le j<k\le n\\ j,k\ne i}} r_{i,j,k}\\ &= \sum_{\substack{j=2\\ j\ne i}}^{n}\sum_{\substack{k=1\\ k\ne i}}^{j-1}(r_{i,j,k}-r_{i,j})+\sum_{\substack{j=2\\ j\ne i}}^{n}\sum_{\substack{k=1\\ k\ne i}}^{j-1} r_{i,j}\\ &:= \sum_{\substack{j=2\\ j\ne i}}^{n} R_{i,j}+\sum_{\substack{j=2\\ j\ne i}}^{n}\big\{j-1-1(j>i)\big\}r_{i,j}.\end{aligned} \tag{C.18}$$

Conditional on (X_i,X_j), $R_{i,j}$ is a sum of independent random variables with zero means. Define $t_{i,j}=v_{i,j}+4b_{i,j}$, where

$$t_{i,j}^2=\sum_{\substack{k=1\\ k\ne i}}^{j-1}(r_{i,j,k}-r_{i,j})^2=\sum_{\substack{k=1\\ k\ne i}}^{j-1}(h_{3,ijk}-h_{2,ij}-h_{1k})^2,$$

$$b_{i,j}^2=\sum_{\substack{k=1\\ k\ne i}}^{j-1}\mathbb{E}\big\{(r_{i,j,k}-r_{i,j})^2|X_i,X_j\big\}=\sum_{\substack{k=1\\ k\ne i}}^{j-1}\big[\mathbb{E}\big\{(h_{3,ijk}-h_{1k})^2|X_i,X_j\big\}-h_{2,ij}^2\big].$$

Applying (C.3) conditional on (X_i, X_j) gives

$$\mathbb{E}\{R_{i,j}^2 I(R_{i,j} \geq \sqrt{y} t_{i,j}) | X_i, X_j\} \leq 23 b_{i,j}^2 e^{-y/4}.$$

Then it follows from Lemma C.2 that

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{i=1}^{n}\left(\sum_{j=2, j\neq i}^{n} R_{i,j}\right)^2 \geq yn\left(\sum_{i=1}^{n}\sum_{j=2, j\neq i}^{n} t_{i,j}^2 + a_3 n^3\right)\right\} \\
&\quad \leq \mathbb{P}\left\{\sum_{i=1}^{n}\sum_{j=2, j\neq i}^{n} R_{i,j}^2 \geq y\left(\sum_{i=1}^{n}\sum_{j=2, j\neq i}^{n} t_{i,j}^2 + a_3 n^3\right)\right\} \\
&\quad \leq C\frac{e^{-y/4}}{a_3 n^3}\sum_{i=1}^{n}\sum_{j=2, j\neq i}^{n}(j-1)\mathbb{E}\big(r_{1,2,3}^2\big) \leq C a_3^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2,3}^2\big).
\end{aligned}$$

This, combined with the inequality $\sum_{i=1}^{n}\sum_{j=2, j\neq i}^{n} t_{i,j}^2 \leq a_3 n^3 (B_1 + B_2 V_n^2)$ implies

$$\mathbb{P}\left\{\sum_{i=1}^{n}\left(\sum_{j=2, j\neq i}^{n} R_{i,j}\right)^2 \geq a_3 y n^4\big(B_1 + 1 + B_2 V_n^2\big)\right\} \leq C a_3^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2,3}^2\big). \tag{C.19}$$

For the second addend in (C.18), consider $\widetilde{r}_{i,j} = \{j - 1 - I(j > i)\} r_{i,j}$ as a new (degenerate) kernel satisfying $\mathbb{E}(\widetilde{r}_{i,j} | X_i) = \mathbb{E}(\widetilde{r}_{i,j} | X_j) = 0$. Then by similar arguments as in step 1, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{n}\left[\sum_{j=2, j\neq i}^{n}\big\{j - 1 - 1(j > i)\big\} r_{i,j}\right]^2 \geq a_3 y n^4\big(B_3 + B_4 V_n^2\big)\right) \\
&\quad \leq C a_3^{-1} e^{-y/4}\mathbb{E}\big(r_{1,2,3}^2\big).
\end{aligned} \tag{C.20}$$

Together, (C.18), (C.19) and (C.20) prove (6.28).

To prove (6.29) for $m = 3$, consider the following decomposition:

$$\begin{aligned}
&\sum_{1\leq i_1 < i_2 < i_3 \leq n} r(X_{i_1}, X_{i_2}, X_{i_3}) \\
&\quad = \sum_{1\leq i_1 < i_2 < i_3 \leq n} r_{i_1,i_2,i_3} \\
&\quad = \sum_{k=3}^{n}\sum_{1\leq i_1 < i_2 < k}(r_{i_1,i_2,k} - r_{i_1,i_2}) + \sum_{k=3}^{n}\sum_{1\leq i_1 < i_2 < k} r_{i_1,i_2} \\
&\quad = \sum_{k=3}^{n}\sum_{1\leq i_1 < i_2 < k}(r_{i_1,i_2,k} - r_{i_1,i_2}) + \sum_{j=2}^{n-1}\sum_{i=1}^{j-1}(n - j) r_{i,j}
\end{aligned} \tag{C.21}$$

$$= \sum_{k=3}^{n}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1}(r_{i,j,k} - r_{i,j} - r_{j,k}) + \sum_{k=3}^{n}\sum_{j=2}^{k-1}(j-1)r_{j,k} + \sum_{j=2}^{n-1}\sum_{i=1}^{j-1}(n-j)r_{i,j}$$

$$:= \sum_{k=3}^{n}\sum_{j=2}^{k-1} r^*_{1,jk} + \sum_{k=3}^{n}\sum_{j=2}^{k-1} r^*_{2,jk} + \sum_{j=2}^{n-1} r^*_j,$$

where

$$r^*_{1,jk} = \sum_{i=1}^{j-1}(r_{i,j,k} - r_{i,j} - r_{j,k}), \qquad r^*_{2,jk} = (j-1)r_{j,k} \quad \text{and} \quad r^*_j = \sum_{i=1}^{j-1}(n-j)r_{i,j}.$$

Put $R^*_k = R^*_{1,k} + R^*_{2,k}$, $R^*_{1,k} = \sum_{j=2}^{k-1} r^*_{1,jk}$ and $R^*_{2,k} = \sum_{j=2}^{k-1} r^*_{2,jk}$. We see that $\{R^*_k, \mathcal{F}_k, k \ge 3\}$ is a sequence of martingale differences, and by (C.1),

$$\mathbb{P}\left(\left|\sum_{k=3}^{n} R^*_k\right| \ge \sqrt{2y}\left[\sum_{k=3}^{n}\left\{R^*_k + 2\mathbb{E}\left(R^{*2}_k|\mathcal{F}_{k-1}\right) + 3\mathbb{E}R^{*2}_k\right\}\right]^{1/2}\right) \le \sqrt{2}e^{-y/4}. \tag{C.22}$$

Note that conditional on (X_j, X_k), $r^*_{1,jk}$ is a sum of independent random variables with zero means, and given X_k, $r^*_{2,jk}$ are independent with zero means. Then it is straightforward to verify that

$$\sum_{k=3}^{n} \mathbb{E}R^{*2}_k \le 2\sum_{k=3}^{n}(k-2)\sum_{j=2}^{k-1}\mathbb{E}r^{*2}_{1,jk} + 2\sum_{k=3}^{n} R^{*2}_{2,k} \le Ca_3n^4. \tag{C.23}$$

Moreover, by noting the resemblance in structure between R^*_k and ψ_i (see (C.18)), it can be shown that

$$\mathbb{P}\left\{\sum_{k=3}^{n} R^{*2}_k \ge a_3yn^4\left(B_5 + B_6V_n^2\right)\right\} \le Ce^{-y/4}, \tag{C.24}$$

which is analogous to (6.28).

It remains to bound the tail probability of $\sum_{k=3}^{n}\mathbb{E}(R^{*2}_k|\mathcal{F}_{k-1})$. In view of (C.21), let $t^*_{j,k} = v^*_{j,k} + 4b^*_{j,k}$ for $2 \le j < k \le n$, where

$$v^{*2}_{j,k} = \sum_{i=1}^{j-1}(r_{i,j,k} - r_{i,j} - r_{j,k})^2, \qquad b^{*2}_{j,k} = \sum_{i=1}^{j-1}\mathbb{E}\left\{(r_{i,j,k} - r_{i,j} - r_{j,k})^2|X_j, X_k\right\},$$

and for $3 \le k \le n$, put

$$t^*_k = v^*_k + 4b^*_k, \qquad v^{*2}_k = \sum_{j=2}^{k-1} r^{*2}_{2,jk}, \qquad b^*_k = \sum_{j=2}^{k-1}\mathbb{E}\left(r^{*2}_{2,jk}|X_k\right).$$

Recall that $R_k^* = R_{1,k}^* + R_{2,k}^* = \sum_{j=2}^{k-1}(r_{1,jk}^* + r_{2,jk}^*)$. We proceed in a similar manner as in (C.14):

$$\begin{aligned}
&\sum_{k=3}^{n}\mathbb{E}\big(R_k^{*2}|\mathcal{F}_{k-1}\big)\\
&\quad\le 2\sum_{k=3}^{n}(k-2)\sum_{j=2}^{k-1}\mathbb{E}\big(r_{1,jk}^{*2}|\mathcal{F}_{k-1}\big)+2\sum_{k=3}^{n}\mathbb{E}\big(R_{2,k}^{*2}|\mathcal{F}_{k-1}\big)\\
&\quad=2\sum_{k=3}^{n}\sum_{j=2}^{k-1}(k-2)\mathbb{E}\big[r_{1,jk}^{*2}\big\{I\big(\big|r_{1,jk}^*\big|\le\sqrt{y}t_{j,k}^*\big)+I\big(\big|r_{1,jk}^*\big|>\sqrt{y}t_{j,k}^*\big)\big\}|\mathcal{F}_{k-1}\big]\\
&\qquad+2\sum_{k=3}^{n}\mathbb{E}\big[R_{2,k}^{*2}\big\{I\big(\big|R_{2,k}^*\big|\le\sqrt{y}t_k^*\big)+I\big(\big|R_{2,k}^*\big|>\sqrt{y}t_k^*\big)\big\}|\mathcal{F}_{k-1}\big].
\end{aligned}$$

By (C.3) and the Markov inequality, we have (recall that $y \ge 2$)

$$\begin{aligned}
&\mathbb{P}\left[\sum_{k=3}^{n}(k-2)\sum_{j=2}^{k-1}\mathbb{E}\big\{r_{1,jk}^{*2}I\big(\big|r_{1,jk}^*\big|>\sqrt{y}t_{j,k}^*\big)|\mathcal{F}_{k-1}\big\}\ge a_3yn^4\right]\\
&\quad\le\big(a_3yn^4\big)^{-1}\sum_{k=3}^{n}(k-2)\sum_{j=2}^{k-1}\mathbb{E}\big\{r_{1,jk}^{*2}I\big(\big|r_{1,jk}^*\big|>\sqrt{y}t_{j,k}^*\big)|\mathcal{F}_{k-1}\big\}\le Ce^{-y/4}
\end{aligned}\tag{C.25}$$

and

$$\begin{aligned}
&\mathbb{P}\left[\sum_{k=3}^{n}\mathbb{E}\big\{R_{2,k}^{*2}I\big(\big|R_{2,k}^*\big|>\sqrt{y}t_k^*\big)|\mathcal{F}_{k-1}\big\}\ge a_3yn^4\right]\\
&\quad\le\big(a_3yn^4\big)^{-1}\sum_{k=3}^{n}\mathbb{E}\big\{R_{2,k}^{*2}I\big(\big|R_{2,k}^*\big|>\sqrt{y}t_k^*\big)|\mathcal{F}_{k-1}\big\}\le Ce^{-y/4}.
\end{aligned}\tag{C.26}$$

However, it follows from (C.16) and (C.17) that

$$\sum_{k=3}^{n}(k-2)\sum_{j=2}^{k-1}\mathbb{E}\big\{r_{1,jk}^{*2}I\big(\big|r_{1,jk}^*\big|\le\sqrt{y}t_{j,k}^*\big)|\mathcal{F}_{k-1}\big\}\le a_3yn^4\big(B_7+B_8V_n^2\big),\tag{C.27}$$

$$\sum_{k=3}^{n}\mathbb{E}\big\{R_{2,k}^{*2}I\big(\big|R_{2,k}^*\big|\le\sqrt{y}t_k^*\big)|\mathcal{F}_{k-1}\big\}\le a_3yn^4\big(B_9+B_{10}V_n^2\big).\tag{C.28}$$

Assembling (C.22)–(C.28), we obtain

$$\mathbb{P}\left\{\left|\sum_{k=3}^{n} R_k^*\right| \geq \sqrt{a_3}\,yn^2\left(B_{11}+B_{12}V_n^2\right)^{1/2}\right\} \leq Ce^{-y/4}.$$

By induction, a similar result holds for $\sum_{j=2}^{n-1} r_j^*$; that is,

$$\mathbb{P}\left\{\left|\sum_{j=2}^{n} r_j^*\right| \geq \sqrt{a_3}\,yn^2\left(B_{13}+B_{14}V_n^2\right)^{1/2}\right\} \leq Ce^{-y/4}.$$

This completes the proof of (6.29) for $m=3$.

Step 3. For a general $3<m<n/2$,

$$r_k^2(x_1,\ldots,x_k) \leq 2a_m\left\{m-k+1+\sum_{j=1}^{k} h_1^2(x_j)\right\}, \tag{C.29}$$

where $r_k(x_1,\ldots,x_k)=E\{r(X_1,\ldots,X_m)|X_1=x_1,\ldots,X_k=x_k\}$ for $k=2,\ldots,m$.

To use the induction, we need the following string of equalities:

$$\begin{aligned}
\psi_i &= \sum_{\substack{1\leq \ell_1<\cdots<\ell_{m-1}\leq n\\ \ell_1,\ldots,\ell_{m-1}\neq i}} r_{\ell_1,\ldots,\ell_{m-1},i}\\
&= \sum_{\substack{\ell_{m-1}=m-1\\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{1\leq \ell_1<\cdots<\ell_{m-2}<\ell_{m-1}\\ \ell_1,\ldots,\ell_{m-2}\neq i}} \left(r_{\ell_1,\ldots,\ell_{m-2},\ell_{m-1},i}-r_{\ell_2,\ldots,\ell_{m-1},i}\right)\\
&\quad+\sum_{\substack{2\leq \ell_2<\cdots<\ell_{m-1}\leq n\\ \ell_2,\ldots,\ell_{m-1}\neq i}} \left\{\ell_2-1-1(i<\ell_2)\right\} r_{\ell_2,\ldots,\ell_{m-1},i}\\
&:= \psi_{1,i}+\psi_{2,i}.
\end{aligned} \tag{C.30}$$

Moreover,

$$\begin{aligned}
\psi_{1,i} &= \sum_{\substack{\ell_{m-1}=m-1\\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{1\leq \ell_1<\cdots<\ell_{m-2}<\ell_{m-1}\\ \ell_1,\ldots,\ell_{m-2}\neq i}} \left(r_{\ell_1,\ldots,\ell_{m-2},\ell_{m-1},i}-r_{\ell_2,\ldots,\ell_{m-1},i}\right)\\
&= \sum_{\substack{\ell_{m-1}=m-1\\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{1\leq \ell_1<\cdots<\ell_{m-2}<\ell_{m-1}\\ \ell_1,\ldots,\ell_{m-2}\neq i}} \check{r}_{\ell_1,\ldots,\ell_{m-1},i}
\end{aligned}$$

$$= \sum_{\substack{\ell_{m-1}=m-1 \\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{\ell_{m-2}=m-2 \\ \ell_{m-2}\neq i}}^{\ell_{m-1}-1} \cdots \sum_{\substack{\ell_2=2 \\ \ell_2\neq i}}^{\ell_3-1} \Bigg(\sum_{\substack{\ell_1=1 \\ \ell_1\neq i}}^{\ell_2-1} \breve{r}_{\ell_1,\dots,\ell_{m-1},i} \Bigg)$$

$$= \sum_{\substack{\ell_{m-1}=m-1 \\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{\ell_{m-2}=m-2 \\ \ell_{m-2}\neq i}}^{\ell_{m-1}-1} \cdots \sum_{\substack{\ell_2=2 \\ \ell_2\neq i}}^{\ell_3-1} \breve{R}_{\ell_2,\dots,\ell_{m-1},i}$$

with

$$\breve{r}_{\ell_1,\dots,\ell_{m-1}} = r_{\ell_1,\dots,\ell_{m-2},\ell_{m-1},i} - r_{\ell_2,\dots,\ell_{m-1},i}, \qquad \breve{R}_{\ell_2,\dots,\ell_{m-1},i} = \sum_{\substack{\ell_1=1 \\ \ell_1\neq i}}^{\ell_2-1} \breve{r}_{\ell_1,\dots,\ell_{m-1},i}.$$

Conditional on $(X_i, X_{\ell_2}, \dots, X_{\ell_{m-1}})$, $\breve{R}_{\ell_2,\dots,\ell_{m-1},i}$ is a sum of independent random variables with zero means. Also, it is straightforward to verify that

$$\psi_{1,i}^2 \leq \binom{n-1}{m-2} \sum_{\substack{\ell_{m-1}=m-1 \\ \ell_{m-1}\neq i}}^{n} \sum_{\substack{\ell_{m-2}=m-2 \\ \ell_{m-2}\neq i}}^{\ell_{m-1}-1} \cdots \sum_{\substack{\ell_2=2 \\ \ell_2\neq i}}^{\ell_3-1} \breve{R}_{\ell_2,\dots,\ell_{m-1},i}^2.$$

Next, let $\breve{t}_\ell = \breve{v}_\ell + 4\breve{b}_\ell$, where

$$\breve{v}_\ell = \sum_{\ell_1=1,\ell_1\neq i}^{\ell-1} \breve{r}_{\ell_1,\dots,\ell_{m-1},i}^2, \qquad \breve{b}_\ell^2 = \sum_{\ell_1=1,\ell_1\neq i}^{\ell-1} \mathbb{E}\big(\breve{r}_{\ell_1,\dots,\ell_{m-1},i}^2 | X_i, X_\ell, X_{\ell_3}, \dots, X_{\ell_{m-1}}\big).$$

Similar to the proof of (C.19), we derive from Lemma C.1 that for every $y \geq 2$,

$$\binom{n-1}{m-2}^{-1} \sum_{i=1}^{n} \psi_{1,i}^2 \leq y \Bigg\{ a_m \binom{n-1}{m-1} + \sum_{i=1}^{n} \sum_{\substack{\ell_{m-1}=m-1 \\ \ell_{m-1}\neq i}}^{n} \cdots \sum_{\substack{\ell_2=2 \\ \ell_2\neq i}}^{\ell_3-1} \breve{t}_{\ell_2}^2 \Bigg\}$$

holds with probability at least $1 - C\exp(-y/4)$. This, together with the following inequality

$$\sum_{i=1}^{n} \sum_{\substack{\ell_{m-1}=m-1 \\ \ell_{m-1}\neq i}}^{n} \cdots \sum_{\substack{\ell_2=2 \\ \ell_2\neq i}}^{\ell_3-1} \breve{t}_{\ell_2}^2 \leq a_m \binom{n}{m} \big(B_{15} + B_{16} V_n^2\big)$$

which can be obtained by using (C.29) repeatedly, gives

$$\mathbb{P}\Bigg\{ \sum_{i=1}^{n} \psi_{1,i}^2 \geq a_m y n^{2m-2} \big(B_{17} + B_{18} V_n^2\big) \Bigg\} \leq C e^{-y/4}. \tag{C.31}$$

For $\psi_{2,i}$, note that the summation is carried out over all $(m-2)$-tuples and

$$\big|\{\ell_2 - 1 - 1(i < \ell_2)\}r_{\ell_2,\dots,\ell_{m-1},i}\big| \le n|r_{\ell_2,\dots,\ell_{m-1},i}|.$$

Regarding $\{\ell_2 - 1 - 1(i < \ell_2)\}r_{\ell_2,\dots,\ell_{m-1},i}$ as a (weighted) degenerate kernel with $(m-1)$ arguments, it follows from induction that

$$\mathbb{P}\left\{\sum_{i=1}^n \psi_{2,i}^2 \ge a_m y n^{2m-2}\big(B_{19} + B_{20}V_n^2\big)\right\} \le Ce^{-y/4}. \tag{C.32}$$

Assembling (C.30), (C.31) and (C.32) yields (6.28).

Similarly, using the decomposition

$$\begin{aligned}
&\sum_{1\le i_1<\cdots<i_m\le n} r(X_{i_1},\dots,X_{i_m}) \\
&\quad = \sum_{1\le i_1<\cdots<i_m\le n} r_{i_1,\dots,i_m} \\
&\quad = \sum_{k=m}^n \sum_{1\le i_1<\cdots<i_{m-1}<k} (r_{i_1,\dots,i_{m-1},k} - r_{i_1,\dots,i_{m-1}}) + \sum_{1\le i_1<\cdots<i_{m-1}\le n-1} (n - i_{m-1}) r_{i_1,\dots,i_{m-1}}.
\end{aligned}$$

Because $\mathbb{E}(r_{i_1,\dots,i_{m-1},k}|\mathcal{F}_{k-1}) = r_{i_1,\dots,i_{m-1}}$,

$$\left\{R_k^* := \sum_{1\le i_1<\cdots<i_{m-1}\le k} (r_{i_1,\dots,i_{m-1},k} - r_{i_1,\dots,i_{m-1}}), \mathcal{F}_k\right\}_{k\ge m}$$

is a martingale difference sequence, such that the following analogue of (C.22) holds:

$$\mathbb{P}\left(\left|\sum_{k=m}^n R_k^*\right| \ge \sqrt{2y}\left[\sum_{k=m}^n \big\{R_k^{*2} + 2\mathbb{E}\big(R_k^{*2}|\mathcal{F}_{k-1}\big) + 3\mathbb{E}R_k^{*2}\big\}\right]^{1/2}\right) \le \sqrt{2}e^{-y/4}.$$

For $m \le k \le n$ fixed, extending (C.21) gives

$$\begin{aligned}
R_k^* &= \sum_{1\le i_1<\cdots<i_{m-1}<k} (r_{i_1,\dots,i_{m-1},k} - r_{i_1,\dots,i_{m-1}}) \\
&= \sum_{i_{m-1}=m-1}^{k-1} \cdots \sum_{i_1=1}^{i_2-1} (r_{i_1,i_2,\dots,i_{m-1},k} - r_{i_1,\dots,i_{m-1}} - r_{i_2,\dots,i_{m-1},k} + r_{i_2,\dots,i_{m-1}}) \\
&\quad + \sum_{i_{m-1}=m-1}^{k-1} \cdots \sum_{i_2=2}^{i_3-1} w_2 (r_{i_2,\dots,i_{m-1},k} - r_{i_2,\dots,i_{m-1}} - r_{i_3,\dots,i_{m-1},k} + r_{i_3,\dots,i_{m-1}}) \\
&\quad + \cdots + \sum_{i_{m-1}=m-1}^{k-1} w_{m-1} r_{i_{m-1},k},
\end{aligned}$$

where $w_j := \binom{i_j-1}{j-2}$ for $2 \le j \le m-1$, and set $w_1 \equiv 1$ for convention. Moreover, for $1 \le j \le m-2$, put

$$r^*_{j,i_{j+1},\dots,i_{m-1},k} = \sum_{i_j=j}^{i_{j+1}-1} w_j(r_{i_j,\dots,i_{m-1},k} - r_{i_j,\dots,i_{m-1}} - r_{i_{j+1},\dots,i_{m-1},k} + r_{i_{j+1},\dots,i_{m-1}})$$

and $r^*_{m-1,k} = \sum_{i_{m-1}=m-1}^{k-1} w_{m-1} r_{i_{m-1},k}$, such that

$$\begin{aligned} R^*_k = & \sum_{2\le i_2<\cdots<i_{m-1}\le k-1} r^*_{1,i_2,\dots,i_{m-1},k} \\ & + \sum_{3\le i_3<\cdots<i_{m-1}\le k-1} r^*_{2,i_3,\dots,i_{m-1},k} + \cdots + r^*_{m-1,k}. \end{aligned} \tag{C.33}$$

For $j = 1,\dots,m-2$, conditional on $(X_{i_{j+1}},\dots,X_{i_{m-1}},X_k)$, $r^*_{j,i_{j+1},\dots,i_{m-1},k}$ is a sum of independent random variables with zero means, and so is $r^*_{m-1,k}$ conditional on X_k.

In particular, we have

$$\begin{aligned} \sum_{k=m}^n \mathbb{E}R^{*2}_k \le{}& (m-1)\sum_{j=m}^n \biggl\{\mathbb{E}\biggl(\sum_{2\le i_2<\cdots<i_{m-1}\le k-1} r^*_{1,i_2,\dots,i_{m-1},k}\biggr)^2 \\ & + \mathbb{E}\biggl(\sum_{3\le i_3<\cdots<i_{m-1}\le k-1} r^*_{2,i_3,\dots,i_{m-1},k}\biggr)^2 + \cdots + \mathbb{E}r^{*2}_{m-1,k}\biggr\} \\ \le{}& (m-1)\sum_{k=m}^n \biggl\{\binom{k-2}{m-2}\sum_{2\le i_2<\cdots<i_{m-1}\le k-1} \mathbb{E}r^{*2}_{1,i_2,\dots,i_{m-1},k} \\ & + \binom{k-3}{m-3}\sum_{3\le i_3<\cdots<i_{m-1}\le k-1} \mathbb{E}r^{*2}_{2,i_3,\dots,i_{m-1},k} + \cdots + \mathbb{E}r^{*2}_{m-1,k}\biggr\} \\ \le{}& C(m-1)\mathbb{E}\{r^2(X_1,\dots,X_m)\}\sum_{k=m}^n \biggl\{\binom{k-2}{m-2}\binom{k-1}{m-1} \\ & + \binom{k-3}{m-3}\sum_{2\le i_2<\cdots<i_{m-1}\le k-1}(i_2-1)^2 + \cdots + \sum_{i=m-1}^{k-1}\binom{i-1}{m-2}^2\biggr\} \\ \le{}& Ca_m n^{2m-2}, \end{aligned}$$

which extends inequality (C.23). In view of (C.33), inequalities (C.24)–(C.28) can be similarly extended by using Lemmas C.1 and C.2 in the same way as in step 2. The proof of Lemma 6.1 is then complete.

Acknowledgements

The authors sincerely thank the Editor, Associate Editor and anonymous referees for their constructive comments that led to substantial improvement of the paper.

Qi-Man Shao was supported by Hong Kong Research Grants Council GRF 603710 and 403513. Wen-Xin Zhou was supported by NIH R01GM100474-4 and a grant from the Australian Research Council.

References

[1] Alberink, I.B. and Bentkus, V. (2001). Berry–Esseen bounds for von Mises and U-statistics. *Lith. Math. J.* **41** 1–16. MR1849804
[2] Alberink, I.B. and Bentkus, V. (2002). Lyapunov type bounds for U-statistics. *Theory Probab. Appl.* **46** 571–588. MR1971830
[3] Arvesen, J.N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100. MR0264805
[4] Bentkus, V. and Götze, F. (1996). The Berry–Esseen bound for student's statistic. *Ann. Probab.* **24** 491–503. MR1387647
[5] Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Ann. Statist.* **2** 1–20. MR0350952
[6] Borovskikh, Y.V. and Weber, N.C. (2003). Large deviations of U-statistics. I. *Lith. Math. J.* **43** 11–33. MR1996751
[7] Borovskikh, Y.V. and Weber, N.C. (2003). Large deviations of U-statistics. II. *Lith. Math. J.* **43** 241–261. MR2019542
[8] Callaert, H. and Janssen, P. (1978). The Berry–Esseen theorem for U-statistics. *Ann. Statist.* **6** 417–421. MR0464359
[9] Chan, Y.-K. and Wierman, J. (1977). On the Berry–Esseen theorem for U-statistics. *Ann. Probab.* **5** 136–139. MR0433551
[10] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2010). *Normal Approximation by Stein's Method*. Berlin: Springer.
[11] Chen, L.H.Y. and Shao, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Related Fields* **120** 236–254. MR1841329
[12] Chen, L.H.Y. and Shao, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. MR2331265
[13] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2003). Donsker's theorem for self-normalized partial sums processes. *Ann. Probab.* **31** 1228–1240. MR1988470
[14] de la Peña, V.H., Lai, T.L. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications. Probability and Its Applications (New York)*. Berlin: Springer. MR2488094
[15] Eichelsbacher, P. and Löwe, M. (1995). A large deviation principle for m-variate von Mises-statistics and U-statistics. *J. Theoret. Probab.* **8** 807–824. MR1353555
[16] Filippova, A.A. (1962). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7** 24–57.
[17] Friedrich, K.O. (1989). A Berry–Esseen bound for functions of independent random variables. *Ann. Statist.* **17** 170–183. MR0981443
[18] Giné, E., Götze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. MR1457629
[19] Grams, W.F. and Serfling, R.J. (1973). Convergence rates for U-statistics and related statistics. *Ann. Statist.* **1** 153–160. MR0336788

[20] Griffin, P. and Kuelbs, J. (1991). Some extensions of the LIL via self-normalizations. *Ann*. *Probab*. **19** 380–395. MR1085343
[21] Griffin, P.S. and Kuelbs, J.D. (1989). Self-normalized laws of the iterated logarithm. *Ann*. *Probab*. **17** 1571–1601. MR1048947
[22] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann*. *Math*. *Statist*. **19** 293–325. MR0026294
[23] Jing, B.-Y., Shao, Q.-M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann*. *Probab*. **31** 2167–2215. MR2016616
[24] Keener, R.W., Robinson, J. and Weber, N.C. (1998). Tail probability approximations for U-statistics. *Statist*. *Probab*. *Lett*. **37** 59–65. MR1622662
[25] Koroljuk, V.S. and Borovskich, Yu.V. (1994). *Theory of U-Statistics*. *Mathematics and Its Applications* **273**. Dordrecht: Kluwer Academic. MR1472486
[26] Lai, T.L., Shao, Q.-M. and Wang, Q. (2011). Cramér type moderate deviations for Studentized U-statistics. *ESAIM Probab*. *Stat*. **15** 168–179. MR2870510
[27] Logan, B.F., Mallows, C.L., Rice, S.O. and Shepp, L.A. (1973). Limit distributions of self-normalized sums. *Ann*. *Probab*. **1** 788–809. MR0362449
[28] Petrov, V.V. (1965). On the probabilities of large deviations for sums of independent random variables. *Theory Probab*. *Appl*. **10** 287–298. MR0185645
[29] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley. MR0595165
[30] Shao, Q.-M. (1997). Self-normalized large deviations. *Ann*. *Probab*. **25** 285–328. MR1428510
[31] Shao, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J*. *Theoret*. *Probab*. **12** 385–398. MR1684750
[32] Shao, Q.-M., Zhang, K. and Zhou, W.-X. (2016). Stein's method for nonlinear statistics: A brief survey and recent progress. *J*. *Statist*. *Plann*. *Inference* **168** 68–89. MR3412222
[33] Shao, Q.-M. and Zhou, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann*. *Probab*. **42** 623–648. MR3178469
[34] Stein, C. (1986). *Approximation Computation of Expectations*. Hayward, CA: IMS.
[35] van Zwet, W.R. (1984). A Berry–Esseen bound for symmetric statistics. *Z*. *Wahrsch*. *Verw*. *Gebiete* **66** 425–440. MR0751580
[36] Vandemaele, M. and Veraverbeke, N. (1985). Cramér type large deviations for Studentized U-statistics. *Metrika* **32** 165–179. MR0824452
[37] Wang, Q. (1998). Bernstein type inequalities for degenerate U-statistics with applications. *Chin*. *Ann*. *Math*. *Ser*. *B* **19** 157–166. MR1655931
[38] Wang, Q., Jing, B.-Y. and Zhao, L. (2000). The Berry–Esseen bound for Studentized statistics. *Ann*. *Probab*. **28** 511–535. MR1756015
[39] Wang, Q. and Weber, N.C. (2006). Exact convergence rate and leading term in the central limit theorem for U-statistics. *Statist*. *Sinica* **16** 1409–1422. MR2327497

Received September 2013 and revised August 2014