Normal Approximation for Non-linear Statistics Using a Concentration Inequality Approach

Louis H.Y. Chen [1]
National University of Singapore
and
Qi-Man Shao[2]
Hong Kong University of Science and Technology, University of Oregon,and Zhejiang University

Abstract. Let T be a general sampling statistic that can be written as a linear statistic plus an error term. Uniform and non-uniform Berry-Esseen type bounds for T are obtained. The bounds are best possible for many known statistics. Applications to U-statistic, multi-sample U-statistic, L-statistic, random sums, and functions of non-linear statistics are discussed.

Nov. 15, 2005

AMS 2000 *subject classification:* Primary 62E20, 60F05; secondary 60G50.

Key words and phrases. Normal approximation, uniform Berry-Esseen bound, non-uniform Berry-Esseen bound, concentration inequality approach, nonlinear statistics, U-statistics, multi-sample U-statistics, L-statistic, random sums, functions of non-linear statistics.

[1]Research is partially supported by grant R-146-000-013-112 at the National University of Singapore.
[2]Research is partially supported by NSF grant DMS-0103487 and grant R-146-000-038-101 at the National University of Singapore.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independent random variables and let $T := T(X_1, \ldots, X_n)$ be a general sampling statistic. In many cases T can be written as a linear statistic plus an error term, say $T = W + \Delta$, where

$$W = \sum_{i=1}^{n} g_i(X_i), \ \Delta := \Delta(X_1, \ldots, X_n) = T - W$$

and $g_i := g_{n,i}$ are Borel measurable functions. Typical cases include U-statistics, multi-sample U-statistics, L-statistics, and random sums. Assume that

$$E(g_i(X_i)) = 0 \text{ for } i = 1, 2, \ldots, n, \text{ and } \sum_{i=1}^{n} E(g_i^2(X_i)) = 1. \tag{1.1}$$

It is clear that if $\Delta \to 0$ in probability as $n \to \infty$, then we have the following central limit theorem

$$\sup_z |P(T \le z) - \Phi(z)| \to 0 \tag{1.2}$$

where Φ denotes the standard normal distribution function, provided that the Lindeberg condition holds:

$$\forall \, \varepsilon > 0, \sum_{i=1}^{n} E g_i^2(X_i) I(|g_i(X_i)| > \varepsilon) \to 0.$$

If in addition, $E|\Delta|^p < \infty$ for some $p > 0$, then by the Chebyshev inequality, one can obtain the following rate of convergence:

$$\sup_z |P(T \le z) - \Phi(z)| \le \sup_z |P(W \le z) - \Phi(z)| + 2(E|\Delta|^p)^{1/(1+p)}. \tag{1.3}$$

The first term on the right hand side of (1.3) is well-understood via the Berry-Esseen inequality. For example, using Stein's method, Chen and Shao (2001) obtained

$$\sup_z |P(W \le z) - \Phi(z)| \le 4.1\Big(\sum_{i=1}^{n} E g_i^2(X_i) I(|g_i(X_i)| > 1) + \sum_{i=1}^{n} E|g_i(X_i)|^3 I(|g_i(X_i)| \le 1)\Big). \tag{1.4}$$

However, the bound $(E|\Delta|^p)^{1/(1+p)}$ is in general not sharp for many commonly used statistics. Many authors have worked towards obtaining better Berry-Esseen bounds. For example, sharp Berry-Esseen bounds have been obtained for general symmetric statistics in van Zwet (1984) and Friedrich (1989). An Edgeworth expansion with remainder $O(n^{-1})$ for symmetric statistics was proved by Bentkus, Götze and Zwet (1997).

The main purpose of this paper is to establish uniform and non-uniform Berry-Esseen bounds for general nonlinear statistics. The bounds are best possible for many known statistics. Our proof is based on a randomized concentration inequality approach to bounding $P(W + \Delta \le z) - P(W \le z)$. Since proofs of uniform and non-uniform bounds for sums of independent random variables can be proved via Stein's method [8], which is much neater and simpler than the traditional Fourier analysis approach, this paper provides a direct and unifying treatment towards the Berry-Esseen bounds for general non-linear statistics.

This paper is organized as follows. The main results are stated in next section, five applications are presented in Section 3 and an example is given in Section 4 to show the sharpness of the main results. Proofs of the main results are given in Section 5, while proofs of other results including Example 4.1 are postponed to Section 6.

Throughout this paper, C will denote an absolute constant whose value may change at each appearance. The L_p norm of a random variable X is denoted by $||X||_p$, i.e., $||X||_p = (E|X|^p)^{1/p}$ for $p \ge 1$.

2 Main results

Let $\{X_i, 1 \le i \le n\}$, T, W, Δ be defined as in Section 1. In the following theorems, we assume that (1.1) is satisfied. Put

$$\beta = \sum_{i=1}^n E|g_i(X_i)|^2 I(|g_i(X_i)| > 1) + \sum_{i=1}^n E|g_i(X_i)|^3 I(|g_i(X_i)| \le 1) \tag{2.1}$$

and let $\delta > 0$ satisfy

$$\sum_{i=1}^n E|g_i(X_i)| \min(\delta, |g_i(X_i)|) \ge 1/2. \tag{2.2}$$

THEOREM **2.1** *For each $1 \le i \le n$, let Δ_i be a random variable such that X_i and $(\Delta_i, W - g_i(X_i))$ are independent. Then*

$$\sup_z |P(T \le z) - P(W \le z)| \le 4\delta + E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta - \Delta_i)| \tag{2.3}$$

for δ satisfying (2.2). In particular, we have

$$\sup_z |P(T \le z) - P(W \le z)| \le 2\beta + E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta - \Delta_i)| \tag{2.4}$$

and

$$\sup_z |P(T \le z) - \Phi(z)| \le 6.1\beta + E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta - \Delta_i)|. \tag{2.5}$$

Next theorem provides a non-uniform bound.

THEOREM **2.2** *For each* $1 \le i \le n$*, let* Δ_i *be a random variable such that* X_i *and* $(\Delta_i, \{X_j, j \ne i\})$ *are independent. Then for* δ *satisfying* (2.2) *and for* $z \in R^1$*,*

$$|P(T \le z) - P(W \le z)| \le \gamma_z + e^{-|z|/3}\tau \tag{2.6}$$

where

$$\begin{aligned} \gamma_z &= P(|\Delta| > (|z|+1)/3) + \sum_{i=1}^n P(|g_i(X_i)| > (|z|+1)/3) \\ &\quad + \sum_{i=1}^n P(|W - g_i(X_i)| > (|z|-2)/3) P(|g_i(X_i)| > 1), \end{aligned} \tag{2.7}$$

$$\tau = 21\delta + 8.1||\Delta||_2 + 3.5 \sum_{i=1}^n ||g_i(X_i)||_2 ||\Delta - \Delta_i||_2. \tag{2.8}$$

In particular, if $E|g_i(X_i)|^p < \infty$ *for* $2 < p \le 3$*, then*

$$\begin{aligned} &|P(T \le z) - \Phi(z)| \\ &\le P(|\Delta| > (|z|+1)/3) + C(|z|+1)^{-p}\Big(||\Delta||_2 + \sum_{i=1}^n ||g_i(X_i)||_2 ||\Delta - \Delta_i||_2 + \sum_{i=1}^n E|g_i(X_i)|^p\Big). \end{aligned} \tag{2.9}$$

A result similar to (2.5) has been obtained by Friedrich (1989) for $g_i = E(T|X_i)$ using the method of characteristic function. Our proof is direct and simpler and the bounds are easier to calculate. The non-uniform bounds in (2.6) and (2.9) for general non-linear statistics are new.

REMARK **2.1** *Assume* $E|g_i(X_i)|^p < \infty$ *for* $p > 2$*. Let*

$$\delta = \Big(\frac{2(p-2)^{p-2}}{(p-1)^{p-1}} \sum_{i=1}^n E|g_i(X_i)|^p\Big)^{1/(p-2)}. \tag{2.10}$$

Then (2.2) *is satisfied. This follows from the inequality*

$$\min(a, b) \ge a - \frac{(p-2)^{p-2} a^{p-1}}{(p-1)^{p-1} b^{p-2}} \tag{2.11}$$

for $a \ge 0$ *and* $b > 0$*.*

REMARK **2.2** *If* $\beta \le 1/2$*, then* (2.2) *is satisfied with* $\delta = \beta/2$*.*

Remark **2.3** *Let $\delta > 0$ be such that*

$$\sum_{i=1}^n E g_i^2(X_i) I(|g_i(X_i)| > \delta) \le 1/2.$$

Then (2.2) *holds. In particular, if $X_1, X_2, \cdots, X_n$ are independent and identically distributed (i.i.d.) random variables and $g_i = g_1$, then* (2.2) *is satisfied with $\delta = c_0/\sqrt{n}$, where c_0 is a constant such that $E(\sqrt{n} g_1(X_1))^2 I(|\sqrt{n} g_1(X_1)| > c_0) \le 1/2$.*

Remark **2.4** *In Theorems 2.1 and 2.2, the choice of Δ_i is flexible. For example, one can choose $\Delta_i = \Delta(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ or $\Delta_i = \Delta(X_1, \ldots, X_{i-1}, \hat{X}_i, X_{i+1}, \ldots, X_n)$, where $\{\hat{X}_i, 1 \le i \le n\}$ is an independent copy of $\{X_i, 1 \le i \le n\}$. The choice of g_i is also flexible. It can be more general than $g_i(x) = E(T|X_i = x)$, which is commonly used by others in the literature.*

Remark **2.5** *Let $X_1, \ldots, X_n$ be independent normally distributed random variables with mean zero and variance $1/n$, and let W, T and Δ be as in Example 4.1. Then*

$$(2.12)\; E|W\Delta| + \sum_{i=1}^n E|X_i|^3 + \sum_{i=1}^n E|X_i(\Delta(X_1, \ldots, X_i, \ldots, X_n) - \Delta(X_1, \ldots, 0, \ldots, X_n))| \le C\varepsilon^{2/3}$$

for $(1/\varepsilon)^{4/3} \le n \le 16(1/\varepsilon)^{4/3}$. This together with (4.5) *shows that the bound in* (2.4) *is achievable. Moreover, the term $\sum E|g_i(X_i)(\Delta - \Delta_i)|$ in* (2.4) *can not be dropped off.*

3 Applications

Theorems 2.1 and 2.2 can be applied to a wide range of different statistics and provide bounds of the best possible order in many instances. To illustrate the usefulness and the generality of these results, we give five applications in this section. The uniform bounds refine many existing results with specifying absolute constants, while the non-uniform bounds are new for many cases.

3.1 U-statistics

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables, and let $h(x_1, \ldots, x_m)$ be a real-valued Borel measurable symmetric function of m variables, where m $(2 \le m < n)$ may depend on n. Consider the Hoeffding (1948) U-statistic

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}).$$

The U-statistic elegantly and usefully generalizes the notion of a sample mean. Numerous investigations on the limiting properties of the U-statistics have been done during the last few decades. A systematic presentation of the theory of U-statistics was given in Koroljuk and Borovskikh (1994). We refer the study on uniform Berry-Esseen bound for U-statistics to Filippova (1962), Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977), Callaert and Janseen (1978), Serfling (1980),Van Zwet (1984), and Friedrich (1989). One can also refer to Wang, Jing and Zhao (2000) on uniform Berry-Esseen bound for studentized U-statistics.

Applying Theorems 2.1 and 2.2 to the U-statistic, we have

THEOREM **3.1** *Assume that* $Eh(X_1,\ldots,X_m)=0$ *and* $\sigma^2=Eh^2(X_1,\ldots,X_m)<\infty$. *Let* $g(x)=E(h(X_1,X_2,\ldots,X_m)|X_1=x)$ *and* $\sigma_1^2=Eg^2(X_1)$. *Assume that* $\sigma_1>0$. *Then*

$$\sup_z|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n\le z\right)-P\Big(\frac{1}{\sqrt{n}\sigma_1}\sum_{i=1}^n g(X_i)\le z\Big)|\le\frac{(1+\sqrt{2})(m-1)\sigma}{(m(n-m+1))^{1/2}\sigma_1}+\frac{c_0}{\sqrt{n}},\tag{3.1}$$

where c_0 *is a constant such that* $Eg^2(X_1)I(|g(X_1)|>c_0\sigma_1)\le\sigma_1^2/2$. *If in addition* $E|g(X_1)|^p<\infty$ *for* $2<p\le3$, *then*

$$\sup_z|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n\le z\right)-\Phi(z)|\le\frac{(1+\sqrt{2})(m-1)\sigma}{(m(n-m+1))^{1/2}\sigma_1}+\frac{6.1E|g(X_1)|^p}{n^{(p-2)/2}\sigma_1^p}\tag{3.2}$$

and for $z\in R^1$,

$$\begin{aligned}|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n\le z\right)-\Phi(z)| &\le \frac{9m\sigma^2}{(1+|z|)^2(n-m+1)\sigma_1^2}+\frac{13.5e^{-|z|/3}m^{1/2}\sigma}{(n-m+1)^{1/2}\sigma_1}\\ &\quad+\frac{CE|g(X_1)|^p}{(1+|z|)^pn^{(p-2)/2}\sigma_1^p}.\end{aligned}\tag{3.3}$$

Moreover, if $E|h(X_1,\ldots,X_m)|^p<\infty$ *for* $2<p\le3$, *then for* $z\in R^1$,

$$\begin{aligned}&|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n\le z\right)-\Phi(z)|\\ &\le \frac{Cm^{1/2}E|h(X_1,\ldots,X_m)|^p}{(1+|z|)^p(n-m+1)^{1/2}\sigma_1^p}+\frac{CE|g(X_1)|^p}{(1+|z|)^pn^{(p-2)/2}\sigma_1^p}.\end{aligned}\tag{3.4}$$

Note that the error in (3.1) is of order $O(n^{-1/2})$ only under the assumption of finite second moment of h. The result appears not known before. The uniform bound given in (3.2) is not new, however, the specifying constant for general m is new. Finite second moment of h is not the weakest assumption for the uniform bound. Friedrich (1989) obtained an order of $O(n^{-1/2})$ when $E|h|^{5/3}<\infty$ which is necessary for the bound as shown by Bentkus, Götze and Zitikis (1994).

For the non-uniform bound, Zhao and Chen (1983) proved that if $m = 2$, $E|h(X_1, X_2)|^3 < \infty$, then

$$|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n \le z\right) - \Phi(z)| \le An^{-1/2}(1+|z|)^{-3} \tag{3.5}$$

for $z \in R^1$, where the constant A does not depend on n and z but the moment of h. Clearly, (3.4) refines Zhao and Chen's result specifying the relationship of the constant A with the moment condition. After we finished proving Theorem 3.1, Wang (2001) informed the second author that he also obtained (3.4) for $m = 2$ and $p = 3$.

Remark **3.1** (3.3) *implies that*

$$|P\left(\frac{\sqrt{n}}{m\sigma_1}U_n \le z\right) - \Phi(z)| \le \frac{Cm^{1/2}\sigma^2}{(1+|z|)^3(n-m+1)^{1/2}\sigma_1^2} + \frac{CE|g(X_1)|^p}{(1+|z|)^p n^{(p-2)/2}\sigma_1^p} \tag{3.6}$$

for $|z| \le ((n-m+1)/m)^{1/2}$. *For* $|z| > ((n-m+1)/m)^{1/2}$, *the bound like* (3.6) *can be easily obtained by using the Chebyshev inequality. On the other hand, if* (3.6) *holds for any* $z \in R^1$, *then it appears necessary to assume* $E|h(X_1, \ldots, X_m)^p < \infty$.

3.2 Multi-sample U-statistics

Consider k independent sequences $\{X_{j1}, \ldots, X_{jn_j}\}$ of i.i.d. random variables, $j = 1, \ldots, k$. Let $h(x_{jl}, l = 1, \ldots, m_j, j = 1, \ldots, k)$ be a measurable function symmetric with respect to m_j arguments of the j-th set, $m_j \ge 1$, $j = 1, \ldots, k$. Let

$$\theta = Eh(X_{jl}, l = 1, \ldots, m_j, j = 1, \ldots, k).$$

The multi-sample U-statistic is defined as

$$U_{\bar{n}} = \Big\{\prod_{j=1}^{k}\binom{n_j}{m_j}^{-1}\Big\}\sum h(X_{jl}, l = i_{j1}, \ldots, i_{jm_j}, j = 1, \ldots, k)$$

where $\bar{n} = (n_1, \ldots, n_k)$ and the summation is carried out over all $1 \le i_{j1} < \ldots < i_{jm_j} \le n_j$, $n_j \ge 2m_j$, $j = 1, \ldots, k$. Clearly, $U_{\bar{n}}$ is an unbiased estimate of θ. The two-sample Wilcoxon statistic and the two-sample ω^2-statistic are two typical examples of the multi-sample U-statistics. Without loss of generality, assume $\theta = 0$. For $j = 1, \ldots, k$, define

$$h_j(x) = E\Big(h(X_{11}, \ldots, X_{1m_1}; \ldots; X_{k1}, \ldots, X_{km_k})|X_{j1} = x)$$

and let $\sigma_j^2 = Eh_j^2(X_{j1})$ and

$$\sigma_{\bar{n}}^2 = \sum_{j=1}^{k} \frac{m_j^2}{n_j} \sigma_j^2.$$

A uniform Berry-Esseen bound with order $O((\min_{1\le j\le k} n_j)^{-1/2})$ for the multi-sample U-statistics was obtained by Helmers and Janssen (1982) and Borovskich (1983) (see, [Koroljuk and Borovskich (1994), pp. 304-311.]). Next theorem refines their results.

THEOREM **3.2** *Assume that* $\theta = 0$, $\sigma^2 := Eh^2(X_{11}, \ldots, X_{1m_1}; \ldots; X_{k1}, \ldots, X_{km_k}) < \infty$ *and* $\max_{1\le j\le k} \sigma_j > 0$. *Then for* $2 < p \le 3$

$$\sup_z |P\left(\sigma_{\bar{n}}^{-1} U_{\bar{n}} \le z\right) - \Phi(z)| \le \frac{(1+\sqrt{2})\sigma}{\sigma_{\bar{n}}} \sum_{j=1}^{k} \frac{m_j^2}{n_j} + \frac{6.6}{\sigma_{\bar{n}}^p} \sum_{j=1}^{k} \frac{m_j^p}{n_j^{p-1}} E|h_j(X_{j1})|^p \tag{3.7}$$

and for $z \in R^1$

$$\begin{aligned} |P\left(\sigma_{\bar{n}}^{-1} U_n \le z\right) - \Phi(z)| \quad \le \quad & \frac{9\sigma^2}{(1+|z|)^2 \sigma_{\bar{n}}^2} \Big(\sum_{j=1}^{k} \frac{m_j^2}{n_j}\Big)^2 + 13.5 e^{-|z|/3} \frac{\sigma}{\sigma_{\bar{n}}} \sum_{j=1}^{k} \frac{m_j^2}{n_j} \\ & + \frac{C}{(1+|z|)^p \sigma_{\bar{n}}^p} \sum_{j=1}^{k} \frac{m_j^p E|h_j(X_{j1})|^p}{n_j^{p-1}}. \end{aligned} \tag{3.8}$$

3.3 L-statistics

Let $X_1, \ldots, X_n$ be i.i.d. random variables with a common distribution function F, and let F_n be the empirical distribution function defined by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x) \quad \text{for} \quad x \in R^1.$$

Let $J(t)$ be a real-valued function on $[0, 1]$ and define

$$T(G) = \int_{-\infty}^{\infty} xJ(G(x))dG(x)$$

for non-decreasing measurable function G. Put

$$\sigma^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J(F(s))J(F(t))F(\min(s,t))(1 - F(\max(s,t))dsdt$$

and

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

The statistic $T(F_n)$ is called an L-statistic (see [Serfling (1980), Chapter 8]). Uniform Berry-Esseen bounds for L-statistic for smoothing J were given by Helmers (1977), and Helmers, Janssen and Serfling (1990). Applying Theorems 2.1 and 2.2 yields the following uniform and non-uniform bounds for L-statistic.

THEOREM **3.3** *Let $n \geq 4$. Assume that $EX_1^2 < \infty$ and $E|g(X_1)|^p < \infty$ for $2 < p \leq 3$. If the weight function $J(t)$ is Lipschitz of order 1 on $[0,1]$, that is, there exists a constant c_0 such that*

$$|J(t) - J(s)| \leq c_0 |t - s| \quad \text{for } 0 \leq s, t \leq 1 \tag{3.9}$$

then

$$\sup_z |P(\sqrt{n}\sigma^{-1}(T(F_n) - T(F)) \leq z) - \Phi(z)| \leq \frac{(1+\sqrt{2})c_0 ||X_1||_2}{\sqrt{n}\sigma} + \frac{6.1 E|g(X_1)|^p}{n^{(p-2)/2}\sigma^p} \tag{3.10}$$

and

$$\begin{aligned} &|P(\sqrt{n}\sigma^{-1}(T(F_n) - T(F)) \leq z) - \Phi(z)| \\ &\leq \frac{9c_0^2 EX_1^2}{(1+|z|)^2 n\sigma^2} + \frac{C}{(1+|z|)^p}\Big(\frac{c_0||X_1||_2}{\sqrt{n}\sigma} + \frac{E|g(X_1)|^p}{n^{(p-2)/2}\sigma^p}\Big) \end{aligned} \tag{3.11}$$

3.4 Random sums of independent random variables with non-random centering

Let $\{X_i, i \geq 1\}$ be i.i.d. random variables with $EX_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, and let $\{N_n, n \geq 1\}$ be a sequence of non-negative integer-valued random variables that are independent of $\{X_i, i \geq 1\}$. Assume that $EN_n^2 < \infty$ and

$$\frac{N_n - EN_n}{\sqrt{\mathrm{Var}(N_n)}} \xrightarrow{d.} N(0,1).$$

Then by Robbins (1948),

$$\frac{\sum_{i=1}^{N_n} X_i - (EN_n)\mu}{\sqrt{\sigma^2 EN_n + \mu^2 \mathrm{Var}(N_n)}} \xrightarrow{d.} N(0,1).$$

This is a special case of limit theorems for random sums with non-random centering. This kind of problems arises in the study, for example, of Galton-Watson branching processes. We refer to Finkelstein, Kruglov and Tucker (1994) and references therein for recent developments in this area.

As another application of our general result, we give a uniform Berry-Esseen bound for the random sum.

THEOREM **3.4** *Let* $\{Y_i, i \geq 1\}$ *be i.i.d. non-negative integer-valued random variables with* $EY_i = \nu$ *and* $Var(Y_i) = \tau^2$. *Put* $N_n = \sum_{i=1}^n Y_i$. *Assume that* $E|X_i|^3 < \infty$ *and that* $\{Y_i, i \geq 1\}$ *and* $\{X_i, i \geq 1\}$ *are independent. Then*

$$\sup_x |P\Big(\frac{\sum_{i=1}^{N_n} X_i - n\mu\nu}{\sqrt{n(\nu\sigma^2 + \tau^2\mu^2)}} \leq x\Big) - \Phi(x)| \leq Cn^{-1/2}\Big(\frac{\tau^2}{\nu^2} + \frac{E|X_1|^3}{\sigma^3} + \frac{\sigma}{\mu\sqrt{\nu}}\Big). \tag{3.12}$$

3.5 Functions of non-linear statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample and $\hat{\Theta}_n = \hat{\Theta}_n(X_1, \ldots, X_n)$ be a weak consistent estimator of an unknown parameter θ. Assume that $\hat{\Theta}_n$ can be written as

$$\hat{\Theta}_n = \theta + \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^n g_i(X_i) + \Delta\Big)$$

where g_i are Borel measurable functions with $Eg_i(X_i) = 0$ and $\sum_{i=1}^n Eg_i^2(X_i) = 1$, and $\Delta := \Delta_n(X_1, \ldots, X_n) \to 0$ in probability. Let h be a real-valued function differentiable in a neighborhood of θ with $g'(\theta) \neq 0$. Then, it is known that

$$\frac{\sqrt{n}(h(\hat{\Theta}_n) - h(\theta))}{h'(\theta)} \xrightarrow{d.} N(0,1)$$

under some regularity conditions. When $\hat{\Theta}_n$ is the sample mean, the Berry-Esseen bound and Edgeworth expansion have been well studied (see Bhattacharya and Ghosh (1978)). The next theorem shows that the results in Section 3 can be extended to functions of non-linear statistics.

THEOREM **3.5** *Assume that* $h'(\theta) \neq 0$ *and* $\delta(c_0) = \sup_{|x-\theta| \leq c_0} |h''(x)| < \infty$ *for some* $c_0 > 0$. *Then for* $2 < p \leq 3$,

$$\begin{aligned} \sup_z |P\Big(\frac{\sqrt{n}(h(\hat{\Theta}_n) - h(\theta))}{h'(\theta)} \leq z\Big) - \Phi(z)| & \\ \leq \Big(1 + \frac{c_0\delta(c_0)}{|h'(\theta)|}\Big)\Big(E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta - \Delta_i)|\Big) & \\ +6.1\sum_{i=1}^n E|g_i(X_i)|^p + \frac{4}{c_0^2 n} + \frac{2E|\Delta|}{c_0 n^{1/2}} + \frac{4.4c_0^{3-p}\delta(c_0)}{|h'(\theta)|n^{(p-2)/2}}, & \end{aligned} \tag{3.13}$$

where $W = \sum_{i=1}^n g_i(X_i)$.

4 An example

In this section we give an example to show that the bound of (2.4) in Theorem 2.1 is achievable. Moreover, the term $\sum E|g_i(X_i)(\Delta - \Delta_i)|$ in (2.4) can not be dropped off. The example also provides a counter-example to a result of Shorack (2000) and of Bolthausen and Götze (1993).

EXAMPLE **4.1** *Let $X_1, \ldots, X_n$ be independent normally distributed random variables with mean zero and variance $1/n$. Define*

$$W = \sum_{i=1}^{n} X_i, \;\; T := T_\varepsilon = W - \varepsilon |W|^{-1/2} + \varepsilon\, c_0 \quad \textit{and} \quad \Delta = T - W = -\varepsilon |W|^{-1/2} + \varepsilon\, c_0,$$

where $c_0 = E(|W|^{-1/2}) = \sqrt{2/\pi} \int_0^\infty x^{-1/2} e^{-x^2/2} dx$. Let $\{\hat{X}_i, 1 \le i \le n\}$ be an independent copy of $\{X_i, 1 \le i \le n\}$ and define

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} E|\Delta(X_1, \ldots, X_i, \ldots, X_n) - \Delta(X_1, \ldots, \hat{X}_i, \ldots, X_n)|. \tag{4.1}$$

Then $ET = 0$ and for $0 < \varepsilon < 1/64$ and $n \ge (1/\varepsilon)^4$

$$P(T \le \varepsilon\, c_0) - \Phi(\varepsilon\, c_0) \ge \varepsilon^{2/3}/6, \tag{4.2}$$

$$E|W\Delta| + E|\Delta| \le 7\varepsilon, \tag{4.3}$$

$$E|\Delta| + \sum_{i=1}^{n} E|X_i|^3 + \sqrt{\alpha} \le C\varepsilon \tag{4.4}$$

where C is an absolute constant.

Clearly, (4.2) implies that

$$\sup_z |P(T_\varepsilon \le z) - \Phi(z)| \ge \varepsilon^{2/3}/6. \tag{4.5}$$

A result of Shorack (2000) (see Lemma 11.1.3, p. 261, [22]) states that for any random variables W and Δ,

$$\sup_z |P(W + \Delta \le z) - \Phi(z)| \le \sup_z |P(W \le z) - \Phi(z)| + 4E|W\Delta| + 4E|\Delta|. \tag{4.6}$$

Another result which is in Theorem 2 of Bolthausen and Götze (1993) states that if $ET = 0$, then

$$\sup_z |P(T \le z) - \Phi(z)| \le C\Big(E|\Delta| + \sum_{i=1}^{n} E|g_i(X_i)|^3 + \sqrt{\alpha}\Big), \tag{4.7}$$

where C is an absolute constant and α is defined in (4.1).

In view of (4.3), (4.4) and (4.5), the result of Shrock and of Bolthausen and Götze can be shown to lead to a contradiction.

5 Proof of Main Theorems

In this section we prove Theorems 2.1 and 2.2 and Remarks 2.1 and 2.2.

Proof of Theorem 2.1. (2.5) follows from (2.4) and (1.4). When $\beta > 1/2$, (2.4) is trivial. For $\beta \le 1/2$, (2.4) is a consequence of (2.3) and Remark 2.2. Thus, we only need to prove (2.3). Note that

$$-P(z-|\Delta| \le W \le z) \le P(T \le z) - P(W \le z) \le P(z \le W \le z+|\Delta|). \tag{5.1}$$

It suffices to show that

$$P(z \le W \le z+|\Delta|) \le 4\delta + E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta-\Delta_i)| \tag{5.2}$$

and

$$P(z-|\Delta| \le W \le z) \le 4\delta + E|W\Delta| + \sum_{i=1}^n E|g_i(X_i)(\Delta-\Delta_i)| \tag{5.3}$$

where δ satisfies (2.2). Let

$$f_\Delta(w) = \begin{cases} -|\Delta|/2-\delta & \text{for } w \le z-\delta, \\ w-\frac{1}{2}(2z+|\Delta|) & \text{for } z-\delta \le w \le z+|\Delta|+\delta, \\ |\Delta|/2+\delta & \text{for } w > z+|\Delta|+\delta. \end{cases} \tag{5.4}$$

Let

$$\begin{aligned} \xi_i &= g_i(X_i),\ \hat{M}_i(t) = \xi_i\{I(-\xi_i \le t \le 0) - I(0 < t \le -\xi_i)\}, \\ M_i(t) &= E\hat{M}_i(t),\ \hat{M}(t) = \sum_{i=1}^n \hat{M}_i(t), M(t) = E\hat{M}(t). \end{aligned}$$

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

$$\begin{aligned} E\{Wf_\Delta(W)\} &= \sum_{1\le i\le n} E\{\xi_i(f_\Delta(W) - f_\Delta(W-\xi_i))\} \\ &\quad + \sum_{1\le i\le n} E\{\xi_i(f_\Delta(W-\xi_i) - f_{\Delta_i}(W-\xi_i))\} \\ &:= H_1 + H_2. \end{aligned} \tag{5.5}$$

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

$$\begin{aligned} H_1 &= \sum_{1\le i\le n} E\Big\{\xi_i \int_{-\xi_i}^0 f'_\Delta(W+t)dt\Big\} \\ &= \sum_{1\le i\le n} E\Big\{\int_{-\infty}^{\infty} f'_\Delta(W+t)\hat{M}_i(t)dt\Big\} \end{aligned} \tag{5.6}$$

$$
\begin{aligned}
&= E\Big\{\int_{-\infty}^{\infty} f'_\Delta(W+t)\hat M(t)dt\Big\}\\
&\ge E\Big\{\int_{|t|\le\delta} f'_\Delta(W+t)\hat M(t)dt\Big\}\\
&\ge E\Big\{I(z\le W\le z+|\Delta|)\int_{|t|\le\delta}\hat M(t)dt\Big\}\\
&= \sum_{1\le i\le n} E\Big\{I(z\le W\le z+|\Delta|)|\xi_i|\min(\delta,|\xi_i|)\Big\}\\
&\ge H_{1,1}-H_{1,2},
\end{aligned}
$$

where

$$
H_{1,1} \;=\; P(z\le W\le z+|\Delta|)\sum_{1\le i\le n}E\eta_i,\;\; H_{1,2}=E|\sum_{1\le i\le n}\eta_i-E\eta_i|,\;\; \eta_i=|\xi_i|\min(\delta,|\xi_i|).
$$

By (2.2),

$$
\sum_{1\le i\le n}E\eta_i\ge 1/2.
$$

Hence

$$
H_{1,1}\ge(1/2)P(z\le W\le z+|\Delta|). \tag{5.7}
$$

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
H_{1,2} &\le \Big(E(\sum_{1\le i\le n}\eta_i-E\eta_i)^2\Big)^{1/2}\\
&\le \Big(\sum_{1\le i\le n}E\eta_i^2\Big)^{1/2}\le\delta.
\end{aligned} \tag{5.8}
$$

As to H_2, it is easy to see that

$$
|f_\Delta(w)-f_{\Delta_i}(w)|\le\Big||\Delta|-|\Delta_i|\Big|/2\le|\Delta-\Delta_i|/2.
$$

Hence

$$
|H_2|\le(1/2)\sum_{i=1}^n E|\xi_i(\Delta-\Delta_i)|. \tag{5.9}
$$

Combining (5.5), (5.7), (5.8) and (5.9) yields

$$
\begin{aligned}
P(z\le W\le z+|\Delta|) &\le 2\Big\{E|Wf_\Delta(W)|+\delta+(1/2)\sum_{i=1}^n E|\xi_i(\Delta-\Delta_i)|\Big\}\\
&\le E|W\Delta|+2\delta E|W|+2\delta+\sum_{i=1}^n E|\xi_i(\Delta-\Delta_i)|\\
&\le 4\delta+E|W\Delta|+\sum_{i=1}^n E|\xi_i(\Delta-\Delta_i)|.
\end{aligned}
$$

This proves (5.2). Similarly, one can prove (5.3) and hence Theorem 2.1. ∎

Proof of Theorem 2.2. First, we prove (2.9). For $|z| \le 4$, (2.9) holds by (2.5). For $|z| > 4$, consider two cases.

Case 1. $\sum_{i=1}^n E|g_i(X_i)|^p > 1/2$.

By the Rosenthal (1970) inequality, we have

$$
\begin{aligned}
P(|W| > (|z|-2)/3) &\le P(|W| > |z|/6) \le (|z|/6)^{-p} E|W|^p \\
&\le C(|z|+1)^{-p}\Big\{\Big(\sum_{i=1}^n Eg_i^2(X_i)\Big)^{p/2} + \sum_{i=1}^n E|g_i(X_i)|^p\Big\} \\
&\le C(|z|+1)^{-p}\sum_{i=1}^n E|g_i(X_i)|^p.
\end{aligned} \tag{5.10}
$$

Hence

$$
\begin{aligned}
|P(T\le z) - \Phi(z)| &\le P(|\Delta| > (|z|+1)/3) + P(|W| > (|z|-2)/3) + P(|N(0,1)| > |z|) \\
&\le P(|\Delta| > (|z|+1)/3) + C(|z|+1)^{-p}\sum_{i=1}^n E|g_i(X_i)|^p,
\end{aligned}
$$

which shows that (2.9) holds.

Case 2. $\sum_{i=1}^n E|g_i(X_i)|^p \le 1/2$.

Similar to (5.10), we have

$$
P(|W - g_i(X_i)| > (|z|-2)/3) \le C(|z|+1)^{-p}\Big\{\Big(\sum_{j=1}^n Eg_j^2(X_j)\Big)^{p/2} + \sum_{j=1}^n E|g_j(X_j)|^p\Big\} \le C(|z|+1)^{-p}
$$

and hence

$$
\begin{aligned}
\gamma_z &\le P(|\Delta| > (|z|+1)/3) + \sum_{i=1}^n ((|z|+1)/3)^{-p} E|g_i(X_i)|^p + \sum_{i=1}^n C(|z|+1)^{-p} E|g_i(X_i)|^p \\
&\le P(|\Delta| > (|z|+1)/3) + C(|z|+1)^{-p}\sum_{i=1}^n E|g_i(X_i)|^p.
\end{aligned}
$$

By Remark 2.1, we can choose

$$
\begin{aligned}
\delta &= \Big(\frac{2(p-2)^{p-2}}{(p-1)^{p-1}}\sum_{i=1}^n E|g_i(X_i)|^p\Big)^{1/(p-2)} \\
&\le \frac{2(p-2)^{p-2}}{(p-1)^{p-1}}\sum_{i=1}^n E|g_i(X_i)|^p.
\end{aligned}
$$

Combining the above inequalities with (2.6) and the non-uniform Berry-Essee bound for independent random variables yields (2.9).

Next we prove (2.6). The main idea of the proof is first to truncate $g_i(X_i)$ and then adopt the proof of Theorem 2.1 to the truncated sum. Without loss of generality, assume $z \ge 0$ as we can simply apply the result to $-T$. By (5.1), it suffices to show that

$$P(z - |\Delta| \le W \le z) \le \gamma_z + e^{-z/3}\tau \tag{5.11}$$

and

$$P(z \le W \le z + |\Delta|) \le \gamma_z + e^{-z/3}\tau. \tag{5.12}$$

Since the proof of (5.12) is similar to that of (5.11), we only prove (5.11). It is easy to see that

$$P(z - |\Delta| \le W \le z) \le P(|\Delta| > (z+1)/3) + P(z - |\Delta| \le W \le z, |\Delta| \le (z+1)/3).$$

Now (5.11) follows directly by Lemmas 5.1 and 5.2 below. This completes the proof of Theorem 2.2. ∎

Lemma **5.1** *Let*

$$\xi_i = g_i(X_i),\ \bar{\xi}_i = \xi_i I(\xi_i \le 1), \quad \bar{W} = \sum_{i=1}^n \bar{\xi}_i.$$

Then

$$\begin{aligned} & P(z - |\Delta| \le W \le z, |\Delta| \le (z+1)/3) \\ \le\ & P(z - |\Delta| \le \bar{W} \le z, |\Delta| \le (z+1)/3) \\ & + \sum_{i=1}^n P(\xi_i > (z+1)/3) + \sum_{i=1}^n P(W - \xi_i > (z-2)/3) P(|\xi_i| > 1). \end{aligned} \tag{5.13}$$

Proof. We have

$$\begin{aligned} & P(z - |\Delta| \le W \le z, |\Delta| \le (z+1)/3) \\ \le\ & P(z - |\Delta| \le W \le z, |\Delta| \le (z+1)/3, \max_{1\le i\le n} |\xi_i| \le 1) \\ & + P(z - |\Delta| \le W \le z, |\Delta| \le (z+1)/3, \max_{1\le i\le n} |\xi_i| > 1) \\ \le\ & P(z - |\Delta| \le \bar{W} \le z, |\Delta| \le (z+1)/3) + \sum_{i=1}^n P(W > (2z-1)/3, |\xi_i| > 1) \end{aligned}$$

and

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

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

as desired. ∎

Lemma **5.2** *We have*

(5.14)
$$P(z - |\Delta| \le \bar W \le z, |\Delta| \le (z+1)/3) \le e^{-z/3}\tau.$$

Proof. Noting that $E\bar\xi_i \le 0$, $e^s \le 1 + s + s^2(e^a - 1 - a)a^{-2}$ for $s \le a$ and $a > 0$ and that $a\bar\xi_i \le a$, we have for $a > 0$

(5.15)
$$
\begin{aligned}
Ee^{a\bar W} &= \prod_{i=1}^n Ee^{a\bar\xi_i} \\
&\le \prod_{i=1}^n \Big(1 + a\, E\bar\xi_i + (e^a - 1 - a)E\bar\xi_i^2\Big) \\
&\le \exp\Big((e^a - 1 - a)\sum_{i=1}^n E\bar\xi_i^2\Big) \\
&\le \exp\Big((e^a - 1 - a)\sum_{i=1}^n E\xi_i^2\Big) = \exp(e^a - 1 - a).
\end{aligned}
$$

In particular, we have $Ee^{\bar W/2} \le \exp(e^{1/2} - 1.5)$. If $\delta \ge 0.07$, then

$$
\begin{aligned}
& P(z - |\Delta| \le \bar W \le z, |\Delta| \le (z+1)/3) \\
&\le \; P(\bar W > (2z-1)/3) \le e^{-z/3+1/6} Ee^{\bar W/2} \\
&\le \; e^{-z/3}\exp(e^{.5} - 4/3) \le 1.38 e^{-z/3} \le 20\delta\, e^{-z/3}.
\end{aligned}
$$

This proves (5.14) when $\delta \ge 0.07$.

For $\delta < 0.07$, let

(5.16)
$$
f_\Delta(w) = \begin{cases} 0 & \text{for } w \le z - |\Delta| - \delta, \\ e^{w/2}(w - z + |\Delta| + \delta) & \text{for } \; z - |\Delta| - \delta \le w \le z + \delta, \\ e^{w/2}(|\Delta| + 2\delta) & \text{for } \; w > z + \delta. \end{cases}
$$

Put

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

By (5.5) and similar to (5.6), we have

$$E\{Wf_\Delta(\bar W)\} = E\Big\{\int_{-\infty}^{\infty} f'_\Delta(\bar W+t)\bar M(t)dt + \sum_{i=1}^n E\Big\{\xi_i(f_\Delta(\bar W-\bar\xi_i) - f_{\Delta_i}(\bar W-\bar\xi_i))\Big\} := G_1+G_2, \tag{5.17}$$

It follows from the fact that $\bar M(t)\ge 0$, $f'_\Delta(w)\ge e^{w/2}$ for $z-|\Delta|-\delta\le w\le z+\delta$ and $f'_\Delta(w)\ge 0$ for all w,

$$\begin{aligned} G_1 &\ge E\Big\{\int_{|t|\le\delta} f'_\Delta(\bar W+t)\bar M(t)dt \\ &\ge E\Big\{e^{\bar W/2}I(z-|\Delta|\le\bar W\le z, |\Delta|\le(z+1)/3)\int_{|t|\le\delta}\bar M(t)dt\Big\} \\ &= E\Big\{e^{\bar W/2}I(z-|\Delta|\le\bar W\le z, |\Delta|\le(z+1)/3)\Big\}\int_{|t|\le\delta}E\bar M(t)dt \\ &\quad +E\Big\{e^{\bar W/2}I(z-|\Delta|\le\bar W\le z, |\Delta|\le(z+1)/3)\int_{|t|\le\delta}(\bar M(t)-E\bar M(t))dt\Big\} \\ &\ge G_{1,1}-G_{1,2}, \end{aligned} \tag{5.18}$$

where

$$\begin{aligned} G_{1,1} &= e^{z/3-1/6}P(z-|\Delta|\le\bar W\le z,\ |\Delta|\le(z+1)/3)\int_{|t|\le\delta}E\bar M(t)dt, \\ G_{1,2} &= E\Big\{\int_{|t|\le\delta} e^{\bar W/2}|\bar M(t)-E\bar M(t)|dt\Big\}. \end{aligned}$$

By (2.2) and the assumption that $\delta\le 0.07$,

$$\begin{aligned} \int_{|t|\le\delta}E\bar M(t)dt &= \sum_{i=1}^n E|\xi_i|\min(\delta,|\bar\xi_i|) \\ &= \sum_{i=1}^n E|\xi_i|\min(\delta,|\xi_i|)\ge 1/2. \end{aligned}$$

Hence

$$G_{1,1}\ge(1/2)e^{z/3-1/6}P\Big(z-|\Delta|\le\bar W\le z, |\Delta|\le(z+1)/3\Big). \tag{5.19}$$

By (5.15), we have $Ee^{\bar W}\le\exp(e-2)<2.06$. It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} G_{1,2} &\le .5\int_{|t|\le\delta}\Big(0.5Ee^{\bar W}+2E|\hat M(t)-M(t)|^2\Big)dt \\ &\le 0.5\Big\{2.06\delta+2\sum_{i=1}^n\int_{|t|\le\delta}E\xi_i^2(I(-\bar\xi_i\le t\le 0)+I(0<t\le-\bar\xi_i))dt\Big\} \end{aligned} \tag{5.20}$$

$$
\begin{aligned}
&= 0.5\Big\{2.06\delta + 2\sum_{i=1}^n E\xi_i^2 \min(\delta, |\bar\xi_i|)\Big\} \\
&\le 0.5\Big\{2.06\delta + 2\delta\sum_{i=1}^n E\xi_i^2\Big\} \le 2.03\delta.
\end{aligned}
$$

As to G_2, it is easy to see that

$$
|f_\Delta(w) - f_{\Delta_i}(w)| \le e^{w/2}\Big||\Delta| - |\Delta_i|\Big| \le e^{w/2}|\Delta - \Delta_i|.
$$

Hence, by the Hölder inequality, (5.15) and the assumption that ξ_i and $\bar W - \bar\xi_i$ are independent

$$
\begin{aligned}
|G_2| &\le \sum_{i=1}^n E|\xi_i e^{(\bar W - \bar\xi_i)/2}(\Delta - \Delta_i)| \\
&\le \sum_{i=1}^n \Big(E\xi_i^2 e^{\bar W - \bar\xi_i}\Big)^{1/2}\Big(E(\Delta-\Delta_i)^2\Big)^{1/2} \\
&= \sum_{i=1}^n \Big(E\xi_i^2 E e^{\bar W - \bar\xi_i}\Big)^{1/2}||\Delta - \Delta_i||_2 \\
&\le 1.44\sum_{i=1}^n ||\xi_i||_2\,||\Delta - \Delta_i||_2.
\end{aligned}
\tag{5.21}
$$

Following the proof of (5.15) and by using $|e^s - 1| \le |s|(e^a - 1)/a$ for $s \le a$ and $a > 0$, we have

$$
\begin{aligned}
EW^2 e^{\bar W} &= \sum_{i=1}^n E\xi_i^2 e^{\bar\xi_i} E e^{\bar W - \bar\xi_i} + \sum_{1\le i\ne j\le n} E\xi_i(e^{\bar\xi_i} - 1)E\xi_j(e^{\bar\xi_j} - 1)Ee^{\bar W - \bar\xi_i - \bar\xi_j} \\
&\le 2.06\,e\sum_{i=1}^n E\xi_i^2 + 2.06(e-1)^2\sum_{1\le i\ne j\le n} E\xi_i^2 E\xi_j^2 \\
&\le 2.06\,e + 2.06(e-1)^2 < 3.42^2.
\end{aligned}
$$

Thus, we obtain

$$
\begin{aligned}
E\{Wf_\Delta(\bar W)\} &\le E|W|e^{\bar W/2}(|\Delta| + 2\delta) \\
&\le \Big\{||\Delta||_2 + 2\delta\Big\}\Big(E(W^2 e^{\bar W})\Big)^{1/2} \\
&\le 3.42(||\Delta||_2 + 2\delta).
\end{aligned}
\tag{5.22}
$$

Combining (5.17), (5.19), (5.20), (5.21) and (5.22) yields

$$
\begin{aligned}
&P(z - |\Delta| \le \bar W \le z, |\Delta| \le (z+1)/3) \\
&\le 2e^{-z/3+1/6}\Big\{3.42(||\Delta||_2 + 2\delta) + 2.03\delta + 1.44\sum_{i=1}^n ||\xi_i||_2\,||\Delta - \Delta_i||_2\Big\} \\
&\le e^{-z/3}\Big\{21\delta + 8.1||\Delta||_2 + 3.5\sum_{i=1}^n ||\xi_i||_2\,||\Delta - \Delta_i||_2\Big\} \\
&= e^{-z/3}\tau.
\end{aligned}
$$

This proves (5.14). ∎

Proof of Remark 2.1. It is known that for $x \ge 0, y \ge 0,\ \alpha > 0, \gamma > 0$ with $\alpha + \gamma = 1$

$$x^\alpha y^\gamma \le \alpha x + \gamma y,$$

which yields with $\alpha = (p-2)/(p-1)$, $\gamma = 1/(p-1)$, $x = b(p-1)/(p-2)$ and $y = \Big((\frac{p-2}{p-1})^{p-2}\frac{a^{p-1}}{b^{p-2}}\Big)^{1/(p-1)}$,

$$a = x^\alpha y^\gamma \le \alpha x + \gamma y = b + \frac{(p-2)^{p-2}}{(p-1)^{p-1}}\frac{a^{p-1}}{b^{p-2}}$$

or

$$b \ge a - \frac{(p-2)^{p-2}}{(p-1)^{p-1}}\frac{a^{p-1}}{b^{p-2}}.$$

On the other hand, it is clear that

$$a \ge a - \frac{(p-2)^{p-2}}{(p-1)^{p-1}}\frac{a^{p-1}}{b^{p-2}}.$$

This proves (2.11). Now (2.2) follows directly from (2.11), (2.10) and the assumption (1.1). ∎

Proof of Remark 2.2. Note that $\delta = \beta/2 \le 1/4$. Applying (2.11) with $p = 3$ yields

$$\begin{aligned}
&\sum_{i=1}^n E|g_i(X_i)|\min(\delta, |g_i(X_i)|) \\
&\ge \sum_{i=1}^n E|g_i(X_i)| I(|g_i(X_i)| \le 1)\min(\delta, |g_i(X_i)|) \\
&\ge \sum_{i=1}^n \Big\{Eg_i^2(X_i)I(|g_i(X_i)| \le 1) - E|g_i(X_i)|^3 I(|g_i(X_i)| \le 1)/(4\delta)\Big\} \\
&= 1 - \Big(4\delta\sum_{i=1}^n Eg_i^2(X_i)I(|g_i(X_i)| > 1) + \sum_{i=1}^n E|g_i(X_i)|^3 I(|g_i(X_i)| \le 1)\Big)/(4\delta) \\
&\ge 1 - \beta/(4\delta) = 1/2.
\end{aligned}$$

This proves Remark 2.2. ∎

6 Proofs of Other Theorems

In this section, we prove Theorems 3.1 - 3.5.

6.1 Proof of Theorem 3.1

For $1 \le k \le m$, let $h_k(x_1, \ldots, x_k) = E(h(X_1, \ldots, X_m)|X_1 = x_1, \ldots, X_k = x_k)$ and $\bar{h}_k(x_1, \ldots, x_k) = h_k(x_1, \ldots, x_k) - \sum_{i=1}^k g(x_i)$. Observing that

$$U_n = n^{-1} m \sum_{i=1}^n g(X_i) + \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} \bar{h}_m(X_{i_1}, \ldots, X_{i_m}),$$

we have

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

where

$$\begin{aligned} W &= \frac{1}{\sqrt{n}\sigma_1} \sum_{i=1}^n g(X_i), \\ \Delta &= \frac{\sqrt{n}}{m\sigma_1} \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} \bar{h}_m(X_{i_1}, \ldots, X_{i_m}). \end{aligned}$$

Let

$$\Delta_l = \frac{\sqrt{n}}{m\sigma_1} \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m, i_j \ne l \text{ for all } j} \bar{h}_m(X_{i_1}, \ldots, X_{i_m}).$$

By Theorems 2.1 and 2.2 (with Remark 2.3 for proof of (3.1)), it suffices to show that

(6.1) $$E\Delta^2 \le \frac{(m-1)^2 \sigma^2}{m(n-m+1)\sigma_1^2}$$

and

(6.2) $$E|\Delta - \Delta_l|^2 \le \frac{2(m-1)^2 \sigma^2}{nm(n-m+1)\sigma_1^2}.$$

It is known that (see, e.g., [18], p.271)

(6.3) $$\begin{aligned} & E\Big(\sum_{1 \le i_1 < \ldots < i_m \le n} \bar{h}_m(X_{i_1}, \ldots, X_{i_m}) \Big)^2 \\ &= \binom{n}{m} \sum_{j=2}^m \binom{m}{j} \binom{n-m}{m-j} E\bar{h}_j^2(X_1, \ldots, X_j). \end{aligned}$$

Note that

(6.4) $$\begin{aligned} & E\bar{h}_j^2(X_1, \ldots, X_j) \\ &= Eh_j^2(X_1, \ldots, X_j) - 2\sum_{i=1}^j E[g(X_i) h_k(X_1, \ldots, X_j)] + E(\sum_{i=1}^j g(X_i))^2 \\ &= Eh_j^2(X_1, \ldots, X_j) - 2jE[g(X_1)E(h(X_1, \ldots, X_m)|X_1, \ldots, X_j)] + kEg^2(X_1) \end{aligned}$$

$$
\begin{aligned}
&= Eh_j^2(X_1,\dots,X_j) - 2jE[g(X_1)h(X_1,\dots,X_m)] + jEg^2(X_1) \\
&= Eh_j^2(X_1,\dots,X_j) - 2jEg^2(X_1) + jEg^2(X_1) \\
&= Eh_j^2(X_1,\dots,X_j) - jEg_1^2(X_1).
\end{aligned}
$$

We next prove that for $2 \le j \le m$

$$
Eh_{j-1}^2(X_1\dots,X_{j-1}) \le \frac{j-1}{j} Eh_j^2(X_1,\dots,X_j) \tag{6.5}
$$

Since $E\bar{h}_2^2(X_1,X_2) \ge 0$, (6.5) holds for $j=2$ by (6.4). Assume that (6.5) is true for j. Then

$$
\begin{aligned}
&E(h_{j+1}(X_1,\dots,X_{j+1}) - h_j(X_1,\dots,X_j) - h_j(X_2,\dots,X_{j+1}))^2 \\
&\quad= Eh_{j+1}^2(X_1,\dots,X_{j+1}) - 4E[h_{j+1}(X_1,\dots,X_{j+1})h_j(X_1,\dots,X_j)] \\
&\qquad +2Eh_j^2(X_1,\dots,X_j) + 2Eh_j(X_1,\dots,X_j)h_j(X_2,\dots,X_{j+1}) \\
&\quad= Eh_{j+1}^2(X_1,\dots,X_{j+1}) - 2Eh_j^2(X_1,\dots,X_j) \\
&\qquad +2E\Big(E(h_j(X_1,\dots,X_j)h_j(X_2,\dots,X_{j+1}) \mid X_2,\dots,X_j)\Big) \\
&\quad= Eh_{j+1}^2(X_1,\dots,X_{j+1}) - 2Eh_j^2(X_1,\dots,X_j) + 2Eh_{j-1}^2(X_1,\dots,X_{j-1}).
\end{aligned} \tag{6.6}
$$

On the other hand, we have

$$
\begin{aligned}
&E(h_{j+1}(X_1,\dots,X_{j+1}) - h_j(X_1,\dots,X_j) - h_j(X_2,\dots,X_{j+1}))^2 \\
&\quad\ge E\Big(E(h_{j+1}(X_1,\dots,X_{j+1}) - h_j(X_1,\dots,X_j) - h_j(X_2,\dots,X_{j+1}) \mid X_1,\dots,X_j)\Big)^2 \\
&\quad= Eh_{j-1}^2(X_1,\dots,X_{j-1}).
\end{aligned} \tag{6.7}
$$

Combining (6.5) and (6.6) yields

$$
\begin{aligned}
2Eh_j^2(X_1,\dots,X_j) &\le Eh_{j+1}^2(X_1,\dots,X_{j+1}) + Eh_{j-1}^2(X_1,\dots,X_{j-1}) \\
&\le Eh_{j+1}^2(X_1,\dots,X_{j+1}) + \frac{j-1}{j} Eh_j^2(X_1,\dots,X_j)
\end{aligned}
$$

by the induction hypothesis, which in turn reduces to (6.4) for $j+1$. This proves (6.4).

It follows from (6.4) that

$$
Eh_j^2(X_1,\dots,X_j) \le \frac{j}{m} Eh_m^2(X_1,\dots,X_m) = \frac{j}{m}\sigma^2. \tag{6.8}
$$

To complete the proof of (6.1), we need the following two inequalities:

$$
\sum_{j=2}^{m} \binom{m}{j}\binom{n-m}{m-j}\frac{j}{m} \le \frac{m(m-1)^2}{(n-m+1)n}\binom{n}{m} \tag{6.9}
$$

and

$$\sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j}\frac{j+1}{m} \le \frac{2(m-1)^2}{(n-m+1)n}\binom{n}{m} \tag{6.10}$$

for $n > m \geq 2$. In fact, we have

$$\begin{aligned}
&\sum_{j=2}^{m} \binom{m}{j}\binom{n-m}{m-j}\frac{j}{m} \\
&= \sum_{j=2}^{m} \binom{m-1}{j-1}\binom{n-m}{m-j} \\
&= \sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j} \\
&= \binom{n-1}{m-1} - \binom{n-m}{m-1} \\
&= \binom{n-1}{m-1}\Big\{1 - \frac{(n-m)!/(n-m-m+1)!}{(n-1)!/(n-m)!}\Big\} \\
&= \binom{n-1}{m-1}\Big\{1 - \prod_{j=n-m+1}^{n-1}(1 - \frac{m-1}{j})\Big\} \\
&\le \binom{n-1}{m-1} \sum_{j=n-m+1}^{n-1} \frac{m-1}{j} \\
&\le \binom{n-1}{m-1}\frac{(m-1)^2}{n-m+1} \\
&= \frac{(m-1)^2 m}{(n-m+1)n}\binom{n}{m}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j}\frac{j+1}{m} \\
&= \frac{m-1}{m}\sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j}\frac{j}{m-1} + \frac{1}{m}\sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j} \\
&= \frac{m-1}{m}\sum_{j=0}^{m-2} \binom{m-2}{j}\binom{n-m}{m-2-j} + \frac{1}{m}\sum_{j=1}^{m-1} \binom{m-1}{j}\binom{n-m}{m-1-j} \\
&= \frac{m-1}{m}\binom{n-2}{m-2} + \frac{1}{m}\{\binom{n-1}{m-1} - \binom{n-m}{m-1}\} \\
&\le \binom{n-1}{m-1}\Big(\frac{(m-1)^2}{m(n-1)} + \frac{(m-1)^2}{(n-m+1)m}\Big) \\
&\le \frac{2(m-1)^2}{(n-m+1)n}\binom{n}{m}.
\end{aligned}$$

From (6.8) and (6.9) we obtain that

$$
\begin{aligned}
E\Delta^2 &= \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-2} E\Big\{\sum_{1\le i_1<\ldots<i_m\le n} \bar{h}_m(X_1,\ldots,X_m)\Big\}^2 \\
&\le \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-1}\sum_{j=2}^{m}\binom{m}{j}\binom{n-m}{m-j} E h_j^2(X_1,\ldots,X_j) \\
&\le \frac{n\sigma^2}{m^2\sigma_1^2}\binom{n}{m}^{-1}\sum_{j=2}^{m}\binom{m}{j}\binom{n-m}{m-j}\frac{j}{m} \\
&\le \frac{(m-1)^2\sigma^2}{(n-m+1)\,m\,\sigma_1^2}.
\end{aligned}
\tag{6.11}
$$

This proves (6.1).

Similarly, by (6.10)

$$
\begin{aligned}
E(\Delta-\Delta_l)^2 &= \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-2} E\Big(\Big\{\sum_{1\le i_1<\ldots<i_m\le n} - \sum_{1\le i_1<\ldots<i_m\le n,\ all\ \ i_j\ne l}\Big\}\bar{h}_m(X_{i_1},\ldots,X_{i_m})\Big)^2 \\
&= \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-2} E\Big(\sum_{1\le i_1<\ldots<i_{m-1}\le n-1} \bar{h}_m(X_{i_1},\ldots,X_{i_{m-1}},X_m)\Big)^2 \\
&= \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-2}\binom{n-1}{m-1}\sum_{j=1}^{m-1}\binom{m-1}{j}\binom{n-m}{m-1-j} E\bar{h}_{j+1}^2(X_1,\ldots,X_j) \\
&\le \frac{n}{m^2\sigma_1^2}\binom{n}{m}^{-2}\binom{n-1}{m-1}\sum_{j=1}^{m-1}\binom{m-1}{j}\binom{n-m}{m-1-j} E h_{j+1}^2(X_1,\ldots,X_j) \\
&\le \frac{\sigma^2}{m\sigma_1^2}\binom{n}{m}^{-1}\sum_{j=1}^{m-1}\binom{m-1}{j}\binom{n-m}{m-1-j}\frac{j+1}{m} \\
&\le \frac{2(m-1)^2\,\sigma^2}{m\,n(n-m+1)\,\sigma_1^2}.
\end{aligned}
\tag{6.12}
$$

This proves (6.2) and hence completes the proof of Theorem 3.1. ∎

6.2 Proof of Theorem 3.2

We follow a similar argument as that in the proof of Theorem 3.1. For $1 \le j \le k$, let $\boldsymbol{X}_j = (X_{j1},\ldots,X_{jm_j})$ and $\boldsymbol{x}_j = (x_{j1},\ldots,x_{jm_j})$ and define

$$\bar{h}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) = h(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) - \sum_{j=1}^{k}\sum_{i=1}^{m_j} h_j(x_{ji}).$$

For the given U-statistic $U_{\bar{n}}$, we define its projection

$$\hat{U}_{\bar{n}} = \sum_{j=1}^{k}\sum_{l=1}^{n_j} E(U_{\bar{n}}|X_{jl}).$$

Since

$$m_j/n_j = \binom{n_j - 1}{m_j - 1} / \binom{n_j}{m_j},$$

we have

$$\hat{U}_{\bar{n}} = \sum_{j=1}^{k} \sum_{l=1}^{n_j} \frac{m_j}{n_j} h_j(X_{jl}).$$

The difference $U_{\bar{n}} - \hat{U}_{\bar{n}}$ can be rewritten as

$$U_{\bar{n}} - \hat{U}_{\bar{n}} = \Big\{ \prod_{j=1}^{k} \binom{n_j}{m_j}^{-1} \Big\} \sum \bar{h}(\boldsymbol{X}_{1\boldsymbol{i}_1}, \ldots, \boldsymbol{X}_{k\boldsymbol{i}_k}),$$

where $\boldsymbol{X}_{j\boldsymbol{i}_j} = (X_{ji_{j1}}, \ldots, X_{ji_{jm_j}})$ and the summation is carried out over all indices $1 \le i_{j1} < i_{j2} < \ldots < i_{jm_j} \le n_j$, $j = 1, 2, \ldots, k$. Thus, we have

$$\sigma_{\bar{n}}^{-1} U_{\bar{n}} = W + \Delta$$

with

$$\begin{aligned} W &= \sigma_{\bar{n}}^{-1} \sum_{j=1}^{k} \sum_{l=1}^{n_j} \frac{m_j}{n_j} h_j(X_{jl}), \\ \Delta &= \sigma_{\bar{n}}^{-1} \Big\{ \prod_{j=1}^{k} \binom{n_j}{m_j}^{-1} \Big\} \sum \bar{h}(\boldsymbol{X}_{1\boldsymbol{i}_1}, \ldots, \boldsymbol{X}_{k\boldsymbol{i}_k}). \end{aligned}$$

Let

$$\Delta_{jl} = \sigma_{\bar{n}}^{-1} \Big\{ \prod_{v=1}^{k} \binom{n_v}{m_v}^{-1} \Big\} \sum{}^{(jl)} \bar{h}(\boldsymbol{X}_{1\boldsymbol{i}_1}, \ldots, \boldsymbol{X}_{k\boldsymbol{i}_k}),$$

where the summation is carried out over all indices $1 \le i_{v1} < i_{v2} < \ldots < i_{vm_v} \le n_v$, $1 \le v \le k, v \ne j$ and $1 \le i_{j1} < i_{j2} < \ldots < i_{jm_j} \le n_j$ with $i_{js} \ne l$ for $1 \le s \le m_j$.

By Theorems 2.1 and 2.2, it suffices to show that

$$E\Delta^2 \le \frac{\sigma^2}{\sigma_{\bar{n}}^2} \Big(\sum_{j=1}^{k} \frac{m_j^2}{n_j} \Big)^2 \tag{6.13}$$

and

$$E|\Delta - \Delta_{jl}|^2 \le \frac{2\sigma^2 \, m_j^2}{n_j^2 \, \sigma_{\bar{n}}^2} \sum_{v=1}^{k} \frac{m_v^2}{n_v} \tag{6.14}$$

For $0 \le d_j \le m_j, 1 \le j \le k$ let

$$Y_{d_1,\ldots,d_k}(x_{ji}, 1 \le i \le d_j, 1 \le j \le k) = E\bar{h}(x_{j1}, \ldots, x_{jd_1}, X_{jd_j+1}, \ldots, X_{jm_j}, 1 \le j \le k)$$

and

$$y_{d_1,\ldots,d_k} = EY^2_{d_1,\ldots,d_k}(X_{ji}, 1 \le i \le d_j, 1 \le j \le k).$$

Noting that

$$E(\bar{h}(\boldsymbol{X}_{1\boldsymbol{i}_1}, \ldots, \boldsymbol{X}_{k\boldsymbol{i}_k})|X_{jl}) = 0 \text{ for every } 1 \le l \le m_j,\ 1 \le j \le k,$$

we have (see (4.5.8) in [18])

$$\begin{aligned}
(6.15)\qquad & E\Big(U_{\bar{n}} - \hat{U}_{\bar{n}}\Big)^2 \\
= & \Big\{\prod_{j=1}^{k} \binom{n_j}{m_j}^{-1}\Big\} \sum_{\substack{d_1 + \ldots + d_k \ge 2 \\ 0 \le d_j \le m_j, 1 \le j \le k}} \prod_{j=1}^{k} \Big\{\binom{m_j}{d_j}\binom{n_j - m_j}{m_j - d_j}\Big\} y_{d_1,\ldots,d_k} \\
\le & \ \sigma^2 \Big\{\prod_{j=1}^{k} \binom{n_j}{m_j}^{-1}\Big\} \sum_{\substack{d_1 + \ldots + d_k \ge 2 \\ 0 \le d_j \le m_j, 1 \le j \le k}} \prod_{j=1}^{k} \Big\{\binom{m_j}{d_j}\binom{n_j - m_j}{m_j - d_j}\Big\} \\
\le & \ \sigma^2 \Big(\sum_{j=1}^{k} \frac{m_j^2}{n_j}\Big)^2,
\end{aligned}$$

where in the last inequality we used the fact that

$$(6.16)\qquad \sum_{\substack{d_1 + \ldots + d_k \ge 2 \\ 0 \le d_j \le m_j, 1 \le j \le k}} \prod_{j=1}^{k} \Big\{\binom{m_j}{d_j}\binom{n_j - m_j}{m_j - d_j}\Big\} \le \Big(\sum_{j=1}^{k} \frac{m_j^2}{n_j}\Big)^2 \prod_{j=1}^{k} \binom{n_j}{m_j}.$$

(See below for proof). This proves (6.13).

As to (6.14), consider $j = 1$ only. Similar to (6.12) and (6.15), we have (with $\boldsymbol{X}^*_{1\boldsymbol{i}_1} = (X_{1i_{1,1}}, \ldots, X_{1i_{1,m_1-1}}, X_{1,m_1})$)

$$\begin{aligned}
& \sigma^2_{\bar{n}} E|\Delta - \Delta_{1l}|^2 \\
= & \Big\{\prod_{v=1}^{k} \binom{n_v}{m_v}\Big\}^{-2} E\Big(\sum_{\substack{1 \le i_{v1} < i_{v2} < \ldots < i_{vm_v} \le n_v, 2 \le v \le k \\ 1 \le i_{1,1} < i_{1,2} < \ldots < i_{1,m_j-1} \le n_j - 1}} \bar{h}(\boldsymbol{X}^*_{1\boldsymbol{i}_1}, \boldsymbol{X}_{2\boldsymbol{i}_2}, \ldots, \boldsymbol{X}_{k\boldsymbol{i}_k})\Big)^2 \\
\le & \ \sigma^2 \Big\{\prod_{v=1}^{k} \binom{n_v}{m_v}\Big\}^{-2} \binom{n_1 - 1}{m_1 - 1} \prod_{v=2}^{k} \binom{n_v}{m_v} \\
& \sum_{\substack{d_1 + \ldots + d_k \ge 1 \\ 0 \le d_1 \le m_1 - 1, 0 \le d_v \le m_v, 2 \le v \le k}} \binom{m_1 - 1}{d_1}\binom{n_1 - m_1}{m_1 - 1 - d_1} \prod_{v=2}^{k} \binom{m_v}{d_v}\binom{n_v - m_v}{m_v - d_v}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{\sigma^2 m_1}{n_1}\Big\{\prod_{v=1}^{k}\binom{n_v}{m_v}\Big\}^{-1}\Big\{\sum_{1\le d_1\le m_1-1} + \sum_{2\le v\le k}\ \sum_{d_1=0,1\le d_v\le m_v}\Big\} \\
&\le \frac{\sigma^2 m_1}{n_1}\Big\{\prod_{v=1}^{k}\binom{n_v}{m_v}\Big\}^{-1}\Big\{\frac{(m_1-1)^2}{n_1-m_1+1}\binom{n_1-1}{m_1-1}\prod_{j=2}^{k}\binom{n_j}{m_j} \\
&\qquad\qquad + \binom{n_1-m_1}{m_1-1}\sum_{2\le v\le k}\frac{m_v^2}{n_v-m_v+1}\prod_{j=2}^{k}\binom{n_j}{m_j}\Big\} \\
&\le \frac{\sigma^2 m_1^2}{n_1^2}\sum_{1\le v\le k}\frac{m_v^2}{n_v-m_v+1} \le \frac{2\sigma^2 m_1^2}{n_1^2}\sum_{1\le v\le k}\frac{m_v^2}{n_v}
\end{aligned}
$$

for $n_1 \ge 2m_1$. This proves (6.14).

Now we prove (6.16). Consider two cases in the summation:

Case 1: At least one of $d_j \ge 2$, say $d_1 \ge 2$. In this case, by (6.9)

$$
\begin{aligned}
&\sum_{\substack{d_1\ge 2\\ 1\le d_j\le m_j,1\le j\le k}} \prod_{j=1}^{k}\Big\{\binom{m_j}{d_j}\binom{n_j-m_j}{m_j-d_j}\Big\} \\
&\le \Big\{\prod_{j=2}^{k}\binom{n_j}{m_j}\Big\}\sum_{2\le d_1\le m_1}\binom{m_1}{d_1}\binom{n_1-m_1}{m_1-d_1} \\
&\le \Big\{\prod_{j=2}^{k}\binom{n_j}{m_j}\Big\}\frac{m}{2}\sum_{2\le d_1\le m_1}\binom{m_1}{d_1}\binom{n_1-m_1}{m_1-d_1}\frac{d_1}{m} \\
&\le \frac{m_1^2(m_1-1)^2}{2n_1(n_1-m_1+1)}\prod_{j=1}^{k}\binom{n_j}{m_j} \\
&\le \frac{m_1^4}{n_1^2}\prod_{j=1}^{k}\binom{n_j}{m_j}
\end{aligned}
$$

for $n_1 \ge 2m_1$.

Case 2. At least two of $\{d_j\}$ are equal to 1, say $d_1 = d_2 = 1$. Then

$$
\begin{aligned}
&\sum_{d_1=d_2=1}\prod_{j=1}^{k}\Big\{\binom{m_j}{d_j}\binom{n_j-m_j}{m_j-d_j}\Big\} \\
&= m_1 m_2\binom{n_1-m_1}{m_1-1}\binom{n_2-m_2}{m_2-1}\prod_{j=3}^{k}\binom{n_j}{m_j} \\
&\le \frac{m_1^2 m_2^2}{n_1 n_2}\prod_{j=1}^{k}\binom{n_j}{m_j}.
\end{aligned}
$$

Thus, we have

$$\begin{aligned} &\sum_{\substack{d_1+\ldots+d_k\ge 2\\ 1\le d_j\le m_j, 1\le j\le k}} \prod_{j=1}^k \Big\{\binom{m_j}{d_j}\binom{n_j-m_j}{m_j-d_j}\Big\} \\ \le\ & \Big(\sum_{j=1}^k \frac{m_j^4}{n_j^2} + \sum_{1\le i\ne j\le k} \frac{m_i^2\, m_j^2}{n_i\, n_j}\Big) \prod_{j=1}^k \binom{n_j}{m_j} \\ =\ & \Big(\sum_{j=1}^k \frac{m_j^2}{n_j}\Big)^2 \prod_{j=1}^k \binom{n_j}{m_j}. \end{aligned}$$

This proves (6.16). Now the proof of Theorem 3.2 is complete. ∎

6.3 Proof of Theorem 3.3

Let $\psi(t)=\int_0^t J(s)ds$. As in [Serfling (1980), p.265], we have

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

and hence

$$\sqrt{n}\sigma^{-1}(T(F_n)-T(F)) = W+\Delta,$$

where

$$\begin{aligned} W &= -\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n \int_{-\infty}^{\infty}(I(X_i\le x)-F(x))J(F(x))dx \\ \Delta &= -\sqrt{n}\sigma^{-1}\int_{-\infty}^{\infty}[\psi(F_n(x))-\psi(F(x))-(F_n(x)-F(x))J(F(x))]dx. \end{aligned}$$

Let

$$\begin{aligned} \eta_i(x) &= I(X_i\le x)-F(x), \\ g_i(X_i) &= -\frac{1}{\sqrt{n}\sigma}\int_{-\infty}^{\infty}(I(X_i\le x)-F(x))J(F(x))dx, \\ \Delta_i &= -\sqrt{n}\sigma^{-1}\int_{-\infty}^{\infty}[\psi(F_{n,i}(x))-\psi(F(x))-(F_{n,i}(x)-F(x))J(F(x))]dx, \end{aligned}$$

where $F_{n,i}(x)=\frac{1}{n}\Big\{F(x)+\sum_{1\le j\le n, j\ne i} I(X_j\le x)\Big\}$. We only need to prove

(6.17) $$\sigma^2 E\Delta^2 \le c_0^2\, n^{-1} EX_1^2$$

and

(6.18) $$\sigma^2 E|\Delta-\Delta_i|^2 \le 2c_0^2 n^{-2} EX_1^2$$

Observe that the Lipschitz condition (3.9) implies

$$|\psi(t)-\psi(s)-(t-s)J(s)| = |\int_0^{t-s}(J(u+s)-J(s))du| \le 0.5\,c_0(t-s)^2 \tag{6.19}$$

for $0 \le s,t \le 1$. Hence

$$\begin{aligned}
\sigma^2\,E\Delta^2 &\le 0.25c_0^2\,nE\Big(\int_{-\infty}^{\infty}(F_n(x)-F(x))^2dx\Big)^2\\
&= 0.25c_0^2\,n^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\Big(\sum_{i=1}^n\sum_{j=1}^n\eta_i(x)\eta_j(y)\Big)^2dxdy\\
&\le 0.25c_0^2\,n^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(3n^2E\eta_1^2(x)E\eta_1^2(y)+nE\{\eta_1^2(x)\eta_1^2(y)\}\Big)dxdy.
\end{aligned}$$

Observe that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\eta_1^2(x)E\eta_1^2(y)dxdy = \Big(\int_{-\infty}^{\infty}F(x)(1-F(x))dx\Big)^2 \le (E|X_1|)^2 \le EX_1^2$$

and

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\{\eta_1^2(x)\eta_1^2(y)\}dxdy\\
&= 2\int\int_{x\le y}E\{\eta_1^2(x)\eta_1^2(y)\}dxdy\\
&= 2\int\int_{x\le y}\Big\{(1-F(x))^2(1-F(y))^2F(x)\\
&\qquad +F^2(x)(1-F(y))^2(F(y)-F(x))+F^2(x)F^2(y)(1-F(y))\Big\}dxdy\\
&\le 2\int\int_{x\le y}F(x)(1-F(y))dxdy\\
&= 2\Big\{\int\int_{x\le y\le 0}+\int\int_{0<x\le y}+\int\int_{x\le 0,y>0}\Big\}F(x)(1-F(y))dxdy\\
&\le 2\Big\{\int_{x\le 0}|x|F(x)dx+\int_{y\ge 0}y(1-F(y))dy+\int_{x\le 0}F(x)dx\int_{y>0}(1-F(y))dy\Big\}\\
&\le 2\Big\{E(X_1^-)^2+E(X_1^+)^2+EX_1^-EX_1^+\Big\}\\
&\le 4EX_1^2
\end{aligned}$$

This proves (6.17).

Next we prove (6.18). Observe that

$$\begin{aligned}
\frac{\sigma}{\sqrt{n}}|\Delta-\Delta_i| &= |\int_{-\infty}^{\infty}[\psi(F_n(x))-\psi(F_{n,i}(x))-(F_n(x)-F_{n,i}(x))J(F_{n,i}(x))dx\\
&\qquad +\int_{-\infty}^{\infty}(F_n(x)-F_{n,i}(x))[J(F_{n,i}(x)-J(F(x))]dx|
\end{aligned}$$

$$
\begin{aligned}
&\le 0.5c_0\int_{-\infty}^{\infty}(F_n(x)-F_{n,i}(x))^2dx \\
&\quad +c_0\int_{-\infty}^{\infty}|F_n(x)-F_{n,i}(x)||F_{n,i}(x)-F(x)|dx \\
&\le 0.5c_0\, n^{-2}\int_{-\infty}^{\infty}(I(X_i\le x)-F(x))^2dx \\
&\quad +c_0\, n^{-2}\int_{-\infty}^{\infty}|I(X_i\le x)-F(x)|\,|\sum_{j\neq i}\{I(X_j\le x)-F(x)\}dx \\
&= 0.5c_0\, n^{-2}\int_{-\infty}^{\infty}\eta_i^2(x)dx + c_0\, n^{-2}\int_{-\infty}^{\infty}|\eta_i(x)|\,|\sum_{j\neq i}\eta_j(x)|dx,
\end{aligned}
$$

$$
E\Big(\int_{-\infty}^{\infty}\eta_i^2(x)dx\Big)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\eta_1^2(x)\eta_1^2(y)dxdy \le 4EX_1^2
$$

and

$$
\begin{aligned}
&E\Big(\int_{-\infty}^{\infty}|\eta_i(x)|\,|\sum_{j\neq i}\eta_j(x)|dx\Big)^2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\{|\eta_i(x)|\,|\sum_{j\neq i}\eta_j(x)||\eta_i(y)|\,|\sum_{j\neq i}\eta_j(y)|\}dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E|\eta_i(x)\eta_i(y)|E\{|\sum_{j\neq i}\eta_j(x)||\sum_{j\neq i}\eta_j(y)|\}dxdy \\
&\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}||\eta_i(x)||_2||\eta_i(y)||_2||\sum_{j\neq i}\eta_j(x)||_2\ ||\sum_{j\neq i}\eta_j(y)||_2dxdy \\
&= (n-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}||\eta_i(x)||_2^2||\eta_i(y)||_2^2dxdy \\
&\le (n-1)(E|X_1|)^2 \le (n-1)EX_1^2.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\sigma^2E|\Delta-\Delta_i|^2 &\le n^{-3}c_0^2E\Big(0.5\int_{-\infty}^{\infty}\eta_i^2(x)dx+\int_{-\infty}^{\infty}|\eta_i(x)|\,|\sum_{j\neq i}\eta_j(x)|dx\Big)^2 \\
&\le n^{-3}c_0^2\Big\{0.75E\Big(\int_{-\infty}^{\infty}\eta_i^2(x)dx\Big)^2+1.5E\Big(\int_{-\infty}^{\infty}|\eta_i(x)|\,|\sum_{j\neq i}\eta_j(x)|dx\Big)^2\Big\} \\
&\le n^{-3}c_0^2\Big\{3EX_1^2+1.5(n-1)EX_1^2\Big\} \\
&\le 2n^{-2}EX_1^2.
\end{aligned}
$$

This proves (6.18) and hence the theorem. ∎

6.4 Proof of Theorem 3.4

Let Z_1 and Z_2 be independent standard normal random variables that are independent of $\{X_i\}$ and $\{Y_i\}$. Put

$$b = \sqrt{\nu\sigma^2 + \tau^2\mu^2},\ T_n = \frac{\sum_{i=1}^{N_n} X_i - n\mu\nu}{\sqrt{n}b}, \quad H_n = \frac{\sum_{i=1}^{N_n} X_i - N_n\mu}{\sqrt{N_n}\sigma}$$

and write

$$T_n \ = \ \frac{\sqrt{N_n}\sigma}{\sqrt{n}b} H_n + \frac{(N_n - n\nu)\mu}{\sqrt{n}b}, \quad T_n(Z_1) = \frac{\sqrt{N_n}\sigma}{\sqrt{n}b} Z_1 + \frac{(N_n - n\nu)\mu}{\sqrt{n}b}.$$

Applying the Berry-Esseen bound to H_n for given N_n yields

$$\begin{aligned} \text{(6.20)} \qquad & \sup_x |P(T_n \le x) - P(T_n(Z_1) \le x)| \\ & \le \ P(|N_n - n\nu| > n\nu/2) + CE\Big(\frac{|X_1|^3}{\sqrt{N_n}\sigma^3} I\{|N_n - n\nu| \le n\nu/2\}\Big) \\ & \le \ 4n^{-1}\nu^{-2}\tau^2 + Cn^{-1/2}\nu^{-1/2}\sigma^{-3}E|X_1|^3. \end{aligned}$$

Let

$$\begin{aligned} x^* &= \begin{cases} .5n\nu & \text{for } x < .5n\nu \\ x & \text{for } .5n\nu \le x \le 1.5n\nu \\ 1.5n\nu & \text{for } x > 1.5n\nu \end{cases} \\ W_n &= \frac{N_n - n\nu}{\sqrt{n}\tau}, \\ \bar{T}_n(Z_1) &:= \frac{\sqrt{N_n^*}\sigma}{\sqrt{n}b} Z_1 + \frac{(N_n - n\nu)\mu}{\sqrt{n}b} \\ &= \frac{\tau\mu}{b}\Big(W_n + \frac{\sigma\sqrt{\nu}}{\tau\mu} Z_1 + \Delta\Big), \end{aligned}$$

where

$$\Delta = \frac{(\sqrt{N_n^*} - \sqrt{n\nu})\sigma Z_1}{\sqrt{n}\tau\mu}.$$

Let

$$\Delta_i = \frac{(\sqrt{(N_n - Y_i + \nu)^*} - \sqrt{n\nu})\sigma Z_1}{\sqrt{n}\tau\mu}.$$

Then

$$E(|W_n\Delta| \mid Z_1) \ \le \ \frac{|Z_1|\sigma}{\sqrt{n}\tau\mu} E\Big(\frac{|W_n(N_n - n\nu)|}{\sqrt{n\nu}} = \frac{|Z_1|\sigma}{\sqrt{n}\mu\sqrt{\nu}}$$

and

$$\frac{1}{\sqrt{n}\tau} E(|(Y_i - \nu)(\Delta - \Delta_i)| \mid Z_1) \ \le \ \frac{|Z_1|}{n^{3/2}\tau^2\mu\sqrt{\nu}} E(Y_i - \nu)^2 = \frac{|Z_1|\sigma}{n^{3/2}\mu\sqrt{\nu}}.$$

Now letting

$$T_n(Z_1, Z_2) = \frac{\tau\mu}{b}\Big(Z_2 + \frac{\sigma\sqrt{\nu}}{\tau\mu}Z_1\Big)$$

and applying Theorem 2.1 for given Z_1 yields

$$\begin{aligned}
(6.21)\qquad & \sup_x |P(T_n(Z_1) \le x) - P(T_n(Z_1, Z_2) \le x)| \\
\le\ & P(|N_n - n\nu| > 0.5n\nu) + \sup_x |P(\bar{T}_n(Z_1) \le x) - P(T_n(Z_1, Z_2) \le x)| \\
\le\ & \frac{4\tau^2}{n\nu^2} + C\Big(\frac{E|X_1|^3}{n^{1/2}\sigma^3} + \frac{E|Z_1|\sigma}{n^{1/2}\mu\sqrt{\nu}}\Big) \\
\le\ & Cn^{-1/2}\Big(\frac{\tau^2}{\nu^2} + \frac{E|X_1|^3}{\sigma^3} + \frac{\sigma}{\mu\sqrt{\nu}}\Big)
\end{aligned}$$

It is clear that $T_n(Z_1, Z_2)$ has a standard normal distribution. This proves (3.12) by (6.20) and (6.21).

∎

6.5 Proof of Theorem 3.5

Since (3.13) is trivial if $\sum_{i=1}^n E|g_i(X_i)|^p > 1/6$, we assume

$$(6.22)\qquad \sum_{i=1}^n E|g_i(X_i)|^p \le 1/6.$$

Let $W = \sum_{i=1}^n g_i(X_i)$. It is known that for $2 < p \le 3$

$$(6.23)\qquad E|W|^p \le 2(EW^2)^{p/2} + \sum_{i=1}^n E|g_i(X_i)|^p \le 2.2.$$

Observe that

$$\begin{aligned}
(6.24)\qquad \frac{\sqrt{n}(h(\hat{\Theta}_n) - h(\theta))}{h'(\theta)} &= \frac{\sqrt{n}}{h'(\theta)}\Big(h'(\theta)(\hat{\Theta}_n - \theta) + \int_0^{\hat{\Theta}_n - \theta}[h'(\theta + t) - h'(\theta)]dt\Big) \\
&= W + \Delta + \frac{\sqrt{n}}{h'(\theta)}\int_0^{n^{-1/2}(W+\Delta)}[h'(\theta + t) - h'(\theta)]dt \\
&:= W + \Lambda + R,
\end{aligned}$$

where

$$\begin{aligned}
\Lambda &= \Delta + \frac{\sqrt{n}}{h'(\theta)}\int_0^{(n^{-1/2}W)^* + (n^{-1/2}\Delta)^*}[h'(\theta + t) - h'(\theta)]dt, \\
R &= \frac{\sqrt{n}}{h'(\theta)}\int_{(n^{-1/2}W)^* + (n^{-1/2}\Delta)^*}^{n^{-1/2}(W+\Delta)}[h'(\theta + t) - h'(\theta)]dt, \\
x^* &= \begin{cases} -c_0/2 & \text{for } x < -c_0/2, \\ x & \text{for } -c_0/2 \le x \le c_0/2, \\ c_0/2 & \text{for } x > c_0/2. \end{cases}
\end{aligned}$$

Clearly, $|n^{-1/2}W| \le c_0/2$ and $|n^{-1/2}\Delta| \le c_0/2$ imply $R = 0$. Hence

$$P(|R| > 0) \le P(|W| > c_0 n^{1/2}/2) + P(|\Delta| > c_0 n^{1/2}/2) \le 4/(c_0^2 n) + 2E|\Delta|/(c_0 n^{1/2}). \tag{6.25}$$

To apply Theorem 2.1, let $W_i = W - g(X_i)$ and

$$\Lambda_i = \Delta_i + \frac{\sqrt{n}}{h'(\theta)} \int_0^{(n^{-1/2}W_i)^* + (n^{-1/2}\Delta_i)^*} [h'(\theta + t) - h'(\theta)]dt.$$

Noting that

$$\begin{aligned} (6.26) \qquad & |\int_0^{(n^{-1/2}W)^* + (n^{-1/2}\Delta)^*} [h'(\theta + t) - h'(\theta)]dt| \\ \le \;& 0.5\delta(c_0)\Big((n^{-1/2}W)^* + (n^{-1/2}\Delta)^*\Big)^2 \\ \le \;& \delta(c_0)\Big((n^{-1/2}W)^{*2} + (n^{-1/2}\Delta)^{*2}\Big) \\ \le \;& \delta(c_0)\Big((c_0/2)^{3-p}(n^{-1/2}W)^{p-1} + (c_0/2)n^{-1/2}|\Delta|\Big), \end{aligned}$$

we have

$$\begin{aligned} (6.27) \qquad E|W\Lambda| \;&\le\; E|W\Delta| + \frac{(c_0/2)^{3-p}\delta(c_0)}{|h'(\theta)|n^{(p-2)/2}} E|W|^p + \frac{c_0\delta(c_0)}{|h'(\theta)|} E|W\Delta| \\ &\le\; \Big(1 + \frac{c_0\delta(c_0)}{|h'(\theta)|}\Big) E|W\Delta| + \frac{2.2c_0^{3-p}\delta(c_0)}{|h'(\theta)|n^{(p-2)/2}}. \end{aligned}$$

Similar to (6.26),

$$\begin{aligned} & |\int_{(n^{-1/2}W_i)^* + (n^{-1/2}\Delta_i)^*}^{(n^{-1/2}W)^* + (n^{-1/2}\Delta)^*} [h'(\theta + t) - h'(\theta)]dt| \\ \le \;& \delta(c_0)\Big((c_0)^{3-p}|(n^{-1/2}W)^* - (n^{-1/2}W_i)^*|\; |(n^{-1/2}W)^* + (n^{-1/2}W_i)^*|^{p-2} \\ & \qquad + c_0|(n^{-1/2}\Delta)^* - (n^{-1/2}\Delta_i)^*|\Big) \\ \le \;& \delta(c_0)\Big(c_0^{3-p} n^{-(p-1)/2}|g(X_i)|(2|W_i|^{p-2} + |g(X_i)|^{p-2}) + c_0 n^{-1/2}|\Delta - \Delta_i|\Big). \end{aligned}$$

From this we obtain

$$\begin{aligned} (6.28) \qquad & \sum_{i=1}^n E|g(X_i)(\Lambda - \Lambda_i)| \\ \le \;& \sum_{i=1}^n E|g(X_i)(\Delta - \Delta_i)| \\ & + \frac{\sqrt{n}\delta(c_0)}{|h'(\theta)|}\Big\{(c_0)^{3-p} n^{-(p-1)/2} \sum_{i=1}^n E\Big(|g(X_i)|^2(2|W_i|^{p-2} + |g(X_i)|^{p-2})\Big) \end{aligned}$$

$$
\begin{aligned}
&\quad +c_0 n^{-1/2}\sum_{i=1}^n E|g(X_i)(\Delta-\Delta_i)|\Big\} \\
\le\ & \Big(1+\frac{c_0\delta(c_0)}{|h'(\theta)|}\Big)\sum_{i=1}^n E|g(X_i)(\Delta-\Delta_i)| \\
&+\frac{c_0^{3-p}\delta(c_0)}{|h'(\theta)|n^{(p-2)/2}}\sum_{i=1}^n (2Eg^2(X_i)+E|g(X_i)|^p) \\
\le\ & \Big(1+\frac{c_0\delta(c_0)}{|h'(\theta)|}\Big)\sum_{i=1}^n E|g(X_i)(\Delta-\Delta_i)|+\frac{2.2c_0^{3-p}\delta(c_0)}{|h'(\theta)|n^{(p-2)/2}}.
\end{aligned}
$$

This proves (3.13) by (2.5), (6.25), (6.27) and (6.28). ∎

6.6 Proofs of Example 4.1 and Remark 2.5

First we prove Example 4.1. Let Z denote a standard normally distributed random variable and $\phi(x)$ be the standard normal density function. Observe that $0 < c_0 < 2$ and

$$
\begin{aligned}
P(T\le \varepsilon\, c_0)-\Phi(\varepsilon\, c_0) &= P(Z-\varepsilon/|Z|^{1/2}\le 0)-\Phi(\varepsilon\, c_0) \\
&= P(Z\le 0)+P(Z^{3/2}\le\varepsilon, Z>0)-\Phi(\varepsilon\, c_0) \\
&= \int_0^{\varepsilon^{2/3}}\phi(t)dt-\int_0^{\varepsilon c_0}\phi(t)dt \\
&\ge \int_{2\varepsilon}^{\varepsilon^{2/3}}\phi(t)dt \ge (\varepsilon^{2/3}-2\varepsilon)/3\ge \varepsilon^{2/3}/6,
\end{aligned}
$$

which proves (4.2).

Clearly, we have

$$
\begin{aligned}
E|W\Delta|+E|\Delta| &= \varepsilon E\Big|c_0 Z - Z|Z|^{-1/2}\Big| + \varepsilon E\Big|c_0-|Z|^{-1/2}\Big| \\
&\le \varepsilon(c_0+1+2c_0)\le 7\varepsilon
\end{aligned}
$$

by the fact that $c_0 < 2$. This proves (4.3).

As to (4.4), observe that

$$
E|\Delta|+\sum_{i=1}^n E|X_i|^3 \le 2c\varepsilon + 4n^{-1/2}\le 8\varepsilon \qquad (6.29)
$$

provided $n > \varepsilon^{-2}$.

Below we bound α. Since $\{X_i\}$ are i.i.d., we have

$$
\alpha=\varepsilon E\Big||W|^{-1/2}-|\hat X_1+X_2+\dots X_n|^{-1/2}\Big|.
$$

Let Y and Z be independent standard normal random variables, and let $r = (n-1)/n$ and $s = \sqrt{1-r^2}$. Noting that $EW(\hat{X}_1 + X_2 + \ldots X_n) = r$ and $EZ(sY + rZ) = r$, we see that $(W, \hat{X}_1 + X_2 + \ldots X_n)$ and $(Z, sY + rZ)$ have the same distribution. Hence

$$
\begin{aligned}
(6.30) \qquad & E\Big||W|^{-1/2} - |\hat{X}_1 + X_2 + \ldots X_n|^{-1/2}\Big| \\
&= E\Big||Z|^{-1/2} - |sY + rZ|^{-1/2}\Big| \\
&= E\Big|\frac{|sY+rZ| - |Z|}{|Z|^{1/2}|sY+rZ|^{1/2}(|Z|^{1/2} + |sY+rZ|^{1/2})}\Big| \\
&\le sE\Big\{\frac{|Y|}{|Z|^{1/2}|sY+rZ|^{1/2}(|Z|^{1/2} + |sY+rZ|^{1/2})}\Big\} \\
&\quad +(1-r)E\Big\{\frac{|Z|}{|Z|^{1/2}|sY+rZ|^{1/2}(|Z|^{1/2} + |sY+rZ|^{1/2})}\Big\} \\
&:= sR_1 + (1-r)R_2.
\end{aligned}
$$

Write

$$
\begin{aligned}
R_1 &= E\{\cdot\}I(|rZ| \le s|Y|/2) + E\{\cdot\}I(s|Y|/2 < |rZ| \le 2s|Y|) + E\{\cdot\}I(|rZ| > 2s|Y|) \\
&:= R_{1,1} + R_{1,2} + R_{1,3}.
\end{aligned}
$$

Let C denote an absolute constant. Then

$$
\begin{aligned}
R_{1,1} &\le 2E\Big\{\frac{|Y|I(|rZ| \le s|Y|/2)}{|Z|^{1/2}|sY|}\Big\} \\
&\le (4/s)E(s|Y|/(2r))^{1/2} \le 4s^{-1/2}, \\
R_{1,2} &\le 4E\Big\{\frac{|Y|I(s|Y|/2 < |rZ| \le 2s|Y|)}{|sY||sY+rZ|^{1/2}}\Big\} \\
&\le Cs^{-1/2}
\end{aligned}
$$

and

$$
\begin{aligned}
R_{1,3} &\le 2E\Big\{\frac{|Y|I(|rZ| > 2s|Y|)}{|Z|^{3/2}}\Big\} \\
&\le CE\Big\{\frac{|Y|}{(s|Y|)^{1/2}}\Big\} \le Cs^{-1/2}.
\end{aligned}
$$

Thus, we have

$$R_1 \le Cs^{-1/2}.$$

As to R_2, we have

$$R_2 \le E(1/|sY+rZ|^{1/2}) = c_0$$

Note that $r \sim 1$ and $s \sim \sqrt{2/n}$ as $n \to \infty$. Combining the above inequalities yields

$$\alpha \le C\varepsilon n^{-1/4}. \tag{6.31}$$

By (6.29) and (6.31), we have

$$E|\Delta| + \sum_{i=1}^{n} E|X_i|^3 + \sqrt{\alpha} \le 8\varepsilon + C\varepsilon^{1/2} n^{-1/8} \le (8+C)\varepsilon$$

provided that $n > (1/\varepsilon)^4$. This proves (4.4).

Finally, we prove (2.12) in Remark 2.5. Let $\theta = \sqrt{(n-1)/n}$, $\rho = (1/n)^{1/2}$, and let Y and Z be independent standard normal random variables. By (6.29),

$$E|\Delta| + \sum_{i=1}^{n} E|X_i|^3 \le 4\varepsilon + 4n^{-1/2} \le 8\varepsilon^{2/3} \tag{6.32}$$

for $n \ge (1/\varepsilon)^{4/3}$. Following the proof of (6.30), we have

$$\begin{aligned}
&\sum_{i=1}^{n} E\Big|X_i(\Delta(X_1,\ldots,X_i,\ldots,X_n) - \Delta(X_1,\ldots,0,\ldots,X_n))\Big| \\
&= n\varepsilon E\Big||X_1|(|X_1+X_2+\ldots X_n|^{-1/2} - |X_2+\ldots X_n|^{-1/2})\Big| \\
&= n\varepsilon E\Big||\rho Y|(|\rho Y+\theta Z|^{-1/2} - |\theta Z|^{-1/2})\Big| \\
&\le n\varepsilon\rho^2 E\Big\{\frac{Y^2}{|\rho Y+\theta Z|^{1/2}|\theta Z|^{1/2}(|\rho Y+\theta Z|^{1/2}+|\theta Z|^{1/2})}\Big\} \\
&\le \varepsilon E\Big\{\frac{Y^2 I(|\theta Z| \le \rho|Y|/2)}{|\rho Y+\theta Z||\theta Z|^{1/2}}\Big\} \\
&\quad + \varepsilon E\Big\{\frac{Y^2 I(\rho|Y|/2 < |\theta Z| \le 2\rho|Y|)}{|\rho Y+\theta Z|^{1/2}|\theta Z|}\Big\} \\
&\quad + \varepsilon E\Big\{\frac{Y^2 I(|\theta Z| > 2\rho|Y|)}{|\rho Y+\theta Z|^{1/2}|\theta Z|}\Big\} \\
&\le 2\varepsilon\rho^{-1} E\Big\{|Y|\ |\theta Z|^{-1/2} I(|\theta Z| \le \rho|Y|/2)\Big\} \\
&\quad + 2\varepsilon\rho^{-1} E\Big\{\frac{|Y| I(\rho|Y|/2 < |\theta Z| \le 2\rho|Y|)}{|\alpha Y+\theta Z|^{1/2}}\Big\} \\
&\quad + 2\varepsilon E\Big\{\frac{Y^2 I(|\theta Z| > 2\rho|Y|)}{|\theta Z|^{3/2}}\Big\} \\
&\le C\varepsilon\rho^{-1/2} = C\varepsilon n^{1/4} \le 2C\varepsilon^{2/3}
\end{aligned} \tag{6.33}$$

for an absolute constant C, provided that $n \le 16(1/\varepsilon)^{4/3}$. This proves (2.12), by (6.32) and (6.33).

∎

Acknowledgments. The authors are thankful to Xuming He for his contribution to the construction of the example in Section 4.

References

[1] Bentickus, V., Götze, F. and Zitikis, M. (1994). Lower estimates of the convergence rate for U-statistics. *Ann. Probab.* **22**, 1701-1714.

[2] Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434-485.

[3] Bickel,P. (1974). Edgeworth expansion in nonparametric statistics. *Ann., Statist.* **2**, 1-20.

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

[5] Borovskich, Yu.V. (1983). Asymptotics of U-statistics and von Mises' functionals. *Sov. Math. Dokl.* **27**, 303-308.

[6] Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Ann., Statist.* **6**, 417-421.

[7] Chan, Y.K. and Wierman, J. (1977). On the Berry-Esseen theorem for U-statistics. *Ann. Probab.* **5**, 136-139.

[8] 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.

[9] Figiel, T., Hitczenko, P., Johnson, W.B., Schechtman, G., and Zinn, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349**, 997-1027.

[10] Filippova, A.A. (1962). Mises's theorem of the asympototic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7**, 24-57.

[11] FINELSTEIN, M., KRUGLOV, V.M., and TUCKER, H.G. (1994). Convergence in law of random sums with non-random centering. *J. Theoret. Probab.* **3**, 565-598.

[12] FRIEDRICH, K. O. (1989). A Berry-Esseen bound for functions of independent random variables. *Ann. Statist.* **17**, 170–183.

[13] GRAMS, W.F. and SERFLING, R.J. (1973). Convergence rates for U-statistics and related statistics. *Ann., Statist.* **1**, 153-160.

[14] HELMERS, R. (1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions. *Ann. Probab.* **5**, 940-953.

[15] HELMERS, R. and JANSSEN, P. (1982). On the Berry-Esseen theorem for multivariate U-statistics. *Math. Cent. Rep. SW* 90/82, Mathematisch Centrum, Amsterdam, pp. 1-22.

[16] HELMERS, R., JANSSEN, P. and SERFLING, R. J. (1990). Berry-Esseen bound and bootstrap results for genalized L-statistics. *Scand. J. Statist.* —bf 17, 65-77.

[17] HOEFFDING, W. (1948). A class of statististics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

[18] KOROLJUK, V.S. and BOROVSKICH, YU. V. (1994). *Theory of U-statistics.* Kluwer Academic Publishers, Boston.

[19] ROBBINS, H. (1948). The asymptotic distribution of the sum of a random numbers of random random variables. *Bull. A.M.S.* **54**, 1151-1161.

[20] ROSENTHAL, H.P. (1970). On the subspaces of $L_p(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8**, 273-303.

[21] SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statisitics.* Wiley, New York.

[22] SHORACK, G.R. (2000). *Probability for Statisticians.* Springer, New York.

[23] WANG, Q. (2001). Non-uniform Berry-Esséen Bound for U-Statistics. *Statist. Sinica* (to appear)

[24] WANG, Q., JING, B.Y. and ZHAO, L. (2000). The Berry-Esseen bound for studentized statistics. *Ann. Probab.* **28**, 511–535.

[25] Zhao, L.C. and Chen, X.R. (1983). Non-uniform convergence rates for distributions of U-statistics. *Sci. Sinica* (Ser. A) **26**, 795-810.

[26] van Zwet, W.R. (1984). A Berry-Esseen bound for symmetric statisics. *Z. Wahrsch. Verw. Gebiete* **66**, 425-440.

Louis H.Y. Chen
Institute for Mathematical Sciences
National University of Singapore
Singapore 118402
Republic of Singapore
lhychen@ims.nus.edu.sg
and
Department of Mathematics
Department of Statistics & Applied Probability
National University of Singapore
Singapore 117543
Republic of Singapore

Qi-Man Shao
Department of Mathematics
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
maqmshao@ust.hk

Department of Mathematics
University of Oregon
Eugene, OR 97403
USA
and
Department of Mathematics
Zhejiang University
Hangzhou, Zhejiang 310027
China