Bounds for the concentration of asymptotically normal statistics

I.B. Alberink[1], V. Bentkus[2,3]

August 8, 1999

Abstract

Let $X_1, \dots, X_n$ be independent, not necessarily identically distributed random variables. An optimal bound is derived for the concentration function of an arbitrary real-valued statistic $\mathbb{T} = \mathbb{T}(X_1, \dots, X_n)$ for which $\mathbb{E}\mathbb{T}^2 < \infty$. Applications are given for Wilcoxon's rank-sum statistic, U-statistics, Student's statistic, the two-sample Student statistic and linear regression.

Key words and phrases: Concentration function, non-i.i.d. sampling, non-symmetric and symmetric statistics, U-statistics, Wilcoxon's rank-sum test, (two-sample) Student's statistic, linear regression.
A.M.S. classification (1991): 62E20.

1 Introduction and results

Let $X_1, \dots, X_n$ be independent, not necessarily identically distributed random variables, taking their values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. We consider arbitrary statistics, that is,

$$\mathbb{T} \;=\; \mathbb{T}(n) \;=\; \mathbb{T}(X_1, \dots, X_n),$$

where $\mathbb{T}$ is a real-valued, measurable function on the product space $\mathcal{X}^n$, assuming henceforth that $\mathbb{E}\,|\mathbb{T}| < \infty$.

A sequence $\mathbb{T} = \mathbb{T}(n)$ of statistics is called *asymptotically normal* if there exist sequences $a_n, b_n \in \mathbb{R}$ such that the sequence $(\mathbb{T} - a_n)/b_n$ converges in distribution to the standard normal distribution. The class of these statistics plays an essential role in random data analysis, although other statistics are of interest as well. Here and throughout the paper we shall write $\mathbb{T}$ instead of $\mathbb{T}(n)$, and in general use notation that does not reflect dependence on n.

[1]University of Nijmegen.
[2]University of Bielefeld. Research supported by the SFB 343 in Bielefeld.
[3]On leave from the Vilnius Institute of Mathematics and Informatics.

It is our goal to develop a sufficiently general and exhausting mathematical method to analyze asymptotically normal statistics, without making further assumptions on their structure. In fact, we are mainly interested in obtaining estimates of the accuracy of normal approximations and bounds for the remainder terms in asymptotic (Edgeworth) expansions of $\mathbb{T}$'s distribution function. Under various assumptions on the structure of $\mathbb{T}$, a large number of results of this kind is already available. Since this number is increasing steadily, it seems worthwhile to have a unifying approach.

As an example of a sufficiently rich class of statistics, we might look at the class of symmetric statistics of independent, identically distributed (i.i.d.) samples. By symmetric we mean that the statistics are invariant under permutations of their arguments. This class was studied, e.g., by Dynkin and Mandelbaum (1983), Mandelbaum and Taqqu (1984), Giraitis and Surgailis (1984) and Mandrekar and Meerschaert (1994), and normal approximations were considered by Van Zwet (1984), Friedrich (1989), Lai and Wang (1993), Guan (1994), Bentkus, Götze and Van Zwet (1997), Putter and Van Zwet (1998) and Wang (1998). Being nice and convenient from a mathematical point of view, the two assumptions are however rather restrictive in applications. In the case of linear rank statistics for example, the assumption on symmetry is not applicable, and in the case of statistics depending on two (or more) samples, the i.i.d. assumption is usually not satisfied. To our knowledge, the only bounds that were obtained without using the i.i.d. and symmetry assumptions were achieved by Friedrich (1989). However, in essence, Friedrich's bounds do not differ too much from the ones obtained in the i.i.d. and symmetric case, since they are stated in terms of quantities like $\max_{1\le j\le n} \mathbb{E} f_j(X_j)$, where the f_j are functions related to $\mathbb{T}$ and the sampling procedure. A more natural way of dropping the two assumptions would for example involve that the bounds for the accuracy of approximations depend on separate observations in an additive way, for example, through quantities of the form $\sum_{j=1}^{n} \mathbb{E} f_j(X_j)$, the f_j being again related to $\mathbb{T}$ and the sampling procedure. A first step in this direction was made by Bentkus, Bloznelis, Götze (1996), who obtained additive bounds in the case of Student's statistic and selfnormalized sums. The methods to obtain such results involve rather refined techniques, based mainly on non-linear Fourier analysis. Sometimes direct methods to analyze the distributions may be applied as well.

In this paper we confine ourselves to finding bounds for the concentration of the distribution of $\mathbb{T}$. The bounds that we obtain will not involve requirements on the structure of $\mathbb{T}$ or the i.i.d. assumption. They have a rather simple additive structure and are precise, up to absolute constants, in the sense that applied to special classes of statistics they provide optimal results. The technique we develop is useful in a more refined analysis of the distribution of $\mathbb{T}$, for example when we are looking for estimates of convergence rates to normality or asymptotic expansions (to appear elsewhere). For the application of modern smoothing techniques (see Prawitz (1972), Bentkus (1994), Aleshkyavichene and Statulevichius (1997), Bentkus, Götze and Van Zwet (1997)), it is as well necessary to have estimates of this form.

The paper is organized as follows. First we shall formulate the results. In Section 2 we provide their proofs. In particular, we describe Hoeffding's decomposition for $\mathbb{T}$, which is extensively used in the proofs. In Section 3 we consider applications to some special classes of statistics. In particular, we provide some simple hints and typical calculations used in order to apply our central results to concrete statistics.

We turn to the formulation of the results. Let $\sigma^2 := \operatorname{var}\mathbb{T}$ (which may be ∞). As will be explained in Section 2.1, we decompose $\mathbb{T}$ as

$$\mathbb{T} \;=\; \textstyle\sum_{A\subset\{1,\dots,n\}} T_A \;=\; \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2 + \mathbf{T}_3, \tag{1}$$

with

$$\mathbb{T}_1 \;:=\; \textstyle\sum_{j=1}^n T_j, \quad \mathbb{T}_2 \;:=\; \textstyle\sum_{1\le i<j\le n} T_{i,j}, \quad \mathbf{T}_3 \;:=\; \textstyle\sum_{A:\,|A|\ge 3} T_A,$$

the so-called Hoeffding decomposition of $\mathbb{T}$. In cases where the sample is i.i.d. and $\mathbb{T}$ is symmetric in its arguments, we set

$$s^2 := \operatorname{var}\mathbb{T}_1, \quad \tilde\beta_3 := \mathbb{E}\,|n^{1/2}\,T_1|^3 \quad \text{and} \quad \tilde\Delta_2^2 := n^3 \sum_{A:\,1,2\in A} \mathbb{E}\,T_A^2. \tag{2}$$

The *concentration function* of our statistic $\mathbb{T}$, defined by

$$Q(\mathbb{T},\lambda) \;:=\; \sup_{x\in\mathbb{R}} \mathbb{P}(x\le\mathbb{T}\le x+\lambda), \quad \text{all } \lambda\ge 0,$$

gives a measure for the concentration of probability mass on (small) intervals. Specialized to the case of symmetric $\mathbb{T}$ and i.i.d. samples, our result reads as follows:

Theorem 1.1. *Suppose that $0<s^2<\infty$, the sample $X_1,\dots,X_n$ is i.i.d. and $\mathbb{T}$ is symmetric. Then there exists an absolute constant c such that, for all $\lambda\ge 0$,*

$$Q(\mathbb{T}/s,\lambda) \;\le\; c\,\max\{\lambda,\ \tilde\beta_3\,s^{-3}\,n^{-1/2},\ \tilde\Delta_2^2\,s^{-2}\,n^{-1}\}. \tag{3}$$

The bound holds as well with s replaced by σ if in addition $\sigma^2<\infty$.

The constant c in (3) satisfies $c\le 850$. The bound (3) shows that the concentration function of $\mathbb{T}$ depends mainly on its linear part $\mathbb{T}_1$. If $\tilde\beta_3\,s^{-3}$ and $\tilde\Delta_2^2\,s^{-2}$ are bounded from above as $n\to\infty$, the contribution of $\mathbb{T}_1$ is of order $\mathcal{O}(n^{-1/2})$, and the contribution of all higher order terms together is $\mathcal{O}(n^{-1})$.

For our main result, we look at the general case in which the sample $X_1,\dots,X_n$ is independent but not necessarily identically distributed, and $\mathbb{T}$ is an arbitrary function, that is, not necessarily symmetric in its arguments. Let

$$s^2 := \operatorname{var}\mathbb{T}_1, \quad \beta := \sum_{j=1}^n \mathbb{E}\,|T_j|^3, \quad \Delta^2 := \sum_{1\le j<k\le n}\ \sum_{A:\,j,k\in A} \mathbb{E}\,T_A^2. \tag{4}$$

We have that $\operatorname{var}(\mathbb{T}_2+\mathbf{T}_3)\le\Delta^2$: for an upper bound for Δ^2, see (18). The result reads as follows:

Theorem 1.2. *Suppose that $0<s^2<\infty$, the observations $X_1,\dots,X_n$ are independent and $\mathbb{T}$ is arbitrary. Then there exists an absolute constant c such that, for all $\lambda\ge 0$,*

$$Q(\mathbb{T}/s,\lambda) \;\le\; c\,\max\big\{\lambda,\ \beta/s^3,\ \Delta^2/s^2\big\}\,.$$

The bound holds as well with s replaced by σ if in addition $\sigma^2 < \infty$.

Note that Theorem 1.2 is a direct generalization of Theorem 1.1: in the case of an i.i.d. sample and a statistic that is symmetric in its arguments, we have that (see (15))

$$\beta = \tilde{\beta}_3 \, n^{-1/2} \quad \text{and} \quad \Delta^2 \le \tfrac{1}{2} \, \tilde{\Delta}_2^2 \, n^{-1}.$$

The proofs of Theorem 1.1 and Theorem 1.2 will be given in Section 2. We start by giving the proof of Theorem 1.1. For reasons of transparancy, in Section 2.2 we first consider the special case of U-statistics of order 2, that is, the case where in the Hoeffding decomposition (1) the part $\mathbf{T}_3 = 0$. In Section 2.3 we then extend this proof to the case of arbitrary statistics. Using the same scheme, the proof of Theorem 1.2 will be given in the Sections 2.4 and 2.5.

In Section 3 we provide applications of Theorem 1.1 and 1.2 to particular statistics, each time obtaining optimal results. Self-normalized statistics are of special interest here. In these applications the exact knowledge of Hoeffding decompositions is often not so much important, since instead of them one can use so-called differences. The *difference* $D_j\mathbb{T}$ of the statistic $\mathbb{T}$ with respect to the observation X_j is defined as $D_j\mathbb{T} := \mathbb{T} - \mathbb{E}_j\mathbb{T}$, with $\mathbb{E}_j\mathbb{T}$ denoting the conditional expectation of $\mathbb{T}$ given $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n$, and higher order differences are defined inductively. In particular,

$$D_j D_k \mathbb{T} \;=\; \mathbb{T} - \mathbb{E}_j\mathbb{T} - \mathbb{E}_k\mathbb{T} + \mathbb{E}\,\mathbb{T},$$

for $j \ne k$, and we have that (see Section 2.1 for the proofs)

$$\beta \;\le\; \textstyle\sum_{j=1}^n \mathbb{E}\,|D_j\mathbb{T}|^3, \quad \Delta^2 \;=\; \textstyle\sum_{1\le j<k\le n} \mathbb{E}\,|D_j D_k \mathbb{T}|^2,$$

In the symmetric i.i.d. case this reads as

$$\tilde{\beta}_3 \;\le\; n^{3/2}\,\mathbb{E}\,|D_1\mathbb{T}|^3 \quad \text{and} \quad \tilde{\Delta}_2^2 \;=\; n^3\,\mathbb{E}\,|D_1 D_2 \mathbb{T}|^2.$$

Finally we gather some general notation. From here on, let

$$N \;:=\; \{1, \dots, n\}.$$

The cardinality of any subset $A \subset N$ will be denoted by $|A|$, and its complement by A^c. By $\mathbb{E}\,(\mathbb{T} \mid A)$ we will mean the conditional expectation of $\mathbb{T}$ given all X_j, $j \in A$, and we let

$$\mathbb{E}_A\mathbb{T} \;:=\; \mathbb{E}\,(\mathbb{T} \mid A^c).$$

For any real number x, by $[x]$ we mean its integer part. For any bounded function f, by

$$\|f\|_\infty \;:=\; \sup_{x\in\mathbb{R}} |f(x)|$$

we denote its supremum norm. Finally, for all $y \in \mathbb{R}$, we shall write

$$e\{y\} \;:=\; \exp\{iy\},$$

with $i = \sqrt{-1}$ denoting the complex root.

2 Proofs of the Theorems

2.1 The Hoeffding decomposition

Let $X_1, \dots, X_n : \Omega \to \mathcal{X}$ be independent random variables, taking their values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. By $\sigma(A)$ we denote the σ-algebra generated by the X_j with $j \in A$, that is to say, the smallest σ-algebra on Ω such that all X_j, $j \in A$ are measurable. We look at an arbitrary statistic

$$\mathbb{T} \;=\; \mathbb{T}(X_1, \dots, X_n),$$

where $\mathbb{T} : \mathcal{X}^n \to \mathbb{R}$ is a measurable function, such that $\mathbb{E}\,|\mathbb{T}| < \infty$.

We describe a canonical decomposition for $\mathbb{T}$, due to Hoeffding (1948). Here we follow the lines of Van Zwet (1984) and Bentkus, Götze, Van Zwet (1997), Section 4, introducing as well the complementary concept of differences. We establish some basic facts and moment inequalities which will be of use later (see as well Lemmas 2.5 and 2.8).

For any subset $A = \{j_1, \dots, j_k\} \subset N$, we have the conditional expectation

$$\mathbb{E}\,(\mathbb{T}\,|\,A) \;=\; \mathbb{E}\,(\mathbb{T}\,|\,j_1, \dots, j_k) \;:=\; \mathbb{E}\,(\mathbb{T}\,|\,X_{j_1}, \dots, X_{j_k}),$$

being the almost surely unique $\sigma(A)$-measurable random variable S for which $\mathbb{E}\,|S| < \infty$ and $\mathbb{E}\,I\{G\}\,\mathbb{T} = \mathbb{E}\,I\{G\}\,S$ for all $G \in \sigma(A)$. By $I\{G\}$ we mean the indicator of the event G. Note that $\mathbb{E}\,(\mathbb{T}\,|\,\emptyset) = \mathbb{E}\,\mathbb{T}$ and that $\mathbb{E}\,(\mathbb{T}\,|\,N) = \mathbb{T}$. In case $\mathbb{E}\,\mathbb{T}^2 < \infty$, the mapping $\mathbb{T} \mapsto \mathbb{E}\,(\mathbb{T}\,|\,A)$ is the orthogonal projection on the Hilbert space $L^2(A)$ of equivalence classes of almost surely equal, square integrable functions $\varphi(X_j \,:\, j \in A)$ (using the standard inner product $\langle S_1, S_2 \rangle := \mathbb{E}\,S_1 S_2$).

For any $A \subset N$, we set

$$T_A \;=\; T_A(X_j \,:\, j \in A) \;:=\; \sum_{B \subset A} (-1)^{|A|-|B|}\, \mathbb{E}\,(\mathbb{T}\,|\,B). \tag{5}$$

As such, we interpret T_A as a function of the X_j, $j \in A$, depending only on the distribution of the sample. For any set $A = \{j_1, \dots, j_l\}$, we shall denote T_A as well by $T_{j_1,\dots,j_l}$. In this way we have that $T_\emptyset = \mathbb{E}\,\mathbb{T}$, and, for any $1 \le j \le n$ and $1 \le k < l \le n$,

$$T_j = \mathbb{E}\,(\mathbb{T}\,|\,j) - \mathbb{E}\,\mathbb{T}, \quad T_{k,l} = \mathbb{E}\,(\mathbb{T}\,|\,k,l) - \mathbb{E}\,(\mathbb{T}\,|\,k) - \mathbb{E}\,(\mathbb{T}\,|\,l) + \mathbb{E}\,\mathbb{T},$$

and so on. It is easily seen that for any two statistics R and S with existing first moments and $A \subset N$,

$$(R+S)_A \;=\; R_A + S_A. \tag{6}$$

We have the following inverse statement to (5):

Lemma 2.1. *For all $B \subset N$,*

$$\mathbb{E}\,(\mathbb{T}\,|\,B) \;=\; \textstyle\sum_{C \subset B} T_C.$$

Proof of Lemma 2.1. It is clear from (5) that $\sum_{C\subset B} T_C$ is a linear combination of $\mathbb{E}(\mathbb{T}\,|\,A)$ with $A \subset B$. Taking any such $A \subset B$, we see that it is encountered as a summand in the definition of T_C precisely if $A \subset C$, and then with the factor $(-1)^{|C|-|A|}$. In this way, using Newton's binomial theorem, we see that the term $\mathbb{E}(\mathbb{T}\,|\,A)$ comes up

$$\textstyle\sum_{j=0}^{|B|-|A|} \binom{|B|-|A|}{j} (-1)^j \;=\; (-1+1)^{|B|-|A|} \;=\; I\{A = B\}$$

times (there being $\binom{|B|-|A|}{j}$ different choices of C for which $A \subset C \subset B$ and $|C| - |A| = j$). Thus we see that indeed $\sum_{C\subset B} T_C = \mathbb{E}(\mathbb{T}\,|\,B)$. $\square$

In particular we have that

$$\mathbb{T} \;=\; \textstyle\sum_{C\subset N} T_C, \tag{7}$$

which gives the so-called *Hoeffding decomposition* of $\mathbb{T}$: cf. (1). Note that T_A is in general a random variable depending on the X_j with $j \in A$. The crucial feature of the Hoeffding decomposition is that it is *orthogonal* in $L^2(N)$, that is to say, in case $\mathbb{E}\,\mathbb{T}^2 < \infty$ we have that $\mathbb{E}\,T_A T_B = 0$ for all $A \neq B$.

Complementary to the Hoeffding decomposition is the concept of so-called *differences*. To this, for any statistic $S = S(X_1, \dots, X_n)$ with $\mathbb{E}\,|S| < \infty$ and any $j \in N$ we define

$$D_j S \;:=\; S - \mathbb{E}(S\,|\,\{X_k \,:\, k \neq j\}) \;=\; S - \mathbb{E}_j S.$$

In this way always $D_j\, D_j S = D_j S$ and $D_l\, D_m S = D_m\, D_l S$, so that one may also speak of

$$D_A S \;:=\; D_{j_1} \cdots D_{j_k} S$$

for any non-empty set $A = \{j_1, \dots, j_k\}$. Note that D_j is linear in the sense that always

$$D_j\, \lambda\, S = \lambda\, D_j S \quad \text{and} \quad D_j\,(S_1 + S_2) = D_j S_1 + D_j S_2, \tag{8}$$

and that for any statistic $S_1 = S_1(X_k \,:\, k \neq j)$ which does not depend on X_j,

$$D_j S_1 S_2 \;=\; S_1\, D_j S_2. \tag{9}$$

Returning to $\mathbb{T}$, we have the following two equalities:

Lemma 2.2. *For any non-empty set $A \subset N$,*

$$D_A \mathbb{T} \;=\; \textstyle\sum_{B\subset A} (-1)^{|B|}\, \mathbb{E}(\mathbb{T}\,|\,B^c) \quad \textit{and} \quad T_A \;=\; \mathbb{E}(D_A \mathbb{T}\,|\,A). \tag{10}$$

Proof of Lemma 2.2. A proof of the first equality is easily found using induction on the number of elements of A. The statement is obvious in the case that $|A| = 1$. Now suppose the statement is correct for all A with $|A| = k$, for some $1 \le k \le n-1$. Looking at any set $A \cup \{m\}$ of size $k+1$, with $m \notin A$, using the inductional hypothesis we see that

$$D_{A\cup\{m\}}\mathbb{T} \;=\; D_m\, D_A\mathbb{T} \;=\; \textstyle\sum_{B\subset A}(-1)^{|B|}\, D_m\mathbb{E}\,(\mathbb{T}\,|\,B^c).$$

Here $D_m\mathbb{E}\,(\mathbb{T}\,|\,B^c) = \mathbb{E}\,(\mathbb{T}\,|\,B^c) - \mathbb{E}\,(\mathbb{T}\,|\,(B\cup\{m\})^c)$, so that

$$\begin{aligned} D_{A\cup\{m\}}\mathbb{T} \;&=\; \textstyle\sum_{C\subset A\cup\{m\},\, m\notin C}(-1)^{|C|}\,\mathbb{E}\,(\mathbb{T}\,|\,C^c) + \sum_{C\subset A\cup\{m\},\, m\in C}(-1)^{|C|}\,\mathbb{E}\,(\mathbb{T}\,|\,C^c) \\ &=\; \textstyle\sum_{C\subset A\cup\{m\}}(-1)^{|C|}\,\mathbb{E}\,(\mathbb{T}\,|\,C^c), \end{aligned}$$

which proves the point.

The second equality is a consequence of the first, as for any $B \subset A$ we have that $\mathbb{E}\,(\mathbb{E}\,(\mathbb{T}\,|\,B^c)\,|\,A) = \mathbb{E}\,(\mathbb{T}\,|\,A\cap B^c)$, so that, using the substitution $C := A \cap B^c$ and (5),

$$\mathbb{E}\,(D_A\mathbb{T}\,|\,A) \;=\; \textstyle\sum_{B\subset A}(-1)^{|B|}\,\mathbb{E}\,(\mathbb{T}\,|\,A\cap B^c) \;=\; \sum_{C\subset A}(-1)^{|A|-|C|}\,\mathbb{E}\,(\mathbb{T}\,|\,C) \;=\; T_A.$$

This finishes the proof. □

The Hoeffding decomposition (7) and the moments $\mathbb{E}\,|T_A|^l$ of its parts are natural tools for the mathematical analysis of the statistic $\mathbb{T}$. However, in applications, the differences $D_A\mathbb{T}$ and their moments $\mathbb{E}\,|D_A\mathbb{T}|^l$ are often more convenient to calculate or estimate. To this: using (10) and Jensen's inequality (conditional version), for any non-empty set $A \subset N$ and $l \ge 1$ we have:

$$\mathbb{E}\,|T_A|^l \;=\; \mathbb{E}\,|\mathbb{E}\,(D_A\mathbb{T}\,|\,A)|^l \;\le\; \mathbb{E}\mathbb{E}\,(|D_A\mathbb{T}|^l\,|\,A) \;=\; \mathbb{E}\,|D_A\mathbb{T}|^l. \tag{11}$$

In the non-identically distributed case we will make use of the following moments corresponding to $\mathbb{T}$:

$$\sigma^2 := \operatorname{var}\mathbb{T}, \quad s_j^2 := \mathbb{E}\,T_j^2, \quad s^2 := \operatorname{var}\mathbb{T}_1 = \textstyle\sum_{j=1}^n s_j^2,$$

$$\beta_j := \mathbb{E}\,|T_j|^3, \quad \beta := \textstyle\sum_{j=1}^n \beta_j, \quad \Delta^2 \;:=\; \sum_{1\le j<k\le n}\mathbb{E}\,|D_j\,D_k\mathbb{T}|^2. \tag{12}$$

In case of an i.i.d. sample and a statistic that is symmetric in its arguments, we will as well use

$$\tilde\beta_3 := \mathbb{E}\,|n^{1/2}\,T_1|^3, \quad \tilde\gamma_p := \mathbb{E}\,|n^{3/2}\,T_{1,2}|^p, \quad \tilde\Delta_m^p := \mathbb{E}\,|n^{m-1/2}\,D_1\cdots D_m\mathbb{T}|^p, \tag{13}$$

with $m \in N$ and $p \ge 0$. As to Δ_2^2 and $\tilde\Delta_m^p$, in general

$$D_{j_1}\cdots D_{j_m}\mathbb{T} = \textstyle\sum_{A:\,\{n_1,\dots,n_m\}\subset A} T_A. \tag{14}$$

In the i.i.d. symmetric case, with (14) we have that

$$\Delta^2 \;=\; \tbinom{n}{2}\,\textstyle\sum_{\{1,2\}\subset A}\mathbb{E}\,T_A^2 \;\le\; \frac{1}{2}\,\tilde\Delta_2^2\,n^{-1}, \tag{15}$$

and we will denote as well

$$\Delta^2 \;=\; \binom{n}{2}\,\mathbb{E}\,T_{1,2}^2 + \binom{n}{2}\,\sum_{\{1,2\}\subset A,\,|A|\geq 3}\mathbb{E}\,T_A^2 \;=:\; \Delta_1^2 + \Delta_2^2. \tag{16}$$

Note that $\Delta_1^2 = \operatorname{var}\mathbb{T}_2$. Obviously (because of orthogonality) we have in general that

$$\sigma^2 \;=\; s^2 + \operatorname{var}\mathbb{T}_2 + \operatorname{var}\mathbf{T}_3,$$

so that for example $s^2 \leq \sigma^2$, while, with (14),

$$\begin{aligned} \operatorname{var}\mathbb{T}_2 + \operatorname{var}\mathbf{T}_3 \;&=\; \mathbb{E}\,\mathbb{T}_2^2 + \mathbb{E}\,\mathbf{T}_3^2 \;=\; \sum_{A:|A|\geq 2}\mathbb{E}\,T_A^2 \\ &\leq\; \sum_{1\leq j<k\leq n}\,\sum_{A:\,j,k\in A}\mathbb{E}\,T_A^2 \;=\; \Delta^2. \end{aligned} \tag{17}$$

With (14) we see that the definitions of Δ^2 and $\tilde{\Delta}_2^2$ are compatible with the ones presented in (4) and (2).

Looking at the definition of Δ^2 in (4), we notice that for any fixed subset C of N with $|C| = l$, $l \geq 2$, the term $\mathbb{E}\,T_C^2$ is encountered $\binom{l}{2}$ times (as this is the number of possible choices for the pair (j,k)). As a result, for all $1 \leq l \leq n$ taking

$$\mathbb{T}_l \;:=\; \sum_{A:\,|A|=l} T_A,$$

we have that

$$\Delta^2 \;=\; \sum_{l=2}^n \binom{l}{2}\,\sum_{C:\,|C|=l}\mathbb{E}\,T_C^2 \;=\; \sum_{l=2}^n \binom{l}{2}\operatorname{var}\mathbb{T}_l. \tag{18}$$

2.2 Proof of Theorem 1.1 for U-statistics of order 2

For all our bounds we will be using the following inequality, which holds for arbitrary $\mathbb{T}$ and may be found in Petrov (1995) as Lemma 1.16: for any $a > 0$,

$$Q(\mathbb{T},\lambda) \;\leq\; 2\cdot\left(\tfrac{96}{95}\right)^2 \max\{\lambda, a^{-1}\}\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt. \tag{19}$$

As promised, first we will prove Theorem 1.1 in the case where $\mathbb{T}$ is a so-called U-statistic of order 2, that is, $\mathbb{T} = \mathbb{E}\,\mathbb{T} + \mathbb{T}_1 + \mathbb{T}_2$ (cf. (1)). Taking $\tilde{\gamma}_2$ as in (13), in this case we have that $\tilde{\Delta}_2^2 = \tilde{\gamma}_2$.

Without loss of generality we may assume that

$$\mathbb{E}\,\mathbb{T} = 0 \quad\text{and}\quad s^2 = 1.$$

Here the expectation value of $\mathbb{T}$ may be taken zero since $Q(\mathbb{T},\lambda) = Q(\mathbb{T}-\mathbb{E}\,\mathbb{T},\lambda)$. In case $s^2 \neq 1$ we may just apply the reduced theorem to $\mathbb{T}/s$. We prove Theorem 1.1 by using (19), taking

$$a := \min\{\tilde{\beta}_3^{-1}\,n^{1/2}, \tilde{\gamma}_2^{-1}\,n\}, \quad a^{-1} = \max\{\tilde{\beta}_3\,n^{-1/2}, \tilde{\gamma}_2\,n^{-1}\}.$$

We may assume that $a > 404$. In case $a \le 404$ it is clear that

$$(\tfrac{96}{95})^2 \int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; 404\, (\tfrac{96}{95})^2 \;\le\; 413,$$

so that the statement of the Theorem clearly follows from (19). Note that via Lyapunov's inequality (see Petrov (1995), p. 7)

$$\tilde{\beta}_3 \ge (s^2)^{3/2} = 1 \quad \text{and} \quad a \le \tilde{\beta}_3^{-1}\, n^{1/2} \le n^{1/2}. \tag{20}$$

To complete the proof it will now suffice (see (19)) to show that

$$\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; 404. \tag{21}$$

From now on, let $c := 404$. The method to obtain (21) is as follows: we divide the interval $[0, a]$ into two parts, $[0, a^{1/4}]$ and $[a^{1/4}, a]$, and use different methods in order to show that the corresponding two integrals both are bounded by a constant. In order to do this we must subdivide the interval $[a^{1/4}, a]$ again into a few intervals, on all of which we may apply the same method. For $p < q$ such that $\frac{1}{4} < q \le 1$ and $p > q - \frac{1}{4}$, we set

$$\varepsilon_1 \;:=\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt, \qquad \varepsilon_2 \;=\; \varepsilon_2(p,q) \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt. \tag{22}$$

We estimate ε_1 and ε_2 from above in the Lemmas 2.3 and 2.4.

Lemma 2.3. *Writing $c_1^2 := 1 - \frac{1}{3}c^{-3/4}$, we have that*

$$\varepsilon_1 \;=\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\le\; 2^{-3/2} + (\tfrac{1}{2}\pi)^{1/2}\, c_1^{-1}.$$

Proof of Lemma 2.3. In order to estimate ε_1 we just remove the non-linear part $\mathbb{T}_2$ from $\mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2$ by a Taylor expansion. As to the latter, we will use the following scheme. For any real or complex-valued function f on $\mathbb{R}$, possessing a continuous $(k+1)^{\text{th}}$ derivative on some open interval containing both 0 and x, we have that

$$f(x) \;=\; \sum_{l=0}^{k} \tfrac{1}{l!}\, f^{(l)}(0)\, x^l + r_{k+1}(x) \tag{23}$$

where

$$r_{k+1}(x) \;:=\; \tfrac{1}{k!} \int_0^1 (1-s)^k\, f^{(k+1)}(sx)\, x^{k+1}\, ds.$$

Introducing a random variable τ which is uniformly distributed on $(0,1)$, we may also write

$$r_{k+1}(x) \;=\; \tfrac{1}{k!}\, \mathbb{E}\, (1-\tau)^k\, f^{(k+1)}(\tau x)\, x^{k+1}.$$

Using this we notice that

$$\varepsilon_1 = \int_0^{a^{1/4}} |\mathbb{E}\, e\{t(\mathbb{T}_1 + \mathbb{T}_2)\}|\, dt \;\le\; \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}_1\}|\, dt + \int_0^{a^{1/4}} t\, \mathbb{E}\, |\mathbb{T}_2|\, dt =: \varepsilon_{11} + \varepsilon_{12},$$

where, with (17) and (15), $\mathbb{E}\,\mathbb{T}_2^2 \le \frac{1}{2}\tilde\gamma_2\, n^{-1} \le \frac{1}{2}a^{-1}$, so that

$$\varepsilon_{12} \;\le\; \tfrac{1}{2}a^{1/2}\,\{\mathbb{E}\,\mathbb{T}_2^2\}^{1/2} \;\le\; 2^{-3/2}.$$

With regard to ε_{11} we need the following inequality:

$$|\mathbb{E}\,e\{tT_1\}| \;\le\; 1 - \tfrac{1}{2}t^2 n^{-1}\, c_1^2 \;\le\; \exp\{-\tfrac{1}{2}t^2 n^{-1}\, c_1^2\}. \tag{24}$$

Indeed, because $e\{tT_1\} = 1 + (it)\,T_1 + \frac{1}{2}(it)^2\,T_1^2 + R_1$ with $|R_1| \le \frac{1}{6}t^3\,|T_1|^3$ (again using (23)), we have that

$$|\mathbb{E}\,e\{tT_1\} - (1 - \tfrac{1}{2}t^2 n^{-1})| \le \tfrac{1}{6}t^3\,\mathbb{E}|T_1|^3 \le \tfrac{1}{6}t^2\,a^{1/4}\,n^{-1}\,(\tilde\beta_3\,n^{-1/2}) \le \tfrac{1}{6}t^2\,a^{-3/4}\,n^{-1},$$

while $1 - \frac{1}{2}t^2 n^{-1} \ge 0$ for $0 \le t \le a^{1/4}$ and $a \ge c$, and (24) is clearly correct. As a result of it we see that

$$\varepsilon_{11} = \int_0^{a^{1/4}} |\mathbb{E}\,e\{tT_1\}|^n\,dt \le \int_0^\infty \exp\{-\tfrac{1}{2}c_1^2\,t^2\}\,dt = (\tfrac{1}{2}\pi)^{1/2}\,c_1^{-1},$$

which completes the proof. □

We turn to the estimation of ε_2.

Lemma 2.4. *Let $\frac{1}{4} < q \le 1$ and $q - \frac{1}{4} < p < q$. Write*

$$r \;:=\; \frac{q - \frac{1}{4}}{p - (q - \frac{1}{4})} \quad \textit{and} \quad \tilde c^2 \;:=\; (1 - c^{2q-5/2})\,(1 - \tfrac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta)$$

for any $\delta > 0$ such that this expression is positive. Then we have that

$$\varepsilon_2(p,q) \;=\; \int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \;\le\; 2^{-3/2} + \delta^{-1}\,c^{-(1-q)} + \tilde c^{-(1+r)} \int_0^\infty y^r\,\exp\{-\tfrac{1}{2}y^2\}\,dy.$$

Proof of Lemma 2.4. The proof of Lemma 2.4 is less straightforward. We proceed by splitting up the sample in two parts, $X_1, \ldots, X_m$ and $X_{m+1}, \ldots, X_n$, for some convenient m. In this way we split up $\mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2$ by writing

$$\mathbb{T}_1 \;=\; \mathbb{T}_{11} + \mathbb{T}_{12} \;:=\; \textstyle\sum_{j=1}^m T_j + \sum_{j=m+1}^n T_j \tag{25}$$

and $\mathbb{T}_2 = \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23}$, with

$$\mathbb{T}_{21} := \textstyle\sum_{1\le i<j\le m} T_{i,j}, \quad \mathbb{T}_{22} := \sum_{1\le i\le m<j\le n} T_{i,j}, \quad \mathbb{T}_{23} := \sum_{m<i<j\le n} T_{i,j}. \tag{26}$$

We choose the number m of the form $[n\,a^{-r}]$, for some $r \in [0,2]$, in such a way that it is as large as possible, while still allowing us to remove $\mathbb{T}_{21}$ by a Taylor expansion. After this removal has taken place we condition the statistic on $X_{m+1}, \ldots, X_n$. The parts of $\mathbb{T}$ that still matter at this moment are $\mathbb{T}_{11}$ and $\mathbb{T}_{22}$, which conditionally are sums of independent random variables, $\mathbb{T}_{11}$ being $T_1 + \ldots + T_m$ with $T_j = T_j(X_j)$, and $\mathbb{T}_{22} = M_1 + \ldots + M_m$ with

$$M_j \;=\; M_j(X_j, X_{m+1}, \ldots, X_n) \;:=\; T_{j,m+1} + \ldots + T_{j,n}, \quad \text{for } 1 \le j \le m. \tag{27}$$

To smoothen the proof we need a truncation argument. Let

$$\varphi \;=\; \varphi(X_{m+1}, \ldots, X_n) \;:=\; \mathbb{E}(M_1^2 \,|\, X_{m+1}, \ldots, X_n). \tag{28}$$

Take any (small) $\delta > 0$ and let $I\{\varphi\} := I\{\varphi \le \delta\, n^{-1}\}$. We condition on the event that $I\{\varphi\} = 1$. To this, note that, by Markov's inequality,

$$\mathbb{P}(\varphi > \delta\, n^{-1}) \;\le\; \delta^{-1} n\, \mathbb{E}\varphi \;=\; \delta^{-1}\, n\,(n-m)\, \mathbb{E}T_{1,2}^2 \;\le\; \delta^{-1}\,(\tilde{\gamma}_2\, n^{-1}) \;\le\; \delta^{-1}\, a^{-1}.$$

As a result, we see that

$$\begin{aligned} \textstyle\int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;&\le\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\}|\, dt + \int_{a^p}^{a^q} |\mathbb{E}\,(1 - I\{\varphi\})\, e\{t\mathbb{T}\}|\, dt \\ &\le\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\}|\, dt + \delta^{-1}\, a^{-1+q}, \end{aligned} \tag{29}$$

as

$$|\mathbb{E}\,(1 - I\{\varphi\})\, e\{t\mathbb{T}\}| \;\le\; \mathbb{E}\,(1 - I\{\varphi\}) \;=\; \mathbb{P}(\varphi > \delta\, n^{-1}) \;\le\; \delta^{-1}\, a^{-1+q}.$$

Now let $\tilde{M}_j := I\{\varphi\}\, M_j$ for $1 \le j \le m$, $\tilde{\mathbb{T}}_{22} := I\{\varphi\}\, \mathbb{T}_{22} = \sum_{j=1}^m \tilde{M}_j$ and

$$\tilde{\mathbb{T}} \;:=\; \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{21} + \tilde{\mathbb{T}}_{22} + \mathbb{T}_{23}.$$

By construction $\mathbb{E}\, I\{\varphi\}\, e\{t\mathbb{T}\} = \mathbb{E}\, I\{\varphi\}\, e\{t\tilde{\mathbb{T}}\}$, so that from here on we may concentrate on

$$\varepsilon_3 \;:=\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t\tilde{\mathbb{T}}\}|\, dt.$$

Now take

$$m \;:=\; [n\, a^{-2q+1/2}],$$

noting that $1 \le m \le n$. We remove $\mathbb{T}_{21}$ by means of a Taylor expansion. As $\mathbb{E}\,\mathbb{T}_{21}^2 = (mn^{-1})^2\, \frac{1}{2}\,(\tilde{\gamma}_2\, n^{-1}) \le \frac{1}{2} a^{-4q+1}\, a^{-1} = \frac{1}{2} a^{-4q}$,

$$\varepsilon_3 \;\le\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\, I\{\varphi\}\, e\{t(\tilde{\mathbb{T}} - \mathbb{T}_{21})\}|\, dt + \int_{a^p}^{a^q} t\, \mathbb{E}\,|\mathbb{T}_{21}|\, dt \;=:\; \varepsilon_{31} + \varepsilon_{32},$$

with $\varepsilon_{32} \le \frac{1}{2} a^{2q}\, \{\mathbb{E}\,\mathbb{T}_{21}^2\}^{1/2} \le 2^{-3/2}$. As to ε_{31}, using a Taylor expansion we see that

$$\begin{aligned} &|\mathbb{E}\,(e\{t(T_1 + \tilde{M}_1)\} \,|\, X_{m+1}, \ldots, X_n)| \\ &\quad\le\; |\mathbb{E}\,(e\{tT_1\} + (it)\, e\{tT_1\}\tilde{M}_1 \,|\, X_{m+1}, \ldots, X_n)| + \tfrac{1}{2} t^2\, \mathbb{E}\,(\tilde{M}_1^2 \,|\, X_{m+1}, \ldots, X_n). \end{aligned}$$

Now by construction $\mathbb{E}\,(\tilde{M}_1^2 \,|\, X_{m+1},\ldots,X_n) = I\{\varphi\}\,\varphi \le \delta\, n^{-1}$. Furthermore,

$$e\{tT_1\} \;=\; 1 + R_1 \tag{30}$$

for a certain $R_1 = R_1(X_1)$ with $|R_1| \le t\,|T_1|$, and moreover

$$\mathbb{E}\,(\tilde{M}_1 \,|\, X_{m+1},\ldots,X_n) \;=\; I\{\varphi\}\,\mathbb{E}\,(M_1 \,|\, X_{m+1},\ldots,X_n) \;=\; 0,$$

so that with Hölder's inequality

$$\begin{aligned} |\mathbb{E}\,(e\{tT_1\}\tilde{M}_1 \,|\, X_{m+1},\ldots,X_n)| &= |\mathbb{E}\,(R_1\,\tilde{M}_1 \,|\, X_{m+1},\ldots,X_n)| \\ &\le\; t\,\mathbb{E}\,(|T_1\tilde{M}_1| \,|\, X_{m+1},\ldots,X_n) \;\le\; n^{-1/2}\,(\delta\, n^{-1})^{1/2}\,t \;=\; \delta^{1/2}\,t\,n^{-1}. \end{aligned}$$

As in (24), we have that $|\mathbb{E}\,e\{tT_1\}| \le 1 - \frac{1}{2}(1 - \frac{1}{3}c^{-(1-q)})\,t^2\,n^{-1}$, and as a result, for any $t \in [a^p, a^q]$,

$$\begin{aligned} |\mathbb{E}\,(e\{t(T_1 + \tilde{M}_1)\} \,|\, X_{m+1},\ldots,X_n)| \;&\le\; 1 - \tfrac{1}{2}(1 - \tfrac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta)\,t^2\,n^{-1} \\ &\le\; \exp\{-\tfrac{1}{2}\hat{c}\,t^2\,n^{-1}\}, \end{aligned} \tag{31}$$

where we write $\hat{c} := 1 - \frac{1}{3}c^{-(1-q)} - 2\,\delta^{1/2} - \delta$. It is now easily seen that

$$\begin{aligned} \varepsilon_{31} \;&\le\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\,|I\{\varphi\}\,\mathbb{E}\,(e\{t(\tilde{\mathbb{T}} - \mathbb{T}_{21})\} \,|\, X_{m+1},\ldots,X_n)|\,dt \\ &=\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\,I\{\varphi\}\,|e\{t(\mathbb{T}_{12} + \mathbb{T}_{23})\}|\,|\mathbb{E}\,(e\{t(\mathbb{T}_{11} + \tilde{\mathbb{T}}_{22})\} \,|\, X_{m+1},\ldots,X_n)|\,dt \\ &\le\; \textstyle\int_{a^p}^{a^q} \mathbb{E}\,|\mathbb{E}\,(e\{t(T_1 + \tilde{M}_1)\} \,|\, X_{m+1},\ldots,X_n)|^m\,dt \\ &\le\; \textstyle\int_{a^p}^{a^q} \exp\{-\tfrac{1}{2}\hat{c}\,t^2\,mn^{-1}\}\,dt. \end{aligned} \tag{32}$$

Since

$$mn^{-1} \;\ge\; a^{-2q+1/2} - n^{-1} \;\ge\; a^{-2q+1/2}(1 - a^{2q-5/2}) \;\ge\; a^{-2q+1/2}(1 - c^{2q-5/2}),$$

(note that by (20) we have that $n^{-1} \le a^{-2}$), it follows that

$$\begin{aligned} \varepsilon_{31} \;&\le\; \textstyle\int_{a^p}^{a^q} \exp\{-\tfrac{1}{2}\tilde{c}^2\,a^{-2q+1/2}\,t^2\}\,dt \\ &\le\; (\tilde{c}^2\,a^{-2q+1/2})^{-1/2}\,\textstyle\int_{(\tilde{c}^2\,a^{-2q+1/2})^{1/2}\,a^p}^{\infty} \exp\{-\tfrac{1}{2}y^2\}\,dy \\ &=\; \tilde{c}^{-1}\,a^{q-1/4}\,\textstyle\int_{\tilde{c}\,a^{p-q+1/4}}^{\infty} y^{-r}\,y^r\,\exp\{-\tfrac{1}{2}y^2\}\,dy \;\le\; \tilde{c}^{-(1+r)}\,\textstyle\int_0^{\infty} y^r\,\exp\{-\tfrac{1}{2}y^2\}\,dy \end{aligned} \tag{33}$$

(note that the final inequality is the point where we need the assumption that $p > q - \frac{1}{4}$), which finishes the proof. □

The proof of Theorem 1.1 for U-statistics is now easily concluded by splitting the interval $[a^{1/4}, a]$ into several intervals of type $[a^p, a^q]$, and applying Lemma 2.4. As the result of Theorem 1.1 is stated for arbitrary statistics, we postpone calculations to the next section.

2.3 Proof of Theorem 1.1 for arbitrary statistics

We go about in the same way as for U-statistics of order 2. Again we may assume without loss of generality that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We put

$$a := \min\{\tilde\beta_3^{-1}\, n^{1/2},\ \tilde\Delta_2^{-2}\, n\}, \quad a^{-1} = \max\{\tilde\beta_3\, n^{-1/2},\ \tilde\Delta_2^{2}\, n^{-1}\},$$

and prove the theorem via inequality (19). Again we assume that $a > c$ and again we will show that

$$\textstyle\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \ \le\ c.$$

In order to do this we need to decompose $\mathbb{T}$ in the following way. Let $m \in N$. As earlier we write

$$\mathbb{T}_{11} \ :=\ \textstyle\sum_{j=1}^m T_j.$$

For $1 \le j \le m$ we write

$$M_j := \sum_{k=m+1}^{n} T_{j,k} \quad \text{and} \quad N_j := \sum_{A:\,|A|\ge 3,\, A\cap\{1,\dots,m\}=\{j\}} T_A.$$

Furthermore, let

$$\Lambda \ =\ \Lambda(m) \ :=\ \sum_{A:\,|A\cap\{1,\dots,m\}|\ge 2} T_A,$$

and $R := \mathbb{E}\,(\mathbb{T}\,|\,X_{m+1},\dots,X_n) = \sum_{A\subset\{m+1,\dots,n\}} T_A$. Now

$$\mathbb{T} \ =\ \mathbb{T}_{11} + \sum_{j=1}^m (M_j + N_j) + \Lambda + R, \tag{34}$$

and we have divided up $\mathbb{T}$'s Hoeffding decomposition into groups of T_A according to the fact whether the A have either zero, one, or more than one elements in common with $\{1,\dots,m\}$. Divided up in this way we obtain

$$R, \quad \mathbb{T}_{11} + \textstyle\sum_{j=1}^n (M_j + N_j) \quad \text{and} \quad \Lambda$$

respectively. In the case of U-statistics of order 2, for example, we obtain:

$$\Lambda = 0, \quad \textstyle\sum_{j=1}^m M_j = \mathbb{T}_{22}, \quad \sum_{j=1}^m N_j = 0 \quad \text{and} \quad R = \mathbb{T}_{12} + \mathbb{T}_{23}.$$

Lemma 2.5 will provide us with bounds for the second moments $\mathbb{E}\,\Lambda^2$ and $\mathbb{E}\,N_1^2$. Our strategy is now as follows: as concerns ε_1 we just take $m = n$ after which we easily remove $\Lambda = \sum_{|A|\ge 3} T_A$. As concerns $\varepsilon_2(p,q)$ we take m as before and remove Λ by a Taylor expansion. After that we take $M_j + N_j$ instead of M_j in the original proof and use the same truncation.

The lemma reads as follows (cf. Bentkus, Götze, Van Zwet (1997), Lemma 4.1):

Lemma 2.5. *We have that*

$$\mathbb{E}\,\Lambda^2 \;\le\; \tfrac{1}{2}(mn^{-1})^2\,(\tilde\Delta_2^2\,n^{-1}) \quad \textit{and} \quad \mathbb{E}\,N_1^2 \;\le\; n^{-1}\,(\tilde\Delta_2^2\,n^{-1}).$$

Proof of Lemma 2.5. We may assume that $m \ge 2$. Notice that with (14) the following equality holds:

$$\tilde\Delta_2^2\,n^{-1} \;=\; n^2 \textstyle\sum_{\{1,2\}\subset A} \mathbb{E}\,T_A^2 \;=\; n^2 \sum_{l=2}^n \binom{n-2}{l-2}\,\mathbb{E}\,T_{1,\dots,l}^2, \tag{35}$$

using the fact that by symmetry $\mathbb{E}\,T_A^2 = \mathbb{E}\,T_{1,\dots,l}^2$ for any set A such that $|A| = l$, and that there are $\binom{n-2}{l-2}$ such sets which satisfy the condition. Using (35), we see that

$$\begin{aligned}\mathbb{E}\,\Lambda^2 \;&=\; \textstyle\sum_{A:\,|A\cap\{1,\dots,m\}|\ge 2} \mathbb{E}\,T_A^2 \;\le\; \sum_{1\le j<k\le m} \sum_{j,k\in A} \mathbb{E}\,T_A^2 \\ &=\; \textstyle\binom{m}{2} \sum_{\{1,2\}\subset A} \mathbb{E}\,T_A^2 \;\le\; \frac{1}{2} m^2\,n^{-2}\,(\tilde\Delta_2^2\,n^{-1}).\end{aligned}$$

As to $\mathbb{E}\,N_1^2$, we may suppose that $m \notin \{n-1, n\}$, since otherwise $N_1 = 0$. Using again (35) we see that

$$\begin{aligned}\mathbb{E}\,N_1^2 \;&=\; \textstyle\sum_{A:\,|A|\ge 2,\, A\subset\{m+1,\dots,n\}} \mathbb{E}\,T_{1,A}^2 \;\le\; (n-m) \sum_{\emptyset\ne A\subset\{m+2,\dots,n\}} \mathbb{E}\,T_{1,m+1,A}^2 \\ &\le\; n \textstyle\sum_{A\subset\{3,\dots,n\}} \mathbb{E}\,T_{1,2,A}^2 \;=\; n^{-1}\,(\tilde\Delta_2^2\,n^{-1}).\end{aligned}$$

This finishes the proof. $\square$

Now we proceed as promised. Let ε_1 and $\varepsilon_2(p,q)$ be as in (22). First we look at ε_1. Let $\mathbb{T}_1 := \sum_{j=1}^n T_j$ and $\mathbb{T}_2 := \sum_{A:\,|A|=2} T_A$ as before. Taking $m = n$ we have that $\Lambda = \sum_{A:\,|A|\ge 3} T_A$, while by Lemma 2.5 it is clear that $\mathbb{E}\,\Lambda^2 \le \frac{1}{2}\tilde\Delta_2^2\,n^{-1} \le \frac{1}{2}a^{-1}$. Therefore

$$\int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\} - \mathbb{E}\,e\{t(\mathbb{T}-\Lambda)\}|\,dt \;\le\; \int_0^{a^{1/4}} t\,\mathbb{E}\,|\Lambda|\,dt \;\le\; (\tfrac{1}{2}a^{-1})^{1/2}\,\tfrac{1}{2}a^{1/2} \;=\; 2^{-3/2},$$

and, via Lemma 2.3,

$$\varepsilon_1 \;\le\; 2^{-3/2} + 2^{-3/2} + (\tfrac{1}{2}\pi)^{1/2}\,(1 - \tfrac{1}{3}c^{-3/4})^{-1/2}.$$

As to $\varepsilon_2(p,q)$, we take $m := [n\,a^{-2q+1/2}]$ as earlier. The cost of removing Λ from $\mathbb{T}$ is (see Lemma 2.5) bounded by

$$\begin{aligned}\int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\} - \mathbb{E}\,e\{t(\mathbb{T}-\Lambda)\}|\,dt \;&\le\; \int_{a^p}^{a^q} t\,\{\mathbb{E}\,\Lambda^2\}^{1/2}\,dt \\ &\le\; \tfrac{1}{2}a^{2q}\,\{\tfrac{1}{2}(mn^{-1})^2\,(\tilde\Delta_2^2\,n^{-1})\}^{1/2} \;\le\; 2^{-3/2}\,a^{2q}\,a^{-2q+1/2}\,a^{-1/2} \;=\; 2^{-3/2}.\end{aligned} \tag{36}$$

Now set

$$\varphi \;=\; \varphi(X_{m+1},\dots,X_n) \;:=\; \mathbb{E}\big((M_1+N_1)^2 \,|\, X_{m+1},\dots,X_n\big). \tag{37}$$

By Lemma 2.5 and the fact that, with (35), $\mathbb{E}\,M_1^2 = (n-m)\,\mathbb{E}T_{1,2}^2 \le n^{-1}\,(\tilde\Delta_2^2\, n^{-1})$,

$$\mathbb{E}\varphi \;\le\; 2n^{-1}\,(\tilde\Delta_2^2\,n^{-1}) \;\le\; 2n^{-1}\,a^{-1},$$

so replacement of

$$\hat{\mathbb{T}} \;:=\; \mathbb{T} - \Lambda \;=\; \mathbb{T}_{11} + \textstyle\sum_{j=1}^m (M_j+N_j) + R$$

by the truncated version

$$\tilde{\mathbb{T}} \;:=\; \mathbb{T}_{11} + I\{\varphi \le \delta\, n^{-1}\}\, \textstyle\sum_{j=1}^m (M_j+N_j) + R$$

will cost us (see (29)) $\le 2\delta^{-1}\,a^{-(1-q)} \le 2\delta^{-1}\,c^{-(1-q)}$. After this we estimate ε_2 in completely the same way as ε_{31} in the proof of Lemma 2.4. This leads us to:

$$\varepsilon_2(p,q) \;\le\; 2^{-3/2} + 2^{-3/2} + 2\delta^{-1}\,c^{-(1-q)} + \tilde c^{-(1+r)} \textstyle\int_0^\infty y^r\, \exp\{-\frac12 y^2\}\, dy,$$

with $\tilde c$ and r as in Lemma 2.4.

Finally we turn to the concrete calculations. What we need is a sequence of intervals $[p_j, q_j]$, for $1 \le j \le k$, such that $q_1 = 1$, $q_{j+1} = p_j$ $(1 \le j \le k-1)$ and $p_k \le \frac14$, after which

$$\textstyle\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\;\le\;\; \varepsilon_1 + \sum_{j=1}^k \varepsilon_2(p_j,q_j)$$

and we are able to prove (21). We make the situation easier to handle by always taking $r=4$ and $\delta = 0.2$. This results in the sequence

$$p_1 = \tfrac{15}{16},\quad p_2 = \tfrac{55}{64},\quad p_3 = \tfrac{195}{264},\quad p_4 = \tfrac{655}{1024},\quad p_5 = \tfrac{1995}{4096},\quad p_6 = \tfrac{4855}{16384},\quad p_7 \le \tfrac14.$$

Notice that

$$\textstyle\int_0^\infty y^4\, \exp\{-\frac12 y^2\}\, dy \;\;=\;\; 3\,(\frac12\pi)^{1/2},$$

so that we get to

$$\begin{aligned}\textstyle\int_0^a |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\;\le\;\; & 16\cdot 2^{-3/2} + (\tfrac12\pi)^{1/2}\,(1-\tfrac13 c^{-3/4})^{-1/2} \\ & + 3\,(\tfrac12\pi)^{1/2} \textstyle\sum_{j=1}^7 \tilde c_j^{-5} + 2\delta^{-1} \sum_{j=1}^7 c^{-(1-q_j)},\end{aligned}$$

which in turn leads us to (21). This finishes the proof of Theorem 1.1 in the case that s is the norming factor.

We conclude by giving the proof of Theorem 1.1 in the case that σ is the norming factor. Here we have that $\sigma^2 = s^2 + R^2$, where (see (17) and (15))

$$R^2 \;\le\; \tfrac12\,\tilde\Delta_2^2\, n^{-1}. \tag{38}$$

In the case that $s^2/\sigma^2 \geq \frac{709}{710}$ the result of the theorem is now easily derived from the equality

$$Q(\mathbb{T}/\sigma, \lambda) \;=\; Q(\mathbb{T}/s, (\sigma/s)\,\lambda),$$

together with the result already obtained. On the other hand: in the case that $s^2/\sigma^2 < \frac{709}{710}$ it is clear that $R^2/\sigma^2 \geq \frac{1}{710}$, so that, with (38), $\frac{1}{2}\tilde{\Delta}_2^2\,\sigma^{-2}\,n^{-1} \geq \frac{1}{710}$. But then the desired result follows from the trivial estimate

$$Q(\mathbb{T}/\sigma, \lambda) \;\leq\; 1.$$

This concludes the proof of Theorem 1.1. □

2.4 Proof of Theorem 1.2 for U-statistics of order 2

We turn to the proof of Theorem 1.2. First we look at the case of U-statistics of order 2. Here we have that $\Delta^2 = \Delta_1^2 = \operatorname{var}\mathbb{T}_2$, cf. (16). As in Section 2.2, we assume that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We set

$$a \;:=\; \min\{\beta^{-1}, \Delta_1^{-2}\}, \quad a^{-1} \;=\; \max\{\beta, \Delta_1^2\}$$

and prove that $\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt$ is bounded by some absolute constant. We may assume that $a \geq 100$.

We pay no attention to concrete values of constants anymore: the reason for this is that the constants involved are getting too big to be of practical use anymore. As a result, it is convenient to introduce the notation

$$b_1 \;\ll\; b_2, \tag{39}$$

for any $b_1, b_2 \geq 0$, meaning by definition that $b_1 \leq c\,b_2$ for some absolute constant c. In this terminology we will prove that $\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1$. In case the constant involved is depending on, say, a, we will just write

$$b_1 \;\ll_a\; b_2.$$

We proceed as earlier. First we will show that as in Section 2.2

$$\varepsilon_1 \;:=\; \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1.$$

After that we prove that

$$\varepsilon_2 \;=\; \varepsilon_2(p,q) \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \;\ll_{p,q}\; 1,$$

for any $\frac{1}{4} < q \leq 1$ and $q - \frac{1}{4} < p \leq q$, and the result of the theorem is an easy consequence. We start with ε_1:

Lemma 2.6. *We have that*

$$\varepsilon_1 \;=\; \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \ll 1.$$

Proof of Lemma 2.6. By a Taylor expansion we see that $\varepsilon_1 \le \varepsilon_{11} + \varepsilon_{12}$ with $\varepsilon_{11} := \int_0^{a^{1/4}} |\mathbb{E}\, e\{t\mathbb{T}_1\}|\, dt$ and $\varepsilon_{12} := \int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2|\, dt \le \frac{1}{2}\Delta_1\, a^{1/2} \le \frac{1}{2}$, so we only need to get a bound for ε_{11}. As to ε_{11} we need the following inequality:

$$|\mathbb{E}\, e\{tT_j\}| \;\le\; \exp\{-\tfrac{1}{2}t^2\,(s_j^2 - \tfrac{1}{3}\beta_j\, \beta^{-1/4})\}. \tag{40}$$

Indeed, since $s_j^2 \le \beta_j^{2/3} \le \beta^{2/3}$ we have that $1 - \frac{1}{2}s_j^2\, t^2 \ge 0$ for $t \in [0, a^{1/4}]$. The Taylor expansion $e\{tT_j\} = 1 + (it)\,T_j + \frac{1}{2}(it)^2\,T_j^2 + R_j$, with $|R_j| \le \frac{1}{6}\,t^3\,|T_j|^3$, now shows us that

$$\mathbb{E}\, e\{tT_j\} \;=\; 1 - \tfrac{1}{2}s_j^2\, t^2 + \mathbb{E}\, R_j\,,$$

so that $|\mathbb{E}\, e\{tT_j\}| \le 1 - \frac{1}{2}s_j^2\, t^2 + \frac{1}{6}\,\beta_j\, t^3$, from which (40) follows easily (as $1 - y \le \exp\{-y\}$ for all $y \in \mathbb{R}$).

Note that as $a \ge 100$ we have that $\beta \le \frac{1}{100}$, and (40) leads us to:

$$\begin{aligned}\varepsilon_{11} \;=\; \int_0^{a^{1/4}} \prod_{j=1}^n |\mathbb{E}\, e\{tT_j\}|\, dt \;&\le\; \int_0^{a^{1/4}} \exp\{-\tfrac{1}{2}t^2 \textstyle\sum_{j=1}^n (s_j^2 - \tfrac{1}{3}\beta_j\, \beta^{-1/4})\}\, dt \\ &=\; \int_0^{a^{1/4}} \exp\{-\tfrac{1}{2}(1 - \tfrac{1}{3}\beta^{3/4})\, t^2\}\, dt \;\ll\; 1,\end{aligned}$$

which concludes the proof of the Lemma. □

Lemma 2.7. *Let $q \in (\frac{1}{4}, 1]$ and $p \in (q - \frac{1}{4}, q)$. Then we have that*

$$\varepsilon_2 \;=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\mathbb{T}\}|\, dt \;\ll_{p,q}\; 1.$$

Proof of Lemma 2.7. In the i.i.d. case we have treated the integral ε_2 by splitting up the sample into two parts, $X_1, \ldots, X_m$ versus $X_{m+1}, \ldots, X_n$, where m is of the form $[n\, a^{-r}]$ for some $r \in [0, 2]$. In this way $\mathbb{T}$ is split up as $\mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2$ with $\mathbb{T}_1 = \mathbb{T}_{11} + \mathbb{T}_{12}$ and $\mathbb{T}_2 = \mathbb{T}_{21} + \mathbb{T}_{22} + \mathbb{T}_{23}$ as in (25) and (26). Now $\mathbb{T}_{21}$ is removed by a Taylor expansion and using conditional expectations with respect to $X_{m+1}, \ldots, X_n$ we get to the result that we want. In the present case this approach is not satisfactory as in a sense it privileges the first m random variables over the remaining $n - m$. This problem is overcome by means of a very useful technique taken from Bentkus, Bloznelis, Götze (1996), to which we will refer as 'randomization'.

To this, let $\overline{X} := (\overline{X}_1, \ldots, \overline{X}_n)$ be an independent copy of $X := (X_1, \ldots, X_n)$, and let $\alpha_1, \ldots, \alpha_n$ be an i.i.d. sample from a Bernoulli distributed population with

$$\mathbb{P}(\alpha_1 = 1) = 1 - \mathbb{P}(\alpha_1 = 0) = m,$$

for some $m \in [0, 1]$ of the form a^{-r} $(r \ge 0)$. The sample $\alpha := (\alpha_1, \ldots, \alpha_n)$ is chosen independently of all other random variables involved. In particular we take

$$m \;:=\; a^{-(2q - \frac{1}{2})}.$$

Now we replace the original sample $X_1, \ldots, X_n$ by the 'mixed' sample

$$\alpha_1 X_1 + (1-\alpha_1)\overline{X}_1,\ \ldots\ , \alpha_n X_n + (1-\alpha_n)\overline{X}_n.$$

Clearly the two samples are identically distributed, as this is so for each fixed outcome of α. Thus, letting

$$\tilde{\mathbb{T}} \ :=\ \mathbb{T}(\alpha_1 X_1 + (1-\alpha_1)\overline{X}_1,\ \ldots,\ \alpha_n X_n + (1-\alpha_n)\overline{X}_n),$$

we have that $\mathbb{T} \overset{d}{=} \tilde{\mathbb{T}}$, where by $\overset{d}{=}$ we mean equality in distribution. As a consequence $\mathbb{E}e\{t\tilde{\mathbb{T}}\} = \mathbb{E}e\{t\mathbb{T}\}$, so it does not matter whether we replace $\mathbb{T}$ by $\tilde{\mathbb{T}}$.

Instead of isolating the first m variables $X_1, \ldots, X_m$ from the original sample, we may now take the set of X_j for which $\alpha_j = 1$ from the mixed sample. Our gain is that in a sense a randomization of the choices of the m variables has taken place.

We wonder what the Hoeffding decomposition of $\tilde{\mathbb{T}}$ looks like. To this, look at the statistic $\tilde{\mathbb{T}}$ for some fixed outcome of α. As an example we take the case in which $\alpha = (0, 1, \ldots, 1)$, that is, the case in which we look at the mixed sample $\overline{X}_1, X_2, \ldots, X_n$. Here $\tilde{\mathbb{T}} = \mathbb{T}(\overline{X}_1, X_2, \ldots, X_n)$, with a Hoeffding decomposition of the form

$$\tilde{T}_1(\overline{X}_1) + \textstyle\sum_{j=2}^n \tilde{T}_j(X_j) + \sum_{j=2}^n \tilde{T}_{1,j}(\overline{X}_1, X_j) + \sum_{2\le i<j\le n} \tilde{T}_{i,j}(X_i, X_j),$$

with the $\tilde{T}_j$ and the $\tilde{T}_{i,j}$ as in (5), using the mixed sample. Since T_j and $\tilde{T}_j$ are functions depending only on the distribution of the underlying sample, and the original and the mixed sample are identically distributed, we may take $T_j = \tilde{T}_j$ for all j. In the same way we may take $T_A = \tilde{T}_A$ for any $A \subset N$. As a result, the random vector

$$\big(\tilde{T}_1(\overline{X}_1),\, \tilde{T}_2(X_2),\, \ldots\, ,\, \tilde{T}_n(X_n),\, \tilde{T}_{1,2}(\overline{X}_1, X_2),\, \ldots\, ,\, \tilde{T}_{n-1,n}(X_{n-1}, X_n)\big) \tag{41}$$

is equal to

$$\big(T_1(\overline{X}_1),\, T_2(X_2),\, \ldots\, ,\, T_n(X_n),\, T_{1,2}(\overline{X}_1, X_2),\, \ldots\, ,\, T_{n-1,n}(X_{n-1}, X_n)\big),$$

and therefore has the same distribution as $(T_1(X_1), \ldots, T_{n-1,n}(X_{n-1}, X_n))$. Using the same type of argument, we see that in general

$$\tilde{\mathbb{T}} \ :=\ \tilde{\mathbb{T}}_1 + \tilde{\mathbb{T}}_2, \tag{42}$$

with

$$\tilde{\mathbb{T}}_1 \ =\ \tilde{\mathbb{T}}_{11} + \tilde{\mathbb{T}}_{12} \ :=\ \textstyle\sum_{j=1}^n \alpha_j\, T_j(X_j) + \sum_{j=1}^n (1-\alpha_j)\, T_j(\overline{X}_j), \tag{43}$$

and $\tilde{\mathbb{T}}_2 = \tilde{\mathbb{T}}_{21} + \tilde{\mathbb{T}}_{22} + \tilde{\mathbb{T}}_{23}$ with

$$\begin{aligned} &\tilde{\mathbb{T}}_{21} := \textstyle\sum_{i<j} \alpha_i\alpha_j\, T_{i,j}(X_i, X_j), \quad \tilde{\mathbb{T}}_{22} := \sum_{i\neq j} \alpha_i(1-\alpha_j)\, T_{i,j}(X_i, \overline{X}_j) \quad \text{and} \\ &\tilde{\mathbb{T}}_{23} := \textstyle\sum_{i<j} (1-\alpha_i)(1-\alpha_j)\, T_{i,j}(\overline{X}_i, \overline{X}_j) \end{aligned} \tag{44}$$

(where, e.g., by $T_{4,3}(X_4, \overline{X}_3)$ we actually mean $T_{3,4}(\overline{X}_3, X_4)$).

We continue with our estimation of $\varepsilon_2 = \int_{a^p}^{a^q} |\mathbb{E}\, e\{t\tilde{\mathbb{T}}\}|\, dt$. For reasons of convenience we drop the tilde from our notation, so that we will just talk about $\mathbb{T}$, $\mathbb{T}_1$ and so on for the decomposition as defined in (42), (43) and (44).

First we get rid of $\mathbb{T}_{21}$: by a Taylor expansion we see that $\varepsilon_2 \le \varepsilon_{21} + \varepsilon_{22}$ with

$$\varepsilon_{21} \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\, e\{t(\mathbb{T} - \mathbb{T}_{21})\}|\, dt \quad \text{and} \quad \varepsilon_{22} \;:=\; \int_{a^p}^{a^q} t\, \mathbb{E}\, |\mathbb{T}_{21}|\, dt,$$

where it is clear that

$$\varepsilon_{22} \;\le\; \tfrac{1}{2}\, \{m^2\, \Delta_1^2\}^{1/2}\, a^{2q} \;\le\; \tfrac{1}{2} a^{-(2q-\frac{1}{2})}\, a^{-1/2}\, a^{2q} \;=\; \tfrac{1}{2},$$

so that we only need to prove that $\varepsilon_{21} \ll_{p,q} 1$. To this, let

$$M_j := \textstyle\sum_{k\neq j} (1-\alpha_k)\, T_{j,k}(X_j, \overline{X}_k) \quad \text{and} \quad \varphi_j^2 := \mathbb{E}\,(M_j^2 \,|\, \overline{X}, \alpha) \tag{45}$$

be the expressions corresponding to (27) and (28). Note that as in (32)

$$\varepsilon_{21} \;\le\; \int_{a^p}^{a^q} \mathbb{E}\, \textstyle\prod_{j=1}^n V_j\, dt \tag{46}$$

with $V_j := |\mathbb{E}\,(e\{t\alpha_j(T_j + M_j)\} \,|\, \overline{X}, \alpha)|$. As in Section 2.2 we proceed by Taylor expanding, but we need to make the situation easier to handle. To this, in the case that $s_j^2 < 2^{-1/2}\, \beta_j\, \beta^{-1}$ we use the trivial bound $V_j \le 1$. Moreover, as before we condition on the event that $\sum_{j=1}^n \alpha_j\, \varphi_j^2 \le \delta\, m$ for some, small, $\delta > 0$. To this end, we set

$$\gamma_j := I\{s_j^2 \ge 2^{-1/2}\, \beta_j\, \beta^{-1}\}, \quad \tilde{\alpha}_j := \gamma_j\, \alpha_j \quad \text{and} \quad I\{\varphi\} := I\{\textstyle\sum_{j=1}^n \alpha_j\, \varphi_j^2 \le \delta\, m\}. \tag{47}$$

Moreover we set

$$W_j \;:=\; V_j^{\tilde{\alpha}_j} \;=\; |\mathbb{E}\,(e\{t\tilde{\alpha}_j(T_j + M_j)\} \,|\, \overline{X}, \alpha)|,$$

for which $V_j \le W_j \le 1$, and

$$\varepsilon_{23} \;:=\; \int_{a^p}^{a^q} \mathbb{E}\, I\{\varphi\}\, \textstyle\prod_{j=1}^n W_j\, dt.$$

It is clear that

$$\varepsilon_{21} \;\le\; \int_{a^p}^{a^q} \mathbb{E}\, \textstyle\prod_{j=1}^n W_j\, dt \;\le\; \varepsilon_{23} + \int_{a^p}^{a^q} \mathbb{E}\,(1 - I\{\varphi\})\, dt. \tag{48}$$

Markov's inequality leads us to the fact that

$$\begin{aligned} \mathbb{E}\,(1 - I\{\varphi\}) \;&=\; \mathbb{P}(\textstyle\sum_{j=1}^n \alpha_j\, \varphi_j^2 > \delta\, m) \;\le\; \delta^{-1}\, m^{-1}\, \mathbb{E}\, \textstyle\sum_{j=1}^n \alpha_j\, \varphi_j^2 \\ &\le\; \delta^{-1}\, m^{-1}\, m\, \textstyle\sum_{j=1}^n \mathbb{E}\, M_j^2 \;\le\; 2\delta^{-1}\, \Delta_1^2, \end{aligned} \tag{49}$$

so that $\int_{a^p}^{a^q} \mathbb{E}\,(1 - I\{\varphi\})\, dt \le 2\delta^{-1}\, a^{-(1-q)}$ and

$$\varepsilon_{21} \;\le\; \varepsilon_{23} + 2\delta^{-1}.$$

We prove that $\varepsilon_{23} \ll_{p,q} 1$.

As to W_j, expansions into Taylor series lead us to

$$\begin{aligned}
W_j &\leq |\mathbb{E}\,(e\{t\tilde{\alpha}_j T_j\} + (it)\, e\{t\tilde{\alpha}_j T_j\}\, \tilde{\alpha}_j M_j \mid \overline{X}, \alpha)| + \tfrac{1}{2}t^2\, \tilde{\alpha}_j\, \mathbb{E}\,(M_j^2 \mid \overline{X}, \alpha) \\
&\leq |\mathbb{E}\,(e\{t\tilde{\alpha}_j T_j\} \mid \alpha_j)| + t^2\, \tilde{\alpha}_j\, \mathbb{E}\,(|T_j\, M_j| \mid \overline{X}, \alpha) + \tfrac{1}{2}t^2\, \tilde{\alpha}_j\, \varphi_j^2 \\
&\leq |1 - \tfrac{1}{2}t^2\, \tilde{\alpha}_j\, s_j^2| + \tfrac{1}{6}t^3\, \tilde{\alpha}_j\, \beta_j + t^2\, \tilde{\alpha}_j\, \mathbb{E}\,(|\delta^{1/4} T_j|\, |\delta^{-1/4} M_j| \mid \overline{X}, \alpha) + \tfrac{1}{2}t^2\, \tilde{\alpha}_j\, \varphi_j^2 \\
&\leq 1 - \tfrac{1}{2}t^2\, \tilde{\alpha}_j\, (s_j^2 - \tfrac{1}{3}\beta_j\, \beta^{-1} - 2(\delta^{1/2}\, s_j^2)^{1/2}\, (\delta^{-1/2}\, \varphi_j^2)^{1/2} - \varphi_j^2) \\
&\leq \exp\{-\tfrac{1}{2}t^2\, \tilde{\alpha}_j\, ((1 - \delta^{1/2})\, s_j^2 - \tfrac{1}{3}\beta_j\, \beta^{-1} - (1 + \delta^{-1/2})\, \varphi_j^2)\}. \qquad (50)
\end{aligned}$$

Here we used Hölder's inequality, the fact that in general both $2\, x^{1/2}\, y^{1/2} \leq x + y$ (for $x, y \geq 0$) and $1 - x \leq \exp\{-x\}$, and the fact that by construction of $\tilde{\alpha}_j$ we have that $1 - \frac{1}{2}t^2\, \tilde{\alpha}_j\, s_j^2 \geq 0$. As to the latter: in case $\tilde{\alpha}_j = 1$ it is clear that $1 \geq \frac{1}{2}\beta^{-2}\, s_j^2$, whereas in case $1 < \frac{1}{2}\beta^{-2}\, s_j^2$ it is clear that $s_j^2 < (\frac{1}{2}\beta^{-2}\, s_j^2)^{1/2}\, s_j^2 \leq 2^{-1/2}\, \beta_j\, \beta^{-1}$ and therefore $\tilde{\alpha}_j = 0$.

Writing $r_j := (1 - \delta^{1/2})\, s_j^2 - \frac{1}{3}\beta_j\, \beta^{-1}$, we see that

$$\begin{aligned}
I\{\varphi\} \textstyle\prod_{j=1}^n W_j &\leq I\{\varphi\}\, \exp\{-\tfrac{1}{2}t^2\, \textstyle\sum_{j=1}^n \tilde{\alpha}_j\, (r_j - (1 + \delta^{-1/2})\, \varphi_j^2)\} \\
&\leq I\{\varphi\}\, \exp\{\tfrac{1}{2}t^2\, \textstyle\sum_{j=1}^n \tilde{\alpha}_j\, (1 + \delta^{-1/2})\, \varphi_j^2\}\, \exp\{-\tfrac{1}{2}t^2\, \textstyle\sum_{j=1}^n \tilde{\alpha}_j\, r_j\}. \qquad (51)
\end{aligned}$$

Now we have that

$$\begin{aligned}
\varepsilon_{23} &\leq \int_{a^p}^{a^q} \exp\{\tfrac{1}{2}m\, t^2\, \delta\, (1 + \delta^{-1/2})\}\, \mathbb{E}\, \exp\{-\tfrac{1}{2}t^2\, \textstyle\sum_{j=1}^n \tilde{\alpha}_j\, r_j\}\, dt \\
&= \int_{a^p}^{a^q} \exp\{\tfrac{1}{2}m\, t^2\, \delta^{1/2}\, (1 + \delta^{1/2})\}\, \textstyle\prod_{j=1}^n \mathbb{E}\, \exp\{-\tfrac{1}{2}t^2\, \tilde{\alpha}_j\, r_j\}\, dt. \qquad (52)
\end{aligned}$$

Taking $Z, Z_1, \ldots, Z_n$ to be an independent sample from the standard normal distribution, we can write

$$v_j \;:=\; \mathbb{E}\, \exp\{-\tfrac{1}{2}t^2\, \tilde{\alpha}_j\, r_j\} \;=\; \mathbb{E}\, e\{t\, \alpha_j\, (\gamma_j\, r_j)^{1/2}\, Z_j\}.$$

Note that $\gamma_j\, r_j \geq 0$ for $j = 1, \ldots, n$ (that is to say: $r_j < 0 \Rightarrow \gamma_j = 0$) if we take δ small enough, which we will. Expansion into Taylor series leads to the following:

$$\begin{aligned}
v_j &\leq 1 - \tfrac{1}{2}m\, t^2\, \gamma_j\, r_j + \tfrac{1}{6}\, m\, t^3\, \mathbb{E}|Z|^3\, (\gamma_j\, r_j)^{3/2} \\
&\leq 1 - \tfrac{1}{2}m\, t^2\, \gamma_j\, (r_j - \tfrac{1}{3}\mathbb{E}|Z|^3\, (1 - \delta^{1/2})^{3/2}\, s_j^3\, \beta^{-1}) \\
&\leq \exp\{-\tfrac{1}{2}m\, t^2\, \gamma_j\, w_j\}, \qquad (53)
\end{aligned}$$

with

$$w_j \;:=\; (1 - \delta^{1/2})\, s_j^2 - \tfrac{1}{3}(1 + 4(2\pi)^{-1/2}\, (1 - \delta^{1/2})^{3/2})\, \beta_j\, \beta^{-1},$$

and as a result

$$\varepsilon_{23} \;\leq\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2}m\, t^2\, (\textstyle\sum_{j=1}^n \gamma_j w_j - \delta^{1/2}\, (1 + \delta^{1/2}))\}\, dt.$$

Finally we note that $\sum_{j=1}^{n} \gamma_j w_j \geq \sum_{j=1}^{n} w_j$ because $\gamma_j = 0 \Rightarrow w_j \leq 0$ for δ small enough, so that

$$\varepsilon_{23} \;\leq\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2} m\, t^2 \,(\textstyle\sum_{j=1}^{n} w_j - \delta^{1/2}\,(1+\delta^{1/2}))\}\,dt \;=\; \int_{a^p}^{a^q} \exp\{-\tfrac{1}{2}\tilde{\delta}\, m\, t^2\}\,dt, \tag{54}$$

with

$$\tilde{\delta} \;:=\; 1 - 2\,\delta^{1/2} - \tfrac{1}{3} - \tfrac{4}{3}\,(2\pi)^{-1/2}\,(1-\delta^{1/2})^{3/2} - \delta. \tag{55}$$

Taking $\delta > 0$ such that $r_j < 0 \Rightarrow \gamma_j = 0$, $\gamma_j = 0 \Rightarrow w_j \leq 0$ and $\tilde{\delta} > 0$, we see as in (33) that $\varepsilon_{23} \ll_{p,q} 1$. Thus $\varepsilon_{21} \ll_{p,q} 1$ and $\varepsilon_2 \ll_{p,q} 1$, which completes the proof. $\square$

2.5 Proof of Theorem 1.2 for arbitrary statistics

We go about in the same way as for U-statistics of order 2. Again we may assume without loss of generality that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. We put

$$a \;:=\; \min\{\beta^{-1},\; \Delta^{-2}\}, \qquad a^{-1} \;=\; \max\{\beta,\; \Delta^2\}$$

and prove that $\int_0^a |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt$ is bounded by some absolute constant. We may assume that $a \geq 100$. We set

$$\varepsilon_1 \;:=\; \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt \quad \text{and} \quad \varepsilon_2 \;=\; \varepsilon_2(p,q) \;:=\; \int_{a^p}^{a^q} |\mathbb{E}\,e\{t\mathbb{T}\}|\,dt$$

with $\frac{1}{4} < q \leq 1$ and $q - \frac{1}{4} < p < q$ and prove that $\varepsilon_1 \ll 1$ and $\varepsilon_2 \ll_{p,q} 1$, after which Theorem 1.2 is an obvious consequence.

We start with ε_1. Here we have that

$$\varepsilon_1 \;\leq\; \int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt + \int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2 + \mathbf{T}_3|\,dt$$

with (see (17))

$$\int_0^{a^{1/4}} t\,\mathbb{E}\,|\mathbb{T}_2 + \mathbf{T}_3|\,dt \;\leq\; \tfrac{1}{2}\,\{\mathbb{E}\,|\mathbb{T}_2 + \mathbf{T}_3|^2\}^{1/2}\,a^{1/2} \;\leq\; \tfrac{1}{2}(\Delta^2)^{1/2}\,a^{1/2} \;\leq\; \tfrac{1}{2},$$

and moreover, taking $\tilde{a} := \min\{\beta^{-1}, \Delta_1^{-2}\}$, by the proof of Lemma 2.6,

$$\int_0^{a^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt \;\leq\; \int_0^{\tilde{a}^{1/4}} |\mathbb{E}\,e\{t\mathbb{T}_1\}|\,dt \;\ll\; 1:$$

as a consequence $\varepsilon_1 \ll 1$.

Using the method which was used in the proof of Lemma 2.7 we show that

$$\varepsilon_2 \;\ll_{p,q}\; 1.$$

To this end we take again an independent copy $\overline{X}_1, \ldots, \overline{X}_n$ of $X_1, \ldots, X_n$, together with an i.i.d. sample $\alpha_1, \ldots, \alpha_n$ with

$$\mathbb{P}(\alpha_1 = 1) \;=\; 1 - \mathbb{P}(\alpha_1 = 0) \;=\; m \;:=\; a^{-(2q - \frac{1}{2})}.$$

Now let for any pair $A, B \subset N$ with $A \cap B = \emptyset$

$$T_{A,\overline{B}} \;:=\; T_{A,B}(\{X_j : j \in A\}, \{\overline{X}_k : k \in B\})$$

(that is to say, by $T_{1,4,\overline{2},\overline{3}}$ in fact we mean $T_{1,2,3,4}(X_1, \overline{X}_2, \overline{X}_3, X_4)$, and so on), and let for any subset $A = \{n_1, \dots, n_k\} \subset N$

$$\alpha_A \;:=\; \prod_{j=1}^k \alpha_{n_j} \quad \text{and} \quad (1-\alpha)_A \;:=\; \prod_{j=1}^k (1-\alpha_{n_j}).$$

We look at $\tilde{\mathbb{T}}$ instead of $\mathbb{T}$, where

$$\tilde{\mathbb{T}} \;:=\; \tilde{\mathbb{T}}_{11} + \sum_{j=1}^n \alpha_j\,(\tilde{M}_j + \tilde{N}_j) + \tilde{\Lambda} + \tilde{R}, \tag{56}$$

with $\tilde{\mathbb{T}}_{11}$ as in (43),

$$\tilde{M}_j \;:=\; \sum_{k:k\neq j}(1-\alpha_k)\,T_{j,\overline{k}}, \quad \tilde{N}_j \;:=\; \sum_{B:\,j\notin B, |B|\geq 2}(1-\alpha)_B\,T_{j,\overline{B}},$$

for $j = 1, \dots, n$,

$$\tilde{\Lambda} \;:=\; \sum_{A:\,|A|\geq 2} \alpha_A \sum_{B:\,A\cap B=\emptyset}(1-\alpha)_B\,T_{A,\overline{B}},$$

and $\tilde{R} := \sum_B (1-\alpha)_B\,T_{\overline{B}}$. This decomposition is the analogue to the one which can be found in (34).

Analogue to Lemma 2.5, we need bounds on the moments of $\tilde{\Lambda}$, the $\tilde{M}_j$ and the $\tilde{N}_j$. The bounds are much similar to the ones obtained before, and read as follows:

Lemma 2.8. *We have:*

$$\mathbb{E}\,\tilde{\Lambda}^2 \leq m^2\,\Delta^2, \quad \sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \leq 2\,\Delta^2, \quad \sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 \leq \Delta^2.$$

Proof of Lemma 2.8. Going about as in (18), we see that

$$\begin{aligned}
\mathbb{E}\,\tilde{\Lambda}^2 &= \sum_{A:\,|A|\geq 2} m^{|A|} \sum_{B:\,A\cap B=\emptyset}(1-m)^{|B|}\,\mathbb{E}\,T^2_{A,\overline{B}} \\
&= \sum_{l=2}^n \sum_{C:\,|C|=l} \Big\{ \sum_{k=2}^l \binom{l}{k} m^k\,(1-m)^{l-k} \Big\}\,\mathbb{E}\,T_C^2.
\end{aligned} \tag{57}$$

Indeed, this corresponds to the fact that, given any subset $C \subset N$ with $|C| = l \geq 2$, restricting ourselves to sets A for which $|A| = k$, $2 \leq k \leq l$, we come $\binom{l}{k}$ times across $\mathbb{E}\,T_C^2$. We notice that

$$\begin{aligned}
\sum_{k=2}^l \binom{l}{k} m^k\,(1-m)^{l-k} &= m^2\,l(l-1) \sum_{k=2}^l \frac{1}{k(k-1)}\,\frac{(l-2)!}{(k-2)!\,(l-k)!}\,m^{k-2}\,(1-m)^{l-k} \\
&\leq m^2 \binom{l}{2} \sum_{r=0}^{l-2} \binom{l-2}{r} m^r\,(1-m)^{(l-2)-r} \;=\; m^2 \binom{l}{2},
\end{aligned}$$

using the substitution $r := k-2$ and Newton's binomial formula, and from (57) and (18) it is clear that

$$\mathbb{E}\,\tilde{\Lambda}^2 \;\leq\; m^2\,\Delta^2.$$

As to the $\tilde{M}_j$, we see that

$$\textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{M}_j^2 \;\le\; \sum_{j=1}^n \sum_{k:\,k\ne j} \mathbb{E}\,T_{j,\overline{k}}^2 \;=\; 2\,\Delta_1^2 \;\le\; 2\,\Delta^2.$$

Finally, using the fact that $l \le \binom{l}{2}$ for $3 \le l \le n$, we see that

$$\textstyle\sum_{j=1}^n \mathbb{E}\,\tilde{N}_j^2 \;\le\; \sum_{j=1}^n \sum_{B:\,j\notin B, |B|\ge 2} \mathbb{E}\,T_{j,\overline{B}}^2 \;=\; \sum_{l=3}^n \sum_{C:\,|C|=l} l\,\mathbb{E}\,T_C^2 \;\le\; \Delta^2,$$

again using (18). This finishes the proof. □

We continue with our estimation of

$$\varepsilon_2 \;=\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\,e\{t\tilde{\mathbb{T}}\}|\,dt.$$

First we remove $\tilde{\Lambda}$ by means of a simple Taylor expansion. This leads us to the following:

$$\varepsilon_2 \;\le\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\,e\{t(\tilde{\mathbb{T}} - \tilde{\Lambda})\}|\,dt + \int_{a^p}^{a^q} t\,\mathbb{E}\,|\tilde{\Lambda}|\,dt \;=:\; \varepsilon_{21} + \varepsilon_{22}.$$

As to ε_{22} we have that $\mathbb{E}\,\tilde{\Lambda}^2 \le m^2\,\Delta^2 \le m^2\,a^{-1}$, so that

$$\varepsilon_{22} \;\le\; \{m^2\,a^{-1}\}^{1/2}\,\tfrac{1}{2}a^{2q} \;\le\; \tfrac{1}{2}m\,a^{2q-\frac{1}{2}} \;=\; \tfrac{1}{2}.$$

This leaves us only the estimation of

$$\varepsilon_{21} \;=\; \textstyle\int_{a^p}^{a^q} |\mathbb{E}\,e\{t\,(\tilde{\mathbb{T}}_{11} + \sum_{j=1}^n \alpha_j\,(\tilde{M}_j + \tilde{N}_j) + \tilde{R})\}|\,dt.$$

We proceed as with the U-statistics of order 2.

Taking $V_j := |\mathbb{E}\,(e\{t\,\alpha_j\,(T_j + \tilde{M}_j + \tilde{N}_j)\}\,|\,\overline{X},\alpha)|$, we have (46), thus removing $\tilde{R}$ from our considerations as well. We go about in the exact same way as in Section 2.4, the only difference being that we take

$$\varphi_j^2 \;:=\; \mathbb{E}\,((\tilde{M}_j + \tilde{N}_j)^2\,|\,\overline{X},\alpha) \;=\; \mathbb{E}\,(\tilde{M}_j^2 + \tilde{N}_j^2\,|\,\overline{X},\alpha) \tag{58}$$

instead of (45) (note that, conditionally on $\overline{X}$ and α, $\tilde{M}_j$ and $\tilde{N}_j$ are uncorrelated). Again we use the indicator $I\{\varphi\}$ as in (47) to ensure that $\sum_{j=1}^n \tilde{\alpha}_j\,\varphi_j^2 \le \delta\,m$. The cost of getting this indicator in is bounded by

$$\begin{aligned} \mathbb{P}(\textstyle\sum_{j=1}^n \tilde{\alpha}_j\,\varphi_j^2 > \delta\,m) \;&\le\; \delta^{-1}\,m^{-1}\,m\,\textstyle\sum_{j=1}^n(\mathbb{E}\,\tilde{M}_j^2 + \mathbb{E}\,\tilde{N}_j^2) \\ &\le\; \delta^{-1}\,(2\Delta^2 + \Delta^2) \;\le\; 3\delta^{-1}\,a^{-1} \end{aligned} \tag{59}$$

(see both (49) and Lemma 2.8). After this the fact that $\varepsilon_2 \ll_{p,q} 1$ is proved in exactly the same way as in Section 2.4. This concludes the proof of Theorem 1.2. □

3 Applications

3.1 Wilcoxon's rank-sum test

Our first application will be on finding a concentration bound for Wilcoxon's rank-sum test.

Let $Y, Y_1, \ldots, Y_m$ and $Z, Z_1, \ldots, Z_n$ be independent i.i.d. samples, with distribution functions F and G respectively. We assume that these two functions are continuous and we are going to test the null hypothesis that they are equal. We do this as follows: let R_j denote the rank number of Y_j among $Y_1, \ldots, Y_m, Z_1, \ldots, Z_n$, that is, let

$$R_j \;:=\; \textstyle\sum_{i=1}^m I\{Y_j \geq Y_i\} + \sum_{i=1}^n I\{Y_j \geq Z_i\}.$$

Note that ties (equal outcomes) occur with probability 0 as we take the distribution functions to be continuous. Wilcoxon's rank-sum statistic is defined by:

$$\mathbb{T} \;:=\; \textstyle\sum_{j=1}^m R_j. \tag{60}$$

It is well-known (see, e.g., Gibbons, Chakraborti (1992) or Pestman (1998)) that for relatively large m and n the distribution of the statistic (60) under the null hypothesis may be approximated by a normal one. Thus $\mathbb{T}$ gives a way to test whether $F = G$. We talk about Wilcoxon's rank-sum test, which is equivalent to the also well-known Mann-Whitney test. Notice that our statistic $\mathbb{T}$ is not symmetric in its arguments.

Theorem 3.1. *Let $\mathbb{T}$ be Wilcoxon's rank-sum statistic, as defined in (60). Under the null hypothesis $F = G$ we have that $s^2 = \frac{1}{12}mn(m+n)$, and an absolute constant c exists such that*

$$Q(\mathbb{T}/s, \lambda) \;\leq\; c \max\left\{\lambda, (m+n)^{-1/2}\left(\sqrt{m/n} + \sqrt{n/m}\right)\right\}.$$

Proof of Theorem 3.1. Let $(X_1, \ldots, X_{m+n}) := (Y_1, \ldots, Y_m, Z_1, \ldots, Z_n)$: under the null hypothesis this constitutes an i.i.d. sample. Therefore we may write $\mathbb{T}$ in the form

$$\mathbb{T} \;=\; \textstyle\sum_{1 \leq i < j \leq m+n} g_{ij}(X_i, X_j),$$

by just taking

$$g_{ij}(x, y) \;:=\; I\{i \leq m < j\}\left(I\{x \geq y\} + \frac{m+1}{2n}\right).$$

This shows that $\mathbb{T}$ is a U-statistic of order 2, and certainly $\Delta^2 = \Delta_1^2$, see (16). Setting

$$\xi_j \;:=\; \mathbb{E}\left(I\{Y_j \geq Z\} \,|\, Y_j\right) - \tfrac{1}{2} \quad \text{and} \quad \eta_k := \mathbb{E}\left(I\{Y \geq Z_k\} \,|\, Z_k\right) - \tfrac{1}{2},$$

for $1 \leq j \leq m$ and $1 \leq k \leq n$, we see that

$$T_j \;=\; I\{1 \leq j \leq m\}\, n\, \xi_j + I\{m < j \leq m+n\}\, m\, \eta_{j-m}$$

for all j. Furthermore, for all $1 \le i < j \le m+n$ we have that

$$T_{ij} = I\{i \le m < j\} \ \big(I\{Y_i \ge Z_{j-m}\} - \mathbb{E}(I\{Y_i \ge Z\} \,|\, Y_i) - \mathbb{E}(I\{Y \ge Z_{j-m}\} \,|\, Z_{j-m}) + \tfrac{1}{2}\big)\,.$$

Now we can determine s^2, β and Δ_1^2. First we note that

$$\mathbb{E}\,|\xi_1|^2 \;=\; \mathbb{E}\,\big|\mathbb{E}(I\{F(Y_1) \ge F(Z)\} \,|\, Y_1) - \tfrac{1}{2}\big|^2 \;=\; \mathbb{E}\,\big|F(Y_1) - \tfrac{1}{2}\big|^2 \;=\; \tfrac{1}{12},$$

since $F(Z)$ and $F(Y_1)$ are independent and uniformly distributed on $(0,1)$. Of course then also $\mathbb{E}\,|\xi_j|^2 = \frac{1}{12}$ for all j. In the same way it is seen that always $\mathbb{E}\,|\eta_k|^2 = \frac{1}{12}$ and $\mathbb{E}\,|\xi_j|^3 = \mathbb{E}\,|\eta_k|^3 = \mathbb{E}\,|F(Y_1) - \frac{1}{2}|^3 = \frac{1}{32}$, and that, for all $1 \le i \le m < j \le m+n$, we have:

$$\mathbb{E}\,\big|I\{Y_i \ge Z_{j-m}\} - \mathbb{E}(I\{Y_i \ge Z\} \,|\, Y_i) - \mathbb{E}(I\{Y \ge Z_{j-m}\} \,|\, Z_{j-m}) + \tfrac{1}{2}\big|^2 \;=\; \tfrac{1}{12}.$$

It turns out that $\mathbb{E}\,\mathbb{T} = \frac{1}{2}m(m+n+1)$,

$$s^2 = \tfrac{1}{12}mn(m+n), \quad \beta = \tfrac{1}{32}mn(m^2+n^2) \quad \text{and} \quad \Delta_1^2 = \tfrac{1}{12}mn. \tag{61}$$

Theorem 3.1 now is an easy consequence of Theorem 1.2. $\square$

3.2 Self-normalized statistics

Next we take a look at so-called self-normalized statistics. What we are in fact thinking of are Student's statistic in a one- and a two-sample setting and two standard statistics associated with the model of linear regression. These will be considered in the following two subsections. The four statistics have a common structure that is asking for a unified approach. The price we pay for this will be that our key result, Theorem 3.2, is of a rather abstract form.

Let $X_1, \ldots, X_n$ be independent, not necessarily identically distributed real-valued random variables. Writing $X := (X_1, \ldots, X_n)$, let $R = R(X)$ be a (supposedly negligible) random variable satisfying

$$R(\lambda X) \;=\; \lambda^2\, R(X) \tag{62}$$

for all $\lambda \ge 0$. We are typically thinking of cases like $R(X) = 0$ or $R(X) = -n\,\overline{X}^2$, with $\overline{X} := n^{-1}\sum_{j=1}^n X_j$ denoting the sample mean. Let $c_1, \ldots, c_n$ moreover be fixed real numbers, for which we will assume that

$$\textstyle\sum_{j=1}^n c_j^2 \;=\; n.$$

An application of Theorem 1.2 may be found for the general self-normalized statistic

$$\mathbb{T} \;:=\; \frac{\sum_{j=1}^n c_j X_j}{\big(\sum_{k=1}^n X_k^2 + R(X)\big)^{1/2}}, \tag{63}$$

if we confine ourselves to the following two cases: either the sample $X_1, \ldots, X_n$ is identically distributed, or all c_j are equal to 1.

We formulate our result. Let $M_j \geq 0$, for $1 \leq j \leq n$, denote fixed positive numbers, usually depending on the probability distribution of X_j, and let

$$s^2 \;:=\; \sum_{j=1}^n \mathbb{E}\, X_j^2\, I\{X_j^2 \leq M_j^2\}, \quad U_j \;:=\; s^{-1} X_j\, I\{X_j^2 \leq M_j^2\} \quad \text{and} \quad s_j^2 \;:=\; \mathbb{E}\, U_j^2,$$

assuming that $s^2 \neq 0$. We write $U := (U_1, \ldots, U_n)$. We assume as well that $M_j \leq s$ for all j, so that by construction $|U_j| \leq 1$ and $\sum_{j=1}^n s_j^2 = 1$. We write

$$A_1 \;:=\; \sum_{j=1}^n \mathbb{P}(X_j^2 > M_j^2), \quad A_2 \;:=\; \sum_{j=1}^n |\mathbb{E}\, U_j|, \quad A_3 \;:=\; \sum_{j=1}^n \mathbb{E}\, |U_j|^3.$$

Taking a random variable α such that $\mathbb{P}(\alpha = s_j^2) = \frac{1}{n}$ for $1 \leq j \leq n$, and using Lyapunov's inequality, we derive that

$$A_3 \;\geq\; \sum_{j=1}^n (s_j^2)^{3/2} \;=\; n\, \mathbb{E}\, |\alpha|^{3/2} \;\geq\; n\, (\mathbb{E}\, |\alpha|)^{3/2} \;=\; n^{-1/2}. \tag{64}$$

For $p > 0$, we write

$$\nu_p \;:=\; n^{-1} \sum_{j=1}^n |c_j|^p.$$

Note that $\nu_2 = 1$, so with Lyapunov's inequality $\nu_1 \leq 1 \leq \nu_3$. In case $c_j = 1$ for all j, we have that $\nu_p = 1$ for all p.

As concerns the concentration bound of $\mathbb{T}$, we have the following general result:

Theorem 3.2. *Assume that $s^2 \neq 0$, $M_j \leq s$ for all j and $\sum_{j=1}^n c_j^2 = n$. Moreover, assume that either the sample $X_1, \ldots, X_n$ is i.i.d. or that $c_j = 1$ for all j. Finally assume that (62) applies and $\mathbb{E}\, |R(U)|^{3/2} \ll A_3^3$. Then there exists an absolute constant c such that*

$$Q(\mathbb{T}, \lambda) \;\leq\; c\, \max\{\lambda,\, A_1,\, A_2,\, \nu_3\, A_3\}. \tag{65}$$

The remainder of the current subsection will constitute the proof of Theorem 3.2. To this, let

$$A \;:=\; \max\{A_1,\, A_2,\, \nu_3\, A_3\}.$$

Taking the c from (65), it is clear that the statement in the Theorem is trivially true in case $A \geq c^{-1}$, so we assume throughout that $A < c^{-1}$. We take $c \geq 1$. We start the proof by turning from our original statistic $\mathbb{T}$ to truncated versions of it. First we replace all X_j by U_j. After this we get rid of the normalization by means of division by a square root, using a smooth function g resembling the function $y \mapsto y^{-1/2}$. After this we are ready to apply Theorem 1.2.

Let

$$\tilde{\mathbb{T}} \;:=\; \mathbb{T}(U) \;=\; (\textstyle\sum_{j=1}^n c_j\, U_j)/(\sum_{k=1}^n U_k^2 + R(U))^{1/2}.$$

Since, for any $x \in \mathbb{R}$ and $\lambda \geq 0$,

$$\begin{aligned}\mathbb{P}(x \leq \mathbb{T} \leq x+\lambda) &\leq \mathbb{P}(x \leq \mathbb{T} \leq x+\lambda,\, X_j^2 \leq M_j^2, \text{ all } j) + \mathbb{P}(X_j^2 > M_j^2, \text{ some } j) \\ &\leq \mathbb{P}(x \leq \tilde{\mathbb{T}} \leq x+\lambda) + A_1 \end{aligned} \tag{66}$$

(this is the reason why we need (62)), it is clear that

$$Q(\mathbb{T},\lambda) \;\leq\; Q(\tilde{\mathbb{T}},\lambda) + A_1, \tag{67}$$

and we may turn our attention to finding $Q(\tilde{\mathbb{T}},\lambda)$.

Our second truncation takes place in order to get rid of the division by the square root. To this, let $g : [0,\infty) \to \mathbb{R}$ be a bounded, infinitely many times differentiable function with bounded derivatives, such that

$$g(y) \;=\; y^{-1/2} \qquad \text{for } y \in [\tfrac{1}{2},\infty). \tag{68}$$

Now let

$$V_k \;:=\; U_k^2 - s_k^2, \tag{69}$$

for $k = 1,\dots,n$. Notice that $\mathbb{E}\, V_k = 0$ for all k and

$$\operatorname{var} \textstyle\sum_{k=1}^n V_k \;=\; \sum_{k=1}^n \mathbb{E}\, V_k^2 \;\leq\; \sum_{k=1}^n \mathbb{E}\, U_k^4 \;\leq\; \sum_{k=1}^n \mathbb{E}\, |U_k|^3 \;=\; A_3. \tag{70}$$

Set

$$\tilde{\tilde{\mathbb{T}}} \;:=\; \left(\textstyle\sum_{j=1}^n c_j\, U_j\right) g\left(1 + \sum_{k=1}^n V_k + R(U)\right).$$

By Chebychev's inequality and (70),

$$\mathbb{P}\left(\left|\textstyle\sum_{k=1}^n V_k\right| \geq \tfrac{1}{4}\right) \;\leq\; 16 \operatorname{var} \textstyle\sum_{k=1}^n V_k \;\ll\; A_3, \tag{71}$$

whereas by assumption $\mathbb{P}(|R(U)| \geq \frac{1}{4}) \leq 4\, \mathbb{E}\, |R(U)| \ll (\mathbb{E}\, |R(U)|^{3/2})^{2/3} \ll A_3^2$. As in (66), thus

$$\mathbb{P}(x \leq \tilde{\mathbb{T}} \leq x+\lambda) \;\ll\; \mathbb{P}(x \leq \tilde{\tilde{\mathbb{T}}} \leq x+\lambda) + A_3 \tag{72}$$

for all $x \in \mathbb{R}$, $\lambda \geq 0$, so that

$$Q(\tilde{\mathbb{T}},\lambda) \;\ll\; Q(\tilde{\tilde{\mathbb{T}}},\lambda) + A_3, \tag{73}$$

and it makes no difference for the theorem whether we prove (65) for $\tilde{\tilde{\mathbb{T}}}$ or $\tilde{\mathbb{T}}$.

For any $C \subset N$ we introduce the following two expressions:

$$\eta_{(C)} \;:=\; \textstyle\sum_{j \notin C} c_j\, U_j \quad \text{and} \quad \xi_{(C)} \;:=\; \sum_{j \notin C} V_j. \tag{74}$$

Moreover, let

$$\eta \;:=\; \sum_{j=1}^{n} c_j\, U_j \quad \text{and} \quad \xi \;:=\; \sum_{j=1}^{n} V_j. \tag{75}$$

As in (70) we see that $\mathbb{E}\xi_{(C)}^2 \le A_3$ for any $C \subset N$. As concerns the moments of $\eta_{(C)}$ we have the following:

Lemma 3.3. *Assume that $s^2 \neq 0$, $M_j^2 \le s^2$ for all j and $\sum_{j=1}^{n} c_j^2 = n$. Assume furthermore that either the sample is i.i.d. or $c_j = 1$ for all j, and that $A_2 \le 1$. Then, for any $C \subset N$,*

$$\mathbb{E}\,|\sum_{j\in C} c_j\, U_j|^2 \;\ll\; 1, \quad \mathbb{E}\,|\sum_{j\in C} c_j\, U_j|^3 \;\ll\; \nu_3. \tag{76}$$

Proof of Lemma 3.3. Let $p \ge 2$ and let $Z_1, \ldots, Z_n$ be a sequence of independent random variables with mean zero. Then Rosenthal's inequality (see Petrov (1995), Theorem 2.9) states that

$$\mathbb{E}\,|\sum_{j=1}^{n} Z_j|^p \;\ll_p\; \sum_{j=1}^{n} \mathbb{E}\,|Z_j|^p + (\sum_{j=1}^{n} \mathbb{E}\,Z_j^2)^{p/2}. \tag{77}$$

Now take any fixed $C \subset N$ and $p \in \{2, 3\}$. We take

$$Z_j \;:=\; c_j\,(U_j - \mathbb{E}\,U_j)\, I\{j \in C\},$$

for $1 \le j \le n$. Using the C_r-inequality (see, for example, Loève (1977), p.157) together with (77), we see that

$$\begin{aligned}
\mathbb{E}\,|\sum_{j\in C} c_j\, U_j|^p \;&\ll\; \mathbb{E}\,|\sum_{j\in C} c_j\,(U_j - \mathbb{E}\,U_j)|^p + \mathbb{E}\,|\sum_{j\in C} c_j\,\mathbb{E}\,U_j\,|^p \\
&=\; \mathbb{E}\,|\sum_{j=1}^{n} Z_j|^p + |\sum_{j\in C} c_j\,\mathbb{E}\,U_j\,|^p \\
&\ll\; \sum_{j=1}^{n} \mathbb{E}\,|Z_j|^p + \big(\sum_{j=1}^{n} \mathbb{E}\,Z_j^2\big)^{p/2} + \big(\sum_{j\in C} |c_j|\,|\mathbb{E}\,U_j\,|\big)^p.
\end{aligned} \tag{78}$$

Using Lyapunov's inequality and the fact that $|U_j| \le 1$ we see that $|\mathbb{E}\,U_j\,|^p \le \mathbb{E}\,|U_j\,|^p \le \mathbb{E}\,U_j^2$ for all j, and a fortiori

$$\mathbb{E}\,|Z_j\,|^p \;\ll\; |c_j|^p\,(\mathbb{E}\,|U_j\,|^p + |\mathbb{E}\,U_j\,|^p) \;\ll\; |c_j|^p\,\mathbb{E}\,U_j^2.$$

Of course $\mathbb{E}\,Z_j^2 \le c_j^2\,\mathbb{E}\,U_j^2$. In case $c_j = 1$ for all j, (78) leads to

$$\mathbb{E}\,|\sum_{j\in C} c_j\, U_j|^p \;\ll\; \sum_{j=1}^{n} s_j^2 + \big(\sum_{j=1}^{n} s_j^2\big)^{p/2} + A_2^p \;\ll\; 1.$$

In the case that the sample is i.i.d. we have that $s_j^2 = n^{-1}$ and $|\mathbb{E}\,U_j\,| = n^{-1}A_2$, so that

$$\mathbb{E}\,|\sum_{j\in C} c_j\, U_j|^p \;\ll\; n^{-1}\,\sum_{j=1}^{n} |c_j|^p + (n^{-1}\,\sum_{j=1}^{n} c_j^2)^{p/2} + (A_2\, n^{-1}\,\sum_{j=1}^{n} |c_j|)^p \;\ll\; \nu_p.$$

This concludes the proof. □

Next we get rid of $R(U)$. To this, let

$$\hat{\mathbb{T}} \;:=\; (\sum_{j=1}^{n} c_j\, U_j)\, g(1 + \sum_{k=1}^{n} V_k) \;=\; \eta\, g(1 + \xi).$$

Using Markov's and Hoeffding's inequality, (76) and the assumption on $\mathbb{E}\,|R(U)|^{3/2}$, we see that

$$\begin{aligned}\mathbb{P}(|\hat{\mathbb{T}} - \tilde{\tilde{\mathbb{T}}}| > A_3) \;&\le\; A_3^{-1}\,\mathbb{E}\,|\hat{\mathbb{T}} - \tilde{\tilde{\mathbb{T}}}| \;\le\; A_3^{-1}\,\|g'\|_\infty\,\mathbb{E}\,|\eta|\,|R(U)|\\ &\ll_g\; A_3^{-1}\,\{\mathbb{E}\,|\eta|^3\}^{1/3}\,\{\mathbb{E}\,|R(U)|^{3/2}\}^{2/3} \;\ll\; \nu_3\,A_3^{-1}\,A_3^2 \;=\; \nu_3\,A_3.\end{aligned}$$

As a result, as in (66),

$$\begin{aligned}\mathbb{P}(x \le \tilde{\tilde{\mathbb{T}}} \le x+\lambda) \;&\le\; \mathbb{P}(x \le \tilde{\tilde{\mathbb{T}}} \le x+\lambda,\, |\hat{\mathbb{T}} - \tilde{\tilde{\mathbb{T}}}| \le A_3) + \mathbb{P}(|\hat{\mathbb{T}} - \tilde{\tilde{\mathbb{T}}}| > A_3)\\ &\le\; \mathbb{P}(x - A_3 \le \hat{\mathbb{T}} \le x + A_3 + \lambda) + \nu_3\,A_3,\end{aligned} \tag{79}$$

for any $x \in \mathbb{R}$ and $\lambda \ge 0$, and it is clear that

$$Q(\tilde{\tilde{\mathbb{T}}}, \lambda) \;\le\; Q(\hat{\mathbb{T}}, \lambda + 2A_3) + \nu_3\,A_3, \tag{80}$$

which reduces the problem to the estimation of $Q(\hat{\mathbb{T}}, \lambda)$.

Now let $\hat{\sigma}^2 := \operatorname{var}\hat{\mathbb{T}}$. First we will show that $\hat{\sigma}^2$ is approximately 1, which will reduce the problem to the estimation of $Q(\hat{\mathbb{T}}/\hat{\sigma}, \lambda)$. After that we derive bounds for $\beta(\hat{\mathbb{T}})$ and $\Delta^2(\hat{\mathbb{T}})$ and apply Theorem 1.2.

Lemma 3.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a two times differentiable function with bounded derivatives. Then, for any $x, y, z \in \mathbb{R}$, $|f(x+y) - f(x)| \le \|f'\|_\infty\,|y|$ and*

$$|f(x+y+z) - f(x+y) - f(x+z) + f(x)| \;\le\; \|f''\|_\infty\,|y|\,|z|.$$

In general: if f is k times differentiable with a bounded k^{th} derivative, and $x, y_1, \dots, y_k \in \mathbb{R}$, then

$$\big|\textstyle\sum_{C\subset\{1,\dots,k\}}(-1)^{|C|}\,f(x + \sum_{j\in\{1,\dots,k\}\setminus C} y_j)\big| \;\le\; \|f^{(k)}\|_\infty\,|y_1|\cdots|y_k|.$$

Proof of Lemma 3.4. For any $1 \le l \le k$ and $u \in \mathbb{R}$, we write

$$h(l,u) \;:=\; \textstyle\sum_{C\subset\{1,\dots,l\}}(-1)^{|C|}\,f(u + \sum_{j\in\{1,\dots,l\}\setminus C} y_j),$$

and we take $\tau_1, \dots, \tau_k$ independently and uniformly distributed on (0,1). All the results are simple implications of the following equality:

$$h(l,u) \;=\; \mathbb{E}\,f^{(l)}(u + \textstyle\sum_{j=1}^{l}\tau_j\,y_j)\,y_1\cdots y_l. \tag{81}$$

The proof of (81) is given using induction. For $l = 1$ we have that

$$h(1,u) \;=\; f(u+y_1) - f(u) \;=\; \mathbb{E}\,f'(u + \tau_1\,y_1)\,y_1$$

following (23), and indeed (81) is correct. Now suppose that (81) applies for all $l \le m-1$ and $u \in \mathbb{R}$. Distinguishing between the sets $C \subset \{1,\dots,m\}$ that contain m and the ones that do not, we see that

$$h(m,u) \;=\; h(m-1,u) - h(m-1, u+y_m).$$

As a consequence of the inductional hypothesis then

$$h(m,u) \;=\; \mathbb{E}\{f^{(m-1)}(u+\textstyle\sum_{j=1}^{m-1}\tau_j\,y_j) - f^{(m-1)}(u+y_m+\sum_{j=1}^{m-1}\tau_j\,y_j)\}\,y_1\cdots y_{m-1}$$

and, again using (23), indeed (81) for $l=m$. This concludes the proof. $\square$

Lemma 3.5. *Assume that $A \le 1$. There exist absolute constants $\tilde{c}_1$ and $\tilde{c}_2$ such that*

$$|\mathbb{E}\,\hat{\mathbb{T}}| \le \tilde{c}_1\,A, \quad |\mathbb{E}\,\hat{\mathbb{T}}^2-1| \le \tilde{c}_2\,A^{1/2}.$$

Proof of Lemma 3.5. First we take a look at $\mathbb{E}\,\hat{\mathbb{T}} = \sum_{j=1}^n c_j\,\mathbb{E}\,U_j\,g(1+\xi)$. Using Lyapunov's inequality, we have that in general

$$|\mathbb{E}\,U_j\,\{g(1+\xi)-g(1+\xi_{(j)})\}| \;\le\; \|g'\|_\infty\,\mathbb{E}\,|U_j|\,|V_j| \;\ll_g\; \mathbb{E}\,|U_j|^3 + s_j^2\,\mathbb{E}\,|U_j| \;\ll\; \mathbb{E}\,|U_j|^3,$$

whereas $|\mathbb{E}\,U_j\,g(1+\xi_{(j)})| = |(\mathbb{E}\,U_j)\,\mathbb{E}\,g(1+\xi_{(j)})| \ll_g |\mathbb{E}\,U_j|$, so that

$$|\mathbb{E}\,U_j\,g(1+\xi)| \;\ll_g\; |\mathbb{E}\,U_j| + \mathbb{E}\,|U_j|^3.$$

In case $c_j=1$ for all j this leads to $|\mathbb{E}\,\hat{\mathbb{T}}| \ll_g A_2+A_3$, which is what we are looking for. In the case that the sample is i.i.d. it follows that

$$|\mathbb{E}\,\hat{\mathbb{T}}| \;=\; |\mathbb{E}\,U_1\,g(1+\xi)|\,|\textstyle\sum_{j=1}^n c_j| \;\ll_g\; n\,\nu_1\,(|\mathbb{E}\,U_1| + \mathbb{E}\,|U_1|^3) \;\le\; A_2+A_3$$

as well, and this proves the first point.

For $\mathbb{E}\,\hat{\mathbb{T}}^2$ we first look at the case in which $c_j=1$ for all j. In this case we have that

$$\mathbb{E}\,\hat{\mathbb{T}}^2 - 1 \;=\; \textstyle\sum_{j=1}^n \mathbb{E}\,U_j^2\,\{g^2(1+\xi)-1\} + \sum_{j\neq l}\mathbb{E}\,U_j U_l\,g^2(1+\xi) \;=:\; d_1+d_2,$$

with $|\mathbb{E}\,\hat{\mathbb{T}}^2-1| \le |d_1|+|d_2|$. As to d_2, writing $d_2^{jl} := \mathbb{E}\,U_j U_l\,g^2(1+\xi)$, we have the following:

$$\begin{aligned}
|d_2^{jl}| \;\le\;& |\mathbb{E}\,U_j U_l\,\{g^2(1+\xi)-g^2(1+\xi_{(j)})-g^2(1+\xi_{(l)})+g^2(1+\xi_{(jl)})\}| \\
&+ |\mathbb{E}\,U_j U_l\,\{g^2(1+\xi_{(j)})-g^2(1+\xi_{(jl)})\}| + |\mathbb{E}\,U_j U_l\,\{g^2(1+\xi_{(l)})-g^2(1+\xi_{(jl)})\}| \\
&+ |\mathbb{E}\,U_j U_l\,g^2(1+\xi_{(jl)})| \;=:\; d_{21}^{jl}+d_{22}^{jl}+d_{23}^{jl}+d_{24}^{jl}.
\end{aligned}$$

Clearly $d_{24}^{jl} = |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|\,|\mathbb{E}\,g^2(1+\xi_{(jl)})| \ll_g |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|$. Using Lemma 3.4,

$$d_{22}^{jl} \;\le\; \|(g^2)'\|_\infty\,|\mathbb{E}\,U_j|\,\mathbb{E}\,|U_l|\,|V_l| \;\ll_g\; |\mathbb{E}\,U_j|\,\mathbb{E}\,|U_l|^3,$$

similarly $d_{23}^{jl} \ll_g |\mathbb{E}\,U_l|\,\mathbb{E}\,|U_j|^3$, and moreover

$$d_{21}^{jl} \;\le\; \|(g^2)''\|_\infty\,\mathbb{E}\,|U_j U_l|\,|V_j V_l| \;\ll_g\; \mathbb{E}\,|U_j|^3\,\mathbb{E}\,|U_l|^3.$$

As a result

$$\begin{aligned}
|d_2| \;\ll_g\;& \textstyle\sum_{j\neq l}\big(\mathbb{E}\,|U_j|^3\,\mathbb{E}\,|U_l|^3 + |\mathbb{E}\,U_j|\,\mathbb{E}\,|U_l|^3 + |\mathbb{E}\,U_l|\,\mathbb{E}\,|U_j|^3 + |\mathbb{E}\,U_j|\,|\mathbb{E}\,U_l|\big) \\
\le\;& \big(\textstyle\sum_{j=1}^n \mathbb{E}\,|U_j|^3\big)^2 + 2\big(\sum_{j=1}^n |\mathbb{E}\,U_j|\big)\big(\sum_{l=1}^n \mathbb{E}\,|U_l|^3\big) + \big(\sum_{j=1}^n |\mathbb{E}\,U_j|\big)^2 \\
=\;& A_3^2 + 2A_2\,A_3 + A_2^2 \;\le\; 4A^2
\end{aligned}$$

and $|d_2| \ll_g A^2$. As to d_1, writing $d_1^j := \mathbb{E}\,U_j^2\,\{g^2(1+\xi)-1\}$, we have that

$$\begin{aligned}|d_1^j| \;&\leq\; \left|\mathbb{E}\,U_j^2\,\{g^2(1+\xi)-g^2(1+\xi_{(j)})\}\right| \\ &\qquad + (\mathbb{E}\,U_j^2)\left|\mathbb{E}\left(g^2(1+\xi_{(j)})-g^2(1)\right)\right| \;=:\; d_{11}^j + d_{12}^j.\end{aligned}$$

With Lemma 3.4 we have that

$$d_{12}^j \;\leq\; s_j^2\,\|(g^2)'\|_\infty\,\mathbb{E}\,|\xi_{(j)}| \;\ll_g\; s_j^2\,\{\mathbb{E}\,\xi_{(j)}^2\}^{1/2} \;\leq\; s_j^2\,A_3^{1/2}$$

(see the remark behind (75)), and that

$$d_{11}^j \;\leq\; \|(g^2)'\|_\infty\,\mathbb{E}\,U_j^2\,|V_j| \;\ll_g\; \mathbb{E}\,|U_j|^3.$$

In turn

$$|d_1| \;\ll_g\; \textstyle\sum_{j=1}^n(\mathbb{E}\,|U_j|^3 + s_j^2\,A_3^{1/2}) \;=\; A_3 + A_3^{1/2} \;\ll\; A_3^{1/2},$$

and as a consequence $|\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \ll_g A^{1/2}$.

Finally we consider $\mathbb{E}\,\hat{\mathbb{T}}^2$ in the case of an i.i.d. sample. Going about in the same way,

$$\begin{aligned}\mathbb{E}\,\hat{\mathbb{T}}^2 - 1 \;&=\; \left(\textstyle\sum_{j=1}^n c_j^2\,\mathbb{E}\,U_j^2\,g^2(1+\xi)\right) - 1 + \textstyle\sum_{j\neq l} c_j c_l\,\mathbb{E}\,U_j U_l\,g^2(1+\xi) \\ &=\; n\,\mathbb{E}\,U_1^2\,(g^2(1+\xi)-1) + \mathbb{E}\,U_1 U_2\,g^2(1+\xi)\,\textstyle\sum_{j\neq l} c_j c_l \;=:\; f_1 + f_2.\end{aligned}$$

As to f_2, using the analysis of d_2^{jl} above,

$$|\mathbb{E}\,U_1 U_2\,g^2(1+\xi)| \;\ll_g\; \left(|\mathbb{E}\,U_1| + \mathbb{E}\,|U_1|^3\right)^2 \;\leq\; n^{-2}\,(A_2+A_3)^2,$$

whereas $|\sum_{j\neq l} c_j c_l| \leq (\sum_{j=1}^n |c_j|)^2 = n^2\,\nu_1^2 \leq n^2$, and $|f_2| \ll_g A^2$. The analysis of d_1^j above is showing us that $\mathbb{E}\,U_1^2\,(g^2(1+\xi)-1) \leq \mathbb{E}\,|U_1|^3 + n^{-1}\,A_3^{1/2} \ll n^{-1}\,A_3^{1/2}$, so that $|f_1| \ll_g A_3^{1/2}$, and as a consequence again $|\mathbb{E}\,\hat{\mathbb{T}}^2 - 1| \ll_g A^{1/2}$. This finishes the proof. □

Taking $\tilde{c}_1$ and $\tilde{c}_2$ as in Lemma 3.5, we take the constant c such that

$$1 - \tilde{c}_2\,c^{-1/2} - \tilde{c}_1^2\,c^{-2} \;\geq\; \tfrac{1}{2}.$$

In this way

$$\hat{\sigma}^2 \;=\; 1 + (\mathbb{E}\,\hat{\mathbb{T}}^2 - 1) - (\mathbb{E}\,\hat{\mathbb{T}})^2 \;\geq\; 1 - \tilde{c}_2\,A^{1/2} - \tilde{c}_1^2\,A^2 \;\geq\; \tfrac{1}{2}.$$

Using Theorem 1.2 we see that

$$\begin{aligned}Q(\hat{\mathbb{T}},\lambda) \;=\; Q(\hat{\mathbb{T}}/\hat{\sigma},\lambda/\hat{\sigma}) \;&\ll\; \max\{\lambda\,\hat{\sigma}^{-1},\,\beta(\hat{\mathbb{T}})\,\hat{\sigma}^{-3},\,\Delta^2(\hat{\mathbb{T}})\,\hat{\sigma}^{-2}\} \\ &\ll\; \max\{\lambda,\,\beta(\hat{\mathbb{T}}),\,\Delta^2(\hat{\mathbb{T}})\},\end{aligned} \qquad (82)$$

and we turn to finding bounds for $\beta(\hat{\mathbb{T}})$ and $\Delta^2(\hat{\mathbb{T}})$.

Lemma 3.6. *Assume that* $A_3 \le c^{-1}$. *We have:*

$$\beta(\hat{\mathbb{T}}) \ll_g \nu_3\, A_3, \quad \Delta^2(\hat{\mathbb{T}}) \ll_g A_3. \tag{83}$$

Proof of Lemma 3.6. Let $W := (W_1, \dots, W_n)$ be an independent copy of U (that is to say, U and W are independent and $U \overset{d}{=} W$), and let $Z_j := W_j^2 - s_j^2$, for $1 \le j \le n$.

First we look at $\beta(\hat{\mathbb{T}})$. Notice that with (11) we have that

$$\beta(\hat{\mathbb{T}}) \;\le\; \textstyle\sum_{j=1}^n \mathbb{E}\,|D_j \hat{\mathbb{T}}|^3. \tag{84}$$

Now take any $1 \le j \le n$ and write $S_1 := c_j\, U_j\, g(1+\xi)$, $S_2 := \eta_{(j)}\, g(1+\xi)$. It is clear that $\hat{\mathbb{T}} = S_1 + S_2$, and using (8) and (9) we see that

$$D_j \hat{\mathbb{T}} \;=\; D_j S_1 + D_j S_2 \;=\; c_j\, D_j U_j\, g(1 + V_j + \xi_{(j)}) + \eta_{(j)}\, D_j g(1 + V_j + \xi_{(j)}),$$

so that with the C_r-inequality

$$|D_j \hat{\mathbb{T}}|^3 \;\ll\; |c_j|^3\, |D_j U_j\, g(1 + V_j + \xi_{(j)})|^3 + |\eta_{(j)}|^3\, |D_j g(1 + V_j + \xi_{(j)})|^3. \tag{85}$$

As to the first term on the right side of (85),

$$\begin{aligned}
|D_j U_j\, g(1 + V_j + \xi_{(j)})| \;&=\; |\mathbb{E}\,(U_j\, g(1 + V_j + \xi_{(j)}) - W_j\, g(1 + Z_j + \xi_{(j)}) \,|\, U)| \\
&\le\; |\mathbb{E}\,(U_j\, \{g(1 + V_j + \xi_{(j)}) - g(1 + Z_j + \xi_{(j)})\} \,|\, U)| \\
&\qquad + |\mathbb{E}\,(g(1 + Z_j + \xi_{(j)})\, \{U_j - W_j\} \,|\, U)| \\
&\le\; |U_j|\, \|g'\|_\infty\, \mathbb{E}\,(|V_j - Z_j| \,|\, V_j) + \|g\|_\infty\, \mathbb{E}\,(|U_j - W_j| \,|\, U_j)| \\
&\ll_g\; |U_j|\,(|V_j| + \mathbb{E}\,|V_j|) + |U_j| + \mathbb{E}\,|U_j| \;\ll\; |U_j| + s_j.
\end{aligned}$$

As to the second one,

$$\begin{aligned}
|D_j g(1 + V_j + \xi_{(j)})| \;&=\; |\mathbb{E}\,(g(1 + V_j + \xi_{(j)}) - g(1 + Z_j + \xi_{(j)}) \,|\, U)| \\
&\le\; \|g'\|_\infty\, \mathbb{E}\,(|V_j - Z_j| \,|\, U_j) \;\ll_g\; U_j^2 + s_j^2 \;\ll\; |U_j| + s_j.
\end{aligned} \tag{86}$$

As a result

$$|D_j \hat{\mathbb{T}}|^3 \;\ll_g\; (|c_j|^3 + |\eta_{(j)}|^3)\,(|U_j|^3 + s_j^3),$$

and, using in turn (84), Lemma 3.3 and Lyapunov's inequality,

$$\beta(\hat{\mathbb{T}}) \;\ll_g\; \textstyle\sum_{j=1}^n (|c_j|^3 + \nu_3)\,(\mathbb{E}\,|U_j|^3 + s_j^3).$$

In case all $c_j = 1$ this means that

$$\beta(\hat{\mathbb{T}}) \;\ll_g\; \textstyle\sum_{j=1}^n \mathbb{E}\,|U_j|^3 = A_3 = \nu_3\, A_3,$$

and in case of an i.i.d. sample

$$\beta(\hat{\mathbb{T}}) \;\ll_g\; (\mathbb{E}\,|U_1|^3 + n^{-3/2})\, n\, \nu_3 \;\ll\; \nu_3\, A_3$$

as well, which proves (83) for $\beta(\hat{\mathbb{T}})$.

We turn to

$$\Delta^2(\hat{\mathbb{T}}) \;=\; \textstyle\sum_{1\le j<k\le n} \mathbb{E}\,|D_{jk}\hat{\mathbb{T}}|^2.$$

We take any combination $1 \le j < k \le n$ and write

$$S_1 := c_j\, U_j\, g(1+\xi), \quad S_2 := c_k\, U_k\, g(1+\xi) \quad \text{and} \quad S_3 := \eta_{(jk)}\, g(1+\xi),$$

so that

$$D_{jk}\hat{\mathbb{T}} \;=\; D_{jk}S_1 + D_{jk}S_2 + D_{jk}S_3 \;=\; \textstyle\sum_{l=j,k} c_l\, D_{jk}U_l\, g(1+\xi) + \eta_{(jk)}\, D_{jk}g(1+\xi).$$

Taking $l = j$, we see that

$$\begin{aligned} &|D_{jk}U_j\, g(1+\xi)| \;=\; |D_jU_j\, D_kg(1+\xi)| \\ &\quad=\; |\mathbb{E}\,(U_j\, D_kg(1+V_j+\xi_{(j)}) - W_j\, D_kg(1+Z_j+\xi_{(j)})\,|\,U)| \\ &\quad\le\; |U_j|\,|D_kg(1+V_j+V_k+\xi_{(jk)})| \;+\; \mathbb{E}\,(|W_j|\,|D_kg(1+Z_j+V_k+\xi_{(jk)})|\,|\,U), \end{aligned} \tag{87}$$

and since, as in (86), $|D_kg(1+X+V_k+\xi_{(jk)})| \ll_g U_k^2 + s_k^2 \le |U_k|^{3/2} + s_k^{3/2}$ for $X = V_j, Z_j$, we see that

$$|D_{jk}S_1| \;\ll_g\; |c_j|\,(|U_j| + s_j)\,(|U_k|^{3/2} + s_k^{3/2}).$$

Of course we have the similar bound $|D_{jk}S_2| \ll_g |c_k|\,(|U_k| + s_k)\,(|U_j|^{3/2} + s_j^{3/2})$. Using Lemma 3.4 we see that

$$\begin{aligned} |D_{jk}g(1+\xi)| \;&=\; |\mathbb{E}\,(g(1+V_j+V_k+\xi_{(jk)}) - g(1+V_j+Z_k+\xi_{(jk)}) \\ &\qquad - g(1+Z_j+V_k+\xi_{(jk)}) + g(1+Z_j+Z_k+\xi_{(jk)}) \mid U)| \\ &\le\; \|g''\|_\infty\, \mathbb{E}\,(|V_j - Z_j|\,|V_k - Z_k|\,|\,U) \\ &\ll_g\; (U_j^2 + s_j^2)\,(U_k^2 + s_k^2) \;\le\; (|U_j|^{3/2} + s_j^{3/2})\,(|U_k|^{3/2} + s_k^{3/2}). \end{aligned} \tag{88}$$

In conclusion,

$$\begin{aligned} \Delta^2(\hat{\mathbb{T}}) \;&\ll_g\; \textstyle\sum_{1\le j<k\le n} \big(c_j^2\, s_j^2\, \mathbb{E}\,|U_k|^3 + c_k^2\, s_k^2\, \mathbb{E}\,|U_j|^3 + \mathbb{E}\,|\eta_{(jk)}|^2\, \mathbb{E}\,|U_j|^3\, \mathbb{E}\,|U_k|^3\big) \\ &\ll\; \big(\textstyle\sum_{j=1}^n c_j^2\, s_j^2\big)\,\big(\textstyle\sum_{k=1}^n \mathbb{E}\,|U_k|^3\big) + \big(\textstyle\sum_{k=1}^n \mathbb{E}\,|U_k|^3\big)^2 \;\ll\; A_3, \end{aligned}$$

using the fact that in both of our cases $\sum_{j=1}^n c_j^2\, s_j^2 = 1$. Thus (83) for $\Delta^2(\hat{\mathbb{T}})$, and this finishes the proof. □

Using in turn (67), (73), (80), (82) and Lemma 3.6 we see that indeed (65), which finishes the proof of Theorem 3.2.

3.3 Student's statistic

Our first application of Theorem 3.2 will be on Student's statistic, in a one-sample and a two-sample setting.

Here and in the next subsection we will need the following lemma, which, together with its proof, can be found as Lemma 1.3 in Bentkus, Götze (1996).

Lemma 3.7. *Let α be any real-valued random variable, with ${\sigma_\alpha}^2 := \mathbb{E}\,\alpha^2$ (which may be ∞). For any integer k, there exists a largest solution $q_0^2 = q_0^2(\alpha, k)$ of the following equation in q^2:*

$$q^2 \;=\; \mathbb{E}\,\alpha^2\, I\{\alpha^2 \le q^2\, k\}. \tag{89}$$

We have that $q_0^2 \le {\sigma_\alpha}^2$, and in case $q_0^2 \le \frac{1}{2}{\sigma_\alpha}^2$ we have that $\mathbb{E}\,\alpha^2\, I\{\alpha^2 \le \frac{1}{2}{\sigma_\alpha}^2\, k\} \le \frac{1}{2}{\sigma_\alpha}^2$.

We first look at the classical Student statistic. To this, let $X_1, \ldots, X_n$ be an independent, identically distributed sample, with $\mathbb{E}\,X_1 = 0$ and $0 < \sigma^2 := \mathbb{E}\,X_1^2 < \infty$. We look at selfnormalized sums

$$\mathbb{T} \;:=\; \frac{X_1 + \ldots + X_n}{\big(X_1^2 + \ldots + X_n^2\big)^{1/2}},$$

or, alternatively, at Student's statistic

$$\mathbb{T}_0 \;:=\; \frac{X_1 + \ldots + X_n}{\big\{\sum_{k=1}^n (X_k - \overline{X})^2\big\}^{1/2}} \;=\; \frac{X_1 + \ldots + X_n}{\big(\sum_{k=1}^n X_k^2 - n\,\overline{X}^2\big)^{1/2}},$$

with $\overline{X}$ denoting the sample mean. In the case of a normal distribution, that is, if the X_j are $N(0, \sigma^2)$-distributed, the statistic $(1 - n^{-1})^{1/2}\,\mathbb{T}_0$ is t-distributed with $n-1$ degrees of freedom, that is to say, almost $N(0,1)$-distributed.

Now let q^2 be the $q_0^2(\alpha, k)$ which we obtain from Lemma 3.7 when taking $\alpha = X_1$ and $k = n$, and let

$$B \;:=\; n^{-1/2}\, q^{-3}\, \mathbb{E}\,|X_1|^3\, I\{X_1^2 \le q^2\, n\} + n^{1/2}\, q^{-1}\, \mathbb{E}\,|X_1|\, I\{X_1^2 > q^2\, n\}.$$

We have the following result:

Corollary 3.8. *Suppose that the sample $X_1, \ldots, X_n$ is i.i.d., $\mathbb{E}\,X_1 = 0$ and $q^2 \ne 0$. Then there exists a constant c such that, for Student's statistic $\mathbb{T}$,*

$$Q(\mathbb{T}, \lambda) \;\le\; c\, \max\{\lambda,\, B\}. \tag{90}$$

In case that the sample is i.i.d., $\mathbb{E}\,X_1 = 0$ and $0 < \sigma^2 < \infty$, this leads us to the following bound:

$$Q(\mathbb{T}, \lambda) \;\le\; c\, \max\{\lambda\,,\, n^{-1/2}\, \sigma^{-3}\, \mathbb{E}\,|X_1|^3\}. \tag{91}$$

The same results hold for $\mathbb{T}_0$.

Notice that the bound achieved in (90) in comparison with the bound (91) has the nice feature that the it is already finite if $\mathbb{E}X_1$ exists.

Next we look at the so-called two-sample Student statistic. To this, let $Y_1, \dots, Y_{n_1}$ and $Z_1, \dots, Z_{n_2}$ be two independent, i.i.d. samples of real-valued variables, for which

$$\mu_1 := \mathbb{E}\,Y_1,\ \mu_2 := \mathbb{E}\,Z_1,\ \sigma_1^2 := \operatorname{var} Y_1 \quad \text{and} \quad \sigma_2^2 := \operatorname{var} Z_1.$$

For any sample $W_1, \dots, W_m$, let

$$\overline{W} := m^{-1} \textstyle\sum_{j=1}^m W_j \quad \text{and} \quad S_W^2 := m^{-1} \textstyle\sum_{j=1}^m (W_j - \overline{W})^2.$$

In cases where one is interested whether the two populations have the same mean, the two-sample Student statistic

$$S \ := \ \frac{\overline{Y} - \overline{Z}}{\sqrt{n_1^{-1} S_Y^2 + n_2^{-1} S_Z^2}}$$

is a natural test statistic. We take q_1^2 and q_2^2 as the $q_0^2(\alpha, k)$ which we obtain from Lemma 3.7 when taking $(\alpha, k) = (Y_1, n_1)$ and $(\alpha, k) = (Z_1, n_2)$ respectively. Writing $V_1 := Y_1 - \mu_1$, $V_2 := Z_1 - \mu_2$, and

$$B_k \ = \ B_k(n_k) \ := \ n_k^{-1/2}\, q_k^{-3}\, \mathbb{E}\,|V_k|^3\, I\{V_k^2 \le q_k^2\, n_k\} + n_k^{1/2}\, q_k^{-1}\, \mathbb{E}\,|V_k|\, I\{V_k^2 > q_k^2\, n_k\},$$

for $k = 1, 2$, we have the following result:

Corollary 3.9. *Assume that* $\mu_1 = \mu_2$ *and that* $q_1^2, q_2^2 \neq 0$. *Then for the two-sample Student statistic* S *there exists a constant* c *such that*

$$Q(S, \lambda) \ \le \ c\, \max\{\lambda,\, B_1,\, B_2\}. \tag{92}$$

In the case that $0 < \sigma_1^2,\, \sigma_2^2 < \infty$, *taking* $\mu := \mu_1 = \mu_2$, *this means that a constant* c *exists such that*

$$Q(S, \lambda) \ \le \ c\, \max\{\lambda\,,\, n_1^{-1/2}\, \sigma_1^{-3}\, \mathbb{E}\,|Y_1 - \mu|^3\,,\, n_2^{-1/2}\, \sigma_2^{-3}\, \mathbb{E}\,|Z_1 - \mu|^3\}. \tag{93}$$

We turn to the proofs of the Corollaries 3.8 and 3.9.

Proof of Corollary 3.8. We want to apply Theorem 3.2, in the case that $c_j = 1$ for all j. Note that under this condition $\mathbb{T}$ and $\mathbb{T}_0$ are of the form (63), with $R(X) = 0$ and $R(X) = -n\,\overline{X}^2$ respectively. We define $M_j^2 := q^2\, n$, for $1 \le j \le n$. In this way

$$s^2 \ = \ n\,\mathbb{E}\,X_1^2\, I\{X_1^2 \le q^2\, n\} \ = \ n\, q^2 \tag{94}$$

(see Lemma 3.7), so that indeed $M_j^2 \le s^2$ for all j. Note that as well by assumption $s^2 = n\,q^2 > 0$, so Theorem 3.2 is indeed applicable. As to A_1:

$$A_1 \;=\; n\,\mathbb{P}(X_1^2 \ge q^2\,n) \;\le\; n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\} \;\le\; B. \tag{95}$$

As $\mathbb{E}\,X_1\,I\{X_1^2 \le q^2\,n\} = -\mathbb{E}\,X_1\,I\{X_1^2 > q^2\,n\}$ (since $\mathbb{E}\,X_1 = 0$), furthermore

$$A_2 \;=\; n\,s^{-1}\,|\mathbb{E}\,X_1\,I\{X_1^2 > q^2\,n\}| \;\le\; n^{1/2}\,q^{-1}\,\mathbb{E}\,|X_1|\,I\{X_1^2 > q^2\,n\} \;\le\; B, \tag{96}$$

and in the same way $A_3 \le n^{-1/2}\,q^{-3}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 \le q^2\,n\} \le B$. Finally note that in the case that $R(X) = n\,\overline{X}^2$, using Lemma 3.3 and (64),

$$\mathbb{E}\,|R(U)|^{3/2} \;=\; n^{-3/2}\,\mathbb{E}\,|\eta|^3 \;\ll\; n^{-3/2} \;\le\; A_3^3$$

as required, and (90) is an easy consequence.

We turn to the proof of (91). Here without loss of generality we may assume that $q^2 > \frac{1}{2}\sigma^2$. Indeed, in the opposite case Lemma 3.7 is telling us that $\mathbb{E}\,X_1^2\,I\{X_1^2 > \frac{1}{2}\sigma^2\,n\} \ge \frac{1}{2}\sigma^2$, so that

$$\begin{aligned} 1 \;&\le\; (\tfrac{1}{2}\sigma^2)^{-1}\,\mathbb{E}\,X_1^2\,I\{X_1^2 > \tfrac{1}{2}\sigma^2\,n\} \\ &\le\; n^{-1/2}\,(\tfrac{1}{2}\sigma^2)^{-3/2}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 > \tfrac{1}{2}\sigma^2\,n\} \;\le\; 2^{3/2}\,n^{-1/2}\,\sigma^{-3}\,\mathbb{E}\,|X_1|^3, \end{aligned} \tag{97}$$

which would make (91) trivially true. Under the assumption

$$\begin{aligned} B \;&\le\; n^{-1/2}\,q^{-3}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 \le q^2\,n\} + n^{1/2}\,q^{-1}\,(q^2\,n)^{-1}\,\mathbb{E}\,|X_1|^3\,I\{X_1^2 > q^2\,n\} \\ &=\; n^{-1/2}\,q^{-3}\,\mathbb{E}\,|X_1|^3 \;\le\; 2^{3/2}\,n^{-1/2}\,\sigma^{-3}\,\mathbb{E}\,|X_1|^3, \end{aligned}$$

and (91) is an obvious consequence of (90). This finishes the proof. □

Proof of Corollary 3.9. Without loss of generality we will suppose that $\mu = 0$. Indeed, if not we may apply the reduced theorem to the situation in which all Y_j and Z_j are replaced by $Y_j - \mu$ and $Z_j - \mu$, and obtain the desired result.

Let

$$X_j \;:=\; n_1^{-1}\,Y_j \quad \text{and} \quad X_{n_1+k} \;:=\; -n_2^{-1}\,Z_k,$$

for $1 \le j \le n_1$ and $1 \le k \le n_2$. Unfortunately we are not able to apply Theorem 3.2 directly to our statistic S, as the appropriate $R(X)$ is not small enough, but the theorem in fact is applicable to the similar statistic

$$\mathbb{T} \;=\; \frac{X_1 + \ldots + X_{n_1+n_2}}{\left(X_1^2 + \ldots + X_{n_1+n_2}^2\right)^{1/2}} \;=\; \frac{\overline{Y} - \overline{Z}}{\left(n_1^{-2}\,\sum_{k=1}^{n_1} Y_k^2 + n_2^{-2}\,\sum_{k=1}^{n_2} Z_k^2\right)^{1/2}}.$$

In order to apply the theorem, we first define

$$M_j^2 \;:=\; q_1^2\,n_1^{-1} \quad \text{and} \quad M_{n_1+k}^2 \;:=\; q_2^2\,n_2^{-1}$$

for $1 \le j \le n_1$ and $1 \le k \le n_2$. We write U_j, s_j^2 and so on as earlier. By construction

$$s^2 \;=\; n_1\, n_1^{-2}\, \mathbb{E}\, Y_1^2\, I\{Y_1^2 \le q_1^2\, n_1\} + n_2\, n_2^{-2}\, \mathbb{E}\, Z_1^2\, I\{Z_1^2 \le q_2^2\, n_2\} \;=\; n_1^{-1}\, q_1^2 + n_2^{-1}\, q_2^2,$$

and clearly $M_j^2 \le s^2$ for all j. As a consequence of the assumption that $q_1^2, q_2^2 \ne 0$ we have that $s^2 \ne 0$. From here on we denote

$$B \;:=\; \max\{B_1, B_2\}.$$

We first take a look at A_1, A_2 and A_3. As in (95) we have that

$$A_1 \;=\; n_1\, \mathbb{P}(Y_1^2 \ge q_1^2\, n_1) + n_2\, \mathbb{P}(Z_1^2 \ge q_2^2\, n_2) \;\le\; B_1 + B_2 \;\ll\; B.$$

As in (96),

$$A_2 = n_1\, |\mathbb{E}\, U_1| + n_2\, |\mathbb{E}\, U_{n_1+1}| = s^{-1}\, |\mathbb{E}\, Y_1\, I\{Y_1^2 > q_1^2\, n_1\}| + s^{-1}\, |\mathbb{E}\, Z_1\, I\{Z_1^2 > q_2^2\, n_2\}| \ll B,$$

and in the same way $A_3 \ll B$.

Now take

$$\overline{X}_{(1)} := n_1^{-1} \textstyle\sum_{j=1}^{n_1} X_j, \quad \overline{X}_{(2)} := n_2^{-1} \textstyle\sum_{j=n_1+1}^{n_1+n_2} X_j,$$

and as well $\overline{U}_{(1)} := n_1^{-1} \sum_{j=1}^{n_1} U_j$ and $\overline{U}_{(2)} := n_2^{-1} \sum_{j=n_1+1}^{n_1+n_2} U_j$. We have that

$$\begin{aligned} S \;&=\; \big(\textstyle\sum_{j=1}^{n_1+n_2} X_j\big) \big/ \big(\textstyle\sum_{k=1}^{n_1} (X_k - \overline{X}_{(1)})^2 + \textstyle\sum_{k=n_1+1}^{n_1+n_2} (X_k - \overline{X}_{(2)})^2\big)^{1/2} \\ &=\; \big(\textstyle\sum_{j=1}^{n_1+n_2} X_j\big) \big/ \big(\textstyle\sum_{k=1}^{n_1+n_2} X_k^2 - n_1\, \overline{X}_{(1)}^2 - n_2\, \overline{X}_{(2)}^2\big)^{1/2}. \end{aligned}$$

We set

$$\tilde{S} \;:=\; \big(\textstyle\sum_{j=1}^{n_1+n_2} U_j\big) \big/ \big(\textstyle\sum_{k=1}^{n_1+n_2} U_k^2 - n_1\, \overline{U}_{(1)}^2 - n_2\, \overline{U}_{(2)}^2\big)^{1/2}$$

and, taking g as in (68) and V_k as in (69),

$$\hat{S} \;:=\; \big(\textstyle\sum_{j=1}^{n_1+n_2} U_j\big)\, g(1 + \textstyle\sum_{k=1}^{n_1+n_2} V_k - n_1\, \overline{U}_{(1)}^2 - n_2\, \overline{U}_{(2)}^2).$$

As in (66) and (67) it is clear that

$$Q(S, \lambda) \;\le\; Q(\tilde{S}, \lambda) + A_1 \;\ll\; Q(\tilde{S}, \lambda) + B,$$

so we may concentrate on $Q(\tilde{S}, \lambda)$. Here we need the following. Because

$$\mathbb{E}\, U_1^2 \;=\; s^{-2}\, n_1^{-2}\, \mathbb{E}\, Y_1^2\, I\{Y_1^2 \le q_1^2\, n_1\} \;=\; s^{-2}\, n_1^{-2}\, q_1^2,$$

by Lyapunov's inequality $\mathbb{E}\, |U_1|^3 \ge s^{-3}\, n_1^{-3}\, q_1^3$, and as a consequence

$$n_1^{-1/2} \;\le\; s^3\, q_1^{-3}\, n_1^{5/2}\, \mathbb{E}\, |U_1|^3 \;=\; q_1^{-3}\, n_1^{-1/2}\, \mathbb{E}\, |Y_1|^3\, I\{Y_1^2 \le q_1^2\, n_1\} \;\le\; 2^{3/2}\, B_1. \qquad (98)$$

Of course as well ${n_2}^{-1/2} \le 2^{3/2}\, B_2$, and using Markov's inequality together with Lemma 3.3,

$$\begin{aligned}\mathbb{P}(n_1\, \overline{U}^2_{(1)} + n_2\, \overline{U}^2_{(2)} \ge \tfrac14) \;&\le\; 4\,(n_1^{-1}\,\mathbb{E}\,|\textstyle\sum_{j=1}^{n_1} U_j|^2 + n_2^{-1}\,\mathbb{E}\,|\sum_{j=n_1+1}^{n_1+n_2} U_j|^2) \\ &\ll\; {n_1}^{-1} + {n_2}^{-1} \;\ll\; B_1^2 + B_2^2 \;\ll\; B^2.\end{aligned}$$

Moreover, as in (71), $\mathbb{P}(|\sum_{k=1}^{n_1+n_2} V_k| \ge \frac14) \le 16 A_3 \ll B$, and as in (72) and (73) we may conclude that

$$Q(\tilde{S}, \lambda) \;\ll\; Q(\hat{S}, \lambda) + B.$$

¿From the proof of Theorem 3.2 it is clear that

$$Q(\hat{\mathbb{T}}, \lambda) \;\ll\; \max\{\lambda,\, A_2,\, A_3\} \;\ll\; \max\{\lambda, B\}.$$

Taking η as in (75), $\eta_{(1)} := \sum_{j=1}^{n_1} U_j$ and $\eta_{(2)} := \sum_{j=n_1+1}^{n_1+n_2} U_j$, using Markov's inequality and the foregoing we see that

$$\begin{aligned}&\mathbb{P}(|\hat{\mathbb{T}} - \hat{S}| > B) \;\le\; B^{-1}\,\mathbb{E}\,|\hat{\mathbb{T}} - \hat{S}| \;\le\; B^{-1}\,\|g'\|_\infty\,\mathbb{E}\,|\eta|\,({n_1}^{-1}\,\eta^2_{(1)} + {n_2}^{-1}\,\eta^2_{(2)}) \\ &\quad\ll_g\; B^{-1}\,(n_1^{-1}\,\mathbb{E}\,|\eta_{(1)}|^3 + n_1^{-1}\,\mathbb{E}\,|\eta_{(2)}|\,\mathbb{E}\,\eta^2_{(1)} + n_2^{-1}\,\mathbb{E}\,|\eta_{(1)}|\,\mathbb{E}\,\eta^2_{(2)} + n_2^{-1}\,\mathbb{E}\,|\eta_{(2)}|^3) \\ &\quad\ll\; (n_1^{-1} + {n_2}^{-1})\,B^{-1} \;\ll\; (B_1^2 + B_2^2)\,B^{-1} \;\ll\; B.\end{aligned}$$

As in (79) and (80) this leads to the fact that

$$Q(\hat{S}, \lambda) \;\ll\; Q(\tilde{\mathbb{T}}, \lambda + 2B) + B \;\ll\; \max\{\lambda, B\}.$$

This concludes the proof of (92).

We turn to the proof of (93). We may assume that

$$q_k^2 \;\ge\; \tfrac12 \sigma_k^2$$

for $k = 1, 2$. Indeed, if this is not the case then as in (97) we have that

$$1 \;\le\; 2^{3/2}\,{n_k}^{-1/2}\,\sigma_k^{-3}\,\mathbb{E}\,|V_k|^3$$

for some $k \in \{1, 2\}$, and the theorem will be trivially true. As in the proof of Corollary 3.8 this easily leads to (93). □

3.4 Linear regression

In the theory of linear regression we look at so-called response variables $Y_1, \ldots, Y_n$ which are of the form

$$Y_j \;:=\; \alpha + \beta\, x_j + E_j \quad (1 \le j \le n),$$

where the α, β, $x_1, \ldots, x_n$ are fixed real numbers, and $E_1, \ldots, E_n$ is a sequence of independent, identically distributed variables with

$$\mathbb{E}\, E_1 = 0 \quad \text{and} \quad \sigma_1^2 \;:=\; \mathbb{E}\, E_1^2 \;<\; \infty.$$

The numbers α and β are supposed to be unknown, the sequence $x_1, \ldots, x_n$ is known, and we are looking for estimators for α and β.

We assume that not $x_1 = \ldots = x_n$, write $\overline{x}$, $\overline{Y}$ and $\overline{E}$ as usual and, for $p = 1, 2, 3$, set

$$\rho_p \;:=\; n^{-1} \textstyle\sum_{j=1}^n |x_j - \overline{x}|^p.$$

With Lyapunov's inequality $\rho_1^2 \le \rho_2$ and $\rho_2^{3/2} \le \rho_3$.

The following two estimators:

$$\hat{\alpha} \;:=\; \overline{Y} - \hat{\beta}\,\overline{x} \quad \text{and} \quad \hat{\beta} \;:=\; (n\rho_2)^{-1} \textstyle\sum_{j=1}^n (x_j - \overline{x})\, Y_j,$$

present the least square estimators for α and β, in the sense that they minimize the sum of squared errors

$$\mathrm{SSE} \;:=\; \textstyle\sum_{j=1}^n (Y_j - (\hat{\alpha} + \hat{\beta}\, x_j))^2.$$

For classical results, see e.g. Pestman (1998), Chapter 4. We have for example that

$$\mathbb{E}\,\hat{\alpha} = \alpha, \quad \mathbb{E}\,\hat{\beta} = \beta, \quad \operatorname{var}\hat{\alpha} = (n\rho_2)^{-1}\,(\rho_2 + \overline{x}^2)\,\sigma_1^2, \quad \operatorname{var}\hat{\beta} = (n\rho_2)^{-1}\,\sigma_1^2,$$

and we look at the statistics

$$\theta_1 \;:=\; \left(\frac{n\rho_2}{\rho_2 + \overline{x}^2}\right)^{1/2} \frac{\hat{\alpha} - \alpha}{\sqrt{\mathrm{SSE}/n}} \quad \text{and} \quad \theta_2 \;:=\; (n\rho_2)^{1/2}\, \frac{\hat{\beta} - \beta}{\sqrt{\mathrm{SSE}/n}}.$$

In the case of normal regression, that is to say, if the E_j are $N(0, \sigma_1^2)$-distributed, the statistics $(1 - 2n^{-1})^{1/2}\,\theta_k$ are t-distributed with $n - 2$ degrees of freedom.

Let from here on

$$B \;=\; B(n) \;:=\; n^{-1/2}\, q^{-3}\, \mathbb{E}\,|E_1|^3\, I\{E_1^2 \le q^2\, n\} + q^{-1}\, \mathbb{E}\,|E_1|\, I\{E_1^2 > q^2\, n\}.$$

We have the following result:

Corollary 3.10. *Assume that $0 < \sigma_1^2 < \infty$. Then, for $k = 1, 2$, there exists a constant c such that*

$$Q(\theta_k, \lambda) \;\le\; c \max\big\{\lambda\,,\, \rho_3\, \rho_2^{-3/2}\, B\big\}.$$

As a result, for $k = 1, 2$,

$$Q(\theta_k, \lambda) \;\le\; c \max\big\{\lambda\,,\, \rho_3\, \rho_2^{-3/2}\, n^{-1/2}\, \sigma^{-3}\, \mathbb{E}\,|E_1|^3\big\}.$$

Letting c be the constant from the theorem, which we will take ≥ 1, without loss of generality we may assume that

$$\rho_3\, \rho_2^{-3/2}\, B \;\le\; c^{-1} \;\le\; 1.$$

Corollary 3.10 will now be proved as an application of Theorem 3.2. To this, we first rewrite θ_1 and θ_2 in terms of $E := (E_1, \dots, E_n)$. Let from here on, for $j = 1, \dots, n$,

$$v_j \;:=\; \rho_2^{1/2}\,(\rho_2 + \overline{x}^2)^{-1/2}\,(1 - \rho_2^{-1}\,\overline{x}\,(x_j - \overline{x})) \quad \text{and} \quad w_j \;:=\; \rho_2^{-1/2}\,(x_j - \overline{x}).$$

Lemma 3.11. *We have the following equalities:*

$$\begin{cases} \hat{\beta} - \beta &=\; n^{-1}\,\rho_2^{-1/2}\,\sum_{j=1}^n w_j E_j, \\ \hat{\alpha} - \alpha &=\; \overline{E} - (\hat{\beta} - \beta)\,\overline{x} \;=\; n^{-1}\,\rho_2^{-1/2}\,(\rho_2 + \overline{x}^2)^{1/2}\,\sum_{j=1}^n v_j E_j, \\ SSE &=\; \sum_{j=1}^n E_j^2 - n\,\overline{E}^2 - n^{-1}\,(\sum_{j=1}^n w_j E_j)^2. \end{cases}$$

Proof of Lemma 3.11. Writing out the definitions we see that

$$\begin{aligned} \hat{\beta} - \beta &= (n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j - \overline{x})\,(\alpha + \beta x_j + E_j) - \beta \\ &= \alpha\,(n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j - \overline{x}) + \beta\,(n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j - \overline{x})\,x_j \\ &\qquad + (n\,\rho_2)^{-1}\,\sum_{j=1}^n (x_j - \overline{x})\,E_j - \beta \\ &= n^{-1}\,\rho_2^{-1/2}\,\sum_{j=1}^n w_j E_j, \end{aligned}$$

since $\sum_{j=1}^n (x_j - \overline{x}) = 0$ and $\sum_{j=1}^n (x_j - \overline{x})\,x_j = \sum_{j=1}^n (x_j - \overline{x})^2 = n\,\rho_2$. Moreover,

$$\begin{aligned} \hat{\alpha} - \alpha &= \overline{Y} - \alpha - \hat{\beta}\,\overline{x} \;=\; \alpha + \beta\,\overline{x} + \overline{E} - \alpha - \hat{\beta}\,\overline{x} \;=\; \overline{E} - (\hat{\beta} - \beta)\,\overline{x} \\ &= n^{-1}\,\sum_{j=1}^n E_j - n^{-1}\,\rho_2^{-1}\,\overline{x}\,\sum_{j=1}^n (x_j - \overline{x})\,E_j = n^{-1}\,\sum_{j=1}^n \rho_2^{-1/2}\,(\rho_2 + \overline{x}^2)^{1/2}\,v_j\,E_j. \end{aligned}$$

Using that $\hat{\alpha} - \alpha = \overline{E} - (\hat{\beta} - \beta)\,\overline{x}$ and

$$\sum_{j=1}^n (x_j - \overline{x})\,(E_j - \overline{E}) \;=\; \sum_{j=1}^n (x_j - \overline{x})\,E_j \;=\; n\rho_2\,(\hat{\beta} - \beta),$$

we finally see that

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^n (\alpha + \beta x_j + E_j - \hat{\alpha} - \hat{\beta}\,x_j)^2 \;=\; \sum_{j=1}^n ((\hat{\beta} - \beta)\,\overline{x} - \overline{E} + E_j - (\hat{\beta} - \beta)\,x_j)^2 \\ &= \sum_{j=1}^n ((E_j - \overline{E}) - (x_j - \overline{x})\,(\hat{\beta} - \beta))^2 \\ &= \sum_{j=1}^n (E_j - \overline{E})^2 - 2\,(\hat{\beta} - \beta)\,\sum_{j=1}^n (x_j - \overline{x})\,(E_j - \overline{E}) + (\hat{\beta} - \beta)^2\,\sum_{j=1}^n (x_j - \overline{x})^2 \\ &= \sum_{j=1}^n (E_j - \overline{E})^2 - 2\,n\,\rho_2\,(\hat{\beta} - \beta)^2 + n\,\rho_2\,(\hat{\beta} - \beta)^2 \\ &= \sum_{j=1}^n E_j^2 - n\,\overline{E}^2 - n\,\rho_2\,(\hat{\beta} - \beta)^2. \end{aligned}$$

Using the foregoing the statement easily follows. □

Using Lemma 3.11 we may express $\hat{\alpha}$, $\hat{\beta}$ and SSE in terms of E, and therefore the θ_k as well. In fact, writing

$$c_{j1} \;:=\; v_j, \quad c_{j2} \;:=\; w_j,$$

for $1 \le j \le n$, we have for $k = 1, 2$ that

$$\theta_k \;=\; \theta_k(E) \;=\; (\textstyle\sum_{j=1}^n c_{jk} E_j)/\sqrt{\mathrm{SSE}(E)}, \tag{99}$$

so that both θ_k are of the form (63), with all X_j replaced by E_j, taking

$$R(E) \;:=\; -n\,\overline{E}^2 - n^{-1}\,(\textstyle\sum_{j=1}^n w_j E_j)^2.$$

Clearly in both cases we have an i.i.d. sample and $R(E)$ satisfies (62), and we want to apply Theorem 3.2.

Let q^2 be the $q_0^2(\alpha, k)$ which we obtain from Lemma 3.7 when taking $\alpha = E_1$ and $k = n$. As in the proof of Corollary 3.8, without loss of generality we may assume that

$$q^2 \;>\; \tfrac{1}{2}\sigma_1^2.$$

Now let $M_j^2 := q^2\, n$ for all j. As in (94) then $s^2 = n\, q^2$, so that $s^2 \neq 0$ and $M_j^2 \le s^2$.

Moreover, we have the following:

Lemma 3.12. *For $k = 1, 2$,*

$$\textstyle\sum_{j=1}^n c_{jk}^2 = n \quad \textit{and} \quad \sum_{j=1}^n |c_{jk}|^3 \le 8n\,\rho_3\,\rho_2^{-3/2}.$$

Proof of Lemma 3.12. Obviously $\sum_{j=1}^n w_j^2 = \rho_2^{-1}\, n\,\rho_2 = n$ and $\sum_{j=1}^n |w_j|^3 = \rho_2^{-3/2}\, n\,\rho_3$. Moreover,

$$\rho_2^{-1}\,(\rho_2 + \overline{x}^2)\,\textstyle\sum_{j=1}^n v_j^2 \;=\; n - 2\rho_2^{-1}\,\overline{x}\,\sum_{j=1}^n (x_j - \overline{x}) + \rho_2^{-2}\,\overline{x}^2\, n\rho_2 \;=\; n\,\rho_2^{-1}\,(\rho_2 + \overline{x}^2),$$

so that $\sum_{j=1}^n v_j^2 = n$, and, using the C_r-inequality and the inequality $|a|^{3/2} + |b|^{3/2} \le 2^{1/2}\,(|a| + |b|)^{3/2}$,

$$\begin{aligned}
\rho_2^{-3/2}\,(\rho_2 + \overline{x}^2)^{3/2}\,\textstyle\sum_{j=1}^n |v_j|^3 \;&\le\; \textstyle\sum_{j=1}^n (1 + \rho_2^{-1}\,|\overline{x}|\,|x_j - \overline{x}|)^3 \\
&\le\; 4\,(n + \rho_2^{-3}\,|\overline{x}|^3\, n\,\rho_3) \;=\; 4n\,\rho_2^{-3/2}\,(\rho_2^{3/2} + \rho_3\,\rho_2^{-3/2}\,|\overline{x}|^3) \\
&\le\; 4n\,\rho_2^{-3/2}\,(\rho_3\,\rho_2^{-3/2})\,(\rho_2^{3/2} + |\overline{x}|^3) \;\le\; 8n\,(\rho_3\,\rho_2^{-3/2})\,\rho_2^{-3/2}\,(\rho_2 + \overline{x}^2)^{3/2},
\end{aligned}$$

so that $\sum_{j=1}^n |v_j|^3 \le 8n\,\rho_3\,\rho_2^{-3/2}$. This finishes the proof. □

The conditions of Lemma 3.3 are satisfied, which, together with the fact that in general $|a + b|^{3/2} \ll |a|^{3/2} + |b|^{3/2}$, leads to the fact that

$$\begin{aligned}
\mathbb{E}\,|R(U)|^{3/2} \;&\ll\; \mathbb{E}\,|n\,\overline{U}^2|^{3/2} + \mathbb{E}\,|n^{-1}\,(\textstyle\sum_{j=1}^n w_j U_j)^2|^{3/2} \\
&=\; n^{-3/2}\,(\mathbb{E}\,|\textstyle\sum_{j=1}^n U_j|^3 + \mathbb{E}\,|\sum_{j=1}^n w_j U_j|^3) \;\ll\; n^{-3/2} \;\le\; A_3^3,
\end{aligned}$$

see as well (64). Now we are ready to apply Theorem 3.2 for both θ_k. As to A_1, A_2 and A_3 we have precisely the situation that we encountered in the proof of Corollary 3.8 (with the X_j replaced by E_j), so that

$$A_1,\, A_2,\, A_3 \;\leq\; 2^{3/2}\, B.$$

Lemma 3.12 finally shows us that

$$\nu_{3k} \;=\; n^{-1} \textstyle\sum_{j=1}^{n} |c_{jk}|^3 \;\ll\; \rho_3\, \rho_2^{-3/2}.$$

Using the theorem, the proof of Corollary 3.10 is now easily concluded. □

References

[1] I.B. Alberink, W.R. Pestman, *Mathematical statistics. Problems and detailed solutions*, Walter De Gruyter, 1998, Berlin.

[2] A.K. Aleshkyavichene, V.A. Statulyavichus, Inversion formulas in the case of a discontinuous limit law, 1997, *Theory Prob. Appl. 42, p.1-16.*

[3] V. Bentkus, On the asymptotical behavior of the constant in the Berry-Esseen inequality, 1994, *J. Theoret. Probab. 7, p.211-224.*

[4] V. Bentkus, M. Bloznelis, F. Götze, A Berry-Esséen Bound for Student's Statistic in the Non-I.I.D. Case, 1996, *Journal of Theoretical Probability 9, p.765-796.*

[5] V. Bentkus, F. Götze, The Berry-Esseen bound for Student's statistic, 1996, *Ann. Prob. 24, p.491-503.*

[6] V. Bentkus, F. Götze and W.R. van Zwet, An Edgeworth expansion for symmetric statistics, 1997, *Ann. Statist. 25, p.851-896.*

[7] G. Blom, Some properties of incomplete U-statistics, 1976, *Biometrika 63, p.573-580.*

[8] M. Bloznelis, A. Rackauskas, A Berry-Esséen bound for least squares estimators of regression parameters, 1999, *Lithuan. Math. J. 39, p.1-7.*

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

[10] Y.-K. Chan, J. Wierman, On the Berry-Esseen theorem for U-statistics, 1977, *Ann. Prob. 5, p.136-139.*

[11] R.J.M.M. Does, Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis, 1982, *Ann. Prob. 10, p.982-991.*

[12] E.B. Dynkin, A. Mandelbaum, Symmetric statistics, Poisson point processes, and multiple Wiener integrals, 1983, *Ann. Statist. 11, p.739-745.*

[13] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd Ed., Wiley, 1971, New York.

[14] K.O. Friedrich, A Berry-Esseen bound for functions of independent random variables, 1989, *Ann. Statist. 17, p.170-183.*

[15] J.D. Gibbons, S. Chakraborti, *Nonparametric statistical inference*, 3rd Ed., Marcel Dekker, Inc., 1992, New York.

[16] L. Giraitis, D. Surgailis, A limit theorem for a triangular array of symmetric statistics, 1984, *Statistics and control of stochastic processes, Moscow, p.147-166.* Transl. Ser. Math. Engrg., Optim. Software, New York, 1985.

[17] Z. Guan, A Berry-Esseen bound for k-sample symmetric statistics, 1994, *Northeast. Math. J. 10, p.411-420.*

[18] R. Helmers, A Berry-Esseen theorem for linear combinations of order statistics, 1981, *Ann. Prob. 9, p.342-347.*

[19] R. Helmers, W.R. van Zwet, The Berry-Esseen bound for U-statistics, *Statistical Decision Theory and Related Topics III*, Vol. 1 (S.S. Gupta and J.O. Berger, eds.), p.497-512, Academic Press, 1982, New York.

[20] W. Hengartner, R. Theodorescu, *Concentration functions*, Academic Press, 1973, New York.

[21] W. Hoeffding, A class of statistics with asymptotically normal distributions, 1948, *Ann. Math. Stat. 19, p.293-325.*

[22] S. Janson, The asymptotic distributions of incomplete U-statistics, 1984, *Z. Wahrsch. verw. Gebiete 66, p.495-505.*

[23] T.L. Lai, J.Q. Wang, Edgeworth expansions for symmetric statistics with applications to bootstrap methods, 1993, *Statist. Sinica 3, p.517-542.*

[24] M. Loève, *Probability theory I*, Springer, 1977, New York.

[25] A. Mandelbaum, M.S. Taqqu, Invariance principle for symmetric statistics, 1984, *Ann. Statist. 12, p.483-496.*

[26] V. Mandrekar, M.M. Meerschaert, Sample moments and symmetric statistics, 1994, *Stochastic analysis on infinite-dimensional spaces* (Baton Rouge, L.A.), p.197-210, Pitman Res. Notes Math. Ser., 310, Longman Sci. Tech., Harlow.

[27] W.R. Pestman, *Mathematical Statistics, An Introduction*, Walter de Gruyter, 1998, Berlin.

[28] V.V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Clarendon Press, 1995, Oxford.

[29] H. Prawitz, Limits for a distribution, if the characteristic function is given in a finite doamin, 1972, *Skand. Aktuar. Tidskr., p.138-154.*

[30] H. Putter, W.R. van Zwet, Empirical Edgeworth expansions for symmetric statistics, 1998, *Ann. Statist. 26, p.1540-1569.*

[31] Q. Wang, L_p rates of convergence for symmetric statistics, 1998, *Nanjing Daxue Xuebao Shuxue Bannian Kan 15, p.41-49.*

[32] W.R. van Zwet, A Berry-Esseen bound for symmetric statistics, 1984, *Z. Wahrsch. verw. Gebiete 66, p.425-440.*

Ivo Alberink
Department of Mathematics
University of Nijmegen
Postbus 9010
6500 GL, Nijmegen
The Netherlands
Email address: alberink@sci.kun.nl

Vidmantas Bentkus
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501, Bielefeld
Germany
Email address: bentkus@mathematik.uni-bielefeld.de