Berry-Esseen bounds for von Mises and U-statistics

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

Abstract

Let $X_1, \dots, X_n$ be i.i.d. random variables. An optimal Berry-Esseen bound is derived for U-statistics of order 2, that is, statistics of the form $\sum_{j<k} H(X_j, X_k)$, where H is a measurable, symmetric function such that $\mathbb{E}\,|H(X_1, X_2)| < \infty$, assuming that the statistic is non-degenerate. The same is done for von Mises statistics, that is, statistics of the form $\sum_{j,k} H(X_j, X_k)$. As a corollary the central limit theorem is derived under optimal moment conditions.

Key words and phrases: U-statistics, von Mises statistics, Berry-Esseen bound, rate of convergence, central limit theorem, normal approximations.
A.M.S. classification (1991): 62E20.

1 Introduction and results

Let $X_1, \dots, X_n$ be independent, identically distributed random variables, taking their values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. By $X, \overline{X}$ we shall denote independent copies of X_1. For $n \geq 2$, we consider real-valued U-statistics of degree 2, that is to say, statistics of the form

$$\mathbb{U} \;=\; \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq n} H(X_j, X_k), \tag{1}$$

where $H : \mathcal{B}^2 \to \mathbb{R}$ is a symmetric kernel such that $\mathbb{E}\,|H(X, \overline{X})| < \infty$. We allow dependence of the sample and the statistic, and as well of the functions g, g_0 and h defined below, on n.[4] Writing $\mathbb{E}\,H := \mathbb{E}\,H(X, \overline{X})$ and introducing the functions

$$g(x) \;:=\; \mathbb{E}\,H(X, x) - \mathbb{E}\,H, \qquad h(x,y) \;:=\; H(x,y) - g(x) - g(y) - \mathbb{E}\,H, \tag{2}$$

[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.

[4]In other words, we write $(\mathcal{X}, \mathcal{B})$, $X_1, \dots, X_n$, H instead of $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)})$, $X_1^{(n)}, \dots, X_n^{(n)}$, $H^{(n)}$ respectively.

for $x, y \in \mathcal{B}$, we may represent the statistic

$$\mathbb{T} := \frac{\sqrt{n}}{2} (\mathbb{U} - \mathbb{E}\,\mathbb{U}) \qquad \text{as} \qquad \mathbb{T} = \mathbb{T}_1 + \mathbb{T}_2, \tag{3}$$

with

$$\mathbb{T}_1 := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(X_j), \qquad \mathbb{T}_2 := \frac{1}{(n-1)\sqrt{n}} \sum_{1 \le j < k \le n} h(X_j, X_k). \tag{4}$$

The representation (4) is the so-called Hoeffding decomposition of $\mathbb{T}$. Here $\mathbb{T}_1$ is the *linear part* of $\mathbb{T}$, whereas $\mathbb{T}_2$ is its *quadratic part*. We will assume everywhere that

$$0 < s^2 := \mathbb{E}\,\mathbb{T}_1^2 = \mathbb{E}\,g^2 < \infty, \tag{5}$$

which means that $\mathbb{T}$ is a non-degenerate U-statistic. All parts of the Hoeffding decomposition (4) are uncorrelated, so that we may write

$$\sigma^2 := \mathbb{E}\,\mathbb{T}^2 = \mathbb{E}\,\mathbb{T}_1^2 + \mathbb{E}\,\mathbb{T}_2^2 = s^2 + \mathbb{E}\,\mathbb{T}_2^2,$$

with

$$\frac{\mathbb{E}\,h^2}{2n} \le \mathbb{E}\,\mathbb{T}_2^2 = \frac{\mathbb{E}\,h^2}{2(n-1)} \le \frac{\mathbb{E}\,h^2}{n}. \tag{6}$$

The (normed) linear part $\mathbb{T}_1/s$ is in fact a sum of i.i.d. random variables with mean 0 and variance 1. Under appropriate conditions its distribution may be approximated by the standard normal distribution Φ (see Feller (1971), Petrov (1995), Bhattacharya and Rao (1986) for classical results). Our result is that the distribution of $\mathbb{T}_1/s + \mathbb{T}_2/s$—which is the distribution of the sum $\mathbb{T}_1/s$ perturbed by a degenerate U-statistic of the second degree—may be approximated by an expansion related to Φ, provided that we correct the classical error bounds, adding to them the variance $\mathbb{E}\,\mathbb{T}_2^2 \le \mathbb{E}\,h^2/n$. This final result refines and includes as partial cases related previous work, see the references and discussion below. Under additional moment assumptions the correction of the error bounds may be improved to $n^2 \mathbb{E}\,h^k/n^{3k/2} + (\mathbb{E}\,h^2/n)^{k/2}$, for even $k \ge 4$. Hence, if $\mathbb{E}\,h^4 < \infty$, the correction would be $\mathcal{O}(n^{-2})$. The results extend as well to higher degree U-statistics and even for arbitrary (non-linear) functions of non-identically distributed samples (to be published elsewhere). Furthermore, similar corrections can be obtained to Edgeworth expansions.

Now we turn to explicit formulations of the results. Write

$$\Delta = \Delta(\mathbb{T}) := \sup_{x \in \mathbb{R}} |\mathbb{P}(\mathbb{T}/s \le x) - \Phi(x)|. \tag{7}$$

By a *Berry-Esseen bound* for $\mathbb{T}$ we will mean a bound for Δ, which, under appropriate conditions on the sequence of statistics and the samples, is expected to satisfy $\Delta = \mathcal{O}(n^{-1/2})$, as $n \to \infty$. Let $s^2 = \mathbb{E}\,g^2$ as in (5),

$$\beta := \mathbb{E}\,|g|^3, \qquad \gamma_p := \mathbb{E}\,|h|^p, \qquad \gamma := \gamma_2,$$

for $p \ge 1$, and

$$\varepsilon := \frac{\beta}{s^3\sqrt{n}}, \qquad \delta := \frac{\gamma}{s^2\, n}. \tag{8}$$

Next define the asymptotic expansion

$$G(x) = \Phi(x) + \Phi_1(x), \tag{9}$$

where

$$\Phi_1(x) := \tfrac{1}{2}\kappa\, \Phi'''(x), \qquad \kappa := n^{-1/2}\, s^{-3}\, \mathbb{E}\, g(X)g(\overline{X})\, h(X,\overline{X}), \tag{10}$$

and write

$$\Gamma = \Gamma(\mathbb{T}) := \sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T}/s \le x) - G(x)|. \tag{11}$$

We shall write $A \ll B$ if there exists an absolute positive constant c such that $A \le cB$. Our main result reads as follows:

Theorem 1. *Assume that $0 < s^2 < \infty$. Then*

$$\Gamma \ll \varepsilon + \delta \qquad \textit{and} \qquad \Delta \ll \varepsilon + \delta + |\kappa|. \tag{12}$$

Moreover, in the case where $\sigma^2 < \infty$, we may replace s by σ.

Hölder's inequality implies that

$$|\kappa| \le n^{-1/2}\, s^{-3}\, (\mathbb{E}\, g^2)\, (\mathbb{E}\, h^2)^{1/2} = n^{-1/2}\, s^{-1}\, (\mathbb{E}\, h^2)^{1/2}, \tag{13}$$

so that

$$|\Phi_1(x)| \ll \sqrt{\frac{\mathbb{E}\, h^2}{s^2\, n}} = \sqrt{\delta}.$$

Therefore an impression may evolve that the Edgeworth correction G has a size $\mathcal{O}(n^{-1/2})$, which is comparable to the bound (12) for the error. This is true asymptotically in the case when the functions g and h are fixed and independent of n. However, in a number of applications, these functions may heavily depend on n so that Φ_1 indeed corrects Φ, see Example 7 below. Example 7 shows as well that bounds of type (12) with $\gamma/(s^2\, n)$ replaced by, say $\big(\gamma/(s^2\, n)\big)^2$, would do much better. Under the assumption that the sequence of moments γ is bounded as $n \to \infty$, the contribution to the error of the non-linear part $\mathbb{T}_2$ is of size $\mathcal{O}(n^{-1})$, which is negligible compared to the contribution of the linear part. This contribution becomes more important under weaker moment assumptions: e.g., of the size $\mathcal{O}(n^{-1/2})$ in the case where the sequence of moments $\gamma_{5/3}$ is bounded, and of the size $\mathcal{O}(1)$ in the case where the sequence of moments $\gamma_{4/3}$ is bounded.

Another advantage of the bounds of Theorem 1 is that they usually provide an optimal dependence on moments. Example 8 demonstrates this. It is interesting to notice that the U-statistic of Example 8 in the case $p = 2$ approximates so-called self-normalized sums, as well as the Student statistic: these statistics usually serve as a touchstone to verify the quality of bounds.

Using truncation methods, Theorem 1 yields a number of bounds under moment assumptions which do not require that $\beta < \infty$ and $\gamma < \infty$, see Theorem 2, Corollaries 3–5, (29) and related comments below. At the same time the result becomes more general since it is applicable now to the so-called von Mises statistics.

A von Mises functional (statistic) of degree 2 is a statistic of the form

$$\mathbb{M} \; = \; \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} M(X_j, X_k), \tag{14}$$

where $M : \mathcal{B}^2 \to \mathbb{R}$ is (usually) supposed to be a symmetric kernel. Writing $H(x, y) := n^{-1}(n-1)M(x, y)$ and rearranging the summands in (14), we have

$$\mathbb{M} \; = \; \frac{1}{n^2} \sum_{j=1}^{n} M(X_j, X_j) + \binom{n}{2}^{-1} \sum_{1 \le j < k \le n} H(X_j, X_k). \tag{15}$$

Now let $\mathbb{T}$ be as in (3), with g and h as in (2). Writing

$$m \; := \; \mathbb{E}\, M(X, X), \qquad g_0(x) \; := \; \tfrac{1}{2}\,(M(x, x) - m),$$

$$S_j \; := \; n^{-3/2}\, g_0(X_j), \qquad \mathbb{S} \; := \; \textstyle\sum_{j=1}^{n} S_j,$$

we can represent $\mathbb{M}$ by (cf. the Hoeffding decomposition (4) for $\mathbb{T}$)

$$\tfrac{1}{2}\sqrt{n}\,(\mathbb{M} - \mathbb{E}\,\mathbb{M}) \; = \; \mathbb{T} + \mathbb{S}, \tag{16}$$

with $\mathbb{E}\,\mathbb{M} = \mathbb{E}\, H + m/n$. Hence, any von Mises statistic (here $\frac{1}{2}\sqrt{n}\,(\mathbb{M} - \mathbb{E}\,\mathbb{M})$) may be interpreted as a perturbation of a U-statistic (here $\mathbb{T}$) by a linear statistic (here $\mathbb{S}$).

Generally speaking, the perturbation $\mathbb{S}$ is stochastically smaller than the linear part $\mathbb{T}_1$. It is immaterial for us that g_0 is related to the von Mises statistic (14). Therefore we will assume throughout only that the $S_1, \dots, S_n$ are i.i.d., and that $S_j = S(X_j)$ is a function of X_j such that $\mathbb{E}\, S_j = 0$.

Now introduce the indicator functions

$$I := \mathbb{I}\{g^2(X) \le s^2 n\}, \quad K := \mathbb{I}\{g_0^2(X) \le s^2 n^3\}, \quad J := \mathbb{I}\{h^2(X, \overline{X}) \le s^2 n^3\}, \tag{17}$$

and their 'complements' $I^c := 1 - I$, $K^c := 1 - K$, $J^c := 1 - J$. Next, define the truncated moments

$$\begin{aligned} \varepsilon_* &:= s^{-2}\, \mathbb{E}\, I^c g^2(X), & \delta_* &:= n^{1/2}\, s^{-1}\, \mathbb{E}\, J^c |h(X, \overline{X})|, \\ \varepsilon^* &:= n^{-1/2}\, s^{-3}\, \mathbb{E}\, I |g(X)|^3, & \delta^* &:= n^{-1}\, s^{-2}\, \mathbb{E}\, J h^2(X, \overline{X}), \end{aligned} \tag{18}$$

and

$$\lambda_* := n\,\mathbb{E}\,K^c, \qquad \lambda^* := n^{-1/2}\,s^{-1}\,\mathbb{E}\,K|g_0(X)|, \qquad \lambda := n^{-1/2}\,s^{-1}\,\mathbb{E}\,|g_0(X)|. \tag{19}$$

The truncated moments are related to the moments ε, δ, cf. (8), and λ as follows:

$$\varepsilon_* + \varepsilon^* \le \varepsilon, \qquad \lambda_* + \lambda^* \le \lambda, \qquad \delta_* + \delta^* \le \delta. \tag{20}$$

Let $\overline{I} := \mathbb{I}\{g^2(\overline{X}) \le s^2 n\}$, and denote

$$\kappa^* \;:=\; n^{-1/2}\,s^{-3}\,\mathbb{E}\,I g(X)\,\overline{I}\,g(\overline{X})\,J\,h(X,\overline{X}). \tag{21}$$

Define Φ_1^* and G^* as Φ_1 and G in (10) and (9), replacing κ by κ^*. Finally, introduce $\Gamma^*(\mathbb{T}+\mathbb{S})$ as Γ in (11), replacing G by G^* and $\mathbb{T}$ by $\mathbb{T}+\mathbb{S}$. The next theorem provides a bound for von Mises' statistic $\mathbb{T}+\mathbb{S}$. If $\mathbb{S}=0$ the bound reduces to a bound for $\mathbb{T}$.

Theorem 2. *Let $0 < s^2 = \mathbb{E}\,\mathbb{T}_1^2 < \infty$. Then*

$$\Gamma^*(\mathbb{T}+\mathbb{S}) \;\ll\; \varepsilon_* + \varepsilon^* + \lambda_* + \lambda^* + \delta_* + \delta^*, \tag{22}$$

and

$$\Delta(\mathbb{T}+\mathbb{S}) \;\ll\; \varepsilon_* + \varepsilon^* + \lambda_* + \lambda^* + \delta_* + \delta^* + |\kappa^*|. \tag{23}$$

A combination of (20), (22), (23), and the fact that $|\kappa^* - \kappa| \ll \varepsilon + \delta$ (see (45) for a proof), yields the following corollary:

Corollary 3. *Let $0 < s^2 < \infty$. Then we have:*

$$\Gamma^*(\mathbb{T}+\mathbb{S}) \;\ll\; \varepsilon + \lambda + \delta, \qquad \Delta(\mathbb{T}+\mathbb{S}) \;\ll\; \varepsilon + \lambda + \delta + |\kappa|. \tag{24}$$

Moreover, in the case where $\sigma^2 < \infty$, we may replace s by σ.

(The replacement of s by σ is taken care of as in the proof of Theorem 1.)

We say that a sequence of statistics $\mathbb{T} = \mathbb{T}(X_1, \dots, X_n)$ and samples $X_1, \dots, X_n$ satisfies the Central Limit Theorem (CLT) if there exist numbers a_n and b_n such that $(\mathbb{T} - a_n)/b_n$ converges weakly to the standard normal distribution Φ.

Corollary 4. *Let $0 < s^2 < \infty$. The sequence of statistics $\mathbb{T}+\mathbb{S}$ satisfies the CLT with $a_n = 0$ and $b_n = s$ if the following conditions are fulfilled. For all $\theta > 0$, and $n \to \infty$,*

$$\begin{aligned}
\varepsilon_*(\theta) \;&:=\; s^{-2}\,\mathbb{E}\,\mathbb{I}\{g^2(X) > \theta s^2 n\}\,g^2(X) \;\to\; 0,\\
\lambda_*(\theta) \;&:=\; n\,\mathbb{E}\,\mathbb{I}\{g_0^2(X) > \theta s^2 n^3\} \;\to\; 0,\\
\delta_*(\theta) \;&:=\; s^{-1}\sqrt{n}\;\mathbb{E}\,\mathbb{I}\{h^2(X,\overline{X}) > \theta s^2 n^3\}\,|h(X,\overline{X})| \;\to\; 0,
\end{aligned} \tag{25}$$

and moreover

$$\lambda^* \to 0, \qquad \delta^* \to 0. \tag{26}$$

The condition (26) *may be replaced by the stronger condition that*

$$\limsup_{n\to\infty} s^{-2/3}\,\mathbb{E}\,K\,|g_0(X)|^{2/3} \;<\; \infty, \quad \limsup_{n\to\infty} s^{-4/3}\,\mathbb{E}\,J\,|h(X,\overline{X})|^{4/3} \;<\; \infty. \tag{27}$$

If the distribution of X_1 and the functions g, h and g_0 are independent of n, then $\mathbb{T}+\mathbb{S}$ already satisfies the CLT if the moments $\mathbb{E}\,g^2$, $\mathbb{E}\,|g_0|^{2/3}$ and $\mathbb{E}\,|h|^{4/3}$ exist.

The following is as well an easy consequence of Theorem 2:

Corollary 5. *Assume that $0 < s^2 < \infty$. Then*

$$\Delta(\mathbb{T}+\mathbb{S}) \;\ll\; \frac{\mathbb{E}\,|g|^3}{s^3\sqrt{n}} + \frac{\mathbb{E}\,|g_0|}{s\,\sqrt{n}} + \frac{\mathbb{E}\,|h|^{5/3}}{s^{5/3}\sqrt{n}}. \tag{28}$$

Analogue to Corollary 5, a spectrum of bounds like (28) may be derived from Theorem 2, assuming that the moments $\mathbb{E}\,|g|^p$, $\mathbb{E}\,|g_0|^q$ and $\mathbb{E}\,|h|^r$ exist for some $2 \le p \le 3$, $\frac{2}{3} \le q \le 1$ and $\frac{4}{3} \le r \le 2$. If X and g, g_0, h are independent of n, these bounds have the order

$$o(n^{-(p-2)/2}) + o(n^{-(3q-2)/2}) + o(n^{-(3r-4)/2}), \tag{29}$$

provided that $p < 3$, $q < 1$ and $r < 2$. If $p = 3$ or $q = 1$ or $r = 2$, the statement in (29) remains true, but then the corresponding o's have to be replaced by $\mathcal{O}$'s. For example: in the case where $p = \frac{5}{2}$ we have that

$$\varepsilon^* \;=\; n^{-1/4}\,\mathbb{E}\,(|g|/s)^{5/2}\,I\,(n^{-1/2}\,|g|/s)^{1/2},$$

where $I\,n^{-1/2}\,|g|/s$ is bounded and pointwise converging to zero. By Lebesgue's theorem on dominated convergence then $n^{1/4}\,\varepsilon^* \to 0$. We omit the further details.

The history related to the CLT and the accuracy of normal approximations to U-statistics is sufficiently rich, see the books Lee (1990), Borovskikh and Korolyuk (1994), Borovskikh (1996). Shapiro and Hubert (1979) considered weighted U-statistics, assuming that $X_1, X_2, \dots$ are independent of n. In the case where the weights are identical their result reads as follows: the statistic $\mathbb{T}/\sigma$ satisfies the CLT if

$$\frac{\mathbb{E}\,h^2}{n\,s^2 + \mathbb{E}\,h^2} \;\to\; 0$$

and g satisfies the condition from (25). This result is clearly implied by Corollary 4. The most advanced result with kernels independent of n — Theorem 4.1.1 in Borovskikh (1996) — says the same as Corollary 4 in the case that $\mathbb{S} = 0$ and g, h are independent of n. Notice that the conditions (25) are Lindeberg type conditions, specialized to the case where all random variables in each series are identically distributed.

The accuracy of normal approximations—improving the bounds for remainders and posing ever lighter conditions on the statistic and the observations—was estimated in a number of papers. We mention Hoeffding (1948), Bickel (1974), Serfling (1980), Chan and Wierman (1977), Callaert and Janssen (1978) only. Helmers and van Zwet (1982) proved (28) with $\mathbb{S} = 0$ and $\frac{5}{3}$ replaced by $\frac{5}{3} + \varepsilon$, $\varepsilon > 0$. Bounds without ε may be found in Borovskikh and Korolyuk (1984, 1985), Friedrich (1989), Bolthausen and Götze (1993). Van Zwet (1984) established (28) with $\mathbb{S} = 0$ and $s^{-2}\,\mathbb{E}\,h^2$ instead of $s^{-5/3}\,\mathbb{E}\,|h|^{5/3}$. We use some methods from Alberink and Bentkus (1999), who obtained bounds for the so-called concentration (of probability mass) of $\mathbb{T}$, comparable in quality with Theorem 1. Using the approach developed in the latter paper, Alberink (2000b) showed that the bound $\Delta(\mathbb{T}) \ll \varepsilon + \delta^{\theta} + |\kappa|$, $\theta < 1$, holds in the non-i.i.d. setting. Alberink (2000a) extended the result of van Zwet (1984) to the non-i.i.d. case and improved the dependence on $s^{-2}\,\mathbb{E}\,h^2$ to $\sqrt{s^{-2}\,\mathbb{E}\,h^2}$. Borovskikh (2000) showed that

$$\Delta(\mathbb{T}) \ \ll\ \varepsilon_* + \varepsilon^* + \delta_* + \delta^*\, \log n + |\kappa^*|. \tag{30}$$

In all cited papers the approximation by the normal law Φ is considered; it seems that this paper is a first one where the corrected normal law $G = \Phi + \Phi_1$ is used.

Compared to n^{-1}, the $\log n$ factor in (30) is of negligible size. However, this factor is a principle obstacle if we wish to establish a sufficiently general, precise and exhausting theory of Edgeworth expansions for non-linear statistics, cf. a related discussion on p. 825 in Bentkus, Götze and van Zwet (1997). Compare as well Lyapunov's bound $\mathcal{O}(\beta\, n^{-1/2} \log n)$ with the Berry-Esseen bound $\mathcal{O}(\beta\, n^{-1/2})$, $s = 1$. We could not prove Theorem 1 extending and improving the existing technique—it seems that the best achievable by this technique is (30). Roughly speaking, this technique usually involves a Fourier transform, which reduces the problem to an estimation of certain integrals of characteristic functions. After this one first removes a relatively small part of $\mathbb{T}_2$ from the characteristic function $\mathbb{E}\,\exp\{it\mathbb{T}\} = \mathbb{E}\,\exp\{it\mathbb{T}_1 + it\mathbb{T}_2\}$. The effect of this is that now the characteristic function contains a (relatively small) product of certain conditional characteristic functions. As a next step, the conditional product is transformed to an unconditional one. The presence of the unconditional product ensures the convergence of the integrals and allows to conclude the proof using rather straightforward Taylor expansions. The proof in this paper starts with a Taylor expansion in powers of $\mathbb{T}_2$ such that the remainder term has the same order as the desired error (which becomes clear at the end of the proof). After this we use the additive structure of the statistic, which allows us to get rid of most of the dependencies. As a consequence we can create a conditional product of approximately $n - 1$ characteristic functions. We can estimate this product (in order to ensure the convergence of the integrals) in the case where certain conditional variances related to kernels are bounded. The general unbounded case is reduced to the bounded one by an application of a non-standard truncation. We would like to notice that such a scheme of the proof has a side effect—a number of technicalities (typical for traditional proofs) disappears.

Another feature of our approach is that we prove the main Theorem 1 for non-truncated kernels. The result with truncated kernels and weaker moment assumptions (Theorem 2 and its corollaries) is derived from Theorem 1 by an application of a rather simple truncation technique. Once again, such a way reduces the number of technicalities considerably.

In conclusion we discuss how one can improve the bounds of Theorem 1. First of all, the term $\varepsilon = \beta/(s^3\sqrt{n})$ is unavoidable, unless we require higher order moments and impose a Cramer type condition on the characteristic function $\mathbb{E}\,\exp\{itg(X)\}$ — see Petrov (1995) for a counter-example: $\mathbb{T} = \mathbb{T}_1$, $g(X)$ being a Rademacher random variable with $\mathbb{P}(g(X) = \pm 1) = \frac{1}{2}$. The term $\delta = \mathbb{E}\,h^2/(s^2 n)$ is optimal in the sense that lower order moments $\mathbb{E}\,|h|^\theta$, $\theta < 2$, do not control $\mathcal{O}(n^{-1})$ behaviour, see Bentkus, Götze and Zitikis (1994), Bentkus and Götze (1993). Another way to see this would be to use Edgeworth type expansions of $\mathbb{P}(\mathbb{T} \le x)$. In particular, the terms with κ are unavoidable as well.

Notation. To conclude the introduction we gather some general notation, to be used throughout the paper. 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. By

$$\mathbb{I}\{A\}$$

we mean the indicator of an event A. The notation

$$b_1 \;\ll\; b_2, \tag{31}$$

for $b_1, b_2 \ge 0$, will mean by definition that $b_1 \le c\, b_2$ for some absolute constant c. We write

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

for $y \in \mathbb{R}$, with $i = \sqrt{-1}$ denoting the complex root.

We will use as well the following, more compact notation for the parts of $\mathbb{T}$: taking

$$T_j \;:=\; g(X_j)/\sqrt{n}, \qquad T_{j,k} \;:=\; h(X_j, X_k)/((n-1)\sqrt{n}),$$

we have

$$\mathbb{T} \;=\; \mathbb{T}_1 + \mathbb{T}_2, \quad \text{with} \quad \mathbb{T}_1 = \sum_{j=1}^n T_j, \qquad \mathbb{T}_2 = \sum_{1\le j<k\le n} T_{j,k}.$$

2 Proof of the corollaries of Theorem 1

Using truncation, in this section we will derive all corollaries of Theorem 1 stated in the introduction.

We shall often use the following simple truncation lemma. Let $Y_1, \dots, Y_k$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$, taking values in any measurable space $(\mathcal{X}, \mathcal{B})$. We do not assume that the Y_j are independent. Now let $T : \mathcal{X}^k \to \mathcal{Y}$ be a measurable function of variables $x_j \in \mathcal{X}$, taking values in a measurable space $(\mathcal{Y}, \mathcal{C})$. For a certain event $A \subset \mathcal{B}$, introduce the truncated random variable

$$Y_1^* \; := \; Y_1 \, \mathbb{I}\{Y_1 \in A\}.$$

Lemma 6. *We have:*

$$\sup_{B \in \mathcal{C}} \left| \mathbb{P}\left(T(Y_1, \dots, Y_k) \in B\right) - \mathbb{P}\left(T(Y_1^*, Y_2, \dots, Y_k) \in B\right) \right| \; \leq \; \mathbb{P}(A^c), \tag{32}$$

where $A^c = \Omega \setminus A$ denotes the complement of A.

Lemma 6 means that if we have a function, say $T(X_1, \dots, X_n)$, of independent random variables $X_1, \dots, X_n$, we can replace any of the sample elements X_j by its truncated version $X_j^* := X_j \, \mathbb{I}\{X_j \in A_j\}$, the error by such a replacement being bounded from above by the sum of $\mathbb{P}(A_j^c)$ for all j such that X_j is replaced somewhere. For example, in the case where $A_j = A$, independently of j, and the function T is a function of $2n$ variables,

$$\sup_{B \in \mathcal{C}} \left| \mathbb{P}\left(T(X_1, X_1, \dots, X_n, X_n) \in B\right) - \mathbb{P}\left(T(X_1^*, X_1, \dots, X_n^*, X_n) \in B\right) \right| \leq n \, \mathbb{P}(A^c). \tag{33}$$

One can use Lemma 6 as well in the opposite direction, for 'de-truncation', that is, for a replacement of any given X_j^* by X_j. Furthermore, we can implement into our statistic indicators $I_j := \mathbb{I}\{X_j \in A_j\}$ in arbitrary places. The error is again bounded from above by the sum of $\mathbb{P}(A_j^c)$ for all j such that I_j is at least once somewhere implemented.

Proof of Lemma 6. Using the abbreviation $T := T(Y_1, \dots, Y_k)$, we have that

$$\begin{aligned} \mathbb{P}(T \in B) \; &= \; \mathbb{P}(Y_1 \in A, \, T(Y_1^*, Y_2, \dots, Y_k) \in B) + \mathbb{P}(Y_1 \in A^c, \, T \in B) \\ &\leq \; \mathbb{P}(T(Y_1^*, Y_2, \dots, Y_k) \in B) + \mathbb{P}(A^c). \end{aligned} \tag{34}$$

Similarly

$$\mathbb{P}(T(Y_1^*, Y_2, \dots, Y_k) \in B) \; \leq \; \mathbb{P}(T \in B) + \mathbb{P}(A^c). \tag{35}$$

The estimates (34) and (35) together yield (32), thus proving the lemma. □

We turn to the proof of Theorem 2.

Proof of Theorem 2. Without loss of generality we shall assume that $s^2 = 1$. We write

$$\alpha \; := \; \varepsilon_* + \varepsilon^* + \lambda_* + \lambda^* + \delta_* + \delta^*,$$

and may assume that $\alpha \le \frac{1}{100}$.

Let I_j, K_j and $J_{j,k}$ denote indicator functions defined as I, K and J in (17), replacing X by X_j and $\overline{X}$ by X_k. We use as well the analogue expressions I_j^c, K_j^c and $J_{j,k}^c$. The idea of the proof is the following. Using Lemma 6, we first replace all T_j, S_j and $T_{j,k}$ by their truncated versions $I_j\,T_j$, $K_j\,S_j$ and $J_{j,k}\,T_{j,k}$. After this we re-center our truncated statistic a bit and apply Theorem 1.

We start with the truncation. Lemma 6 is telling us that the error introduced by truncating the T_j, S_j and $T_{j,k}$ as described is bounded by

$$n\,\mathbb{E}\,I^c + n\,\mathbb{E}\,K^c + \binom{n}{2}\,\mathbb{E}\,J^c \;\le\; \varepsilon_* + \lambda_* + \tfrac{1}{2}\delta_*, \tag{36}$$

which is as small as we need. Hence we may look at the statistic

$$\tilde{\mathbb{T}} \;:=\; \textstyle\sum_{j=1}^n I_j\,T_j + \sum_{j=1}^n K_j\,S_j + \sum_{1\le j<k\le n} J_{j,k}\,T_{j,k},$$

and confine ourselves to the proof of the fact that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\mathbb{T}} \le x) - G^*(x)| \;\ll\; \alpha. \tag{37}$$

To this, we need to determine the Hoeffding decomposition of $\tilde{\mathbb{T}}$. Let

$$T_j^* := I_j\,T_j - \mathbb{E}\,I_1\,T_1, \qquad S_j^* := K_j\,S_j - \mathbb{E}\,K_1 S_1,$$

$$T_{j,k}^* \;:=\; J_{j,k}\,T_{j,k} - \mathbb{E}\,(J_{j,k}\,T_{j,k}\,|\,X_j) - \mathbb{E}\,(J_{j,k}\,T_{j,k}\,|\,X_k) + \mathbb{E}\,J_{1,2}\,T_{1,2}.$$

Next to this, let $\xi_1 := (n-1)\big(\mathbb{E}\,(J_{1,2}\,T_{1,2}\,|\,X_1) - \mathbb{E}\,J_{1,2}\,T_{1,2}\big)$, and, for $2 \le j \le n$,

$$\xi_j \;:=\; (n-1)\big(\mathbb{E}\,(J_{1,j}\,T_{1,j}\,|\,X_j) - \mathbb{E}\,J_{1,2}\,T_{1,2}\big).$$

The Hoeffding decomposition of $\tilde{\mathbb{T}}$ is given by

$$\tilde{T}_j \;=\; T_j^* + S_j^* + \xi_j, \qquad \tilde{T}_{j,k} \;=\; T_{j,k}^*,$$

whereas moreover

$$\mathbb{E}\,\tilde{\mathbb{T}} \;=\; n\,\mathbb{E}\,I_1 T_1 + n\,\mathbb{E}\,K_1 S_1 + \binom{n}{2}\,\mathbb{E}\,J_{1,2}\,T_{1,2}.$$

Since for example $\mathbb{E}\,I_1 T_1 = -\mathbb{E}\,I_1^c T_1$, it is easily checked that

$$|\mathbb{E}\,\tilde{\mathbb{T}}| \;\le\; \varepsilon_* + \lambda^* + \tfrac{1}{2}\delta_*.$$

As $\|G'\|_\infty \ll 1$, the latter leads to the fact that we may prove (37) for $\tilde{\mathbb{T}} - \mathbb{E}\,\tilde{\mathbb{T}}$ instead of $\tilde{\mathbb{T}}$.

Another truncation is in place here. In fact, we more or less need that $|\xi_j| \le 1$ for all j. To this we introduce the indicators

$$L_j \;:=\; \mathbb{I}\{\xi_j^2 \le 1\}.$$

The substitution of all ξ_j by $L_j\,\xi_j$ has a cost that, using Markov's inequality, is bounded by

$$n\,\mathbb{E}\,L_1^c \;=\; n\,\mathbb{P}(|\xi_1|>1) \;\le\; n\,\mathbb{E}\,|\xi_1| \;\ll\; n^2\,\mathbb{E}\,J_{1,2}\,|T_{1,2}| \;\le\; \delta^*, \tag{38}$$

and hence is acceptable to us, so that we may change all ξ_j into

$$\xi_j^* \;:=\; L_j\,\xi_j - \mathbb{E}\,L_1\,\xi_1.$$

Again this involves a shift in the mean, of $n\,\mathbb{E}\,L_1\,\xi_1$, but because of (38) it is negligible.

In short, we are now looking at the statistic

$$\hat{\mathbb{T}} \;:=\; \textstyle\sum_{j=1}^n (T_j^* + S_j^* + \xi_j^*) + \sum_{1\le j<k\le n} T_{j,k}^*,$$

and if we can prove (37) with $\hat{\mathbb{T}}$ instead of $\tilde{\mathbb{T}}$ the proof will be finished. Note that, for all j and $k<l$,

$$|T_j^*|\le 2, \quad |S_j^*|\le 2, \quad |\xi_j^*|\le 2, \quad |T_{k,l}^*|\le 4. \tag{39}$$

Now we look what Theorem 1 can do for us. In fact it is telling us that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\hat{\mathbb{T}}/\hat{s}\le x) - \hat{G}(x)| \;\ll\; \hat{\varepsilon} + \hat{\delta},$$

with $\hat{s}^2 = s^2(\hat{\mathbb{T}})$, $\hat{\varepsilon} = \varepsilon(\hat{\mathbb{T}})$, $\hat{\delta} = \delta(\hat{\mathbb{T}})$ and $\hat{G} = G(\hat{\mathbb{T}})$, cf. (5), (8) and (9). We prove that

$$|\hat{s}^2 - 1| \;\ll\; \lambda^* + \varepsilon^*, \qquad |\hat{s}^2-1| \;\le\; \tfrac12, \tag{40}$$

that

$$\hat{\varepsilon} \;\ll\; \varepsilon^* + \lambda^* + \delta^*, \qquad \hat{\delta} \;\ll\; \delta^*, \tag{41}$$

and that

$$\|G^* - \hat{G}\|_\infty \;\le\; \delta_* + \varepsilon^* + \lambda^* + \delta^*. \tag{42}$$

As a result then

$$\begin{aligned}
|\mathbb{P}(\hat{\mathbb{T}}\le x) - G^*(x)| \;&\le\; |\mathbb{P}(\hat{\mathbb{T}}/\hat{s}\le x/\hat{s}) - \hat{G}(x/\hat{s})| + |\hat{G}(x/\hat{s}) - G^*(x/\hat{s})| \\
&\qquad + |G^*(x/\hat{s}) - G^*(x)| \\
&\ll\; \alpha + |G^*(x/\hat{s}) - G^*(x)|,
\end{aligned}$$

whereas, using a simple Taylor expansion,

$$|G^*(x/\hat{s}) - G^*(x)| \;\ll\; |\hat{s}^{-1}-1| \;\ll\; |1-\hat{s}^2| \;\ll\; \alpha,$$

so that indeed (37) for $\hat{\mathbb{T}}$ instead of $\tilde{\mathbb{T}}$, and the proof of Theorem 2 is finished.

We turn to the proofs of (40), (41) and (42). First we look at

$$\hat{s}^2 \;=\; n\,\mathbb{E}\,(T_1^* + S_1^* + \xi_1^*)^2.$$

Here, using (39) and the fact that,

$$\mathbb{E}\,|S_1^*| \;\le\; 2n^{-1}\,\lambda^*, \qquad \mathbb{E}\,|\xi_1^*| \;\le\; 2\,\mathbb{E}\,L_1\,|\xi_1| \;\le\; 4n^{-1}\,\delta^*, \tag{43}$$

cf. (38), we see that

$$\begin{aligned} |\hat{s}^2 - n\,\mathbb{E}\,(T_1^*)^2| \;&\le\; 2n\,\mathbb{E}\,|T_1^*|\,|S_1^* + \xi_1^*| + n\,\mathbb{E}\,|S_1^* + \xi_1^*|^2 \\ &\le\; 8n\,\mathbb{E}\,|S_1^* + \xi_1^*| \;\le\; 48\,(\lambda^* + \delta^*). \end{aligned}$$

Using that $\mathbb{E}\,|I_1\,T_1| \le 1$, in the same way we see that

$$|n\,\mathbb{E}\,(T_1^*)^2 - n\,\mathbb{E}\,I_1\,T_1^2| \;\le\; 3n\,\mathbb{E}\,I_1^c\,|T_1| \;\le\; 3\varepsilon_*,$$

whereas $|n\,\mathbb{E}\,I_1\,T_1^2 - 1| = n\,\mathbb{E}\,I_1^c\,T_1^2 = \varepsilon_*$. As a result of all of this, (40) easily follows. As to $\hat{\varepsilon}$ we have:

$$\begin{aligned} \hat{\varepsilon} \;&=\; \hat{s}^{-3}\,n\,\mathbb{E}\,|T_1^* + S_1^* + \xi_1^*|^3 \;\ll\; n\,(\mathbb{E}\,|T_1^*|^3 + \mathbb{E}\,|S_1^*|^3 + \mathbb{E}\,|\xi_1^*|^3) \\ &\ll\; n\,\mathbb{E}\,I_1\,|T_1|^3 + n\,\mathbb{E}\,|S_1^*| + n\,\mathbb{E}\,|\xi_1^*| \;\ll\; \varepsilon^* + \lambda^* + \delta^*, \end{aligned}$$

cf. (43). As to $\hat{\delta}$ on the other hand

$$\hat{\delta} \;=\; \hat{s}^{-2}\,n^2\,\mathbb{E}\,|T_{1,2}^*|^2 \;\ll\; n^2\,\mathbb{E}\,J_{1,2}\,T_{1,2}^2 \;\le\; \delta^*,$$

and together indeed (41). Finally we look at

$$\|G^* - \hat{G}\|_\infty \;\le\; \tfrac{1}{2}\,\|\Phi'''\|_\infty\,|\kappa^* - \hat{\kappa}|, \tag{44}$$

with

$$|\kappa^* - \hat{\kappa}| \;=\; n^2\,|\mathbb{E}\,T_1^*\,T_2^*\,T_{1,2}^* - \hat{s}^{-3}\,\mathbb{E}\,\hat{T}_1\,\hat{T}_2\,\hat{T}_{1,2}|.$$

Since $\mathbb{E}\,\hat{T}_1\,\hat{T}_2\,\hat{T}_{1,2} = \mathbb{E}\,\hat{T}_1\,\hat{T}_2\,T_{1,2}^*$, we see that

$$\begin{aligned} |\kappa^* - \hat{\kappa}| \;&\le\; n^2\,|\mathbb{E}\,T_1^*\,T_2^*\,T_{1,2}^* - \mathbb{E}\,\hat{T}_1\,T_2^*\,T_{1,2}^*| + n^2\,|\mathbb{E}\,\hat{T}_1\,T_2^*\,T_{1,2}^* - \mathbb{E}\,\hat{T}_1\,\hat{T}_2\,T_{1,2}^*| \\ &\quad +\, n^2\,|1 - \hat{s}^{-3}|\,|\mathbb{E}\,\hat{T}_1\,\hat{T}_2\,T_{1,2}^*| \;=:\; \rho_1 + \rho_2 + \rho_3. \end{aligned} \tag{45}$$

Here

$$\rho_1 \;=\; n^2\,|\mathbb{E}\,(S_1^* + \xi_1^*)\,T_2^*\,T_{1,2}^*|,$$

whereas, cf. (39) and (43),

$$\mathbb{E}\,(S_1^* + \xi_1^*)^2 \;\ll\; \mathbb{E}\,|S_1^* + \xi_1^*| \;\ll\; n^{-1}\,(\lambda^* + \delta^*).$$

Using Hoeffding's inequality together with the fact that $\mathbb{E}\,(T^*_{1,2})^2 \le 16\,\mathbb{E}\,J_{1,2}\,T^2_{1,2} \ll n^{-2}\,\delta^*$, we have

$$\begin{aligned} \rho_1 &\le n^2\,(\mathbb{E}\,(S^*_1+\xi^*_1)^2)^{1/2}\,(\mathbb{E}\,(T^*_2)^2)^{1/2}\,(\mathbb{E}\,(T^*_{1,2})^2)^{1/2} \\ &\le n^2\,n^{-1/2}\,(\lambda^*+\delta^*)^{1/2}\,n^{-1/2}\,n^{-1}\,(\delta^*)^{1/2} \;\ll\; \lambda^*+\delta^*. \end{aligned}$$

In the same way $\rho_2 \ll \lambda^*+\delta^*$, and since

$$|1-\hat{s}^{-3}| \;=\; \hat{s}^{-3}\,|1-\hat{s}|\,(1+\hat{s}+\hat{s}^2) \;\ll\; |1-\hat{s}^2| \;\ll\; \lambda^*+\varepsilon^*,$$

we see that $\rho_3 \le (\lambda^*+\varepsilon^*)\,n^2\,\mathbb{E}\,|\hat{T}_1\,\hat{T}_2\,T^*_{1,2}| \ll (\lambda^*+\varepsilon^*)\,(\delta^*)^{1/2}$, and indeed

$$|\kappa^*-\hat{\kappa}| \;\ll\; \delta_*+\lambda^*+\varepsilon^*+\delta^*.$$

Looking at (44), this finishes the proof of (42). □

Proof of Corollary 4. The statement that (25) and (26) yield the CLT will be proved as an application of Theorem 2, using as well that, see (13), $|\kappa^*| \le (\delta^*)^{1/2}$. In fact, taking $\theta = 1$ in (25) we see that ε_*, λ_*, $\delta_* \to 0$, whereas by assumption $\lambda^*, \delta^* \to 0$, so we only need to prove that $\varepsilon^*(n) \to 0$. To this, let $0 < \theta \le 1$ be any small positive number. Because of (25) there exists an integer m_θ such that $\varepsilon_*(\theta) = \varepsilon_*(n,\theta) \le \theta^{1/2}$ for all $n \ge m_\theta$. Writing

$$I_1 := \mathbb{I}\{g^2(X) \le \theta s^2 n\}, \qquad I_2 := \mathbb{I}\{\theta s^2 n < g^2(X) \le s^2 n\},$$

then, for $n \ge m_\theta$,

$$\begin{aligned} \varepsilon^*(n) &= n^{-1/2}\,s^{-3}\,\mathbb{E}\,(I_1+I_2)\,|g(X)|^3 \\ &\le (\theta\,s^2 n)^{1/2}\,n^{-1/2}\,s^{-3}\,\mathbb{E}\,g^2 + (s^2 n)^{1/2}\,n^{-1/2}\,s^{-3}\,\mathbb{E}\,I_2\,g(X)^2 \\ &\le \theta^{1/2}+\varepsilon_*(n,\theta) \;\le\; 2\,\theta^{1/2}. \end{aligned}$$

As a consequence $\varepsilon^*(n) \to 0$, which finishes the proof of the first statement.

We show that (27) is stronger than (26). Indeed, from (27) it follows that

$$\lambda^*(n) \;=\; s^{-2/3}\,\mathbb{E}\,|g_0|^{2/3}\,K\,(s^{-1}\,n^{-3/2}\,|g_0|)^{1/3},$$

where $K\,s^{-1}\,n^{-3/2}\,|g_0|$ is bounded and pointwise converging to zero. Since the sequence $s^{-2/3}\,\mathbb{E}\,K\,|g_0(X)|^{2/3}$ is uniformly bounded, as a consequence $\lambda^*(n) \to 0$. In the same way $\delta^*(n) \to 0$, which proves that (27) indeed implies (26).

We turn to the final statement: suppose that the functions g, g_0 and h are independent on n and that $\mathbb{E}\,g^2$, $\mathbb{E}\,|g_0|^{2/3}$, $\mathbb{E}\,|h|^{4/3} < \infty$, and let $\theta > 0$ be fixed. Since $\mathbb{I}\{g^2(X) > \theta\,s^2\,n\}\,g^2(X) \to 0$ pointwise, by Lebesgue's theorem on dominated convergence we immediately see that $\varepsilon_*(\theta) \to 0$. By the same argument

$$\begin{aligned} \lambda_*(\theta) &= n\,\mathbb{E}\,\mathbb{I}\{|g_0(X)|^{2/3} > \theta^{1/3}\,s^{2/3}\,n\} \\ &\le \theta^{-1/3}\,s^{-2/3}\,\mathbb{E}\,\mathbb{I}\{|g_0(X)|^{2/3} > \theta^{1/3}\,s^{2/3}\,n\}\,|g_0(X)|^{2/3} \;\to\; 0, \end{aligned}$$

and

$$\begin{aligned}\delta_*(\theta) &= s^{-1}\,n^{1/2}\,\mathbb{E}\,\mathbb{I}\{|h(X,\overline{X})|^{4/3} > \theta^{2/3}\,s^{4/3}\,n^2\}\,|h(X,\overline{X})| \\ &\le s^{-1}\,n^{1/2}\,(\theta^{2/3}\,s^{4/3}\,n^2)^{-1/4}\,\mathbb{E}\,\mathbb{I}\{|h(X,\overline{X})|^{4/3} > s^{4/3}\,n^2\}\,|h(X,\overline{X})|^{4/3} \;\to\; 0.\end{aligned}$$

It is furthermore clear that the upper limits from (26) are bounded, and as an application of the first part we have the CLT for $\mathbb{T}+\mathbb{S}$. □

Proof of Corollary 5. We apply Theorem 2. Using (20) and the definitions we see that $\varepsilon_*+\varepsilon^*$ and $\lambda_*+\lambda^*$ are small enough for our purposes, so we may concentrate on δ_*, δ^* and $|\kappa^*|$. Here, taking $J=\mathbb{I}\{h^2\le s^2\,n^3\}$ and $J^c = 1-J$ as before,

$$J^c\,|h| \;\le\; J^c\,|h|^{5/3}\,(s^2\,n^3)^{-1/3} \;=\; s^{-2/3}\,n^{-1}\,J^c\,|h|^{5/3}$$

and

$$J\,h^2 \;\le\; J\,|h|^{5/3}\,(s^2\,n^3)^{1/6} \;=\; s^{1/3}\,n^{1/2}\,J\,|h|^{5/3},$$

and it easily follows that

$$\delta_*+\delta^* \;\le\; s^{-5/3}\,n^{-1/2}\,\mathbb{E}\,|h|^{5/3}.$$

Using Hölder's inequality we see that

$$|\kappa^*| \;\le\; n^{-1/2}\,\big(\mathbb{E}\,|s^{-1}\,J\,h(X,\overline{X})|^{5/3}\big)^{3/5}\,\big(\mathbb{E}\,|s^{-1}\,I\,g(X)|^{5/2}\big)^{4/5} \;=:\; n^{-1/2}\,\alpha_1^{3/5}\,\alpha_2^{2/5},$$

introducing the appropriate $\alpha_1,\,\alpha_2\ge 0$. As in general $\alpha_1^{3/5}\,\alpha_2^{2/5}\le\alpha_1+\alpha_2$ for $\alpha_1,\,\alpha_2\ge 0$, as a consequence

$$|\kappa^*| \;\le\; n^{-1/2}\,s^{-5/3}\,\mathbb{E}\,|h|^{5/3} + n^{-1/2}\,\alpha_2,$$

whereas, using again Hölder's inequality,

$$\begin{aligned}\alpha_2 &\le (s^{-5/2}\,\mathbb{E}\,|g|^{5/2})^2 \;=\; s^{-5}\,(\mathbb{E}\,|g|\,|g|^{3/2})^2 \\ &\le s^{-5}\,(\mathbb{E}\,g^2)\,\mathbb{E}\,|g|^3 \;=\; s^{-3}\,\mathbb{E}\,|g|^3,\end{aligned}$$

and the proof of the corollary is easily finished. □

3 Examples

Next we provide some examples which demonstrate the optimality of the bounds.

Example 7. Let $X, X_1, \ldots, X_n$ be real-valued, i.i.d. random variables, with $\mu := \mathbb{E}\,X$. We look at the statistic

$$\mathbb{U} \;:=\; \binom{n}{2}^{-1}\sum_{1\le j<k\le n} X_j X_k,$$

cf. (1). It is clear that $\mathbb{U}$ is an unbiased estimator of the unknown parameter μ^2, that is, $\mathbb{E}\,\mathbb{U} = \mu^2$. The functions g and h from the Hoeffding decomposition (4) are now given by

$$g(x) = \mu\,(x-\mu) \qquad \text{and} \qquad h(x,y) = (x-\mu)\,(y-\mu).$$

Hence, writing

$$a^2 := \mathbb{E}\,(X-\mu)^2 \qquad \text{and} \qquad b := \mathbb{E}\,|X-\mu|^3,$$

the moments corresponding to our model are

$$s^2 = \mu^2\,a^2, \quad \beta = |\mu|^3\,b, \quad \gamma = a^4, \quad \kappa = n^{-1/2}\,\mu^{-1}\,a.$$

By Theorem 1, under the assumption that $0 < \mu^2\,a^2 < \infty$ we have that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}\,\frac{(\mathbb{U}-\mu^2)}{2s} \le x \right) - \Phi(x) - \tfrac{1}{2}\kappa\,\Phi'''(x) \right| \ll \frac{b}{a^3\sqrt{n}} + \frac{a^2}{\mu^2\,n}, \tag{46}$$

a bound depending on μ^2. This bound is precisely of order $\mathcal{O}(n^{-1/2})$ in the case where the sequence b/a^3 is bounded and the sequence $\mu^{-2}\,a^2$ is of order $\mathcal{O}(n^{-1/2})$. The latter will mean moreover that $|\kappa| \le n^{-1/4}$, so that the Edgeworth correction Φ_1 then is of order $\mathcal{O}(n^{-1/4})$. In case $\mu^{-2}\,a^2 = c\,n^{-1/2}$ for some fixed constant $c > 0$, we have indeed that $\|\Phi_1\|_\infty \ge \tilde{c}\,n^{-1/4}$ for some constant $\tilde{c} > 0$, in which case the correction really makes sense.

Example 8. Let $X, X_1, \dots, X_n$ be real-valued, i.i.d. random variables such that $\mathbb{E}\,X = 0$ and $\mathbb{E}\,X^2 = 1$. Now consider the kernel $H(x,y) = xy^2 + x^2\,y$. The corresponding statistic $\mathbb{U}$ is closely connected to Student's statistic. The expectation $\mathbb{E}\,H$ and functions g and h from (4) are of the form

$$\mathbb{E}\,H = 0, \qquad g(x) = x, \qquad h(x,y) = x(y^2-1) + (x^2-1)y.$$

Hence, $\mathbb{T} = \frac{1}{2}\sqrt{n}\,\mathbb{U}$, so that in fact

$$\mathbb{T} \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j + \frac{1}{(n-1)\sqrt{n}} \sum_{1\le j<k\le n} \{X_j(X_k^2-1) + (X_j^2-1)X_k\}. \tag{47}$$

Now let

$$I := \mathbb{I}\{X^2 \le n\}, \quad I^c := 1 - I, \quad \varepsilon_* := \mathbb{E}\,I^c X^2, \qquad \varepsilon^* := n^{-1/2}\,\mathbb{E}\,I|X|^3.$$

From Theorem 1 we derive that

$$\Delta(\mathbb{T}) \;\ll\; \varepsilon_* + \varepsilon^* \;\le\; n^{-1/2}\,\mathbb{E}\,|X|^3. \tag{48}$$

Indeed, to this, we first truncate all X_j by $I_j\,X_j$, I_j being the analogue of I if we replace X by X_j. Looking at Lemma 6 this has a cost of

$$n\,\mathbb{P}(X^2 > n) \;=\; n\,\mathbb{P}(I^c X^2 > n) \;\le\; \mathbb{E}\,I^c X^2 \;=\; \varepsilon_*,$$

and thus is okay. Next we are looking at a statistic $\tilde{\mathbb{T}} = \mathbb{T}(I_1 X_1, \dots, I_n X_n)$, which we need to rewrite. Here we go about as in the proof of Theorem 2. First we decompose $\tilde{\mathbb{T}}$ as

$$\tilde{\mathbb{T}} \;=\; \mathbb{E}\,\tilde{\mathbb{T}} + \textstyle\sum_{j=1}^{n}(T_j^* + \xi_j) + \sum_{1\le j<k\le n} T_{j,k}^*,$$

with $T_j^* = n^{-1/2}\,(I_j\,X_j - \mathbb{E}\,I X)$ and so on. Then we proceed by making the ξ_j bounded by means of indicators $K_j := \mathbb{I}\{|\xi_j| \le 1\}$, changing the ξ_j into $\xi_j^* = K_j\,\xi_j - \mathbb{E}\,K_1\,\xi_1$. To the then obtained statistic $\hat{\mathbb{T}}$ we apply Theorem 1. Since

$$|\hat{s}^2 - 1| \le \tfrac{1}{2}, \quad |\hat{s}^2 - 1| \ll \varepsilon_* + \varepsilon^*, \quad \varepsilon(\hat{\mathbb{T}}) \ll \varepsilon^*, \quad \delta(\hat{\mathbb{T}}) \ll \varepsilon^*,$$

combining the arguments it follows that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T} \le x) - \Phi(x)| \;\ll\; \varepsilon_* + \varepsilon^* + |\kappa(\hat{\mathbb{T}})|,$$

which, since

$$|\kappa(\hat{\mathbb{T}})| \;=\; n^2\,\hat{s}^{-3}\,|\mathbb{E}\,\hat{T}_1\hat{T}_2\,\hat{T}_{1,2}| \;\le\; 8n^2\,\mathbb{E}\,|T_1^* + \xi_1^*|^2\,(\mathbb{E}\,(T_{1,2}^*)^2)^{1/2} \;\ll\; \varepsilon^*,$$

concludes the proof of (48).

An application of Corollary 5, with $\mathbb{E}\,|h|^{5/3}$ in the bound instead of $\mathbb{E}\,h^2$, leads to the fact that

$$\Delta(\mathbb{T}) \;\ll\; \varepsilon_* + n^{-1/2}\,\mathbb{E}\,I|X|^{10/3} \;\le\; n^{-1/2}\,\mathbb{E}\,|X|^{10/3}. \tag{49}$$

Hence, for $\mathbb{T}$ defined by (47), the corollary (48) yields $\Delta(\mathbb{T}) = \mathcal{O}(n^{-1/2})$ under the optimal condition that $\mathbb{E}\,|X|^3 < \infty$, whereas (49) requires that $\mathbb{E}\,|X|^{10/3} < \infty$. Another advantage of Theorem 1 and its corollary (48) is that it implies the CLT under the optimal condition that $0 < \mathbb{E}\,X^2 < \infty$. More generally, we have that $\Delta(\mathbb{T}) = o(n^{-\theta})$ if $0 < \mathbb{E}\,|X|^{2(1+\theta)} < \infty$, for $0 \le \theta < \frac{1}{2}$.

One can introduce statistics similar to (47) corresponding to the kernels

$$g(x) = x, \quad h(x,y) = x\,(y^p - \mathbb{E}\,X^p) + (x^p - \mathbb{E}\,X^p)\,y,$$

with $p = 2, 3, 4, \dots$, just replacing in (47) the 2^{nd} by p^{th} powers. Using a truncation determined by the indicator $I := \mathbb{I}\{|X| \le n^{1/p}\}$, from Theorem 1 it is now derived that

$$\Delta(\mathbb{T}) \;\ll\; n^{-1/2}\,\mathbb{E}\,|X|^{3p/2}, \tag{50}$$

whereas Corollary 5 leads us to the bound

$$\Delta(\mathbb{T}) \;\ll\; n^{-1/2}\,\mathbb{E}\,|X|^{5p/3}, \tag{51}$$

corresponding to (48) and (49) respectively. For the CLT the condition $0 < \mathbb{E}\,|X|^p < \infty$ is sufficient. We omit the calculations leading to (50) and (51) since they are similar to the ones for $p = 2$. It is clear that the result (50) is again better than (51).

4 Proof of Theorem 1

We turn to the proof of Theorem 1, using the general notation described at the end of Section 1. Without loss of generality we may assume that $\mathbb{E}\,\mathbb{T} = 0$ and $s^2 = 1$. From here on, let

$$a \;:=\; \min\{\beta^{-1}\,n^{1/2},\;\gamma^{-1}\,n\}. \tag{52}$$

Under these assumptions, using Lyapunov's inequality,

$$\beta \;\geq\; s^3 \;=\; 1 \quad \text{and} \quad a \;\leq\; \beta^{-1}\,n^{1/2} \;\leq\; n^{1/2}, \tag{53}$$

whereas furthermore

$$\mathbb{E}\,|T_1|^3 = n^{-3/2}\,\beta \;\leq\; n^{-1}\,a^{-1} \quad \text{and} \quad \mathbb{E}\,T_{1,2}^2 = n^{-3}\,\gamma \;\leq\; n^{-2}\,a^{-1}. \tag{54}$$

We will use the inequalities (53) and (54) extensively.

In order to prove the theorem we are going to apply Esseen's smoothing lemma. Before doing this it appears to be wise to look at the conditional variance connected to the kernel h. Let

$$\varphi_1^2 \;:=\; \mathbb{E}\,(T_{1,2}^2\,|\,X_1) \quad \text{and} \quad \varphi_j^2 \;:=\; \mathbb{E}\,(T_{1,j}^2\,|\,X_j),$$

for $2 \leq j \leq n$, and let $0 < \theta \leq 10^{-4}$ be some fixed positive number. Without loss of generality we assume as well that

$$a \;\geq\; 10^3\,\theta^{-1},$$

and we will prove Theorem 1 under the condition that, for any j,

$$\varphi_j^2 \;\leq\; \theta\,n^{-1}. \tag{55}$$

Using a truncation argument on $\mathbb{T}$, in the following subsection we will show that for all our purposes we may indeed make this final assumption.

4.1 Truncation of $\mathbb{T}$

We show that it suffices to prove Theorem 1 in the case where $\varphi_j \leq \theta\,n^{-1}$ for all j, for some fixed $0 < \theta \leq 10^{-4}$. Indeed, assume that we have the mentioned result. We want to prove that, for some constant $c(\theta) > 0$, we have the bound

$$\Gamma(\mathbb{T}) \;\leq\; c(\theta)\,a^{-1}. \tag{56}$$

To this let, for $1 \leq j \leq n$,

$$I_j \;=\; I_j(X_j) \;:=\; \mathbb{I}\{\varphi_j^2 \leq \tfrac{1}{4}\theta\,n^{-1}\}, \qquad I_j^c \;:=\; 1 - I_j,$$

and define

$$\hat{\mathbb{T}} \;:=\; \mathbb{T}_1 + \textstyle\sum_{1\le j<k\le n} I_j I_k\, T_{j,k}$$

with Hoeffding decomposition

$$\hat{\mathbb{T}} \;=\; \mathbb{E}\,\hat{\mathbb{T}} + \textstyle\sum_{j=1}^n \hat{T}_j + \sum_{1\le j<k\le n} \hat{T}_{j,k} \;=:\; \mathbb{E}\,\hat{\mathbb{T}} + \hat{\mathbb{T}}_1 + \hat{\mathbb{T}}_2.$$

Let $\hat{s}^2 := \operatorname{var}\hat{\mathbb{T}}_1$. Without loss of generality we may assume again that

$$\mathbb{E}\,\mathbb{T} = 0, \quad s^2 = 1, \qquad \text{and} \qquad a \;\ge\; 10^3\,\theta^{-1}. \tag{57}$$

We have the following:

Lemma 9. *Assume that* (57) *applies. Then*

$$\Gamma(\mathbb{T}) \;\ll\; \theta^{-1}\,a^{-1} + \Gamma(\hat{\mathbb{T}}). \tag{58}$$

Moreover we have $|\mathbb{E}\,\hat{\mathbb{T}}| \le 2\,\theta^{-1}\,a^{-3/2}$*, and, for all* $1 \le j \le n$ *and* $1 \le j < k \le n$*,*

$$\tfrac{1}{2} \le\; \hat{s}^2 \;=\; n\,\mathbb{E}\,\hat{T}_j^2 \;\le \tfrac{3}{2}, \qquad \mathbb{E}\,|\hat{T}_j|^3 \;\le\; 4\,\mathbb{E}\,|T_j|^3 + 12\,n^{-1}\,a^{-3/2}, \tag{59}$$

$$\mathbb{E}\,\hat{T}_{j,k}^2 \;\le\; 4\,\mathbb{E}\,T_{j,k}^2 \qquad \text{and} \qquad \mathbb{E}\,(\hat{T}_{j,k}^2\,|\,X_j),\; \mathbb{E}\,(\hat{T}_{j,k}^2\,|\,X_k) \;\le\; \theta\,n^{-1}. \tag{60}$$

As a consequence of the right part of (60), the restricted theorem is applicable to $\hat{\mathbb{T}}$, thus yielding a bound of the form

$$\Gamma(\hat{\mathbb{T}}) \;\le\; c(\theta)\,\big(n\,\mathbb{E}\,|\hat{T}_1|^3\,\hat{s}^{-3} + n^2\,\mathbb{E}\,\hat{T}_{1,2}^2\,\hat{s}^{-2}\big).$$

In turn, using (59) and (60), this bound is seen to be bounded from above by (a constant times) the desired bound a^{-1} for $\Gamma(\mathbb{T})$, plus an extra term of order $a^{-3/2} \le a^{-1}$. From (58) the desired bound (56) then immediately follows, and this shows that we may indeed restrict our attention to the theorem under the condition.

We turn to the proof of the lemma.

Proof of Lemma 9. We denote the moment κ and expansion G, cf. (10) and (9), corresponding to $\hat{\mathbb{T}}$ by $\hat{\kappa}$ and $\hat{G}$. First we notice that, for all j, by Markov's inequality

$$\mathbb{P}(I_j = 0) \;\le\; (\tfrac{1}{4}\theta\,n^{-1})^{-1}\,\mathbb{E}\,T_{1,2}^2 \;\le\; 4\,\theta^{-1}\,n^{-1}\,a^{-1}.$$

This implies that

$$\mathbb{P}(\mathbb{T} \ne \hat{\mathbb{T}}) \;\le\; \mathbb{P}(I_j = 0,\ \text{some } j) \;\le\; 4n\,\theta^{-1}\,n^{-1}\,a^{-1} \;=\; 4\,\theta^{-1}\,a^{-1},$$

so that, using Lemma 6, for any $x \in \mathbb{R}$,

$$|\mathbb{P}(\mathbb{T} \le x) - G(x)| \ \ll\ \theta^{-1}\, a^{-1} + |\mathbb{P}(\hat{\mathbb{T}} \le x) - G(x)|. \tag{61}$$

Now we turn to $\hat{\mathbb{T}}$'s Hoeffding decomposition. First note that

$$|\mathbb{E}\,\hat{\mathbb{T}}| \ =\ \tbinom{n}{2}\, |\mathbb{E}\, I_1 I_2\, T_{1,2}| \ =\ \tbinom{n}{2}\, |\mathbb{E}\, I_1^c\, I_2^c\, T_{1,2}|,$$

since for example $\mathbb{E}\, I_2\, T_{1,2} = \mathbb{E}\, I_2\, \mathbb{E}\,(T_{1,2}\,|\,X_2) = 0$, and with Hölder's inequality

$$\begin{aligned} |\mathbb{E}\,\hat{\mathbb{T}}| \ &\le\ \tfrac{1}{2} n^2\, (\mathbb{E}\, I_1^c\, \mathbb{E}\, I_2^c)^{1/2}\, (\mathbb{E}\, T_{1,2}^2)^{1/2} \\ &\le\ \tfrac{1}{2} n^2\, \mathbb{P}(I_1 = 0)\, (n^{-2}\, a^{-1})^{1/2} \\ &\le\ 2n^2\, \theta^{-1}\, n^{-1}\, a^{-1}\, n^{-1}\, a^{-1/2} \ =\ 2\, \theta^{-1}\, a^{-3/2}. \end{aligned} \tag{62}$$

Now let

$$\xi_j \ :=\ (n-1)\, \big(\mathbb{E}\,(I_1 I_j\, T_{1,j}\,|\,X_j) - \mathbb{E}\, I_1 I_2\, T_{1,2}\big)$$

for $2 \le j \le n$, and $\xi_1 := (n-1)\, \big(\mathbb{E}\,(I_1\, I_2\, T_{1,2}\,|\,X_1) - \mathbb{E}\, I_1 I_2\, T_{1,2}\big)$. Following the definitions, it is easily seen that

$$\hat{T}_j \ =\ T_j + \xi_j.$$

Looking for example at the moments of ξ_1, we see that

$$\begin{aligned} \mathbb{E}\, \xi_1^2 \ &=\ (n-1)^2\, \big(\mathbb{E}\, \mathbb{E}^{\,2}(I_1 I_2\, T_{1,2}\,|\,X_1) - \mathbb{E}^{\,2} I_1 I_2\, T_{1,2}\big) \\ &\le\ n^2\, \mathbb{E}\, I_1\, \mathbb{E}^{\,2}(I_2^c\, T_{1,2}\,|\,X_1) \end{aligned}$$

(again using the fact that $\mathbb{E}\,(T_{1,2}\,|\,X_1) = 0$). Here with Hölder's inequality

$$\mathbb{E}^{\,2}(I_2^c\, T_{1,2}\,|\,X_1) \ \le\ \mathbb{E}\,(T_{1,2}^2\,|\,1)\, \mathbb{E}\, I_2^c \ \le\ 4\theta^{-1}\, n^{-1}\, a^{-1}\, \mathbb{E}\,(T_{1,2}^2\,|\,X_1), \tag{63}$$

so that

$$\mathbb{E}\, \xi_1^2 \ \le\ 4\, \theta^{-1}\, n^{-1}\, a^{-1}\, n^2\, \mathbb{E}\, T_{1,2}^2 \ \le\ 4\, \theta^{-1}\, n^{-1}\, a^{-2}.$$

As a consequence we have that

$$\begin{aligned} |\mathbb{E}\, \hat{T}_1^2 - n^{-1}| \ &=\ |2\, \mathbb{E}\, T_1\, \xi_1 + \mathbb{E}\, \xi_1^2| \\ &\le\ 2\, (n^{-1})^{1/2}\, (4\, \theta^{-1}\, n^{-1}\, a^{-2})^{1/2} + 4\, \theta^{-1}\, n^{-1}\, a^{-2} \\ &\le\ 4\, \theta^{-1}\, n^{-1}\, a^{-1}\, (\theta^{1/2} + a^{-1}) \ \le\ \theta^{-1}\, n^{-1}\, a^{-1} \ \le\ 10^{-3} n^{-1}, \end{aligned}$$

which proves the first part of (59). As a result we also have that

$$|\hat{s}^2 - 1| \ \le\ \textstyle\sum_{j=1}^n |\mathbb{E}\, \hat{T}_j^2 - n^{-1}| \ \le\ 3\, \theta^{-1}\, a^{-1}, \tag{64}$$

and using (62), (64) and (61), a standard Slutzky type argument is leading us to (58), with G instead of $\hat{G}$. In order to get to (58) it will now suffice to show that

$$\|G-\hat{G}\|_\infty \;\ll\; \theta^{-1}\,a^{-1}. \tag{65}$$

We postpone the proof of (65) until the end of the proof.

As to the third moments: using the general inequality

$$(|a|+|b|)^r \;\le\; 2^{r-1}\,(|a|^r+|b|^r), \tag{66}$$

for $r\ge 1$, we see that

$$\begin{aligned}\mathbb{E}\,|\xi_1|^3 &\le n^3\,4\left(\mathbb{E}\,|\mathbb{E}\,(I_1I_2\,T_{1,2}\,|\,X_1)|^3+|\mathbb{E}\,I_1I_2\,T_{1,2}|^3\right)\\ &\le 8\,n^3\,\mathbb{E}\,I_1\,\mathbb{E}^{\,2}(I_2\,T_{1,2}\,|\,X_1)\,|\mathbb{E}\,(I_2\,T_{1,2}\,|\,X_1)|.\end{aligned}$$

With Hölder's inequality and I_1's definition,

$$\begin{aligned}I_1\,|\mathbb{E}\,(I_2\,T_{1,2}\,|\,X_1)| &= I_1\,|\mathbb{E}\,(I_2^c\,T_{1,2}\,|\,X_1)| \;\le\; I_1\,\mathbb{E}\,(T_{1,2}^2\,|\,X_1)^{1/2}\,\mathbb{E}^{\,1/2}I_2^c\\ &\le (4\theta^{-1}\,n^{-1}\,a^{-1})^{1/2}\,(\theta\,n^{-1})^{1/2} \;=\; 2n^{-1}\,a^{-1/2},\end{aligned}$$

so that, using as well (63),

$$\begin{aligned}\mathbb{E}\,|\xi_1|^3 &\le 2^6\,n^3\,n^{-1}\,a^{-1/2}\,\theta^{-1}\,n^{-1}\,a^{-1}\,\mathbb{E}\,T_{1,2}^2\\ &= 2^6\,\theta^{-1}\,n^{-1}\,a^{-5/2} \;\le\; 2^6\cdot\tfrac{1}{10^3}\,n^{-1}\,a^{-3/2} \;\le\; \tfrac{1}{16}n^{-1}\,a^{-3/2}.\end{aligned}$$

Using again (66) it is then easily seen that the second part of (59) is correct as well.

We turn to the non-linear part of $\hat{\mathbb{T}}$. For all $1\le j<k\le n$ it is easily seen that

$$\hat{T}_{j,k} \;=\; I_jI_k\,T_{j,k}-\mathbb{E}\,(I_jI_k\,T_{j,k}\,|\,X_j)-\mathbb{E}\,(I_jI_k\,T_{j,k}\,|\,X_k)+\mathbb{E}\,I_1I_2\,T_{1,2}.$$

As to $\hat{T}_{1,2}$ we have the following. Taking

$$p_{12}:=I_1I_2\,T_{1,2},\quad p_1:=\mathbb{E}\,(p_{12}\,|\,X_1),\quad p_2:=\mathbb{E}\,(p_{12}\,|\,X_2)\quad\text{and}\quad p:=\mathbb{E}\,p_{12},$$

we see that

$$\begin{aligned}\mathbb{E}\,\hat{T}_{1,2}^2 &= \mathbb{E}\,(p_{12}-p_1-p_2+p)^2\\ &= \mathbb{E}\,(p_{12}-p_1)^2-2\,\mathbb{E}\,(p_{12}-p_1)\,(p_2-p)+\mathbb{E}\,(p_2-p)^2\\ &= \mathbb{E}\,(p_{12}-p_1)^2-\mathbb{E}\,(p_2-p)^2\\ &\le 2\,(\mathbb{E}\,p_{12}^2+\mathbb{E}\,p_1^2) \;\le\; 4\,\mathbb{E}\,p_{12}^2 \;\le\; 4\,\mathbb{E}\,T_{1,2}^2.\end{aligned}$$

Concentrating on $\mathbb{E}\,(\hat{T}_{1,2}^2\,|\,X_1)$ we see in a similar way that

$$\begin{aligned}\mathbb{E}\,(\hat{T}_{1,2}^2\,|\,X_1) &= \mathbb{E}\,((p_{12}-p_2)^2-2\,(p_{12}-p_2)\,(p_1-p)+(p_1-p)^2\,|\,X_1)\\ &= \mathbb{E}\,((p_{12}-p_2)^2\,|\,X_1)-(p_1-p)^2 \;\le\; 2\left(\mathbb{E}\,(p_{12}^2\,|\,X_1)+\mathbb{E}\,p_2^2\right)\\ &\le 2\left(\mathbb{E}\,(p_{12}^2\,|\,X_1)+\mathbb{E}\,\mathbb{E}\,(p_{12}^2\,|\,X_1)\right) \;\le\; 4\cdot\tfrac{1}{4}\theta\,n^{-1} \;=\; \theta\,n^{-1},\end{aligned}$$

using the fact that by definition $\mathbb{E}\,(p_{12}^2 \,|\, X_1) \le I_1\,\mathbb{E}\,(\hat{T}_{12}^2 \,|\, X_1) \le \frac{1}{4}\theta\, n^{-1}$. The same applies for any $\mathbb{E}\,\hat{T}_{j,k}^2$, $\mathbb{E}\,(\hat{T}_{j,k} \,|\, X_j)$ and $\mathbb{E}\,(\hat{T}_{j,k} \,|\, X_k)$, which proves (60).

Finally we give a proof of (65). To this we take a look at

$$\begin{aligned}
|\kappa - \hat{\kappa}| &= n^2\,|\mathbb{E}\,T_1T_2\,T_{1,2} - \mathbb{E}\,\hat{T}_1\hat{T}_2\,\hat{T}_{1,2}| \\
&\le n^2\,|\mathbb{E}\,(T_{1,2} - \hat{T}_{1,2})\,T_1T_2| + n^2\,|\mathbb{E}\,\hat{T}_{1,2}\,(T_1 - \hat{T}_1)\,T_2| \\
&\quad + n^2\,|\mathbb{E}\,\hat{T}_{1,2}\,\hat{T}_1\,(T_2 - \hat{T}_2)| + \ n^2\,|1 - \hat{s}^{-3}|\,|\mathbb{E}\,\hat{T}_1\,\hat{T}_2\,\hat{T}_{1,2}| \\
&=: \eta_1 + \eta_2 + \eta_3 + \eta_4.
\end{aligned}$$

Here

$$\begin{aligned}
\eta_1 &= n^2\,|\mathbb{E}\,(1 - I_1I_2)\,T_1T_2\,T_{1,2}| \\
&\le n^2\,\textstyle\sum_{k=1}^2 |\mathbb{E}\,I_k^c\,T_1T_2\,T_{1,2}| + n^2\,|\mathbb{E}\,I_1^c\,I_2^c\,T_1T_2\,T_{1,2}|,
\end{aligned}$$

where with Hölder's inequality

$$\begin{aligned}
|\mathbb{E}\,I_1^c\,T_1T_2\,T_{1,2}| &\le (\mathbb{E}\,I_1^c\,T_1^2)^{1/2}\,(\mathbb{E}\,T_2^2)^{1/2}\,(\mathbb{E}\,T_{1,2}^2)^{1/2} \\
&\le (\mathbb{E}\,I_1^c)^{1/6}\,(\mathbb{E}\,|T_1|^3)^{1/3}\,(n^{-1})^{1/2}\,(n^{-2}\,a^{-1})^{1/2} \\
&\le (4\theta^{-1}\,n^{-1}\,a^{-1})^{1/6}\,n^{-1/2}\,\beta^{1/3}\,n^{-3/2}\,a^{-1/2} \\
&\ll \theta^{-1/6}\,n^{-2}\,a^{-1/6}\,a^{-1/3}\,a^{-1/2} \ = \ \theta^{-1/6}\,n^{-2}\,a^{-1}
\end{aligned}$$

and, in the same way,

$$\begin{aligned}
|\mathbb{E}\,I_1^c\,I_2^c\,T_1T_2\,T_{1,2}| &\le (\mathbb{E}\,I_1^c\,T_1^2)\,(\mathbb{E}\,T_{1,2}^2)^{1/2} \\
&\ll (\theta^{-1}\,n^{-1}\,a^{-1})^{1/3}\,n^{-1}\,\beta^{2/3}\,(n^{-2}\,a^{-1})^{1/2} \ \le \ \theta^{-1/3}\,n^{-2}\,a^{-1},
\end{aligned}$$

so that $\eta_1 \ll \theta^{-1/3}\,a^{-1}$. Moreover, using the above we have that

$$\begin{aligned}
\eta_3 &\le n^2\,(\mathbb{E}\,\hat{T}_{1,2}^2)^{1/2}\,(\mathbb{E}\,\hat{T}_1^2)^{1/2}\,(\mathbb{E}\,\xi_2^2)^{1/2} \\
&\ll n^2\,(n^{-2}\,a^{-1})^{1/2}\,(n^{-1})^{1/2}\,(\theta^{-1}\,n^{-1}\,a^{-2})^{1/2} \ = \ \theta^{-1/2}\,a^{-3/2},
\end{aligned}$$

and in the same way $\eta_2 \ll \theta^{-1/2}\,a^{-1}$. As to η_4, using (64) we see that

$$|1 - \hat{s}^{-3}| \ = \ \hat{s}^{-3}\,(\hat{s}^2 + \hat{s} + 1)\,|\hat{s} - 1| \ \ll \ |\hat{s}^2 - 1| \ \ll \ \theta^{-1}\,a^{-1},$$

and with Hölder's inequality

$$\eta_4 \ \ll \ \theta^{-1}\,a^{-1}\,a^{-1/2} \ \ll \ \theta^{-1}\,a^{-3/2}.$$

As a result

$$|\kappa - \hat{\kappa}| \ \ll \ \theta^{-1}\,a^{-1},$$

and

$$\|G - \hat{G}\|_\infty \ \ll \ \theta^{-1}\,a^{-1}\,\|\Phi'''\|_\infty \ \ll \ \theta^{-1}\,a^{-1}.$$

This finishes the proof of (65), and hence of the lemma. $\square$

4.2 Expansion of the characteristic function on $[0,1]$

From here on, let g be the Fourier transform of G, and let

$$f(t) \;:=\; \mathbb{E}\, e\{t\mathbb{T}\}.$$

In order to prove (12) we note that Esseen's smoothing lemma is telling us that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(\mathbb{T}\le x) - G(x)| \;\ll\; a^{-1} + \textstyle\int_0^a |f(t)-g(t)|\, t^{-1}\, dt. \tag{67}$$

See, for example, Feller (1971), Chapter XVI, Lemma 3.2. This leaves us the estimation of the integral on the right side of the inequality. To this, taking τ uniformly distributed on $(0,1)$ and independent of the sample, we will expand our characteristic function f as

$$f(t) \;=\; f_1(t) + f_2(t) + f_3(t), \tag{68}$$

where

$$f_1(t) \;:=\; \mathbb{E}\, e\{t\mathbb{T}_1\}, \qquad f_2(t) \;:=\; (it)\,\mathbb{E}\,\mathbb{T}_2\, e\{t\mathbb{T}_1\} \tag{69}$$

and

$$f_3(t) \;:=\; (it)^2\,\mathbb{E}\,(1-\tau)\,\mathbb{T}_2^2\, e\{t\,(\mathbb{T}_1+\tau\,\mathbb{T}_2)\}. \tag{70}$$

We concentrate first on the case where t is a real number on the interval $[0,1]$, which may be dealt with in a rather simple way.

Lemma 10. *Assume that $\mathbb{E}\,\mathbb{T}=0$, $s^2=1$, and $a\ge 10^3\,\theta^{-1}$. In case $0\le t\le 1$ we have that*

$$|f(t)-g(t)| \;\ll\; a^{-1}\, t. \tag{71}$$

Proof of Lemma 10. First we remark that, for $0\le t\le 1$,

$$|f_3(t)| \;\le\; \tfrac12 t^2\,\mathbb{E}\,\mathbb{T}_2^2 \;\le\; \tfrac12 a^{-1}\, t^2, \tag{72}$$

which is small enough for our purposes. Now let $Z_1,\dots,Z_n$ be an i.i.d. sample of random variables that are $N(0,\frac1n)$-distributed, independent of the sample. Writing $Z := \sum_{j=1}^n Z_j$, we have that Z is standard normally distributed. Now

$$g(t) \;=\; \mathbb{E}\, e\{tZ\} + \tfrac12\,(it)^3\,\kappa\,\mathbb{E}\, e\{tZ\} \;=:\; g_1(t)+g_2(t),$$

and we will prove that $f_1(t)\approx g_1(t)$ and $f_2(t)\approx g_2(t)$.

The first approximation is a standard result, which may for example be derived by gradually changing the $\sum_{j=1}^n T_j$ in the exponent into $\sum_{j=1}^n Z_j$, one term at a time. In this way one has to perform n estimations, which via Taylor expansions may be proven to be bounded by

$$\tfrac{1}{6}\,(\beta\, n^{-1/2})\, n^{-1}\, t^3\, \exp\{-\tfrac{1}{4}t^2\} \;\le\; \tfrac{1}{6}\, a^{-1}\, n^{-1}\, t^3\, \exp\{-\tfrac{1}{4}t^2\},$$

so that in turn

$$|f_1(t) - g_1(t)| \;\ll\; a^{-1}\, t^3\, \exp\{-\tfrac{1}{4}t^2\}, \tag{73}$$

a bound which is very useful as well for large t. See, for example, Alberink (1999), Lemma 3.4, or Bentkus et al. (2000). Moreover we have that

$$\begin{aligned} f_2(t) &= \binom{n}{2}\,(it)\,\mathbb{E}\, T_{1,2}\, e\{t\mathbb{T}_1\} \\ &\approx \tfrac{1}{2}\,(it)^3\,\kappa\,\mathbb{E}\, e\{t(\mathbb{T}_1 - (T_1 + T_2))\} \;=:\; \chi(t). \end{aligned}$$

In order to get from f_2 to χ, here one needs to exchange the term $\mathbb{E}\, T_{1,2}\, e\{t(T_1+T_2)\}$ by $(it)^2\,\mathbb{E}\, T_1 T_2\, T_{1,2}$. To this one may use that

$$e\{tT_j\} \;=\; 1 + (it)\, T_j + R_j, \qquad \text{with } \; |R_j| \le t^{3/2}\, |T_j|^{3/2},$$

and it is easily seen that

$$|f_2(t) - \chi(t)| \;\ll\; a^{-1}\, t^3\, (1 + t^{1/2})\, \exp\{-\tfrac{1}{4}t^2\},$$

see again Alberink (1999) (Lemma 3.5). As for f_1 it is seen that

$$|\chi(t) - g_2(t)| \;\ll\; a^{-1}\, t^4\, \exp\{-\tfrac{1}{4}t^2\},$$

and as a result we then have that

$$|f_2(t) - g_2(t)| \;\ll\; a^{-1}\, t^3\, (1 + t)\, \exp\{-\tfrac{1}{4}t^2\}, \tag{74}$$

after which (71) is an easy consequence of (72), (73) and (74). □

4.3 Expansion of the characteristic function on $[1, a^{1/6}]$

We turn to the expansion of the characteristic function on the interval $[1, a^{1/6}]$. We have the following result.

Lemma 11. *Assume that $\mathbb{E}\,\mathbb{T} = 0$, $s^2 = 1$, $a \ge 10^3\,\theta^{-1}$ and $t \ge 1$. Then there exists a constant $c > 0$ such that*

$$|f(t) - g(t)| \;\ll\; a^{-3/2}\, t^3 + a^{-1}\, t^4\, \exp\{-c\, t^2\}.$$

Proof of Lemma 11. We start again from the expansion (68). Since from (73) and (74) we already have that

$$|f_1(t) + f_2(t) - g(t)| \;\ll\; a^{-1}\,(1 + t^4)\, \exp\{-\tfrac{1}{4}t^2\},$$

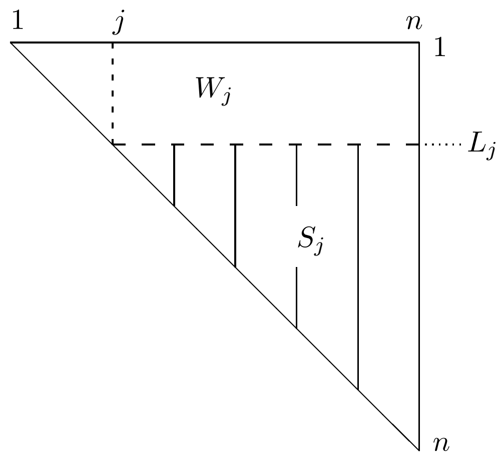

Figure 1: Decomposition of $\mathbb{T}_2$ into lines L_j, triangles S_j and complements W_j.

it will suffice to prove that

$$|f_3(t)| \ll a^{-3/2}\, t^3 + a^{-1}\, t^2 \exp\{-c_0\, t^2\} \tag{75}$$

for some constant $c_0 > 0$. To this, for $1 \le j \le n$, we introduce the lines, triangles and complements

$$L_j := \textstyle\sum_{k=j+1}^{n} T_{j,k}, \quad S_j := \sum_{k=j}^{n-1} L_k \quad \text{and} \quad W_j := \sum_{k=1}^{j} L_k = \mathbb{T}_2 - S_{j+1}, \tag{76}$$

see Figure 1.

Now let

$$\mathbb{T}_2^2 \;=\; U_1 + U_2 + U_3, \tag{77}$$

with

$$U_1 \;:=\; \textstyle\sum_{A:\,|A|=2} T_A^2, \quad U_2 \;:=\; \sum_{A,B:\,|A|=|B|=2,\,|A\cap B|=1} T_A\, T_B$$

and

$$U_3 \;:=\; \textstyle\sum_{A,B:\,|A|=|B|=2,\,A\cap B=\emptyset} T_A\, T_B.$$

For $k = 1, 2, 3$, we look in turn at

$$h_k(t) \;:=\; |(it)^2\, \mathbb{E}\, (1-\tau)\, U_k\, e\{t\, (\mathbb{T}_1 + \tau\, \mathbb{T}_2)\}|.$$

We start by looking at

$$h_1(t) \;=\; t^2 \tbinom{n}{2}\, |\mathbb{E}\, (1-\tau)\, T_{1,2}^2\, e\{t\, (\mathbb{T}_1 + \tau\, \mathbb{T}_2)\}|.$$

Using a short Taylor expansion in terms of S_3, we see that

$$\begin{aligned} h_1(t) &\le \tfrac{1}{2}n^2\,t^2\,|\mathbb{E}\,(1-\tau)\,T_{1,2}^2\,e\{t\,(\mathbb{T}_1+\tau\,(L_1+L_2))\}|+h_5(t) \\ &=: h_4(t)+h_5(t), \end{aligned} \tag{78}$$

with

$$\begin{aligned} |h_5(t)| &\le \tfrac{1}{2}n^2\,t^3\,\mathbb{E}\,\tau(1-\tau)\,T_{1,2}^2\,|S_3| \qquad (79) \\ &\le \tfrac{1}{4}n^2\,t^3\,(\mathbb{E}\,T_{1,2}^2)\,(\mathbb{E}\,S_3^2)^{1/2} \;\ll\; a^{-1}\,a^{-1/2}\,t^3 \;=\; a^{-3/2}\,t^3, \end{aligned}$$

which is small enough for our purposes, so that we may concentrate on h_4. As to h_4 we have that

$$\begin{aligned} &|\mathbb{E}\,(1-\tau)\,T_{1,2}^2\,e\{t\,(\mathbb{T}_1+\tau\,(L_1+L_2))\}| \\ &\quad\le \mathbb{E}\,(1-\tau)\,T_{1,2}^2\,|e\{t\,(T_1+T_2)+t\,\tau\,T_{1,2}\}| \\ &\qquad\qquad |\mathbb{E}\,(e\{t\,\textstyle\sum_{j=3}^{n}(T_j+\tau\,T_{1,j}+\tau\,T_{2,j})\}\,|\,X_1,X_2,\tau)| \\ &\quad\le \mathbb{E}\,T_{1,2}^2\,|Y_{1,2,\tau}|^{n-2}, \end{aligned} \tag{80}$$

with

$$Y_{1,2,\tau} \;:=\; \mathbb{E}\,(e\{t\,(T_3+\tau\,T_{1,3}+\tau\,T_{2,3})\}\,|\,X_1,X_2,\tau). \tag{81}$$

Now let $\tilde{c}:=10^3\,\theta^{-1}$, for which we typically have that $a\ge\tilde{c}$. We take

$$c_1 \;:=\; 1-3\theta^{-1}\,\tilde{c}^{-1}-\tfrac{4}{3}\,\tilde{c}^{-5/6}\,(1+8\,\theta^{-1}\,\tilde{c}^{-3/2}) \tag{82}$$

and

$$c_2 \;:=\; \tfrac{1}{2}\,(c_1-8\,(\theta+\tilde{c}^{-2})^{1/2}-16\,\theta), \tag{83}$$

and note that under the assumptions

$$\tfrac{99}{100}\le c_1\le 1 \qquad\text{and}\qquad \tfrac{45}{100}\le c_2\le\tfrac{1}{2}.$$

We will prove that, as a result of the truncation,

$$|Y_{1,2,\tau}| \;\le\; 1-c_2\,n^{-1}\,t^2 \;\le\; \exp\{-c_2\,n^{-1}\,t^2\}. \tag{84}$$

As a result of (84) then

$$\mathbb{E}\,T_{1,2}^2\,|Y_{1,2,\tau}|^{n-2} \;\le\; \exp\{-c_3\,t^2\}\,4\,n^{-2}\,a^{-1}, \tag{85}$$

taking $c_3:=(1-2n^{-1})\,c_2$, and, combining (78), (79), (80) and (85),

$$h_1(t) \;\ll\; a^{-3/2}\,t^3+a^{-1}\,t^2\,\exp\{-c_3\,t^2\}.$$

We turn to the proof of (84). In fact, using a Taylor expansion in terms of $\tau(T_{1,3}+T_{2,3})$ we see that

$$\begin{aligned} |Y_{1,2,\tau}| \;&\le\; |\mathbb{E}\, e\{t\, T_3\}| + t\,\tau\, |\mathbb{E}\,((T_{1,3}+T_{2,3})\, e\{tT_3\} \,|\, 1,2)| \\ &\qquad + \tfrac12\, t^2\, \tau^2\, \mathbb{E}\,((T_{1,3}+T_{2,3})^2 \,|\, X_1, X_2) \;=:\; \psi_1(t) + \psi_2(t) + \psi_3(t). \end{aligned}$$

Here clearly (see Lemma 9)

$$\psi_3(t) \;\le\; \tfrac12 t^2\, 4\, \mathbb{E}\,(T_{1,3}^2 \,|\, X_1) \;\le\; 8\, t^2\, \theta\, n^{-1}. \tag{86}$$

As to $\psi_1(t)$ a three-term Taylor expansion is telling us that

$$\psi_1(t) \;\le\; |1 - \tfrac12 t^2\, \mathbb{E}\, T_3^2| + \tfrac16 t^3\, \mathbb{E}\, |T_3|^3.$$

By Lemma 9 we have that

$$\begin{aligned} \tfrac12 t^2\, \mathbb{E}\, T_3^2 \;&\le\; \tfrac12\, (1 + 3\,\theta^{-1}\, a^{-1})\, n^{-1}\, t^2 \\ &\le\; \tfrac12\, (1 + 3\,\theta^{-1}\, a^{-1})\, n^{-1}\, a^{1/3} \;\le\; \tfrac12\, (1 + 3\,\theta^{-1}\, a^{-1})\, n^{-5/6} \;\le\; 1, \end{aligned}$$

and using again the lemma it is then easily seen that

$$\begin{aligned} \psi_1(t) \;&\le\; 1 - \tfrac12\, (1 - 3\,\theta^{-1}\, a^{-1})\, n^{-1}\, t^2 + \tfrac16\, a^{1/6}\, t^2\, 4\, (\mathbb{E}\, |T_j|^3 + \mathbb{E}\, |\xi_j|^3) \\ &\le\; 1 - \tfrac12\, n^{-1}\, t^2\, (1 - 3\,\theta^{-1}\, a^{-1} - \tfrac43\, \tilde c^{1/6}\, (\tilde c^{-1} + 8\,\theta^{-1}\, \tilde c^{-5/2})) \\ &\le\; 1 - \tfrac12\, c_1\, n^{-1}\, t^2. \end{aligned} \tag{87}$$

As to $\psi_2(t)$, using Hölder's inequality we easily see that

$$\begin{aligned} \psi_2(t) \;&\le\; t^2\, \mathbb{E}\,(|T_{1,3}+T_{2,3}|\, |T_3| \,|\, X_1, X_2) \\ &\le\; t^2\, \mathbb{E}\,(|T_{1,3}+T_{2,3}|^2 \,|\, 1,2)^{1/2}\, (\mathbb{E}\, T_3^2)^{1/2} \\ &\le\; t^2\, (4^2\, \theta\, n^{-1})^{1/2}\, (n^{-1} + \theta^{-1}\, n^{-1}\, a^{-2})^{1/2} \;\le\; 4\, t^2\, n^{-1}\, (\theta + a^{-2})^{1/2}, \end{aligned} \tag{88}$$

and using (86), (87) and (88), the proof of (84) is easily concluded.

We turn to h_2. To this, let

$$U_{2,1} \;:=\; \textstyle\sum_{A,B:\, |A|=|B|=2,\, A\cap B=\{1\}} T_A\, T_B.$$

Now

$$h_2(t) \;=\; n\, t^2\, |\mathbb{E}\, (1-\tau)\, U_{2,1}\, e\{t\, (\mathbb{T}_1 + \tau\, \mathbb{T}_2)\}|. \tag{89}$$

We notice that

$$\mathbb{E}\,(U_{2,1}^2 \,|\, X_1) \;=\; \tbinom{n-1}{2}\, \mathbb{E}\,(T_{1,2}^2\, T_{1,3}^2 \,|\, X_1) \;\le\; \tfrac12 n^2\, \mathbb{E}\,(T_{1,2}^2 \,|\, X_1)^2, \tag{90}$$

since for example $\mathbb{E}\, T_{1,2}\, T_{1,3}\, T_{1,2}\, T_{1,4} \;=\; \mathbb{E}\, T_{1,2}\, T_{1,3}\, T_{1,2}\, \mathbb{E}\,(T_{1,4} \,|\, X_1, X_2, X_3) \;=\; 0$, so that removal of S_2 from the expression $\mathbb{T}_2$ in (89) will cost no more than

$$\begin{aligned} n\, t^3\, \mathbb{E}\, \tau(1-\tau)\, |U_{2,1}|\, |S_2| &\le \tfrac{1}{2} n\, t^3\, \mathbb{E}\, \mathbb{E}\,(U_{2,1}^2 \,|\, X_1)^{1/2}\, \mathbb{E}\,(S_2^2 \,|\, X_1)^{1/2} \\ &\le 2^{-3/2}\, n^2\, t^3\, (\mathbb{E}\, S_2^2)^{1/2}\, \mathbb{E}\, \mathbb{E}\,(T_{1,2}^2 \,|\, X_1) \\ &\le 2^{-3/2}\, n^2\, a^{-1/2}\, 4n^{-2}\, a^{-1}\, t^3 \;\le\; 2^{1/2}\, a^{-3/2}\, t^3, \end{aligned} \tag{91}$$

which is acceptable. Hence we look at (89) with $\mathbb{T}_2$ replaced by L_1, and taking

$$U_{2,1}^* \;:=\; \textstyle\sum_{1<j,k\le[\frac{1}{2}n]\,,\, j\ne k} T_{1,j}\, T_{1,k},$$

we see that

$$\begin{aligned} h_2(t) &\le 2^{1/2}\, a^{-3/2}\, t^3 + n\, 2 \tbinom{n-1}{2} \tfrac{1}{2} \tbinom{[\frac{1}{2}n]-1}{2}^{-1} t^2\, |\mathbb{E}\,(1-\tau)\, U_{2,1}^*\, e\{t\,(\mathbb{T}_1 + \tau\, L_1)\}| \\ &=: 2^{1/2}\, a^{-3/2}\, t^3 + h_6(t). \end{aligned} \tag{92}$$

Taking

$$Y_{1,\tau} \;:=\; \mathbb{E}\,(e\{t\,(T_n + \tau\, T_{1,n})\} \,|\, X_1, \tau), \tag{93}$$

here

$$\begin{aligned} h_6(t) &\ll n\, t^2\, \mathbb{E}\,(1-\tau)\, |\mathbb{E}\,(U_{2,1}^*\, e\{t \textstyle\sum_{j=2}^{[\frac{1}{2}n]} (T_j + \tau\, T_{1,j})\} \,|\, X_1, \tau)|\, |Y_{1,\tau}|^{n-[\frac{1}{2}n]} \\ &\ll n\, t^2\, \mathbb{E}\, \mathbb{E}\,(|U_{2,1}^*| \,|\, X_1)\, |Y_{1,\tau}|^{n-[\frac{1}{2}n]}, \end{aligned}$$

where, see (90),

$$\mathbb{E}\,(|U_{2,1}^*| \,|\, X_1) \;\le\; \mathbb{E}\,((U_{2,1}^*)^2 \,|\, X_1)^{1/2} \;\le\; n\, \mathbb{E}\,(T_{1,2}^2 \,|\, X_1),$$

and, as in (84), $|Y_{1,\tau}| \le \exp\{-c_2\, n^{-1}\, t^2\}$,

$$\begin{aligned} h_6(t) &\ll n^2\, t^2\, \mathbb{E}\, \mathbb{E}\,(T_{1,2}^2 \,|\, X_1)\, \exp\{-c_2\, (n - [\tfrac{1}{2}n])\, n^{-1}\, t^2\} \\ &\ll a^{-1}\, t^2\, \exp\{-\tfrac{1}{2} c_2\, t^2\}. \end{aligned} \tag{94}$$

From (92) and (94) we conclude that

$$h_2(t) \;\ll\; a^{-3/2}\, t^3 + a^{-1}\, t^2\, \exp\{-\tfrac{1}{2} c_2\, t^2\}.$$

As to $h_3(t)$ we have that

$$U_3 \;=\; \textstyle\sum_{1\le j<k<p<q\le n} V_{j,k,p,q} \tag{95}$$

with

$$V_{j,k,p,q} \;:=\; 2\,(T_{j,k}\, T_{p,q} + T_{j,p}\, T_{k,q} + T_{j,q}\, T_{k,p}).$$

The $V_{j,k,p,q}$ are mutually uncorrelated, so that, using the inequality (66) and Lemma 9,

$$\begin{aligned}\mathbb{E}\,U_3^2 &= \textstyle\sum_{1\le j<k<p<q\le n}\mathbb{E}\,V_{j,k,p,q}^2 \;=\; \binom{n}{4}\,\mathbb{E}\,V_{1,2,3,4}^2 \\ &\le \tfrac14 n^4\,4\cdot 3^2\,\mathbb{E}\,T_{1,2}^2\,T_{3,4}^2 \;\ll\; a^{-2}. \end{aligned}\tag{96}$$

Because of this we may remove $\tau\,\mathbb{T}_2$ from the exponent directly, the cost of this removal being bounded by

$$t^3\,\mathbb{E}\,\tau\,(1-\tau)\,|U_3|\,|\mathbb{T}_2| \;\ll\; t^3\,(\mathbb{E}\,U_3^2)^{1/2}\,(\mathbb{E}\,\mathbb{T}_2^2)^{1/2} \;\ll\; a^{-3/2}\,t^3,$$

so that

$$h_3(t) \;\ll\; a^{-3/2}\,t^3 + t^2\,|\mathbb{E}\,(1-\tau)\,U_3\,e\{t\,\mathbb{T}_1\}| \;=:\; a^{-3/2}\,t^3 + h_7(t).$$

Taking

$$U_3^* \;:=\; \textstyle\sum_{1\le j<k<p<q\le[\frac12 n]} V_{j,k,p,q},$$

we have that

$$h_7(t) \;=\; t^2\,\binom{n}{4}\binom{[\frac12 n]}{4}^{-1}|\mathbb{E}\,(1-\tau)\,U_3^*\,e\{t\,\mathbb{T}_1\}|,$$

which as for $h_6(t)$ leads to the fact that

$$h_7(t) \;\ll\; t^2\,(\mathbb{E}\,|U_3^*|)\,|\mathbb{E}\,e\{t\,T_n\}|^{n-[\frac12 n]} \;\ll\; a^{-1}\,t^2\,\exp\{-\tfrac12 c_2\,t^2\}.$$

In turn

$$h_3(t) \;\ll\; a^{-3/2}\,t^3 + a^{-1}\,t^2\,\exp\{-\tfrac12 c_2\,t^2\},$$

and the proof of (75) is easily concluded. This finishes the proof of the lemma. □

4.4 Expansion of the characteristic function on $[a^{1/6}, a]$

Finally we turn to the interval $[a^{1/6}, a]$. Here we have the following:

Lemma 12. *Assume that $\mathbb{E}\,\mathbb{T} = 0$, $s^2 = 1$, $a \ge 10^3\,\theta^{-1}$ and $t \ge 1$. There exists a constant $c > 0$ such that*

$$|g(t)| \;\ll\; (1 + a^{-1/2}\,t^3)\,\exp\{-\tfrac12\,t^2\}$$

and

$$|f(t)| \;\ll\; t^{-6} + a^{-1/2}\,(\log t)\,t^{-5} + a^{-1}\,(\log^{3/2} t)\,t^{-1} + a^{-3/2}\,\log^{3/2} t.$$

Proof of Lemma 12. We first notice that

$$|\kappa| \;\le\; (\gamma\,n^{-1})^{1/2} \;\le\; a^{-1/2},$$

from which it easily follows that

$$|g(t)| \ll (1 + a^{-1/2}\,t^3)\,\exp\{-\tfrac{1}{2}\,t^2\},$$

as desired.

Now let L_j, S_j and W_j be as in (76), c_1 and c_2 as in (82) and (83), and

$$m \;=\; m(t) \;:=\; [6\,c_2^{-1}\,(\log t)\,t^{-2}\,n].$$

We notice that under the assumptions $100 \le m \le \frac{1}{4}n$. Since $c_2 \le \frac{1}{2}c_1$ and $c_1 \le 1$, using (87) we have that

$$\begin{aligned} |\mathbb{E}\,e\{t\,T_1\}|^{m-2} &\le \exp\{-\tfrac{1}{2}\,c_1\,(m-2)\,n^{-1}\,t^2\} \\ &\le \exp\{-6\log t + \tfrac{3}{2}c_1\,n^{-1}\,t^2\} \;\ll\; t^{-6}. \end{aligned} \tag{97}$$

Instead of (68), we use an expansion for $f(t)$ in powers of W_m, thus looking at

$$f(t) \;=\; f_1(t) + f_2(t) + f_3(t)$$

with

$$f_1(t) \;:=\; \mathbb{E}\,e\{t\,(\mathbb{T}_1 + S_{m+1})\}, \qquad f_2(t) \;:=\; (it)\,\mathbb{E}\,W_m\,e\{t\,(\mathbb{T}_1 + S_{m+1})\},$$

and

$$f_3(t) \;:=\; (it)^2\,\mathbb{E}\,(1-\tau)\,W_m^2\,e\{t\,(\mathbb{T}_1 + S_{m+1} + \tau\,W_m)\}.$$

As to f_1 it is easy to see that, as a result of (97),

$$\begin{aligned} |f_1(t)| &= |\mathbb{E}\,e\{t\,T_1\}|^m\,|\mathbb{E}\,e\{t\,(\textstyle\sum_{j=m+1}^n T_j + S_{m+1})\}| \\ &\le |\mathbb{E}\,e\{t\,T_1\}|^{m-2} \;\ll\; t^{-6}, \end{aligned}$$

which is small enough for our purposes. As to f_2, for $l = 2, m+1$, let

$$V_l \;:=\; e\{t\,(\mathbb{T}_1 - (T_1 + T_l) + S_{m+1})\}.$$

Using the Taylor expansion $e\{t\,T_j\} = 1 + R_j$ with $|R_j| \le t\,|T_j|$, together with independence, we see that

$$\begin{aligned} f_2(t) &= (it)\,\tbinom{m}{2}\,\big(\mathbb{E}\,T_{1,2}\,e\{t\,(T_1 + T_2)\}\big)\,\mathbb{E}\,V_2 \\ &\qquad + (it)\,m\,(n-m)\,\mathbb{E}\,T_{1,m+1}\,e\{t\,(T_1 + T_{m+1})\}\,V_{m+1} \\ &= (it)\,\tbinom{m}{2}\,(\mathbb{E}\,T_{1,2}\,R_1\,R_2)\,\mathbb{E}\,V_2 + (it)\,m\,(n-m) \\ &\qquad \times \mathbb{E}\,T_{1,m+1}\,R_1\,(1 + R_{m+1})\,V_{m+1} \;=:\; f_4(t) + f_5(t). \end{aligned} \tag{98}$$

Using Hölder's inequality, Lemma 9, (97) and independence, we see that

$$\begin{aligned} |f_4(t)| &\ll m^2\,t\,\mathbb{E}\,|T_{1,2}\,R_1\,R_2|\,|\mathbb{E}\,V_2| \\ &\ll m^2\,t^3\,(\mathbb{E}\,T_{1,2}^2)^{1/2}\,(\mathbb{E}\,T_1^2)\,|\mathbb{E}\,e\{t\,T_3\}|^{m-2} \\ &\ll (mn^{-1})^2\,a^{-1/2}\,t^{-3} \;\ll\; a^{-1/2}\,(\log^2 t)\,t^{-7}. \end{aligned} \tag{99}$$

As to

$$
\begin{aligned}
f_5(t) &= (it)\, m\,(n-m)\,\big(\mathbb{E}\, T_{1,m+1}\, R_1\, V_{m+1} + \mathbb{E}\, T_{1,m+1}\, R_1\, R_{m+1}\, V_{m+1}\big) \\
&=: f_6(t) + f_7(t) \qquad (100)
\end{aligned}
$$

we have that

$$
\begin{aligned}
|f_7(t)| &\le mn\, t\, |\mathbb{E}\, e\{t\, T_2\}|^{m-1}\, t^2\, \mathbb{E}\, |T_{1,m+1}\, T_1\, T_{m+1}| \\
&\ll mn^{-1}\, a^{-1/2}\, t^{-3} \ll a^{-1/2}\, (\log t)\, t^{-5}, \qquad (101)
\end{aligned}
$$

cf. (99). On the other hand

$$
f_6(t) = (it)\, m(n-m)\, (\mathbb{E}\, e\{t\, T_2\})^{m-1}\, \mathbb{E}\, T_{1,m+1}\, R_1\, e\{t\, (L_{m+1} + S_{m+2} + \textstyle\sum_{j=m+2}^{n} T_j)\},
$$

an expression which would vanish if the term L_{m+1} were not there in the exponent. Removal by means of an expansion thus shows us that, see (97),

$$
\begin{aligned}
|f_6(t)| &\ll mn\, t\, t^{-6}\, t^2\, \mathbb{E}\, |T_{1,m+1}\, T_1\, L_{m+1}| \\
&\le mn\, t^{-3}\, (\mathbb{E}\, T_{1,m+1}^2)^{1/2}\, (\mathbb{E}\, T_1^2)^{1/2}\, (\mathbb{E}\, L_{m+1}^2)^{1/2} \\
&\le mn^{-1}\, t^{-3}\, a^{-1} \ll a^{-1}\, (\log t)\, t^{-5}, \qquad (102)
\end{aligned}
$$

and following (98), (99), (100), (101) and (102) it is easily seen that $|f_2(t)|$ is small enough for our purposes.

We turn to f_3, for which we will do the following. We use the decompositions

$$
W_m = P_1 + P_2 + P_3 \qquad \text{and} \qquad S_{m+1} = P_4 + P_5 + P_6,
$$

laid down by

$$
P_1 := \textstyle\sum_{1 \le j < k \le m} T_{j,k}, \qquad P_2 := \sum_{j=1}^{m} \sum_{k=m+1}^{n-m} T_{j,k},
$$

$$
P_4 := \textstyle\sum_{j=m+1}^{n-m} \sum_{k=n-m+1}^{n} T_{j,k}, \qquad P_5 := \sum_{n-m+1 \le j < k \le n} T_{j,k}\, :
$$

see Figure 2. Moreover, in order to denote the tails of a line L_j, for all $1 \le j < k \le n$ we write

$$
L_{j,k} := \textstyle\sum_{l=k}^{n} T_{j,l}.
$$

Now we go about as follows. First we will remove P_1 from the expression W_m altogether. After this we write out $(W_m - P_1)^2$ like $\mathbb{T}_2^2$, see (77), and perform three separate estimations.

As to the removal of P_1 we note that, writing

$$
Q := (1-\tau)\, e\{t\, (\mathbb{T}_1 + S_{m+1} + \tau\, W_m)\},
$$

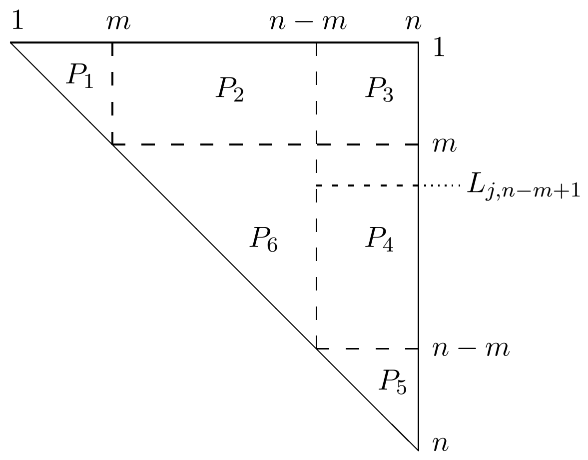

Figure 2: Decomposition of $\mathbb{T}_2$ into $P_1 + \ldots + P_6$.

we have a decomposition of form

$$f_3(t) \;=\; f_8(t) + f_9(t) + f_{10}(t)$$

with

$$f_8(t) \;:=\; (it)^2\,\mathbb{E}\,(P_2+P_3)^2\,Q, \quad f_9(t) \;:=\; (it)^2\,\mathbb{E}\,P_1\,(P_2+P_3)\,Q$$

and

$$|f_{10}(t)| \;\le\; \tfrac{1}{2}t^2\,\mathbb{E}\,P_1^2 \;\le\; \tfrac{1}{2}t^2\,(mn^{-1})^2\,a^{-1} \;\ll\; a^{-1}\,(\log^2 t)\,t^{-2}.$$

Furthermore

$$\begin{aligned} |f_9(t)| \;&\le\; \tfrac{1}{2}t^2\,(\mathbb{E}\,P_1^2)^{1/2}\,(\mathbb{E}\,(P_2+P_3)^2)^{1/2} \\ &\le\; \tfrac{1}{2}\,(mn^{-1})^{3/2}\,t^2\,a^{-1} \;\ll\; a^{-1}\,(\log^{3/2} t)\,t^{-1}, \end{aligned}$$

so that we may indeed concentrate on f_8.

As to f_8 we have that

$$(P_2+P_3)^2 \;=\; U_1 + U_2 + U_3,$$

taking

$$\begin{aligned} U_1 \;&:=\; \textstyle\sum_{j=1}^{m}\sum_{k=m+1}^{n} T_{j,k}^2, \\ U_2 \;&:=\; \textstyle\sum_{j=1}^{m}\sum_{k=m+1}^{n}\sum_{1\le p\le m<q\le n\,:\,\text{either } p=j \text{ or } q=k} T_{j,k}\,T_{p,q}, \\ U_3 \;&:=\; \textstyle\sum_{j=1}^{m}\sum_{k=m+1}^{n}\sum_{1\le p\le m<q\le n\,:\,p\ne j \text{ and } q\ne k} T_{j,k}\,T_{p,q}, \end{aligned}$$

cf. (77). Now let, for $k = 1,2,3$,

$$h_k(t) \;:=\; (it)^2\,\mathbb{E}\,U_k\,Q.$$

We have that

$$f_8(t) \;=\; h_1(t) + h_2(t) + h_3(t),$$

and start by looking at

$$h_1(t) \;=\; (it)^2\, m(n-m)\, \mathbb{E}\,(1-\tau)\, T_{1,m+1}^2\, e\{t\,(\mathbb{T}_1 + S_{m+1} + \tau\, W_m)\}.$$

Parallel to (78), here we remove $P_4 + P_5 - L_{m+1,n-m+1}$ and $P_3 - L_{1,n-m+1}$ from the S_{m+1} and W_m in the exponent by means of a Taylor expansion, thus obtaining a decomposition of the form

$$h_1(t) \;=\; h_4(t) + h_5(t),$$

where, using independence,

$$\begin{aligned} |h_5(t)| \;&\le\; t^3\, mn\, (\mathbb{E}\, T_{1,m+1}^2) \left(\mathbb{E}\,|P_3 - L_{1,n-m+1}| + \mathbb{E}\,|P_4 + P_5 - L_{m+1,n-m+1}|\right) \\ &\ll\; t^3\, mn^{-1}\, a^{-1}\, (mn^{-1}\, a^{-1})^{1/2} \;\ll\; a^{-3/2}\, \log^{3/2} t, \end{aligned}$$

which is small enough. As to h_4, conditioning on the variables $\tau, X_1, \ldots, X_{n-m}$ leads us to the fact that

$$\begin{aligned} |h_4(t)| \;&\le\; t^2\, mn\, \mathbb{E}\, T_{1,m+1}^2\, |\mathbb{E}\,(e\{t \textstyle\sum_{j=n-m+1}^{n} (T_j + \tau\, T_{1,j} + T_{m+1,j})\} \,|\, \tau, 1, \ldots, n-m)| \\ &=\; mn\, t^2\, \mathbb{E}\, T_{1,m+1}^2\, |Y_{1,m+1}|^m \end{aligned}$$

with

$$Y_{1,m+1} \;:=\; \mathbb{E}\,(e\{t\,(T_n + \tau\, T_{1,n} + T_{m+1,n})\} \,|\, \tau, X_1, X_{m+1}),$$

cf. (81). As in (84) we have $|Y_{1,m+1}| \le \exp\{-c_2\, n^{-1}\, t^2\}$, and hence

$$|h_4(t)| \;\le\; mn^{-1}\, t^2\, a^{-1}\, \exp\{-c_2\, mn^{-1}\, t^2\} \;\ll\; a^{-1}\, (\log t)\, t^{-6},$$

and it follows that $|h_1(t)|$ is small enough for our purposes.

We turn to h_2. Here let

$$U_{2,1} \;:=\; 2 \textstyle\sum_{m+1 \le k < l \le n} T_{1,k}\, T_{1,l}, \quad U_{2,m+1} \;:=\; 2 \textstyle\sum_{1 \le k < l \le m} T_{k,m+1}\, T_{l,m+1},$$

so that

$$h_2(t) \;=\; m\,(it)^2\, \mathbb{E}\, U_{2,1}\, Q + (n-m)\,(it)^2\, \mathbb{E}\, U_{2,m+1}\, Q.$$

Taking

$$U_{2,1}^* \;:=\; 2 \textstyle\sum_{m+1 \le k < l \le n-m} T_{1,k}\, T_{1,l},$$

we have that

$$\begin{aligned} h_2(t) &= \binom{n-m}{2}\binom{n-2m}{2}^{-1} m\,(it)^2\,\mathbb{E}\,U^*_{2,1}\,Q + (n-m)\,(it)^2\,\mathbb{E}\,U_{2,m+1}\,Q \\ &=: h_6(t) + h_7(t). \end{aligned}$$

First we look at $h_6(t)$. Noting that, cf. (90),

$$\mathbb{E}\,((U^*_{2,1})^2 \,|\, X_1) \;\le\; \tfrac{1}{2}n^2\,\mathbb{E}\,(T^2_{1,2}\,|\,X_1)^2,$$

as for h_1 we may remove $P_4 + P_5$ from S_{m+1} and $P_3 - L_{1,n-m+1}$ from W_m respectively. Indeed, the removal is corresponding to a decomposition

$$h_6(t) \;=\; h_8(t) + h_9(t),$$

with, cf. (91),

$$\begin{aligned} |h_9(t)| &\ll m\,t^3\,\mathbb{E}\,\mathbb{E}\,((U^*_{2,1})^2\,|\,X_1)^{1/2}\,\mathbb{E}\,((P_4 + P_5 + \tau\,(P_3 - L_{1,n-m+1}))^2\,|\,X_1)^{1/2} \\ &\le m\,t^3\,(\mathbb{E}\,(P_4 + P_5 + \tau\,(P_3 - L_{1,n-m+1}))^2)^{1/2}\,n\,\mathbb{E}\,T^2_{1,2} \\ &\le mn^{-1}\,a^{-1}\,t^3\,(mn^{-1}\,a^{-1})^{1/2} \;\ll\; a^{-3/2}\,\log^{3/2} t. \end{aligned}$$

As to h_8 on the other hand we may condition on the random variables τ, $X_1, \ldots, X_{n-m}$, thus removing $P_2 + P_6$ as well from the exponent: taking $Y_{1,\tau}$ as in (93) we see that

$$\begin{aligned} |h_8(t)| &\ll m\,t^2\,\mathbb{E}\,(1-\tau)\,|U^*_{2,1}|\,|Y_{1,\tau}|^m \\ &\ll m\,t^2\,\exp\{-c_2\,mn^{-1}\,t^2\}\,\mathbb{E}\,|U^*_{2,1}| \\ &\ll mn^{-1}\,a^{-1}\,t^2\,t^{-6} \;=\; a^{-1}\,(\log t)\,t^{-6}, \end{aligned}$$

cf. (84). As to h_7 we may use the same arguments, first removing $P_4 + P_5 - L_{m+1,n-m+1}$ from S_{m+1} and P_3 from W_m respectively, and then using independence of $T_{n-m+1}, \ldots, T_n$ of the remaining variables. Using the fact that

$$\mathbb{E}\,(U^2_{2,m+1}\,|\,X_{m+1}) \;\ll\; m^2\,\mathbb{E}\,(T^2_{1,m+1}\,|\,X_{m+1})^2,$$

we see that

$$|h_7(t)| \;\ll\; a^{-3/2}\,\log^{3/2} t + a^{-1}\,(\log t)\,t^{-6},$$

which finishes the estimation of $|h_2(t)|$.

We finally turn to the estimation of h_3. Taking

$$V_{j,k,p,q} \;:=\; 2\,(T_{j,p}\,T_{k,q} + T_{j,q}\,T_{k,p}),$$

here we have

$$U_3 \;=\; \textstyle\sum_{1\le j<k\le m} \sum_{m+1\le p<q\le n} V_{j,k,p,q},$$

cf. (95). The $V_{j,k,p,q}$ are (again) mutually uncorrelated, and, as in (96),

$$\begin{aligned}\mathbb{E}\, U_3^2 &\le 4 \binom{m}{2} \binom{n-m}{2} 2^2\, \mathbb{E}\, T_{1,m+1}^2\, T_{2,m+2}^2 \\ &\le 4\, m^2\, n^2\, (n^{-2}\, a^{-1})^2 \;=\; 4\, (mn^{-1})^2\, a^{-2}.\end{aligned}$$

Taking

$$U_3^* \;:=\; \textstyle\sum_{1\le j<k\le m} \sum_{m+1\le p<q\le n-m} V_{j,k,p,q},$$

instead of at h_3, we may look at h_{10} with U_3 replaced by U_3^*. As to h_{10}, as before we may remove $P_4 + P_5 + \tau\, P_3$ from the exponent by means of a simple Taylor expansion, in fact using a decomposition

$$h_{10}(t) \;=\; h_{11}(t) + h_{12}(t),$$

with

$$\begin{aligned}|h_{12}(t)| &\ll t^3\, \mathbb{E}\, |U_3^*|\, |P_4 + P_5 + \tau\, P_3| \\ &\ll t^3\, mn^{-1}\, a^{-1}\, (mn^{-1})^{1/2}\, a^{-1/2} \;\ll\; a^{-3/2} \log^{3/2} t.\end{aligned}$$

The independence of the linear parts $T_{n-m+1}, \ldots, T_n$ and the remaining expressions U_3^*, P_2 and P_6 leads us to the fact that

$$|h_{11}(t)| \;\ll\; t^{-6}\, \mathbb{E}\, |U_3^*| \;\ll\; a^{-1}\, (\log t)\, t^{-8},$$

and it is easily seen that $|h_{10}(t)|$, and hence $|h_3(t)|$, is small enough. Collecting the estimates, the proof of the lemma is now easily completed. □

Collecting the results of the Lemmas 10, 11 and 12, and using Esseen's smoothing lemma (67), the proof of (12) is now easily concluded. The replacement of s by σ, in the case where $\sigma^2 < \infty$, is easily obtained from the original bound. In fact, writing $\tilde{\mathbb{T}} := \mathbb{T} - \mathbb{E}\, \mathbb{T}$, since $\Gamma \ll \varepsilon + \delta$ we have that

$$\begin{aligned}|\mathbb{P}(\tilde{\mathbb{T}}/\sigma \le x) - G(x)| &\le |\mathbb{P}(\tilde{\mathbb{T}}/s \le s^{-1}\sigma\, x) - G(s^{-1}\sigma\, x)| + |G(s^{-1}\sigma\, x) - G(x)| \\ &\ll \varepsilon + \delta + |s^{-1}\sigma - 1|.\end{aligned}$$

Here

$$|s^{-1}\sigma - 1| \;=\; s^{-1}\, (\sigma + s)^{-1}\, |\sigma^2 - s^2| \;\le\; s^{-2}\, n^{-1}\, \mathbb{E}\, h^2 \;=\; \delta,$$

cf. (6). Moreover, denoting G and κ with s replaced by σ by G_σ and κ_σ, using Hölder's inequality and (6), it is easily seen that

$$\|G - G_\sigma\| \;\ll\; \delta^{3/2}.$$

As a consequence, if $\delta \le 1$ then $s^{-1}\sigma \le 1 + \delta \le 2$, and denoting Γ, ε and δ with s replaced by σ by Γ_σ, ε_σ and δ_σ, it easily follows that

$$\Gamma_\sigma \;\ll\; \varepsilon + \delta \;\le\; 8\, (\varepsilon_\sigma + \delta_\sigma).$$

In case $\delta > 1$ on the other hand, we have that $(\mathbb{E}\, \mathbb{T}_2^2)/\sigma^2 \ge \frac{1}{4}\, s^2/\sigma^2$. Since $\sigma^2 = s^2 + \mathbb{E}\, \mathbb{T}_2^2$ this means that $(\mathbb{E}\, \mathbb{T}_2^2)/\sigma^2 \ge \frac{1}{5}$, so that in turn, cf. (6), $\delta_\sigma \ge \frac{1}{5}$, and hence $\Gamma_\sigma \le 1 \le 5\, \delta_\sigma$, which concludes the argument.

References

[1] I.B. Alberink, A Berry-Esseen bound for U-statistics in the non-i.i.d. case, to appear in: *J. Theoret. Probab.* (2000a).

[2] I.B. Alberink, *Berry-Esseen bounds for arbitrary statistics*, 2000b (dissertation).

[3] I.B. Alberink, V. Bentkus, Bounds for the concentration of asymptotically normal statistics, 1999, *Acta Applic. Math. 58, p.11-59.*

[4] V. Bentkus, F. Götze, On minimal moment assumptions in Berry-Esseen theorems for U-statistics, 1995, *Theory Probab. Appl. 40, no. 3, p.430-455.*

[5] V. Bentkus, F. Götze, Optimal results of convergence in the CLT for quadratic forms, 1996, *Ann. Prob. 24, p.466-490.*

[6] V. Bentkus, F. Götze and R. Zitikis, Lower estimate for the convergence rate for U-statistics, 1994, *Ann. Prob. 22, p.1707-1714.*

[7] V. Bentkus, F. Götze, V. Paulauskas, A. Račkauskas, The accuracy of Gaussian approximation in Banach spaces, *Limit Theorems of Probability Theory*, (Yu.V. Prokhorov, V.A. Statulevičius, eds.), p.25-111, Springer Verlag, 2000, Berlin.

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

[9] P. Bickel, Edgeworth expansion in nonparametric statistics, 1974, *Ann. Statist. 4, p.108-156.*

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

[11] R.N. Bhattacharya, R. Ranga Rao, *Normal approximation and asymptotic expansions*, Robert E. Krieger Publishing Co., 1986, Melbourne.

[12] Yu.V. Borovskikh, *U-statistics in Banach spaces*, VSP, 1996, Utrecht.

[13] Yu.V. Borovskikh, On normal approximations to U-statistics, to appear in: *Probab. Theory and Appl.* (2000).

[14] Yu.V. Borovskikh, V.S. Korolyuk, *Asymptotic Analysis of Distributions of Statistics*, Naukova Dumka, 1984, Kiev (in Russian).

[15] Yu.V. Borovskikh, V.S. Korolyuk, Approximation of non-degenerate U-statistics, 1985, *Theory Probab. Appl. 30, p.439-450.*

[16] Yu.V. Borovskikh, V.S. Korolyuk, *Theory of U-statistics*, Kluwer Academic Publishers, 1994, Dordrecht.

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

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

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

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

[21] 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.

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

[23] A.J. Lee, *U-statistics. Theory and practice*, Marcel Dekker, 1990, New York.

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

[25] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, 1980, New York.

[26] C.P. Shapiro, L. Hubert, Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions, 1979, *Ann. Stat. 7, p.788-794.*

[27] 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