Math-Net.Ru

Общероссийский математический портал

I. B. Alberink, V. Yu. Bentkus, Lyapunov-Type
Bounds for U -Statistics, *Теория вероятн. и ее при-
мен.*, 2001, том 46, выпуск 4, 724–743

DOI: <https://doi.org/10.4213/tvp3797>

Использование Общероссийского математического портала Math-
Net.Ru подразумевает, что вы прочитали и согласны с пользователь-
ским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 52.6.47.48

26 апреля 2023 г., 14:43:48

© 2001 г.

ALBERINK I. B.*, BENTKUS V.**

LYAPUNOV TYPE BOUNDS FOR U -STATISTICS¹⁾

Пусть $X_1, \dots, X_n$ — независимые одинаково распределенные случайные величины. Оптимальная оценка Ляпунова (или Берри–Эссеена) получена для U -статистики степени 2, т.е. статистики вида $\sum_{j < k} H(X_j, X_k)$, где H — измеримая, симметричная функция, такая, что $\mathbf{E} |H(X_1, X_2)| < \infty$, в предположении, что эта статистика невырождена.

Ключевые слова и фразы: U -статистика, оценка типа Ляпунова, оценка Берри–Эссеена, скорость сходимости, нормальная аппроксимация.

1. Introduction and results. Let $X_1, \dots, X_n$ be independent, identically distributed (i.i.d.) random variables taking their values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. By $X, \bar{X}$ we 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

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

where $H: \mathcal{X}^2 \rightarrow \mathbf{R}$ is a symmetric kernel such that $\mathbf{E} |H(X, \bar{X})| < \infty$. We allow dependence of the sample and the statistic, and as well of the functions g and h defined below, on n . (In other words, we write $(\mathcal{X}, \mathcal{B})$, $X_1, \dots, X_n$, H and so on instead of $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)})$, $X_1^{(n)}, \dots, X_n^{(n)}$, $H^{(n)}$, respectively.) Writing $\mathbf{E} H := \mathbf{E} H(X, \bar{X})$ and introducing the functions

$$g(x) := \mathbf{E} H(X, x) - \mathbf{E} H, \quad h(x, y) := H(x, y) - g(x) - g(y) - \mathbf{E} H, \quad (2)$$

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

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

*Department of Mathematics, University of Nijmegen, Postbus 9010, 6500 GL, Nijmegen, The Netherlands, e-mail: alberink@sci.kun.nl

**Institute of Mathematics and Informatics, Akademijos 4, 2600 Vilnius, Lithuania, e-mail: bentkus@takas.lt

¹⁾ Research supported by the SFB 343 in Bielefeld.

with

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

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

$$0 < s^2 := \mathbf{E} \mathbf{T}_1^2 = \mathbf{E} g^2 < \infty, \quad (5)$$

which means that $\mathbf{T}$ is a nondegenerate U -statistic. All parts of the Hoeffding decomposition (4) are uncorrelated, so that we may write $\sigma^2 := \mathbf{E} \mathbf{T}^2 = \mathbf{E} \mathbf{T}_1^2 + \mathbf{E} \mathbf{T}_2^2 = s^2 + \mathbf{E} \mathbf{T}_2^2$, with

$$\frac{\mathbf{E} h^2}{2n} \leq \mathbf{E} \mathbf{T}_2^2 = \frac{\mathbf{E} h^2}{2(n-1)} \leq \frac{\mathbf{E} h^2}{n}. \quad (6)$$

The (normed) linear part $\mathbf{T}_1/s$ 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 [18], [23], [10] for classical results). Our result is that the distribution of $\mathbf{T}_1/s + \mathbf{T}_2/s$, which is the distribution of the sum $\mathbf{T}_1/s$ 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 $\mathbf{E} \mathbf{T}_2^2 \leq \mathbf{E} h^2/n$. 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 $n^2 \mathbf{E} h^k / n^{3k/2} + (\mathbf{E} h^2/n)^{k/2}$, for even $k \geq 4$. Hence, if $\mathbf{E} h^4 < \infty$, the correction would be $O(n^{-2})$. The results extend as well to higher degree U -statistics and even for arbitrary (nonlinear) functions of nonidentically 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

$$\Delta := \sup_{x \in \mathbb{R}} \left| \mathbf{P} \left\{ \mathbf{P} \frac{\mathbf{T}}{s} \leq x \right\} - \Phi(x) \right|. \quad (7)$$

By a *Lyapunov* (or *Berry–Esseen*) *bound* for $\mathbf{T}$ we will mean a bound for Δ , which, under appropriate conditions on the sequence of statistics and the samples, is expected to satisfy $\Delta = O(n^{-1/2})$ as $n \rightarrow \infty$.

Let $s^2 = \mathbf{E} g^2$ as in (5), $\beta := \mathbf{E} |g|^3$, $\gamma_p := \mathbf{E} |h|^p$, $\gamma := \gamma_2$, $\varepsilon := \beta/(s^3 \sqrt{n})$, $\delta := \gamma/(s^2 n)$. Next define the asymptotic expansion

$$G(x) = \Phi(x) + \Phi_1(x), \quad (8)$$

where

$$\Phi_1(x) := \tfrac{1}{2}\kappa \Phi'''(x), \quad \kappa := n^{-1/2} s^{-3} \mathbf{E} g(X) g(\overline{X}) h(X, \overline{X}), \quad (9)$$

and write

$$\Gamma = \Gamma(\mathbf{T}) := \sup_{x \in \mathbf{R}} \left| \mathbf{P} \left\{ \frac{\mathbf{T}}{s} \leq x \right\} - G(x) \right|. \quad (10)$$

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

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

$$\Gamma \ll \varepsilon + \delta \quad \text{and} \quad \Delta \ll \varepsilon + \delta + |\kappa|. \quad (11)$$

Moreover, in the case where $\sigma^2 < \infty$, we can replace the scaling by s by the scaling by σ (see Section 3.4 for details).

Hölder's inequality implies that

$$|\kappa| \leq n^{-1/2} s^{-3} (\mathbf{E} g^2) (\mathbf{E} h^2)^{1/2} = \delta^{1/2} \quad (12)$$

so that $|\Phi_1(x)| \ll \sqrt{\delta} = \sqrt{\mathbf{E} h^2 / (s^2 n)}$. Therefore an impression may evolve that the Edgeworth correction G has a size $O(n^{-1/2})$, which is comparable to the bound (11) 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 2 below. Example 2 shows as well that bounds of type (11) with $\gamma/(s^2 n)$ replaced by, say $(\gamma/(s^2 n))^2$, would do much better. Under the assumption that the sequence of moments γ is bounded as $n \rightarrow \infty$, the contribution to the error of the nonlinear part $\mathbf{T}_2$ is of size $O(n^{-1})$, 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 $O(n^{-1/2})$ in the case where the sequence of moments $\gamma_{5/3}$ is bounded, and of the size $O(1)$ 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 2 demonstrates this. It is interesting to notice that the U -statistic of Example 2 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 $\gamma < \infty$, for example:

$$\Delta \ll \frac{\mathbf{E} |g|^3}{s^3 \sqrt{n}} + \frac{\mathbf{E} |h|^{5/3}}{s^{5/3} \sqrt{n}}, \quad (13)$$

see [4].

The results may be extended to so-called von Mises functionals (statistics) of degree 2, that is, statistics of the form $n^{-2} \sum_{j,k} M(X_j, X_k)$, where $M: \mathcal{X}^2 \rightarrow \mathbf{R}$ is (usually) supposed to be a symmetric kernel. Finally, the central limit theorem (CLT) may be derived under optimal (Lindeberg type) conditions (see [4]).

The history related to the CLT and the accuracy of normal approximations to U -statistics is sufficiently rich, see [22], [15], [11]²⁾. 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 [21], [8], [24], [17], [16] only. Helmers and van Zwet [20] proved (13) with $\frac{5}{3}$ replaced by $\frac{5}{3} + \varepsilon$, $\varepsilon > 0$. Bounds without ε may be found in [13], [14], [19], [9]. Van Zwet (see [25]) established (13) with $s^{-2} \mathbf{E} h^2$ instead of $s^{-5/3} \mathbf{E} |h|^{5/3}$. We use some methods from [3], where the bounds for the so-called concentration (of probability mass) of $\mathbf{T}$, comparable in quality with Theorem 1, were obtained. Using the approach developed in the latter paper, Alberink ([2]) showed that the bound $\Delta \ll \varepsilon + \delta^\theta + |\kappa|$, $\theta < 1$, holds in the non-i.i.d. setting. Alberink ([1]) extended the result of van Zwet ([25]) to the non-i.i.d. case and improved the dependence on $s^{-2} \mathbf{E} h^2$ to $\sqrt{s^{-2} \mathbf{E} h^2}$. A result of Borovskikh ([12]) implies that $\Delta \ll \varepsilon + \delta \log n + |\kappa|$. 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 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 nonlinear statistics, see a related discussion in [7]. Compare as well Lyapunov's bound $O(\beta n^{-1/2} \log n)$ with the Berry–Esseen bound $O(\beta n^{-1/2})$. We could not prove Theorem 1 extending and improving the existing technique — it seems that the best achievable by this technique is a bound with a log factor. 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 $\mathbf{T}_2$ from the characteristic function $\mathbf{E} \exp\{it \mathbf{T}\} = \mathbf{E} \exp\{it \mathbf{T}_1 + it \mathbf{T}_2\}$. The effect of this is that now it 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

²⁾ Вопросы точности нормальной аппроксимации распределений U -статистик в схеме серий рассматривались также в работах Rao, Janson (Ann. Probab. 1986, v. 14, № 4), Weber (Math. Proc. Cambridge Philos. Soc., 1983, v. 94, № 2), de Jong (Probab. Theory Related Fields, 1987, v. 75, № 2), Ронжин (Матем. заметки, 1986, т. 40, № 5). — Прим. рецензента.

with a Taylor expansion in powers of $\mathbf{T}_2$ 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 a conditional variance of the kernels is bounded. The general unbounded case is reduced to the bounded one by an application of a nonstandard 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 nontruncated kernels. The result with truncated kernels and weaker moment assumptions is derived from Theorem 1 by an application of a rather simple truncation technique (see [4]). 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 $\mathbf{E} \exp\{itg(X)\}$ — see [23] for a counter-example: $\mathbf{T} = \mathbf{T}_1$, $g(X)$ being a Rademacher random variable with $\mathbf{P}\{g(X) = \pm 1\} = \frac{1}{2}$. The term $\delta = \mathbf{E} h^2/(s^2n)$ is optimal in the sense that lower order moments $\mathbf{E} |h|^\theta$, $\theta < 2$, do not control $O(n^{-1})$ behavior, (see [6], [5]). Another way to see this would be to use Edgeworth type expansions of $\mathbf{P}\{\mathbf{T} \leq x\}$. In particular, the terms with κ are unavoidable as well.

2. Examples. In this section $X, X_1, \dots, X_n$ denote real-valued i.i.d. random variables.

Example 1. Let $\mu := \mathbf{E} X$. Define $\mathbf{U}$ by (1) with the kernel $H(x, y) = xy$, for $x, y \in \mathbf{R}$. It is clear that $\mathbf{U}$ is an unbiased estimator of the unknown parameter μ^2 , that is, $\mathbf{E} \mathbf{U} = \mu^2$. The functions g and h from the Hoeffding decomposition (4) are now given by $g(x) = \mu(x - \mu)$ and $h(x, y) = (x - \mu)(y - \mu)$. Hence, writing $a^2 := \mathbf{E}(X - \mu)^2$ and $b := \mathbf{E}|X - \mu|^3$, the moments corresponding to our model are $s^2 = \mu^2 a^2$, $\beta = |\mu|^3 b$, $\gamma = a^4$, $\kappa = n^{-1/2} \mu^{-1} a$. By Theorem 1, under the assumption that $0 < \mu^2 a^2 < \infty$ we have that

$$\sup_{x \in \mathbf{R}} \left| \mathbf{P} \left\{ \sqrt{n} \frac{(\mathbf{U} - \mu^2)}{2s} \leq x \right\} - \Phi(x) - \frac{1}{2} \kappa \Phi'''(x) \right| \ll \frac{b}{a^3 \sqrt{n}} + \frac{a^2}{\mu^2 n}, \quad (14)$$

a bound depending on μ^2 . This bound is precisely of order $O(n^{-1/2})$ in the case where the sequence b/a^3 is bounded and the sequence $\mu^{-2} a^2$ is of order $O(n^{-1/2})$. The latter will mean moreover that $|\kappa| \leq n^{-1/4}$, so that the Edgeworth correction Φ_1 then is of order $O(n^{-1/4})$. In case $\mu^{-2} a^2 = c n^{-1/2}$

for some fixed constant $c > 0$, we have indeed that $\|\Phi_1\|_\infty \geq \tilde{c}n^{-1/4}$ for some constant $\tilde{c} > 0$, which shows a fortiori that the correction really can make a difference.

Example 2. Let $\mathbf{E}X = 0$ and $\mathbf{E}X^2 = 1$ and consider the kernel $H(x, y) = xy^2 + x^2y$. The corresponding statistic $\mathbf{U}$ is closely connected to Student's statistic. The expectation $\mathbf{E}H$ and functions g and h from (4) are of the form $\mathbf{E}H = 0$, $g(x) = x$, $h(x, y) = x(y^2 - 1) + (x^2 - 1)y$. Hence, $\mathbf{T} = \frac{1}{2}\sqrt{n}\mathbf{U}$, so that in fact

$$\mathbf{T} = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j + \frac{1}{(n-1)\sqrt{n}} \sum_{1 \leq j < k \leq n} \{X_j(X_k^2 - 1) + (X_j^2 - 1)X_k\}. \quad (15)$$

Using a truncation method (see [4] for details) Theorem 1 yields

$$\Delta \ll n^{-1/2} \mathbf{E}|X|^3. \quad (16)$$

An application of (13) leads to $\Delta \ll n^{-1/2} \mathbf{E}|X|^{10/3}$. Hence, for $\mathbf{T}$ defined by (15), the bound (16) yields $\Delta = O(n^{-1/2})$ under the optimal condition that $\mathbf{E}|X|^3 < \infty$, whereas (13) leads to the stronger condition $\mathbf{E}|X|^{10/3} < \infty$. Another advantage of Theorem 1 and (16) is that (using truncation arguments) they imply the CLT under the optimal condition that $0 < \mathbf{E}X^2 < \infty$. More generally, we have that $\Delta = o(n^{-\theta})$ if $0 < \mathbf{E}|X|^{2(1+\theta)} < \infty$, for $0 \leq \theta < \frac{1}{2}$.

One can introduce statistics similar to (15) corresponding to the kernels

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

with $p = 2, 3, 4, \dots$, just replacing in (15) the 2nd by p th powers. Theorem 1 now yields $\Delta \ll n^{-1/2} \mathbf{E}|X|^{3p/2}$, whereas (13) leads us to the bound $\Delta \ll n^{-1/2} \mathbf{E}|X|^{5p/3}$. For the CLT the condition $0 < \mathbf{E}|X|^p < \infty$ is sufficient.

3. Proof of Theorem 1. We use the following notation. For any bounded function f , by $\|f\|_\infty := \sup_{x \in \mathbf{R}} |f(x)|$ we denote its supremum norm. By $\mathbf{I}\{A\}$ we mean the indicator of an event A . We write $e\{y\} := \exp\{iy\}$, for $y \in \mathbf{R}$, with $i = \sqrt{-1}$ denoting the complex root.

We will use a more compact notation for the parts of $\mathbf{T}$: taking

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

we have $\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2$, with $\mathbf{T}_1 = \sum_{j=1}^n T_j$, $\mathbf{T}_2 = \sum_{1 \leq j < k \leq n} T_{j,k}$.

To prove the theorem, it suffices to verify that

$$\Gamma \ll \varepsilon + \delta \quad (17)$$

since the bound $\Delta \ll \varepsilon + \delta + |\kappa|$ easily follows from (17) using $\|\Phi_1\|_\infty \ll |\kappa|$. We shall prove (17) for the statistic $\mathbf{T}$ normalized by s ; for a proof of the result with the normalizing factor σ see subsection 3.4 below.

From here on, let

$$a := \min\{\beta^{-1}n^{1/2}, \gamma^{-1}n\}. \tag{18}$$

In the proof of (17), without loss of generality we may assume that

$$\mathbf{E} \mathbf{T} = 0, \quad s^2 = 1, \quad a \geqslant 10^3 \theta^{-1}, \quad \text{and} \quad n \geqslant \theta^{-1}, \tag{19}$$

where $0 < \theta \leqslant 10^{-4}$ is some fixed positive number. Indeed, if at least one of the assumptions (19) does not hold, we may use the trivial bound $|\Gamma| \ll 1 + |\kappa|$ in order to derive (17).

Using Lyapunov’s inequality and assumptions (19), we have

$$\beta \geqslant s^3 = 1 \quad \text{and} \quad a \leqslant \beta^{-1}n^{1/2} \leqslant n^{1/2}, \tag{20}$$

whereas furthermore

$$\mathbf{E} |T_1|^3 = n^{-3/2} \beta \leqslant n^{-1}a^{-1} \quad \text{and} \quad \mathbf{E} T_{1,2}^2 = n^{-3} \gamma \leqslant n^{-2}a^{-1}. \tag{21}$$

We will use the inequalities (20) and (21) extensively.

3.1. Truncation of $\mathbf{T}$. In order to prove the theorem we are going to apply Esseen’s smoothing lemma. Before doing this it appears to be wise to perform a (nonstandard) truncation of the kernel h . Consider the function $\varphi^2(x) = n^{-1}(n-1)^{-2} \mathbf{E} h^2(X, x)$, for $x \in \mathcal{X}$. Define the random variables $\varphi_j^2 = \varphi^2(X_j)$. We may realize the random variables as conditional variances related to the kernel h , $\varphi_1^2 := \mathbf{E} (T_{1,2}^2 | X_1)$ and $\varphi_j^2 := \mathbf{E} (T_{1,j}^2 | X_j)$, $2 \leqslant j \leqslant n$. We will prove (17) under the condition that, for any j ,

$$\varphi_j^2 \leqslant \theta n^{-1}. \tag{22}$$

The assumption (22) does not restrict the generality. Indeed, assume that we have proved (17) under the assumption (22). Let us prove that then the bound (17) holds without the assumption (22). To this, consider the indicators

$$I_j = I_j(X_j) := \mathbf{I}\left\{\varphi_j^2 \leqslant \frac{1}{4} \theta n^{-1}\right\}, \quad I_j^c := 1 - I_j,$$

and define the truncated version of $\mathbf{T}$ as $\hat{\mathbf{T}} := \mathbf{T}_1 + \sum_{1 \leqslant j < k \leqslant n} I_j I_k T_{j,k}$ with the Hoeffding decomposition

$$\hat{\mathbf{T}} = \mathbf{E} \hat{\mathbf{T}} + \sum_{j=1}^n \hat{\mathbf{T}}_j + \sum_{1 \leqslant j < k \leqslant n} \hat{\mathbf{T}}_{j,k} =: \mathbf{E} \hat{\mathbf{T}} + \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2.$$

Lemma 1. *Let $\hat{s}^2 := \text{var} \hat{\mathbf{T}}_1$. Assume that (19) applies. Then*

$$\Gamma(\mathbf{T}) \ll \theta^{-1}a^{-1} + \Gamma(\hat{\mathbf{T}}). \tag{23}$$

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

$$\frac{1}{2} \leq \hat{s}^2 = n \mathbf{E} \hat{\mathbf{T}}_j^2 \leq \frac{3}{2}, \quad \mathbf{E} |\hat{\mathbf{T}}_j|^3 \leq 4 \mathbf{E} |T_j|^3 + 12 n^{-1} a^{-3/2}, \quad (24)$$

$$\mathbf{E} \hat{\mathbf{T}}_{j,k}^2 \leq 4 \mathbf{E} T_{j,k}^2 \quad \text{and} \quad \mathbf{E} (\hat{\mathbf{T}}_{j,k}^2 | X_j), \mathbf{E} (\hat{\mathbf{T}}_{j,k}^2 | X_k) \leq \theta n^{-1}. \quad (25)$$

As a consequence of the last inequality in (25), the bound (17) is applicable to $\hat{\mathbf{T}}$, thus yielding a bound of the form

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

In turn, using (24) and (25), this bound is seen to be bounded from above by (a constant times) the desired bound a^{-1} for $\Gamma(\mathbf{T})$, plus an extra term of order $a^{-3/2} \leq a^{-1}$. From (23) the desired bound (17) then immediately follows, and this shows that indeed it suffices to prove (17) assuming the condition (22).

Sketch of the proof of Lemma 1. The proof requires rather lengthy, in essence elementary calculations, see [4] for a detailed exposition. First we notice that, for all j , by Markov's inequality

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

This implies that

$$\mathbf{P}\{\mathbf{T} \neq \hat{\mathbf{T}}\} \leq \mathbf{P}\{I_j = 0, \text{ some } j\} \leq 4n \theta^{-1} n^{-1} a^{-1} = 4 \theta^{-1} a^{-1},$$

so that, for any $x \in \mathbf{R}$, we have $|\mathbf{P}\{\mathbf{T} \leq x\} - G(x)| \ll \theta^{-1} a^{-1} + |\mathbf{P}\{\hat{\mathbf{T}} \leq x\} - G(x)|$.

Let us prove, e.g., that $|\mathbf{E} \hat{\mathbf{T}}| \leq 2 \theta^{-1} a^{-3/2}$. Note that

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

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

$$\begin{aligned} |\mathbf{E} \hat{\mathbf{T}}| &\leq \tfrac{1}{2} n^2 (\mathbf{E} I_1^c \mathbf{E} I_2^c)^{1/2} (\mathbf{E} T_{1,2}^2)^{1/2} \leq \tfrac{1}{2} n^2 \mathbf{P}\{I_1 = 0\} (n^{-2} a^{-1})^{1/2} \\ &\leq 2 n^2 \theta^{-1} n^{-1} a^{-1} n^{-1} a^{-1/2} = 2 \theta^{-1} a^{-3/2}. \end{aligned} \quad (26)$$

Following the definitions, it is easily seen that $T_j = T_j + \xi_j$ with

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

for $2 \leq j \leq n$, and $\xi_1 := (n-1) (\mathbf{E}(I_1 I_2 T_{1,2} | X_1) - \mathbf{E} I_1 I_2 T_{1,2})$. The further proof is based, in particular, on the bounds $\mathbf{E} \xi_1^2 \leq 4 \theta^{-1} n^{-1} a^{-2}$ and $\mathbf{E} |\xi_1|^3 \leq \frac{1}{16} n^{-1} a^{-3/2}$, which show that the statistics $\mathbf{T}$ and $\hat{\mathbf{T}}$ are almost equal.

3.2. A reduction to bounds for the characteristic functions.

From here on, let g , g_1 , and g_2 be the Fourier–Stieltjes transforms of G , Φ , and Φ_1 , respectively. We have $g = g_1 + g_2$ since $G = \Phi + \Phi_1$, cf. (8). Let $f(t) := \mathbf{E} e\{t\mathbf{T}\}$. Taking τ uniformly distributed on $(0, 1)$ and independent of the sample, we expand our characteristic function f as

$$f(t) = f_1(t) + f_2(t) + f_3(t), \tag{27}$$

where $f_1(t) := \mathbf{E} e\{t\mathbf{T}_1\}$,

$$f_2(t) := (it) \mathbf{E} \mathbf{T}_2 e\{t\mathbf{T}_1\}, \quad f_3(t) := (it)^2 \mathbf{E} (1 - \tau) \mathbf{T}_2^2 e\{t(\mathbf{T}_1 + \tau \mathbf{T}_2)\}. \tag{28}$$

Due to Esseen’s smoothing lemma ([18, Chap. XVI, Lemma 3.2])

$$\Gamma = \sup_{x \in \mathbf{R}} |\mathbf{P}\{\mathbf{T} \leq x\} - G(x)| \ll \theta^{-1} a^{-1} + \int_0^{\theta a} |f(t) - g(t)| t^{-1} dt, \tag{29}$$

so that we have to estimate $|f(t) - g(t)|$. The easy part of this estimation we formulate as the next Lemma 2. The real mathematical contents of this paper represent Lemmas 3 and 4, where $|f(t) - g(t)|$ is estimated for $1 \leq t \leq \theta a$. In order to prove (17) (and, hence, to conclude the proof of Theorem 1), it suffices to combine the smoothing inequality (29) with the bounds of the Lemmas 2–4.

Lemma 2. *Let assumptions (19) be fulfilled. Then we have*

$$|f(t) - g(t)| \ll a^{-1} t, \quad \text{for } 0 \leq t \leq 1, \tag{30}$$

$$|f_1(t) - g_1(t)| \ll a^{-1} t^3 \exp \left\{ -\frac{1}{4} t^2 \right\}, \quad \text{for } 0 \leq t \leq \theta a, \tag{31}$$

$$|f_2(t) - g_2(t)| \ll a^{-1} t^3 (1 + t) \exp \left\{ -\frac{1}{4} t^2 \right\}, \quad \text{for } 0 \leq t \leq \theta a. \tag{32}$$

$$|f_1(t) + f_2(t) - g(t)| \ll a^{-1} t^3 (1 + t) \exp \left\{ -\frac{1}{4} t^2 \right\}, \quad \text{for } 0 \leq t \leq \theta a. \tag{33}$$

Lemma 3. *Let assumptions (19) and (22) be fulfilled. Then there exists a constant $c > 0$ such that*

$$|f(t) - g(t)| \ll a^{-3/2} t^3 + a^{-1} t^4 \exp\{-ct^2\} \quad \text{for } 1 \leq t \leq \theta a.$$

Lemma 4. *Let assumptions (19) and (22) be fulfilled. Then there exists a constant $c > 0$ such that*

$$|g(t)| \ll (1 + a^{-1/2} t^3) \exp \left\{ -\frac{1}{2} t^2 \right\}$$

and

$$\begin{aligned} |f(t)| &\ll t^{-6} + a^{-1/2} (\log t) t^{-5} + a^{-1} (\log^{3/2} t) t^{-1} \\ &\quad + a^{-3/2} \log^{3/2} t \quad \text{for } 1 \leq t \leq \theta a. \end{aligned}$$

3.3. Proofs of Lemmas 2–4.

P r o o f o f L e m m a 2. The bound (31) is a classical one since the statistic $\mathbf{T}_1$ is a sum of i.i.d. random variables with the characteristic function f_1 (see, e.g., [23]). We turn to (32). Using that $\mathbf{T}_2$ is a sum of $T_{j,k}$, the i.i.d. assumption, and the symmetry of the kernel h , we obtain

$$f_2(t) = \binom{n}{2} (it) \mathbf{E} T_{1,2} e\{t\mathbf{T}_1\} = \binom{n}{2} (it) h_1(t) h_2(t)$$

with $h_1(t) := \mathbf{E} T_{1,2} e\{tT_1 + tT_2\}$ and $h_2(t) := \mathbf{E} e\{t(\mathbf{T}_1 - T_1 - T_2)\}$. We may replace $h_2(t)$ by the characteristic function of the standard normal distribution since $\mathbf{T}_1 - T_1 - T_2$ is a sum of $n - 2$ i.i.d. random variables, and the variance of $\mathbf{T}_1 - T_1 - T_2$ is approximately 1 (e.g., using standard bounds from [23]). Now, applying twice the expansion $e\{tT_j\} = 1 + (it)T_j + R_j$ with $|R_j| \leq t^{3/2} |T_j|^{3/2}$, we may replace $h_1(t)$ by $(it)^2 \kappa$ (up to acceptable remainder terms), and (32) follows.

The bound (33) is obviously implied by (31) and (32), while clearly $|f_3(t)| \leq \frac{1}{2} t^2 \mathbf{E} \mathbf{T}_2^2 \leq \frac{1}{2} a^{-1} t^2$, and we derive (30).

P r o o f o f L e m m a 3. The expansion (27) and the bound (33) show that it will suffice to prove that

$$|f_3(t)| \ll a^{-3/2} t^3 + a^{-1} t^2 \exp\{-c_0 t^2\} \quad (34)$$

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

$$L_j := \sum_{k=j+1}^n T_{j,k}, \quad S_j := \sum_{k=j}^{n-1} L_k, \quad \text{and} \quad W_j := \sum_{k=1}^j L_k = \mathbf{T}_2 - S_{j+1}, \quad (35)$$

see Fig. 1. In the figure the big triangle symbolizes the array of pairs $\{j, k\}$ such that $1 \leq j < k \leq n$. Now let

$$\mathbf{T}_2^2 = U_1 + U_2 + U_3, \quad (36)$$

with (we write $T_{j,k} = T_A$ with $A = \{j, k\}$)

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

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

Fig. 1. Decomposition of $\mathbf{T}_2$ into lines L_j , triangles S_j , and complements W_j .

For $k = 1, 2, 3$, we look in turn at $h_k(t) := |(it)^2 \mathbf{E}(1 - \tau) U_k e\{t(\mathbf{T}_1 + \tau \mathbf{T}_2)\}|$.
We start by looking at

$$h_1(t) = t^2 \binom{n}{2} \left| \mathbf{E}(1 - \tau) T_{1,2}^2 e\{t(\mathbf{T}_1 + \tau \mathbf{T}_2)\} \right|.$$

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

$$\begin{aligned} h_1(t) &\leq \frac{1}{2} n^2 t^2 \left| \mathbf{E}(1 - \tau) T_{1,2}^2 e\{t(\mathbf{T}_1 + \tau(L_1 + L_2))\} \right| + h_5(t) \\ &=: h_4(t) + h_5(t), \end{aligned} \quad (37)$$

with

$$\begin{aligned} |h_5(t)| &\leq \frac{1}{2} n^2 t^3 \mathbf{E} \tau (1 - \tau) T_{1,2}^2 |S_3| \\ &\leq \frac{1}{4} n^2 t^3 (\mathbf{E} T_{1,2}^2) (\mathbf{E} S_3^2)^{1/2} \ll a^{-1} a^{-1/2} t^3 = a^{-3/2} t^3, \end{aligned} \quad (38)$$

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

$$\begin{aligned} &\left| \mathbf{E}(1 - \tau) T_{1,2}^2 e\{t(\mathbf{T}_1 + \tau(L_1 + L_2))\} \right| \\ &\leq \mathbf{E}(1 - \tau) T_{1,2}^2 \left| e\{t(T_1 + T_2) + t\tau T_{1,2}\} \right| \\ &\quad \times \left| \mathbf{E} \left(e \left\{ t \sum_{j=3}^n (T_j + \tau T_{1,j} + \tau T_{2,j}) \right\} \mid X_1, X_2, \tau \right) \right| \\ &\leq \mathbf{E} T_{1,2}^2 |Y_{1,2,\tau}|^{n-2}, \end{aligned} \quad (39)$$

with

$$Y_{1,2,\tau} := \mathbf{E} \left(e\{t(T_3 + \tau T_{1,3} + \tau T_{2,3})\} \mid X_1, X_2, \tau \right). \quad (40)$$

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

$$c_1 := 1 - 3\theta^{-1}\tilde{c}^{-1} - \frac{4}{3}\tilde{c}^{-5/6}(1 + 8\theta^{-1}\tilde{c}^{-3/2}) \quad (41)$$

and

$$c_2 := \frac{1}{2}(c_1 - 8(\theta + \tilde{c}^{-2})^{1/2} - 16\theta), \quad (42)$$

and note that under the assumptions $\frac{99}{100} \leq c_1 \leq 1$ and $\frac{45}{100} \leq c_2 \leq \frac{1}{2}$. We will prove that, as a result of the truncation,

$$|Y_{1,2,\tau}| \leq 1 - c_2 n^{-1} t^2 \leq \exp\{-c_2 n^{-1} t^2\}. \quad (43)$$

As a result of (43) then

$$\mathbf{E} T_{1,2}^2 |Y_{1,2,\tau}|^{n-2} \leq \exp\{-c_3 t^2\} 4 n^{-2} a^{-1}, \quad (44)$$

taking $c_3 := (1 - 2n^{-1})c_2$, and, combining (37), (38), (39), and (44), $h_1(t) \ll a^{-3/2} t^3 + a^{-1} t^2 \exp\{-c_3 t^2\}$.

We turn to the proof of (43). In fact, using a Taylor expansion in terms of $\tau(T_{1,3} + T_{2,3})$ we see that

$$\begin{aligned} |Y_{1,2,\tau}| &\leq |\mathbf{E} e\{t T_3\}| + t \tau \left| \mathbf{E} ((T_{1,3} + T_{2,3}) e\{t T_3\} | 1, 2) \right| \\ &\quad + \frac{1}{2} t^2 \tau^2 \mathbf{E} ((T_{1,3} + T_{2,3})^2 | X_1, X_2) =: \psi_1(t) + \psi_2(t) + \psi_3(t). \end{aligned}$$

Here clearly (see Lemma 1)

$$\psi_3(t) \leq \frac{1}{2} t^2 4 \mathbf{E} (T_{1,3}^2 | X_1) \leq 8 t^2 \theta n^{-1}. \quad (45)$$

As to $\psi_1(t)$ a three-term Taylor expansion is telling us that $\psi_1(t) \leq |1 - \frac{1}{2} t^2 \mathbf{E} T_3^2| + \frac{1}{6} t^3 \mathbf{E} |T_3|^3$. By Lemma 1 we have

$$\begin{aligned} \frac{1}{2} t^2 \mathbf{E} T_3^2 &\leq \frac{1}{2} (1 + 3\theta^{-1} a^{-1}) n^{-1} t^2 \\ &\leq \frac{1}{2} (1 + 3\theta^{-1} a^{-1}) n^{-1} a^{1/3} \leq \frac{1}{2} (1 + 3\theta^{-1} a^{-1}) n^{-5/6} \leq 1, \end{aligned}$$

and using again the lemma it is then easily seen that

$$\begin{aligned} \psi_1(t) &\leq 1 - \frac{1}{2} (1 - 3\theta^{-1} a^{-1}) n^{-1} t^2 + \frac{1}{6} a^{1/6} t^2 4 (\mathbf{E} |T_j|^3 + \mathbf{E} |\xi_j|^3) \\ &\leq 1 - \frac{1}{2} n^{-1} t^2 \left(1 - 3\theta^{-1} a^{-1} - \frac{4}{3} \tilde{c}^{1/6} (\tilde{c}^{-1} + 8\theta^{-1} \tilde{c}^{-5/2}) \right) \\ &\leq 1 - \frac{1}{2} c_1 n^{-1} t^2. \end{aligned} \quad (46)$$

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

$$\begin{aligned} \psi_2(t) &\leq t^2 \mathbf{E} (|T_{1,3} + T_{2,3}| |T_3| | X_1, X_2) \\ &\leq t^2 \mathbf{E} (|T_{1,3} + T_{2,3}|^2 | 1, 2)^{1/2} (\mathbf{E} T_3^2)^{1/2} \\ &\leq t^2 (4^2 \theta n^{-1})^{1/2} (n^{-1} + \theta^{-1} n^{-1} a^{-2})^{1/2} \leq 4 t^2 n^{-1} (\theta + a^{-2})^{1/2}, \end{aligned} \quad (47)$$

and using (45), (46), and (47), the proof of (43) is easily concluded.

We turn to h_2 . To this, let

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

Now

$$h_2(t) = n t^2 \left| \mathbf{E} (1 - \tau) U_{2,1} e\{t (\mathbf{T}_1 + \tau \mathbf{T}_2)\} \right|. \tag{48}$$

We notice that

$$\mathbf{E} (U_{2,1}^2 \mid X_1) = \binom{n-1}{2} \mathbf{E} (T_{1,2}^2 T_{1,3}^2 \mid X_1) \leq \frac{1}{2} n^2 \mathbf{E} (T_{1,2}^2 \mid X_1)^2, \tag{49}$$

since for example $\mathbf{E} T_{1,2} T_{1,3} T_{1,2} T_{1,4} = \mathbf{E} T_{1,2} T_{1,3} T_{1,2} \mathbf{E} (T_{1,4} \mid X_1, X_2, X_3) = 0$, so that removal of S_2 from the expression $\mathbf{T}_2$ in (48) will cost no more than

$$\begin{aligned} n t^3 \mathbf{E} \tau (1 - \tau) |U_{2,1}| |S_2| &\leq \tfrac{1}{2} n t^3 \mathbf{E} \mathbf{E} (U_{2,1}^2 \mid X_1)^{1/2} \mathbf{E} (S_2^2 \mid X_1)^{1/2} \\ &\leq 2^{-3/2} n^2 t^3 (\mathbf{E} S_2^2)^{1/2} \mathbf{E} \mathbf{E} (T_{1,2}^2 \mid X_1) \\ &\leq 2^{-3/2} n^2 a^{-1/2} 4 n^{-2} a^{-1} t^3 \leq 2^{1/2} a^{-3/2} t^3, \end{aligned} \tag{50}$$

which is acceptable. Hence we look at (48) with $\mathbf{T}_2$ replaced by L_1 , and taking

$$U_{2,1}^* := \sum_{1 < j, k \leq [\frac{1}{2} n], j \neq k} T_{1,j} T_{1,k},$$

we see that

$$\begin{aligned} h_2(t) &\leq 2^{1/2} a^{-3/2} t^3 + n 2 \binom{n-1}{2} \frac{1}{2} \binom{[n/2]-1}{2}^{-1} t^2 \\ &\times \left| \mathbf{E} (1 - \tau) U_{2,1}^* e\{t (\mathbf{T}_1 + \tau L_1)\} \right| =: 2^{1/2} a^{-3/2} t^3 + h_6(t). \end{aligned} \tag{51}$$

Taking

$$Y_{1,\tau} := \mathbf{E} \left(e\{t (T_n + \tau T_{1,n})\} \mid X_1, \tau \right), \tag{52}$$

here

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

where, see (49),

$$\mathbf{E} (|U_{2,1}^*| \mid X_1) \leq \mathbf{E} ((U_{2,1}^*)^2 \mid X_1)^{1/2} \leq n \mathbf{E} (T_{1,2}^2 \mid X_1),$$

and, as in (43), $|Y_{1,\tau}| \leq \exp\{-c_2 n^{-1} t^2\}$,

$$\begin{aligned} h_6(t) &\ll n^2 t^2 \mathbf{E} \mathbf{E}(T_{1,2}^2 | X_1) \exp\left\{-c_2 \left(n - \left\lfloor \frac{1}{2} n \right\rfloor\right) n^{-1} t^2\right\} \\ &\ll a^{-1} t^2 \exp\left\{-\frac{1}{2} c_2 t^2\right\}. \end{aligned} \quad (53)$$

From (51) and (53) we conclude that $h_2(t) \ll a^{-3/2} t^3 + a^{-1} t^2 \exp\{-\frac{1}{2} c_2 t^2\}$.

As to $h_3(t)$ we have that

$$U_3 = \sum_{1 \leq j < k < p < q \leq n} V_{j,k,p,q} \quad (54)$$

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 $(|a| + |b|)^r \leq 2^{r-1} (|a|^r + |b|^r)$, and Lemma 1,

$$\begin{aligned} \mathbf{E} U_3^2 &= \sum_{1 \leq j < k < p < q \leq n} \mathbf{E} V_{j,k,p,q}^2 = \binom{n}{4} \mathbf{E} V_{1,2,3,4}^2 \\ &\leq \frac{1}{4} n^4 4 \cdot 3^2 \mathbf{E} T_{1,2}^2 T_{3,4}^2 \ll a^{-2}. \end{aligned} \quad (55)$$

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

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

so that

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

Taking $U_3^* := \sum_{1 \leq j < k < p < q \leq \lfloor \frac{1}{2} n \rfloor} V_{j,k,p,q}$, we have

$$h_7(t) = t^2 \binom{n}{4} \binom{\lfloor n/2 \rfloor}{4}^{-1} |\mathbf{E} (1 - \tau) U_3^* e\{t \mathbf{T}_1\}|,$$

which as for $h_6(t)$ leads to

$$h_7(t) \ll t^2 (\mathbf{E} |U_3^*|) |\mathbf{E} e\{t T_n\}|^{n - \lfloor n/2 \rfloor} \ll a^{-1} t^2 \exp\left\{-\frac{1}{2} c_2 t^2\right\}.$$

In turn

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

and the proof of (34) is easily concluded. This finishes the proof of the lemma.

P r o o f o f L e m m a 4. First notice that, because of (12), $|\kappa| \leq a^{-1/2}$, from which the bound for $|g(t)|$ easily follows.

Now let L_j , S_j , and W_j be as in (35), c_1 and c_2 as in (41) and (42), and

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

where by $[x]$ we mean the integer part of $[x]^3$. We notice that under the assumptions $100 \leq m \leq \frac{1}{4} n$. Since $c_2 \leq \frac{1}{2} c_1$ and $c_1 \leq 1$, using (46) we have

$$\begin{aligned} |\mathbf{E} e\{t T_1\}|^{m-2} &\leq \exp \left\{ -\frac{1}{2} c_1 (m-2) n^{-1} t^2 \right\} \\ &\leq \exp \left\{ -6 \log t + \frac{3}{2} c_1 n^{-1} t^2 \right\} \ll t^{-6}. \end{aligned} \quad (56)$$

Instead of (27), 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 $f_1(t) := \mathbf{E} e\{t(\mathbf{T}_1 + S_{m+1})\}$, $f_2(t) := (it) \mathbf{E} W_m e\{t(\mathbf{T}_1 + S_{m+1})\}$, $f_3(t) := (it)^2 \mathbf{E} (1 - \tau) W_m^2 e\{t(\mathbf{T}_1 + S_{m+1} + \tau W_m)\}$. As to f_1 it is easy to see that, as a result of (56),

$$|f_1(t)| = |\mathbf{E} e\{t T_1\}|^m \left| \mathbf{E} e\left\{t \left(\sum_{j=m+1}^n T_j + S_{m+1} \right)\right\} \right| \ll t^{-6},$$

which is small enough for our purposes. As to f_2 , for $l = 2, m+1$, let $V_l := e\{t(\mathbf{T}_1 - (T_1 + T_l) + S_{m+1})\}$. Using the Taylor expansion $e\{t T_j\} = 1 + R_j$ with $|R_j| \leq t |T_j|$ together with independence we see that

$$\begin{aligned} f_2(t) &= (it) \binom{m}{2} (\mathbf{E} T_{1,2} e\{t(T_1 + T_2)\}) \mathbf{E} V_2 \\ &\quad + (it) m(n-m) \mathbf{E} T_{1,m+1} e\{t(T_1 + T_{m+1})\} V_{m+1} \\ &= (it) \binom{m}{2} (\mathbf{E} T_{1,2} R_1 R_2) \mathbf{E} V_2 + (it) m(n-m) \\ &\quad \times \mathbf{E} T_{1,m+1} R_1 (1 + R_{m+1}) V_{m+1} =: f_4(t) + f_5(t). \end{aligned} \quad (57)$$

Using Hölder's inequality, Lemma 1, (56), and independence, we see that

$$\begin{aligned} |f_4(t)| &\ll m^2 t \mathbf{E} |T_{1,2} R_1 R_2| |\mathbf{E} V_2| \ll m^2 t^3 (\mathbf{E} T_{1,2}^2)^{1/2} (\mathbf{E} T_1^2) |\mathbf{E} e\{t T_3\}|^{m-2} \\ &\ll (mn^{-1})^2 a^{-1/2} t^{-3} \ll a^{-1/2} (\log^2 t) t^{-7}. \end{aligned} \quad (58)$$

As to

$$\begin{aligned} f_5(t) &= (it) m(n-m) (\mathbf{E} T_{1,m+1} R_1 V_{m+1} + \mathbf{E} T_{1,m+1} R_1 R_{m+1} V_{m+1}) \\ &=: f_6(t) + f_7(t) \end{aligned} \quad (59)$$

³⁾ We have chosen the number m as in traditional proofs. Notice however that one may prove the result choosing m to be considerably larger.

we have that

$$\begin{aligned} |f_7(t)| &\leq mn t |\mathbf{E} e\{t T_2\}|^{m-1} t^2 \mathbf{E} |T_{1,m+1} T_1 T_{m+1}| \\ &\ll mn^{-1} a^{-1/2} t^{-3} \ll a^{-1/2} (\log t) t^{-5}, \end{aligned} \quad (60)$$

cf. (58). On the other hand

$$\begin{aligned} f_6(t) &= (it) m(n-m) (\mathbf{E} e\{t T_2\})^{m-1} \\ &\quad \times \mathbf{E} T_{1,m+1} R_1 e\left\{t \left(L_{m+1} + S_{m+2} + \sum_{j=m+2}^n T_j\right)\right\}, \end{aligned}$$

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 (56),

$$\begin{aligned} |f_6(t)| &\ll mn t t^{-6} t^2 \mathbf{E} |T_{1,m+1} T_1 L_{m+1}| \\ &\leq mn t^{-3} (\mathbf{E} T_{1,m+1}^2)^{1/2} (\mathbf{E} T_1^2)^{1/2} (\mathbf{E} L_{m+1}^2)^{1/2} \\ &\leq mn^{-1} t^{-3} a^{-1} \ll a^{-1} (\log t) t^{-5}, \end{aligned} \quad (61)$$

and following (57), (58), (59), (60), and (61) it is easily seen that $|f_2(t)|$ is small enough for our purposes.

We turn to f_3 , for which we do the following. We use the decompositions $W_m = P_1 + P_2 + P_3$ and $S_{m+1} = P_4 + P_5 + P_6$, laid down by

$$\begin{aligned} P_1 &:= \sum_{1 \leq j < k \leq m} T_{j,k}, & P_2 &:= \sum_{j=1}^m \sum_{k=m+1}^{n-m} T_{j,k}, \\ P_4 &:= \sum_{j=m+1}^{n-m} \sum_{k=n-m+1}^n T_{j,k}, & P_5 &:= \sum_{n-m+1 \leq j < k \leq n} T_{j,k} : \end{aligned}$$

see Fig. 2 (cf. Fig. 1). Moreover, in order to denote the tails of a line L_j , for all $1 \leq j < k \leq n$ we write $L_{j,k} := \sum_{l=k}^n T_{j,l}$.

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 $\mathbf{T}_2^2$, see (36), and perform three separate estimations.

As to the removal of P_1 we note that, writing

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

we have a decomposition of form $f_3(t) = f_8(t) + f_9(t) + f_{10}(t)$ with $f_8(t) := (it)^2 \mathbf{E} (P_2 + P_3)^2 Q$, $f_9(t) := (it)^2 \mathbf{E} P_1 (P_2 + P_3) Q$ and $|f_{10}(t)| \leq \frac{1}{2} t^2 \mathbf{E} P_1^2 \leq \frac{1}{2} t^2 (mn^{-1})^2 a^{-1} \ll a^{-1} (\log^2 t) t^{-2}$. Furthermore,

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

so that we may indeed concentrate on f_8 .

Fig. 2. Decomposition of T_2 into $P_1 + \dots + P_6$.

As to f_8 we have that $(P_2 + P_3)^2 = U_1 + U_2 + U_3$, taking

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

cf. (36). Now let $h_k(t) := (it)^2 \mathbf{E} U_k Q$, for $k = 1, 2, 3$. We have $f_8(t) = h_1(t) + h_2(t) + h_3(t)$, and start by looking at $h_1(t) = (it)^2 m(n - m) \mathbf{E} (1 - \tau) T_{1,m+1}^2 e\{t(\mathbf{T}_1 + S_{m+1} + \tau W_m)\}$. Parallel to (37), 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,

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

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

$$\begin{aligned} |h_4(t)| &\leq t^2 mn \mathbf{E} T_{1,m+1}^2 \\ &\quad \times \left| \mathbf{E} \left(e \left\{ t \sum_{j=n-m+1}^n (T_j + \tau T_{1,j} + T_{m+1,j}) \right\} \middle| \tau, 1, \dots, n - m \right) \right| \\ &= mn t^2 \mathbf{E} T_{1,m+1}^2 |Y_{1,m+1}|^m \end{aligned}$$

with $Y_{1,m+1} := \mathbf{E}(e\{t(T_n + \tau T_{1,n} + T_{m+1,n})\} \mid \tau, X_1, X_{m+1})$, cf. (40). As in (43) we have $|Y_{1,m+1}| \leq \exp\{-c_2 n^{-1}t^2\}$, and hence

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

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

We turn to h_2 . Here let

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

so that $h_2(t) = m(it)^2 \mathbf{E} U_{2,1} Q + (n-m)(it)^2 \mathbf{E} U_{2,m+1} Q$. Taking

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

we have that

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

First we look at $h_6(t)$. Noting that, cf. (49),

$$\mathbf{E}((U_{2,1}^*)^2 \mid X_1) \leq \frac{1}{2}n^2 \mathbf{E}(T_{1,2}^2 \mid X_1)^2,$$

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. (50),

$$\begin{aligned} |h_9(t)| &\ll m t^3 \mathbf{E} \mathbf{E}((U_{2,1}^*)^2 \mid X_1)^{1/2} \\ &\quad \times \mathbf{E} \left((P_4 + P_5 + \tau(P_3 - L_{1,n-m+1}))^2 \mid X_1 \right)^{1/2} \\ &\leq m t^3 \left(\mathbf{E}(P_4 + P_5 + \tau(P_3 - L_{1,n-m+1}))^2 \right)^{1/2} n \mathbf{E} T_{1,2}^2 \\ &\leq mn^{-1} a^{-1} t^3 (mn^{-1} a^{-1})^{1/2} \ll a^{-3/2} \log^{3/2} t. \end{aligned}$$

As to h_8 on the other hand we may condition on the random variables $\tau, X_1, \dots, X_{n-m}$, thus removing $P_2 + P_6$ as well from the exponent: taking $Y_{1,\tau}$ as in (52) we see that

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

cf. (43). 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 $T_{n-m+1}, \dots, T_n$ of the remaining variables. Using the fact that

$$\mathbf{E}(U_{2,m+1}^2 \mid X_{m+1}) \ll m^2 \mathbf{E}(T_{1,m+1}^2 \mid X_{m+1})^2,$$

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 $U_3 = \sum_{1 \leq j < k \leq m} \sum_{m+1 \leq p < q \leq n} V_{j,k,p,q}$, cf. (54). The $V_{j,k,p,q}$ are (again) mutually uncorrelated, and, as in (55),

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

Taking

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

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

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

The independence of the linear parts $T_{n-m+1}, \dots, T_n$ and the remaining expressions U_3^* , P_2 , and P_6 leads us to the fact that $|h_{11}(t)| \ll t^{-6} \mathbf{E} |U_3^*| \ll a^{-1}(\log t) t^{-8}$, 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.

3.4. The replacement of s by σ . Assume that $\sigma^2 < \infty$. Let us show that the bound (17) for $\mathbf{T}/s$ implies the same bound for $\mathbf{T}/\sigma$. In fact, writing $\tilde{\mathbf{T}} := \mathbf{T} - \mathbf{E} \mathbf{T}$, since $\Gamma \ll \varepsilon + \delta$ we have that

$$\begin{aligned} \left| \mathbf{P} \left\{ \frac{\tilde{\mathbf{T}}}{\sigma} \leq x \right\} - G(x) \right| &\leq \left| \mathbf{P} \left\{ \frac{\tilde{\mathbf{T}}}{s} \leq s^{-1} \sigma x \right\} - G(s^{-1} \sigma x) \right| \\ &\quad + |G(s^{-1} \sigma x) - G(x)| \ll \varepsilon + \delta + |s^{-1} \sigma - 1|. \end{aligned}$$

Here

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

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

$$\Gamma_\sigma \ll \varepsilon + \delta \leq 8(\varepsilon_\sigma + \delta_\sigma).$$

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

REFERENCES

1. Alberink I. B. A Berry–Esseen bound for U -statistics in the non-i.i.d. case. — J. Theoret. Probab., 2000, v. 13, № 2, p. 519–533.
2. Alberink I. B. Berry–Esseen bounds for arbitrary statistics, 2000b (dissertation).
3. Alberink I. B., Bentkus V. Bounds for the concentration of asymptotically normal statistics. — Acta Applic. Math., 1999, v. 58, p. 11–59.
4. Alberink I. B., Bentkus V. Berry–Esseen bounds for von Mises and U -statistics. (To appear in Lith. Math. J.)
5. Bentkus V., Götze F. On minimal moment assumptions in Berry–Esseen theorems for U -statistics. — Теория вероятн. и ее примен., 1995, т. 40, в. 3, с. 596–614.
6. Bentkus V., Götze F., Zitikis R. Lower estimate for the convergence rate for U -statistics. — Ann. Probab., 1994, v. 22, p. 1707–1714.
7. Bentkus V., Götze F., van Zwet W. R. An Edgeworth expansion for symmetric statistics. — Ann. Statist., 1997, v. 25, p. 851–896.
8. Bickel P. Edgeworth expansion in nonparametric statistics. — Ann. Statist., 1974, v. 4, p. 108–156.
9. Bolthausen E., Götze F. The rate of convergence for multivariate sampling statistics. — Ann. Statist., 1993, v. 21, p. 1692–1710.
10. Bhattacharya R. N., Ranga Rao R. Normal Approximation and Asymptotic Expansions. Melbourne: Krieger, 1986.
11. Borovskikh Yu. V. U -statistics in Banach spaces. Utrecht: VSP, 1996.
12. Боровский Ю. В. О нормальной аппроксимации U -статистик. — Теория вероятн. и ее примен., 2000, т. 45, в. 3, с. 469–488.
13. Боровский Ю. В., Королюк В. С. Асимптотический анализ распределений статистики. Киев: Наукова Думка, 1984, 304 с.
14. Королюк В. С., Боровский Ю. В. Аппроксимация невырожденных U -статистик. — Теория вероятн. и ее примен., 1985, т. 30, в. 3, с. 417–426.
15. Borovskikh Yu. V., Korolyuk V. S. Theory of U -statistics. Dordrecht: Kluwer, 1994.
16. Callaert H., Janssen P. The Berry–Esseen theorem for U -statistics. — Ann. Statist., 1978, v. 6, p. 417–421.
17. Chan Y.-K., Wierman J. On the Berry–Esseen theorem for U -statistics. — Ann. Probab., 1977, v. 5, p. 136–139.
18. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. М.: Мир, 1984, 752 с.
19. Friedrich K. O. A Berry–Esseen bound for functions of independent random variables. — Ann. Statist., 1989, v. 17, p. 170–183.
20. Helmers R., van Zwet W. R. The Berry–Esseen bound for U -statistics. — Statistical Decision Theory and Related Topics III, Vol. 1 (S. S. Gupta and J. O. Berger, eds.), New York, Academic Press, 1982, p. 497–512.
21. Hoeffding W. A class of statistics with asymptotically normal distributions. — Ann. Math. Statist., 1948, v. 19, p. 293–325.
22. Lee A. J. U -statistics. Theory and practice. New York: Marcel Dekker, 1990.
23. Petrov V. V. Limit Theorems of Probability Theory: Sequences of Independent Random Variables, Clarendon Press, 1995, Oxford.
24. Serfling R. J. Approximation Theorems of Mathematical Statistics. New York: Wiley, 1980.
25. van Zwet W. R. A Berry–Esseen bound for symmetric statistics. — Z. Wahrscheinlichkeitstheor. verw. Geb., 1984, B. 66, S. 425–440.

Поступила в редакцию
26.I.2000