The Annals of Probability
1999, Vol. 27, No. 1, 454–521

OPTIMAL BOUNDS IN NON-GAUSSIAN LIMIT THEOREMS FOR U-STATISTICS[1]

BY V. BENTKUS AND F. GÖTZE

University of Bielefeld

Let $X, X_1, X_2, \ldots$ be i.i.d. random variables taking values in a measurable space $\mathscr{X}$. Let $\phi(x, y)$ and $\phi_1(x)$ denote measurable functions of the arguments $x, y \in \mathscr{X}$. Assuming that the kernel ϕ is symmetric and that $\mathbf{E}\phi(x, X) = 0$, for all x, and $\mathbf{E}\phi_1(X) = 0$, we consider U-statistics of type

$$T = N^{-1} \sum_{1 \le j < k \le N} \phi(X_j, X_k) + N^{-1/2} \sum_{1 \le j \le N} \phi_1(X_j).$$

It is known that the conditions $\mathbf{E}\phi^2(X, X_1) < \infty$ and $\mathbf{E}\phi_1^2(X) < \infty$ imply that the distribution function of T, say F, has a limit, say F_0, which can be described in terms of the eigenvalues of the Hilbert–Schmidt operator associated with the kernel $\phi(x, y)$. Under optimal moment conditions, we prove that

$$\Delta_N = \sup_x |F(x) - F_0(x) - F_1(x)| = \mathscr{O}(N^{-1}),$$

provided that at least nine eigenvalues of the operator do not vanish. Here F_1 denotes an Edgeworth-type correction. We provide explicit bounds for Δ_N and for the concentration functions of statistics of type T.

1. Introduction. Let $X, \overline{X}, X_1, \ldots, X_N$ be independent identically distributed (i.i.d.) random variables taking values in an arbitrary measurable space $(\mathscr{X}, \mathscr{B})$. Let ϕ_1: $\mathscr{X} \to \mathbb{R}$ and ϕ: $\mathscr{X}^2 \to \mathbb{R}$ be real-valued measurable functions. Assume that ϕ is symmetric, that is, $\phi(x, y) = \phi(y, x)$, for all $x, y \in \mathscr{X}$. Consider the U-statistic

$$T = \frac{1}{N} \sum_{1 \le i < j \le N} \phi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \phi_1(X_i), \tag{1.1}$$

assuming that

$$\mathbf{E}\phi_1(X) = 0, \qquad \mathbf{E}\phi(x, X) = 0, \quad \text{for all } x \in \mathscr{X}. \tag{1.2}$$

Our aim is to establish an Edgeworth expansion with remainder $\mathscr{O}(N^{-1})$, for the distribution function of T. We shall show that moment conditions on ϕ_1 and ϕ together with assumption that the limiting weighted χ^2 distribution has at least nine degrees of freedom, are sufficient for the result to hold.

Received August 1997; revised June 1998.

[1]Supported by the SFB 343 in Bielefeld.

AMS 1991 *subject classifications*. Primary 62E20; secondary 60F05.

Key words and phrases. U-statistics, degenerate U-statistics, von Mises statistics, symmetric statistics, central limit theorem, convergence rates, Berry–Esseen bounds, Edgeworth expansions, second order efficiency.

We shall establish as well explicit bounds of order $\mathscr{O}(N^{-1})$ for concentration functions of a somewhat larger class of statistics than (1.1).

We shall write

$$\begin{aligned} &\beta_s = \mathbf{E}|\phi_1(X)|^s, &&\gamma_s = \mathbf{E}|\phi(X, \overline{X})|^s, \\ &\sigma^2 = \gamma_2, &&\gamma_{s,r} = \mathbf{E}\Big(\mathbf{E}\big\{|\phi(X, \overline{X})|^s | X\big\}\Big)^r, \end{aligned} \tag{1.3}$$

and assume throughout that

$$\beta_2 < \infty, \qquad 0 < \sigma^2 < \infty. \tag{1.4}$$

The variance of T splits as

$$\mathbf{E}T^2 = \beta_2 + \frac{N-1}{2N}\sigma^2. \tag{1.5}$$

The statistic T is called degenerate since the condition $\sigma^2 > 0$ ensures that the quadratic part of the statistic is not asymptotically negligible and therefore T is not asymptotically normal. More precisely, the asymptotic distribution of T is non-Gaussian and is given by the distribution of the random variable

$$T_0 = \tfrac{1}{2}\sum_{j\ge 1} q_j\big(\eta_j^2 - 1\big) + \sum_{j\ge 0} a_j\eta_j. \tag{1.6}$$

Here η_j is a sequence of i.i.d. standard normal random variables, $a_0, a_1, \dots$ denotes a (nonrandom) sequence of square summable weights and $q_1, q_2, \dots$ denote eigenvalues of the Hilbert–Schmidt operator, say $\mathbb{Q}$, associated with the kernel ϕ (see Section 2 for detailed definitions). Without loss of generality, we shall assume throughout that $|q_1| \ge |q_2| \ge \cdots$.

Introduce the distribution functions

$$F(x) = \mathbf{P}\{T \le x\}, \qquad F_0(x) = \mathbf{P}\{T_0 \le x\}. \tag{1.7}$$

Write

$$\Delta_N = \sup_x |\Delta_N(x)|, \qquad \Delta_N(x) = F(x) - F_0(x) - F_1(x), \tag{1.8}$$

where $F_1(x)$ denotes an Edgeworth correction defined by (4.1). Let us notice here only that F_1 vanishes if $\phi_1 = 0$, or if

$$\mathbf{E}\phi_1^3(X) = \mathbf{E}\phi_1^2(X)\phi(X, x) = \mathbf{E}\phi_1(X)\phi^2(X, x) = \mathbf{E}\phi^3(X, x) = 0 \tag{1.9}$$

holds for all $x \in \mathscr{X}$.

We shall denote by $c, c_1, \dots$ absolute generic constants. If a constant depends on, say s, we shall write $c(s)$ or c_s.

THEOREM 1.1.

(i) *Assume that* $q_{13} \ne 0$. *Then we have*

$$\Delta_N \le \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\beta_3^2}{\sigma^6} + \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,2}}{\sigma^4}\right) \quad \textit{where } C \le \exp\left\{\frac{c\sigma}{|q_{13}|}\right\}. \tag{1.10}$$

(ii) *Assume that* (1.9) *holds and* $q_9 \neq 0$. *Then we have*

$$\Delta_N \le \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,2}}{\sigma^4}\right) \quad \textit{where } C \le \exp\left\{\frac{c\sigma}{|q_9|}\right\}. \tag{1.11}$$

Notice that $\gamma_{2,2} \le \gamma_4$. Thus, for Δ_N in (1.10) and (1.11) we have

$$\Delta_N \le \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\beta_3^2}{\sigma^6} + \frac{\gamma_4}{\sigma^4}\right) \quad \text{and} \quad \Delta_N \le \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\gamma_4}{\sigma^4}\right), \tag{1.12}$$

respectively. See Example 9.4 for a kernel ϕ such that $\gamma_{2,2} < \infty$ and $\gamma_4 = \infty$.

Theorem 1.1 follows from more general bounds for Δ_N in Section 4.

Unfortunately, the bound (1.10) is not applicable when $q_{13} = 0$. In Section 9 we prove Theorem 9.2, which holds for statistics of finite dimension, say d, such that $q_9 \neq 0$. However, comparing with (1.10), the bound of Theorem 9.2 depends on the smallest nonzero eigenvalue q_d. Furthermore, in Section 9 we extend Theorem 1.1 to von Mises statistics; see Theorem 9.1.

Let us formulate a related result for concentration functions of statistics

$$T_* = \sum_{1 \le j < k \le N} \phi(X_j, X_k) + f_1 + f_2, \tag{1.13}$$

where $f_1 = f_1(X_1, \dots, X_M)$ is an arbitrary statistic depending only on $X_1, \dots, X_M$, and $f_2 = f_2(X_{M+1}, \dots, X_N)$ is as well arbitrary but independent of $X_1, \dots, X_M$, for some $1 \le M \le N/2$. Note that the class of statistics (1.13) is slightly more general than the class of statistics (1.1). In applications, the most interesting case is $M \ge cN$, with some $c > 0$ independent of N, and in this case the bound of Theorem 1.2 is $\mathscr{O}(N^{-1})$. Write

$$Q(T_*; \lambda) = \sup_x \mathbf{P}\{x \le T_* \le x + \lambda\}, \qquad \lambda \ge 0. \tag{1.14}$$

THEOREM 1.2. *Let* $q_9 \neq 0$. *Then*

$$Q(T_*; \lambda) \le \frac{C}{M} \max\left\{\frac{\lambda}{\sigma}; \frac{\gamma_3^2}{\sigma^6}\right\} \quad \textit{where } C \le \exp\left\{\frac{c\sigma}{|q_9|}\right\}. \tag{1.15}$$

See Section 3 for refinements of Theorem 1.2.

In particular cases Theorems 1.1, 1.2, 9.1 and 9.2 are closely related to the well-known lattice point problem in number theory. Let $\mathbb{A}: \mathbb{R}^d \to \mathbb{R}^d$ denote a positive linear operator. The relative lattice point remainder for ellipsoids $E_r = \{x \in \mathbb{R}^d: \langle \mathbb{A}x, x\rangle \le r^2\}$ is defined as

$$\delta(r) = (\mathrm{vol}_{\mathbb{Z}} E_r - \mathrm{vol}\, R_r)/\mathrm{vol}\, E_r,$$

where $\langle x, y\rangle$ denotes the standard scalar product in $\mathbb{R}^d$, and $\mathrm{vol}_{\mathbb{Z}} E_r$ and $\mathrm{vol}\, E_r$ are the number of points of the standard lattice $\mathbb{Z}^d$ in E_r and the volume of E_r, respectively. Let $Y, Y_1, Y_2, \dots$ be a sequence of i.i.d. bounded centered lattice random vectors taking values in $\mathbb{Z}^d$ and such that Y has covariance $\mathbb{A}^{-1}$. Yarnold (1972) showed that

$$\mathbf{P}\{\langle \mathbb{A}S_N, S_N\rangle \le t\} = F_0(t) + e^{-t/2}\delta(\sqrt{tN}) + \mathscr{O}(N^{-1}), \tag{1.16}$$

where $S_N = N^{-1/2}(Y_1 + \cdots + Y_N)$ and $F_0(t)$ denotes the χ^2 distribution function with d degrees of freedom. The statistic $\langle \mathbb{A} S_N, S_N \rangle$ is a degenerate U-statistic of a very special type since

$$\langle \mathbb{A} S_N, S_N \rangle - d = \frac{2}{N} \sum_{1 \le j < k \le N} \langle \mathbb{A} Y_j, Y_k \rangle + \frac{1}{N} \sum_{1 \le j \le N} \big(\langle \mathbb{A} Y_j, Y_j \rangle - d \big).$$

Therefore, a comparison of (1.16) with expansions of Theorems 1.1 and 9.2 shows that $\delta(r) = \mathscr{O}(r^{-2})$, as $r \to \infty$, for $d \ge 9$. See Bentkus and Götze (henceforth called BG) (1995, 1997a, 1997d) for directly proved optimal and explicit bounds for the lattice remainder. This result solves the lattice remainder problem for general ellipsoids, which has been open since 1915 when Landau obtained the upper bound $\delta(r) = \mathscr{O}(r^{-2+2/(d+1)})$. Esseen's (1945) bound $\mathscr{O}(N^{-d/(d+1)})$ in the CLT for ellipsoids in $\mathbb{R}^d$ in conjunction with (1.16) provides an alternative proof of Landau's result.

The bounds of Theorems 1.1, 1.2, 9.1 and 9.2 are optimal with respect to the dependence on N in view of the following example.

EXAMPLE 1.3. Assume that $\phi_1 = 0$ and that the kernel ϕ is bounded and takes integer values only. The convergence of T in distribution to a limit T_0 implies

$$\mathbf{P}\{|T| \le 1\} = \mathbf{P}\{|NT| \le N\} \ge c > 0, \tag{1.17}$$

for sufficiently large N. The statistic NT assumes integer values only and the interval $[-N, N]$ contains $2N + 1$ integers. Hence, there exists an integer $|j| \le N$ such that $\mathbf{P}\{|NT| = j\} \ge c_1/N$ and

$$\sup_t \mathbf{P}\{T = t\} \ge c_1/N. \tag{1.18}$$

The estimate (1.18) indicates that an $\mathscr{O}(N^{-1})$ bound for the concentration function in Theorem 1.2 is optimal. Similarly, the dependence on N in Theorems 1.1, 9.1 and 9.2 is optimal since the function $F_0 + F_1$ is continuous. Thus, in order to obtain bounds $o(N^{-1})$ we have either to impose additional smoothness conditions on the distributions involving ϕ, ϕ_1 and X or to replace F_1 by some highly distribution dependent discontinuous function.

Notice that the construction of the example extends to the case of bounded $\phi_1 \ne 0$ taking integer values only. Here (1.17) and (1.18) hold for sufficiently large $N = n^2$, $n \in \mathbb{N}$. Furthermore, one can choose ϕ such that the associated Hilbert–Schmidt operator $\mathbb{Q}$ has a given number, say s, $1 \le s \le \infty$, of nonzero eigenvalues.

It is likely that the dimensional dependence (that is, the dependence on the eigenvalues) of our results is not optimal. To prove $\mathscr{O}(N^{-1})$ we require that $q_9 \ne 0$. An optimal condition would be $q_5 \ne 0$. To prove (or disprove) that $q_5 \ne 0$ is sufficient for rates $\mathscr{O}(N^{-1})$ seems to be a difficult problems, since its solution combined with (1.16) would imply a solution of the corresponding unsolved problem for the lattice point remainder. The question about precise convergence rates in lower dimensions when $q_d \neq 0$ and $q_{d+1} = 0$, for some $2 \leq d \leq 4$, still remains completely open. For instance, in the case $q_2 \neq 0$ and $q_3 = 0$, a precise convergence rate would imply a solution of the circle problem. Known lower bounds in the circle problem correspond to $\mathscr{O}(N^{-3/4} \log^{\delta} N)$ in our set-up. A famous conjecture by Hardy (1916) says that, up to logarithmic factors, this is the true order.

The dependence on moments in Theorem 1.1 seems to be optimal since this theorem implies optimal bounds in the CLT for conic sections in $\mathbb{R}^d$, $d \geq 13$, and Hilbert spaces (see Section 9).

For degenerate U-statistics, rates of convergence and Edgeworth expansions were proved by Götze (1979, 1984) and by Koroljuk and Borovskich (1994). Assuming infinitely many nonzero eigenvalues and certain moment conditions, Götze (1979) proved a bound $\mathscr{O}(N^{-1+\varepsilon})$, $\varepsilon > 0$, using a Weyl-type symmetrization inequality [Weyl (1915/16)], In their monograph, Koroljuk and Borovskich prove bounds of $o(N^{-1/2})$. Götze and Zitikis (1995) obtained approximations with remainders $\mathscr{O}(N^{-1})$ and better under certain smoothness assumptions related to the kernel ϕ and the distribution of X.

Degenerate U-statistics (1.1) include as partial cases a number of other statistics [see, for example, Lee (1990)]. In particular, they include quadratic forms of sums of independent random vectors in finite-dimensional and Hilbert spaces. A special choice of these random vectors leads to empirical processes and, in particular, to the uniform empirical process on [0, 1] and its L^2-norm, the so-called ω^2-statistics of Cramér and von Mises. For the ω^2-statistic, the relevant kernel is given by

$$\phi(x, y) = \tfrac{1}{2}(x^2 + y^2) - \max\{x; y\} + \tfrac{1}{3} \quad \text{for } 0 \leq x, y \leq 1,$$

with eigenvalues $\lambda_k = (k\pi)^{-2}$. Related examples are Watson's goodness of fit statistics on a circle. The limiting distribution of such statistics has been studied by Smirnov (1937) and Anderson and Darling (1952). Rates of convergence have been proved, to name a few results, starting with rates like $\mathscr{O}((\log n)^{-1}$ [Kandelaki (1965)] to rates $\mathscr{O}(n^{-\alpha})$: $\alpha = 1/6$, Sazonov (1969); $\alpha = 1/4$, Kiefer (1972); $\mathscr{O}(n^{-1/2} \log n)$ via the KMT-approximation. Csörgö (1976). For random vectors in Hilbert spaces, Paulauskas (1976) obtained $\alpha = 1/6$. Assuming infinitely many positive eigenvalues, Götze (1979, 1984) proved $\alpha = 1 - \varepsilon$, $\varepsilon > 0$. Eigenvalue and moment conditions were improved, for example, by Yurinskii (1982, 1995), Bentkus (1984), Nagaev and Chebotarev (1986), Zalesskii, Sazonov and Ulyanov (1988), Senatov (1989). For ω^2-statistics, bounds of $\mathscr{O}(N^{-1})$ and better were obtained by Bentkus, Götze and Zitikis (1993). The rate $\alpha = 1$ for diagonal quadratic forms was proved in BG (1996) using a discretization method and the Hardy–Littlewood circle method from analytic number theory. In BG (1997b) the result was extended to the case of nondiagonal quadratic forms again via discretization techniques and a useful extension of Weyl's inequality (henceforth called multiplicative inequality) for characteristic functions of lattice distributions.

The class of U-statistics is much more general than that of quadratic forms. Using inequalities of the form $|\phi(x, y)| \le c\mathscr{N}(x)\mathscr{N}(y)$, where $\mathscr{N}$ is a norm function satisfying $\mathbf{E}\mathscr{N}^4(X) \le c\gamma_4$, one would be able to reduce estimates for U-statistics to estimates for quadratic forms like those derived in BG (1997b). Unfortunately, such inequalities for *general* U-statistics do not hold (e.g., for the kernel ϕ of Example 9.4). Therefore, we had to develop new versions of random selections, discretization and multiplicative inequalities. The bounds in terms of the given stochastic set-up involve moment and eigenvalue conditions only. In this sense they are more natural compared with those for quadratic forms in BG (1996, 1997b), where geometric conditions are superimposed on the given stochastic problem. In spite of the large generality, our results are sharp enough to recover optimal moment conditions in the case of quadratic forms; see Section 9. The methods developed turn out to be useful for investigations of Poissonian approximations of stable laws [Bentkus, Götze and Paulauskas (1996)], and for infinite divisible approximations of sums of independent random vectors [see Bentkus, Götze and Zaitsev (1997)].

The paper is organized as follows.

In Section 2 we introduce the notation used throughout and describe a representation of T as a sum of a quadratic form and a linear functional of random vectors taking values in $\mathbb{R}^\infty$. This simplifies the notation and certain algebraic operations like symmetrization.

In Section 3 we derive explicit bounds for the concentration functions of statistics (1.13). The proofs in Section 3 are simpler than those of bounds for Δ_N, although they already involve the main principal techniques.

In Section 4 we obtain the main result—explicit bounds for Δ_N, using some auxiliary results, which are proved in Sections 5–8. In particular, these explicit bounds are applicable in cases when random variables and the kernels ϕ_1 and ϕ depend on N and other parameters.

Section 5 contains certain facts related to the nondegeneracy condition (3.2), as well as a symmetrization lemma.

In Section 6 we introduce a randomization which allows to replace the statistic (1.1) by statistics of a special discrete type and establish upper bounds for their characteristic functions.

In Section 7 we prove the basic multiplicative inequality.

Section 8 contains asymptotic expansions of the characteristic functions.

In Section 9 we extend the bounds to von Mises statistics and prove Theorem 9.2. An application to the distributions of quadratic forms of sums of random vectors in Euclidean spaces shows that our bounds are indeed sharp, up to certain constants.

Finally, in Section 10 we show that the multiplicative inequalities imply desired bounds for the integrals of characteristic functions.

2. Special representations of bivariate U-statistics, notation and auxiliary results. Let us introduce notation related to the operator $\mathbb{Q}$. Consider the measurable space $(\mathscr{X}, \mathscr{B}, \mu)$ with measure $\mu = \mathscr{L}(X)$. Let $L^2 =$

$L^2(\mathscr{X}, \mathscr{B}, \mu)$ denote the real Hilbert space of square integrable real functions. The Hilbert–Schmidt operator $\mathbb{Q}$: $L^2 \to L^2$ is defined via

$$(2.1)\quad \mathbb{Q}f(x) = \int_{\mathscr{X}} \phi(x, y) f(y) \mu(dy) = \mathbf{E}\phi(x, X) f(X), \qquad f \in L^2.$$

Let $\{e_j\colon j \geq 1\}$ denote an orthonormal complete system of eigenfunctions of $\mathbb{Q}$ ordered by decreasing absolute values of the corresponding eigenvalues $q_1, q_2, \dots$, that is, $|q_1| \geq |q_2| \geq \cdots$. Then

$$(2.2)\quad \sigma^2 = \mathbf{E}\phi^2(X, \overline{X}) = \sum_{j \geq 1} q_j^2 < \infty, \qquad \phi(x, y) = \sum_{j \geq 1} q_j e_j(x) e_j(y)$$

since $\mathbb{Q}$ is the Hilbert–Schmidt operator and the kernel ϕ is degenerate. The series in (2.2) converges in $L^2(\mathscr{X}^2, \mathscr{B}^2, \mu \times \mu)$. Consider the subspace $L^2(\phi_1, \phi) \subset L^2$ generated by ϕ_1 and eigenfunctions e_j corresponding to nonzero eigenvalues $q_j \neq 0$. Introducing, if necessary, a normalized eigenfunction, say e_0, such that $\mathbb{Q}e_0 = 0$, we can assume that $e_0, e_1, \dots$ is an orthonormal basis of $L^2(\phi_1, \phi)$. Thus, we can write

$$(2.3)\qquad \phi_1(X) = \sum_{j \geq 0} a_j e_j(X) \quad \text{in } L^2, \qquad \beta_2 = \mathbf{E}\phi_1^2(X) = \sum_{j \geq 0} a_j^2,$$

with $a_j = \mathbf{E}\phi_1(X) e_j(X)$. It is easy to see that $\mathbf{E}e_j(X) = 0$, for all j. Therefore $(e_j(X))_{j \geq 0}$ is an orthonormal system of mean zero random variables.

For a random variable X, we shall denote by $\overline{X}$, X_j, $\overline{X}_j$, its independent copies. Throughout we shall assume as well that all random variables and vectors are independent in aggregate, if the contrary is not clear from the context.

A special representation of U-statistics. Let $\mathbb{R}^\infty$ denote the space of all real sequences $x = (x_0, x_1, x_2, \dots)$, $x_j \in \mathbb{R}$. The Hilbert space $l_2 \subset \mathbb{R}^\infty$ consists of $x \in \mathbb{R}^\infty$, such that

$$|x|^2 =_{\text{def}} \langle x, x \rangle, \quad |x| < \infty, \qquad \langle x, y \rangle = \sum_{j \geq 0} x_j y_j.$$

Consider the random vector

$$(2.4)\qquad \mathbf{X} =_{\text{def}} (e_0(X), e_1(X), e_2(X), \dots),$$

which takes values in $\mathbb{R}^\infty$. Since $\{e_j(X)\}_{j \geq 0}$ is a system of mean zero uncorrelated random variables with variances 1, the random vector $\mathbf{X}$ has identity covariance and mean zero. Using (2.2) and (2.3), we can write

$$(2.5)\qquad \phi(X, \overline{X}) = \langle \mathbb{Q}\mathbf{X}, \overline{\mathbf{X}} \rangle, \qquad \phi_1(X) = \langle a, \mathbf{X} \rangle,$$

where we define $\mathbb{Q}x = (0, q_1 x_1, q_2 x_2, \dots)$, for $x \in \mathbb{R}^\infty$, and $a = (a_j)_{j \geq 0} \in \mathbb{R}^\infty$. The equalities (2.5) allow us to assume throughout that the measurable space $\mathscr{X}$ is $\mathbb{R}^\infty$, the random variable X is a random vector taking values in $\mathbb{R}^\infty$ with mean zero and identity covariance and that

$$(2.6)\qquad \phi(X, \overline{X}) = \langle \mathbb{Q}X, \overline{X} \rangle, \qquad \phi_1(X) = \langle a, X \rangle.$$

In particular, without loss of generality we shall assume throughout that the kernels $\phi(x, y)$ and $\phi_1(x)$ are linear functions in each of their arguments. This assumption is of a technical nature since our proofs are independent of it. However, it is technically very convenient since it saves lots of additional notation and yields more intuitive arguments involving sums of random vectors. There is a simple example illustrating the advantages: in particular, we have to consider $\phi(\tilde{S}, \tilde{\bar{S}})$, where $S = X_1 + \cdots + X_m$ is a sum, $\tilde{S}$ denotes a symmetrization of random vector S and $\bar{S}$ is an independent copy of S. Alternatively, we would have to deal with

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \Big(\phi(X_i, X_j) - \phi\big(X_i, \bar{X}_j\big) - \phi\big(\bar{X}_i, X_j\big) + \phi\big(\bar{X}_i, \bar{X}_j\big)\Big).$$

instead of $\phi(\tilde{S}, \tilde{\bar{S}})$.

Introduce the Gaussian random vector $G = (\eta_0, \eta_1, \dots)$ with values in $\mathbb{R}^\infty$, where $\eta_0, \eta_1, \dots$ denote i.i.d. standard normal random variables. The random vectors G and X have mean zero and equal covariances, that is,

$$\begin{gathered} \mathbf{E}\phi_1(G) = \mathbf{E}\phi(G, x) = 0, \qquad \mathbf{E}\phi_1^2(G) = \mathbf{E}\phi_1^2(X), \\ \mathbf{E}\phi_1(G)\phi(G, x) = \mathbf{E}\phi_1(X)\phi(X, x), \\ \mathbf{E}\phi(G, x)\phi(G, y) = \mathbf{E}\phi(X, x)\phi(X, y). \end{gathered} \tag{2.7}$$

Furthermore, we have

$$\begin{aligned} \mathbf{E}|\phi_1(G)|^s &\le c_s \mathbf{E}|\phi_1(X)|^s, \\ \mathbf{E}|\phi(G, x)|^s &\le c_s \mathbf{E}|\phi(X, x)|^s, \qquad s \ge 2. \end{aligned} \tag{2.8}$$

Indeed, the random variable $\phi_1(G)$ [or $\phi(G, x)$] is Gaussian and therefore

$$\mathbf{E}|\phi_1(G)|^s \le c_s\big(\mathbf{E}\phi_1^2(G)\big)^{s/2} = c_s\big(\mathbf{E}\phi_1^2(X)\big)^{s/2} \le c_s \mathbf{E}|\phi_1(X)|^s.$$

Introduce the statistic

$$T(G_1, \dots, G_N) = \frac{1}{N} \sum_{1 \le i < j \le N} \phi\big(G_i, G_j\big) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \phi_1(G_i) \tag{2.9}$$

with

$$\phi_1(G) = \langle a, G \rangle, \qquad \phi(G, \bar{G}) = \langle \mathbb{Q}G, \bar{G} \rangle.$$

Define the random variable

$$\phi_0(G, G) = \sum_{j \ge 1} q_j\big(\eta_j^2 - 1\big). \tag{2.10}$$

The series converges in L^2 since $\sum_j q_j^2 < \infty$. Furthermore, using finite-dimensional approximations and substituting $\sqrt{N}G =_{\mathscr{D}} G_1 + \cdots + G_N$ in (2.10), we see that

$$\phi_0(G, G) =_{\mathscr{D}} \frac{2}{N} \sum_{1 \le i < j \le N} \phi\big(G_i, G_j\big) + \frac{1}{N} \sum_{1 \le j \le N} \phi_0\big(G_j, G_j\big). \tag{2.11}$$

Consequently, for the statistic T_0 defined by (1.6), we have

$$\begin{aligned} T_0 &= \frac{1}{2}\sum_{j\ge 1} q_j\big(\eta_j^2 - 1\big) + \sum_{j\ge 0} a_j\eta_j = \frac{1}{2}\phi_0(G,G) + \phi_1(G) \\ &=_{\mathscr{D}} T(G_1,\dots,G_N) + \frac{1}{2N}\sum_{1\le k\le N}\phi_0(G_k,G_k). \end{aligned} \tag{2.12}$$

Let us introduce some additional notation which will be used throughout the paper. We write $[b]$ for the integer part of a number b. We denote $\mathrm{e}\{x\} = \exp\{ix\}$. We write $A \ll B$ (resp. $A \ll_s B$) if $A \le cB$ (resp. $A \le c_s B$). Furthermore, we write $A \asymp B$ (resp. $A \asymp_s B$) if $A \ll B \ll A$ (resp. $A \ll_s B \ll_s A$).

For a matrix $\mathbb{A} = (a_{ij})$, $1 \le i, j \le s$ (resp. for the related linear operator), we denote by $|\mathbb{A}| = \sup_{|x|=1}|\mathbb{A}x|$ its operator norm. We shall use as well the norms

$$|\mathbb{A}|_2^2 = \sum_{1\le i,j\le s} a_{ij}^2, \qquad |\mathbb{A}|_\infty = \max_{1\le i,j\le s}|a_{ij}|.$$

Throughout $\mathbb{I}$ stands for the identity matrix (resp. identity operator).

Introduce the function

$$\begin{aligned} \mathscr{M}(t;N) &= 1/\sqrt{|t|N} \quad \text{for } |t| \le N^{-1/2}, \\ \mathscr{M}(t;N) &= \sqrt{|t|} \quad \text{for } |t| \ge N^{-1/2}. \end{aligned} \tag{2.13}$$

Notice that, for $s > 0$,

$$2^{-1}\big(|tN|^{-s/2} + |t|^{s/2}\big) \le \mathscr{M}^s(t;N) \le |tN|^{-s/2} + |t|^{s/2} \tag{2.14}$$

and

$$\inf_{t\in\mathbb{R}}\mathscr{M}^s(t;N) = N^{-s/4}. \tag{2.15}$$

For positive constants $c_1(s)$ and $c_2(s)$, the inequalities (2.14) imply

$$\mathscr{M}^s(c_1(s)t;c_2(s)N) \ll_s \mathscr{M}^s(t;N) \ll_s \mathscr{M}^s(c_1(s)t;c_2(s)N). \tag{2.16}$$

For subsets of the set $\mathbb{Z}$ of integer numbers, we write

$$[1,N] = \{1,\dots,N\}, \qquad (a,b] = \{j\in\mathbb{Z}: a < j \le b\}$$

and so on. By $|A|$ we denote the number of elements of a set $A \subset \mathbb{Z}$. We denote by $\mathbb{N}$ the set of natural numbers, and $\mathbb{Z}_+ = \mathbb{N}\cup\{0\}$.

If a random vector takes values in a linear space, then $\tilde{X} = X - \overline{X}$ denotes its symmetrization. For conditional expectations we shall use the following notation (and its natural extensions):

$$\mathbf{E}\{f(X,Y)|Y\} = \mathbf{E}_X f(X,Y) = \mathbf{E}^Y f(X,Y).$$

Let $\xi_1,\dots,\xi_N$ denote independent real random variables which have mean zero. Then the following Rosenthal type inequality holds:

$$\mathbf{E}\Big|\sum_{j=1}^N \xi_j\Big|^q \ll_q \sum_{j=1}^N \mathbf{E}|\xi_j|^q + \bigg(\sum_{j=1}^N \mathbf{E}|\xi_j|^2\bigg)^{q/2}, \qquad 2 \le q < \infty. \tag{2.17}$$

The inequality (2.17), combined with Hölder's inequality, implies that

$$\mathbf{E}\Big|\sum_{j=1}^{N} \xi_j\Big|^q \ll_q N^{-1+q/2} \sum_{j=1}^{N} \mathbf{E}|\xi_j|^q, \qquad 2 \le q < \infty. \tag{2.18}$$

Consider a sufficiently smooth function $f\colon \mathbb{E} \to \mathbb{B}$, where $\mathbb{E}$ and $\mathbb{B}$ denote real Banach spaces. Let $\tau \in [0,1]$ be a random variable uniformly distributed on $[0,1]$ and independent of all other random variables. For $r = 0, 1, \dots,$ we shall use the Taylor formula

$$\begin{aligned} f(x) &= f(0) + f'(0)x + \cdots \\ &\quad + \frac{1}{r!} f^{(r)}(0)x^r + \frac{1}{r!}\mathbf{E}(1-\tau)^r f^{(r+1)}(\tau x)x^{r+1}. \end{aligned} \tag{2.19}$$

3. Bounds for concentration functions. In this section we shall consider the statistic T_* defined by (1.13). In the case $M \ge cN$, with some $c > 0$, the bounds of Theorem 3.1 are $\mathscr{O}(N^{-1})$.

We shall use the following nondegeneracy condition $\mathscr{N}(p, \delta, s, Z)$ for the distribution of a random vector Z, a kernel ϕ, parameters $0 < p \le 1$, $\delta \ge 0$ and $s \in \mathbb{N}$. Define the (random) matrix

$$\mathbb{A} = \mathbb{A}(Z) = (a_{ij})_{1 \le i, j \le s} \quad \text{with entries } a_{ij} = \phi\big(Z_i, \overline{Z}_j\big), \tag{3.1}$$

where Z_i and $\overline{Z}_j$ are independent copies of Z. We say that Z and ϕ satisfy the nondegeneracy condition $\mathscr{N}(p, \delta, s, Z)$ if

$$\mathbf{P}\{|\mathbb{A}(Z) - \mathbb{I}|_\infty \le \delta\} = \mathbf{P}\Big\{\max_{1 \le i, j \le s} \Big|\phi\big(Z_i, \overline{Z}_j\big) - \delta_{ij}\Big| \le \delta\Big\} \ge p, \tag{3.2}$$

where $\mathbb{I}$ denotes the identity matrix and δ_{ij} is Kronecker's symbol. A Gaussian random vector G satisfies this condition for any $s \ge 1$ and $\delta > 0$ with some $p > 0$, provided that at least s eigenvalues of the operator $\mathbb{Q}$ are nonzero (see Lemma 5.3). Another example of a random vector which satisfies the condition is

$$Y = (2m)^{-1/2}\big(\tilde{X}_1 + \cdots + \tilde{X}_m\big) \tag{3.3}$$

provided that $q_s \ne 0$ and m is sufficiently large. For the random variable Y defined by (3.3), the condition $\mathscr{N}(p, \delta, s, Y)$ can be expressed in terms of ϕ and $X, X_1, \dots, X_N$, $N \ge 2ms^2$, as follows. Let $X_i, \overline{X}_j, 1 \le i,\ j \le s,$ denote independent copies of X. Consider $2m$ independent copies $X_{ik}, \overline{X}_{jl}$ and $U_{ik}, \overline{U}_{jl}$, $1 \le k, l \le m$, of the sequence $X_i, \overline{X}_j, 1 \le i,\ j \le s$. Then the entries a_{ij} of the random matrix $\mathbb{A}(Y)$ can be written as

$$a_{ij} = \frac{1}{2m} \sum_{k,l=1}^{m} \Big(\phi\big(X_{ik}, \overline{X}_{jl}\big) - \phi\big(X_{ik}, \overline{U}_{jl}\big) - \phi\big(U_{ik}, \overline{X}_{jl}\big) + \phi\big(U_{ik}, \overline{U}_{jl}\big)\Big).$$

THEOREM 3.1. *Let $s \geq 9$ and $\lambda \geq 0$. Then, for the statistic T_* and its concentration function $Q(T_*;\lambda)$ defined by* (1.13) *and* (1.14) *respectively, we have*:

(i) *Let $q_s \neq 0$. Then*

$$(3.4)\qquad Q(T_*;\lambda) \ll_s \exp\left\{\frac{c_s\sigma}{|q_s|}\right\}\frac{\max\{\lambda/\sigma;m_0\}}{M},\qquad m_0 = \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,3/2}^2}{\sigma^6}.$$

(ii) *Assume that the condition $\mathscr{N}(p,(4s)^{-1},s,G)$ holds. Then*

$$(3.5)\qquad Q(T_*,\lambda) \ll_s \frac{\max\{\lambda;m_0\}}{pM}\quad \text{with } m_0 = \frac{\gamma_3}{p} + \frac{\gamma_{2,3/2}^2}{p^2}.$$

(iii) *Let $m \in \mathbb{N}$. Assume that the random vector Y defined by* (3.3) *satisfies the condition $\mathscr{N}(p,(2s)^{-1},s,Y)$. Then we have*

$$(3.6)\qquad Q(T_*;\lambda) \ll_s \frac{\max\{\lambda;m\}}{pM}.$$

REMARK. The bound (3.6) of Theorem 3.1 is valid without any moment assumptions. The bounds (3.4) and (3.5) imply $Q(T_*;\lambda) = \mathscr{O}((\lambda+1)/M)$ provided that $\gamma_3 < \infty$. An inspection of the proofs shows that the condition $\gamma_3 < \infty$ can be relaxed to $\gamma_{2+\varepsilon} < \infty$, for some $\varepsilon > 0$.

To prove Theorem 3.1, we shall apply a well-known inequality (3.12) which provides an upper bound for the concentration function through an integral of the characteristic function. One usually proceeds by showing that the characteristic function $|\mathbf{E}e\{tT_*\}|$ allows an upper bound such that it is integrable over the interval of interest. Unfortunately such upper bounds are not available in our situation since the characteristic function may have relatively high peaks. For example, it may happen that $|\mathbf{E}e\{t_0T_*\}| \asymp 1$, for some t_0 in the interval of integration. Moreover, a description of the location of peaks is not available. Nevertheless, we can estimate the integral, using the fact that the peaks are very narrow and are separated by wide stretches of "lowland." An analytical tool which implicitly involves all these features is a multiplicative inequality of type (3.7). The inequality holds for characteristic functions of the statistics of a special discrete type only. Therefore we need to apply a discretization procedure first (see Section 6). The discretization is achieved writing the distribution of T as an average over random discrete statistics. In particular, the integral of interest is bounded from above by $\mathbf{E}\int_a^b\psi(t)(dt/|t|)$ with some $a, b \in \mathbb{R}$ and a random function ψ which is the characteristic function of a discrete random statistic. The function ψ satisfies the multiplicative inequality (3.7) with high probability. Thus, the integral can be bounded using the following Lemma 3.2 [see BG (1997b)]. This lemma slightly extends Lemma 6.1 in BG (1997a). Its proof is given in Section 10.

LEMMA 3.2. *Let $\varphi(t)$, $t \geq 0$, denote a continuous function such that $\varphi(0) = 1$ and $0 \leq \varphi \leq 1$. Assume that*

$$(3.7)\qquad \varphi(t)\varphi(t+\tau) \leq \Theta\mathscr{M}^s(\tau;N)\quad \text{for all } t \geq 0 \text{ and } \tau \geq 0,$$

with some $\Theta \geq 1$ *independent of* t *and* τ. *Then, for any* $A \geq 1$, $0 < B \leq 1$ *and* $N \geq 1$,

$$\int_{B/\sqrt{N}}^{A} \varphi(t)\,\frac{dt}{t} \ll_s \frac{\Theta^2(1+\log A)}{N} + \Theta^2 B^{-s/2} N^{-s/4} \quad \textit{for } s > 8.$$

For $A \geq t_0$, $t_1 \geq 0$, define the integrals

$$I_0 = \int_{-t_1}^{t_1} \bigl|\hat{\Psi}(t)\bigr|\,dt, \qquad I_1 = \int_{t_0 \leq |t| \leq A} \bigl|\hat{\Psi}(t)\bigr|\,\frac{dt}{|t|},$$

where $\hat{\Psi} = \int_{\mathbb{R}} \mathrm{e}\{tx\}\,d\Psi(x)$ denotes the Fourier–Stieltjes transform of the distribution function $\Psi(x) = \mathbf{P}\{T_* \leq x\}$.

LEMMA 3.3. *Let* $m \in \mathbb{N}$. *Assume that the random vector* Y *defined by* (3.3) *satisfies the condition* $\mathscr{N}(p,\delta,s,Y)$ *with some* $0 \leq \delta \leq 1/(2s)$ *and* $s \geq 9$. *Let*

$$k = \frac{pM}{m}, \qquad t_0 = \frac{c_0(s)}{m} k^{-1+2/s}, \qquad t_1 = \frac{c_1(s)}{m} k^{-1/2},$$
$$\frac{c_2(s)}{m} \leq A \leq \frac{c_3(s)}{m}$$

with M *as in* (1.13) *and some positive constants* $c_j(s)$, $0 \leq j \leq 3$. *Then*

$$I_0 \ll_s (pM)^{-1}, \qquad I_1 \ll_s m(pM)^{-1}. \tag{3.8}$$

PROOF. Without loss of generality, we shall assume that $k \geq c_s$, for a sufficiently large constant c_s. Indeed, if $k \leq c_s$, then we can derive (3.8) using $|\hat{\Psi}| \leq 1$. Another consequence of $k \geq c_s$ is that $1/(km) \leq t_0 \leq t_1 \leq A$.

Let us prove (3.8) for I_0. By Theorem 6.2 we have $|\hat{\Psi}(t)| \ll_s \mathscr{M}^s(tm;k)$. Using the obvious inequality $|\hat{\Psi}| \leq 1$, we get $|\hat{\Psi}(t)| \ll_s \min\{1;\mathscr{M}^s(tm;k)\}$. Furthermore, denoting $t_2 = m^{-1}k^{-1/2}\max\{1;c_1(s)\}$ and using the definition of the function $\mathscr{M}$, we obtain

$$\begin{aligned} I_0 &\ll_s \int_0^{1/(km)} dt + \int_{1/(km)}^{\infty} \frac{dt}{(tmk)^{s/2}} + \int_0^{t_2} (tm)^{s/2}\,dt \\ &= \frac{1}{km} + \frac{c_s}{km} + \frac{c_s}{k^{(s+2)/4} m} \ll_s \frac{1}{km}, \end{aligned}$$

thus proving (3.8) for I_0.

It remains to estimate I_1. We shall reduce the proof to an application of Lemma 3.2. Using Lemma 6.3, we have

$$I_1 = \int_{t_0 \leq |t| \leq A} \bigl|\hat{\Psi}(t)\bigr|\,\frac{dt}{|t|} \leq \mathbf{E}\int_{t_0 \leq |t| \leq A} \psi(t)\,\frac{dt}{|t|}, \tag{3.9}$$

where a random function $\psi = |\mathbf{E}_\vartheta\, \mathrm{e}\{T^\vartheta\}|$ (defined in Lemma 6.3) may depend on $X_1,\dots,X_N$. By Lemma 7.1, there exists a (random) event D such that its complement D^c satisfies

$$\mathbf{P}\{D^c\} \ll_{s,d} k^{-d}, \qquad d \geq 0,$$

and

$$\mathbf{1}\{D\}\psi(t-\gamma)\psi(t+\gamma) \ll_{s,d} \mathscr{M}^s(\gamma m; k) \qquad \text{for all } t, \gamma \in \mathbb{R}. \tag{3.10}$$

Thus, (3.9) yields

$$\begin{aligned} &I_1 \ll_{s,d} k^{-d} \log(A/t_0) + \mathbf{E} I, \\ &I = \int_{t_0 \leq |t| \leq A} \varphi_0(t) \frac{dt}{|t|} \qquad \varphi_0 = \mathbf{1}\{D\}\psi. \end{aligned} \tag{3.11}$$

Clearly, $k^{-d}\log(A/t_0) \ll_s k^{-d}\log k \ll k^{-1}$ provided that we choose $d = 2$. We have

$$I = \int_{mt_0 \leq |t| \leq mA} \varphi(t) \frac{dt}{|t|} \quad \text{where } \varphi(t) = \varphi_0(t/m).$$

The inequality (3.10) yields $\varphi(t-\gamma)\varphi(t+\gamma) \ll_{s,d} \mathscr{M}^s(\gamma; k)$. Hence, in order to estimate I in (3.11), we can use Lemma 3.2. Replacing in this lemma N by k and A by mA respectively, choosing $B = \sqrt{k}\, mt_0$, we obtain $I \ll_s 1/k$. Thus, (3.11) implies $I_1 \ll_s k^{-1}$, proving (3.8) for I_1. □

Proof of Theorem 3.1. First we shall prove (iii) and then (iii) ⇒ (ii) ⇒ (i).

Let us prove (iii). Without loss of generality, we shall assume that $m < pM$, since otherwise the result follows from the trivial estimate $Q(T_*; \lambda) \leq 1$.

It is well known that

$$Q(T_*; \lambda) \leq 2\max\left\{\lambda; \frac{1}{A}\right\} \int_{-A}^{A} \left|\hat{\Psi}(t)\right| dt, \tag{3.12}$$

for any $A > 0$ [see Petrov (1975), Lemma 3 of Chapter 3]. Choose

$$A = c_s m^{-1}, \qquad t_1 = m^{-1}(pM/m)^{-1/2}, \qquad t_0 = c_0(s) m^{-1}(pM/m)^{1+2/s}.$$

We can assume that $t_0 \leq t_1$ since otherwise (iii) follows from $Q(T_*; \lambda) \leq 1$. The estimate (3.12) yields (iii) provided we show that

$$J = \int_{|t| \leq A} \left|\hat{\Psi}(t)\right| dt \ll_s (pM)^{-1}, \tag{3.13}$$

assuming that the condition $\mathscr{N}(p, (2s)^{-1}, s, Y)$ is fulfilled. Using the obvious inequality $1 \ll_s |mt|^{-1}$, for $|t| \leq c_s/m = A$, we have

$$\begin{aligned} &J \ll_s I_0 + m^{-1} I_1 \quad \text{with } I_0 = \int_{-t_1}^{t_1} \left|\hat{\Psi}(t)\right| dt, \\ &I_1 = \int_{t_0 \leq |t| \leq A} \left|\hat{\Psi}(t)\right| \frac{dt}{|t|}, \end{aligned} \tag{3.14}$$

since $t_0 \leq t_1$. Lemma 3.3 yields $I_0 \ll_s (pM)^{-1}$ and $I_1 \ll_s m(pM)^{-1}$, proving (3.13) and (iii).

Let us prove that (iii) ⇒ (ii). By Lemma 5.4, the condition $\mathscr{N}(2p, \delta, s, G)$ implies $\mathscr{N}(p, 2\delta, s, Y)$, provided that

$$m \geq \frac{c_s}{\delta^3 p}\left(\frac{\gamma_{2,3/2}^2}{\delta^3 p} + \gamma_3\right). \tag{3.15}$$

Choosing in (3.15) the minimal possible m, we derive (ii).

Let us prove that (ii) ⇒ (i). Write $\varepsilon^2 = \delta/(4s\sigma)$ and $\delta = (4s)^{-1}$. Then

$$Q(T_*; \lambda) = \sup_x \mathbf{P}\{\varepsilon^2 x \leq \varepsilon^2 T_* \leq \varepsilon^2 x + \varepsilon^2\lambda\}. \tag{3.16}$$

The statistic $\varepsilon^2 T_*$ is a statistic of type (3.1) with kernel $\varepsilon^2\phi$. By Lemma 5.3, the condition $\mathscr{N}(p, \delta, s, \varepsilon G)$ is fulfilled with $p = \exp\{-c_s/(\varepsilon^2|q_s|)\}$. Therefore we can apply (ii) to the statistic $\varepsilon^2 T_*$. Using (3.16) and $p \leq 1$, we get

$$\begin{aligned} Q(T_*; \lambda) &= Q(\varepsilon^2 T_*; \varepsilon^2\lambda) \\ &\ll_s \frac{\max\{\varepsilon^2\lambda; m_0\}}{p^3 M} \quad \text{with } m_0 = \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,3/2}^2}{\sigma^6}. \end{aligned} \tag{3.17}$$

The bound (3.17) implies (i) since by our choice of ε and δ we have $p = \exp\{-c_s\sigma/|q_s|\}$. Finally $c(s) \leq \exp\{c_1(s)\sigma/|q_s|\}$ since $\sigma/|q_s| \geq 1$. □

PROOF OF THEOREM 1.2. The result follows from (i) of Theorem 3.1 choosing $s = 9$, using $\gamma_3/\sigma^3 \geq 1$ as well as the inequality $\gamma_{2,3/2} \leq \gamma_3$. □

4. Proof of Theorem 1.1. The Edgeworth correction $F_1(x) = F_1(x; \mathscr{L}(X), \phi_1, \phi)$ is defined as a function of bounded variation [provided $|q_s| > 0$, for some $s \geq 9$; see (4.14) and Lemma 8.5] satisfying $F_1(-\infty) = 0$ and with the Fourier–Stieltjes transform given by

$$\hat{F}_1(t) = \frac{(it)^3}{6\sqrt{N}}\mathbf{E}(\phi_1(X) + \phi(X, G))^3 \mathrm{e}\{tT_0\}. \tag{4.1}$$

LEMMA 4.1. *Assume that the condition* $\mathscr{N}(p, (4s)^{-1}, s, G)$ *is fulfilled.*

(i) *Let* $s \geq 13$. *Then*

$$\begin{aligned} \Delta_N \ll_s &\frac{1}{N}\left(\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6} + \frac{1}{p^6}\right) \\ &\times(\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2\gamma_3 + \gamma_{2,2} + \sigma^2\gamma_{2,2}). \end{aligned} \tag{4.2}$$

(ii) *Assume that the condition* (1.9) *holds and that* $s \geq 9$. *Then*

$$\Delta_N \ll_s \frac{1}{N}\left(\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6} + \frac{1}{p^6}\right)(\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}). \tag{4.3}$$

PROOF OF THEOREM 1.1. We shall derive the result from Lemma 4.1. We shall prove (1.10) only. The proof of (1.11) is similar, provided we choose $s = 9$

and use (4.3) instead of (4.2). Write

$$\varepsilon^2 = 1/(16 s^2 \sigma) \quad \text{and} \qquad (4.4)$$
$$\Delta_N = \sup_x \left| F(x/\varepsilon^2) - F_0(x/\varepsilon^2) - F_1(x/\varepsilon^2) \right|.$$

Let us estimate Δ_N in (4.4) using Lemma 4.1. We have $F(x/\varepsilon^2) = \mathbf{P}\{T^{(\varepsilon)} \le x\}$, where the statistic $T^{(\varepsilon)}$ is defined similarly as T just replacing ϕ_1 and ϕ by $\varepsilon^2\phi_1$ and $\varepsilon^2\phi$, respectively. Similar representations are valid for $F_0(x/\varepsilon^2)$ and $F_1(x/\varepsilon^2)$ as well. Notice that $\varepsilon^2\phi(x,y) = \phi(\varepsilon x, \varepsilon y)$. By Lemma 5.3, the condition $\mathscr{N}(p, (4s)^{-1}, s, \varepsilon G)$ holds with $p = \exp\{-c_s/(\varepsilon^2 |q_s|)\}$. Therefore, we can use (4.2). Choosing $s = 13$ and replacing moments and eigenvalues by those corresponding to the kernel $\varepsilon^2\phi$ and $\varepsilon^2\phi_1$, we obtain

$$\Delta_N \ll \frac{1}{N}\left(\frac{\sigma^{13}}{|q_{13}|^{13}} + \frac{\sigma^6}{|q_{13}|^6} + \exp\left\{ \frac{c\sigma}{|q_{13}|} \right\} \right) \qquad (4.5)$$
$$\times \left(\frac{\beta_4}{\sigma^4} + \frac{\beta_3^2}{\sigma^6} + 1 + \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,2}}{\sigma^4} \right).$$

The bound (4.5) implies (1.10) since $1 \le \gamma_3/\sigma^3$ and $\sigma^{13}/|q_{13}|^{13} \le \exp\{c\sigma/|q_{13}|\}$, in view of $\sigma/|q_{13}| \ge 1$. □

Write $l = [(N-2)/20]$ and introduce the following bound

$$\varkappa(t) = \varkappa(t, N, \phi, \mathscr{L}(X)) = \varkappa_1(t) + \varkappa_2(t) \qquad (4.6)$$

of characteristic functions, where

$$\varkappa_1(t) = \sup_L \left| \mathbf{E}\mathrm{e}\left\{ tN^{-1} \sum_{1 \le j < k \le l} \phi(X_j, X_k) + L(X_1, \dots, X_l) \right\} \right|, \qquad (4.7)$$
$$\varkappa_2(t) = \sup_L \left| \mathbf{E}\mathrm{e}\left\{ tN^{-1} \sum_{1 \le j < k \le l} \phi(G_j, G_k) + L(G_1, \dots, G_l) \right\} \right|,$$

where supremum is taken over all linear statistics L, that is, over all functions L which can be represented as $L(x_1, \dots, x_l) = \sum_{j=1}^{l} f_j(x_j)$ with some functions $f_1, \dots, f_l$.

For $r \in \mathbb{Z}_+$ and functions f_i, introduce the statistic

$$T^{(r)} = \frac{1}{N} \sum_{1 \le i < j \le N} \phi(Z_i, Z_j) + \sum_{1 \le i \le N} f_i(Z_i), \qquad (4.8)$$

where

$$Z_j = X_j \quad \text{for } 1 \le j \le r \quad \text{and} \quad Z_j = G_j \quad \text{for } r < j \le N.$$

Notice that any of statistics $T, T(G_1, \dots, G_N)$ can be represented as $T^{(r)}$ in (4.8), with some r and f_i. For the statistic T_0, we have $\mathscr{L}(T_0) = \mathscr{L}(T^{(0)})$ with some $T^{(0)}$ [to see this combine (2.9) and (2.12)].

LEMMA 4.2. *Let $m \in \mathbb{N}$, $s \geq 9$ and $t_0 = m^{-1}(pN/m)^{-1+2/s}$. Assume that the random vector Y defined by* (3.3) *satisfies the condition $\mathscr{N}(p, (2s)^{-1}, s, Y)$. Furthermore, assume that $\mathscr{N}(p, (4s)^{-1}, s, G)$ holds. Then, for $pN > m$ and $m^{-1} \geq t_* \geq t_0$, the distribution function $F^{(r)}$ of $T^{(r)}$ satisfies*

$$F^{(r)}(x) = \frac{1}{2} + \frac{i}{2\pi}\text{V.P.}\int_{-Nt_*}^{Nt_*} \mathrm{e}\{-xt\}\hat{F}^{(r)}(t)\frac{dt}{t} + R, \tag{4.9}$$

where $|R| \ll_s m/(pN)$.

PROOF. We shall use the following approximate formula for the Fourier–Stieltjes inversion. For any $B > 0$ and any distribution function F with characteristic function $\hat{F}$, we have [see, for example, BG (1997b), Section 4]

$$F(x) = \frac{1}{2} + \frac{i}{2\pi}\text{V.P.}\int_{-B}^{B} \mathrm{e}\{-xt\}\hat{F}(t)\frac{dt}{t} + R, \tag{4.10}$$

where

$$|R| \leq \frac{1}{B}\int_{-B}^{B}\left|\hat{F}(t)\right| dt.$$

Here V.P. $\int f(t)\,dt = \lim_{\varepsilon \to 0} \int_{|t| > \varepsilon} f(t)\,dt$ denotes Cauchy's principal value of the integral.

First let us consider the case $r \geq N/2$. Choosing $B = N/m$ and applying the approximate Fourier inversion formula (4.10), we have

$$F^{(r)}(x) = \frac{1}{2} + \frac{i}{2\pi}\text{V.P.}\int_{-Nt_*}^{Nt_*} \mathrm{e}\{-xt\}\hat{F}^{(r)}(t)\frac{dt}{t} + J_1 + J_2 \tag{4.11}$$

with

$$J_1 = \frac{m}{N}\int_{-N/m}^{N/m}\left|\hat{F}^{(r)}(t)\right| dt, \qquad J_2 = \int_{Nt_0 \leq |t| \leq N/m}\left|\hat{F}^{(r)}(t)\right|\frac{dt}{|t|}.$$

Let us show that

$$J_1 \ll_s m/(pN) \quad \text{and} \quad J_2 \ll_s m/(pN). \tag{4.12}$$

Conditioning on $Z_{r+1}, \ldots, Z_N$, changing variables in the integral and rearranging summands in (4.8), we have

$$J_1 \leq m\mathbf{E}\int_{-1/m}^{1/m}\left|\mathbf{E}_{X_1,\ldots,X_r}\,\mathrm{e}\{tT_*^{(r)}\}\right| dt \tag{4.13}$$

with

$$T_*^{(r)} = \sum_{1 \leq i < j \leq r}\phi(X_i, X_j) + \sum_{1 \leq i \leq r} f_i^*(X_i),$$

with some f_j^* which can depend on Z_j, $j > r$. The statistic $T_*^{(r)}$ is a statistic type T_* [see (1.13)] with N replaced by $r \geq N/2$. For $T_*^{(r)}$ the number M in the definition (1.13) of T_* may be chosen arbitrary such that $M \leq N/2$. Let us take $M \asymp N/4$. Now an application of (3.13) to the right-hand side of (4.13) shows that J_1 satisfies (4.12).

Arguments similar to those for J_1 show that J_2 satisfies (4.12); however, instead of (3.13) one has to apply an analog of the second inequality in (3.8). Combining (4.11) and (4.12), we conclude the proof of the lemma in the case $r \ge N/2$.

In the case $r < N/2$, the statistic $T^{(r)}$ depends on at least $N/2$ Gaussian random vectors G_j. Therefore, we can repeat the proof for the case $r \ge N/2$, replacing X_j by G_j. Instead of $\mathscr{N}(p/2, (2s)^{-1}, s, Y)$ we shall use the condition $\mathscr{N}(p, (2s)^{-1}, s, Y')$ with $Y' = (2m)^{-1/2}(\tilde{G}_1 + \cdots + \tilde{G}_m)$. But this condition holds since $Y' =_{\mathscr{D}} G$ and $\mathscr{N}(p, (4s)^{-1}, s, G)$ is fulfilled. □

PROOF OF LEMMA 4.1. We shall represent the Edgeworth correction by the following Fourier–Stieltjes inversion formula. For any function $F\colon \mathbb{R} \to \mathbb{R}$ of bounded variation such that $F(-\infty) = 0$ and $2F(x) = F(x+) + F(x-)$, for all $x \in \mathbb{R}$, we have [see, e.g., Chung (1974)]

$$F(x) = \frac{1}{2}F(\infty) + \frac{i}{2\pi}\lim_{M\to\infty} \text{V.P.}\int_{|t|\le M} \mathrm{e}\{-xt\}\hat{F}(t)\frac{dt}{t}. \tag{4.14}$$

The formula is well known for distribution functions. For functions of bounded variation, it extends by linearity.

Let $m \in \mathbb{N}$. Assuming in addition that the random vector Y defined by (3.3) satisfies the condition $\mathscr{N}(p, (2s)^{-1}, s, Y)$, we shall prove that

$$\begin{aligned}\Delta_N \ll_s \frac{m}{pN} + \alpha + \frac{1}{p^6 N} \\ \times\big(\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2\gamma_3 + \gamma_{2,2} + \sigma^2\gamma_{2,2}\big),\end{aligned} \tag{4.15}$$

provided that $s \ge 13$, where

$$\alpha = \frac{\beta_3^2 + \sigma^2\gamma_{2,2}}{N}\left(\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6}\right).$$

Furthermore, in the case that the condition (1.9) holds and $s \ge 9$, we shall prove

$$\Delta_N \ll_s \frac{m}{pN} + \alpha + \frac{1}{p^4 N}\big(\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}\big). \tag{4.16}$$

The bounds (4.15) and (4.16) yield (4.2) and (4.3), respectively. Indeed, using Lemma 5.4 and estimating $\gamma_{2,3/2}^2 \le \sigma^2\gamma_{2,2}$ we see that

$$\mathscr{N}\big(p, (4s)^{-1}, s, G\big) \quad \text{implies}\, \mathscr{N}\big(2^{-1}p, (2s)^{-1}, s, Y\big),$$

provided that

$$m \ge c_s p^{-2}\sigma^2\gamma_{2,2} + c_s p^{-1}\gamma_3. \tag{4.17}$$

Hence, choosing the minimal m in (4.17), we obtain (4.2) and (4.3).

In the proof of (4.15) and (4.16) we may assume that $pN > m$. Indeed, if $pN \le m$, then using the obvious bounds $|F| \le 1$, $|F_0| \le 1$ as well as (4.14) and the bound (8.109) of Lemma 8.5 for $|F_1|$, we have

$$\Delta_N \ll_s 1 + \frac{\left(\beta_3^2 + \sigma^2\gamma_{2,2}\right)^{1/2}}{N^{1/2}|q_s|^3} \le \frac{m}{pN} + \frac{\left(\beta_3^2 + \sigma^2\gamma_{2,2}\right)^{1/2}}{N^{1/2}|q_s|^3}\left(\frac{m}{pN}\right)^{1/2}$$

$$\ll \frac{m}{pN} + \frac{\beta_3^2 + \sigma^2\gamma_{2,2}}{N|q_s|^6},$$

and there is nothing to prove.

Choose $t_* = m^{-1}(pN/m)^{-1/2}$. Applying Lemma 4.2 to the distribution functions F and F_0, Lemma 8.5 with $\lambda = Nt_*$ and (4.14) to the Edgeworth correction F_1, we obtain

$$\text{(4.18)} \qquad \Delta_N \ll_s J + R, \qquad J =_{\text{def}} \int_{-Nt_*}^{Nt_*} \left|\hat{F}(t) - \hat{F}_0(t) - \hat{F}_1(t)\right| \frac{dt}{|t|},$$

where

$$|R| \ll_s \frac{m}{pN} + N^{-1/2}\left(\beta_3^2 + \sigma^2\gamma_{2,2}\right)^{1/2}|q_s|^{-s/2}(Nt_*)^{3-s/2}.$$

We have $Nt_* = (N/m)(pN/m)^{-1/2}$. Therefore $ab \ll a^2 + b^2$ combined with $pN \ge m$, $0 < p \le 1$ and $s \ge 9$ yields

$$\text{(4.19)} \qquad |R| \ll_s \frac{m}{pN} + \frac{\beta_3^2 + \sigma^2\gamma_{2,2}}{|q_s|^s}.$$

Below we shall prove that

$$\text{(4.20)} \qquad I_r =_{\text{def}} \int_{-Nt_*}^{Nt_*} |t|^r \varkappa(t) \frac{dt}{|t|} \ll_s p^{-r}, \qquad 1 \le r < s/2.$$

Using (4.20) and $p \le 1$, integrating the bounds for $\hat{\Delta}_N(t) = |\hat{F}(t) - \hat{F}_0(t) - \hat{F}_1(t)|$ of Lemma 8.1, we obtain

$$\text{(4.21)} \qquad J \ll_s \frac{1}{p^6 N}\left(\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2\gamma_3 + \gamma_{2,2} + \sigma^2\gamma_{2,2}\right),$$

provided that $s \ge 13$, and

$$\text{(4.22)} \qquad J \ll_s \frac{1}{p^4 N}\left(\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}\right)$$

if (1.9) is fulfilled and $s \ge 9$. Collecting the bounds (4.18)–(4.22), we obtain (4.15) and (4.16).

It remains to prove (4.20). To estimate $\varkappa(t)$, we shall apply Theorem 6.2. The statistic in the definition (4.7) of $\varkappa_1$ is a statistic of type (1.13) with N replaced by $l = [(N-2)/20]$. There M can be chosen arbitrarily such that $1 \le M \le l$; for example, let $M = [l/2]$. Thus, using (2.14), we have $\varkappa_1(t) \ll_s$

$\min\{1; \mathscr{M}^s(tm/N; pN/m)\}$. A similar bound holds for $\varkappa_2(t)$. Therefore, applying (2.13), we obtain

$$\begin{aligned} I_r &\ll_s \int_{-Nt_*}^{Nt_*} |t|^r \min\{1; \mathscr{M}^s(tm/N; pN/m)\} \frac{dt}{|t|} \\ &= 2N^r \int_0^{t_*} t^r \min\{1; \mathscr{M}^s(tm; pN/m)\} \frac{dt}{t} \\ &\ll N^r \int_0^{1/(pN)} t^r \frac{dt}{t} + N^r \int_{1/(pN)}^{t_*} t^r \frac{1}{(tpN)^{s/2}} \frac{dt}{t} \ll p^{-r}. \qquad \square \end{aligned}$$

5. Auxiliary results.

Estimates of the convergence rate for expectations of smooth functions. Recall that X, X_j and G, G_j, $j \geq 1$, denote sequences of i.i.d. mean zero random vectors taking values in $\mathbb{R}^\infty$. We assume that X and G have equal covariances, that is, that (2.7) holds and that G is Gaussian. Using the linearization (2.6), we can assume that $\phi(x, y)$ is linear in its arguments $x, y \in \mathbb{R}^\infty$. For $S_m = (X_1 + \cdots + X_m)/\sqrt{m}$, we may write

$$\phi(S_m, \overline{S}_m) = m^{-1} \sum_{j=1}^{m} \sum_{k=1}^{m} \phi(X_j, \overline{X}_k). \tag{5.1}$$

LEMMA 5.1. *For any three times continuously differentiable function H: $\mathbb{R} \to \mathbb{R}$, we have*

$$\begin{aligned} &\left|\mathbf{E} H(\phi(S_m, \overline{S}_m)) - \mathbf{E} H(\phi(G, \overline{G}))\right| \\ &\qquad \ll m^{-1/2} |H'''|_\infty \left(\gamma_{2,3/2} + m^{-1/2}\gamma_3\right), \end{aligned} \tag{5.2}$$

where $|H|_\infty = \sup_{x \in \mathbb{R}} |H(x)|$.

PROOF. Notice that

$$\left|\mathbf{E} H(\phi(S_m, \overline{S}_m)) - \mathbf{E} H(\phi(G, \overline{G}))\right| \leq \mathbf{E} J_1 + \mathbf{E} J_2,$$

where

$$\begin{aligned} J_1 &= \left|\mathbf{E}_{S_m} H(\phi(S_m, \overline{S}_m)) - \mathbf{E}_G H(\phi(G, \overline{S}_m))\right|, \\ J_2 &= \left|\mathbf{E}_{\overline{S}_m} H(\phi(G, \overline{S}_m)) - \mathbf{E}_{\overline{G}} H(\phi(G, \overline{G}))\right|. \end{aligned}$$

In order to prove the lemma, it is sufficient to show that both $\mathbf{E} J_1$ and $\mathbf{E} J_2$ are bounded by the right-hand side of (5.2). For example, we shall bound $\mathbf{E} J_1$. Let us fix the value $z = \overline{S}_m$. Split

$$\begin{aligned} \phi(S_m, \overline{S}_m) &= \phi(S_m, z) = \xi_1 + \cdots + \xi_m \quad \text{where } \xi_j = \phi(X_j, z)/\sqrt{m} \\ \phi(G, \overline{S}_m) &= \phi(G, z) =_{\mathscr{D}} \zeta_1 + \cdots + \zeta_m \quad \text{where } \zeta_j = \phi(G_j, z)/\sqrt{m}. \end{aligned}$$

Notice that random variables $\xi_1, \dots, \xi_m$ (resp. $\zeta_1, \dots, \zeta_m$) are i.i.d. Writing

$$\delta_j = \left|\mathbf{E}H(W_j + \xi_j) - \mathbf{E}H(W_j + \zeta_j)\right|, \qquad W_j = \zeta_1 + \cdots + \zeta_{j-1} + \xi_{j+1} + \cdots + \xi_m,$$

we obtain

$$J_1 \le \delta_1 + \cdots + \delta_m. \tag{5.3}$$

Expanding in powers of ξ_j and of ζ_j and using (5.1), we get

$$\delta_j \le c\left(\mathbf{E}|\xi_j|^3 + \mathbf{E}|\zeta_j|^3\right)|H'''|_\infty. \tag{5.4}$$

The random variable $\zeta_j = \phi(G_j, z)/\sqrt{m}$ is centered and Gaussian. Therefore,

$$\mathbf{E}|\zeta_j|^3 \ll \left(\mathbf{E}\zeta_j^2\right)^{3/2} = \left(\mathbf{E}\xi_j^2\right)^{3/2} \le \mathbf{E}|\xi_j|^3. \tag{5.5}$$

Collecting (5.3)–(5.5), we obtain $J_1 \ll m|H'''|_\infty \mathbf{E}|\xi_1|^3$, whence, using $z = \overline{S}_m$, we have

$$\mathbf{E}J_1 \ll m^{-1/2}|\mathbf{H}'''|_\infty\, \mathbf{E}\left|\phi\left(X, \overline{S}_m\right)\right|^3. \tag{5.6}$$

Consequently, $\mathbf{E}J_1$ is bounded by the right-hand side of (5.2) since Rosenthal's inequality implies

$$\begin{aligned}\mathbf{E}\left|\phi\left(x, \overline{S}_m\right)\right|^3 &= m^{-3/2}\mathbf{E}\left|\sum_{j=1}^m \phi\left(x, \overline{X}_j\right)\right|^3 \\ &\ll m^{-1/2}\mathbf{E}\left|\phi(x, \overline{X})\right|^3 + \left(\mathbf{E}\phi^2(x, \overline{X})\right)^{3/2},\end{aligned}$$

for any $x \in \mathbb{R}^k$. □

Let $Y_1, \dots, Y_s$ and $\overline{Y}_1, \dots, \overline{Y}_s$ be independent copies of S_m. For a sufficiently smooth function $H\colon \mathbb{R} \to \mathbb{R}$, we write

$$D = \left|\mathbf{E}\prod_{1\le i,j\le s} H\left(\phi\left(Y_i, \overline{Y}_j\right) - \delta_{ij}\right) - \mathbf{E}\prod_{1\le i,j\le s} H\left(\phi\left(G_i, \overline{G}_j\right) - \delta_{ij}\right)\right|, \tag{5.7}$$

where δ_{ij} is the Kronecker symbol.

Lemma 5.2. *For D defined by* (5.7) *we have*

$$D \le C\left(m^{-1/2}\gamma_{2,3/2} + m^{-1}\gamma_3\right) \tag{5.8}$$

with

$$C \ll_s |H|_\infty^{s^2-3}\left(|H|_\infty^2|H'''|_\infty + |H|_\infty|H'|_\infty|H''|_\infty + |H'|_\infty^3\right).$$

Proof. The proof consists of 2s steps. In each step we shall replace one of the random vectors Y_j (or $\overline{Y}_j$) by a corresponding Gaussian random vector G_j (respectively, $\overline{G}_j$), for $1 \le j \le s$. We shall show that the errors of these replacements are bounded from above by the right-hand side of (5.8), thus proving the lemma.

For example, let us replace Y_1 by G_1 in

$$\mathbf{E} \prod_{1 \le i, j \le s} H\left(\phi\left(Y_i, \overline{Y}_j\right) - \delta_{ij}\right). \tag{5.9}$$

Similarly as in the proof of Lemma 5.1, fix

$$y_i = Y_i,\ \bar{y}_j = \overline{Y}_j \quad \text{and} \quad z_i = G_i,\ \bar{z}_j = \overline{G}_j \quad \text{for } 2 \le i \le s,\ 1 \le j \le s. \tag{5.10}$$

Conditioning on the random vectors in (5.10), we see that it is sufficient to bound the difference

$$I = \left| \mathbf{E} \prod_{1 \le j \le s} H\left(\phi\left(Y_1, \bar{y}_j\right) - \delta_{ij}\right) - \mathbf{E} \prod_{1 \le j \le s} H\left(\phi\left(G_1, \bar{y}_j\right) - \delta_{1j}\right) \right|. \tag{5.11}$$

For $x = (x_1, \dots, x_s) \in \mathbb{R}^s$, define the function

$$\psi(x) = H(x_1 - 1) H(x_2) \cdots H(x_s). \tag{5.12}$$

This function is Frechet differentiable and

$$\|\psi'''\|_\infty \ll_s |H|_\infty^{s-3} \left(|H|_\infty^2 |H'''|_\infty + |H|_\infty |H'|_\infty |H''|_\infty + |H'|_\infty^3 \right), \tag{5.13}$$

where $\|\psi'''\|_\infty = \sup_x \|\psi'''(x)\|$, and $\|\psi'''(x)\| = \sup\{|\psi'''(x)h^3|\colon h \in \mathbb{R}^s,\ |h| = 1\}$ denotes the norm of the Frechet derivative $\psi'''(x)$.

Introduce the random vectors

$$\mathbf{Y} = \left(\phi(Y_1, \bar{y}_1), \dots, \phi(Y_1, \bar{y}_s)\right), \qquad \mathbf{G} = \left(\phi(G_1, \bar{y}_1), \dots, \phi(G_1, \bar{y}_s)\right), \tag{5.14}$$

which take values in $\mathbb{R}^s$. Using the notation (5.12) and (5.14), we can rewrite I defined by (5.11) as $I = |\mathbf{E}\psi(\mathbf{Y}) - \mathbf{E}\psi(\mathbf{G})|$. Recall that $Y_1 =_{\mathscr{D}} (X_1 + \cdots + X_m)/\sqrt{m}$. A similar representation is valid for G_1. Thus

$$\mathbf{Y} =_{\mathscr{D}} \xi_1 + \cdots + \xi_m \quad \text{where } \xi_j = \left(\phi\left(X_j, \bar{y}_l\right)/\sqrt{m}\right)_{1 \le l \le s},$$

$$\mathbf{G} =_{\mathscr{D}} \zeta_1 + \cdots + \zeta_m \quad \text{where } \zeta_j = \left(\phi\left(G_j, \bar{y}_l\right)/\sqrt{m}\right)_{1 \le l \le s}.$$

Notice that the random vectors $\xi_1, \dots, \xi_m$ (resp. $\zeta_1, \dots, \zeta_m$) take values in $\mathbb{R}^s$ and are i.i.d. Moreover, the random vectors ζ_j are Gaussian, have mean zero and their covariances are equal to those of ξ_j. Writing

$$\delta_j = \left| \mathbf{E}\psi(\mathbf{W}_j + \xi_j) - \mathbf{E}\psi(\mathbf{W}_j + \zeta_j) \right|,$$

$$\mathbf{W}_j = \zeta_1 + \cdots + \zeta_{j-1} + \xi_{j+1} + \cdots + \xi_m,$$

we obtain

$$\mathbf{I} = \left| \mathbf{E}\psi(\mathbf{Y}) - \mathbf{E}\psi(\mathbf{G}) \right| \le \delta_1 + \cdots + \delta_m.$$

Expanding into a Taylor series in powers of ξ_j and ζ_j and using the equality of the covariances of ξ_j and ζ_j, we get

$$\delta_j \ll \left(\mathbf{E}|\xi_j|^3 + \mathbf{E}|\zeta_j|^3 \right) \|\psi'''\|_\infty.$$

Here

$$|\xi_j|^3 = m^{-3/2}\left(\sum_{l=1}^{s} \phi^2\big(X_j, \bar{y}_l\big)\right)^{3/2} \ll_s m^{-3/2} \sum_{l=1}^{s} \left|\phi\big(X_j, \bar{y}_l\big)\right|^3,$$

and $\|\psi'''\|_\infty$ was estimated in (5.13). A similar bound holds for $|\zeta_j|^3$. Averaging and proceeding similarly as in the proof of Lemma 5.1, we see that Y_1 in (5.9) can be replaced by G_1 with an error bounded from above by the right-hand side of (5.8). □

The condition $\mathscr{N}(p, \delta, s, X)$. Recall [see (3.1) and (3.2)] that the random $s \times s$ matrix $\mathbb{A}(X)$ has entries $a_{ij} = \phi(X_i, \overline{X}_j)$, where X_j and $\overline{X}_j$ denote independent copies of X. The matrix $\mathbb{A}(G)$ is defined similarly using independent copies of G. Furthermore, we write $\sigma^2 = q_1^2 + q_2^2 + \cdots$ and $|\mathbb{A}|_\infty = \max\{|a_{ij}|: 1 \le i, j \le s\}$.

LEMMA 5.3. *Let $q_s \neq 0$, for some $s \ge 1$. Then, for any $\delta > 0$ and $\varepsilon > 0$, the condition $\mathscr{N}(p, \delta, s, \varepsilon G)$ holds with some $p > 0$; that is,*

$$\nu =_{\text{def}} \mathbf{P}\big\{|\mathbb{A}(\varepsilon G) - \mathbb{I}|_\infty < \delta\big\} \ge p > 0. \tag{5.15}$$

Moreover, there exists a positive constant c_s, such that the condition $\mathscr{N}(p, \delta, s, \varepsilon G)$ is fulfilled with $p = \exp\{-c_s/(\varepsilon^2|q_s|)\}$; that is,

$$\nu \ge \exp\big\{-c_s/\big(\varepsilon^2|q_s|\big)\big\} \tag{5.16}$$

provided that $\delta \le 1$ and $4s\sigma\varepsilon^2 \le \delta$.

PROOF. Notice that $\phi(x, y) = \sum_{j \ge 1} q_j x_j y_j$ [cf. (2.5) and (2.6)]. Therefore, in the proof of the lemma we can assume that $G = (0, \eta_1, \eta_2, \dots)$. Let us split the identity operator $\mathbb{J}: \mathbb{R}^\infty \to \mathbb{R}^\infty$ into the sum $\mathbb{J} = \mathbb{P}_s + \mathbb{P}_s^\perp$ of the projector $\mathbb{P}_s x = (0, x_1, \dots, x_s, 0, 0, \dots)$, $x = (x_0, x_1, \dots)$, and its orthogonal complement $\mathbb{P}_s^\perp$. Then $x = \mathbb{P}_s x + \mathbb{P}_s^\perp x$ and we have $\phi(\mathbb{P}_s x, \mathbb{P}_s^\perp x) = \langle \mathbb{Q}\mathbb{P}_s x, \mathbb{P}_s^\perp x\rangle \equiv 0$, for all $x \in \mathbb{R}^\infty$. Thus the Gaussian vectors $\mathbb{P}_s \varepsilon G$ and $\mathbb{P}_s^\perp \varepsilon G$ are orthogonal and independent since εG has independent coordinates. Therefore we can write

$$\mathbb{A}(\varepsilon G) = \mathbb{A}(\mathbb{P}_s \varepsilon G) + \mathbb{A}\big(\mathbb{P}_s^\perp \varepsilon G\big) =_{\text{def}} \mathbb{U} + \mathbb{V} \tag{5.17}$$

with independent random matrices $\mathbb{U} = \mathbb{A}(\mathbb{P}_s \varepsilon G)$ and $\mathbb{V} = \mathbb{A}(\mathbb{P}_s^\perp \varepsilon G)$.

Let v_{ij} denote entries of the matrix $\mathbb{V}$. The random variables v_{ij} are identically distributed and we have

$$\mathbf{P}\{|\mathbb{V}|_\infty \ge b\} \le \frac{1}{b^2}\mathbf{E} \max_{1 \le i, j \le s} v_{ij}^2 \le \frac{s^2}{b^2}\mathbf{E} v_1^2 \le \frac{s^2\varepsilon^4\sigma^2}{b^2}, \qquad b > 0. \tag{5.18}$$

Let us prove (5.15). Replacing q_j by $\varepsilon^{1/2} q_j$, for all j, we can assume that $\varepsilon = 1$. Choose $b = 2s^2\sigma$. Then (5.18) yields $\mathbf{P}\{|\mathbb{V}|_\infty < b\} \ge 1/2$. Using (5.17), the independence of $\mathbb{U}$ and $\mathbb{V}$, and conditioning on $\mathbb{U}$, we have

$$\begin{aligned}\nu &\geq \mathbf{E}\mathbf{1}\{|\mathbb{V}|_\infty < b\}\mathbf{E}_{\mathbb{U}}\mathbf{1}\{|\mathbb{U} + \mathbb{V} - \mathbb{I}|_\infty < \delta\}\\ &\geq \tfrac{1}{2}\inf\{\mathbf{P}\{|\mathbb{U} - \mathbb{C}|_\infty < \delta\} : |\mathbb{C}|_\infty \leq b + 1\}.\end{aligned} \tag{5.19}$$

Introduce the standard basis vectors $e_j = (\delta_{0j}, \delta_{1j}, \dots)$ in $\mathbb{R}^\infty$, where δ_{ij} is the Kronecker symbol. Let $\lambda > 0$ be arbitrary. Write $\rho = \delta/(2|q_1|\lambda)$ and consider the events

$$A = \left\{|\mathbb{P}_s \overline{G}_j - e_j| \leq \rho \text{ for } 1 \leq j \leq s\right\}, \qquad B = \left\{|\mathbb{P}_s G_i| \leq \lambda \text{ for } 1 \leq i \leq s\right\}.$$

The events A and B are independent. Furthermore,

$$\mathbf{P}\{A\} = \prod_{j=1}^{s} \mathbf{P}\left\{|\mathbb{P}_s \overline{G}_j - e_j| \leq \rho\right\} > 0 \quad \text{for any } \rho > 0.$$

Notice that the event $A \cap B$ implies that $|u_{ij} - \langle \mathbb{Q}\mathbb{P}_s G_i, e_j\rangle| \leq |q_1|\,\rho\lambda \leq \delta/2$ since $u_{ij} = \langle \mathbb{Q}\mathbb{P}_s G_i, P_s \overline{G}_j\rangle$ and $\rho = \delta/(2|q_1|\lambda)$. Therefore, denoting by c_{ij} the entries of the matrix $\mathbb{C}$, we obtain

$$\begin{aligned}&\mathbf{P}\{|\mathbb{U} - \mathbb{C}|_\infty < \delta\}\\ &\quad\geq \mathbf{P}\{A \cap B \cap \{|\mathbb{U} - \mathbb{C}|_\infty < \delta\}\}\\ &\quad\geq \mathbf{P}\left\{A \cap B \cap \left\{\max_{1 \leq i, j \leq s} \left|\langle \mathbb{Q}\mathbb{P}_s G_i, e_j\rangle - c_{ij}\right| < \delta/2\right\}\right\}.\end{aligned} \tag{5.20}$$

The random vectors G, G_i and $\overline{G}_j$ are i.i.d. Thus (5.19) and (5.20) together imply

$$\nu \geq \frac{1}{2}\mathbf{P}\{A\} \inf_{|c| \leq b+1} \left(\mathbf{P}\left\{|\mathbb{P}_s G| \leq \lambda \text{ and } \max_{1 \leq j \leq s} \left|\langle \mathbb{Q}\mathbb{P}_s G, e_j\rangle - c\right| < \frac{\delta}{2}\right\}\right)^s. \tag{5.21}$$

The Gaussian random vector G has independent standard normal coordinates and the eigenvalues q_j of $\mathbb{Q}$ are nonzero, for $1 \leq j \leq s$. Therefore we can choose a sufficiently large λ depending on s, $\mathbb{Q}$ and δ such that the right-hand side of (5.21) is positive, completing the proof of (5.15).

Let us prove (5.16). The condition $4s\varepsilon^2\sigma \leq \delta$ and the estimate (5.18) with $b = \delta/2$ yield $\mathbf{P}\{|\mathbb{V}|_\infty \leq \delta/2\} \geq 1/2$. Therefore, applying (5.17) and $|\mathbb{A} + \mathbb{B}|_\infty \leq |\mathbb{A}|_\infty + |\mathbb{B}|_\infty$, using the independence of $\mathbb{U}$ and $\mathbb{V}$, we have

$$\nu \geq \mathbf{P}\left\{|\mathbb{V}|_\infty < \frac{\delta}{2}\right\}\nu_1 \geq \frac{1}{2}\nu_1 \quad \text{where } \nu_1 = \mathbf{P}\left\{|\mathbb{U} - \mathbb{I}|_\infty < \frac{\delta}{2}\right\}. \tag{5.22}$$

Let henceforth Z_i, $\overline{Z}_j$ denote independent copies of the standard normal random vector $Z = (\eta_1, \dots, \eta_s)$ in $\mathbb{R}^s$. Then we can represent the entries u_{ij} of the matrix $\mathbb{U}$ as

$$u_{ij} =_{\mathscr{D}} \varepsilon^2 \langle \mathbb{Q}Z_j, \overline{Z}_k\rangle, \qquad \mathbb{Q}x = (q_1 x_1, \dots, q_s x_s) \in \mathbb{R}^s.$$

Let $e_1, \dots, e_s$ be the standard orthonormal basis in $\mathbb{R}^s$ and $B = \{x \in \mathbb{R}^s\colon |x| \le 1\}$ denote the unit ball. For $\rho > 0$, consider the subsets

$$D_i = \frac{1}{\varepsilon\sqrt{|q_i|}}(e_i + \rho B), \qquad \overline{D}_j = \frac{\operatorname{sign} q_j}{\varepsilon\sqrt{|q_j|}}(e_j + \rho B), \qquad 1 \le i, j \le s,$$

of $\mathbb{R}^s$. It is easy to verify that

$$\text{(5.23)} \qquad \begin{aligned} &\varepsilon^2 \langle \mathbb{Q}x, y\rangle = \delta_{ij} + r_{ij}, \\ &\qquad |r_{ij}| \le (2\sigma\rho + \sigma\rho^2)/|q_s| \quad \text{for } x \in D_i \text{ and } y \in \overline{D}_j. \end{aligned}$$

where δ_{ij} is the Kronecker symbol. Choose $\rho = \delta|q_s|/6\sigma$. Then $\rho \le 1$ since $\delta \le 1$. Thus (5.23) shows that the independent events

$$A_i = \{Z_i \in D_i\}, \qquad \overline{A}_j = \left\{\overline{Z}_j \in \overline{D}_j\right\}, \qquad 1 \le i, j \le s,$$

satisfy

$$\text{(5.24)} \qquad \nu_1 \ge \prod_{j=1}^{s} \mathbf{P}\{A_j\}\mathbf{P}\left\{\overline{A}_j\right\} = \prod_{j=1}^{s} \left(\mathbf{P}\{A_j\}\right)^2,$$

since $Z_i, \overline{Z}_j$ are i.i.d. and symmetric.

The random vector Z is centered and Gaussian. Let $e \in \mathbb{R}^s$. According to a result of Borell (1974), we have $\mathbf{P}\{Z \in ae + \rho B\} \ge \mathbf{P}\{Z \in be + \rho B\}$ provided that $|a| \le |b|$. Thus, using $\sigma \ge |q_j| \ge |q_s|$ and shrinking the radius of the balls, we get

$$\mathbf{P}\{A_j\} \ge \mathbf{P}\{Z \in B_s\}, \qquad B_s =_{\text{def}} \frac{e_s}{\varepsilon\sqrt{|q_s|}} + \frac{\rho B}{\varepsilon\sqrt{\sigma}}.$$

The distribution of Z has density $p(x) = (2\pi)^{-s/2} \exp\{-|x|^2/2\}$. Estimating $p(x) \ge \inf\{p(x)\colon x \in B_s\}$ and writing $t = \rho/(\varepsilon\sqrt{\sigma})$, we obtain

$$\text{(5.25)} \quad \mathbf{P}\{A_j\} \ge c_1(s) t^s \exp\left\{-\frac{1}{2}\left(\frac{1}{\varepsilon\sqrt{|q_s|}} + t\right)^2\right\} \ge c_1(s) t^s \exp\left\{-\frac{2}{\varepsilon^2|q_s|}\right\},$$

since the assumption $\delta \le 1$ and the definitions of t and ρ imply that $1/\varepsilon\sqrt{|q_s|} \ge t$. Using the assumptions $4s\varepsilon^2\sigma \le \delta$ and $\delta \le 1$, it is easy to verify that $t \ge \varepsilon^2 q_s/2$ and that $\varepsilon^2|q_s| \le 1$. Therefore the inequality $z \ge \exp\{-1/z\}$ with $z = \varepsilon^2|q_s| \le 1$ combined with (5.24) and (5.25) implies

$$\nu_1 \ge c_2(s)\exp\left\{-\frac{c_3(s)}{\varepsilon^2|q_s|}\right\},$$

whence (5.16) follows by (5.22) and since $\varepsilon^2|q_s| \le 1$. □

LEMMA 5.4. *Let $\delta > 0$. Then there exists a positive constant c_s such that the condition $\mathscr{N}(2p, \delta, s, G)$ implies condition $\mathscr{N}(p, 2\delta, s, S_m)$, provided that*

$$m \ge c_s\, \delta^{-3} p^{-1}\left(\delta^{-3} p^{-1} \gamma_{2,3/2}^2 + \gamma_3\right).$$

PROOF. It suffices to show that

$$2\mathbf{P}\{|\mathbb{A}(S_m) - \mathbb{I}|_\infty \le 2\delta\} \ge \mathbf{P}\{|\mathbb{A}(G) - \mathbb{I}|_\infty \le \delta\} =_{\text{def}} 2p, \tag{5.26}$$

for m specified in the condition of the lemma. We shall apply Lemma 5.2. Consider an infinitely differentiable function $H: \mathbb{R} \to \mathbb{R}$ with bounded derivatives such that

$$0 \le H \le 1, \qquad H(x) = 1 \quad \text{for } |x| \le 1, \qquad H(x) = 0 \quad \text{for } |x| \ge 2.$$

Let $Y_1, \dots, Y_s$ and $\overline{Y}_1, \dots, \overline{Y}_s$ denote independent copies of S_m. Applying Lemma 5.2, we have

$$\begin{aligned} 2\mathbf{P}\{|\mathbb{A}(S_m) - \mathbb{I}|_\infty \le 2\delta\} &\ge 2\mathbf{E} \prod_{1 \le i, j \le s} H\left(\frac{\phi(Y_i, \overline{Y}_j) - \delta_{ij}}{\delta} \right) \\ &\ge 2\mathbf{E} \prod_{1 \le i, j \le s} H\left(\frac{\phi(G_i, \overline{G}_j) - \delta_{ij}}{\delta} \right) \\ &\quad - c(s)\left(m^{-1/2}\gamma_{2,3/2} + m^{-1}\gamma_3\right)\delta^{-3} \\ &\ge 2\mathbf{P}\{|\mathbb{A}(G) - \mathbb{I}|_\infty \le \delta\} - 2p = 2p \end{aligned}$$

provided m satisfies the condition of the lemma. □

A symmetrization lemma.

LEMMA 5.5. *Assume that the set $[1, N] = \bigcup_{j=1}^4 \Omega_j$ is a union of disjoint subsets. Let $\theta_1, \dots, \theta_N$ be arbitrary real-valued random variables such that the families of random variables*

$$\{\theta_j, j \in \Omega_1\}, \{\theta_j, j \in \Omega_2\}, \{\theta_j, j \in \Omega_3\}, \{\theta_j, j \in \Omega_4\}$$

are independent. Let a_{jk}, $1 \le j, k \le N$, be arbitrary real numbers. Assume that a statistic T can be represented as

$$T = \sum_{1 \le j < k \le N} \theta_j \theta_k a_{jk} + f_1 + f_2,$$

where f_1 (resp. f_2) is a function of θ_j, $j \in \Omega_1 \cup \Omega_4$ (resp. of θ_j, $j \in [1, N] \setminus \Omega_1$). Denote by $\overline{\theta}_1, \dots, \overline{\theta}_N$ an independent copy of $\theta_1, \dots, \theta_N$. Then

$$2|\mathbf{E}e\{T\}|^2 \le \mathbf{E}e\{T'\} + \mathbf{E}e\{T''\},$$

where

$$T' = \sum_{j \in \Omega_1} \sum_{k \in \Omega_2} \tilde{\theta}_j \tilde{\theta}_k a_{jk}, \qquad T'' = \sum_{j \in \Omega_1} \sum_{k \in \Omega_3} \tilde{\theta}_j \tilde{\theta}_k a_{jk},$$

and $\tilde{\theta}_j = \theta_j - \overline{\theta}_j$ for $1 \le j \le N$.

PROOF. For a two-point set $A = \{j, k\} \subset [1, N]$ we shall denote $T_A = \theta_j \theta_k a_{jk}$. In this notation we can write

$$T = \sum_{|A|=2} T_A + f_1 + f_2.$$

Furthermore, excluding certain terms T_A from the sum and adding them to f_1 or f_1, we can represent T as follows:

$$T = R + R^* + f_1^* + f_2^*, \tag{5.27}$$

where

$$R = \sum_{A:\,|A\cap\Omega_1|=1,\,|A\cap\Omega_2|=1} T_A, \qquad R^* = \sum_{A:\,|A\cap\Omega_1|=1,\,|A\cap\Omega_3|=1} T_A,$$

and where f_1^* (resp. f_2^*) have the same structure as f_1 (resp. f_2), which means that f_1^* depends on θ_j, $j \in \Omega_1 \cup \Omega_4$ only resp. f_2^* is independent of θ_j, $j \in \Omega_1$. Indeed, if $A \subset \Omega_1 \cup \Omega_4$, then we can include T_A in f_1; if $A \subset \Omega_2 \cup \Omega_3 \cup \Omega_4$, then we can add T_A to f_2; thus only terms T_A such that $|A \cap \Omega_1| = 1$ and $|A \cap (\Omega_2 \cup \Omega_3)| = 1$ will remain; collecting them into sums R and R^*, we obtain (5.27).

Finally, for a subset $B \subset [1, N]$ and a statistic $f = f(\theta_j;\, j \in [1, N])$, introduce the symmetrization with respect to the random variables θ_k, $k \in B$, as

$$\mathscr{S}_B f = f\big(\theta_j;\, j \in [1, N]\big) - f\Big(\overline{\theta}_j, \theta_k;\, j \in B,\, k \in [1, N] \setminus B\Big).$$

Using (5.27), conditioning on θ_j, $j \notin \Omega_1$, writing $\mathbf{E}_r$ for the partial integration with respect to the distributions of θ_j and $\overline{\theta}_j$, $j \in \Omega_r$, and applying Hölder's inequality we have

$$\begin{aligned} |\mathbf{E}\mathrm{e}\{T\}|^2 &\le \mathbf{E}\left|\mathbf{E}_1\mathrm{e}\{R + R^* + f_1^*\}\right|^2 \\ &= \mathbf{E}e\big\{\mathscr{S}_{\Omega_1}R + \mathscr{S}_{\Omega_1}R^* + \mathscr{S}_{\Omega_1}f_1^*\big\} \\ &\le \mathbf{E}\Big|\mathbf{E}_2\mathbf{E}_3\mathrm{e}\big\{\mathscr{S}_{\Omega_1}R + \mathscr{S}_{\Omega_1}R^*\big\}\Big|. \end{aligned} \tag{5.28}$$

For given $\tilde{\theta}_j$, $j \in \Omega_1$, the random variables $\mathscr{S}_{\Omega_1}R$ and $\mathscr{S}_{\Omega_1}R^*$ are independent. Thus (5.28) together with $2ab \le a^2 + b^2$ implies

$$\begin{aligned} 2|\mathbf{E}\mathrm{e}\{T\}|^2 &\le 2\mathbf{E}\Big|\mathbf{E}_2\mathrm{e}\big\{\mathscr{S}_{\Omega_1}R\big\}\mathbf{E}_3\mathrm{e}\big\{\mathscr{S}_{\Omega_1}R^*\big\}\Big| \\ &\le \mathbf{E}\Big|\mathbf{E}_2\mathrm{e}\big\{\mathscr{S}_{\Omega_1}R\big\}\Big|^2 + \mathbf{E}\Big|\mathbf{E}_3\mathrm{e}\big\{\mathscr{S}_{\Omega_1}R^*\big\}\Big|^2 \\ &= \mathbf{E}\mathrm{e}\{T'\} + \mathbf{E}\mathrm{e}\{T''\}. \end{aligned}$$

□

6. Bounds for characteristic functions of *U*-statistics. In this section we shall consider the statistic T_* defined by (1.13). The main result of the section is the following

LEMMA 6.1. *Assume that the distribution of a Gaussian random vector G satisfies the nondegeneracy condition $\mathscr{N}(p, (4s)^{-1}, s, G)$. Write $m_0 = p^{-2}\gamma_{2,3/2}^2 + p^{-1}\gamma_3$. For any statistic T_* of type* (1.13) *we have*

$$|\mathbf{E}\mathrm{e}\{tT_*\}| \ll_s \mathscr{M}^s(tm_0; pM/m_0),$$

where the function $\mathscr{M}$ is defined by (2.13).

Lemma 6.1 is a corollary of the following Theorem 6.2, which is valid without any moment assumptions.

THEOREM 6.2. *Let* $m \in \mathbb{N}$. *Assume that the sum* $Y = (2m)^{-/2}(\tilde{X}_1 + \cdots + \tilde{X}_m)$ *satisfies the nondegeneracy condition* $\mathscr{N}(p, \delta, (2s)^{-1}, Y)$. *Then, for any statistic* T_* *of type* (1.13), *we have*

$$|\mathbf{E}e\{tT_*\}| \ll_s \mathscr{M}^s(tm; pM/m).$$

PROOF OF LEMMA 6.1. Using Lemma 5.4, we have that

$$\mathscr{N}\big(p, (4s)^{-1}, s, G\big) \quad \text{implies} \, \mathscr{N}\big(2^{-1}p, (2s)^{-1}, s, Y\big),$$

provided that

$$m \geq c_s m_0, \qquad m_0 = p^{-2}\gamma_{2,3/2}^2 + p^{-1}\gamma_3. \tag{6.1}$$

Thus, by Theorem 6.2, we obtain

$$|\mathbf{E}e\{tT_*\}| \ll_s \mathscr{M}^s(c_1(s)tm; c_2(s)pM/m). \tag{6.2}$$

Using (2.16) and choosing the minimal $m \asymp c_s m_0$ such that (6.1) holds, Lemma 6.1 follows from (6.2). □

To prove Theorem 6.2 we need the auxiliary Lemmas 6.3–6.6. Lemma 6.3 allows replacing of the characteristic function of T_* by the characteristic function of a statistic of discrete random variables. Lemmas 6.4 and 6.5 are necessary steps for an application of the double large sieve-type bound of Lemma 6.6. In Section 7 we extend the methods of this section for the proof of a multiplicative inequality.

Using randomization and conditioning techniques, we shall bound the characteristic function of T_* by the characteristic function of certain statistics of discrete random variables. To this end we need the following notation. Let $\varepsilon_1, \varepsilon_2, \ldots$ denote i.i.d. Rademacher random variables, $\mathbf{P}\{\varepsilon_j = -1\} = \mathbf{P}\{\varepsilon_j = 1\} = 1/2$. As usual we assume that all random variables and vectors are independent in aggregate, if the contrary is not clear from the context. In particular, the sequences introduced below are independent of $X_1, \ldots, X_N$. Let m be an arbitrary natural number. Define the random variables $\vartheta_1 = \varepsilon_1, \ldots, \vartheta_m = \varepsilon_1$. Similarly, the random variables ϑ_j, $j > m$, are defined blockwise, that is,

$$\vartheta_j = \varepsilon_l \quad \text{for } j \in I(l) =_{\text{def}} (ml - m, ml]. \tag{6.3}$$

Furthermore, introduce the sequence $\alpha_j = \vartheta_j/2 + 1/2$ of random variables taking values 0 and 1 with probability 1/2.

For natural numbers m and s, we introduce the nonnegative integers

$$L = [M/(2ms)], \qquad K = sL \quad \text{and} \quad K_0 = mK = smL \tag{6.4}$$

and the disjoint sets

$$\Omega_1 = [1, K_0], \Omega_2 = (K_0, 2K_0], \Omega_3 = (2K_0, 3K_0], \Omega_4 = (3K_0, N]. \tag{6.5}$$

LEMMA 6.3. *Assume that* m, $s \in \mathbb{N}$. *Then the statistic* T_* *defined by* (1.13) *satisfies*

$$|\mathbf{E}e\{tT_*\}| \leq \mathbf{E}\big|\mathbf{E}_\vartheta e\{tT^\vartheta\}\big|, \tag{6.6}$$

where $\mathbf{E}_\vartheta = \mathbf{E}_{\{\vartheta_j,\, 1\le j\le 3K_0\}}$,

$$T^\vartheta = \sum_{1\le j<k\le 3K_0} \vartheta_j\vartheta_k a_{jk} + F_1 + F_2, \qquad a_{jk} =_{\mathrm{def}} \tfrac{1}{4}\phi\Big(\tilde{X}_j, \tilde{X}_k\Big), \tag{6.7}$$

the number K_0 *is defined in* (6.4), *and where* F_1 (*resp.* F_2) *is a function of* ϑ_j, $j\in[1,K_0]$, (*resp. of* ϑ_j, $j\in(K_0,3K_0]$).

Notice, that in (6.7) both statistics F_1 and F_2 may depend on $X_1,\dots,X_N$ and $\overline{X}_1,\dots,\overline{X}_N$.

PROOF OF LEMMA 6.3. Define the disjoint sets $\Omega_1',\dots,\Omega_4'\subset[1,N]$ as follows:

$$\Omega_1' = [1, K_0], \qquad \Omega_2' = (M, M+K_0], \Omega_3' = (M+K_0, M+2K_0], \tag{6.8}$$

and $\Omega_4' = [1,N]\setminus(\Omega_1'\cup\Omega_2'\cup\Omega_3')$. Without loss of generality we shall assume that

$$\Omega_1' = \Omega_1, \qquad \Omega_2' = \Omega_2, \qquad \Omega_3' = \Omega_3, \qquad \Omega_4' = \Omega_4. \tag{6.9}$$

Indeed, we have to estimate $|\mathbf{E}e\{tT_*\}|$, and using the i.i.d. assumption we can reorder $X_1,\dots,X_N$ so that (6.9) holds.

Using (6.8), we can write

$$T_* = \sum_{\{j,k\}\subset\Omega_1\cup\Omega_2\cup\Omega_3} \phi\big(X_j, X_k\big) - f_1^* + f_2^*, \tag{6.10}$$

where f_1^* is a function of X_j with $j\in\Omega_1\cup\Omega_4$, and f_2^* is a function of X_j with $j\in\Omega_2\cup\Omega_3\cup\Omega_4$. Indeed, the terms $\phi(X_j,X_k)$ with $\{j,k\}\cap\Omega_4\ne\varnothing$ can be included in f_1^* or in f_2^*.

Introduce the statistic T^α replacing the random vectors X_j in (6.10) by the vectors $\alpha_jX_j + (1-\alpha_j)\overline{X}_j$, for $1\le j\le 3K_0$. The statistics T_* and T^α have the same distribution since, for given $\alpha_1,\alpha_2,\dots$, the conditional distribution of T^α equals the distribution of T_*. Indeed, if $\alpha_j=1$, then X_j remains unchanged; if $\alpha_j=0$, then X_j are just exchanged with an independent copy $\overline{X}_j$. In any case, this does not change the distribution of T_*. Therefore (6.10) implies

$$\begin{aligned} T_* &=_{\mathscr{D}} T^\alpha \\ &= \sum_{\{j,k\}\subset\Omega_1\cup\Omega_2\cup\Omega_3} \phi\Big(\alpha_jX_j + (1-\alpha_j)\overline{X}_j, \alpha_kX_k + (1-\alpha_k)\overline{X}_k\Big) \\ &\quad + f_1^{**} + f_2^{**}, \end{aligned} \tag{6.11}$$

where f_1^{**} is a function of α_j with $j\in\Omega_1\cup\Omega_4$, and f_2^{**} is a function of α_j with $j\in\Omega_2\cup\Omega_3\cup\Omega_4$. The equality $\alpha_j = 1/2+\vartheta_j/2$ implies

$$\alpha_jX_j + (1-\alpha_j)\overline{X}_j = \tfrac{1}{2}\vartheta_j\tilde{X}_j + \tfrac{1}{2}\Big(X_j+\overline{X}_j\Big), \qquad \tilde{X}_j = X_j - \overline{X}_j. \tag{6.12}$$

Applying (6.12) to (6.11), using the linearity of $\phi(x, y)$ and rearranging the terms of the statistic, we obtain that $T_* =_{\mathscr{D}} T^\vartheta$ with T^ϑ defined by (6.7). Conditioning on all random variables and vectors except ϑ_j, $1 \le j \le 3K_0$, concludes the proof of (6.6). $\square$

LEMMA 6.4. *Let* $m \in \mathbb{N}$. *Write*

$$Y = (2m)^{-1/2} \sum_{j=1}^{m} \tilde{X}_j, \qquad \Lambda = \sum_{j=1}^{K} \tilde{\varepsilon}_j Y_j, \tag{6.13}$$

where $Y_1, Y_2, \dots$ *are independent copies of* Y *and* K *is given by* (6.4). *Then we have*

$$|\mathbf{E}e\{tT_*\}|^2 \le \mathbf{E}e\{2^{-1}tm\phi(\Lambda, \overline{\Lambda})\}.$$

PROOF. We shall apply the symmetrization Lemma 5.5 to the characteristic function $\mathbf{E}_\vartheta e\{tT^\vartheta\}$ appearing in the estimate of Lemma 6.3. Consider the sets Ω_l, $1 \le l \le 3$, defined by (6.5). Notice that the families of random variables

$$\{\vartheta_j, j \in \Omega_1\}, \qquad \{\vartheta_j, j \in \Omega_2\}, \qquad \{\vartheta_j, j \in \Omega_3\}$$

are independent.

Using the estimate of Lemma 6.3, conditioning on $X_1, \dots, X_M$ and $\overline{X}_1, \dots, \overline{X}_M$, writing $\mathbf{E}_\vartheta = \mathbf{E}_{\vartheta_1, \vartheta_2, \dots}$ and applying the symmetrization Lemma 5.5, we have

$$\begin{aligned} 2|\mathbf{E}e\{tT_*\}|^2 &\le 2\mathbf{E}\big|\mathbf{E}_\vartheta e\{tT^\vartheta\}\big|^2 \le \mathbf{E}\mathbf{E}_\vartheta e\{tT'\} + \mathbf{E}\mathbf{E}_\vartheta e\{tT''\} \\ &= \mathbf{E}e\{tT'\} + \mathbf{E}e\{tT''\} = 2\mathbf{E}e\{tT'\}, \end{aligned} \tag{6.14}$$

since both statistics

$$T' = \sum_{j \in \Omega_1} \sum_{k \in \Omega_2} \tilde{\vartheta}_j \tilde{\vartheta}_k a_{jk} \quad \text{and} \quad T'' = \sum_{j \in \Omega_1} \sum_{k \in \Omega_3} \tilde{\vartheta}_j \tilde{\vartheta}_k a_{jk} \tag{6.15}$$

[with a_{jk} defined by (6.7)] have the same distribution. Finally, notice that $\mathscr{L}(T') = \mathscr{L}(m\phi(\Lambda, \overline{\Lambda})/2)$, due to the definition (6.3) of ϑ_j and the linearity of $\phi(x, y)$. Indeed, we have

$$T' = \tfrac{1}{4}\phi\bigg(\sum_{j \in \Omega_1} \tilde{\vartheta}_j \tilde{X}_j, \sum_{k \in \Omega_2} \tilde{\vartheta}_k \tilde{X}_k\bigg) \tag{6.16}$$

and

$$\sum_{j \in \Omega_1} \tilde{\vartheta}_j \tilde{X}_j = \sum_{j=1}^{mK} \tilde{\vartheta}_j \tilde{X}_j = \sum_{l=1}^{K} \tilde{\varepsilon}_l \sum_{j \in I(l)} \tilde{X}_j = \sqrt{2m}\,\Lambda. \qquad \square$$

In the proof of Lemma 6.5 we shall combine random selections and the geometric-arithmetic mean inequality. This will reduce the problem to the case of sums of i.i.d. random vectors of a very special type.

By $\tau, \tau_1, \tau_2, \dots$ and τ_{kl} we shall denote independent copies of a symmetric random variable τ with nonnegative characteristic function and such that

$$1 \le \mathbf{E}\tau^2 \le 2, \qquad \mathbf{P}\{|\tau| \le 2\} = 1. \tag{6.17}$$

An example of such τ is $\tilde{\varepsilon}$. Define the random s-dimensional vector $W = (\tau_1, \dots, \tau_s)$. Let $W_1, \dots, W_q$ and $\overline{W}_1, \dots, \overline{W}_q$ denote independent copies of W. Introduce the sums

$$U = W_1 + \cdots + W_q, \qquad V = \overline{W}_1 + \cdots + \overline{W}_q. \tag{6.18}$$

LEMMA 6.5. *Let $s \in \mathbb{N}$ and $L \in \mathbb{Z}_+$. Assume that a random vector Y takes values in $\mathbb{R}^\infty$ and satisfies the condition $\mathscr{N}(p, \delta, s, Y)$. Write*

$$\Lambda = \sum_{j=1}^{sL} \tau_j Y_j, \qquad \overline{\Lambda} = \sum_{j=1}^{sL} \bar{\tau}_j \overline{Y}_j, \qquad q = [pL/4], \tag{6.19}$$

where Y_j and $\overline{Y}_j$ are independent copies of Y. Then we have

$$\mathbf{E}e\{t\phi(\Lambda, \overline{\Lambda})\} \le c_d(s)(pL)^{-d} + \sup_{\mathbb{A}} \mathbf{E}e\{t\langle \mathbb{A}U, V\rangle\}, \qquad t \in \mathbb{R}, d \ge 0, \tag{6.20}$$

where $\sup_{\mathbb{A}}$ denotes the supremum over all $s \times s$ (nonrandom) matrices $\mathbb{A}$ such that $|\mathbb{A} - \mathbb{I}|_\infty \le \delta$.

PROOF. Introduce the sets $\mathscr{I}(j) = (js - s, js] \subset [1, sL]$ of natural numbers, for $1 \le j \le L$. Split the sum Λ into a sum of blocks of length s, that is, write

$$\Lambda = Z(1) + \cdots + Z(L) \quad \text{where } Z(j) = \sum_{k \in \mathscr{I}(j)} \tau_k Y_k. \tag{6.21}$$

Similarly, write

$$\overline{\Lambda} = \overline{Z}(1) + \cdots + \overline{Z}(L) \quad \text{where } \overline{Z}(j) = \sum_{k \in \mathscr{I}(j)} \bar{\tau}_k \overline{Y}_k. \tag{6.22}$$

Let us describe a random event, say D, which will play an essential role in our proof. For $1 \le j, k \le L$, define the $s \times s$-matrices

$$\mathbb{A}(j,k) = \big(\phi\big(Y_l, \overline{Y}_r\big)\big), \qquad l \in \mathscr{I}(j), r \in \mathscr{I}(k).$$

Introduce the random events and variables

$$B(j,k) = \{|\mathbb{A}(j,k) - \mathbb{I}|_\infty \le \delta\}, \qquad \xi_{jk} = \mathbf{1}\{B(j,k)\}. \tag{6.23}$$

Condition $\mathscr{N}(p, \delta, s, Y)$ guarantees that $\mathbf{E}\xi_{jk} = \mathbf{P}\{B(j,k)\} \ge p$. Without loss of generality we shall assume that $\mathbf{E}\xi_{jk} = \mathbf{P}\{B(j,k)\} = p$. Furthermore, the events $B(j,k)$ occur with equal probabilities. Write

$$\eta_j = \sum_{k=1}^{L} \xi_{jk}, \qquad D = \left\{\sum_{j=1}^{L} \eta_j \ge \frac{pL^2}{2}\right\}, \qquad D_j = \left\{\eta_j \ge \frac{pL}{4}\right\}.$$

We shall use the following properties of D and D_j: the complement D^c of D satisfies

$$\mathbf{P}\{D^c\} = \mathbf{P}\left\{\sum_{j=1}^{L} \sum_{k=1}^{L} \xi_{jk} < \frac{pL^2}{2}\right\} \ll_d (pL)^{-d} \quad \text{for } d \ge 0 \tag{6.24}$$

and

(6.25) D occurs implies that at least $\dfrac{pL}{4}$ of the events $D_1, \dots, D_L$ occur.

Let us prove (6.24). If $pL \le 1$ then (6.24) follows from $\mathbf{P}\{D^c\} \le 1$. Therefore, we suppose without loss of generality that $pL \ge 1$. Assume as well that $d \ge 1$ since the desired result for $0 \le d \le 1$ follows from (6.24) with $d = 1$. Introduce a random vector ψ such that $\mathscr{L}(\psi) = \mathscr{L}(Y_1, \dots, Y_s)$, and let ψ_j and $\overline{\psi}_k$ denote its independent copies. Then we can represent ξ_{jk} as a function, say u, of ψ_j and $\overline{\psi}_k$, that is $\xi_{jk} = u(\psi_j, \overline{\psi}_k)$. Define the function $\rho(x) = \mathbf{E}u(x, \psi)$. Then we can write

$$-pL^2 + \sum_{j=1}^{L} \eta_j = -pL^2 + \sum_{j=1}^{L}\sum_{k=1}^{L} \xi_{jk} =_{\text{def}} R + R_1 + \overline{R}_1, \tag{6.26}$$

where

$$R = \sum_{j=1}^{L}\sum_{k=1}^{L}\left(\xi_{jk} - \rho(\psi_j) - \rho(\overline{\psi}_k) + p\right), \qquad R_1 = L\sum_{j=1}^{L}\left(\rho(\psi_j) - p\right),$$

and where $\overline{R}_1$ is defined similarly to R_1 with ψ_j replaced by $\overline{\psi}_j$. Using this notation and Chebyshev's inequality, we have

$$\mathbf{P}\{D^c\} \ll_d (pL^2)^{-2d}\left(\mathbf{E}|R|^{2d} + \mathbf{E}|R_1|^{2d} + \mathbf{E}|\overline{R}_1|^{2d}\right). \tag{6.27}$$

To bound $\mathbf{E}|R_1|^{2d} = \mathbf{E}|\overline{R}_1|^{2d}$ we use the inequality (2.17). We obtain

$$\mathbf{E}|R_1|^{2d} \ll_d pL + (pL)^d \ll_d (pL)^d \quad \text{since } pL \ge 1,\ 0 \le \rho \le 1,\ \mathbf{E}\rho(\psi) = p.$$

Using (2.18) twice conditionally, we derive

$$\begin{aligned}
\mathbf{E}|R|^{2d} &= \mathbf{E}\mathbf{E}_{\psi}|R|^{2d}\\
&\ll_d L^d\mathbf{E}\mathbf{E}_{\psi}\left|\sum_{k=1}^{L}\left(\xi_{1k} - \rho(\psi_1) - \rho(\overline{\psi}_k) + p\right)\right|^{2d}\\
&= L^d\mathbf{E}\mathbf{E}_{\overline{\psi}}\left|\sum_{k=1}^{L}\left(\xi_{1k} - \rho(\psi_1) - \rho(\overline{\psi}_k) + p\right)\right|^{2d}\\
&\ll_d L^{2d}\mathbf{E}\left|\xi_{11} - \rho(\psi_1) - \rho(\overline{\psi}_1) + p\right|^{2d} \ll_d L^{2d}.
\end{aligned} \tag{6.28}$$

In the proof of (6.28) we used the i.i.d. assumption and the notation $\mathbf{E}_{\psi} = \mathbf{E}_{\psi_1, \dots, \psi_L}$. Collecting the bounds (6.27) and (6.28), we obtain (6.24).

Let us prove (6.25). Assume the contrary, that is, that D occurs and strictly less than $pL/4$ of events D_j occur. Then at least $L - pL/4$ of D_j^c occur, and we have

$$\frac{pL^2}{2} \le \left(L - \frac{pL}{4}\right)\frac{pL}{4} + \frac{pL}{4}L = \frac{pL^2}{2} - \frac{p^2L^2}{16},$$

which is impossible.

Let us return to the proof of (6.20). Define the random variable

$$J = \mathbf{E}_\tau \mathrm{e}\{t\phi(\Lambda, \overline{\Lambda})\}, \tag{6.29}$$

where $\mathbf{E}_\tau$ denotes the partial integration with respect to the distributions of the random variables which are independent copies of τ. Note that J is random since it depends on Y_j and $\overline{Y}_j$, $1 \le j \le sL$. Due to (6.24), the lemma follows provided that we can verify that

$$\mathbf{1}\{D\} J \le \sup_{\mathbb{A}} \mathbf{E}\mathrm{e}\{t\langle \mathbb{A}U, V\rangle\}. \tag{6.30}$$

So it remains to prove (6.30). If $\mathbf{1}\{D\} = 0$, then (6.30) is obviously fulfilled. Thus we can assume that $\mathbf{1}\{D\} = 1$ and hence [see (6.25)] that at least $pL/4$ of random variables $\eta_j = \sum_{k=1}^L \xi_{jk}$ satisfy $\eta_j \ge pL/4$. Thus there exist indices $j_1, \dots, j_q$ such that

$$\eta_{j_1} \ge pL/4, \dots, \eta_{j_q} \ge pL/4 \quad \text{with } q = [\,pL/4]. \tag{6.31}$$

Given the random vectors Y_k, let $S(j)$ denote the sum of q independent copies of $Z(j)$, that is,

$$S(j) = \sum_{l=1}^{q} \sum_{k \in \mathscr{I}(j)} \tau_{kl} Y_k, \tag{6.32}$$

where the random variables τ_{kl}, $1 \le l \le q$, $k \in \mathscr{I}(j)$, are i.id. Finally, let $\mathbf{E}_{\tau, Z(j)}$ (resp. $\mathbf{E}_{\tau, S(j)}$) denote the integration with respect to the distributions of the random variables τ_j (resp. τ_{kl}) occurring in the sum $Z(j)$ [resp. in $S(j)$].

Using the bilinearity of ϕ, conditioning, estimating

$$\mathbf{E}_{\tau, Z(j)} \mathrm{e}\{t\phi(Z(j), \overline{\Lambda})\} \le 1 \quad \text{for } j \notin \{j_1, \dots, j_q\},$$

and applying the geometric-arithmetic mean inequality, we have

$$\begin{aligned} J &= \mathbf{E}_\tau \prod_{j=1}^{L} \mathbf{E}_{\tau, Z(j)} \mathrm{e}\{t\phi(Z(j), \overline{\Lambda})\} \\ &\le \mathbf{E}_\tau \prod_{r=1}^{q} \mathbf{E}_{\tau, Z(j_r)} \mathrm{e}\{t\phi(Z(j_r), \overline{\Lambda})\} \\ &\le \mathbf{E}_\tau \frac{1}{q} \sum_{r=1}^{q} \left(\mathbf{E}_{\tau, Z(j_r)} \mathrm{e}\{t\phi(Z(j_r), \overline{\Lambda})\}\right)^q \\ &= \frac{1}{q} \sum_{r=1}^{q} \mathbf{E}_\tau \mathbf{E}_{\tau, S(j_r)} \mathrm{e}\{t\phi(S(j_r), \overline{\Lambda})\} \\ &= \frac{1}{q} \sum_{r=1}^{q} \mathbf{E}_\tau \prod_{k=1}^{L} \mathbf{E}_{\tau, \overline{Z}(k)} \mathrm{e}\{t\phi(S(j_r), \overline{Z}(k))\}. \end{aligned} \tag{6.33}$$

Recall that we assume that the event D occurs. Therefore (6.31) implies that $\eta_{j_r} \ge pL/4$, for any j_r. In other words, there exist (random) indices

$k_1(j_r),\dots,k_q(j_r)$ such that the events $B(j_r,k_l(j_r))$ occur [notice that this yields $|\mathbb{A}(j_r,k_l(j_r))-\mathbb{I}|_\infty\le\delta$, for all $1\le l\le q$]. Therefore, using (6.33), (6.18) and arguing similarly as in the proof of (6.33), we have

$$\begin{aligned} J &\le \frac{1}{q}\sum_{r=1}^{q}\mathbf{E}_\tau\prod_{l=1}^{q}\mathbf{E}_{\tau,\,\overline{Z}(k_l(j_r))}\mathrm{e}\big\{t\phi\big(S(j_r),\overline{Z}(k_l(j_r))\big)\big\}\\ &\le \frac{1}{q}\sum_{r=1}^{q}\mathbf{E}_\tau\frac{1}{q}\sum_{l=1}^{q}\Big(\mathbf{E}_{\tau,\,\overline{Z}(k_l(j_r))}\mathrm{e}\big\{t\phi\big(S(j_r),\overline{Z}(k_l(j_r))\big)\big\}\Big)^q\\ &= \frac{1}{q^2}\sum_{r=1}^{q}\sum_{l=1}^{q}\mathbf{E}_\tau\mathrm{e}\big\{t\phi\big(S(j_r),\overline{S}(k_l(j_r))\big)\big\} \qquad (6.34)\\ &= \frac{1}{q^2}\sum_{r=1}^{q}\sum_{l=1}^{q}\mathbf{E}_\tau\mathrm{e}\{t\langle\mathbb{A}(j_r,k_l(j_r))U,V\rangle\}\\ &\le \frac{1}{q^2}\sum_{r=1}^{q}\sum_{l=1}^{q}\sup_{\mathbb{A}}\mathbf{E}\mathrm{e}\{t\langle\mathbb{A}U,V\rangle\} = \sup_{\mathbb{A}}\mathbf{E}\mathrm{e}\{t\langle\mathbb{A}U,V\rangle\}, \end{aligned}$$

thus proving (6.30) and the lemma. □

LEMMA 6.6 [BG (1997b), Lemma 4.7]. *Let $\mathbb{A}$ be a $s\times s$ matrix. Let X denote a random vector taking values in $\mathbb{R}^s$ with covariance matrix $\mathbb{C}$. Assume that there exists a constant c_s such that*

$$\mathbf{P}\{|X|\le c_s\}=1,\ |\mathbb{A}|\le c_s,\ |\mathbb{A}^{-1}|\le c_s,\ |\mathbb{C}^{-1}|\le c_s. \qquad (6.35)$$

Let U and V denote independent random vectors which are sums of N independent copies of X. Then

$$\big|\mathbf{E}\mathrm{e}\{t\langle\mathbb{A}U,V\rangle\}\big|\ll_s \mathscr{M}^{2s}(t;N)\quad \text{for } t\in\mathbb{R}, \qquad (6.36)$$

where the function $\mathscr{M}$ is defined by (2.13).

LEMMA 6.7. *Let U and V denote independent random vectors in $\mathbb{R}^s$ which are sums of q independent copies of $W=(\tau_1,\dots,\tau_s)$ [see* (6.17) *and* (6.18) *for the definition]. Assume that an $s\times s$ (nonrandom) matrix $\mathbb{A}$ satisfies $|\mathbb{A}-\mathbb{I}|_\infty\le 1/(2s)$. Then*

$$\mathbf{E}\mathrm{e}\{t\langle\mathbb{A}U,V\rangle\}\ll_s \mathscr{M}_s^2(t;q)\quad \text{for } t\in\mathbb{R}. \qquad (6.37)$$

PROOF. The result follows by an application of Lemma 6.6 replacing N by q and X by W. We have to verify the condition (6.35). The covariance of W clearly satisfies (6.35). The matrix $\mathbb{A}$ satisfies (6.35) as well, since

$$2^{-1}|x|\le|\mathbb{A}x|\le 2|x|\quad \text{for all } x\in\mathbb{R}^s. \qquad (6.38)$$

For example, let us verify the left-hand side inequality in (6.38). Using $|\mathbb{A}-\mathbb{I}|_\infty\le 1/(2s)$, writing $\mathbb{A}=\mathbb{I}+\mathbb{B}$ with some $\mathbb{B}$ such that $|\mathbb{B}|_\infty\le 1/(2s)$ and introducing the Hilbert–Schmidt norm $|\mathbb{B}|_2\le 1/2$, we have

$$|\mathbb{A}x|\ge|x|-|\mathbb{B}|\,|x|\ge(1-|\mathbb{B}|_2)|x|\ge|x|/2. \qquad □$$

PROOF OF THEOREM 6.2. We shall apply Lemmas 6.4, 6.5 and 6.7.

Without loss of generality we can assume that $pM \geq c_s m$ with a sufficiently large constant c_s. Indeed, otherwise $pM \leq c_s m$ and the result follows from the trivial bound $|\mathbf{E}e\{tT_*\}| \leq 1$ combined with (2.15). The assumption $pM > c_s m$ implies that the natural numbers L, K, K_0/m introduced in (6.4) are well defined. Moreover,

$$L \asymp_s K \asymp_s K_0/m \asymp_s M/m \asymp_s q/p, \qquad q = [pL/4]. \tag{6.39}$$

Lemma 6.4 implies

$$|\mathbf{E}e\{tT_*\}|^2 \leq \mathbf{E}e\{\beta\phi(\Lambda, \overline{\Lambda})\}, \qquad \beta = tm/2. \tag{6.40}$$

Lemma 6.5 yields

$$\mathbf{E}e\{\beta\phi(\Lambda, \overline{\Lambda})\} \leq_{d,s} (pL)^{-d} + \sup_{\mathbb{A}} \mathbf{E}e\{\beta\langle \mathbb{A}U, V\rangle\}. \tag{6.41}$$

An application of Lemma 6.7 yields

$$\mathbf{E}e\{\beta\langle \mathbb{A}U, V\rangle\} \ll_s \mathscr{M}^{2s}(\beta; q). \tag{6.42}$$

Collecting the bounds (6.40)–(6.42), substituting $\beta = tm/2$, using (2.16) and (6.39), we obtain

$$|\mathbf{E}e\{tT_*\}| \ll_{d,s} (pM/m)^{-d/2} + \mathscr{M}^s(tm; pM/m). \tag{6.43}$$

By (2.15) we have $\inf_x \mathscr{M}^s(x; pM/m) = (pM/m)^{-s/4}$. Thus (6.43) implies the desired result provided that we choose $d = s/2$. $\square$

7. A multiplicative inequality for U-statistics. In this section we prove a multiplicative inequality for characteristic function of the statistic T^ϑ defined by (6.7). This inequality yields the desired bound $\mathscr{O}(N^{-1})$ for an integral of the characteristic function of a U-statistic. Write

$$\psi(t) = |\mathbf{E}_\vartheta e\{tT^\vartheta\}|. \tag{7.1}$$

Notice that $\psi(t)$ is random since it depends on $X_1, \dots, X_N$ and $\overline{X}_1, \dots, \overline{X}_N$.

LEMMA 7.1. *Let $d \geq 0$ and $m, s \in \mathbb{N}$. Assume that $Y = (2m)^{-1/2}\sum_{k=1}^m \tilde{X}_k$ satisfies the nondegeneracy condition $\mathscr{N}(p, (2s)^{-1}, s, Y)$. Then there exist constants $c_1(s,d)$ and $c_2(s,d)$ such that the event*

$$D = \{\psi(t-\gamma)\psi(t+\gamma) \leq c_1(s,d)\mathscr{M}^s(\gamma m; pM/m) \quad \text{for all } t, \gamma \in \mathbb{R}\} \tag{7.2}$$

satisfies

$$\mathbf{P}\{D\} \geq 1 - c_2(s,d)(pM/m)^{-d}. \tag{7.3}$$

We shall prove Lemma 7.1 a little bit later. Lemma 7.1 holds for "discrete" statistics T^ϑ only. Nevertheless it can be applied to estimate integrals of arbitrary U-statistics. As an example of such application we provide the following Corollary 7.2. In order to simplify notation we restrict ourselves to degenerate U-statistics without lower order terms and choose $m = 1$.

COROLLARY 7.2. *Let* $s \geq 9$, $s \in \mathbb{N}$, $0 < p \leq 1$ *and*

$$T_* = \sum_{1 \leq j < k \leq N} \phi(X_j, X_k). \tag{7.4}$$

Assume that ϕ *and* $Y = \tilde{X}_1/\sqrt{2}$ *satisfy the nondegeneracy condition* $\mathscr{N}(p, (2s)^{-1}, s, Y)$ [*see* (3.2)]. *Then*

$$J =_{\text{def}} \int_{1/\sqrt{pN}}^{A} |\mathbf{E}e\{tT_*\}| \frac{dt}{t} \ll_s \frac{1 + \log A}{pN} \quad \textit{for any } A \geq 1. \tag{7.5}$$

PROOF OF COROLLARY 7.2. Without loss of generality we shall assume that $pN \geq 1$ since otherwise integrating the trivial bound $|\mathbf{E}e\{tT_*\}| \leq 1$ we prove (7.5).

By Lemma 7.1 we have $|\mathbf{E}e\{tT_*\}| \leq \mathbf{E}\psi(t)$. Furthermore, in the case of the statistic (7.4) we can choose $1 \leq M \leq N$ arbitrary, say $M \sim N/2$. Applying Lemma 7.1 with $d = 2$ and $m = 1$, we see that the complement D^c of the vent

$$D = \{\psi(t)\psi(t + \tau) \leq c_3(s, d)\mathscr{M}^s(\tau; pN) \text{ for all } t, \tau \in \mathbb{R}\} \tag{7.6}$$

satisfies $\mathbf{P}\{D^c\} \ll_s (pN)^{-2}$. Hence, writing

$$J_1 = \mathbf{E}\int_{1/\sqrt{pN}}^{A} \mathbf{1}\{D\}\psi(t)\frac{dt}{t}$$

and combining Fubini's theorem with $\psi(t) \leq 1$, we have

$$\begin{aligned} J &\leq J_1 + \mathbf{P}\{D^c\}\int_{1/\sqrt{pN}}^{A} \frac{dt}{t} \\ &\ll_s J_1 + (pN)^{-2}\log\big(A\sqrt{pN}\big) \ll J_1 + \frac{1 + \log A}{pN}. \end{aligned} \tag{7.7}$$

To estimate J_1 we shall apply Lemma 3.2. If the event D occurs, then (7.6) holds and therefore the condition (3.7) is fulfilled with N replaced by pN. Choosing $B = 1$, we see that Lemma 3.2 yields

$$J_1 \ll_s \frac{1 + \log A}{pN}. \tag{7.8}$$

The bounds (7.7) and (7.8) yield (7.5). □

PROOF OF LEMMA 7.1. Recall that L, K, K_0 are defined by (6.4). Let $\rho_1, \rho_2, \ldots$ be i.i.d. random variables such that $\mathbf{P}\{\rho_1 = 1\} = \mathbf{P}\{\rho_1 = 0\} = 1/2$. Write

$$Y_j = (2m)^{-1/2} \sum_{k \in I(j)} \tilde{X}_k, \qquad I(j) = (mj - m, mj], \tag{7.9}$$

and [cf. (6.19)]

$$\Lambda_r = \sum_{(r-1)sL < j \leq rsL} \rho_j \tilde{\varepsilon}_j Y_j \quad \text{for } 1 \leq r \leq 3, \tag{7.10}$$

where $\tilde{\varepsilon}_j$ denote symmetrizations of i.i.d Rademacher random variables.

We shall derive the lemma from the following bound:

(7.11)
$$2\psi^2(\tau-\gamma)\psi^2(\tau+\gamma) \le \mathbf{E}_{\rho,\varepsilon}\mathrm{e}\{m\gamma\phi(\Lambda_1,\Lambda_2)\} + \mathbf{E}_{\rho,\varepsilon}\mathrm{e}\{m\gamma\phi(\Lambda_1,\Lambda_3)\},$$

where $\mathbf{E}_{\rho,\varepsilon}$ denotes the expectation with respect to the random variables $\rho_1, \rho_2, \dots$ and $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots$.

In order to prove (7.11), let us start with a simple identity. Let $\zeta_1, \zeta_2, \dots$ denote an independent copy of the sequence $\vartheta_1, \vartheta_2, \dots$ introduced in (6.3). Then, for any function f and given $\vartheta_1, \vartheta_2, \dots$, we have

(7.12)
$$\mathbf{E}_\zeta f(\zeta_1\vartheta_1, \zeta_2\vartheta_2 \dots) = \mathbf{E} f(\zeta_1, \zeta_2, \dots),$$

where $\mathbf{E}_\zeta = \mathbf{E}_{\zeta_1,\zeta_2,\dots}$. Indeed, for any $j \in \mathbb{N}$, the conditional distribution of $\zeta_j\vartheta_j$ given ϑ_j is independent of ϑ_j and is equal to the distribution of ζ_j.

Define

(7.13)
$$Q^\vartheta = \sum_{1\le j<k\le 3K_0} \vartheta_j\vartheta_k a_{jk},$$

and $T^{\zeta\vartheta}$ just as T^ϑ with ϑ_j replaced by $\vartheta_j\zeta_j$. Using (7.1), (7.12), (7.13), Fubini's theorem and Hölder's inequality, we get

(7.14)
$$\begin{aligned}\psi^2(\tau-\gamma)\psi^2(\tau+\gamma) &= \left|\mathbf{E}_\vartheta \psi(\tau-\gamma)\mathrm{e}\{-(\tau+\gamma)T^\vartheta\}\right|^2\\ &= \left|\mathbf{E}_\vartheta\mathbf{E}_\zeta\,\mathrm{e}\{(\tau-\gamma)T^{\zeta\vartheta} - (\tau+\gamma)T^\vartheta\}\right|^2\\ &\le \mathbf{E}_\zeta\left|\mathbf{E}_\vartheta\,\mathrm{e}\{(\tau-\gamma)T^{\zeta\vartheta} - (\tau+\gamma)T^\vartheta\}\right|^2\\ &= \mathbf{E}_\zeta J,\end{aligned}$$

where

$$J = \left|\mathbf{E}_\vartheta\mathrm{e}\{\tau(Q^{\zeta\vartheta} - Q^\vartheta) - \gamma(Q^{\zeta\vartheta} + Q^\vartheta) + f_1 + f_2\}\right|^2,$$

with a statistic f_1 depending only on ϑ_j, $j \in [1, K_0]$, and a statistic f_2 independent of ϑ_j, $j \in [1, K_0]$; of course, f_1 and f_2 may depend on τ, γ, ζ_j, X_j and $\overline{X}_j$.

It is easy to see that

(7.15)
$$\begin{aligned}Q^{\zeta\vartheta} - Q^\vartheta &= \sum_{1\le j<k\le 3K_0} (\zeta_j\zeta_k - 1)\vartheta_j\vartheta_k a_{jk},\\ Q^{\zeta\vartheta} + Q^\vartheta &= \sum_{1\le j<k\le 3K_0} (\theta_j\theta_k + 1)\vartheta_j\vartheta_k a_{jk}.\end{aligned}$$

Notice that

(7.16)
$$\zeta_j\zeta_k - 1 = 0 \quad\text{and}\quad \zeta_j\zeta_k + 1 = 2 \text{ whenever } \zeta_j = \zeta_k = 1.$$

For given $\zeta_1, \zeta_2, \dots$ introduce the sets

$$\Omega_1 = \{j \in [1, K_0]: \zeta_j = 1\}, \qquad \Omega_2 = \{j \in (K_0, 2K_0]: \zeta_j = 1\},$$
$$\Omega_3 = \{j \in (2K_0, 3K_0]: \zeta_j = 1\}, \qquad \Omega_4 = [1, 3K_0] \setminus (\Omega_1 \cup \Omega_2 \cup \Omega_3).$$

The sets Ω_j, $1 \le j \le 4$, are random since they depend on $\zeta_1, \zeta_2, \dots$, but they are independent of $\vartheta_1, \vartheta_2, \dots$. Write $\mathbf{E}_1 = \mathbf{E}_{\vartheta_j,\, j \in \Omega_1 \cup \Omega_2 \cup \Omega_3}$. Then, conditioning on ϑ_j, $j \notin \Omega_1 \cup \Omega_2 \cup \Omega_3$, regrouping the summands, and using (7.15) and (7.16), we obtain

$$(7.17)\qquad \begin{aligned} J &\le \mathbf{E}_\vartheta |\mathbf{E}_1 \mathrm{e}\{-2\gamma Q' + f_1^* + f_2^*\}|^2, \\ Q' &= \sum_{\{j,k\} \subset \Omega_1 \cup \Omega_2 \cup \Omega_3} \vartheta_j \vartheta_k a_{jk}, \end{aligned}$$

with a statistic f_1^* depending on ϑ_j only such that $j \in \Omega_1 \cup \Omega_4$, and with a statistic f_2^* independent of ϑ_j such that $j \in \Omega_1$. In order to bound the expectation $\mathbf{E}_1 \dots$ in (7.17), let us apply the symmetrization Lemma 5.5. Then we have

$$(7.18)\qquad 2J \le J_1 + J_2,$$

where

$$(7.19)\qquad J_1 = \mathbf{E}_\vartheta \mathbf{E}_1 \mathrm{e}\{-2\gamma Q_1\}, \qquad Q_1 = \sum_{j \in \Omega_1,\, k \in \Omega_2} a_{jk} \tilde{\vartheta}_j \tilde{\vartheta}_k,$$

$$(7.20)\qquad J_2 = \mathbf{E}_\vartheta \mathbf{E}_1 \mathrm{e}\{-2\gamma Q_2\}, \qquad Q_2 = \sum_{j \in \Omega_1,\, k \in \Omega_3} a_{jk} \tilde{\vartheta}_j \tilde{\vartheta}_k,$$

Let us rewrite the conditional expectations in (7.19) and (7.20) as unconditional ones. To this end introduce the random variables $\sigma_j = (\zeta_j + 1)/2$. Notice that $\sigma_j = 1$, if $\zeta_j = 1$, and $\sigma_j = 0$, if $\zeta_j = -1$. Therefore we can write

$$Q_1 = \sum_{j=1}^{K_0} \sum_{k=K_0+1}^{2K_0} a_{jk} \sigma_j \sigma_k \tilde{\vartheta}_j \tilde{\vartheta}_k, \qquad Q_2 = \sum_{j=1}^{K_0} \sum_{k=2K_0+1}^{3K_0} a_{jk} \sigma_j \sigma_k \tilde{\vartheta}_j \tilde{\vartheta}_k,$$

and

$$(7.21)\qquad J_1 = \mathbf{E}_\vartheta \mathrm{e}\{-2\gamma Q_1\}, \qquad J_2 = \mathbf{E}_\vartheta \mathrm{e}\{-2\gamma Q_2\}.$$

By the definition [see (6.3)], we have $\vartheta_j = \varepsilon_l$, for $j \in I(l)$. Since $\zeta_1, \zeta_2, \dots$ is an independent copy of $\vartheta_1, \vartheta_2, \dots$, we can introduce the random variables $\rho_l = \sigma_j$, for $j \in I(l)$. Due to the definition of σ_j as a function of ζ_j, the random variables $\rho_1, \rho_2, \dots$ are i.i.d. and assume only values 0 and 1, each with probability $1/2$. Thus, using $4a_{jk} = \phi(\tilde{X}_j, \tilde{X}_k)$, $K_0 = msL$ and the linearity of $\phi(x, y)$ in its arguments, we can write

$$(7.22)\qquad Q_1 = \frac{1}{4} \phi\left(\sum_{j=1}^{K_0} \sigma_j \tilde{\vartheta}_j \tilde{X}_j, \sum_{k=K_0+1}^{2K_0} \sigma_k \tilde{\vartheta}_k \tilde{X}_k \right) = \frac{m}{2} \phi(\Lambda_1, \Lambda_2)$$

since, for example,

$$\sum_{j=1}^{K_0} \sigma_j \tilde{\vartheta}_j \tilde{X}_j = \sum_{j=1}^{sL} \rho_l \tilde{\varepsilon}_l \sum_{j \in I(l)} \tilde{X}_j = \sqrt{2m}\, \Lambda_1.$$

Similarly, $Q_2 = (m/2)\phi(\Lambda_1, \Lambda_3)$. Using the symmetry of the random variables $\tilde{\varepsilon}_j$, writing $\mathbf{E}_{\zeta,\vartheta} = \mathbf{E}_{\rho,\varepsilon}$ and combining (7.14), (7.18), (7.21) and (7.22), we obtain (7.11).

Let us return to the proof of the lemma. Define the $s \times s$-matrices

$$\mathbb{A}(j,k) = (\phi(Y_l, Y_r)), \qquad l \in (js - s, js],\ r \in (ks - s, ks].$$

Introduce the random events and variables

$$B(j,k) = \left\{ |\mathbb{A}(j,k) - \mathbb{I}|_\infty \le (2s)^{-1} \right\}, \qquad \xi_{jk} = \mathbf{1}\{B(j,k)\}.$$

The condition $\mathscr{N}(p, (2s)^{-1}, s, Y)$ guarantees that $\mathbf{E}\,\xi_{jk} = \mathbf{P}\{B(j,k)\} \ge p$, for $j \ne k$. Furthermore, the events $B(j,k)$ occur with equal probabilities, provided that $j \ne k$. Introduce the events

$$D_1 = \left\{ \sum_{j=1}^{L} \sum_{k=L+1}^{2L} \xi_{jk} \ge \frac{pL^2}{2} \right\}, \qquad D_2 = \left\{ \sum_{j=1}^{L} \sum_{k=2L+1}^{3L} \xi_{jk} \ge \frac{pL^2}{2} \right\}.$$

and $D = D_1 \cap D_2$. Then

$$\mathbf{P}\{D\} \ge 1 - c_d (pL)^{-d} \ge 1 - c_d(s)(pK)^{-d}, \tag{7.23}$$

for any $d \ge 0$, since by (6.24) each of events D_1 and D_2 satisfies an inequality similar to (7.23). Due to (7.23), it remains to show that the multiplicative inequality

$$\psi(t-\gamma)\psi(t+\gamma) \ll_{s,d} \mathscr{M}^s(\gamma m; pM/m)$$

is fulfilled whenever the event D occurs. The estimate (7.11) shows that it is sufficient to verify that

$$\mathbf{1}\{D_1\}\mathbf{E}e\{m\gamma\phi(\Lambda_1, \Lambda_2)\} \le \sup_{\mathbb{A}} \mathbf{E}e\{m\gamma\langle \mathbb{A}U, V\rangle\}, \tag{7.24}$$

$$\mathbf{1}\{D_2\}\mathbf{E}e\{m\gamma\phi(\Lambda_1, \Lambda_3)\} \le \sup_{\mathbb{A}} \mathbf{E}e\{m\gamma\langle \mathbb{A}U, V\rangle\}, \tag{7.25}$$

$$\sup_{\mathbb{A}} \mathbf{E}e\{m\gamma\langle \mathbb{A}U, V\rangle\} \ll_{s,d} \mathscr{M}^{2s}(m\gamma; pM/m). \tag{7.26}$$

The inequality (7.26) follows from Lemma 6.7, using $pL \asymp pM/m$. Inequalities (7.24) and (7.25) differ from (6.3) in notation only. For example, we get (7.24) by replacing in (6.30) Λ, $\overline{\Lambda}$ and τ_j by Λ_1, Λ_2 and $\rho_j \tilde{\varepsilon}_j$, respectively. □

8. Expansions of characteristic functions. In this section we shall obtain expansions for the difference of the characteristic functions of the statistics T and T_0 defined by (1.1) and (1.6). Recall that

$$\hat{F}(t) = \mathbf{E}e\{tT\}, \qquad \hat{F}_0(t) = \mathbf{E}e\{tT_0\},$$

where F and F_0 are distribution functions of T and T_0. The Edgeworth correction F_1 is defined by (4.1). The bound $\varkappa = \varkappa(t)$ is given by (4.6) and (4.7). The moments β_s, γ_s, and $\gamma_{s,r}$ are defined by (1.3). Write

$$\hat{\Delta}_N = \left| \hat{F}(t) - \hat{F}_0(t) - \hat{F}_1(t) \right|.$$

The result of this section is the following lemma.

LEMMA 8.1. *We have*

$$\hat{\Delta}_N \ll \varkappa N^{-1} \big(t^4 \beta_4 + t^6 \beta_3^2 + t^2 \gamma_2 + |t|^3 \gamma_3 + |t|^5 \gamma_2 \gamma_3 + t^2 \gamma_{2,2} + t^6 \gamma_2 \gamma_{2,2} \big). \tag{8.1}$$

If, in addition, the condition (1.9) *is fulfilled, then* $\hat{\Delta}_N = |\hat{F}(t) - \hat{F}(t)|$, *and*

$$\hat{\Delta}_N \ll \varkappa N^{-1}\big(t^4\beta_4 + t^2\gamma_2 + |t|^3\gamma_3 + t^4\gamma_{2,2}\big). \tag{8.2}$$

The result follows from Lemmas 8.2–8.4 by an obvious application of (8.9)–(8.17). □

This section is organized as follows. First we shall introduce Bergström-type identities, which are valid for arbitrary statistics of arbitrary samples. Then we shall consider a specialized version of these identities for symmetric statistics of i.i.d. samples. Combined with Taylor expansions, these identities provide the Edgeworth expansions of Lemmas 8.3 and 8.4.

Bergström-type identities. Consider two arbitrary samples

$$X_1, \dots, X_N \quad \text{and} \quad G_1, \dots, G_N. \tag{8.3}$$

Let

$$S = S(X_1, \dots, X_N)$$

be an (eventually nonsymmetric) statistic based on the sample $X_1, \dots, X_N$. For a subset $A \subset [1, N]$, introduce the operation $\mathscr{R}_A$ as replacement of the random variables X_j, $j \in A$, by G_j, $j \in A$. For example,

$$\mathscr{R}_{[1,j]}S = S\big(G_1, \dots, G_j, X_{j+1}, \dots, X_N\big). \tag{8.4}$$

Using this notation, we have the obvious identity

$$S = \mathscr{R}_{[1,N]}S + \sum_{j=1}^{N} \big\{\mathscr{R}_{[1,j-1]}S - \mathscr{R}_{[1,j]}S\big\}. \tag{8.5}$$

Repeating the procedure with the statistic in braces $\{\cdots\}$ in (8.5), we can replace $X_{j+1}, \dots, X_N$ by $G_{j+1}, \dots, G_N$ and obtain

$$\begin{aligned} S - \mathscr{R}_{[1,N]}S - \sum_{j=1}^{N} \big\{\mathscr{R}_{[1,N]\setminus\{j\}}S - \mathscr{R}_{[1,N]}S\big\} \\ = \sum_{j=1}^{N} \sum_{k=j+1}^{N} \big\{\mathscr{R}_{[1,k]\setminus\{j,k\}}S - \mathscr{R}_{[1,k]\setminus\{k\}}S \\ -\mathscr{R}_{[1,k]\setminus\{j\}}S + \mathscr{R}_{[1,k]}S\big\}. \end{aligned} \tag{8.6}$$

It is easy to extend (8.5) and (8.6) to higher order differences. The identities (8.5) and (8.6) will be called Bergström expansions (cf. with the classical Bergström expansions for sums of independent random variables).

In particular, if the statistic S is a *symmetric* function of its arguments and each of the (independent) samples (8.3) is *i.i.d.*, then (8.5) implies

$$\mathbf{E}S = \mathbf{E}\mathscr{R}_{[1,N]}S + \sum_{j=1}^{N} \big\{\mathbf{E}\mathscr{R}_{[2,j]}S - \mathbf{E}\mathscr{R}_{[1,j]}S\big\}. \tag{8.7}$$

Similarly, changing the order of summation and collecting equal terms, we see that under the same conditions (8.6) turns into

$$\begin{aligned}\mathbf{E}S - \mathbf{E}\mathscr{R}_{[1,N]}S - N\{\mathbf{E}\mathscr{R}_{[2,N]}S - \mathbf{E}\mathscr{R}_{[1,N]}S\} \\ = \sum_{k=2}^{N}(k-1)\{\mathbf{E}\mathscr{R}_{[3,k]}S - \mathbf{E}\mathscr{R}_{[2,k]}S \\ -\mathbf{E}\mathscr{R}_{[1,k]\setminus\{2\}}S + \mathbf{E}\mathscr{R}_{[1,k]}S\}.\end{aligned} \tag{8.8}$$

In order to simplify the extensive calculations we introduce the abbreviations

$$f = t\phi_1/\sqrt{N}, \qquad h = t\phi/N, \qquad h_0 = t\phi_0/(2N). \tag{8.9}$$

Then, for the statistic T defined by (1.1), we can write

$$tT = tT(X_1,\dots,X_N) = \sum_{1\le j<k\le N} h(X_j,X_k) + \sum_{1\le j\le N} f(X_j). \tag{8.10}$$

Using (1.6), (2.9) and (2.12), we may assume that

$$tT_0 = tT(G_1,\dots,G_N) + \sum_{1\le j\le N} h_0(G_j,G_j), \tag{8.11}$$

$$t\mathscr{R}_{[1,N]}T = tT(G_1,\dots,G_N) = \sum_{1\le j<k\le N} h(G_j,G_k) + \sum_{1\le j\le N} f(G_j), \tag{8.12}$$

$$\begin{aligned}\hat{F}_1(t) &= \frac{N}{6}\mathbf{E}(iL + ih(X,G_1))^3\, \mathrm{e}\{tT_0\},\\ L &= f(X) + \sum_{2\le j\le N} h(X,G_j).\end{aligned} \tag{8.13}$$

Define as well as the following modification of $\hat{F}_1$:

$$\begin{aligned}\hat{F}_1^*(t) &= \frac{N}{6}\mathbf{E}(iL)^3\mathrm{e}\{U\},\\ U &= \sum_{2\le j<k\le N} h(G_j,G_k) + \sum_{2\le j\le N} f(G_j).\end{aligned} \tag{8.14}$$

Introducing

$$\begin{aligned}\bar{\beta}_s &= \mathbf{E}|f(X)|^s, \qquad \bar{\gamma}_s = \mathbf{E}\left|h(X,\bar{X})\right|^s,\\ \bar{\gamma}_{s,q} &= \mathbf{E}\left(\mathbf{E}_X\left|h(X,\bar{X})\right|^s\right)^q,\end{aligned} \tag{8.15}$$

we have

$$\bar{\beta}_s = |t|^s N^{-s/2}\beta_s, \qquad \bar{\gamma}_s = |t|^s N^{-s}\gamma_s, \qquad \bar{\gamma}_{s,q} = |t|^{sq}N^{-sq}\gamma_{s,q}. \tag{8.16}$$

Furthermore, we can rewrite the definition of $\varkappa = \varkappa_1 + \varkappa_2$ as

$$\begin{aligned}\varkappa_1(t) &= \sup_A \left|\mathbf{E}\mathrm{e}\left\{\sum_{1\le j<k\le l} h(X_j,X_k) + A(X_1,\dots,X_l)\right\}\right|,\\ \varkappa_2(t) &= \sup_A \left|\mathbf{E}\mathrm{e}\left\{\sum_{1\le j<k\le l} h(G_j,G_k) + A(G_1,\dots,G_l)\right\}\right|,\end{aligned} \tag{8.17}$$

where $\sup_A$ is taken over all linear statistics A, that is, such that $A(x_1, \dots, x_l) = \sum_{j=1}^{l} f_j(x_j)$ with some functions $f_1, \dots, f_l$ and $l = [(N-2)/20]$. For the sake of brevity, in this section we shall write $\varkappa$ instead of $\varkappa(t)$.

Furthermore, we shall use the inequalities

$$\mathbf{E}|h_0(G,G)|^p \ll_p \overline{\gamma}_2^{p/2} \le \overline{\gamma}_{2,p/2} \le \overline{\gamma}_p, \qquad p \ge 2. \tag{8.18}$$

It suffices to prove (8.18) assuming that p is even. Using the representation (2.10) of $\phi_0(G,G)$ and applying the well-known Zygmund–Marcinkievicz inequality, we have

$$\mathbf{E}|\phi_0(G,G)|^p \ll_p \mathbf{E}\Big|\sum_{j\ge 1} q_j^2\big(\eta_j^4 + 1\big)\Big|^{p/2} \ll_p \sum_{j_1,\dots,j_{p/2}\ge 1} q_{j_1}^2 \cdots q_{j_{p/2}}^2 \ll_p \gamma_2^{p/2},$$

whence (8.18) follows.

The proofs of Lemmas 8.2 and 8.3 are simpler than the proof of the main Lemma 8.4. Therefore in these proofs we demonstrate in more detail certain technicalities, which are widely used in the proof of Lemma 8.4. The proof of Lemma 8.2 is based on the standard splitting and conditioning techniques. It is possible to considerably simplify the proofs of Lemmas 8.3 and 8.4 allowing the bounds to depend on $\overline{\gamma}_6$ instead of $\overline{\gamma}_{2,2}$. In order to prove bounds in terms of $\overline{\gamma}_{2,2}$, we have to make extensive use of identities of type (8.44) combined with the inequalities (8.45). A rough scheme of the proof of these lemmas is to expand the characteristic function in Taylor's series with a remainder which contains only squares, say $h^2(X_1, X_2)$, of the kernel h as the highest power of h. In case the order of the remaining terms is still not $\mathscr{O}(N^{-1})$, we have to apply Taylor's expansions again involving $h(X_i, X_j)$ such that, say X_i, is independent of at least one of X_1 or X_2.

Lemma 8.2. *The statistics* (8.11) *and* (8.12) *satisfy*

$$\big|\mathbf{E}e\{\mathscr{R}_{[1,N]}tT\} - \mathbf{E}e\{tT_0\}\big| \ll \varkappa\big(N\overline{\beta}_4 + N\overline{\gamma}_2 + N^2\overline{\gamma}_3\big). \tag{8.19}$$

For the Edgeworth corrections (8.13) *and* (8.14), *we have*

$$\big|\hat{F}_1(t) - \hat{F}_1^*(t)\big| \ll \varkappa\big(N\overline{\beta}_4 + N\overline{\gamma}_2 + N^2\overline{\gamma}_3 + N^3\overline{\gamma}_{2,2}\big). \tag{8.20}$$

Proof. Let us prove (8.19). Using (8.11) and (8.12), we can write

$$tT_0 = \mathscr{R}_{[1,N]}tT + d, \qquad d =_{\mathrm{def}} \sum_{j=1}^{N} h_0(G_j, G_j), \tag{8.21}$$

Expanding in powers of d in a Taylor series with remainder $\mathscr{O}(d^2)$ [that is, using (2.19) with $r = 1$], we have

$$\big|\mathbf{E}e\{\mathscr{R}_{[1,N]}tT\} - \mathbf{E}e\{tT_0\}\big| \ll J_1 + \mathbf{E}J_2, \tag{8.22}$$

with

$$J_1 = \big|\mathbf{E}\, d e\{\mathscr{R}_{[1,N]}tT\}\big|, \qquad J_2 = \big|\mathbf{E}^\tau\, d^2 e\{\mathscr{R}_{[1,N]}tT + \tau d\}\big|,$$

where τ is a random variable uniformly distributed on $[0,1]$ and independent of all other random variables. Recall that $\mathbf{E}^\tau$ denotes the conditional expectation given τ and $\mathbf{E}_\tau$ stands for the conditional expectation given all random variables except τ. In order to prove (8.19), it suffices to verify that J_1 and J_2 in (8.22) satisfy

$$|J_1| \ll \varkappa\Big(N\overline{\beta}_4 + N\overline{\gamma}_2 + N^2\overline{\gamma}_3\Big), \qquad |J_2| \ll \varkappa N\overline{\gamma}_2. \tag{8.23}$$

Let us prove (8.23) for J_2. Split $d = d_1 + d_2 + d_3$, where

$$d_1 = \sum_{1\le j\le N/3} h_0(G_j, G_j), \qquad d_2 = \sum_{N/3<j\le 2N/3} h_0(G_j, G_j),$$

and $d_3 = d - d_1 - d_2$. Then

$$|J_2| \le \sum_{1\le j,\,k\le 3} \Big|\mathbf{E}^\tau\, d_j d_k\, \mathrm{e}\big\{\mathscr{R}_{[1,N]}tT + \tau d\big\}\Big|. \tag{8.24}$$

Any one of the products $d_j d_k$ does not contains at least one of the terms d_1, d_2, d_3. For example, let us consider the case of $d_1 d_2$. Conditioning on d_1 and d_2, using the definition of $\varkappa$, the i.i.d. assumption and the inequality $ab \le a^2 + b^2$, we have

$$\begin{aligned}
&\Big|\mathbf{E}^\tau\, d_1 d_2\, \mathrm{e}\big\{\mathscr{R}_{[1,N]}tT + \tau d\big\}\Big| \\
&\quad\le \mathbf{E}^\tau |d_1|\,|d_2|\,\Big|\mathbf{E}_{d_3}\mathrm{e}\big\{\mathscr{R}_{[1,N]}tT + \tau d\big\}\Big| \\
&\quad\le \big(\mathbf{E}d_1^2 + \mathbf{E}d_2^2\big)\varkappa \ll N\mathbf{E}\big(h_0(G,G)\big)^2\varkappa \ll N\overline{\gamma}_2\varkappa,
\end{aligned} \tag{8.25}$$

since $\mathbf{E}h_0(G,G) = 0$ and, by (8.18), we have $\mathbf{E}(h_0(G,G))^2 \ll \overline{\gamma}_2$. Combining (8.24) and (8.25), we obtain (8.23) for $|J_2|$.

Let us verify (8.32) for J_1. Using the i.i.d assumption and (8.21), we have

$$J_1 = N\Big|\mathbf{E}h_0(G_1, G_1)\mathrm{e}\big\{\mathscr{R}_{[1,N]}tT\big\}\Big|. \tag{8.26}$$

Split

$$R_{[1,N]}tT = K + U \quad \text{where } K = f(G_1) + \sum_{j=2}^{N} h(G_1, G_j) \tag{8.27}$$

and where U is defined by (8.14) and is independent of G_1. Notice that

$$\mathbf{E}h_0(G_1, G_1) = 0 \quad \text{and} \quad \mathbf{E}_{G_1}h_0(G_1, G_1)K = 0, \tag{8.28}$$

due to the symmetry of G_1 [that is, $\mathscr{L}(G_1) = \mathscr{L}(-G_1)$]. Using (8.26), (8.28) and expanding in powers of K, we obtain

$$|J_1| \ll N\Big|\mathbf{E}^\tau h_0(G_1, G_1)K^2\mathrm{e}\{U + \tau K\}\Big|. \tag{8.29}$$

In order to bound the expectations in (8.29), let us repeat the arguments leading to (8.25): split the sum K into a sum of $f(G_1)$ and of three parts with approximately equal number of summands; condition on the part which is absent in the product and so forth. We get

$$\begin{aligned}
|J_1| &\ll N\varkappa\mathbf{E}\big|h_0(G,G)\big|f^2(G) \\
&\quad + N\varkappa \max_{1\le m\le N} \mathbf{E}\big|h_0(G,G)\big|\bigg(\sum_{j=2}^{m} h(G, G_j)\bigg)^2
\end{aligned} \tag{8.30}$$

since $\mathscr{L}(G_1) = \mathscr{L}(G)$. Using $ab \le a^2 + b^2$, applying (8.18) and replacing $\mathbf{E} f^4(G)$ by $\overline{\beta}_4$ by an application of (2.8), we have

$$\mathbf{E}|h_0(G,G)|f^2(G) \le \mathbf{E}\big(h_0(G,G)\big)^2 + \mathbf{E} f^4(G) \ll \overline{\gamma}_2 + \overline{\beta}_4. \tag{8.31}$$

Furthermore, using the i.i.d. assumption, Hölder's inequality, relations (2.8), (8.15), (8.18) and $\mathscr{L}(G_j) = \mathscr{L}(\overline{G})$, we obtain

$$\begin{aligned} \max_{1 \le m \le N} \mathbf{E}|h_0(G,G)| \bigg(\sum_{j=1}^{m} h(G,G_j) \bigg)^2 \\ \le N \mathbf{E}|h_0(G,G)| h^2(G,\overline{G}) \ll N\overline{\gamma}_3. \end{aligned} \tag{8.32}$$

Collecting (8.30)–(8.32), we obtain (8.23) for J_1, which concludes the proof of (8.19).

Let us prove (8.20). Using the notation (8.14), (8.13), (8.27) and (8.21), we can write

$$\hat{F}_1^*(t) = \frac{N}{6} \mathbf{E}(iL)^3 \mathrm{e}\{U\}, \qquad \hat{F}_1(t) = \frac{N}{6} \mathbf{E}\big(iL + ih(X,G_1)\big)^3 \mathrm{e}\{U + K + d\}.$$

Introducing

$$\hat{F}_2(t) = \frac{N}{6} \mathbf{E}(iL)^3 \mathrm{e}\{U + K + d\},$$

we reduce the proof of (8.20) to showing

$$\big|\hat{F}_1(t) - \hat{F}_2(t)\big| \ll \varkappa\big(N\overline{\beta}_4 + N\overline{\gamma}_2 + N^2\overline{\gamma}_3\big), \tag{8.33}$$

$$\big|\hat{F}_2(t) - \hat{F}_1^*(t)\big| \ll \varkappa\big(N\overline{\beta}_4 + N^3\overline{\gamma}_{2,2}\big). \tag{8.34}$$

Decompose $[1, N] = \bigcup_{s=1}^{5} \Omega_s$ into five disjoint sets with approximately equal cardinalities. Furthermore, split each of the sums L, K and d into the parts

$$L = \sum_{s=1}^{6} L_s, \qquad K = \sum_{s=1}^{6} K_s, \qquad d = \sum_{s=1}^{6} d_s,$$

where, for example,

$$L_6 = f(X),\ K_6 = f(G_1),\ d_6 = 0,\ K_s = \sum_{j \in \Omega_s \setminus \{1\}} h(G_1, G_j),\ 1 \le s \le 5.$$

Using the splitting and conditioning arguments, applying $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$ and several times $a^p b^q \ll a + b$, for $p, q \ge 0$ such that $p + q = 1$, we have

$$\begin{aligned} &\big|\hat{F}_1(t) - \hat{F}_2(t)\big| \\ &\quad \ll \varkappa N \max_{1 \le s \le 5} \mathbf{E}|h(X,G_1)|\big(h^2(X,G_1) + f^2(X) + L_s^2\big) \\ &\quad \ll \varkappa N\big(\overline{\beta}_4 + \overline{\gamma}_2\big) + \varkappa N^2 \mathbf{E}|h(X,G_1)|^3 \\ &\qquad + \varkappa N^{1/2} \max_{1 \le s \le 5} \mathbf{E}|L_s|^3. \end{aligned} \tag{8.35}$$

The inequality (8.35) implies (8.33) since, by an application of (2.8) and (2.18), the last two expectations in (8.35) are bounded from above by $c\overline{\gamma}_3$ and $cN^{3/2}\overline{\gamma}_3$, respectively.

Let us prove (8.34). Expanding in powers of $K + d$ with remainder $\mathscr{O}(K + d)$, splitting and conditioning again, we have

$$
\begin{aligned}
&\left|\hat{F}_2(t) - \hat{F}_1^*(t)\right| \\
&\quad \ll \varkappa N \max_{1 \le s_1, s_2, s_3 \le 6} \mathbf{E}|L_{s_1}|^3\big(|K_{s_2}| + |d_{s_3}|\big) \\
&\quad \ll \varkappa N \max_{1 \le s_1, s_2, s_3 \le 6} \mathbf{E}\big(L_{s_1}^4 + K_{s_2}^4 + d_{s_3}^4\big) \\
&\quad \ll \varkappa N \overline{\beta}_4 + \varkappa N\Big(\max_{1 \le s \le 5} \mathbf{E}L_s^4 + \max_{1 \le s \le 5} \mathbf{E}K_s^4 + \max_{1 \le s \le 5} \mathbf{E}d_s^4\Big).
\end{aligned} \tag{8.36}
$$

An application of (2.18), (2.8), and (8.18) shows that all expectations on the right-hand side of (8.36) are bounded from above by $cN^2\overline{\gamma}_{2,2}$ proving (8.34). □

In the proofs of Lemmas 8.3 and 8.4, we shall use the following notation and facts. Write

$$\varphi(x) = \mathbf{E}h^2(X, x), \qquad \varphi_3(x) = \mathbf{E}|h(X, x)|^3. \tag{8.37}$$

Then, for $0 \le \alpha \le 1$, Hölder's inequality implies

$$\mathbf{E}|h(X, x)|^{\alpha+2} \le \varphi^{1-\alpha}(x)\varphi_3^{\alpha}(x), \qquad \mathbf{E}\varphi^{\alpha+1}(X) \le \overline{\gamma}_2^{1-\alpha}\,\overline{\gamma}_{2,2}^{\alpha}, \tag{8.38}$$

$$
\begin{aligned}
\mathbf{E}|h(z, X)|^{1/2}h^2(y, X) &\le \big(\mathbf{E}|h(z, X)|h^2(y, X)\big)^{1/2}\varphi^{1/2}(y) \\
&\le \varphi_3^{1/6}(z)\varphi_3^{1/3}(y)\varphi^{1/2}(y).
\end{aligned} \tag{8.39}
$$

For a random vector Z, write

$$\hat{Z} = f(Z) + \sum_{k=1}^{N-1} h(Z, X_k). \tag{8.40}$$

The sum $\hat{Z}$ without summands $h(Z, X_i), h(Z, X_j), \ldots$ we shall denote as $\hat{Z}(i, j, \ldots)$. For example, $\hat{Z}(k) = \hat{Z} - h(Z, X_k)$, for $k \ge 1$. We write as well $\hat{Z}(0) = \hat{Z} - f(Z)$ and $\hat{Z}(0,1) = \hat{Z} - f(Z) - h(Z, X_1)$. Introducing the statistic

$$S = \sum_{1 \le j < k \le N-1} h(X_j, X_k) + \sum_{j=1}^{N-1} f(X_j), \tag{8.41}$$

we can write

$$tT = \mathscr{R}_{[2,1]}tT =_{\mathscr{D}} \hat{X} + S \quad \text{and} \quad \mathscr{R}_{[1,1]}tT =_{\mathscr{D}} \hat{G} + S. \tag{8.42}$$

To obtain (8.42), it suffices to change the numeration of random variables replacing X_1, G_1 and $X_2, \ldots, X_N$ by X, G and $X_1, \ldots, X_{N-1}$, respectively. This is possible due to the i.i.d. assumption.

Finally, split $[N/2, N-2] \cap \mathbb{N} = \bigcup_{s=1}^{7} \Omega_s$ into a union of seven disjoint subsets such that their cardinalities are approximately equal and bounded from below by $N/20$. Write

$$d_s(Z) = \sum_{j \in \Omega_s} h(Z, X_j), \; 1 \le s \le 7,$$
$$d_0(Z) = \sum_{j \notin [N/2, N-2]} h(Z, X_j). \tag{8.43}$$

We shall often use relations such as the following:

$$\mathbf{E} d_s(Z) \mathrm{e}\{S + \hat{Z}\} = \operatorname{card} \Omega_s \mathbf{E} h(Z, X_1) \mathrm{e}\{S + \hat{Z}\} \tag{8.44}$$

and

$$\begin{aligned} \left|\mathbf{E}\hat{Z}(0)\mathrm{e}\{S + \hat{Z}\}\right| &= (N-1)\left|\mathbf{E} h(Z, X_1)\mathrm{e}\{S + \hat{Z}\}\right| \\ &\ll \left|\mathbf{E} d_s(Z)\mathrm{e}\{S + \hat{Z}\}\right| \ll \left|\mathbf{E}\hat{Z}(0)\mathrm{e}\{S + \hat{Z}\}\right|. \end{aligned} \tag{8.45}$$

Both (8.44) and (8.45) are consequences of the i.i.d. assumption. We have as well

$$\mathbf{E}^Z |d_1(Z)| \le \left(\mathbf{E}^Z d_1^2(Z)\right)^{1/2} \ll N^{1/2} \varphi^{1/2}(Z). \tag{8.46}$$

LEMMA 8.3. *Assume that the third-order moments of X and G coincide, that is, that* (1.9) *holds. Then we have*

$$\left|\mathbf{E}\mathrm{e}\{tT\} - \mathbf{E}\mathrm{e}\{\mathscr{R}_{[1,N]} tT\}\right| \ll \varkappa\left(N\overline{\beta}_4 + N^2\overline{\gamma}_3 + N^3\overline{\gamma}_{2,2}\right).$$

PROOF. Let us apply the Bergström-type identity (8.7) replacing S by $\mathrm{e}\{tT\}$. Then we have

$$\left|\mathbf{E}\mathrm{e}\{tT\} - \mathbf{E}\mathrm{e}\{\mathscr{R}_{[1,N]} tT\}\right| \le \sum_{j=1}^{N} \delta_j,$$
$$\delta_j =_{\mathrm{det}} \left|\mathbf{E}\mathrm{e}\{\mathscr{R}_{[2,j]} tT\} - \mathbf{E}\mathrm{e}\{\mathscr{R}_{[1,j]} tT\}\right|. \tag{8.47}$$

Due to (8.47), in order to prove the lemma it suffices to verify that

$$\delta_j \ll \varkappa\left(\overline{\beta}_4 + N\overline{\gamma}_3 + N^2\overline{\gamma}_{2,2}\right). \tag{8.48}$$

To simplify notation, we shall estimate δ_1 only. An estimation of δ_j, $j > 1$, is similar since here $X_2, \dots, X_j$ are just replaced by the Gaussian random vectors $G_2, \dots, G_j$. Nevertheless, we can repeat the proof using inequalities of type (2.8) for the estimation of moments involving Gaussian random vectors by moments involving X_k. Another unessential complication is that formulas similar to (8.52) can include factors of the type $h^i(Z, G_j)$, which again leads to unessential although tedious additional technicalities. Using (8.42), we can represent δ_1 as

$$\delta_1 = \left|\mathbf{E}\mathrm{e}\{\hat{X} + S\} - \mathbf{E}\mathrm{e}\{\hat{G} + S\}\right|. \tag{8.49}$$

Expanding in (8.49) in a Taylor series with remainders $\mathscr{O}(\hat{X}^3)$ and $\mathscr{O}(\hat{G}^3)$, we get

$$\delta_1 \ll \mathbf{E}\big|\mathbf{E}^\tau a(X) - \mathbf{E}^\tau a(G)\big|, \qquad a(Z) =_{\text{def}} \hat{Z}^3 \mathrm{e}\{S + \tau\hat{Z}\}. \tag{8.50}$$

Here τ denotes a random variable uniformly distributed on $[0,1]$ independent of all other random variables, and $\mathbf{E}^\tau$ stands for the conditional expectation given τ. The terms of the expansion depending on the lower order derivatives cancel, since by (2.7) moments of X and G up to the second order coincide.

Writing $\hat{X}^3$ and $\hat{G}^3$ in (8.50) as triple sums, using the i.i.d. assumption and collecting identical summands, we obtain

$$\begin{aligned} &\mathbf{E}\delta_1 \ll \sum \mathbf{E} I(i_2), \\ &I(i_2) =_{\text{def}} \big|\mathbf{E}^\tau b(X)\mathrm{e}\{S + \hat{X}\} - \mathbf{E}^\tau b(G)\mathrm{e}\{S + \hat{G}\}\big|, \end{aligned} \tag{8.51}$$

with

$$b(Z) = N^{\theta(i_2)+\theta(i_3)+\theta(i_4)} f^{i_1}(Z) h^{i_2}(Z, X_1) h^{i_3}(Z, X_2) h^{i_4}(Z, X_3). \tag{8.52}$$

Here the sum in (8.51) extends over all nonnegative integers indices i_1, i_2, i_3, i_4 such that $i_1 + i_2 + i_3 + i_4 = 3$, and $i_2 \ge i_3 \ge i_4$. Furthermore, we write $\theta(x) = 1$, for $x > 0$ and $\theta(x) = 0$ for $x \le 0$. It is clear that $I(i_2)$ depends on i_1, i_3 and i_4 as well. However, the further estimation of this quantity depends on the value of i_2 only.

Due to (8.51), in order to prove (8.48) it suffices to show that

$$I(i_2) \ll \varkappa\Big(\overline{\beta}_4 + N\overline{\gamma}_3 + N^2\overline{\gamma}_{2,2}\Big), \tag{8.53}$$

for all allowable values of the indices i_1, i_2, i_3, i_4.

We shall write, for example,

$$I(i_2) \sim \big\{\hat{Z}(0); b'(Z)\big\} \quad \text{if } I(i_2) \ll I'(i_2) + \varkappa\Big(\overline{\beta}_4 + N\overline{\gamma}_3 + N^2\overline{\gamma}_{2,2}\Big),$$

where $I'(i_2)$ is defined similarly to $I(i_2)$ in (8.51), just replacing $\hat{Z}$ and b by $\hat{Z}(0)$ and b', respectively. In order to verify (8.53), we have to prove that $I(i_2) \sim \{\dots; 0\}$.

Let us prove (8.53) for $I(3)$. Then $i_1 = i_3 = i_4 = 0$ and using (8.15), (8.17), we have

$$I(3) \ll \max_{Z=X,\, Z=G} N\big|\mathbf{E}^\tau h^3(Z, X_1)\mathrm{e}\{S + \tau\hat{Z}\}\big| \ll \varkappa N\overline{\gamma}_3.$$

Let us prove (8.53) for $I(2)$. Then $i_4 = 0$ and we have

$$I(2) \sim \big\{\hat{Z}; b(Z)\big\}, \qquad b(Z) = N^{1+\theta(i_3)} f^{i_1}(Z) h^2(Z, X_1) h^{i_3}(Z, X_2). \tag{8.54}$$

The indices i_1 and i_3 in (8.54) satisfy $i_1 + i_3 = 1$. Hence, there are only two possibilities: either $i_1 = 1$, $i_3 = 0$ or $i_1 = 0$, $i_3 = 1$. Using an identity type (8.44), we can rewrite (8.54) as

$$I(2) \sim \big\{\hat{Z}; b_1(Z)\big\}, \qquad b_1(Z) = N f^{i_1}(Z) h^2(Z, X_1) d_1^{i_3}(Z) \tag{8.55}$$

(we write $0^0 = 1$). Let us show that in (8.55) we can remove $f(Z)$ and $h(Z, X_1)$ from the corresponding exponents, that is, that

$$I(2) \sim \{\hat{Z}(0,1); b_1(Z)\}. \tag{8.56}$$

We derive (8.56) by expanding the exponents in Taylor series with remainders $\mathscr{O}(f(Z))$ and $\mathscr{O}(\sqrt{|h(Z, X_1)|})$. The errors are bounded from above by

$$\rho_1(i_1) =_{\text{def}} \varkappa N \mathbf{E} \big| f^{i_1+1}(Z) h^2(Z, X_1) d_1^{i_3}(Z) \big|$$

and

$$\rho_2(i_1) =_{\text{def}} \varkappa N \mathbf{E} \big| f^{i_1}(Z) \big| \, \big| h(Z, X_1) \big|^{5/2} \big| d_1^{i_3}(Z) \big|,$$

respectively. The estimation of $\rho_j(i_1)$ will depend on i_1 only. We have to show that $\rho_1(i_1)$ and $\rho_2(i_1)$ are bounded from above by the right-hand side of (8.53). Assume first that $i_1 = 1$. Then $i_3 = 0$. Using (8.15), (8.37), Hölder's inequality and $ab \ll a^2 + b^2$, we have

$$\begin{aligned} \rho_1(1) &\ll \varkappa N \mathbf{E} f^2(Z) h^2(Z, X_1) = \varkappa N \mathbf{E} f^2(Z) \varphi(Z) \\ &\le \varkappa N \overline{\beta}_4^{1/2} \, \overline{\gamma}_{2,2}^{1/2} \ll \varkappa \overline{\beta}_4 + \varkappa N^2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.57}$$

Furthermore, using (8.15), (8.38) and Hölder's inequality, we obtain

$$\begin{aligned} \rho_2(1) &\ll \varkappa N \mathbf{E} | f(Z) | \mathbf{E}_{X_1} | h(Z, X_1) |^{5/2} \\ &\ll \varkappa N \mathbf{E} | f(Z) | \varphi^{1/2}(Z) \varphi_3^{1/2}(Z) \\ &\ll \varkappa N \big(\mathbf{E} f^2(Z) \varphi(Z) \big)^{1/2} \overline{\gamma}_3^{1/2} \ll \varkappa N \overline{\beta}_4^{1/4} \, \overline{\gamma}_{2,2}^{1/4} \, \overline{\gamma}_3^{1/2} \\ &\ll \varkappa N \overline{\beta}_4 + \varkappa N \overline{\gamma}_3 + \varkappa N^2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.58}$$

In the case $i_1 = 0$ we can proceed similarly as in the proof of (8.57) and (8.58). Using (8.15), (8.37), (8.38), (8.46) we have

$$\begin{aligned} \rho_1(0) &\ll \varkappa N^{3/2} \mathbf{E} | f(Z) | h^2(Z, X_1) \varphi^{1/2}(Z) = \varkappa N^{3/2} \mathbf{E} | f(Z) | \varphi^{3/2}(Z) \\ &\le \varkappa N^{3/2} \overline{\beta}_4^{1/4} \, \overline{\gamma}_{2,2}^{3/4} \ll \varkappa \overline{\beta}_4 + \varkappa N^2 \overline{\gamma}_{2,2}, \\ \rho_2(0) &\ll \varkappa N^{3/2} \mathbf{E} \mathbf{E}_{X_1} | h(Z, X_1) |^{5/2} \varphi^{1/2}(Z) \ll \varkappa N^{3/2} \mathbf{E} \varphi(Z) \varphi_3^{1/2}(Z) \\ &\ll \varkappa N^{3/2} \overline{\gamma}_3^{1/2} \, \overline{\gamma}_{2,2}^{1/2} \ll \varkappa N \overline{\gamma}_3 + \varkappa N^2 \overline{\gamma}_{2,2}. \end{aligned}$$

proving (8.56).

Let us show that (8.56) implies $I(2) \sim \{\hat{Z}(0, 1); 0\}$. To this end it suffices to expand the exponents in powers of $\hat{Z}(0, 1)$ with remainder $\mathscr{O}(\hat{Z}(0, 1))$. Lower order terms cancel since by the assumption (1.9) the third-order moments of X and G are equal, that is,

$$\begin{aligned} \mathbf{E} f^3(X) &= \mathbf{E} f^2(X) h(X, x) = \mathbf{E} f(X) h^2(X, x) \\ &= \mathbf{E} h(X, x) h(X, y) h(X, z) = 0, \end{aligned}$$

for all $x, y, z \in \mathscr{X}$. Using splitting of the sum $\hat{Z}(0,1)$ of type (8.43) such that the sum $d_0(Z)$ is independent of X_1, we get a remainder term which is bounded from above by

$$\rho(i_1) =_{\text{def}} \max_{0 \le s \le 7} \varkappa N \mathbf{E} \big| f^{i_1}(Z) h^2(Z, X_1) d_1^{i_3}(Z) d_s(Z) \big|$$
$$\ll \max_{0 \le s \le 7} \varkappa N \mathbf{E} \big| f^{i_1}(Z) \varphi(Z) d_1^{i_3}(Z) d_s(Z) \big|,$$

by an application of (8.44). We have

$$\rho(1) \ll \max_{0 \le s \le 7} \varkappa N \mathbf{E} \big| f(Z) \varphi(Z) d_s(Z) \big|$$
$$\ll \varkappa N^{3/2} \mathbf{E} | f(Z) | \varphi^{3/2}(Z) \ll \varkappa \overline{\beta}_4 + \varkappa N^2 \overline{\gamma}_{2,2}$$

and

$$\rho(0) = \max_{0 \le s \le 7} \varkappa N \mathbf{E} h^2(Z, X_1) | d_1(Z) d_s(Z) |$$
$$\ll \max_{0 \le s \le 7} \varkappa N \mathbf{E} \varphi(z) \big(\mathbf{E}^Z d_1^2(Z) \mathbf{E}^Z d_s^2(Z) \big)^{1/2} \ll \varkappa N^2 \mathbf{E} \varphi^2(Z) = \varkappa N^2 \overline{\gamma}_{2,2},$$

which completes the estimation of $I(2)$.

Let us prove (8.53) for $I(1)$. In this case

$$b(Z) = N^{1+\theta(i_3)+\theta(i_4)} f^{i_1}(Z) h(Z, X_1) h^{i_3}(Z, X_2) h^{i_4}(Z, X_3). \tag{8.59}$$

The relation (8.59) and identities of type (8.44) imply

$$I(1) \sim \big\{ \hat{Z}; b(Z) \big\}, \qquad b(Z) = f^{i_1}(Z) d_1(Z) d_2^{i_3}(Z) d_3^{i_4}(Z).$$

Expanding the exponents with remainder $\mathscr{O}(\hat{Z})$, we obtain $I(1) \sim \{\hat{Z}; 0\}$. Lower order terms cancel again. Using splitting of the sum $\hat{Z}(0,1)$ of type (8.43), we get a remainder term which is bounded from above by

$$\rho(i_1) =_{\text{def}} \max_{0 \le s \le 7} \varkappa \mathbf{E} \big| f^{i_1}(Z) d_1(Z) d_2^{i_3}(Z) d_3^{i_4}(Z) d_s(Z) \big|.$$

Using (8.43), (8.46) and Hólder's inequality, we have

$$\rho(0) = \max_{0 \le s \le 7} \varkappa \mathbf{E} \mathbf{E}^Z | d_1(Z) d_2(Z) d_3(Z) d_s(Z) |$$
$$\le \max_{0 \le s \le 7} \varkappa \mathbf{E} \big(\mathbf{E}^Z d_1^2(Z) \mathbf{E}^Z d_2^2(Z) \mathbf{E}^Z d_3^2(Z) \mathbf{E}^Z d_s^2(Z) \big)^{1/2}$$
$$\ll \varkappa N^2 \mathbf{E} \varphi^2(Z) \big) \ll \varkappa N^2 \overline{\gamma}_{2,2}.$$

The estimation of

$$\rho(1) = \max_{0 \le s \le 7} \varkappa \mathbf{E} | f(Z) d_1(Z) d_2(Z) d_s(Z) |,$$
$$\rho(2) = \max_{0 \le s \le 7} \varkappa \mathbf{E} f^2(Z) | d_1(Z) d_s(Z) |$$

is similar that of $\rho(0)$.

It remains to prove (8.53) for $I(0)$. In this case $b(Z) = f^3(Z)$. Similarly to the case in $I(1)$, we can expand exponents with a remainder $\mathscr{O}(\hat{Z})$. We obtain $I(0) \sim \{\hat{Z}\colon 0\}$. The remainder is bounded from above by $\max_{0 \le s \le 7} \varkappa \mathbf{E}|f^3(Z) d_s(Z)|$, which can be estimated similarly as in (8.79). □

LEMMA 8.4. *We have*

$$\begin{aligned} &\Big|\mathbf{E}e\{tT\} - \mathbf{E}e\{\mathscr{R}_{[1,N]}tT\} - \hat{F}_1^*(t)\Big| \\ &\quad \ll \varkappa\Big(N\overline{\beta}_4 + N^2\overline{\beta}_3^2 + N^2\overline{\gamma}_3 + N^4\overline{\gamma}_2\,\overline{\gamma}_3 + N^3\overline{\gamma}_{2,2} + N^5\overline{\gamma}_2\,\overline{\gamma}_{2,2}\Big), \end{aligned} \tag{8.60}$$

where the Edgeworth correction $\hat{F}_1^*$ *is defined by* (8.14).

PROOF. We shall use the Bergström-type expansion (8.8). Substituting $S = e\{T\}$ in (8.8) reduces the proof of (8.60) to a verification of

$$\Big|N\mathbf{E}e\{\mathscr{R}_{[2,N]}tT\} - N\mathbf{E}e\{\mathscr{R}_{[1,N]}tT\} - \hat{F}_1^*(t)\Big| \ll \varkappa\Big(N\overline{\beta}_4 + N^3\overline{\gamma}_{2,2}\Big) \tag{8.61}$$

and

$$\begin{aligned} &\max_{2 \le k \le N} |J_1 - J_2 - J_3 + J_4| \\ &\quad \ll \varkappa\Big(\overline{\beta}_3^2 + \overline{\gamma}_3 + N^2\overline{\gamma}_2\,\overline{\gamma}_3 + N\overline{\gamma}_{2,2} + N^3\overline{\gamma}_2\,\overline{\gamma}_{2,2}\Big) \end{aligned} \tag{8.62}$$

with

$$\begin{aligned} J_1 &= \mathbf{E}e\{\mathscr{R}_{[3,k]}tT\}, \qquad J_2 = \mathbf{E}e\{\mathscr{R}_{[2,k]}tT\}, \\ J_3 &= \mathbf{E}e\{\mathscr{R}_{[1,k]\setminus\{2\}}tT\}, \qquad J_4 = \mathbf{E}e\{\mathscr{R}_{[1,k]}tT\}. \end{aligned} \tag{8.63}$$

Let us prove (8.61). Setting $L(Z) = f(Z) + \sum_{j=2}^N h(Z, G_j)$ and using the notation (8.14), we can rewrite (8.61) as

$$\begin{aligned} &\Big|6\mathbf{E}e\{U + L(X)\} - 6\mathbf{E}e\{U + L(G)\} - \mathbf{E}(iL(X))^3 e\{U\}\Big| \\ &\quad \ll \varkappa\overline{\beta}_4 + \varkappa N^2\overline{\gamma}_{2,2}. \end{aligned} \tag{8.64}$$

The proof of (8.64) is simple—it suffices to expand in powers of $L(Z)$ with $Z = X, G$ and remainders $\mathscr{O}(Z^4)$, and to apply splitting and conditioning arguments as proof of Lemma 8.2. More precisely, split $L(Z) = \sum_{s=1}^6 L_j$ into approximately equal parts, for $1 \le j \le 5$, and put $L_6 = f(Z)$. Expanding and conditioning, we reduce the proof of (8.64) to the bound $\mathbf{E}L_s^4 \ll \overline{\beta}_4 + N^2\overline{\gamma}_{2,2}$. But using (2.8), we have $\mathbf{E}L_6^4 \le \mathbf{E}f^4(Z) \ll \overline{\beta}_4$. For $1 \le s \le 5$, an application of (2.18) yields $\mathbf{E}L_s^4 \ll N^2\mathbf{E}h^4(Z, G)$. Using (2.8) again, we have $\mathbf{E}h^4(Z, G) \ll \mathbf{E}\varphi^2(Z) \ll \overline{\gamma}_{2,2}$. Collecting these estimates, we obtain (8.64), proving (8.61).

To complete the proof of the lemma, we have to prove (8.62). To simplify the notation we shall assume henceforth that $k = 2$. That will not restrict the generality since in the case $k > 2$, more Gaussian random vectors are present in statistics $\mathscr{R}_{[\dots]}tT$ in (8.63). However, while estimating moments, we can use (2.8) and replace them by random vectors distributed as X. Introduce the notation

$$\hat{Z} = \sum_{1 \le j \le N-2} h(Z, X_j),$$
$$W = \sum_{1 \le l < m \le N-2} h(X_l, X_m) + \sum_{1 \le l \le N-2} f(X_l). \tag{8.65}$$

Denote $\hat{Z}(i, j, \dots)$ the sum $\hat{Z}$ without summands with indices $i, j, \dots$. For example, $\hat{Z}(1) = \hat{Z} - h(Z, X_1)$. Notice that the notation (8.65) differs slightly from that used earlier in (8.40). Using (8.65) and replacing

$$X_1, X_2, G_1, G_2, X_3, \dots, X_N, G_3, \dots, G_N$$

by

$$X, \overline{X}, G, \overline{G}, X_1, \dots, X_{N-2}, G_1, \dots, G_{N-2},$$

respectively, we can rewrite J_s, $1 \le s \le 4$, in (8.62) as follows:

$$J_1 = \mathbf{E}\mathrm{e}\Big\{W + f(X) + f(\overline{X}) + \hat{X} + \hat{\overline{X}} + h(X, \overline{X})\Big\},$$
$$J_2 = \mathbf{E}\mathrm{e}\Big\{W + f(G) + f(\overline{X}) + \hat{G} + \hat{\overline{X}} + h(G, \overline{X})\Big\},$$
$$J_3 = \mathbf{E}\mathrm{e}\Big\{W + f(X) + f(\overline{G}) + \hat{X} + \hat{\overline{G}} + h(X, \overline{G})\Big\},$$
$$J_4 = \mathbf{E}\mathrm{e}\Big\{W + f(G) + f(\overline{G}) + \hat{G} + \hat{\overline{G}} + h(G, \overline{G})\Big\}.$$

Expanding in powers of $h(\cdot, \cdot)$ and conditioning, we obtain

$$|J_2 - J_{s0} - J_{s1} - \tfrac{1}{2} J_{s2}| \ll \varkappa \overline{\gamma}_3, \qquad 1 \le s \le 4, \tag{8.66}$$

where, for example,

$$J_{1r} = \mathbf{E}\big(ih(X, \overline{X})\big)^r \mathrm{e}\Big\{W + f(X) + f(\overline{X}) + \hat{X} + \hat{\overline{X}}\Big\},$$

for $0 \le r \le 2$. Writing

$$I_r = J_{1r} - J_{2r} - J_{3r} + J_{4r}, \qquad 0 \le r \le 2,$$

and using (8.66), we have

$$|J_1 - J_2 - J_3 + J_4| \ll |I_0| + |I_1| + |I_2| + \varkappa \overline{\gamma}_3. \tag{8.67}$$

Due to (8.67), instead of (8.62) it suffices to verify that

$$|I_r| \ll \varkappa\Big(\overline{\beta}_3^2 + \overline{\gamma}_3 + N^2 \overline{\gamma}_2 \overline{\gamma}_3 + N \overline{\gamma}_{2,2} + N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}\Big) \quad \text{for } r = 0, 1, 2. \tag{8.68}$$

It is easy to see that

$$|I_r| \ll \Big|\mathbf{E}\mathrm{e}\{W\}\big(a_r(X, \overline{X}) - a_r(G, \overline{X}) - a_r(X, \overline{G}) + a_r(G, \overline{G})\big)\Big|, \qquad r = 0, 1, 2, \tag{8.69}$$

with

$$a_r(Z,Y) = h^r(Z,Y)\mathrm{e}\big\{f(Z) + \hat{Z}\big\}\mathrm{e}\big\{f(Y) + \hat{Y}\big\}. \tag{8.70}$$

Here and below Z denotes one of X or G, and Y stands for $\overline{X}$ or $\overline{G}$.

Let us introduce the notation used throughout the rest of the proof. We shall write

$$\Gamma \sim \big\{\hat{Z}, \hat{Y}; b(Z,Y)\big\} \tag{8.71}$$

if

$$\Gamma \ll \Big|\mathbf{E}\mathrm{e}\{W\}\big(a(X,\overline{X}) - a(G,\overline{X}) - a(X,\overline{G}) + a(G,\overline{G})\big)\Big| + R$$

with some R bounded (up to an absolute constant) by the right-hand side of (8.68) and some function a such that

$$a(Z,Y) = b(Z,Y)\mathrm{e}\{\tau\hat{Z}\}\mathrm{e}\{\bar{\tau}\hat{Y}\},$$

for some independent $0 \le \tau, \bar{\tau} \le 1$ which are independent of all other random vectors. Furthermore we write

$$\Gamma \sim \Gamma_1 \cup \Gamma_2 \quad \text{if } \Gamma \ll \Gamma_1 + \Gamma_2 + R. \tag{8.72}$$

In the notation (8.71), the bound (8.68) means that we have to prove that $I_r \sim \{\cdots; 0\}$.

Let τ and $\bar{\tau}$ be independent random variables uniformly distributed on $[0,1]$. Using expansions with remainders $\mathscr{O}(f^3(Z))$ and $\mathscr{O}(f^3(Y))$, from (8.69) and (8.70) we derive that

$$\begin{aligned} |I_0| &\sim \bigcup_{0 \le k,l \le 3} I_{0,k,l}, \\ I_{0,k,l} &\sim \Big\{\delta_{k,3}\tau f(Z) + \hat{Z},\, \delta_{l,3}\bar{\tau} f(Y) + \hat{Y};\, f^k(Z) f^l(Y)\Big\}, \end{aligned} \tag{8.73}$$

where $\delta_{i,j}$ denotes the Kronecker symbol. Similarly to (8.73), expanding with remainders $\mathscr{O}(f^{3-r}(Z))$ and $\mathscr{O}(f^{3-r}(Y))$ (resp. $\mathscr{O}(f(Z))$ and $\mathscr{O}(f(Y))$), we obtain

$$|I_r| \sim \bigcup_{0 \le k,l \le 3-r} I_{r,k,l}, \qquad r = 1,2, \tag{8.74}$$

where

$$\begin{aligned} I_{r,k,l} \sim \Big\{&\delta_{k,3-r}\tau f(Z) + \hat{Z},\, \delta_{l,3-r}\bar{\tau} f(Y) + \hat{Y}; \\ &f^k(Z) f^l(Y) h^r(Z,Y)\Big\}. \end{aligned} \tag{8.75}$$

Relations (8.68)–(8.75) reduce the proof of the lemma to showing that

$$I_{r,k,l} \sim \{\cdots; 0\}, \tag{8.76}$$

for all allowable values of the indices r, k and l. The proof of (8.76) is rather tedious and its complete exposition is 20 pages long; see preprint BG (1997c), which can be obtained from the authors. (The text is available as well by ftp at `ftp://ftp.uni-bielefeld.de/pub/pages/sfb343/pr97077.ps`. Related papers of the authors are listed in the files `.../index` or `index.html`.) In this paper we estimate only several typical terms which are related to $I_{0,0,0}$; estimation of other terms related to $I_{r,k,l}$ with at least one index nonzero is similar although in some cases it requires rather long and complicated calculations.

PROOF OF (8.76) FOR $I_{0,0,0}$. Expanding in powers of $\hat{Z}$ and $\hat{Y}$ with $Z = X, G$ and $Y = \overline{X}, \overline{G}$ with remainders $\mathscr{O}(\hat{Z}^2)$ and $\mathscr{O}(\hat{Y}^2)$, we obtain

$$I_{0,0,0} \sim \{\hat{Z}, \hat{Y}; \hat{Z}^2\hat{Y}^2\}. \tag{8.77}$$

Write $\hat{Z}^2\hat{Y}^2$ as a fourfold sum. Using (8.77), the i.i.d. assumption of X_j, identities of type (8.44) and classifying summands according to multiplicities of the indices, we obtain

$$I_{0,0,0} \sim \bigcup_{s=1}^{9} \Gamma_s \tag{8.78}$$

with [see (8.43) for the definition of $d_s(Z)$]

$$\Gamma_1 \sim \Big\{\hat{Z}, \hat{Y}; Nh^2(Z, X_1)h^2(Y, X_1)\Big\}, \tag{8.79}$$

$$\begin{aligned}\Gamma_2 &\sim \Big\{\hat{Z}, \hat{Y}; N^2h^2(Z, X_1)h(Y, X_1)h(Y, X_2)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; Nh^2(Z, X_1)h(Y, X_1)d_1(Y)\Big\},\end{aligned} \tag{8.80}$$

$$\begin{aligned}\Gamma_3 &\sim \Big\{\hat{Z}, \hat{Y}; N^2h^2(Y, X_1)h(Z, X_1)h(Z, X_2)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; Nh^2(Y, X_1)h(Z, X_1)d_1(Z)\Big\},\end{aligned} \tag{8.81}$$

$$\Gamma_4 \sim \Big\{\hat{Z}, \hat{Y}; N^2h^2(Z, X_1)h^2(Y, X_2)\Big\}, \tag{8.82}$$

$$\Gamma_5 \sim \Big\{\hat{Z}, \hat{Y}; N^2h(Z, X_1)h(Z, X_2)h(Y, X_1)h(Y, X_2)\Big\}, \tag{8.83}$$

$$\begin{aligned}\Gamma_6 &\sim \Big\{\hat{Z}, \hat{Y}; N^3h^2(Z, X_1)h(Y, X_2)h(Y, X_3)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; Nh^2(Z, X_1)d_1(Y)d_2(Y)\Big\},\end{aligned} \tag{8.84}$$

$$\begin{aligned}\Gamma_7 &\sim \Big\{\hat{Z}, \hat{Y}; N^3h(Z, X_2)h(Z, X_3)h^2(Y, X_1)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; Nd_1(Z)d_2(Z)h^2(Y, X_1)\Big\},\end{aligned} \tag{8.85}$$

$$\begin{aligned}\Gamma_8 &\sim \Big\{\hat{Z}, \hat{Y}; N^3h(Z, X_1)h(Z, X_2)h(Y, X_1)h(Y, X_3)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; Nh(Z, X_1)d_1(Z)h(Y, X_1)d_2(Y)\Big\},\end{aligned} \tag{8.86}$$

$$\begin{aligned}\Gamma_9 &\sim \Big\{\hat{Z}, \hat{Y}; N^4h(Z, X_1)h(Z, X_2)h(Y, X_3)h(Y, X_4)\Big\}\\ &\sim \Big\{\hat{Z}, \hat{Y}; d_1(Z)d_2(Z)d_3(Y)d_4(Y)\Big\}.\end{aligned} \tag{8.87}$$

Let us prove that $\Gamma_1 \sim \{\cdots;0\}$. Using (2.8) and (8.37), we have $\Gamma_1 \sim \{\cdots;0\}$ with an error bounded from above by

$$\varkappa N \mathbf{E} h^2(Z, X_1) h^2(Y, X_1) \ll \varkappa N \mathbf{E} \varphi^2(X_1) = \varkappa N \overline{\gamma}_{2,2}.$$

Below we shall use systematically without referring (8.15), (8.37), (8.38) and (8.46).

Let us prove that $\Gamma_2 \sim \{\cdots;0\}$. Using a Taylor expansion with remainder $\mathscr{O}(|h(Y, X_1)|^{1/2})$ and Hölder's inequality, we obtain

$$\Gamma_2 \sim \left\{\hat{Z}, \hat{Y}(1); N h^2(Z, X_1) h(Y, X_1) d_1(Y)\right\} \tag{8.88}$$

with an error bounded from above by

$$\begin{aligned} &\varkappa N \mathbf{E} h^2(Z, X_1) \left| h(Y, X_1)\right|^{3/2} |d_1(Y)| \\ &\quad \ll \varkappa N^{3/2} \mathbf{E} \varphi(X_1) \varphi^{1/2}(Y) \left| h(Y, X_1)\right|^{3/2} \\ &\quad \ll \varkappa N^{3/2} \left(\overline{\gamma}_2 \overline{\gamma}_3 \overline{\gamma}_{2,2}\right)^{1/2} \le \varkappa \overline{\gamma}_3 + \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.89}$$

Expanding in powers of $\hat{Y}(1)$, using (8.88) and a splitting of $\hat{Y}(1)$ of type (8.43) with d_0 independent of X_1, conditioning and applying $ab \le a^2 + b^2$, we obtain $\Gamma_2 \sim \{\cdots;0\}$ [notice that lower order terms cancel since the covariances of X and G, resp. of $\overline{X}$ and $\overline{G}$, are equal; here the decisive role plays a second order type difference; cf. (8.69) and the notation of (8.71)] with a remainder bounded from above by

$$\begin{aligned} &\varkappa N \max_{0 \le s \le 7} \mathbf{E} h^2(Z, X_1) \left| h(Y, X_1) d_1(Y) d_s(Y)\right| \\ &\quad \ll \varkappa N^2 \mathbf{E} \varphi(X_1) \left| h(Y, X_1)\right| \varphi(Y) \\ &\quad \ll \varkappa N^2 \overline{\gamma}_{2,2} \overline{\gamma}_2^{1/2} \ll \varkappa N \overline{\gamma}_{2,2} + \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.90}$$

Notice that Γ_3 differs from Γ_2 only in that Z and Y are exchanged. Thus, we can apply to Γ_3 the bound for Γ_2.

Estimation of Γ_4 and Γ_5 is similar to that of Γ_2, and we omit related details.

Let us estimate Γ_6. Expanding with remainders $\mathscr{O}(h(Z, X_1))$ and $\mathscr{O}(h(Y, X_1))$ we obtain

$$\Gamma_6 \sim \left\{\hat{Z}(1), \hat{Y}(1); N h^2(Z, X_1) d_1(Y) d_2(Y)\right\} \tag{8.91}$$

with errors bounded by

$$\begin{aligned} &\varkappa N \mathbf{E} h^2(Z, X_1) \left| h(Z, X_1) d_1(Y) d_2(Y)\right| \\ &\qquad \ll \varkappa N^2 \mathbf{E} \left| h(Z, X_1)\right|^3 \varphi(Y) \ll \varkappa N^2 \overline{\gamma}_2 \overline{\gamma}_3, \\ &\varkappa N \mathbf{E} h^2(Z, X_1) \left| h(Y, X_1) d_1(Y) d_2(Y)\right| \\ &\qquad \ll \varkappa N^2 \mathbf{E} \varphi(X_1) \sqrt{h(Y, X_1)}\, \varphi(Y) \ll \varkappa N \overline{\gamma}_{2,2} + \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.92}$$

The i.i.d. assumption and (8.91) yield

$$\Gamma_6 \sim \{\hat{Z}(1), \hat{Y}(1); N^3 h^2(Z, X_1) h(Y, X_2) h(Y, X_3)\}. \tag{8.93}$$

Expanding with remainder $\mathscr{O}(\hat{Z}(1))$, we get

$$\begin{aligned} \Gamma_6 &\sim \{\hat{Z}(1), \hat{Y}(1); N^3 h^2(Z, X_1) h(Y, X_2) h(Y, X_3) \hat{Z}(1)\} \\ &\sim \Gamma_{6,1} \cup \Gamma_{6,2}, \end{aligned} \tag{8.94}$$

where

$$\begin{aligned} \Gamma_{6,1} &\sim \{\hat{Z}(1), \hat{Y}(1); N^3 h^2(Z, X_1) h(Y, X_2) h(Z, X_2) h(Y, X_3)\} \\ &\sim \{\hat{Z}(1), \hat{Y}(1); N^2 h^2(Z, X_1) h(Y, X_2) h(Z, X_2) d_1(Y)\}, \end{aligned} \tag{8.95}$$

$$\begin{aligned} \Gamma_{6,2} &\sim \{\hat{Z}(1), \hat{Y}(1); N^4 h^2(Z, X_1) h(Y, X_2) h(Y, X_3) h(Z, X_4)\} \\ &\sim \{\hat{Z}(1), \hat{Y}(1); N h^2(Z, X_1) d_1(Y) d_2(Y) d_3(Z)\}, \end{aligned} \tag{8.96}$$

Expanding with remainder $\mathscr{O}(|h(Y, X_2)|^{1/2})$, we have

$$\Gamma_{6,1} \sim \{\hat{Z}(1), \hat{Y}(1,2); N^2 h^2(Z, X_1) h(Y, X_2) h(Z, X_2) d_1(Y)\}, \tag{8.97}$$

with an error bounded by

$$\begin{aligned} &\varkappa N^2 \mathbf{E} h^2(Z, X_1) |h(Z, X_2)| |h(Y, X_2)|^{3/2} d_1(Y) \\ &\quad \ll \varkappa N^{5/2} \mathbf{E} \varphi(Z) |h(Y, X_2)|^{3/2} |h(Z, X_2)| \varphi^{1/2}(Y) \\ &\quad \ll \varkappa N^{5/2} \overline{\gamma}_{2,2}^{1/2} \overline{\gamma}_3^{1/2} \overline{\gamma}_2 \ll \varkappa N^2 \overline{\gamma}_2 \overline{\gamma}_3 + \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.98}$$

Using (8.97) and expanding with remainder $\mathscr{O}(\hat{Y}(1,2))$, we have $\Gamma_{6,1} \sim \{\cdots; 0\}$ with a remainder bounded by

$$\begin{aligned} &\varkappa N^2 \mathbf{E} h^2(Z, X_1) |h(Z, X_2) h(Y, X_2) d_1(Y) d_s(Y)| \\ &\quad \ll \varkappa N^3 \mathbf{E} \varphi(Z) |h(Y, X_2)| |h(Z, X_2)| \varphi(Y) \ll \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned} \tag{8.99}$$

Let us estimate $\Gamma_{6,2}$ defined by (8.96). Expanding with remainder $\mathscr{O}(\hat{Z}(1))$, we obtain $\Gamma_{6,2} \sim \{\cdots; 0\}$ with an error bounded by

$$\begin{aligned} &\varkappa N \mathbf{E} h^2(Z, X_1) |d_s(Z) d_1(Y) d_2(Y) d_3(Y)| \\ &\quad \ll \varkappa N^3 \mathbf{E} \varphi^{3/2}(Z) \varphi^{3/2}(Y) \ll \varkappa N^3 \overline{\gamma}_2, \overline{\gamma}_{2,2}. \end{aligned}$$

The estimation of Γ_7 is similar to that of Γ_6 since Z and Y are just exchanged.

Estimation of Γ_8 recalls estimation of Γ_2, and we omit related details.

Let us show that $\Gamma_9 \sim \{\cdots; 0\}$. Expanding in powers of $\hat{Z}$ with remainder $\mathscr{O}(\hat{Z})$ and using the i.i.d. assumption, we have [cf. (8.93)–(8.96)]

$$\Gamma_9 \ll \Gamma_{9,1} \cup \Gamma_{9,2} \cup \Gamma_{9,3} \tag{8.100}$$

with

$$\Gamma_{9,1} \sim \Big\{\hat Z, \hat Y; d_1(Z)d_2(Z)d_3(Y)d_4(Y)d_5(Z)\Big\}, \tag{8.101}$$

$$\Gamma_{9,2} \sim \Big\{\hat Z, \hat Y; Nh^2(Z,X_1)d_1(Z)d_2(Y)d_3(Y)\Big\}, \tag{8.102}$$

$$\Gamma_{9,3} \sim \Big\{\hat Z, \hat Y; Nh(Y,X_1)h(Z,X_1)d_1(Z)d_2(Z)d_3(Y)\Big\}. \tag{8.103}$$

Expanding in powers of $\hat Y$ with remainder $\mathscr{O}(\hat Y)$ we get $\Gamma_{9,1} \sim \{\cdots;0\}$ with an error bounded by

$$\begin{aligned} &\varkappa \mathbf{E}|d_1(Z)d_2(Z)d_3(Y)d_4(Y)d_5(Z)d_s(Y)| \\ &\qquad \ll \varkappa N^3 \mathbf{E}\varphi^{3/2}(Z)\varphi^{3/2}(Y) \ll \varkappa N^3 \overline{\gamma}_2 \overline{\gamma}_{2,2}. \end{aligned}$$

In order to estimate $\Gamma_{9,s}$, for $s = 2, 3$, we first remove $h(Y, X_1)$ from the exponents. Expansions with remainders $\mathscr{O}(|h(Y,X_1)|^{1/2})$ produce errors bounded by

$$\begin{aligned} &\varkappa N \mathbf{E} h^2(Z,X_1)|h(Y,X_1)|^{1/2}|d_1(Z)d_2(Y)d_3(Y)| \\ &\quad \ll \varkappa N^{5/2}\mathbf{E}\varphi(Y)\mathbf{E}^Y|h(Z,X_1)|^{3/2} \\ &\qquad \times\big(|h(Z,X_1)h(Y,X_1)|\varphi(Z)\big)^{1/2} \\ &\quad \ll \varkappa N^{5/2}\mathbf{E}\varphi(Y)\overline{\gamma}_3^{1/2}\big(\mathbf{E}^Y|h(Z,X_1)h(Y,X_1)|\varphi(Z)\big)^{1/2} \\ &\quad \ll \varkappa N^{5/2}\mathbf{E}\varphi(Y)\overline{\gamma}_3^{1/2}\big(\mathbf{E}^Y\varphi^{1/2}(Y)\varphi^{3/2}(Z)\big)^{1/2} \\ &\quad = \varkappa N^{5/2}\mathbf{E}\varphi^{5/4}(Y)\overline{\gamma}_3^{1/2}\big(\mathbf{E}\varphi^{3/2}(Z)\big)^{1/2} \\ &\quad \ll \varkappa N^{5/2}\overline{\gamma}_2\overline{\gamma}_{2,2}^{1/2}\,\overline{\gamma}_3^{1/2} \\ &\quad \ll \varkappa N^3\overline{\gamma}_2\overline{\gamma}_{2,2} + \varkappa N^2\overline{\gamma}_2\overline{\gamma}_3 \end{aligned} \tag{8.104}$$

and

$$\begin{aligned} &\varkappa N \mathbf{E}|h(Z,X_1)||h(Y,X_1)|^{3/2}|d_1(Z)d_2(Z)d_3(Y)| \\ &\quad \ll \varkappa N^{5/2}\mathbf{E}\varphi^{1/2}(Z)\Big(\mathbf{E}^Y|h(Y,X_1)|^3\Big)^{1/2}\varphi(Z)\varphi^{1/2}(Y) \\ &\quad \ll \varkappa N^{5/2}\mathbf{E}\varphi^{3/2}(Z)\overline{\gamma}_2^{1/2}\,\overline{\gamma}_3^{1/2} \ll \varkappa N^3\overline{\gamma}_2\overline{\gamma}_{2,2} + \varkappa N^2\overline{\gamma}_2\overline{\gamma}_3. \end{aligned} \tag{8.105}$$

Once terms depending on Y and X_1 are removed from the exponents, we can expand in powers of $\hat Y(1)$. Using splittings and analyzing various cases, we obtain $\Gamma_{9,s} \sim \{\cdots;0\}$ with remainders bounded by

$$\varkappa N^3\mathbf{E}\varphi^{3/2}(Z)\varphi^{3/2}(Y) \ll \varkappa N^3\overline{\gamma}_2\overline{\gamma}_{2,2}, \tag{8.106}$$

completing the estimation of $I_{0,0,0}$. □

LEMMA 8.5. *The Edgeworth correction F_1 satisfies*

$$\big|\hat{F}_1(t)\big| \ll N^{-1/2}|t|^3\big(\beta_3^2 + \sigma^2\gamma_{2,2}\big)^{1/2} \prod_{j\geq 1}\big(1 + 2t^2q_j^2/25\big)^{-1/4}. \tag{8.107}$$

Furthermore, for $s \geq 7$ we have

$$\int_{|t|\geq\lambda}\big|\hat{F}_1(t)\big|\frac{dt}{|t|} \ll_s N^{-1/2}\big(\beta_3^2 + \sigma^2\gamma_{2,2}\big)^{1/2}|q_s|^{-s/2}\lambda^{3-s/2} \quad \textit{for } \lambda > 0, \tag{8.108}$$

$$\int_{\mathbb{R}}\big|\hat{F}_1(t)\big|\frac{dt}{|t|} \ll_s N^{-1/2}\big(\beta_3^2 + \sigma^2\gamma_{2,2}\big)^{1/2}|q_s|^{-3}. \tag{8.109}$$

PROOF. We can write $G =_{\mathscr{D}} (G_1 + \cdots + G_5)/\sqrt{5}$. Using the linearity of ϕ_1, the relation (2.11) with $N = 5$ and (2.12), we have

$$T_0 =_{\mathscr{D}} \frac{1}{5}\sum_{1\leq i<j\leq 5}\phi(G_i, G_j) + \sum_{j=1}^{5} f_j(G_j) \tag{8.110}$$

with some statistics f_j. Using the explicit formula (4.1) for $\hat{F}_1(t)$, sorting the terms of the statistic (8.110) and conditioning [cf., for example, the proof of (8.23)], we obtain

$$\big|\hat{F}_1(t)\big| \ll N^{-1/2}|t|^3\big(\beta_3 + \gamma_{2,3/2}\big)\chi(t), \tag{8.111}$$

where

$$\chi(t) = \sup_{f_1, f_2}\big|\mathbf{E}e\{t\phi(G_1, G_2)/5 + f_1(G_1) + f_2(G_2)\}\big|. \tag{8.112}$$

Using $G_1 - G_1 =_{\mathscr{D}} \sqrt{2}\,G$ and $\mathbf{E}e\{\phi(x, G_2)\} \geq 0$, we obtain

$$\begin{aligned} \chi^2(t) &\leq \sup_{f_1, f_2}\mathbf{E}\big|\mathbf{E}_{G_1}e\{t\phi(G_1, G_2)/5 + f_1(G_1)\}\big|^2 \\ &= \sup_{f_1}\mathbf{E}e\big\{\sqrt{2}\,t\phi(G, G_2)/5 + f_1(G_1) - f_1\big(\overline{G}_1\big)\big\} \\ &\leq \mathbf{E}\big|\mathbf{E}_{G_2}e\big\{\sqrt{2}\,t\phi(G, G_2)/5\big\}\big| = \mathbf{E}e\big\{\sqrt{2}\,t\phi(G, \overline{G})/5\big\}. \end{aligned} \tag{8.113}$$

Recall that $\phi(x, y) = \sum_{j\geq 1}q_j x_j y_j$ [see (2.6)]. Hence

$$\mathbf{E}e\big\{\theta\phi(G, \overline{G})\big\} = \prod_{j\geq 1}\big(1 + \theta^2q_j^2\big)^{-1/2}, \qquad \theta \in \mathbb{R}. \tag{8.114}$$

Estimating $\gamma_{2,3/2}^2 \ll \sigma^2\gamma_{2,2}$ and collecting (8.111)–(8.114), we obtain (8.107).

Using (8.107) and estimating $1 + 2t^2q_j^2/25 \geq 1$, for $j > s$, and $q_s^2t^2 \ll 1 + 2t^2q_j^2/25$, for $j \leq s$, we obtain (8.108). For the proof of (8.109) we use $1 + q_s^2t^2 \ll 1 + 2t^2q_j^2/25$, for $j \leq s$. □

9. An extension of bounds to von Mises statistics. Applications. Assuming that the kernels ϕ and ϕ_1 are degenerate, consider the von Mises statistic

$$M = \frac{1}{2N} \sum_{1 \le i, j \le N} \phi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \phi_1(X_i). \tag{9.1}$$

Introducing the function $\psi(x) = (\phi(x, x) - \nu)/2$ with $\nu = \mathbf{E}\phi(X, X)$, we can rewrite (9.1) as

$$M - \frac{\nu}{2} = T + \frac{1}{N} \sum_{1 \le i \le N} \psi(X_i) \tag{9.2}$$

with T defined by (1.1). In this section we shall extend the bounds to statistics of type (9.2), assuming that $\mathbf{E}\psi(X) = 0$ and $\varrho = \mathbf{E}\psi^2(X) < \infty$.

Similarly to the case of T, we can represent the kernel ϕ (respectively, ϕ_1 and ψ) as a bilinear (respectively, linear) function defined on $\mathbb{R}^\infty$. However in this case we have to assume that $\mathbb{R}^\infty$ has an additional coordinate since ψ can be linearly independent of ϕ_1 and of the eigenfunctions of $\mathbb{Q}$. To fix notation, we shall assume that $\mathbb{R}^\infty$ consists of vectors $x = (x_{-1}, x_0, x_1, \dots)$. Then all representations and results of Section 2 concerning ϕ and ϕ_1 still hold, and for ψ we have $\psi(x) = \langle b, x \rangle$ with some $b = (b_{-1}, b_0, b_1, \dots)$ such that $\sum_{j \ge -1} b_j^2 < \infty$. Write $\psi_0(x) = \sum_{j \ge 0} b_j x_j$.

Introduce the function F_* of bounded variation (provided that $q_3 \ne 0$ with the Fourier–Stieltjes transform

$$\hat{F}_*(t) = \frac{it}{\sqrt{N}} \mathbf{E}\psi(G)\mathrm{e}\{tT_0\} = \frac{it}{\sqrt{N}} \mathbf{E}\psi_0(G)\mathrm{e}\{tT_0\}$$

and such that $F_*(-\infty) = 0$. Below we shall show that (see Lemma 9.3)

$$\hat{F}_*(t) = \frac{(it)^2}{\sqrt{N}} \mathbf{E}\psi(X)(\phi_1(X) + \phi(X, G))\mathrm{e}\{tT_0\}. \tag{9.3}$$

Notice that $F_* = 0$ whenever $\phi_1 = 0$.

Write $H_1 = F_1 + F_*$, and let H denote the distribution function of $M - \nu/2$. Define

$$\delta_N = \sup_x |\delta_N(x)|, \qquad \delta_N(x) = H(x) - F_0(x) - H_1(x).$$

THEOREM 9.1.

(i) *Assume that* $q_{13} \ne 0$. *Then we have*

$$\delta_N \le \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\beta_3^2}{\sigma^6} + \frac{\gamma_3}{\sigma^3} + \frac{\gamma_{2,2}}{\sigma^4} + \frac{\varrho}{\sigma^2} \right) \quad \text{where } C \le \exp\left\{ \frac{c\sigma}{|q_{13}|} \right\}. \tag{9.4}$$

(ii) *Assume that* (1.9) *is fulfilled and* $q_9 \ne 0$. *Then* (9.4) *holds with* $C \le \exp\{c\sigma/|q9|\}$.

PROOF. We shall use the following estimates. Write

$$\xi = \frac{t}{N}\sum_{1\le j\le N}\psi(X_j), \qquad \zeta = \frac{t}{N}\sum_{1\le j\le N}\psi(G_j). \tag{9.5}$$

Expanding with remainder $\mathscr{O}(\xi)$, splitting the sum ξ in parts and conditioning [cf. the proof of (8.32)], we have

$$\left|\mathbf{E}e\{tT+\xi\} - \mathbf{E}e\{tT\} - i\mathbf{E}\xi e\{tT\}\right| \ll \varkappa N^{-1}t^2\varrho. \tag{9.6}$$

Proceeding similarly to the proof of Lemma 8.2, we obtain

$$\left|\hat{F}_*(t) - i\mathbf{E}\zeta e\{\mathscr{R}_{[1,N]}T\}\right| \ll \varkappa N^{-1}t^2(\varrho+\sigma^2). \tag{9.7}$$

Applying the Bergström-type identity (8.7) with $S = \xi e\{tT\}$ and proceeding similarly to the proof of Lemma 8.3, we get (we omit cumbersome technical details)

$$\begin{aligned}&\left|\mathbf{E}\xi e\{tT\} - \mathbf{E}\zeta e\{\mathscr{R}_{[1,N]}T\}\right|\\ &\qquad \ll \varkappa N^{-1}(t^2+t^4)\left(\varrho+\beta_4+\beta_3^2+\gamma_3+\gamma_{2,2}+\gamma_2\Gamma_{2,2}\right).\end{aligned} \tag{9.8}$$

Arguments similar to the proof of Lemma 8.5 allow proving

$$\left|\hat{F}_*(t)\right| \ll N^{-1.2}|t|\varrho^{1/2}\prod_{j\ge 1}\left(1+2t^2q_j^2/25\right)^{-1/4}, \tag{9.9}$$

and, for $s \ge 3$,

$$\int_{|t|\ge\lambda}\left|\hat{F}_*(t)\right|\frac{dt}{|t|} \ll_s N^{-1/2}\varrho^{1/2}|q_s|^{-s/2}\lambda^{1-s/2}, \qquad \lambda > 0, \tag{9.10}$$

$$\int_{\mathbb{R}}\left|\hat{F}_*(t)\right|\frac{dt}{|t|} \ll_s N^{-1/2}\varrho^{1/2}|q_s|^{-1}. \tag{9.11}$$

The estimates (9.6)–(9.11) allow proceeding similarly to the proof of Theorem 1.1, using a lemma similar to Lemma 4.1. Proving such a lemma, we have to apply Lemma 4.2 to the distribution function H. This is possible since that statistic $M - \nu/2$ is a statistic of type (4.8). The estimates (9.10) and (9.11) allow application of the Fourier inversion (4.14) to the function F_*. As a result [cf. integral J in (4.18)], we arrive at

$$\int_{-Nt_*}^{Nt_*}\left|\hat{H}(t) - \hat{F}_0(t) - \hat{H}_1(t)\right|\frac{dt}{|t|}.$$

Here, however, we have $\hat{H}(t) = \mathbf{E}e\{tT+\xi\}$, and

$$\begin{aligned}&\left|\hat{H}(t) - \hat{F}_0(t) - \hat{H}_1(t)\right|\\ &\quad\le\left|\hat{F}(t) - \hat{F}_0(t) - \hat{F}_1(t)\right|\\ &\qquad+\left|\mathbf{E}e\{tT+\xi\} - \mathbf{E}e\{tT\} - i\mathbf{E}\xi e\{tT\}\right|\\ &\qquad+\left|\hat{F}_*(t) - i\mathbf{E}\zeta e\{\mathscr{R}_{[1,N]}T\}\right|\\ &\qquad+\left|\mathbf{E}\xi e\{tT\} - \mathbf{E}\zeta e\{\mathscr{R}_{[1,N]}T\}\right|.\end{aligned} \tag{9.12}$$

Therefore, using (9.6)–(9.8), we can proceed as in the proof of Lemma 4.1. As a final result we get bounds similar to those of Theorem 1.1, with the additional summand ϱ/σ^2. □

THEOREM 9.2. *Assume that $q_{d+1} = 0$ and $q_d \neq 0$, for some $d \geq 9$, and that the weights a_j in* (1.6) *satisfy $a_0 = 0$ and $a_j = 0$, for $j > d$. Then we have*

$$(9.13)\quad \Delta_N \ll_d \frac{C}{N}\left(\frac{\beta_4}{\sigma^4} + \frac{\beta_3^2}{\sigma^6} + \frac{\gamma_{2,2}}{\sigma^4} + \frac{\beta_2^{3/2}\gamma_{2,2}}{\sigma^7}\right) \quad \textit{where } C \leq \exp\left\{\frac{c\sigma}{|q_d|}\right\}.$$

PROOF. The assumptions of the theorem allow assuming that the random vectors X and G take values in $\mathbb{R}^d$, have mean zero and identity covariances, and that

$$\phi(x, y) = \langle \mathbb{Q}x, y\rangle = \sum_{j=1}^{d} q_j x_j y_j, \qquad \phi_1(x) = \langle \mathbb{Q}x, a\rangle = \sum_{j=1}^{d} a_j x_j,$$

where $\mathbb{Q}$: $\mathbb{R}^d \to \mathbb{R}^d$ is a linear diagonal operator with eigenvalues q_j. Introducing $S_N = N^{-1/2}(X_1 + \cdots + X_N)$, we can write

$$T = T_* - \frac{1}{N}\sum_{j=1}^{N} \psi(X_j), \qquad 2T_* = \langle \mathbb{Q}(S_N - v), (S_N - v)\rangle - \langle \mathbb{Q}u, v\rangle - \nu,$$

where $2\psi(x) = \langle \mathbb{Q}x, x\rangle - \nu$, $v = \mathbb{Q}^{-1}a$ and $\nu = \sum_{j=1}^d q_j$. Similarly,

$$2T_0 = \langle \mathbb{Q}(G - v), (G - v)\rangle - \langle \mathbb{Q}v, v\rangle - \nu.$$

We have $\Delta_N \leq \Delta_{N,1} + \Delta_{N,2}$ where

$$\Delta_{N,1} = \sup_x \left|\mathbf{P}\{T_* \leq x\} - \mathbf{P}\{T_0 \leq x\} - F_*(x) - F_1(x)\right|,$$

$$\Delta_{N,2} = \sup_x \left|\mathbf{P}\{T \leq x\} - \mathbf{P}\{T_* \leq x\} + F_*(x)\right|.$$

To estimate $\Delta_{N,1}$ we shall apply Theorem 1.5 in BG (1997b). Let $\mathbb{C}$ be the diagonal operator with eigenvalues $|q_j|$. Then

$$(9.14)\quad \Delta_{N,1} \ll_d CN^{-1}\theta^{-4}\mathbf{E}|\mathbb{C}^{1/2}X|^4(1 + \theta^{-3}|\mathbb{C}^{1/2}v|^3), \qquad C \leq \exp\left\{\frac{c\theta^2}{|q_d|}\right\}.$$

where $\theta^2 = \sum_{j=1}^d |q_j|$. However, using

$$\theta^4 \asymp_d \sigma^2, \qquad |\mathbb{C}^{1/2}X|^2 \leq |q_d|^{-1}\mathbf{E}_{\overline{X}}\phi^2(X, \overline{X}), \qquad |\mathbb{C}^{1/2}v|^2 \leq |q_d|^{-1}\beta_2$$

and (9.14), we see that $\Delta_{N,1}$ is bounded from above by the right-hand side of (9.13).

The estimation of of $\Delta_{N,2}$ is similar to the proof of Theorem 9.1. Using the notation (9.5) and writing $tT_* = tT + \xi$ we reduce the estimation of $\Delta_{N,2}$ to the estimation of $\Delta_{N,2} = |\mathbf{E}e\{tT\} - \mathbf{E}e\{tT + \xi\} + \hat{F}_*(t)|$. It is easy to see that $\hat{\Delta}_{N,2}$ is bounded from above by the right-hand side of (9.12). Thus, using $\varrho \leq q_d^{-2}\gamma_{2,2} \leq C\gamma_{2,2}/\sigma^2$, we see that $\Delta_{N,2}$ can be estimated by the right-hand side of (9.13). □

LEMMA 9.3. *Let $f\colon \mathbb{R} \to \mathbb{C}$ be a bounded infinitely many times differentiable function with bounded derivatives. Then we have*

$$\mathbf{E}\psi(G)f(T_0) = \mathbf{E}\psi(X)\big(\phi_1(X) + \phi(X,G)\big)f'(T_0) \tag{9.15}$$

provided that $\beta_2 < \infty$, $\varrho < \infty$ and $\gamma_{2,2} < \infty$.

PROOF. Using $G =_{\mathscr{D}} N^{-1/2}(G_1 + \cdots + G_N)$, the i.i.d. assumption, representations (2.12) and (2.9), the linearaty of ψ and the symmetry of T_0, we obtain

$$\mathbf{E}\psi(G)f(T_0) = \sqrt{N}\,\mathbf{E}\psi(G_1)f\big((2N^{-1})\phi_0(G_1) + L(G_1) + W\big), \tag{9.16}$$

where $L(Z) = N^{-1/2}\phi(Z) + N^{-1}\sum_{j=2}^{N}\phi(Z,G_j)$, and the statistic W is defined as the right-hand side of (2.12), just omitting summands depending on G_1. Expanding in (9.16) with remainders $\mathscr{O}((2N)^{-1}\phi_0(G_1))$ and $\mathscr{O}(L^2(G_1))$, we get

$$\mathbf{E}\psi(G)f(T_0) = \sqrt{N}\,\mathbf{E}\psi(G_1)L(G_1)f'(W) + \mathscr{O}(N^{-1/2}) \tag{9.17}$$

since $\mathbf{E}\psi(G_1) = 0$. The equality of the covariances of G_1 and X yields

$$\mathbf{E}\psi(G_1)L(G_1)f'(W) = \mathbf{E}\psi(X)L(X)f'(W). \tag{9.18}$$

Writing now $W = T_0 - L(G_1) - (2N)^{-1}\phi_0(G_1)$, expanding with remainders $\mathscr{O}(L(G_1))$ and $\mathscr{O}((2N)^{-1}\phi_0(G_1))$ and using $L(X) =_{\mathscr{D}} N^{-1/2}\phi_1(X) + N^{-1}\sqrt{N-1}\,\phi(X,G)$, we obtain

$$\begin{aligned}&\sqrt{N}\,\mathbf{E}\psi(X)L(X)f'(W)\\ &\qquad = \mathbf{E}\psi(X)\big(\phi_1(X) + \phi(X,G)\big)f'(T_0) + \mathscr{O}(N^{-1/2}).\end{aligned} \tag{9.19}$$

Collecting (9.16)–(9.19) and passing to the limit as $N \to \infty$, we get (9.15). □

Estimates of the convergence rate in the CLT for conic sections in $\mathbb{R}^d$ and Hilbert spaces. Let $\mathbb{R}^d$ denote the standard d-dimensional Euclidean space with scalar product $\langle x, y\rangle$ and norm $|x|^2 = \langle x, x\rangle$. In the case $d = \infty$ in this subsection, we shall understand $\mathbb{R}^d$ as the Hilbert space l_2. Consider i.i.d. random vectors $Y, Y_1, \ldots, Y_N$ taking values in $\mathbb{R}^d$ such that $\mathbf{E}Y = 0$. Write $S_N = N^{-1/2}(Y_1 + \cdots + Y_N)$. Let $\mathbb{A}\colon \mathbb{R}^d \to \mathbb{R}^d$ be a symmetric bounded linear operator. Introduce the quadratic form $\mathbb{A}[x] = \langle \mathbb{A}x, x\rangle$. We are interested in approximations of the probability $\mathbf{P}\{\mathbb{A}[S_N - w] \le x\}$, $w \in \mathbb{R}^d$. Without loss of generality, we can replace the aforementioned probability by $\mathbf{P}\{M - \nu/2 \le x\}$, where $M - \nu/2 = \mathbb{A}[S_N] - 2\langle \mathbb{A}S_N, w\rangle - \nu$ is the statistic (9.2) with

$$\phi(x,y) = 2\langle \mathbb{A}x, y\rangle, \quad \phi_1(x) = -2\langle \mathbb{A}x, w\rangle,$$
$$\psi(x) = \langle \mathbb{A}x, x\rangle - \nu, \quad \nu = \mathbf{E}\langle \mathbb{A}Y, Y\rangle.$$

We shall use the following fact: for any bounded linear operators $\mathbb{A}, \mathbb{B}\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\mathbb{A} \ge 0$, the collections of nonzero eigenvalues of $\mathbb{A}\mathbb{B}$, $\mathbb{B}\mathbb{A}$ and $\sqrt{\mathbb{A}}\,\mathbb{B}\sqrt{\mathbb{A}}$ coincide.

Introducing the covariance operator $\mathbb{C}$ of Y, we have $\nu = \operatorname{Tr}\mathbb{C}^{1/2}\mathbb{A}\mathbb{C}^{1/2} = \operatorname{Tr}\mathbb{A}\mathbb{C}$.

Let u_j, $j \geq 1$, be a complete orthonormal system of eigenvectors of the operator $2\mathbb{C}^{1/2}\mathbb{A}\mathbb{C}^{1/2}$ with the corresponding eigenvalues q_j. Consider the system of functions

$$e_j(x) = \langle \mathbb{C}^{-1/2}u_j, x\rangle = \frac{2}{q_j}\langle \mathbb{A}\mathbb{C}^{1/2}u_j, x\rangle, \qquad q_j \neq 0,$$

in $L^2(\mathbb{R}^d, \mathscr{L}(Y))$. It is easy to verify that this system is an orthonormal system of eigenfunctions of the operator $\mathbb{Q}$ associated with the kernel $\phi(x, y)$ in $L^2(\mathbb{R}^d, \mathscr{L}(Y))$ and that the corresponding eigenvalues are q_j. Moreover, we have $\sigma^2 = 4\,\mathrm{Tr}(\mathbb{A}\mathbb{C})^2 = \sum_{j\geq 1} q_j^2$.

Obviously,

$$\gamma_s \ll_s \mathbf{E}|\langle \mathbb{A}Y, \overline{Y}\rangle|^s, \qquad \gamma_{2,2} \ll \mathbf{E}\langle \mathbb{A}\mathbb{C}\mathbb{A}Y, Y\rangle^2,$$
$$\beta_s \ll_s \mathbf{E}|\langle \mathbb{A}Y, w\rangle|^s, \qquad \varrho \ll \mathbf{E}\langle \mathbb{A}Y, Y\rangle^2.$$

Thus we can apply to δ_N the bounds of Theorem 9.1. In particular, if $\mathbb{A} = \mathbb{I}$ then we have

$$\delta_N \ll \frac{C}{N}\left(\frac{\mathbf{E}|Y|^4}{\sigma^2} + \frac{\left(|\mathbf{E}\langle Y, w\rangle|^3\right)^2}{\sigma^6} + \frac{\mathbf{E}\langle Y, w\rangle^4}{\sigma^4}\right)$$

with C as in Theorem 9.1 and q_s being the eigenvalues of $2\mathbb{C}$. Comparing with the bounds in BG (1997b), the bounds for δ_N implied by Theorem 9.1 show an improved dependence on w and moments.

Introduce a centered Gaussian random vector Z taking values in $\mathbb{R}^d$ and such that the covariances of Y and Z are equal. Then we can write

$$T_0 =_{\mathscr{D}} M_0 = \langle \mathbb{A}Z, Z\rangle - \langle \mathbb{A}Z, w\rangle - \nu.$$

Furthermore, we have $H_1 = F_1 + F_*$, where [using (9.3)]

$$\hat{F}_1(t) = \frac{4(it)^3}{3\sqrt{N}}\mathbf{E}\langle \mathbb{A}Y, Z - w\rangle^3 \mathrm{e}\{tM_0\},$$

$$\hat{F}_*(t) = \frac{2(it)^2}{\sqrt{N}}\mathbf{E}\langle \mathbb{A}Y, Y\rangle\langle \mathbb{A}Y, Z - w\rangle \mathrm{e}\{tT_0\}.$$

Approximations of multiple Wiener–Itô integrals. Consider a probability distribution μ on a measurable space $(\mathscr{X}, \mathscr{B})$. Let $I_n(f)$ be the n-fold multiple Wiener–Itô integral of a symmetric function $f\colon \mathscr{X}^n \to \mathbf{R}$ with respect to an orthogonal Gaussian random measure with control measure μ [see Itô (1951), Major (1981)]. The Itô formula yields

$$\tfrac{1}{2}I_2(\phi) + I_1(\phi_1) =_{\mathscr{D}} \tfrac{1}{2}\sum_{j\geq 1} q_j\left(\eta_j^2 - 1\right) + \sum_{j\geq 0} a_j\eta_j = T_0.$$

Thus, assuming that the sample $X_1, \dots, X_N$ is taken from the distribution μ, that is, $\mathscr{L}(X) = \mu$, we can use the statistic T defined by (1.1) to approximate the random variable $\frac{1}{2}I_2(\phi) + I_1(\phi_1)$ in distribution. In particular, Theorem 1.1. shows that approximating $I_2(\phi)$ by a corresponding statistic T of type (1.1), we get an error of the order $\mathscr{O}(N^{-1})$, provided that $\phi \in L^4$ and $q_{13} \neq 0$.

EXAMPLE 9.4. Let X denote a random variable uniformly distributed on $\mathscr{X} = [0, 1]$. For $0 \le x \le 1$ and $\alpha < 1$, consider the functions

$$f(x, y) = |x - y|^{-\alpha},$$
$$f_1(x) = \mathbf{E} f(x, X) = (1 - \alpha)^{-1}\bigl(x^{1-\alpha} + (1 - x)^{1-\alpha}\bigr).$$

Introduce the degenerate kernel $\phi(x, y) = f(x, y) - f_1(x) - f_1(y) + \mathbf{E} f(X, \overline{X})$. Since the function f_1 is bounded, we have

$$\gamma_s < \infty \Leftrightarrow \int_0^1\int_0^1 |x - y|^{-s\alpha}\, dx\, dy < \infty \Leftrightarrow \alpha < \frac{1}{s},$$

$$\gamma_{2,2} < \infty \Leftrightarrow \int_0^1\biggl(\int_0^1 |x - y|^{-2\alpha}\, dx\biggr)^2 dy < \infty \Leftrightarrow \alpha < \tfrac{1}{2}.$$

Consequently, the bounds of Theorem 1.1 are applicable when $\alpha < 1/3$, whereas the bounds (1.12) depending on γ_4 are applicable only when $\alpha < 1/4$.

10. An integration procedure.

Recall that [see (2.13)]

$$\mathscr{M}(t; N) = (|t|N)^{-1/2}\mathbf{I}\{|t| \le N^{-1/2}\} + |t|^{1/2}\mathbf{I}\{|t| > N^{-1/2}\}, \tag{10.1}$$

where $N > 0$ will be a positive large parameter.

For a number $A \ge 1$ and a family of functions $\varphi(\cdot) = \varphi(\cdot; N)\colon \mathbb{R} \to \mathbb{R}$, introduce

$$\gamma = \gamma(N, A) =_{\text{def}} \sup\bigl\{|\varphi(t)|\colon N^{-1/2} \le t \le A\bigr\} \tag{10.2}$$

The following Theorem 10.1 sharpens Theorem 6.1 of BG (1997) in cases where $\gamma < 1$

THEOREM 10.1. *Let $\varphi(t)$, $t \ge 0$ denote a continuous function such that $\varphi(0) = 1$ and $0 \le \varphi \le 1$. Assume that, for some $s > 8$ and $\Theta \ge 1$,*

$$\varphi(t)\varphi(t + \tau) \le \Theta\mathscr{M}^s(\tau; N) \quad \text{for all } t \ge 0 \text{ and } \tau \ge 0. \tag{10.3}$$

Let $A \ge 1$. Assume that the number γ defined by (10.2) *satisfies*

$$\gamma > 4^{s/(s-8)}N^{-s/8} \quad \text{if } -1 < \alpha \le 0,$$

$$\gamma\biggl(1 + \log\frac{1}{\gamma}\biggr)^{s/(s-8)} > 4^{s/(s-8)}N^{-s/8}(1 + \log N)^{s/(s-8)} \quad \text{if } \alpha = -1. \tag{10.4}$$

Then the integral

$$J = \int_{N^{-1/2}}^{A}\varphi(t)t^{\alpha}\, dt$$

can be bounded as follows:

$$J \ll_{\alpha,s} \left(\frac{\gamma}{\Theta}\right)^{1-8/s} A^{\alpha+1}\frac{\Theta}{N} \quad \textit{for}\ -1<\alpha\le 0 \tag{10.5}$$

and

$$J \ll_{\alpha,s} \left(\frac{\gamma}{\Theta}\right)^{1-8/s}\left(1+\log\frac{\Theta}{\gamma}\right)(1+\log A)\frac{\Theta}{N} \quad \textit{for}\ \alpha=-1.$$

If (10.4) *is not fulfilled then*

$$\begin{aligned} &J \ll_{s} \gamma(1+\log N)(1+\log A) \quad \textit{for}\ \alpha=-1,\\ &J \ll_{\alpha,s} \gamma A^{\alpha+1} \quad \textit{for}\ \alpha>-1. \end{aligned} \tag{10.6}$$

PROOF. The proof is similar to Lebesgue integration by partitioning the range of φ in intervals $[2^{-l-1}, 2^{-l}]$. We have to estimate the Lebesgue measure of the corresponding sets $B_l = \{t\colon 2^{-l-1} \le \varphi(t) \le 2^{-l}\} \cap [N^{-1/2}, A]$. Using inequality (10.3), we shall show that two points in B_l are either very "close" or far apart. This means that the set B_l consists of "small" clusters of size $\mathscr{O}(N^{-1})$ separated by "large" gaps of size $\mathscr{O}_l(1)$. These constraints on the structure of B_l suffice to bound the measure of B_l well enough in order to estimate the size of J for $s>8$ as claimed in Theorem 10.1.

Throughout the proof we shall write $\ll$ instead of $\ll_{\alpha,s}$.

To prove (10.6) it suffices to use $\varphi(t)\le\gamma$ and to notice that

$$\int_{N^{-1/2}}^{A}\frac{dt}{t} \ll (1+\log N)(1+\log A),$$

$$\int_{N^{-1/2}}^{A} t^{\alpha}\,dt \ll A^{\alpha+1} \quad \text{for}\ \alpha>-1.$$

Let us prove (10.5). Inequality (10.3) implies that (set $t=0$, use $\varphi(0)=1$ and note that $\Theta\ge 1$)

$$\varphi(t)\le \Theta\mathscr{M}^s(t;N) \quad \text{and} \quad \varphi(t)\varphi(t+\tau)\le \Theta^2\mathscr{M}^s(\tau;N). \tag{10.7}$$

Starting the proof of (10.5) with (10.7), we may assume without loss of generality that $\Theta=1$, that is,

$$\varphi(t)\le \mathscr{M}^s(t;N) \quad \text{and} \quad \varphi(t)\varphi(t+\tau)\le \mathscr{M}^s(\tau;N). \tag{10.8}$$

Indeed, we may replace φ in (10.7) (resp. γ) by φ/Θ (resp. by γ/Θ), and we may integrate over φ/Θ instead of φ. Notice, that now $\varphi(0)\le 1$ and the case $\varphi(0)<1$ is not excluded.

Thus assuming (10.8) we have to prove that

$$\int_{N^{-1/2}}^{A}\varphi(t)t^{\alpha}\,dt \ll \gamma^{1-8/s}\frac{\Lambda_\alpha}{N} \quad \text{for}\ s>8, \tag{10.9}$$

with $\Lambda_\alpha = A^{\alpha+1}$, for $-1<\alpha\le 0$, and $\Lambda_{-1}=(1+\log A)(1+\log(1/\gamma))$. While proving (10.9) we may assume that $1\ll N$. Otherwise, (10.9) obviously holds since $\varphi\le\gamma\le 1$.

Let l_γ denote the largest integer such that $2^{-l_\gamma} \ge \gamma$. For the integers $l \ge l_\gamma$, introduce the sets

$$B_l = [N^{-1/2}, A] \cap \{t\colon 2^{-l-1} \le \varphi(t) \le 2^{-l}\},$$
$$D_l = [N^{-1/2}, A] \cap \{t\colon \varphi(t) \le 2^{-l-1}\}.$$

Since the function φ satisfies $0 \le \varphi(t) \le \gamma$, for $N^{-1/2} \le t \le A$, the sets B_l and D_l are closed and $D_m \cup \bigcup_{l=l_\gamma}^m B_l = [N^{-1/2}, A]$. Furthermore, (10.8) implies that $\varphi(t) \le t^{s/2}$, for $t \ge N^{-1/2}$, whence $B_l \subset [L_l^{-1}, A]$, where $L_l = 2^{2(l+1)/s}$.

Recall that $\varphi(t) \le 2^{-l}$, for $t \in B_l$, and $\varphi(t) \le 2^{-m-1}$, for $t \in D_m$. Therefore the relation $D_m \subset [N^{-1/2}, A]$ yields

$$\begin{aligned} \int_{N^{-1/2}}^A \varphi(t) t^\alpha\, dt &\le \int_{D_m} \varphi(t) t^\alpha\, dt + \sum_{l=l_\gamma}^m \int_{B_l} \varphi(t) t^\alpha\, dt \\ &\ll 2^{-m} G_\alpha + \sum_{l=l_\gamma}^m 2^{-l} \int_{B_l} t^\alpha\, dt, \end{aligned} \tag{10.10}$$

where $G_\alpha = \log A + \log N$, for $\alpha = -1$, and $G_\alpha = A^{\alpha+1}$, for $-1 < \alpha \le 0$. We shall choose m such that

$$2^{-m} G_\alpha \le \gamma^{1-8/s} \Lambda_\alpha N^{-1} \quad \text{for } -1 \le \alpha \le 0. \tag{10.11}$$

More precise, we choose the minimal m such that

$$m \ge \frac{1}{\log 2} \log \frac{N G_\alpha}{\gamma^{1-8/s} \Lambda_\alpha} \quad \text{for } -1 \le \alpha \le 0. \tag{10.12}$$

Using (10.10), (10.11), we see that the estimate (10.9) follows provided that we show that

$$\sum_{l=l_\gamma}^m I_l \ll \gamma^{1-8/s} \frac{\Lambda_\alpha}{N} \quad \text{where } I_l = 2^{-l} \int_{B_l} t^\alpha\, dt. \tag{10.13}$$

Below we shall prove the inequalities

$$\begin{aligned} I_l &\ll (l + \log A) N^{-1} 2^{-2l+8l/s} \quad \text{for } \alpha = -1, \\ I_l &\ll A^{\alpha+1} N^{-1} 2^{-2l+8l/s} \quad \text{for } -1 < \alpha \le 0 \end{aligned} \tag{10.14}$$

for $l \le m$. These inequalities imply (10.13). Indeed, in both cases $\alpha = -1$ and $\alpha > -1$, we can apply the bound

$$\sum_{l=l_\gamma}^\infty 2^{-2l+8l/s} \ll \left(2^{-l_\gamma}\right)^{1-8/s} \ll \gamma^{1-8/s}$$

since $s > 8$ ensures the convergence of the series, and, according to the definitions of γ and l_γ, we have $2^{-l_\gamma - 1} \le \gamma$. In the case $\alpha = -1$ one needs in addition the following estimates. For $l_\gamma \ll 1$, we have

$$\sum_{l=l_\gamma}^{\infty} l 2^{-2l+8l/s} \le \sum_{l=0}^{\infty} l 2^{-2l+8l/s} \ll 1 \ll \left(2^{-l_\gamma}\right)^{1-8/s} \ll \gamma^{1-8/s},$$

and, for $l_\gamma \ge c_s$ with a sufficiently large constant c_s, we obtain

$$\sum_{l=l_\gamma}^{\infty} l 2^{-2l+8l/s} \ll \int_{l_\gamma - 1}^{\infty} x 2^{-2x+8x/s}\, dx \ll l_\gamma \left(2^{-l_\gamma}\right)^{1-8/s} \ll \gamma^{1-8/s}\left(1 + \log\frac{1}{\gamma}\right).$$

It remains to prove inequalities (10.14). For the estimation of I_l we need a description of the structure of the sets B_l with $l \le m$. Let $t, t' \in B_l$ denote points such that $t' > t$. The inequality (10.8) and the definition of B_l imply

$$4^{-l-1} \le \mathscr{M}^s(t'-t; N). \tag{10.15}$$

If $t' - t \le N^{-1/2}$, then by (10.15) and the definition of $\mathscr{M}(\tau; N)$ we get

$$t' - t \le \delta \quad \text{where } \delta = N^{-1} 4^{2(l+1)/s}. \tag{10.16}$$

If $t' - t \ge N^{-1/2}$ then by (10.15) and the definition of $\mathscr{M}(\tau; N)$ we have

$$t' - t \ge \rho \quad \text{where } \rho = 4^{-2(l+1)/s}. \tag{10.17}$$

For $s > 8$ and sufficiently large $N \gg 1$ note that

$$\delta < \rho \quad \text{provided } l \le m. \tag{10.18}$$

Indeed, using the definitions of δ and ρ, we see that inequality (10.18) follows from the inequality $4N^{-s/8} < 2^{-m}$, which is implies by (10.12), the assumption (10.4) and $N \gg 1$.

The estimate (10.18) implies that either $t' - t \le \delta$ or $t' - t \ge \rho$. Therefore, it follows from (10.16)–(10.18) that

$$t \in B_l \Rightarrow B_l \cap (t + \delta, t + \rho) = \varnothing, \tag{10.19}$$

that is, that in the interval $(t + \delta, t + \rho)$ the function φ takes values lying outside of the interval $[2^{-l-1}, 2^{-l}]$.

Let us return to the proof of (10.14). If the set B_l is empty then (10.14) is obviously fulfilled. If B_l is nonempty then define $e_1 = \min\{t\colon t \in B_l\}$. Choosing $t = e_1$ and using (10.19), we see that the interval $(e_1 + \delta, e_1 + \rho)$ does not intersect B_l. Similarly, let e_2 denote the smallest $t \ge e_1 + \rho$ such that $t \in B_l$. Then the interval $(e_2 + \delta, e_2 + \rho)$ does not intersect B_l. Repeating this procedure, we construct a sequence $L_l^{-1} \le e_1 < e_2 < \cdots < e_k \le A$ such that

$$B_l \subset \bigcup_{j=1}^{k} \left[e_j, e_j + \delta\right] \quad \text{and} \quad e_{j+1} \ge e_j + \rho. \tag{10.20}$$

The sequence $e_1 < \cdots < e_k$ cannot be infinite. Indeed, due to (10.20), we have

$$A \geq e_k \geq e_1 + (k-1)\rho \geq L_l^{-1} + (k-1)\rho \geq k\rho,$$

and therefore $k \leq A/\rho$.

Using (10.20) we can finally prove (10.14). We start with the case $\alpha = -1$. Using $\log(1+x) \leq x$, for $x \geq 0$, we have

$$\begin{aligned} I_l &\leq 2^{-l} \sum_{j=1}^{k} \int_{e_j}^{e_j+\delta} \frac{dt}{t} = 2^{-l} \sum_{j=1}^{k} \log\left\{1 + \frac{\delta}{e_j}\right\} \\ &\leq 2^{-l} \sum_{j=1}^{k} \frac{\delta}{e_j} \ll (l + \log A) 2^{-2l+8l/s} N^{-1} \end{aligned}$$

since $e_1 \geq L_l^{-1} \geq \rho$, $k \leq A/\rho$, and

$$\begin{aligned} \sum_{j=1}^{k} \frac{1}{e_j} &\leq \sum_{j=1}^{k} \frac{1}{e_1 + (j-1)\rho} \\ &\leq \frac{1}{\rho} \sum_{j=1}^{k} \frac{1}{j} \ll \frac{\log A + \log \rho^{-1}}{\rho} \ll (1 + \log A) 4^{2l/s}. \end{aligned}$$

Finally, let us prove (10.14) for $-1 < \alpha \leq 0$. We have

$$\begin{aligned} I_l &\leq \frac{1}{2^l} \sum_{j=1}^{k} \int_{e_j}^{e_j+\delta} t^{\alpha}\, dt \leq \frac{1}{2^l} \sum_{j=1}^{k} \delta e_j^{\alpha} \\ &\leq \frac{\delta \rho^{\alpha}}{2^l} \sum_{j=1}^{k} j^{\alpha} \ll \frac{\delta A^{\alpha+1}}{2^l \rho} \ll N^{-1} 2^{-l+8l/s} A^{\alpha+1}. \end{aligned}$$ □

PROOF OF LEMMA 3.2. We shall apply Theorem 10.1. With $\alpha = -1$ this theorem implies

$$\begin{aligned} J &=_{\text{def}} \int_{N^{-1/2}}^{A} \varphi(t) \frac{dt}{t} \\ &\ll_s \left(\frac{\gamma}{\Theta}\right)^{1-8/s} \left(1 + \log \frac{\Theta}{\gamma}\right) (1 + \log A) \frac{\Theta}{N} \end{aligned} \tag{10.21}$$

in the case that γ satisfies (10.4), and

$$J \ll_s N^{-s/8} (1 + \log N)(1 + \log A) \tag{10.22}$$

if γ does not satisfy (10.4). The bounds (10.21) and (10.22) together yield

$$J \ll_s \frac{\Theta^2 (1 + \log A)}{N} \tag{10.23}$$

since we assume that $s > 8$, $\Theta \geq 1$ and since $\gamma \leq 1$. The bound (10.23) reduces the proof to verification that

$$\int_{BN^{-1/2}}^{N^{-1/2}} \varphi(t) \frac{dt}{t} \ll_s \Theta^2 B^{-s/2} N^{-s/4}, \qquad 0 < B \leq 1. \tag{10.24}$$

Substituting $t = 0$ in (3.7) and using the definition of the function $\mathscr{M}$, we have $\varphi(\tau) \le \Theta(\tau N)^{-s/2}$, for $BN^{-1/2} \le \tau \le N^{-1/2}$. Integrating now the bound for $\varphi(\tau)$, we obtain (10.24). □

Acknowledgments. The authors thank A. Yu. Zaitsev for useful discussions and V. Yurinsky for suggestions related to the exposition of the material.

REFERENCES

ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

BENTKUS, V. (1984). Asymptotic expansions for distributions of sums of independent random elements of a Hilbert space. *Lithuanian Math. J.* **24** 305–319.

BENTKUS, V. and GÖTZE, F. (1995). On the lattice point problem for ellipsoids. *Russian Acad. Sci. Dokl. Mach.* **343** 439–440.

BENTKUS, V. and GÖTZE, F. (1996). Optimal rates of convergence in the CLT for quadratic forms. *Ann. Probab.* **24** 466–490.

BENTKUS, V. and GÖTZE, F. (1997a). On the lattice point problem for ellipsoids. *Acta Arith.* **80** 101–125.

BENTKUS, V. and GÖTZE, F. (1997b). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Prob. Theory and Related Fields* **109** 367–416.

BENTKUS, V. and GÖTZE, F. (1997c). Optimal bounds in non-Gaussian limit theorems for U-statistics. Preprint 97-077 SFB 343, Univ. Bielefeld (http://www.mathematik.uni-bielefeld.de/sfb343/preprints.html).

BENTKUS, V. and GÖTZE, F. (1997d). Lattice point problems and distribution of values of quadratic forms. Preprint 96-111 SFB 343, Univ. Bielefeld (http://www.mathematik.uni-bielefeld.de/sfb343/preprints.html).

BENTKUS, V., GÖTZE, F. and PAULAUSKAS, V. (1996). Bounds for the accuracy of Poissonian approximations of stable laws. *Stochastic Process. Appl.* **65** 55–58.

BENTKUS, V., GÖTZE, F. and ZAITSEV, A. YU. (1997). Approximation of distributions of quadratic forms of independent random vectors by accompanying laws. *Theory Probab. Appl.* **42** 308–335.

BENTKUS, V., GÖTZE, F. and ZITIKIS, R. (1993). Asymptotic expansions in the integral and local limit theorems in Banach spaces with applications to ω-statistics. *J. Theoret. Probab.* **6** 727–780.

CHUNG, K. L. (1974). A course in probability theory. Academic Press, New York.

BORELL, C. (1974). Convex measure on locally convex spaces. *Arkiv. Mat.* **12** 239–252.

CSÖRGÖ, S. (1976). On an asymptotic expansion for the von Mises ω^2-statistics. *Acta. Sci. Math.* **38** 45–67.

ESSEEN, C.-G. (1945). Fourier analysis of distribution functions. *Acta. Math.* **77** 1–125.

GÖTZE, F. (1979). Asymptotic expansion for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50** 333–355.

GÖTZE, F. (1984). Expansions for von Mises functions. *Z. Wahrsch. Verw. Gebiete* **65** 599–625.

GÖTZE, F. and ZITIKIS, R. (1995). Edgeworth expansions and bootstrap for degenerate von Mises statistics. *Probab. Math. Statist.* **15** 327–351.

HARDY, G. H. (1916). The average order of the arithmetical functions $P(x)$ and $\Delta(x)$. *Proc. London Math. Soc.* **15** 192–213.

ITÔ, K. (1951). Multiple Wiener integral. *J. Math. Soc. Japan* **3** 157–169.

KANDELAKI, N. P. (1965). On limit theorem in Hilbert space. *Trudy Vychisl. Centra Akad. Nauk Gruzin. SSR* **11** 46–55 (in Russian).

KIEFER, J. (1972). Skorohod embedding of multivariate r.v.'s and the sample d.f. *Z. Wahrsch. Verw Gebiete* **24** 1–35.

KOROLJUK, V. S. and BOROVSKICH, YU. V. (1994). *Theory of U-statistics*. Kluwer, Dordrecht.

LANDAU, E. (1915). Zur analytischen Zahlentheorie der definiten quadratischen Formen. *Sitzber. Preuss. Akad. Wiss.* **31** 458–476.

LEE, A. J. (1990). *U*-statistics, Theory and practice. Dekker, New York.

MAJOR, P. (1981). *Multiple Wiener Integral. Lecture Notes in Math.* **849** 157–169. Springer, New York.

NAGAEV, S. V. and CHEBOTAREV, V. I. (1986). A refinement of the error: estimate of the normal approximation in a Hilbert space. *Siberian Math. J.* **27** 434–450.

PAULAUSKAS, V. (1976). On the rate of convergence in the central limit theorem in certain Banach spaces. *Theory Probab. Appl.* **21** 754–769.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

SAZONOV, V. V. (1969). An improvement of a convergence rate estimate. *Theory Probab. Appl.* **14** 640–651.

SENATOV, V. V. (1989). On the estimate of the rate of convergence in the central limit theorem in Hilbert space. *Stability Problems for Stochastic Models. Lecture Notes in Math.* **1412** 309–327. Springer, Berlin.

SMIRNOV, N. V. (1937). On the distribution of the ω^2 criterion of von Mises. *Rec. Math.* (*NS*) **2** 973–993 (in Russian).

WEYL, H. (1915). Über die Gleichverteilung der Zahlen mod-Eins. *Mathem. Ann.* **77** 313–352.

YARNOLD, J. K. (1972). Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Ann. Math. Statist.* **43** 1566–1580.

YURINSKII, V. V. (1982). On the accuracy of normal approximation of the probability of hitting a ball. *Theory Probab. Appl.* **27** 280–289.

YURINSKII, V. V. (1995). *Sums and Gaussian Vectors. Lecture Notes in Math.* **1617**. Springer, Berlin.

ZALESSKII, B. A., SAZONOV, V. V. and ULYANOV, V. V. (1988). A sharp estimate for the accuracy of the normal approximation in a Hilbert space. *Theory Probab. Appl.* **33** 700–701.

FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
POSTFACH 100131
33501 BIELEFELD 1
GERMANY
E-MAIL: bentkus@mathematik.uni-bielefeld.de
goetze@mathematik.uni-bielefeld.de