The Annals of Statistics
1997, Vol. 25, No. 2, 851–896

AN EDGEWORTH EXPANSION FOR SYMMETRIC STATISTICS

BY V. BENTKUS,[1] F. GÖTZE[1] AND W. R. VAN ZWET[1,2]

University of Bielefeld, University of Leiden and University of North Carolina, Chapel Hill

We consider asymptotically normal statistics which are symmetric functions of N i.i.d. random variables. For these statistics we prove the validity of an Edgeworth expansion with remainder $O(N^{-1})$ under Cramér's condition on the linear part of the statistic and moment assumptions for all parts of the statistic. By means of a counterexample we show that it is generally not possible to obtain an Edgeworth expansion with remainder $o(N^{-1})$ without imposing additional assumptions on the structure of the nonlinear part of the statistic.

1. Introduction and results. A second-order asymptotic theory for sums of independent and identically distributed (i.i.d.) random variables was established in the 1930's, mainly through the work of Esseen and Cramér. If $X_1, \dots, X_N$ are i.i.d. random variables with $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 = 1$ and $S = (X_1 + \cdots + X_N)/\sqrt{N}$, then, as $N \to \infty$, the distribution function F of S will converge to the standard normal distribution function Φ by the central limit theorem. If also $\mathbf{E}|X_1|^3 < \infty$, then the speed of convergence is given by the Berry–Esseen bound

$$\sup_x |F(x) - \Phi(x)| = O(N^{-1/2}).$$

If, moreover, Cramér's condition (C) holds, that is,

$$\sup_{|t|>\delta} |\mathbf{E}\exp\{itX_1\}| < 1 \quad \text{for some } \delta > 0,$$

and $\mathbf{E}|X_1|^4 < \infty$, then we have an Edgeworth expansion for F of the form

$$G(x) = \Phi(x) - \frac{\varkappa}{6\sqrt{N}}\Phi'''(x)$$

and

$$\sup_x |F(x) - G(x)| = O(N^{-1}).$$

Results of this type are called second-order asymptotic results. The Berry–Esseen bound is mainly of theoretical interest but clearly a necessary first step if one wishes to obtain Edgeworth expansions. Edgeworth expansions with

Received September 1994; revised April 1996.
[1]Research supported by the SFB 343 at Bielefeld.
[2]Research supported by the Netherlands Organization for Scientific Research.
AMS 1991 *subject classifications*. Primary 62E20; secondary 60F05.
Key words and phrases. Asymptotic expansion, Edgeworth expansions, symmetric statistics, Hoeffding's decomposition, U-statistics, functions of sample means, functionals of empirical distribution functions, linear combinations of order statistics, Student's statistic.

remainders $o(N^{-1/2})$ and $o(N^{-1})$ often provide excellent approximations, and also form the basis of refined asymptotic investigations in statistical theory.

Of course, most of the random variables occurring in statistics are not sums of i.i.d. random variables, but many may be approximated by such sums. To prove asymptotic normality for such statistics T, one usually proceeds by writing

$$T = S + R,$$

where S is a properly normalized sum of i.i.d. random variables and R a remainder term which tends to 0 in probability. Over the past decades, many authors have worked to extend this argument to obtain Berry–Esseen bounds and Edgeworth expansions for such more general statistics T. A great many results have been obtained for special cases, but as yet there is no satisfactory general second-order theory for asymptotically normal statistics. As long as we discuss statistics T that are functions of i.i.d. random variables, we may as well assume also that T is a symmetric function of $X_1, \dots, X_N$, thus restricting ourselves to symmetric statistics. For this class of statistics, one would like to have a general second-order asymptotic theory.

Much of the work on second-order asymptotics is based on Taylor expansion of the statistic T in terms of the underlying i.i.d. random variables $X_1, \dots, X_N$ [see, e.g., Chibisov (1972), Pfanzagl (1973), Bhattacharya and Ghosh (1978) and Bai and Rao (1991)]. This has been successful for a number of special cases, but is not likely to lead to a satisfactory general theory. The reason for this is that smoothness of the statistic as a function of $X_1, \dots, X_N$, which is needed for this type of argument, has very little to do with the existence of the Edgeworth expansion. We believe that Hoeffding's decomposition, which expands the statistic in a series of U-statistics of increasing order thus

$$T - \mathbf{E}T = \sum_{1 \le i \le N} \psi_1(X_i) + \sum_{1 \le i < j \le N} \psi_2(X_i, X_j) + \cdots,$$

is the appropriate tool for this problem.

A program with the ambitious aim of establishing a general theory of Edgeworth expansions for symmetric functions of i.i.d. random variables, should, of course, start by obtaining a Berry–Esseen bound for such statistics. If such a bound is to serve as preparation for obtaining Edgeworth expansions as a next step, it is essential that it be of the correct order $O(N^{-1/2})$. If one can only prove a Berry–Esseen bound of order, say, $O(N^{-1/2+\epsilon})$ or $O(N^{-1/2} \log N)$, how can one hope to establish an Edgeworth expansion where the first correction term is of exact order $O(N^{-1/2})$? This point is apparently not generally understood, witness the numerous papers providing Berry–Esseen bounds of larger order than $O(N^{-1/2})$. Such bounds are not only usually trivial, but also useless as a starting point for a next step.

After more than two decades of work on the Berry–Esseen bound for U-statistics [see, e.g., Filippova (1962), Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977), Callaert and Janssen (1978) and Serfling (1980)], the Berry-Esseen bound for general symmetric statistics was established in van Zwet (1984). Friedrich (1989), Götze (1991) and Bolthausen and Götze

(1993) extended this result to certain non-symmetric statistics, multivariate symmetric statistics and multivariate sampling statistics. The assumptions needed seem natural and are thought to be almost minimal. All that is required is a finite third moment of the leading term in Hoeffding's decomposition of T and a moment assumption controlling the influence of the remaining terms of the decomposition. It is important to stress that in order to check these two conditions, one does not need to calculate Hoeffding's decomposition explicitly. Berry–Esseen bounds were also obtained separately for many special cases of symmetric statistics such as linear functions of order statistics [Helmers (1982)], Studentized U-statistics [Zhao (1983) and Helmers (1985)], multivariate U-statistics [Götze (1987)], multivariate L-statistics [Zitikis (1993)] and so on.

As a next step, an Edgeworth expansion with remainder $o(N^{-1})$ was established for U-statistics of degree 2 in Bickel, Götze and van Zwet (1986) after an earlier and less explicit result of Callaert, Janssen and Veraverbeke (1980). The expansion was proved under Cramér's condition on the leading term in Hoeffding's decomposition and moment assumptions that were to be expected. However, a condition on the eigenvalues of the second-order term in the decomposition was also needed. This condition had occurred in earlier work on degenerate U-statistics [Götze (1979)] and was therefore not entirely unexpected. However, the authors expressed their uncertainty about the necessity of this assumption and this has led a number of authors to propose proofs that the eigenvalue assumption could be omitted, but all were found to be in error. To end this discussion, we provide an example in Theorem 1.4 which shows that the Edgeworth expansion with error $o(N^{-1})$ is not necessarily valid if the eigenvalue assumption does not hold.

The present authors will confess that they optimistically believed that it would be straightforward to generalize the result in Bickel, Götze and van Zwet to symmetric statistics, thereby completing the program for obtaining an Edgeworth expansion for such statistics. However, it turns out that the interplay between the higher-order terms in Hoeffding's decomposition—which are, of course, not present for a U-statistic of degree 2—is so complex that the proof becomes extremely hard. We do have a theorem on the Edgeworth expansion for general symmetric statistics, but this result is as yet not in a form which is fit for publication [Götze and van Zwet (1991)].

Instead, we shall backtrack a little in the present paper. The problem with obtaining the general result lies in the contributions to the Edgeworth expansion which are of approximate order $O(N^{-1})$. In this paper we obtain an Edgeworth expansion with remainder of order $O(N^{-1})$, hence not including in the expansion the term of order $O(N^{-1})$ but establishing that it is indeed of this order. As we pointed out in our discussion of Berry–Esseen bounds, it is essential to get exactly the right order $O(N^{-1})$ for the remainder; anything larger will simply not do as a preparation for the final step, which will be to establish the term of order $O(N^{-1})$.

The conditions of the result of this paper seem natural and are probably very close to optimal: we need Cramér's condition on the leading term, the obvious moments, as well as a moment assumption controlling the influence of the terms involving functions of three or more variables in Hoeffding's decomposition of T. The latter condition is a natural counterpart of the corresponding condition for the Berry–Esseen bound in van Zwet (1984).

Having sketched the general context in which the results of this paper should be viewed, we now define the notation necessary to formulate these results.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed (i.i.d.) random variables taking values in an arbitrary measurable space $(\mathscr{X}, \mathscr{B})$ with common distribution $\mathbf{P}$. Let the measurable function $t\colon \mathscr{X}^N \to \mathbf{R}$ be symmetric in its N arguments, in the sense that it is invariant under permutation of these arguments. Consider the symmetric statistic

$$T = t(X_1, \ldots, X_N),$$

and assume throughout that

$$\mathbf{E}T^2 < \infty.$$

Our aim is to establish an Edgeworth expansion with remainder $O(N^{-1})$ for the distribution of T. Define

$$T_i = \mathbf{E}(T|X_i) - \mathbf{E}T, \tag{1.1}$$

$$T_{ij} = \mathbf{E}(T|X_i, X_j) - \mathbf{E}(T|X_i) - \mathbf{E}(T|X_j) + \mathbf{E}T \tag{1.2}$$

and write

$$T = \mathbf{E}T + \sum_{i=1}^{N} T_i + \sum_{1 \le i < j \le N} T_{ij} + R. \tag{1.3}$$

Notice that the terms on the right-hand side of (1.3) are all pairwise uncorrelated. Define

$$\sigma^2 = \sigma^2(T) = \mathbf{E}(T - \mathbf{E}T)^2 \tag{1.4}$$

and

$$\varkappa = \alpha_3 + 3\delta,$$

where

$$\alpha_3 = \mathbf{E}(\sqrt{N}T_1)^3, \qquad \delta = \mathbf{E}(\sqrt{N}T_1\,\sqrt{N}T_2\,N^{3/2}T_{12}).$$

It is easy to verify that the third moment of

$$\sum_{i=1}^{N} T_i + \sum_{1 \le i < j \le N} T_{ij}$$

equals $\varkappa N^{-1/2} + O(N^{-1})$, provided that third moments of $\sqrt{N}T_1$ and $N^{3/2}T_{12}$ are bounded. The formal Edgeworth expansion for the distribution function of a random variable with expectation 0, variance 1 and third moment $\varkappa$ equals

$$\Phi(x) - \frac{\varkappa}{6\sqrt{N}}\Phi'''(x).$$

Hence, if F and G denote the distribution functions of $(T - \mathbf{E}T)/\sigma$ and its two-term Edgeworth expansion, then

$$(1.5) \qquad F(x) = \mathbf{P}\{T - \mathbf{E}T \le x\sigma\}, \qquad G(x) = \Phi(x) - \frac{\varkappa}{6\sigma^3\sqrt{N}}\Phi'''(x).$$

We shall show that under appropriate conditions

$$\sup_x |F(x) - G(x)| = O(N^{-1}).$$

To arrive at the result, we shall obviously have to make an assumption to control the remainder term R in (1.3). Define

$$\mathbf{E}_i\, T = (T|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N),$$
$$D_i T = T - \mathbf{E}_i\, T \quad \text{for } 1 \le i \le N.$$

Repeating the operation D_i, we obtain, for distinct i, j and k,

$$D_i D_j T = T - \mathbf{E}_i\, T - \mathbf{E}_j\, T + \mathbf{E}_{ij}\, T,$$
$$D_i D_j D_k T = T - \mathbf{E}_i\, T - \mathbf{E}_j\, T - \mathbf{E}_k\, T + \mathbf{E}_{ij}\, T + \mathbf{E}_{ik}\, T + \mathbf{E}_{jk}\, T - \mathbf{E}_{ijk}\, T,$$

where, for brevity, we write $\mathbf{E}_{ij}$ instead of $\mathbf{E}_i\, \mathbf{E}_j$ and $\mathbf{E}_{ijk}$ instead of $\mathbf{E}_i\, \mathbf{E}_j\, \mathbf{E}_k$. Note that the order in which these operations are carried out is immaterial. Also, the distributions of

$$D_i T, \quad D_i D_j T, \quad D_i D_j D_k T, \quad T_i, \quad T_{ij}$$

for distinct i, j and k do not depend on i, j and k.

For $s > 0$ and $m = 1, 2, 3$, define the moments

$$\beta_s = \mathbf{E}|\sqrt{N} T_1|^s, \qquad \gamma_s = \mathbf{E}|N^{3/2} T_{12}|^s$$

and

$$\Delta_m^s = \mathbf{E}|N^{m-1/2} D_1 D_2 \cdots D_m T|^s.$$

By Jensen's inequality it follows that

$$\beta_s \le \Delta_1^s, \qquad \gamma_s \le \Delta_2^s \quad \text{for } s \ge 1.$$

Define the number $q(\sigma) \in [0, 1]$ by

$$1 - q(\sigma) = \sup\left\{ |\mathbf{E}\exp\{it\sqrt{N}T_1\}| \colon \; \frac{\sigma^2}{2\beta_3} \le |t| \le \frac{\sqrt{N}}{\sigma} \right\}.$$

Note that $q(\sigma) > 0$ if the characteristic function of $\sqrt{N} T_1$ satisfies Cramér's condition

$$(\mathrm{C}) \qquad \limsup_{|t| \to \infty} |\mathbf{E}\exp\{it\sqrt{N}T_1\}| < 1.$$

THEOREM 1.1 (Edgeworth expansion). *There exists an absolute constant c such that, for any symmetric statistic T, any* **P** *and any* $N = 1, 2, \dots,$

$$\sup_x |F(x) - G(x)| \le \frac{c}{q^2(\sigma)N}\left(\frac{\beta_4}{\sigma^4} + \frac{\gamma_3}{\sigma^3} + \frac{\Delta_3^2}{\sigma^2}\right).$$

Sometimes it is more convenient to have an Edgeworth expansion for the distribution of $(T - \mathbf{E}T)/\tau$, where τ^2 is the variance of the linear part $\sum T_i$ of $T - \mathbf{E}T$. Thus

$$\tau^2 = \mathbf{E}(\sqrt{N}T_1)^2.$$

Let

$$\overline{F}(x) = \mathbf{P}\{T - \mathbf{E}T \le x\tau\}, \qquad \overline{G}(x) = \Phi(x) - \frac{\varkappa}{6\tau^3\sqrt{N}}\Phi'''(x)$$

and

$$1 - \bar{q}(\tau) = \sup\left\{|\mathbf{E}\exp\{it\sqrt{N}T_1\}|: \ \frac{\tau^2}{\beta_3} \le |t| \le \frac{\sqrt{N}}{\tau}\right\}.$$

THEOREM 1.2. *There exists an absolute constant c such that, for any symmetric statistic T, any* **P** *and any* $N = 1, 2, \dots,$

$$\sup_x |\overline{F}(x) - \overline{G}(x)| \le \frac{c}{\bar{q}^2(\tau)N}\left(\frac{\beta_4}{\tau^4} + \frac{\gamma_3}{\tau^3} + \frac{\Delta_3^2}{\tau^2}\right).$$

REMARK 1.3. Let $0 < \rho \le 1$. Theorem 1.1 remains valid if $q^2(\sigma)$ is replaced by $\rho^2 q_\rho^2(\sigma)$, where

$$q_\rho(\sigma) = 1 - \sup\left\{|\mathbf{E}\exp\{it\sqrt{N}T_1\}|: \ \frac{\sigma^2}{2\beta_3} \le |t| \le \frac{\rho\sqrt{N}}{\sigma}\right\}.$$

Furthermore, the proof shows that, for every $\varepsilon > 0$, there exists $c(\varepsilon)$ such that

$$\sup_x |F(x) - G(x)| \le \frac{c(\varepsilon)}{q^2(\sigma)N}\left(\frac{\beta_4}{\sigma^4} + \frac{\gamma_{2+\varepsilon}}{\sigma^{2+\varepsilon}} + \frac{\Delta_3^2}{\sigma^2}\right).$$

We conjecture that ε may be removed. Such an estimate would be optimal in view of the lower bounds of Bentkus, Götze and Zitikis (1994). The dependence on $q^2(\sigma)$ may be improved to $q^{1+\varepsilon}(\sigma)$, $\varepsilon > 0$. It is not clear whether the bound remains valid with $q^2(\sigma)$ replaced by $q(\sigma)$ or an even lower power of $q(\sigma)$. The same improvements and comments hold for the bound on $\sup_x\left|\overline{F}(x) - \overline{G}(x)\right|$ in Theorem 1.2 with σ replaced by τ.

In the proof of our result an essentially new idea appears to be needed, and we shall introduce a technique which we call *data-dependent smoothing*.

It is based on a nonstandard smoothing inequality [see Prawitz (1972)]. Such inequalities were suggested by Beurling and have been used in number theory and analysis [see, e.g., Graham and Vaaler (1981)]. We did not succeed in proving our result using the conventional Esseen inequality (3.4).

The estimates in Theorems 1.1 and 1.2 are formulated for any fixed sample size N, but since the constant c is not specified, these are purely asymptotic results concerned with a sequence of statistics $(T_N - \mathbf{E}T_N)/\sigma(T_N)$ as $N \to \infty$. If β_4/σ^4, γ_3/σ^3 and Δ_3^2/σ^2 are bounded and $q(\sigma)$ is bounded away from 0, as $N \to \infty$, then the theorems establish Edgeworth expansions with remainders $O(N^{-1})$. For given bounds on

$$\beta_4/\sigma^4, \quad \gamma_3/\sigma^3, \quad \Delta_3^2/\sigma^2 \quad \text{and} \quad q(\sigma),$$

the result is uniform for all symmetric functions of the i.i.d. variables $X_1, \dots, X_N$ and for all underlying distributions $\mathbf{P}$.

In Section 2 we shall apply our general results to several special cases—sample means, U-statistics, sample variances, Student's statistic, functions of sample means, functionals of empirical distribution functions and linear combinations of order statistics—to see whether we can obtain results comparable to the best available ones for these well-studied cases. It turns out that this is indeed the case. In some instances, such as functions of sample means of random functions, we derive stronger and more general results than those available so far. Usually our conditions seem to be optimal, with the sole exception of Student's t-statistic, where we need $4 + \varepsilon$ moments instead of 4, which should suffice.

Edgeworth expansions can easily be used to establish the accuracy of bootstrap approximations since the estimates of the remainder are explicit, and the expansions are valid for a sufficiently general class of symmetric statistics. Bootstrap or Studentized versions of a symmetric statistic are again symmetric, and in order to show the validity of Edgeworth expansions, it suffices to estimate several lower-order conditional differences. In this paper we shall do this for the simplest example of the Studentized mean only. More general results will be published elsewhere.

Similar estimates to those in Theorems 1.1 and 1.2 do not hold for higher-order Edgeworth expansions, even if we assume the existence of moments of arbitrarily high order of all parts of the statistic. In Theorem 1.4 we shall provide an example of this phenomenon. We show that under Cramér's condition (C) on the linear part of the statistic and arbitrarily restrictive moment conditions, it is not possible to establish an Edgeworth expansion with remainder $o(N^{-1})$, even for U-statistics of order 2. Hence in this sense our results are best possible.

In order to formulate Theorem 1.4, we need some additional notation. A U-statistic T of order 2 can be written as

$$T = \sum_{i=1}^{N} T_i + \sum_{1 \le i < j \le N} T_{ij}, \tag{1.6}$$

where $T_1 = g_1(X_1)$ and $T_{12} = g_2(X_1, X_2)$, with some symmetric functions g_1 and g_2 such that $\mathbf{E}T_1 = 0$ and $\mathbf{E}(T_{12}|X_2) = 0$. In this case $\mathbf{E}T = 0$,

$$D_1 T = T_1 + \sum_{j=2}^{N} T_{1j}, \qquad D_1 D_2 T = T_{12}, \qquad D_1 D_2 D_3 T = 0.$$

Let us assume that the components of the statistic T are uniformly bounded by some nonrandom $A < \infty$, that is,

$$\sup_N |\sqrt{N} T_1/\sigma| \le A, \qquad \sup_N |N^{3/2} T_{12}/\sigma| \le A. \tag{1.7}$$

Notice that (1.7) ensures that the Edgeworth approximation G, as well as any higher-order Edgeworth expansion, must satisfy

$$\sup_N \sup_x |G'(x)| < \infty. \tag{1.8}$$

THEOREM 1.4. *There exists a sequence $T = T_N$ of symmetric statistics* (1.6) *such that*

$$\mathbf{E}(\sqrt{N} T_1)^2 = 1, \qquad |\sqrt{N} T_1| \le 4, \qquad |N^{3/2} T_{12}| \le 2, \tag{1.9}$$

$$\sup_N \sup_{|t|\ge\delta} |\mathbf{E}\{it\sqrt{N} T_1\}| < 1 \quad \text{for each } \delta > 0 \tag{1.10}$$

and

$$\limsup_N N \sup_{x\in\mathbf{R}} |F(x) - G_N(x)| > 0$$

for any given sequence of functions G_N: $\mathbf{R} \to \mathbf{R}$ such that

$$\sup_N \sup_x |G_N'(x)| < \infty.$$

Moreover, for any given $A > 0$,

$$\limsup_{N\to\infty} N \inf_G \sup_x |F(x) - G(x)| > 0, \tag{1.11}$$

where the inf *is taken over all G: $\mathbf{R} \to \mathbf{R}$ such that* $\sup_x |G'(x)| \le A$.

2. Applications. In this section we use the following notation. Let $\mathbf{B}$ denote a real Banach space. Let $X_1, \ldots, X_N$ be i.i.d. mean-zero random variables taking values in $\mathbf{B}$. Denote

$$b_s = \mathbf{E}\|X_1\|^s, \qquad \overline{X} = (X_1 + \cdots + X_N)/N,$$
$$E_N = \sqrt{N}\,\overline{X}, \qquad Q_s = \sup_{N\ge1} \mathbf{E}\|E_N\|^s,$$

where we shall assume implicitly that all random variables are well defined (e.g., assume that $\mathbf{B}$ is separable). If $\mathbf{B} = \mathbf{R}$ is the real line, then we write $\mu_s = \mathbf{E}X_1^s$.

For a function $H\colon \mathbf{B} \to \mathbf{R}$, let $H^{(s)}(x)$ denote the sth Frechét derivative of H at the point $x \in \mathbf{B}$. Define

$$H^{(s)}(x) h_1 \cdots h_s$$

as the value of the s-linear continuous symmetric form $H^{(s)}(x)$ with arguments $h_j \in \mathbf{B}$, $1 \le j \le s$, as well as

$$\|H^{(s)}\|_\infty = \sup_{x \in \mathbf{B}} \|H^{(s)}(x)\|, \qquad M_0 = M_0(H) = L_0 + \|H'\|_\infty + \|H''\|_\infty,$$

where $\|H^{(s)}(x)\|$ is the sup-norm of the s-linear form $H^{(s)}(x)$, and

$$L_0 = L_0(H) = \sup_{x \in \mathbf{B}} \|H''(x) - H''(0)\| / \|x\|.$$

Write

$$M = M(H) = L + \sum_{i=1}^{3} \|H^{(i)}\|_\infty, \qquad L = L(H) = \sup_{x,\, y \in \mathbf{B}} \frac{\|H'''(x) - H'''(y)\|}{\|x - y\|^{1/2}}.$$

2.1. *Sample means.* Let $X_1, \dots, X_N$ be real i.i.d. random variables. Consider the statistic $T = \overline{X}$, where $\overline{X}$ is the sample mean. It follows from Theorem 1.2 that the distribution function of the statistic $\sqrt{N} T / \sqrt{b_2}$ can be uniformly approximated by a two-term Edgeworth expansion, and that the error does not exceed $c b_4 / (q^2 b_2^2 N)$. This is a well-known result. Note that the dependence on

$$q = 1 - \sup_{|t| \ge b_2 / b_3} |\mathbf{E} \exp\{itX_1\}|$$

can be improved; see Petrov (1975).

2.2. *Sample variances.* Let $X_1, \dots, X_N$ be real i.i.d. random variables. Consider the sample variance

$$T = \widehat{b}_2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2 = \frac{1}{N^2} \sum_{1 \le i < j \le N} (X_i - X_j)^2.$$

In this case we may write in the decomposition (1.3) for $T - \mathbf{E}T$,

$$T_1 = (N-1) N^{-2} (X_1^2 - b_2), \qquad T_{12} = -2N^{-2} X_1 X_2, \qquad R = 0.$$

Furthermore,

$$\tau^2 = (N-1)^2 N^{-3} D^2, \qquad D^2 = b_4 - b_2^2, \qquad \sigma^2 = \tau^2 + 2(N-1) N^{-3} b_2^2.$$

The distribution function of

$$(T - \mathbf{E}T)/\tau \quad \big[\text{or of } (T - \mathbf{E}T)/\sigma\big]$$

is uniformly approximable by its Edgeworth expansion, and the error does not exceed $cb_8/(q^2D^4N)$, where

$$q = 1 - \sup_{|t|\geq D^2/\beta_3} |\mathbf{E}\exp\{itX_1^2\}|$$

and

$$\beta_3 = \mathbf{E}|X_1^2 - b_2|^3.$$

2.3. *On the estimation of differences.* For the calculation and estimation of differences and terms of the Hoeffding decomposition, the following simple inequalities are useful. If V, W are (not necessarily symmetric) statistics depending on $X_1, \dots, X_N$, then

$$\mathbf{E}|D_1 \cdots D_k V|^s \leq 2^{k+s}\mathbf{E}|V|^s \quad \text{for all } s \geq 1$$

and

$$\mathbf{E}\big|D_1 \cdots D_k(V+W)|^s \leq 2^s\mathbf{E}|D_1 \cdots D_k V|^s + 2^{k+s}\mathbf{E}|W|^s \quad \text{for all } s \geq 1.$$

If V is independent of X_i, then $D_iV = 0$. If V is independent of at least one of the random variables $X_1, \dots, X_k$, then $D_1 \dots D_k V = 0$.

For example, let us consider the case when $T = H(\overline{X})$ is a sufficiently smooth function of the sample mean. Let $\tau_1, \dots, \tau_k$ denote i.i.d. random variables uniformly distributed on $[0, 1]$ and independent of all other random variables. Then

$$\begin{aligned} &D_1 \cdots D_k H(\overline{X}) \\ (2.1)\qquad &= \frac{1}{N^k} D_1 \cdots D_k \mathbf{E}_\tau H^{(k)}\bigg(\frac{\tau_1 X_1}{N} + \cdots + \frac{\tau_k X_k}{N} + B\bigg) X_1 \cdots X_k, \end{aligned}$$

where $\mathbf{E}_\tau$ denotes the conditional expectation given all r.v.'s but $\tau_1, \dots, \tau_k$, and where B is defined by $B = (X_{k+1} + \cdots + X_N)/N$.

Indeed, split

$$\overline{X} = N^{-1}X_1 + A \quad \text{where } A = (X_2 + \cdots + X_N)/N.$$

Expanding into the Taylor series, we have

$$H(\overline{X}) = H(A) + N^{-1}\mathbf{E}_\tau H'(N^{-1}\tau_1 X_1 + A)X_1.$$

Since the random variable $H(A)$ is independent of X_1, we have $D_1H(A) = 0$. Repeating this procedure for $X_2, \dots, X_k$, we obtain (2.1).

Thus the right-hand side of (2.1) is bounded from above by

$$c(k)N^{-k}\|H^{(k)}\|_\infty (\|X_1\| + b_1) \cdots (\|X_k\| + b_1),$$

and in many applications this yields a satisfactory estimate.

For a polynomial H of order k, relation (2.1) yields $D_1 \cdots D_{k+1} H(\overline{X}) = 0$.

2.4. *Student's statistic and the bootstrap.* Let $X_1, \dots, X_N$ denote real i.i.d. random variables. Consider Student's t-statistic and its bootstrap version t^*,

$$t = \overline{X}/\sqrt{\widehat{b}_2}, \qquad t^* = (\overline{X}^* - \overline{X})/\sqrt{\widehat{b}_2^*},$$

where $\overline{X}$ denotes the sample mean and $\widehat{b}_2$ denotes the sample variance. The bootstrap version t^* is obtained by replacing the arguments

$$X_1, \dots, X_N \quad \text{of } t \quad \text{by } X_1^* - \overline{X}, \dots, X_N^* - \overline{X},$$

where the i.i.d. random variables $X_1^*, \dots, X_N^*$ are drawn from $F_N = N^{-1}\sum_{j=1}^N \delta_{X_j}$ and are independent of $X_1, \dots, X_N$.

We shall prove the following estimate:

$$\sup_x \left| \mathbf{P}\{\sqrt{N}t \le x\} - G_N\left(x, \frac{\mu_3}{b_2^{3/2}}\right) \right| \le \frac{c(\varepsilon)}{q^2 N}\left(\frac{b_3 b_{4+\varepsilon}}{b_2^{7/2+\varepsilon/2}} + \frac{b_4^3}{b_2^6}\right) \tag{2.2}$$

for arbitrary $\varepsilon > 0$, where

$$G_N(x, y) = \Phi(x) + \frac{y(2x^2+1)}{6\sqrt{N}}\Phi'(x)$$

and

$$1 - q = \sup\{|\mathbf{E}\exp\{i\tau X_1\}|: b_2/(2b_3) \le |\tau| \le \sqrt{N/b_2}\}.$$

For fixed $X_1, \dots, X_N$, let $\mathbf{P}^*$ and $\mathbf{E}^*$ denote the conditional probability and expectation with respect to F_N. Substituting in (2.2) X_i by $X_i^* - \overline{X}$, we obtain

$$\sup_x |\mathbf{P}^*\{\sqrt{N}t^* \le x\} - G_N(x, \widehat{\mu}_3 \widehat{b}_2^{-3/2})| \le \frac{c(\varepsilon)}{\widehat{q}^{\,2} N}(\widehat{b}_2^{-7/2-\varepsilon/2}\widehat{b}_3\widehat{b}_{4+\varepsilon} + \widehat{b}_2^{-6}\widehat{b}_4^3)$$

for arbitrary $\varepsilon > 0$, where

$$1 - \widehat{q} = \sup\left\{|\mathbf{E}^*\exp\{i\tau X_1^*\}|: \widehat{b}_2/(2\widehat{b}_3) \le |\tau| \le \sqrt{N/\widehat{b}_2}\right\}$$

and

$$\widehat{\mu}_3 = \mathbf{E}^*(X_1^* - \overline{X})^3 = \frac{1}{N}\sum_{i=1}^N (X_i - \overline{X})^3,$$

$$\widehat{b}_s = \mathbf{E}^*|X_1^* - \overline{X}|^s = \frac{1}{N}\sum_{i=1}^N |X_i - \overline{X}|^s$$

denote the sample moments.

Let us fix a sequence $X_1, X_2, \dots$ of random variables. Then $\mathbf{P}\{\widehat{b}_2 < b_2/2\} \to 0$ as $N \to \infty$. If X_1 satisfies Cramér's condition (C), then $\mathbf{P}\{\widehat{q} < q/2\} \to 0$ as $N \to \infty$. Indeed,

$$\mathbf{E}^*\exp\{i\tau X_1^*\} = N^{-1}\sum_{j=1}^N \exp\{i\tau X_j\},$$

and applying Chebyshev's inequality, we obtain

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{j=1}^{N}\exp\{i\tau X_j\}-\mathbf{E}\exp\{i\tau X_1\}\right|>\frac{q}{2}\right\}\le\frac{16}{q^2N}.$$

Thus a comparison of Edgeworth expansions for t and t^* shows that

$$\sup_x|\mathbf{P}\{\sqrt{N}t\le x\}-\mathbf{P}^*\{\sqrt{N}t^*\le x\}|=O_P(N^{-1}),$$

provided that $b_6<\infty$, where the notation $\xi_N=O_{\mathbf{P}}(\delta_N)$ means that

$$\lim_{\lambda\to\infty}\limsup_{N\to\infty}\mathbf{P}\{|\xi_N|>\lambda\delta_N\}=0.$$

Proof of (2.2). Without loss of generality, we may assume that $b_2=1$. We will apply the estimate of Remark 1.3. Let $f\colon\mathbf{R}\to\mathbf{R}$ be a bounded infinitely many times differentiable function such that $f(x)=1/\sqrt{|x|}$ for $|x|\ge 1/2$. Note that all derivatives of f are bounded. We may regularize t by replacing it by $t_f=\overline{X}f(\widehat{b}_2)$. Indeed,

$$\mathbf{P}\{t\ne\overline{X}f(\widehat{b}_2)\}\le\mathbf{P}\{\widehat{b}_2\le 1/2\}\le c\mathbf{E}(\widehat{b}_2-\mathbf{E}\widehat{b}_2)^2\le cb_4/N,\qquad N\ge 3.$$

In order to truncate t_f define

$$Y_i=N^{-1/2}X_i\mathbb{1}\{X_i^2\le N\},\qquad S_N=Y_1+\cdots+Y_N,\qquad s_N^2=Y_1^2+\cdots+Y_N^2.$$

Thus the complement of the event $\{X_i^2\le N,\quad 1\le i\le N\}$ occurs with probability less than cb_4/N, and hence we may replace $\sqrt{N}t_f$ by

$$T=S_Nf(s_N^2-N^{-1}S_N^2).$$

It is well known (and easy to verify) that

$$\mathbf{E}|S_N|^p\le c(p)\quad\text{for } p\ge 1.$$

Note that

$$\mathbf{E}T=-\frac{\mu_3}{2\sqrt{N}}+R\quad\text{where } |R|\le\frac{cb_4}{N}.$$

Now split $S_N=Y_1+\cdots+Y_i+S_{N-i}$ and use the independence of the random variables $Y_1,\dots,Y_i$ and S_{N-i}. This yields the bounds

$$\begin{aligned}|\sqrt{N}D_1T|&\le c(1+|S_{N-1}|)(1+|X_1|),\\ |N^{3/2}D_1D_2T|&\le c(1+S_{N-2}^2)(1+X_1^2+X_2^2+X_1^2|X_1|+|X_1|X_2^2),\qquad(2.3)\\ |N^{5/2}D_1D_2D_3T|&\le c(1+S_{N-3}^4)\sum|X_1|^k|X_1|^l|X_1|^m,\end{aligned}$$

where the last sum is taken over all nonnegative integers $k,l,m\le 2$. The inequalities (2.3) imply

$$\beta_4\le cb_4,\qquad\gamma_{2+\varepsilon}\le cb_3b_{4+\varepsilon},\qquad\Delta_3^2\le cb_4^3.$$

We shall prove the first inequality in (2.3). The proofs of the other inequalities are similar but somewhat more tedious. Split $s_N^2 = Y_1^2 + s_{N-1}^2$. We may split T into

$$T = Q_1 + Q_2 \quad \text{where } Q_1 = Y_1 f(s_N^2 - N^{-1} S_N^2)$$

and

$$Q_2 = S_{N-1} f(Y_1^2 + s_{N-1}^2 - N^{-1} Y_1^2 - 2N^{-1} Y_1^2 S_{N-1} - N^{-1} S_{N-1}^2).$$

We have $Q_1 \le |Q_1|$. In order to estimate Q_2 one should expand first in powers of Y_1^2 and then in those of $2N^{-1} Y_1^2 S_{N-1}$, using $Y_1^2 \le |Y_1|$.

For the estimation of q notice that $\sqrt{N} T_1 = \sqrt{N} Y_1 + R$, where $|R|$ is bounded by $cN^{-1} b_4(|X_1| + X_1^2)$. This completes the sketch of the proof of (2.2). □

2.5. *Functions of sample means.* Assume that the independent identically distributed random variables $X_1, \dots, X_N$ take values in a real Banach space $\mathbf{B}$. Consider the statistics

$$T^0 = \sqrt{N}(H(\overline{X}) - H(0))$$

and

$$T = \sqrt{N}(H(\overline{X}) - \mathbf{E} H(\overline{X})),$$

where H denotes a function $H: \mathbf{B} \to \mathbf{R}$. The statistics T^0 and T are equivalent for our purposes since

$$\left| \mathbf{E} H(\overline{X}) - H(0) - \frac{1}{2N} \mathbf{E} H''(0) X_1^2 \right| \le L_0 Q_3 N^{-3/2},$$

where we write $H''(0)XX = H''(0)X^2$. This follows immediately by expanding H in powers of $\overline{X}$. Similarly,

$$\left| \mathbf{E} H(\overline{X}) - H(0) - \frac{1}{2N} \mathbf{E} H''(0) X_1^2 \right| \le M N^{-3/2} (b_{7/2} + b_{5/2}),$$

which is obtained by expansions in powers of X_i, $1 \le i \le N$.

Therefore it suffices to prove Edgeworth expansions for T.

Define

$$s_1^2 = \mathbf{E}(H'(0) X_1)^2, \qquad a_3 = s_1^{-3} \mathbf{E}(H'(0) X_1)^3,$$

$$d = s_1^{-3} \mathbf{E}(H'(0) X_1)(H'(0) X_2)(H''(0) X_1 X_2), \qquad k = a_3 + 3d$$

and

$$q = 1 - \sup_{|t| \ge s_1^2/(16 M_0^3 b_3)} |\mathbf{E} \exp\{it H'(0) X_1\}|.$$

Then we have the following result.

THEOREM 2.1. *The distribution function of T/s_1 can be uniformly approximated by $\{\Phi(x) - k\Phi'''(x)\}/(6\sqrt{N})$, and the error does not exceed*

$$cN^{-1}q^{-2}(b_3 + b_4 + b_4^2 + Q_2 + Q_2^4)(s_1^{-2}M_0^2 + s_1^{-4}M_0^4)$$

or, in case $M < \infty$,

$$cN^{-1}q^{-2}(b_3 + b_4 + b_4^2)(s_1^{-2}M^2 + s_1^{-4}M^4).$$

Theorem 2.1 follows from Theorem 1.2 and Lemmas 2.2 and 2.3.

Bounds for the moments Q_s are well known; see, for instance, de Acosta (1981).

Edgeworth expansions for functions of sample means in $\mathbf{B} = \mathbf{R}^k$ have been extensively studied; see, for instance, Bhattacharya and Ghosh (1978) and Bai and Rao (1991). The smoothness condition $M_0 < \infty$ for the function H seems to be optimal and is comparable with conditions in Bai and Rao (1991). Although Theorem 2.1 holds for infinite-dimensional spaces $\mathbf{B}$, our requirement that the random variable $H'(0)X$ satisfies Cramér's condition (C) is weaker than corresponding conditions used in the literature for $\mathbf{B} = \mathbf{R}^k$. For instance, it allows X_1 to have discrete coordinates since we need Cramér's condition (C) for one coordinate only. The moment assumption $b_4 < \infty$ is natural and seems unimprovable.

LEMMA 2.2. *We have*

$$\mathbf{E}|N^{k-1/2}D_1 \cdots D_k T|^s \le c(k, s)\|H^{(k)}\|_\infty^s b_s^k \quad \textit{for all } s \ge 1,\ 1 \le k \le N.$$

In particular,

$$s_1^{-4}\beta_4 + s_1^{-3}\gamma_3 + s_1^{-2}\Delta_3^2 \le c(s_1^{-4}M_0^4 + s_1^{-2}M_0^2)(b_4 + b_3^2 + b_2^3).$$

PROOF. The second estimate of the lemma follows from the first one. The first estimate is a consequence of (2.1).

We omit the tedious proof of Lemma 2.3 since it is similar to that of Lemma 2.2.

LEMMA 2.3. *For the statistic $T = \sqrt{N}(H(\overline{X}) - \mathbf{E}H(\overline{X}))$, we have*

$$\sqrt{N}T_1 = H'(0)X_1 + R,$$

where $|R|$ is bounded by either of the two following quantities:

$$cM_0N^{-1}(b_2 + Q_2\|X_1\| + \|X_1\|^2), \qquad cMN^{-1}(b_2 + b_2\|X_1\| + \|X_1\|^2).$$

Furthermore,

$$\tau^2 = \mathbf{E}(H'(0)X_1)^2 + R_1,$$

$$|R_1| \le c\min\{M_0^2 N^{-1}(b_3 + Q_2 b_2);\ M^2 N^{-1}(b_3 + b_4)\},$$

$$\mathbf{E}(\sqrt{N}T_1)^3 = \mathbf{E}(H'(0)X_1)^3 + R_2,$$

where

$$|R_2| \le cN^{-1/2}\min\{M_0^3(b_4 + Q_1 b_3);\ M^3(b_4 + b_3 b_{3/2})\},$$

$$N^{5/2}\mathbf{E}T_1T_2T_{12} = \mathbf{E}(H'(0)X_1)(H'(0)X_2)(H''(0)X_1X_2) + R_3,$$

with

$$|R_3| \le cN^{-1/2}\min\{M_0^3(b_3b_2 + b_2^2 Q_1);\ M^3(b_2b_3 + b_2^2 b_{3/2})\}.$$

2.6. *Functionals of empirical distribution functions.* The results of this section differ only in notation from those for functions of sample means. We consider functionals of empirical distribution functions, and, for simplicity of notation, we shall assume that they have bounded derivatives and that they are functionals of distribution functions on the real line (but not empirical measures in general spaces). A simple example satisfying all those assumptions is a linear combination of order statistics.

Throughout this section we shall use the following notation. By $\eta_1, \dots, \eta_N$ we denote i.i.d. real random variables with common distribution function P. By P_N we denote the empirical distribution function corresponding to the sample $\eta_1, \dots, \eta_N$. Define the random processes

$$X_i(t), \qquad t \in \mathbf{R}, \qquad 1 \le i \le N$$

by $X_i(t) = \mathbb{1}\{\eta_i < t\} - P(t)$. Let

$$E_N = \sqrt{N}(P_N - P) = (X_1 + \cdots + X_N)/\sqrt{N}$$

denote the empirical process.

Assume that a functional T takes real values and that $T(P)$ and $T(P_N)$ are well defined. Define

$$S^0 = \sqrt{N}(T(P_N) - T(P)), \qquad S = \sqrt{N}(T(P_N) - \mathbf{E}T(P_N)).$$

We may write $P_N - P = E_N/\sqrt{N}$. Denoting $H_P(h) = T(P + h)$, we have

$$S = \sqrt{N}(H_P(E_N/\sqrt{N}) - \mathbf{E}H_P(E_N/\sqrt{N})).$$

Let us define the derivatives of T by the derivatives of H_P via $T^{(s)}(P + h) = H_P^{(s)}(h)$. In order to define the derivatives of H_P, we introduce a Banach space $\mathbf{B}$, which may depend on P and should be chosen in dependence on T and the particular problem. We shall assume that H_P: $\mathbf{B} \to \mathbf{R}$ admits Frechét derivatives, and we require that $\mathbf{B}$ contains the sample functions $X_1(t) =$

$\mathbb{1}\{\eta_1 < t\}$ a.s. Furthermore, we assume that the random variables X_1, E_N are well defined and take values in $\mathbf{B}$.

It is sufficient to prove Edgeworth expansions for S since

$$\left|\mathbf{E}T(P_N) - T(P) - \frac{1}{2N}\mathbf{E}T''(P)X_1^2\right| \le N^{-3/2}M_0(T)Q_3,$$

$$\left|\mathbf{E}T(P_N) - T(P) - \frac{1}{2N}\mathbf{E}T''(P)X_1^2\right| \le cN^{-3/2}M(T)(b_{7/2} + b_{5/2}).$$

In order to formulate Edgeworth expansions for S, we need some additional notation. Write

$$s_1^2 = \mathbf{E}(T'(P)X_1)^2, \qquad k = a_3 + 3d,$$

$$a_3 = s_1^{-3}\mathbf{E}(T'(P)X_1)^3, \qquad d = s_1^{-3}\mathbf{E}(T'(P)X_1)(T'(P)X_2)(T''(P)X_1X_2)$$

and

$$q = 1 - \sup_{|t| \ge s_1^2/(16M_0^3 b_3)} |\mathbf{E}\exp\{itT'(P)X_1\}|.$$

THEOREM 2.4. *The distribution function of S/s_1 may be uniformly approximated by $\{\Phi(x) - k\Phi'''(x)\}/(6\sqrt{N})$, and the error does not exceed*

$$cN^{-1}q^{-2}(b_3 + b_4 + b_4^2 + Q_2 + Q_2^4)(s_1^{-2}M_0^2(T) + s_1^{-4}M_0^4(T))$$

or, if $M(T) < \infty$,

$$cN^{-1}q^{-2}(b_3 + b_4 + b_4^2)(s_1^{-2}M^2(T) + s_1^{-4}M^4(T)).$$

2.7. *L-statistics.* We are going to apply the result for functionals of empirical distribution functions.

Let $\eta_1, \dots, \eta_N$ be i.i.d. real random variables with common distribution function P, and let P_N be the empirical distribution function corresponding to the sample $\eta_1, \dots, \eta_N$. Consider the statistic

$$l_N = N^{-1}\sum_{i=1}^{N} c_{iN}\eta_{i;N},$$

where $\eta_{1;N} \le \cdots \le \eta_{N;N}$ are the order statistics of $\eta_1, \dots, \eta_N$, and the coefficients $c_{1N}, \dots, c_{NN}$ are generated by a weight function J: $[0,1] \to \mathbf{R}$,

$$c_{iN} = N\int_{(i-1)/N}^{i/N} J(u)\,du.$$

Define

$$S^0 = \sqrt{N}(l_N - \mu), \qquad S = \sqrt{N}(l_N - \mathbf{E}l_N),$$

where

$$\mu = \int_{-\infty}^{\infty} xJ(P(x))\,dP(x).$$

It follows from (2.4) and (2.5) below, that

$$\left|\mathbf{E}l_N - \mu - \frac{\nu}{2N}\right| \le cM_0(J)N^{-3/2}(\mathbf{E}|\eta_1|^4 + \mathbf{E}|\eta_1|),$$

$$\left|\mathbf{E}l_N - \mu - \frac{\nu}{2N}\right| \le cM(J)N^{-3/2}(\mathbf{E}|\eta_1|^4 + \mathbf{E}|\eta_1|^{5/6}),$$

where

$$\nu = \int_{\mathbf{R}} J'(P(x))P(x)(1 - P(x))\,dx.$$

Therefore we may restrict ourselves to Edgeworth expansions for S.

Denote

$$b_s = \mathbf{E}|\eta_1|^s, \qquad x \wedge y = \min\{x, y\},$$

$$s_1^2 = \int_{\mathbf{R}^2} J(P(x))J(P(y))\{P(x \wedge y) - P(x)P(y)\}\,dx\,dy.$$

Furthermore, write $k = a_3 + 3d$, and

$$a_3 = s_1^{-3} \int_{\mathbf{R}^3} J(P(x))J(P(y))J(P(z))\{P(x \wedge y \wedge z) + U\}\,dx\,dy\,dz,$$

$$U = -P(x)P(y \wedge z) - P(y)P(x \wedge z) - P(z)P(x \wedge y) + 2P(x)P(y)P(z),$$

$$d = s_1^{-3} \int_{\mathbf{R}^3} J(P(x))J(P(y))J'(P(z))V(x, z)V(y, z)\,dx\,dy\,dz,$$

$$V(x, y) = P(y \wedge z) - P(y)P(z),$$

$$q = 1 - \sup\{|\mathbf{E}\exp\{itY\}|\colon\ |t| \ge c_0 s_1^2/(M_0^3(J)(b_3 + b_1))\},$$

where

$$Y = \int_{\mathbf{R}} J(P(x))(\mathbb{1}\{x < \eta_1\} - P(x))\,dx.$$

THEOREM 2.5. *There exists an absolute positive constant* c_0 *(see the definition of* q*) such that the distribution function of* S/s_1 *can be uniformly approximated by* $\{\Phi(x) - k\Phi'''(x)\}/(6\sqrt{N})$, *and the error does not exceed*

$$cN^{-1}q^{-2}(b_4 + b_4^2 + b_{2/3})(s_1^{-2}M_0^2(J) + s_1^{-4}M_0^4(J))$$

or, if $M(J) < \infty$,

$$cN^{-1}q^{-2}(b_4 + b_4^2 + b_1)(s_1^{-2}M^2(J) + s_1^{-4}M^4(J)).$$

The result of Theorem 2.5 is a consequence of Theorem 2.4. To see this, we represent S as a differentiable functional of P_N, and provide some useful estimates of moments.

If $\mathbf{E}|\eta_1| < \infty$ the boundedness of J is sufficient for the following representation [see Govindarajulu and Mason (1983)]:

$$l_N - \mu = \int_{-\infty}^{\infty} [\Psi(P_N(t)) - \Psi(P(t))]\, dt,$$

where

$$\Psi(x) = \int_x^1 J(u)\, du.$$

Therefore we may write

$$l_N - \mu = T(P_N) - T(P),$$

where

$$T(h) = \int_{-\infty}^{0} [\Psi(h(t)) - \Psi(0)]\, dt + \int_0^{\infty} \Psi(h(t))\, dt.$$

Let $\|\cdot\|_p$ denote the norm of the space $\mathbf{L}^p(\mathbf{R})$. Let $\mathbf{B}$ be the Banach space of functions with norm

$$\|x\| = \|x\|_1 + \|x\|_2 + \|x\|_3.$$

The functional $h \mapsto T(P+h)\colon \mathbf{B} \to \mathbf{R}$ is twice [or thrice if $M(J) < \infty$] Frechét differentiable and

$$M_0(T) \le M_0(J), \qquad M(T) \le M(J).$$

Define the random processes

$$X_i(t) = \mathbb{1}\{\eta_i < t\} - P(t), \qquad t \in \mathbf{R},\ 1 \le i \le N.$$

It is easy to verify that

$$\mathbf{E}\|X_1\|_p^s \le c(p,s)(\mathbf{E}|\eta_1|^{s/p} + (\mathbf{E}|\eta_1|)^{s/p}) \quad \text{for all} \quad p \ge 1, \quad s > 0.$$

The space $\mathbf{L}^p$, $p \ge 2$, is of type 2 and therefore [see de Acosta (1981)]

$$\sup_{N\ge 1} \mathbf{E}\|E_N\|_p^s \le c(p,s)\mathbf{E}\|X_1\|_p^s \le c(p,s)(\mathbf{E}|\eta_1|^{s/p} + (\mathbf{E}|\eta_1|)^{s/p}) \tag{2.4}$$

for $p, s \ge 2$. In the case of $\mathbf{L}^1$ we have

$$\sup_{N\ge 1} \mathbf{E}\|E_N\|_1^s \le c(p,\varepsilon)(\mathbf{E}|\eta|_1 + \mathbf{E}|\eta_1|^{s+\varepsilon}) \quad \text{for } s \ge 2,\ \varepsilon > 0. \tag{2.5}$$

Indeed, by means of the weight function $1 + |t|^\alpha$ with $\alpha s > s - 1$ and using Hölder's inequality, we obtain

$$\mathbf{E}\|E_N\|_1^s \le c(\alpha, s)\mathbf{E}\int_{\mathbf{R}} |E_N(t)|^s (1 + |t|^\alpha)^s\, dt.$$

Now we may use Fubini's theorem. Integrating by parts, we arrive at (2.5).

2.8. *U-statistics.* In this section we again consider the general case of i.i.d. random variables $X_1, \dots, X_N$ taking values in a measurable space. An example of a symmetric statistic is a U-statistic of fixed order r, which is given by

$$U = \sum_{1 \le i_1 < \cdots < i_r \le N} h(X_{i_1}, X_{i_2}, \dots, X_{i_r}),$$

where h is a fixed kernel function with

$$\mathbf{E}h(X_1, \dots, X_r) = 0, \qquad \mathbf{E}h^2(X_1, \dots, X_r) < \infty.$$

If Hoeffding's decomposition (see section 4) of $h(X_1, \dots X_r)$ is written as

$$h(X_1, \dots, X_r) = \sum_{k=1}^{r} \sum_{1 \le i_1 < \cdots < i_k \le r} h_k(X_{i_1}, X_{i_2}, \dots, X_{i_k}),$$

where

$$h_k(X_1, \dots, X_k) = \sum_{A \subset \Omega_k} (-1)^{k-|A|} \mathbf{E}(h(X_1, \dots, X_r)|A),$$

then Hoeffding's decomposition of U is given by

$$U = \sum_{k=1}^{r} \binom{N-k}{r-k} \sum_{1 \le i_1 < \cdots < i_k \le N} h_k(X_{i_1}, X_{i_2}, \dots, X_{i_k}).$$

We have $h_1(X_1) = \mathbf{E}(h(X_1, \dots, X_r)|X_1)$,

$$T_1 = \binom{N-1}{r-1} h_1(X_1), \qquad \tau^2 = N \binom{N-1}{r-1}^2 \mathbf{E}h_1^2(X_1),$$

$$\sigma^2 = \sum_{k=1}^{r} \binom{N-k}{r-k}^2 \binom{N}{k} \mathbf{E}h_k^2(X_1, \dots, X_k),$$

$$\varkappa = \alpha_3 + 3\delta, \qquad \alpha_3 = N^{3/2} \binom{N-1}{r-1}^3 \mathbf{E}h_1^3(X_1),$$

$$\delta = N^{5/2} \binom{N-2}{r-2} \binom{N-1}{r-1}^2 \mathbf{E}h_1(X_1)h_1(X_2)h_2(X_1, X_2).$$

The following theorem is a consequence of Theorems 1.1 and 1.2.

THEOREM 2.6. *The distribution function of U/σ or U/τ can be uniformly approximated by the corresponding Edgeworth expansion, and the error does not exceed*

$$\frac{c}{Nq^2(\sigma)} \left(\frac{\beta_4}{\sigma^4} + \frac{\gamma_3}{\sigma^3} + \frac{\Delta_3^2}{\sigma^2} \right)$$

or

$$\frac{c}{N\bar{q}^2(\tau)} \left(\frac{\beta_4}{\tau^4} + \frac{\gamma_3}{\tau^3} + \frac{\Delta_3^2}{\tau^2} \right),$$

where [*we write* $h_k = h_k(X_1, \dots, X_k)$]

$$\frac{\beta_4}{\sigma^4} \le \frac{\beta_4}{\tau^4} \le \frac{\mathbf{E}h_1^4}{(\mathbf{E}h_1^2)^2},$$

$$\frac{\gamma_3}{\sigma^3} \le \frac{\gamma_3}{\tau^3} \le \frac{8(r-1)^2\mathbf{E}|h_2|^3}{(\mathbf{E}h_1^2)^{3/2}},$$

$$\frac{\Delta_3^2}{\sigma^2} \le \frac{\Delta_3^2}{\tau^2} \le \sum_{k=3}^{r} \frac{c(r)(N-k)!\mathbf{E}h_k^2}{(N-3)!\mathbf{E}h_1^2}.$$

Note that the terms T_A of the statistic in Theorem 2.6 corresponding to cardinalities $|A| > 3$ are of much smaller order than required.

3. A smoothing inequality. Let F be a distribution function with characteristic function f. It is known [Prawitz (1972)] that, for all $x \in \mathbf{R}$ and $H > 0$,

$$F(x+) \le \frac{1}{2} + \text{V.P.} \int_{\mathbf{R}} \exp\{-ixt\} \frac{1}{H} K\left(\frac{t}{H}\right) f(t)\, dt \tag{3.1}$$

and

$$F(x-) \ge \frac{1}{2} - \text{V.P.} \int_{\mathbf{R}} \exp\{-ixt\} \frac{1}{H} K\left(-\frac{t}{H}\right) f(t)\, dt. \tag{3.2}$$

Notice that all integrals are real and that the integrands vanish unless $|t| \le H$. Here we write

$$F(x+) = \lim_{z \downarrow x} F(z), \qquad F(x-) = \lim_{z \uparrow x} F(z),$$

and V.P. denotes Cauchy's principal value,

$$\text{V.P.} \int_{\mathbf{R}} = \lim_{h \downarrow 0} \left(\int_{-\infty}^{-h} + \int_{h}^{\infty} \right).$$

Furthermore, $2K(s) = K_1(s) + iK_2(s)/(\pi s)$, where

$$K_1(s) = \begin{cases} 1 - |s|, & \text{for } |s| \le 1, \\ 0, & \text{for } |s| \ge 1, \end{cases}$$

$$K_2(s) = \begin{cases} \pi s(1 - |s|) \cot \pi s + |s|, & \text{for } |s| \le 1, \\ 0, & \text{for } |s| \ge 1. \end{cases}$$

It is known [see, e.g., Chung (1974), page 159] that if we redefine a distribution function G at discontinuity points (say x) as $2G(x) = G(x+) + G(x-)$, then

$$G(x) = \frac{1}{2} + \frac{i}{2\pi} \lim_{M \to \infty} \text{V.P.} \int_{|t| \le M} \exp\{-itx\} g(t) \frac{dt}{t}, \tag{3.3}$$

where g denotes the characteristic function of G. One can generalize (3.3) to functions of bounded variation.

The following lemma is elementary.

LEMMA 3.1. *For* $0 \le s \le 1$ *we have*

$$K_2(0) = 1, \qquad K_2(1) = 0, \qquad K_2(\tfrac{1}{2}) = \tfrac{1}{2},$$
$$K_2'(s) \le 0, \qquad K_2(s) + K_2(1-s) = 1.$$

Furthermore,

$$1 - 2(1-s)\sin^2\frac{\pi s}{2} \le K_2(s) \le 1 \quad \textit{for } 0 \le s \le \frac{1}{2},$$

$$0 \le K_2(s) \le 2s\,\sin^2\frac{\pi(1-s)}{2} \quad \textit{for } \frac{1}{2} \le s \le 1.$$

Let us compare the smoothing inequality (3.1) with the classical Esseen inequality for characteristic functions, namely,

$$|F(x) - G(x)| \le \frac{cA}{H} + c\int_{|t|\le H} |f(t) - g(t)|\frac{dt}{|t|}, \tag{3.4}$$

where $A = \sup_x |G'(x)|$. First of all the inequality (3.1) does not refer to a comparison of distribution functions. If a limit distribution or an Edgeworth approximation G is already known, then no regularity conditions on G [like $\sup_x |G'(x)| < \infty$] are needed. It follows from (3.1) and (3.3) that

$$F(x+) - G(x) \le I_1 + I_2 + I_3, \tag{3.5}$$

where

$$I_1 = \frac{1}{2}\int_{\mathbf{R}} \exp\{-ixt\}\frac{1}{H}K_1\left(\frac{t}{H}\right)f(t)\,dt,$$
$$I_2 = \frac{i}{2\pi}\lim_{M\to\infty} \text{V.P.}\int_{|t|\le M} \exp\{-ixt\}K_2\left(\frac{t}{H}\right)(f(t) - g(t))\frac{dt}{t},$$
$$I_3 = \frac{i}{2\pi}\lim_{M\to\infty} \text{V.P.}\int_{|t|\le M} \exp\{-ixt\}\left(K_2\left(\frac{t}{H}\right) - 1\right)g(t)\frac{dt}{t}.$$

The term I_1 corresponds to the concentration function of F, the term I_2 corresponds to the integral in (3.4) and I_3 corresponds to the first summand on the right-hand side in (3.4). It is essential for our purposes that one may interchange the order of integration and expectation $\mathbf{E}\exp\{itZ\} = f(t)$ in (3.5), where Z has d.f. F; this allows us to choose a random H depending on the sample and apply (3.5) conditionally (we call this procedure "data-dependent smoothing").

4. Hoeffding's decomposition and moment inequalities. In this section we decompose the symmetric statistic T into several parts and derive bounds for the variances of their conditional Hoeffding decompositions. The approach is similar to that of van Zwet (1984), but the calculations are more involved since the variances are not completely monotone, which was essential for the short proof in van Zwet (1984). We conclude this section with bounds for moments of any part of a U-statistic of order 2. These bounds are similar up to constants to those for sums of independent random variables.

Let $X_1, X_2, \dots, X_N$ denote i.i.d. random variables taking values in an arbitrary measurable space with common distribution $\mathbf{P}$. Consider the symmetric statistic $T = t(X_1, \dots, X_N)$. Without loss of generality, we shall assume that $\mathbf{E}T = 0$.

Let $\Omega = \{1, \dots, N\}$. For any $A \subset \Omega$, $|A|$ will denote the cardinality of A, and we shall write

$$\mathbf{E}(T|A) = \mathbf{E}(T|X_i,\ i \in A).$$

Thus $\mathbf{E}(T|A)$ denotes the conditional expectation of T given those X_i with index $i \in A$. In particular, $\mathbf{E}(T|\varnothing) = \mathbf{E}T = 0$ and $\mathbf{E}(T|\Omega) = T$.

Let us define T_A as an alternating sum of $\mathbf{E}(T|B)$ over all subsets $B \subset A$, including the empty set and A itself,

$$T_A = \sum_{B\subset A} (-1)^{|A|-|B|}\mathbf{E}(T|B). \tag{4.1}$$

Equation (4.1) expresses T_A in terms of the conditional expectations. It is easy to see that there is also an inverse relation

$$\mathbf{E}(T|B) = \sum_{A\subset B} T_A, \tag{4.2}$$

and, in particular,

$$T = \sum_{A\subset\Omega} T_A, \tag{4.3}$$

which is called Hoeffding's decomposition of T. The components T_A of T have the special property that

$$\mathbf{E}(T_A|C) = 0 \quad \text{unless } A \subset C.$$

Since $\mathbf{E}T = 0$, it follows that

$$\mathbf{E}T_A = 0 \quad \text{for all } A \subset \Omega,$$

$$\mathbf{E}T_A T_C = 0 \quad \text{if } A \neq C.$$

Hence (4.2) and (4.3) represent $\mathbf{E}(T|B)$ and T as sums of uncorrelated mean-zero random variables T_A, and, as a result,

$$\sigma^2(\mathbf{E}(T|B)) = \sum_{A\subset B} \mathbf{E}T_A^2, \tag{4.4}$$

$$\sigma^2 = \sigma^2(T) = \sum_{A\subset\Omega} \mathbf{E}T_A^2. \tag{4.5}$$

So far we have not used the fact that t is symmetric and $X_1, \dots, X_N$ are i.i.d. This implies that, for $A = \{i_1, \dots, i_k\}$,

$$T_A = g_k(X_{i_1}, \dots, X_{i_k}),$$

where g_k is symmetric in its arguments and depends on A only through its cardinality $|A| = k$. Writing T_i and T_{ij} as alternative notation for $T_{\{i\}}$ and $T_{\{i,j\}}$, we have, for example,

$$T_i = T_{\{i\}} = g_1(X_i) = \mathbf{E}(T|X_i),$$

$$T_{ij} = T_{\{i,j\}} = g_2(X_i, X_j) = \mathbf{E}(T|X_i, X_j) - \mathbf{E}(T|X_i) - \mathbf{E}(T|X_j)$$

and so forth. It follows that whenever $|A| = k$, then T_A has the same distribution as T_{Ω_k}, where $\Omega_k = \{1, \dots, k\}$. Hoeffding's decomposition now assumes the form

$$T = \sum_{A \subset \Omega} T_A = \sum_{1 \le i \le N} g_1(X_i) + \sum_{1 \le i < j \le N} g_2(X_i, X_j) + \sum_{1 \le i < j < k \le N} g_3(X_i, X_j, X_k) + \cdots + g_k(X_1, \dots, X_N),$$

and the variance decompositions (4.4) and (4.5) become

$$\sigma^2(\mathbf{E}(T|B)) = \sum_{k=1}^{|B|} \binom{|B|}{k} \mathbf{E}T^2_{\Omega_k} = \sum_{k=1}^{|B|} \binom{|B|}{k} \mathbf{E}g_k^2(X_1, \dots, X_k),$$

$$\sigma^2 = \sum_{k=1}^{N} \binom{N}{k} \mathbf{E}T^2_{\Omega_k} = \sum_{k=1}^{N} \binom{N}{k} \mathbf{E}g_k^2(X_1, \dots, X_k).$$

For $i \in \Omega$ and for a (not necessarily symmetric) statistic T, we define the difference

$$D_i T = T - \mathbf{E}(T|\Omega \setminus \{i\}). \tag{4.6}$$

Obviously, $D_i(D_i T) = D_i T$, $D_i D_j T = D_j D_i T$ and, for $A = \{i_1, \dots, i_k\} \subset \Omega$, the difference

$$D_A T = D_{i_1} \cdots D_{i_k} T$$

is well defined. We have $D_A D_B T = D_{A \cup B} T$,

$$D_A T = \sum_{B \subset A} (-1)^{|B|} \mathbf{E}(T|\Omega \setminus B).$$

The random variable T_A is related to the difference $D_A T$, and

$$T_A = \mathbf{E}(D_A T|A).$$

Denote $\Omega_m = \{1, \dots, m\}$ and consider repeated differences (4.6),

$$D_1 D_2 \cdots D_m T = \sum_{A \subset \Omega,\, \Omega_m \subset A} T_A \quad \text{for } m = 1, 2, \dots, N. \tag{4.7}$$

It follows from (4.7) that

$$\Delta_m^2 = N^{2m-1}\mathbf{E}(D_1 D_2 \cdots D_m T)^2 = N^{2m-1} \sum_{k=m}^{N} \binom{N-m}{k-m} \mathbf{E} T_{\Omega_k}^2,$$

$$\tfrac{1}{3} N^2 \sum_{k=3}^{N} k^3 \binom{N}{k} \mathbf{E} T_{\Omega_k}^2 \le \Delta_3^2 \le 3N^2 \sum_{k=3}^{N} k^3 \binom{N}{k} \mathbf{E} T_{\Omega_k}^2.$$

Denote

$$\Lambda_1 = \sum_{|A|=2,\ |A\cap\Omega_m|=2} T_A, \qquad \Lambda_2 = \sum_{|A|\ge 3,\ |A\cap\Omega_m|=2} T_A, \qquad \Lambda_3 = \sum_{A:\ |A\cap\Omega_m|\ge 3} T_A.$$

Define identically distributed random variables $\eta_1, \dots, \eta_m$ by

$$\eta_i = \sum_{|A|\ge 3,\ A\cap\Omega_m=\{i\}} T_A.$$

LEMMA 4.1. *We have*

$$\mathbf{E}\Lambda_3^2 \le m^3 N^{-5} \Delta_3^2, \qquad 3 \le m \le N, \tag{4.8}$$

$$\mathbf{E}\Lambda_2^2 \le m^2 N^{-4} \Delta_3^2, \qquad 2 \le m \le N, \tag{4.9}$$

$$\mathbf{E}|\Lambda_1|^3 \le c N^{-9/2} m^3 \gamma_3, \qquad 2 \le m \le N, \tag{4.10}$$

$$\mathbf{E}\eta_1^2 \le N^{-3} \Delta_3^2, \qquad 1 \le m \le N. \tag{4.11}$$

PROOF. Let us prove (4.8). Define the function $\varphi(m) = \mathbf{E}\Lambda_3^2$, for $2 \le m \le N$, and $\varphi(2) = 0$. The random variables T_A and T_B are uncorrelated unless $A = B$. Therefore the definition of Λ_3 implies

$$\varphi(m) = \sum_{r=3}^{N} c(r, m, N) \mathbf{E} T_{\Omega_r}^2, \qquad 2 \le m \le N,$$

where $c(r, m, N)$ is the number of subsets $A \subset \Omega$ such that $|A| = r$ and $|A \cap \Omega_m| \ge 3$. We have

$$\varphi(3) = \sum_{r=3}^{N} \binom{N-3}{r-3} \mathbf{E} T_{\Omega_r}^2 = \mathbf{E}(D_1 D_2 D_3 T)^2 = N^{-5}\Delta_3^2.$$

Define the difference operator d acting on the variable m as $d\varphi(m) = \varphi(m+1) - \varphi(m)$. Then

$$d\varphi(2) = \varphi(3) = \mathbf{E}(D_1 D_2 D_3 T)^2,$$
$$\varphi(m) = \varphi(2) + d\varphi(2) + \cdots + d\varphi(m-1),$$

and (4.8) will follow if we show that

$$d\varphi(m) \le m^2 \mathbf{E}(D_1 D_2 D_3 T)^2, \qquad 3 \le m \le N. \tag{4.12}$$

For (4.12) it is sufficient to verify that

$$dc(r, m, N) \le m^2 \binom{N-3}{r-3}. \tag{4.13}$$

Let us consider the class $\mathscr{A}_m$ of sets,

$$\mathscr{A}_m = \{A\colon A \subset \Omega,\ |A| = r,\ |A \cap \Omega_m| \le 2\}.$$

Then $\mathscr{A}_{m+1} \subset \mathscr{A}_m$ and

$$c(r, m, N) = |\{A\colon A \subset \Omega\}| - |\mathscr{A}_m|. \tag{4.14}$$

From (4.14) we see that

$$dc(r, m, N) = |\mathscr{A}_m| - |\mathscr{A}_{m+1}| = |\mathscr{A}_m \setminus \mathscr{A}_{m+1}|.$$

Under conditions $|A| = r$ and $A \subset \{1, \dots, N\}$, we have

$$\begin{aligned} A \in \mathscr{A}_m \setminus \mathscr{A}_{m+1} \quad &\Leftrightarrow \quad |A \cap \Omega_m| \le 2, |A \cap \Omega_{m+1}| \ge 3 \\ &\Leftrightarrow \quad |A \cap \Omega_m| = 2, \qquad m+1 \in A, \end{aligned}$$

and we have

$$dc(r, m, N) = \binom{m}{2}\binom{N-m-1}{r-3}.$$

However,

$$\binom{m}{2} \le m^2, \qquad \binom{N-m-1}{r-3} \le \binom{N-3}{r-3}, \quad \text{for } 3 \le m,$$

and (4.13) follows.

Let us prove (4.9). We have

$$\mathbf{E}\Lambda_2^2 = \sum_{r=3}^{N} \sum_{|A|=r, |A\cap\Omega_m|=2} \mathbf{E}T_{\Omega_r}^2 = \sum_{r=3}^{N} \binom{m}{2}\binom{N-m}{r-2}\mathbf{E}T_{\Omega_r}^2,$$

and (4.9) follows since

$$\binom{m}{2} \le m^2, \qquad \binom{N-m}{r-2} \le N\binom{N-3}{r-3} \quad \text{for } m \ge 3.$$

Let us prove (4.11). Obviously,

$$\mathbf{E}\eta_1^2 = \sum_{r=3}^{N} \binom{N-m}{r-1}\mathbf{E}T_{\Omega_r}^2 \le \sum_{r=3}^{N} \binom{N-1}{r-1}\mathbf{E}T_{\Omega_r}^2 \le N^2\mathbf{E}(D_1 D_2 D_3 T)^2,$$

since

$$\binom{N-m}{r-1} \le \binom{N-1}{r-1} \le N^2\binom{N-3}{r-3} \quad \text{for } r \ge 3.$$

Inequality (4.10) is a consequence of the following lemma.

In Lemma 4.2 we consider a symmetric function ψ taking values in a Banach space $(\mathbf{B}, \|\cdot\|)$ of type 2 (e.g., B is finite dimensional).

LEMMA 4.2. *Assume that* $\mathbf{E}\{\psi(X, X_1)|X\} = 0$. *Let* A *be a finite subset of the set of all integer pairs* (i, j) *such that* $1 \le i < j$. *Then*

$$\mathbf{E}\bigg\| \sum_{(i,j)\in A} \psi(X_i, X_j)\bigg\|^p \le |A| c(p)\gamma_p \qquad \text{for } 1 \le p \le 2, \tag{4.15}$$

$$\mathbf{E}\bigg\| \sum_{(i,j)\in A} \psi(X_i, X_j)\bigg\|^p \le |A|^{p/2} c(p)\gamma_p \quad \text{for } 2 \le p \le \infty, \tag{4.16}$$

where $\gamma_p = \mathbf{E}\|\psi(X, X_1)\|^p$, *and the constant* $c(p)$ *in* (4.15) *and* (4.16) *depends only on* p *and* $\mathbf{B}$.

PROOF. We use a decoupling inequality due to de la Peña (1992). For any convex increasing function Φ: $[0, \infty) \to [0, \infty)$, we have

$$\mathbf{E}\Phi\bigg(\bigg\| \sum_{(i,j)\in A} \psi(X_i, X_j)\bigg\|\bigg) \le \mathbf{E}\Phi\bigg(8\bigg\| \sum_{(i,j)\in A} \psi(X_i, Z_j)\bigg\|\bigg), \tag{4.17}$$

where the sequence $Z_1, Z_2, \ldots$ is an independent copy of $X_1, X_2, \ldots$.

In the proof of the lemma we shall write $c(p)$ instead of $c(p, \mathbf{B})$. Let $\eta_1, \ldots, \eta_m \in \mathbf{B}$, $\mathbf{E}\eta_j = 0$ be independent random variables assuming values in $\mathbf{B}$. A result of de Acosta (1981) yields

$$\mathbf{E}\bigg\| \sum_{j=1}^m \eta_j\bigg\|^p \le c(p) \sum_{j=1}^m \mathbf{E}\|\eta_j\|^p \quad \text{for } 1 \le p \le 2, \tag{4.18}$$

$$\mathbf{E}\bigg\| \sum_{j=1}^m \eta_j\bigg\|^p \le c(p)\bigg(\sum_{j=1}^m \mathbf{E}\|\eta_j\|^2\bigg)^{p/2} + c(p)\sum_{j=1}^m \mathbf{E}\|\eta_j\|^p, \quad 2 \le p < \infty, \tag{4.19}$$

provided the Banach space B is finite dimensional or of type 2.

Let us prove (4.15). Let $N < \infty$ be a natural number such that $(i, j) \in A \Rightarrow j \le N$. Define the sets $A_i = \{(i, j): (i, j) \in A\}$ for $i \ge 1$. Without loss of generality, we may assume that $A_i \ne \emptyset$ for $1 \le i \le K$, where a number $K \le N$, and that $A_i = \emptyset$ for $i > K$. Denote

$$V = \sum_{(i,j)\in A} \psi(X_i, Z_j) = \sum_{i=1}^K \xi_i,$$

where $\xi_i = \sum_{(i,j)\in A_i} \psi(X_i, Z_j)$. Due to the decoupling inequality, it is sufficient to estimate $\mathbf{E}\|V\|^p$. By a conditioning argument followed by an application of (4.18) to

$$\mathbf{E}_X\bigg\| \sum_{i=1}^K \xi_i\bigg\|^p \quad \text{and} \quad \mathbf{E}_Z\|\xi_i\|^p,$$

we easily obtain (4.15), where we write $\mathbf{E}_X = \mathbf{E}_{X_1, X_2, \ldots}$ for the conditional expectation given all r.v.'s but $X_1, X_2, \ldots$.

The proof of (4.16) is a little bit more tedious. Applying (4.19), we have

$$\mathbf{E}\|V\|^p \leq c(p)\mathbf{E}_Z\biggl(\sum_{i=1}^{K}\mathbf{E}_X\|\xi_i\|^2\biggr)^{p/2} + c(p)\sum_{i=1}^{K}\mathbf{E}\|\xi_i\|^p,$$

$$\mathbf{E}\|\xi_i\|^p \leq c(p)\mathbf{E}_X\biggl(\sum_{(i,j)\in A_i}\mathbf{E}_Z\|\psi(X_i, Z)\|^2\biggr)^{p/2} + c(p)m_i\gamma_p \leq c(p)m_i^{p/2}\gamma_p, \tag{4.20}$$

where $m_i = |A_i|$. Summing, we get

$$\sum_{i=1}^{K}\mathbf{E}\|\xi_i\|^p \leq |A|^{p/2}c(p)\gamma_p,$$

since $m_1 + \cdots + m_K = |A|$. Let us define an r.v. σ by $\mathbf{P}\{\sigma = i\} = m_i/|A|$ for $1 \leq i \leq K$. We assume that σ is independent of all other r.v.'s Then

$$\sum_{i=1}^{K}\mathbf{E}_X\|\xi_i\|^2 = |A|\mathbf{E}_{X,\sigma}\|\xi_\sigma\|^2/m_\sigma,$$

and applying Hölder's inequality, we get

$$\mathbf{E}_Z\biggl(\sum_{i=1}^{K}\mathbf{E}_X\|\xi_i\|^2\biggr)^{p/2} \leq |A|^{p/2}\mathbf{E}m_\sigma^{-p/2}\|\xi_\sigma\|^p \leq c(p)|A|^{p/2}\gamma_p$$

if we estimate $\mathbf{E}_Z\|\xi_\sigma\|^p$ by (4.13). Collecting these estimates concludes the proof of (4.16).

5. Proofs. We shall derive Theorem 1.1 from Theorem 1.2. The proof of Theorem 1.2 is laborious and needs much effort. One may obtain the result of Remark 1.3 by inspecting the proof of Theorem 1.2 and making relatively simple obvious changes. We conclude this section with the construction of a counterexample in Theorem 1.4.

By $c, c_1, \ldots$ we shall denote generic absolute positive constants, by $C, C_1, \ldots,$ generic sufficiently large absolute positive constants. If a constant depends on a parameter, say A, then we write $c(A)$. By $\mathbf{E}_{i,j,\ldots} = \mathbf{E}(\cdot|\Omega \setminus \{i, j, \ldots\})$ we shall denote the expectation with respect to random variables with pointed out indices $i, j, \ldots$; in other words, we condition on all random variables with indices different from $i, j, \ldots$.

PROOF OF THEOREM 1.1. It follows from (4.5) that

$$\sigma^2 = \tau^2 + R^2 \quad \text{where } 0 \leq R^2 \leq \frac{\gamma_2}{N} + \frac{\Delta_3^2}{N^2}.$$

We have

$$\sup_x |F(x) - G(x)| \le 1 + \frac{|\varkappa|}{\sigma^3\sqrt{N}} \le c\left(1 + \frac{\beta_3}{\sigma^3\sqrt{N}} + \frac{\gamma_2^{1/2}}{\sigma\sqrt{N}}\right). \tag{5.1}$$

We may assume that

$$\frac{\beta_3}{\sigma^3\sqrt{N}} + \frac{\gamma_2^{1/2}}{\sigma\sqrt{N}} \le 2. \tag{5.2}$$

Indeed, otherwise

$$1 \le \max\left\{\frac{\beta_3}{\sigma^3\sqrt{N}}, \frac{\gamma_2^{1/2}}{\sigma\sqrt{N}}\right\} \le \max\left\{\left(\frac{\beta_3}{\sigma^3\sqrt{N}}\right)^2, \left(\frac{\gamma_2^{1/2}}{\sigma\sqrt{N}}\right)^3\right\},$$

and Theorem 1.1 follows.

Now we consider the cases $2\tau^2 \ge \sigma^2$ and $2\tau^2 \le \sigma^2$.

The case $2\tau^2 \le \sigma^2$ is trivial, since the result is an easy consequence of (5.1) and (5.2). Indeed, the inequality $R^2/\sigma^2 \ge 1/2$ (since $2\tau^2 \le \sigma^2$) implies that at least one of the inequalities

$$\frac{\gamma_2}{\sigma^2 N} \ge \frac{1}{4}, \qquad \frac{\Delta_3^2}{\sigma^2 N^2} \ge \frac{1}{4}$$

is fulfilled. If the first inequality holds, then (5.1) and (5.2) imply

$$\sup_x |F(x) - G(x)| \le c \le c\left(\frac{\gamma_2}{\sigma^2 N}\right)^{3/2} \le \frac{c\gamma_3}{\sigma^3 N}.$$

If the first inequality is not fulfilled, then the second holds and

$$\sup_x |F(x) - G(x)| \le c \le \frac{c\Delta_3^2}{\sigma^2 N^2},$$

which concludes the proof in the case $2\tau^2 \le \sigma^2$.

In the case $2\tau^2 \ge \sigma^2$ the result follows from Theorem 1.2. Using the estimate of this theorem and (5.2), we reduce the problem to the estimation of

$$\sup_x \left|\Phi(x) - \Phi\left(\frac{\tau x}{\sigma}\right)\right|, \qquad \left|1 - \frac{\tau^3}{\sigma^3}\right|, \qquad \sup_x \left|\Phi'''(x) - \Phi'''\left(\frac{\tau x}{\sigma}\right)\right|. \tag{5.3}$$

The inequality $2\tau^2 \ge \sigma^2$ implies $\tau^2/\sigma^2 = 1 - R^2/\sigma^2$, where $R^2/\sigma^2 \le 1/2$, and one may estimate the quantities (5.3) by expansion in Taylor series. □

PROOF OF THEOREM 1.2. Roughly speaking, the proof of Theorem 1.2 is based on a reduction to sums of (conditionally) independent random variables and on elimination of the influence of the nonlinear part of the statistic. For this we use Taylor expansions and moment inequalities from Section 4 for symmetric statistics or their parts.

Without loss of generality, we shall assume that

$$\tau^2 = 1.$$

We shall prove that

$$\sup_x |F(x) - G(x)| \le \frac{c(\beta_4 + \gamma_3 + \Delta_3^2)}{q^2 N}, \tag{5.4}$$

where

$$q = q_N(\tau) = 1 - \sup\{|\mathbf{E}\exp\{itN^{1/2}T_1\}|\colon 1/\beta_3 \le |t| \le \sqrt{N}\},$$

and where, for brevity, we write $q_N(\tau) = q$, $\overline{F} = F$ and $\overline{G} = G$. Now Theorem 1.2 is a consequence of (5.4).

Without loss of generality in the proof of (5.4), we may assume that

$$\begin{gathered} \beta_4 \le cN, \qquad \beta_3 \le c\sqrt{N}, \qquad \gamma_2 \le cN, \\ q \ge CN^{-1}\ln N, \qquad N \ge C, \end{gathered} \tag{5.5}$$

where c is a sufficiently small positive absolute constant and where C is a sufficiently large absolute constant. Indeed, if at least one of inequalities (5.5) is not fulfilled, then the result of the theorem follows from the obvious estimate $\sup_x |F(x) - G(x)| \le c$, which we may assume as in (5.2).

The result of Theorem 1.2 follows from Lemmas 5.1–5.3. In Lemmas 5.2 and 5.3 we show that the integrals $I_{k,r}$, I and J from Lemma 5.1 do not exceed the right-hand side of (5.4).

LEMMA 5.1. *Let $H_1 = \sqrt{N}/\beta_3$, and let $\widehat{F}$ resp. $\widehat{G}$ denote the ch.f. of F resp. G. For some natural number $k > r \ge N/2$, we have*

$$\sup_x |F(x) - G(x)| \le cI_{k,r} + cI + cJ + R, \tag{5.6}$$

where R does not exceed the right-hand side of (5.4), *and*

$$I = \int_{|t|\le H_1} |\widehat{F}(t) - \widehat{G}(t)| \frac{dt}{|t|},$$

$$J = cq^{-2}\ln^2 N \int_{|t|\le H_1} |\mathbf{E}T_{N-1N}\exp\{itT\}|\,dt,$$

$$I_{k,r} = \frac{1}{qN}\int_{|t|\le H_1} \mathbf{E}(1+\theta)|\mathbf{E}_{1,\dots,r}\exp\{itT\}|\,dt, \qquad \theta = N\mathbf{E}_1\Bigl|\sum_{j=k}^{N} T_{1j}\Bigr|.$$

In the proofs of Lemmas 5.1–5.3 we use the conditional Hoeffding decomposition of certain parts of T. Fix a number m, $1 \le m \le N$, and denote $\Omega_m = \{1, \dots, m\}$. We may decompose T:

$$T = V + W, \tag{5.7}$$

where

$$W = \mathbf{E}(T|X_{m+1}, \dots, X_N) = \sum_{A\colon A\cap\Omega_m = \varnothing} T_A,$$

and where the sum $\sum_A$ is taken over all nonempty subsets $A \subset \Omega = \{1, \dots, N\}$.

For fixed $X_{m+1}, \dots, X_N$,

$$V = T - W = \sum_{A:\ A\cap\Omega_m \neq \emptyset} T_A$$

is a symmetric statistic of $X_1, \dots, X_m$. Let us consider its conditional Hoeffding decomposition:

$$V = \sum_{B\subset\Omega_m} V_B \quad \text{where } V_B = \sum_{A\subset\Omega,\ A\cap\Omega_m = B} T_A.$$

Writing V_i instead of $V_{\{i\}}$, we obtain

$$V = \sum_{i=1}^{m} V_i + \Lambda_1 + \Lambda_2 + \Lambda_3, \tag{5.8}$$

where $V_i = \sum_{A:\ A\cap\Omega_m = \{i\}} T_A$,

$$\Lambda_1 = \sum_{|A|=2,\ |A\cap\Omega_m|=2} T_A, \qquad \Lambda_2 = \sum_{|A|\geq 3,\ |A\cap\Omega_m|=2} T_A, \qquad \Lambda_3 = \sum_{A:\ |A\cap\Omega_m|\geq 3} T_A.$$

Let us decompose V_i:

$$V_i = T_i + \xi_i + \eta_i, \tag{5.9}$$

where

$$\xi_i = \sum_{j=m+1}^{N} T_{ij}, \qquad \eta_i = \sum_{|A|\geq 3,\ A\cap\Omega_m = \{i\}} T_A.$$

For fixed $X_{m+1}, \dots, X_N$ the random variables T_i, $1 \leq i \leq m$ (or ξ_i, $1 \leq i \leq m$ or η_i, $1 \leq i \leq m$), are i.i.d.

PROOF OF LEMMA 5.1. Let

$$m \approx Cq^{-1} \ln N,$$

be an integer, where the positive constant C is sufficiently large and is chosen later. Because of (5.5) such an m, $1 \leq m \leq N$, exists.

Before applying the Fourier transform [i.e., the smoothing inequality (3.1)], we shall remove certain parts of the statistic T in order to obtain a relatively small sum of (conditionally) independent summands. For each m, $3 \leq m \leq N$, we may write [see (5.7), (5.8) and Lemma 4.1]

$$T = \widetilde{T} + \Lambda_1 + \Lambda_2 + \Lambda_3, \qquad \widetilde{T} = \sum_{j=1}^{m} V_j + W,$$

where

$$\mathbf{E}|\Lambda_1|^3 \leq cN^{-9/2} m^3 \gamma_3, \qquad \mathbf{E}\Lambda_2^2 \leq m^2 N^{-4} \Delta_3^2, \qquad \mathbf{E}\Lambda_3^2 \leq m^3 N^{-5} \Delta_3^2. \tag{5.10}$$

We are going to apply a simple Slutzky argument. Due to (5.5), $\sup_x |G'(x)| \le c$. Therefore, instead of

$$\sup_x |F(x) - G(x)|, \quad \text{where } F(x) = \mathbf{P}\{T < x\},$$

we shall estimate

$$\sup_x |\widetilde{F}(x) - G(x)|, \quad \text{where } \widetilde{F}(x) = \mathbf{P}\{\widetilde{T} < x\}.$$

The error by this replacement does not exceed [use (5.10) and note that $m \le N$, $m = Cq^{-1}\ln N$]

$$\begin{aligned} &cq^{-1}N^{-1} + 2\mathbf{P}\{qN|\Lambda_1| \ge 1\} + 2\mathbf{P}\{qN|\Lambda_2| \ge 1\} + 2\mathbf{P}\{qN|\Lambda_3| \ge 1\} \\ &\quad \le cq^{-1}N^{-1} + cN^{-3/2}\gamma_3 \ln^3 N + cN^{-2}\ln^2 N\Delta_3^2, \end{aligned}$$

which is bounded by the right-hand side of (5.4).

In order to prove (5.6) we shall apply the smoothing inequality (3.1). The estimate (5.6) follows from

$$2\widetilde{F}(x+) \le 2G(x) + cI_{k,r} + cI + cJ + R, \tag{5.11}$$

$$2\widetilde{F}(x-) \ge 2G(x) - cI_{k,r} - cI - cJ + R \tag{5.12}$$

if the remainder $|R|$ is bounded by the right-hand side of (5.4). We shall prove only (5.11). The proof of (5.12) is similar; a minor difference being that instead of the smoothing inequality (3.1) one should use (3.2). Alternatively, one can derive (5.12) from (5.11), applying (5.11) to $-T$ (instead of T) and using symmetry arguments.

Define

$$\theta_1 = N\mathbf{E}_1\left|\sum_{j=m+1}^{k} T_{1j}\right|, \qquad \theta_2 = N\mathbf{E}_1\left|\sum_{j=k+1}^{N} T_{1j}\right|,$$

where the number k is chosen as $k \approx (N+m)/2$, and write

$$H = \frac{qN}{4(1+\theta_1+\theta_2)}.$$

Notice that H depends only on $X_{m+1}, \dots, X_N$.

The random variables $X_1, \dots, X_m$ are independent of $X_{m+1}, \dots, X_N$. Thus we may apply the smoothing inequality (3.1) conditionally with $H = qN/(4 + 4\theta_1 + 4\theta_2)$, which is random but independent of $X_1, \dots, X_m$. We get

$$2\widetilde{F}(x+) \le 1 + \mathbf{E}I_1 + \mathbf{E}I_2,$$

where

$$\begin{aligned} I_1 &= \frac{1}{H}\int_{\mathbf{R}} \exp\{-ixt\}K_1\left(\frac{t}{H}\right)f(t)\,dt, \\ I_2 &= \frac{i}{\pi}\int_{\mathbf{R}} \exp\{-ixt\}K_2\left(\frac{t}{H}\right)f(t)\frac{dt}{t}, \end{aligned}$$

and where $f(t) = \mathbf{E}_{1,\dots,m} \exp\{it\widetilde{T}\}$. Therefore (5.11) follows from

$$|\mathbf{E}I_1| \le cI_{k,r} + R, \tag{5.13}$$

$$|\mathbf{E}I_2 + 1 - 2G(x)| \le cI + cJ + R, \tag{5.14}$$

with the remainder terms $|R|$ bounded by the right-hand side of (5.4).

Estimation of $|\mathbf{E}I_1|$. In order to estimate $|\mathbf{E}I_1|$ we shall replace the random bound H in the integral I_1 by a nonrandom one, and $K_1(t/H)$ by 1. Furthermore, due to the special structure of H, we can interchange the integral and expectation with respect to a sufficiently large group of r.v.'s $X_{m+1}, \dots, X_N$.

We have

$$|\mathbf{E}I_1| \le |\mathbf{E}I_3| + \mathbf{E}I_4,$$

where

$$I_3 = \frac{1}{H}\int_{|t|\le H_1} \exp\{-ixt\} K_1\left(\frac{t}{H}\right) f(t)\,dt, \qquad H_1 = \frac{\sqrt{N}}{\beta_3},$$

$$I_4 = \frac{1}{H}\int_{H_1\le|t|} K_1\left(\frac{t}{H}\right)|f(t)|\,dt \le \frac{1}{H}\mathbb{1}\{H \ge H_1\}\int_{H_1\le|t|\le H}|f(t)|\,dt.$$

Let us show that

$$\mathbf{E}I_4 \le c(1+\Delta_3^2)/N. \tag{5.15}$$

The random variable W is independent of $X_1, \dots, X_m$, and we have

$$|f(t)| = |\mathbf{E}_1 \exp\{itV_1\}|^m$$

for fixed $X_{m+1}, \dots, X_N$. We have [see (5.9) and Lemma 4.1]

$$V_1 = T_1 + \xi_1 + \eta_1, \qquad \xi_1 = \sum_{j=m+1}^{N} T_{1j}$$

and

$$\mathbf{E}\eta_1^2 \le N^{-3}\Delta_3^2. \tag{5.16}$$

Let us consider the indicator function

$$\mathbb{1} = \mathbb{1}\{4\mathbf{E}_1|t\eta_1| \le q\}.$$

Estimating $|f(t)| \le 1$, we have

$$\mathbf{E}I_4 \le \mathbf{E}J_1 + \mathbf{E}J_2,$$

where

$$J_1 = \frac{1}{H}\mathbb{1}\{H \ge H_1\}\int_{H_1\le|t|\le H}\mathbb{1}|f(t)|\,dt,$$

$$J_2 = \frac{1}{H}\mathbb{1}\{H \ge H_1\}\int_{H_1\le|t|\le H}(1-\mathbb{1})\,dt.$$

Using Chebyshev's inequality and (5.16), we get

$$\mathbf{E}J_2 \le c\mathbf{E}q^{-2}H^2(\mathbf{E}_1|\eta_1|)^2 \le c\Delta_3^2/N,$$

since $H \le qN$.

In order to estimate J_1 we use a Taylor expansion

$$\mathbb{1}\,|\mathbf{E}_1 \exp\{itV_1\}| \le |\mathbf{E}\,\exp\{itT_1\}| + |t|\mathbf{E}_1|\xi_1| + \mathbb{1}\,\mathbf{E}_1|t\eta_1|.$$

However, the first summand does not exceed $1-q$ because of Cramér's condition (C). The second summand does not exceed $q/4$ for $|t| \le H$ due to the definition of H. The third summand is obviously less than $q/4$. Therefore

$$\mathbf{E}J_1 \le (1-q/2)^m \le \exp\{-mq/2\} \le c/N,$$

because of our choice of $m = Cq^{-1}\ln N$ with a sufficiently large C. Collecting the estimates, we get (5.15).

Let us estimate $\mathbf{E}I_3$. Note that $|K_1(s)-1| \le |s|$ for all $s \in \mathbf{R}$. Therefore

$$|\mathbf{E}I_3| \le |\mathbf{E}I_5| + \mathbf{E}I_6,$$

where

$$I_5 = \frac{1}{H}\int_{|t|\le H_1} \exp\{-ixt\} f(t)\,dt$$

and

$$\mathbf{E}I_6 \le \mathbf{E}\frac{2}{H}\int_0^{H_1} \frac{t}{H}\,dt \le \mathbf{E}\frac{H_1^2}{H^2} \le \frac{cq^{-2}\beta_3^{-2}(1+\gamma_2)}{N}.$$

It remains to estimate $|\mathbf{E}I_5|$. Let us replace $\widetilde{T}$ by T in $\mathbf{E}I_5$. We have [see (5.10)] $\widetilde{T} = T - \Lambda_1 - \Lambda_2 - \Lambda_3$. Expanding in a Taylor series, we get

$$\mathbf{E}I_5 = \mathbf{E}I_7 + R,$$

where

$$I_7 = \frac{1}{H}\int_{|t|\le H_1} \exp\{-ixt\}\exp\{itT\}\,dt,$$

$$|R| \le c\mathbf{E}\frac{H_1^2}{H}(|\Lambda_1| + |\Lambda_2| + |\Lambda_3|).$$

Now, applying Hölder's inequality several times together with Lemma 4.1, we obtain

$$|R| \le cq^{-2}\beta_3^{-2}(1+\gamma_3+\Delta_3^2)/N.$$

It remains to estimate $\mathbf{E}I_7$. Recall that $1/H = q^{-1}N^{-1}(1+\theta_1+\theta_2)$, and that θ_1 depends on $X_{m+1},\dots,X_k$ only, and θ_2 on $X_{k+1},\dots,X_N$ only. Therefore we may write $\mathbf{E}I_7$ as the sum of three integrals, and we can interchange the integrals and the expectation with respect to r.v.'s independent of θ_1 (or θ_2).

Due to symmetry and the i.i.d. assumption, the last two integrals are bounded by

$$q^{-1}N^{-1}\int_{|t|\le H_1}\mathbf{E}|\theta_2|\,|\mathbf{E}_{1,\dots,r}\exp\{itT\}|\,dt,$$

where r, $1\le r<(N+m)/2$, is arbitrary, $\theta_2=N\mathbf{E}_1|\sum_{j=k+1}^{N}T_{1j}|$ and $k\ge(N+m)/2$. Therefore

$$|\mathbf{E}I_7|\le cq^{-1}N^{-1}\int_{|t|\le H_1}\mathbf{E}(1+|\theta_2|)|\mathbf{E}_{1,\dots,r}\exp\{itT\}|\,dt,$$

and (5.13) follows.

Estimation of $\mathbf{E}I_2$. We shall proceed as in the estimation of $\mathbf{E}I_1$. We have

$$\mathbf{E}I_2=\mathbf{E}I_3+R,\qquad |R|\le\mathbf{E}I_4,$$

where

$$I_3=\frac{i}{\pi}\mathrm{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}K_2\left(\frac{t}{H}\right)f(t)\frac{dt}{t},\qquad H_1=\frac{\sqrt{N}}{\beta_3},$$

$$I_4=\mathbb{1}\{H\ge H_1\}\int_{H_1\le|t|\le H}|f(t)|\frac{dt}{|t|}.$$

The estimation of $\mathbf{E}I_4$ is similar to the proof of (5.15) and $\mathbf{E}I_4\le c(1+\Delta_3^2)/N$. Note that $|K_2(\tau)-1|\le 5\tau^2$. Therefore

$$\mathbf{E}I_3=\mathbf{E}I_5+R,\qquad |R|\le\mathbf{E}I_6,$$

where

$$I_5=\frac{i}{\pi}\mathrm{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}f(t)\frac{dt}{t},$$

$$\mathbf{E}I_6=c\mathbf{E}\int_{|t|\le H_1}\frac{|t|}{H^2}\,dt\le\frac{cq^{-2}\beta_3^2(1+\gamma_2)}{N}.$$

Collecting these estimates, we get $\mathbf{E}I_2=\mathbf{E}I_5+R$, where $|R|$ does not exceed the right-hand side of (5.4), and we may rewrite $\mathbf{E}I_5$ as

$$\mathbf{E}I_5=\frac{i}{\pi}\mathrm{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}\mathbf{E}\,\exp\{it\widetilde{T}\}\frac{dt}{t}.$$

As in the estimation of $\mathbf{E}I_1$, we may replace $\widetilde{T}$ by $T-\Lambda_1$ and $\mathbf{E}I_5$ by

$$I_6=\frac{i}{\pi}\mathrm{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}\mathbf{E}\,\exp\{it(T-\Lambda_1)\}\frac{dt}{t}.$$

In order to remove the term Λ_1, we expand $\exp\{-it\Lambda_1\}=1-it\Lambda_1+O(t^2\Lambda_1^2)$. Due to symmetry and the i.i.d. assumption,

$$I_6=I_7+I_8+R,$$

where R is bounded by the right-hand side of (5.4), and

$$I_7 = \frac{i}{\pi}\text{V.P.}\int_{|t|\le H_1} \exp\{-ixt\}\widehat{F}(t)\frac{dt}{t},$$

$$|I_8| \le m^2 \int_{|t|\le H_1} |\mathbf{E}T_{N-1N}\exp\{itT\}|\,dt = J.$$

We have

$$2G(x) = 1 + \frac{i}{\pi}\text{V.P.}\int_{\mathbf{R}} \exp\{-itx\}\widehat{G}(t)\frac{dt}{t},$$

where

$$\widehat{G}(t) = \left(1 + \frac{(it)^3\varkappa}{6\sqrt{N}}\right)\exp\left\{-\frac{t^2}{2}\right\}.$$

Estimating $|\varkappa| \le c\sqrt{N}$, we get

$$|1 + I_7 - 2G(x)| \le I + c\beta_4/N.$$

Collecting these estimates, we get (5.14), which concludes the proof of the lemma. □

The proof of Lemma 5.2 is a simplified version of that of Lemma 5.3.

LEMMA 5.2. *The integrals J and $I_{k,r}$ in Lemma* 5.1 *are bounded by the right-hand side of* (5.4).

PROOF. We shall estimate $I_{k,r}$ only since the estimation of J is similar and somewhat simpler. Using Hoeffding's decomposition, we may write $T = U + Z$, where U is a U-statistic:

$$U = \sum_{i=1}^{N} T_i + \sum_{1\le i<j\le N} T_{ij}, \tag{5.17}$$

and Z is the remaining part of T. According to Lemma 4.1 (with $m = N$), we have $\mathbf{E}Z^2 \le N^{-2}\Delta_3^2$. Expanding in a Taylor series, we get

$$I_{k,r} \le \mathbf{E}I_1 + \mathbf{E}I_2,$$

where

$$I_1 = q^{-1}N^{-1}\int_{|t|\le H_1} (1+\theta)|\mathbf{E}_{1,\dots,r}\exp\{itU\}|\,dt$$

and

$$\begin{aligned}\mathbf{E}I_2 &= \mathbf{E}q^{-1}N^{-1}\int_{|t|\le H_1}(1+\theta)|t|(\mathbf{E}_{1,\dots,r}Z^2)^{1/2}\,dt\\ &\le cq^{-1}N^{-2}H_1^2\mathbf{E}^{1/2}(1+\theta)^2\Delta_3 \le cq^{-1}N^{-1}(1+\gamma_2+\Delta_3^2),\end{aligned}$$

which does not exceed the right-hand side of (5.4).

Let us estimate $\mathbf{E}I_1$. Estimating the characteristic function in the integral by 1, we have

$$|\mathbf{E}I_1| \le cq^{-1}N^{-1}(1+\gamma_2) + cI_3,$$

where

$$I_3 = q^{-1}N^{-1}\mathbf{E}\int_{C\le|t|\le H_1}(1+|\theta|)|\mathbf{E}_{1,\dots,m}\exp\{itU\}|\,dt,$$

and where C is an absolute sufficiently large constant, and $m \le N/2 \le r$ is arbitrary. We choose

$$m = C_1\frac{N\ln|t|}{t^2},$$

with a sufficiently large absolute constant C_1. For fixed $X_{m+1},\dots,X_N$ we consider the conditional Hoeffding decomposition of U:

$$U = \sum_{i=1}^{m} V_i + \Lambda_1 + W, \qquad V_i = T_i + \xi_i,$$

$$\xi_i = \sum_{j=m+1}^{N} T_{ij}, \qquad \Lambda_1 = \sum_{1\le i<j\le m} T_{ij}.$$

Expanding in a Taylor series, we may replace U in I_3 by $\sum_{i=1}^m V_i + W$; the error does not exceed

$$cq^{-1}N^{-1}\int_{C\le|t|\le H_1}\mathbf{E}(1+|\theta|)|t\Lambda_1|\,dt,$$

and is bounded from above by the right-hand side of (5.4) due to our choice of m. Therefore, instead of I_3, we have to estimate

$$I_4 = q^{-1}N^{-1}\mathbf{E}\int_{C\le|t|\le H_1}(1+|\theta|)|h(t)|^m\,dt,$$

where $h(t) = \mathbf{E}_1\exp\{itV_1\}$. Let us consider the indicator function

$$\mathbb{I} = \mathbb{I}\{6N\mathbf{E}_1|\xi_1| \le |t|\}.$$

Then

$$\frac{1}{qN}\mathbf{E}\int_{C\le|t|\le H_1}(1+|\theta|)(1-\mathbb{I})\,dt \le \frac{cN}{q}\mathbf{E}\int_{C\le|t|\le H_1}\mathbf{E}(1+|\theta|)(\mathbf{E}_1|\xi_1|)^2\frac{dt}{t^2},$$

which is bounded from above by the right-hand side of (5.4). Therefore, instead of I_4, we have to estimate

$$I_5 = q^{-1}N^{-1}\mathbf{E}\int_{C\le|t|\le H_1}(1+|\theta|)\,\mathbb{I}\,|h(t)|^m\,dt.$$

However,

$$\mathbb{I}\,|h(t)| \le |\mathbf{E}\exp\{itT_1\}| + \mathbb{I}\,|t|\mathbf{E}_1|\xi_1| \le 1 - \frac{t^2}{6N}, \tag{5.18}$$

since

$$|\mathbf{E}\exp\{itT_1\}| \le 1 - \frac{t^2}{3N}, \qquad \mathbb{1}\,|t|\mathbf{E}|\xi_1| \le \frac{t^2}{6N}.$$

The estimate (5.18) implies $\mathbb{1}\,\big|h(t)\big|^m \le |t|^{-C}$ with a sufficiently large C, and therefore I_5 is bounded by the right-hand side of (5.4), which concludes the proof of the lemma. □

LEMMA 5.3. *The integral I in Lemma* 5.1 *does not exceed the right-hand side of* (5.4).

PROOF. Let us show that

$$I \le I_1 + I_2, \tag{5.19}$$

where

$$I_1 = \int_{C\le|t|\le H_1} |\widehat{F}(t) - \widehat{G}(t)|\frac{dt}{|t|},$$

$$I_2 = \int_{|t|\le C} |\widehat{F}(t) - \widehat{G}(t)|\frac{dt}{|t|} \le c\frac{\beta_4 + \gamma_2 + \Delta_3^2}{N}.$$

Let us start with the (unconditional) Hoeffding decomposition $T = U + Z$, where U is the U-statistic (5.17), and, according to Lemma 4.1, $\mathbf{E}Z^2 \le \Delta_3^2/N^2$. Expanding in a Taylor series, we get

$$I_2 \le I_3 + c(\mathbf{E}Z^2)^{1/2}, \qquad (\mathbf{E}Z^2)^{1/2} \le c(1 + \Delta_3^2)/N,$$

where

$$I_3 = \int_{|t|\le C} |\mathbf{E}\exp\{itU\} - \widehat{G}(t)|\frac{dt}{|t|}.$$

For $|t| \le C$ the function $f(x) = \exp\{itx\}$ is smooth and has bounded derivatives. Therefore, applying Lemma 6.1, we get $I_3 \le c(\beta_4 + \gamma_2)/N$, and (5.19) follows.

Henceforth we shall write $R \prec D$ if there is an absolute constant c such that

$$\int_{C\le|t|\le H_1} |R|\frac{dt}{|t|} \le cD,$$

as well as $A \simeq B$ if $A - B \prec (\beta_4 + \gamma_3 + \Delta_3^2)/N$. In view of (5.19), we have to show that $\widehat{F}(t) \simeq \widehat{G}(t)$.

Let us choose

$$m = C_1\frac{N\ln|t|}{t^2},$$

with a sufficiently large absolute constant C_1. Using the conditional Hoeffding decomposition (5.10), we have

$$T = \widetilde{T} + \Lambda_1 + \Lambda_2 + \Lambda_3, \qquad \widetilde{T} = \sum_{j=1}^{m} V_j + W,$$

where W is independent of $X_1, \dots, X_m$. Expanding in a Taylor series and using (5.10) for the estimation of variances of Λ_2 and Λ_3, we obtain

$$\widehat{F}(t) \simeq \mathbf{E} \exp\{it(\widetilde{T} + \Lambda_1)\}.$$

The error by this replacement does not exceed

$$c|t| m N^{-2} \Delta_3 \prec (1 + \Delta_3^2)/N.$$

We have $\Lambda_1 = \sum_{1 \le j < k \le m} T_{jk}$. Expanding in powers of Λ_1 and using the symmetry and the i.i.d. property, we obtain

$$\widehat{F}(t) \simeq \mathbf{E}\left(1 + it\binom{m}{2} T_{12}\right) \exp\{it\widetilde{T}\}, \tag{5.20}$$

with an error not exceeding $cN^{-3} t^2 m^2 \gamma_2 \prec \gamma_2/N$.

Recall that $V_j = T_j + \xi_j + \eta_j$ [see (5.9)], and introduce the notation

$$\begin{aligned} h_1(t) &= \mathbf{E} \exp\{itT_1\}, \\ h_2(t) &= \mathbf{E}_1 \exp\{itT_1 + it\xi_1\}, \\ h_3(t) &= \mathbf{E}_1 \exp\{itV_1\}, \\ h_4(t) &= it\binom{m}{2} \mathbf{E}_{1,2} T_{12} \exp\{itV_1 + itV_2\}. \end{aligned}$$

Then we may rewrite (5.20) as

$$\widehat{F}(t) \simeq \mathbf{E} \exp\{itW\}(h_3^m(t) + h_4(t) h_3^{m-2}(t)).$$

Let us show that

$$h_4(t) \simeq h_5(t) = it\binom{m}{2} \mathbf{E}_{1,2} T_{12} \exp\{itT_1 + itT_2\}, \tag{5.21}$$

and that

$$|h_5(t)| \le |t|^3 m^2 N^{-5/2} \gamma_2^{1/2} \le c \quad \text{for } C \le |t| \le H_1. \tag{5.22}$$

Expanding in powers of η_1, η_2 and using

$$\binom{m}{2} \le m^2, \qquad m = CNt^{-2} \ln|t|,$$

we get

$$h_4(t) \simeq it\binom{m}{2} \mathbf{E}_{1,2} T_{12} \exp\{itT_1 + it\xi_1 + itT_2 + it\xi_2\}.$$

Now we expand in powers of $\xi_1+\xi_2$. Using symmetry and the i.i.d. assumption, we get $h_4(t) \simeq h_5(t) + h_6(t)$, where

$$h_6(t) = 2(it)^2 \binom{m}{2} \mathbf{E}_{1,2}\xi_1 T_{12} \exp\{itT_1 + itT_2\}.$$

Expanding in powers of T_2, we have

$$\mathbf{E}|h_6(t)| \le |t|^3 m^2 \mathbf{E}_{1,2}|\xi_1 T_{12} T_2| \prec \gamma_2/N,$$

and (5.21) follows. Finally, we derive (5.22) from (5.21) by expansions, using $\gamma_2 \le cN$.

It follows from (5.21) that

$$\widehat{F}(t) \simeq \mathbf{E}\, \exp\{itW\}(h_3^m(t) + h_5(t)h_3^{m-2}(t)). \tag{5.23}$$

Let us consider the following indicator functions:

$$\mathbb{I}_1 = \mathbb{I}\{N\mathbf{E}_1\eta_1^2 < c_1\}, \qquad \mathbb{I}_2 = \mathbb{I}\{N\mathbf{E}_1\xi_1^2 < c_1\},$$

choosing a sufficiently small absolute positive constant c_1. Notice that the functions $\mathbb{I}_1$, $\mathbb{I}_2$ depend on $X_{m+1}, \dots, X_N$ only.

Using (5.22), Chebyshev's inequality and estimates of moments of Lemma 4.1, we get

$$\mathbf{E}(1 - \mathbb{I}_1 + 1 - \mathbb{I}_2)(1 + |h_5(t)|) \prec c(\gamma_3 + \Delta_3^2)/N.$$

Therefore (5.23) yields

$$\widehat{F}(t) \simeq \mathbf{E}\,\mathbb{I}_1\mathbb{I}_2 \exp\{itW\}(h_3^m(t) + h_5(t)h_3^{m-2}(t)). \tag{5.24}$$

Let us show that

$$\max\{\mathbb{I}_1\mathbb{I}_2|h_3(t)|;\ \mathbb{I}_2|h_2(t)|;\ |h_1(t)|\} \le \exp\left\{-\frac{t^2}{6N}\right\} \quad \text{for } |t| \le H_1. \tag{5.25}$$

We shall estimate $\mathbb{I}_1\mathbb{I}_2\big|h_3(t)\big|$ only. We have

$$V_1 = T_1 + P, \quad \text{where } P = \xi_1 + \eta_1.$$

Expanding in powers of P, we get

$$\mathbb{I}_1\mathbb{I}_2|h_3(t)| \le |h_1(t)| + \mathbb{I}_1\mathbb{I}_2|\mathbf{E}_1 tP\, \exp\{itT_1\}| + \mathbb{I}_1\mathbb{I}_2 t^2\mathbf{E}_1P^2.$$

However,

$$\mathbb{I}_1\mathbb{I}_2\mathbf{E}_1P^2 \le 2\,\mathbb{I}_2\mathbf{E}_1\xi_1^2 + \mathbb{I}_1\mathbf{E}_1\eta_1^2 \le \frac{1}{12N},$$

provided that the constant c_1 in the definitions of $\mathbb{I}_1$ and $\mathbb{I}_2$ is sufficiently small, and

$$\mathbb{I}_1\mathbb{I}_2|\mathbf{E}_1 tP\exp\{itT_1\}| \le t^2\,\mathbb{I}_2\mathbf{E}_1|\xi_1T_1| + \mathbb{I}_1 t^2\mathbf{E}_1|\eta_1T_1| \le \frac{t^2}{12N},$$

$$|h_1(t)| \le 1 - \frac{t^2}{2N} + \frac{\beta_3|t|^3}{6N\sqrt{N}} \le 1 - \frac{t^2}{3N}.$$

Now (5.25) follows.

For a natural number $k \geq m/4$ by (5.25) we have

$$|h_i^k(t)| \leq \exp\{-kt^2/6\} \leq |t|^{-A} \quad \text{for } i = 1, 2, 3 \text{ and } |t| \leq H_1, \tag{5.26}$$

where the number A can be chosen arbitrarily large by choice of the constant in the definition of m. This means that while estimating (5.24) we may neglect powers of t, provided we keep at least $m/4$ factors $h_i(t)$ and the product $\mathbb{I}_1 \mathbb{I}_2$ as balancing factors.

In view of (5.26), we derive from (5.24), expanding in a Taylor series,

$$\widehat{F}(t) \simeq \mathbf{E}\,\mathbb{I}_1\,\mathbb{I}_2 \exp\{itW\}\left(h_3^m(t) + i\,\delta t^3 N^{-5/2}\binom{m}{2}h_3^{m-2}(t)\right), \tag{5.27}$$

where $\delta = N^{5/2}\mathbf{E}T_1T_2T_{12}$.

Let us now replace $h_3(t)$ in (5.27) by $h_2(t)$. Choose k approximately equal to $m/2$. Then

$$h_3^k(t) = \mathbf{E}_{1,\dots,k} \exp\left\{it\sum_{j=1}^k V_j\right\} = h_2^k(t) + R,$$

where

$$\mathbf{E}|R| \leq c\mathbf{E}|t|\mathbf{E}_1\left|\sum_{j=1}^k \eta_j\right| \leq c|t|\sqrt{k}(\mathbf{E}\eta_1^2)^{1/2} \leq c(\ln|t|)^{1/2}\Delta_3/N.$$

Therefore we may replace $h_3^k(t)$ by $h_2^k(t)$. Arguing similarly, we replace the remaining factors $h_3(t)$ by $h_2(t)$, obtaining

$$\widehat{F}(t) \simeq \mathbf{E}\,\mathbb{I}_1\,\mathbb{I}_2 \exp\{itW\}\left(h_2^m(t) + it^3\delta N^{-5/2}\binom{m}{2}h_2^{m-2}(t)\right). \tag{5.28}$$

Now we shall replace $h_2^k(t)$ by $h_1^k(t)$ in (5.28). We have

$$h_2^k(t) = \mathbf{E}_{1,\dots,k} \exp\left\{it\sum_{j=1}^k (T_j + \xi_j)\right\} = h_1^k(t) + itkh_1^{k-1}(t)\mathbf{E}_1\xi_1 \exp\{itT_1\} + R,$$

where

$$\mathbf{E}|R| \leq ct^2\mathbf{E}\mathbf{E}_{1,\dots,k}(\xi_1 + \cdots + \xi_k)^2 \leq kt^2\gamma_2/N^2.$$

Furthermore,

$$itkh_1^{k-1}(t)\mathbf{E}_1\xi_1 \exp\{itT_1\} = -t^2kh_1^{k-1}(t)\mathbf{E}_1\xi_1T_1 + R,$$

where

$$\mathbf{E}|R| \leq |t|^3kN^{-1}\mathbf{E}(\mathbf{E}_1\xi_1^2)^{1/2} \leq |t|^3k\gamma_2^{1/2}/N^2.$$

By similar arguments we may replace the remaining factors $h_2(t)$ by $h_1(t)$, obtaining

$$\widehat{F}(t) \simeq h_1^m(t)G_1(t) + h_1^{m-1}(t)G_2(t) + h_1^{m-2}(t)G_3(t), \tag{5.29}$$

where

$$G_1(t) = \mathbf{E}\,\mathbb{1}_1\mathbb{1}_2\exp\{itW\}, \tag{5.30}$$

$$G_2(t) = (it)^2 m\mathbf{E}T_1\xi_1\mathbb{1}_1\mathbb{1}_2\exp\{itW\}, \tag{5.31}$$

$$G_3(t) = (it)^3\delta N^{-5/2}\binom{m}{2}\mathbf{E}\,\mathbb{1}_1\mathbb{1}_2\exp\{itW\}. \tag{5.32}$$

Using Chebyshev's inequality, we can finally remove $\mathbb{1}_1\mathbb{1}_2$ in (5.30)–(5.32).

Since the symmetric statistic W depends on $X_{m+1},\dots,X_N$ only, we may write (applying a simplified variation of Lemma 4.1)

$$W = W_1 + \cdots + W_{N-m},$$

where

$$W_1 = \sum_{j=m+1}^{N} T_j, \qquad W_2 = \sum_{m+1\le j<k\le N} T_{jk},$$

$$\mathbf{E}(W_2+\cdots+W_{N-m})^2 \le (\gamma_2+\Delta_3^2)/N, \qquad \mathbf{E}(W_3+\cdots+W_{N-m})^2 \le \Delta_3^2/N^2.$$

Expanding in a Taylor series, we may assume that (5.29) holds with

$$\begin{aligned} G_1(t) &= \mathbf{E}\exp\{itW_1 + itW_2\},\\ G_2(t) &= (it)^2 m\mathbf{E}T_1\xi_1\exp\{itW_1\}\\ G_3(t) &= (it)^3\delta N^{-5/2}\binom{m}{2}\mathbf{E}\,\exp\{itW_1\}. \end{aligned} \tag{5.33}$$

To prove the lemma, it is sufficient to show that, for G_j defined in (5.33),

$$h_1^m(t)G_1(t) \backsimeq \left(1 + \frac{(it)^3\alpha_3}{6\sqrt{N}} + \frac{(it)^3(N-m)^2\delta}{2N^2\sqrt{N}}\right)\exp\left\{-\frac{t^2}{2}\right\}, \tag{5.34}$$

$$h_1^{m-1}(t)G_2(t) \backsimeq (it)^3 m(N-m)\delta N^{-5/2}\exp\left\{-\frac{t^2}{2}\right\}, \tag{5.35}$$

$$h_1^{m-2}(t)G_3(t) \backsimeq \frac{(it)^3\delta m^2}{2N^2\sqrt{N}}\exp\left\{-\frac{t^2}{2}\right\}, \tag{5.36}$$

which, together with (5.29), finally implies the desired result $\widehat{F}(t)\backsimeq\widehat{G}(t)$.

For the proof of (5.34) we apply the expansion (6.2) of Lemma 6.1. We get

$$\begin{aligned} h_1^m(t)G_1(t) \backsimeq\ & h_1^m(t)\left(1 + \frac{(it)^3}{6\sqrt{N}}\left(\alpha_3\left(1-\frac{m}{N}\right) + 3\delta\left(1-\frac{m}{N}\right)^2\right)\right)\\ &\times\exp\left\{-\frac{(1-m/N)t^2}{2}\right\}. \end{aligned}$$

Similarly, applying Lemma 6.2 to $h_1^m(t)$, we obtain (5.34).

It remains to prove (5.35) and (5.36). We shall give the proof of (5.35) only [the proof of (5.36) is similar but simpler]. Using the i.i.d. assumption, we have

$$h_1^{m-1}(t)G_2(t) = (it)^2 m(N-m)h_1^{N-2}\mathbf{E}T_1T_{12}\exp\{itT_2\}.$$

Expanding the exponential function, we get

$$h_1^{m-1}G_2(t) \backsimeq (it)^3 m(N-m)\delta N^{-5/2}h_1^{N-2}(t).$$

An application of Lemma 6.2 now establishes (5.35). □

PROOF OF THEOREM 1.4. Let us choose

$$T = T_N = \sum_{i=1}^{N} T_i + \sum_{1\le i<j\le N} T_{ij}$$

as a sequence of U-statistics of second order. To prove the theorem, it is sufficient to verify (1.11) for a subsequence of N's. Therefore we have to construct T_1 and T_{12}, verify (1.9) and (1.10) and prove

$$N \inf_{|G'|\le A} \sup_x |F(x) - G(x)| \ge c > 0 \tag{5.37}$$

for sufficiently large $N = m^2$ only, where m are odd natural numbers.

One gets the result, dividing T by $\tau = (N\mathbf{E}T_1^2)^{1/2}$ in the following construction. Let $\{x\}$ denote the difference of x to the nearest integer:

$$\{x\} = x - k \quad \text{if } k \text{ is an integer such that } |x-k| < \tfrac{1}{2}$$

(define $\{x\} = 0$ for $x = k + 1/2$). Let $[x]$ be the nearest integer to x, that is, $\{x\} + [x] = x$. The functions $[x]$ and $\{x\}$ are odd.

Define the statistic

$$T = \frac{1}{12\sqrt{N}} + \sum_{j=1}^{N} \frac{X_j}{\sqrt{N}}\left(1 - \sum_{k=1}^{N} \frac{\{\sqrt{N}X_k\}}{N}\right),$$

where $X_1, X_2, \ldots$ are i.i.d. random variables having uniform distribution on the interval $[-1/2, 1/2]$. It is easy to see that

$$T_1 = \frac{X_1}{\sqrt{N}} - \frac{X_1\{mX_1\}}{N\sqrt{N}} + \frac{1}{12N^2}$$

and

$$N^{3/2}T_{12} = -X_1\{mX_2\} - \{mX_1\}X_2$$

satisfy $|\sqrt{N}T_1| \le 1$ and $|N^{3/2}T_{12}| \le 1/2$. Writing

$$V_N = \sum_{k=1}^{N} \frac{\{mX_k\}}{\sqrt{N}}, \qquad W_N = \sum_{k=1}^{N} \frac{[mX_k]}{N},$$

we may decompose T as follows:

$$T = \frac{1}{12\sqrt{N}} + \left(W_N + \frac{V_N}{\sqrt{N}}\right)\left(1 - \frac{V_N}{\sqrt{N}}\right).$$

It is easy to verify that the random variables $\{mX_1\}$ and $[mX_1]$, as well as V_N and W_N are independent.

Let $A_N = \{u_N < T \le v_N\}$,

$$u_N = 1 + \frac{1}{12\sqrt{N}} - \frac{\delta}{N}, \qquad v_N = 1 + \frac{1}{12\sqrt{N}}, \qquad 0 < \delta < 1.$$

Due to the independence of V_N and W_N, we have

$$\mathbf{P}\{A_N\} \ge \mathbf{P}\{A_N \text{ and } W_N = 1\} = \mathbf{P}\{|V_N| < \sqrt{\delta}\}\mathbf{P}\{W_N = 1\}.$$

The random variable $\{mX_1\}$ is uniformly distributed. Therefore the local limit theorem for densities implies [see Petrov (1975), Chapter 7]

$$\mathbf{P}\{|V_N| < \sqrt{\delta}\} \ge c\sqrt{\delta} \quad \text{for some } c > 0.$$

The random variable $[mX_1]$ assumes integer values. Using the explicit formula for the ch.f. of $[mX_1]$ and proceeding as in the proof of Theorem 2 in Chapter 7 of Petrov (1975), we have

$$\mathbf{P}\{W_N = 1\} \ge \frac{c}{N} \quad \text{for some } c > 0.$$

Therefore $\mathbf{P}\{A_N\} \ge c\sqrt{\delta}/N$. Let

$$\Delta_1 = F(v_N) - G(v_N), \qquad \Delta_2 = G(u_N) - F(u_N).$$

Thus, for any given function $G: \mathbf{R} \to \mathbf{R}$ such that $\sup_x |G'(x)| \le A$, we get

$$\Delta_1 + \Delta_2 \ge (c\sqrt{\delta} - A\delta)/N \ge c_1\sqrt{\delta}/N, \qquad c_1 > 0,$$

if $\delta = \delta(A)$ is sufficiently small. Therefore $\max\{\Delta_1, \Delta_2\} \ge c_1\sqrt{\delta}/(2N)$, which implies (5.37).

The Cramér condition (1.10) may be verified using Taylor expansions and separating the cases $|t| \le \delta_1$ and $\delta_1 \le |t| \le \delta_2 N$ for some appropriate δ_1 and δ_2. The case $|t| \ge \delta_2 N$ again needs special arguments. □

6. Auxiliary results Let T denote a symmetric statistic with Hoeffding decomposition (4.3). For $1 \le m \le N$ consider the statistic

$$V = \sum_{i=1}^{m} T_i + \sum_{1 \le i < j \le m} T_{ij}.$$

LEMMA 6.1. *Let $f: \mathbf{R} \to \mathbf{C}^1$ be a sufficiently smooth function. Then*

$$|\mathbf{E}f(V/\tau) - \mathbf{E}f(\sqrt{m/N}\eta)| \le c(f)mN^{-3/2}(\tau^{-3}\beta_3 + \tau^{-5/3}\gamma_{5/3}), \tag{6.1}$$

where $c(f) = \|f'\| + \|f''\| + \|f'''\|$ *and* $\|f\| = \sup_x |f(x)|$. *Furthermore,*

$$\mathbf{E}f\left(\frac{V}{s}\right) = \mathbf{E}f\left(\sqrt{\frac{m}{N}}\eta\right) + \frac{\varkappa_m}{6\tau^3\sqrt{N}}\mathbf{E}f'''\left(\sqrt{\frac{m}{N}}\eta\right) + R, \tag{6.2}$$

where η *is a standard* $(0,1)$ *normal r.v.,*

$$\varkappa_m = mN^{-1}\alpha_3 + 3m^2N^{-2}\delta, \qquad |R| \le c(f)mN^{-2}(\tau^{-4}\beta_4 + \tau^{-2}\gamma_2),$$

and $c(f) = \|f^{(2)}\| + \|f^{(4)}\| + \|f^{(6)}\|$.

PROOF. Without loss of generality, we may assume that $\tau^2 = 1$. We shall give the proof of (6.2) only since the proof of (6.1) is simpler. Expanding, we get

$$\mathbf{E}f(V) = \mathbf{E}f(A) + \mathbf{E}f'(A)B + R,$$

where

$$A = \sum_{i=1}^{m} T_i, \qquad B = \sum_{1\le i<j\le m} T_{ij},$$

$$|R| \le c\|f''\|\mathbf{E}B^2, \qquad \mathbf{E}B^2 \le cm\gamma_2/N^2.$$

It is easy to show [see Bentkus, Götze, Paulauskas and Račkauskas (1991)] that

$$\left|\mathbf{E}f(A) - \mathbf{E}f\left(\sqrt{\frac{m}{N}}\eta\right) - \frac{m\alpha_3}{6N\sqrt{N}}\mathbf{E}f'''\left(\sqrt{\frac{m}{N}}\eta\right)\right| \le \frac{c(\|f^{(4)}\| + \|f^{(6)}\|)m\beta_4}{N^2}.$$

We have

$$\mathbf{E}f'(A)B = \binom{m}{2}\mathbf{E}T_{12}f'(A).$$

Let us write $A = T_1 + A_1$. Then

$$\mathbf{E}N^{3/2}T_{12}f'(A) = \mathbf{E}T_1N^{3/2}T_{12}f''(A_1) + R_1,$$

where

$$|R_1| \le c|\mathbf{E}(1-\theta)T_1^2N^{3/2}T_{12}f'''(A_1 + \tau T_1)|,$$

and where θ denotes an r.v. uniformly distributed in $(0,1)$, independent from all other r.v.'s. Let us write $A_1 = T_2 + A_2$. Expanding in powers of T_2, we have

$$|R_1| \le c\delta_1 N^{-3/2}\|f^{(4)}\|,$$

$$\delta_1 = N^3\mathbf{E}T_1^2|T_2T_{12}| \le \beta_4^{1/2}\gamma_2^{1/2} \le \beta_4 + \gamma_2.$$

Write $\delta = N^{5/2}\mathbf{E}T_1T_2T_{12}$. We have

$$\mathbf{E}T_1N^{3/2}T_{12}f''(A_1) = N^{-1}\delta\mathbf{E}f'''(A_2) + R,$$

$$|R| \le cN^{-3/2}\|f^{(4)}\|\delta_1 \le cN^{-3/2}\|f^{(4)}\|(\beta_4 + \gamma_2).$$

Furthermore,

$$\mathbf{E}f'''(A_2) = \mathbf{E}f'''(A) + R, \qquad |R| \le c\|f^{(4)}\|/\sqrt{N},$$
$$|\mathbf{E}f'''(A) - \mathbf{E}f'''(\sqrt{m/N}\eta)| \le cmN^{-3/2}\|f^{(6)}\|\beta_3.$$

Collecting the estimates concludes the proof of the lemma. □

The result of the following lemma is well known; see Petrov (1975) and Bhattacharya and Ranga Rao (1986).

LEMMA 6.2. *Let* $\tau^2 = 1$. *For* $1 \le m \le N$ *and* $|t| \le \sqrt{N}/\beta_3$, *the characteristic function* $h(t) = \mathbf{E}\exp\{itT_1\}$ *satisfies*

$$|h^m(t)| \le \exp\left\{-\frac{mt^2}{4N}\right\},$$
$$\left|h^m(t) - \exp\left\{-\frac{mt^2}{2N}\right\}\right| \le cN^{-1/2}\beta_3|t|^3\exp\left\{-\frac{mt^2}{4N}\right\},$$
$$h^m(t) = \left(1 + \frac{(it)^3 m\alpha_3}{6N\sqrt{N}}\right)\exp\left\{-\frac{mt^2}{2N}\right\} + R,$$

where

$$|R| \le cN^{-1}\beta_4(t^4 + t^6)\exp\left\{-\frac{mt^2}{4N}\right\}.$$

REFERENCES

BAI, Z. D. and RAO, C. R. (1991). Edgeworth expansion of a function of sample means. *Ann. Statist.* **19** 1295–1315.

BENTKUS, V., GÖTZE, F., PAULAUSKAS, V. and RAČKAUSKAS, A. (1991). The accuracy of Gaussian approximation in Banach spaces. *Probab. Theory* **6** 39–139.

BENTKUS, V., GÖTZE, F. and ZITIKIS, R. (1994). Lower estimates of the convergence rate for U-statistics. *Ann. Probab.* **22** 1707–1714.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

BICKEL, P. (1974). Edgeworth expansion in nonparametric statistics. *Ann. Statist.* **2** 1–20.

BICKEL, P., GÖTZE, F. and VAN ZWET, W. (1986). The Edgeworth expansion for U-statistics. *Ann. Statist.* **14** 1463–1484.

BOLTHAUSEN, E. and GÖTZE, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21** 1692–1710.

CALLAERT, H. and JANSSEN, P. (1978). The Berry–Esseen theorem for U-statistics. *Ann. Statist.* **6** 417–421.

CALLAERT, H., JANSSEN, P. and VERAVERBEKE, N. (1980). An Edgeworth expansion for U-statistics. *Ann. Statist.* **8** 299–312.

CHAN, Y. K. and WIERMAN, J. (1977). On the Berry–Esséen theorem for U-statistics. *Ann. Probab.* **5** 136–139.

CHIBISOV, D. M. (1972). Asymptotic expansion for the distribution of a statistic admitting a stochastic expansion. *Theory Probab. Appl.* **17** 620–630.

CHUNG, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.

DE ACOSTA, A. (1981). Inequalities for B-valued random vectors with applications to the strong law of large numbers. *Ann. Probab.* **9** 157–161.

DE LA PEÑA, V. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** 1877–1892.

FILIPPOVA, A. A. (1962). Mises' theorem of the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7** 24–57.

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

GÖTZE, F. (1979). Asymptotic expansions for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50** 333–355.

GÖTZE, F. (1987). Approximations for multivariate U-statistics. *J. Multivariate Anal.* **22** 212–229.

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739.

GÖTZE, F. and VAN ZWET, W. (1991). Edgeworth expansions for asymptotically linear statistics. Preprint 91-034, Univ. Bielefeld.

GOVINDARAJULU, Z. and MASON, D. M. (1983). A strong representation for linear combinations of order statistics with application to fixed-width confidence intervals for location and scale parameters. *Scand. J. Statist.* **10** 97–115.

GRAHAM, S. W. and VAALER, J. D. (1981). A class of extremal functions for the Fourier transform. *Trans. Amer. Math. Soc.* **265** 283–302.

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

HELMERS, R. (1982). *Edgeworth Expansions for Linear Combinations of Order Statistics*. *Math. Centre Tracts* **105**. Math. Centrum, Amsterdam.

HELMERS, R. (1985). The Berry–Esséen bound for studentized U-statistics. *Canad. J. Statist.* **13** 79–82.

HELMERS, R. and VAN ZWET, W. (1982). The Berry–Esseen bound for U-statistics. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. O. Berger, eds.) **1** 497–512. Academic Press, New York.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

PFANZAGL, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1** 993–1026.

PRAWITZ, H. (1972). Limits for a distribution, if the characteristic function is given in a finite domain. *Skand. Aktuartidskr.* 138–154.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

VAN ZWET, W. R. (1984). A Berry–Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440.

ZHAO, L. C. (1983). The rate of the normal approximation for a studentized U-statistic. *Sci. Exploration* **3** 45–52.

ZITIKIS, R. (1993). A Berry–Esseen bound for multivariate L-estimates with explicit dependence on dimensions. *Stability Problems for Stochastic Models. Lecture Notes in Math.* **1546** 197–211. Springer, Berlin.

V. BENTKUS
F. GÖTZE
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
POSTFACH 100131
33501 BIELEFELD 1
GERMANY
E-MAIL: bentkus@mathematik.uni-bielefeld.de
goetze@mathematik.uni-bielefeld.de

W. R. VAN ZWET
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LEIDEN
PO BOX 9512
2300 RA LEIDEN
NETHERLANDS
E-MAIL: vanzwet@wi.leidenuniv.nl