AN EDGEWORTH EXPANSION FOR SYMMETRIC STATISTICS

V. Bentkus[1] F. Götze[1] W. van Zwet[1,2]

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

January 1994

Abstract. 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 $O(1/N)$ under Cramér's condition on the linear part of the statistic and moment assumptions for all parts of the statistic. By means of a counter-example we show that it is generally not possible to obtain an Edgeworth expansion with remainder $o(1/N)$ without imposing additional assumptions on the structure of the non-linear part of the statistic

1. Introduction and results

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

$$T = t(X_1, \dots, X_N).$$

[1] Research supported by the SFB 343 at Bielefeld.

[2] Research supported by the Netherlands' Organization for Scientific Research.

1991 *Mathematics Subject Classification.* 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.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

We shall derive Edgeworth expansions for the distribution of T. Expansions terms and conditions are expressed in terms of 'influence functions' of the statistic. These functions measure the dependence of T on up to three observations. They are usually easy to compute or estimate in applications by means of Taylor expansions. See section 2 for examples including applications to the Bootstrap. The basis of the mathematical treatment of the problem will be Hoeffding's (1948) decomposition of the statistic in terms of conditional means of influence functions (see section 4).

Denote by

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

the expectation with respect to X_i, that is the conditional expectation given all other random variables. Define the first order influence function (of X_i) as the difference

$$D_i T = T - \mathbf{E}_i T. \tag{1.1}$$

For distinct i, j, k, define the second and the third order influence functions (with respect to observations X_i, X_j and X_k) as repeated differences

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

where we write for brevity $\mathbf{E}_{ij\dots}$ instead of $\mathbf{E}_i \mathbf{E}_j \dots$. Notice that $D_i D_i T = D_i T$ and $D_i D_j T = D_i D_j T$.

Furthermore, define the random variables

$$T_i = \mathbf{E}(D_i T|X_i) = \mathbf{E}(T|X_i) - \mathbf{E}T, \tag{1.2}$$

$$T_{ij} = \mathbf{E}(D_i D_j T|X_i, X_j) = \mathbf{E}(T|X_i, X_j) - \mathbf{E}(T|X_i) - \mathbf{E}(T|X_j) + \mathbf{E}T. \tag{1.3}$$

These expectations are the well known Hoeffding's decomposition terms of first and second orders. We may write

$$T - \mathbf{E}T = \sum_{i=1}^{N} T_i + \sum_{1 \le i < j \le N} T_{ji} + R, \tag{1.4}$$

where all random variables on the right hand side of (1.4) have mean zero and are uncorrelated. The terms

$$\sum_{i=1}^{N} T_i \qquad \text{and} \qquad \sum_{1 \le i < j \le N} T_{ji} + R$$

are called the linear and non-linear parts of the statistic. Our moment assumptions imply that asymptotically the linear part dominates distribution of the statistic and yields a normal limit distribution. The remainder terms in R are rather complex but will be analyzed in section 4 using a general Hoeffding's decomposition of the statistic. This decomposition (sometimes also called ANOVA-type decomposition or Hajek's projection method) is essential for the proof of our results.

So far, we have not used the fact that $t(X_1, \dots, X_N)$ is symmetric and that $X_1, \dots, X_N$ are i.i.d. This implies that, for distinct i, j, k, the distributions of

$$D_i T, \qquad D_i D_j T, \qquad D_i D_j D_k T, \qquad T_i, \quad T_{ij} \tag{1.5}$$

do not depend on i, j, k. Note that T_i is a function of X_i, and T_{ij} is symmetric function of X_i and X_j only. Random variables T_i, $1 \le i \le N$ are i.i.d. In particular, the moments of the random variables in (1.5) do not depend on i, j, k.

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

$$\beta_s = \mathbf{E}|\sqrt{N} T_1|^s, \quad \gamma_s = \mathbf{E}|N^{3/2} T_{12}|^s, \quad \Delta_m^s = \mathbf{E}(N^{m-1/2} D_1 D_2 \dots D_m T)^s.$$

By Jensen's inequality it follows from (1.1)–(1.3) that

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

Furthermore, let $\kappa = \alpha_3 + 3\delta$, where

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

Let $\sigma^2 = \sigma^2(T) = \mathbf{E}(T - \mathbf{E}T)^2$ denote the variance of T. Define the number $q = q(s)$, $0 \le q \le 1$, by

$$1 - q = \sup\Big\{ |\mathbf{E} \exp\{itN^{1/2} T_1\}| : \ \sigma^2/(2\beta_3) \le |t| \le \sqrt{N}/\sigma \Big\}.$$

Note that $q > 0$ if the characteristic function satisfies Cramér's condition (C),

$$\limsup_{|t| \to \infty} |\mathbf{E} \exp\{itN^{1/2} T_1\}| < 1. \tag{C}$$

Let F and G denote the distribution function of $(T - \mathbf{E}T)/\sigma$ and its two term Edgeworth expansion, thus

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

where $\Phi(x)$ is the standard normal distribution function. We shall prove

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

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

Sometimes it is more convenient to use Edgeworth expansion with σ^2 replaced by the variance $s^2 = \mathbf{E}(\sqrt{N}T_1)^2$ of the linear part of T. Denote

$$F_s(x) = \mathbf{P}(T - \mathbf{E}T \le xs), \qquad G_s(x) = \Phi(x) - \frac{\kappa}{6s^3\sqrt{N}}\Phi'''(x),$$

$$1 - q = 1 - q(s) = \sup\Bigl\{|\mathbf{E}\exp\{itN^{1/2}T_1\}| : \ s^2/\beta_3 \le |t| \le \sqrt{N}/s\Bigr\}.$$

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

$$\sup_x \bigl|F_s(x) - G_s(x)\bigr| \le \frac{c}{q^2 N}\Bigl(\frac{\beta_4}{s^4} + \frac{\gamma_3}{s^3} + \frac{\Delta_3^2}{s^2}\Bigr).$$

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

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

Accordingly, in estimate of Theorem 1.1, $q^2(\sigma)$ should be replaced by $\rho^2 q_\rho^2(\sigma)$. Furthermore, the proof yields that

$$\sup_x \bigl|F(x) - G(x)\bigr| \le \frac{c(\varepsilon)}{q^2 N}\Bigl(\frac{\beta_4}{\sigma^4} + \frac{\gamma_{2+\varepsilon}}{\sigma^{2+\varepsilon}} + \frac{\Delta_3^2}{\sigma^2}\Bigr), \quad \text{for all} \quad \varepsilon > 0.$$

We conjecture that ε may be removed. Such estimate would be optimal in view of lower bounds due to Bentkus, Götze and Zitikis (1992). The dependence on q^2 may be improved to $q^{1+\varepsilon}$, $\varepsilon > 0$. Whether the bound remain valid with q^2 replaced by q or some lower power of q, remains unclear.

The same improvements and comments hold for the bound of $\sup_x \bigl|F_s(x) - G_s(x)\bigr|$ in Theorem 1.2 with σ replaced by s.

If

$$\sigma^{-2}\mathbf{E}(T - \mathbf{E}T - \sum_{i=1}^N T_i)^2 \to 0, \quad \text{as} \quad N \to \infty,$$

then $(T - \mathbf{E}T)/\sigma$ is asymptotically equivalent to its linear part, and if this satisfies the Lindeberg condition, $(T - \mathbf{E}T)/\sigma$ will be asymptotically standard normal as

$N \to \infty$. Theorems 1.1 and 1.2 are natural extensions of the following Berry-Esseen bound due to van Zwet (1984):

$$\sup_x \bigl|F(x) - \Phi(x)\bigr| \le \frac{c}{\sqrt{N}}\Bigl(\frac{\beta_3}{\sigma^3} + \frac{\Delta_2^2}{\sigma^2}\Bigr).$$

Friedrich (1989), Bolthausen and Götze (1993) extended this bound for certain classes of non-symmetric statistics and improved the dependence on moments. It seems that this paper is the first dealing with Edgeworth expansions for *general symmetric* statistics.

Berry-Esseen bounds and Edgeworth expansions for U and other statistics were considered by Filippova (1962), Bickel (1974), Serfling (1980), Callaert, Janssen and Veraverbeke (1980), Helmers and van Zwet (1982), Helmers (1985), Pfanzagl (1985), Götze (1987, 1991), Hall (1987), Bhattacharya and Denker (1990), Zitikis (1993), etc. Chibisov (1980), Bickel, Götze and van Zwet (1986), Bhattacharya and Ghosh (1986), Bai and Rao (1990) Götze and van Zwet (1992), Lai and Wang (1993) have obtained expansions with errors of order $o(1/N)$ assuming certain Cramér type conditions *involving the nonlinear part* of the statistic. These conditions are not always easily verifyable. Let us remark that the methods of Bickel, Götze and van Zwet (1986), Götze and van Zwet (1992) may be used to prove under Cramér's condition (C) the validity of Edgeworth expansions with remainders $O(N^{\varepsilon-1})$, $\varepsilon > 0$ assuming only moment conditions on the non-linear part. In order to remove ε one needs some new ideas like 'data depending smoothing' and a non-standard smoothing inequality (see Prawitz (1972)). Such inequalities suggested by Beurling have been used in number theory and analysis, see, for example, Graham and Vaaler (1981). This inequality allows to handle data depending smoothing, and is precise in some other aspects (see Bentkus (1994)). We did not succeed to prove the result using the conventional Esseen inequality (3.4) for characteristic functions.

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, it is a purely asymptotic result concerned with a sequence of statistics $(T_N - \mathbf{E}T)/\sigma(T_N)$ as $N \to \infty$. If β_4/σ^4, γ_3/σ^3 and Δ_3^2/σ^2 are bounded and q is bounded away from 0, as $N \to \infty$, then the theorems establish Edgeworth expansions with remainders $O(1/N)$. For given bounds on β_4/σ^4, γ_3/σ^3, Δ_3^2/σ^2 and q, the result is uniform for all symmetric functions of the i.i.d. variables $X_1, \dots, X_N$ and for all underlying distributions $\mathbf{P}$.

Let us briefly discuss the nature of the assumption that β_4/σ^4, γ_3/σ^3 and Δ_3^2 are bounded. If we measure the order of magnitude of a mean zero random variable by its standard deviation, the boundedness of β_4/σ^4 means that T_1/σ is $O(N^{-1/2})$ and $N^{1/2}T_1/\sigma$ has a bounded fourth moment. This is certainly sufficient for the asymptotic normality of $\sum_{i=1}^N T_i/\sigma$, and since the fourth moment would occur in the term of order $O(N^{-1})$ of the Edgeworth expansion, its boundedness is a logical requirement. Similarly, the assumption that γ_3/σ^3 is bounded means that T_{12}/σ is of order $O(N^{-3/2})$ and that $N^{3/2}T_{12}/\sigma$ has a bounded absolute third moment. The boundedness of Δ_3^2/σ^2 ensures that the remaining part R of the decomposition (1.4) is negligible for our purposes.

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 our conditions are comparable with the those known so far. In some cases, like 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 are easily applicable for the estimation of the accuracy of Bootstrap approximations since estimates of remainders are explicit, and the expansions are valid for a sufficiently general class of symmetric statistics. Bootstraped or studentized versions of a statistic are again symmetric, and in order to show validity of Edgeworth expansions, it just suffices to estimate several lower order conditional differences. In this paper we shall do this for the simplest example of a studentized mean only, more general results will be published elsewhere.

Similar estimates as those in Theorems 1.1 and 1.2 do not hold for higher order Edgeworth expansions even if we assume arbitrary high moments on all parts of the statistic. In Theorem 1.4 we shall provide an example for this phenomenon. Thus assuming Cramér's (C) condition on the linear part T_1 of the statistic, the estimate $O(N^{-1})$ of the remainder term is best possible unless further conditions (apart from moments) for the non-linear part are used, see Götze and van Zwet (1992). But the verification of such conditions usually requires a detailed knowledge of higher order terms in the stochastic expansion of the statistic.

In order to formulate Theorem 1.4 we need some further notation. Any U-statistic T of second order 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 $T_{1,2}=g_2(X_1,X_2)$, 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 $\mathbf{E}T=0$,

$$D_1T=T_1+\sum_{j=2}^{N}T_{1j},\qquad D_1D_2T=T_{12},\qquad \text{and}\quad D_1D_2D_3T=0.$$

Let us assume that components of T are uniformly bounded by some non-random $A < \infty$,

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

Notice that (1.7) ensures that Edgeworth approximation G, as well as higher order Edgeworth approximations, satisfy

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

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

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

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

and

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

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

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

Moreover, for any given $A > 0$,

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

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

2. Applications

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

$$b_s = \mathbf{E}||X_1||^s, \quad \bar{X} = (X_1 + \cdots + X_N)/N, \quad E_N = \sqrt{N}\bar{X}, \quad Q_s = \sup_{N\ge 1} \mathbf{E}||E_N||^s,$$

where we shall assume implicitly that all random variables are well defined (for example, 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 $H : \mathbf{B} \to \mathbf{R}$, let $H^{(s)}(x)$ denote the s-th Frechét derivative of H at the point $x \in \mathbf{B}$. Define $H^{(s)}(x)h_1 \dots h_s$ as the value of s-linear continuous symmetric form $H^{(s)}(x)$ with arguments $h_1, \dots, h_s \in \mathbf{B}$, as well as

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

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

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

Define furthermore $M = M(H)$ and $L = L(H)$

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

2.1. Sample means. Let $X_1, \dots, X_N$ be real i.i.d. random variables. Consider the statistic $T = \bar{X}$, where $\bar{X}$ 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 $cb_4/(q^2 b_2^2 N)$. This is a well known result. Note that the dependence on

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

may 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

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

In this case we may write in the decomposition (1.4) for $T - \mathbf{E}T$,

$$T_1 = (N-1)N^{-2}(X_1^2 - b_2), \qquad T_{12} = -2N^{-2}X_1X_2, \quad \text{and} \quad R = 0.$$

Furthermore,

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

The distribution function of $(T - \mathbf{E}T)/s$ (or of $(T-\mathbf{E}T)/\sigma$) is uniformly approximable by its Edgeworth expansion, and the error does not exceed $cb_8/(q^2 D^4 N)$, where

$$q = 1 - \sup_{|t|\geq D^2/\beta_3} |\mathbf{E}\exp\{itX_1^2\}|, \quad \text{and} \quad \beta_3 = \mathbf{E}|X_1^2 - b_2|^3.$$

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

$$\mathbf{E}|D_1 \dots D_k V|^s \le 2^{k+s}\mathbf{E}|V|^s, \quad \text{for all} \quad s \ge 1,$$

$$\mathbf{E}|D_1 \dots D_k(V+W)|^s \le 2^s\mathbf{E}|D_1 \dots D_k V|^s + 2^{k+s}\mathbf{E}|W|^s, \quad \text{for all} \quad s \ge 1.$$

If V is independent of X_i then $D_i V = 0$. If V is independent of at least of 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 $T = H(\bar{X})$ is a sufficiently smooth function of the sample mean. Let $\tau_1, \dots, \tau_k$ denote i.i.d. random variables uniformly distributed on $[0,1]$ independent of all other random variables. Then

$$D_1 \dots D_k H(\bar{X}) =$$

$$N^{-k} D_1 \dots D_k \mathbf{E}_\tau H^{(k)}(\tau_1 N^{-1}X_1 + \dots + \tau_k N^{-1}X_k + B)X_1 \dots X_k, \tag{2.1}$$

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

Indeed, split $\bar{X} = N^{-1}X_1 + A$, where $A = (X_2 + \dots + X_N)/N$. Expanding into the Taylor series we have

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

Since the random variable $H(A)$ is independent of X_1 we have $D_1 H(A) = 0$. 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

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

and in many applications this yields a satisfactory estimate.

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

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^*,

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

where $\bar{X}$ denotes the sample mean, and $\hat{b}_2$ denotes the sample variance. The Bootstrap version t^* is obtained by replacing the arguments $X_1, \dots, X_N$ of t by

$X_1^* - \bar{X}, \dots, X_N^* - \bar{X}$, 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

(2.2)

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

for arbitrary $\varepsilon > 0$, where

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

and

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

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 $X_i^* - \bar{X}$ we obtain

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

for arbitrary $\varepsilon > 0$, where

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

and

$$\hat{\mu}_3 = \mathbf{E}^*(X_1^* - \bar{X})^3 = N^{-1} \sum_{i=1}^N (X_i - \bar{X})^3, \quad \hat{b}_s = \mathbf{E}^* |X_1^* - \bar{X}|^s = N^{-1} \sum_{i=1}^N |X_i - \bar{X}|^s$$

denote the sample moments.

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

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

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

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

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 $f : \mathbf{R} \to \mathbf{R}$ be a bounded infinitely many times differentiable function such that $f(x) = 1/\sqrt{|x|}$, for $|x| \geq 1/2$. Note that all derivatives of f are bounded. We may regularize t by replacing it by $t_f = \bar{X} f(\hat{b}_2)$. Indeed,

$$\mathbf{P}\{t \neq \bar{X} f(\hat{b}_2\} \leq \mathbf{P}\{\hat{b}_2 \leq 1/2\} \leq c\mathbf{E}(\hat{b}_2 - \mathbf{E}\hat{b}_2)^2 \leq cb_4/N, \quad N \geq 3.$$

In order to truncate t_f define

$$Y_i = N^{-1/2} X_i \mathbf{I}\{X_i^2 \leq N\}, \quad S_N = Y_1 + \cdots + Y_N, \quad s_N^2 = Y_1^2 + \cdots + Y_N^2.$$

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

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

It is well known (and easy to verify) that

$$\mathbf{E}|S_N|^p \leq c(p), \quad \text{for} \quad p \geq 1.$$

Note that

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

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

$$\begin{aligned} &|\sqrt{N} D_1 T| \leq c(1 + |S_{N-1}|)(1 + |X_1|), \\ &|N^{3/2} D_1 D_2 T| \leq c(1 + S_{N-2}^2)(1 + X_1^2 + X_2^2 + X_1^2|X_1| + |X_1|X_2^2), \\ &|N^{5/2} D_1 D_2 D_3 T| \leq c(1 + S_{N-3}^4) \sum |X_1|^k |X_1|^l |X_1|^m, \end{aligned} \tag{2.3}$$

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

$$\beta_4 \leq cb_4, \quad \gamma_{2+\varepsilon} \leq cb_3 b_{4+\varepsilon}, \quad \Delta_3^2 \leq cb_4^3.$$

We shall prove the first inequality in (2.3) only since the proofs of other ones are similar but somewhat more tedious. Split $s_N^2 = Y_1^2 + s_{N-1}^2$. We may split T into $T = Q_1 + Q_2$, where $Q_1 = Y_1 f(s_N^2 - N^{-1} S_N^2)$, and

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

Here we estimate Q_1 by $|Q_1|$. 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 $Y_1^2 \leq |Y_1|$.

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 i.i.d. random variables $X_1, \dots, X_N$ take values in a real Banach space $\mathbf{B}$. Consider the statistics

$$T^0 = \sqrt{N}(H(\bar{X}) - H(0)) \quad \text{and} \quad T = \sqrt{N}(H(\bar{X}) - \mathbf{E}H(\bar{X})),$$

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

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

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

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

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

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

$$d = s_1^{-3}\mathbf{E}\big(H'(0)X_1\big)\big(H'(0)X_2\big)\big(H''(0)X_1X_2\big), \quad k = a_3 + 3d$$

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

Then we have

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 (1986), 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*

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

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* $T=\sqrt{N}(H(\bar X)-\mathbf{E}H(\bar X))$ *we have*

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

where $|R|$ *is bounded by any of the two following quantities*

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

Furthermore,

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

$$|R_1| \le c\min\Big\{M_0^2N^{-1}(b_3+Q_2b_2); \quad M^2N^{-1}(b_3+b_4)\Big\},$$

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

where

$$|R_2| \le cN^{-1/2}\min\Big\{M_0^3(b_4+Q_1b_3); \quad M^3(b_4+b_3b_{3/2})\Big\},$$

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

with

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

2.6. Functionals of empirical distribution functions. The results of this subsection 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 subsection 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 random processes $X_i(t)$, $t \in \mathbf{R}$, $1 \le i \le N$, by $X_i(t) = \mathbf{I}\{\eta_i < t\} - P(t)$. Let

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

denote the empirical process.

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

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

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 $X_1(t) = \mathbf{I}\{\eta_1 < t\}$ a.s. Furthermore, we assume that 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

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

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

In order to formulate Edgeworth expansions for S we need some notation. Let

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

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

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

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 $\eta_{1;N} \leq \cdots \leq \eta_{N;N}$ are the order statistics of $\eta_1, \dots, \eta_N$, and 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

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

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

$$|\mathbf{E} l_N - \mu - \frac{\nu}{2N}| \leq c M_0(J) N^{-3/2} \big(\mathbf{E}|\eta_1|^4 + \mathbf{E}|\eta_1| \big),$$
$$|\mathbf{E} l_N - \mu - \frac{\nu}{2N}| \leq c M(J) N^{-3/2} \big(\mathbf{E}|\eta_1|^4 + \mathbf{E}|\eta_1|^{5/6} \big),$$

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

$$b_s = \mathbf{E}|\eta_1|^s, \quad x \wedge y = \min\{x, y\},$$
$$s_1^2 = \int_{\mathbf{R}^2} J(P(x)) J(P(y)) \{P(x \wedge y) - P(x)P(y)\} dx dy.$$

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

$$a_3 = s_1^{-3} \int_{\mathbf{R}^3} J(P(x)) J(P(y)) J(P(z)) \big\{ P(x \wedge y \wedge z) + U \big\} dx dy dz,$$
$$U = -P(x)P(y \wedge z) - P(y)P(x \wedge z) - P(z)P(x \wedge y) + 2P(x)P(y)P(z),$$
$$d = s_1^{-3} \int_{\mathbf{R}^3} J(P(x)) J(P(y)) J'(P(z)) V(x,z) V(y,z) dx dy dz,$$
$$V(x,y) = P(y \wedge z) - P(y)P(z),$$
$$q = 1 - \sup\{ |\mathbf{E} \exp\{itY\}| : \ |t| \geq c_0 s_1^2 / (M_0^3(J)(b_3 + b_1)) \},$$

where $Y = \int_{\mathbf{R}} J(P(x))(\mathbf{I}\{x < \eta_1\} - P(x)) dx$.

Theorem 2.5. *There exists an absolute positive constant c_0 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))

$$l_N - \mu = \int_{-\infty}^{\infty} \Big[\Psi\big(P_N(t)\big) - \Psi\big(P(t)\big)\Big]\,dt, \quad \text{where} \quad \Psi(x) = \int_x^1 J(u)du.$$

Therefore we may write

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

where

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

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 $h \mapsto T(P+h) : \mathbf{B} \to \mathbf{R}$ 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

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

It is easy to verify that

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

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

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

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

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

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

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

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

2.8. ***U*-statistics.** In this subsection 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 symmetric statistic is a U-statistic of fixed order r, which is given by

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

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

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

where

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

then Hoeffding's decomposition of U is given by

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

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

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

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

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

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

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

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

$$\frac{c}{Nq^2(\sigma)}\Big(\frac{\beta_4}{\sigma^4}+\frac{\gamma_3}{\sigma^3}+\frac{\Delta_3^2}{\sigma^2}\Big), \quad or \quad \frac{c}{Nq^2(s)}\Big(\frac{\beta_4}{s^4}+\frac{\gamma_3}{s^3}+\frac{\Delta_3^2}{s^2}\Big),$$

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

$$\frac{\beta_4}{\sigma^4} \le \frac{\beta_4}{s^4} \le \frac{\mathbf{E}h_1^4}{(\mathbf{E}h_1^2)^2},$$
$$\frac{\gamma_3}{\sigma^3} \le \frac{\gamma_3}{s^3} \le \frac{8(r-1)^2\mathbf{E}|h_2|^3}{(\mathbf{E}h_1^2)^{3/2}},$$
$$\frac{\Delta_3^2}{\sigma^2} \le \frac{\Delta_3^2}{s^2} \le \sum_{k=3}^{r} \frac{c(r)(N-k)!\mathbf{E}h_k^2}{(N-3)!\mathbf{E}h_1^2}.$$

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 $H>0$ (note that all integrals are real and that the expressions under the signs of integrals vanish unless $|t| \le H$),

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

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

where $F(x+) = \lim_{z\downarrow x} F(z)$, $F(x-) = \lim_{z\uparrow x} F(z)$, and V.P. denotes Cauchy's Principal Value,

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

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

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

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

It is known (see, for instance, Chung (1974) p. 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 the bounded variation.

The following lemma is elementary.

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

$$K_2(0) = 1, \quad K_2(1) = 0, \quad K_2\big(\frac{1}{2}\big) = \frac{1}{2}, \quad K_2'(s) \le 0, \quad K_2(s) + K_2(1-s) = 1.$$

Furthermore,

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

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

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

where $A = \sup_x \big|G'(x)\big|$. 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 $\sup_x \big|G'(x)\big| < \infty$) are needed. It follows from (3.1) and (3.3) that

$$
\begin{aligned}
(3.5)\qquad & F(x+) - G(x) \le I_1 + I_2 + I_3, \\
& I_1 = \frac{1}{2}\int_{\mathbf{R}} \exp(-ixt)\frac{1}{H}K_1\big(\frac{t}{H}\big)f(t)dt, \\
& I_2 = \frac{i}{2\pi}\text{V.P.}\int_{\mathbf{R}} \exp(-ixt)K_2\big(\frac{t}{H}\big)\big(f(t) - g(t)\big)\frac{dt}{t}, \\
& I_3 = \frac{i}{2\pi}\text{V.P.}\int_{\mathbf{R}} \exp(-ixt)\Big(K_2\big(\frac{t}{H}\big) - 1\Big)g(t)\frac{dt}{t},
\end{aligned}
$$

where V.P. means as well that one should take $\lim_{M\to\infty}$, if it is necessary. 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 chose random H depending on the sample and apply (3.5) conditionally (we call this procedure "data dependent smoothing").

4. HOFFDING'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. 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, $\mathbf{E}(T|\emptyset) = \mathbf{E}T = 0$ and $\mathbf{E}(T|\Omega) = T$.

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

$$D_i T = T - \mathbf{E}(T|\Omega \setminus \{i\}). \tag{4.1}$$

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

$$D_AT = D_{i_1} \dots D_{i_k}T$$

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

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

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,

$$T_A = \sum_{B\subset A} (-1)^{|A|-|B|}\mathbf{E}(T|B). \tag{4.2}$$

The random variable T_A is related to the difference D_AT, and

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

Equation (4.2) expresses T_A in terms of the conditional expectations. It is easy to see that there is also an inverse relation

$$\mathbf{E}(T|B) = \sum_{A\subset B} T_A, \tag{4.3}$$

and in particular

$$T = \sum_{A\subset\Omega} T_A, \tag{4.4}$$

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

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

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

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

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

Hence (4.3) and (4.4) represent $\mathbf{E}(T|B)$ and T as sums of uncorrelated mean zero random variables T_A, and as a result

$$\sigma^2(\mathbf{E}(T|B)) = \sum_{A\subset B} \mathbf{E}T_A^2, \tag{4.5}$$

$$\sigma^2 = \sigma^2(T) = \sum_{A\subset\Omega} \mathbf{E}T_A^2. \tag{4.6}$$

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,

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

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

etc. 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

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

and the variance decompositions (4.5) and (4.6) become

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

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

Consider repeated differences (4.1),

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

It follows from (4.7) that

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

$$\frac{1}{3}N^2\sum_{k=3}^{N}k^3\binom{N}{k}\mathbf{E}T^2_{\Omega_k}\le\Delta_3^2\le 3N^2\sum_{k=3}^{N}k^3\binom{N}{k}\mathbf{E}T^2_{\Omega_k}.$$

Denote $\Omega_m=\{1,\dots,m\}$,

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

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*

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

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

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

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

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

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

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

$$\varphi(3)=\sum_{r=3}^{N}\binom{N-3}{r-3}\mathbf{E}T^2_{\Omega_r}=\mathbf{E}(D_1D_2D_3T)^2=N^{-5}\Delta_3^2.$$

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

$$d\varphi(2)=\varphi(3)=\mathbf{E}(D_1D_2D_3T)^2,$$
$$\varphi(m)=\varphi(2)+d\varphi(2)+\dots+d\varphi(m-1),$$

and (4.8) will follow if we show that

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

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 $\mathcal{A}_m$ of sets,

$$\mathcal{A}_m = \Big\{ A: \;\; A \subset \Omega, |A| = r, \big|A \cap \Omega_m\big| \le 2 \Big\}.$$

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

$$c(r,m,N) = \Big| \{A: \;\; A \subset \Omega\} \Big| - |\mathcal{A}_m|, \tag{4.14}$$

From (4.14) we see that

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

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

$$\begin{aligned} A \in \mathcal{A}_m \setminus \mathcal{A}_{m+1} &\Longleftrightarrow \big|A \cap \Omega_m\big| \le 2 \text{ and } \big|A \cap \Omega_{m+1}\big| \ge 3 \Longleftrightarrow \\ &\Longleftrightarrow \big|A \cap \Omega_m\big| = 2 \text{ and } m+1 \in A, \end{aligned}$$

and

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

But

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

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

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

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

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

Inequality (4.10) is a consequence of the following Lemma 4.2.

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

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

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

where $\gamma_p = \mathbf{E}\big|\big|\psi(X, X_1)\big|\big|^p$.

Proof. We use a decoupling inequality (due to Peña (1992), see (4.17) below) and the well-known estimates (see (4.18) and (4.19) below) for moments of sums of independent r.v., which follow from inequalities of Acosta (1981).

For any convex increasing function $\Phi : [0,\infty) \to [0,\infty)$,

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

where $Z_1, Z_2, \dots$ is an independent copy of $X_1, X_2, \dots$.

If $\eta_1, \dots, \eta_m \in \mathbf{B}$, $\mathbf{E}\eta_j = 0$ are independent, then

$$\mathbf{E}||\eta_1 + \dots + \eta_m||^p \le c(p)\big(\mathbf{E}||\eta_1||^p + \dots + ||\eta_m||^p\big), \qquad \text{for} \quad 1 \le p \le 2, \tag{4.18}$$

and, for $\quad 2 \le p \le \infty$,

$$\mathbf{E}||\eta_1 + \dots + \eta_m||^p \le c(p)\mathbf{E}\big(||\eta_1||^2 + \dots + ||\eta_m||^2\big)^{p/2} + c(p)\big(\mathbf{E}||\eta_1||^p + \dots + \mathbf{E}||\eta_m||^p\big), \tag{4.19}$$

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 $(i,j)\in A \Longrightarrow j\le N$. Define the sets $A_i=\big\{(i,j):\ (i,j)\in A\big\}$, 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

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

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

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

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

(4.13)
$$\mathbf{E}||\xi_i||^p\le c(p)\mathbf{E}_X(\sum_{(i,j)\in A_i}\mathbf{E}_Z||\psi(X_i,Z)||^2)^{p/2}+c(p)m_i\gamma_p\le c(p)m_i^{p/2}\gamma_p,$$

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

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

since $m_1+\dots+m_K=|A|$. Let us define a r.v. σ by $\mathbf{P}\{\sigma=i\}=m_i/|A|$, for $1\le i\le K$. We assume, that σ is independent from all other r.v. 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

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

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

Proof of Theorem 1.1. It follows from (4.6) that

$$\sigma^2 = s^2 + R^2, \quad \text{where} \quad 0 \le R^2 \le \frac{\gamma_2}{N} + \frac{\Delta_3^2}{N^2}.$$

We have

$$\sup_x \bigl| F(x) - G(x) \bigr| \le 1 + \frac{|\kappa|}{\sigma^3\sqrt{N}} \le c\Bigl(1 + \frac{\beta_3}{\sigma^3\sqrt{N}} + \frac{\gamma_2^{1/2}}{\sigma\sqrt{N}}\Bigr). \tag{5.1}$$

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

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

and Theorem 1.1 follows.

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

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

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

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

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

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

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

$$\sup_x \Bigl|\Phi(x) - \Phi\Bigl(\frac{sx}{\sigma}\Bigr)\Bigr|, \quad \Bigl|1 - \frac{s^3}{\sigma^3}\Bigr|, \quad \sup_x \Bigl|\Phi'''(x) - \Phi'''\Bigl(\frac{sx}{\sigma}\Bigr)\Bigr|. \tag{5.3}$$

The inequality $2s^2 \ge \sigma^2$ implies $s^2/\sigma^2 = 1 - R^2/\sigma^2$, 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 non-linear 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

$$s^2 = 1.$$

We shall prove that

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

where

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

and where for brevity we write $q_N(s) = q$, $F_s = F$ and $G_s = G$. Now Theorem 1.2 is a consequence of (5.4).

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

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

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 $\sup_x \bigl|F(x) - G(x)\bigr| \le c$, which we may assume like (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 $\hat{F}$ resp. $\hat{G}$ denote c.f. of F resp. G. For some natural number $k > r \geq N/2$, we have*

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

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

$$I = \int_{|t|\leq H_1} \bigl|\hat{F}(t) - \hat{G}(t)\bigr| \frac{dt}{|t|},$$
$$J = cq^{-2}\ln^2 N \int_{|t|\leq H_1} \bigl|\mathbf{E}T_{N-1\,N}\exp\{itT\}\bigr| dt,$$
$$I_{k,r} = \frac{1}{qN}\int_{|t|\leq H_1} (1+\theta)\bigl|\mathbf{E}_{1,\dots,r}\exp\{itT\}\bigr| dt, \quad \theta = N\mathbf{E}_1\Bigl|\sum_{j=k}^{N} T_{1j}\Bigr|.$$

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

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

where

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

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

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

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

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

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

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

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

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)

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

where

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

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

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

we shall estimate

$$\sup_x\bigl|\tilde{F}(x)-G(x)\bigr|, \qquad \text{where} \qquad \tilde{F}(x)=\mathbf{P}\{\tilde{T}<x\}.$$

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

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

which is majorized 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

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

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

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

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

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 $X_{m+1}, \dots, X_N$.

Random variables $X_1, \dots, X_m$ are independent of $X_{m+1}, \dots, X_N$. Thus we may apply 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

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

where

$$I_1 = \frac{1}{H}\int_{\mathbf{R}} \exp\{-ixt\} K_1\Big(\frac{t}{H}\Big) f(t)\,dt,$$
$$I_2 = \frac{i}{\pi}\int_{\mathbf{R}} \exp\{-ixt\} K_2\Big(\frac{t}{H}\Big) f(t)\frac{dt}{t},$$

and where $f(t) = \mathbf{E}_{1,\dots,m}\exp\{it\tilde{T}\}$. 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 integral I_1 by a non-random 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. $X_{m+1}, \dots, X_N$.

We have

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

where

$$I_3 = \frac{1}{H}\int_{|t|\le H_1} \exp\{-ixt\} K_1\Big(\frac{t}{H}\Big) f(t)\,dt, \qquad H_1 = \sqrt{N}/\beta_3,$$
$$I_4 = \frac{1}{H}\int_{H_1\le |t|} K_1\Big(\frac{t}{H}\Big)|f(t)|\,dt \le \frac{1}{H}\mathbf{I}\{H \ge H_1\}\int_{H_1\le|t|\le H} |f(t)|\,dt.$$

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

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

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

$$\mathbf{I} = \mathbf{I}\big\{4\mathbf{E}_1|t\eta_1| \le q\big\}.$$

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

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

where

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

$$J_2 = \frac{1}{H}\mathbf{I}\{H \geq H_1\} \int_{H_1 \leq |t| \leq H} (1 - \mathbf{I})dt.$$

Using Chebyshev's inequality and (5.16) we get

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

since $H \leq qN$.

In order to estimate J_1 we use a Taylor expansion,

$$\mathbf{I}\Big|\mathbf{E}_1 \exp\{itV_1\}\Big| \leq \Big|\mathbf{E}\exp\{itT_1\}\Big| + |t|\mathbf{E}_1\big|\xi_1\big| + \mathbf{I}\mathbf{E}_1\big|t\eta_1\big|.$$

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

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

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 $|K_1(s) - 1| \leq |s|$, for all $s \in \mathbf{R}$. Therefore

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

where

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

and

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

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

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

where

$$I_7 = \frac{1}{H}\int_{|t| \leq H_1} \exp\{-ixt\}\exp\{itT\}dt, \quad |R| \leq c\mathbf{E}\frac{H_1^2}{H}\Big(|\Lambda_1| + |\Lambda_2| + |\Lambda_3|\Big).$$

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

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

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 sum of three integrals, and we can interchange integrals and the expectation with respect to r.v. independent of θ_1 (or θ_2). Due to symmetry and the i.i.d. assumption, the last two integrals are bounded by

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

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

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

and (5.13) follows.

ESTIMATION OF $\mathbf{E}I_2$. We shall proceed similarly 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

$$I_3 = \frac{i}{\pi}\text{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}K_2\Big(\frac{t}{H}\Big)f(t)\frac{dt}{t}, \qquad H_1=\sqrt{N}/\beta_3,$$
$$I_4 = \mathbf{I}\{H\ge H_1\}\int_{H_1\le|t|\le H}\big|f(t)\big|\frac{dt}{|t|},$$

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 $\big|K_2(\tau)-1\big|\le 5\tau^2$. Therefore

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

where

$$I_5 = \frac{i}{\pi}\text{V.P.}\int_{|t|\le H_1}\exp\{-ixt\}f(t)\frac{dt}{t},$$
$$\mathbf{E}I_6 = c\mathbf{E}\int_{|t|\le H_1}\frac{|t|}{H^2}\ dt\le cq^{-2}\beta_3^2(1+\gamma_2)/N.$$

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

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

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

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

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

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

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

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

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

We have

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

where

$$\hat{G}(t) = \Big(1 + \frac{(it)^3\kappa}{6\sqrt{N}}\Big)\exp\Big\{-\frac{t^2}{2}\Big\}.$$

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

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

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 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 Taylor's series we get

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

where

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

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

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

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,

$$U=\sum_{i=1}^m V_i+\Lambda_1+W,\quad V_i=T_i+\xi_i,\quad \xi_i=\sum_{j=m+1}^N T_{ij},\quad \Lambda_1=\sum_{1\le i<j\le m}T_{ij}.$$

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

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

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

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

Then

$$q^{-1}N^{-1}\mathbf{E}\int_{C\le|t|\le H_1}(1+|\theta|)(1-\mathbf{I})dt \le cq^{-1}N\mathbf{E}\int_{C\le|t|\le H_1}\mathbf{E}(1+|\theta|)\Big(\mathbf{E}_1\big|\xi_1\big|\Big)^2\frac{dt}{t^2},$$

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

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

But

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

since

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

The estimate (5.18) implies $\mathbf{I}\big|h(t)\big|^m \le |t|^{-C}$ with a sufficiently large C, and therefore I_5 is majorized 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

$$I_1 = \int_{C\le|t|\le H_1}\big|\hat F(t)-\hat G(t)\big|\frac{dt}{|t|},$$
$$I_2 = \int_{|t|\le C}\big|\hat F(t)-\hat G(t)\big|\frac{dt}{|t|} \le c(\beta_4+\gamma_2+\Delta_3^2)/N.$$

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

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

For $|t| \le C$ the function $f(x) = \exp\{itx\}$ is smooth and has bounded derivatives. Therefore applying Lemma 5.1 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 $A \backsimeq B$ if $A - B \prec (\beta_4 + \gamma_3 + \Delta_3^2)/N$. In view of (5.19) we have to show that $\hat F(t) \backsimeq \hat G(t)$.

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

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

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

$$\hat F(t) \backsimeq \mathbf{E}\exp\{\tilde T + \Lambda_1\}.$$

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

$$\hat F(t) \backsimeq \mathbf{E}\biggl(1 + it\binom{m}{2} T_{12}\biggr)\exp\{it\tilde T\} \tag{5.20}$$

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

Recall that $V_j = T_j + \xi_j + \eta_j$, and and introduce the notation

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

Then we may rewrite (5.20) as

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

Let us show that

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

and that

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

Expanding in powers of η_1, η_2 and using $\binom{m}{2} \le m^2$, $m = CNt^{-2}\ln|t|$, we get

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

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

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

Expanding in powers of T_2 we have

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

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

$$\hat{F}(t) \backsimeq \mathbf{E}\exp\{itW\}\Big(h_3^m(t) + h_5(t)h_3^{m-2}(t)\Big). \tag{5.23}$$

Let us consider the following indicator functions

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

choosing a sufficiently small absolute positive absolute constant c_1. Notice that the functions $\mathbf{I}_1$, $\mathbf{I}_2$ 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

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

Therefore (5.23) yields

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

Let us show that

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

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

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

Expanding in powers of P we get

$$\mathbf{I}_1\mathbf{I}_2\big|h_3(t)\big| \le \big|h_1(t)\big| + \mathbf{I}_1\mathbf{I}_2\Big|\mathbf{E}_1 tP\exp\big\{itT_1\big\}\Big| + \mathbf{I}_1\mathbf{I}_2 t^2\mathbf{E}_1 P^2.$$

But

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

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

$$\mathbf{I}_1\mathbf{I}_2\Big|\mathbf{E}_1 tP\exp\big\{itT_1\big\}\Big| \le t^2\mathbf{I}_2\mathbf{E}_1\big|\xi_1 T_1\big| + \mathbf{I}_1 t^2\big|\eta_1 T_1\big| \le \frac{t^2}{12N},$$

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

Now (5.25) follows.

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

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

where the number A can be chosen arbitrary large by choosing 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 $\mathbf{I}_1\mathbf{I}_2$ as balancing factors.

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

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

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

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

where

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

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

$$\hat{F}(t) \backsimeq \mathbf{E}\mathbf{I}_1\mathbf{I}_2\exp\{itW\}\Big(h_2^m(t) + it^3\delta N^{-5/2}\binom{m}{2}h_2^{m-2}(t)\Big). \tag{5.28}$$

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

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

where

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

Furthermore,

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

where

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

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

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

where

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

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

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

Using Chebyshev's inequality we can finally remove $\mathbf{I}_1\mathbf{I}_2$ 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),

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

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

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

which together with (5.29) finally implies the desired result $\hat F(t)\simeq\hat G(t)$.

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

$$h_1^m(t)G_1(t) \simeq h_1^m(t)\Big(1+\frac{(it)^3}{6\sqrt{N}}\Big(\alpha_3\big(1-\frac{m}{N}\big)+3\delta\big(1-\frac{m}{N}\big)^2\Big)\Big)\exp\Big\{-\frac{(1-m/N)t^2}{2}\Big\}.$$

Similarly applying Lemma 5.2 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), (1.10) and prove

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

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

One gets the result dividing T by $s = (N\mathbf{E}T_1^2)^{1/2}$ in the following construction.

Let $\{x\}$ denote the difference of x to the nearest integer,

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

(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 interval $[-\frac{1}{2}, \frac{1}{2}]$. It is easy to see that

$$T_1 = \frac{X_1}{\sqrt{N}} - \frac{X_1\{mX_1\}}{N\sqrt{N}} + \frac{1}{12N^2} \qquad \text{and} \qquad N^{3/2}T_{12} = -X_1\{mX_2\} - \{mX_1\}X_2$$

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

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

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 independence of V_N and W_N we have

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

The random variable $\{mX_1\}$ is uniformly distributed. Therefore the local limit Theorem for densities implies (see Petrov (1975), Ch. VII)

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

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

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

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 $\sup_x \big|G'(x)\big| \le A$, 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 $|t| \ge \delta_2 N$ needs again special arguments.

6. Auxiliary results

Let T denote a symmetric statistic with Hoeffding decomposition (4.4). 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 a function $f : \mathbf{R} \to \mathbf{C}^1$ be sufficiently smooth. Then*

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

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

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

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

$$\kappa_m = m N^{-1} \alpha_3 + 3 m^2 N^{-2} \delta, \quad |R| \le c(f) m N^{-2} (s^{-4}\beta_4 + s^{-2}\gamma_2),$$

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

Proof. Without loss of generality we may assume that $s^2 = 1$. 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

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

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

It is not hard to show (see Bentkus, Götze, Paulauskas and Račkauskas (1990)) that

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

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

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

and where τ denotes a r.v. uniformly distributed in $(0,1)$, independent from all other r.v. 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,

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

Collecting the estimates concludes the proof of the Lemma.

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

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

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

where

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

References

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

Bai, Z.D. and Rao, C.R., *Edgeworth Expansion of a Function of Sample Means*, Ann. Statist. **19** (1991), 1295–1315.

Bentkus V., *On the asymptotical behavior of the constant in the Berry-Esseen inequality* , J. Th. Prob. **7** (1994), 211–224.

Bentkus V., Götze F., Paulauskas V. and Račkauskas A., *The accuracy of Gaussian approximation in Banach spaces*, SFB 343 "Diskrete Structuren in der Mathematik," Preprint 90 -100, Universität Bielefeld (In Russian: Itogi Nauki i Tehniki, ser. Sovr. Probl. Matem., Moskva, VINITI **81** (1991), 39–139; to appear in English in Encyclopedia of Mathematical Sciences, Springer), 1990.

Bentkus V., Götze F. and Zitikis R., *Lower estimates of the convergence rate for U-statistics*, Preprint Univ. Bielefeld (to appear in Ann. Prob.) **92-075** (1992).

Bhattacharya R.N. and Denker M., *Asymptotic statistics*, Birkhäuser, Basel–Boston–Berlin, 1990.

Bhattacharya, R.N. and Ghosh, J.K., *On the validity of the formal Edgeworth expansion*, Ann. Statist. **6** (1986), 434–451.

Bhattacharya, R.N. and Ranga Rao, R., *Normal Approximation and Asymptotic Expansions*, Wiley, New York (1986).

Bickel, P., *Edgeworth expansion in nonparametric statistics*, Ann. Statist. **4** (1974), 108–156.

Bickel, P., Götze, F. and van Zwet, W., *The Edgeworth expansion for U-statistics*, Ann. Statist. **14** (1986), 1463–1484.

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

Callaert, H., Janssen, P. and Veraverbeke, N., *An Edgeworth expansion for U-statistics*, Ann. Statist. **8** (1980), 299–312.

Chibisov, D.M., *Asymptotic expansion for the distribution of a statistic admitting a stochastic expansion I*, Theor. Probab. Appl. **25** (1980), 732–744.

Chung K.L., *A course in probability theory*, Academic Press, New York and London (1974).

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

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

Götze, F., *Approximations for Multivariate U-Stastistics*, J. Mult. An. **22** (1987), 212–229.

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

Götze F. and van Zwet W., *Edgeworth expansions for asymptotically linear statistics*, Preprint 91–034 of SFB 343 "Diskrete Strukturen in der mathematik", Universität Bielefeld, 1991, revised version 1992.

Graham S.W. and Vaaler J.D., *A class of extremal functions for the Fourier transform*, Trans. of the Amer. Math. Soc. **265** (1981).

Hall P., *Edgeworth expansion for Student's t statistic under minimal moment conditions*, Ann. Probab. **15** (1987), 920–931.

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

Helmers, R. and van Zwet, W., *The Berry-Esseen bound for U-statistics*, Statist. Dec. Theory Rl. Top. ((S.S. Gupta and J.O. Berger), eds.), vol. 1, Academic Press,, New York, 1982, pp. 497–512.

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

Lai, T.L., Wang, J.Q., *Edgeworth expansions for symmetric statistics with applications to Bootstrap methods*, Statistica Sinica **3** (1993), 517–542.

de la Peña V., *Decoupling and Khintchine's inequalities for U-statistics*, Ann. Prob. **20** (1992), 1877–1892.

Petrov, V.V., *Sums of independent random variables*, Springer, New York (1975).

Pfanzagl, J. (Wefelmeyer, W.), *Asymptotic expansions for general statistical models*, Lect. Notes in Statist. **31** (1985), Springer-Verlag Heidelberg/New York.

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

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

Zitikis R., *A Berry-Esseen bound for multivariate L-estimates with explicit dependence on dimensions*, Lect. Notes in Math. Stability problems for stochastic models. V.V.Kalashnikov and V.M.Zolotariov (Eds.) Proceedings, Suzdal-1991 **1546** (1993), Springer-Verlag Heidelberg/New York, 197–211.

van Zwet, W.R., *A Berry-Esseen bound for symmetric statistics*, Z. Wahrsch. verw. Gebiete **66** (1984), 425–440.

Vidmantas Bentkus
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld 1
Germany
E-mail address: bentkus@mathematik.uni-bielefeld.de

Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld 1
Germany
E-mail address: goetze@mathematik.uni-bielefeld.de

Willem R. van Zwet
Department of Mathematics
University of Leiden
PO Box 9512
Leiden 2300 RA
Netherlands
E-mail address: vanzwet@rulcri.leidenuniv.nl