EMPIRICAL EDGEWORTH EXPANSIONS

FOR SYMMETRIC STATISTICS[1]

BY HEIN PUTTER AND WILLEM R. VAN ZWET

University of Leiden and University of North Carolina, Chapel Hill

In this paper the validity of a one-term Edgeworth expansion for Studentized symmetric statistics is proved. We propose jackknife estimates for the unknown constants appearing in the expansion and prove their consistency. As a result we obtain the second-order correctness of the empirical Edgeworth expansion for a very general class of statistics, including U-statistics, L-statistics and smooth functions of the sample mean. We illustrate the application of the bootstrap in the case of a U-statistic of degree two.

1. Introduction

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and P a probability measure on $(\mathcal{X}, \mathcal{A})$. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, taking values in $\mathcal{X}$ with unknown common distribution P. Let $t_N : \mathcal{X}^N \times \mathcal{P} \to \mathbb{R}$ be symmetric as a function on $\mathcal{X}^N$, i.e. for every $x_1, \ldots, x_N \in \mathcal{X}$ and every permutation $\{\alpha_1, \ldots, \alpha_N\}$ of $\{1, \ldots, N\}$, we have

$$t_N(x_1, \ldots, x_N; P) = t_N(x_{\alpha_1}, \ldots, x_{\alpha_N}; P) \ .$$

It will be assumed throughout this paper that

$$T_N = t_N(X_1, \ldots, X_N; P) \ , \tag{1.1}$$

is a random variable with expectation

$$ET_N = 0 \qquad \text{for all } N \tag{1.2}$$

[1] The research for this paper was supported by the Netherlands Organization for Scientific Research (NWO) and the Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik" at the University of Bielefeld, Germany.

AMS 1991 subject classifications. Primary 62 E 20, 62 G 09.

Key words and phrases. Edgeworth expansion, Hoeffding's decomposition, jackknife, bootstrap.

and variance $\sigma^2(T_N) = \sigma_N^2$, satisfying

$$0 < c \leq \sigma_N^2 \leq C < \infty \qquad \text{for all } N \tag{1.3}$$

for finite positive constants c and C. Suppose that T_N/σ_N converges in distribution to a standard normal distribution. Typically the accuracy of the normal approximation is of the order $N^{-\frac{1}{2}}$ as N tends to infinity. In this paper we shall focus on second-order approximations, i.e. on approximations with error $o(N^{-\frac{1}{2}})$ to the distribution functions of T_N/σ_N and also of the Studentized version T_N/S_N, where S_N^2 is an estimator of σ_N^2. We shall accomplish this by first proving the validity of a (one-term) Edgeworth expansion with remainder $o(N^{-\frac{1}{2}})$ and then estimating the unknown constants in the expansion. The procedure is therefore called an *empirical Edgeworth expansion.*

As an estimate S_N^2 of σ_N^2 we shall use the jackknife estimator of variance, introduced by Quenouille (1949, 1956) and Tukey (1958). Let us suppose that we have one additional observation X_{N+1} at our disposal, and define for $i = 1, \ldots, N$,

$$T_N^{(i)} = t_N(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{N+1}; P) \ , \qquad T_N^{(N+1)} = T_N \ , \tag{1.4}$$

and

$$\overline{T}_N = \frac{1}{N+1} \sum_{i=1}^{N+1} T_N^{(i)} \ . \tag{1.5}$$

The jackknife estimator of variance S_N^2 is then defined as

$$S_N^2 = \sum_{i=1}^{N+1} \left(\overline{T}_N - T_N^{(i)}\right)^2 \ . \tag{1.6}$$

At first sight it may seem somewhat awkward to try to estimate the variance of T_N with $N+1$ observations from P, since for practical purposes we would then wish to approximate the distribution of T_{N+1} instead of T_N, and the variance of T_{N+1} would be the object of interest. With σ_N^2 satisfying (1.3), however, σ_N^2 will typically converge to a constant and the difference between σ_N^2 and σ_{N+1}^2 will vanish. In practically all cases of interest it will not influence the one-term Edgeworth expansions either. We shall therefore not be concerned with this difference.

We shall make extensive use of the properties of Hoeffding's decomposition, introduced by Hoeffding (1948). For $k = 1, 2, \ldots$, let Ω_k denote the set of integers $\{1, \ldots, k\}$ and define for a set $A \subset \Omega_N$,

$$E(T_N|A) = E(\,T_N \mid X_i, i \in A\,) \tag{1.7}$$

Next, for $D \subset \Omega_N$, define

$$T_{N,D} = \sum_{A \subset D} (-1)^{|D|-|A|} \ E(T_N|A) \ . \tag{1.8}$$

Here $|A|$ denotes the cardinality of a set A and the summation in (1.8) is over all subsets A of D including the empty set. In this way we obtain for instance

$$\begin{aligned} T_{N,\emptyset} &= ET_N \\ T_{N,\{i\}} &= E(T_N|X_i) - ET_N \ , \\ T_{N,\{i,j\}} &= E(T_N|X_i, X_j) - E(T_N|X_i) - E(T_N|X_j) + ET_N \ . \end{aligned} \tag{1.9}$$

In the following lemma we have collected some useful facts about the random variables defined in (1.8). The first of these asserts that T_N can be written as a sum of all $T_{N,D}$, which is called Hoeffding's decomposition of T_N. A proof of these facts can be found in van Zwet (1984, 1990, 1995).

Lemma 1.1.

(1.10) $T_N = \sum_{D \subset \Omega_N} T_{N,D}$,

(1.11) $E(T_{N,D}|D') = 0$ a.s. *unless* $D \subset D'$,

(1.12) *If* T_N *depends only on those* X_i *with* $i \in D'$, *then* $T_{N,D} = 0$ a.s. *unless* $D \subset D'$,

(1.13) $T_{N,\emptyset} = ET_N$ *and* $ET_{N,D} = 0$ *for* $D \neq \emptyset$,

(1.14) $\text{cov}(T_{N,D}, T_{N,D'}) = ET_{N,D}T_{N,D'} = 0$, *for* $D \neq D'$,

(1.15) $ET_N^2 = \sum_{D \subset \Omega_N} ET_{N,D}^2$; $\sigma^2(T_N) = \sum_{\emptyset \neq D \subset \Omega_N} ET_{N,D}^2$,

(1.16) $E(T_N|A) = \sum_{D \subset A} T_{N,D}$; $E(E(T_N|A))^2 = \sum_{D \subset A} ET_{N,D}^2$,

(1.17) $ET_{N,D}^2 = \sum_{A \subset D} (-1)^{|D|-|A|} E(E(T_N|A))^2$,

(1.18) *Let* $D_1, \ldots, D_l$ *be any collection of subsets of* Ω_N. *If there exists at least one index* i *that appears in exactly one of the sets* $D_1, \ldots, D_l$, *then* $ET_{N,D_1}T_{N,D_2} \cdots T_{N,D_l} = 0$.

The decomposition and its properties remain valid when $X_1, \ldots, X_N$ are independent but not identically distributed. However, when $X_1, \ldots, X_N$ are i.i.d. and T_N is a symmetric statistic, $T_{N,D}$'s where D's have the same number of elements have a similar form and we may write

$$T_{N,D} = \psi_{N,k}(X_{i_1}, X_{i_2}, \ldots, X_{i_k}), \qquad \text{for } D = \{i_1, i_2, \ldots, i_k\} \ , \tag{1.19}$$

where the function $\psi_{N,k}$ is symmetric and depends on D only through $|D| = k$. The Hoeffding decomposition of T_N now looks like

$$\begin{aligned} T_N = ET_N &+ \sum_{i=1}^{N} \psi_{N,1}(X_i) + \sum\sum_{i<j} \psi_{N,2}(X_i, X_j) \\ &+ \sum\sum\sum_{i<j<k} \psi_{N,3}(X_i, X_j, X_k) + \ldots + \psi_{N,N}(X_1, \ldots, X_N) \ . \end{aligned} \tag{1.20}$$

Choosing Ω_k as a representative subset of Ω_N with k elements, we see that the variance of T_N becomes

$$\sigma^2(T_N) = \sum_{k=1}^{N} \binom{N}{k} ET_{N,\Omega_k}^2 = \sum_{k=1}^{N} \binom{N}{k} E\psi_{N,k}^2(X_1, \ldots, X_k) \ . \tag{1.21}$$

Let us write the Hoeffding decomposition of T_N as

$$T_N = \sum_{D \subset \Omega_N} T_{N,D} = \sum_{i=1}^{N} T_{N,\{i\}} + \sum\sum_{1 \le i < j \le N} T_{N,\{i,j\}} + \sum\sum\sum_{1 \le i < j < k \le N} T_{N,\{i,j,k\}} + \dots$$

For notational convenience we shall write T_{Ni} instead of $T_{N,\{i\}}$ and T_{Nij} instead of $T_{N,\{i,j\}}$. Define two real numbers λ_1 and λ_2 as

$$\lambda_1 = N^{\frac{3}{2}} \sigma_N^{-3} E T_{N1}^3 \ , \quad \lambda_2 = N^{\frac{5}{2}} \sigma_N^{-3} E T_{N1} T_{N2} T_{N12} \ . \tag{1.22}$$

Then the Edgeworth expansion for the distribution function of T_N/σ_N is given by

$$G_N(x) = \Phi(x) - \frac{\lambda_1 + 3\lambda_2}{6\sqrt{N}} (x^2 - 1)\phi(x) \ . \tag{1.23}$$

Note that $(\lambda_1 + 3\lambda_2)N^{-\frac{1}{2}}$ serves as an approximation to the third cumulant of T_N/σ_N. The Edgeworth expansion to the distribution function of T_N/S_N is given by

$$H_N(x) = \Phi(x) + \frac{\phi(x)}{6\sqrt{N}} \big((2x^2 + 1)\lambda_1 + 3(x^2 + 1)\lambda_2 \big) \ . \tag{1.24}$$

The empirical Edgeworth expansions are obtained by replacing the constants λ_1 and λ_2 in (1.23)–(1.24) by estimates. We shall use estimates similar to the jackknife estimator of variance. A description of these estimates, which are much like the ones used by Beran (1982), may be found in Section 4.

The typical situation to which these empirical Edgeworth expansions may be applied is the following. Let $\theta = \theta(P)$ be a parameter of interest and suppose that $U_N = u_N(X_1, \dots, X_N)$ is an unbiased estimator of θ. As $X_1, \dots, X_N$ are i.i.d., we may safely restrict attention to symmetric functions u_N. Let

$$T_N = \sqrt{N}\,(U_N - \theta) \ , \tag{1.25}$$

and suppose that with $\sigma_N^2 = \sigma^2(T_N)$, $T_N/\sigma_N \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. We wish to obtain a second-order correct confidence interval for θ. For the special case of the sample mean it is well known that it is important to base inference on a pivotal random variable, i.e. on a random variable whose limiting distribution does not depend on any unknown quantities. If σ_N^2 is known, we might take T_N/σ_N; if σ_N^2 is unknown, we have to use T_N/S_N, with S_N^2 an appropriate estimator of σ_N^2. The confidence interval for θ may then be based on the quantiles of the empirical Edgeworth expansion of the distribution function of T_N/σ_N or T_N/S_N.

Note that $T_N^{(i)}$ and $\overline{T}_N$ in (1.4)–(1.5) are not directly computable, since they depend on the unknown distribution P. However, in the present case, where $T_N = \sqrt{N}\,(U_N - \theta)$, and U_N is a true statistic (i.e. not depending on P) , the jackknife estimator of variance is computable. We have the relation

$$T_N^{(i)} = \sqrt{N} \left(U_N^{(i)} - \theta \right) \ ,$$

where $U_N^{(i)}$ is U_N computed for the sample with the i^{th} element replaced by the additional observation X_{N+1}. The jackknife estimator is defined in terms of differences of these $T_N^{(i)}$'s, so with $\overline{T}_N$ equal to $\sqrt{N}(\overline{U}_N - \theta)$, where $\overline{U}_N = \frac{1}{N+1}\sum_{i=1}^{N+1} U_N^{(i)}$, we have

$$S_N^2 = \sum_{i=1}^{N+1}\left(\overline{T}_N - T_N^{(i)}\right)^2 = N\sum_{i=1}^{N+1}\left(\overline{U}_N - U_N^{(i)}\right)^2 \;,$$

and this is certainly computable. The jackknife estimates of λ_1 and λ_2, to be proposed in Section 4 are also independent of θ, as we shall see later.

In one of the very first papers on the bootstrap it was recognized by Singh (1981) that Efron's bootstrap gives an approximation to the true distribution function of the standardized sample mean that is better then the traditional normal approximation. The empirical Edgeworth expansion is to second order equivalent to the bootstrap. Beran (1982, 1984) showed that for locally quadratic functionals $T(F)$, the bootstrap approximation and the empirical Edgeworth expansion to the distribution function of $\frac{T(F_N)-T(F)}{\sigma(T(F_N))}$ agree to second order and that both are asymptotically minimax. General though Beran's approach is, from a practical point of view his results and those of Singh (1981) are only a partial solution to the problem, because no Studentization is employed, and therefore their methods do not provide a second-order correct procedure when the variance of the estimator involved is unknown. Babu & Singh (1983, 1984) and later Bhattacharya & Qumsiyeh (1989) showed that both bootstrap and empirical Edgeworth are better than the normal approximation for Student's t statistic.

Results, similar to ours were obtained earlier in a paper by Helmers (1991) in the special case of Studentized U-statistics of degree two. To put his results in a proper perspective, let $h : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a symmetric function with $Eh(X_1, X_2) = \theta$ and $Eh^2(X_1, X_2) < \infty$, and let

$$U_N = \binom{N}{2}^{-1} \sum\sum_{1\le i<j\le N} h(X_i, X_j) \tag{1.26}$$

be a U-statistic of degree two. Define $T_N = \sqrt{N}(U_N - \theta)$, $g(x) = E(h(X_1, X_2)|X_1 = x) - \theta$ and $\psi(x, y) = h(x, y) - g(x) - g(y) - \theta$. If $\sigma_g^2 = Eg^2(X_1) > 0$, then U_N is called a non-degenerate U-statistic, and we have asymptotic normality of T_N (cf. Hoeffding (1948)). The Hoeffding decomposition of T_N is given by

$$T_N = \frac{2}{\sqrt{N}}\sum_{i=1}^{N} g(X_i) + \frac{2}{(N-1)\sqrt{N}}\sum\sum_{1\le i<j\le N}\psi(X_i, X_j) \;, \tag{1.27}$$

so that in this case λ_1 and λ_2 are asymptotically equivalent to

$$\lambda_1 = \frac{Eg^3(X_1)}{(Eg^2(X_1))^{\frac{3}{2}}} \quad , \quad \lambda_2 = \frac{Eg(X_1)g(X_2)\psi(X_1, X_2)}{(Eg^2(X_1))^{\frac{3}{2}}} \;.$$

Let S_N^2 be the jackknife estimator of variance of T_N and let H_N be defined as in (1.24). Helmers (1991) proved that if the distribution of $g(X_1)$ is non-lattice and if there exists $r > 4$ with $E|h(X_1, X_2)|^r < \infty$, then

$$N^{\frac{1}{2}} \sup_{x \in \mathbb{R}} |P(T_N/S_N \leq x) - H_N(x)| \to 0 \ , \qquad \text{as } N \to \infty \ .$$

Helmers uses bootstrap estimates for λ_1 and λ_2, which in the case of a U-statistic allow an explicit expression. In general however, resampling would be needed to compute them. He shows that both the Edgeworth expansion $H_N(x)$ with λ_1 and λ_2 replaced by bootstrap estimates, and the bootstrap approximation approximate $P(T_N/S_N \leq x)$ up to $o(N^{-\frac{1}{2}})$ almost surely. In Section 5 we shall apply our results to U-statistics, showing that Helmers' moment conditions may be relaxed. Section 6 is devoted to bootstrapping a U-statistic of degree two.

The remainder of this paper is organized as follows. In Section 2 a useful representation of the jackknife estimator of variance is proved. In Section 3 we present a proof of the validity of a one-term Edgeworth expansion for Studentized symmetric statistics. Section 4 contains the jackknife estimators of the quantities appearing in the Edgeworth expansion and we conclude in Sections 5 and 6 with some important applications.

2. The jackknife estimator of variance

Consider the set-up of the previous section. The jackknife estimator of variance of $\sigma_N^2 = \sigma^2(T_N)$ is given by (1.6). To let the estimation of σ_N^2 make sense we have assumed (1.3). Let

$$T_N = \sum_{D \subset \Omega_N} T_{ND} = \sum_{i=1}^N T_{Ni} + \sum\sum_{1 \leq i < j \leq N} T_{Nij} + \dots$$

denote Hoeffding's decomposition of T_N. If, with one additional observation, we want to express $T_N^{(i)}$ in terms of functions appearing in the Hoeffding decomposition of T_N we are confronted with a small problem of notation that we have to deal with. The random variable T_N has a Hoeffding decomposition $T_N = \sum_{D \subset \Omega_N} T_{ND}$, where $\Omega_N = \{1, \dots, N\}$ as before, and for $D = \{i_1, \dots, i_k\} \subset \Omega_N$,

$$T_{ND} = \sum_{A \subset D} (-1)^{|D|-|A|} \ E(T_N|A) = \psi_{Nk}(X_{i_1}, \dots, X_{i_k}) \ .$$

Similarly, $T_N^{(i)} = \sum_{D \subset \Omega_{N+1}-\{i\}} T_{ND}^{(i)}$, and for $D \subset \Omega_N, i \notin D$, we have $T_{ND}^{(i)} = T_{ND}$. However, if $N+1 \in D, i \notin D$ (or $i \in D, N+1 \notin D$), then $T_{ND} = 0$ (or $T_{ND}^{(i)} = 0$) by (1.12), and $T_{ND}^{(i)}$ is not the same as T_{ND}! Nevertheless, it is still true that $T_{ND}^{(i)} = \psi_{Nk}(X_{i_1}, \dots, X_{i_k})$ for $D = \{i_1, \dots, i_k\} \subset \Omega_{N+1} - \{i\}$, even if $D \not\subset \Omega_N$. Hence we define

$$\tilde{T}_{ND} = \begin{cases} \psi_{Nk}(X_{i_1}, \dots, X_{i_k}), & \text{for } D = \{i_1, \dots, i_k\} \subset \Omega_{N+1}, |D| \leq N \ , \\ 0, & \text{for } |D| > N \ . \end{cases}$$

This way all $T_N^{(i)}$, including $T_N = T_N^{(N+1)}$, have Hoeffding decompositions in terms of the same random variables $\tilde{T}_{ND}$:

$$\begin{aligned} T_N^{(i)} &= \sum_{\substack{j=1\\ j\neq i}}^{N+1} \psi_{N1}(X_j) + \sum\sum_{\substack{1\leq j<k\leq N+1\\ j,k\neq i}} \psi_{N2}(X_j, X_k) + \ldots \\ &= \sum_{k=1}^{N} \sum_{\substack{D\subset\Omega_{N+1}\\ |D|=k, i\notin D}} \tilde{T}_{ND} = \sum_{\substack{D\subset\Omega_{N+1}\\ i\notin D}} \tilde{T}_{ND} \ . \end{aligned}$$

Note that $\tilde{T}_{ND} = T_{ND}$ for $D \subset \Omega_N$ and that $\tilde{T}_{ND}$ still have the basic property (cf. (1.11)) : $E(\tilde{T}_{ND}|D') = 0$, if $D \neq D'$. For notational convenience we shall even go a step further and write T_{ND} again instead of $\tilde{T}_{ND}$, if this does not cause confusion. The reader should keep in mind however that T_{ND} is now defined by $T_{ND} = \psi_{Nk}(X_{i_1}, \ldots, X_{i_k})$, if $D = \{i_1, \ldots, i_k\} \subset \Omega_{N+1}$ with $|D| \leq N$. Thus we obtain

$$\overline{T}_N = \frac{1}{N+1}\sum_{i=1}^{N+1} T_N^{(i)} = \sum_{i=1}^{N+1} \frac{1}{N+1} \sum_{\substack{D\subset\Omega_{N+1}\\ i\notin D}} T_{ND} = \sum_{D\subset\Omega_{N+1}} \left(1 - \frac{|D|}{N+1}\right) T_{ND} \ ,$$

so that

$$\begin{aligned} \overline{T}_N - T_N^{(i)} = \sum_{D\subset\Omega_{N+1}} \left(1 - \frac{|D|}{N+1}\right) T_{ND} &- \sum_{\substack{D\subset\Omega_{N+1}\\ i\notin D}} T_{ND} \\ &= \sum_{D\subset\Omega_{N+1}} \left(\mathbf{1}_D(i) - \frac{|D|}{N+1}\right) T_{ND} \ . \end{aligned}$$

Necessary and sufficient conditions for consistency of the jackknife estimator of variance are given in the following result, due to van Zwet (1990, 1995).

Theorem 2.1. *Assume (1.2) and (1.3). Then S_N^2 is an asymptotically unbiased and consistent estimator of $\sigma^2(T_N)$ if and only if*

$$\lim_{N\to\infty} N \sum_{k=2}^{N} \binom{N-1}{k-1} ET^2_{N\Omega_k} = 0 \tag{2.1}$$

and

$$NET^2_{N1}\mathbf{I}_{\{|T_{N1}|\geq\varepsilon\}} \to 0 \ , \qquad \text{for every } \varepsilon > 0 \ . \tag{2.2}$$

Note that, since $N\sum_{k=2}^{N}\binom{N-1}{k-1}ET^2_{N\Omega_k} \geq \sum_{k=2}^{N}\binom{N}{k}ET^2_{N\Omega_k}$, (2.1) and (2.2) imply that T_N is asymptotically linear and asymptotically normal.

For the Edgeworth expansion of T_N/S_N it will be necessary to have more precise information concerning the behaviour of S_N^2. We shall prove the following theorem.

Theorem 2.2. *Suppose that there exist numbers $p > 3$, $r > 2$, $c > 0$ and $C > 0$, such that for $N = 2, 3, \ldots$, (1.2) and (1.3) are satisfied and*

$$E|N^{\frac{1}{2}} T_{N1}|^p \leq C \ , \tag{2.3}$$

$$E|N^{\frac{3}{2}} T_{N12}|^r \leq C \ ; \tag{2.4}$$

$$\begin{cases} \sum_{k=3}^{N} \binom{N-1}{k-1} ET^2_{N\Omega_k} \leq CN^{-3} \\ \sum_{k=3}^{N} \binom{N-2}{k-2} ET^2_{N\Omega_k} \leq CN^{-\frac{7}{2}} \ . \end{cases} \tag{2.5}$$

Then there exist sequences $\delta_N \searrow 0$ and $\varepsilon_N \searrow 0$, depending on T_N and P only through p, r, c and C, such that for $N = 2, 3, \ldots$,

$$P\left(\left| S_N^2 - \sigma_N^2 - \sum_{i=1}^{N+1} (K_{Ni} + L_{Ni}) - \sum\sum_{1 \leq i < j \leq N+1} M_{Nij} \right| \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \delta_N N^{-\frac{1}{2}} \ , \tag{2.6}$$

where

$$K_{Ni} = f_{N1}(X_i) = 2NE\,(T_{Nij} T_{Nj} \,|\, X_i) \ , \quad j \neq i \ , \tag{2.7}$$

$$L_{Ni} = f_{N2}(X_i) = T^2_{Ni} - ET^2_{N1} \ , \tag{2.8}$$

and for $i \neq j$

$$\begin{aligned} M_{Nij} = f_{N3}(X_i, X_j) = 2T_{Nij}(T_{Ni} + T_{Nj}) &- 2E\,(T_{Nij} T_{Nj} \,|\, X_i) \\ &- 2E\,(T_{Nij} T_{Ni} \,|\, X_j) \ . \end{aligned} \tag{2.9}$$

Theorem 2.2 should be compared for instance with identity (A8) in Callaert & Veraverbeke (1981). It asserts essentially that the difference between S_N^2 and σ_N^2 can be expressed as a U-statistic plus a remainder term which is of negligible order for our purposes.

Theorem 2.2, as well as the results in Section 3 about the Edgeworth expansions, are formulated as inequalities for fixed, but arbitrary N. Since the constants in the conclusions are not specified however, they should be viewed as purely asymptotic results. The reason for phrasing the assumptions and conclusions in these results in such a laborious way is that we want to define uniformity classes. The constants and sequences appearing in the conclusions of the results depend only on the constants and sequences appearing in the assumptions and in particular *not on* N. This allows us to consider a sequence of problems, indexed by N, where for every N the random variables X_i may be different, as well as their distributions P, the functions $T_N = t_N(X_1, \ldots, X_N; P)$ etc., as long as the conditions continue to be satisfied for the same fixed constants and sequences for every N. The conclusions of the theorems are then also true for every N and the asymptotic assertion follows, uniformly in P and $\{T_N\}$.

In the proofs, we should then formally also be working with inequalities that are true for every N. Phrases like $o(N^{-1})$ are strictly speaking not allowed, because there is no N

tending to infinity. To work with inequalities throughout the proofs would however become extremely tedious. To avoid these laborious formulations from occurring throughout the proofs and the more informal passages in the text, we shall use o and $\mathcal{O}$ symbols all the same, agreeing that they are uniform in everything that satisfies the assumptions for fixed constants and sequences. Thus, the statement $a_N = o(N^{-1})$ e.g. should be read as $Na_N \leq \varepsilon_N$, for some sequence ε_N tending to zero, depending only on the constants and sequences in the assumptions. Similarly, $a_N = \mathcal{O}(N^{-1})$ will stand for $Na_N \leq C$, for some constant C, depending only on the constants and sequences in the assumptions.

Before setting out to prove Theorem 2.2, we establish a truncation lemma.

Lemma 2.3. (Truncation Lemma) *Let Y be a random variable with $EY = 0$ and $E|Y|^s = \nu_s < \infty$ for some $s \geq 1$. Then, for every $\eta > 0$, there exists $Y' = \varrho(Y)$ such that*

$$\begin{aligned}
&|Y'| \leq \eta \quad \text{a.s. ;} \\
&P(Y' \neq Y) \leq \frac{2\nu_s}{\eta^s} \ ; \\
&EY' = 0 \ ; \\
&E|Y' - Y|^t \leq \frac{2^{2t+1}\nu_s}{\eta^{s-t}} \ , \qquad \text{for every } 0 < t \leq s \ .
\end{aligned}$$

Proof. Choose $\eta > 0$ and define

$$Y'' = \begin{cases} -\eta \ , & \text{if } Y < -\eta \\ Y \ , & \text{if } -\eta \leq Y \leq \eta \\ \eta \ , & \text{if } Y > \eta \ . \end{cases}$$

We have $|Y''| \leq \eta$ a.s. ,

$$P(Y'' \neq Y) = P(|Y| > \eta) \leq \frac{\nu_s}{\eta^s} \ ,$$

and

$$E|Y'' - Y|^t = E\left(|Y| - \eta\right)^t \mathbf{I}_{\{|Y| \geq \eta\}} \leq E|Y|^t \left(\frac{|Y|}{\eta}\right)^{s-t} = \frac{\nu_s}{\eta^{s-t}} \ , \ \text{for } 0 < t \leq s \ .$$

Next we change Y'' slightly to make its expectation vanish. By the above we have $|EY''| \leq \frac{\nu_s}{\eta^{s-1}}$ because $EY = 0$. Assume without loss of generality that $EY'' < 0$ and change the value of Y'' on a set of probability $\leq \frac{\nu_s}{\eta^s}$ where $-\eta \leq Y'' < 0$ to the value $+\eta$ until the expectation equals zero. This can actually be done, since otherwise this process would end with a nonnegative random variable with negative expectation. Call the resulting variable $Y' = \varrho(Y)$.

Obviously, $|Y'| \leq \eta$ and $EY' = 0$. Also, $P(Y'' \neq Y') \leq \frac{\nu_s}{\eta^s}$ so $P(Y' \neq Y) \leq \frac{2\nu_s}{\eta^s}$. Since we have obtained Y' from Y'' by changing the value of Y'' by at most 2η on a set of probability at most $\frac{\nu_s}{\eta^s}$, we have $E|Y' - Y''|^t \le \frac{\nu_s}{\eta^s}(2\eta)^t$. Hence

$$E|Y' - Y|^t \le (2^{t-1} \vee 1)\left\{E|Y'' - Y|^t + E|Y' - Y|^t\right\}$$
$$\le \frac{\nu_s}{\eta^{s-t}}(1+2^t)(2^{t-1} \vee 1) \le 2^{(2t+1)}\frac{\nu_s}{\eta^{s-t}} \ .$$

This proves the lemma. □

In Sections 2 and 3 we want to apply this lemma in situations where, as in the conclusion of Theorem 2.2, events of probability $o(N^{-\frac{1}{2}})$ are negligible. We may then replace random variables Y_{Ni} by Y'_{Ni} as long as the probability that any of them are different is of this negligible order $o(N^{-\frac{1}{2}})$. The first application concerns the random variables T_{Ni}. We have $E|T_{Ni}|^p \le CN^{-p/2}$ for some $p > 3$, by (2.3). In Lemma 2.3 we therefore choose $s = p > 3$ and $\eta = N^{-\frac{p-3}{4p}}$ to obtain

Corollary 2.4. *Suppose that* $E|N^{\frac{1}{2}}T_{N1}|^p \le C$ *for some* $p > 3$. *Then there exist* i.i.d. $T'_{N1}, \dots, T'_{N,N+1}$ *with* $T'_{Ni} = \varrho_N(T_{Ni})$ *satisfying*

$$|T'_{Ni}| \le N^{-\frac{p-3}{4p}} = o(1) \ ; \tag{2.10}$$

$$P(T'_{Ni} = T_{Ni}\ , i = 1, \dots, N+1) \ge 1 - 2C\frac{N+1}{N}N^{-\frac{p-1}{4}} \ge \ge 1 - 4CN^{-\frac{p-1}{4}} = 1 - o(N^{-\frac{1}{2}}) \ ; \tag{2.11}$$

$$ET'_{Ni} = 0 \ ; \tag{2.12}$$

$$N^{\frac{t}{2}}E|T'_{Ni} - T_{Ni}|^t \le 2^{2t+1}CN^{-\frac{(p+3)(p-t)}{4p}} = \begin{cases} o(N^{-\frac{3(p-t)}{2p}})\ , & 0 < t < p \\ \mathcal{O}(1)\ , & t = p\ . \end{cases} \tag{2.13}$$

Throughout the proofs of Theorems 2.2 and 3.3 we shall assume that the T_{Ni} have been replaced by their truncated versions T'_{Ni}. For simplicity we delete the prime in T'_{Ni}, thus in effect assuming that (2.10), (2.12), and (2.13) are satisfied for the original T_{Ni}.

In the formulation of the theorems however, both the assumptions and the conclusions are stated in terms of the original T_{Ni}. We should therefore make sure in the first place that the assumptions for the original T_{Ni} imply the same assumptions for the truncated T_{Ni}. Secondly, having conducted the proof with the truncated random variables, we obtain a conclusion for the truncated random variables, and hence we also have to check that the conclusion of the theorem is still true when we replace the truncated T_{Ni} by the original T_{Ni}.

To see that all this is justified in Theorems 2.2 and 3.3, we note that the probability that the substitution affects the values of any of the T_{Ni}, and hence of T_N, S_N^2, T_{Ni}^2 or $T_{Nij}(T_{Ni} + T_{Nj})$, is $o(N^{-\frac{1}{2}})$ uniformly in P and T_N satisfying (2.3). Exceptional events with probabilities of this order of magnitude are allowed in all our results.

It remains to check that the substitution does not affect the assumptions or conclusions of our results in any other way. Clearly, (2.13) guarantees that (2.3) implies that $E|N^{\frac{1}{2}}T'_{N1}|^p \le 2^{p-1}C(2^{2p+1}+1)$, so that assumption (2.3) is satisfied for T'_{N1}. In Theorem

3.3 we shall also encounter the assumption that there exist a sequence $\tau_N \to \infty$ and a positive function χ on $(0,\infty)$, such that $|E\exp\{itN^{\frac{1}{2}}T_{N1}\}| \le 1-\chi(t)$ for all $t\in(0,\tau_N)$ for all N. Now (2.11) and (2.13) imply

$$\left|E\exp\{itN^{\frac{1}{2}}T'_{N1}\} - E\exp\{itN^{\frac{1}{2}}T_{N1}\}\right| \le 8CN^{-\frac{p-1}{4}} \le 8CN^{-\frac{1}{2}} \qquad \text{for all } t \ ,$$

$$\left|E\exp\{itN^{\frac{1}{2}}T'_{N1}\}\right| \le 1 - 1/3\ NET'^{2}_{N1}t^2 \ , \qquad \text{if } |t| \le \frac{E(N^{\frac{1}{2}}T'_{N1})^2}{E|N^{\frac{1}{2}}T'_{N1}|^3} \ .$$

Since the remaining assumptions ensure that $NET^2_{N1} \sim \sigma^2_N \ge c > 0$, and (2.13) yields

$$N|ET'^{2}_{N1} - ET^2_{N1}| \le 32CN^{-\frac{(p+3)(p-2)}{4p}} \ , \tag{2.14}$$

we find that $E(N^{\frac{1}{2}}T'_{N1})^2$ and $E|N^{\frac{1}{2}}T'_{N1}|^3$ are bounded away from zero and infinity. Hence, for some $t_0>0$ and integer N_0, $\frac{E(N^{\frac{1}{2}}T'_{N1})^2}{E|N^{\frac{1}{2}}T'_{N1}|^3} \ge t_0$ for all $N\ge N_0$. Now, for $0<t<\tau_N$,

$$\left|E\exp\{itN^{\frac{1}{2}}T'_{N1}\}\right| \le 1-\chi(t)+8CN^{-\frac{1}{2}} \le 1-\chi(t)/2 \ ,$$

if t is such that $\chi(t)\ge 16CN^{-\frac{1}{2}}$. Since χ is a positive continuous function, it has a positive minimum on every closed interval $K=[t_0,\tau]$, and hence we may choose a $\tau=\tau'_N<\tau_N$ for sufficiently large N, so that the minimum over K of $\chi(t)$ is still larger than $16CN^{-\frac{1}{2}}$. The fact that $16CN^{-\frac{1}{2}}\to 0$ as $N\to\infty$, allows us to choose τ'_N tending to infinity, as N tends to infinity. Defining a function

$$\tilde{\chi}(t) = \min(\frac{\chi(t)}{2}, \frac{c}{4}t^2) \ ,$$

it follows that there exists a sequence $\tau'_N\to\infty$, such that for sufficiently large $N\ge N_0$, $|E\exp\{itN^{\frac{1}{2}}T'_{N1}\}| \le 1-\tilde{\chi}(t)$ for all $t\in(0,\tau'_N)$. It is easy to see that the function $\tilde{\chi}$, the sequence $\{\tau'_N\}$ and N_0 depend only on P and $\{T_N\}$ through the various constants and sequences in our assumptions (including $\{\tau_N\}$) and this is therefore all we need.

In the conclusion of Theorem 2.2 we encounter the random variables T^2_{Ni} an $T_{Nij}(T_{Ni}+T_{Nj})$ that we discussed above, but also ET^2_{N1} and $NE(T_{Nij}T_{Nj}\,|\,X_i)$. We find that L_{Ni} in (2.6) contains NET^2_{N1}. In view of (2.14), substituting ET^2_{N1} for ET'^{2}_{N1} here makes no difference. Writing $\tilde{K}_{Ni}=2NE(T_{Nij}T'_{Nj}\,|\,X_i)$ for $j\ne i$ we have, uniformly in P and $\{T_N\}$ satisfying our assumptions,

$$\begin{aligned}\left(\tilde{K}_{Ni}-K_{Ni}\right)^2 &= \left[2NE\left(T_{Nij}(T'_{Nj}-T_{Nj})\,|\,X_i\right)\right]^2 \\ &\le 4N^2E(T^2_{Nij}\,|\,X_i)E(T'_{Nj}-T_{Nj})^2 = o\left(N^{1-\frac{3(p-2)}{2p}}\cdot E(T^2_{Nij}\,|\,X_i)\right)\end{aligned}$$

by (2.13). Hence, by (2.4),

$$E\left[\sum_{i=1}^{N+1}(\tilde{K}_{Ni}-K_{Ni})\right]^2=(N+1)E(\tilde{K}_{N1}-K_{N1})^2=\mathcal{O}(N^{-1-\frac{3(p-2)}{2p}})=\mathcal{O}(N^{-3/2})\ ,$$

so that

$$P\left(\left|\sum_{i=1}^{N+1}(\tilde{K}_{Ni}-K_{Ni})\right|\geq\varepsilon_N N^{-\frac{1}{2}}\right)\leq\delta_N N^{-\frac{1}{2}}$$

for sequences $\varepsilon_N\searrow 0$ and $\delta_N\searrow 0$ depending only on c, C, p and r. Hence (2.6) will hold with K_{Ni} if it holds with $\tilde{K}_{Ni}$.

Finally, in the results of Sections 2 and 3, the quantities σ_N^2, $\lambda_1=N^{\frac{3}{2}}\sigma_N^{-3}ET_{N1}^3$ and $\lambda_2=N^{\frac{5}{2}}\sigma_N^{-3}ET_{N1}T_{N2}T_{N12}$ occur in the conclusion. Application of (2.13) shows that changing T'_{Ni} into T_{Ni} affects σ_N^2, λ_1 and λ_2 only to order $\mathcal{O}(N^{-\frac{1}{2}})$, $\mathcal{O}(1)$ and $\mathcal{O}(1)$ respectively, which does not alter our conclusions.

Summarizing, we conclude that in all proofs of Sections 3.2 and 3.3 we may assume without loss of generality that (2.10), (2.12) and (2.13) hold for the original T_{Ni}.

We are now equipped to prove Theorem 2.2.

Proof of Theorem 2.2. In the course of the proof we shall assume (cf. (2.10))

$$|T_{Ni}|\leq a_N=N^{-\frac{p-3}{4p}}\ ,\qquad i=1,\ldots,N+1\ . \tag{2.15}$$

We write

$$\overline{T}_N-T_N^{(i)}=V_{Ni}+W_{Ni}+Z_{Ni}\ ,$$

where

$$V_{Ni}=\sum_{\substack{D\subset\Omega_{N+1}\\|D|=1}}\left(\mathbf{1}_D(i)-\frac{|D|}{N+1}\right)T_{ND}\ ,$$

$$W_{Ni}=\sum_{\substack{D\subset\Omega_{N+1}\\|D|=2}}\left(\mathbf{1}_D(i)-\frac{|D|}{N+1}\right)T_{ND}\ ,$$

and

$$Z_{Ni}=\sum_{\substack{D\subset\Omega_{N+1}\\|D|\geq 3}}\left(\mathbf{1}_D(i)-\frac{|D|}{N+1}\right)T_{ND}\ .$$

Then

$$S_N^2=\sum_{i=1}^{N+1}(\overline{T}_N-T_N^{(i)})^2=\sum_{i=1}^{N+1}\left[V_{Ni}^2+2V_{Ni}W_{Ni}+W_{Ni}^2+2W_{Ni}Z_{Ni}+2V_{Ni}Z_{Ni}+Z_{Ni}^2\right]\ .$$

By (2.5) we have

$$E \sum_{i=1}^{N+1} Z_{Ni}^2 = N \sum_{k=3}^{N} \binom{N-1}{k-1} ET_{N\Omega_k}^2 \leq CN^{-2} \;, \tag{2.16}$$

so that

$$P\left(\sum_{i=1}^{N+1} Z_{Ni}^2 \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \frac{CN^{-2}}{\varepsilon_N N^{-\frac{1}{2}}} \leq \delta_N N^{-\frac{1}{2}} \;, \tag{2.17}$$

for appropriate $\varepsilon_N \searrow 0, \delta_N \searrow 0$. Since

$$\sum_{i=1}^{N+1} V_{Ni} Z_{Ni} = \sum_{\substack{D \subset \Omega_{N+1} \\ |D|=1}} \sum_{\substack{D' \subset \Omega_{N+1} \\ |D'| \geq 3}} \left[|D \cap D'| - \frac{|D||D'|}{N+1} \right] T_{ND} T_{ND'} \;,$$

we have $E \sum_{i=1}^{N+1} V_{Ni} Z_{Ni} = 0$, and straightforward but tedious calculations show (cf. Putter (1994)) that, using (2.15),

$$E\Big(\sum_{i=1}^{N+1} V_{Ni} Z_{Ni} \Big)^2 \leq \varepsilon_N^* N^{-\frac{3}{2}} \;,$$

for some $\varepsilon_N^* \searrow 0$. Consequently, for $\varepsilon_N \searrow 0$, satisfying $\varepsilon_N^*/\varepsilon_N^2 \to 0$, there exists $\delta_N \to 0$ such that

$$P\left(\left| \sum_{i=1}^{N+1} V_{Ni} Z_{Ni} \right| \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \frac{\varepsilon_N^* N^{-\frac{3}{2}}}{\varepsilon_N^2 N^{-1}} \leq \delta_N N^{-\frac{1}{2}} \;. \tag{2.18}$$

Similar arguments show (for details we refer to Putter (1994)) that there exist $\varepsilon_N \searrow 0$ and $\delta_N \searrow 0$, such that

$$P\left(\left| \sum_{i=1}^{N+1} W_{Ni} Z_{Ni} \right| \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \delta_N N^{-\frac{1}{2}} \;,\; P\left(\sum_{i=1}^{N+1} W_{Ni}^2 \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \delta_N N^{-\frac{1}{2}} \;, \tag{2.19}$$

We proceed with $2 \sum_{i=1}^{N+1} V_{Ni} W_{Ni}$:

$$2 \sum_{i=1}^{N+1} V_{Ni} W_{Ni} = 2 \sum_{\substack{D \subset \Omega_{N+1} \\ |D|=1}} \sum_{\substack{D' \subset \Omega_{N+1} \\ |D'|=2 \\ D \subset D'}} T_{ND} T_{ND'} - \frac{4}{N+1} \sum_{\substack{D \subset \Omega_{N+1} \\ |D|=1}} \sum_{\substack{D' \subset \Omega_{N+1} \\ |D'|=2}} T_{ND} T_{ND'} \;.$$

The second part has mean zero and variance

$$\begin{aligned}
&\frac{16}{(N+1)^2} \sum_{|D|=1} \sum_{|D'|=2} \sum_{|\tilde D|=1} \sum_{|\tilde D'|=2} E T_{ND} T_{ND'} T_{N\tilde D} T_{N\tilde D'} \\
&\quad \le \frac{16}{(N+1)^2} \binom{N+1}{2} (5N-1) \left(E|T_{N12}|^r\right)^{2/r} \left(E|T_{N1}|^{\frac{2r}{r-2}}\right)^{\frac{r-2}{r}} \\
&\quad \le 40\, C^{2/r} a_N^2 N^{-2} = o(N^{-2}) \ .
\end{aligned}$$

and can therefore be neglected. The first part equals

$$\begin{aligned}
2 \sum_{\substack{D\subset \Omega_{N+1} \\ |D|=1}} \ \sum_{\substack{D'\subset \Omega_{N+1} \\ |D'|=2 \\ D\subset D'}} T_{ND} T_{ND'} &= 2 \sum\sum_{1\le i<j\le N+1} T_{Nij}(T_{Ni}+T_{Nj}) \\
&= \sum_{i=1}^{N+1} K_{Ni} + \sum\sum_{1\le i<j\le N+1} M_{Nij} \ ,
\end{aligned}$$

where (cf. (2.7) and (2.9))

$$K_{Ni} = 2NE\left(T_{Nij}T_{Nj} \mid X_i\right) \ ,$$

and

$$M_{Nij} = 2T_{Nij}(T_{Ni}+T_{Nj}) - 2E\left(T_{Nij}T_{Nj} \mid X_i\right) - 2E\left(T_{Nij}T_{Ni} \mid X_j\right) \ .$$

As a result we obtain

$$P\left(\left|2 \sum_{i=1}^{N+1} V_{Ni}W_{Ni} - \sum_{i=1}^{N+1} K_{Ni} - \sum\sum_{1\le i<j\le N+1} M_{Nij}\right| \ge \varepsilon_N N^{-\frac12}\right) \le \delta_N N^{-\frac12} \tag{2.20}$$

for appropriate $\varepsilon_N \searrow 0, \delta_N \searrow 0$.

Finally we consider

$$\begin{aligned}
\sum_{i=1}^{N+1} V_{Ni}^2 &= \sum_{\substack{D,D'\subset\Omega_{N+1} \\ |D|=|D'|=1}} \sum \left[|D\cap D'| - \frac{1}{N+1}\right] T_{ND}T_{ND'} \\
&= \sum_{i=1}^{N+1} T_{Ni}^2 - \frac{1}{N+1}\left(\sum_{i=1}^{N+1} T_{Ni}\right)^2 \ .
\end{aligned}$$

We have, by (2.12) and a result of Dharmadhikari, Fabian & Jogdeo (1968),

$$(N+1)^{-\frac32} E\left|\sum_{i=1}^{N+1} T_{Ni}\right|^3 \le (N+1)^{-\frac32} C' N^{\frac32} E|T_{N1}|^3 \le C'' N^{-\frac32} \ ,$$

and hence

$$(2.21) \qquad P\left(\left| \sum_{i=1}^{N+1} V_{Ni}^2 - \sum_{i=1}^{N+1} T_{Ni}^2 \right| \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \delta_N N^{-\frac{1}{2}}$$

By (2.17), (2.18)–(2.21), and since

$$\sigma^2(T_N) = NET_{N1}^2 + \sum_{k=2}^{N} \binom{N}{k} ET_{N\Omega_k}^2 = (N+1)ET_{N1}^2 + \mathcal{O}(N^{-1}) \ ,$$

we obtain the desired result. □

3. Edgeworth expansions

The results of this section rely heavily on a result on Edgeworth expansions for U-statistics, obtained by Bickel, Götze & van Zwet (1986) . Let $X_1, \ldots, X_N$ be i.i.d. random variables assuming values in a measurable space $(\mathcal{X}, \mathcal{A})$ with common distribution P, and let $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a measurable function which is symmetric in its two arguments, with $Eh(X_1, X_2) = \theta$ and $Eh^2(X_1, X_2) < \infty$. Let

$$U_N = \binom{N}{2}^{-1} \sum\sum_{1 \leq i < j \leq N} h(X_i, X_j)$$

be a U-statistic of degree two. Define

$$g(x) = E\left(h(X_1, X_2) \,|\, X_1 = x\right) - \theta \ , \quad \psi(x, y) = h(x, y) - g(x) - g(y) - \theta$$

and

$$(3.1) \qquad \lambda_1 = \sigma_g^{-3} Eg^3(X_1) \ , \quad \lambda_2 = \sigma_g^{-3} Eg(X_1)g(X_2)\psi(X_1, X_2) \ , \quad \kappa_3 = \lambda_1 + 3\lambda_2 \ .$$

Then $\kappa_3 N^{-\frac{1}{2}}$ serves as an approximation to the third cumulant of $U_N/\sigma(U_N)$. The (one-term) Edgeworth expansion for the distribution function of $U_N/\sigma(U_N)$ is given by

$$(3.2) \qquad G_N(x) = \Phi(x) - \frac{\kappa_3}{6\sqrt{N}}(x^2 - 1)\phi(x) \ .$$

The validity of this expansion was first proved by Janssen (1978) and Callaert, Janssen & Veraverbeke (1980) . The best result to date has been obtained by Bickel, Götze & van Zwet (1986), who proved the following theorem.

Theorem 3.1. (Bickel, Götze & van Zwet (1986)) *Suppose that there exist real numbers $C > 0$, $p > 3$, $r > 2$, and a positive continuous function χ on $(0, \infty)$, such that*

$$(3.3) \qquad E|g(X_1)|^p \leq C \ ;$$

$$(3.4) \qquad E|\psi(X_1, X_2)|^r \leq C \ ;$$

and

$$\left|Ee^{itg(X_1)}\right| \leq 1 - \chi(t) < 1 \;, \qquad \forall\, t > 0 \;. \tag{3.5}$$

Then there exists a sequence $\varepsilon_N \searrow 0$*, depending only on* $p, C, r,$ *and* χ*, such that for* $N = 2, 3, \ldots$

$$\sup_{x \in \mathbb{R}} \left| P(U_N/\sigma(U_N) \leq x) - G_N(x) \right| \leq \varepsilon_N N^{-\frac{1}{2}} \;. \tag{3.6}$$

Bickel, Götze and van Zwet prove this theorem under a slightly weaker moment assumption on g, namely $E|g(X_1)|^3 \mathbf{I}_{\{|g(X_1)| \geq t\}} \to 0$ as $t \to \infty$. This is implied by (3.3), since for $t > 0$ and $p > 3$,

$$E|g(X_1)|^3 \mathbf{I}_{\{|g(X_1)| \geq t\}} \leq E|g(X_1)|^3 \left(\frac{|g(X_1)|}{t} \right)^{p-3} = \frac{E|g(X_1)|^p}{t^{p-3}} \;.$$

Note that like Theorem 2.2, the present theorem is formulated in such a way that the conclusion is valid uniformly for any class (h, P) for which the assumptions are satisfied for fixed $\delta_1, C, p, r,$ and χ. The uniformity is important when we want to apply this result to symmetric statistics where the functions g and ψ in the Hoeffding decomposition of the statistic depend on N. It allows us to consider sequences $\{g_N\}$, $\{\psi_N\}$ and $\{P_N\}$ as long as (3.3)–(3.5) are satisfied for fixed constants C, p and r, and for a fixed function χ for every N.

Assumption (3.5) ensures that the distribution of $g(X_1)$ is non-lattice. However, it is clear from the proof of Bickel, Götze and van Zwet that the behaviour of the characteristic function of $g = g_N$ is irrelevant for $t > \tau_N$, if $\tau_N \to \infty$. Thus, assumption (3.5) can actually be relaxed. Sufficient is the existence of a sequence $\tau_N \to \infty$ and a positive continuous function χ on $(0, \infty)$, such that

$$\left|Ee^{itg_N(X_1)}\right| \leq 1 - \chi(t) \;, \quad \forall\, t \in (0, \tau_N) \;, \quad \text{for all } N \;. \tag{3.7}$$

Of course the sequence $\{\varepsilon_N\}$ in Theorem 3.1 will then also depend on the sequence $\{\tau_N\}$. In particular in Theorem 3.3 we shall actually need the less restrictive assumption (3.7). This is because truncation of the linear term $g_N(X_1)$ (corresponding to $N^{\frac{1}{2}} T_{N1}$) destroys (3.5), while (3.7) is still fulfilled for the truncated $g_N(X_1)$ if it is fulfilled for the original $g_N(X_1)$ (cf. the discussion after Corollary 2.4).

Generalization of Theorem 3.1 to symmetric statistics is fairly straightforward. Let $T_N = t_N(X_1, \ldots, X_N; P)$, where $X_1, \ldots, X_N$ are i.i.d. with distribution P, be a random variable with $ET_N = 0$ and with Hoeffding decomposition

$$T_N = \sum_{D \subset \Omega_N} T_{ND} = \sum_{i=1}^{N} T_{Ni} + \sum\sum_{1 \leq i < j \leq N} T_{Nij} + \ldots \;.$$

Let σ_N^2 denote the variance of T_N, suppose that there exist constants c and C such that $0 < c \le \sigma_N^2 \le C < \infty$, and define, as in (3.1),

$$\lambda_1 = N^{\frac{3}{2}} \sigma_N^{-3} E T_{N1}^3 \ , \quad \lambda_2 = N^{\frac{5}{2}} \sigma_N^{-3} E T_{N1} T_{N2} T_{N12} \ , \quad \kappa_3 = \lambda_1 + 3\lambda_2 \ . \tag{3.8}$$

Let $G_N(x)$ be defined as in (3.2). If for some sequence $\delta_N \searrow 0$,

$$\sum_{k=3}^{N} \binom{N}{k} E T_{N\Omega_k}^2 \le \delta_N N^{-\frac{3}{2}} \ ,$$

then

$$P\left(\Big| \sum_{|D|\ge 3} T_{ND} \Big| \ge \delta_N^{\frac{1}{3}} N^{-\frac{1}{2}} \right) \le \frac{\delta_N N^{-\frac{3}{2}}}{\delta_N^{\frac{2}{3}} N^{-1}} \le \delta_N^{\frac{1}{3}} N^{-\frac{1}{2}} \ ,$$

so that this part can be neglected and Theorem 3.1 can be applied to the U-statistic $\sum_{|D|\le 2} T_{ND}$. As a result we obtain

Theorem 3.2. *Suppose that there exist real numbers $c > 0$, $C > 0$, $p > 3$, $r > 2$, sequences $\{\delta_N\}_{N=1}^{\infty}$ with $\delta_N \searrow 0$, $\{\tau_N\}$ with $\tau_N \to \infty$ and a positive continuous function χ on $(0, \infty)$, such that (1.2) and (1.3) are satisfied and*

$$E|N^{\frac{1}{2}} T_{N1}|^p \le C \ ; \tag{3.9}$$

$$E|N^{\frac{3}{2}} T_{N12}|^r \le C \ ; \tag{3.10}$$

$$\sum_{k=3}^{N} \binom{N}{k} E T_{N\Omega_k}^2 \le \delta_N N^{-\frac{3}{2}} \ ; \tag{3.11}$$

and

$$\left| E e^{itN^{\frac{1}{2}} T_{N1}} \right| \le 1 - \chi(t) < 1 \ , \qquad \forall\ t \in (0, \tau_N) \ , \text{ for } N = 1, 2, \ldots, \tag{3.12}$$

Then there exists a sequence $\varepsilon_N \searrow 0$, depending only on $c, C, p, r, \{\delta_N\}, \{\tau_N\}$ and χ, such that for $N = 2, 3, \ldots$

$$\sup_{x\in\mathbb{R}} \left| P(T_N/\sigma_N \le x) - G_N(x) \right| \le \varepsilon_N N^{-\frac{1}{2}} \ . \tag{3.13}$$

The result of Bickel, Götze and van Zwet is also the basis of the Edgeworth expansion of the Studentized T_N. Let S_N^2 be the jackknife estimator of variance, given by (1.6), and define, with λ_1 and λ_2 as in (3.8),

$$H_N(x) = \Phi(x) + \frac{\phi(x)}{6\sqrt{N}} \left[(2x^2+1)\lambda_1 + 3(x^2+1)\lambda_2 \right] \ . \tag{3.14}$$

Then $H_N(x)$ is the one-term Edgeworth expansion for the distribution function of T_N/S_N. If $T_N = \sqrt{N}(\overline{X} - EX)$, then T_N/S_N is well known as Student's t statistic. In that case the one-term Edgeworth expansion for H_N is given by (3.14) with $\lambda_2 = 0$ and $\lambda_1 = EX_1^3/\sigma^3$, where σ^2 is the variance of X_1. The validity of the one-term Edgeworth expansion for Student's t under minimal moment conditions (viz. $E|X_1|^3 < \infty$) was first proved by Chibisov (1980).

Theorem 3.3. *Suppose that there exist real numbers $c > 0$, $C > 0$, $p > 3$, $r > 2$, a sequence $\{\tau_N\}$ with $\tau_N \to \infty$ and a positive continuous function χ on $(0, \infty)$ such that (1.2) and (1.3) are satisfied and*

$$E|N^{\frac{1}{2}} T_{N1}|^p \leq C \; ; \tag{3.15}$$

$$E|N^{\frac{3}{2}} T_{N12}|^r \leq C \; ; \tag{3.16}$$

$$\begin{cases} \sum_{k=3}^{N} \binom{N-1}{k-1} ET^2_{N\Omega_k} \leq CN^{-3} \\ \sum_{k=3}^{N} \binom{N-2}{k-2} ET^2_{N\Omega_k} \leq CN^{-\frac{7}{2}} \; ; \end{cases} \tag{3.17}$$

and

$$\left| Ee^{itN^{\frac{1}{2}} T_{N1}} \right| \leq 1 - \chi(t) < 1 \; , \qquad \forall \; t \in (0, \tau_N) \; , \text{ for } N = 1, 2, \ldots, \tag{3.18}$$

Then there exists a sequence $\varepsilon_N \searrow 0$, depending only on $c, C, p, r, \{\tau_N\}$ and χ, such that for $N = 2, 3, \ldots$

$$\sup_{x \in \mathbb{R}} \left| P\left(T_N / S_N \leq x\right) - H_N(x) \right| \leq \varepsilon_N N^{-\frac{1}{2}} \; . \tag{3.19}$$

Proof. Recall from the explanation following Theorem 2.2 that all o and $\mathcal{O}$ symbols used are uniform, i.e. $o(b_N)$ and $\mathcal{O}(b_N)$ mean $\leq \varepsilon_N b_N$ and $\leq C b_N$ for a fixed $\varepsilon_N \searrow 0$ or a fixed C, uniformly in T_N and P satisfying the assumptions. Also recall that, in view of Corollary 2.4, we assume that $|T_{Ni}| \leq a_N = N^{-\frac{p-3}{4p}}$. Let V_N denote the U-statistic

$$V_N = \sum_{i=1}^{N+1} (K_{Ni} + L_{Ni}) + \sum\sum_{1 \leq i < j \leq N+1} M_{Nij} \; .$$

Under the assumptions of the present theorem, Theorem 2.2 asserts that

$$P\left(\left| S_N^2 - \sigma_N^2 - V_N \right| \geq \varepsilon_N N^{-\frac{1}{2}} \right) \leq \delta_N N^{-\frac{1}{2}} \tag{3.20}$$

for sequences $\varepsilon_N \searrow 0$ and $\delta_N \searrow 0$.

We begin by bounding the moments of the terms in V_N. We have

$$\begin{aligned} |K_{Ni}| &\leq 2N \left[E\left(|T_{Nij}|^{\frac{p}{p-1}} \mid X_i \right) \right]^{\frac{p-1}{p}} \left[E|T_{Nj}|^p \right]^{\frac{1}{p}} \\ &\leq 2C^{1/p} N^{\frac{1}{2}} \left[E\left(|T_{Nij}|^{\frac{p}{p-1}} \mid X_i \right) \right]^{\frac{p-1}{p}} \; , \qquad \text{for } i \neq j \; . \end{aligned}$$

Since $r(p-1)/p > 1$ for $r > 2$ and $p > 3$, this implies that

$$E|K_{N1}|^r \leq 2^r C^{r/p} N^{r/2} E|T_{N12}|^r \leq 2^r C^{1+r/p} N^{-r} \; , \tag{3.21}$$

and in view of an inequality of Dharmadhikari et al. (1968),

$$E\Big|\sum_{i=1}^{N+1} K_{Ni}\Big|^r \leq C' N^{-r/2} \ , \tag{3.22}$$

for an appropriate constant C'. Similarly,

$$E|L_{N1}|^{p/2} \leq 2^{p/2} E|T_{N1}|^p \leq 2^{p/2} C N^{-p/2} \ , \tag{3.23}$$

$$E|L_{N1}|^t \leq 2^t E|T_{N1}|^{2t} \leq 2^t C a_N^{2t-p} N^{-p/2} \ , \qquad \text{for } t > p/2 \ , \tag{3.24}$$

$$E\left(\sum_{i=1}^{N+1} L_{Ni}\right)^2 \leq C' a_N^{4-p} N^{-\frac{p-2}{2}} \ , \qquad \text{if } p < 4 \ , \tag{3.25}$$

$$E|M_{N12}|^r \leq 2^{3r} E|T_{N1}T_{N12}|^r \leq 2^{3r} C a_N^r N^{-\frac{3r}{2}} \ , \tag{3.26}$$

$$E\Big|\sum\sum_{1\leq i<j\leq N+1} M_{Nij}\Big|^r \leq C' a_N^r N^{-r/2} \ . \tag{3.27}$$

Now, first of all, (3.22), (3.25) and (3.27) imply that $EV_N^2 = o(N^{-\frac{1}{2}})$, so by (1.3) and (3.20) we find that for every $\varepsilon > 0$,

$$P\left(\left|\frac{S_N^2 - \sigma_N^2}{\sigma_N^2}\right| \geq \varepsilon\right) = o(N^{-\frac{1}{2}}) \ .$$

Since also

$$P\left(\Big|\sum_{|D|\geq 3} T_{ND}\Big| \geq \varepsilon_N N^{-\frac{1}{2}}\right) \leq \frac{\sum_{k=3}^N \binom{N}{k} ET^2_{N\Omega_k}}{\varepsilon_N^2 N^{-1}} = \mathcal{O}(\frac{1}{\varepsilon_N^2 N}) \ ,$$

we find that there exist $\varepsilon_N \searrow 0$, such that

$$P\left(\left|\frac{T_N}{S_N} - \frac{\sum_{|D|\leq 2} T_{ND}}{S_N}\right| \geq \varepsilon_N N^{-\frac{1}{2}}\right) = o(N^{-\frac{1}{2}}) \ ,$$

and

$$F_N(x) = P(\frac{T_N}{S_N} \leq x) \leq P\left(\frac{\sum_{|D|\leq 2} T_{ND}}{S_N} \leq x + o(N^{-\frac{1}{2}})\right) + o(N^{-\frac{1}{2}}) \ ,$$

$$F_N(x) = P(\frac{T_N}{S_N} \leq x) \geq P\left(\frac{\sum_{|D|\leq 2} T_{ND}}{S_N} \leq x - o(N^{-\frac{1}{2}})\right) + o(N^{-\frac{1}{2}}) \ .$$

We have to show that $\sup_x |F_N(x) - H_N(x)| = o(N^{-\frac{1}{2}})$, and as $H_N(x)$ has bounded derivative, this is equivalent to showing the same thing with F_N replaced by the distribution function of $\sum\limits_{|D|\leq 2} T_{ND}/S_N$. Hence T_N may be replaced by the U-statistic

$$U_N = \sum_{i=1}^N T_{Ni} + \sum\sum_{1\leq i<j\leq N} T_{Nij} \ ,$$

and $F_N(x)$ by $P(U_N/S_N \leq x)$.

Next we write

$$P\left(\frac{U_N}{S_N} \leq x\right) = P\left(\frac{U_N}{\sigma_N} \leq x\frac{S_N}{\sigma_N}\right) \ .$$

Because $xH'_N(x)$ is bounded, $H_N\left(x(1+o(N^{-\frac{1}{2}}))\right) = H_N(x) + o(N^{-\frac{1}{2}})$, and in view of (3.20) we may replace S_N^2 in $P\left(\frac{U_N}{\sigma_N} \leq x\frac{S_N}{\sigma_N}\right)$ by $\sigma_N^2 + V_N$. Therefore we may replace $F_N(x)$ by

$$P\left(\frac{U_N}{\sigma_N} \leq x\frac{(\sigma_N^2+V_N)^{\frac{1}{2}}}{\sigma_N}\right) = P\left(\frac{U_N}{\sigma_N} - x\left\{(1+\frac{V_N}{\sigma_N^2})^{\frac{1}{2}} - 1\right\} \leq x\right) \ .$$

For $|z| \leq \frac{4}{5}$ we have $1 + \frac{z}{2} - \frac{z^2}{4} \leq (1+z)^{\frac{1}{2}} \leq 1 + \frac{z}{2}$ and hence we have, with probability $1 - o(N^{-\frac{1}{2}})$,

$$\frac{V_N}{2\sigma_N^2} - \frac{V_N^2}{4\sigma_N^4} \leq \left(1+\frac{V_N}{\sigma_N^2}\right)^{\frac{1}{2}} - 1 \leq \frac{V_N}{2\sigma_N^2} \ .$$

As

$$V_N^2 \leq 3\left\{\left(\sum_{i=1}^{N+1} K_{Ni}\right)^2 + \left(\sum_{i=1}^{N+1} L_{Ni}\right)^2 + \left(\sum\sum_{1\leq i<j\leq N+1} M_{Nij}\right)^2\right\} \ ,$$

and

$$P\left(\left(\sum_{i=1}^{N+1} K_{Ni}\right)^2 + \left(\sum\sum_{1\leq i<j\leq N+1} M_{Nij}\right)^2 \geq \varepsilon_N N^{-\frac{1}{2}}\right) \leq \delta_N N^{-\frac{1}{2}}$$

for some $\varepsilon_N \searrow 0$ and $\delta_N \searrow 0$ by (3.22) and (3.27), we see that

$$\frac{V_N}{2\sigma_N^2} - 3\frac{(\sum_{i=1}^{N+1} L_{Ni})^2}{4\sigma_N^4} + o(N^{-\frac{1}{2}}) \leq \left(1+\frac{V_N}{\sigma_N^2}\right)^{\frac{1}{2}} - 1 \leq \frac{V_N}{2\sigma_N^2}$$

with probability $1 - o(N^{-\frac{1}{2}})$. By the same argument as before, it follows that we have to show that

$$\text{(3.28)} \qquad \sup_x \left| P\left(\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2} \leq x\right) - H_N(x)\right| = o(N^{-\frac{1}{2}}) \ ,$$

and

$$\text{(3.29)} \qquad \sup_x \left| P\left(\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2} + \frac{3x(\sum_{i=1}^{N+1} L_{Ni})^2}{4\sigma_N^4} \leq x\right) - H_N(x)\right| = o(N^{-\frac{1}{2}}) \ .$$

With (3.28) established, to prove (3.29) we need only check that under the assumptions of the present theorem the term $\frac{3x(\sum_{i=1}^{N+1} L_{Ni})^2}{4\sigma_N^4}$ does not influence the expansion. For the sake of brevity we omit this and refer to Putter (1994). Here we prove only (3.28). For this we start by noting that $\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2}$ is a U-statistic of degree 2 to which we can apply Theorem 3.1. Let us therefore check the assumptions (3.3), (3.4) and (3.7). First, because $a_N = N^{-\frac{p-3}{4p}}$ and $p > 3$, (3.26) implies

$$E|N^{\frac{3}{2}} x M_{N12}|^r \leq C' a_N^r (\log N)^r = o(1) , \qquad \text{if } |x| \leq \log N .$$

Because of (3.24), we find that for $|x| \leq \log N$,

$$E\left|N^{\frac{1}{2}} x L_{N1}\right|^p \leq 2^p C a_N^p (\log N)^p = o(1) .$$

At this point we apply Lemma 2.3 again to $K_{Ni} = 2NE(T_{Nij}T_{Nj} \mid X_i)$. By (3.21), $E|K_{N1}|^r \leq C' N^{-r}$. In Lemma 2.3 we now choose $s = r > 2$ and $\eta = N^{-\frac{1}{4} - \frac{r-2}{2r}}$. It follows that there exist i.i.d. $K'_{N1}, \ldots, K'_{N,N+1}$ with $K'_{Ni} = \varrho_N(K_{Ni})$ satisfying

(3.30) $\qquad |K'_{Ni}| \leq N^{-\frac{1}{4} - \frac{r-2}{2r}}$;

(3.31) $$P(K'_{Ni} = K_{Ni} , i = 1, \ldots, N+1) \geq 1 - 2C' \frac{N+1}{N} N^{-r/4} \geq$$
$$\geq 1 - 4C' N^{-r/4} = 1 - o(N^{-\frac{1}{2}}) ;$$

(3.32) $\qquad EK'_{Ni} = 0$;

(3.33) $$N^t E|K'_{Ni} - K_{Ni}|^t \leq 2^{2t+1} C N^{-\frac{(r+4)(r-t)}{4r}} = \begin{cases} o(N^{-\frac{3(r-t)}{2r}}) , & 0 < t < r \\ \mathcal{O}(1) , & t = r . \end{cases}$$

We shall treat the K_{Ni} as we did the T_{Ni}, i.e. we shall delete the prime and assume throughout the proof of this theorem that the original K_{Ni} satisfy (3.30), (3.32), and (3.33). The probability that this makes any difference is $o(N^{-\frac{1}{2}})$. The only place where the K_{Ni} occur in the conclusion of the proof is in $\lambda_2 = N^{\frac{5}{2}} ET_{N1}T_{N2}T_{N12} = N^{\frac{3}{2}} ET_{N1}K_{N1}$, and we have to check that

$$N^{\frac{3}{2}} \left| ET_{N1}K_{N1} - ET_{N1}K'_{N1} \right| \to 0 .$$

This is a consequence of (3.15) and (3.33), since

$$\begin{aligned} N^{\frac{3}{2}} \left| ET_{N1}K_{N1} - ET_{N1}K'_{N1} \right| &\leq N^{\frac{3}{2}} \left| ET_{N1}(K_{N1} - K'_{N1}) \right| \\ &\leq N^{\frac{3}{2}} \left(ET_{N1}^2 \right)^{\frac{1}{2}} \left(E(K_{N1} - K'_{N1})^2 \right)^{\frac{1}{2}} = o(N^{-\frac{3(r-2)}{4r}}) . \end{aligned}$$

Now we find that $E|N^{\frac{1}{2}} x K_{Ni}|^{p'} = o(1)$, for $|x| \leq \log N$ and any $p' \leq 3r$. It follows that for $|x| \leq \log N$, $\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2}$ satisfies assumptions (3.3) and (3.4) of Theorem 3.1.

To check assumption (3.7) for $\left(U_N - \frac{x}{2\sigma_N} V_N\right)$, we have to consider the characteristic function $\psi(t) = E\exp\{itN^{\frac{1}{2}}\tilde{T}_{N1}\}$, where $\tilde{T}_{N1} = T_{N1} - \frac{x}{2\sigma_N}(K_{N1} + L_{N1})$. Of course

$$|\psi(t)| \le 1 - 1/3\ E(N^{\frac{1}{2}}\tilde{T}_{N1})^2 t^2 \qquad \text{if} \quad |t| \le \frac{E(N^{\frac{1}{2}}\tilde{T}_{N1})^2}{E|N^{\frac{1}{2}}\tilde{T}_{N1}|^3} \ .$$

We have shown above that $E|N^{\frac{1}{2}}x(K_{N1} + L_{N1})|^p \to 0$ uniformly in $|x| \le \log N$ for some $p > 3$. Also $E(N^{\frac{1}{2}}T_{N1})^2$ and $E|N^{\frac{1}{2}}T_{N1}|^3$ are bounded away from zero and infinity by (1.2), (3.15) and (3.17). Hence there exist $t_0 > 0$ and N_0 such that

$$|\psi(t)| \le 1 - c/4\ t^2 \qquad \text{for } 0 < |t| \le t_0 \text{ and } N \ge N_0 \ .$$

Next we note that

$$|\psi(t)| \le 1 - \chi(t) + \frac{|t|}{2\sigma_N} E|N^{\frac{1}{2}}x(K_{N1} + L_{N1})| \ .$$

The expectation on the right tends to zero uniformly for $x \le \log N$. Since $\chi(t)/t$ is positive and continuous on $(0, \infty)$, we find that there exist $\tau_N' \to \infty$, $\tau_N' < \tau_N$, such that

$$|\psi(t)| \le 1 - 1/2\ \chi(t) \qquad \text{for } t_0 \le |t| \le \tau_N' \text{ and } N \ge N_1 \ .$$

It follows that assumption (3.7) is satisfied uniformly in $|x| \le \log N$. Hence, by Theorem 3.1,

$$\sup_{|x| \le \log N} \left| P\left(\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2} \le x\right) - \tilde{G}_N(x)\right| = o(N^{-\frac{1}{2}}) \ ,$$

uniformly in all T_N and P satisfying our assumptions. Here

$$\begin{aligned}
&\tilde{G}_N(x) = \Phi\left(\frac{x}{\sigma_x}\right) - \frac{\kappa_3}{6} N^{-\frac{1}{2}}(x^2 - 1)\phi(x) \ , \\
(3.34)\qquad &\sigma_x^2 = \sigma^2\left(\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2}\right) = 1 - \frac{Nx}{\sigma_N^3}\left(ET_{N1}^3 + 2NET_{N12}T_{N1}T_{N2}\right) + o(N^{-\frac{1}{2}}) \ , \\
&\kappa_3 = \sigma_N^{-3}\left(N^{\frac{3}{2}}ET_{N1}^3 + 3N^{\frac{5}{2}}ET_{N12}T_{N1}T_{N2}\right) \ ,
\end{aligned}$$

and it follows that $\tilde{G}_N(x)$ may be replaced by

$$\begin{aligned}
H_N(x) &= \Phi(x) + \frac{\phi(x)}{6\sqrt{N}}\left[\lambda_1(2x^2 + 1) + 3\lambda_2(x^2 + 1)\right] \ , \\
\lambda_1 &= N^{\frac{3}{2}}\sigma_N^{-3}ET_{N1}^3 \ , \qquad \lambda_2 = N^{\frac{5}{2}}\sigma_N^{-3}ET_{N12}T_{N1}T_{N2} \ .
\end{aligned}$$

For $x \le -\log N$, $H_N(x) = \mathcal{O}(N^{-c})$, and for $x \ge \log N$, we have that $1 - H_N(x) = \mathcal{O}(N^{-c})$ for every $c > 0$, so monotonicity of a distribution function implies (3.28),

$$\sup_x \left| P\left(\frac{U_N}{\sigma_N} - \frac{xV_N}{2\sigma_N^2} \le x\right) - H_N(x)\right| = o(N^{-\frac{1}{2}}) \ .$$

The theorem is proved. □

4. Empirical Edgeworth expansions

In this section we prove the consistency of jackknife estimates of the quantities λ_1 and λ_2 appearing in the Edgeworth expansions $G_N(x)$ and $H_N(x)$ of the distribution functions of $\frac{T_N}{\sigma_N}$ and $\frac{T_N}{S_N}$ respectively, thus validating the empirical Edgeworth expansion. The estimation of λ_1 is straightforward and very similar to the estimation of $\sigma^2(T_N)$. Recall from Section 3.1 that, with one additional observation X_{N+1} from P, $T_N^{(i)}$ and $\overline{T}_N$ are defined as in (1.4) and (1.5).

To estimate λ_2 we assume that we have two additional observations from P: X_{N+1} and X_{N+2}. Let T_N be as in (1.1). Define, for $1 \leq i < j \leq N+2$,

$$T_N^{(i,j)} = T_N^{(j,i)} = t_N(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{N+2}; P) , \tag{4.1}$$

with X_i and X_j replaced by X_{N+1} and X_{N+2}. Furthermore, let

$$\overline{T}_N^{(i)} = \frac{1}{N+1} \sum_{\substack{j=1 \\ j \neq i}}^{N+2} T_N^{(i,j)} , \tag{4.2}$$

$$\overline{\overline{T}}_N = \frac{1}{N+2} \sum_{i=1}^{N+2} \overline{T}_N^{(i)} = \binom{N+2}{2}^{-1} \sum\sum_{1 \leq i < j \leq N+2} T_N^{(i,j)} . \tag{4.3}$$

We propose the following jackknife estimates for λ_1 and λ_2.

$$\hat{\lambda}_1 = \sqrt{N} \sum_{i=1}^{N+1} \left(\overline{T}_N - T_N^{(i)} \right)^3 / S_N^3 \tag{4.4}$$

and

$$\hat{\lambda}_2 = 2\sqrt{N} \sum\sum_{1 \leq i < j \leq N+2} \left(\overline{\overline{T}}_N - \overline{T}_N^{(i)} - \overline{T}_N^{(j)} + T_N^{(i,j)} \right) \cdot \\ \cdot \left(\overline{T}_N - T_N^{(i)} \right) \left(\overline{T}_N - T_N^{(j)} \right) / S_N^3 . \tag{4.5}$$

For the proofs we again assume that $T_N = t_N(X_1, \ldots, X_N; P)$ is a symmetric random variable with $ET_N = 0$, $ET_N^2 < \infty$ and $0 < c \leq \sigma^2(T_N) \leq C < \infty$ for some positive numbers c and C.

Lemma 4.1. *Suppose that there exists constants $p > 3$, $c > 0$, and $C > 0$ such that (1.2) and (1.3) are satisfied, $E|N^{\frac{1}{2}} T_{N1}|^p \leq C$, and $\sum_{k=2}^{N} \binom{N-1}{k-1} ET_{N\Omega_k}^2 \leq CN^{-2}$. Then there exist sequences $\delta_N \searrow 0$ and $\varepsilon_N \searrow 0$, which depend only on p, c and C, such that for $N = 2, 3, \ldots$,*

$$P\left(N^{\frac{1}{2}} \left| \sum_{i=1}^{N+1} \left(\overline{T}_N - T_N^{(i)} \right)^3 - NET_{N1}^3 \right| \geq \varepsilon_N \right) \leq \delta_N .$$

Proof. Obviously all we have to prove is that for every $\varepsilon > 0$,

$$P\left(N^{\frac{1}{2}}\left|\sum_{i=1}^{N+1}\left(\overline{T}_N - T_N^{(i)}\right)^3 - NET_{N1}^3\right| \geq \varepsilon\right) \to 0$$

uniformly for fixed p, c and C. Write $\overline{T}_N - T_N^{(i)} = V_{Ni} + W_{Ni}$, with

$$V_{Ni} = \sum_{\substack{D\subset\Omega_{N+1}\\ |D|=1}}\left(\mathbf{1}_D(i) - \frac{|D|}{N+1}\right)T_{ND}$$

and

$$W_{Ni} = \sum_{\substack{D\subset\Omega_{N+1}\\ |D|\geq 2}}\left(\mathbf{1}_D(i) - \frac{|D|}{N+1}\right)T_{ND}\ .$$

V_{Ni} can be expressed as $T_{Ni} - \frac{1}{N+1}\sum_{i=1}^{N+1} T_{Ni} = T_{Ni} - \Delta_N$, and an inequality of Dharmadhikari et al. (1968) ensures that $E|\Delta_N|^3 = \mathcal{O}(N^{-3})$. Expansion of $\sum_{i=1}^{N+1}(\overline{T}_N - T_N^{(i)})^3$ yields

$$\sum_{i=1}^{N+1}(\overline{T}_N - T_N^{(i)})^3 = \sum_{i=1}^{N+1} V_{Ni}^3 + 3\sum_{i=1}^{N+1} V_{Ni}^2 W_{Ni} + 3\sum_{i=1}^{N+1} V_{Ni}W_{Ni}^2 + \sum_{i=1}^{N+1} W_{Ni}^3\ .$$

First we show that

$$N^{\frac{1}{2}}\left(\sum_{i=1}^{N+1} V_{Ni}^3 - \sum_{i=1}^{N+1} T_{Ni}^3\right) \xrightarrow{P} 0\ .$$

This follows easily from the fact that for $i = 1, \ldots, N+1$,

$$\begin{aligned} E|V_{Ni}^3 - T_{Ni}^3| &= E|-3T_{Ni}^2\Delta_N + 3T_{Ni}\Delta_N^2 - \Delta_N^3| \\ &\leq 3\left(E|T_{N1}|^3\right)^{\frac{2}{3}}\left(E|\Delta_N|^3\right)^{\frac{1}{3}} + 3\left(E|T_{N1}|^3\right)^{\frac{1}{3}}\left(E|\Delta_N|^3\right)^{\frac{2}{3}} + E|\Delta_N|^3 = \mathcal{O}(N^{-2})\ . \end{aligned}$$

Since $E|N^{\frac{1}{2}}T_{N1}|^3 \leq C^{3/p}$, it also follows that $E|N^{\frac{1}{2}}V_{N1}|^3 \leq C_1$ for some $C_1 > 0$ and all N.

Next, define

$$R_{N1} = N^{\frac{1}{2}}\sum_{i=1}^{N+1} W_{Ni}^3\ , \qquad R_{N2} = N^{\frac{1}{2}}\sum_{i=1}^{N+1} V_{Ni}W_{Ni}^2\ , \qquad R_{N3} = N^{\frac{1}{2}}\sum_{i=1}^{N+1} V_{Ni}^2 W_{Ni}\ .$$

Since $E|N^{\frac{1}{2}}V_{N1}|^3 \leq C_1$ and $E(NW_{N1})^2 \leq N^2\sum_{k=2}^{N}\binom{N-1}{k-1}ET_{N\Omega_k}^2 \leq C$, two applications of Lemma 3.2.3 yield that for every $\eta_N \to \infty$ there exist V'_{Ni} and W'_{Ni} for $i = 1, \ldots, N+1$, such that

$$\begin{aligned} &|V'_{Ni}| \leq \eta_N N^{-\frac{1}{6}}\ , \qquad |W'_{Ni}| \leq \eta_N N^{-\frac{1}{2}}\ , \\ &P\left(V'_{Ni} = V_{Ni}\ , i = 1, \ldots, N+1\right) = 1 - o(1)\ , \end{aligned}$$

and

$$P\left(W'_{Ni} = W_{Ni}\ , i = 1, \ldots, N+1\right) = 1 - o(1)\ ,$$

Hence we may replace V_{Ni} by V'_{Ni} and W_{Ni} by W'_{Ni} in R_{N1}, R_{N2} and R_{N3} and for these remainder terms to go to zero in probability, it suffices to show that $E|R'_{Nj}| \to 0$, for $j = 1, 2, 3$, with

$$R'_{N1} = N^{\frac{1}{2}} \sum_{i=1}^{N+1} W'_{Ni} W^2_{Ni}\ , \quad R'_{N2} = N^{\frac{1}{2}} \sum_{i=1}^{N+1} V'_{Ni} W^2_{Ni}\ , \quad R'_{N3} = N^{\frac{1}{2}} \sum_{i=1}^{N+1} V'_{Ni} V_{Ni} W_{Ni}\ .$$

This is easy, since for $\eta_N = o(N^{1/6})$,

$$\begin{aligned} E|R'_{N1}| &\leq \eta_N N E W^2_{N1} \to 0\ , \\ E|R'_{N2}| &\leq \eta_N N^{\frac{4}{3}} E W^2_{N1} \to 0\ , \\ E|R'_{N3}| &\leq \eta_N N^{\frac{4}{3}} \left(E V^2_{N1}\right)^{\frac{1}{2}} \left(E W^2_{N1}\right)^{\frac{1}{2}} \to 0\ . \end{aligned}$$

Finally, using an inequality of von Bahr & Esséen (1965), we have for every $\varepsilon > 0$, as N tends to infinity,

$$P\left(\Big|N^{\frac{1}{2}} \sum_{i=1}^{N+2} (T^3_{Ni} - ET^3_{N1})\Big| > \varepsilon\right) \leq \frac{C'(N+1)E|T_{N1}|^p}{(\varepsilon N^{-\frac{1}{2}})^{p/3}} \to 0\ .$$

Since the uniformity of the convergence of the various terms is easily checked, the lemma is proved. $\square$

Lemma 4.2. *Suppose that there exist constants $c > 0$ and $C > 0$, such that (1.2) and (1.3) are satisfied, $E|N^{\frac{1}{2}} T_{N1}|^3 \leq C$, $E(N^{\frac{3}{2}} T_{N12})^2 \leq C$, and $\sum_{k=3}^{N} \binom{N-2}{k-2} E T^2_{N\Omega_k} \leq C N^{-4}$. Then there exist sequences $\delta_N \searrow 0$ and $\varepsilon_N \searrow 0$, which depend only on c and C, such that for $N = 2, 3, \ldots$,*

$$P\Big(N^{\frac{1}{2}}\Big| \sum\sum_{1 \leq i < j \leq N+2} \Big\{\big(\overline{\overline{T}}_N - \overline{T}^{(i)}_N - \overline{T}^{(j)}_N + T^{(i,j)}_N\big)\big(\overline{T}_N - T^{(i)}_N\big)\big(\overline{T}_N - T^{(j)}_N\big) - E T_{N1} T_{N2} T_{N12}\Big\}\Big| \geq \varepsilon_N\Big) \leq \delta_N\ .$$

Proof. Write

$$\begin{aligned} \overline{\overline{T}}_N - \overline{T}^{(i)}_N - \overline{T}^{(j)}_N + T^{(i,j)}_N &= T_{Nij} - \frac{1}{N+1}(T_{Ni} + T_{Nj}) + \Delta_{Nij}\ , \\ \overline{T}_N - T^{(i)}_N &= T_{Ni} + Z_{Ni}\ . \end{aligned}$$

Then

$$\begin{aligned} &\left(\overline{\overline{T}}_N - \overline{T}^{(i)}_N - \overline{T}^{(j)}_N + T^{(i,j)}_N\right) \left(\overline{T}_N - T^{(i)}_N\right) \left(\overline{T}_N - T^{(i)}_N\right) \\ &\quad = \big(T_{Nij} - \frac{1}{N+1}(T_{Ni} + T_{Nj})\big)(T_{Ni} + Z_{Ni})(T_{Nj} + Z_{Nj}) \\ &\quad\quad + \Delta_{Nij}(T_{Ni} + Z_{Ni})(T_{Nj} + Z_{Nj})\ . \end{aligned}$$

Straightforward calculations show (cf. Putter (1994)) that under the conditions of the present lemma, $E\Delta_{Nij}^2 = \mathcal{O}(N^{-4})$. It also follows from the proof of Lemma 4.1 that $EZ_{Ni}^2 = \mathcal{O}(N^{-2})$.

First we shall show that

$$N^{\frac{1}{2}} \sum_{i<j}\sum \Delta_{Nij}(T_{Ni} + Z_{Ni})(T_{Nj} + Z_{Nj})$$
$$= N^{\frac{1}{2}} \sum_{i<j}\sum \Delta_{Nij}\left(T_{Ni}T_{Nj} + T_{Ni}Z_{Nj} + T_{Nj}Z_{Ni} + Z_{Ni}Z_{Nj}\right)$$

tends to zero in probability. Applying Lemma 2.3 we see that there exists T'_{Ni} such that $|T'_{Ni}| \leq N^{-\frac{1}{8}}$ and

$$P\left(T'_{Ni} = T_{Ni}\ , \quad \text{for } i = 1, \ldots, N+2\right) = o(1)\ ,$$

so it suffices to show that

$$E\Big|N^{\frac{1}{2}} \sum_{i<j}\sum \Delta_{Nij}\left(T'_{Ni}T_{Nj} + T'_{Ni}Z_{Nj} + T'_{Nj}Z_{Ni} + Z_{Ni}Z_{Nj}\right)\Big| \to 0\ .$$

This is true, as the above expression is less than or equal to

$$\tfrac{1}{2}N^{\frac{5}{2}}N^{-\frac{1}{8}}E\Big|\Delta_{N12}(T_{N1} + Z_{N1} + Z_{N2})\Big| + \tfrac{1}{2}N^{\frac{5}{2}}E|\Delta_{N12}Z_{N1}Z_{N2}|$$
$$\leq N^{\frac{19}{8}}\left(E\Delta_{N12}^2\right)^{\frac{1}{2}}\left\{(ET_{N1}^2)^{\frac{1}{2}} + (EZ_{N1}^2)^{\frac{1}{2}}\right\}$$
$$+ \tfrac{1}{2}N^{\frac{5}{2}}\left(E\Delta_{N12}^2\right)^{\frac{1}{2}} EZ_{N1}^2 = \mathcal{O}\left(N^{-1/8} + N^{-3/2}\right)\ .$$

Next, write

$$N^{\frac{1}{2}} \sum_{i<j}\sum \Big(T_{Nij} - \frac{1}{N+1}(T_{Ni} + T_{Nj})\Big)(T_{Ni} + Z_{Ni})(T_{Nj} + Z_{Nj})$$
$$= N^{\frac{1}{2}}\binom{N+2}{2}ET_{N1}T_{N2}T_{N12} + N^{\frac{1}{2}}(N+1)\sum_{i=1}^{N+2} L_{Ni} + N^{\frac{1}{2}} \sum_{1\leq i<j\leq N+2}\sum M_{Nij} + R_N\ ,$$

with

$$R_N = N^{\frac{1}{2}} \sum_{i<j}\sum \Big(T_{Nij} - \frac{1}{N+1}(T_{Ni} + T_{Nj})\Big)(T_{Ni}Z_{Nj} + T_{Nj}Z_{Ni})$$
$$+ N^{\frac{1}{2}} \sum_{i<j}\sum \Big(T_{Nij} - \frac{1}{N+1}(T_{Ni} + T_{Nj})\Big)Z_{Ni}Z_{Nj}\ ,$$

$$L_{Ni} = E\left(T_{Ni}T_{Nj}T_{Nij} \mid X_i\right) - \frac{1}{N+1}E\left(T_{Ni}T_{Nj}(T_{Ni} + T_{Nj}) \mid X_i\right) - ET_{N1}T_{N2}T_{N12}\ ,$$

and

$$M_{Nij} = T_{Ni}T_{Nj}T_{Nij} - \frac{1}{N+1}T_{Ni}T_{Nj}(T_{Ni}+T_{Nj}) - L_{Ni} - L_{Nj} - ET_{N1}T_{N2}T_{N12} \ .$$

Similar arguments as before show that $R_N \xrightarrow{P} 0$. Also note that under the moment assumptions of the present lemma,

$$E|T_{N1}T_{N2}T_{N12}|^{6/5} \le \left(E|T_{N1}T_{N2}|^3\right)^{2/5} \left(ET_{N12}^2\right)^{3/5} \le C^{7/5}N^{-3} \ ,$$
$$E\left|\frac{1}{N+1}T_{N1}^2T_{N2}\right|^{6/5} \le N^{-6/5}\left(E|T_{N1}|^3\right)^{6/5} \le C^{6/5}N^{-3} \ .$$

Hence we have $E|L_{N1}|^{6/5} \le C'N^{-3}$ and $E|M_{N12}|^{6/5} \le C'N^{-3}$ for an appropriate C' and it follows that

$$P\left(\left|N^{\frac{1}{2}}(N+1)\sum_{i=1}^{N+2} L_{Ni}\right| > \varepsilon\right) \le \frac{C''(N+2)E|L_{N1}|^{6/5}}{(\varepsilon N^{-\frac{1}{2}}(N+1)^{-1})^{6/5}} \to 0 \ ,$$
$$P\left(\left|N^{\frac{1}{2}}\sum\sum_{1\le i<j\le N+2} M_{Nij}\right| > \varepsilon\right) \le \frac{C''\binom{N+2}{2}E|M_{N12}|^{6/5}}{(\varepsilon N^{-\frac{1}{2}})^{6/5}} \to 0 \ ,$$

by the inequality of von Bahr & Esséen (1965). One easily checks the uniformity of the various convergences and the lemma is proved. □

The consistency of the estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ follows from Lemma's 4.1 and 4.2, together with Theorem 2.2 and Slutsky's lemma. By substituting the estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ for λ_1 and λ_2 in the Edgeworth expansions G_N and H_N, defined in (1.23) and (1.24), we obtain the empirical Edgeworth expansions

$$\hat{G}_N(x) = \Phi(x) - \frac{\hat{\lambda}_1 + 3\hat{\lambda}_2}{6\sqrt{N}}(x^2-1)\phi(x) \ , \tag{4.6}$$

and

$$\hat{H}_N(x) = \Phi(x) + \frac{\phi(x)}{6\sqrt{N}}\left((2x^2+1)\hat{\lambda}_1 + 3(x^2+1)\hat{\lambda}_2\right) \ . \tag{4.7}$$

Combining everything we have obtained so far leads to

Theorem 4.3. *Let $X_1,\dots,X_N$ be an* i.i.d. *sample from P, let T_N be a symmetric random variable, defined as in (1.1). Let the jackknife estimator of variance S_N^2 be defined as in (1.6), $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as in (4.4) and (4.5), and $\hat{G}_N$ and $\hat{H}_N$ as in (4.6) and (4.7). Suppose that there exist real numbers $c>0$, $C>0$, $p>3, r>2$, a positive continuous function χ on $(0,\infty)$ and a sequence $\tau_N \to \infty$, such that (1.2) and (1.3) are satisfied and*

$$E|N^{\frac{1}{2}}T_{N1}|^p \le C \ ; \tag{4.8}$$

$$E|N^{\frac{3}{2}}T_{N12}|^r \le C \ ; \tag{4.9}$$

$$\sum_{k=3}^{N}\binom{N-2}{k-2}ET_{N\Omega_k}^2 \le CN^{-4} \ ; \tag{4.10}$$

and

$$\left|Ee^{itN^{\frac{1}{2}}T_{N1}}\right| \le 1 - \chi(t) < 1 \ , \qquad \text{for } N = 1, 2, \ldots, \text{ and for all } t \in (0, \tau_N) \ . \tag{4.11}$$

Then there exist sequences $\delta_N \searrow 0$ and $\varepsilon_N \searrow 0$, which depend only on c, C, p, r, χ and the sequence $\{\tau_N\}$, such that for $N = 2, 3, \ldots,$

$$P\left(\sup_{x\in\mathbb{R}}\left|P(T_N/\sigma_N \le x) - \hat{G}_N(x)\right| \ge \varepsilon_N N^{-\frac{1}{2}}\right) \le \delta_N \ , \tag{4.12}$$

$$P\left(\sup_{x\in\mathbb{R}}\left|P(T_N/S_N \le x) - \hat{H}_N(x)\right| \ge \varepsilon_N N^{-\frac{1}{2}}\right) \le \delta_N \ . \tag{4.13}$$

5. Applications

We shall consider three important applications of Theorem 4.3; U-statistics, L-statistics and smooth functions of the sample mean. Berry-Esseen bounds for the first two of these statistics were established in van Zwet (1984) and in the second example we have made use of the results in that paper. Recall that the jackknife estimator of variance S_N^2 is defined as in (1.6), $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as in (4.4) and (4.5), and $\hat{G}_N$ and $\hat{H}_N$ as in (4.6) and (4.7).

Application 1 U-statistics.

Let $X_1, \ldots, X_N$ be i.i.d. random variables assuming values in a measurable space $(\mathcal{X}, \mathcal{A})$ with common distribution P, and let $h : \mathcal{X}^m \to \mathbb{R}$ be a measurable function which is symmetric in its arguments, with

$$Eh(X_1, \ldots, X_m) = \theta \qquad \text{and} \qquad Eh^2(X_1, \ldots, X_m) < \infty \ .$$

Let

$$U_N = \binom{N}{m}^{-1} \sum_{1\le i_1<\ldots<i_m\le N}\cdots\sum h(X_{i_1}, \ldots, X_{i_m})$$

be a U-statistic of degree m and put

$$T_N = \sqrt{N}(U_N - \theta) \ .$$

Define

$$g(x) = E\left(h(X_1, \ldots, X_m) \,|\, X_1 = x\right) - \theta \ , \tag{5.1}$$

$$\psi(x, y) = E\left(h(X_1, \ldots, X_m) \,|\, X_1 = x, X_2 = x\right) - g(x) - g(y) - \theta \ . \tag{5.2}$$

Theorem 4.3 implies

Corollary 5.1. *Suppose that the distribution of $g(X_1)$ is non-lattice and suppose that there exist $p>3$ and $r>2$ such that $E|g(X_1)|^p<\infty$ and $E|\psi(X_1,X_2)|^r<\infty$. Then*

$$\sqrt{N}\sup_{x\in\mathbb{R}}\left|P\left(T_N/\sigma_N\le x\right)-\hat{G}_N(x)\right|\xrightarrow{P}0\ , \tag{5.3}$$

$$\sqrt{N}\sup_{x\in\mathbb{R}}\left|P\left(T_N/S_N\le x\right)-\hat{H}_N(x)\right|\xrightarrow{P}0\ , \tag{5.4}$$

Proof. First note that the non-lattice assumption also implies that $Eg^2(X_1)>0$. From (1.12) it follows that $T_{ND}=0$ a.s., if $|D|\ge m+1$. For $k=0,1,\ldots,m$, define

$$g_k(X_1,\ldots,X_k)=(h(X_1,\ldots,X_k)-\theta)_{\Omega_k}=\sum_{A\subset\Omega_k}(-1)^{k-|A|}E\left(h(X_1,\ldots,X_m)-\theta\,|\,A\right)\ .$$

In particular, $g_0=0$, $g_1=g$, and $g_2=\psi$ as in (5.1) and (5.2). It follows from (1.15) that

$$\sigma_h^2=E\left(h(X_1,\ldots,X_m)-\theta\right)^2=\sum_{k=0}^{m}\binom{m}{k}Eg_k^2(X_1,\ldots,X_k)\ .$$

Note that, again by (1.12),

$$(h(X_{i1},\ldots,X_{i_m}))_{\Omega_k}=\begin{cases}g_k(X_1,\ldots,X_k)\ , & \text{if } \Omega_k\subset\{i_1,\ldots,i_m\}\ ,\\ 0\ , & \text{otherwise}\ .\end{cases}$$

Hence, for $k=0,1,\ldots,m$,

$$T_{N\Omega_k}=\sqrt{N}\binom{N}{m}^{-1}\binom{N-k}{m-k}g_k(X_1,\ldots,X_k)\ .$$

For $k=0$ this yields zero, as it should. For $k=1$ we get

$$T_{N1}=\frac{m}{\sqrt{N}}g(X_1)\ ,$$

and for $k=2$

$$T_{N12}=\frac{m(m-1)}{\sqrt{N}(N-1)}\psi(X_1,X_2)\ .$$

It follows that assumptions (4.8) and (4.9) are satisfied. The fact that $g(X_1)$ is non-lattice ensures the validity of (4.11). Finally, (4.10) holds because

$$\begin{aligned}\sum_{k=3}^{N}\binom{N-2}{k-2}ET^2_{N\Omega_k}&=N\sum_{k=3}^{m}\binom{N-2}{k-2}\frac{\binom{N-k}{m-k}^2}{\binom{N}{m}^2}Eg_k^2(X_1,\ldots,X_k)\\&=\binom{N}{m}^{-1}N\sum_{k=3}^{m}\frac{k(k-1)}{N(N-1)}\binom{N-k}{m-k}\binom{m}{k}Eg_k^2(X_1,\ldots,X_k)\\&\le\binom{N}{m}^{-1}N\frac{m(m-1)}{N(N-1)}\binom{N-3}{m-3}\sigma_h^2\\&=\frac{m^2(m-1)^2(m-2)}{N(N-1)^2(N-2)}\le C'N^{-4}\ .\end{aligned}$$

The proof is complete. $\square$

Application 2 L-statistics.

Let $X_1, \ldots, X_N$ be i.i.d. random variables with common distribution function F. Let $c_1, \ldots, c_N$ be a sequence of real numbers, let $X_{1:N}, \ldots, X_{N:N}$ denote the order statistics of $X_1, \ldots, X_N$, and define the L-statistic

$$L_N = N^{-1} \sum_{i=1}^{N} c_i X_{i:N} \ .$$

Suppose that L_N is an unbiased estimate of θ and define

$$T_N = \sqrt{N}\,(L_N - \theta) \ .$$

Suppose that there exist real numbers a, b and c such that

$$\max_{1\le i\le N} |c_i| \le a \ , \qquad N \max_{2\le i\le N} |c_i - c_{i-1}| \le b \ , \qquad N^2 \max_{3\le i\le N} |c_i - 2c_{i-1} + c_{i-2}| \le c. \tag{5.5}$$

This corresponds to the case of smooth weights. Assumption (5.5) is fulfilled if there exists a function $J : (0,1) \to \mathbb{R}$ with bounded second derivative such that $c_i = J\left(\frac{i}{N+1}\right)$. Theorem 4.3 implies

Corollary 5.2. *Suppose that (5.5) is satisfied, $E|X_1|^p < \infty$ for some $p > 3$, $\sigma^2(T_N) \ge c'$ for some $c' > 0$ and all N and $E(N^{\frac{1}{2}} T_N \mid X_1)$ satisfies (4.11). Then (5.3) and (5.4) hold.*

To prove corollary 5.2 we start by deriving a representation for the terms in the Hoeffding decomposition of T_N which is of interest in its own right. Define i.i.d. random variables $U_1, \ldots, U_N$ with a uniform distribution on $(0,1)$, and take $X_i = F^{-1}(U_i)$, where $F^{-1}(t) = \inf\{x : F(x) \ge t\}$ denotes the left-continuous version of the inverse of F. We let $U_{1:N}, \ldots, U_{N:N}$ denote the order statistics of $U_1, \ldots, U_N$. Then we have the relation

$$X_{j+1:N} - X_{j:N} = \sum_{\substack{A\subset\Omega_N \\ |A|=j}} \int_0^1 \prod_{i\in A} \left(\mathbf{1}_{[U_i,1)}(t)\right) \cdot \prod_{i\in A^c} \left(\mathbf{1}_{(0,U_i)}(t)\right) dF^{-1}(t) \ .$$

To see why this relation holds, note that the integrand is zero for a fixed t unless t is between the largest of the U_i's with i in A and the smallest of the U_i's with i not in A, and this can only occur if the U_i's with i in A happen to be the j smallest among $U_1, \ldots, U_N$. For the only A for which the integrand is not identically equal to zero, the integral yields $\int_{[U_{j:N}, U_{j+1:N})} dF^{-1}(t) = X_{j+1:N} - X_{j:N}$. Now we obtain

$$\begin{aligned}
[X_{j+1:N} - X_{j:N}]_D = \sum_{\substack{A\subset\Omega_N \\ |A|=j}} \int_0^1 &\prod_{i\in A\cap D} \left(\mathbf{1}_{[U_i,1)}(t) - t\right) t^{|A\cap D^c|} \\
&\cdot \prod_{i\in A^c\cap D} \left(\mathbf{1}_{(0,U_i)}(t) - (1-t)\right) (1-t)^{|A^c\cap D^c|} dF^{-1}(t) \ = \qquad (5.6)\\
= \sum_{\substack{A\subset\Omega_N \\ |A|=j}} \int_0^1 &\prod_{i\in D} \left(\mathbf{1}_{[U_i,1)}(t) - t\right) (-1)^{|A^c\cap D|}\ t^{|A\cap D^c|}\ (1-t)^{|A^c\cap D^c|} dF^{-1}(t).
\end{aligned}$$

Next, write

$$\begin{aligned}\sum_{i=1}^{N} c_i X_{i:N} &= \overline{c}\sum_{i=1}^{N} X_i + \sum_{i=1}^{N}(c_i-\overline{c})X_{i:N}\\ &= \sum_{j=1}^{N-1} a_j(X_{j+1:N}-X_{j:N}) + \overline{c}\sum_{i=1}^{N} X_i \ ,\end{aligned} \tag{5.7}$$

where $a_j = -\sum_{i=1}^{j}(c_i-\overline{c})$, for $j=1,\ldots,N-1$, and 0 otherwise. Write $|A\cap D^c| = l$ and $|A\cap D| = m$, with $j=l+m$. For the first term on the right in (5.7), (5.6) yields

$$\begin{aligned}\left[\sum_{j=1}^{N-1} a_j(X_{j+1:N}-X_{j:N})\right]_D &=\\ &= \int_0^1 \prod_{i\in D}\left(\mathbf{1}_{[U_i,1)}(t)-t\right)\cdot \sum_{l=0}^{N-|D|}\binom{N-|D|}{l} t^l(1-t)^{N-|D|-l}\\ &\qquad \cdot \sum_{m=0}^{|D|}\binom{|D|}{m}(-1)^{|D|-m} a_{l+m}\, dF^{-1}(t)\\ &= \int_0^1 \prod_{i\in D}\left(\mathbf{1}_{[U_i,1)}(t)-t\right)\cdot \sum_{l=0}^{N-|D|}\mathcal{B}(l;N-|D|,t)\,\Delta^{|D|}(a_l)\, dF^{-1}(t) \ ,\end{aligned} \tag{5.8}$$

where $\mathcal{B}(l;N-|D|,t)$ denotes the probability that a binomial random variable with parameters $N-|D|$ and t equals l and $\Delta^{|D|}(a_l)$ is the $|D|^{th}$ difference of a_l, defined recursively by

$$\Delta(a_l) = a_{l+1}-a_l \ , \qquad \Delta^{\nu}(a_l) = \Delta\left(\Delta^{\nu-1}(a_l)\right) \ .$$

Taking $D=\{i\}$ and $D=\{i,j\}$ and using (5.7) we find that

$$T_{Ni} = -N^{-\frac{1}{2}}\sum_{l=1}^{N} c_l \int_0^1 \left(\mathbf{1}_{[U_i,1)}(t)-t\right)\binom{N-1}{l-1} t^{l-1}(1-t)^{N-l} dF^{-1}(t) \ , \tag{5.9}$$

$$\begin{aligned}T_{Nij} = -N^{-\frac{1}{2}}\sum_{l=2}^{N}(c_l-c_{l-1}) \int_0^1 &\left(\mathbf{1}_{[U_i,1)}(t)-t\right)\left(\mathbf{1}_{[U_j,1)}(t)-t\right)\\ &\cdot\binom{N-2}{l-2} t^{l-2}(1-t)^{N-l} dF^{-1}(t) \ .\end{aligned} \tag{5.10}$$

By (5.5) and using $\sum_{l=1}^{N}\mathcal{B}(l-1;N-1,t)=1$, we find after straightforward calculations,

$$E|T_{N1}|^p \le a^p\, 2^{p-1}\, E|X_1|^p\, N^{-\frac{1}{2}p} \tag{5.11}$$

$$E|T_{N12}|^r \le b^r\, 2^{2r}\, E|X_1|^r N^{-\frac{3}{2}r} \tag{5.12}$$

In Putter (1994) it is shown that

$$\sum_{k=3}^{N} \binom{N-2}{k-2} ET^2_{N\Omega_k} \le 45c^2 N^{-4} EX_1^2 \ . \tag{5.13}$$

Since by (5.11), (5.12) and (5.13), $\sigma^2(T_N)$ is bounded, application of Theorem 4.3 completes the proof of the corollary. □

Application 3 Smooth functions of the sample mean.

Let $X_1, \ldots, X_N$ be i.i.d. random variables taking values in a real Banach space $\mathbf{B}$. Let H be a map from $\mathbf{B}$ to $\mathbb{R}$ and define

$$T_N = \sqrt{N}\left(H(\overline{X}_N) - EH(\overline{X}_N)\right) \ .$$

Let $H^{(s)}(x)$ denote the s-th Fréchet derivative of H at the point $x \in \mathbf{B}$, and define

$$\|H^{(s)}\|_\infty = \sup_{x\in\mathbf{B}} \|H^{(s)}(x)\|.$$

Since $\sum_{k=3}^{N} \binom{N-3}{k-3} ET^2_{N\Omega_k} \le CN^{-5}$ implies $\sum_{k=3}^{N} \binom{N-2}{k-2} ET^2_{N\Omega_k} \le CN^{-4}$, the following is a direct consequence of the results in Bentkus, Götze & van Zwet (1994) and Theorem 4.3.

Corollary 5.3. *Suppose that the distribution of X_1 is non-lattice and suppose that $E|X_1|^p < \infty$ for some $p > 3$. Suppose furthermore that $\|H''\|_\infty < \infty$ and*

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

Then (5.3) and (5.4) hold.

6. The bootstrap

The results in Sections 2, 3 and 4 have been formulated in such a way that the conclusions hold uniformly for all T_N and P satisfying the assumptions of the theorems for fixed constants and sequences (cf. the discussion following the formulations of Theorems 2.2 and 3.1). This allows an almost immediate application to the bootstrap. For every N, we take as our underlying distribution the empirical distribution P_N based on the observed sample $X_1, \ldots, X_N$. Then we need to check the moment assumptions of the various terms in the Hoeffding decomposition under P_N. When the structure of T_N is not too complicated, it is possible to relate these moment assumptions under P_N to moment assumptions of the corresponding terms in the Hoeffding decomposition of T_N under P. We shall illustrate this for a U-statistic of degree two. This case has been studied earlier by Helmers (1991).

Let $X_1, \ldots, X_N$ be a sequence of i.i.d. random variables with common distribution P, let $h : \mathbb{R}^2 \to \mathbb{R}$ be a symmetric kernel with $Eh(X_1, X_2) = \theta$ and $Eh^2(X_1, X_2) < \infty$, and define the U-statistic

$$U_N = \binom{N}{2}^{-1} \sum\sum_{1 \le i < j \le N} h(X_i, X_j) \ .$$

Define T_N, the functions g and ψ and Hoeffding's decomposition of T_N as in (1.26)–(1.27). Let S_N^2 be the jackknife estimator of variance of T_N and define $F_N(x) = P(T_N/S_N \le x)$. To define the bootstrap approximation to F_N, let $X_1^*, \ldots, X_N^*$ be an i.i.d. sample from the empirical distribution P_N and define

$$\begin{aligned} U_N^* &= \binom{N}{2}^{-1} \sum\sum_{1 \le i < j \le N} h(X_i^*, X_j^*) \\ \theta_N^* &= E\left(U_N^* \,|\, X_1, \ldots, X_N\right) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N h(X_i, X_j) \ , \\ T_N^* &= \sqrt{N}\,(U_N^* - \theta_N^*) \ , \ F_N^*(x) = P(T_N^*/S_N^* \le x \,|\, X_1, \ldots, X_N) \ , \end{aligned}$$

where ${S_N^*}^2$ is given by

$$\begin{aligned} U_N^{(i)*} &= \binom{N}{2}^{-1} \sum\sum_{\substack{1 \le j < k \le N+1 \\ j,k \ne i}} h(X_j^*, X_k^*) \ , \quad T_N^{(i)*} = \sqrt{N}\left(U_N^{(i)*} - \theta_N^*\right) \ , \\ \overline{T}_N^* &= \frac{1}{N+1} \sum_{i=1}^{N+1} T_N^{(i)^*} \ , \quad {S_N^*}^2 = \sum_{i=1}^{N+1} (\overline{T}_N^* - T_N^{(i)^*})^2 \ . \end{aligned}$$

The Hoeffding decomposition of T_N^* can be expressed as

$$T_N^* = \frac{2}{\sqrt{N}} \sum_{i=1}^N g_N(X_i^*) + \frac{2}{\sqrt{N}(N-1)} \sum\sum_{1 \le i < j \le N} \psi_N(X_i^*, X_j^*) \ ,$$

where the functions g_N and ψ_N are defined by

$$\begin{aligned} g_N(X_1^*) &= E^*\left(h(X_1^*, X_2^*) \,|\, X_1^*\right) - \theta_N^* \ , \\ \psi_N(X_1^*, X_2^*) &= h(X_1^*, X_2^*) - g_N(X_1^*) - g_N(X_2^*) - \theta_N^* \ . \end{aligned}$$

Here E^* denotes expectation under P_N, conditionally given $X_1, \ldots, X_N$. It is easily seen that the functions g_N and ψ_N may be expressed in terms of the functions g and ψ as follows.

$$\begin{aligned} g_N(X_1^*) &= g(X_1^*) - \overline{g} + \overline{\psi}(X_1^*) - \overline{\overline{\psi}} \ , \\ \psi_N(X_1^*, X_2^*) &= \psi(X_1^*, X_2^*) - \overline{\psi}(X_1^*) - \overline{\psi}(X_2^*) + \overline{\overline{\psi}} \ , \end{aligned}$$

where

$$\overline{g} = N^{-1} \sum_{i=1}^{N} g(X_i) \ , \ \overline{\psi}(x) = N^{-1} \sum_{i=1}^{N} \psi(x, X_i) \ , \ \overline{\overline{\psi}} = N^{-1} \sum_{j=1}^{N} \overline{\psi}(X_j) \ .$$

Corollary 6.1. *Suppose that the distribution of $g(X_1)$ is non-lattice, and suppose that there exist $p > 3$ and $r > 2$ such that $E|g(X_1)|^p < \infty$, $E|\psi(X_1, X_2)|^r < \infty$, and $E|\psi(X_1, X_1)|^{r/2} < \infty$. Then*

$$\sqrt{N} \sup_{x \in \mathbb{R}} |F_N^*(x) - F_N(x)| \xrightarrow{P} 0 \ .$$

Proof. It is clear from the proof of Theorem 3 of Helmers (1991) that the non-lattice condition on the distribution of $g(X_1)$ implies almost surely a non-lattice condition on the distribution of $g_N(X_1^*)$, which is uniform for large N, i.e. for every $0 < a < A < \infty$, there exists $\varepsilon > 0$ such that

$$\limsup_{N} \sup_{a \le |t| \le A} |E^* e^{itg_N(X_1^*}| \le 1 - \varepsilon \text{ a.s. } .$$

We proceed to show that $E^*|g_N(X_1^*)|^p$ is bounded in probability. Note that

$$E^*|g_N(X_1^*)|^p \le c_p \left(E^*|g(X_1^*)|^p + |\overline{g}|^p + E^*|\overline{\psi}(X_1^*)|^p + |\overline{\overline{\psi}}|^p \right) \ .$$

Since $E|g(X_1)|^p < \infty$, we have by the law of large numbers,

$$E^*|g(X_1^*)|^p \xrightarrow{P} E|g(X_1)|^p \ , \quad |\overline{g}|^p \xrightarrow{P} 0 \ .$$

Next,

$$\begin{aligned} E^*|\overline{\psi}(X_1^*)|^p + |\overline{\overline{\psi}}|^p &= N^{-1} \sum_{i=1}^{N} |N^{-1} \sum_{j=1}^{N} \psi(X_i, X_j)|^p + |N^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} \psi(X_i, X_j)|^p \\ &\le (N^{-1} + N^{-p}) \sum_{i=1}^{N} |N^{-1} \sum_{j=1}^{N} \psi(X_i, X_j)|^p \ . \end{aligned} \tag{6.1}$$

It suffices therefore to show that $N^{-1} \sum_{i=1}^{N} |N^{-1} \sum_{j=1}^{N} \psi(X_i, X_j)|^p$ goes to zero in probability. Write

$$N^{-1} \sum_{i=1}^{N} |N^{-1} \sum_{j=1}^{N} \psi(X_i, X_j)|^p = N^{-p-1} \sum_{i=1}^{N} |\psi(X_i, X_i)|^p + N^{-1} \sum_{i=1}^{N} |N^{-1} \sum_{j \ne i} \psi(X_i, X_j)|^p \ .$$

The first of these terms tends to zero almost surely, and a fortiori in probability, by the Marcinkiewicz strong law of large numbers (if $E|\psi(X_1, X_1)|^{p/(p+1)} < \infty$). Applying

Lemma 2.3 we see that there exists $\delta > 0$ and random variables $\tilde{\psi}(X_i, X_j)$, for $1 \le i < j \le N$ such that $|\tilde{\psi}(X_i, X_j)| \le N^{1-\delta}$ and

$$P(\tilde{\psi}(X_i, X_j) = \psi(X_i, X_j) \ , \ 1 \le i < j \le N) \ = \ 1 - o(1) \ .$$

It follows that

$$\begin{aligned}
P&\left(N^{-1}\sum_{i=1}^{N}\left|N^{-1}\sum_{j\neq i}\psi(X_i, X_j)\right|^p \ge \varepsilon\right) \\
&= P\left(N^{-1}\sum_{i=1}^{N}\left|N^{-1}\sum_{j\neq i}\tilde{\psi}(X_i, X_j)\right|^p \ge \varepsilon\right) + o(1) \\
&= P\left(N^{(1-\delta)(p-2)-3}\sum_{i=1}^{N}(\sum_{j\neq i}\tilde{\psi}(X_i, X_j))^2 \ge \varepsilon\right) + o(1) \\
&= P\left(N^{(1-\delta)(p-2)-3}\sum_{i=1}^{N}(\sum_{j\neq i}\psi(X_i, X_j))^2 \ge \varepsilon\right) + o(1) \\
&\le \varepsilon^{-1}N^{(1-\delta)(p-2)-3}E\sum_{i=1}^{N}(\sum_{j\neq i}\psi(X_i, X_j))^2 + o(1) \\
&\le \varepsilon^{-1}N^{(1-\delta)(p-2)-1}E\psi^2(X_1, X_2) + o(1) \to 0
\end{aligned}$$

if $3 < p < 2 + (1-\delta)^{-1}$. It follows that there exist $p > 3$ such that

$$E^*|\psi(X_1^*, \cdot)|^p + |\psi(\cdot, \cdot)|^p \xrightarrow{P} 0 \ . \tag{6.2}$$

Next, we show that $E^*|\psi_N(X_1^*, X_2^*)|^r$ is bounded in probability. Note that

$$E^*|\psi_N(X_1^*, X_2^*)|^r \le c_r\left(E^*|\psi(X_1^*, X_2^*)|^r + 2E^*|\overline{\psi}(X_1^*)|^r + |\overline{\overline{\psi}}|^r\right) \ .$$

First of all,

$$\begin{aligned}
E^*|\psi(X_1^*, X_2^*)|^r &= N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\psi(X_i, X_j)|^r \\
&= N^{-2}\sum_{i=1}^{N}|\psi(X_i, X_i)|^r + \binom{N}{2}^{-1}\sum\sum_{1\le i<j\le N}|\psi(X_i, X_j)|^r \ .
\end{aligned}$$

The first of these terms tends to zero almost surely by the Marcinkiewicz strong law of large numbers, the second to $E|\psi(X_1, X_2)|^r < \infty$ by the strong law of large numbers for U-statistics. Finally, for $r \le p$, we have by (6.2),

$$2E^*|\overline{\psi}(X_1^*)|^r + |\overline{\overline{\psi}}|^r \xrightarrow{P} 0 \ .$$

Application of Theorem 3.3 shows that

$$F_N^*(x) = \Phi(x) + \frac{\phi(x)}{6\sqrt{N}} \left[(2x^2+1)\lambda_1^* + 3(x^2+1)\lambda_2^*\right] \ ,$$

where λ_1^* and λ_2^* are given by

$$\lambda_1^* = E^* g_N^3(X_1^*) \ , \ \lambda_2^* = E^* g_N(X_1^*) g_N(X_2^*) \psi_N(X_1^*, X_2^*) \ .$$

Arguments, similar to the ones employed for the terms $E^*|g_N(X_1^*)|^p$ and $E^*|\psi_N(X_1^*, X_2^*)|^r$ show that $\lambda_i^* \xrightarrow{P} \lambda_i$ for $i = 1, 2$. Since the functions $(2x^2+1)\phi(x)$ and $(x^2+1)\phi(x)$ are bounded in x, the theorem is proved. $\square$

Acknowledgement. The authors are indebted to D. Chibisov for his help.

References

Babu, G. & Singh, K. (1983), 'Inference on means using the bootstrap', *Ann. Statist.* **11**, 999–1003.

Babu, G. & Singh, K. (1984), 'On one-term Edgeworth correction by Efron's bootstrap', *Sankhyā Ser. A* **46**, 219–232.

Bentkus, V., Götze, F. & van Zwet, W. R. (1994), 'An Edgeworth expansion for symmetric statistics'. Preprint 94-020, Sonderforschungsbereich 343, University of Bielefeld.

Beran, R. (1982), 'Estimated sampling distributions: The bootstrap and competitors', *Ann. Statist.* **10**, 212–225.

Beran, R. (1984), 'Bootstrap methods in statistics', *Jber. d. Dt. Math. Verein.* **86**, 14–30.

Bhattacharya, R. N. & Qumsiyeh, M. (1989), 'Second order and L^p-comparisons between the bootstrap and empirical Edgeworth expansion methodologies', *Ann. Statist.* **17**, 160–169.

Bickel, P. J., Götze, F. & van Zwet, W. R. (1986), 'The Edgeworth expansion for U-statistics of degree 2', *Ann. Statist.* **14**, 1463–1484.

Callaert, H. & Veraverbeke, N. (1981), 'The order of the normal approximation for a Studentized U-statistic', *Ann. Statist.* **9**, 194–200.

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

Dharmadhikari, S. W., Fabian, V. & Jogdeo, K. (1968), 'Bounds on the moments of martingales', *Ann. Math. Statist.* **39**, 1719–1723.

Helmers, R. (1991), 'On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic', *Ann. Statist.* **19**, 470–484.

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

Janssen, P. (1978), The Berry-Esséen theorem and an asymptotic expansion for U-statistics, PhD thesis, Limburg University Centre. (in Dutch).

Putter, H. (1994), Consistency of Resampling Methods, PhD thesis, University of Leiden.

Quenouille, M. H. (1949), 'Approximate tests of correlation in time series', *J. Roy. Statist. Soc. Ser. B* **11**, 68–84.

Quenouille, M. H. (1956), 'Notes on bias in estimation', *Biometrika* **43**, 353–360.

Singh, K. (1981), 'On the asymptotic accuracy of Efron's bootstrap', *Ann. Statist.* **9**, 1187–1195.

Tukey, J. W. (1958), 'Bias and confidence in not-quite large samples', *Ann. Math. Statist.* **29**, 614. Abstract.

van Zwet, W. R. (1984), 'A Berry-Esséen bound for symmetric statistics', *Z. Wahrsch. Verw. Geb.* **66**, 425–440.

van Zwet, W. R. (1990), *Asymptotic statistics*, Unpublished lecture notes, Chapel Hill.

van Zwet, W. R. (1995), 'Resampling, I: The jackknife and the naïve bootstrap'. To be submitted to the Annals of Statistics.

von Bahr, B. & Esséen, C. G. (1965), 'Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$', *Ann. Math. Statist.* **36**, 299–303.