*The Annals of Statistics*
1981, Vol. 9, No 1, 194–200

# THE ORDER OF THE NORMAL APPROXIMATION FOR A STUDENTIZED $U$-STATISTIC

BY HERMAN CALLAERT AND NOËL VERAVERBEKE

*Limburgs Universitair Centrum, Belgium*

Let $U_N$ be a one sample $U$-statistic with kernel $h$ of degree two, such that $Eh(X_1, X_2) = \vartheta$ and $\text{Var}\, E[h(X_1, X_2) \mid X_1] > 0$. It is shown that for a studentized $U$-statistic $\sup_x |P(N^{1/2} S_N^{-1}(U_N - \vartheta) \leq x) - \Phi(x)| = O(N^{-1/2})$ as $N \to \infty$, where $N^{-1} S_N^2 = 4N^{-1}(N-1)(N-2)^{-2} \sum_{i=1}^N [(N-1)^{-1} \sum_{j \neq i} h(X_i, X_j) - U_N]^2$ is the jackknife estimator of $\text{Var}\, U_N$. The condition needed to obtain this order bound is the existence of the 4.5th absolute moment of the kernel $h$. As in Helmers' Ph.D. thesis on linear combinations of order statistics, the analogous result for a studentized sum of i.i.d. random variables arises as a special case.

**1. Introduction.** Let $X_1, X_2, \ldots, X_N$, $N \geq 2$, be i.i.d. random variables with common distribution function $F$. Let $h(x, y)$ be a real-valued function, symmetric in its arguments, and with $Eh(X_1, X_2) = \vartheta$. Define a $U$-statistic by

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

and suppose that $g(X_1) = E[h(X_1, X_2) - \vartheta \mid X_1]$ has a positive variance $\sigma_g^2$. Then it is known that the distribution function (df) of $(\text{Var}\, U_N)^{-1/2} (U_N - \vartheta)$ converges, for $N \to \infty$, to the standard normal df $\Phi$ if $Eh^2(X_1, X_2) < \infty$ (Hoeffding (1948)). Further, the rate of this convergence to normality has been found to be $O(N^{-1/2})$ if $E|h(X_1, X_2)|^3 < \infty$ (Callaert and Janssen (1978)) or, more sharply, if $E|g(X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^2 < \infty$ (Helmers and van Zwet [6]). Further, if the variance of $U_N$ is unknown, let us denote the jackknife estimator of $\text{Var}\, U_N$ by $N^{-1} S_N^2$ (for information on jackknifing variances in general, see, e.g., Miller (1968)). Then the df of the studentized $U$-statistic $N^{1/2} S_N^{-1}(U_N - \vartheta)$ tends asymptotically to $\Phi$ if $Eh^2(X_1, X_2) < \infty$ (Arvesen (1969)). In the present paper the rate of convergence to normality is studied for this studentized $U$-statistic. Our study is inspired by part of the Ph.D. thesis of Helmers (1978) where an analogous result is found for linear combinations of order statistics.

**2. Preliminary results.** For ease of notation we will write from now on $\sum_{i \neq j}$, $\sum_{i<j}$ and $\sum_{i<j}^{(k)}$ for $\sum_{i=1;\, i \neq j}^N$, $\sum_{1 \leq i < j \leq N}$ and $\sum_{1 \leq i < j \leq N;\, i \neq k;\ j \neq k}$. We also systematically consider the $U$-statistic centered at its mean $\vartheta$. Further, let

$$\hat{U}_N = \sum_{k=1}^N E(U_N - \vartheta \mid X_k) = (2/N) \sum_{i=1}^N g(X_i)$$

be the projection of the $U$-statistic, and let

$$(U_N - \vartheta) - \hat{U}_N = \binom{N}{2}^{-1} \sum_{i<j} [h(X_i, X_j) - \vartheta - g(X_i) - g(X_j)] = \binom{N}{2}^{-1} \sum_{i<j} \varphi(X_i, X_j)$$

be its orthogonal complement. Note that $\hat{U}_N$ is a sum of i.i.d random variables with zero mean and that $(U_N - \vartheta) - \hat{U}_N$ is a sum of uncorrelated random variables with $E\varphi(X_1, X_2) = 0$ and also $E[\varphi(X_1, X_2) \mid X_1] = 0$. For the variance one has

$$\begin{aligned} \text{Var}\, U_N &= \text{Var}\, \hat{U}_N + \text{Var}\,(U_N - \hat{U}_N) \\ &= \frac{4\sigma_g^2}{N} + \binom{N}{2}^{-1} E\varphi^2(X_1, X_2). \end{aligned} \tag{1}$$

---

Received February 1979; revised August 1979.
AMS 1970 subject classifications. Primary 60F05; secondary 62E20.
*Key words and phrases.* Order of normal approximation, Studentized $U$-statistic, jackknife statistic, Student's $t$-statistic.

The following lemmas will be used frequently.

LEMMA 1. *Let*

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

*be a U-statistic with* $Eh(X_1, X_2) = \vartheta$. *Then, for any* $r \geq 1$, $\nu_r = E|h(X_1, X_2) - \vartheta|^r < \infty$ *implies that* $E|g(X_1)|^r < \nu_r$ *and* $E|\varphi(X_1, X_2)|^r < 3^r\nu_r$.

PROOF. Since $g(X_1) = E[h(X_1, X_2) - \vartheta | X_1]$, an application of Jensen's inequality for conditional expectations yields $E|g(X_1)|^r \leq \nu_r$. Further, using $E|\sum_{i=1}^N Z_i|^r \leq N^{r-1}\sum_{i=1}^N E|Z_i|^r$ for $r \geq 1$, one obtains for $\varphi(X_1, X_2) = h(X_1, X_2) - \vartheta - g(X_1) - g(X_2)$ that $E|\varphi(X_1, X_2)|^r \leq 3^r\nu_r$.

Note that, by Hölder's inequality, $\nu_r < \infty$ also assures the existence of mixed moments, such as $E|g^s(X_1)\varphi^t(X_1, X_2)|$ and $E|\varphi^s(X_1, X_2)\varphi^t(X_1, X_2)|$ whenever $s + t \leq r$.

LEMMA 2. *Let*

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

*be a U-statistic with* $Eh(X_1, X_2) = \vartheta$ *and* $\nu_r = E|h(X_1, X_2) - \vartheta|^r < \infty$ *for* $r \geq 1$. *Then* $E|U_N - \vartheta|^r \leq C\nu_r N^{-s}$ *where C is a constant,* $s = r - 1$ *for* $1 \leq r \leq 2$, *and* $s = r/2$ *for* $r \geq 2$.

PROOF. Since

$$U_N - \vartheta = \frac{2}{N}\sum_{i=1}^N g(X_i) + \binom{N}{2}^{-1} \sum_{i<j} \varphi(X_i, X_j),$$

one has

$$E|U_N - \vartheta|^r \leq 2^{r-1} E|2N^{-1}\sum_{i=1}^N g(X_i)|^r + 2^{r-1} E|N^{-1}(N-1)^{-1}\sum_{i=1}^N \sum_{j\neq i} \varphi(X_i, X_j)|^r. \tag{2}$$

We first consider the case $1 \leq r \leq 2$.

Application of a theorem of von Bahr and Esseén (1965) yields

$$E|2N^{-1}\sum_{i=1}^N g(X_i)|^r \leq 2^{r+1} E|g(X_1)|^r N^{1-r}. \tag{3}$$

Further

$$\begin{aligned} E|N^{-1}\sum_{i=1}^N (N-1)^{-1}\sum_{j\neq i} \varphi(X_i, X_j)|^r &\leq N^{-1}\sum_{i=1}^N E|(N-1)^{-1}\sum_{j\neq i}\varphi(X_i, X_j)|^r \\ &= N^{-1}\sum_{i=1}^N E[E(|(N-1)^{-1}\sum_{j\neq i}\varphi(X_i, X_j)|^r | X_i)] \end{aligned} \tag{4}$$

where, given $X_i$, $\sum_{j\neq i}\varphi(X_i, X_j)$ is a sum of i.i.d random variables with zero mean. Hence, using again the von Bahr and Esseén theorem, one has

$$\begin{aligned} &E[E(|(N-1)^{-1}\sum_{j\neq i}\varphi(X_i, X_j)|^r X_i)] \\ &\leq 2(N-1)^{1-r} E|\varphi(X_1, X_2)|^r \leq 2^r N^{1-r} E|\varphi(X_1, X_2)|^r. \end{aligned} \tag{5}$$

Applying Lemma 1 and combining (2), (3), (4) and (5), the lemma is proved for $1 \leq r \leq 2$.

If $r \geq 2$, a result by Chung (1951) together with (4) yield

$$E|2N^{-1}\sum_{i=1}^N g(X_i)|^r \leq CE|g(X_1)|^r N^{-r/2} \leq C\nu_r N^{-r/2} \tag{6}$$

and

$$E|N^{-1}(N-1)^{-1}\sum_{i=1}^N \sum_{j\neq i}\varphi(X_i, X_j)|^r \leq C\nu_r N^{-r/2}. \tag{7}$$

Hence, from (2), (6) and (7), the lemma holds for $r \geq 2$.

LEMMA 3. *Let*

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

*be a U-statistic with* $Eh(X_1, X_2) = \vartheta$ *and such that* $g(X_1) = E[h(X_1, X_2) - \vartheta \mid X_1]$ *has a positive variance* $\sigma_g^2$. *If* $E|g(X_1)|^3 < \infty$ *and* $E|h(X_1, X_2)|^2 < \infty$ *then, for* $N \to \infty$,

$$\sup_x |P(N^{1/2}2^{-1}\sigma_g^{-1}(U_N - \vartheta) \le x) - \Phi(x)| = O(N^{-1/2}). \tag{8}$$

PROOF. The lemma relies on the proof of the theorem in Callaert and Janssen (1978), where, putting $\vartheta = 0$, it is shown that for $\nu_3 = E|h(X_1, X_2)|^3 < \infty$, $\sup_x |P(N^{1/2}2^{-1}\sigma_g^{-1}U_N \le x) - \Phi(x)| \le C\nu_3\sigma_g^{-3}N^{-1/2}$ for $N \ge 2$ and $C$ an absolute constant. Recently Helmers and van Zwet [6] were able to refine the condition $\nu_3 < \infty$ by imposing $E|g(X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^2 < \infty$.

It should be noted that (8) also applies if $2\sigma_g N^{-1/2}$ is replaced by $(\text{Var } U_N)^{1/2}$, giving the exact standardization. In this paper, however, we will use the fact that a standardization by the variance of the projection of the $U$-statistic also yields the right order bound, as written in (8).

## 3. Main result.

THEOREM. *Let*

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

*be a U-statistic with* $Eh(X_1, X_2) = \vartheta$ *and such that* $\sigma_g^2 > 0$. *If* $E|h(X_1, X_2)|^{4.5} < \infty$ *then for* $N \to \infty$,

$$\sup_x |P(N^{1/2}S_N^{-1}(U_N - \vartheta) \le x) - \Phi(x)| = O(N^{-1/2}),$$

*where*

$$S_N^2 = 4(N-1)(N-2)^{-2} \sum_{i=1}^N [(N-1)^{-1} \sum_{j\ne i} h(X_i, X_j) - U_N]^2. \tag{9}$$

Before proceeding to the proof we give some comments on the variance estimator. Note that $N^{-1}S_N^2$ is nothing but the jackknife estimator of Var $U_N$, i.e., $S_N^2$ is the sample variance of the "pseudo-values" $\hat\theta_i = NU_N - (N-1)U_{N-1}^i$ with $U_{N-1}^i$ the statistic based upon $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N$. As shown in (A9) of the appendix $ES_N^2 = 4\sigma_g^2 + 4(N-2)^{-1} E\varphi^2(X_1, X_2)$ and hence, $S_N^2$ is not an unbiased estimator of $N$ Var $U_N$. However $S_N^2$ does have the following desirable properties:

(i) for any sample size, $S_N^2$ is nonnegative,

(ii) for $U$-statistics with vanishing orthogonal complement, $S_N^2$ equals $4(N-1)^{-1} \sum_{i=1}^N [g(X_i) - N^{-1}\sum_{k=1}^N g(X_k)]^2$, in compliance with the classical theory for i.i.d. random variables. Note that, if $U_N = 2N^{-1}\sum_{i=1}^N g(X_i)$, then $N^{1/2}S_N^{-1}(U_N - \vartheta)$ becomes the well-known $t$-statistic for which our theorem sharpens the moment conditions of Chung's result (1946).

Another way for arriving at (9) consists in considering (8) of Lemma 3 and noting that $4\sigma_g^2 N^{-1}$ is the variance of $2N^{-1}\sum_{i=1}^N g(X_i)$. Hence one might start with $4(N-1)^{-1}\sum_{i=1}^N [g(X_i) - N^{-1}\sum_{k=1}^N g(X_k)]^2$ as an estimator for $4\sigma_g^2$. However, the kernel $h(x, y)$ is the only function one really knows since, the underlying distribution $F$ being unavailable, the function $g(x)$ cannot be computed explicitly. Therefore, based upon the fact that $g(x) + \vartheta = \int h(x, y)\, dF(y)$ and replacing $F$ by the empirical distribution function $F_N$, one arrives at precisely the same expression for $S_N^2$ as given in (9).

Essentially, $S_N^2$ has the structure $4\sigma_g^2 + T_N + R_N$ where $T_N$ equals $1/N$ times a sum of i.i.d random variables with zero mean and involving no higher than second moments of the kernel $h$, and $R_N$ is a quantity which is $O(N^{-1/2}/\ln N)$ except on a set of probability $O(N^{-1/2})$. As will be seen from the proof of the theorem, any other proposal for $S_N^2$ which has the just mentioned structure and which does not violate the imposed 4.5th moment condition, will lead to the right order bound for the studentized $U$-statistic. In this context, we constructed several other estimators for $N$ Var $U_N$ or for $4\sigma_g^2$ in order to reduce or even to eliminate the bias of our proposal. All of them resulted in variance estimators which for finite sample sizes could be negative with positive probability. We illustrate this by just one example: the $U$-statistic estimator of $4\sigma_g^2 = \int\int\int\int h(x, y)[h(x, z) - h(z, u)]\, dF(x)\, dF(y)\, dF(z)\, dF(u)$ leads to the following proposal (the kernel can easily be symmetrized)

$$S_N^* = \frac{4}{N(N-1)(N-2)(N-3)} \sum_{i=1}^N \sum_{j\neq i} \sum_{k\neq i,j} \sum_{m\neq i,j,k} h(X_i, X_j)[h(X_i, X_k) - h(X_k, X_m)]$$

which is an unbiased estimator for $4\sigma_g^2$ but with the undesirable property mentioned above.

**4. Proof of the theorem.** The main idea of the proof consists in writing

$$\frac{N^{1/2}(U_N - \vartheta)}{S_N} = \frac{N^{1/2}(U_N - \vartheta)}{2\sigma_g} \cdot 2\sigma_g S_N^{-1}$$

and using a stochastic expansion for $2\sigma_g S_N^{-1}$.

From (A8) and the fact that $(1 + x)^{-1/2} = 1 - \frac{1}{2} x + O(x^2)$ as $x \to 0$, it follows that $2\sigma_g S_N^{-1} = 1 - \frac{1}{8}\sigma_g^{-2} T_N + \tilde{R}_N$, where $\tilde{R}_N = O(N^{-1/2}/\ln N)$ except on a set with probability $O(N^{-1/2})$ as $N \to \infty$. Indeed, $\tilde{R}_N = -\frac{1}{8}\sigma_g^{-2} R_N + O(T_N + R_N)^2$ which is $O(N^{-1/2}/\ln N)$ because of (A10) and (A11). Hence, except on a set with probability $O(N^{-1/2})$ as $N \to \infty$, we may work with $W_N = N^{1/2} 2^{-1} \sigma_g^{-1}(U_N - \vartheta)(1 - \frac{1}{8}\sigma_g^{-2} T_N)$ instead of $S_N^{-1} N^{1/2}(U_N - \vartheta)$ since

$$\begin{aligned} P[\,|N^{1/2}2^{-1}\sigma_g^{-1}(U_N - \vartheta)\tilde{R}_N| \geq N^{-1/2}] &\leq P[\,|\tilde{R}_N| \geq N^{-1/2}/\ln N] + P[\,|N^{1/2}2^{-1}\sigma_g^{-1}(U_N - \vartheta)| \geq \ln N] \\ &\leq O(N^{-1/2}) + 2[1 - \Phi(\ln N)] \text{ as } N \to \infty, \text{ using Lemma 3.} \end{aligned}$$

Writing $W_N$ as

$$\frac{N^{1/2}}{2\sigma_g}\left[\frac{2}{N}\sum_{i=1}^N g(X_i) + \binom{N}{2}^{-1} \sum_{i<j} \varphi(X_i, X_j)\right]\left[1 - \frac{1}{8}\sigma_g^{-2}\frac{1}{N}\sum_{i=1}^N f(X_i)\right]$$

and working this out, we obtain the following decomposition:

$$W_N = \frac{N^{1/2}}{2\sigma_g} U_N^* + Z_{N1} + Z_{N2} + Z_{N3}, \tag{10}$$

where

$$U_N^* = \frac{2}{N}\sum_{i=1}^N g(X_i) + \binom{N}{2}^{-1} \sum_{i<j}\left\{\varphi(X_i, X_j) - \frac{1}{8}\sigma_g^{-2}\frac{N-1}{N}[g(X_i)f(X_j) + g(X_j)f(X_i)]\right\}$$

$$Z_{N1} = -\frac{1}{8}\sigma_g^{-3}N^{-3/2}\sum_{i=1}^N \{g(X_i)f(X_i) - E[g(X_i)f(X_i)]\} - \frac{1}{8}\sigma_g^{-3}N^{-1/2}E[g(X_1)f(X_1)]$$

$$Z_{N2} = -\frac{1}{16}\sigma_g^{-3}N^{-1/2}\binom{N}{2}^{-1} \sum_{i<j}\varphi(X_i, X_j)[f(X_i) + f(X_j)]$$

$$Z_{N3} = -\frac{1}{16}\sigma_g^{-3}(N-2)N^{-3/2}\sum_{i=1}^N f(X_i)\left[\binom{N-1}{2}^{-1} \sum_{k<m}^{(i)} \varphi(X_k, X_m)\right].$$

We now show that, for $i = 1, 2, 3$, $Z_{Ni} = O(N^{-1/2})$ except on a set with probability $O(N^{-1/2})$ as $N \to \infty$. The first term of $Z_{N1}$ and also $Z_{N2}$ can be considered as $N^{-1/2}$ times a $U$-statistic with kernel having zero mean and finite absolute moment of order $\frac{3}{2}$. Hence the Markov inequality together with Lemma 2 provide the desired order bound for $Z_{N1}$ and $Z_{N2}$.

We point out that $Z_{N1}$ and $Z_{N2}$ are the only terms which require the full 4.5th moment condition. A straightforward computation (see (A12)) shows that $E(Z_{N3}^2) = O(N^{-2})$ which is sufficient for our purpose.

The quantity $U_N^*$ in (10) is easily seen to be a $U$-statistic with kernel

$$g(X_i) + g(X_j) + \varphi(X_i, X_j) - \frac{1}{8}\sigma_g^{-2}\frac{N-1}{N}[g(X_i)f(X_j) + g(X_j)f(X_i)]$$

having zero mean. Since its projection is $2N^{-1}\sum_{i=1}^N g(X_i)$, the first term in (10) is a properly normalized $U$-statistic, to which Lemma 3 can be applied. Note that by the independence of $g(X_i)$ and $f(X_j)$, no higher moment than $Eh^4$ is required here.

A classical argument finally yields

$$\sup_x |P(N^{1/2}S_N^{-1}(U_N - \vartheta) \leq x) - \Phi(x)| \leq \sup_x |P(N^{1/2}2^{-1}\sigma_g^{-1}U_N^* \leq x) - \Phi(x)| + O(N^{-1/2}) = O(N^{-1/2}) \text{ as } N \to \infty.$$

**5. Remark.** From the proof of the theorem and the computations in the appendix, it is seen that one arrives at the desired result by almost exclusively using decompositions and order bounds of quantities which have the structure of a $U$-statistic. We therefore conjecture that the theorem, formulated in this paper for a one sample $U$-statistic of degree two, is also valid for multisample $U$-statistics of arbitrary order. As a hint we refer to the method of proof for the Berry-Esseén theorem of general $U$-statistics as worked out in the Ph.D. thesis of Janssen (1978). The computations, however, might be rather involved.

## APPENDIX

We draw special attention to the second term in (1) which shows that for the orthogonal complement of a $U$-statistic

$$\text{(A1)} \qquad E\left[\binom{N}{2}^{-1} \sum_{i<j} \varphi(X_i, X_j)\right]^2 = O(N^{-2}) \text{ as } N \to \infty.$$

Also note that if $k$ is any integrable function such that $E|k\varphi| < \infty$, then for any $i$

$$\text{(A2)} \qquad E[k(X_i)\varphi(X_1, X_2)] = 0,$$

either by $E\{k(X_1)E[\varphi(X_1, X_2) \mid X_1]\} = 0$ or by independence and $E\varphi(X_1, X_2) = 0$.

We now construct a decomposition of $S_N^2$. Remarking that $N^{-1}\sum_{i=1}^N [(N-1)^{-1}\sum_{j\neq i}(h(X_i, X_j) - \vartheta)] = U_N - \vartheta$, we can write (9) as

$$\text{(A3)} \qquad S_N^2 = \frac{8}{(N-1)(N-2)^2}\{\sum_{i<j}(h(X_i, X_j) - \vartheta)^2 + \sum_{i=1}^N \sum_{k<m}^{(i)} (h(X_i, X_k) - \vartheta)(h(X_i, X_m) - \vartheta)\} - 4N(N-1)(N-2)^{-2}(U_N - \vartheta)^2.$$

Noting that $h(X_i, X_j) - \vartheta = g(X_i) + g(X_j) + \varphi(X_i, X_j)$, we arrive at

$$\text{(A4)} \qquad \sum_{i<j}(h(X_i, X_j) - \vartheta)^2 = (N-1)\sum_{i=1}^N g^2(X_i) + \sum_{i<j}[2g(X_i)g(X_j) + 2[g(X_i) + g(X_j)]\varphi(X_i, X_j) + \varphi^2(X_i, X_j)]$$

$$\text{(A5)} \qquad \sum_{i=1}^N \sum_{k<m}^{(i)} (h(X_i, X_k) - \vartheta)(h(X_i, X_m) - \vartheta) = \frac{(N-1)(N-2)}{2}\sum_{i=1}^N g^2(X_i) + (N-2)\sum_{i<j}\{3g(X_i)g(X_j) + [g(X_i) + g(X_j)]\varphi(X_i, X_j)\} + \sum_{i=1}^N \sum_{k<m}^{(i)} \varphi(X_i, X_k)\varphi(X_i, X_m) + 2\sum_{i=1}^N g(X_i)\sum_{k<m}^{(i)}\varphi(X_k, X_m)$$

$$(U_n - \vartheta)^2 = \frac{4}{N^2}\sum_{i=1}^N g^2(X_i) + \frac{8}{N^2}\sum_{i<j} g(X_i)g(X_j) + \frac{4}{N^2(N-1)^2}[\sum_{i<j}\varphi(X_i, X_j)]^2$$

(A6) $$+\frac{8}{N^2(N-1)}\textstyle\sum_{i<j}[g(X_i)+g(X_j)]\varphi(X_i, X_j) + \frac{8}{N^2(N-1)}\sum_{i=1}^{N} g(X_i)\sum_{k<m}^{(i)}\varphi(X_k, X_m).$$

Inserting (A4), (A5) and (A6) into (A3) and introducing the function $\tilde{g}(x) = \int g(y)\varphi(x, y)\, dF(y)$, we obtain

(A7)
$$\begin{aligned}
S_N^2 = {} & 4\sigma_g^2 \\
& + \frac{1}{N}\textstyle\sum_{i=1}^{N}[4(g^2(X_i) - \sigma_g^2) + 8\tilde{g}(X_i)] && (T_N) \\
& - 4\binom{N}{2}^{-1}\textstyle\sum_{i<j} g(X_i)g(X_j) && (R_{N1}) \\
& + 4\binom{N}{2}^{-1}\textstyle\sum_{i<j}[(g(X_i)+g(X_j))\varphi(X_i, X_j) - \tilde{g}(X_i) - \tilde{g}(X_j)] && (R_{N2}) \\
& - \frac{8}{N}\textstyle\sum_{i=1}^{N}\left[g(X_i)\binom{N-1}{2}^{-1}\sum_{k<m}^{(i)}\varphi(X_k, X_m)\right] && (R_{N3}) \\
& + \frac{4}{N-2}\textstyle\sum_{i=1}^{N}\left[\binom{N-1}{2}^{-1}\sum_{k<m}^{(i)}\varphi(X_i, X_k)\varphi(X_i, X_m)\right] && (R_{N4}) \\
& - \frac{4N(N-1)}{(N-2)^2}\left[\binom{N}{2}^{-1}\textstyle\sum_{i<j}\varphi(X_i, X_j)\right]^2 && (R_{N5}) \\
& + \frac{4N}{(N-2)^2}\binom{N}{2}^{-1}\textstyle\sum_{i<j}\varphi^2(X_i, X_j) && (R_{N6})
\end{aligned}$$

or

(A8) $$S_N^2 = 4\sigma_g^2 + T_N + R_N$$

where $T_N = N^{-1}\sum_{i=1}^{N} f(X_i)$ with $f(X_i) = 4(g^2(X_i) - \sigma_g^2) + 8\tilde{g}(X_i)$ and $R_N = \sum_{i=1}^{6} R_{Ni}$. Since $Eg(X_1) = E\tilde{g}(X_1) = E\varphi(X_1, X_2) = 0$ and using the independence and (A2) we arrive at

(A9) $$ES_N^2 = 4\sigma_g^2 + \frac{4}{N-2}E\,\varphi^2(X_1, X_2).$$

We now show that

(A10) $$P(|R_N| \geqslant N^{-1/2}/\ln N) = O(N^{-1/2}) \text{ as } N \to \infty,$$

imposing a no higher than fourth moment condition on the kernel $h$. First remark that $E(R_{Ni}) = 0$ for $i = 1, 2, 3, 4$. We then apply (A1) to $R_{N1}$ and $R_{N2}$ directly and to $R_{N3}$ after writing

$$E(R_{N3}^2) \leqslant \frac{64}{N}\sum_{i=1}^{N}\left\{Eg^2(X_i)E\left[\binom{N-1}{2}^{-1}\sum_{k<m}^{(i)}\varphi(X_k, X_m)\right]^2\right\}.$$

To $R_{N4}$ we perform the same operation and, remarking that the summands in $\sum_{k<m}^{(i)}$ are uncorrelated with mean zero, also find

$$E(R_{N4}^2) = O(N^{-2}) \text{ as } N \to \infty.$$

That $E|R_{N5}| = O(N^{-2})$ is immediate from (A1) whereas $E(R_{N6}^2) \leqslant 16N^2(N-2)^{-4}E\varphi^4(X_1, X_2) = O(N^{-2})$. Hence, using Markov's inequality we arrive at the desired order bound for $R_N$. That

(A11) $$P(T_N^2 \geqslant N^{-1/2}/\ln N) = O(N^{-1/2}) \text{ as } N \to \infty$$

relies on the fact that $T_N$ consists of a sum of i.i.d. random variables with $Ef(X_1) = 0$. Using (6) in the proof of Lemma 2 with $r = 2 + 2\epsilon$ and $0 < \epsilon < 1/8$, one finds that $E|T_N|^{2+2\epsilon} \leqslant C'N^{-1-\epsilon}$. Hence, from Markov's inequality for $T_N^2$ we obtain $P(T_N^2 \geqslant N^{-1/2}/\ln N) \leqslant C''N^{-1/2-\epsilon/2}(\ln N)^{1+\epsilon}$. Note that, with $0 < \epsilon < 1/8$, the existence of $E|h(X_1, X_2)|^{4.5}$ suffices for our proof. Finally we give a sketch of the proof that

$$EZ_{N3}^2 = O(N^{-2}). \tag{A12}$$

Write $E(Z_{N3}^2) = 1/256\, \sigma_g^{-6}(N-2)^2 N^{-3} E(\sum_{i=1}^N B_i)^2$. Then, by independence,

$$\sum_{i=1}^N EB_i^2 = \sum_{i=1}^N Ef^2(X_i)E\left[\binom{N-1}{2}^{-1} \sum_{k<m}^{(i)} \varphi(X_k, X_m)\right]^2 = O(N^{-1})$$

using (A1). Further

$$\sum_{u<v} E(B_u B_v) = \binom{N-1}{2}^{-2} \sum_{u<v} E[f(X_u) \sum_{k<m}^{(u)} \varphi(X_k, X_m) f(X_v) \sum_{r<s}^{(v)} \varphi(X_r, X_s)]$$

where

$$E[f(X_1)\varphi(X_2, X_3)f(X_2)\varphi(X_1, X_4)] = E[f(X_1)\varphi(X_1, X_4)]E[f(X_2)\varphi(X_2, X_3)] = 0$$

by independence and (A2). And also

$$E[f(X_1)\varphi(X_2, X_3)f(X_2)\varphi(X_3, X_4)] = E[f(X_1)]E[\varphi(X_2, X_3)f(X_2)\varphi(X_3, X_4)] = 0$$

since $Ef(X_1) = 0$. Hence $\sum_{u<v} E(B_u B_v) = O(N^{-1})$.

Note that the nonzero terms, which necessarily are of the form $E[f(X_1)\varphi(X_2, X_3)f(X_2)\varphi(X_1, X_3)]$, do not involve moments of order higher than $Eh^4$ using Hölder's inequality and the independence present.

**Acknowledgment.** The authors thank the referee for his valuable comments leading to a considerable improvement in the presentation of the paper.

## REFERENCES

[1] ARVESEN, J. N. (1969). Jackknifing $U$-statistics. *Ann. Math. Statist.* **40** 2076–2100.

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

[3] CHUNG, K. L. (1946). The approximate distribution of Student's statistic. *Ann. Math. Statist.* **17** 442–465.

[4] CHUNG, K. L. (1951). The strong law of large numbers. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 341–352. Univ. California Press.

[5] HELMERS, R. (1978). Edgeworth expansions for linear combinations of order statistics. Ph.D. thesis. Math. Centre, Amsterdam. Math. Centre Track 105.

[6] HELMERS, R. and VAN ZWET, W. R. (1979). The Berry-Esseén bound for $U$-statistics. Unpublished manuscript.

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

[8] JANSSEN, P. (1978). The Berry-Esseén theorem and an asymptotic expansion for $U$-statistics. Ph.D. thesis (in Dutch).

[9] MILLER, R. G. JR. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 567–582.

[10] VON BAHR, B. and ESSEÉN, C. G. (1965). Inequalities for the $r$th absolute moment of a sum of random variables, $1 \leqslant r \leqslant 2$. *Ann. Math. Statist.* **36** 299–303.

DEPARTMENT OF MATHEMATICS
LIMBURGS UNIVERSITAIR CENTRUM
3610 Diepenbeek, Belgium