The Annals of Statistics
1991, Vol. 19, No. 1, 470–484

ON THE EDGEWORTH EXPANSION AND THE BOOTSTRAP APPROXIMATION FOR A STUDENTIZED U-STATISTIC

BY R. HELMERS

Centre for Mathematics and Computer Science

The asymptotic accuracy of the estimated one-term Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic is investigated. It is shown that both the Edgeworth expansion estimate and the bootstrap approximation are asymptotically closer to the exact distribution of a Studentized U-statistic than the normal approximation. The conditions needed to obtain these results are weak moment assumptions on the kernel h of the U-statistic and a nonlattice condition for the distribution of $g(X_1) = E[h(X_1, X_2)|X_1]$. As an application improved Edgeworth and bootstrap based confidence intervals for the mean of a U-statistic are obtained.

1. Introduction and main results. Let $X_1, X_2, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables (r.v.) with common distribution function (d.f.) F. Let h be a real-valued symmetric function of its two arguments with

$$Eh(X_1, X_2) = \theta. \tag{1.1}$$

Define a U-statistic of degree 2 by

$$U_n = \binom{n}{2}^{-1} \sum\sum_{1 \le i < j \le n} h(X_i, X_j) \tag{1.2}$$

and suppose that

$$g(X_1) = E\big[h(X_1, X_2) - \theta | X_1\big] \tag{1.3}$$

has a positive variance σ_g^2. Let

$$S_n^2 = 4(n-1)(n-2)^{-2} \sum_{i=1}^{n} \left[(n-1)^{-1} \sum_{j=1}^{n} h(X_i, X_j) - U_n\right]^2 \tag{1.4}$$

and note that $n^{-1}S_n^2$ is the jackknife estimator of the variance of U_n. Let, for each $n \ge 2$ and real x,

$$F_n(x) = P\big(\{n^{1/2}S_n^{-1}(U_n - \theta) \le x\}\big). \tag{1.5}$$

It is well-known that F_n converges in distribution to the standard normal d.f. Φ, as $n \to \infty$, provided $Eh^2(X_1, X_2) < \infty$ and $\sigma_g^2 > 0$ [cf. Arvesen (1969)]. The

Received July 1987; revised March 1990.

AMS 1980 *subject classifications*. Primary 62E20, 62G05; secondary 60F05.

Key words and phrases. Edgeworth expansions, bootstrap approximations, studentized U-statistics, bootstrap confidence intervals, Edgeworth based confidence intervals.

speed of this convergence to normality is of the classical order $n^{-1/2}$ [cf. Callaert and Veraverbeke (1981), Zhao (1983) and Helmers (1985)].

The traditional way to improve upon the normal approximation is to establish a one-term Edgeworth expansion for F_n. Let, for $n \geq 2$ and real x,

$$\tilde{F}_n(x) = \Phi(x) + 6^{-1}n^{-1/2}\sigma_g^{-3}\phi(x)\{(2x^2+1)Eg^3(X_1) + 3(x^2+1)Eg(X_1)g(X_2)h(X_1,X_2)\}. \tag{1.6}$$

THEOREM 1. *Suppose that*

$$E|h(X_1,X_2)|^{4+\varepsilon} < \infty \quad \textit{for some } \varepsilon > 0 \tag{1.7}$$

and

$$\textit{the d.f. of } g(X_1) \textit{ is nonlattice.} \tag{1.8}$$

Then, as $n \to \infty$,

$$\sup_x |F_n(x) - \tilde{F}_n(x)| = o(n^{-1/2}). \tag{1.9}$$

Note that the nondegeneracy condition $\sigma_g^2 > 0$, which is already needed to ensure asymptotic normality, is easily implied by assumption (1.8).

The proof of Theorem 1 (cf. Section 2) depends heavily on the results of Callaert, Janssen and Veraverbeke (1980), Callaert and Veraverbeke (1981) and Helmers (1985). In this connection I also want to mention the paper of Bickel, Götze and van Zwet (1986), which contains the best result concerning two-term Edgeworth expansions for normalized U-statistics of degree 2 so far obtained.

In a non- or semiparametric framework, F is completely unknown, and one does not know the quantities

$$a = \frac{Eg^3(X_1)}{\left(Eg^2(X_1)\right)^{3/2}}, \qquad b = \frac{Eg(X_1)g(X_2)h(X_1,X_2)}{\left(Eg^2(X_1)\right)^{3/2}} \tag{1.10}$$

appearing in the expansion (1.6). These moments depend on the underlying d.f. F and must be estimated from the observations $X_1,\dots,X_n$. One way of doing this is to compute bootstrap estimates for a and b; i.e., we replace a and b by their empirical counterparts. Let $\hat{F}_n$ denote the empirical d.f. based on $X_1,\dots,X_n$. Conditionally given $X_1,\dots,X_n$, let $X_1^*,\dots,X_n^*$ be n independent r.v.'s with common d.f. $\hat{F}_n$, the bootstrap sample of size n drawn with replacement from $\hat{F}_n$. Bootstrap estimates a_n and b_n of a and b are obtained by simply replacing X_1, X_2, g and E by X_1^*, X_2^*, g_n and E^*, where

$$g_n(X_i^*) = E^*\left[h(X_1^*,X_2^*) - \theta_n \middle| X_i^*\right] \tag{1.11}$$

for $i = 1, 2$, and

$$\theta_n = E^*h(X_1^*,X_2^*) = n^{-2}\sum_{i=1}^n\sum_{j=1}^n h(X_i,X_j). \tag{1.12}$$

E^* of course refers to the conditional expectation w.r.t. $\hat{F}_n$, conditionally given that $X_1, \dots, X_n$ are observed. A simple calculation yields

$$(1.13)\quad a_n = a_n(X_1, \dots, X_n) = \frac{n^{-1}\sum_{i=1}^{n}\left(n^{-1}\sum_{j=1}^{n} h(X_i, X_j) - \theta_n\right)^3}{\left(n^{-1}\sum_{i=1}^{n}\left(n^{-1}\sum_{j=1}^{n} h(X_i, X_j) - \theta_n\right)^2\right)^{3/2}}$$

and

$$(1.14)\quad \begin{aligned} b_n &= b_n(X_1, \dots, X_n) \\ &= \frac{\begin{aligned} &n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(n^{-1}\sum_{k=1}^{n} h(X_i, X_k) - \theta_n\right) \\ &\qquad \times\left(n^{-1}\sum_{l=1}^{n} h(X_j, X_l) - \theta_n\right) h(X_i, X_j) \end{aligned}}{\left(n^{-1}\sum_{i=1}^{n}\left(n^{-1}\sum_{j=1}^{n} h(X_i, X_j) - \theta_n\right)^2\right)^{3/2}}. \end{aligned}$$

Thus easily computable expressions for the bootstrap estimates a_n and b_n are available and no Monte Carlo simulations are required for the evaluation of these estimates.

In our second theorem we shall show that we may replace the quantities a and b in the expansion (1.6) by the bootstrap estimates a_n and b_n, without affecting the asymptotic accuracy of the expansion. Let, for $n \geq 2$ and real x,

$$(1.15)\quad \tilde{E}_n(x) = \Phi(x) + 6^{-1}n^{-1/2}\phi(x)\{(2x^2 + 1)a_n + 3(x^2 + 1)b_n\}$$

denote the resulting one-term estimated Edgeworth expansion for F_n. In contrast with $\tilde{F}_n$ [cf. (1.6)], the expansion $\tilde{E}_n$ can be computed from the observations $X_1, \dots, X_n$.

THEOREM 2. *Suppose that the assumptions of Theorem 1 are satisfied, and, in addition,*

$$(1.16)\quad E|h(X_1, X_1)|^3 < \infty.$$

Then, with probability 1, *as* $n \to \infty$,

$$(1.17)\quad \sup_x |F_n(x) - \tilde{E}_n(x)| = o(n^{-1/2}).$$

Theorem 2 tells us that the *Edgeworth expansion estimate* $\tilde{E}_n$ is asymptotically $\tilde{E}$ closer to the exact d.f. F_n than the classical normal approximation Φ. In a way E_n adapts itself to the possible asymmetry present in the exact d.f. F_n; the normal approximation of course fails to achieve this.

Another possibility to obtain an improved approximation for F_n is to employ bootstrap methods in a more direct way. We consider the bootstrapped Studentized U-statistic, corresponding to $n^{1/2}S_n^{-1}(U_n - \theta)$, based on the bootstrap sample $X_1^*, \dots, X_n^*$, which is given by

$$(1.18)\quad n^{1/2}S_n^{*-1}(U_n^* - \theta_n).$$

Here U_n^* and S_n^* are obtained from U_n and S_n simply by replacing the X_i's by the X_i^*'s in (1.2) and (1.4); the parameter θ [cf. (1.1)] is replaced by its natural estimator θ_n [cf. (1.12)]. The bootstrap approximation

$$F_n^*(x) = P^*\left(n^{1/2}S_n^{*-1}\left(U_n^* - \theta_n\right) \le x\right) \tag{1.19}$$

for $n \ge 2$ and real x, is nothing else but the conditional distribution of $n^{1/2}S_n^{*-1}(U_n^* - \theta_n)$, conditionally given the observed values of $X_1, \dots, X_n$; P^* of course refers to the conditional probability measure corresponding to $\hat{F}_n$.

Athreya, Ghosh, Low and Sen (1984) recently showed that

$$\sup_x \left| F_n(x) - F_n^*(x) \right| \to 0 \quad \text{as } n \to \infty, \tag{1.20}$$

with probability 1, provided, in addition to the assumptions already needed to guarantee asymptotic normality of F_n [cf. Arvesen (1969)], the requirement $Eh^2(X_1, X_1) < \infty$ is imposed. We also refer to Bickel and Freedman (1981) for a closely related result for normalized U-statistics.

THEOREM 3. *Suppose that the assumptions of Theorem* 2 *are satisfied. Then, with probability* 1, *as* $n \to \infty$,

$$\sup_x |F_n(x) - F_n^*(x)| = o(n^{-1/2}). \tag{1.21}$$

We see that the *bootstrap approximation* F_n^* shares with the Edgeworth based estimate $\tilde{E}_n$ the property of being asymptotically closer to the exact d.f. F_n than the normal approximation Φ. [See Beran (1982, 1984) for some related results suggesting that F_n^*, like $\tilde{E}_n$, should be locally asymptotically minimax among all possible estimates of F_n.] Both $\tilde{E}_n$ as well as F_n^* reflect—at least to first order—the asymmetry present in F_n. In contrast to $\tilde{E}_n$, the bootstrap approximation F_n^* cannot be evaluated explicitly, and Monte Carlo simulations are of course needed to obtain numerical approximations to F_n^*.

Results, similar to our Theorems 2 and 3, were obtained for the simpler case of smooth functions of Studentized sample means by Babu and Singh (1983, 1984). For the important special case of the Student t-statistic these authors proved (1.21), provided F is continuous and $EX_1^6 < \infty$. If we take $h(x, y) = \frac{1}{2}(x + y)$ in Theorem 3, we obtain the same result, requiring only that F is nonlattice and $E|X_1|^{4+\varepsilon} < \infty$, for some $\varepsilon > 0$. In addition, we extend the results of Babu and Singh (1983, 1984) to an important class of nonlinear statistics, i.e., to Hoeffding's class of U-statistics. This opens a way to obtain a similar result for Studentized statistical functions of a more general type. Such an extension will be considered elsewhere.

It should be noted that, without Studentization, the improved accuracy of order $o(n^{-1/2})$ of the Edgeworth and bootstrap based estimates does not hold true any more. This is a consequence of the fact that the leading terms in the asymptotic expansions for the exact d.f. of $n^{1/2}(U_n - \theta)$ and the corresponding bootstrap approximation [i.e., the conditional d.f. of $n^{1/2}(U_n^* - \theta_n)$] are no longer identical, but are respectively equal to $\Phi(x2^{-1}\sigma_g^{-1})$ and $\Phi(xs_n^{-1})$, which differ typically by an amount of order $n^{-1/2}$ in probability. The interesting phenomenon that Studentization enables us to obtain more accurate bootstrap estimates for the d.f. of a statistical function is also discussed in Babu and Singh (1984) [see also Hartigan (1986)].

Next we indicate very briefly an important application of our results to the problem of obtaining *better confidence intervals* than the classical jackknife confidence intervals based on the normal approximation, by employing Edgeworth and bootstrap based approximations.

We wish to establish confidence intervals for the mean $\theta = Eh(X_1, X_2)$ of a U-statistic. Let $u_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. The normal approximation yields an approximate two-sided confidence interval

$$(1.22)\qquad \left(U_n - S_n n^{-1/2} u_{\alpha/2}, U_n + S_n n^{-1/2} u_{\alpha/2}\right)$$

for θ. Though the difference between true and nominal confidence level is of order $o(n^{-1/2})$, the upper and lower confidence limits in (1.22) have error rates equal to $\alpha/2 + O(n^{-1/2})$. Thus, in the case of two-sided normal based confidence intervals of the form (1.22), we find a coverage probability $1 - \alpha + o(n^{-1/2})$, while for the corresponding one-sided intervals we obtain a coverage probability $1 - \alpha/2 + O(n^{-1/2})$. The reason behind this is that it is easily checked from (1.6) that the skewness terms of order $n^{-1/2}$ in an asymptotic expansion for the coverage probability cancel in the two-sided case, but give rise to an error term of order $n^{-1/2}$ in the coverage probability for one-sided intervals. A clear exposition of this issue was recently given by Hall and Singh in their contributions to the discussion of a paper by Wu (1986) on resampling methods in regression models.

Improved confidence intervals for θ can be obtained by using either the estimated Edgeworth expansion $\tilde{E}_n$ [cf. (1.15)] or the bootstrap approximation F_n^* [cf. (1.19)]. Inverting $\tilde{E}_n$ yields an Edgeworth based confidence interval for θ given by

$$(1.23)\qquad \left(U_n - S_n n^{-1/2}\hat{c}_{nE,\alpha/2}-, U_n + S_n n^{-1/2}\hat{c}_{nE,\alpha/2}+\right),$$

where

$$(1.24)\quad \hat{c}_{nE,\alpha/2}\pm = u_{\alpha/2} \pm 6^{-1}n^{-1/2}\left\{u_{\alpha/2}^2(2a_n + 3b_n) + (a_n + 3b_n)\right\}$$

with a_n and b_n as in (1.13) and (1.14).

Similarly, a bootstrap based confidence interval for θ is given by

$$(1.25)\qquad \left(U_n - n^{-1/2}S_n C_{nB,1-\alpha/2}^*, U_n - n^{-1/2}S_n C_{nB,\alpha/2}^*\right),$$

where $C_{nB,\alpha/2}^*$ and $C_{nB,1-\alpha/2}^*$ denote the $(\alpha/2)$th and $(1 - \alpha/2)$th percentile of the (simulated) bootstrap approximation F_n^*. Though, asymptotically, the lengths of each of the three intervals (1.22), (1.23) and (1.25) are the same, the Edgeworth and bootstrap based intervals (1.23) and (1.25) are more accurate than the usual normal based jackknife confidence interval (1.22) in the sense that not only the error in the coverage probability for these corrected two-sided intervals is of a lower order than $n^{-1/2}$, but also the upper and lower confidence limits in (1.23) and (1.25) have error rates equal to $\alpha/2 + o(n^{-1/2})$. Accordingly the intervals (1.23) and (1.25) are asymmetric around the point estimate U_n of θ, in contrast with the symmetric interval (1.22). In this way, the asymmetry present in F_n is reflected in our improved interval estimates for θ. We note in passing that the one-sided Edgeworth based intervals suggested by Beran (1984), page 103, do *not* have the desirable property of having error rates equal to $\alpha/2 + o(n^{-1/2})$. This is due to the fact that no Studentization is employed.

Our results can be viewed as a mathematical contribution to the asymptotic distribution theory for bootstrapping Studentized U-statistics. In a way the only thing we do is prove for U-statistics what statisticians expect to be true about bootstrapping in nice asymptotically normal situations.

To conclude this section, we remark that improved confidence intervals of the form (1.23) or (1.25) are also discussed in Hinkley and Wei (1984) for a large class of Studentized statistical functions. However, these authors use formal expansions only to arrive at their Edgeworth and bootstrap based confidence intervals, whereas in the present paper such improved interval estimates are derived rigorously for the case of Studentized U-statistics of degree 2.

Second-order correct bootstrap confidence intervals for a real-valued parameter θ based on maximum likelihood estimators in a parametric framework are also considered by Efron (1987), but his approach is of a different flavor. We also refer to Hall (1988) where a detailed higher-order comparison is made of various types of bootstrap confidence intervals for real-valued parameter θ based on statistics which are smooth functions of sums of i.i.d. random vectors.

2. Proof of Theorem 1. We begin by writing

$$n^{1/2}S_n^{-1}(U_n - \theta) = 2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)2\sigma_g S_n^{-1}, \tag{2.1}$$

where

$$2\sigma_g S_n^{-1} = 1 - \tfrac{1}{8}\sigma_g^{-2}n^{-1}\sum_{i=1}^{n} f(X_i) + R_n \tag{2.2}$$

with

$$\begin{aligned} f(x) &= 4\left(g^2(x) - \sigma_g^2\right) \\ &\quad + 8\int_{-\infty}^{\infty} g(y)\big(h(x,y) - \theta - g(x) - g(y)\big)\,dF(y) \end{aligned} \tag{2.3}$$

for real x, and R_n is a remainder term, satisfying

$$P\left(\left\{|R_n| \ge n^{-1/2}(\log n)^{-1}\right\}\right) = o(n^{-1/2}) \quad \text{as } n \to \infty. \tag{2.4}$$

To establish (2.1)–(2.4) we inspect the proof given by Callaert and Veraverbeke (1981) of their relation (A.10) [which is precisely (2.2)–(2.4) with $o(n^{-1/2})$

replaced by $O(n^{-1/2})$] to find that (2.2)–(2.4) is true under the assumptions $\sigma_g^2 > 0$ and $E|h(X_1, X_2)|^{4+\varepsilon} < \infty$, for some $\varepsilon > 0$. Recall that $\sigma_g^2 > 0$ is implied by assumption (1.8).

Define

$$V_n = 2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)\left(1 - \tfrac{1}{8}\sigma_g^{-2}n^{-1}\sum_{i=1}^{n} f(X_i)\right) \tag{2.5}$$

and let

$$G_n(x) = P(V_n \le x) \quad \text{for } -\infty < x < \infty. \tag{2.6}$$

A simple argument involving (2.4) now yields:

$$\begin{aligned} &P\left(\left\{\left|2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)R_n\right| \ge n^{-1/2}(\log n)^{-1/2}\right\}\right) \\ &\quad \le P\left(\left\{|R_n| \ge n^{-1/2}(\log n)^{-1}\right\}\right) \\ &\qquad + P\left(\left\{\left|2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)\right| \ge (\log n)^{1/2}\right\}\right) \\ &\quad = o(n^{-1/2}) + P\left(\left\{\left|2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)\right| \ge (\log n)^{1/2}\right\}\right). \end{aligned} \tag{2.7}$$

Application of the theorem of Malevich and Abdalimov (1979) directly gives us:

$$P\left(\left\{\left|2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)\right| \ge (\log n)^{1/2}\right\}\right) = o(n^{-1/2}), \tag{2.8}$$

provided $\sigma_g^2 > 0$ and $E|h(X_1, X_2)|^{3+\varepsilon} < \infty$, for some $\varepsilon > 0$.

Together the relations (2.7) and (2.8) imply that

$$P\left(\left\{\left|2^{-1}\sigma_g^{-1}n^{1/2}(U_n - \theta)R_n\right| \ge n^{-1/2}(\log n)^{-1/2}\right\}\right) = o(n^{-1/2}). \tag{2.9}$$

In view of the preceding argument it remains to prove that

$$\sup_x \left|G_n(x) - \tilde{F}_n(x)\right| = o(n^{-1/2}) \quad \text{as } n \to \infty, \tag{2.10}$$

i.e., we must prove (1.9) with F_n replaced by G_n. To prove (2.10) we remark that [cf. Helmers (1985)]

$$V_n = V_{n1} + V_{n2}, \tag{2.11}$$

where $2\sigma_g n^{-1/2}V_{n1} + \theta$ is a U-statistic with varying kernel h_n of the form $h_n = \alpha + n^{-1}\beta$, where α and β are given by

$$\alpha(x,y) = h(x,y) - \tfrac{1}{8}\sigma_g^{-2}(g(x)f(y) + g(y)f(x)) \tag{2.12}$$

and

$$\begin{aligned} \beta(x,y) = {}& -\tfrac{1}{8}\sigma_g^{-2}((h(x,y) - \theta)(f(x) + f(y)) \\ & -2(g(x)f(y) + g(y)f(x)) - 2\mu), \end{aligned} \tag{2.13}$$

where $\mu = \int g(x) f(x)\, dF(x)$, with f given by (2.3). It is easily verified that V_{n2} can be written as [cf. Callaert and Veraverbeke (1981), where this quantity is denoted by $EZ_{n1} + Z_{n3}$]:

$$
\begin{aligned}
V_{n2} = {} & -\tfrac{1}{8}\sigma_g^{-3} n^{-1/2} E(g(X_1) f(X_1)) \\
& -\tfrac{1}{16}\sigma_g^{-3}(n-2) n^{-3/2} \sum_{i=1}^{n} f(X_i)\left[\binom{n-1}{2}^{-1} \sum_{j<k}^{(i)} \psi(X_j, X_k)\right],
\end{aligned} \tag{2.14}
$$

where the function ψ is given by

$$\psi(x,y) = h(x,y) - \theta - g(x) - g(y) \tag{2.15}$$

for real x and y and $\sum_{j<k}^{(i)}$ denotes $\sum_{1 \le j < k \le n,\ j \ne i,\ k \ne i}$.

Callaert and Veraverbeke (1981) proved that the second moment of the second term on the r.h.s. of (2.14) is $O(n^{-2})$, using only $Eh^4(X_1, X_2) < \infty$. It follows directly that

$$P\left(\{|V_{n2} - EV_{n2}| \ge n^{-1/2}(\log n)^{-1}\}\right) = O\left(n^{-1}(\log n)^2\right) \tag{2.16}$$

so that we can replace, for our purposes, V_n by $V_{n1} + EV_{n2}$.

Note that EV_{n2} is a nonrandom term of the critical order $n^{-1/2}$. By an argument like (2.7)–(2.9) we easily verify that it suffices now to prove

$$\sup_x |H_n(x) - \tilde{F}_n(x)| = o(n^{-1/2}) \quad \text{as } n \to \infty, \tag{2.17}$$

where

$$H_n(x) = P(\{V_{n1} + EV_{n2} \le x\}) \tag{2.18}$$

for real x and $n \ge 2$, instead of proving (2.10). Note that

$$
\begin{aligned}
V_{n1} + EV_{n2} = {} & -2^{-1}\sigma_g^{-1} n^{1/2}\binom{n}{2}^{-1} \\
& \times \sum\sum_{1 \le i < j \le n} \{\alpha(X_i, X_j) - \theta + n^{-1}\beta(X_i, X_j)\} \\
& -\tfrac{1}{8}\sigma_g^{-3} n^{-1/2} E[g(X_1) f(X_1)],
\end{aligned} \tag{2.19}
$$

where

$$E\alpha(X_1, X_2) = \theta, \qquad E\beta(X_1, X_2) = 0 \tag{2.20}$$

and

$$Eg(X_1) f(X_1) = 4Eg^3(X_1) + 8Eg(X_1) g(X_2) h(X_1, X_2), \tag{2.21}$$

where we have used (2.3). Clearly, V_{n1} is a suitably normalized U-statistic of degree 2 with kernel $\alpha + n^{-1}\beta$ and $EV_{n2} = O(n^{-1/2})$.

In view of the result of Bickel, Götze and Van Zwet (1986) (see their Theorem 1.2) [cf. also Callaert, Janssen and Veraverbeke (1980)] we easily deduce from (2.18)–(2.21) that

$$\begin{aligned} H_n(x) &= P(\{V_{n1} + EV_{n2} \le x\}) \\ &= P(\{V_{n1} \le x - EV_{n2}\}) \\ &= P\Big(\big\{V_{n1} \le x + \tfrac{1}{8}\sigma_g^{-3} n^{-1/2}\big(4Eg^3(X_1) + 8Eg(X_1)g(X_2)h(X_1, X_2)\big)\big\}\Big) \\ &= \Phi(x) \\ &\quad + \tfrac{1}{6} n^{-1/2}\sigma_g^{-3}\phi(x)\big(Eg^3(X_1) + 3Eg(X_1)g(X_2)\alpha(X_1, X_2)\big)(1 - x^2) \\ &\quad + \tfrac{1}{6} n^{-1/2}\sigma_g^{-3}\phi(x)\big(3Eg^3(X_1) + 6Eg(X_1)g(X_2)h(X_1, X_2)\big) \\ &\quad + o(n^{-1/2}), \end{aligned} \tag{2.22}$$

where we have used the assumptions (1.7) and (1.8) to validate the application of Theorem 1.2 of Bickel, Götze and Van Zwet (1986). In addition, we have employed the fact that under the (weak) moment assumptions of Theorem 1 the term in (2.19) involving $n^{-1}\beta$ is negligible for our purposes. This can be achieved by an analysis closely resembling the proof of Theorem 4.1 of Helmers and Van Zwet (1982). A simple calculation yields

$$\begin{aligned} &3Eg(X_1)g(X_2)\alpha(X_1, X_2) \\ &\qquad = -3Eg^3(X_1) - 3Eg(X_1)g(X_2)h(X_1, X_2). \end{aligned} \tag{2.23}$$

Combining now (2.22) and (2.23), we easily check (2.17) and the proof of Theorem 1 is complete. □

3. Proof of Theorem 2. In view of Theorem 1 it suffices clearly to show that, with probability 1,

$$E^* g_n^k(X_1^*) \to Eg^k(X_1) \quad \text{for } k = 2, 3 \tag{3.1}$$

and

$$E^* g_n(X_1^*) g_n(X_2^*) h(X_1^*, X_2^*) \to Eg(X_1)g(X_2)h(X_1, X_2). \tag{3.2}$$

We first prove (3.1) for $k = 2$. A simple calculation yields that [cf. (1.11)]

$$\begin{aligned} E^* g_n^2(X_1^*) &= E^*\left[n^{-1}\sum_{j=1}^{n} h(X_1^*, X_j) - \theta_n\right]^2 \\ &= n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h(X_i, X_j)h(X_i, X_k) \\ &\quad - n^{-4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} h(X_i, X_j)h(X_k, X_l). \end{aligned} \tag{3.3}$$

To proceed we note that the first term on the r.h.s. of (3.3) can be written as

$$\begin{aligned} n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h(X_i, X_j)h(X_i, X_k) &= \theta^2 + n^{-1}\sum_{i=1}^{n} g^2(X_i) + 3n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} g(X_i)g(X_j) \\ &\quad + 2\theta n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(g(X_i) + g(X_j) + \psi(X_i, X_j)\big) \\ &\quad + 2n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} g(X_i)\psi(X_i, X_j) \\ &\quad + 2n^{-1}\sum_{i=1}^{n} g(X_i)n^{-2}\sum_{j=1}^{n}\sum_{k=1}^{n}\psi(X_j, X_k) \\ &\quad + n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\psi(X_i, X_j)\psi(X_i, X_k), \end{aligned} \tag{3.4}$$

where the functions g and ψ are given in (1.3) and (2.15) and $\theta = Eh(X_1, X_2)$. With the aid of the SLLN and the easily verified fact that the last five terms on the r.h.s. of (3.4) $\to 0$ a.s. as $n \to \infty$, by the moment assumptions of Theorem 2 and some well-known arguments involving conditional expectations, we find that

$$n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h(X_i, X_j)h(X_i, X_k) \to \theta^2 + Eg^2(X_1) \quad \text{a.s.} \tag{3.5}$$

as $n \to \infty$. Similarly, we also find for the second term on the r.h.s. of (3.3):

$$n^{-4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} h(X_i, X_j)h(X_l, X_k) \to \theta^2 \quad \text{a.s. as } n \to \infty. \tag{3.6}$$

Together (3.3)–(3.6) yields (3.1) for the case $k = 2$. The proof of (3.1) for $k = 3$ is similar and therefore omitted.

It remains to establish (3.2). An argument like (3.2)–(3.6) yields

$$\begin{aligned} &E^*g_n(X_1^*)g_n(X_2^*)h(X_1^*, X_2^*) \\ &\quad = n^{-4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} h(X_i, X_k)h(X_j, X_l)h(X_i, X_j) \\ &\qquad - 2n^{-5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} h(X_i, X_k)h(X_l, X_m)h(X_i, X_j) \\ &\qquad + n^{-6}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{p=1}^{n} h(X_k, X_l)h(X_m, X_p)h(X_i, X_j) \\ &\quad \to Eg(X_1)g(X_2)h(X_1, X_2) \quad \text{a.s. as } n \to \infty. \end{aligned} \tag{3.7}$$

This completes the proof of Theorem 2. □

4. Proof of Theorem 3. To prove Theorem 3, we proceed in a number of steps.

To begin with, we shall show that the arguments leading to (2.10) in Section 2 can be repeated to establish a parallel result for the bootstrapped quantities corresponding to the Studentized U-statistics $n^{1/2}S_n^{-1}(U_n - \theta)$ and its approximand V_n [cf. (2.5)]. Let $n^{1/2}S_n^{*-1}(U_n^* - \theta_n)$ be the bootstrapped Studentized U-statistic [cf. (1.20)], and let

$$V_n^* = 2^{-1}\sigma_{g_n}^{-1}n^{1/2}\big(U_n^* - \theta_n\big)\left(1 - \tfrac{1}{8}\sigma_{g_n}^{*2}n^{-1}\sum_{i=1}^{n} f_n\big(X_i^*\big)\right), \tag{4.1}$$

where

$$\sigma_{g_n}^{2*} = E^*g_n^2\big(X_1^*\big) \tag{4.2}$$

with g_n given by (1.13) and [cf. (2.3)]

$$\begin{aligned} f_n(x) &= 4\big(g_n(x) - \sigma_{g_n}^2\big) \\ &\quad + 8\int_{-\infty}^{\infty} g_n(y)(h(x,y) - \theta_n - g_n(x) - g_n(y))\,d\hat{F}_n(y) \quad \text{for real } x. \end{aligned} \tag{4.3}$$

Recall that $\hat{F}_n$ is the empirical d.f. based on $X_1, \dots, X_n$. Define

$$G_n^*(x) = P^*(V_n^* \le x) \tag{4.4}$$

for $-\infty < x < \infty$ and $n \ge 2$. Analogous to (2.10) we must now show that

$$\sup_x \big|G_n^*(x) - F_n^*(x)\big| = o(n^{-1/2}) \quad \text{a.s.} \tag{4.5}$$

with F_n^* as in (1.21). To check (4.5), we simply follow the argument leading to the parallel result (2.10), to find that (4.5) holds, provided

$$\begin{aligned} E^*\big|h\big(X_1^*, X_2^*\big)\big|^{4+\varepsilon} &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big|h(X_i, X_j)\big|^{4+\varepsilon} \\ &= 2n^{-2}\sum\sum_{1\le i<j\le n}\big|h(X_i, X_j)\big|^{4+\varepsilon} + n^{-2}\sum_{i=1}^{n}\big|h(X_i, X_i)\big|^{4+\varepsilon} \\ &< \infty \quad \text{a.s.} \end{aligned} \tag{4.6}$$

This is a direct consequence of the SLLN for U-statistics and the Marcinkievitz–Zygmund SLLN for sums of i.i.d. r.v.'s using the moment requirements $E|h(X_1, X_2)|^{4+\varepsilon} < \infty$ and $E|h(X_1, X_1)|^{2+\varepsilon/2} < \infty$, for some $\varepsilon > 0$.

Also note that

$$\begin{aligned} \sigma_{g_n}^{2*} &= E^*g_n^2\big(X_1^*\big) = E^*h\big(X_1^*, X_2^*\big)h\big(X_1^*, X_3^*\big) \\ &= n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h\big(X_i, X_j\big)h\big(X_i, X_k\big) \\ &= n^{-1}\sum_{i=1}^{n} g^2(X_i)(1 + o(1)) \to \sigma_g^2 \quad \text{a.s. as } n \to \infty, \end{aligned} \tag{4.7}$$

by a simple calculation, similar to the one given in Section 3, using the moment assumptions of Theorem 3 and Kolmogorov's strong law. Together these results easily yield (4.5) by following the argument leading to (2.10).

It remains to establish

$$\sup_x \left| P^*(V_n^* \le x) - \tilde{F}_n(x) \right| = o(n^{-1/2}) \quad \text{a.s.} \tag{4.8}$$

with $\tilde{F}_n$ as in (1.6). To prove this, we begin by noting that [cf. (2.11)]

$$V_n^* = V_{n1}^* + V_{n2}^*, \tag{4.9}$$

where $2\sigma_{g_n}^* n^{-1/2} V_{n1}^* + \theta_n$ is a U-statistic with varying kernel $h_n^{(n)}$ of the form $h_n^{(n)} = \alpha_n + n^{-1}\beta_n$, where α_n and β_n are given by (2.12) and (2.13), with g, f, θ and μ replaced by g_n, f_n, θ_n and μ_n, where

$$\mu_n = \int g_n(x) f_n(x)\, d\hat{F}_n(x) = n^{-1} \sum_{i=1} g_n(X_i) f_n(X_i).$$

Note that V_{n2}^* is obtained from V_{n2} by replacing f and g by f_n and g_n. The function ψ [cf. (2.15)] should be replaced by ψ_n, which is given by

$$\psi_n(x, y) = h(x, y) - \theta_n - g_n(x) - g_n(y) \tag{4.10}$$

for real x and y. By an argument like the one given in (2.16) we easily check that we can replace, for our purpose, V_n^* by $V_{n1}^* + E^* V_{n2}^*$. The assumptions $Eh^4(X_1, X_2) < \infty$ and $Eh^2(X_1, X_1) < \infty$ are needed to establish the result corresponding to (2.16). We can conclude, similarly as in (2.17), that it suffices now to establish

$$\sup_x \left| H_n^*(x) - \tilde{F}_n(x) \right| = o(n^{-1/2}) \quad \text{a.s.}, \tag{4.11}$$

where

$$H_n^*(x) = P^*\left(\left\{V_{n1}^* + E^* V_{n2}^* \le x\right\}\right) \tag{4.12}$$

for real x and $n \ge 2$, instead of proving (4.8). Note that [cf. (2.19)]

$$\begin{aligned} V_{n1}^* + E^* V_{n2}^* = {} & -2^{-1} \sigma_{g_n}^{-1*} n^{1/2} \binom{n}{2}^{-1} \\ & \times \sum\sum_{1 \le i < j \le n} \left\{ \alpha_n\left(X_i^*, X_j^*\right) - \theta_n + n^{-1} \beta_n\left(X_i^*, X_j^*\right) \right\} \\ & - \tfrac{1}{8} \sigma_{g_n}^{-3*} n^{-1/2} E^*\left[g_n\left(X_1^*\right) f_n\left(X_1^*\right) \right], \end{aligned} \tag{4.13}$$

where

$$E^* \alpha_n\left(X_1^*, X_2^*\right) = \theta_n \tag{4.14}$$

and

$$E^* \beta_n\left(X_1^*, X_2^*\right) = 0 \tag{4.15}$$

and [cf. (2.21), (3.1) and (3.2)], as $n \to \infty$,

$$
\begin{aligned}
&E^* g_n(X_1^*) f_n(X_1^*) \\
(4.16) \qquad &= 4E^* g_n^3(X_1^*) + 8E^* g_n(X_1^*) g_n(X_2^*) h(X_1^*, X_2^*) \\
&\to 4Eg^3(X_1) + 8Eg(X_1)g(X_2)h(X_1, X_2)
\end{aligned}
$$

with probability 1. Note that V_{n1}^* is a suitable normalized U-statistic of degree 2 with kernel $\alpha_n + n^{-1}\beta_n$, based on the X_i^*'s, $1 \le i \le n$, and $E^* V_{n2}^* = O(n^{-1/2})$ a.s. We can now simply repeat the calculations given in (2.22) and (2.23), to find that (4.11) [cf. (2.17)] holds true, provided the assumptions of Theorem 3 remain valid, if we replace E, X_1 and X_2 by E^*, X_1^* and X_2^* and g by g_n. Since the resulting moment assumptions $E^*|h(X_1^*, X_2^*)|^{4+\varepsilon} < \infty$, for some $\varepsilon > 0$, and $E^*|h(X_1^*, X_1^*)|^3 < \infty$, are already shown to be satisfied a.s., it remains to prove that, with probability 1,

$$(4.17) \qquad \text{the d.f. of } g_n(X_1^*) \text{ is nonlattice}$$

for all sufficiently large n. To check (4.17), we note first that, because of assumption (1.8), it suffices to show that, for any fixed $a > 0$,

$$(4.18) \qquad \Delta_n = \sup_{|t| \le a} \left| E^* e^{itg_n(X_i^*)} - Ee^{itg(X_1)} \right| \to 0 \quad \text{a.s.}$$

as $n \to \infty$. To see this we begin by remarking that

$$(4.19) \quad E^* e^{itg_n(X_1^*)} = n^{-1} \sum_{i=1}^n e^{itg_n(X_i)} = n^{-1} \sum_{i=1}^n e^{\,it(n^{-1}\sum_{j=1}^n h(X_i, X_j) - \theta_n)}$$

so that

$$
\begin{aligned}
\Delta_n &\le \sup_{|t| \le a} \left| n^{-1} \sum_{i=1}^n \left\{ e^{it(n^{-1}\sum_{j=1}^n h(X_i, X_j) - \theta_n)} - e^{itg(X_1)} \right\} \right| \\
(4.20) \qquad &\quad + \sup_{|t| \le a} \left| n^{-1} \sum_{i=1}^n e^{itg(X_1)} - Ee^{itg(X_1)} \right| \\
&= \Delta_{n1} + \Delta_{n2}.
\end{aligned}
$$

Because $|e^{ix} - e^{iy}| \le |x - y|$ we get

$$
\begin{aligned}
\Delta_{n1} &\le an^{-1} \sum_{i=1}^n \left| n^{-1} \sum_{j=1}^n h(X_i, X_j) - \theta_n - g(X_i) \right| \\
(4.21) \quad &\le an^{-1} \sum_{i=1}^n \left| n^{-1} \sum_{j=1}^n \{ g(X_i) + g(X_j) + \psi(X_i, X_j) + \theta \} - \theta_n - g(X_i) \right| \\
&\le a \left| n^{-1} \sum_{j=1}^n g(X_j) \right| + an^{-1} \sum_{i=1}^n \left| n^{-1} \sum_{j=1}^n \psi(X_i, X_j) \right| + a|\theta_n - \theta|.
\end{aligned}
$$

Now $n^{-1}\sum_{j=1}^{n} g(X_j) \to 0$ a.s. as $n \to \infty$, by the strong law, and similarly, $\theta_n \to \theta$ a.s. by the SLLN for U-statistics and the strong law. To show finally that

$$n^{-1} \sum_{i=1}^{n} \left| n^{-1} \sum_{j=1}^{n} \psi(X_i, X_j) \right| \to 0 \quad \text{a.s.},$$

we note first that $n^{-2}\sum_{i=1}^{n} \psi(X_i, X_i) \to 0$ a.s., again by the strong law, whereas

$$n^{-1} \sum_{i=1}^{n} \left| n^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \psi(X_i, X_j) \right| \to 0 \quad \text{a.s.} \tag{4.22}$$

because of Lemma 5 on page 157 of Dehling, Denker and Philipp (1984). In the latter paper it is shown that, for any fixed i,

$$\left| n^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \psi(X_i, X_j) \right| \to 0 \quad \text{a.s.}, \tag{4.23}$$

provided $E\psi^2(X_1, X_2)\log^2 \psi(X_1, X_2) < \infty$, which directly yields (4.22). We note in passing that the latter moment assumption may be relaxed [cf. Dehling (1989)]. Thus we have proved that $\Delta_{n1} \to 0$ a.s. as $n \to \infty$.

It remains to show that $\Delta_{n2} \to 0$ a.s. as $n \to \infty$. This is a direct consequence of a theorem of Feuerverger and Mureika (1977). This completes the proof of Theorem 3. □

Acknowledgments. I would like to thank P. Janssen, the editor W. R. van Zwet, an associate editor and a referee for their valuable comments.

REFERENCES

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

ATHREYA, K. B. GHOSH, M. LOW, L. Y. and SEN, P. K. (1984). Laws of large numbers for bootstrapped U-statistics. *J. Statist. Plann. Inference* **9** 185–194.

BABU, G. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BABU, G. and SINGH, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā, Ser. A* **46** 219–232.

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

BERAN, R. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.* **12** 101–118.

BICKEL, P. J. and FREEDMAN, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1986). The Edgeworth expansion for U-statistics of degree 2. *Ann. Statist.* **14** 1463–1484.

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

CALLAERT, H. and VERAVERBEKE, N. (1981). The order of the normal approximation to a Studentized U-statistics. *Ann. Statist.* **9** 194–200.

DEHLING, H. (1989). Complete convergence of triangular arrays and the law of the iterated logarithm for U-statistics. *Statist. Probab. Lett.* **7** 319–321.

DEHLING, H. DENKER, M. and PHILIPP, W. (1984). Invariance principles for von Mises and U-statistics. *Z. Wahrsch. Verw. Gebiete* **67** 139–167.

EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82** 171–200.

FEUERVERGER, A. and MUREIKA, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5** 88–97.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16** 927–981.

HARTIGAN, J. A. (1986). Comment on "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy" by B. Efron and R. J. Tibshirani. *Statist. Sci.* **1** 75–77.

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

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

HINKLEY, D. and WEI, B. C. (1984). Improvements of jackknife confidence limit methods. *Biometrika* **71** 331–119.

MALEVICH, T. L. and ABDALIMOV, B. (1979). Large deviation probabilities for U-statistics. *Theory Probab. Appl.* **24** 215–219.

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

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

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

CENTRE FOR MATHEMATICS AND COMPUTER SCIENCE
P.O. BOX 4079
1009 AB AMSTERDAM
THE NETHERLANDS