Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R. Helmers

# On the Edgeworth expansion and the bootstrap approximation for a studentized $U$-statistic

Department of Mathematical Statistics Report MS-R8708 July

# On the Edgeworth Expansion and the Bootstrap Approximation for a Studentized $U$-statistic

R. Helmers
*Centre for Mathematics and Computer Science*
*P.O. Box 4079, 1009 AB Amsterdam, The Netherlands*

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 non-lattice 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. Extensions to Studentized statistical functions admitting an second order von Mises expansion, such as Studentized $L$-statistics with smooth weights and Studentized $M$-estimators of maximum likelihood type, are also briefly discussed.

*1980 Mathematics Subject Classification:* Primary: 62E20, 62G05. Secondary: 60F05.
*Key Words & Phrases:* Edgeworth expansions, bootstrap approximations, studentized U-statistics, bootstrap confidence intervals, Edgeworth based confidence intervals, studentized L-statistics, studentized M-estimators.

## 1. Introduction and main results

Let $X_1,X_2,\cdots,X_n$ be independent and identically distributed (i.i.d.) random variables (r.v.) with common distribution function (df) $F$. 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

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

and suppose that

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

has a positive variance $\sigma_g^2$. 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 $n\geq 2$ and real $x$,

$$F_n(x)=P(\{n^{\frac{1}{2}}s_n^{-1}(U_n-\theta)\le x\}) \tag{1.5}$$

It is well-known that $F_n$ converges in distribution to the standard normal df $\Phi$, as $n\to\infty$, provided $Eh^2(X_1,X_2)<\infty$ and $\sigma_g^2>0$ (cf. Arvesen (1969)). The speed of this convergence to normality is of the classical order $n^{-\frac{1}{2}}$ (cf Callaert & Veraverbeke (1981), Zhao Lincheng (1983), Helmers (1985)).

Report MS-R8708
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

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

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

THEOREM 1. *Suppose that*

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

*and*

$$\text{the df of } g(X_1) \text{ is non}-\text{lattice} \tag{1.8}$$

*then, as* $n \to \infty$,

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

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 a recent 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 semi-parametric framework, $F$ is completely unknown, and one does not know the quantities

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

appearing in the expansion (1.6). These moments depend on the underlying df $F$ and must be estimated from the observations $X_1, \cdots, X_n$. 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 df based on $X_1, \cdots, X_n$. Conditionally given $X_1, \cdots, X_n$, let $X_1^*, \cdots, X_n^*$ be $n$ independent r.v.'s with common df $\hat{F}_n$, 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 given by

$$a_n=\frac{E^*g_n^3(X_1^*)}{(E^*g_n^2(X_1^*))^{\frac{3}{2}}} \tag{1.11}$$

and

$$b_n=\frac{E^*g_n(X_1^*)g_n(X_2^*)h(X_1^*,X_2^*)}{(E^*g_n^2(X_1^*))^{\frac{3}{2}}} \tag{1.12}$$

where

$$g_n(X_i^*)=E^*[h(X_1^*,X_2^*)-\theta_n|X_i^*] \tag{1.13}$$

for $i=1,2$, and

$$\theta_n=E^*h(X_1^*,X_2^*)=n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}h(X_i,X_j) \tag{1.14}$$

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

$$a_n = a_n(X_1, \cdots, X_n) = \frac{n^{-1}\sum_{i=1}^{n}(n^{-1}\sum_{j=1}^{n} h(X_i,X_j) - \theta_n)^3}{(n^{-1}\sum_{i=1}^{n}(n^{-1}\sum_{j=1}^{n} h(X_i,X_j) - \theta_n)^2)^{\frac{3}{2}}} \tag{1.15}$$

and

$$b_n = b_n(X_1, \cdots, X_n) = \tag{1.16}$$

$$= \frac{n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(n^{-1}\sum_{k=1}^{n} h(X_i,X_k) - \theta_n)\,(n^{-1}\sum_{l=1}^{n} h(X_j,X_l) - \theta_n) h(X_i,X_j)}{(n^{-1}\sum_{i=1}^{n}(n^{-1}\sum_{j=1}^{n} h(X_i,X_j) - \theta_n)^2)^{\frac{3}{2}}}.$$

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 $n \geqslant 2$ and real $x$,

$$\tilde{E}_n(x) = \Phi(x) + 6^{-1} n^{-\frac{1}{2}} \phi(x)\{(2x^2+1)a_n + 3(x^2+1)b_n\} \tag{1.17}$$

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 $X_1, \cdots X_n$.

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

$$E|h(X_1,X_1)|^3 < \infty \tag{1.18}$$

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

$$\sup_x |F_n(x) - \tilde{E}_n(x)| = o(n^{-\frac{1}{2}}) \tag{1.19}$$

Theorem 2 tells us that the *Edgeworth expansion estimate* $\tilde{E}_n$ is asymptotically closer to the exact df $F_n$ than the classical normal approximation $\Phi$. In a way $\tilde{E}_n$ adapts itself to the possible asymmetry present in the exact df $F_n$; 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 $n^{\frac{1}{2}} S_n^{-1}(U_n - \theta)$, based on the bootstrap sample $X_1^*, \cdots, X_n^*$, which is given by

$$n^{\frac{1}{2}} S_n^{*-1}(U_n^* - \theta_n) \tag{1.20}$$

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 the formula's (1.2) and (1.4); the parameter $\theta$ (cf. (1.1)) is replaced by its natural estimator $\theta_n$ (cf. (1.14)). The bootstrap approximation

$$F_n^*(x) = P^*(n^{\frac{1}{2}} S_n^{*^{-1}}(U_n^* - \theta_n) \leqslant x) \tag{1.21}$$

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

ATHREYA et. al. (1984) recently showed that

$$\sup_x |F_n(x) - F_n^*(x)| \to 0, \text{ as } n \to \infty, \tag{1.22}$$

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* $n \to \infty$,

$$\sup_x |F_n(x) - F_n^*(x)| = o(n^{-\frac{1}{2}}) \tag{1.23}$$

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 df $F_n$ then the normal approximation $\Phi$. (see BERAN (1982), (1984) for some related results suggesting that $F_n^*$, like $\tilde{E}_n$, should be locally asymptotically minimax among all possible estmates 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 JOGESH BABU and SINGH (1983; 1984). For the important special case of the Student t-statistic these authors proved (1.23), 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 non-lattice and $E|X_1|^{4+\epsilon} < \infty$, for some $\epsilon > 0$. In addition we extend the results of JOGESH BABU and SINGH (1983; 1984) to an important class of non-linear 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. A brief sketch of such an extension is given section 5.

It should be noted that, without studentization, the improved accuracy of order $o(n^{-\frac{1}{2}})$ 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 df of $n^{\frac{1}{2}}(U_n - \theta)$ and the corresponding bootstrap approximation (i.e. the conditional df of $n^{\frac{1}{2}}(U_n^* - \theta_n)$) 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 $n^{-\frac{1}{2}}$ in probability. The interesting phenomenon that Studentization enables us to obtain more accurate bootstrap estimates for the df of a statistical function, is also discussed in JOGESH 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 $u_{\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$. The normal approximation yields an approximate two-sided confidence interval

$$(U_n - s_n n^{-\frac{1}{2}} u_{\frac{\alpha}{2}},\ U_n + s_n n^{-\frac{1}{2}} u_{\frac{\alpha}{2}}) \tag{1.24}$$

for $\theta$. Though, the difference between true and nominal confidence level is of order $o(n^{-\frac{1}{2}})$, the upper and lower confidence limits in (1.24) have error rates equal to $\frac{\alpha}{2} + O(n^{-\frac{1}{2}})$. Thus, in the case of two-sided normal based confidence intervals of the form (1.24), we find an coverage probability $1 - \alpha + o(n^{-\frac{1}{2}})$, while for the corresponding one-sided intervals we obtain an coverage probability $1 - \frac{\alpha}{2} + O(n^{-\frac{1}{2}})$. The reason behind this is that it is easily checked from (1.6) that the skewness terms of order $n^{-\frac{1}{2}}$ in an asymptotic expansion for the coverage probability cancell in the two-sided case, but give rise to an error term of order $n^{-\frac{1}{2}}$ in the coverage probability for one-sided intervals. A clear exposition of this issue was recently given by P. HALL and K. SINGH in their contributions to the discussion of a paper by WU (1986) on resampling methods in regression models.

Improved confidence intervals for $\theta$ can be obtained by using either the estimated Edgeworth expansion $\tilde{E}_n$ (cf. (1.17)) or the bootstrap approximation $F_n^*$ (cf. (1.21)). Inverting $\tilde{E}_n$ yields an

Edgeworth based confidence interval for $\theta$ given by

$$(U_n - s_n n^{-\frac{1}{2}} \hat{c}_{nE,\frac{\alpha}{2}} -, \; U_n + s_n n^{-\frac{1}{2}} \hat{c}_{nE,\frac{\alpha}{2}} +) \tag{1.25}$$

where

$$\hat{c}_{nE,\frac{\alpha}{2}} \pm = u_{\frac{\alpha}{2}} \pm 6^{-1} n^{-\frac{1}{2}} \left\{ u^2_{\frac{\alpha}{2}} (2a_n + 3b_n) + (a_n + 3b_n) \right\} \tag{1.26}$$

with $a_n$ and $b_n$ as in (1.15) and (1.16)

Similarly, a bootstrap based confidence interval for $\theta$ is given by

$$\left[ (U_n - n^{-\frac{1}{2}} s_n C^*_{nB,1-\frac{\alpha}{2}} \; U_n - n^{-\frac{1}{2}} s_n C^*_{nB,\frac{\alpha}{2}} \right] \tag{1.27}$$

where $C^*_{nB,\frac{\alpha}{2}}$ and $C^*_{nB,1-\frac{\alpha}{2}}$ denote the $\frac{\alpha}{2}$th and $(1-\frac{\alpha}{2})$th percentile of the (simulated) bootstrap approximation $F^*_n$. Though, asymptotically, the lengths of each of the three intervals (1.24), (1.25) and (1.27) are the same, the Edgeworth and bootstrap based intervals (1.25) and (1.27) are more accurate than the usual normal based jackknife confidence interval (1.24) in the sense that not only the error in the coverage probability for these corrected two-sided intervals is of a lower order then $n^{-\frac{1}{2}}$, but also the upper and lower confidence limits in (1.25) and (1.27) have error rates equal to $\frac{\alpha}{2} + o(n^{-\frac{1}{2}})$. Accordingly the intervals (1.25) and (1.27) are asymmetric around the point estimate $U_n$ of $\theta$, in contrast with the symmetric interval (1.24). In this way, the asymmetry present in $F_n$ is reflected in our improved interval estimates for $\theta$. 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 $\frac{\alpha}{2} + o(n^{-\frac{1}{2}})$. This is due to the fact that no studentization is employed.

To conclude this section we remark that improved confidence intervals of the form (1.25) or (1.27) 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. Extensions to Studentized statistical functions admitting an second-order von Mises expansion, such as Studentized $L$-statistics with smooth weights and Studentized $M$-estimators of maximum likelihood type, are discussed in section 5.

Second order correct bootstrap confidence intervals for a real-valued parameter $\theta$ based on maximum likelihood estimators in a parametric framework are also considered by EFRON (1987), but his approach is of a different flavour. The asymptotic accuracy of the bootstrap approximation of the distribution of least squares estimators in the context of a linear regression model was recently investigated by NAVIDI (1986).

## 2. PROOF OF THEOREM 1

We begin by writing

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

where

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

with

$$f(x) = 4(g^2(x) - \sigma_g^2) + 8 \int_{-\infty}^{\infty} g(y)(h(x,y) - \theta - g(x) - g(y)) dF(y) \tag{2.3}$$

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

$$P(\{|R_n| \geqslant n^{-\frac{1}{2}}(\log n)^{-1}\}) = o(n^{-\frac{1}{2}}), \text{ as } n \to \infty . \tag{2.4}$$

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

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

and let

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

A simple argument involving (2.4) now yields:

$$\begin{aligned} &P(\{|2^{-1}\sigma_g^{-1} n^{\frac{1}{2}} (U_n - \theta) R_n| \geqslant n^{-\frac{1}{2}} (\log n)^{-\frac{1}{2}}\}) \leqslant \\ &\leqslant P(\{|R_n| \geqslant n^{-\frac{1}{2}} (\log n)^{-1}\}) \\ &+ P(\{|2^{-1}\sigma_g^{-1} n^{\frac{1}{2}} (U_n - \theta)| \geqslant (\log n)^{\frac{1}{2}}\}) \\ &= o(n^{-\frac{1}{2}}) + P(\{|2^{-1}\sigma_g^{-1} n^{\frac{1}{2}} (U_n - \theta)| \geqslant (\log n)^{\frac{1}{2}}\}) . \end{aligned} \tag{2.7}$$

Application of the Theorem of MALEVICH and ABDALIMOV (1979) directly gives us:

$$P(\{\, | \, 2^{-1}\sigma_g^{-1} n^{\frac{1}{2}} (U_n - \theta) \,| \geqslant (\log n)^{\frac{1}{2}}\}) = o(n^{-\frac{1}{2}}) \tag{2.8}$$

provided $\sigma_g^2 > 0$ and $E|h(X_1, X_2)|^{3+\epsilon} < \infty$, for some $\epsilon > 0$.
Together the relations (2.7) and (2.8) imply that

$$P(\{|2^{-1}\sigma_g^{-1} n^{\frac{1}{2}} (U_n - \theta) R_n| \geqslant n^{-\frac{1}{2}} (\log n)^{-\frac{1}{2}}\}) = o(n^{-\frac{1}{2}}) . \tag{2.9}$$

In view of the preceeding argument it remains to prove that

$$\sup_x |G_n(x) - \tilde{F}_n(x)| = o(n^{-\frac{1}{2}}), \textit{ as } n \to \infty , \tag{2.10}$$

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 $2\sigma_g n^{-\frac{1}{2}} V_{n1} + \theta$ is a $U$-statistic with varying kernel $h_n$ of the form $h_n = \alpha + n^{-1}\beta$, where $\alpha$ and $\beta$ 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 $EZ_{n1} + Z_{n3}$):

$$V_{n2} = -\tfrac{1}{8}\sigma_g^{-3} n^{-\frac{1}{2}} E(g(X_1)f(X_1)) + \tag{2.14}$$

$$-\tfrac{1}{16}\sigma_g^{-3}(n-2)n^{-\frac{3}{2}}\sum_{i=1}^{n} f(X_i)\left[\binom{n-1}{2}^{-1}\sum_{j<k}^{(i)}\psi(X_i,X_k)\right]$$

where the function $\psi$ is given by

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

for real $x$ and $y$ and $\sum\limits_{j<k}^{(i)}$ denotes $\sum\limits_{1\leq j<k\leq n,\ j\neq i,\ k\neq i}$.

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

$$P(\{|V_{n2}-EV_{n2}|\geq n^{-\frac{1}{2}}(\log n)\})=\mathcal{O}(n^{-1}(\log n)^2) \tag{2.16}$$

so that we can replace, for our purposes, $V_n$ by $V_{n1}+EV_{n2}$.
Note that $EV_{n2}$ is a non-random term of the critical order $n^{-\frac{1}{2}}$. By an argument like (2.7) - (2.9) we easily verify that it suffices now to prove

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

where

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

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

$$V_{n1}+EV_{n2}=-2^{-1}\sigma_g^{-1}n^{\frac{1}{2}}\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\sum\Big\{\alpha(X_i,X_j)-\theta+ \tag{2.19}$$

$$+n^{-1}\beta(X_i,X_j)\Big\}-\tfrac{1}{8}\sigma_g^{-3}n^{-\frac{1}{2}}E\left[g(X_1)f(X_1)\right]$$

where

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

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 $EV_{n2}=\mathcal{O}(n^{-\frac{1}{2}})$.

In view of the result of BICKEL, GÖTZE and VAN ZWET (1986) (see their Theorem 2.1) (cf. also CALLAERT, JANSSEN and VERAVERBEKE (1980)) we easily deduce from (2.18) - (2.21) that

$$\begin{aligned} H_n(x)&=P(\{V_{n1}+EV_{n2}\leq x\}) \\ &=P(\{V_{n1}\leq x-EV_{n2}\})= \\ &=P(\{V_{n1}\leq x+\tfrac{1}{8}\sigma_g^{-3}n^{-\frac{1}{2}}\left[4Eg^3(X_1)+8Eg(X_1)g(X_2)h(X_1,X_2)\right]\}) \\ &=\Phi(x)+6^{-1}n^{-\frac{1}{2}}\sigma_g^{-3}\phi(x)\left[Eg^3(X_1)+3Eg(X_1)g(X_2)\alpha(X_1,X_2)\right] \\ &(1-x^2)+6^{-1}n^{-\frac{1}{2}}\sigma_g^{-3}\phi(x)\left[3Eg^3(X_1)+6Eg(X_1)g(X_2)h(X_1,X_2)\right] \\ &+o(n^{-\frac{1}{2}}) \end{aligned} \tag{2.22}$$

where we have used the assumption (1.7) and (1.8) to validate the application of Theorem 2.1 of

BICKEL, GÖTZE and VAN ZWET (1986). A simple calculation yields

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

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,

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

and

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

We first prove (3.1) for $k=2$. A simple calculation yields that (cf. (1.14))

$$E^* g_n^2(X_1^*)=E^*\left[n^{-1}\textstyle\sum_{j=1}^n h(X_1^*,X_j)-\theta_n\right]^2= \tag{3.3}$$

$$=n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n h(X_i,X_j)h(X_i,X_k)-$$

$$-n^{-4}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n h(X_i,X_j)h(X_k,X_l)\,.$$

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

$$n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n h(X_i,X_j)h(X_i,X_k)= \tag{3.4}$$

$$=\theta^2+n^{-1}\sum_{i=1}^n g^2(X_i)+3n^{-2}\sum_{i=1}^n\sum_{j=1}^n g(X_i)g(X_j)+$$

$$+2\theta n^{-2}\sum_{i=1}^n\sum_{j=1}^n(g(X_i)+g(X_j)+\psi(X_i,X_j))+$$

$$+2n^{-2}\sum_{i=1}^n\sum_{j=1}^n g(X_i)\psi(X_i,X_j)+2n^{-1}\sum_{i=1}^n g(X_i)\cdot n^{-2}\sum_{j=1}^n\sum_{k=1}^n\psi(X_j,X_k)$$

$$+n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\psi(X_i,X_j)\psi(X_i,X_k)\,,$$

where the function $g$ and $\psi$ 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) $\underset{a.s.}{\rightarrow}0$, as $n\rightarrow\infty$, by the moment assumptions of Theorem 2 and some well-known arguments involving conditional expectations, we find that

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

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

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

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

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

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 results for the bootstrapped quantities corresponding to the Studentized $U$-statistics $n^{\frac{1}{2}} S_n^{-1}(U_n - \theta)$ and its approximand $V_n$ (cf. (2.5)). Let $n^{\frac{1}{2}} S_n^{*-1}(U_n^* - \theta_n)$ be the bootstrapped Studentized $U$-statistic (cf. (1.20)), and let

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

where

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

with $g_n$ given by (1.13) and (cf. (2.3))

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

Recall that $\hat{F}_n$ is the empirical df based on $X_1, \cdots, X_n$.
Define

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

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

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

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

$$E^* |h(X_1^*, X_2^*)|^{4+\epsilon} = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} |h(X_i, X_j)|^{4+\epsilon} = \tag{4.6}$$
$$= 2n^{-2} \sum_{1 \leqslant i < j \leqslant n} \sum |h(X_i, X_j)|^{4+\epsilon} + n^{-2} \sum_{i=1}^{n} |h(X_i, X_i)|^{4+\epsilon}$$
$$< \infty \ \text{a.s.}$$

This 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 $E|h(X_1, X_2)|^{4+\epsilon} < \infty$ and $E|h(X_1, X_1)|^{2+\epsilon/2} < \infty$, for some $\epsilon > 0$.

Also note that

$$\sigma_{g_n^*}^2 = E^* g_n^2(X_1^*) = E^* h(X_1^*,X_2^*)h(X_1^*,X_3^*) =$$

$$= n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h(X_i,X_j)h(X_i,X_k) \tag{4.7}$$

$$= n^{-1}\sum_{i=1}^{n} g^2(X_i)(1+o(1)) \to \sigma_g^2 \text{ a.s. as}, n\to\infty,$$

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

$$\sup_x |P^*(V_n^* \leqslant x) - \tilde{F}_n(x)| = o(n^{-\frac{1}{2}}), \text{a.s.} \tag{4.8}$$

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 $2\sigma_{g_n}{}^* n^{-\frac{1}{2}} V_{n1}^* + \theta_n$ is a $U$-statistic with varying kernel $h_n^{(n)}$ of the form $h_n^{(n)} = \alpha_n + n^{-1}\beta_n$, where $\alpha_n$ and $\beta_n$ are given by (2.12) and (2.13), with $g, f, \theta$ and $\mu$ replaced by $g_n, f_n, \theta_n$ and $\mu_n$, 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 $\psi$ (cf. (2.15)) should be replaced by $\psi_n$, 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

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

where

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

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

$$V_{n1}^* + E^* V_{n2}^* = -2^{-1}\sigma_{g_n}^{-1} {}^* n^{\frac{1}{2}} \binom{n}{2}^{-1} \sum \sum_{1 \leqslant i < j \leqslant n} \{\alpha_n(X_i^*, X_j^*) + \tag{4.13}$$

$$-\theta_n + n^{-1}\beta_n(X_i^*, X_j^*)\} - \tfrac{1}{8}\sigma_{g_n}^{-3} {}^* n^{-\frac{1}{2}} E^*[g_n(X_1^*) f_n(X_1^*)]$$

where

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

and

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

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

$$E^* g_n(X_1^*) f_n(X_1^*) = 4E^* g_n^3(X_1^*) + \tag{4.16}$$

$$+ 8E^* g_n(X_1^*) g_n(X_2^*) h(X_1^*, X_2^*) \to$$
$$\to 4Eg^3(X_1) + 8Eg(X_1)g(X_2)h(X_1, X_2)$$

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 $(1 \leqslant i \leqslant n)$ and $E^* V_{n2}^* = \Theta(n^{-\frac{1}{2}})$ a.s. We can now simply repeat the calculations given in (2.22)-(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 momentassumptions $E^*|h(X_1^*, X_2^*)|^{4+\epsilon} < \infty$, for some $\epsilon > 0$, 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,

$$\text{the df of } g_n(X_1^*) \text{ is non-lattice} \tag{4.17}$$

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 $a > 0$,

$$\Delta_n = \sup_{|t| \leqslant a} |E^* e^{itg_n(X_1^*)} - Ee^{itg(X_1)}| \underset{\text{a.s}}{\to} 0 \tag{4.18}$$

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

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

so that

$$\Delta_n ]( \leqslant \sup_{|t| \leqslant a} |n^{-1} \sum_{i=1}^n \{e^{it(n^{-1}\Sigma_{j=1}^n h(X_i, X_j) - \theta_n)} - \tag{4.20}$$
$$- e^{itg(X_i)}\}| + \sup_{|t| \leqslant a} |n^{-1}\Sigma_{i=1}^n e^{itg(X_i)} - Ee^{itg(X_1)}|$$
$$= \Delta_{n1} + \Delta_{n2}$$

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

$$\Delta_{n1} \leqslant an^{-1}\Sigma_{i=1}^n |n^{-1}\Sigma_{j=1,}^n h(X_i, X_j) - \theta_n - g(X_i)| \tag{4.21}$$
$$\leqslant an^{-1}\Sigma_{i=1}^n |n^{-1}\Sigma_{j=1}^n \{g(X_i) + g(X_j) + \psi(X_i, X_j) + \theta\} - \theta_n - g(X_i)|$$
$$\leqslant a|n^{-1}\Sigma_{j=1}^n g(X_j)| + an^{-1}\Sigma_{i=1}^n |n^{-1}\Sigma_{j=1}^n \psi(X_i, X_j)| +$$
$$+ a|\theta_n - \theta|.$$

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

$$n^{-1}\Sigma_{i=1}^n |n^{-1}\Sigma_{j=1}^n \psi(X_i, X_j)| \to_{\text{a.s.}} 0,$$

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

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

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

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

provided $E\psi^2(X_1, X_2) \log^2 \psi(X_1, X_2) < \infty$, which directly yields (4.22). 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 Theorem of FEUERVERGER and MUREIKA (1977). This completes the proof of Theorem 3. □

5. EXTENSIONS

It is relatively straightforward to extend the results of Section 1 for Studentized $U$-statistics of degree 2 to other classes of Studentized statistical functions, such as Studentized $L$-statistics or linear combinations of order statistics and Studentized $M$-estimators of maximum likelihood type, admitting an second order von Mises expansion. Without going into details we briefly sketch the main argument of such an extension. Let $T$ denote a fixed real-valued functional on the space of all df's on the real line. Suppose that $T$ is twice differentiable at $F$, i.e. there exists real-valued functions $T'(\cdot,F)$ and $T''(\cdot,\cdot,F)$ such that

$$T(G) = T(F) + \int T'(x;F)dG(x) + \frac{1}{2}\int\int T''(x,y;F)dG(x)dG(y) + r(G,F) \tag{5.1}$$

The functions $T'(\cdot;F)$ and $T''(\cdot,\cdot;F)$ obey the requirements $E_F T'(X_1;F)=0, E_F T''(X_1,x;F)=0$ for any real $x$, and $T''(\cdot,\cdot;F)$ is symmetric in its two arguments. In addition we shall assume that $r(G,F)$ is of negligible order for our purpose; i.e.

$$|r(G,F)| = \mathcal{O}(\|G-F\|_\infty^{2+\delta}) \tag{5.2}$$

for some $\delta>0$, uniformly for all $G$ in a neighbourhood of $F$. Here $\| \ \|_\infty$ denotes the usual supremum distance.

Let, for each $n \geq 2, T_n = T(\hat{F}_n)$ and let $\theta = T(F)$. Because of (5.1) we may write

$$\begin{aligned} T_n - \theta &= n^{-1}\Sigma_{i=1}^n T'(X_i;F) + n^{-2}\sum_{i=1}^n\sum_{j=1}^n T''(X_i,X_j;F) + \\ &+ r(\hat{F}_n,F) \\ &= V_n + R_n \end{aligned} \tag{5.3}$$

where $V_n$ can be viewed as a $U$-statistic of degree 2 with varying kernel $\gamma + n^{-1}\delta$, which is given by

$$\gamma(x,y) = \frac{1}{2}(T'(x;F) + T'(y;F)) + T''(x,y;F) \tag{5.4}$$

and

$$\delta(x,y) = \frac{1}{2}(T''(x,x;F) + T''(y,y;F)) - T''(x,y;F) \tag{5.5}$$

and $R_n = r(\hat{F}_n,F)$ is a remainder term of negligibly order of magnitude; i.e.

$$P(\{|R_n| \geq n^{-1}(\ln n)^{-1}\}) = o(n^{-\frac{1}{2}}) \tag{5.6}$$

The bound (5.6) is easily inferred from (5.2) and the well known Dvoretzky, Kiefer, Wolfowitz exponential bound for the Kolmogorov-Smirnov statistic.

We now consider Studentized statistical functions of the form $n^{\frac{1}{2}} s_n^{-1}(T_n - \theta)$, with $T_n - \theta$ satisfying (5.3)-(5.6), and

$$s_n^2 = 4(n-1)(n-2)^{-2}\Sigma_{i=1}^n[(n-1)^{-1}\Sigma_{j=1}^n \gamma(X_i,X_j) - \binom{n}{2}^{-1}\sum_{1 \leq k < l \leq n}\sum \gamma(X_k,X_l)]^2 \tag{5.7}$$

Note that (cf. the remark following (1.4)) that $n^{-1}s_n^2$ is nothing else but the jackknife estimator of the variance of a $U$-statistic of degree 2 with kernel $\gamma$. Let

$$G_n(x) = P(\{n^{\frac{1}{2}} s_n^{-1}(T_n-\theta)\leq x\}) \tag{5.8}$$

and

$$\tilde{G}_n(x) = \Phi(x) + 6^{-1}n^{-\frac{1}{2}}\phi(x)\{(2x^2+1)\frac{E(T'(X_1;F))^3}{(E(T'(X_1;F))^2)^{\frac{3}{2}}} + \tag{5.9}$$

$$+ 3(x^2+1)\frac{ET'(X_1;F)T'(X_2;F)T''(X_1,X_2;F)}{(E(T'(X_1;F))^2)^{\frac{3}{2}}}$$

$$+ 3\frac{ET''(X_1,X_1;F)}{(E(T'(X_1;F))^2)^{\frac{1}{2}}}\}$$

In the very special case that $T_n=U_n$, and $\theta=Eh(X_1,X_2)$, then $T_n$ is a $U$-statistic of degree 2 with kernel $h$, $V_n=U_n, \gamma=h-\theta, \delta\equiv 0, R_n\equiv 0$, and we obtain

$$T'(x;F) = 2g(x), \text{ and } T''(x,y;F) = h(x,y)-\theta-g(x)-g(y)$$

if $x\neq y$ and zero elsewhere. It is now easily checked that (5.9) reduces to (1.6) in this very special case. A fairly standard argument involving the proof of Theorem 1 and the relations (5.1)-(5.7) directly yields

$$\sup_x|G_n(x)-\tilde{G}_n(x)| = o(n^{-\frac{1}{2}}) \tag{5.10}$$

as $n\to\infty$, provided suitable momentassumptions are imposed upon $T'(X_1;F), T''(X_1,X_2;F)$ and $T''(X_1,X_1;F)$ and, in addition, the df of $T'(X_1;F)$ is non-lattice. Thus (5.10) extends Theorem 1 to a general class of Studentized statistical functions. Of course $T$ has to be "sufficiently smooth" in order to validate the stochastic expansion (5.3)-(5.6). In fact the statement (5.10) fails for such non-smooth functionals like $T(F)=F^{-1}(\frac{1}{2})$ (see SINGH (1981); we also refer to PFANZAGL (1985), page 775, where it is noted that $T(F)=F^{-1}(\frac{1}{2})$ does *not* satisfy (5.1) on the space of all sufficiently smooth df's $F$). Parallel extensions to Theorem 2 and 3 are easily established as well. We only have to repeat the argument leading to (1.19) and (1.23) to find that the *Edgeworth expansion estimate*

$$\tilde{E}_n(x) = \Phi(x)+6^{-1}n^{-\frac{1}{2}}\phi(x)\{(2x^2+1)a_n+3(x^2+1)b_n+3c_n\} \tag{5.11}$$

is *asymptotically closer* (in the sense of (1.19)) to the exact df $G_n$ cf. (5.8)) of a Studentized statistical function $n^{\frac{1}{2}}S_n^{-1}(T_n-\theta)$ then the standard normal distribution. Recall that the estimates $a_n$ and $b_n$ are given in (1.15) and (1.16), but replace $h(x,y)$ by $\frac{1}{2}(T'(x;F)+T'(y;F))+T''(x,y;F)$; the estimate $c_n$ is given by

$$c_n=c_n(X_1,\ldots,X_n)=\frac{n^{-1}\Sigma_{i=1}^n T''(X_i,X_i;F)}{(n^{-1}\Sigma_{i=1}^n(n^{-1}\Sigma_{j=1}^n h(X_i,X_j)-\theta_n)^2)^{\frac{1}{2}}} \tag{5.12}$$

with the same choice for $h$ and $\theta_n=T(\hat{F}_n)$.

Similarly, the *bootstrap approximation*

$$F_n^*(x)=P^*(\{n^{\frac{1}{2}}S_n^{*-1}(T_n^*-\theta_n)\leq x\}) \tag{5.13}$$

also has the desirable property of being *asymptotically closer* (cf. (1.23)) to the exact df $G_n$ (cf. (5.8)) of $n^{\frac{1}{2}}S_n^{-1}(T_n-\theta)$ then the standard normal distribution. Note that $T_n^*=T(\hat{F}_n^*)$, where $\hat{F}_n^*$ is the empirical df based on the $X_i^*$'s $(1\leq i\leq n)$, $\theta_n=T(\hat{F}_n)$, and $S_n^*$ is nothing else than $S_n$ (cf. (5.7)) with the $X_i$'s replaced by the $X_i^*$'s.

To conclude this section we briefly discuss two special cases - $L$-statistics and $M$-estimators - to which our results can be applied. We begin with $L$-statistics. Let $T(F)=\int_0^1 F^{-1}(t)J(t)dt$. If $J:(0,1)\to\mathbb{R}$ is differentiable, with a derivative $J'$ on $(0,1)$ which satisfies a Lipschitz condition of order $\alpha>0$ (i.e. $|J(t)-J(s)|\leqslant K|s-t|^\alpha$ for all $0<s,t<1$ and some constants $\alpha,K>0$). then the stochastic expansion (5.3), with a remainder satisfying (5.6), holds true with

$$T'(x;F)=-\int_0^1 J(t)(I_{(0,t)}(F(x)-t)dF^{-1}(t)$$

$$\text{and } T''(x,y;F)=-\int_0^1 J'(t)(I_{(0,t)}(F(x))-t)(I_{(0,t)}(F(y))-t)dF^{-1}(t).$$

Suppose that $J$ and $J'$ are bounded on $(0,1)$, $F$ is non-lattice, $0<E|X_1|^{4+\epsilon}<\infty$, for some $\epsilon>0$, then $\sigma^2(T'(X_1;F))>0$ implies that (5.9) yields an one-term Edgeworth expansion for Studentized $L$-statistics with smooth weights. Related results concerning Edgeworth expansions for normalized $L$-statistics are summarized in [13]; this latter monograph also contains a Berry-Esseen bound for a Studentized $L$-statistics (though with a Studentization, different from the one employed in the present paper). The parallel results validating an asymptotic accuracy of order $o(n^{-\frac{1}{2}})$ a.s., for the Edgeworth expansion estimate and the bootstrap approximation for a Studentized $L$-statistic are also easily obtained under the above mentioned set of assumptions. In fact, arguing along the lines of the proofs of the Theorems 2 and 3, no new difficulties will be encountered, while establishing these results. In [22] SINGH indicated the desirability of obtaining such results for Studentized $L$-statistics.

We now turn to our second example which is concerned with $M$-estimators. Let $\lambda_F(t)=\int\psi(x-t)dF(x)$, where $\psi$ is strictly monotone with $\psi(-\infty)<0$ and $\psi(\infty)>0$. The functional $T(F)$ solves the equation $\lambda_F(T(F))=0$. If $\psi$ is twice differentiable with derivatives $\psi'$ and $\psi''$, and $\psi''$ satisfies a Lipschitz condition of order $\alpha>0$, then the stochastic expansion (5.3) holds true with $T'(x;F)=-(\lambda'_F(T(F)))^{-1}\psi(x-T(F))$ and

$$\begin{aligned} T''(x,y;F)=&T'(x;F)[1+\{2\lambda'_F(T(F))\}^{-1}\psi'(y-T(F))]+ \\ &+T'(y;F)[1+\{2\lambda'_F(T(F))\}^{-1}\psi'(x-T(F))]- \\ &-\{2\lambda'_F(T(F))\}^{-1}\lambda''_F(T(F))T'(x;F)T'(y;F) \end{aligned}$$

(cf. BERAN (1984), p. 104). Again, if suitable momentassumptions are imposed upon $T'(X_1,F)$, $T''(X_1,X_2;F)$ and $T''(X_1,X_1;F)$, and, in addition, the df of $T'(X_1;F)$ is non-lattice, then (5.10) holds true and parallel extensions to Theorem 2 and Theorem 3 for Studentized $M$-estimators are also easily obtained.

## REFERENCES


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

[2] K.B. ATHREYA, M. GHOSH, L.Y. LOW, P.K. SEN (1984). *Laws of Large Numbers for Bootstrapped U-statistics,* Journal of Statistical Planning and Inference **9**, 185-194.

[3] R. BERAN (1982). *Estimated Sampling Distributions: the Bootstrap and Competitors,* Ann. Statist. **10**, 212-225.

[4] R. BERAN (1984). *Jackknife Approximations to Bootstrap Estimates,* Ann. Statist., **12**, 101-118.

[5] P.J. BICKEL and D. FREEDMAN (1981). *Some Asymptotic Theory for the Bootstrap,* Ann. Statist. **9**, 1196-1217.

[6] P.J. BICKEL, F. GÖTZE and W.R. VAN ZWET (1986). *The Edgeworth Expansion for U-statistics of Degree 2,* Ann. Statist. **14**, 1463-1484.

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

[8] H. CALLAERT and N. VERAVERBEKE (1981). *The Order of the Normal Approximation to a Studentized U-statistics,* Ann. Statist. **9**, 194-200.

[9] H. DEHLING, M. DENKER and W. PHILIPP (1984). *Invariance Principles for von Mises and U-statistics,* Z. Wahrscheinlichkeitstheorie verw. Gebiete, **67**, 139-167.

[10] B. EFRON (1987). *Better Bootstrap Confidence Intervals, (with discussion)* J. Amer. Statist. Assoc. *82,* 171-200.

[11] A. FEUERVERGER and R.A. MUREIKA (1977). *The Empirical Characteristic Function and its Applications,* Ann. Statist. **5**, 88-97.

[12] J.A. HARTIGAN (1986). *Comment on "Bootstrap Methods for measures of Statistical Accuracy",* by B. EFRON and R.J. TIBSHIRANI", Statistical Science, Vol. 1, no. 1, 75-77.

[13] R. HELMERS (1982). *Edgeworth Expansions for Linear Combinations of Order Statistics,* Mathematical Centre Tract 105, Amsterdam.

[14] R. HELMERS (1985). *The Berry-Esseen bound for Studentized U-statistics,* The Canadian Journal of Statistics, Vol. 13, no. 1, 79-82.

[15] D. HINKLEY and B. CHENG WEI (1984). *Improvements of Jackknife Confidence Limit Methods,* Biometrika, **71**, 2, 331-119.

[16] G. JOGESH BABU and K. SINGH (1983). *Inference on means using the Bootstrap,* Ann. Statist. **11**, 999-1003.

[17] G. JOGESH BABU and K. SINGH (1984) *On One Term Edgeworth Correction by Efron's Bootstrap,* Sankhyá, **46**, Series A, Part 2, 219-232.

[18] T.L. MALEVICH and B. ABDALIMOV (1979). *Large Deviation Probabilities for U-statistics,* Theory Prob. Applications, **24**, 215-219.

[19] W.C. NAVIDI (1986). *Edgeworth expansions for bootstrapping regression models,* PH. D. Thesis, University of California, Berkeley.

[20] J. PFANZAGL (1985). *On second-order asymptotics for differentiable functionals,* Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, *Volume II,* 769-786, eds. Lucien M. Le Cam and Richard A. Olshen.

[21] K. SINGH (1981). *On Asymptotic Accuracy of Efron's Bootstrap,* Ann. Statist. **9**, 1187-1195.

[22] K. SINGH (1983). Review of: *Edgeworth Expansions for Linear Combinations of Order Statistics* by R. HELMERS in: Sankhyá, **45**, Series A, Part 3, 395-396.

[23] C.F.J. WU (1986). *Jackknife, bootstrap and other resampling methods in regression analysis. (with discussion),* Ann. Stat. **14**, 1261-3150.

[24] ZHAO LINCHENG (1983). *The Rate of the Normal Approximation for a Studentized U-statistic,* Science Exploration, **3**, 45-52.