Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R. Helmers, P. Janssen, N. Veraverbeke

Bootstrapping U-quantiles

Department of Operations Research, Statistics, and System Theory   Report BS-R9021   July

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Bootstrapping U-quantiles

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

P. Janssen and N. Veraverbeke
Limburgs Universitair Centrum
Diepenbeek, Belgium

The asymptotic consistency of the bootstrap approximation for generalized quantiles of U-statistic structure (U-quantiles for short) is established. The same method of proof also yields the asymptotic accuracy of the bootstrap approximation in this case. Applications to location and spread estimators, such as the classical sample quantile, the Hodges-Lehmann estimator of location and a spread estimator proposed by Bickel and Lehmann are given. Our method of proof for the asymptotic consistency relies on an idea of Sheehy and Wellner (1988), who treat the sample median

(1980) Mathematics Subject Classification: 62E20, 62G30, 60F05.
Key words & Phrases: bootstrap approximation, generalized quantiles, U-statistics, consistency, asymptotic accuracy, empirical processes.
Note: This paper is submitted for publication to the Proceedings Volume of Special Topics Meeting on the Bootstrap, 214 th IMS meeting, 15-16 May 1990, East Lansing, Michigan, U.S.A.

1. Introduction

Let $X_1, X_2, \ldots$ be independent random variables defined on a common probability space (Ω, A, P), having common unknown distribution function (df) F. Let $h(x_1, \ldots, x_m)$ be a kernel of degree m (i.e. a real-valued measurable function symmetric in its m arguments) and let

$$H_F(y) = P(h(X_1, \ldots, X_m) \leq y), \quad y \in \mathbb{R} \tag{1.1}$$

denote the df of the random variable $h(X_1, \ldots, X_m)$. Define, for each $n \geq m$ and real y,

$$H_n(y) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \cdots \sum I(h(X_{i_1}, \ldots, X_{i_m}) \leq y) \tag{1.2}$$

the empirical df of U-statistic structure.

Let, for $0 < p < 1$, $\xi_p = H_F^{-1}(p)$, denote the p-th quantile corresponding to H_F, and let $\hat{\xi}_{pn} = H_n^{-1}(p)$ denote its empirical counterpart. Generalized quantiles of the form $\hat{\xi}_{pn} = H_n^{-1}(p)$ are called U-quantiles. Choudhury and Serfling (1988) note that $\hat{\xi}_{pn} \to \xi_p$, a.s. $[P]$, as $n \to \infty$, and, in addition, that, as $n \to \infty$,

$$n^{\frac{1}{2}}(\hat{\xi}_{pn} - \xi_p) \xrightarrow{d} N(0, \sigma^2) \tag{1.3}$$

where

$$\sigma^2 = m^2 \zeta_p h_F^{-2}(\xi_p) \tag{1.4}$$

with

$$\zeta_p = Var(g_p(X_1)) > 0 \tag{1.5}$$

and

Report BS-R9021
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

$$g_p(X_1) = E(I(h(X_1, \ldots, X_m) \leq \xi_p) \mid X_1) - p \tag{1.6}$$

provided H_F has density h_F positive at ξ_p.

In applications one often wishes to establish a confidence interval for $\xi_p = H_F^{-1}(p)$ and a studentized version of (1.3) is required. An estimator of the asymptotic variance σ^2 was proposed by CHOUDHURY and SERFLING (1988). Their estimator was shown to be strongly consistent, but appears to be rather unpleasant to work with.

The aim of this paper is to employ bootstrap methods for the construction of a confidence interval for $\xi_p = H_F^{-1}(p)$. We establish a bootstrap analog of (1.3), under a slightly more stringent smoothness condition on H_F. Our proof is inspired by the argument given in SHEEHY and WELLNER (1988). These authors treat the classical case of the ordinary sample median (i.e. the case $p = \frac{1}{2}, m = 1, h(x) = x$) and obtain an insightful proof of Proposition 5.1 of BICKEL and FREEDMAN (1981). Here we extend their result to the more general case of U-quantiles of the form $\hat{\xi}_{pn} = H_n^{-1}(p)$, $0 < p < 1$.

In Section 2 we state and prove the bootstrap CLT for U-quantiles, whereas in Section 2 we establish the asymptotic accuracy of the bootstrap approximation in this case. Applications to certain estimators of location and spread, such as the classical sample quantile, the Hodges-Lehmann estimator of location and a spread estimator proposed in BICKEL and LEHMANN (1979) are discussed in Section 4.

2. CONSISTENCY OF THE BOOTSTRAP FOR U-QUANTILES

Let $\hat{F}_n$ denote the empirical df based on $X_1, \ldots, X_n$. Define $\hat{\xi}^\star_{pn} = H_n^{\star-1}(p)$, $0 < p < 1$, where $H_n^\star$ denotes the empirical df of U-statistic structure based on the bootstrap sample $X_1^\star, \ldots, X_n^\star$. Here and elsewhere $X_1^\star, \ldots, X_n^\star$ denotes a random sample of size n from $\hat{F}_n$, conditionally given $X_1, \ldots, X_n$.

Our first main result is as follows:

THEOREM 2.1. *Suppose that H_F is continuously differentiable (with density h_F) on a neighbourhood of ξ_p with $h_F(\xi_p) > 0$. Then, for almost every sample sequence $X_1, X_2, \ldots$*

$$n^{\frac{1}{2}}(\hat{\xi}^\star_{pn} - \hat{\xi}_{pn}) \xrightarrow{d} N(o, \sigma^2) \tag{2.1}$$

with σ^2 as in (1.4).

PROOF. Let $\hat{P}_n$ denote the probability measure corresponding to $\hat{F}_n$. Similarly as in SHEEHY and WELLNER (1988), we write

$$\begin{aligned} &\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}^\star_{pn} - \hat{\xi}_{pn}) \leq x) \\ &= \hat{P}_n(n^{\frac{1}{2}}(H_n^{\star-1}(p) - H_n^{-1}(p)) \leq x) \\ &= \hat{P}_n(H_n^{\star-1}(p) \leq H_n^{-1}(p) + xn^{-\frac{1}{2}}) \\ &= \hat{P}_n(H_n^\star(H_n^{-1}(p) + xn^{-\frac{1}{2}}) \geq p) \\ &= \hat{P}_n(W_n^\star \geq -\overline{D}_n) \end{aligned} \tag{2.2}$$

where

$$W_n^\star = n^{\frac{1}{2}}\{H_n^\star(H_n^{-1}(p) + xn^{-\frac{1}{2}}) - \overline{H}_n(H_n^{-1}(p) + xn^{-\frac{1}{2}})\} \tag{2.3}$$

with, for each $n \geqslant m$ and real y,

$$\overline{H}_n(y) = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} I(h(X_{i_1}, \ldots, X_{i_m}) \leqslant y) \tag{2.4}$$

the empirical df of von Mises structure, and

$$\overline{D}_n = n^{\frac{1}{2}} \{\overline{H}_n(H_n^{-1}(p) + xn^{-\frac{1}{2}}) - p\}. \tag{2.5}$$

We first consider $\overline{D}_n$. Note that

$$\overline{D}_n = \sum_{i=1}^{3} \overline{D}_{in} \tag{2.6}$$

where

$$\overline{D}_{1n} = n^{\frac{1}{2}} \{\overline{H}_n(H_n^{-1}(p) + xn^{-\frac{1}{2}}) - H_F(H_n^{-1}(p) + xn^{-\frac{1}{2}})\} - n^{\frac{1}{2}} \{\overline{H}_n(H_n^{-1}(p)) - H_F(H_n^{-1}(p))\} \tag{2.7}$$

$$\overline{D}_{2n} = n^{\frac{1}{2}} \{H_F(H_n^{-1}(p) + xn^{-\frac{1}{2}}) - H_F(H_n^{-1}(p))\} \tag{2.8}$$

and

$$\overline{D}_{3n} = n^{\frac{1}{2}} \{\overline{H}_n(H_n^{-1}(p)) - p\}. \tag{2.9}$$

To treat $\overline{D}_{1n}$ note first that $\overline{D}_{1n} = D_{1n} + O(n^{-\frac{1}{2}})$ a.s. $[P]$, as $n \to \infty$, where D_{1n} is obtained from $\overline{D}_{1n}$ by replacing $\overline{H}_n$ by H_n, with H_n as in (1.2). Suppose without loss of generality that $x > 0$. Clearly, for n sufficiently large,

$$|D_{1n}| \leqslant \sup_{\substack{|t-s| \leqslant xn^{-\frac{1}{2}} \\ s,t \in J}} |U_n(t) - U_n(s)| \quad \text{a.s. } [P] \tag{2.10}$$

where J is the neighborhood of ξ_p on which H_F is continuously differentiable, and

$$U_n(t) = n^{\frac{1}{2}} (H_n(t) - H_F(t)), \quad t \in \mathbb{R} \tag{2.11}$$

denotes the empirical process of U-statistic structure.

Similarly as in SILVERMAN (1983) it is easy to see that

$$|D_{1n}| \leqslant (n!)^{-1} \sum_{\alpha} \sup_{\substack{|t-s| \leqslant xn^{-\frac{1}{2}} \\ t,s \in J}} |U^{\alpha}_{[\frac{n}{m}]}(t) - U^{\alpha}_{[\frac{n}{m}]}(s)| \tag{2.12}$$

where, for any given permutation α of $\{1,2,\ldots,n\}$ $U^{\alpha}_{[\frac{n}{m}]}(t)$ denotes the empirical df (evaluated at the point t) of the $[\frac{n}{m}]$ independent random variables
$h(X_{\alpha(mj+1)}, \ldots, X_{\alpha(mj+m)}), j = 0, 1, \ldots, [\frac{n}{m}] - 1$, all with common df H_F. Application of relation (2.13) of STUTE (1982) to each of the $n!$ terms appearing on the r.h.s. of (2.12) directly yields that $D_{1n} = O(n^{-\frac{1}{4}} (ln\, n)^{\frac{1}{2}})$ a.s. $[P]$, as $n \to \infty$, hence,

$$\overline{D}_{1n} = O(n^{-\frac{1}{4}} (\ln n)^{\frac{1}{2}}) \quad \text{a.s.}[P], \text{ as } n \to \infty. \tag{2.13}$$

Here we have used the smoothness of H_F as well as the inequality (2.12).

Next we consider $\overline{D}_{2n}$. Using again the smoothness assumption on H_F and employing the a.s. $[P]$ convergence of ξ_{pn} to ξ_p, as $n \to \infty$, we easily obtain from the mean value theorem that

$$\overline{D}_{2n} \to x h_F(\xi_p) \quad \text{a.s. } [P], \text{ as } n \to \infty. \tag{2.14}$$

Finally note that $\overline{D}_{3n} = O(n^{-\frac{1}{2}})$. We can conclude that

$$\overline{D}_n \to x h_F(\xi_p) \quad \text{a.s. } [P], \text{ as } n \to \infty. \tag{2.15}$$

Next we consider the limit behaviour of W_n^*, $n = m, m+1, \ldots$ (cf (2.3)), conditionally given $\hat{F}_n$. Obviously, given $\hat{F}_n$, W_n^* is a normalized U-statistic of degree m, with bounded kernel, depending on n, of the form

$$h_n(x_1, \ldots, x_m) = I(h(x_1, \ldots, x_m) \leq \hat{\xi}_{pn} + xn^{-\frac{1}{2}}) - \overline{H}_n(\hat{\xi}_{pn} + xn^{-\frac{1}{2}}) \tag{2.16}$$

Of course $E_{\hat{F}_n} W_n^* = E_{\hat{F}_n} h_n(X_1^*, \ldots, X_m^*) = 0$, a.s. $[P]$, whereas it is easily checked that

$$Var_{\hat{F}_n}(W_n^*) \sim m^2 E_{\hat{F}_n} g_n^2(X_1^*) \quad \text{a.s. } [P], \text{ as } n \to \infty, \tag{2.17}$$

where

$$g_n(X_1^*) = E_{\hat{F}_n}(h_n(X_1^*, \ldots, X_m^*) | X_1^*). \tag{2.18}$$

A simple argument involving the strong law for U-statistics with estimated parameters (Theorem 2.9 of IVERSON and RANDLES (1989)) directly yields that

$$E_{\hat{F}_n} g_n^2(X_1^*) \to \zeta_p \quad \text{a.s. } [P], \text{ as } n \to \infty, \tag{2.19}$$

with ζ_p as in (1.5).

At this point we apply the Berry-Esseen bound for U-statistics of degree m of VAN ZWET (1984) to find that

$$\sup_x |\hat{P}_n(W_n^* \leq x) - \Phi(xm^{-1}\zeta_p^{-\frac{1}{2}})| \tag{2.20}$$

$$= O\{(\frac{E_{\hat{F}_n}|g_n(X_1^*)|^3}{(E_{\hat{F}_n} g_n^2(X_1^*))^{\frac{3}{2}}} + \frac{E_{\hat{F}_n} h_n^2(X_1^*, \ldots, X_m^*)}{E_{\hat{F}_n} g_n^2(X_1^*)})n^{-\frac{1}{2}}\}.$$

Note that, in contrast to Corollary 4.1 of VAN ZWET (1984), the asymptotic variance instead of the exact variance of W_n^* is employed. It is easy to see that this does not affect the bound (2.20). The different standardization will give rise to an additional term of type

$$\frac{E_{\hat{F}_n} h_n^2(X_1^*, \ldots, X_m^*)}{E_{\hat{F}_n} g_n^2(X_1^*)} n^{-\frac{1}{2}} \tag{2.21}$$

which is already present in van Zwet's bound. Because h_n is bounded by 1, for all n, and combining (2.19) with the fact that $\zeta_p > 0$, we easily see that the moments appearing on the r.h.s. of (2.20) are $O(1)$ a.s. $[P]$, as $n \to \infty$. Hence the r.h.s. of (2.20) is $O(n^{-\frac{1}{2}})$ a.s. $[P]$, as $n \to \infty$.

From (2.2), (2.15) and (2.20) we obtain

$$\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}_{pn}^* - \hat{\xi}_{pn}) \leq x) \tag{2.22}$$

$$=1-\Phi(-\bar{D}_n m^{-1}\zeta_p^{-\frac{1}{2}})+O(n^{-\frac{1}{2}})$$
$$=\Phi(x\sigma^{-1})+o(1)$$

a.s. $[P]$, as $n\to\infty$. This completes the proof of theorem. □

For the special case $m=1$, $h(x)=x, p=\frac{1}{2}$, the classical sample median, SHEEHY and WELLNER (1988) obtained the same result. It should also be noted that instead of the Berry-Esseen bound (2.20) we could have used a CLT for a triangular scheme of U-statistics like the one given in Theorem 2.1 of JAMMALAMADAKA and JANSON (1986). However, for our purpose (see Section 3) it will be convenient to have the order bound (2.22).

3. Asymptotic accuracy of the bootstrap for U-quantiles

From (1.3) and (2.1) we know that the bootstrap approximation for a normalized U-quantile is asymptotically consistent. In this section we investigate the a.s. rate at which the difference between the bootstrap approximation and the exact distribution of a normalized U-quantile tends to zero, as the sample size gets large.

THEOREM 3.1. *Suppose that the assumptions of Theorem 2.1 are satisfied. Suppose, in addition, that h_F satisfies a Lipschitz condition of order $\geq\frac{1}{2}$ on a neighborhood of ξ_p. Then*

$$\sup_x |\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}^*_{pn}-\hat{\xi}_{pn})\leq x)-P(n^{\frac{1}{2}}(\hat{\xi}_{pn}-\xi_p)\leq x)|=O(n^{-\frac{1}{4}}\ln n) \tag{3.1}$$

a.s. $[P]$, *as* $n\to\infty$.

For the special case $m=1, h(x)=x$, the classical p-th sample quantile, SINGH (1981) obtained a sligthly better a.s. rate: the factor $\ln n$ in (3.1) is replaced by $(\ln\ln n)^{\frac{1}{2}}$ in this case. Whether the same improvement holds true for U-quantiles appears to be an interesting open problem.

PROOF. First note that

$$\sup_x |\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}^*_{pn}-\hat{\xi}_{pn})\leq x) - P(n^{\frac{1}{2}}(\hat{\xi}_{pn}-\xi_p)\leq x)|\leq \sum_{i=1}^{3} I_{in} \tag{3.2}$$

where, for some constant $K>0$,

$$I_{1n}=\sup_{|x|\leq K(\ln n)^{\frac{1}{2}}} |\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}^*_{pn}-\hat{\xi}_{pn})\leq x)-\Phi(x\sigma^{-1})| \tag{3.3}$$

and

$$I_{2n}=\sup_{|x|> K(\ln n)^{\frac{1}{2}}} |\hat{P}_n(n^{\frac{1}{2}}(\hat{\xi}^*_{pn}-\hat{\xi}_{pn})\leq x)-\Phi(x\sigma^{-1})| \tag{3.4}$$

and

$$I_{3n}=\sup_x|P(n^{\frac{1}{2}}(\hat{\xi}_{pn}-\xi_p)\leq x)-\Phi(x\sigma^{-1})|.$$

We first consider I_{1n}. Going through the proof of Theorem 2.1 we easily verify that

$$\sup_{|x|\leq K(\ln n)^{\frac{1}{2}}} |\bar{D}_n-xh_F(\xi_p)|=0(n^{-\frac{1}{4}}\ln n) \quad \text{a.s. } [P], \text{ as } n\to\infty. \tag{3.5}$$

Here we have used (see (2.12)) that

$$\sup_{|x|\leqslant K(\ln n)^{\frac{1}{2}}} |D_{1n}| \leqslant (n!)^{-1} \sum_{\alpha} \sup_{\substack{|t-s|\leqslant Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}} \\ t,s\in J}} |U^{\alpha}_{[\frac{n}{m}]}(t) - U^{\alpha}_{[\frac{n}{m}]}(s)|$$

$$= O(n^{-\frac{1}{4}}(\ln n)^{\frac{3}{4}}) \quad \text{a.s.}[P],\ n\to\infty$$

by application of relation (2.13) in STUTE (1982). Also (2.14) is replaced by the stronger assertion that

$$\sup_{|x|\leqslant K(\ln n)^{\frac{1}{2}}} |\overline{D}_{2n} - xh_F(\xi_p)| = O(n^{-\frac{1}{4}}\ln n) \quad \text{a.s.}[P],\ \text{as } n\to\infty.$$

For this we used Lemma 3. of CHOUDHURY and SERFLING (1988) and the Lipschitz condition on h_F. Combining (3.5) with (2.22) directly yields

$$I_{1n} = O(n^{-\frac{1}{4}}\ln n) \quad \text{a.s. } [P],\ \text{as } n\to\infty. \tag{3.6}$$

For the quantity I_{2n} we have

$$\begin{aligned} I_{2n} \leqslant\ & \hat{P}_n(\hat{\xi}^{*}_{pn} - \hat{\xi}_{pn} > Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ & + \hat{P}_n(\hat{\xi}^{*}_{pn} - \hat{\xi}_{pn} < -Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ & + 2(1-\Phi(K(\ln n)^{\frac{1}{2}}\sigma^{-1}). \end{aligned} \tag{3.7}$$

The third term is $O(n^{-\frac{1}{2}})$ by taking K large enough. It remains to estimate the two other terms. Since the argument is the same for both, we only deal with the first term of the r.h.s. of (3.7). Similarly as in (2.2) we write

$$\begin{aligned} & \hat{P}_n(\hat{\xi}^{*}_{pn} - \hat{\xi}_{pn} > Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ & = P_n(H^{*}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ & \qquad < p - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}})) \end{aligned} \tag{3.8}$$

Application of Lemma 3.1 of CHOUDHURY and SERFLING (1988) directly yields that for all n sufficiently large,

$$p - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \leqslant p - \overline{H}_n(\xi_p + \frac{K}{2}n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \tag{3.9}$$

a.s. $[P]$, provided we take K large enough. A simple argument involving Corollary 2.1 of HELMERS, JANSSEN and SERFLING (1988) and the a.s. closeness of $\overline{H}_n$ and H_n gives us (with C_m as in the corollary)

$$\begin{aligned} & p - \overline{H}_n(\xi_p + \frac{K}{2}n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ & \leqslant p - H_F(\xi_p + \frac{K}{2}n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) + C_m n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}} + O(n^{-1}) \end{aligned} \tag{3.10}$$

a.s. $[P]$. The smoothness assumption of the theorem directly implies that

$$p - H_F(\xi_p + \frac{K}{2}n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \tag{3.11}$$

$$= -\frac{K}{2} h_F(\xi_p) n^{-\frac{1}{2}} (\ln n)^{\frac{1}{2}} (1+o(1)) \qquad \text{a.s.}[P], \text{ as } n \to \infty.$$

Together (3.9), (3.10) and (3.11) yield that $p - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) < 0$, for all n sufficiently large, a.s. $[P]$, provided we take K large enough

We can now apply an exponential bound for U-statistics with bounded kernels of HOEFFDING (1963) (see also SERFLING (1980), p. 201) to find that

$$\begin{aligned} &\hat{P}_n(H_n^{\star}(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}) \\ &\qquad\qquad < p - \overline{H}_n(\hat{\xi}_{pn} + Kn^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}})) \\ &\leq \exp\{-\tfrac{1}{8}[\frac{n}{m}] n^{-1} \ln n K^2 h_F^2(\xi_p)\} \\ &= O(n^{-\frac{1}{2}}) \qquad\qquad \text{a.s.}[P], \text{ as } n \to \infty, \end{aligned} \tag{3.12}$$

provided K is taken sufficiently large. This together with (3.7) and (3.8) implies that

$$I_{2n} = O(n^{-\frac{1}{2}}) \qquad \text{a.s.}[P], \text{ as } n \to \infty. \tag{3.13}$$

Hence I_{2n} is of negligible order for our purposes. It remains to consider I_{3n}. Clearly, as $n \to \infty$,

$$I_{3n} = \sup_x |P(n^{-\frac{1}{2}}(\hat{\xi}_{pn} - \xi_p) \leq x) - \Phi(x\sigma^{-1})| = O(n^{-\frac{1}{2}}), \tag{3.14}$$

i.e. the Berry-Esseen bound for U-quantiles is valid. To check (3.14) is an easy matter in view of the classical proof of a Berry-Esseen bound for ordinary sample quantiles (see, e.g. SERFLING (1980), p. 81-84). We have to apply instead of the Lemma on p. 75 of SERFLING (1980), the exponential bound of HOEFFDING (1963) for U-statistics with bounded kernels. Also a Berry-Esseen bound for U-statistics is needed.

Combining (3.6), (3.13) and (3.14) with (3.2), we find that the theorem is proved.

□

4. APPLICATIONS

In this section we indicate briefly applications of our results to the problem of obtaining confidence intervals for $\xi_p = H_F^{-1}(p)$. Let $u_{\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$. The normal approximation (1.3) yields an approximate two-sided confidence interval

$$(\hat{\xi}_{pn} - n^{-\frac{1}{2}} \hat{\sigma}_n u_{\frac{\alpha}{2}},\ \hat{\xi}_{pn} + n^{-\frac{1}{2}} \hat{\sigma}_n u_{\frac{\alpha}{2}}) \tag{4.1}$$

for ξ_p. Here $\hat{\sigma}_n^2$ denotes a consistent estimator (e.g., the one proposed by CHOUDHURY and SERFLING (1988)) of the asymptotic variance σ^2. Clearly, the error rates corresponding to the upper and lower confidence limits in (4.1) will depend on the rate at which $\hat{\sigma}_n^2$ approaches σ^2.

A bootstrap based confidence interval for ξ_p is given by

$$(\hat{\xi}_{pn} - n^{-\frac{1}{2}} c^{\star}_{n,1-\frac{\alpha}{2}},\ \hat{\xi}_{pn} - n^{-\frac{1}{2}} c^{\star}_{n,\frac{\alpha}{2}}) \tag{4.2}$$

where $c^{\star}_{n,\frac{\alpha}{2}}$ and $c^{\star}_{n,1-\frac{\alpha}{2}}$ denote the $\frac{\alpha}{2}$-th and $(1-\frac{\alpha}{2})$-th percentile of the (simulated) bootstrap approximation. It is easily verified that the upper and lower confidence limits in (4.2) have error rates equal to $\frac{\alpha}{2}+O(n^{-\frac{1}{4}}\ln n)$.

We discuss a few specific examples of U-quantiles. In the first of these we take $m=1$, $h(x)=x$ and obtain the classical p-th sample quantile $\hat{\xi}_{pn}=\hat{F}_n^{-1}(p)$, $0<p<1$. Our second example is obtained by taking $p=\frac{1}{2}$, $m=2$, $h(x_1,x_2)=(x_1+x_2)/2$. In this case $\hat{\xi}_{\frac{1}{2}n}=H_n^{-1}(\frac{1}{2})$ becomes the well-known Hodges-Lehmann location estimator. In the third and final example we take $p=\frac{1}{2}$, $m=2$, $h(x_1,x_2)=|x_1-x_2|$. In this case $\hat{\xi}_{\frac{1}{2}n}=H_n^{-1}(\frac{1}{2})$ reduces to an estimator of spread proposed by Bickel and Lehmann (1979).

A further investigation into the relative merits of the normal and bootstrap based confidence intervals (4.1) and (4.2) for U-quantiles appears to be worthwhile. The authors hope to report on these matters elsewhere.

References

P.J. Bickel and D. Freedman (1981). Some asymptotic theory for the bootstrap. Ann. Statist. **9**, 1196-1217.

P.J. Bickel and E.L. Lehmann (1979). Descriptive statistics for nonparametric models IV. Spread. Contributions to Statistics (J. Hájek Memorial Volume), 33-40. (ed. J. Jurečková). Academia, Prague.

J. Choudhury and R. Serfling (1988). Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals. J. Statist. Planning Inf. **19**, 269-282.

R. Helmers, P. Janssen and R. Serfling (1988). Glivenko-Cantelli properties of some generalized empirical df's and strong convergence of generalized L-statistics. Probab. Th. Rel. Fields **79**, 75-93.

W. Hoeffding (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statis. Assoc. **58**, 13-30.

H. Iverson and R. Randles (1989). The effect on convergence of substituting parameter estimates into U-statistics and other families of statistics. Probab. Th. Rel. Fields **81**, 453-471.

S.R. Jammalamadaka and S. Janson (1986). Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. Ann. Probability **14**, 1347-1358.

R.J. Serfling (1980). Approximation theorems of mathematical statistics. J. Wiley, New York.

A. Sheehy and J. Wellner (1988). Almost sure convergence of the bootstrap for the median. Unpublished manuscript.

B.W. Silverman (1983). Convergence of a class of empirical distribution functions of dependent random variables. Ann. Probability **11**, 745-751.

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

W. Stute (1982). The oscillation behavior of empirical processes. Ann. Probability **10**, 86-107.

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