The Annals of Statistics
1993, Vol. 21, No. 4, 1811–1823

CONSISTENCY OF THE GENERALIZED BOOTSTRAP FOR DEGENERATE U-STATISTICS

BY MARIE HUŠKOVÁ[1] AND PAUL JANSSEN[2]

Charles University and Limburgs Universitair Centrum

A generalized bootstrap version is defined for degenerate U-statistics. Our main result shows that the (conditional) distribution function of the bootstrapped degenerate U-statistic provides a consistent estimator for the unknown distribution function of the degenerate U-statistic under consideration. For the proof we rely on a rank statistic approach.

1. Introduction and main result. Let $X_1, \dots, X_n$ be iid random variables defined on a probability space $(\Omega, \mathscr{A}, P)$ and let F denote the common d.f. Let h be a symmetric *kernel* of degree 2. We assume

$$Eh^2(X_1, X_2) < \infty, \tag{U.1}$$

and without loss of generality (w.l.o.g.) we can then assume $Eh(X_1, X_2) = 0$. Consider the U-statistic

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

and assume that we are in the situation where

$$Eh(X_1, x) = 0 \quad \text{for (almost) all } x. \tag{U.2}$$

Then the U-statistic is *degenerate*. According to Serfling (1980), subsection 5.5.2,

$$nU_n(h) \to_d \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1),$$

with $Z_1^2, Z_2^2, \dots$ independent $\chi^2(1)$ random variables and with $\lambda_1, \lambda_2, \dots$ the eigenvalues corresponding to the orthonormal eigenfunctions $\varphi_1, \varphi_2, \dots$ associated with the kernel h. Recall that

$$E\varphi_j(X_1)\varphi_k(X_1) = \delta_{jk},$$

$$Eh^2(X_1, X_2) = \sum_{k=1}^{\infty} \lambda_k^2 < \infty,$$

$$\lim_{K \to \infty} E\left(h(X_1, X_2) - \sum_{k=1}^{K} \lambda_k \varphi_k(X_1)\varphi_k(X_2)\right)^2 = 0.$$

Received April 1992; revised January 1993.
[1]Supported in part by ZWO, the Netherlands.
[2]Supported in part by Grant DMS-8723011.
AMS 1991 *subject classifications*. Primary 62E20; secondary 60F05.
Key words and phrases. Bootstrap consistency, degenerate U-statistics, weighted bootstrapping.

For a recent survey on degenerate U-statistics and for a list of examples including *nonparametric* tests such as Cramér–von Mises statistics and goodness-of-fit statistics, we refer to de Wet (1987).

In general, it is hard to obtain for a given kernel h the corresponding sequence of eigenfunctions and eigenvalues. Therefore the *bootstrap* approximation to the distribution function of nU_n provides a useful alternative for the asymptotic distribution given previously. For a degenerate U-statistic we define the corresponding bootstrapped U-statistic as

$$U_{\mathbb{W}_n}(h) = \sum_{i \neq j}\sum \left(W_{ni} - \frac{1}{n}\right)\left(W_{nj} - \frac{1}{n}\right) h(X_i, X_j),$$

where $\mathbb{W}_n = (W_{n1}, \dots, W_{nn})$ is a vector of *random weights* defined on a probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ and independent of the data $X_1, \dots, X_n$. We assume that the components of $\mathbb{W}_n$ are *exchangeable* and that

(W.1) $$W_{ni} \geq 0, \qquad i = 1, \dots, n,$$

(W.2) $$\sum_{i=1}^{n} W_{ni} = 1,$$

(W.3) $$n \sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \to_{\tilde{P}} c^2, \qquad n \to \infty \quad \text{for some } c > 0,$$

(W.4) $$n \max_{1 \leq i \leq n} \left(W_{ni} - \frac{1}{n}\right)^2 \to_{\tilde{P}} 0, \qquad n \to \infty.$$

Typical choices of weights include, among others, *multinomial weights* (resulting in *Efron's* resampling scheme) and *Dirichlet weights* (*Bayesian* resampling). See Section 2 for a more detailed discussion. Relevant references for further examples are Praestgaard and Wellner (1993) and Hall and Mammen (1992). We also note that for degenerate U-statistics the Efron bootstrap is considered in Arcones and Giné (1992) and that iid weights are studied in Dehling and Mikosch (1992). Since

$$\begin{aligned} U_{\mathbb{W}_n}(h) = {} & \sum_{i \neq j}\sum W_{ni} W_{nj} h(X_i, X_j) - \sum_{i=1}^{n} W_{ni} \frac{1}{n} \sum_{j \neq i} h(X_i, X_j) \\ & - \sum_{j=1}^{n} W_{nj} \frac{1}{n} \sum_{i \neq j} h(X_i, X_j) + \frac{1}{n^2} \sum_{i \neq j}\sum h(X_i, X_j), \end{aligned}$$

we see that for multinomial weights $U_{\mathbb{W}_n}(h)$ coincides [up to terms $h(X_i, X_i)$] to what Arcones and Giné (1992) denote as $U_2^n(\bar{h}_n, P_n)$.

Our main result states that, given $\mathbb{X}_n = (X_1, \dots, X_n)$, the distribution function of $nc^{-1}U_{\mathbb{W}_n}(h)$ provides a uniformly consistent estimator for the unknown distribution of $nU_n(h)$. It reads as follows.

THEOREM. *Assume a symmetric kernel satisfying* (U.1) *and* (U.2) *and weights satisfying* (W.1)–(W.4). *Then, with P-probability* 1,

$$\sup_{x \in \mathbb{R}} \left| P\left\{ nc^{-1} U_{\mathbb{W}_n}(h) \le x | \mathbb{X}_n \right\} - P\{ nU_n(h) \le x \} \right| \to 0, \qquad n \to \infty.$$

2. Examples of weights.

Bayesian resampling scheme. Let $\zeta_1, \dots, \zeta_n$ be any sequence of strictly positive iid random variables with mean μ and finite variance σ^2. With $\bar{\zeta} = n^{-1} \sum_{i=1}^n \zeta_i$ consider weights of the form

$$W_{ni} = \frac{\zeta_i}{n\bar{\zeta}}. \tag{2.1}$$

Note that in the case of exponentially distributed random variables with mean 1, these weights are distributional equivalent with Dirichlet weights. See Mason and Newton (1992) and Janssen (1993) for further details. Weights of form (2.1) trivially satisfy (W.1) and (W.2). Also (W.3) is valid, since

$$n \sum_{i=1}^n \left(W_{ni} - \frac{1}{n} \right)^2 = \frac{1}{n\bar{\zeta}^2} \sum_{i=1}^n \left(\zeta_i - \bar{\zeta} \right)^2 \to_{\tilde{P}} \frac{\sigma^2}{\mu^2}, \qquad n \to \infty.$$

Finally, we have

$$\begin{aligned} 0 \le n \max \left(W_{ni} - \frac{1}{n} \right)^2 &= \frac{1}{n\bar{\zeta}^2} \max_{1 \le i \le n} \left(\zeta_i - \bar{\zeta} \right)^2 \\ &\le \frac{2}{n\bar{\zeta}^2} \left[\left(\bar{\zeta} - \mu \right)^2 + \max_{1 \le i \le n} \left(\zeta_i - \mu \right)^2 \right]. \end{aligned} \tag{2.2}$$

Since $\sigma^2 < \infty$ if and only if

$$\frac{\max_{1 \le i \le n} \left(\zeta_i - \bar{\zeta} \right)}{n^{1/2}} \to 0, \qquad n \to \infty \quad \text{a.s. } [\tilde{P}].$$

It follows that the r.h.s. in (2.2) converges to 0 a.s. $[\tilde{P}]$. Hence (W.4) is satisfied.

Efron's resampling scheme. Multinomial weights satisfy

$$(nW_{n1}, \dots, nW_{nn}) \sim \text{Mult}\left(n; \frac{1}{n}, \dots, \frac{1}{n} \right).$$

Then our generalized bootstrap procedure corresponds to the classical resampling from the empirical d.f. $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i \le x\}$. For multinomial weights (W.1) and (W.2) are trivially satisfied. The validity of (W.3) follows from Lemma 4.1 in Mason and Newton (1992) with $r = 2$. To establish the validity of (W.4), note that the fourth central moment of a binomial distributed random variable with parameters n and p is given by $3(np(1+p))^2 + np(1-p)(1-6p(1-p))$. Since, for $i = 1, \dots, n$, nW_{ni} has a binomial distribution with parameters n and n^{-1}, we have

$$E(nW_{ni} - 1)^4 \le 4$$

and hence, for any $\varepsilon > 0$,

$$\begin{aligned}&\tilde{P}\Big\{\max_{1\le i\le n} n\big(W_{ni} - n^{-1}\big)^2 \ge \varepsilon\Big\}\\ &\quad\le n\tilde{P}\big\{(nW_{ni} - 1)^2 \ge \varepsilon n\big\} \le n\frac{4}{\varepsilon^2 n^2} \to 0, \qquad n \to \infty.\end{aligned}$$

3. Proof of the main result. The proof of our main theorem requires a good understanding of the proof of Theorem 2.1 in Mason and Newton (1992). We assume w.l.o.g. that $c^2 = 1$ in (W.3). We then have to show that

$$\sup_{x\in\mathbb{R}}\big|P\{nU_{\mathbb{W}_n}(h) \le x|\mathbb{X}_n\} - P\{nU_n(h) \le x\}\big| \to 0, \quad n\to\infty \quad \text{a.s. } [P]. \tag{3.1}$$

From Section 2A in Arcones and Giné (1992) [see also the related Lemma 4.1 in Dehling (1989)], we have the following result for a mean-zero kernel satisfying (U.1) and (U.2). For any $\eta > 0$ there exists a simple kernel

$$h_l(x,y) = \sum_{j=1}^{s_l} t_{lj}\varphi_{lj}(x)\varphi_{lj}(y), \tag{3.2}$$

with $t_{l1}, \dots, t_{ls_l}$ real numbers and $\varphi_{l1}, \dots, \varphi_{ls_l}$ bounded functions with $E\varphi_{lj}(X_1) = 0$, such that

$$E(h(X_1,X_2) - h_l(X_1,X_2))^2 \le \eta.$$

At this point we direct the reader to the remark at the end of this section. Since, for any $\varepsilon > 0$,

$$\begin{aligned}&P\{nU_{\mathbb{W}_n}(h_l) \le x - \varepsilon|\mathbb{X}_n\} - P\{n|U_{\mathbb{W}_n}(h - h_l)| \ge \varepsilon|\mathbb{X}_n\}\\ &\quad\le P\{nU_{\mathbb{W}_n}(h) \le x|\mathbb{X}_n\}\\ &\quad\le P\{nU_{\mathbb{W}_n}(h_l) \le x + \varepsilon|\mathbb{X}_n\} + P\{n|U_{\mathbb{W}_n}(h - h_l)| \ge \varepsilon|\mathbb{X}_n\}\end{aligned} \tag{3.3}$$

and since similar inequalities hold for the distribution of $nU_n(h)$, we have

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\big|P\{nU_{\mathbb{W}_n}(h) \le x|\mathbb{X}_n\} - P\{nU_n(h) \le x\}\big|\\ &\quad\le T_{n1}(\mathbb{X}_n) + T_{n2} + T_{n3} + T_{n4}(\mathbb{X}_n),\end{aligned} \tag{3.4}$$

with

$$\begin{aligned}T_{n1}(\mathbb{X}_n) &= \sup_{y\in\mathbb{R}}\big|P\{nU_{\mathbb{W}_n}(h_l) \le y|\mathbb{X}_n\} - P\{nU_n(h_l) \le y\}\big|,\\ T_{n2} &= \sup_{y\in\mathbb{R}}\big[P\{nU_n(h_l) \le y + \varepsilon\} - P\{nU_n(h_l) \le y - \varepsilon\}\big],\\ T_{n3} &= P\{n|U_n(h - h_l)| \ge \varepsilon\},\\ T_{n4}(\mathbb{X}_n) &= P\big\{n\big|U_{\mathbb{W}_n}(h - h_l)\big| \ge \varepsilon|\mathbb{X}_n\big\}.\end{aligned}$$

For the degenerate kernel $g_l(x, y) = h(x, y) - h_l(x, y)$, we apply Lemma 2.1(a) in Arcones and Giné (1992) to obtain

$$T_{n3} \le \frac{2n}{(n-1)\varepsilon^2} E g_l^2(X_1, X_2) \le \frac{4\eta}{\varepsilon^2}.$$

To study the behavior of T_{n2}, note that

$$\begin{aligned} nU_n(h_l) &= \frac{n}{n-1} \frac{1}{n} \sum_{i \ne j} \sum_{j=1}^{s_l} t_{lj} \varphi_{lj}(X_i) \varphi_{lj}(X_j) \\ &= \frac{n}{n-1} \sum_{j=1}^{s_l} t_{lj} \big(A_{n,lj}^2 - B_{n,lj} \big), \end{aligned}$$

with

$$\begin{aligned} A_{n,lj} &= n^{-1/2} \sum_{i=1}^{n} \varphi_{lj}(X_i), \\ B_{n,lj} &= n^{-1} \sum_{i=1}^{n} \varphi_{lj}^2(X_i). \end{aligned}$$

Furthermore, note that $E(A_{n,lj}) = 0$ and

$$\operatorname{Cov}(A_{n,lj_1}, A_{n,lj_2}) = E\big(\varphi_{lj_1}(X_1)\varphi_{lj_2}(X_1)\big) \equiv \sigma_{l, j_1 j_2}.$$

Therefore by the Lindeberg–Lévy central limit theorem

$$(A_{n,l1}, \dots, A_{n,ls_l}) \to_d (V_{l1}, \dots, V_{ls_l}),$$

where $(V_{l1}, \dots, V_{ls_l})$ has a multivariate normal distribution with mean vector 0 and variance–covariance matrix

$$\Sigma_l = \big(\sigma_{l, j_1 j_2} \colon j_1, j_2 = 1, \dots, s_l\big).$$

Moreover, by the classical SLLN we have

$$B_{n,lj} \to \sigma_{l,jj}, n \to \infty \quad \text{a.s. } [P].$$

Consequently,

$$nU_n(h_l) \to_d Y_l = \sum_{j=1}^{s_l} t_{lj} \big(V_{lj}^2 - \sigma_{l,jj} \big)$$

(note that Y_l is continuous) and

$$T_{n2} \to \sup_{y \in \mathbb{R}} P\{y - \varepsilon \le Y_l \le y + \varepsilon\}.$$

The discussion just given also implies that instead of handling $T_{n1}(\mathbb{X}_n)$ it suffices to study the behavior of

$$T_{n5}(\mathbb{X}_n) = \sup_{y \in \mathbb{R}} \Big| P\big\{ nU_{\mathbb{W}_n}(h_l) \le y | \mathbb{X}_n \big\} - P\{Y_l \le y\} \Big|.$$

We now handle $T_{n4}(\mathbb{X}_n)$ and $T_{n5}(\mathbb{X}_n)$. Define

$$U_{\mathbb{R}_n}(h) = \sum\sum_{i \neq j} \left(W_{nR_i} - \frac{1}{n} \right) \left(W_{nR_j} - \frac{1}{n} \right) h(X_i, X_j)$$

and similarly

$$U_{\mathbb{R}_n}(h_l) = \sum\sum_{i \neq j} \left(W_{nR_i} - \frac{1}{n} \right) \left(W_{nR_j} - \frac{1}{n} \right) h_l(X_i, X_j),$$

where $(R_1, \dots, R_n)$ is a random vector taking each permutation of $(1, \dots, n)$ with equal probability and independent of $\mathbb{X}_n$ and $\mathbb{W}_n$. The rowwise exchangeability of the W_{ni}'s implies that $U_{\mathbb{W}_n}(h)$ and $U_{\mathbb{R}_n}(h)$ and also $U_{\mathbb{W}_n}(h_l)$ and $U_{\mathbb{R}_n}(h_l)$ have the same distribution. Therefore

$$\sup_{y \in \mathbb{R}} \left| P\{nU_{\mathbb{R}_n}(h_l) \le y | \mathbb{X}_n\} - P\{Y_l \le y\} \right| \tag{3.5}$$

is distributional equivalent to $T_{n5}(\mathbb{X}_n)$ and

$$P\{n|U_{\mathbb{R}_n}(h - h_l)| \ge \varepsilon | \mathbb{X}_n\} \tag{3.6}$$

is distributional equivalent to $T_{n4}(\mathbb{X}_n)$. If we add in (3.5) and (3.6) conditioning w.r.t. $\mathbb{W}_n$ the randomness in the probability statements only comes in through the ranks $R_1, \dots, R_n$. It then follows from arguments similar to the ones in the proof of Theorem 2.1 in Mason and Newton (1992) that, to control the behavior of (3.5) and (3.6), it is sufficient to show

$$\sup_{y \in \mathbb{R}} \left| P\{nU_{\mathbb{R}_n}(h_l) \le y | \mathbb{X}_n, \mathbb{W}_n\} - P\{Y_l \le y\} \right| \to_{\tilde{P}} 0, \qquad n \to \infty, \tag{3.7}$$

with P-probability 1, and to establish the inequality

$$\limsup_{n \to \infty} P\left\{n \left| U_{\mathbb{R}_n}(h - h_l) \right| \ge \varepsilon | \mathbb{X}_n, \mathbb{W}_n \right\} \le \frac{\eta C}{\varepsilon^2} \quad \text{a.s. } [P], \tag{3.8}$$

for some absolute constant C. Note once more that by adding the conditioning on $\mathbb{W}_n$ the problem completely reduces to a rank statistic problem (upon the conditioning). This idea is taken from Mason and Newton (1992), where it is used to handle generalized bootstrapped means.

We first show the validity of (3.8). We have

$$\begin{aligned} &P\left\{n \left| U_{\mathbb{R}_n}(h - h_l) \right| \ge \varepsilon | \mathbb{X}_n, \mathbb{W}_n \right\} \\ &\quad \le \frac{n^2}{\varepsilon^2} E_R \left[\left(\sum\sum_{i \neq j} g_l(X_i, X_j) \left(W_{nR_i} - \frac{1}{n} \right) \left(W_{nR_j} - \frac{1}{n} \right) \right)^2 | \mathbb{X}_n, \mathbb{W}_n \right], \end{aligned} \tag{3.9}$$

where E_R means expectation w.r.t. the ranks. An upper bound for the expectation in the r.h.s. of (3.9) is given by

$$E_R\left[\prod_{l=1}^{4}\left(W_{nR_l}-\frac{1}{n}\right)\right]A_{n1}(\mathbb{X}_n)$$
$$+\,4E_R\left[\left(W_{nR_1}-\frac{1}{n}\right)^2\prod_{l=2}^{3}\left(W_{nR_l}-\frac{1}{n}\right)\right]A_{n2}(\mathbb{X}_n) \tag{3.10}$$
$$+\,2E_R\left[\prod_{l=1}^{2}\left(W_{nR_l}-\frac{1}{n}\right)^2\right]A_{n3}(\mathbb{X}_n),$$

with

$$A_{n1}(\mathbb{X}_n)=\sum\sum\sum\sum_{i\neq j\neq s\neq t} g_l(X_i,X_j)g_l(X_s,X_t),$$
$$A_{n2}(\mathbb{X}_n)=\sum\sum\sum_{i\neq j\neq s} g_l(X_i,X_j)g_l(X_i,X_s),$$
$$A_{n3}(\mathbb{X}_n)=\sum\sum_{i\neq j} g_l^2(X_i,X_j).$$

Upper bounds for the expectations in (3.10) are obtained from the following lemma, which is essentially inequality (2.9) in Mason (1981).

LEMMA. *Let* $(W_{n1},\dots,W_{nn})$ *satisfy* (W.1) *and* (W.2). *For* $l\in\mathbb{N}_0$ *and for* $\alpha_1,\dots,\alpha_l\in\mathbb{N}_0$, *there exists a constant* $C=C_l(\alpha_1,\dots,\alpha_l)$ *such that*

$$\left|E_R\left[\prod_{i=1}^{l}\left(W_{nR_i}-\frac{1}{n}\right)^{\alpha_i}\right]\right|\le Cn^{-l}\left(\sum_{i=1}^{n}\left(W_{ni}-\frac{1}{n}\right)^2\right)^{\sum_{i=1}^{l}\alpha_i/2}. \tag{3.11}$$

Combining (3.9)–(3.11), we obtain

$$P\left\{n\left|U_{\mathbb{R}_n}(h-h_l)\right|\ge\varepsilon|\mathbb{X}_n,\mathbb{W}_n\right\}$$
$$\le\frac{C}{\varepsilon^2}\left(n\sum_{i=1}^{n}\left(W_{ni}-\frac{1}{n}\right)^2\right)^2 \tag{3.12}$$
$$\times\left[n^{-4}A_{n1}(\mathbb{X}_n)+n^{-3}A_{n2}(\mathbb{X}_n)+n^{-2}A_{n3}(\mathbb{X}_n)\right].$$

From (U.2) and $E\varphi_{lj}(X_1)=0$, $j=1,\dots,s_l$, we have

$$Eg_l(X_i,X_j)=0,$$
$$Eg_l(X_i,X_j)g_l(X_i,X_s)=0$$

and

$$n^{-4}A_{n1}(\mathbb{X}_n) \to 0, \qquad n \to \infty \quad \text{a.s.}\,[P], \tag{3.13}$$

$$n^{-3}A_{n2}(\mathbb{X}_n) \to 0, \qquad n \to \infty \quad \text{a.s.}\,[P]. \tag{3.14}$$

We also have

$$n^{-2}A_{n3}(\mathbb{X}_n) \to Eg_l^2(X_1, X_2) \quad \text{a.s.}\,[P]. \tag{3.15}$$

We therefore can conclude (3.8) from (3.12)–(3.15) and (W.3).

To establish (3.7), define S_{nj} and Z_{nj} as follows:

$$S_{nj}(R_1, \dots, R_n) \equiv S_{nj} = n^{1/2} \sum_{i=1}^{n} \left(W_{nR_i} - \frac{1}{n} \right) \varphi_{lj}(X_i),$$

$$Z_{nj}(R_1, \dots, R_n) \equiv Z_{nj} = n \sum_{i=1}^{n} \left(W_{nR_i} - \frac{1}{n} \right)^2 \varphi_{lj}^2(X_i).$$

Then note that

$$nU_{\mathbb{R}_n}(h_l) = \sum_{j=1}^{s_l} t_{lj}\left(S_{nj}^2 - Z_{nj}\right).$$

Therefore, similar to (3.4), we have the following upper bound for (3.7):

$$\begin{aligned} &\sup_{y \in \mathbb{R}} \left| P\{nU_{\mathbb{R}_n}(h_l) \le y | \mathbb{X}_n, \mathbb{W}_n\} - P\{Y_l \le y\} \right| \\ &\qquad \le T_{n6}(\mathbb{X}_n, \mathbb{W}_n) + T_{n2} + T_{n7}(\mathbb{X}_n, \mathbb{W}_n), \end{aligned} \tag{3.16}$$

with T_{n2} as in (3.4) and

$$\begin{aligned} &T_{n6}(\mathbb{X}_n, \mathbb{W}_n) \\ &\quad = \sup_{y \in \mathbb{R}} \left| P\left\{ \sum_{j=1}^{s_l} t_{lj}\left(S_{nj}^2 - \sigma_{l,jj}\right) \le y | \mathbb{X}_n, \mathbb{W}_n \right\} - P\{Y_l \le y\} \right|, \end{aligned} \tag{3.17}$$

$$T_{n7}(\mathbb{X}_n, \mathbb{W}_n) = P\left\{ \left| \sum_{j=1}^{s_l} t_{lj}(Z_{nj} - \sigma_{l,jj}) \right| \ge \varepsilon | \mathbb{X}_n, \mathbb{W}_n \right\}. \tag{3.18}$$

Since s_l is a fixed integer, showing

$$T_{n7}(\mathbb{X}_n, \mathbb{W}_n) \to_{\tilde{P}} 0, \qquad n \to \infty \quad \text{a.s.}\,[P]$$

is equivalent to showing that

$$P\{|Z_{nj} - \sigma_{l,jj}| \ge \varepsilon | \mathbb{X}_n, \mathbb{W}_n\} \to_{\tilde{P}} 0, \qquad n \to \infty \quad \text{a.s.}\,[P], \tag{3.19}$$

for all $j = 1, \dots, s_l$. Since Z_{nj} is a simple linear rank statistic, we easily obtain [use (W.3) and (W.4)]

$$E_R(Z_{nj}) = n \sum_{k=1}^{n} \left(W_{nk} - \frac{1}{n} \right)^2 \frac{1}{n} \sum_{i=1}^{n} \varphi_{lj}^2(X_i) \to_{\tilde{P}} \sigma_{l,jj}, \qquad n \to \infty \quad \text{a.s. } [P], \tag{3.20}$$

$$\begin{aligned} \mathrm{Var}_R(Z_{nj}) &\le n^2 \sum_{k=1}^{n} \left(W_{nk} - \frac{1}{n} \right)^4 \frac{1}{n-1} \sum_{i=1}^{n} \varphi_{lj}^4(X_i) \\ &\le n \max_{1 \le k \le n} \left(W_{nk} - \frac{1}{n} \right)^2 n \sum_{k=1}^{n} \left(W_{nk} - \frac{1}{n} \right)^2 \\ &\quad \times \frac{1}{n-1} \sum_{i=1}^{n} \varphi_{lj}^4(X_i) \to_{\tilde{P}} 0, \qquad n \to \infty \quad \text{a.s. } [P]. \end{aligned} \tag{3.21}$$

From (3.20) and (3.21) it follows that (3.19) and therefore also (3.18) are immediate.

We finally handle $T_{n6}(\mathbb{X}_n, \mathbb{W}_n)$ defined in (3.17). The convergence of $(S_{n1}, \dots, S_{ns_l})$ as well as the identification of the limit follows directly from the Cramér–Wold device. We indeed have for any set of constants $\gamma_1, \dots, \gamma_{s_l}$ that, conditioned on $\mathbb{X}_n$ and $\mathbb{W}_n$,

$$\sum_{j=1}^{s_l} \gamma_j S_{nj} = n^{1/2} \sum_{i=1}^{n} \left(W_{nR_i} - \frac{1}{n} \right) \sum_{j=1}^{s_l} \gamma_j \varphi_{lj}(X_i)$$

converges in distribution (w.r.t. the randomness due to the ranks) to a normal distribution with mean 0 and variance

$$E\left(\sum_{j=1}^{s_l} \gamma_k \varphi_{lj}(X_1) \right)^2 = \sum_{j=1}^{s_l} \sum_{k=1}^{s_l} \gamma_j \gamma_k \sigma_{l,jk}$$

[see the proof of Corollary 2.2 in Mason and Newton (1992)]. Since $\sum_{j=1}^{s_l} t_{lj}(S_{nj}^2 - \sigma_{l,jj})$ is a continuous function, we immediately obtain that

$$T_{n6}(\mathbb{X}_n, \mathbb{W}_n) \to_{\tilde{P}} 0, \qquad n \to \infty \quad \text{a.s. } [P]. \qquad \square$$

REMARK. In an earlier version of the manuscript, we used

$$\sum_{k=1}^{K} \lambda_k \varphi_k(x) \varphi_k(y)$$

to approximate $h(x, y)$, where the λ_k's and the φ_k's are the eigenvalues and the eigenfunctions mentioned in Section 1.

We then can follow the same type of proof, but an extra truncation argument is needed to handle $T_{n7}(\mathbb{X}_n, \mathbb{W}_n)$ as defined in (3.18) but now in terms of the λ_k's and the φ_k's. Using the present approach, this truncation argument is not needed due to the boundedness of the φ_{lj}'s. Another advantage of the present method is that it provides the appropriate approach to extend our main theorem to degenerate U-statistics with kernels of degree more than 2.

4. Appendix: proof of the lemma. The proof of the lemma is essentially contained in Mason (1981). We include it for completeness. First note that for $\beta = 2, 4, 6, \dots$ we obtain by direct calculation

$$\sum_{i=1}^{n}\left(W_{ni} - \frac{1}{n}\right)^{\beta} \le \left(\sum_{i=1}^{n}\left(W_{ni} - \frac{1}{n}\right)^{2}\right)^{\beta/2}.$$

For $\beta = 3, 5, \dots$ we have, with $\beta = 2\tilde{\beta} + 1$,

$$\begin{aligned}\sum_{i=1}^{n}\left|W_{ni} - \frac{1}{n}\right|^{2\tilde{\beta}+1} &= \sum_{i=1}^{n}\left(W_{ni} - \frac{1}{n}\right)^{2\tilde{\beta}}\left|W_{ni} - \frac{1}{n}\right| \\ &\le \sum_{i=1}^{n}\left(W_{ni} - \frac{1}{n}\right)^{2\tilde{\beta}}\left(\sum_{j=1}^{n}\left(W_{nj} - \frac{1}{n}\right)^{2}\right)^{1/2} \\ &\le \left(\sum\left(W_{ni} - \frac{1}{n}\right)^{2}\right)^{\tilde{\beta}+1/2}.\end{aligned}$$

So we obtain, for $k = 2, 3, \dots,$

$$\sum_{i=1}^{n}\left|W_{ni} - \frac{1}{n}\right|^{k} \le \left(\sum_{i=1}^{n}\left(W_{ni} - \frac{1}{n}\right)^{2}\right)^{k/2}. \tag{4.1}$$

Direct calculation yields

$$\begin{aligned}&\left|E_R\left[\prod_{i=1}^{l}\left(W_{nR_i} - \frac{1}{n}\right)^{\alpha_i}\right]\right| \\ &\quad = \frac{1}{n(n-1)\cdots(n-l+1)}\left|\sum_{\substack{i_1,\dots,i_l \\ \text{distinct}}}\left(W_{ni_1} - \frac{1}{n}\right)^{\alpha_1}\cdots\left(W_{ni_l} - \frac{1}{n}\right)^{\alpha_l}\right|.\end{aligned}$$

Case 1: $\alpha_1, \dots, \alpha_l$ *are all even integers.*

$$\begin{aligned}
&\sum_{\substack{i_1,\dots,i_l\\ \text{distinct}}} \left(W_{ni_1} - \frac{1}{n}\right)^{\alpha_1} \cdots \left(W_{ni_l} - \frac{1}{n}\right)^{\alpha_l} \\
&\quad = \sum_{i_1=1}^{n} \left(W_{ni_1} - \frac{1}{n}\right)^{\alpha_1} \Bigg\{ \cdots \sum_{i_{l-1}\neq i_1,\dots,i_{l-2}} \left(W_{ni_{l-1}} - \frac{1}{n}\right)^{\alpha_{l-1}} \\
&\qquad\qquad \times \sum_{i_l \neq i_1,\dots,i_{l-1}} \left(W_{ni_l} - \frac{1}{n}\right)^{\alpha_l} \Bigg\} \\
&\quad \le \sum_{i_1=1}^{n} \left(W_{ni_1} - \frac{1}{n}\right)^{\alpha_1} \sum_{i_2=1}^{n} \left(W_{ni_2} - \frac{1}{n}\right)^{\alpha_2} \cdots \sum_{i_l=1}^{n} \left(W_{ni_l} - \frac{1}{n}\right)^{\alpha_l} \\
&\quad \le \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{\sum_{i=1}^{l} \alpha_i/2},
\end{aligned}$$

where the last inequality follows from (4.1).

Case 2: $\alpha_1, \dots, \alpha_l$ *are* $l - q$ *even integers* (*say* $\gamma_1, \dots, \gamma_{l-q}$) *and* q *odd integers greater than or equal to* 3 (*say* $\eta_1, \dots, \eta_q$). Note: relabel if necessary. Then

$$\begin{aligned}
&\left| \sum_{\substack{i_1,\dots,i_l\\ \text{distinct}}} \left(W_{ni_1} - \frac{1}{n}\right)^{\gamma_1} \cdots \left(W_{ni_{l-q}} - \frac{1}{n}\right)^{\gamma_{l-q}} \left(W_{ni_{l-q+1}} - \frac{1}{n}\right)^{\eta_1} \cdots \left(W_{ni_l} - \frac{1}{n}\right)^{\eta_q} \right| \\
&\quad \le \sum_{\substack{i_1,\dots,i_l\\ \text{distinct}}} \left(W_{ni_1} - \frac{1}{n}\right)^{\gamma_1} \cdots \left(W_{ni_{l-q}} - \frac{1}{n}\right)^{\gamma_{l-q}} \\
&\quad \times \left(W_{ni_{l-q+1}} - \frac{1}{n}\right)^{\eta_1 - 1} \left|W_{ni_{l-q+1}} - \frac{1}{n}\right| \left(W_{ni_l} - \frac{1}{n}\right)^{\eta_q - 1} \left|W_{ni_l} - \frac{1}{n}\right| \\
&\quad \le \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{q/2} \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{(\sum_{j=1}^{l-q} \gamma_j + \sum_{\nu=1}^{q} (\eta_\nu - 1))/2} \\
&\quad = \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{\sum_{i=1}^{l} \alpha_i/2}.
\end{aligned}$$

Case 3: $\alpha_1 = \cdots = \alpha_q = 1$, $\alpha_{q+1} > 1, \dots, \alpha_l > 1$. Note: relabel if necessary. Define

$$A(q; \alpha_{q+1}, \dots, \alpha_l) = \left| \sum_{\substack{i_1, \dots, i_l \\ \text{distinct}}} \left(W_{ni_1} - \frac{1}{n}\right) \cdots \left(W_{ni_q} - \frac{1}{n}\right)\left(W_{ni_{q+1}} - \frac{1}{n}\right)^{\alpha_{q+1}} \cdots \left(W_{ni_l} - \frac{1}{n}\right)^{\alpha_l} \right|.$$

For $q = 0$ we are in one of the previous cases and for $q = 1$ we use

$$\sum_{i_1 \neq i_2, \dots, i_l} \left(W_{ni_1} - \frac{1}{n}\right) = - \sum_{j=2}^{l} \left(W_{ni_j} - \frac{1}{n}\right)$$

and the fact that

$$\left| \sum_{j=2}^{l} \left(W_{ni_j} - \frac{1}{n}\right) \right| \le (l-1) \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{1/2}$$

to obtain

$$A(1; \alpha_2, \dots, \alpha_l) \le (l-1) \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{1/2} A(0; \alpha_2, \dots, \alpha_l). \tag{4.2}$$

For $q > 1$ we have

$$\sum_{i_1 \neq i_2, \dots, i_l} \left(W_{ni_1} - \frac{1}{n}\right) = - \sum_{j=2}^{q} \left(W_{ni_j} - \frac{1}{n}\right) - \sum_{j=q+1}^{l} \left(W_{ni_j} - \frac{1}{n}\right) \tag{4.3}$$

and

$$\left| \sum_{j=q+1}^{l} \left(W_{ni_j} - \frac{1}{n}\right) \right| \le (l-q) \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{1/2}. \tag{4.4}$$

Using (4.3) and (4.4), we obtain by direct calculation

$$\begin{aligned} A(q; \alpha_{q+1}, \dots, \alpha_l) &\le (q-1) A(q-2; 2, \alpha_{q+1}, \dots, \alpha_l) \\ &\quad + (l-q) \left(\sum_{i=1}^{n} \left(W_{ni} - \frac{1}{n}\right)^2 \right)^{1/2} A(q-1; \alpha_{q+1}, \dots, \alpha_l). \end{aligned} \tag{4.5}$$

Since we already obtained the appropriate bound for $q = 0$ and $q = 1$, the proof follows from the recursive relation (4.5). □

Acknowledgments. The authors thank one of the referees for the suggestion to use a set of bounded functions rather than a set of eigenfunctions to approximate the kernel. They thank both referees for their constructive reports and David Mason for providing a preprint of Mason and Newton (1992).

REFERENCES

ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674.

DEHLING, H. (1989). The functional law of the iterated logarithm for von Mises functionals and multiple Wiener integrals. *J. Multivariate Anal.* **28** 177–189.

DEHLING, H. and MIKOSCH, T. (1992). Random quadratic forms and the bootstrap for U-statistics. Technical report.

DE WET, T. (1987). Degenerate U- and V-statistics. *South African Statist. J.* **21** 99–130.

HALL, P. and MAMMEN, E. (1992). On general resampling algorithms and their performance in distribution estimation. Preprint 682, Univ. Heidelberg.

JANSSEN, P. (1993). Weighted bootstrapping of U-statistics. *J. Statist. Plann. Inference* **34**. To appear.

MASON, D. M. (1981). On the use of a statistic based on sequential ranks to prove limit theorems for simple linear rank statistics. *Ann. Statist.* **9** 424–436.

MASON, D. M. and NEWTON, M. A. (1992). A rank statistic approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611–1624.

PRAESTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF STATISTICS
CHARLES UNIVERSITY
SOKOLOVSKÁ 83
18600 PRAHA
CZECHOSLOVAKIA

DEPARTMENT OF STATISTICS
LIMBURGS UNIVERSITAIR CENTRUM
UNIVERSITAIRE CAMPUS
B-3590 DIEPENBEEK
BELGIUM