The Annals of Statistics
1992, Vol. 20, No. 2, 655–674

ON THE BOOTSTRAP OF U AND V STATISTICS[1]

BY MIGUEL A. ARCONES AND EVARIST GINÉ

City University of New York and University of Connecticut

Bootstrap distributional limit theorems for U and V statistics are proved. They hold a.s., under weak moment conditions and without restrictions on the bootstrap sample size (as long as it tends to ∞), regardless of the degree of degeneracy of U and V. A testing procedure based on these results is outlined.

1. Introduction. 1. Let $h(x_1, \ldots, x_m)$ be a symmetric measurable function and let P be a probability measure on $\mathbb{R}$ (it is not relevant here that h be defined on $\mathbb{R}^m$ and P on $\mathbb{R}$; in fact, $\mathbb{R}$ could be replaced by any measurable space S throughout this paper.) Let $\{X_i\}$ be i.i.d. (P). The U and the V statistics based on h and P are defined as

$$U_m^n(h, P) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}) \tag{1.1}$$

and

$$V_m^n(h, P) = n^{-m} \sum_{i_1, \ldots, i_m = 1}^{n} h(X_{i_1}, \ldots, X_{i_m}). \tag{1.2}$$

Under moment conditions on h, these statistics, properly centered and normalized, converge in distribution (with some abuse of notation we say that they satisfy a central limit theorem, CLT). If h is nondegenerate for P, the limits are normal (depending on h and P), but if h is P-degenerate (definition in Section 2), then the limits are functionals of a Gaussian process, sometimes complicated ones. Therefore, bootstrap approximations to the laws of these statistics, or of their limits, are potentially useful. Bickel and Freedman (1981) show that the CLT for nondegenerate U and V statistics of order 2 ($m = 2$) can be bootstrapped just by replacing X_i by the bootstrap sample $X_{n1}, \ldots, X_{nn}$ [i.i.d. (P_n) with $P_n = n^{-1}\sum_{i=1}^{n} \delta_{X_i}$] at stage n. Bretagnolle (1983) studies the same type of bootstrap in the general case. In particular, he proves that in the degenerate case the bootstrap works in probability if the bootstrap sample size N_n satisfies $N_n/n \to 0$ and a.s. if $N_n (\log n)^b/n \to 0$ for some $b > 1$, under quite strong moment conditions on h. He also observes that for $N_n = n$, the usual bootstrap does not work for $h(x, y) = xy$ if $EX_1 = 0$. This example gives a clue to both, why the usual bootstrap fails and how to proceed. Let $EX_1 = 0$.

Received February 1990; revised June 1991.

[1]Research partially supported by NSF Grant DMS-90-00132. Most of this research was carried out at the Graduate Center and at the College of Staten Island of the City University of New York.

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

Key words and phrases. Bootstrap, U-statistics, von Mises functionals.

Then

$$(1.3)\qquad \begin{aligned} nU_2^n(h,P) &= n\binom{n}{2}^{-1}\sum_{1\le i<j\le n} X_iX_j \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i\right)^2 - \frac{1}{n-1}\sum_{i=1}^n X_i^2, \end{aligned}$$

but

$$(1.4)\qquad nU_2^n(h,P_n) = \frac{1}{n-1}\left(\sum_{i=1}^n X_{ni}\right)^2 - \frac{1}{n-1}\sum_{i=1}^n X_{ni}^2,$$

which, even after centering, is not a natural replica of $U_2^n(h,P)$. It is clearly more natural to apply the bootstrap CLT and the bootstrap law of large numbers to the right side of (1.3) to obtain that the bootstrap statistic that works is

$$(1.5)\qquad \frac{1}{n-1}\left(\sum_{i=1}^n (X_{ni}-E^*X_{ni})\right)^2 - \frac{1}{n-1}\sum_{i=1}^n (X_{ni}-E^*X_{ni})^2,$$

where E^* denotes conditional expectation given the sample. Expression (1.5) equals $nU_2^n(\bar{h}_n,P_n)$, where $\bar{h}_n(x,y) = h(x,y) - P_nh(x,\cdot) - P_nh(\cdot,y) + P_n^2h$. $\bar{h}_n$ is the orthogonal projection [in $L_2(P_n^2)$] of h onto the subspace of P_n-degenerate functions. This suggests how to proceed in the general case: If h is P-degenerate of order $r-1$ and if $\bar{h}_n$ is the P_n-orthogonal projection of h onto the subspace of P_n-degenerate functions of order $r-1$, then take as the bootstrap statistic not $U_m^n(h,P_n) - U_m^n(h,P)$, but $U_m^n(\bar{h}_n,P_n)$. Still better, since only the leading term in the Hoeffding expansion of $U_m^n(\bar{h}_n,P_n)$ contributes to the limit, take this term as the bootstrap statistic, which is

$$(1.6)\qquad \binom{m}{r}U_r^n(\pi_{r,m}^{P_n}h, P_n),$$

where, for a probability measure Q we write $(\pi_{r,m}^{Q}h)(x_1,\dots,x_r) = (\delta_{x_1}-Q)\cdots(\delta_{x_r}-Q)Q^{m-r}h$ (this notation is explained below).

The same idea applies for V statistics. Although the bootstrap results for V statistics can be deduced from those for U statistics via their Hoeffding decomposition, it seems more convenient to use in this case the decomposition not into U but into V statistics with P-canonical kernels. In this way, it follows very easily that the important V statistics that appear in the Taylor expansion of von Mises functionals, namely $(P_n - P)^m h$, can be naively bootstrapped.

The law of large numbers for U and V statistics can be bootstrapped under weak integrability conditions. This is proved in Section 4; see Athreya, Ghosh, Low and Sen (1984) for a different approach leading to a somewhat weaker result.

In Section 2 we prove the bootstrap CLT for U statistics, Theorem 2.4 (and Corollary 2.6). We find it convenient to give a short review of the CLT for U

statistics, and for this we follow, in essence, Bretagnolle's (1983) arguments (although we make no explicit reference to symmetrized tensor products or to the Wiener chaos). V statistics are treated in Section 3. The main result here is Theorem 3.5 (and Remark 3.6). Corollary 3.8 gives the bootstrap version of Filippova's (1961) limit theorem for $\{n^{m/2}(P_n - P)^m h\}_{n=1}^{\infty}$. In the case of U statistics we try to obtain results under minimal integrability conditions, but no such attempt is made for V statistics: in this case we just use Filippova's conditions. In Section 5, we outline a possible way in which the previous results can be used for testing hypotheses. Sometimes testing a composite hypothesis for P reduces to testing (H_0): A certain function $h(x_1, \dots, x_m)$ is P-centered and P-degenerate of a certain order versus (H_1): h is not P-centered or has a lower order of degeneracy. The above results can be applied to the construction of natural bootstrap tests of this sort which satisfy desirable properties.

2. The bootstrap CLT for U-statistics. In Section 2A we set up notation and sketch Bretagnolle's (1983) proof of the CLT for U-statistics. In Section 2B, we obtain the bootstrap limit theorem.

2A. *A short review of U-statistics.* A function $h: \mathbb{R}^m \to \mathbb{R}$ is *symmetric* if $h(x_1, \dots, x_m) = h(x_{\sigma(1)}, \dots, x_{\sigma(m)})$ for all $(x_1, \dots, x_m) \in \mathbb{R}^m$ and all permutations σ of $N_m = \{1, \dots, m\}$. Let P be a probability measure on $\mathbb{R}$ and let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with law P. Given a measurable symmetric function h on $\mathbb{R}^m$, the U statistic of order m based on h and P is

$$U_m^n(h, P) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \dots, X_{i_m}). \tag{2.1}$$

Given the notational complexity of U statistics, it is convenient (but not essential) to define the following two linear operators on (symmetric) functions $h: \mathbb{R}^m \to \mathbb{R}$. For all $x_1, \dots, x_m \in \mathbb{R}$, $n \in \mathbb{N}$, we let

$$\begin{gathered}(\sigma_m^n h)(x_1, \dots, x_n) = \sum_{1 \le i_1 < \cdots < i_m \le n} h(x_{i_1}, \dots, x_{i_m}) \quad \text{for } m \le n,\\ \sigma_m^n(h) \equiv 0 \quad \text{for } m > n \quad \text{and} \quad \sigma_0^n(c) = c \quad \text{for } c \in \mathbb{R}.\end{gathered} \tag{2.2}$$

The projection operator $\pi_{k,m}^P$ is defined on functions h of m variables in $L_1(\mathbb{R}^m, \mathscr{B}^m, P^m)$ and takes values in $L_1(\mathbb{R}^k, \mathscr{B}^k, P^k)$, $0 \le k \le m$, as follows: For $(x_1, \dots, x_m) \in \mathbb{R}^m$, $m \ne 0$,

$$(\pi_{k,m}^P h)(x_1, \dots, x_k) = (\delta_{x_1} - P) \cdots (\delta_{x_k} - P) P^{m-k} h \tag{2.3}$$

and $\pi_{0,0}^P(c) = c$ for $c \in \mathbb{R}$, where for measures Q_i on $\mathbb{R}$ we let $Q_1 \cdots Q_m h = \int_{\mathbb{R}^m} h(x_1, \dots, x_m)\, dQ_1(x_1) \cdots dQ_m(x_m)$. We will let $\pi_k^P = \pi_{k,k}^P$. If a function h on $\mathbb{R}^m$ is not necessarily symmetric, we will write $S_m h$ for its symmetrization, that is,

$$(S_m h)(x_1, \dots, x_m) = (m!)^{-1} \sum h(x_{\sigma(1)}, \dots, x_{\sigma(m)}), \tag{2.4}$$

where the sum extends over all permutations σ of N_m. Important types of functions in this theory are the *P-canonical functions*: $h\colon \mathbb{R}^m \to \mathbb{R}$ is P-canonical if h is symmetric and $Ph(x_1, \dots, x_{m-1}, \cdot) = \delta_{x_1} \cdots \delta_{x_{m-1}} Ph \equiv 0$ for P-almost all $x_1, \dots, x_{m-1} \in \mathbb{R}$. Obviously, square integrable P-canonical functions of different number of variables are orthogonal (with respect to P). Note that the functions $\pi_{k,m}^P h$, $1 \le k \le m$, are P-canonical and in particular orthogonal. Next we list some obvious properties of the σ and π operators which will be used below:

Lemma 2.1. *σ and π are linear. Moreover, for $h\colon \mathbb{R}^m \to \mathbb{R}$ symmetric:*

(a) *If h is P-canonical, then $E(\sigma_m^n h(X_1, \dots, X_n))^2 = \binom{n}{m} Eh^2(X_1, \dots, X_m)$ for $1 \le m \le n$.*

(b) *$E(\pi_{k,m}^P h(X_1, \dots, X_k))^2 \le Eh^2(X_1, \dots, X_m)$ for $1 \le k \le m$.*

(c) *$\sigma_m^n \circ \sigma_k^m(h) = \binom{n}{m}\binom{m}{k}\binom{n}{k}^{-1} \sigma_k^n(h)$ for $0 \le k \le m \le n$.*

(d) *$\pi_k^Q \circ \pi_{k,m}^P(h) = \pi_k^Q P^{m-k}(h)$, $0 \le k \le m$, $Q \in \mathscr{P}(\mathbb{R})$, where $P^{m-k}h$ is identified with $\delta_{x_1} \cdots \delta_{x_k} P^{m-k} h$.*

(e) *If h is P-canonical and $0 \le m, n \le r$, then $\pi_{n,r}^P \sigma_m^r h = h$ if $m = n$ and $\pi_{n,r}^P \sigma_m^r h = 0$ if $m \ne n$.*

Proof. (a), (c) and (d) are trivial. (b) follows using $E(X - EX)^2 \le EX^2$ conditionally. As for (e), we have

$$
\begin{aligned}
&\left(\pi_{n,r}^P \sigma_m^r h\right)(x_1, \dots, x_n) \\
&\quad = \sum_{1 \le i_1 < \cdots < i_m \le r} (\delta_{x_1} - P) \cdots (\delta_{x_n} - P) P^{r-n} g_{i_1, \dots, i_m},
\end{aligned}
$$

where $g_{i_1, \dots, i_m}(x_1, \dots, x_r) = h(x_{i_1}, \dots, x_{i_m})$. So, if no i_j is k, then $(\delta_{x_k} - P) g_{i_1, \dots, i_m} = 0$ (P is understood here as integrating x_k), that is, the terms of this sum for which $\{1, \dots, n\} \not\subset \{i_1, \dots, i_m\}$ are all zero. On the other hand, if $\{i_1, \dots, i_m\} \not\subset \{1, \dots, n\}$, that is, if some i_j is larger than n, then $P^{r-n} g_{i_1, \dots, i_m} = 0$ because h is P-canonical. Finally, if $\{i_1, \dots, i_m\} = \{1, \dots, n\}$, then $(\delta_{x_1} - P) \cdots (\delta_{x_n} - P) P^{r-n} h(x_1, \dots, x_n) = h(x_1, \dots, x_n)$ since h is P-canonical. (e) is thus proved. □

Using the $\sigma - \pi$ notation we obviously have that for h symmetric,

$$
\begin{aligned}
h(x_1, \dots, x_m) &= \delta_{x_1} \cdots \delta_{x_m} h = (\delta_{x_1} - P + P) \cdots (\delta_{x_m} - P + P) h \\
&= \sum_{k=0}^{m} \sigma_k^m \pi_{k,m}^P h(x_1, \dots, x_m), \qquad x_i \in \mathbb{R}.
\end{aligned} \tag{2.5}
$$

Then, since $U_m^n(h, P) = \binom{n}{m}^{-1} (\sigma_m^n h)(X_1, \dots, X_n)$ it follows from Lemma 2.1(c) and (2.5) that

$$U_m^n(h,P) = \sum_{k=0}^{m} \binom{m}{k}\binom{n}{k}^{-1} (\sigma_k^n \pi_{k,m}^P h)(X_1,\dots,X_n) \tag{2.6}$$
$$= \sum_{k=0}^{m} \binom{m}{k} U_k^n(\pi_{k,m}^P h, P).$$

This is the *Hoeffding decomposition* of U_m^n into a sum of U statistics associated to P-canonical functions (recall $\pi_{k,m}^P h$ is P-canonical for $k \ge 1$) plus $EU_m^n = P^m h = \sigma_0^n \pi_{0,m}^P h$.

The symmetric function $h(x_1,\dots,x_m)$ or the U-statistic $U_m^n(h,P)$ is P-degenerate of order $r-1$, $1 \le r \le m$, if $\delta_{x_1}\dots\delta_{x_{r-1}} P^{m-r+1}h \equiv P^m h$ a.s. but $\delta_{x_1}\cdots\delta_{x_r} P^{m-r} h$ is not a constant a.s.; it is nondegenerate (or degenerate of order zero) if $\delta_{x_1} P^{m-1} h$ is not a.s. equal to $P^m h$. It follows that h is P-degenerate of order $r-1$ if and only if $\pi_{1,m}^P h = \cdots = \pi_{r-1,m}^P h \equiv 0$ and $\pi_{r,m}^P h \not\equiv 0$. So, if h is degenerate of order $r-1$, $1 \le r \le m$, the Hoeffding decomposition of U is

$$U_m^n(h,P) - EU_m^n(h,P) = \sum_{k=r}^{m} \binom{m}{k}\binom{n}{k}^{-1} \sigma_k^n \pi_{k,m}^P h(X_1,\dots,X_m) \tag{2.7}$$
$$= \sum_{k=r}^{m} \binom{m}{k} U_k^n(\pi_{k,m}^P h, P).$$

In the proof of the bootstrap CLT we will require the law of large numbers and the central limit theorem for U statistics. The first, due to Hoeffding (1961) [see also Berk (1966)], uses the martingale structure (or alternatively the reverse martingale structure) of U-statistics, and is as follows:

$$E|h(X_1,\dots,X_m)| < \infty \text{ implies that } U_m^n(h,P) \to EU_m^n(h,P) \text{ a.s. as } n \to \infty. \tag{2.8}$$

We will also require an easy complement of this theorem [Sen (1974); see also Giné and Zinn (1990)]. Let $g\colon \mathbb{R}^d \to \mathbb{R}$ be a symmetric function; then

$$r > d \text{ and } E|g(X_1,\dots,X_d)|^{d/r} < \infty \text{ implies that } n^{-r} \sum_{1 \le i_1 < \cdots < i_d \le n} g(X_{i_1},\dots,X_{i_d}) \to 0 \text{ a.s. as } n \to \infty. \tag{2.9}$$

A consequence for V statistics is that, letting $\#(C)$ denote the cardinality of the set C,

$$\text{if for each } i_1,\dots,i_d\ E\left|g(X_{i_1},\dots,X_{i_d})\right|^{\#\{i_1,\dots,i_d\}/d} < \infty, \text{ then } n^{-d} \sum_{i_1,\dots,i_d=1}^{n} g(X_{i_1},\dots,X_{i_d}) \to Eg(X_1,\dots,X_d) \text{ a.s.} \tag{2.10}$$

The proof of the central limit theorem is crucial for the bootstrap. So, we sketch it here essentially following Bretagnolle (1983). We borrow some notation from Dynkin and Mandelbaum (1983) who have another approach to the

CLT for symmetric statistics. For ϕ: $\mathbb{R} \to \mathbb{R}$ measurable and bounded and with $P\phi = 0$, let $h_k^\phi(x_1,\dots,x_k) = \phi(x_1)\cdots\phi(x_k)$, $k \in \mathbb{N}$. Then it is classical [the Newton's identities; see, e.g., Jacobson (1951), page 110] that for each $k \in \mathbb{N}$ there is a polynomial R_k of k variables and of degree k, independent of n, such that for all $n \geq k$,

$$\sigma_k^n h_k^\phi(x_1,\dots,x_n) = R_k\left(\sum_1^n \phi(x_i), \sum_1^n \phi^2(x_i),\dots,\sum_1^n \phi^k(x_i)\right). \tag{2.11}$$

Let $\{G_P(\phi)$: $\phi \in L_2(\mathbb{R},\mathscr{B},P)$ and $P\phi = 0\}$, be the isonormal Gaussian process, that is, $G_P(\phi)$ is $N(0, P\phi^2)$ and $EG_P(\phi)G_P(\varphi) = P\phi\varphi$ for all P-mean zero $\phi,\varphi \in L_2(\mathbb{R},\mathscr{B},P)$. Let ϕ_j, $j = 1,\dots,J < \infty$, be P-centered and bounded real functions. Then the central limit theorem and the law of large numbers give

$$\begin{aligned}&\mathscr{L}\big(n^{-k/2}\sigma_k^n h_k^{\phi_j}(X_1,\dots,X_n): j = 1,\dots,J\big)\\ &\quad\to_w \mathscr{L}\Big(R_k\big(G_P(\phi_j),\sigma_{\phi_j}^2,0,\dots,0\big): j = 1,\dots,J\Big),\end{aligned} \tag{2.12}$$

where $\sigma_{\phi_j}^2 = P\phi_j^2$. It follows by orthogonality considerations that

$$R_k\big(G_P(\phi),\sigma_\phi^2,0,\dots,0\big) = (k!)^{-1/2}\sigma_\phi^k H_k\big(G_P(\phi)/\sigma_\phi\big),$$

where H_k is the kth Hermite polynomial [see, e.g., Bretagnolle (1983)]. If $h \in L_2(\mathbb{R}^k,\mathscr{B}^k,P^k)$, then h can be approximated in L_2 by functions g of the form $g = \sum_{i=1}^s c_i I_{A_{i1}}(x_1)\cdots I_{A_{ik}}(x_k)$. If moreover h is symmetric, since $h = S_k h$ and S_k has norm bounded by 1 as an operator acting on L_2, then h is approximated in $L_2(P^k)$ by functions of the form $\sum_{i=1}^s c_i S_k(I_{A_{i1}}(x_1)\cdots I_{A_{ik}}(x_k))$. This last sum, by polarization, can be written as $\sum b_i\psi_i(x_i)\cdots\psi_i(x_k)$ for functions ψ_i which are sums of some of the $I_{A_{i,j}}$. (Given $\varepsilon_1,\dots,\varepsilon_k$ i.i.d. with $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$ the polarization formula can be written as

$$\begin{aligned}k!S_k(f_1(x_1)\cdots f_k(x_k)) = E_\varepsilon\varepsilon_1\cdots\varepsilon_k(\varepsilon_1 f_1(x_1) + \cdots + \varepsilon_k f_k(x_1))\cdots\\ (\varepsilon_1 f_1(x_k) + \cdots + \varepsilon_k f_k(x_k)).)\end{aligned}$$

So we obtain that for any symmetric h there are simple functions ψ_{lj} and real numbers t_{lj} such that

$$h = \lim_{l\to\infty}\sum_{j=1}^{s_l} t_{lj}h_k^{\psi_{lj}} \quad \text{in } L_2(P^k). \tag{2.13}$$

Let $\phi_{lj} = \psi_{lj} - P\psi_{lj}$. If moreover h is P-canonical, then $h = \pi_k^P h = \lim_{l\to\infty}\sum_{j=1}^{s_l} t_{lj}h_k^{\phi_{lj}}$ in $L_2(P^k)$ because π_k^P is an operator of norm 1 on L_2. By (2.12) for each l the sequence of random variables $n^{-k/2}\sigma_k^n(\sum_{j=1}^{s_l} t_{lj}h_k^{\psi_{lj}}) = n^{-k/2}\sum_{j=1}^{s_l} t_{lj}\sigma_k^n(h_k^{\psi_{lj}})$ converges in distribution to

$$\sum_{j=1}^{s_l} t_{lj}R_k\big(G_P(\psi_{lj}),\sigma_{\psi_{lj}}^2,0,\dots,0\big).$$

On the other hand, by Lemma 2.1(a), as $l \to \infty$,

$$E\left[n^{-k/2}\left(\sigma_k^n h(X_1,\dots,X_n) - \sigma_k^n\left(\sum_{j=1}^{s_l} t_{lj} h_k^{\psi_{l,j}}\right)(X_1,\dots,X_n)\right)\right]^2$$

$$= n^{-k}\binom{n}{k} E\left(h - \sum_{j=1}^{s_l} t_{lj} h_k^{\psi_{lj}}\right)^2 \to 0.$$

These two observations yield weak convergence of the laws of

$$n^{-k/2}\sigma_k^n h(X_1,\dots,X_n)$$

by an easy triangle inequality argument [showing that if $Y_{nl} \to_d Y_l$ as $n \to \infty$ and $\sup_n d(Z_n, Y_{nl}) \to 0$ as $l \to \infty$ for some distance d metrizing weak convergence, then $Y_l \to_d Y$ for some Y and $Z_n \to_d Y$). The limit is obviously a functional of the process G_P in its Wiener chaos of order k.

Combining the CLT just proved with the Hoeffding decomposition (2.7) of a degenerate U-statistic, we obtain the CLT for U statistics:

(2.14) If $P^m h^2 < \infty$ and $U_m^n(h, P)$ is degenerate of order $r - 1$, $1 \le r \le m$, then the sequence $\{n^{r/2}(U_m^n(h, P) - EU_m^n(h, P))\}_{n=1}^{\infty}$ converges in distribution and its limit coincides with the limit of the sequence

$$\{r!\binom{m}{r} n^{-r/2}\sigma_r^n \pi_{r,m}^P h(X_1,\dots,X_n)\}_{n=1}^{\infty}.$$

[We are not interested here in obtaining the exact form of this limit, which may be quite complicated for $r > 1$ (it is normal for $r = 1$): The bootstrap limit theorem to be proved below provides approximations for it.] The limit theorem (2.14) is due to Rubin and Vitale (1980), in general and, for $m = 2$, to Hoeffding (1948a) (nondegenerate case) and Serfling (1980). Von Mises (1947) and Filippova (1961) obained a similar result for V statistics.

2B. *The bootstrap* CLT. Given $\{X_i\}_{i=1}^{\infty}$ i.i.d. (P), let $P_n(\omega) = n^{-1}\sum_{i=1}^{n}\delta_{X_i(\omega)}$ be their empirical distribution. Let $X_{n1}^{\omega},\dots,X_{nn}^{\omega}$ be i.i.d. random variables with law $P_n(\omega)$. In what follows the superscript ω in X_{ni}^{ω} will be omitted and we will also write P_n for $P_n(\omega)$. Let h be a kernel (i.e., a symmetric measurable function on $\mathbb{R}^m$) degenerate of order $r - 1$ for P.

As mentioned in the Introduction, there are some difficulties with the naive bootstrap $U_m^n(h, P_n) - U_m^n(h, P)$ of $U_m^n(h, P) - EU_m^n(h, P)$ [Bretagnolle (1983)]. The central limit theorem (2.14) suggests bootstrapping instead the first nonnull term of the Hoeffding expansion of $U_m^n(h, P) - EU_m^n(h, P)$ (see Remark 2.5 for an alternative, equivalent heuristics and bootstrap procedure).

Letting $\mathscr{L}^*$ denote conditional law with respect to the sample $\{X_i\}_{i=1}^{\infty}$, we will prove that, under mild integrability conditions,

$$\lim_{n\to\infty}\mathscr{L}^*\left(r!\binom{m}{r} n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n} h(X_{n1},\dots,X_{nn})\right) = \lim_{n\to\infty}\mathscr{L}\left(n^{r/2}(U_m^n(h,P) - EU_m^n(h,P))\right) \quad (2.15)$$

ω-almost surely. (Here and in what follows, the limit of a sequence of laws refers to its weak limit.) Note that the bootstrapped statistic is obtained from the leading term

$$r!\binom{m}{r} n^{-r/2} \sum_{1 \le i_1 < \cdots < i_r \le n} \left(\delta_{X_{i_1}} - P\right) \cdots \left(\delta_{X_{i_r}} - P\right) P^{m-r} h$$

by two substitutions: $\{X_i\}$ is replaced by $\{X_{ni}\}$ and P is replaced by P_n, to get

$$r!\binom{m}{r} n^{-r/2} \sum_{1 \le i_1 < \cdots < i_r \le n} \left(\delta_{X_{ni_1}} - P_n\right) \cdots \left(\delta_{X_{ni_r}} - P_n\right) P_n^{m-r} h,$$

a statistic that does not involve P. These two statistics are the rth terms in the Hoeffding decompositions of $n^{r/2} U_m^n(h, P)$ and $n^{r/2} U_m^n(h, P_n)$, respectively. We first show that

$$\begin{aligned} &\lim_{n \to \infty} \mathscr{L}^*\left(n^{-r/2} \sigma_r^n \pi_{r,m}^{P_n} g(X_{n1}, \dots, X_{nn})\right) \\ (2.16) \qquad &= \lim_{n \to \infty} \mathscr{L}\left(n^{-r/2} \sigma_r^n \pi_{r,m}^{P} g(X_1, \dots, X_n)\right) \quad \text{a.s.} \end{aligned}$$

for functions g of the form $\sum_{i=1}^s t_i h_m^{\phi_i}$ with ϕ_i bounded. Then we show that (2.16) also holds for functions g in $L_2(\mathbb{R}^m, \mathscr{B}^m, P^m)$ satisfying certain moment conditions.

Lemma 2.2. *Let $\phi_j \in L_\infty(\mathbb{R}, \mathscr{B}, P)$, $j = 1, \dots, J$. Then ω-a.s.,*

$$\begin{aligned} &\lim_{n \to \infty} \mathscr{L}^*\left(n^{-r/2} \sigma_r^n \pi_{r,m}^{P_n} h_m^{\phi_j}(X_{n1}, \dots, X_{nn}) \colon j = 1, \dots, J\right) \\ (2.17) \qquad &= \lim_{n \to \infty} \mathscr{L}\left(n^{-r/2} \sigma_r^n \pi_{r,m}^{P} h_m^{\phi_j}(X_1, \dots, X_n) \colon j = 1, \dots, J\right). \end{aligned}$$

Proof. We have

$$\begin{aligned} \pi_{r,m}^{P_n} h_m^{\phi_j}(x_1, \dots, x_r) &= \left(P_n(\phi_j)\right)^{m-r} (\delta_{x_1} - P_n)(\phi_j) \cdots (\delta_{x_r} - P_n)(\phi_j) \\ &= \left(P_n(\phi_j)\right)^{m-r} h_r^{\phi_j - P_n \phi_j}. \end{aligned}$$

Hence, by (2.11),

$$\begin{aligned} &n^{-r/2} \sigma_r^n \pi_{r,m}^{P_n} h_m^{\phi_j}(X_{n1}, \dots, X_{nn}) \\ &\quad = \left(P_n(\phi_j)\right)^{m-r} R_r\Bigg(n^{-1/2} \sum_{i=1}^n \left(\phi_j(X_{ni}) - P_n \phi_j\right), \\ &\qquad\qquad n^{-1} \sum_{i=1}^n \left(\phi_j(X_{ni}) - P_n \phi_j\right)^2, \dots, \\ &\qquad\qquad n^{-r/2} \sum_{i=1}^n \left(\phi_j(X_{ni}) - P_n \phi_j\right)^r\Bigg). \end{aligned}$$

Similarly,

$$n^{-r/2}\sigma_r^n \pi_{r,m}^P h_m^{\phi_j}(X_1,\dots,X_n)$$
$$= \left(P(\phi_j)\right)^{m-r} R_r\Bigg(n^{-1/2}\sum_{i=1}^n (\phi_j(X_i)-P\phi_j),$$
$$n^{-1}\sum_{i=1}^n (\phi_j(X_i)-P\phi_j)^2,\dots,$$
$$n^{-r/2}\sum_{i=1}^n (\phi_j(X_i)-P\phi_j)^r\Bigg).$$

By the bootstrap CLT in $\mathbb{R}^J$ [Bickel and Freedman (1981)], ω-a.s.,

$$\lim_{n\to\infty} \mathscr{L}^*\left(n^{-1/2}\sum_{i=1}^n (\phi_j(X_{ni})-P_n\phi_j): j=1,\dots,J\right)$$
$$= \lim_{n\to\infty} \mathscr{L}\left(n^{-1/2}\sum_{i=1}^n (\phi_j(X_i)-P\phi_j): j=1,\dots,J\right),$$

and by the strong law of large numbers and the bootstrap LLN in $\mathbb{R}$ (loc. cit.),

$$P_n\phi_j \to P\phi_j \quad \text{a.s.} \quad \text{and} \quad n^{-1}\sum_{i=1}^n \phi_j^s(X_{ni}) \to P\phi_j^s \quad \text{in } \Pr^*,\ \omega\text{-a.s.}$$

Since polynomials commute with weak limits, we obtain that a.s.,

$$\lim_{n\to\infty} \mathscr{L}^*\left(n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n} h_m^{\phi_j}: j=1,\dots,J\right)$$
$$= \lim_{n\to\infty} \mathscr{L}\left(n^{-r/2}\sigma_r^n \pi_{r,m}^{P} h_m^{\phi_j}: j=1,\dots,J\right). \qquad \square$$

LEMMA 2.3. *Let h be a symmetric function in $L_2(\mathbb{R}^m, \mathscr{B}^m, P^m)$ satisfying the integrability condition: For each $(i_1,\dots,i_m)\in N_m^m$, if $d = \#\{i_1,\dots,i_m\}$, then*

$$E\left|h(X_{i_1},\dots,X_{i_m})\right|^{2d/m} < \infty. \tag{2.18}$$

Then, ω-a.s.,

$$\lim_{n\to\infty} \mathscr{L}^*\left(n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n} h(X_{n1},\dots,X_{nn})\right)$$
$$= \lim_{n\to\infty} \mathscr{L}\left(n^{-r/2}\sigma_r^n \pi_{r,m}^{P} h(X_1,\dots,X_n)\right). \tag{2.19}$$

PROOF. By (2.13), $h = \lim_{l\to\infty} g_l$ in $L_2(P^m)$, where g_l are functions of the form $\sum_{\text{finite}} t_j h_m^{\phi_j}$. By Lemma 2.2, the identity (2.19) holds for g_l. Now Lemma

2.1(a) and (b), with P_n replacing P, gives

$$\begin{aligned} E^*\left(n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n}(h-g_l)\right)^2 &= n^{-r}\binom{n}{r} E^*\left(\pi_{r,m}^{P_n}(h-g_l)\right)^2 \\ &\le (r!)^{-1} E^*(h-g_l)^2(X_{n1},\dots,X_{nm}) \\ &= (r!)^{-1} n^{-m} \sum_{i_1,\dots,i_m=1}^{n} (h-g_l)^2(X_{i_1},\dots,X_{i_m}) \\ &\to (r!)^{-1} E(h-g_l)^2(X_1,\dots,X_m) \end{aligned}$$

as $n \to \infty$ by (2.10) [note that $h - g_l$ satisfies (2.18) as g_l is bounded]. These two observations give the identity (2.19) for h by an easy triangle inequality [e.g., using a distance that metrizes weak convergence; see the paragraph prior to (2.14)]. □

The CLT for U-processes together with Lemma 2.3 immediately gives:

THEOREM 2.4. *Let P be a probability measure on $\mathbb{R}$ and let $h\colon \mathbb{R}^m \to \mathbb{R}$ be a measurable symmetric function P-degenerate of order $r-1$ for some $1 \le r \le m$. Let $\{X_i\}_{i=1}^\infty$, be i.i.d (P) and for each n, let $X_{nj}^\omega = X_{nj}$, $j = 1,\dots,n$, be i.i.d. $(P_n(\omega))$, where $P_n(\omega) = n^{-1}\sum_{i=1}^n \delta_{X_i}(\omega)$. Let h satisfy the integrability condition (2.18). Then the following sequences of probability distributions converge weakly and all have the same limits*:

(a) $\{\mathscr{L}(n^{r/2}(U_m^n(h,P) - EU_m^n(h,P)))\}_{n=1}^\infty$,

(b) $\{\mathscr{L}(r!\binom{m}{r} n^{-r/2}\sigma_r^n \pi_{r,m}^P h(X_1,\dots,X_n))\}_{n=1}^\infty$,

(c) $\{\mathscr{L}^*(r!\binom{m}{r} n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n} h(X_{n1},\dots,X_{nn}))\}_{n=1}^\infty$, *$\omega$-a.s.*

REMARK 2.5. A slightly different (but essentially equivalent to the above) rationale for the bootstrap of degenerate U-statistics goes as follows: Since $U_n^m(h,P)$ is P-degenerate of order $r-1$ and degeneracy is so crucial that it must be preserved when bootstrapping, the bootstrap statistic should not be $U_m^n(h,P_n)$ but instead $U_m^n(\bar{h}_n,P_n)$, where $\bar{h}_n$ is the P_n-orthogonal projection of h onto the subspace of functions which are P_n-degenerate of order $r-1$, namely,

$$\bar{h}_n(\omega) = h - \sigma_0^m \pi_{0,m}^{P_n(\omega)} h - \cdots - \sigma_{r-1}^m \pi_{r-1,m}^{P_n(\omega)} h. \tag{2.20}$$

Notice that by Lemma 2.1(e), for $0 \le k < r$, $\pi_{k,m}^{P_n}\bar{h}_n = 0$ and for $k \ge r$, $\pi_{k,m}^{P_n}\bar{h}_n = \pi_{k,m}^{P_n} h$, so that

$$U_m^n(\bar{h}_n, P_n) = \sum_{k=r}^{m} \binom{m}{k}\binom{n}{k}^{-1} (\sigma_k^n \pi_{k,m}^{P_n} h)(X_{n1},\dots,X_{nn}). \tag{2.21}$$

The leading term in (2.21), multiplied by $n^{r/2}$, is equivalent to the nth term of the sequence (c) in Theorem 2.4 and the remaining terms are $o(n^{-r/2})$ by Lemma 2.3 provided that (2.18) holds. Hence, we have:

COROLLARY 2.6. *If h is as in Theorem 2.4, then*

$$\{n^{r/2}U_m^n(\bar{h}_n, P_n)\}_{n=1}^{\infty} \tag{d}$$

has a.s. the same weak limit as the sequences (a)–(c) *in Theorem* 2.4.

REMARK 2.7 (The case $r = 1$). It is worth noting that for $r = 1$ the statistics (d) in Corollary 2.6 and (c) in Theorem 2.4 take, respectively, the forms

$$n^{1/2}(U_m^n(h, P_n) - P_n^m h) \tag{2.22}$$

and

$$mn^{-1/2} \sum_{i_1,\dots,i_m \le n} \left(h(X_{ni_1}^*, X_{i_2}, \dots, X_{i_m}) - h(X_{i_1}, \dots, X_{i_m})\right). \tag{2.23}$$

The centering $P_n^m h$ in (2.22) can be replaced by $U_m^n(h, P)$.

REMARK 2.8 (Relation to previous work). The bootstrap in the nondegenerate case ($r = 1$ in Theorem 2.4), under stronger integrability conditions, was obtained by Bickel and Freedman (1981) for $m = 2$ and by Bretagnolle (1983) for general m. Our bootstrap is different from Bretagnolle's in the degenerate case as explained in the Introduction. Babu (1984) shows that, under high moment conditions, if H is a twice differentiable function, then $n(H(\bar{X}_n) - H(EX_1) - H'(EX_1)(\bar{X}_n - EX_1))$ is asymptotically equivalent to $n(H(\bar{X}_n^*) - H(\bar{X}_n) - H'(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n))$ and, closer in spirit to our work, observes that the bootstrap linear term cannot be dropped even if $H'(EX_1) = 0$. (Here $\bar{X}_n$ is the sample mean and $\bar{X}_n^*$ is the mean of the bootstrap sample.)

REMARK 2.9 (The case $m = 2$). For $m = 2$ and $r = 2$ (the degenerate case) the bootstrap statistics corresponding to the sequences (c) and (d) coincide (up to multiplicative constants tending to 1) with

$$\begin{aligned} nU_2^n(\bar{h}_n, P_n) &= n\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \Bigg[h(X_{ni}, X_{nj}) - n^{-1}\sum_{k=1}^n h(X_{ni}, X_k) \\ &\qquad - n^{-1}\sum_{k=1}^n h(X_k, X_{nj}) + n^{-1}\sum_{k,l=1}^n h(X_k, X_l)\Bigg]. \end{aligned} \tag{2.24}$$

REMARK 2.10 (Different bootstrap sample sizes and other extensions). (i) Since the CLT and LLN in $\mathbb{R}^J$ can be bootstrapped for any bootstrap sample size $N = N_n \to \infty$ and since the $L_2(P_n)$ estimates in Lemmas 2.2 and 2.3 also work in this case, it follows that Theorem 2.4 and Corollary 2.6 also hold for bootstrap sample size $N \to \infty$ not necessarily equal to n. The sequences (c) and (d) are now, for $X_{n1}, \dots, X_{nN}$ (i.i.d. (P_n)), respectively,

$$\left\{\mathscr{L}^*\left(r!\binom{m}{r}N^{-r/2}\sigma_r^N\pi_{r,m}^{P_n}(h(X_{n1}, \dots, X_{nN}))\right)\right\}_{n=1}^{\infty}$$

and

$$\left\{\mathscr{L}^*\left(N^{r/2}U_m^N(\bar{h}_n, P_n)\right)\right\}_{n=1}^{\infty}.$$

(ii) The limit (2.12) holds also jointly in $k = 1, \dots, K < \infty$, hence the argument below (2.12) gives that if $h_1, \dots, h_K$ are P-degenerate kernels of order $r_1 - 1, \dots, r_K - 1$, respectively, then the vector of U statistics $(n^{r_1/2}U_m^n(h_1, P), \dots, n^{r_K/2}U_m^n(h_K, P))$ converges in distribution in $\mathbb{R}^K$ and the analogues of Theorem 2.4 and Corollary 2.6 hold. In other words, U statistics can be bootstrapped jointly.

(iii) Finally, the bootstrap also works for multisample U statistics in an analogous way.

3. The bootstrap CLT for V statistics. Given a symmetric measurable function $h: \mathbb{R}^m \to \mathbb{R}$ and a probability measure P on $\mathbb{R}$, the V statistic of order m based on h and P is defined as

$$V_m^n(h, P) = n^{-m} \sum_{i_1, \dots, i_m = 1}^{n} h(X_{i_1}, \dots, X_{i_m}) = P_n^m h, \tag{3.1}$$

where $\{X_i\}$ are i.i.d. (P). Every symmetric statistic, and $V_m^n(h, P)$ is one, admits a Hoeffding expansion into U-statistics with P-canonical kernels. So, the CLT and the bootstrap CLT for V_m^n can be deduced from the results in the previous section. This is done for a different bootstrap CLT in Bretagnolle (1983). Here, we will apply the same principles of Section 2 but not the results themselves. It is somewhat easier to decompose a V statistic into a sum of V statistics with P-canonical kernels (instead of U-statistics) and work with these. As in (2.6), we have

$$V_m^n(h, P) = P_n^m h = ((P_n - P) + P)^m h = \sum_{j=0}^{m}\binom{m}{j}(P_n - P)^j P^{m-j}h$$

$$= \sum_{j=0}^{m}\binom{m}{j} n^{-j} \sum_{i_1, \dots, i_j = 1}^{n} \left(\delta_{X_{i_1}} - P\right) \cdots \left(\delta_{X_{i_j}} - P\right)P^{m-j}h \tag{3.2}$$

$$= \sum_{j=0}^{m}\binom{m}{j} V_j^n\left(\pi_{j,m}^P h, P\right).$$

Now we must introduce some extra notation in order to account for repetition of indices. Given a partition $Q = (A_1, \dots, A_q)$ of $N_m = \{1, \dots, m\}$, where $q = \#Q$ and $A_i \neq \varnothing$, we let for $h\colon \mathbb{R}^m \to \mathbb{R}$,

$$J_Q(h)(x_1,\dots,x_q) = h(x_{i_1},\dots,x_{i_q}), \quad \text{where } i_j = k \text{ if } j \in A_k,\ k = 1,\dots,q. \tag{3.3}$$

With this notation we obviously have

$$V_m^n(h,P) = n^{-m} \sum_Q \sigma_q^n(q!S_q J_Q h)(X_1,\dots,X_n), \tag{3.4}$$

where the summation runs over all partitions Q of N_m. This is a decomposition of V_m^n into U statistics [each of which has in turn its Hoeffding decomposition: (3.2) is simpler]. Following Filippova (1961), we let $\tilde{L}_2(\mathbb{R}^m, \mathscr{B}^m, P^m) := \{h \in L_2(\mathbb{R}^m, \mathscr{B}^m, P^m)\colon J_Q h \in L_2(\mathbb{R}^q, \mathscr{B}^q, P^q)$ for each partition Q of $N_m\}$ and

$$\|h\|_{\tilde{L}^2} := \left(\sum_Q E(J_Q h)^2(X_1,\dots,X_q)\right)^{1/2}. \tag{3.5}$$

LEMMA 3.1. *Let $h\colon \mathbb{R}^m \to \mathbb{R}$ be a symmetric measurable function. Then:*
(a) $P^n\sigma_m^n h = \binom{n}{m} P^m h$ *for* $m \le n$,
(b) $V_k^n(\pi_{k,m}^P h, P) = P_n^k(\pi_{k,m}^P h) = (P_n - P)^k P^{m-k} h,\ k \le m$,
(c) *If* $h \in \tilde{L}_2(P^m)$, *then* $E[n^{m/2}(P_n - P)^m h]^2 \le c_m^2 \|h\|_{\tilde{L}_2}^2$, *where* c_m *depends only on* m.

PROOF. (a) and (b) are trivial [see (3.2)]. (c) is proved in Filippova (1961), Lemma 1 (von Mises lemma). □

Our goal is to bootstrap the CLT for $V_m^n(h,P)$ for h P-degenerate of order $r-1$, $1 \le r \le m$. In this case the sum (3.2) becomes

$$V_m^n(h,P) - EV_m^n(h,P) = \sum_{j=r}^m \binom{m}{j} V_j^n(\pi_{j,m}^P h, P). \tag{3.6}$$

We show that the first term is the leading term:

LEMMA 3.2. *If* $h \in \tilde{L}_2(P^m)$ *is P-degenerate of order* $r-1$, $1 \le r \le m$, *then*

$$n^r E\left[V_m^n(h,P) - EV_m^n(h,P) - \binom{m}{r} V_r^n(\pi_{r,m}^P h, P)\right]^2 = O(n^{-1}). \tag{3.7}$$

PROOF. Note that by (3.6) and by Lemma 3.1(b, c),

$$n^{r/2}\left\| V_m^n(h,P) - EV_m^n(h,P) - \binom{m}{r} V_r^n(\pi_{r,m}^P h, P)\right\|_2$$
$$\le \sum_{j=r+1}^m n^{r/2}\binom{m}{j}\left\|V_j^n(\pi_{j,m}^P h, P)\right\|_2 \le \sum_{j=r+1}^m n^{(r-j)/2}\binom{m}{j} c_j \|P^{m-j}h\|_{\tilde{L}_2}.$$

□

Filippova [(1961), Theorem 4] proves that if $f \in \tilde{L}_2(P^r)$, then

$$\begin{aligned}&\mathscr{L}\big(n^{r/2}(P_n - P)^r f\big)\\ (3.8)\qquad &\to_w \mathscr{L}\bigg(\int_0^1 \cdots \int_0^1 f(F^{-1}(x_1), \dots, F^{-1}(x_r))\, d\beta(x_1) \cdots d\beta(x_r)\bigg),\end{aligned}$$

where β is the Brownian bridge and F is the cumulative distribution function of P. We denote $\mu(f, P, r)$ the limit law in (3.8). Then, Lemma 3.1(b), (3.7) and (3.8) give the CLT for V statistics:

If $h \in \tilde{L}_2(P^m)$ is P-degenerate of order $r - 1$, then

$$\begin{aligned}(3.9)\qquad &\lim_{n\to\infty} \mathscr{L}(n^{r/2}(V_m^n(h, P) - EV_m^n(h, P)))\\ &= \lim_{n\to\infty} \mathscr{L}\bigg(n^{r/2}\binom{m}{r}V_r^n(\pi_{r,m}^P h, P)\bigg) = \mu\bigg(\binom{m}{r}\pi_{r,m}^P h, P, r\bigg).\end{aligned}$$

So, we have a situation analogous to that of Section 2 and will bootstrap in the same way. Let, as before, $X_{n1}^{\omega}, \dots, X_{nn}^{\omega}$ be i.i.d. $(P_n(\omega))$ and let $P_n^* = n^{-1}\sum_{i=1}^n \delta_{X_{ni}^{\omega}}$ be the empirical measure of the bootstrap sample. We drop the variable ω.

LEMMA 3.3. *Let* $-\infty = t_0 < t_1 < \cdots < t_{k-1} < t_k = \infty$ *and let* $A_j = (t_{j-1}, t_j]$, $j = 1, \dots, k$, *be the associated partition of* $\mathbb{R}$. *For constants* $g_{j_1,\dots,j_m}$, *let* $g(x_1, \dots, x_m)$ *be the function*

$$(3.10)\qquad g(x_1, \dots, x_m) = \sum_{j_1,\dots,j_m=1}^{k} g_{j_1,\dots,j_m} I_{A_{j_1}}(x_1) \cdots I_{A_{j_m}}(x_m).$$

Then

$$\begin{aligned}(3.11)\qquad &\lim_{n\to\infty} \mathscr{L}^*\big(n^{r/2}(P_n^* - P_n)^r P_n^{m-r} g\big)\\ &= \lim_{n\to\infty} \mathscr{L}\big(n^{r/2}(P_n - P)^r P^{m-r} g\big) \quad a.s.\end{aligned}$$

PROOF. By the bootstrap CLT and LLN in $\mathbb{R}^k$ [Bickel and Freedman (1981)], $\lim_{n\to\infty} \mathscr{L}^*(n^{1/2}(P_n^* - P_n)(A_i): i = 1, \dots, k) = \lim_{n\to\infty} \mathscr{L}(n^{1/2}(P_n - P)(A_i): i = 1, \dots, k)$ a.s. and $P_n(A_i) \to P(A_i)$ a.s. So the result follows because

$$\begin{aligned}n^{r/2}(P_n - P)^r P^{m-r} g = \sum_{j_1,\dots,j_m=1}^{k} &g_{j_1,\dots,j_m} n^{1/2}(P_n - P)(A_{j_1}) \cdots\\ &n^{1/2}(P_n - P)(A_{j_r}) P(A_{j_{r+1}}) \cdots P(A_{j_m}),\end{aligned}$$

$$\begin{aligned}n^{r/2}(P_n^* - P_n)^r P_n^{m-r} g = \sum_{j_1,\dots,j_m=1}^{k} &g_{j_1,\dots,j_m} n^{1/2}(P_n^* - P_n)(A_{j_1}) \cdots\\ &n^{1/2}(P_n^* - P_n)(A_{j_r}) P_n(A_{j_{r+1}}) \cdots P_n(A_{j_m})\end{aligned}$$

and polynomials commute with weak limits. □

LEMMA 3.4. *Let $f \in \tilde{L}_2(P^m)$. Then*

$$\lim_{n\to\infty} \mathscr{L}^*\left(n^{r/2}(P_n^* - P_n)^r P_n^{m-r} f\right) = \lim_{n\to\infty} \mathscr{L}\left(n^{r/2}(P_n - P)^r P^{m-r} f\right) \quad a.s. \tag{3.12}$$

PROOF. The set of functions g of the form (3.10) is dense in $\tilde{L}_2(P^r)$. Let then $f = \lim_{l\to\infty} g_l$ in $\tilde{L}_2(P^r)$ with g_l of the form (3.10). By Lemma 3.1(c),

$$E\left(n^{r/2}(P_n - P)^r P^{m-r}(f - g_l)\right)^2 \le c_r^2 \| P^{m-r}(f - g_l)\|^2_{\tilde{L}_2(P^r)} \le c_r^2 \| f - g_l\|^2_{\tilde{L}_2(P^m)}. \tag{3.13}$$

Moreover, again by Lemma 3.1(c),

$$E^*\left(n^{r/2}(P_n^* - P_n)^r P_n^{m-r}(f - g_l)\right)^2 \le c_r^2 \| f - g_l\|^2_{\tilde{L}_2(P_n^m)} = c_r^2 \sum_Q \left\| J_Q(f - g_l)\right\|^2_{L_2(P_n^q)}. \tag{3.14}$$

The last random variable in (3.14) is a sum of V statistics that satisfy the integrability condition in (2.10) (actually, with some room left). Hence, the law of large numbers for V statistics shows that

$$\limsup_{n\to\infty} E^*\left(n^{r/2}(P_n^* - P_n)^r P_n^{m-r}(f - g_l)\right)^2 \le c_r^2 \| f - g_l\|_{\tilde{L}_2(P^m)} \quad \text{a.s.} \tag{3.15}$$

Now the result follows from Lemma 3.3, (3.13) and (3.15). □

The CLT (3.9) and Lemma 3.4 give the bootstrap CLT for V statistics.

THEOREM 3.5. *Let $h(x_1, \dots, x_m)$ be a measurable symmetric function P-degenerate of order $r - 1$, $1 \le r \le m$, such that $h \in \tilde{L}_2(P^m)$. Then the following sequences of laws converge weakly to the same limit:*

(a) $\{\mathscr{L}(n^{r/2}(V_m^n(h, P) - EV_m^n(h, P)))\}_{n=1}^\infty$,

(b) $\{\mathscr{L}(n^{r/2}\binom{m}{r}V_r^n(\pi_{r,m}^P h, P))\}_{n=1}^\infty$,

(c) $\{\mathscr{L}^*(n^{r/2}\binom{m}{r}V_r^n(\pi_{r,m}^{P_n} h, P_n))\}_{n=1}^\infty$, *$\omega$-a.s.*

REMARK 3.6. As in Remark 2.5, if $\bar{h}_n(\omega)$ is the P_n-degenerate component of h of order $r - 1$ [(2.20)], then it follows that the sequence

$$\left\{\mathscr{L}^*\left(n^{r/2}V_m^n\left(\bar{h}_n, P_n\right)\right)\right\}_{n=1}^\infty \tag{d}$$

converges weakly ω-a.s. to the same limit as the sequences (a)–(c), under the hypothesis of Theorem 3.5.

REMARK 3.7. As in the case of U statistics, extensions of Theorem 3.5 are possible. In particular, the bootstrap sample size can be any $N_n \to \infty$.

A particular V-statistics of interest is the mth term in the Taylor expansion of a von Mises statistical functional, namely $(P_n - P)^m f$, for $f(x_1, \dots, x_m)$ symmetric, $f \in \tilde{L}_2(P^m)$. By Lemma 3.1(b), $(P_n - P)^m f = V_m^n(\pi_m^P f)$, so that this is the special case of (3.6) for $r = m$. Since we also have $(P_n^* - P_n)^m f = V_m^n(\pi_m^{P_n} f)$ (by the same lemma), the equivalence between (a) and (c) in Theorem 3.5 gives:

COROLLARY 3.8 (The bootstrap of Filippova's CLT). *Let $f \in \tilde{L}_2(P^m)$ be a symmetric function. Then ω-a.s.,*

$$\lim_{n\to\infty} \mathscr{L}^*\left(n^{m/2}(P_n^* - P_n)^m f\right) = \lim_{n\to\infty} \mathscr{L}\left(n^{m/2}(P_n - P)^m f\right) = \mu(f, P, m). \tag{3.16}$$

4. The bootstrap of the law of large numbers for U and V statistics.

The purpose of this section is to prove the bootstrap law of large numbers for U and V statistics under as weak moment conditions as possible. In particular the following result strictly contains the bootstrap law of large numbers in Athreya, Ghosh, Low and Sen (1984).

THEOREM 4.1. *Suppose that for each possible combination of integers $i_1, \dots, i_m$, $E|h(X_{i_1}, \dots, X_{i_m})|^{\#\{i_1,\dots,i_m\}/m} < \infty$. Then*

$$U_m^n(h, P_n) \to_{\Pr^*} Eh(X_1, \dots, X_m) \quad a.s.$$

and

$$V_m^n(h, P_n) \to_{\Pr^*} Eh(X_1, \dots, X_m) \quad a.s.$$

PROOF. First consider the U statistics case. The 0th term in the Hoeffding decomposition is $P_n^m h$ and, by the law of large numbers (2.10) for V statistics, $P_n^m h \to P^m h$ a.s. For $1 \le j \le m$, $U_j^m(\pi_{j,m}^{P_n} h, P_n) = U_j^m(\pi_{j,m}^{P_n} h I_{|h| \le n^{1/2}}, P_n) + U_j^m(\pi_{j,m}^{P_n}(h I_{|h| > n^{1/2}}), P_n)$. Now

$$\begin{aligned}
&\operatorname{Var}^*\left(U_j^n\left(\pi_{j,m}^{P_n}(h I_{|h| \le n^{1/2}}), P_n\right)\right) \\
&= \binom{n}{j}^{-1} E^*\left(\pi_{j,m}^{P_n}(h I_{|h| \le n^{1/2}})\right)^2 \\
&\le \binom{n}{j}^{-1} n^{-(2m-j)} \sum_{\alpha,\beta,\gamma=1}^{n} \left| h I_{|h| \le n^{1/2}}\left(X_{\alpha_1}, \dots, X_{\alpha_j}, X_{\beta_1}, \dots, X_{\beta_{m-j}}\right) \right. \\
&\qquad \left. \times h I_{|h| \le n^{1/2}}\left(X_{\alpha_1}, \dots, X_{\alpha_j}, X_{\gamma_1}, \dots, X_{\gamma_{m-j}}\right) \right| \\
&\le \binom{n}{j}^{-1} n^{(1/2)-m} \sum_{\alpha,\beta=1}^{n} \left| h\left(X_{\alpha_1}, \dots, X_{\alpha_j}, X_{\beta_1}, \dots, X_{\beta_{m-j}}\right) \right| \to 0 \quad \text{a.s.},
\end{aligned}$$

by the law of large numbers for V statistics. We also have

$$E^*\left|U_j^n\left(\pi_{j,m}^{P_n} hI_{|h|>n^{1/2}}, P_n\right)\right|$$

$$\le 2^j E^*|h(X_{n1},\dots,X_{nm})|I_{|h|>n^{1/2}}$$

$$= 2^j n^{-m} \sum_{i_1,\dots,i_m=1}^{n} \left|h(X_{i_1},\dots,X_{i_m})\right| I_{|h|>n^{1/2}} \to 0 \quad \text{a.s.},$$

again by the law of the large numbers.

As for the LLN for V statistics, we note that, using the decomposition (3.4) of a V-statistics into U statistics, it is enough to show that, for $q < m$, $n^{-(m-q)}U_q^n(J_Q h, P_n) \to_{\Pr^*} 0$ a.s. But $E^*|n^{-(m-q)}U_q^n(J_Q h, P_n)| \le n^{-m}\sum_{i_1,\dots,i_q=1}^{n}|(J_Q h)(X_{i_1},\dots,X_{i_q})|$, which converges to zero a.s. because of (2.9). □

5. A remark on applications. Usually a kernel h is degenerate only for a small class of probability measures. So, only some extra care is needed in order to apply the bootstrap CLT in the usual way to obtain bootstrap confidence intervals for $\theta(P) = P^m h$ or for related characteristics of P. A different application that exploits the above results for degenerate statistics could go as follows. Suppose we have a class $\mathscr{P}$ of probability measures, a class $\overline{\mathscr{P}} \subset \mathscr{P}$ and a kernel h such that h is P-centered and is P-degenerate of exact order $r-1$ for $P \in \mathscr{P}$ if and only if $P \in \overline{\mathscr{P}}$, and it is not centered or it is P-degenerate of lower order otherwise [such a class $\overline{\mathscr{P}}$ could be, e.g., the class of product probability measures on $\mathbb{R}^2$, for the kernel h in Hoeffding (1948b), or the class of symmetric measures about the origin for a suitable function h —as we see below]. Then the previous results could be useful in testing H_0: $P \in \overline{\mathscr{P}}$ against H_1: $P \in \mathscr{P} - \overline{\mathscr{P}}$. We could take as critical region $\{n^{r/2}|U_m^n(h, P)| > c\}$ or $\{n^{r/2}|V_m^n(h, P)| > c\}$ (often these two statistics are equivalent). Then, given the data $X_1,\dots,X_n$, we could estimate c using the quantiles of the bootstrap statistic

$$T_n^* = r!\binom{m}{r} n^{-r/2}\sigma_r^n \pi_{r,m}^{P_n} h(X_{n1},\dots,X_{nn}).$$

Under the null hypothesis and if the limit distribution is continuous, by Theorem 2.4,

$$\Pr\left\{n^{r/2}|(U_m^n(h,P)| > c\right\} \approx \Pr^*\{|T_n^*| > c\}$$

asymptotically, that is, the test would have asymptotically the correct level. Moreover, by Lemma 2.3, T_n^* converges weakly ω-a.s. for any P whereas $n^{r/2}|U_m^n(h, P)| \to \infty$ in probability for all $P \in \mathscr{P} - \overline{\mathscr{P}}$ and therefore the test would be consistent against all fixed alternatives. (Actually, local alternatives could also be considered.) The same comment applies to tests based on V statistics, by virtue of the results in Section 3. We give two simple examples.

EXAMPLE 5.1 (A bootstrap test for symmetry about zero). Take $\mathscr{P} = \{P \in \mathscr{P}(\mathbb{R}): \int x^2\,dP < \infty\}$ and $\overline{\mathscr{P}} = \{P \in \mathscr{P}: P$ is symmetric about zero$\}$. Consider the statistic

$$
\begin{aligned}
T_n &= \int_{-\infty}^{0} (F_n(x) + F_n(-x-) - 1)^2\,dx \\
&= \int_{-\infty}^{0} \big(P_n(I_{(-\infty,x]} - I_{[-x,\infty)})\big)^2\,dx
\end{aligned}
\tag{5.1}
$$

(a modification of an example in Filippova). Then if $P \in \overline{\mathscr{P}}$,

$$
T_n = \int_{-\infty}^{0} [(F_n - F)(x) + (F_n - F)(-x-)]^2\,dx = (P_n - P)^2 h, \tag{5.2}
$$

where

$$
h(u,v) = (|u| \wedge |v|)\big(I_{\{u,v>0\}} + I_{\{u,v<0\}} - I_{\{u>0,v<0 \text{ or } u<0,v>0\}}\big), \quad u, v \in \mathbb{R}. \tag{5.3}
$$

If P is not symmetric

$$
\begin{aligned}
T_n = (P_n - P)^2 h &+ \int_{-\infty}^{0} (F(x) + F(-x-) - 1)^2\,dx \\
&- 2\int_{-\infty}^{0} [(F_n - F)(x) + (F_n - F)(-x-)] \\
&\quad \times [F(x) + F(-x-) - 1]\,dx,
\end{aligned}
\tag{5.4}
$$

with the second summand different from zero by right continuity of F and the third tending to zero as $n \to \infty$ by Gilvenko–Cantelli. The limiting distribution of $n(P_n - P)^2 h$ is the law of a shift of an infinite linear combination of centered independent chi-square random variables of order 1, hence it is continuous. So, if we define $c_n^*(\alpha)$ by

$$
\Pr{}^*\big\{n\big|(P_n^* - P_n)^2 h\big| > c_n^*(\alpha)\big\} = \alpha, \tag{5.5}
$$

we have

$$
\Pr\{|nT_n| > c_n^*(\alpha)\} \to \alpha \quad \text{if } P \in \overline{\mathscr{P}}
$$

and

$$
\Pr\{|nT_n| > c_n^*(\alpha)\} \to 1 \quad \text{if } P \in \mathscr{P} - \overline{\mathscr{P}}
$$

by Corollary 3.8, (5.2) and (5.4). (Obviously this test is also consistent against local alternatives of the form $P + n^{-\lambda}\Delta$, P symmetric, Δ nonsymmetric and $\lambda < 1/2$.)

A similar comment applies for the test of symmetry based on the more interesting statistic

$$
T_n = \int_{-\infty}^{\infty} (F_n(x) + F_n(-x-) - 1)^2\,dF_n(x)
$$

which is a V statistic of order 3. If we take here $\mathscr{P} = \{P \in \mathscr{P}(\mathbb{R}): P$ is continuous} and $\bar{\mathscr{P}} = \{P \in \mathscr{P}$: P is symmetric about the origin}, then the associated function h satisfies $P^3 h = 0$ if and only if $P \in \bar{\mathscr{P}}$ and in this case it is degenerate of order 1. Hence Theorem 3.5 applies.

EXAMPLE 5.2 (Hoeffding's test for independence). Here we take $\mathscr{P} = \{P \in \mathscr{P}(\mathbb{R}^2)$: P has continuous joint and marginal densities}, $\bar{\mathscr{P}} = \{P \in \mathscr{P}$: P is the product of its marginals} and

$$T_n = U_5^n(h, P),$$

where

$$h((x_1, y_1), \dots, (x_5, y_5)) \\ = 4^{-1} S_5[\varphi(x_1, x_2, x_3)\varphi(x_1, x_4, x_5)\varphi(y_1, y_2, y_3)\varphi(y_1, y_4, y_5)]$$

with $\varphi(u, v, w) = I_{\{v \le u\}} - I_{\{w \le u\}}$. Let F be the cdf of P and $\Delta = ET_n = \int_{\mathbb{R}^2}[F(x,y) - F(x,\infty)F(\infty,y)]^2\, dF(x,y)$. Then Hoeffding (1948b) shows that $\Delta = 0$ if and only if $P \in \bar{\mathscr{P}}$, that if $P \in \bar{\mathscr{P}}$, then h is P-degenerate of order 1 and that the limiting distribution of nT_n is continuous (an infinite convolution with at least one absolutely continuous component). Hence if $c_n^*(\alpha)$ is defined as

$$Pr^*\left\{\left|2!\binom{5}{2} n^{-1}\sigma_2^n \pi_{2,5}^{P_n} h(X_{n1}, \dots, X_{nn})\right| > c_n^*(\alpha)\right\} = \alpha, \tag{5.6}$$

we have

$$\Pr^*\{|nT_n| > c_n^*(\alpha)\} \to \alpha \quad \text{if } P \in \bar{\mathscr{P}} \tag{5.7}$$

and

$$\Pr\{|nT_n| > c_n^*(\alpha)\} \to 1 \quad \text{if } P \in \mathscr{P} - \bar{\mathscr{P}}. \tag{5.8}$$

By Theorem 2.4, under $P \in \bar{\mathscr{P}}$, nT_n is asymptotically equivalent in distribution to the statistic in (5.6), so that (the limit being continuous) (5.7) follows; for any $P \in \mathscr{P} - \bar{\mathscr{P}}$, by Lemma 2.3, $c_n^*(\alpha)$ converges a.s. to a finite quantity, whereas $n^{1/2}(T_n - \Delta)$ converges in distribution (by the CLT for U-statistics) and $\Delta \neq 0$, hence (5.8) follows.

Generalizing the above examples given a von Mises differentiable functional $T = T(P)$, a test of the hypothesis $P \in \bar{\mathscr{P}}$, where $\bar{\mathscr{P}}$ is the set of all probability measures for which the first (or the several first) derivative (in the von Mises–Filippova sense) is zero, could be constructed just by proceeding as in the examples. The only problem is that the second (or the first nonzero) derivative, say h_P, may depend on P. The appropriate smoothness conditions on T at P should, however, allow for the replacement of h_P by h_{P_n} and therefore justify taking $n^{1/2}(P_n^* - P_n)^2(h_{P_n})$ as the basis for the computation of critical numbers [see, e.g., Dudley (1990) for a general framework on this for $r = 1$].

REFERENCES

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

BABU, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā Ser. A* **46** 86–93.

BERK, R. H. (1966). Limiting behavior of posterior distributions where the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

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

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré Sec. B (N.S.)* **3** 281–296.

DUDLEY, R. M. (1990). Nonlinear functionals of empirical processes and the bootstrap. In *Probability in Banach Spaces* **7** (E. Eberlein, J. Kuelbs and M. Maray eds.). Birkhäuser, Boston.

DYNKIN, E. B and MANDELBAUM, A. (1983). Symmetric statistics, Poisson point processes and multiple Wiener integrals. *Ann. Statist.* **11** 739–745.

FILIPPOVA, A. A. (1961). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7** 24–57.

GINÉ, E. and ZINN, J. (1990). Marcinkiewicz type laws of large numbers and convergence of moments for U-statistics. Unpublished manuscript.

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

HOEFFDING, W. (1948b). A non-parametric test of independence. *Ann. Math. Statist.* **19** 546–557.

HOEFFDING, W. (1961). The strong law of large numbers for U-statistics. Inst. Statistics, Univ. North Carolina, Mimeo Series 302.

JACOBSON, N. (1951). *Lectures on Abstract Algebra* **1**. Van Nostrand, Princeton, N.J.

RUBIN, H. and VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170.

SEN, P. K. (1974). On L_p-convergence of U-statistics. *Ann. Inst. Statist. Math.* **26** 55–60.

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

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

M.S.R.I.
1000 CENTENNIAL DRIVE
BERKELEY, CALIFORNIA 94720

DEPARTMENT OF MATHEMATICS, U-9
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269