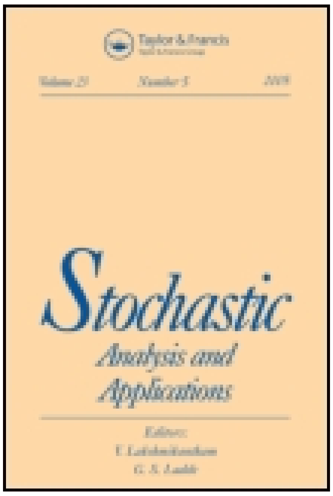

Stochastic Analysis and Applications
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsaa20

Strong convergence' for u-statistics in arrays of.row-wise exchangeable random variables

Ronald Frank Patterson [a]

[a] Georgia State University , Atlanta, Georgia 30303

Published online: 03 Apr 2007.

STOCHASTIC ANALYSIS AND APPLICATIONS, 7(1), 89-102 (1989)

STRONG CONVERGENCE FOR U-STATISTICS IN ARRAYS OF ROW-WISE EXCHANGEABLE RANDOM VARIABLES

Ronald Frank Patterson
Georgia State University
Atlanta, Georgia 30303

Key Words and Phrases: Exchangeable, U-Statistics, almost sure convergence.

ABSTRACT

Let $\left\{ X_{nk}: 1 \leq k \leq n,\ n \geq 1 \right\}$ be an array of row-wise exchangeable random variables. Strong convergence results are obtained for U-statistics, using suitable moment conditions and martingale techniques. Statistics considered for exemplification are the sample variance and Spearman's Rank Correlation Coefficient.

1. INTRODUCTION AND PRELIMINARIES.

Grams and Serfling [1] obtained strong convergence rates for U-statistics based on independent, identically distributed random variables. These results were obtained by using the projection of U, moment conditions on the "kernel" and by exploiting

Research partially supported by National Science Foundation Grant R11-8809494.

AMS 1979 Subject Classifications: Primary 60B12; secondary 60B99.

the reverse martingale property of U-statistics. The kernel $h(X_{i_1},\ldots, X_{i_m})$ is a symmetric function of its argument

where $Eh(X_{i_1},\ldots,X_{i_m}) = 0$.

The U-statistic for estimation of θ is

$U = \binom{n}{m}^{-1} \sum_c h(X_{i_1},\ldots, X_{i_m})$. Where $(i_1,\ldots,i_m)$ are taken

from the integers $(1, 2,\ldots, n)$. The projection of U is given by

$$\hat{U} = \sum_{i=1}^{n} E(U|X_i) - (n - 1)\,\theta.$$

In this paper, results similar to those produced by Grams and Serfling [1] will be obtained for arrays of row-wise exchangeable random variables.

In general, consider the following triangular array, $\left\{ X_{nk}: 1 \leq k \leq n,\ n \geq 1\right\}$ of row-wise exchangeable random variables. Let $h(X_{ni_1},\ldots,X_{ni_m})$ be a Symmetric function of its argument (i.e. taken from the n^{th} row of the array) where $(i_1,\ldots,i_m)$ is one of the $\binom{n}{m}$ combinations of the first n positive integers and $1 \leq m \leq n$. Next, define

the U-statistic $$U_{nn} = \binom{n}{m}^{-1} \sum_c h(X_{ni_1},\ldots,X_{ni_m}) \qquad (1.1)$$

The projection of U_{nn} is given by

$$\hat{U}_{nn} = \sum_{i=1}^{n} E(U_{nn}|X_{ni}) - (n - 1)\theta_n \qquad (1.2)$$

Where $$\theta_n = Eh(X_{n1},\ldots,X_{nm}) \qquad (1.3)$$

Next, define $h^*(X_{ni}) = E[h(X_{n1},\ldots X_{nm})|X_{ni}] - \theta_n$ (1.4)

Finally let C_{nn} be the σ-field of permutable events

generated by $\left\{U_{nn}, U_{n+1,n+1},\ldots\right\}$. (1.5)

In addition, strong convergence results are obtained for U-statistics based on triangular arrays of row-wise exchangeable random variables using moment conditions on the U-statistics and reverse martingale methods. The two main results will be given in Section 2. Section 3 will consist of two examples applying the theory. The first example expresses the sample variance in terms of a U-statistic. The second example, Spearman's rank correlation coefficient, is of considerable interest because the estimate can be written as the sum of a U-statistic and a remainder term that converges to zero almost surely.

2. Strong Convergence For Triangular Arrays

The main results of this section are two strong convergence results for U-statistics. The first result is a generalization of the Gram, Serfling result, for triangular arrays of independent, identically distributed random variables. The second strong convergence result is based on triangular arrays of exchangeable random variable. In the independent case, the U-statistic based on each row has the reverse martingale property. However, this condition does not exist between the rows of the triangular array. Hence, second moment and reverse martingale conditions are assumed on the kernel,

$h(X_{ni_1},\ldots,X_{ni_m})$.

For the second case dealing with exchangeability, p^{th} moment and reverse martingale conditions will be assumed on the U-statistic and its kernel, $h(X_{ni_1},\ldots,X_{ni_m})$, respectively.

First, five preliminary results will be presented for later use in sections 2 and 3. The first result, Lemma 2.1 shows that the U-statistic U_{nn}, is the expected value of $h(X_{n1},\ldots,X_{nn})$ conditioned on C_{nn}, the σ-field defined in (1.5). Lemma 2.3 expresses the difference of $\hat{U}_{nn}$ and θ_n as a weighted average of the $h^*(X_{ni})$'s, $1 \leq i \leq n$. Lemma 2.3 gives a convergence rate for the second moment of $|U_{nn} - \hat{U}_{nn}|$ and shows that $|\hat{U}_{nn} - \theta_n|$ and consequently $|U_{nn} - \theta_n|$ converge to 0 in the quadratic mean. Finally, Lemma 2.4 is a restatement of the reverse martingale inequality by Tucker [5] and will be given without proof.

Lemma 2.1. Let $\left\{x_{nk}: 1 \leq k \leq n,\ n \geq 1\right\}$ be an array of random variable that are row-wise exchangeable for each n. Let C_{nn} be the σ-field defined in (1.5). Then

$$U_{nn} = E[h(X_{n1},\ldots,X_{nm})|C_{nn}] \quad \text{a.s.} \qquad (2.1)$$

Proof Kingman [2] showed that

$$E[X_{ni}|g(X_{n1},\ldots,X_{nn})] = E[X_{n1}g(X_{n1},\ldots,X_{nn})]$$

Where $g(X_{n1},\ldots,X_{nn})$ is a symmetric function of $(X_{n1},\ldots,X_{nn})$. In particular, let $g(X_{n1},\ldots,X_{nn}) = I_A$, where $A \in C_{nn}$. Hence, by exchangeability

$$E[U_{nn}\, I_A] = E\Big[\binom{n}{m}^{-1}\sum_{c} h(X_{ni_1},\ldots,X_{ni_m})I_A\Big]$$

$$= \binom{n}{m}^{-1}\sum_{c} E[h(X_{ni_1},\ldots,X_{ni_m})I_A]$$

$$= \binom{n}{m}^{-1} \sum_{c} E[h(X_{n1}, \ldots, X_{nm}) I_A]$$

$$= E[h(X_{n1}, \ldots, X_{nm}) I_A].$$

Thus,

$$\int_A U_{nn} dP = \int_A h(X_{n1}, \ldots, X_{nm}) dP$$

$$= \int_A E[h(X_{n1}, \ldots, X_{nm}) | C_{nn}] dP$$

for all $A \varepsilon C_{nn}$. Hence, it follows that

$$U_{nn} = E[h(X_{n1}, \ldots, X_{nm}) | C_{nn}] \quad \text{a.s.} \qquad ///$$

<u>Lemma 2.2</u>. Let U_{nn}, $\hat{U}_{nn}$,and $h^*(X_{ni})$ be defined as in (1.1), (1.2), and (1.4) respectively. If $\left\{X_{nk}: 1 \leq k \leq n,\ n \geq 1\right\}$ are row-wise independent, identically distributed variables, then

$$\hat{U}_{nn} - \theta_n = \sum_{i=1}^{n} \frac{m}{n} h^*(X_{ni}). \tag{2,2}$$

<u>Proof</u>: By independence,

$$\hat{U}_{nn} - \theta_n = \sum_{i=1}^{n} E(U_{nn} | X_{ni}) - (n - 1)\theta_n - \theta_n$$

$$= \sum_{i=1}^{n} E[\binom{n}{m}^{-1} \sum_{c} h(X_{ni_1}, \ldots, X_{ni_m}) | X_{ni}] - n\theta_n$$

$$= \sum_{i=1}^{n} [\binom{n}{m}^{-1} \left\{ [\binom{n}{m} - \binom{n-1}{n-m}] E[h(X_{n1}, \ldots, X_{nm})] \right.$$

$$\left. + \binom{n-1}{n-m} E[h(X_{n1}, \ldots, X_{nm}) | X_{ni}] \right\} - \theta_n]$$

$$= \sum_{i=1}^{n} \left\{ \left[1 - \binom{n}{m}^{-1} \binom{n-1}{n-m} \right] \theta_n + \binom{n}{m}^{-1} \binom{n-1}{n-m} E[h(X_{n1},\ldots,X_{nm})|X_{ni}] - \theta_n \right\}$$

$$= \sum_{i=1}^{n} \left\{ [1 - \tfrac{m}{n}]\theta_n - \theta_n + \tfrac{m}{n} E[\, X_{n1},\ldots,X_{nm}|X_{ni}] \right\}$$

$$= \sum_{i=1}^{n} \left(\tfrac{m}{n} E[h(X_{n1},\ldots,X_{nm})|X_{ni}] - \tfrac{m}{n}\theta_n \right)$$

$$= \sum_{i=1}^{n} \tfrac{m}{n} h^*(X_{ni}).$$

<u>Lemma</u> <u>2.3.</u> Let $[X_{nk}: 1 \leq k \leq n,\ n \geq 1\}$ be an array of random variables that are row-wise independent for each n. Let $h(X_{ni_1},\ldots,X_{ni_m})$ be a symmetric function of distinct random variables from $(X_{n1},\ldots X_{nn})$. Let U_{nn}, $\hat{U}_{nn}$ be defined as in (1.1) and (1.2) respectively. If

$$E[h(X_{n1},\ldots,X_{nm})]^2 = o(n), \tag{2.3}$$

then

(i) $E|U_{nn} - \hat{U}_{nn}|^2 = o(n^{-2})$

(ii) $E|\hat{U}_{nn} - \theta_n|^2 \to 0$ a $n \to \infty$; (i) and (ii) imply that

(iii) $E|U_{nn} - \theta_n|^2 \to 0$ as $n \to \infty$

<u>Proof</u> The proof of (i) is virtually identical to the proof given in Grams and Serfling [1]. By (1.4), (2.2), independence and the conditional form of Jensen's inequality,

$$E|\hat{U}_{nn} - \theta_n|^2 = \frac{m^2}{n^2} E[\sum_{i=1}^{n} h^*(X_{ni})]^2$$

$$= \frac{m^2}{n^2} \left\{ \sum_{i=1}^{n} E[h^*(X_{ni})]^2 + n(n-1)E[h^*(X_{n1})h^*(X_{n2})] \right\}$$

$$= \frac{m^2}{n^2} \left\{ nE[|E(h(X_{n1},\ldots,X_{nm})|X_{n1}) - \theta_n]^2 + n(n-1)[E[h^*(X_{n1})]^2 \right\}$$

$$\le 2\frac{m^2}{n} E[E(h^2(X_{n1},\ldots,X_{nm})|X_{n1}) + \theta_n^2] + 0$$

$$\le 2\frac{m^2}{n}[Eh^2(X_{n1},\ldots,X_{nm}) + \theta_n^2]$$

$$\to 0 \quad \text{as } n \to \infty.$$

Now by (i) and (ii)

$$E|U_{nn} - \theta_n|^2 \le 4\,(E|U_{nn} - \hat{U}_{nn}|^2 + E|\hat{U}_{nn} - \theta_n|^2)$$

$$\to 0 \quad \text{as} \quad n \to \infty.$$

Lemma 2.4. If the random variables $\{X_n\}$ form a reverse martingale, then for every $\varepsilon > 0$

$$P[\sup_n |X_n| > \varepsilon] \le \frac{1}{\varepsilon}E|X_1| \tag{2.4}$$

Theorem 2.5. Let $\left\{ X_{nk}: 1 \le k \le n,\ n \ge 1 \right\}$ be a triangular array of random variables that are row-wise independent and identically distributed. Let $h(X_{ni_1},\ldots,X_{ni_m})$ be a symmetric function of m distinct random variables from $\left\{X_{n1},\ldots,X_{nn}\right\}$ where $1 \le m \le n$. Let C_{nn} denote the σ-field defined in (1.5). Let U_{nn} be defined as in (1.1).

If

(i) $E[h(X_{n1},\ldots,X_{nm})]^2 = o(n)$ a.s.

(ii) $\left\{E[h(X_{n1},\ldots,X_{nn})|C_{nn}]\right\}_{n\varepsilon\mathbf{N}}$ is a reverse martingale

then

$$U_{nn} \to 0 \quad \text{a.s.}$$

Proof: By (2.1), (2.4) and for each $\varepsilon > 0$,

$$P[\sup_{n \geq \ell} |U_{nn}| > \varepsilon] = P[\sup_{n \geq \ell} |E[h(_{n},\ldots,X_{nm})|C_{nn}|] > \varepsilon]$$

$$\leq E[|E[h(X_{\ell\ell},\ldots,X_{\ell m})|C_{\ell\ell}]|] \qquad (2.5)$$

Now, again by (2.1) and Holder's inequality,

$$E[|E[h(X_{\ell\ell},\ldots,X_{\ell m})|C_{\ell\ell}]|] = E|U_{\ell\ell}|$$

$$\leq \left(E|U_{\ell\ell}|^2\right)^{1/2} \qquad (2.6)$$

Thus, it follows from (i) and Lemma 2.4 that

$$P[\sup_{n \geq \ell} |U_{nn}| > \varepsilon] \leq (E|U_{\ell\ell}|^2)^{1/2}$$
$$\to 0 \quad \text{as} \quad n \to \infty. \qquad ///$$

Remarks: Since $\left\{E[h(X_{n1},\ldots,X_{nm})|C_{nn}]\right\}$ is a reverse martingale,

$$E[h(X_{n1},\ldots,X_{nm})] = E\left\{E[h(X_{n1}\ldots,X_{nm})|C_{n+1,n+1}]\right\}$$

$$= E[h(X_{n+1,1},\ldots,X_{n+1,m})].$$

This implies that $\Theta_n = \Theta_{n+1} = \Theta$ for all n. W.L.O.G., $E[h(X_{n1},\ldots,X_{nm})] = \Theta = 0$ for all n. It is to be noted that the moment condition (i) in Theorem 2.5 can be replaced by $E|hX_{n1},\ldots,X_{nm})|^p = 0(n)$ where $1 \leq p \leq 2$. By Hölder's inequality,

$(E|h(X_{n1},\ldots,X_{nm}|^p)^{1/p} \leq (E|h(X_{n1},\ldots,X_{nm})|^2)^{1/2}$ and the result follows from Theorem 2.5

The next strong convergence result is for triangular arrays of random variables that are row-wise exchangeable. In this setting, the p^{th} moment condition on the kernel is not sufficient to guarantee a strong law of large numbers. This is because conditions (i) and (ii) of Lemma 2.4 are no longer sufficient without independence. In particular, the nonorthogonal term in $E|\hat{U} - \Theta_n|^2$, $E[h^*(X_{n1})h^*(X_{n2})]$, is not zero which is necessary for the mean square convergence of $|\hat{U} - \Theta_n|$. Thus, the moment condition on the kernel is replaced by the p^{th} moment condition on $\mathbf{U}_{nn}$ in the next result, where $p \geq 1$.

<u>Theorem 2.6</u>. Let $\left\{ X_{nk}: 1 \leq k \leq n,\ n \geq 1 \right\}$ be a triangular array of random variables that are row-wise exchangeable. Let $h(X_{ni_1},\ldots,X_{ni_m})$ be a symmetric function of random variables from $\left\{ X_{n1},\ldots,X_{nn} \right\}$ where $1 \leq m \leq n$. Let C_{nn} denote the σ-field defined in (1.5). Consider the U-statistic defined for any integer m, $1 \leq m \leq n$ by

$$\mathbf{U}_{nn} = \binom{n}{m}^{-1} \sum_{c} h(X_{ni_1},\ldots, X_{ni_m})$$

If for some $p \geq 1$,

(i) $E(|\mathbf{U}_{nn}|^p) \to 0$ as $n \to \infty$

(ii) $\left\{E[h(X_{n1},\ldots,X_{nm})|C_{nn}]\right\}_{n\varepsilon\mathbf{N}}$ is a reverse martingale

Then $\mathbf{U}_{nn} \to 0$ a.s.

Proof: From (2.11) and (2.8), for each $\varepsilon > 0$

$$P[\sup_{n \geq \ell}|U_{nn}| > \varepsilon] = P[\sup_{n \geq \ell}|\binom{n}{m}^{-1}\sum h(X_{ni_1},\ldots,X_{ni_m})| > \varepsilon]$$

$$= P[\sup_{n \geq \ell}|E[h(X_{n1},\ldots,X_{nm})|C_{nn}]| > \varepsilon]$$

$$\leq \frac{1}{\varepsilon} E[|E[h(X_{\ell 1},\ldots,X_{\ell m})|C_{\ell\ell}]|]$$

$$= \frac{1}{\varepsilon}E|U_{\ell\ell}|.$$

By Holders's inequality,

$E|U_{\ell\ell}| \leq (E|U_{\ell\ell}|^p)^{1/p}$. Thus,

$$P[\sup_{n \geq \ell}|U_{nn}| > \varepsilon] \leq \frac{1}{\varepsilon}(E|U_{\ell\ell}|^p)^{1/p}$$

$$\to 0 \quad \text{as} \quad n \to \infty \qquad ///.$$

Remark: For $m = 1$ and $p = 2$, Theorem 2.6 becomes the Strong Laws of large numbers for triangular arrays of row-wise exchangeable random variables in Taylor and Patterson [4]. The condition

$E\left(\frac{1}{n}\sum_{i=1}^{n} X_{ni}\right)^2 \to 0$ is equivalent to the second moment

condition on X_{ni} and the condition relating to nonorthogonality.

Similarly $E|U_{nn}|^2 \to 0$ as $n \to \infty$ implies that

$E|h(X_{n1},\ldots,X_{nm})| = o(n)$ and that

$E[h(X_{ni_1},\ldots,X_{ni_m})h(X_{nj_1},\ldots,X_{nj_m})] \to 0$ as $n \to \infty$.

3. <u>Examples</u> Some typical examples of U-statistics are given by the sample variance and Spearman rank correlation coefficient.

<u>Example 1</u>. Let $X_1, \ldots, X_n$ be independent, identically distributed normal random variables where $EX_1 = 0$ and $EX_1^2 = \sigma^2$ for each n the sample variance is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \binom{n}{2}^{-1}\sum_{i<j} h(X_i, X_j)$$

Where $h(X_1, X_2) = \dfrac{(X_1 - X_2)^2}{2}$.

Let $X_{nk} = X_k$, $1 \le k \le n$ and $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_{nk}$

Define $H(X_{ni}, X_{nj}) = h(X_{ni}, X_{nj}) - \sigma^2$. The corresponding statistic is

$$\mathbf{U}_{nn} = \binom{n}{2}^{-1}\sum_{i<j} H(X_{ni}, X_{nj}). \quad \text{For } p = 2$$

$$E[\mathbf{U}_{nn}]^2 = E[X_1^2 - \sigma^2]^2 = \frac{2\sigma^4}{n-1} \to 0 \text{ as } n \to \infty.$$

By the reverse martingale condition of U-Statistics, $\left\{E[H(X_{n1}, \ldots, X_{nm}) | C_{nn}]\right\}_{n\varepsilon\mathbf{N}}$ is a reverse martingale and the hypothesis of Theorem 2.6 are satisfied.

<u>Remark</u>: For independent, identically distributed random variables with $E|X|^4 < \infty$ the hypothesis of Theorem 2.6 are also satisfied.

Example 2. Let $(Y_1, Z_1), \ldots (Y_n, Z_n)$ be a random sample from a continuous, bivariate population. Let $X_1 = (Y_1, Z_1), \ldots, X_n = (Y_n, Z_n)$. The rank correlation coefficient is given by

$$r_s = \frac{12}{n^3 - n} \sum [R_i - \tfrac{1}{2}(n + 1)][S_i - \tfrac{1}{2}(n + 1)]$$

To relate r_s to a U-statistic Lehmann [3] expressed r_s in the following manner:

$$r_s = \frac{3}{n^3 - n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} S(Y_i - Y_j)S(Z_i - Z_k) \tag{3.1}$$

$$S(u) = \begin{cases} -1 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases}$$

For m = 3, define the statistic

$U = \binom{n}{3}^{-1} \sum S(Y_i - Y_j)S(Z_i - Z_k)$ where the summation extends over $n(n - 1)(n - 2)$ triples of distinct subscripts. Thus, U can be expressed as the U-statistic

$$U_{nn} = \binom{n}{3}^{-1} \sum_{c} h(X_{ni}, X_{nj}, X_{nk}) \tag{3.2}$$

Where $X_{ni} = X_i$, $1 \leq i \leq n$. Here

$$\begin{aligned} h(X_{ni}, X_{nj}, X_{nk}) &= S(y_{ni} - y_{nj})S(Z_{ni} - Z_{nk}) + S(y_{ni} - y_{nk})S(Z_{ni} - Z_{nj}) \\ &+ S(y_{nj} - y_{ni})S(Z_{nj} - Z_{nk}) + S(y_{nj} - y_{nk})S(Z_{nj} - Z_{ni}) \\ &+ S(y_{nk} - y_{ni})S(Z_{nk} - Z_{nj}) + S(y_{nk} - y_{nj})S(Z_{nk} - Z_{ni}). \end{aligned}$$

Which is clearly a bounded symmetric function of its arguments. Thus, by (3.1) and (3.2)

$$r_s = \frac{1}{2} \left[\frac{n-2}{n+1} U_{nn} + \frac{3}{n^3 - n} D_{nn} \right] \tag{3.3}$$

Where $D_{nn} = \sum S(Y_{ni} - Y_{nj})S(Z_{ni} - Z_{nk})$ summed over all $3n^2 - 2n$ triples that are not distinct. It suffices to show that

$$\frac{D_{nn} - ED_{nn}}{n^3 - n} \to 0 \quad \text{a.s.}$$ since $U_{nn} - EU_{nn}$ clearly satisfy the hypothesis of Theorem 2.6. For $\varepsilon > 0$ and n sufficiently large,

$$P\left[\sup_{n \geq m} \left| \frac{D_{nn} - ED_{nn}}{n^3 - n} \right| > \varepsilon\right] \leq P\left[\frac{2(3_n^2 - 2_n)}{n^3 - n} > \varepsilon \right] = 0.$$

This implies that $\dfrac{D_n - ED_n}{n^3 - n} \to 0$ a.s ///

References

1. Grams, W. F. and Serfling, R.J. (1973) Convergence rates for U-statistics and related statistics. Ann. Statist., 1, 153-160.

2. Kingman, J.F.C. (1978) Uses of Exchangeability. Ann., Prob., 6, 183-197.

3. Lehmann, E.L. (1975) NonParametrics. Statistical Methods Based on Ranks. McGraw-Hill, New York.

4. Patterson, R. F. and Taylor, R. L. (1985) Strong Laws of Large Numbers For Triangular Arrays of Exchangeable Random Variables. Stochastic Annalysis and Appl's 3(2), 171-187

5. Tucker, Lt. G. (1967) A Graduate Course In Probability. Academic Press, New York.