

RATE OF CONVERGENCE IN DISTRIBUTION OF A LINEAR COMBINATION OF U -STATISTICS FOR A DEGENERATE KERNEL

Yamato, Hajime

Department of Mathematics and Computer Science, Kagoshima University

Kondo, Masao

Department of Mathematics and Computer Science, Kagoshima University

<https://doi.org/10.5109/13509>

出版情報 : Bulletin of informatics and cybernetics. 34 (1), pp.61-73, 2002-10. Research
Association of Statistical Sciences

バージョン :

権利関係 :

RATE OF CONVERGENCE IN DISTRIBUTION OF A LINEAR COMBINATION OF U-STATISTICS FOR A DEGENERATE KERNEL

By

Hajime YAMATO* and Masao KONDO†

Abstract

Associated with an estimable parameter, we consider a linear combination of U-statistics (Toda and Yamato, 2001) which includes V-statistic and LB-statistic. For a degenerate kernel, its asymptotic distribution (Yamato et al., 2001) is easily derived by the same method as Yamato and Toda (2001). We give the rate of this convergence in distribution.

Key Words and Phrases: Linear combination of U-statistics, order of degeneracy, rate of convergence.

1. Introduction

Let $\theta(F)$ be an estimable parameter of an unknown distribution function F which has a symmetric kernel $g(x_1, \dots, x_k)$ of degree $k (\geq 2)$ and X_1, \dots, X_n be a random sample of size n from F .

As an estimator of $\theta(F)$, Toda and Yamato (2001) introduced a linear combination Y_n of U-statistics as follows. Let $w(r_1, \dots, r_j; k)$ be a nonnegative and symmetric function of positive integers r_1, \dots, r_j such that $j = 1, \dots, k$ and $r_1 + \dots + r_j = k$, where k is the degree of the kernel g and fixed. We assume that at least one of $w(r_1, \dots, r_j; k)$'s is positive. For $j = 1, \dots, k$, let $g_{(j)}(x_1, \dots, x_j)$ be the kernel given by

$$g_{(j)}(x_1, \dots, x_j) = \frac{1}{d(k, j)} \sum_{r_1 + \dots + r_j = k}^+ w(r_1, \dots, r_j; k) g(x_1^{r_1}, \dots, x_j^{r_j}), \quad (1.1)$$

where

$$g(x_1^{r_1}, \dots, x_n^{r_n}) = g(\underbrace{x_1, \dots, x_1}_{r_1}, \dots, \underbrace{x_n, \dots, x_n}_{r_n})$$

and the summation $\sum_{r_1 + \dots + r_j = k}^+$ is taken over all positive integers r_1, \dots, r_j satisfying $r_1 + \dots + r_j = k$ with j and k fixed and $d(k, j) = \sum_{r_1 + \dots + r_j = k}^+ w(r_1, \dots, r_j; k)$ for $j = 1, 2, \dots, k$.

Let $U_n^{(j)}$ be the U-statistic associated with this kernel $g_{(j)}(x_1, \dots, x_j; k)$ for $j =$

* Department of Mathematics and Computer Science, Kagoshima University, Kagoshima 890-0065, Japan.

† Department of Mathematics and Computer Science, Kagoshima University, Kagoshima 890-0065, Japan.

$1, \dots, k$. The kernel $g_{(j)}(x_1, \dots, x_j; k)$ is symmetric because of the symmetry of $w(r_1, \dots, r_j; k)$. If $d(k, j)$ is equal to zero for some j , then the associated $w(r_1, \dots, r_j; k)$'s are equal to zero. In this case, we let the corresponding statistic $U_n^{(j)}$ be zero. The statistics Y_n is given by

$$Y_n = \frac{1}{D(n, k)} \sum_{j=1}^k d(k, j) \binom{n}{j} U_n^{(j)}, \quad (1.2)$$

where $D(n, k) = \sum_{j=1}^k d(k, j) \binom{n}{j}$. Since w 's are nonnegative and at least one of them is positive, $D(n, k)$ is positive.

Y_n generalizes well-known statistics. Four examples are given to show this. First, let w be the function given by $w(1, 1, \dots, 1; k) = 1$ and $w(r_1, \dots, r_j; k) = 0$ for positive integers r_1, \dots, r_j such that $j = 1, \dots, k-1$ and $r_1 + \dots + r_j = k$. Then Y_n is equal to U-statistic U_n , which is given by

$$U_n = \binom{n}{k}^{-1} \sum_{1 \leq j_1 < \dots < j_k \leq n} g(X_{j_1}, \dots, X_{j_k}), \quad (1.3)$$

where $\sum_{1 \leq j_1 < \dots < j_k \leq n}$ denotes the summation over all integers j_1, \dots, j_k satisfying $1 \leq j_1 < \dots < j_k \leq n$.

Second, let w be the function given by $w(r_1, \dots, r_j; k) = 1$ for positive integers r_1, \dots, r_j such that $j = 1, \dots, k$ and $r_1 + \dots + r_j = k$. Then Y_n is equal to the LB-statistic B_n given by

$$B_n = \binom{n+k-1}{k}^{-1} \sum_{r_1 + \dots + r_n = k} g(X_1^{r_1}, \dots, X_n^{r_n}), \quad (1.4)$$

where $\sum_{r_1 + \dots + r_n = k}$ denotes the summation over all non-negative integers r_1, \dots, r_n satisfying $r_1 + \dots + r_n = k$.

Third, let w be the function given by $w(r_1, \dots, r_j; k) = k!/(r_1! \dots r_j!)$ for positive integers r_1, \dots, r_j such that $j = 1, \dots, k$ and $r_1 + \dots + r_j = k$. In case of $k = 3$, Y_n is equal to the V-statistic V_n given by

$$V_n = \frac{1}{n^k} \sum_{j_1=1}^n \dots \sum_{j_k=1}^n g(X_{j_1}, \dots, X_{j_k}). \quad (1.5)$$

(See Toda and Yamato, 2001).

The last, let w be the function given by $w(r_1, \dots, r_j; k) = k!/(r_1 \dots r_j)$ for positive integers r_1, \dots, r_j such that $j = 1, \dots, k$ and $r_1 + \dots + r_j = k$. In the case of $k = 3$, Y_n for the third central moment of the distribution F is given by

$$S_n = \frac{n}{n^2 + 1} \sum_{i=1}^n (X_i - \bar{X})^3,$$

where \bar{X} is the sample mean of X_1, \dots, X_n (see Nomachi et al., 2002).

Now, for the kernel $g(x_1, \dots, x_k)$, we put

$$\psi_j(x_1, \dots, x_j) = E[g(X_1, \dots, X_k) \mid X_1 = x_1, \dots, X_j = x_j], \quad j = 1, \dots, k$$

and

$$\sigma_j^2 = \text{Var}[\psi_j(X_1, \dots, X_j)], \quad j = 1, \dots, k.$$

In this paper, we suppose that

$$\sigma_1^2 = \dots = \sigma_{d-1}^2 = 0 \quad \text{and} \quad \sigma_d^2 > 0,$$

that is, the U-statistic and/or the kernel g is degenerate of order $d-1$. Hence, $E\psi_d(X_1, \dots, X_d) = \theta$ and with probability one (w.p.1) $\psi_1(X_1) = \theta, \dots, \psi_{d-1}(X_1, \dots, X_{d-1}) = \theta$.

The asymptotic distribution of U_n was studied by Lee (1990), Koroljuk and Borovskich (1994) and Borovskich (1996). We may summarize their results as follows. Let W be the Gaussian random measure associated with the distribution F on the real line $(-\infty, \infty)$ such that $EW(A) = 0$ and $EW(A)W(B) = F(A \cap B)$ for any Borel sets A, B . Denote the stochastic integral of the function $f(x_1, \dots, x_c)$ by $J_c(f)$, that is,

$$J_c(f) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_c) W(dx_1) \dots W(dx_c).$$

Alternatively, $J_c(f)$ can be represented by using an orthonormal basis e_1, e_2, \dots of $L_2(F)$. For any function f_1 and f_2 such that $\int_{R^c} f_i(x_1, \dots, x_c)^2 \prod_{j=1}^c dF(x_j) < \infty$ ($i = 1, 2$), their inner product is given by $(f_1, f_2) = \int_{R^c} f_1(x_1, \dots, x_c) f_2(x_1, \dots, x_c) \prod_{j=1}^c dF(x_j)$. $J_c(f)$ is also written as

$$J_c(f) = \sum_{i_1=1}^{\infty} \dots \sum_{i_c=1}^{\infty} (f, e_{i_1} \dots e_{i_c}) \prod_{l=1}^{\infty} H_{r_l(i)}(Z_l),$$

where H_r is the r -th Hermite polynomial, $\{Z_l\}_{l=1}^{\infty}$ is a sequence of independent standard normal random variables and $r_l(i)$ is the number of indices among $\mathbf{i} = (i_1, \dots, i_d)$ equal to l . For the degenerate kernel g of order $d-1$, under the conditions $E |g(X_1, \dots, X_k)|^2 < \infty$, the asymptotic distribution of U_n is given by

$$n^{d/2}(U_n - \theta) \xrightarrow{D} \binom{k}{d} J_d(\xi_{d,d}), \quad (1.6)$$

where \xrightarrow{D} means the convergence in distribution as $n \rightarrow \infty$ and $\xi_{d,d}(x_1, \dots, x_d) = \psi_d(x_1, \dots, x_d) - \theta$.

The convergence (1.6) also holds under the following conditions (i) or (ii): (i) $E |g^{(c)}(X_1, \dots, X_d)|^{2c/(2c-d)} < \infty$ for $c = d, d+1, \dots, k$ (see Koroljuk and Borovskich, 1994). (ii) $E |g^{(d)}(X_1, \dots, X_d)|^2 < \infty$ and $t^{2c/(2c-d)} P[|g^{(c)}| > t] \rightarrow 0$ ($t \rightarrow \infty$) for $c = d+1, \dots, k$ (see Borovskich, 1996).

Yamato et al. (2001) studied an invariance principle (functional limit theorem) of Y_n and derived its asymptotic distribution as a special case of the principle. Also the asymptotic distribution of Y_n may be obtained by the method of Yamato and Toda (2001).

The purpose of this paper is to give the asymptotic distribution of Y_n directly by the later method and to evaluate the rate of its convergence, for the degenerate kernel. In Section 2, we give the asymptotic distribution of a linear combination Y_n of U-statistics given by (1.2). In Section 3, we give its rate of convergence in distribution.

2. Asymptotic distribution of Y-statistic

For kernel $g_{(j)}(x_1, \dots, x_j)$ given by (1.1), we put for $c = 1, \dots, j$ and $j = 1, \dots, k$

$$\begin{aligned} \psi_{(j),c}(x_1, \dots, x_c) &= E[g_{(j)}(X_1, \dots, X_j) \mid X_1 = x_1, \dots, X_c = x_c] \\ &= \frac{1}{d(k, j)} \sum_{r_1 + \dots + r_j = k}^+ w(r_1, \dots, r_j; k) E g(x_1^{r_1}, \dots, x_c^{r_c}, X_{c+1}^{r_{c+1}}, \dots, X_j^{r_j}), \end{aligned}$$

where on the right-hand side we use the notation used in (1.1).

The U-statistic $U_n^{(j)}$, $j = 1, \dots, k$ corresponding to the kernel $g_{(j)}$ have the following properties.

LEMMA 2.1. (Yamato et al., 2001)

$$E[U_n^{(j)}] = \theta, \quad k - \frac{d-1}{2} \leq j \leq k$$

or

$$E[U_n^{(k-j)}] = \theta, \quad 0 \leq j \leq \frac{d-1}{2}.$$

LEMMA 2.2. (Yamato et al., 2001) *The order of degeneracy of $U_n^{(k-j)}$ is at least $d - 2j - 1$ for $1 \leq j \leq (d-1)/2$ and*

$$\psi_{(k-j),d-2j}(x_1, \dots, x_{d-2j}) = \theta + \binom{k-d+j}{j} \frac{w(1^{k-2j}, 2^j; k)}{d(k, k-j)} [\varphi_{d,d-2j}(x_1, \dots, x_{d-2j}) - \theta], \quad (2.1)$$

where for $1 \leq j \leq (d-1)/2$

$$\varphi_{d,d-2j}(x_1, \dots, x_{d-2j}) = E[\psi_d(x_1, \dots, x_{d-2j}, X_{d-2j+1}, X_{d-2j+1}, \dots, X_{d-j}, X_{d-j})] \quad (2.2)$$

and

$$w(1^r, 2^s; k) = w(\underbrace{1, 1, \dots, 1}_r, \underbrace{2, 2, \dots, 2}_s; k).$$

By the similar method to the above Lemma, we can show that for $1 \leq j \leq (d-1)/2$,

$$E\varphi_{d,d-2j}(X_1, \dots, X_{d-2j}) = \theta. \quad (2.3)$$

From Lemmas 2.1 and 2.2, it follows that If $d = 2l + 1$ and l is a positive integer, then $EU_n^{(k)} = EU_n^{(k-1)} = \dots = EU_n^{(k-l+1)} = EU_n^{(k-l)} = \theta$. The orders of degeneracy of $U_n^{(k-1)}, \dots, U_n^{(k-l+1)}, U_n^{(k-l)}$ are at least $2(l-1), \dots, 2, 0$, respectively. If $d = 2l$ and l is a positive integer, then $EU_n^{(k)} = EU_n^{(k-1)} = \dots = EU_n^{(k-l+2)} = EU_n^{(k-l+1)} = \theta$. The orders of degeneracy of $U_n^{(k-1)}, \dots, U_n^{(k-l+2)}, U_n^{(k-l+1)}$ are at least $2l-3, \dots, 3, 1$, respectively.

LEMMA 2.3. (Yamato et al., 2001) *In case of $d = 2l$,*

$$EU_n^{(k-l)} - \theta = \frac{1}{d(k, k-l)} \binom{k-l}{k-d} w(1^{k-d}, 2^l; k) [E\psi_d(X_1, X_1, \dots, X_l, X_l) - \theta]. \quad (2.4)$$

Now we assume $d(k, k) = w(1, \dots, 1; k) > 0$, then, it follows that $U_n^{(k)} = U_n$, because of $g_{(k)} = g$. Under this assumption, there exists a constant $b(\geq 0)$ such that

$$\frac{d(k, k)n^{(k)}}{D(n, k)k!} = 1 - \frac{b}{n} + O\left(\frac{1}{n^2}\right). \quad (2.5)$$

Thus

$$\frac{n^k}{D(n, k)} \rightarrow \frac{k!}{d(k, k)} \text{ as } n \rightarrow \infty. \quad (2.6)$$

For the U-statistic U_n , $d(k, k)n^{(k)}/[D(n, k)k!] = 1$. For the V-statistic V_n and the S-statistic S_n , $b = k(k-1)$. For the LB-statistic B_n , $b = k(k-1)/2$. (see Nomachi et al., 2001). It is assumed that $n^k/D(n, k)$ is nondecreasing in Yamato et al. (2001) for invariance principle, but the convergence in distribution needs only the convergence (2.6).

Yamato et al. (2001) obtained the asymptotic distribution of Y_n from invariance principle. Alternatively, it can be obtained directly by the same method as Yamato and Toda (2001). We state here the idea of how to compute directly the asymptotic distribution of Y_n . Noting the degeneracy of $U_n^{(k-j)}$, $0 \leq j \leq (d-1)/2$, stated above, we expand Y_n as follows: For $d = 2l + 1$

$$n^{d/2}(Y_n - \theta) = \sum_{j=0}^l T_{j,n} + R_{1n}, \quad (2.7)$$

where

$$T_{j,n} = d(k, k-j) \frac{\binom{n}{k-j}}{D(n, k)} n^{d/2} (U_n^{(k-j)} - \theta), \quad j = 0, 1, \dots, l,$$

$$R_{1n} = \sum_{j=1}^{k-l-1} d(k, j) \frac{\binom{n}{j}}{D(n, k)} n^{d/2} (U_n^{(j)} - \theta).$$

By Minkowski's inequality we have

$$[ER_{1n}^2]^{1/2} \leq \sum_{j=1}^{k-l-1} d(k, j) \frac{\binom{n}{j}}{D(n, k)} n^{d/2} [E(U_n^{(j)} - \theta)^2]^{1/2}.$$

By the relation (2.6), $\binom{n}{j} n^{d/2}/D(n, k) \sim cn^{j-k+d/2}$ ($n \rightarrow \infty$), where c is a positive constant. For $1 \leq j \leq k-l-1$, it holds that $j-k+d/2 \leq -l-1+d/2 \leq -1/2$. Under the condition (2.11) given in the following theorem 2.4, we have $E(U_n^{(j)} - \theta)^2 < \infty$ and therefore it follows that $ER_{1n}^2 = O(n^{-1})$. In case of $d = 2l$, it holds that $j-k+d/2 \leq -l-1+d/2 \leq -1$ for $1 \leq j \leq k-l-1$. Thus, for $d = 2l$,

$$n^{d/2}(Y_n - \theta) = \sum_{j=0}^{l-1} T_{j,n} + T_{l,n} + R_{2n}, \quad (2.8)$$

where $ER_{2n}^2 = O(n^{-2})$ under the condition (2.11) given in the following theorem 2.4. R_{1n} , R_{2n} do not affect the convergence in distribution of $n^{d/2}(Y_n - \theta)$.

We note that $g_{(k)} = g$ and the kernel g is assumed to be degenerate of order $d-1$. We put $g_{(k)}^{(d)}(x_1, \dots, x_d) = \psi_d(x_1, \dots, x_d) - \theta (= \varphi_{d,d}(x_1, \dots, x_d) - \theta)$. Then, by (1.6) it follows

$$n^d(U_n^{(k)} - \theta) \xrightarrow{\mathcal{D}} \binom{k}{d} J(g_{(k)}^{(d)}). \quad (2.9)$$

Similarly, noting the degeneracy of the kernel $g_{(k-j)}$ given by Lemma 2.2, we put

$$g_{(k-j)}^{(d-2j)}(x_1, \dots, x_{d-2j}) = \psi_{(k-j), d-2j}(x_1, \dots, x_{d-2j}) - \theta, \quad 1 \leq j \leq (d-1)/2.$$

For $1 \leq j \leq (d-1)/2$, at first we assume that the order of degeneracy of $U_n^{(k-j)}$ is $d-2j-1$. Then, by Lemma 2.1 and (1.6), it follows that

$$n^{d-2j}(U_n^{(k-j)} - \theta) \xrightarrow{\mathcal{D}} \binom{k-j}{d-2j} J_{d-2j}(g_{(k-j)}^{(d-2j)}), \quad 1 \leq j \leq (d-1)/2. \quad (2.10)$$

By (2.1), for $1 \leq j \leq (d-1)/2$

$$g_{(k-j)}^{(d-2j)}(x_1, \dots, x_{d-2j}) = \frac{1}{d(k, k-j)} \binom{k-d+j}{j} w(1^{k-2j}, 2^j; k) \xi_{d, d-2j}(x_1, \dots, x_{d-2j}),$$

where

$$\xi_{d, d-2j}(x_1, \dots, x_{d-2j}) = \varphi_{d, d-2j}(x_1, \dots, x_{d-2j}) - \theta.$$

If for some j_0 ($1 \leq j_0 \leq (d-1)/2$) the order of degeneracy of $U_n^{(k-j_0)}$ is less than $d-2j_0-1$, then the limit in distribution of $U_n^{(k-j_0)}$ vanish because of $g_{(k-j_0)}^{(d-2j_0)} = 0$. In the case of $d = 2l$, $T_{l,n}$ converges to a constant as $n \rightarrow \infty$. Applying these convergence of $U_n^{(k-j)}$ ($0 \leq j \leq (d-1)/2$) to (2.7) and (2.8), we get the following proposition.

THEOREM 2.4. (Yamato et al., 2001) *We suppose that*

$$E[g(X_{j_1}, X_{j_2}, \dots, X_{j_k})^2] < \infty \quad (2.11)$$

for all j_1, j_2, \dots, j_k such that $1 \leq j_1 \leq j_2 \leq \dots \leq j_k \leq k$. We assume $d(k, k) > 0$. Then in case of $d = 2l + 1$ ($l = 1, 2, \dots$), we have

$$n^{d/2}(Y_n - \theta) \xrightarrow{\mathcal{D}} \sum_{j=0}^l \tau_j, \quad (2.12)$$

where

$$\tau_j = \frac{k!}{(k-d)! (d-2j)! j!} \cdot \frac{w(1^{k-2j}, 2^j; k)}{w(1^k; k)} J_{d-2j}(\xi_{d, d-2j}), \quad j = 0, 1, \dots, l$$

and

$$\xi_{d, d-2j} = \varphi_{d, d-2j}(x_1, \dots, x_{d-2j}) - \theta, \quad j = 0, 1, \dots, l.$$

In case of $d = 2l$ ($l = 1, 2, \dots$) we have

$$n^{d/2}(Y_n - \theta) \xrightarrow{\mathcal{D}} \sum_{j=0}^{l-1} \tau_j + \alpha_l, \quad (2.13)$$

where

$$\alpha_l = \frac{k!w(1^{k-d}, 2^l; k)}{(k-d)!l!w(1^k; k)} [E\psi_d(X_1, X_1, \dots, X_l, X_l) - \theta].$$

The value $w(1^{k-2j}, 2^j; k)/w(1^k; k)$ is equal to $1/2^j$ for the V-statistic V_n and the S-statistic S_n , and 1 for the LB-statistic B_n , respectively. The asymptotic distributions of the V-statistic V_n and the LB-statistic B_n are given by Yamato and Toda (2001) and Yamato et al. (2001). The asymptotic distribution of the S-statistic S_n is given by the followings: In case of $d = 2l + 1$ ($l = 1, 2, \dots$), we have

$$n^{d/2}(S_n - \theta) \xrightarrow{D} \frac{k!}{(k-d)!} \sum_{j=0}^l \frac{1}{(d-2j)!j!2^j} J_{d-2j}(\xi_{d,d-2j}).$$

In case of $d = 2l$ ($l = 1, 2, \dots$) we have

$$n^{d/2}(S_n - \theta) \xrightarrow{D} \frac{k!}{(k-d)!} \left\{ \sum_{j=0}^{l-1} \frac{1}{(d-2j)!j!2^j} J_{d-2j}(\xi_{d,d-2j}) + \frac{1}{l!2^l} [E\psi_d(X_1, X_1, \dots, X_l, X_l) - \theta] \right\}.$$

3. Rate of convergence

To get the rate of convergence in distribution of Y_n , we use the following rate of convergence in distribution of the U-statistic U_n .

LEMMA 3.1. (Theorem 6.5.2 of Borovskikh, 1996) *Assume that*

$$|E \exp(itJ_k(\xi_{d,d}))| = O(|t|^{-\gamma})$$

as $|t| \rightarrow \infty$, where γ is some sufficiently large number. Then if d is odd and

$$E |g(X_1, \dots, X_k)|^3 < \infty,$$

it follows that

$$\sup_{-\infty < t < \infty} |P(n^{d/2}(U_n - \theta) \leq t) - P(\tau_0 \leq t)| = O(n^{-\frac{1}{2}}). \quad (3.1)$$

Furthermore, if d is even and

$$E |g(X_1, \dots, X_k)|^4 < \infty$$

then,

$$\sup_{-\infty < t < \infty} |P(n^{d/2}(U_n - \theta) \leq t) - P(\tau_0 \leq t)| = o(n^{-\frac{1}{2}}). \quad (3.2)$$

This lemma is also given by Theorem 6.6.3 of Koroljuk and Borovskikh (1994). Under the condition about the characteristic functions of Lemma 3.1, the corresponding distribution function has a bounded continuous derivative (see, for example, Corollary 11.6.1 of Kawata, 1972).

THEOREM 3.2. We assume that $d(k, k) > 0$ and

$$|E \exp(itJ_{k-2j}(\xi_{d,d-2j}))| = O(|t|^{-\gamma_j}), \quad 0 \leq j \leq (d-1)/2$$

as $|t| \rightarrow \infty$, where γ_j is some sufficiently large number. Then if $d = 2l + 1$ ($l = 1, 2, \dots$) and

$$E |g(X_{i_1}, \dots, X_{i_k})|^3 < \infty \quad \text{for } 1 \leq i_1 \leq \dots \leq i_k \leq k,$$

it follows that

$$\sup_{-\infty < t < \infty} |P(n^{d/2}(Y_n - \theta) \leq t) - P\left(\sum_{j=0}^l \tau_j \leq t\right)| = O(n^{-\frac{1}{2}}). \quad (3.3)$$

If $d = 2l$ ($l = 1, 2, \dots$) and

$$E |g(X_{i_1}, \dots, X_{i_k})|^4 < \infty \quad \text{for } 1 \leq i_1 \leq \dots \leq i_k \leq k$$

then, we have

$$\sup_{-\infty < t < \infty} |P(n^{d/2}(Y_n - \theta) \leq t) - P\left(\sum_{j=0}^{l-1} \tau_j + \alpha_l \leq t\right)| = O(n^{-\frac{1}{2}}). \quad (3.4)$$

In case of $d = 2l$ ($l = 1, 2, \dots$), the rate of convergence given by (3.4) is weaker than (3.2). This is due to the evaluation of l_{31} given later by (3.13). In order to prove this theorem, we need the following lemma.

LEMMA 3.3. For any random variables W_{11}, W_{12}, W_{21} and W_{22} , we have

$$\begin{aligned} I_{11} &= \sup_{-\infty < t < \infty} |P(W_{11} + W_{21} \leq t) - P(W_{12} + W_{22} \leq t)| \\ &\leq \sup_{-\infty < t < \infty} |P(W_{11} \leq t) - P(W_{12} \leq t)| + \sup_{-\infty < t < \infty} |P(W_{21} \leq t) - P(W_{22} \leq t)|. \end{aligned} \quad (3.5)$$

We suppose that $W_n \xrightarrow{D} W$, EW_n^2 is uniformly finite in $n = 1, 2, \dots$, and that $a_n = 1 - (a/n) + O(n^{-2})$, where a is a constant. Then

$$\begin{aligned} I_{12} &= \sup_{-\infty < x < \infty} |P(a_n W_n \leq x) - P(W \leq x)| \\ &\leq \sup_{-\infty < x < \infty} |P(W_n \leq x) - P(W \leq x)| + O\left(\frac{1}{n}\right). \end{aligned} \quad (3.6)$$

Proof of Lemma 3.3. For I_{11} , we have the following inequality.

$$\begin{aligned} I_{11} &\leq \sup_t |E_{W_{21}}\{P(W_{11} + W_{21} \leq t | W_{21}) - P(W_{12} + W_{21} \leq t | W_{21})\}| \\ &\quad + \sup_t |E_{W_{12}}\{P(W_{12} + W_{21} \leq t | W_{12}) - P(W_{12} + W_{22} \leq t | W_{12})\}|, \end{aligned}$$

where E_W denotes the expectation with respect to the random variable W . The right-hand side of the above inequality is less than or equal to

$$E_{W_{21}} \sup_t |P(W_{11} + W_{21} \leq t | W_{21}) - P(W_{12} + W_{21} \leq t | W_{21})|$$

$$+EW_{12} \sup_t | P(W_{12} + W_{21} \leq t \mid W_{12}) - P(W_{12} + W_{22} \leq t \mid W_{12}) |.$$

Thus we can obtain (3.5).

Next we shall prove the inequality (3.6). For I_{12} , we have

$$I_{12} \leq \sup_x | P(a_n W_n \leq x) - P(W_n \leq x) | + \sup_x | P(W_n \leq x) - P(W \leq x) |.$$

For any random variables W and Δ , it holds that

$$\sup_x | P(W + \Delta \leq x) - P(W \leq x) | \leq 4(E | W\Delta | + E | \Delta |) \quad (3.7)$$

(see p.261 of Shorack, 2000). Because $a_n = 1 - (a/n) + O(n^{-2})$ and EW_n^2 is uniformly finite, we have from (3.7)

$$\sup_x | P(a_n W_n \leq x) - P(W_n \leq x) | \leq 4 \left[\frac{|a|}{n} (EW_n^2 + E | W_n |) + O\left(\frac{1}{n^2}\right) \right] = O\left(\frac{1}{n}\right).$$

□

Proof of Theorem 3.2. At first we consider the case of $d = 2l + 1$ ($l = 1, 2, \dots$). Put

$$\begin{aligned} I_2 &= \sup_t | P(n^{d/2}(Y_n - \theta) \leq t) - P\left(\sum_{j=0}^l \tau_j \leq t\right) | \\ &= \sup_t | P\left(\sum_{j=0}^l T_{j,n} + R_{1n} \leq t\right) - P\left(\sum_{j=0}^l \tau_j \leq t\right) |. \end{aligned}$$

Then we have

$$I_2 \leq I_{21} + I_{22}, \quad (3.8)$$

where

$$I_{21} = \sup_t | P\left(\sum_{j=0}^l T_{j,n} \leq t\right) - P\left(\sum_{j=0}^l \tau_j \leq t\right) |$$

and

$$I_{22} = \sup_t | P\left(\sum_{j=0}^l T_{j,n} + R_{1n} \leq t\right) - P\left(\sum_{j=0}^l T_{j,n} \leq t\right) |.$$

Since $ET_{j,n}^2 = O(1)$, $j = 0, 1, \dots, l$ and $ER_{1n}^2 = O(n^{-1})$, applying (3.7) to I_{22} we have

$$I_{22} = O(n^{-\frac{1}{2}}). \quad (3.9)$$

By (1.7) we can write $T_{j,n}$, $j = 0, \dots, l$ of the right-hand side of (2.5) as

$$T_{j,n} = c_1 \left(1 - \frac{c_2}{n} + O\left(\frac{1}{n^2}\right)\right) n^{\frac{d-2j}{2}} (U_n^{(k-j)} - \theta), \quad j = 0, \dots, l, \quad (3.10)$$

where $c_1(> 0)$ and c_2 are generic constants. By Lemma 3.3 we have

$$I_{21} \leq \sum_{j=0}^l \sup_t | P(T_{j,n} \leq t) - P(\tau_j \leq t) |.$$

Using the relation $E | U_n^{(k-j)} - \theta |^2 = O(n^{-(d-2j)})$, $j = 0, \dots, l$ (see p.185 of Serfling, 1980) and applying (3.1), (3.6) and (3.10) to the right-hand side of the above inequality, we have

$$I_{21} = O(n^{-\frac{1}{2}}). \quad (3.11)$$

Because of (3.9) and (3.11), we get from (3.8)

$$I_2 = O(n^{-\frac{1}{2}}),$$

which is (3.3).

Next we consider the case of $d = 2l$ ($l = 1, 2, \dots$). On the right-hand side of (2.6), $T_{l-1,n}$ converges in distribution to τ_{l-1} and T_l converges in probability to α_l . We shall compute

$$\begin{aligned} I_3 &= \sup_x | P(T_{l-1,n} + T_{l,n} \leq x) - P(\tau_{l-1} + \alpha_l \leq x) | \\ &= \sup_x | P(T_{l-1,n} + T_{l,n} - \alpha_l \leq x) - P(\tau_{l-1} \leq x) |. \end{aligned}$$

It follows that

$$I_3 \leq I_{31} + I_{32}$$

where

$$I_{31} = \sup_x | P(T_{l-1,n} + T_{l,n} - \alpha_l \leq x) - P(T_{l-1,n} \leq x) |$$

and

$$I_{32} = \sup_x | P(T_{l-1,n} \leq x) - P(\tau_{l-1} \leq x) |.$$

For I_{32} , by (3.2), (3.6) and (3.10) with $j = l - 1$ we have

$$I_{32} = o(n^{-\frac{1}{2}}). \quad (3.12)$$

Next we consider I_{31} . By (3.7), we have

$$I_{31} \leq 4E | T_{l-1,n}(T_{l,n} - \alpha_l) | + 4E | T_{l,n} - \alpha_l |, \quad (3.13)$$

where

$$T_{l,n} - \alpha_l = \frac{d(k, k-l)k!}{(k-l)!d(k, k)}(A + B),$$

$$A = \frac{d(k, k)}{D(n, k)k!} n^{(k-l)} n^l U_n^{(k-l)} - EU_n^{(k-l)} \quad \text{and} \quad B = \left[1 - \frac{d(k, k)}{D(n, k)k!} n^{(k-l)} n^l \right] \theta.$$

By (1.7), we have

$$A = U_n^{(k-l)} - EU_n^{(k-l)} + \left[\frac{c}{n} + O\left(\frac{1}{n^2}\right) \right] U_n^{(k-l)} \quad \text{and} \quad B = O\left(\frac{1}{n}\right),$$

for some constant c . Because of $E[U_n^{(k-l)} - EU_n^{(k-l)}]^2 = O(n^{-1})$ (see, for example, Lee, 1990, p.21), we have $E | A | = O(n^{-1/2})$. Thus we get $E | T_{l,n} - \alpha_l | = O(n^{-1/2})$ and $E | T_{l-1,n}(T_{l,n} - \alpha_l) | = O(n^{-1/2})$. Thus by (3.13) we have $I_{31} = O(n^{-\frac{1}{2}})$. From this and (3.12), we get

$$I_3 = O(n^{-\frac{1}{2}}).$$

Since $ET_{j,n}^2 = O(1)$, $j = 0, 1, \dots, l$ and $ER_{2n}^2 = O(n^{-2})$, by the similar discussion to the first part of the proof in case of $d = 2l + 1$ we have

$$\begin{aligned} I_4 &= \sup_t | P(n^{d/2}(Y_n - \theta) \leq t) - P(\sum_{j=0}^{l-1} \tau_j + \alpha_l \leq t) | \\ &\leq \sup_t | P(\sum_{j=0}^{l-1} T_{j,n} + T_{1n} \leq t) - P(\sum_{j=0}^{l-1} \tau_j + \alpha_l \leq t) | + O(n^{-1}). \end{aligned}$$

By (3.5), we have

$$\begin{aligned} I_4 &\leq \sum_{j=0}^{l-2} \sup_t | P(T_{j,n} \leq t) - P(\tau_j \leq t) | + I_3 + O(n^{-1}) \\ &= \sum_{j=0}^{l-2} \sup_t | P(T_{j,n} \leq t) - P(\tau_j \leq t) | + O(n^{-\frac{1}{2}}). \end{aligned}$$

By the similar discussion to the case of $d = 2l + 1$, by (3.2), (3.6) and (3.10) we have

$$I_4 = O(n^{-\frac{1}{2}}),$$

which is (3.4). □

In particular, for the case of $k = 2$ and $d = 2$ the rate of convergence of the U-statistic is given by Theorem 6.5.1 of Borovskikh (1996) and Theorem 6.6.2 of Koroljuk and Borovskikh (1994) as follows: Suppose the condition (A) that for S-operator

$$S : f \rightarrow E[g(X_1, X_2)f(X_2) | X_1 = x],$$

there exists infinitely many indices j for which the eigenvalues λ_j of this operator S are nonzero. Furthermore we assume that $E | g(X_1, X_2) |^3 < \infty$. Then

$$\sup_{-\infty < t < \infty} | P(n(U_n - \theta) \leq t) - P(\tau_0 \leq t) | = o(n^{-\frac{1}{2}}), \quad (3.14)$$

where

$$\tau_0 = \sum_{j=1}^{\infty} \lambda_j (Z_j - 1)$$

and Z_1, Z_2, \dots are independent standard normal random variables. In the proof of Theorem 3.2, we consider the case of $d = 2l$ and $l = 1$. By (3.6), (3.10) and (3.14), we have $I_{32} = o(n^{-\frac{1}{2}})$. Under the condition $E | g^{(1)}(X_1, X_1) |^2 < \infty$, we have $I_{31} = O(n^{-\frac{1}{2}})$. Hence we get $I_3 = O(n^{-\frac{1}{2}})$. Thus we have the following proposition.

PROPOSITION 3.4. *Suppose that the condition (A), $E |g(X_1, X_2)|^3 < \infty$ and $E |g^{(1)}(X_1, X_1)|^2 < \infty$. Then we have,*

$$\sup_{-\infty < t < \infty} |P(n(Y_n - \theta) \leq t) - P(\tau_0 + \alpha_1 \leq t)| = O(n^{-\frac{1}{2}}), \quad (3.15)$$

where $\tau_0 = J_2(\xi_{2,2})$ and $\alpha_1 = 2[w(2; 2)/w(1, 1; 2)] \cdot [Eg(X_1, X_1) - \theta]$.

Specially, for the LB-statistic B_n , by (2.13) we have

$$\sup_{-\infty < t < \infty} |P(n(B_n - \theta) \leq t) - P(\tau_0 + 2[Eg(X_1, X_1) - \theta] \leq t)| = O(n^{-\frac{1}{2}}).$$

For the V-statistic V_n , by (2.103) we have

$$\sup_{-\infty < t < \infty} |P(n(V_n - \theta) \leq t) - P(\tau_0 + Eg(X_1, X_1) - \theta \leq t)| = O(n^{-\frac{1}{2}}). \quad (3.16)$$

For the statistic S_n , (3.16) holds with S_n instead of V_n in the first P on the left-hand side. For the V-statistic, Theorem 6.6.1 of Koroljuk and Borovskikh, 1994 shows that the rate of convergence is $o(n^{-1/2})$ under the condition (A), $E[|g(X_1, X_2)|^3] < \infty$ and $E[|g^{(1)}(X_1, X_1)|^{3/2}] < \infty$. This is stronger than our result given (3.16). The reason why our result is weaker is due to the evaluation $I_{31} = O(n^{-1/2})$.

Acknowledgement

The authors wish to express their thanks to the referee for his careful reading and kind comments.

References

- Borovskikh, Yu.V. (1996). *U-statistics in Banach spaces*. VSP, Utrecht.
- Kawata, T. (1972). *Fourier analysis in probability theory*. Academic Press, New York.
- Koroljuk, V.S. and Borovskikh, Yu.V. (1994). *Theory of U-statistics*, Kluwer Academic Publishers, Dordrecht.
- Lee, A. J. (1990). *U-statistics*, Marcel Dekker, New York.
- Nomachi, T., Kondo, M. and Yamato, H. (2001). Higher order efficiency of linear combinations of U-statistics as estimators of estimable parameters, *Scientiae Mathematicae Japonicae*, **56**, 95–106.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley, New York.
- Shorack, G. R. (2000). *Probability for statisticians*. Springer, New York.
- Toda, K. and Yamato, H. (2001). Berry-Esseen bounds for some statistics including LB-statistic and V-statistic, *J. Japan Statist. Soc.* **31**, 225–237.
- Yamato, H. and Toda, K. (2001). Asymptotic distributions of LB-statistics and V-statistics for degenerate kernel, *Bulletin of Informatics and Cybernetics*, **33**, 27–42.

Yamato, H. , Kondo, M. and Toda, K. (2001). Asymptotic properties of linear combinations of U-statistics with degenerate kernels, (Submitted)

Received March 11, 2002

Revised September 24, 2002

Re-revised October 10, 2002