Acta Sci. Math., 53 (1989), 385—394

The accuracy of the normal approximation for U-statistics with a random summation index converging to a random variable

M. AERTS and H. CALLAERT

Introduction

The exact order of the normal approximation has been obtained in [2] for U-statistics with a random summation index L_n where $L_n/n \to \tau$ with τ a constant. In this paper it is shown that the same order bounds can be obtained in the situation that the random index L_n satisfies $L_n/n \to \tau$ where now τ is a positive random variable. Moreover, a sharpening of the moment condition on the kernel is included. The results are valid for U-statistics with kernel of general degree r but in order to avoid a cumbersome notation, the proofs of the main theorems are given for the case that $r=2$. Tools for passing from $r=2$ to an arbitrary degree r are given in the preliminary lemmas which are formulated and proved for general r. For further information we refer to the Ph. D. thesis of one of the authors [1].

The results obtained in this paper are an extension of earlier results for random sums of i.i.d. random variables, proved in [6] and [3]. The proofs of these results use some methods which heavily rely on the i.i.d. structure. However, if one makes use of the structure of a U-statistic together with some technical fine-tuning, it is possible to obtain order bounds which are as sharp as in the i.i.d. case without imposing any additional condition. We also note that an asymptotic normality result contained in Theorem 1 below could in principle be obtained from Theorem 1 of [4]. However, this derivation would require some extra assumptions on the kernel function and no information on the rate of convergence could be gained.

Received December 17, 1985 and in final revised form July 20, 1987.

Preliminary lemmas

In order to create some flexibility in the renormalization of the statistics under consideration we formulate some general lemmas, special cases of which will be needed in the proof of our main theorem. The proof of Lemma 1 is elementary and is left to the reader. Throughout the paper we use the convention $[x]=\min\{k\in\mathbf{N}, x\leqq k\}$.

Lemma 1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and X_n and Y_n two sequences of random variables defined on Ω. Let C be a positive constant and d_n a sequence of nonnegative real numbers. If for some $k\geqq 0$ and some $\alpha>1$, $S_n^{k,\alpha}$ denotes the set on which $Y_n>k\alpha/(\alpha-1)$, then*

$$S_n^{k,\alpha}\cap\left\{\left|\frac{X_n-k}{Y_n-k}-1\right|>\alpha C d_n\right\}\subset S_n^{k,\alpha}\cap\left\{\left|\frac{X_n}{Y_n}-1\right|>C d_n\right\}.$$

Lemma 2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and X_n and Y_n two sequences of positive random variables defined on Ω. If there exist positive constants c_1 and c_2 and a sequence of positive numbers ε_n with $\varepsilon_n\to 0$ for $n\to\infty$, such that*

$$P\left(\left|\frac{X_n}{[Y_n]}-1\right|>c_1\varepsilon_n\right)=O(\sqrt{\varepsilon_n}),\quad n\to\infty \tag{1}$$

and

$$P(Y_n<c_2\varepsilon_n^{-1/\delta})=O(\sqrt{\varepsilon_n}),\quad n\to\infty,\quad 0<\delta\leqq 1, \tag{2}$$

then, for every integer $k\geqq 0$, there exists a constant M_k such that

$$P\left(\left|\frac{\sqrt{[Y_n]}}{\sqrt{X_n}}\,\frac{X_n(X_n-1)\dots(X_n-k)}{[Y_n]([Y_n]-1)\dots([Y_n]-k)}-1\right|>M_k\sqrt{\varepsilon_n}\right)=O(\sqrt{\varepsilon_n}),\quad n\to\infty. \tag{3}$$

Proof. The proof is by induction. For $k=0$, (3) follows by taking $M_0=\sqrt{c_1}$. Assume that (3) holds true when $k=r-1$ for some $r\in\mathbf{N}_0$. Putting

$$Z_n=\frac{\sqrt{[Y_n]}}{\sqrt{X_n}}\,\frac{X_n(X_n-1)\dots(X_n-r+1)}{[Y_n]([Y_n]-1)\dots([Y_n]-r+1)},$$

the induction hypothesis yields that $P(|Z_n-1|>M_{r-1}\sqrt{\varepsilon_n})=O(\sqrt{\varepsilon_n})$, $n\to\infty$, for some constant M_{r-1}. Now choose M_r such that $M_r\geqq\max(3M_{r-1}, 6c_1)$ and then take n so large that

$$\varepsilon_n<\min\{1, (c_2/2r)^\delta, 9/(M_r^2)\} \tag{4}$$

is satisfied. Since (4) implies that $[c_2\varepsilon_n^{-1/\delta}]>2r$, one has, using the Bonferroni inequality and (2)

$$P\left(\left|Z_n\frac{X_n-r}{[Y_n]-r}-1\right|>M_r\sqrt{\varepsilon_n}\right)\leqq P(Y_n<c_2\varepsilon_n^{-1/\delta})+$$

$$+P\left(\left|(Z_n-1)\left(\frac{X_n-r}{[Y_n]-r}-1\right)+\left(\frac{X_n-r}{[Y_n]-r}-1\right)+(Z_n-1)\right|>M_r\sqrt{\varepsilon_n},\ [Y_n]>2r\right)\leqq$$

$$\leqq O(\sqrt{\varepsilon_n})+P\left(|Z_n-1|>\frac{M_r}{3}\sqrt{\varepsilon_n}\right)+P\left(\left|\frac{X_n-r}{[Y_n]-r}-1\right|>\frac{M_r}{3}\sqrt{\varepsilon_n},\ [Y_n]>2r\right)+$$

$$+P\left(\left|(Z_n-1)\left(\frac{X_n-r}{[Y_n]-r}-1\right)\right|>\frac{M_r}{3}\sqrt{\varepsilon_n},\ [Y_n]>2r\right).$$

It is easy to see, using the choice of M_r, the induction hypothesis, (1), (4), and Lemma 1 with $C=M_r/6$, $\alpha=2$, $d_n=\sqrt{\varepsilon_n}$ and $k=r$, that the second and third terms here are $O(\sqrt{\varepsilon_n})$. But by (4) the fourth term is not greater than

$$P\left(\left|(Z_n-1)\left(\frac{X_n-r}{[Y_n]-r}-1\right)\right|>\frac{M_r}{3}\sqrt{\varepsilon_n},\ [Y_n]>2r,\ |Z_n-1|\leqq 1\right)+P(|Z_n-1|>1)\leqq$$

$$\leqq P\left(\left|\frac{X_n-r}{[Y_n]-r}-1\right|>\frac{M_r}{3}\sqrt{\varepsilon_n},\ [Y_n]>2r\right)+P\left(|Z_n-1|>\frac{M_r}{3}\sqrt{\varepsilon_n}\right),$$

and the lemma follows.

The next lemma, which states the rate of convergence to normality for non-stochastically indexed U-statistics, plays a crucial role in the proof of the main theorem. It determines, together with the asymptotic behaviour of the random index L_n, the final approximation order for random U-statistics.

Lemma 3. *Let $(\Omega,\mathcal{A},P)$ be a probability space and $X_1, X_2, \ldots$ i.i.d. random variables defined on Ω. Let $U_n=\binom{n}{r}^{-1}\sum^{(n,r)}h(X_{i_1},\ldots,X_{i_r})$ be a U-statistic with $Eh(X_1,\ldots,X_r)=\theta$ and put $g(X_1)=E(h(X_1,\ldots,X_r)-\theta|X_1)$. Assume that $\sigma^2=\operatorname{Var}g(X_1)$ is strictly positive, and that for some δ, $0<\delta\leqq 1$, one has that $E|g(X_1)|^{2+\delta}<\infty$ and $E|h(X_1,\ldots,X_r)|^{(4+\delta)/3}<\infty$. Then, one has:*

$$\sup_x\left|P\left\{\frac{\sqrt{n}(U_n-\theta)}{r\sigma}\leqq x\right\}-\Phi(x)\right|=O(n^{-\delta/2}),\quad n\to\infty.$$

Proof. The proof is essentially based on an improvement of a Berry—Esseen bound for more general non-parametric statistics (see [5]). For details of the proof we refer to [1], where it is also shown that the result of Lemma 3 is valid for statistics with structure $\sum_{i=1}^{p}g(X_i)+Y_k$ as used in the proof of our main theorem.

Main result

Theorem 1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X_1, X_2, \ldots$ i.i.d. random variables defined on Ω. Let $U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j)$ be a U-statistic with $Eh(X_1, X_2) = \theta$ and put $g(X_1) = E(h(X_1, X_2) - \theta | X_1)$. Assume that $\sigma^2 = \operatorname{Var} g(X_1)$ is strictly positive, and that for some δ, $0 < \delta \le 1$, one has that $E|g(X_1)|^{2+\delta} < \infty$ and $E|h(X_1, X_2)|^{(4+\delta)/3} < \infty$. Further, let ε_n be a sequence of positive numbers tending to zero and such that, for n large, $n^{-\delta} \le \varepsilon_n$. Let $L_n: \Omega \to \{2, 3, 4, \ldots\}$ and $\tau: \Omega \to (0, \infty)$ be random variables satisfying, for some constants $c_1, c_2 > 0$:*

$$P\left(\left|\frac{L_n}{[n\tau]} - 1\right| > c_1 \varepsilon_n\right) = O(\sqrt{\varepsilon_n}), \quad n \to \infty \tag{5}$$

$$P\left(\tau < \frac{c_2}{n} \varepsilon_n^{-1/\delta}\right) = O(\sqrt{\varepsilon_n}), \quad n \to \infty \tag{6}$$

(7) τ *is independent of* X_k, $k = 1, 2, \ldots$

then, one has:

(i) $$\sup_x \left| P\left(\frac{\sqrt{n\tau}}{2\sigma} \binom{n\tau}{2}^{-1} \sum_{1 \le i < j \le L_n} (h(X_i, X_j) - \theta) \le x\right) - \Phi(x)\right| = O(\sqrt{\varepsilon_n}), \quad n \to \infty$$

(ii) $$\sup_x \left| P\left(\frac{\sqrt{L_n}}{2\sigma} (U_{L_n} - \theta) \le x\right) - \Phi(x)\right| = O(\sqrt{\varepsilon_n}), \quad n \to \infty$$

and, if $\sigma_n^2 = \operatorname{Var} U_n$ *exists:*

(iii) $$\sup_x \left| P(\sigma_{L_n}^{-1}(U_{L_n} - \theta) \le x) - \Phi(x)\right| = O(\sqrt{\varepsilon_n}), \quad n \to \infty.$$

Proof. W.l.g. we assume that $\theta = 0$. The following notation will be used:

$$\mathbf{N}_1 = \{2, 3, 4, \ldots\}$$

$$I_n^{**} = I_n^{**}(\omega) =$$

$$= \{j \in \mathbf{N}_1 \,|\, [n\tau(\omega)](1 - c_1\varepsilon_n) \le j \le L_n(\omega) \quad \text{or} \quad L_n(\omega) < j < [n\tau(\omega)](1 - c_1\varepsilon_n)\},$$

$$I_n^* = I_n^*(\omega) = \{j \in \mathbf{N}_1 \,|\, j < [n\tau(\omega)](1 - c_1\varepsilon_n)\},$$

$$I_k = \{j \in \mathbf{N}_1 \,|\, j < k(1 - c_1\varepsilon_n)\},$$

$$J_n^* = J_n^*(\omega) = \{j \in \mathbf{N}_1 \,|\, [n\tau(\omega)](1 - c_1\varepsilon_n) \le j \le [n\tau(\omega)](1 + c_1\varepsilon_n)\},$$

$$J_k = \{j \in \mathbf{N}_1 \,|\, k(1 - c_1\varepsilon_n) \le j \le k(1 + c_1\varepsilon_n)\},$$

$$\delta_n = \delta_n(\omega) = \begin{cases} 1 & \text{if } [n\tau(\omega)](1 - c_1\varepsilon_n) \le L_n(\omega), \\ -1 & \text{otherwise.} \end{cases}$$

Proof of (i). We first prove (i) with $n\tau$ replaced by $[n\tau]$. Choose n large enough so that $\varepsilon_n < c_2^\delta$ and, using (6), remark that

$$P([n\tau]<[c_2\varepsilon_n^{-1/\delta}]) \leqq P\left(\tau < \frac{c_2}{n}\varepsilon_n^{-1/\delta}\right) = O(\sqrt{\varepsilon_n}). \tag{8}$$

Hence

$$\sup_x \left| P\left(\frac{\sqrt{[n\tau]}}{2\sigma}\binom{[n\tau]}{2}^{-1} \sum_{1\leqq i<j\leqq L_n} h(X_i, X_j) \leqq x\right) - \Phi(x)\right| \leqq$$

$$\leqq \sup_x \left| P\left(\frac{\sqrt{[n\tau]}}{2\sigma}\binom{[n\tau]}{2}^{-1} \sum_{1\leqq i<j\leqq L_n} h(X_i, X_j) \leqq x,\ [n\tau] \geqq [c_2\varepsilon_n^{-1/\delta}]\right) - \right.$$
$$\left. - \Phi(x)P([n\tau]\geqq [c_2\varepsilon_n^{-1/\delta}])\right| + O(\sqrt{\varepsilon_n}).$$

Putting $\psi(X_i, X_j) = h(X_i, X_j) - g(X_i) - g(X_j)$, the following decomposition holds on the set where $[n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}]$:

$$\frac{\sqrt{[n\tau]}}{2\sigma}\binom{[n\tau]}{2}^{-1} \sum_{1\leqq i<j\leqq L_n} h(X_i, X_j) =$$

$$= \frac{1}{\sigma\sqrt{[n\tau]}}\sum_{i=1}^{L_n} g(X_i) + \frac{1}{\sigma\sqrt{[n\tau]}([n\tau]-1)} \sum_{j\in I_n^*}\sum_{i=1}^{j-1} \psi(X_i, X_j) +$$

$$+ \left(\frac{L_n-1}{[n\tau]-1} - 1\right)\frac{1}{\sigma\sqrt{[n\tau]}}\sum_{i=1}^{L_n} g(X_i) + \frac{\delta_n}{\sigma\sqrt{[n\tau]}([n\tau]-1)} \sum_{j\in I_n^{**}}\sum_{i=1}^{j-1} \psi(X_i, X_j) =$$

$$= \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}.$$

Using a Slutsky argument and the Bonferroni inequality, it suffices to prove that

(i.A)

$$\sup_x |P(\mathrm{I}+\mathrm{II}\leqq x,\ [n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}]) - \Phi(x)P([n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}])| = O(\sqrt{\varepsilon_n}),\quad n\to\infty$$

$$P\left(|\mathrm{III}| > \frac{\sqrt{\varepsilon_n}}{2},\ [n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}]\right) = O(\sqrt{\varepsilon_n}),\quad n\to\infty \tag{i.B}$$

$$P\left(|\mathrm{IV}| > \frac{\sqrt{\varepsilon_n}}{2},\ [n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}]\right) = O(\sqrt{\varepsilon_n}),\quad n\to\infty. \tag{i.C}$$

Proof of (i.A)

$$\sup_x |P(\mathrm{I}+\mathrm{II}\leqq x,\ [n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}]) - \Phi(x)P([n\tau]\geqq[c_2\varepsilon_n^{-1/\delta}])| \leqq \tag{9}$$

$$\leqq \sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P([n\tau]=k)\sup_x \left|P\left(\sum_{i=1}^{L_n} g(X_i)+Y_k \leqq b_k(x)\,|\,[n\tau]=k\right) - \Phi(x)\right|$$

with $b_k(x)=x\sigma\sqrt{k}$ and $Y_k=\frac{1}{k-1}\sum_{j\in I_k}\sum_{i=1}^{j-1}\psi(X_i, X_j)$. On the summands in the r.h.s. of (9) we use the following inequality:

$$\sup_x\left|P\left(\sum_{i=1}^{L_n} g(X_i)+Y_k \leqq b_k(x)\mid[n\tau]=k\right)-\Phi(x)\right| \leqq \tag{10}$$

$$\leqq \sup_x\left|P\left(\sum_{i=1}^{k} g(X_i)+Y_k \leqq b_k(x)\right)-\Phi(x)\right|+$$

$$+\sup_x\left|P\left(\sum_{i=1}^{L_n} g(X_i)+Y_k \leqq b_k(x),\ L_n\in J_k\mid[n\tau]=k\right)-P\left(\sum_{i=1}^{k} g(X_i)+Y_k \leqq b_k(x)\right)\right|+$$
$$+P(L_n\notin J_k\mid[n\tau]=k).$$

Putting

$$r_k(x)=P\left(\sum_{i=1}^{L_n} g(X_i)+Y_k \leqq b_k(x),\ L_n\in J_k\mid[n\tau]=k\right),$$

$$s_k(x)=P\left(\sum_{i=1}^{k} g(X_i)+Y_k \leqq b_k(x)\right),$$

$$A_k(x)=\left\{\omega\,\Big|\max_{m\in J_k}\sum_{i=1}^{m} g(X_i)+Y_k \leqq b_k(x)\right\},$$

$$B_k(x)=\left\{\omega\,\Big|\min_{m\in J_k}\sum_{i=1}^{m} g(X_i)+Y_k \leqq b_k(x)\right\},$$

one has that $P(A_k(x))\leqq s_k(x)\leqq P(B_k(x))$ and $P(A_k(x), L_n\in J_k|[n\tau]=k)\leqq r_k(x)\leqq \leqq P(B_k(x))$, where we have used (7) to obtain the last inequality. Since $P(A_k(x))= =P(A_k(x), L_n\in J_k|[n\tau]=k)+P(A_k(x), L_n\notin J_k|[n\tau]=k)$ it follows that $|r_k(x)-s_k(x)|\leqq \leqq P(B_k(x))-P(A_k(x))+P(A_k(x), L_n\notin J_k|[n\tau]=k)$ and hence that

$$\sup_x|r_k(x)-s_k(x)| \leqq \sup_x\left(P(B_k(x))-P(A_k(x))\right)+P(L_n\notin J_k\mid[n\tau]=k). \tag{11}$$

An application of Lemma 3 yields that there exists a constant C such that

$$\sup_x\left|P\left(\sum_{i=1}^{k} g(X_i)+Y_k \leqq b_k(x)\right)-\Phi(x)\right| \leqq Ck^{-\delta/2}. \tag{12}$$

Applying (11) and (12) on the r.h.s. of (10) and using the obtained inequality in (9) leads to:

$$\sup_x|P(\mathrm{I}+\mathrm{II} \leqq x,\ [n\tau] \geqq [c_2\varepsilon_n^{-1/\delta}])-\Phi(x)P([n\tau] \geqq [c_2\varepsilon_n^{-1/\delta}])| \leqq$$

$$\leqq \sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P([n\tau]=k)\sup_x\left(P(B_k(x))-P(A_k(x))\right)+$$

$$+2\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P(L_n\notin J_k\mid[n\tau]=k)P([n\tau]=k)+C\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} k^{-\delta/2}P([n\tau]=k).$$

Now, remark that

(13) $$\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} k^{-\delta/2}P([n\tau]=k) \leqq c_2^{-\delta/2}\sqrt{\varepsilon_n}\,P([n\tau] \geqq [c_2\varepsilon_n^{-1/\delta}]) = O(\sqrt{\varepsilon_n})$$

and that, using (5),

(14) $$\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P(L_n \notin J_k \mid [n\tau]=k)P([n\tau]=k) \leqq P(L_n \notin J_n^*) = O(\sqrt{\varepsilon_n}).$$

Hence, it suffices to show that

(15) $$\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P([n\tau]=k)\sup_x (P(B_k(x))-P(A_k(x))) = O(\sqrt{\varepsilon_n}).$$

Putting $p=\min J_k$, $q=\max J_k$, $r=\max I_k$ and remarking that $r=p-1$, it follows from Lemma 2 in [2] that

$$P(B_k(x))-P(A_k(x)) \leqq c\Big\{P\Big(\sum_{i=1}^{p} g(X_i) \leqq b_k(x)-Y_k,\ \sum_{i=1}^{q} g(X_i) \geqq b_k(x)-Y_k\Big)+$$

$$+P\Big(\sum_{i=1}^{p} g(X_i) \geqq b_k(x)-Y_k,\ \sum_{i=1}^{q} g(X_i) \leqq b_k(x)-Y_k\Big)\Big\}$$

for some constant c. We now use Lemma 3 from [2] with X replaced by

$\sum_{i=1}^{p} g(X_i)+Y_k$; Y by $\sum_{i=p+1}^{q} g(X_i)$; b by $\sigma\sqrt{k}$; d by $Ck^{-\delta/2}$ and t by $b_k(x)$. We then obtain that for constants K and L:

$$\sup_x (P(B_k(x))-P(A_k(x))) \leqq Kk^{-\delta/2}+Lk^{-1/2}E\Big|\sum_{i=p+1}^{q} g(X_i)\Big| \leqq Kk^{-\delta/2}+\sigma L\sqrt{\frac{q-p}{k}}$$

where the last inequality follows from $E\Big|\frac{1}{\sqrt{q-p}}\sum_{i=p+1}^{q} g(X_i)\Big| \leqq \sigma$ by the moment inequality and the independence of the X_i's together with $Eg(X_1)=0$. Inserting this result into the l.h.s. of (15), after remarking that $\sqrt{\frac{q-p}{k}} \leqq \sqrt{2c_1\varepsilon_n}$, and using (13), one arrives at the desired order bound $O(\sqrt{\varepsilon_n})$, completing the proof of (i.A).

Proof of (i.B). From (8) and (14) it follows that

(16) $$P\Big(|\mathrm{III}| > \frac{\sqrt{\varepsilon_n}}{2},\ [n\tau] \geqq [c_2\varepsilon_n^{-1/\delta}]\Big) \leqq O(\sqrt{\varepsilon_n})+$$

$$+\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty} P\Big(\Big|\Big(\frac{L_n-1}{k-1}-1\Big)\frac{1}{\sigma\sqrt{k}}\sum_{i=1}^{L_n} g(X_i)\Big| > \frac{\sqrt{\varepsilon_n}}{2},\ L_n \in J_k \mid [n\tau]=k\Big)P([n\tau]=k).$$

Using (7) and the fact that $\max_{m\in J_k}|m-k|\leq kc_1\varepsilon_n$, one obtains:

$$P\left(\left|\frac{L_n-1}{k-1}-1\right|\left|\sum_{i=1}^{L_n}g(X_i)\right|>\frac{\sigma\sqrt{k\varepsilon_n}}{2},\ L_n\in J_k|[n\tau]=k\right)\leq$$

$$\leq P\left(\max_{m\in J_k}\left|\frac{m-k}{k-1}\right|\max_{m\in J_k}\left|\sum_{i=1}^{m}g(X_i)\right|>\frac{\sigma\sqrt{k\varepsilon_n}}{2}\right)\leq P\left(\max_{m\in J_k}\left|\sum_{i=1}^{m}g(X_i)\right|>\frac{\sigma(k-1)}{2c_1\sqrt{k\varepsilon_n}}\right).$$

Since $\sum_{i=1}^{m}g(X_i)$, $m=1,2,\ldots$, forms a martingale, the Kolmogorov inequality yields that

$$P\left(\max_{p\leq m\leq q}\left|\sum_{i=1}^{m}g(X_i)\right|>\frac{\sigma(k-1)}{2c_1\sqrt{k\varepsilon_n}}\right)\leq\frac{4c_1^2k\varepsilon_n}{\sigma^2(k-1)^2}E\left(\sum_{i=1}^{q}g(X_i)\right)^2=$$

$$=4c_1^2qk\varepsilon_n/(k-1)^2=O(\sqrt{\varepsilon_n})$$

showing that the r.h.s. of (16) is of the order $O(\sqrt{\varepsilon_n})$.

Proof of (i.C). Using the same reasoning as in the proof of (i.B) and remembering that $\delta_n=1$ if $[n\tau](1-c_1\varepsilon_n)\leq L_n$, one has:

$$P\left(|\mathrm{IV}|>\frac{\sqrt{\varepsilon_n}}{2},\ [n\tau]\geq[c_2\varepsilon_n^{-1/\delta}]\right)\leq$$

$$\leq\sum_{k=[c_2\varepsilon_n^{-1/\delta}]}^{\infty}P\left(\max_{m\in J_k}\left|\sum_{j=p}^{m}\sum_{i=1}^{j-1}\psi(X_i,X_j)\right|>\frac{\sigma\sqrt{k\varepsilon_n}(k-1)}{2}\right)P([n\tau]=k)+O(\sqrt{\varepsilon_n}).$$

Further, it is well-known that $V_m=\sum_{j=p}^{m}\sum_{i=1}^{j-1}\psi(X_i,X_j)$, $m=p,p+1,\ldots,q$, and also $W_k=\sum_{i=1}^{k}\psi(X_i,X_j)$, $k=1,2,\ldots,j-1$ are martingales. An application of the Kolmogorov inequality and a theorem in [8] lead to $\left(\text{denote } \frac{4+\delta}{3} \text{ by } s\right)$

$$P\left(\max_{p\leq m\leq q}\left|\sum_{j=p}^{m}\sum_{i=1}^{j-1}\psi(X_i,X_j)\right|>\frac{\sigma\sqrt{k\varepsilon_n}(k-1)}{2}\right)\leq 2^s\sigma^{-s}(k-1)^{-s}(k\varepsilon_n)^{-s/2}E|V_q|^s\leq$$

$$\leq K(k-1)^{-s}(k\varepsilon_n)^{-s/2}\sum_{j=p}^{q}E|W_{j-1}|^s\leq K'(k-1)^{-s}(k\varepsilon_n)^{-s/2}(q-p+1)q$$

where K and K' are constants. A short computation, using $q-p\leq 2kc_1\varepsilon_n$ and $q\leq k(1+c_1\varepsilon_n)$, yields the desired order bound $O(\sqrt{\varepsilon_n})$. To complete the proof of (i), we have to show that $[n\tau]$ can be replaced by $n\tau$. An application of Lemma 1 of [7] yields that it is sufficient to prove that for some constant C

$$P\left(\left|\frac{\sqrt{n\tau(n\tau-1)}}{\sqrt{[n\tau]([n\tau]-1)}}-1\right|>C\sqrt{\varepsilon_n}\right)=O(\sqrt{\varepsilon_n}). \tag{17}$$

That (17) holds follows from Lemma 2 with $X_n=Y_n=n\tau$ and $k=1$, checking by (6) that (1) and (2) are satisfied.

Proof of (ii). As above, it can be shown that $[n\tau]$ may also be replaced by L_n. We take $X_n=L_n$, $Y_n=n\tau$ and $k=1$ in Lemma 2. Since (1) and (2) then coincide with (5) and (6), the proof of (ii) is complete.

Proof of (iii). We first show that

$$P\left(\left|\frac{L_n\sigma_{L_n}^2}{4\sigma^2}-1\right|>C^2\varepsilon_n\right)=O(\sqrt{\varepsilon_n}) \quad \text{with} \quad C^2=\frac{2E\psi^2(X_1,X_2)}{c_2\sigma^2}.$$

Using that $n\sigma_n^2=4\sigma^2+\frac{2}{n-1}E\psi^2(X_1,X_2)$, this follows from condition (6) after easy manipulation. Since

$$P\left(\left|\frac{\sigma_{L_n}\sqrt{L_n}}{2\sigma}-1\right|>C\sqrt{\varepsilon_n}\right)\leqq P\left(\left|\frac{L_n\sigma_{L_n}^2}{4\sigma^2}-1\right|>C^2\varepsilon_n\right)=O(\sqrt{\varepsilon_n})$$

a lemma of [7] makes it possible to go from (ii) to (iii). This finishes the proof of the theorem.

We close with a result concerning the case when the indices are independent of the basic sequence. The details of the proof are of course simpler than in the general case (for instance, there is no need for the decomposition of U_n) and therefore are not given here.

Theorem 2. *Let the assumptions of Theorem 1 be fulfilled with* (5) *deleted and* (7) *replaced by:* L_n, τ and X_k, $k=1,2,\ldots$ *are independent for each* $n=1,2,\ldots$. *Then*

(a) *if* $P\left(\left|\frac{L_u}{[n\tau]}-1\right|>c_1\sqrt{\varepsilon_n}\right)=O(\sqrt{\varepsilon_n})$, *the results* (i), (ii) *and* (iii) *of Theorem* 1 *hold;*

(b) *if* $P\left(\frac{L_n}{[n\tau]}<1-\alpha\right)=O(\sqrt{\varepsilon_n})$ *for some constant* $\alpha<1$, *the results* (ii) *and* (iii) *of Theorem* 1 *hold.*

Acknowledgement. The authors thank the editor and a referee for careful reading and precise remarks which have resulted in a considerable improvement of the presentation of the paper.

References

[1] M. Aerts, *Asymptotic distribution theory for random U- and GM-statistics, with application in sequential estimation*, Ph. D. Dissertation, Limburgs Universitair Centrum (1988).

[2] M. Aerts and H. Callaert, The exact approximation order in the central limit theorem for random U-statistics, *Sequential Analysis*, **5** (1986), 19—35.

[3] H. Callaert and P. Janssen, A note on the convergence rate of random sums, *Rev. Roumaine Math. Pures Appl.*, **28** (1983), 147—151.

[4] S. Csörgő, On weak convergence of the empirical process with random sample size, *Acta Sci. Math.*, **36** (1974), 17—25, 375—376.

[5] K. O. Friedrich, A Berry—Esseen bound for functions of independent random variables, *Ann. Statistics*, **17** (1989), 170—183.

[6] D. Landers and L. Rogge, A counterexample in the approximation theory of random summation, *Ann. Probability*, **5** (1977), 1018—1023.

[7] R. Michel and J. Pfanzagl, The accuracy of the normal approximation for minimum contrast estimates, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, **18** (1971), 73—84.

[8] B. von Bahr and C. G. Esseen, Inequalities for the rth absolute moment of a sum of random variables, $1 \leqq r \leqq 2$, *Ann. Math. Statist.*, **36** (1965), 299—303.

LIMBURGS UNIVERSITAIR CENTRUM
DEPARTMENT OF MATHEMATICS
UNIVERSITAIRE CAMPUS
B-3610 DIEPENBEEK, BELGIUM