ON L^p-CONVERGENCE OF U-STATISTICS*

PRANAB KUMAR SEN

(Received Oct. 21, 1971; revised Dec. 16, 1972)

For a kernel belonging to L^p-space, $p>1$, the rate of convergence of Hoeffding's [5] U-statistic to its expectation is studied; this includes as a special case a similar result on the sample mean previously studied by Chung [3]. Also, an L^p-convergence result of Pyke and Root [8] on the sample partial sum is extended to U-statistics.

1. Statement of the results

Let $\{X_i, i\geqq 1\}$ be a sequence of independent and identically distributed random variables (i.i.d.r.v.) with a distribution function $F(x)$, defined on the real line R. For the sequence $\{\bar{X}_n=n^{-1}\sum_{i=1}^{n} X_i, n\geqq 1\}$ of sample means, using an inequality of Marcinkiewicz and Zygmund [7], Chung [3] has shown that if $X_i \in L^p$, $p>1$, then for every $\varepsilon>0$, there exists a positive $C(\varepsilon)<\infty$, such that

$$\text{(1.1)}\qquad \mathrm{P}\{|\bar{X}_n-\mathrm{E}\,\bar{X}_n|>\varepsilon\}<C(\varepsilon)n^{-s}\,;\quad s=\begin{cases} p-1\,, & 1<p<2\,,\\ p/2\,, & p\geqq 2\,.\end{cases}$$

It is also known [viz., Pyke and Root [8]] that for $X_i \in L^p$, $0<p<2$,

$$\text{(1.2)}\qquad n^{(p-1)/p}|\bar{X}_n-\alpha|\to 0 \text{ almost surely (a.s.) and in } L^p,$$

as $n\to\infty$, where $\alpha=0$ for $p<1$, and $\alpha=\mathrm{E}\,\bar{X}_n$ for $p\geqq 1$. Our contention is to show that similar results also hold for Hoeffding's [5] U-statistics.

For a Borel-measurable *kernel* $\phi(X_1,\cdots,X_m)$, symmetric in its arguments, of degree m $(\geqq 1)$, the U-statistic based on a sample of size n $(\geqq m)$ is defined by

$$\text{(1.3)}\qquad U_n=\binom{n}{m}^{-1}\sum_{C_{n,m}}\phi(X_{i_1},\cdots,X_{i_m})\,;\quad C_{n,m}=\{1\leqq i_1<\cdots<i_m\leqq n\}\,.$$

Thus, if $\phi\in L^{-1}$, then U_n unbiasedly estimates the functional of the dis-

* Work supported by the Army Research Office, Durham, Grant DA-ARO-D-31-124-70-G6.

tribution F defined by

$$\theta(F)=\int_{R^m}\cdots\int \phi(x_1,\cdots,x_m)dF(x_1)\cdots dF(x_m)\,. \tag{1.4}$$

Also, for $\phi\in L^1$, $\{U_n, n\geqq m\}$ forms a reverse martingale sequence [cf. Berk [1]], and hence, $U_n\to\theta(F)$ a.s. as $n\to\infty$; however, no specific rate of convergence is known. If $\phi\in L^2$, it follows from Hoeffding [5] that if $\theta(F)$ is stationary of order d ($\geqq 0$), then $\operatorname{var}\{U_n\}=O(n^{-(d+1)})$, so that by the Chebychev inequality, $\mathrm{P}\{|U_n-\theta(F)|>\varepsilon\}\leqq C(\varepsilon)n^{-s}$, $s=(d+1)/2$. On the other hand, when p is not an even integer, the usual difficulties associated with the computation of the p-th moment of $|U_n-\theta(F)|$ introduces complications in the proof of (1.1) [or (1.2)]. This leads us to prove formally the following theorem which is useful in the study of sequential procedures based on U-statistics.

THEOREM 1. *If $\phi\in L^p$, $p>1$, then for every $\varepsilon>0$, there exists a positive $C(\varepsilon)$ ($<\infty$), such that for all $n\geqq m$,*

$$\mathrm{P}\{|U_n-\theta(F)|>\varepsilon\}\leqq C(\varepsilon)n^{-s}\,;\quad s=\begin{cases} p-1\,, & 1<p<2\,,\\ p/2\,, & p\geqq 2\,.\end{cases} \tag{1.5}$$

Also, for $\phi\in L^p$, $0<p<2$, as $n\to\infty$,

$$n^{r(p)}|U_n-\alpha|\to 0 \text{ a.s. and in } L^p, \tag{1.6}$$

where $r(p)=m(1-p^{-1})$, $\alpha=0$ for $p<1$, and $r(p)=1-p^{-1}$, $\alpha=\theta(F)$ for $p\geqq 1$.

Note that for $p<1$, whereas $|\bar{X}_n|=o(n^{p^{-1}-1})$ a.s. (and in L^p), our $|U_n|=o(n^{m(p^{-1}-1)})$ a.s. (and in L^p); we have not been able to equalize the rates for $p<1$. Also, for $1<p<2$, using the L^p-convergence in (1.6) along with the Markov inequality, we have for every $\varepsilon>0$, $\mathrm{P}\{|U_n-\theta(F)|>\varepsilon\}=o(n^{-p+1})$, whereas in (1.5), we have $o(n^{-p+1})$. Our proof is based on a decomposition of U_n, essentially due to Hoeffding [6], and certain other L^p-convergence results on martingales in Chatterji [2] and Dharmadhikari, Fabian and Jogdeo [4]. The theorem remains good when the X_i are q-dimensional stochastic vectors, for $q\geqq 1$.

2. The proof of the theorem

For every h ($1\leqq h\leqq m$), we let for $\phi\in L^p$, $p\geqq 1$,

$$\phi_h(x_1,\cdots,x_h)=\mathrm{E}\,\phi(x_1,\cdots,x_h,X_{h+1},\cdots,X_m)\,;\quad \phi_0=\theta(F)\,; \tag{2.1}$$

$$W_n^{(h)}=\binom{n}{h}^{-1}\sum_{C_{n,h}}\phi_h(X_{i_1},\cdots,X_{i_h})\,;\quad W_n^{(0)}=\theta(F)\,; \tag{2.2}$$

$$U_n^{(h)}=\sum_{k=0}^{h}(-1)^k\binom{h}{k}W_n^{(h-k)}\,; \quad U_n^{(0)}=\theta(F)\,. \tag{2.3}$$

Then, by (1.3), (1.4), (2.2) and (2.3), for $\phi\in L^p$, $p\geqq 1$, we have

$$[U_n-\theta(F)]=\sum_{h=1}^{m}\binom{m}{h}U_n^{(h)}\,, \qquad n\geqq m\,. \tag{2.4}$$

Since $\phi\in L^p$, $p\geqq 1\Rightarrow\phi_h\in L^p$, $p\geqq 1$ for all $h\leqq m$, in order to prove (1.5), it suffices to show that for every $h(1\leqq h\leqq m)$ and $\varepsilon_h>0$, there exists a positive $C(\varepsilon_h)$ $(<\infty)$, such that for all $n\geqq h$,

$$\mathrm{P}\{|U_n^{(h)}|>\varepsilon_h\}\leqq C(\varepsilon_h)n^{-s}\,; \quad h=1,\cdots,m\,, \tag{2.5}$$

where s is defined in (1.5). Since $U_n^{(1)}=n^{-1}\sum_{i=1}^{n}[\phi_1(X_i)-\theta(F)]$ involves i.i.d.r.v., by (1.1), (2.5) holds for $h=1$. So, we require only to prove (2.5) for $h\geqq 2$. Let us now define

$$V_n^{(h)}=\binom{n}{h}U_n^{(h)}\,, \qquad h=2,\cdots,m\,, \tag{2.6}$$

and let $\mathcal{B}_n$ be the σ-field generated by the order statistics corresponding to $X_1,\cdots,X_n$, so that $\mathcal{B}_n$ is $\uparrow$ in n $(\geqq 1)$.

LEMMA 2. *For every h $(2\leqq h\leqq m)$, $\{V_n^{(h)},\mathcal{B}_n\,;\ n\geqq h\}$ is a martingale.*

The proof readily follows from (2.2), (2.3), (2.6) and some standard computations; for details, we may refer to Hoeffding [6].

LEMMA 3. *If $\phi\in L^p$, $p\geqq 1$, then $\mathrm{E}\,|V_{n+1}^{(h)}-V_n^{(h)}|^p<Cn^{p(h-1)}$, $2\leqq h\leqq m$, where C $(<\infty)$ does not depend on n.*

PROOF. By (2.2), (2.3) and (2.6), $V_{n+1}^{(h)}-V_n^{(h)}$ involves a sum over $2^h\binom{n}{h-1}$ terms of the type $\phi_g(X_{i_1},\cdots,X_{i_g})$, $1\leqq i_1<\cdots<i_g\leqq n+1$, $0\leqq g\leqq h$, and hence, the lemma follows directly by using the generalized C_p-inequality $\left|\sum_{j=1}^{N}C_j\right|^p\leqq N^{p-1}\sum_{j=1}^{N}|C_j|^p$, for $p\geqq 1$ along with $\mathrm{E}\,|\phi_g|\leqq\mathrm{E}\,|\phi|<\infty$ for all $0\leqq g\leqq h\leqq m$, when $\phi\in L^p$.

LEMMA 4. *If $\phi\in L^p$, $1<p\leqq 2$, then*

$$\mathrm{E}\,|U_n^{(h)}|^p\leqq C^*n^{-(p-1)}\,, \quad C^*<\infty\,, \quad \text{for all}\ \ 2\leqq h\leqq m\,. \tag{2.7}$$

PROOF. By our Lemma 3 and Lemma 1 of Chatterji [2], we have for every h $(2\leqq h\leqq m)$ and $n\geqq h$,

$$\mathrm{E}\,|V_n^{(h)}|^p\leqq 2\Big\{\mathrm{E}\,|V_h^{(h)}|^p+\sum_{j=h+1}^{n}\mathrm{E}\,|V_j^{(h)}-V_{j-1}^{(h)}|^p\Big\} \tag{2.8}$$

$$\leqq 2C\left\{1+\sum_{j=h}^{n-1} j^{p(h-1)}\right\} \leqq C^* n^{ph-p+1}; \quad C^*<\infty .$$

Hence, the lemma directly follows from (2.6) and (2.8).

LEMMA 5. *If $\phi \in L^p$, $p \geqq 2$, then for every h $(2 \leqq h \leqq m)$,*

$$\mathrm{E}|U_n^{(h)}|^p \leqq C^* n^{-p/2}, \qquad C^*<\infty . \tag{2.9}$$

PROOF. By our Lemma 3 and a theorem in Dharmadhikari, Fabian and Jogdeo [4], for $\phi \in L^p$, $p \geqq 2$,

$$\mathrm{E}|V_n^{(h)}|^p \leqq C_p n^{p/2} \beta_{n,p}^{(h)}, \quad C_p<\infty ; \quad 2 \leqq h \leqq m , \tag{2.10}$$

where

$$\beta_{n,p}^{(h)} = \left\{\mathrm{E}|V_h^{(h)}|^p + \sum_{j=h}^{n-1} \mathrm{E}|V_{j+1}^{(h)} - V_j^{(h)}|^p\right\} \Big/ n \leqq C^* n^{ph-p}, \tag{2.11}$$

$$C^*<\infty \text{ (as in (2.8))}, \qquad 2 \leqq h \leqq m .$$

Hence the lemma follows directly from (2.6), (2.10) and (2.11).

Now, by the preceding two lemmas, we have for every h: $2 \leqq h \leqq m$,

$$\mathrm{E}|U_n^{(h)}|^p \leqq C^* n^{-s}, \quad \phi \in L^p, \quad p>1 , \tag{2.12}$$

where s is defined in (1.5). Hence, (2.5) directly follows from (2.12) and the Markov inequality.

To prove (1.6), we first consider the case $0<p<1$. We let $Z_1 = \phi(X_1, \cdots, X_m)$, and for $k>m$, the set of $\binom{k-1}{m-1}$ kernels $\phi(X_k, X_{i_2}, \cdots, X_{i_m})$, $1 \leqq i_2 < \cdots < i_m \leqq k-1$ is indexed in an arbitrary order as $Z_{\binom{k-1}{m}+j}$, $j=1, \cdots, \binom{k-1}{m-1}$, so that

$$U_N = N^{-1} \sum_{j=1}^{N} Z_j ; \quad N = \binom{n}{m}, \quad n \geqq m . \tag{2.13}$$

The Z_j are marginally identically distributed and $\int |Z_j|^p < \infty$ for $\phi \in L^p$. So the conditions of the theorem of Chatterji [2] are satisfied, and hence,

$$N^{-1/p} \sum_{j=1}^{N} Z_j = \binom{n}{m}^{1-p^{-1}} U_n \to 0 \text{ a.s. and in } L^p, \text{ as } n \to \infty. \tag{2.14}$$

This completes the proof for $p<1$. The case of $p=1$ follows readily by using the reverse martingale property of U_n, as mentioned in Section 1. So, in the remaining, we confine ourselves only to $1<p<2$.

We are entitled here to use (2.4), where by Pyke and Root [8], (2.15) $$\lim_{n\to\infty}\{n^{1-p^{-1}}U_n^{(1)}\}=0 \text{ a.s. and in } L^p,\ 1<p<2.$$

Also, by (2.2) and (2.3), for every h $(2\leqq h\leqq m)$, $n\leqq h$,

(2.16) $$U_n^{(h)}=\binom{n}{h}^{-1}\sum_{C_{n,h}}\phi_h^*(X_{i_1},\cdots,X_{i_h});$$

(2.17) $$\begin{aligned}&\phi_h^*(X_1,\cdots,X_h)\\ &=\phi_h(X_1,\cdots,X_h)-\sum_{j=1}^{h}\phi_{h-1}(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_h)\\ &\quad+\sum_{1\leqq j<k\leqq h}\phi_{h-2}(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{k-1},X_{k+1},\cdots,X_h)\cdots\\ &\quad+(-1)^h\theta(F).\end{aligned}$$

Note that $\phi\in L^p\Rightarrow\phi_h^*\in L^p$ for all $h\leqq m$, and by (2.17),

(2.18) $$\mathrm{E}|\phi_h^*|^p\leqq 2^{h(p-1)}\mathrm{E}|\phi|^p<\infty \quad\text{for}\quad \phi\in L^p,\quad 1\leqq p<2;\quad h\leqq m.$$

Thus, if we write for $n\geqq h$

(2.19) $$f_n^{(h)}=\binom{n-1}{h-1}^{-1}\sum_n^{*}\phi_n^*(X_n,X_{i_2},\cdots,X_{i_h})=\binom{n-1}{h-1}^{-1}[V_n^{(h)}-V_{n-1}^{(h)}],$$

where the summation $\sum_n^{*}$ extends over all $1\leqq i_2<\cdots<i_h\leqq n-1$, we have

(2.20) $$\sup_n \mathrm{E}|f_n^{(h)}|^p\leqq \mathrm{E}|\phi_h^*(X_1,\cdots,X_h)|^p<2^h\cdot\mathrm{E}|\phi|^p<\infty,\qquad \phi\in L^p;$$

(2.21) $$\mathrm{E}\{f_n^{(h)}|\mathcal{B}_{n-1}\}=0 \text{ a.s., by Lemma 2 and (2.19).}$$

Finally, writing for $j\leqq n$ $(\geqq h)$

(2.22) $$b_{nj}^{(h)}=h\binom{j-1}{h-1}\binom{n-1}{h-1}^{-1},\qquad 2\leqq h\leqq m,$$

we have from (2.16), (2.19) and (2.22),

(2.23) $$n^{1-p^{-1}}U_n^{(h)}=n^{-p^{-1}}\sum_{j=h}^{n}b_{nj}^{(h)}f_j^{(h)},\quad n\geqq h,\quad 2\leqq h\leqq m.$$

Let us then consider the following lemma.

LEMMA 6. *Let $\{Z_n, n\geqq 1\}$ be a martingale difference sequence, $\{a_{nj}, 1\leqq j\leqq n\}$, $n\geqq 1$ be sequences of real numbers such that* $\sup_n\max_{1\leqq j\leqq n}|a_{nj}|<\infty$, *and for* $1<p<2$, $\sup_n \mathrm{E}|Z_n|^p<\infty$. *Then*

(2.24) $$\lim_{n\to\infty}\left\{n^{-1/p}\sum_{j=1}^{n}a_{nj}Z_j\right\}=0 \text{ a.s. and in } L^p.$$

The proof follows virtually on the same lines as in the second part of the proof of the theorem in Chatterji ([2], p. 1069), and hence, is omitted.

Now, by (2.22), $\sup_n \max_{1\leq j\leq n} |b_{nj}^{(h)}|=h$, and hence, by (2.20), (2.21), (2.23) and Lemma 6, we obtain that for $\phi \in L^p$, $1<p<2$

$$\lim_{n\to\infty} \{n^{1-p^{-1}} U_n^{(h)}\}=0 \text{ a.s. and in } L^p, \text{ for all } h\leqq m. \tag{2.25}$$

From (2.4), (2.15) and (2.25) the a.s. result in (1.6) follows directly, while using, in addition, the Minkowski inequality, the L^p-convergence in (1.6) follows. Q.E.D.

Acknowledgement

The author is grateful to the referee for his critical reading of the manuscript and useful comments.

REFERENCES

[1] Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect, *Ann. Math. Statist.*, **37**, 51-58.

[2] Chatterji, S. D. (1969). An L^p-convergence theorem, *Ann. Math. Statist.*, **40**, 1068-1070.

[3] Chung, K. L. (1951). The strong law of large numbers, *Proc. Second Berkeley Symp. Math. Statist. Prob.*, 341-352. (Univ. Calif. Press.)

[4] Dharmadhikari, S. W., Fabian, V., and Jogdeo, K. (1968). Bounds on moments of martingales, *Ann. Math. Statist.*, **39**, 1719-1723.

[5] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.*, **19**, 293-325.

[6] Hoeffding, W. (1961). The strong law of large numbers for *U*-statistics, *Inst. Statist., Univ. North Carolina, Mimeo Report*, No. 302.

[7] Marcinkiewicz, J., and Zygmund, A. (1937). Sur les fonctions independantes, *Fund. Math.*, **29**, 60-90.

[8] Pyke, R., and Root, D. (1968). On convergence in *r*-means of normalized partial sums, *Ann. Math. Statist.*, **39**, 379-381.