UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

Mathematical Sciences Directorate
Air Force Office of Scientific Research
Washington 25, D. C.

AFOSR Report No.

THE STRONG LAW OF LARGE NUMBERS FOR U-STATISTICS

by

Wassily Hoeffding

July, 1961

Contract No. AF 49(638)-261

Let X_1, X_2, ... be independent, identically distributed random variables, let $f(X_1, \ldots, X_r)$ be a function whose expected value, Θ, exists, and let $\bar{f}_n$ be the arithmetic (with equal weights) of the values $f(X_{i_1}, \ldots, X_{i_r})$, for all r-tuples $(i_1, \ldots, i_r)$ of distinct positive integers not exceeding n. It is shown that $\bar{f}_n \longrightarrow \Theta$ almost surely.

Institute of Statistics
Mimeograph Series No. 302

THE STRONG LAW OF LARGE NUMBERS FOR U-STATISTICS

by Wassily Hoeffding[1]

University of North Carolina

1. Introduction. Let $X_1, X_2, \ldots$ be a sequence of mutually independent, identically distributed random variables taking values in a space $\mathcal{X}$. Let r be a positive integer and f a real-valued measurable function on $\mathcal{X}^r$. For $n \geq r$ define $\bar{f}_n$ as the arithmetic mean

$$(1) \qquad \bar{f}_n = \frac{1}{n(n-1)\ldots(n-r+1)} \Sigma f(X_{i_1}, \ldots, X_{i_r}) \qquad ,$$

where the sum is extended over all r-tuples $i_1, \ldots, i_r$ of distinct positive integers not exceeding n. The following theorem will be proved. (We write a.s. for almost surely or almost sure.)

Theorem. If $E[\ |f(X_1, \ldots, X_r)|\ ] < \infty$, then

$$(2) \qquad \bar{f}_n \longrightarrow E\ [f(X_1, \ldots, X_r)] \qquad \text{a.s.}$$

The theorem contains, for r = 1, the sufficiency part of Kolmogorov's strong law of large numbers. The asymptotic behavior of random variables of the form (1) (which have been called U-statistics) has been studied, among others, by the author in [2], where a central limit theorem for such sums has been proved, and by P. K. Sen [3], who obtained the result (2) under the assumption that a moment of order higher than $2 - r^{-1}$ is finite.

The proof of the theorem makes use of the fact that $\bar{f}_n$ can be represented as a linear combination of r random variables each of which has a martingale

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

property (Lemma 1). The summands are suitably truncated and a version of Doob's extension to martingales of Kolmogorov's inequality (see Lemma 2) is applied.

2. Preliminaries. We first observe that since the sum in (1) is symmetric, each term may be replaced by the arithmetic mean of the $r!$ terms whose subscripts are permutations of the same set of integers. We thus may, and shall assume that f is invariant under permutations:

(3) $$f(x_{i_1}, \ldots, x_{i_r}) = f(x_1, \ldots, x_r)$$

for all permutations $(i_1, \ldots, i_r)$ of $(1, \ldots, r)$. We now can write

(4) $$\bar{f}_n = \binom{n}{r}^{-1} S_n \quad , \qquad S_n = \sum_{1 \le i_1 < \ldots < i_r \le n} f(X_{i_1}, \ldots, X_{i_r}) \quad .$$

We shall also assume with no loss of generality that $E\,[f(X_1, \ldots, X_r)]$ $= 0$ or

(5) $$\int \cdots \int f(x_1, \ldots, x_r)\, dF(x_1) \ldots dF(x_r) = 0 \quad ,$$

where F denotes the common probability measure of the random variables X_n.

Lemma 1. If condition (5) is satisfied, we have

(6) $$\binom{n}{r}^{-1} S_n = \sum_{h=1}^{r} \binom{r}{h} \binom{n}{h}^{-1} S_{h,n} \quad ,$$

where

(7) $$S_{h,n} = \sum_{1 \le i_1 < \ldots < i_h \le n} f_h(X_{i_1}, \ldots, X_{i_h}) \quad ,$$

the function f_h $(h = 1, \ldots, r)$ is invariant under permutations of its arguments and satisfies the condition

(8) $$\int f_h(x_1, \ldots, x_{h-1}, x)\, dF(x) \equiv 0 \quad ;$$

and $S_{h,n}$ has the martingale property

$$E\left[S_{h,n} \mid X_1, \ldots, X_m\right] = S_{h,m} \quad , \quad h \leq m \leq n \quad . \tag{9}$$

Proof. We shall write $f(x_1, \ldots, x_{r-1}, *)$ for $\int f(x_1, \ldots, x_{r-1}, x)\, dF(x)$, etc. We define the functions f_h as follows.

$$\begin{aligned} f_1(x_1) &= f(x_1, *, \ldots, *) \quad , \\ f_2(x_1,x_2) &= f(x_1,x_2,*,\ldots,*) - f_1(x_1) - f_1(x_2) \quad , \\ f_3(x_1,x_2,x_3) &= f(x_1,x_2,x_3,*,*,*) - f_1(x_1) - f_1(x_2) - f_1(x_3) \\ &\quad - f_2(x_1,x_2) - f_2(x_1,x_3) - f_2(x_2,x_3) \\ \ldots\ldots &\quad \ldots\ldots \\ f_r(x_1, \ldots, x_r) &= f(x_1, \ldots, x_r) - f_1(x_1) - \ldots - f_1(x_r) \\ &\quad - f_2(x_1,x_2) - f_2(x_1,x_3) - \ldots - f_2(x_{r-1},x_r) \\ &\quad \ldots - f_{r-1}(x_1,\ldots,x_{r-1}) - \ldots - f_{r-1}(x_2,\ldots,x_r) \quad . \end{aligned} \tag{10}$$

The last equation expresses f in terms of $f_1, \ldots, f_r$. Inserting this expression in (4), we obtain equation (6). The invariance of f_h under permutations and equation (8) follow from the definition of the f_h and assumption (5). Finally, for $h \leq m < n$, $S_{h,n} - S_{h,m}$ is a sum of terms $f_h(X_{i_1}, \ldots, X_{i_h})$ with $i_j > m$ for some j. Hence, due to (8), the conditional expectation of $S_{h,n} - S_{h,m}$ when $X_1, \ldots, X_m$ are fixed is 0, and the martingale property (9) follows.

Due to Lemma 1 we may assume that

$$\int f(x_1, \ldots, x_{r-1}, x)\, dF(x) \equiv 0 \tag{11}$$

and hence

(12) $$E\left[S_n \mid X_1, \ldots, X_m \right] = S_m \quad , \quad r \leq m \leq n \quad .$$

The theorem will be proved if we show that $n^{-r} S_n \longrightarrow 0$ a.s. To this end we first truncate the random variables $f(X_{i_1}, \ldots, X_{i_r})$ as follows. For $j > 0$ let

(13) $$f^{(j)}(x_1, \ldots, x_r) = \begin{cases} f(x_1, \ldots, x_r) \text{ if } |f(x_1, \ldots, x_r)| \leq j^r \\ 0 \qquad\qquad \text{otherwise} \end{cases} ,$$

(14) $$S_n' = \sum_{1 \leq i_1 < \ldots < i_r \leq n} f^{(i_r)}(X_{i_1}, \ldots, X_{i_r}) \quad .$$

It will be shown that $n^{-r}(S_n - S_n') \longrightarrow 0$ a.s. Now S_n' does not have the martingale property. However, by using a device similar to that of Lemma 1, S_n' can be written as $T_n + R_n$, where T_n has the martingale property and R_n is a sum of terms having less than r arguments. By induction on r it is shown that $n^{-r} R_n \longrightarrow 0$ a.s., and Lemma 2 below (which, in essence, is known) will imply that $n^{-r} T_n \longrightarrow 0$ a.s. The proof will be given only for $r = 2$, in such a way that the extension to the case $r > 2$ can be left to the reader.

<u>Lemma 2</u>. Let $X_1, X_2, \ldots$ be a sequence of random variables, and for each positive integer n let Z_n be a real-valued function of $X_1, \ldots, X_n$ such that, for some $a \geq 1$, $E\left[|Z_n|^a \right] < \infty$ and

(15) $$E\left[Z_n \mid X_1, \ldots, X_m \right] = Z_m \quad , \quad 1 \leq m \leq n \quad .$$

If, for some $b > 0$,

(16) $$\sum_{k=1}^{\infty} 2^{-abk} E\left[\left| Z_{2^k} \right|^a \right] < \infty \quad ,$$

then $n^{-b} Z_n \longrightarrow 0$ a.s.

Proof. We have for $t > 0$

(17) $$P\left[\left| Z_m \right| \geq t \text{ for some } m \leq n \right] \leq t^{-a} E\left[\left| Z_n \right|^a \right] \quad .$$

This is a trivial extension of Doob's generalization ([1], p. 314) of Kolmogorov's inequality. By a standard argument, (16) and (17) imply $n^{-b} Z_n \longrightarrow 0$ a.s.

3. Proof of the theorem for $r = 2$. We now let

(18) $$S_n = \sum_{1 \leq i < j \leq n} f(X_i, X_j) \quad ,$$

where

(19) $$f(x,y) = f(y,x) \quad , \quad \int f(x,y)\, dF(y) \equiv 0 \quad ,$$

and define

(20) $$f^{(j)}(x,y) = \begin{cases} f(x,y) \quad , & |f(x,y)| \leq j^2 \\ 0 & \text{otherwise} \quad , \end{cases}$$

(21) $$S_n' = \sum_{1 \leq i < j \leq n} f^{(j)}(X_i, X_j) \quad .$$

Lemma 3. $n^{-2}(S_n - S_n') \longrightarrow 0$ a.s.

Proof. Let $Y_{ij} = f(X_i, X_j)$, $Y_{ij}' = f^{(j)}(X_i, X_j)$. We have for $i \neq j$

$$P[\, Y_{ij} \neq Y'_{ij} \,] = \iint_{|f(x,y)|>j^2} dF(x)dF(y)$$

$$= \sum_{\nu=j^2}^{\infty} \iint_{\nu<|f(x,y)|\leq\nu+1} dF(x)dF(y) \leq \sum_{\nu=j^2}^{\infty} \nu^{-1} a_{\nu+1} \quad ,$$

where

(22) $$a_n = \iint_{n-1<|f(x,y)|\leq n} |\, f(x,y) \,|\, dF(x)dF(y) \quad .$$

Hence

$$\sum_{1\leq i<j<\infty} P\,[Y_{ij} \neq Y'_{ij}] \leq \sum_{1\leq i<j<\infty} \sum_{\nu=j^2}^{\infty} \nu^{-1} a_{\nu+1}$$

$$= \sum_{\nu=4}^{\infty} \sum_{2\leq j\leq\nu^{1/2}} (j-1)\nu^{-1} a_{\nu+1} \leq \sum_{\nu=4}^{\infty} a_{\nu+1} < \infty \quad .$$

This implies

(23) $$\lim_{m\to\infty} P\,[Y_{ij} \neq Y'_{ij} \text{ for some } (i,j),\ m < i < j < \infty\,] = 0 \quad .$$

Now let m be a fixed integer. For $n > m$,

(24) $$n^{-2}(S_n - S'_n) = n^{-2}(S_m - S'_m) + \sum_{i=1}^{m} n^{-2} \sum_{j=m+1}^{n} (Y_{ij} - Y'_{ij})$$

$$+ n^{-2} \sum_{m<i<j\leq n} (Y_{ij} - Y'_{ij}) \quad .$$

By (23) we can choose m so large that with a probability arbitrarily close to 1 the last sum in (24) is 0 for all n. Clearly $n^{-2}(S_m - S'_m) \longrightarrow 0$ a.s. as $n \longrightarrow \infty$. It remains to show that for a fixed $i \leq m$, $n^{-2}\Sigma_{j=m+1}^{n}(Y_{ij} - Y'_{ij}) \longrightarrow 0$ a.s. It will suffice to prove this for $i = m = 1$.

Let $g^{(j)}(x,y) = f(x,y)$ if $|f(x,y)| > j^2$, $g^{(j)}(x,y) = 0$ otherwise. Then

$$(25) \qquad n^{-2} \sum_{j=2}^{n} (Y_{1j} - Y'_{1j}) = n^{-2} \sum_{j=2}^{n} g^{(j)}(X_1,X_j) = n^{-2} V_n + n^{-2} \sum_{j=2}^{n} g^{(j)}(X_1,*) \quad ,$$

where $V_n = \Sigma_{j=2}^{n} [g^{(j)}(X_1,X_j) - g^{(j)}(X_1,*)]$. Since $|g^{(j)}(x,*)| \leq \int |f(x,y)| dF(y)$, it is clear that the last term in (25) tends to 0 a.s. Also $E[V_n | X_1, \ldots, X_m] = V_m$ for $1 < m < n$. We apply Lemma 2 with $Z_n = V_n$, $a = 1$, $b = 2$. Since $E[\, |V_n| \,] \leq c\, n$, condition (16) is satisfied and hence $n^{-2} V_n \longrightarrow 0$ a.s. The proof of Lemma 3 is complete.

We now show that $n^{-2} S'_n \longrightarrow 0$ a.s. Let

$$(26) \qquad f_2^{(j)}(x,y) = f^{(j)}(x,y) - f^{(j)}(x,*) - f^{(j)}(y,*) + f^{(j)}(*,*) \quad .$$

Then

$$(27) \qquad S'_n = \sum_{1 \leq i < j \leq n} f_2^{(j)}(X_i,X_j) + \sum_{1 \leq i < j \leq n} [f^{(j)}(X_i,*) + f^{(j)}(X_j,*)]$$

$$- \sum_{j=2}^{n} (j-1) f^{(j)}(*,*) \quad .$$

Now

$$| f^{(j)}(*,*) | = \Big| \iint_{|f(x,y)| \leq j^2} f(x,y) dF(x) dF(y) \Big|$$

$$= \Big| \iint_{|f(x,y)| > j^2} f(x,y) dF(x) dF(y) \Big| \leq \sum_{\nu = j^2+1}^{\infty} a_\nu \quad .$$

Hence

$$\left| \sum_{j=2}^{n} (j-1) f^{(j)}(*,*) \right| \leq \sum_{j=2}^{n} (j-1) \sum_{\nu=j^2+1}^{\infty} a_\nu$$

$$\leq \sum_{\nu=5}^{\infty} \sum_{2 \leq j \leq \min(n, \nu^{1/2})} (j-1) a_\nu \leq \sum_{\nu=5}^{n} n a_\nu + \sum_{\nu=n+1}^{\infty} n^2 a_\nu ,$$

so that

$$n^{-2} \sum_{j=2}^{n} (j-1) f^{(j)}(*,*) \longrightarrow 0 . \tag{28}$$

Next, making use of (19),

$$f^{(j)}(x,*) = \int_{|f(x,y)| \leq j^2} f(x,y) dF(y) = - \int_{|f(x,y)| > j^2} f(x,y) dF(y) ,$$

so that

$$| f^{(j)}(x,*) | \leq \int_{|f(x,y)| > j^2} | f(x,y) | dF(y) = h_j(x) , \quad \text{say.} \tag{29}$$

We note that $h_j(x) \geq 0$, $h_j(x) \leq h_m(x)$ for $j > m$, and $\int h_m(x) dF(x) \longrightarrow 0$ as $m \longrightarrow 0$. Now for $n > m$,

$$\left| \sum_{1 \leq i < j \leq n} [f^{(j)}(X_i,*) + f^{(j)}(X_j,*)] \right| \leq \sum_{1 \leq i < j \leq m} [h_o(X_i) + h_o(X_j)]$$

$$+ \sum_{i=1}^{m} \sum_{j=m+1}^{n} [h_o(X_i) + h_m(X_j)] + \sum_{m < i < j \leq n} [h_m(X_i) + h_m(X_j)]$$

$$= (n-1) \sum_{i=1}^{m} h_o(X_i) + (n-1) \sum_{i=m+1}^{n} h_m(X_i) .$$

For m fixed and $n \longrightarrow \infty$, by Kolmogorov's strong law of large numbers (our theorem with $r = 1$), this upper bound, divided by n^2, converges a.s. to $\int h_m(x) dF(x)$, which is arbitrarily small for m large enough. Hence

(30) $$n^{-2} \sum_{1\le i<j\le n} [f^{(j)}(X_i,*) + f^{(j)}(X_j,*)] \longrightarrow 0 \quad \text{a.s.}$$

Finally, consider the first sum in (27),

(31) $$T_n = \sum_{1\le i<j\le n} f_2^{(j)}(X_i,X_j) \quad .$$

By (26), $f_2^{(j)}(x,y) = f_2^{(j)}(y,x)$ and

(32) $$\int f_2^{(j)}(x,y)dF(y) = 0 \quad .$$

Hence

(33) $$E[T_n^2] = \sum_{1\le i<j\le n} \iint [f_2^{(j)}(x,y)]^2 dF(x)dF(y)$$

$$= \sum_{j=2}^{n} (j-1) \iint [f_2^{(j)}(x,y)]^2 dF(x)dF(y) \quad .$$

It follows from (26) that

$$\iint [f_2^{(j)}(x,y)]^2 dF(x)dF(y) \le \iint [f^{(j)}(x,y)]^2 dF(x)dF(y)$$

$$= \sum_{\nu=1}^{j^2} \iint_{\nu-1<|f(x,y)|\le\nu} [f(x,y)]^2 dF(x)dF(y) \le \sum_{\nu=1}^{j^2} \nu a_\nu \quad .$$

Hence

(34) $$E[T_n^2] \le \sum_{j=2}^{n} (j-1) \sum_{\nu=1}^{j^2} \nu a_\nu \le n^2 \sum_{\nu-1}^{n^2} \nu a_\nu \quad .$$

Now, by (31) and (32), $E[T_n \mid X_1, \ldots, X_m] = T_m$ for $m < n$. We apply Lemma 2 with $Z_n = T_n$, $a = b = 2$. We have

$$\sum_{k=1}^{\infty} 2^{-4k} E[T_{2^k}^2] \le \sum_{k=1}^{\infty} 2^{-2k} \sum_{\nu=1}^{2^{2k}} \nu a_\nu = \sum_{\nu=1}^{\infty} \sum_{2^{2k}\ge\nu} 2^{-2k} \nu a_\nu < 2 \sum_{\nu=1}^{\infty} a_\nu < \infty \quad .$$

Hence

(35) $$n^{-2} T_n \longrightarrow 0 \qquad \text{a.s.}$$

It now follows from(27), (28), (30), (31) and (35) that $n^{-2} S'_n \longrightarrow 0$ a.s. By Lemma 3 this implies $n^{-2} S_n \longrightarrow 0$ a.s. The proof of the theorem for $r = 2$ is complete.

REFERENCES

[1] J. L. Doob, Stochastic Processes, John Wiley and Sons, New York, 1953.

[2] Wassily Hoeffding, "A class of statistics with asymptotically normal distribution," Ann. Math. Stat., Vol. 19 (1948), pp. 293-325.

[3] Pranab Kumar Sen, "On some convergence properties of U-statistics," Calcutta Stat. Assn. Bull., Vol. 10 (1960), pp. 1-18.