ON A THEOREM OF HSU AND ROBBINS

BY P. ERDÖS

Syracuse University

Let $f_1(x), f_2(x), \cdots$ be an infinite sequence of measurable functions defined on a measure space X with measure m, $m(X) = 1$, all having the same distribution function $G(t) = m(x; f_k(x) \leq t)$. In a recent paper Hsu and Robbins[1] prove the following theorem: *Assume that*

$$\int_{-\infty}^{\infty} t\,dG(t) = 0, \tag{1}$$

$$\int_{-\infty}^{\infty} t^2\,dG(t) < \infty. \tag{2}$$

Denote by S_n *the set* $\left(x; \left| \sum_{k=1}^{n} f_k(x) \right| > n\right)$, *and put* $M_n = m(S_n)$. *Then* $\sum_{n=1}^{\infty} M_n$ *converges.*

It is clear that the same holds if $\left| \sum_{k=1}^{n} f_k(x) \right| > n$ is replaced by $\left| \sum_{k=1}^{n} f_k(x) \right| > c \cdot n$ (replace $f_k(x)$ by $c \cdot f_k(x)$).

It was conjectured that the conditions (1) and (2) are necessary for the convergence of $\sum_{n=1}^{\infty} M_n$. Dr. Chung pointed it out to me that in this form the conjecture is inaccurate; to see this it suffices to put $f_k(x) = \frac{1}{4}(1 + r_k(x))$ where $r_k(x)$ is the kth Rademacher function. Clearly $| f_k(x) | < 1$; thus $M_n = 0$, thus $\sum_{n=1}^{\infty} M_n$ converges, but $\int_{-\infty}^{\infty} t\,dG(t) \neq 0$. On the other hand we shall show in the present note that the conjecture of Hsu and Robbins is essentially correct. In fact we prove

THEOREM I. *The necessary and sufficient condition for the convergence of* $\sum_{n=1}^{\infty} M_n$ *is that*

$$\left| \int_{-\infty}^{\infty} t\,dG(t) \right| < 1, \tag{1'}$$

and (2) *should hold.*

In proving the sufficiency of Theorem I, we can assume without loss of generality that (1) holds. It suffices to replace $f_k(x)$ by $(f_k(x) - C)$ where $C = \int_{-\infty}^{\infty} t\,dG(t)$. The following proof of the sufficiency of Theorem I (in other words essentially for the theorem of Hsu and Robbins) is simpler and quite different from theirs. Put

$$a_i = m(x; | f_k(x) | > 2^i), \tag{3}$$

[1] *Proc. Nat. Acad. Sciences*, 1947, pp. 25–31.

since the f_k's all have the same distribution, a_i clearly does not depend on k. We evidently have

$$\sum_{i=0}^{\infty} 2^{2i-1} a_i \leq \sum_{i=0}^{\infty} 2^{2i}(a_i - a_{i+1}) \leq \int_{-\infty}^{\infty} t^2\, dG(t) \leq \sum_{i=0}^{\infty} 2^{2i+2}(a_i - a_{i+1}) \leq \sum_{i=0}^{\infty} 2^{2i+2} a_i .$$

Thus (2) is equivalent to

$$\sum_{i=0}^{\infty} 2^{2i} a_i < \infty. \tag{4}$$

Let $2^i \leq n < 2^{i+1}$. Put

$S_n^{(1)} = (x; |f_k(x)| > 2^{i-2}$, for at least one $k \leq n)$,

$S_n^{(2)} = (x; |f_{k_1}(x)| > n^{4/5}, |f_{k_2}(x)| > n^{4/5}$, for at least two $k_1 \leq n, k_2 \leq n)$,

$S_n^{(3)} = \left(x; \left|\sum_{k=1}^{n}{}' f_k(x)\right| > 2^{i-2}\right)$,

where the dash indicates that the k with $|f_k(x)| > n^{4/5}$ are omitted. We evidently have

$$S_n \subset S_n^{(1)} \cup S_n^{(2)} \cup S_n^{(3)}.$$

For if x is not in $S_n^{(1)} \cup S_n^{(2)} \cup S_n^{(3)}$, then clearly

$$\left|\sum_{k=1}^{n} f_k(x)\right| \leq 2^{i-2} + 2^{i-2} < n.$$

Thus to prove the convergence of $\sum_{n=1}^{\infty} M_n$ it will suffice to show that

$$\sum_{n=1}^{\infty} (m(S_n^{(1)}) + m(S_n^{(2)}) + m(S_n^{(3)})) < \infty. \tag{5}$$

From (3) we obtain that $m(S_n^{(1)}) \leq n \cdot a_{i-2} < 2^{i+1} \cdot a_{i-2}$. Thus from (4)

$$\sum_{n=1}^{\infty} m(S_n^{(1)}) = \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} m(S_n^{(1)}) < \sum_{i=0}^{\infty} 2^{2i+3} a_i < \infty. \tag{6}$$

From (4) we evidently have that for large u

$$m(x; |f_k(x)| > u) < 1/u^2.$$

Thus since the f's are independent and have the same distribution function it follows that for sufficiently large n,

$$m(S_n^{(2)}) \leq \sum_{1 \leq k_1 < k_2 \leq n} m(x; |f_{k_1}(x)| > n^{4/5}, |f_{k_2}(x)| > n^{4/5})$$

$$\leq \binom{n}{2} m(x; |f_1(x)| > n^{4/5}), m(x; |f_2(x)| > n^{4/5}) < n^2 \cdot n^{-16/5} = n^{-6/5}.$$

Hence

$$\sum_{n=1}^{\infty} m(S_n^{(2)}) < \infty. \tag{7}$$

Put

$$f_k^+(x) = \begin{cases} f_k(x) \text{ for } |f_k(x)| < n^{4/5}; \\ 0 \text{ otherwise.} \end{cases}$$

Clearly the $f_k^+(x)$ are independent and have the same distribution function $G^+(t)$. Put

(8) $$\int_{-\infty}^{\infty} t\, dG^+(t) = \epsilon, \qquad g_k(x) = f_k^+(x) - \epsilon.$$

We have from (8) that $\int_X g_k(x)\, dm = 0$, and by (1) that $\epsilon \to 0$ as $n \to \infty$. We evidently have

$$\int_X \left(\sum_{k=1}^n g_k(x)\right)^4 dm = \int_X \sum_{k=1}^n g_k^4(x)\, dm + 6 \int_X \sum_{1 \le k < l \le n} g_k^2(x) \cdot g_l^2(x)\, dm.$$

Now since $\max |g_k(x)| < n^{4/5} + \epsilon$,

$$\int_X g_k^4(x)\, dm < (n^{4/5} + \epsilon)^2 \cdot \int_X g_k^2(x)\, dm < c_1 \cdot n^{8/5},$$

and

$$\int_X g_k^2(x) \cdot g_l^2(x)\, dm = \int_X g_k^2(x)\, dm \int_X g_l^2(x)\, dm < c_2.$$

Thus

$$\int_X \left(\sum_{k=1}^n g_k(x)\right)^4 dm < c_3\, n^{13/5}.$$

Hence

(9) $$m\left(x; \left|\sum_{k=1}^n g_k(x)\right| > n/16\right) < c_4\, n^{-(7/5)}.$$

Thus from (8), (9), $|f_k^+(x)| < |g_k(x)| + 1/16$ (for $\epsilon < 1/16$) and $n/8 < 2^{i-2}$ we have

$$m\left(x; \left|\sum_{k=1}^n f'(x)\right| > 2^{i-2}\right) = m\left(x; \left|\sum_{k=1}^n f_k^+(x)\right| > 2^{i-2}\right)$$
$$< m\left(x; \left|\sum_{k=1}^n g_k(x)\right| > n/16\right) < c_4 n^{-(7/5)},$$

or

(10) $$m(S_n^{(3)}) < c_4\, n^{-(7/5)}.$$

Thus finally from (6), (7) and (10) we obtain (5) and this completes the proof of the sufficiency of Theorem I.

Next we prove the necessity of Theorem I, in other words we shall show that if $\sum_{n=1}^{\infty} M_n$ converges then (1′) and (2) hold.

First we prove (2). The following proof was suggested by Dr. Chung, who simplified my original proof. By a simple rearrangement we see that (2) is equivalent to

$$\sum_{n=1}^{\infty} n \int_{|t|>cn} dG(t) < \infty \tag{11}$$

for any $c > 0$; while

$$\int_{-\infty}^{\infty} |t| \, dG(t) < \infty \tag{12}$$

is equivalent to

$$\sum_{n=1}^{\infty} \int_{|t|>cn} dG(t) < \infty \tag{13}$$

for any $c > 0$. Now we have clearly,

$$(x; |f_n(x)| > 2n) \subset S_{n-1} \cup S_n .$$

Hence

$$\sum_n \int_{|t|>2n} dG(t) \leqq \sum_n (m(S_{n-1}) + m(S_n)) < \infty.$$

Thus we obtain (12). Since the terms of this series is non-increasing it follows that

$$n \int_{|t|>2n} dG(t) \to 0. \tag{14}$$

Our assumption being that $\Sigma M_n < \infty$ we have $M_n \to 0$ as $n \to \infty$. It follows that there is a constant $\rho > 0$ independent of k and n such that

$$m\left(x; \left| \sum_{\substack{l=1 \\ l \neq k}}^{n} f_l(x) \right| < n \right) \geqq \rho.$$

Now, writing set intersections as products, we have

$$\bigcup_{k=1}^{n} (x; |f_k(x)| > 2n) \cdot \left(x; \left| \sum_{\substack{l=1 \\ l \neq k}}^{n} f_l(x) \right| < n \right) \subset S_n .$$

Writing this for a moment as

$$\bigcup_{k=1}^{n} (R_k T_k) \subset S_n ,$$

where $R_k = (x; |f_k(x)| > 2n)$ etc. and denoting by R' the complement of R, we have

$$\begin{aligned} M_n = m(S_n) &\geqq m\left(\bigcup_{k=1}^{n} (R_k \cdot T_k)\right) \\ &= m\left(\bigcup_{k=1}^{n} (R_1 T_1)' \cdots (R_{k-1} T_{k-1})' R_k T_k\right) \\ &= \sum_{k=1}^{n} m((R_1 T_1)' \cdots (R_{k-1} T_{k-1})' R_k T_k) \\ &\geqq \sum_{k=1}^{n} m(R_1' \cdots R_{k-1}' R_k T_k) \\ &\geqq \sum_{k=1}^{n} \{m(R_k \cdot T_k) - m((R_1 \cup \cdots \cup R_{k-1}) R_k)\} \\ &\geqq \sum_{k=1}^{n} \{m(T_k) - (k-1) m(R_1)) m(R_k)\} \\ &\geqq \sum_{k=1}^{n} \{\rho - nm(R_1)\} m(R_k) \geqq \sum_{k=1}^{m} (\rho - \sigma(1)) m(R_k). \\ &\geqq \rho' \sum_{k=1}^{n} m(R_k) = n\rho' \int_{|t|>2n} dG(t) \end{aligned}$$

by (14) since $m(R_1) = \int_{|t|>2n} dG(t)$, $nm(R_1) \to 0$ as $n \to \infty$.

Thus

$$\sum_n n \int_{|t|>2n} dG(t) \leqq \frac{1}{\rho'} \sum_n M_n < \infty.$$

Hence we have (11), which is equivalent to (2). The proof of (1′) is quite easy. By virtue of (2) we can put

$$\int_{-\infty}^{\infty} tG(t) = C.$$

If $C > 1$, then it follows from (2) and Tschebycheff inequality that $M_n \to 1$ as $n \to \infty$, thus $C \leqq 1$. But if $C = 1$, we conclude from (2) and the central limit theorem that M_n does not tend to 0. Hence $C < 1$, and (1′) is proved.

By similar methods we can prove the following results: Let $2 < c < 4$. Put

$$M_n^{(c)} = m\left(x; \left|\sum_{k=1}^{n} f_k(x)\right| > n^{2/c}\right).$$

Then the necessary and sufficient condition for the convergence of $\sum_{k=1}^{\infty} M_n^{(c)}$

is that

$$\int_{-\infty}^{\infty} t\,dG(t) = 0, \qquad \int_{-\infty}^{\infty} | t |^{c}\,dG(t) < \infty.$$

If $c < 2$ then the necessary and sufficient condition for the convergence of $\sum_{n=1}^{\infty} M_n^{(c)}$ is that $\int_{-\infty}^{\infty} | t |^{c}\,dG(t) < \infty$.

Finally we can prove the following result: Assume that $\int_{-\infty}^{\infty} t\,dG(t) = 0$ and $\int_{-\infty}^{\infty} t^4\,dG(t) < \infty$. Then there exists a constant r so that

$$\sum_{n=1}^{\infty} m\left[x; \left| \sum_{k=1}^{n} f_k(x) \right| > n^{1/2} \cdot (\log n)^r \right] < \infty. \tag{17}$$

The case of the Rademacher functions shows that (17) can not be improved very much, in fact only the value of r could be improved.

pp. 85–88) which appeared in *Mathematical Reviews*, Vol. 9 (1948), p. 470, I stated that "a more simple and elegant, and equally general, expression is obtainable by a simple adaptation of formula (41), p. 215, of J. F. Steffensen's book, *Interpolation*."

This statement is not entirely correct and is also misleading in its implications since Dr. Kincaid's expressions are actually more general in certain respects, and simplicity and generality are not the only considerations nor, in this case, the most important ones. In setting up an expression for the remainder in an interpolation formula, the primary objective is to secure an efficient appraisal of the remainder. In this respect, Dr. Kincaid's expressions are superior as they involve only the higher derivatives of the function it is desired to represent, whereas Steffensen's method would always involve a first derivative term in such a way as to prevent any refinement of estimates of the error by introducing additional given values.

REMARK ON MY PAPER "ON A THEOREM OF HSU AND ROBBINS"

By P. Erdös

Syracuse University

Professor Robbins kindly pointed out that in my paper mentioned in the title (*Annals of Math. Stat.*, Vol. 20 (1949), p. 286–291) I have misquoted a statement in the paper of Hsu and Robbins ("Complete Convergence and the Law of Large Numbers" *Proc. Nat. Acad. of Sci.*, Vol. 33 (1947), p. 25–31). I attribute to Hsu and Robbins the conjecture (notations of my paper) that if $\sum_{n=1}^{\infty} M_n < \infty$ then (1) and (2) hold, and proceed to give a counter example. However, the conjecture of Hsu and Robbins is not the above false one but the following: If $\sum_{n=1}^{\infty} M_n < \infty$ and (1) holds then (2) also holds. This conjecture is true and is in fact proved in my paper.

Professor Robbins also points out that a slight modification of my theorem can be stated in a more concise form as follows: Let $X_1, X_2, \cdots$ be a sequence of independent random variables having the same distribution function $F(x)$, and let

$$Y_n = (1/n)\ (X_1 + \cdots + X_n)$$

Then the necessary and sufficient condition that

$$\sum_{n=1}^{\infty} P_r\{|Y_n| > \epsilon\} < \infty, \qquad \text{for every } \epsilon > 0,$$

is that

$$\int_{-\infty}^{\infty} x\, dF(x) = 0, \qquad \int_{-\infty}^{\infty} x^2\, dF(x) < \infty.$$