hence, using the fact that $B > B_4$,

$$\frac{|R(\xi, z)|}{\log^{\varkappa+1}\xi} \leqslant \frac{BL}{\log z}\left\{(\nu-1)^{\varkappa+1}\left(1+\frac{\varkappa}{\nu}\right)+\frac{1}{\nu}\right\} < \frac{BL}{\log z}(\nu-1)^{\varkappa+1}\left(1+\frac{\varkappa+1}{\nu}\right),$$

and since $(\nu-1)^{\varkappa+1}\left(1+\dfrac{\varkappa+1}{\nu}\right) \leqslant \nu^{\varkappa+1}$ (as may easily be verified), we obtain

$$\frac{|R(\xi, z)|}{\log^{\varkappa+1}\xi} \leqslant \frac{BL}{\log z}\nu^{\varkappa+1} \quad \text{if} \quad z^{\nu} < \xi \leqslant z^{\nu+1},$$

and thereby confirm the truth of (6.11) with $\nu+1$ in place of ν.

This completes the proof of (6.11) and hence also of (6.9).

To complete the proof of Theorem 2, we substitute (6.9) in (6.7), and use this composite relation, with $\xi = x$ and $\xi = x/z$ in turn, to evaluate $G(x, z)$ from (6.5) (with $t = x$): we obtain

$$G(x, z) = (\varkappa+1)C_0\frac{\bar{\sigma}_{\varkappa}(2\tau)}{2\tau}\log^{\varkappa}z - \varkappa C_0\frac{\bar{\sigma}_{\varkappa}(2\tau-2)}{2\tau}\log^{\varkappa}z + O(L\tau^{2\varkappa+1}\log^{\varkappa-1}z)$$

$$= C_0\sigma_{\varkappa}(2\tau)\log^{\varkappa}z + O(L\tau^{2\varkappa+1}\log^{\varkappa-1}z)$$

by (5.5). Theorem 2 follows at once from this and (2.12).

References

[1] N. C. Ankeny and H. Onishi, *The general sieve*, Acta Arith. 10 (1964), pp. 31–62.

[2] H. Halberstam and K. F. Roth, *Sequences*, Oxford 1966.

[3] Б. В. Левин и А. С. Фейнлейб, *Применение некоторых интегральных уравнений к вопросам теории чисел*, УМН 22 (1967), pp. 119–197.

[4] E. Wirsing, *Über die Zahlen, deren Primteiler einer gegebenen Menge angehören*, Arch. der Math. 7 (1956), pp. 263–272.

Received on 13. 3. 70

ACTA ARITHMETICA
XVIII (1971)

A theorem on chains of finite sets, II

by

R. Rado (Reading)

Dedicated to the memory of Harold Davenport

1. Introduction. E. Harzheim [1] proved the following theorem:

Theorem A. *Given a positive integer n, there is a positive integer n^* such that the following statement holds. If S is a set of n^* elements, and if $f(X)$, for every non-empty subset X of S, is an element of X, then there always are subsets $X_0, X_1, \ldots, X_n$ of S such that*(1) $X_0 \subset X_1 \subset \ldots \subset X_n$ *and*

$$f(X_0) = f(X_1) = \ldots = f(X_n).$$

The following theorem is a generalization of Theorem A ([4], Theorem 3):

Theorem B. *Given a positive integer n, there is a positive integer n^* such that the following statement holds. If S is a set of n^* elements, and if $f(X)$, for every subset X of S, is a subset of X, then there always are subsets $X_0, \ldots, X_n$ of S such that $X_0 \subset \ldots \subset X_n$ and $f(X_0) \subseteq \ldots \subseteq f(X_n)$.*

In the present note Theorem B will be further generalized. No knowledge of the earlier papers [1], [4] will be assumed. In fact, the proof of the still more general Theorem C given below is simpler than that of Theorem B as given in [4], thanks to an application of an idea used by D. J. White [6] which makes it unnecessary to appeal to a theorem of G. Higman [3] which was needed in [4].

2. Notation and terminology. We put $N = \{0, 1, 2, \ldots\}$. Lower case letters other than $f, g, h, \varphi, \psi, \chi, \pi$ denote elements of N, and capital letters denote subsets of N. If nothing is said to the contrary these sets are finite. The cardinal of A is denoted by $|A|$, and for every S, finite or infinite, we put

$$[S]^r = \{X: X \subseteq S;\ |X| = r\}.$$

Also, $[0, m) = \{0, 1, \ldots, m-1\}$.

(1) $A \subset B$ denotes set inclusion in the strict sense.

A *kernel function* is a function f such that $X \supseteq f(X)$ for every X. The letters $f, g, h, \varphi, \psi, \chi$ denote kernel functions; the functional symbols are always placed to the right of the argument. The function φ is *divergent* if $|X\varphi| \to \infty$ as $|X| \to \infty$, i.e. if given a, there is b such that $|X| \geqslant b$ implies $|X\varphi| \geqslant a$. Thus the identity function φ_0, for which $X\varphi_0 = X$, is divergent.

The extension of Theorem B will take place in two directions. We shall consider several kernel functions simultaneously, and we shall impose lower bounds on the rate of growth along the sequence of sets X_ν. In [2] Harzheim has extended [1] to infinite increasing chains of sets. It is convenient to state and prove our result in terms of decreasing rather than increasing sequences.

3. The main theorem.

THEOREM C. *Let $k, n \in N$, and let φ be a divergent kernel function. Then there is n^* such that the following statement holds. If $|S| \geqslant n^*$ and if $f_0, \ldots, f_k$ are any kernel functions then there are sets $A_0, \ldots, A_n$ such that either*

(i) $S \supset A_0 \supseteq A_0 f_\varkappa \supseteq A_0 f_\varkappa \varphi \supset A_1 \supseteq A_1 f_\varkappa \supseteq A_1 f_\varkappa \varphi \supset \ldots \supset A_n$ *for some* $\varkappa \leqslant k$,

or

(ii) $\quad S \supset A_0 \supset \ldots \supset A_n \supseteq A_0 f_\varkappa = \ldots = A_n f_\varkappa \quad$ *for every* $\varkappa \leqslant k$.

Theorems A and B are weaker than the simplest case $k = 0$; $X\varphi = X$ of Theorem C. If $n \geqslant 1$ then there are cases when only (i) is possible (all $Xf_\varkappa = X$) and cases when only (ii) is possible (all $Xf_\varkappa = \emptyset$).

4. Lemmas.

LEMMA 1 (Ramsey's Theorem). *If S is infinite and $[S]^r = K_0 \cup \ldots \cup K_m$ then there is an infinite $S' \subseteq S$ such that $[S']^r \subseteq K_\mu$ for some $\mu < m$.*

See [5].

LEMMA 2. *If $f_0, f_1, \ldots$ is an infinite sequence of functions then there are a function g and an infinite sequence $\nu_0 < \nu_1 < \nu_2 < \ldots$ such that, for every X, $Xf_{\nu_\lambda} = Xg$ whenever λ is sufficiently large.*

This is a well-known compactness proposition.

LEMMA 3. *If $x_0, x_1, \ldots \in N$ then there is a sequence $\nu_0 < \nu_1 < \ldots$ such that $x_{\nu_0} \leqslant x_{\nu_1} \leqslant \ldots$*

Proof. There is a least ν_0 such that $x_{\nu_0} = \min\{x_\nu\colon \nu \geqslant 0\}$, and then there is a least $\nu_1 > \nu_0$ such that $x_{\nu_1} = \min\{x_\nu\colon \nu > \nu_0\}$, and a least $\nu_2 > \nu_1$ such that $x_{\nu_2} = \min\{x_\nu\colon \nu > \nu_1\}$, and so on. Then the assertion holds(2).

(2) The trivial Lemma 3 replaces Higman's theorem in [3] which was needed in [4].

5. Proof of Theorem C. Denote by

$$\mathscr{C}(n, S, \varphi, f_0, \ldots, f_k)$$

the statement that there are $A_0, \ldots, A_n$ such that either (i) or (ii) of Theorem C is true. The precise meaning of $\mathscr{C}$ will not be required for a large part of the proof.

Let k, n, φ be fixed, and let φ be divergent. We assume that given any a, there are $S, f_0, \ldots, f_k$ such that $|S| \geqslant a$ and $\mathscr{C}(n, S, \varphi, f_0, \ldots, f_k)$ is false. We have to deduce a contradiction.

Let $m \in N$. Then there are $S_m, f_{m0}, \ldots, f_{mk}$ with $|S_m| = m$ and such that $\mathscr{C}(n, S_m, \varphi, f_{m0}, \ldots, f_{mk})$ is false. By applying a suitable permutation π to N, which takes S into $[0, m)$, and by transferring the functions φ and $f_{m\varkappa}$ accordingly we find functions φ_m and $g_{m\varkappa}$ such that $X\varphi = X\pi\varphi_m$ and $Xf_{m\varkappa} = X\pi g_{m\varkappa}$ and the statement

$$\mathscr{C}(n, [0, m), \varphi_m, g_{m0}, \ldots, g_{mk})$$

is false. Then, for every X, there is Y such that $|X| = |Y|$ and $|X\varphi_m| = |Y\varphi|$. Hence φ_m is divergent. By Lemma 2 we find $\psi, g_0, \ldots, g_k$ such that the statement $\mathscr{C}(n, N, \psi, g_0, \ldots, g_k)$ is false. Then ψ is divergent. For let $a \in N$. Then there is b such that $|X| \geqslant b$ implies $|X\varphi| \geqslant a$. Now let $|X| \geqslant b$. Then $X\psi = X\varphi_m$ for some m. There is Y such that $|X| = |Y|$ and $|X\varphi_m| = |Y\varphi|$. Then $|Y| = |X| \geqslant b$ and hence $|X\varphi| = |X\varphi_m| = |Y\varphi| \geqslant a$ which shows that ψ is divergent.

Case 1. Given a there is b such that whenever $|B| \geqslant b$ then there are A and $\varkappa$ such that $A \subset B$; $\varkappa \leqslant k$; $|Ag_\varkappa| \geqslant a$. Let $r = (k+1)(n+1)+1$. Then, by repeated application of this condition, we find sets $S, X_0, \ldots, X_r$ and numbers $\varkappa_0, \ldots, \varkappa_r \leqslant k$ such that

$$S \supset X_0 \supseteq X_0 g_{\varkappa_0} \supseteq X_0 g_{\varkappa_0}\psi \supset X_1 \supseteq X_1 g_{\varkappa_1} \supseteq X_1 g_{\varkappa_1}\psi \supset \ldots \supseteq X_r g_{\varkappa_r}\psi.$$

By the pigeon hole principle there is $\varkappa \leqslant k$ and there are numbers $a_0 < a_1 < \ldots < a_n \leqslant r$ such that $\varkappa_{a_0} = \ldots = \varkappa_{a_n} = \varkappa$, say. Then

$$X_{a_0} \supseteq X_{a_0} g_\varkappa \supseteq X_{a_0} g_\varkappa \psi \supset \ldots \supset X_{a_n} \supseteq X_{a_n} g_\varkappa \supseteq X_{a_n} g_\varkappa \psi,$$

so that $\mathscr{C}(n, N, \psi, g_0, \ldots, g_k)$ is true, which is a contradiction.

Case 2. There is a_0 such that, for every b, there is a set B satisfying $|B| \geqslant b$ and $|Ag_\varkappa| < a_0$ for all $A \subset B$ and all $\varkappa \leqslant k$. Let $m \in N$. Then there is B_m such that $|B_m| = m$ and $|Ag_\varkappa| < a_0$ for all $A \subset B_m$ and all $\varkappa \leqslant k$. Transfer, just as near the beginning of the proof, B_m to $[0, m)$ and change ψ and $g_\varkappa$ accordingly. We find $\psi_m, h_{m0}, \ldots, h_{mk}$ such that $|Ah_{m\varkappa}| < a_0$ for all $A \subset [0, m)$ and all $\varkappa \leqslant k$, and $\mathscr{C}(n, N, \psi_m, h_{m0}, \ldots, h_{mk})$ is false.

By Lemma 2 we find $\chi, h_0, \ldots, h_k$ such that

(1) $$|Xh_\varkappa| < a_0 \quad \text{for all } X \text{ and all } \varkappa \leqslant k,$$

and $\mathscr{C}(n, N, \chi, h_0, \ldots, h_k)$ is false[3].

We introduce the equivalence relation

$$(X_0, \ldots, X_m) \sim (Y_0, \ldots, Y_m)$$

to denote the fact that there is an order preserving bijection

$$X_0 \cup \ldots \cup X_m \to Y_0 \cup \ldots \cup Y_m$$

such that $X_\mu \to Y_\mu$ for all $\mu \leqslant m$. For $A \supseteq B$ we define a vector $v(A, B)$ which describes the position of B within A. If $A = \{a_1, \ldots, a_s\}_<$ and $B = \{a_{p_1}, \ldots, a_{p_t}\}_<$ then we put

$$v(A, B) = (p_1 - 1, p_2 - p_1 - 1, p_3 - p_2 - 1, \ldots, p_t - p_{t-1} - 1, s - p_t).$$

Here the right-hand side is interpreted in the obvious way when $s = 0$ or $t = 0$. The components of the vector $v(A, B)$ are, in a sense, the sizes of the $t+1$ connected components of $A \setminus B$ in A. By Lemma 1 there are infinite sets $N_0, N_1, \ldots$ such that $N = N_0 \supseteq N_1 \supseteq N_2 \supseteq \ldots$ and, for $r \in N$ and $A, B \in [N_r]^r$, we have

(2) $$(A, Ah_0, \ldots, Ah_k) \sim (B, Bh_0, \ldots, Bh_k).$$

By (1) there is an infinite R such that, if $r, s \in R$ and $A \in [N_r]^r$ and $B \in [N_s]^s$, then

(3) $$(Ah_0, \ldots, Ah_k) \sim (Bh_0, \ldots, Bh_k).$$

This follows from the fact that by (2) the equivalence class of $(Ah_0, \ldots, Ah_k)$ depends on r only and not on A when A ranges over $[N_r]^r$. Put $Ah = Ah_0 \cup \ldots \cup Ah_k$. Then the number $t = |Ah|$ is independent of r and A when $r \in R$ and $A \in [N_r]^r$, and we have $v(A, Ah) = (x_{r0}, \ldots, x_{rt})$ where, by (2), the $x_{r\tau}$ are independent of A. By Lemma 3 there is an infinite set $R' \subseteq R$ such that, whenever $\{r, s\}_< \subset R'$, then $x_{r\tau} \leqslant x_{s\tau}$ for all $\tau \leqslant t$.

Now we are ready for the final step of the argument. Let $\{r, s\}_< \subset R'$, and choose $B \in [N_s]^s$. Then $v(B, Bh) = (x_{s0}, \ldots, x_{st})$. Since $x_{s\tau} \geqslant x_{r\tau}$ for all $\tau \leqslant t$, there is $A \subseteq B$ such that

$$v(A, Bh) = (x_{r0}, \ldots, x_{rt}).$$

Let $A' \in [N_r]^r$. Then $v(A', A'h) = (x_{r0}, \ldots, x_{rt})$. Hence $|A| - |Bh| = x_{r0} + \ldots + x_{rt} = |A'| - |A'h|$. By (3) we have

$$(Bh_0, \ldots, Bh_k) \sim (A'h_0, \ldots, A'h_k).$$

Therefore $|Bh| = |A'h|$ and so $|A| = |A'| = r$.

[3] In fact, χ is divergent but we do not need this fact.

We have $v(A, Ah) = (x_{r0}, \ldots, x_{rt})$. Hence $v(A, Bh) = v(A, Ah)$ and therefore $Bh = Ah$. We also have, by (3),

$$(Bh_0, \ldots, Bh_k) \sim (Ah_0, \ldots, Ah_k).$$

This implies that $Bh_\varkappa = Ah_\varkappa$ for all $\varkappa \leqslant k$.

To sum up: if $s \in R'$ and $B \in [N_s]^s$ then, for every $r \in R'$ with $r < s$ there is $A \in [B]^r \subset [N_s]^r \subseteq [N_r]^r$ such that $Bh_\varkappa = Ah_\varkappa$ for all $\varkappa \leqslant k$ (Proposition (*)).

We now choose $\{r_0, \ldots, r_n\}_> \subset R'$. Let $A_0 \in [N_{r_0}]^{r_0}$. Then, by repeated application of (*), we find $A_0 \supset \ldots \supset A_n$ such that $A_0 h_\varkappa = \ldots = A_n h_\varkappa$ for all $\varkappa \leqslant k$, and thus the statement $\mathscr{C}(n, N, \chi, h_0, \ldots, h_k)$ is true which is the desired contradiction.

6. In conclusion I mention the following two problems.

(α) To decide whether (ii) of Theorem C can be sharpened by inserting the divergent function φ in the same way as it is inserted in (i).

(β) The proof of Theorem C is highly non-constructive. Harzheim proved his theorem in [1] by a constructive argument which actually yields the best possible value $n^* = 2^n$. One should try to modify the proof of Theorem C in such a way that an upper estimate for n^* is obtained in terms of k and n.

References

[1] E. Harzheim, *Ein kombinatorisches Problem über Auswahlfunktionen*, Publicationes Math. Debrecen 15 (1968), pp. 19–22.

[2] — *Kombinatorische Betrachtungen über die Struktur der Potenzmenge*, Math. Nachrichten 34 (1967), pp. 123–141.

[3] G. Higman, *Ordering by divisibility in abstract algebra*, Proceed. London Math. Soc. (3) 2 (1952), pp. 326–336.

[4] R. Rado, *A theorem on chains of finite sets*, Journ. London Math. Soc. 42 (1967), pp. 101–106.

[5] F. P. Ramsey, *On a problem in formal logic*, Proceed. London Math. Soc. (2) 30 (1930), pp. 246–286.

[6] D. J. White, *On finite nests of intervals of integers*, Journ. London Math. Soc. 42 (1967), pp. 501–503.

THE UNIVERSITY
Reading, England

Received on 13. 3. 1970

ACTA ARITHMETICA
XVIII (1971)

The average of the least primitive root modulo p^2

by

D. A. BURGESS (Nottingham)

1. In 1968 Dr. Elliott and I [3] obtained the estimate

$$\pi(X)^{-1}\sum_{p\leqslant X} g(p) \ll (\log X)^2(\log\log X)^4 \tag{1}$$

for the average over all primes $p\leqslant X$ of the least primitive root $g(p)$ to the modulus p. Professor Heilbronn proposed to me the problem of the similar estimation of the least primitive root $h(p)$ to the modulus p^2. The argument of [3] remains applicable with slight modifications but yields only the weaker estimate

$$\pi(X)^{-1}\sum_{p\leqslant X} h(p) \ll (\log X)^4(\log\log X)^8. \tag{2}$$

The argument of [3] was based on the Large Sieve inequality which may be stated as

$$\sum_{m\leqslant X}\ \sum_{\substack{a=1\\(a,m)=1}}^{m}\Big|\sum_{n=1}^{N} e(an/q)\,a_n\Big|^2 \ll (X^2+N)\sum_{n=1}^{N}|a_n|^2 \tag{3}$$

where as usual $e(x) = e^{2\pi i x}$. In the estimation of $g(p)$ m in (3) ranged over the primes. In the estimation of $h(p)$ however m ranges over the $p^2\leqslant X$ (together with the $p\leqslant X^{1/2}$) and it is this decrease in the size of the set of m that gives rise to the loss in effectiveness seen on comparing (2) with (1). The purpose of this paper is to regain in part this effectiveness by producing a modified form of the Large Sieve which will reflect such restrictions on the set of sieving moduli m. The resultant estimation for the average of $h(p)$ is contained in the following theorem:

THEOREM. *For large X*

$$\pi(X)^{-1}\sum_{p\leqslant X} h(p) \ll (\log X)^3(\log\log X)^6\ .$$

the summation being extended over prime numbers p.