ACTA ARITHMETICA
XLIX (1988)

Prime-like sequences

by

HUGH L. MONTGOMERY* (Ann Arbor, Mich.)

Dedicated to Pál Erdős on the occasion of his 75th birthday

In this note we answer in the affirmative a question raised by Erdős.

THEOREM. *Let $\mathscr{A}$ be a set of positive integers,*

$$A(x) = \operatorname{card}\{a \in \mathscr{A}:\ a \leqslant x\},$$

and suppose that

$$A(x) \sim x/\log x \quad \text{as} \quad x \to \infty. \tag{1}$$

Put

$$f(n) = \sum_{\substack{a<n\\ a\in\mathscr{A}}} 1/(n-a).$$

Then the number 1 is a limit point of the sequence $\{f(n)\}$.

Under the stated hypotheses it is easy to see that $f(n)$ has mean value 1, since

$$\sum_{n=1}^{N} f(n) = \sum_{\substack{a<N\\ a\in\mathscr{A}}} \sum_{m=1}^{N-a} 1/m = \sum_{\substack{a<N\\ a\in\mathscr{A}}} \log(N-a) + O\big(A(N)\big) \sim N. \tag{2}$$

Proof. We argue by contradiction. Suppose that there is a $\delta > 0$ such that

$$|f(n)-1| \geqslant \delta \quad \text{for all large } n. \tag{3}$$

First we show that there exists an arbitrarily large n_0 such that

$$f(n_0) \leqslant 1-\delta \tag{4}$$

and

$$f(n) \geqslant 1+\delta \quad \text{for} \quad n_0 - n_0^{\delta/2} \leqslant n < n_0. \tag{5}$$

* Research supported in part by National Science Foundation Grant NSF DMS 85-02804.

To see this, let $f(n, u) = \sum_{\substack{a<n-u\\ a\in\mathscr{A}}} 1/(n-a)$. Then

$$\sum_{n=1}^{N} f(n, n^{\delta/2}) = \sum_{\substack{a<N-N^{\delta/2}\\ a\in\mathscr{A}}} \sum_{\substack{n\leqslant N\\ n-n^{\delta/2}>a}} 1/(n-a).$$

The root of the equation $x-x^{\delta/2} = a$ lies in the interval $(a+a^{\delta/2},\ a+a^{\delta/2}+1)$. Hence by the integral test the inner sum above is

$$\int_{a^{\delta/2}}^{N-a} (1/u)\,du + O(a^{-\delta/2}) = \log(N-a) - \tfrac{1}{2}\delta\log a + O(a^{-\delta/2}).$$

Then by partial summation and (1) we deduce that

$$\sum_{n=1}^{N} f(n, n^{\delta/2}) \sim (1-\delta/2)N \quad \text{as} \quad N\to\infty.$$

Thus there exist arbitrarily large values of n for which $f(n, n^{\delta/2}) > 1-\delta$, say $f(n_1, n_1^{\delta/2}) > 1-\delta$. If $n_1 - n_1^{\delta/2} \leqslant n \leqslant n_1$ then

$$f(n) \geqslant \sum_{\substack{a<n_1-n_1^{\delta/2}\\ a\in\mathscr{A}}} 1/(n-a) \geqslant f(n_1, n_1^{\delta/2}) > 1-\delta.$$

Hence by (3) we see that $f(n)\geqslant 1+\delta$ when $n_1-n_1^{\delta/2}\leqslant n\leqslant n_1$. Now let n_0 be the least $n>n_1$ such that $f(n_0)<1$. Such an n_0 must exist, in view of (2) and (3). Then $f(n_0)\leqslant 1-\delta$, and $f(n)\geqslant 1+\delta$ for $n_1-n_1^{\delta/2}\leqslant n<n_0$. Hence we have (4) and (5).

We now show that if $f(n_0)\leqslant 1-\delta$, if $f(n)\geqslant 1+\delta$ for $n_0-2x\leqslant n\leqslant n_0-x$, and if $x\geqslant 4/\delta$, then

(6) $$A(n_0-x) - A(n_0-4x/\delta) \geqslant \delta x/(4\log x).$$

To derive this, we first note that if $n\leqslant n_0$, $u>0$, and $a<n-u$ then

$$\frac{1}{n-a} \leqslant \frac{n_0-n+u}{u}\cdot\frac{1}{n_0-a}.$$

On summing this over $a<n-u$, $a\in\mathscr{A}$, we deduce that

$$f(n, u) \leqslant \frac{n_0-n+u}{u} f(n_0, n_0-n+u).$$

But $f(n_0, v)\leqslant f(n_0)$ for any $v\geqslant 0$, so by (4) the above is

$$\leqslant \frac{n_0-n+u}{u}(1-\delta).$$

If we take $u=(n_0-n)/\delta$ then this is

$$(1+\delta)(1-\delta) = 1-\delta^2 < 1,$$

so that if $f(n)\geqslant 1+\delta$ then

$$\sum_{\substack{n-(n_0-n)/\delta\leqslant a<n\\ a\in\mathscr{A}}} 1/(n-a) \geqslant \delta.$$

We sum this over all $n\in(n_0-2x,\ n_0-x)$ to see that

$$(x-1)\delta \leqslant \sum_{n_0-2x\leqslant n<n_0-x}\ \sum_{\substack{n-(n_0-n)/\delta\leqslant a<n\\ a\in\mathscr{A}}} \frac{1}{n-a} = \sum_{a\in\mathscr{A}}\sum_{n}\frac{1}{n-a}$$

where the outer sum, over a, is subject to the constraint $n_0-2(1+1/\delta)x\leqslant a<n_0-x$, and the inner sum, over n, is subject to the two constraints $a<n\leqslant(\delta a+n_0)/(1+\delta)$, $n_0-2x\leqslant n<n_0-x$. If we drop the latter of these two constraints then the inner sum is made larger (or at least not decreased). By appeal to the inequality $\sum_{k\leqslant v} 1/k\leqslant 1+\log v$, which holds for all $v\geqslant 1$, we conclude that the inner sum above is

$$\leqslant 1+\log\frac{n_0-a}{1+\delta} \leqslant 1+\log(2x/\delta).$$

Thus the double sum is

$$\leqslant \big(A(n_0-x) - A(n_0-4x/\delta)\big)\log(2ex/\delta),$$

which gives (6).

We now use (5) and (6) to derive a lower bound for $f(n_0)$. Let $K = [(\delta\log n_0)/(4\log(4/\delta))]$. The intervals $I_k = (n_0-(4/\delta)^{k+1},\ n_0-(4/\delta)^k]$, $1\leqslant k\leqslant K$, are disjoint and lie in the range $[n_0-n_0^{\delta/2},\ n_0)$. Also, if $a\in I_k$ then $1/(n_0-a)\geqslant(\delta/4)^{k+1}$. Hence

$$f(n_0) \geqslant \sum_{k=1}^{K}\sum_{a\in I_k\cap\mathscr{A}} 1/(n_0-a)$$
$$\geqslant \sum_{k=1}^{K} (\delta/4)^{k+1}\big(A(n_0-(4/\delta)^k) - A(n_0-(4/\delta)^{k+1})\big).$$

From (5) and (6) we see that this is

$$\geqslant \frac{\delta^2}{16\log(4/\delta)}\sum_{k=1}^{K}\frac{1}{k} \geqslant \frac{\delta^2}{16\log(4/\delta)}\log\left(\frac{\delta\log n_0}{4\log(4/\delta)}\right).$$

That is, $f(n_0)\gg_\delta \log\log n_0$ for large n_0. This contradicts (4), so the proof is complete.

From the hypothesis (1) alone it is not possible to derive a quantitative upper bound for the rate at which the limit point 1 is approached. For as was observed by Erdős, if $a_n = [(1-\varepsilon_n)n\log n]$ and $\varepsilon_n\to 0$ very slowly, then the limit point 1 is also approached very slowly. Also, from (1) it does not follow that

(7) $$\sum_{n\le x} f(n)^2 \sim x \quad \text{as} \quad x\to\infty.$$

On the other hand, it is not hard to show that if there is an $h = h(x)$ such that $\log h = o((\log x)^{1/2})$ and

(8) $$\int_2^x \left(A(u+h)-A(u)-\frac{h}{\log u}\right)^2 du = o(h^2 x(\log x)^{-2}),$$

then both (1) and (7) hold. From (1) and (7) we see that

(9) $$\sum_{n\le x} (f(n)-1)^2 = o(x),$$

from which it follows that $f(n)$ is near 1 for almost all n.

If we take $\mathscr{A}$ to be the set of prime numbers then we have (1), since this is the prime number theorem. If the Riemann Hypothesis is assumed, then (8) holds for prime numbers with $h = \exp((\log x)^{1/3})$, for example. (See [1].)

I am happy to thank Professor Pál Erdős for his comments, and also Professor Carl Pomerance, who pointed out an error in and a simplification of my original argument.

Reference

[1] B. Saffari and R. C. Vaughan, *On the fractional parts of x/n and related sequences II*, Ann. Inst. Fourier (Grenoble) 27 (2) (1977), pp. 1–30.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
Ann Arbor, MI 48109

Received on 18.4.1986 (1623)

ACTA ARITHMETICA
XLIX (1988)

On an additive property of squares and primes

by

IMRE Z. RUZSA (Budapest)

To Uncle Paul, with epsilonial love

1. Introduction. The additive property in the title is that of being an essential component. Essential components are traditionally defined via the Schnirelmann density. The *Schnirelmann density* $\sigma(A)$ of a set A of integers is defined by the formula

$$\sigma(A) = \inf A(n)/n,$$

where n runs over the natural numbers and we use $A(n)$ to denote the *counting function* of our set A, that is, the number of its elements between 1 and n (the nonpositive elements are not taken into account).

This concept of density was introduced by and named after L. G. Schnirelmann [10], who proved the inequality

(1.1) $$\sigma(A+B) \geqslant \sigma(A)+\sigma(B)-\sigma(A)\sigma(B)$$

and used it to show that every set of positive density is a basis, and that the set P of primes is an asymptotic basis (that is, the sumset $P+\ldots+P$ with a sufficiently large number of summands contains all large integers), which was the first unconditional result concerning the Goldbach conjecture.

A set H is called a (Schnirelmann) *essential component* if $\sigma(A+H) > \sigma(A)$ whenever $0<\sigma(A)<1$. By (1.1), sets of positive density always have this property. The first essential component of density 0 was discovered by Khintchine [4]; it was the set Q of squares. A few years later Erdős [1] found that every basis is an essential component; he proved this in the effective form

(1.2) $$\sigma(A+H) \geqslant \sigma(A)+\sigma(A)(1-\sigma(A))/(2h),$$

if H is a basis of order h. A much stronger version of (1.2) was found by Plünnecke [8]; he proved

(1.3) $$\sigma(A+H) \geqslant \sigma(A)^{1-1/h},$$

which is, in this generality, the best possible order of magnitude.