On local distribution of elements of sets of positive integers

Sergei Konyagin

April 4, 2025, Moscow

Let $\mathcal{A} \subset \mathbb{N}$ be an infinite set, $\mathcal{A} = \{a_1 < a_2 < \dots\}$. We say that $\mathcal{A}$ is syndetic if $\sup_k (a_{k+1} - a_k) < \infty$. Also, we consider that a finite subset of $\mathbb{N}$ is not syndetic.

Let $\mathcal{A} \subset \mathbb{N}$ be an infinite set, $\mathcal{A} = \{a_1 < a_2 < \dots\}$. We say that $\mathcal{A}$ is syndetic if $\sup_k(a_{k+1} - a_k) < \infty$. Also, we consider that a finite subset of $\mathbb{N}$ is not syndetic.

For example, the set $\mathcal{A}$ of composite numbers is syndetic since for any k one of the numbers $a_k + 1$, $a_k + 2$ is even, and it is an element of $\mathcal{A}$. Hence, $\sup_k(a_{k+1} - a_k) \le 2$.

Let $\mathcal{A} \subset \mathbb{N}$ be an infinite set, $\mathcal{A} = \{a_1 < a_2 < \dots\}$. We say that $\mathcal{A}$ is syndetic if $\sup_k(a_{k+1} - a_k) < \infty$. Also, we consider that a finite subset of $\mathbb{N}$ is not syndetic.

For example, the set $\mathcal{A}$ of composite numbers is syndetic since for any k one of the numbers $a_k + 1$, $a_k + 2$ is even, and it is an element of $\mathcal{A}$. Hence, $\sup_k(a_{k+1} - a_k) \leq 2$.

Similarly, let $\mathcal{A}$ be the complement to the set of square free numbers in $\mathbb{N}$. Then $\mathcal{A}$ contains all positive integers divisible by 4, and for any k at least one of the numbers $a_k + 1$, $a_k + 2$, $a_k + 3$, $a_k + 4$ is an element of $\mathcal{A}$. Hence, $\sup_k(a_{k+1} - a_k) \leq 4$.

Now we will talk about large gaps between consecutive elements of subsets. However, sometimes we do not know if a set $\mathcal{A}$ is finite or infinite, and it is better to talk about large intervals not containing elements from $\mathcal{A}$. For $x > 0$, define the number $\rho(x; \mathcal{A})$ as the size of the largest interval $(\alpha, \beta) \subset (0, x)$ without elements of $\mathcal{A}$. Clearly, $\rho(x; \mathcal{A})$ is a nondecreasing function of x.

Now we will talk about large gaps between consecutive elements of subsets. However, sometimes we do not know if a set $\mathcal{A}$ is finite or infinite, and it is better to talk about large intervals not containing elements from $\mathcal{A}$. For $x > 0$, define the number $\rho(x;\mathcal{A})$ as the size of the largest interval $(\alpha,\beta) \subset (0,x)$ without elements of $\mathcal{A}$. Clearly, $\rho(x;\mathcal{A})$ is a nondecreasing function of x.

Let $|\mathcal{A}| = \infty$. Any interval without elements of $\mathcal{A}$ is contained either in $(0,a_1)$ or in (a_k,a_{k+1}) for some k. Thus, if $\mathcal{A}$ is syndetic then for any $x > 0$ we have $\rho(x;\mathcal{A}) \le \max(\max_k(a_{k+1}-a_k), a_1) < \infty$.

Now we will talk about large gaps between consecutive elements of subsets. However, sometimes we do not know if a set $\mathcal{A}$ is finite or infinite, and it is better to talk about large intervals not containing elements from $\mathcal{A}$. For $x > 0$, define the number $\rho(x; \mathcal{A})$ as the size of the largest interval $(\alpha, \beta) \subset (0, x)$ without elements of $\mathcal{A}$. Clearly, $\rho(x; \mathcal{A})$ is a nondecreasing function of x.
Let $|\mathcal{A}| = \infty$. Any interval without elements of $\mathcal{A}$ is contained either in $(0, a_1)$ or in (a_k, a_{k+1}) for some k. Thus, if $\mathcal{A}$ is syndetic then for any $x > 0$ we have $\rho(x; \mathcal{A}) \le \max(\max_k(a_{k+1} - a_k), a_1) < \infty$.
If $|\mathcal{A}| = \infty$ and $\mathcal{A}$ is not syndetic then $a_{k+1} - a_k$ can be arbitrarily large. Since $\rho(x; \mathcal{A}) \ge a_{k+1} - a_k$ for $x \ge a_{k+1}$ we see that $\lim_{x\to\infty} \rho(x; \mathcal{A}) = \infty$. If $\mathcal{A}$ is finite and a is the maximal element of $\mathcal{A}$ then for $x > a$ we have $\rho(x; \mathcal{A}) \ge x - a$. Again, $\lim_{x\to\infty} \rho(x; \mathcal{A}) = \infty$.

Now we will talk about large gaps between consecutive elements of subsets. However, sometimes we do not know if a set $\mathcal{A}$ is finite or infinite, and it is better to talk about large intervals not containing elements from $\mathcal{A}$. For $x > 0$, define the number $\rho(x;\mathcal{A})$ as the size of the largest interval $(\alpha, \beta) \subset (0, x)$ without elements of $\mathcal{A}$. Clearly, $\rho(x;\mathcal{A})$ is a nondecreasing function of x.
Let $|\mathcal{A}| = \infty$. Any interval without elements of $\mathcal{A}$ is contained either in $(0, a_1)$ or in (a_k, a_{k+1}) for some k. Thus, if $\mathcal{A}$ is syndetic then for any $x > 0$ we have $\rho(x;\mathcal{A}) \le \max(\max_k(a_{k+1} - a_k), a_1) < \infty$.
If $|\mathcal{A}| = \infty$ and $\mathcal{A}$ is not syndetic then $a_{k+1} - a_k$ can be arbitrarily large. Since $\rho(x;\mathcal{A}) \ge a_{k+1} - a_k$ for $x \ge a_{k+1}$ we see that $\lim_{x\to\infty} \rho(x;\mathcal{A}) = \infty$. If $\mathcal{A}$ is finite and a is the maximal element of $\mathcal{A}$ then for $x > a$ we have $\rho(x;\mathcal{A}) \ge x - a$. Again, $\lim_{x\to\infty} \rho(x;\mathcal{A}) = \infty$.
Therefore, $\mathcal{A}$ is syndetic if and only if $\lim_{x\to\infty} \rho(x;\mathcal{A}) = \infty$.

Let $\mathcal{P}$ be the set of the primes. It is well-known that $\mathcal{P}$ is not a syndetic set.

Let $\mathcal{P}$ be the set of the primes. It is well-known that $\mathcal{P}$ is not a syndetic set.
E. Westzynthius (1930) proved that $\rho(x;\mathcal{P})/\log x \to \infty$ as $x \to \infty$.
Lower estimates for $\rho(x;\mathcal{P})$ were found by P. Erdős, R.A. Rankin and other mathematicians. K. Ford, B. Green, SK, J. Maynard, T.Tao (2018) proved that for large x

$$\rho(x;\mathcal{P}) \gg \log x(\log\log x)(\log\log\log x)^{-1}(\log\log\log\log x).$$

Heuristically, $\rho(x;\mathcal{P}) \asymp (\log x)^2$. Best known upper estimate is much larger $\rho(x;\mathcal{P}) \le x^{0.525}$ for $x \ge x_0$ (R.C. Baker, G. Harman, J. Pintz, 2001).

We say that a set of k consecutive integers is an integer segment of size k. Thus, a set $\mathcal{A}$ is syndetic if there is k such that any integer segment of size k contains an element of $\mathcal{A}$.

We say that a set of k consecutive integers is an integer segment of size k. Thus, a set $\mathcal{A}$ is syndetic if there is k such that any integer segment of size k contains an element of $\mathcal{A}$.
M. Beigldőck, V. Bergelson, N. Hondman, and D. Strauss (2018) defined a wider class of piecewise syndetic sets. We say that a set $\mathcal{A} \subset \mathbb{N}$ is a piecewise syndetic if there is $k \in \mathbb{N}$ such that there are arbitrary long integer segments such that any k consecutive integers in it contains an element of $\mathcal{A}$.

We say that a set of k consecutive integers is an integer segment of size k. Thus, a set $\mathcal{A}$ is syndetic if there is k such that any integer segment of size k contains an element of $\mathcal{A}$.
M. Beigldőck, V. Bergelson, N. Hondman, and D. Strauss (2018) defined a wider class of piecewise syndetic sets. We say that a set $\mathcal{A} \subset \mathbb{N}$ is a piecewise syndetic if there is $k \in \mathbb{N}$ such that there are arbitrary long integer segments such that any k consecutive integers in it contains an element of $\mathcal{A}$.
We have seen that the function $\rho(x; \mathcal{A})$ indicates whether $\mathcal{A}$ is a syndetic set. In some sense, the rate of tending this function to infinity as $x \to \infty$ shows how far the set $\mathcal{A}$ differs from syndetic sets. We define a similar function related to piecewise syndetic sets.

Now define the function $\rho(x, y; \mathcal{A})$, where $x > 0, y \geq 0$, as the size of the largest interval $(\alpha, \beta) \subset (y, x + y)$ without elements of $\mathcal{A}$. Let

$$\rho_*(x; \mathcal{A}) = \inf_{y \geq 0} \rho(x, y; \mathcal{A}).$$

Clearly, $\rho_*(x; \mathcal{A}) \leq \rho(x; \mathcal{A})$ and $\rho_*(x; \mathcal{A})$ is a nondecreasing function of x.

Now define the function $\rho(x, y; \mathcal{A})$, where $x > 0, y \geq 0$, as the size of the largest interval $(\alpha, \beta) \subset (y, x + y)$ without elements of $\mathcal{A}$. Let

$$\rho_*(x; \mathcal{A}) = \inf_{y \geq 0} \rho(x, y; \mathcal{A}).$$

Clearly, $\rho_*(x; \mathcal{A}) \leq \rho(x; \mathcal{A})$ and $\rho_*(x; \mathcal{A})$ is a nondecreasing function of x.
It is easy to see that a set $\mathcal{A} \subset \mathbb{N}$ is not a piecewise syndetic set if and only if

$$\lim_{x \to \infty} \rho_*(x; \mathcal{A}) = \infty.$$

Now define the function $\rho(x, y; \mathcal{A})$, where $x > 0, y \geq 0$, as the size of the largest interval $(\alpha, \beta) \subset (y, x + y)$ without elements of $\mathcal{A}$. Let

$$\rho_*(x; \mathcal{A}) = \inf_{y \geq 0} \rho(x, y; \mathcal{A}).$$

Clearly, $\rho_*(x; \mathcal{A}) \leq \rho(x; \mathcal{A})$ and $\rho_*(x; \mathcal{A})$ is a nondecreasing function of x.
It is easy to see that a set $\mathcal{A} \subset \mathbb{N}$ is not a piecewise syndetic set if and only if

$$\lim_{x \to \infty} \rho_*(x; \mathcal{A}) = \infty.$$

It turns out that sometimes the technique of the proof of lower estimates for $\rho(x; \mathcal{A})$ gives actually more, namely, the same lower estimate for $\rho_*(x; \mathcal{A})$. This gives not only a stronger result, but also a better understanding of contemporary methods to study $\rho(x; \mathcal{A})$. We discuss briefly the case $\mathcal{A} = \mathcal{P}$.

Let $z \geq 2$. We denote by $P(z)$ the product of primes up to z and by $Y(z)$ the maximal number Y such that there exist Y consecutive integers each of them is divisible by some prime not exceeding z. In the above mentioned paper by K. Ford, B. Green, SK, J. Maynard, T.Tao (2018) it was proved that

$$Y(z) \gg z(\log z)(\log\log z)^{-1}(\log\log\log z). \tag{1}$$

Let $z \geq 2$. We denote by $P(z)$ the product of primes up to z and by $Y(z)$ the maximal number Y such that there exist Y consecutive integers each of them is divisible by some prime not exceeding z. In the above mentioned paper by K. Ford, B. Green, SK, J. Maynard, T.Tao (2018) it was proved that

$$Y(z) \gg z(\log z)(\log\log z)^{-1}(\log\log\log z). \quad (1)$$

The inequality

$$\rho(x;\mathcal{P}) \gg \log x(\log\log x)(\log\log\log x)^{-1}(\log\log\log\log x).$$

can be deduced from (1) and standard arguments. The same arguments give a stronger inequality

$$\rho_*(x;\mathcal{P}) \gg (\log x)(\log\log x)(\log\log\log x)^{-1}(\log\log\log\log x). \quad (2)$$

Take a large positive integer x, and let z be the maximal positive integer satisfying the inequality $x \geq 2P(z)$. It follows from the well-known estimate $\log P(z) \asymp z$ that

$$z \gg \log x. \tag{3}$$

Take a large positive integer x, and let z be the maximal positive integer satisfying the inequality $x \geq 2P(z)$. It follows from the well-known estimate $\log P(z) \asymp z$ that

$$z \gg \log x. \tag{3}$$

Let $Y = Y(z)$. Then there exists an integer u such that any of the numbers $u + 1, \ldots, u + Y$ has a prime factor less than or equal to z. Next we take an arbitrary $y \geq 0$ and a number v such $v \equiv u (\bmod P(z))$ and $v + Y \in (x + y - P(z), x + y]$. then all numbers $v + j, j = 1, \ldots, Y$. are composite numbers from $(y, y + x]$. Since $y \geq 0$ is an arbitrary integer, we conclude that

$$\rho_*(x; \mathcal{P}) \geq Y(z).$$

Due to (1) and (3) we get the required inequality (2).

There is a similar situation for the set $\mathcal{S}$ of positive integers representable as a sum of two squares and for the set $\mathcal{Q}$ of squarefree positive integers. Richards (1982) established that $\rho(x; \mathcal{S}) \gg \log x$ for $x \geq 2$. His arguments show that actually $\rho_*(x; \mathcal{S}) \gg \log x$. The order $\log x$ remain unbeaten.

There is a similar situation for the set $\mathcal{S}$ of positive integers representable as a sum of two squares and for the set $\mathcal{Q}$ of squarefree positive integers. Richards (1982) established that $\rho(x;\mathcal{S}) \gg \log x$ for $x \geq 2$. His arguments show that actually $\rho_*(x;\mathcal{S}) \gg \log x$. The order $\log x$ remain unbeaten.

Elementary arguments give the lower bounds

$$\rho(x;\mathcal{Q}) \geq \rho_*(x;\mathcal{Q}) \gg \log x/\log\log x \,(x \geq 3).$$

There is a similar situation for the set $\mathcal{S}$ of positive integers representable as a sum of two squares and for the set $\mathcal{Q}$ of squarefree positive integers. Richards (1982) established that $\rho(x; \mathcal{S}) \gg \log x$ for $x \geq 2$. His arguments show that actually $\rho_*(x; \mathcal{S}) \gg \log x$. The order $\log x$ remain unbeaten.

Elementary arguments give the lower bounds

$$\rho(x; \mathcal{Q}) \geq \rho_*(x; \mathcal{Q}) \gg \log x / \log \log x \, (x \geq 3).$$

We know that the upper estimate for $\rho(x; \mathcal{P})$ has a power type order, and it is much larger than known lower estimates and even the heuristic estimate. The same situation takes place for $\mathcal{S}$ and $\mathcal{Q}$ as well.

There is a similar situation for the set $\mathcal{S}$ of positive integers representable as a sum of two squares and for the set $\mathcal{Q}$ of squarefree positive integers. Richards (1982) established that $\rho(x;\mathcal{S}) \gg \log x$ for $x \geq 2$. His arguments show that actually $\rho_*(x;\mathcal{S}) \gg \log x$. The order $\log x$ remain unbeaten.
Elementary arguments give the lower bounds

$$\rho(x;\mathcal{Q}) \geq \rho_*(x;\mathcal{Q}) \gg \log x/\log\log x\,(x \geq 3).$$

We know that the upper estimate for $\rho(x;\mathcal{P})$ has a power type order, and it is much larger than known lower estimates and even the heuristic estimate. The same situation takes place for $\mathcal{S}$ and $\mathcal{Q}$ as well.
Moreover, we do not know how to estimate from above the function $\rho_*(x;\mathcal{A})$ better than $\rho(x;\mathcal{A})$ for $\mathcal{A} = \mathcal{P}$ and $\mathcal{A} = \mathcal{S}$. In these cases the distinction between current lower and upper bounds for $\rho_*(x\mathcal{A})$ is terrible.

However, for $\mathcal{A} = \mathcal{Q}$ our understanding of the behavior of $\rho_*(x\mathcal{A})$ is much better. The main result of the talk is the following theorem.

Theorem

We have

$$\rho_*(x;\mathcal{Q}) \asymp \log x/\log\log x \quad (x \geq 3).$$

However, for $\mathcal{A} = \mathcal{Q}$ our understanding of the behavior of $\rho_*(x\mathcal{A})$ is much better. The main result of the talk is the following theorem.

Theorem

We have

$$\rho_*(x; \mathcal{Q}) \asymp \log x / \log\log x \quad (x \geq 3).$$

Actually we can get the same order for $\rho_*(x; \mathcal{A})$ if $\mathcal{A}$ behaves like $\mathcal{Q}$. In particular, for any positive integer $k \geq 2$ this holds for the set $\mathcal{A}$ of the k-th powers free positive integers.

Sometimes it happens that we can estimate nontrivially the function $\rho(x;\mathcal{A})$ but can not do so for $\rho_*(x;\mathcal{A})$, or vice versa.

Sometimes it happens that we can estimate nontrivially the function $\rho(x;\mathcal{A})$ but can not do so for $\rho_*(x;\mathcal{A})$, or vice versa.
Let $k\in\mathbb{N}$, $k\geq 2$ and $\mathcal{Q}_k$ be the set of positive integers representable as $\sum_{j=1}^k a_j^k$, $a_j\in\mathbb{Z}_+$. In particular, $\mathcal{Q}_2=\mathcal{Q}$. We know that $\rho(x;\mathcal{Q})\gg\log x$ for $x\geq 2$. The proof is based on local restrictions for the set $\mathcal{Q}$: the elements of $\mathcal{Q}$ must avoid a series of congruences modulo small numbers. For $k\geq 3$ there are not many local restrictions for the set $\mathcal{Q}_k$, and the question whether $\mathcal{Q}_k$ is a syndetic is highly nontrivial. For $k=3$ and $k=4$ the negative answer was given by L. Chidelli (2020). However, we do not know whether the sets $\mathcal{Q}_k$ are piecewise syndetic.

Now we consider the set $\mathcal{T}$ of the values of the totient function φ. It is very likely that for large x the number $\rho(x;\mathcal{T})$ is much larger than $\rho_*(x;\mathcal{T})$. Consider the counting function

$$\pi(x;\mathcal{A}) = |\{n \in \mathcal{A} : \ n \le x\}|.$$

Since for most x the number $\varphi(n)$ is divisible by all small primes, we can deduce that $\pi(x;\mathcal{T}) = o(x)$ as $x \to \infty$. iThis implies that $\rho(x;\mathcal{T}) \to \infty$ as $x \to \infty$. Hence, $\mathcal{T}$ is not a syndetic set.

Now we consider the set $\mathcal{T}$ of the values of the totient function φ. It is very likely that for large x the number $\rho(x;\mathcal{T})$ is much larger than $\rho_*(x;\mathcal{T})$. Consider the counting function

$$\pi(x;\mathcal{A}) = |\{n \in \mathcal{A} : \ n \le x\}|.$$

Since for most x the number $\varphi(n)$ is divisible by all small primes, we can deduce that $\pi(x;\mathcal{T}) = o(x)$ as $x \to \infty$. iThis implies that $\rho(x;\mathcal{T}) \to \infty$ as $x \to \infty$. Hence, $\mathcal{T}$ is not a syndetic set.
K. Ford (1998) determined the exact order of $\pi(x;\mathcal{T})$.

Now we consider the set $\mathcal{T}$ of the values of the totient function φ. It is very likely that for large x the number $\rho(x;\mathcal{T})$ is much larger than $\rho_*(x;\mathcal{T})$. Consider the counting function

$$\pi(x;\mathcal{A}) = |\{n \in \mathcal{A} :\ n \le x\}|.$$

Since for most x the number $\varphi(n)$ is divisible by all small primes, we can deduce that $\pi(x;\mathcal{T}) = o(x)$ as $x \to \infty$. iThis implies that $\rho(x;\mathcal{T}) \to \infty$ as $x \to \infty$. Hence, $\mathcal{T}$ is not a syndetic set.
K. Ford (1998) determined the exact order of $\pi(x;\mathcal{T})$.
The magnitude $\rho_*(x;\mathcal{T})$ is evaluated for large x under a very plausible conjecture.

Conjecture

(Dickson's conjecture) Let s be a positive integer and let $F_1, \dots, F_s$ be s linear polynomials with integral coefficients and positive linear coefficient such that their product has no fixed prime divisor. Then there exist infinitely many positive integers l such that $F_1(l), \dots, F_s(l)$ are all primes.

Conjecture

(Dickson's conjecture) Let s be a positive integer and let $F_1, \ldots, F_s$ be s linear polynomials with integral coefficients and positive linear coefficient such that their product has no fixed prime divisor. Then there exist infinitely many positive integers l such that $F_1(l), \ldots, F_s(l)$ are all primes.

Assuming the validity of Dickson's conjecture, J. – M. Deshouillers, P. Eyyumni, and C. Gun (2021) actually proved that $\rho_*(x; \mathcal{T}) = 4$ for sufficiently large x.

Let me observe that assuming the validity of Dickson's conjecture and using the results of H. L. Montgomery and R. C. Vaughan one can prove that $\rho(x;\mathcal{P}) = x^{o(1)}$ as $x \to \infty$. Probably, the last estimate can be sharpen to $\rho(x;\mathcal{P}) = (\log x)^{O(1)}$.

Let me observe that assuming the validity of Dickson's conjecture and using the results of H. L. Montgomery and R. C. Vaughan one can prove that $\rho(x;\mathcal{P}) = x^{o(1)}$ as $x \to \infty$. Probably, the last estimate can be sharpen to $\rho(x;\mathcal{P}) = (\log x)^{O(1)}$.
However, we do not know whether there are infinitely many positive integers x such that $\rho_*(x;\mathcal{P}) < \rho(x;\mathcal{P})$ or whether there are infinitely many positive integers x such that $\rho_*(x;\mathcal{P}) = \rho(x;\mathcal{P})$. We can prove that there are infinitely many positive integers x such that $\rho_*(x;\mathcal{Q}) < \rho(x;\mathcal{Q})$.

Let me observe that assuming the validity of Dickson's conjecture and using the results of H. L. Montgomery and R. C. Vaughan one can prove that $\rho(x; \mathcal{P}) = x^{o(1)}$ as $x \to \infty$. Probably, the last estimate can be sharpen to $\rho(x; \mathcal{P}) = (\log x)^{O(1)}$.
However, we do not know whether there are infinitely many positive integers x such that $\rho_*(x; \mathcal{P}) < \rho(x; \mathcal{P})$ or whether there are infinitely many positive integers x such that $\rho_*(x; \mathcal{P}) = \rho(x; \mathcal{P})$. We can prove that there are infinitely many positive integers x such that $\rho_*(x; \mathcal{Q}) < \rho(x; \mathcal{Q})$.

THANK YOU!