PRODUCT-FREE SETS WITH HIGH DENSITY

PÄR KURLBERG, JEFFREY C. LAGARIAS, AND CARL POMERANCE

Dedicated to Professor Andrzej Schinzel on his 75th birthday

Abstract. We show that there are sets of integers with asymptotic density arbitrarily close to 1 in which there is no solution to the equation $ab = c$, with a, b, c in the set. We also consider some natural generalizations, as well as a specific numerical example of a product-free set of integers with asymptotic density greater than 1/2.

1. Introduction

We say a set of integers $\mathcal{S}$ is *product-free* if whenever $a, b, c \in \mathcal{S}$ we have $ab \neq c$. Similarly, if $\mathcal{S} \subset \mathbb{Z}/n\mathbb{Z}$, we say $\mathcal{S}$ is product-free if $ab \not\equiv c \pmod{n}$, whenever $a, b, c \in \mathcal{S}$. Clearly, if $\mathcal{S}$ is a product-free subset of $\mathbb{Z}/n\mathbb{Z}$, then the set of integers congruent modulo n to some member of $\mathcal{S}$ is a product-free set of integers. For a positive integer n, let $D(n)$ denote the maximum value of $|\mathcal{S}|/n$ where $\mathcal{S}$ runs over all product-free subsets of $\mathbb{Z}/n\mathbb{Z}$. (Here $|\mathcal{S}|$ denotes the cardinality of a set $\mathcal{S}$.)

In a recent paper, the third author and Schinzel [9] obtained an upper bound on $D(n)$ valid for a large set of n. They showed that $D(n) < 1/2$ whenever n is not divisible by a square with at least 6 distinct prime factors. Further, those numbers which are divisible by a square with at least 6 distinct prime factors form a set of asymptotic density about 1.56×10^{-8}. Originally they suspected that $D(n) < 1/2$ might hold for all n.

In this paper we show that for each real number $\epsilon > 0$ there is some number n with $D(n) > 1 - \epsilon$. Thus, there are product-free sets of integers with asymptotic density arbitrarily close to 1. Stated this way, the result is best possible, since no product-free set can have density 1. Indeed, if $\mathcal{S}$ is a product free set of positive integers and a is the least member of $\mathcal{S}$, then it is easy to see that the upper density of $\mathcal{S}$ is at most $1 - 1/(2a)$; see Remark 2.7.

A consequence of our main result is that the set of integers n having $D(n) > 1 - \epsilon$ has a positive lower density. This follows using the property that $D(mn) \geq D(n)$ for all positive integers m, n. If $D(n_0) > 1 - \epsilon$, then

2010 *Mathematics Subject Classification.* Primary 11B05, 11B75.
Key words and phrases. product-free, asymptotic density.

it shows that $D(n) > 1 - \epsilon$ holds for every multiple of n_0, and so it holds for a set of positive integers n of positive lower density. Furthermore the set $\mathcal{N}(u) = \{n \geq 1 : D(n) > u\}$ has a well-defined logarithmic density $\delta(u)$ which is positive for $0 \leq u < 1$. In Theorem 2.1 we obtain a quantitative rate at which $D(n)$ approaches 1, which yields a lower bound for $\delta(u)$ as $u \to 1^-$, given as (5.1) in Sec. 5.

We also compute a numerical example of a number n with $D(n) > 1/2$ and we consider some generalizations of the equation $ab = c$.

It is interesting to note that while there are product-free subsets with density arbitrarily close to 1, the density of *sum-free* subsets of finite abelian groups (written additively) is easily seen to be bounded by 1/2 (see [4] for a complete characterization of the maximum density of sum-free subsets of various types of finite abelian groups).

2. The main theorem

In this section we show that there can be product-free sets of integers of density arbitrarily close to one, but not equal to one. Our main result is as follows.

Theorem 2.1. *There is a positive constant C and infinitely many integers n with*

$$D(n) > 1 - \frac{C}{(\log\log n)^{1-\frac{1}{2}\mathrm{e}\log 2}}.$$

Here the exponent $1 - \frac{1}{2}\mathrm{e}\log 2 \approx 0.057915$.

Corollary 2.2. *For each real number $\epsilon > 0$ there is a positive integer n with $D(n) > 1 - \epsilon$.*

We first sketch the idea of the proof. Let $\Omega(m)$ denote the number of prime factors of m counted with multiplicity. Clearly for any fixed z, the set of numbers m with $z < \Omega(m) < 2z$ is product-free. Further, after Hardy and Ramanujan, we know that $\Omega(m)$ for numbers $m \leq x$ is usually concentrated near $\log\log x$. So if $z \approx \frac{2}{3}\log\log x$ (actually $\frac{\mathrm{e}}{4}$ works out a little better than $\frac{2}{3}$), we have a product-free set that has the great preponderance of integers in $[1, x]$. With an extra device (see Lemma 2.3) for creating such a set that is periodic modulo some particular large number n, we obtain the result. The idea used bears some resemblance to that of Remark 2 and its proof in Hajdu, Schinzel, and Skalba [5].

Before giving the proof, we establish some preliminary lemmas. Let φ denote Euler's function and let $\mathrm{rad}(n)$ denote the largest squarefree divisor of the positive integer n.

Lemma 2.3. *Suppose that n is a positive integer and $\mathcal{D}$ is a product-free set of divisors of $n/\mathrm{rad}(n)$. Then*

$$\mathcal{S}_{\mathcal{D}} := \{s \in \mathbb{Z}/n\mathbb{Z} : \gcd(s,n) \in \mathcal{D}\}$$

is product-free and

$$|\mathcal{S}_{\mathcal{D}}| = \varphi(n) \sum_{d\in\mathcal{D}} \frac{1}{d}.$$

Proof. Suppose $s_1, s_2 \in \mathcal{S}_{\mathcal{D}}$ with $\gcd(s_i, n) = d_i \in \mathcal{D}$ for $i = 1, 2$. We have $\gcd(s_1s_2, n) = \gcd(d_1d_2, n) = d_3$, say. If $d_3 \nmid n/\mathrm{rad}(n)$, then by hypothesis $d_3 \notin \mathcal{D}$, so $s_1s_2 \notin \mathcal{S}_{\mathcal{D}}$. On the other hand, if $d_3 \mid n/\mathrm{rad}(n)$, then $d_3 = d_1d_2$, so again by hypothesis, $d_3 \notin \mathcal{D}$ and $s_1s_2 \notin \mathcal{S}_{\mathcal{D}}$. Thus, $\mathcal{S}_{\mathcal{D}}$ is product-free and it remains to compute its cardinality. For $d \in \mathcal{D}$, we have

$$\{s \in \mathbb{Z}/n\mathbb{Z} : \gcd(s,n) = d\} = \{jd : j \in \mathbb{Z}/(n/d)\mathbb{Z},\ \gcd(j, n/d) = 1\}.$$

Thus, $|\mathcal{S}_{\mathcal{D}}| = \sum_{d\in\mathcal{D}} \varphi(n/d)$. But, by hypothesis, we have $\mathrm{rad}(n/d) = \mathrm{rad}(n)$ for $d \in \mathcal{D}$, so that $\varphi(n/d) = \varphi(n)/d$. This completes the proof. □

For an integer $n > 1$, let $P(n)$ denote the largest prime factor of n and let $P(1) = 1$. As above, we let $\Omega(n)$ denote the number of prime factors of n, counted with multiplicity. We use the notation $f(x) \asymp g(x)$ to mean there are positive constants c_1, c_2 such that $c_1g(x) \le f(x) \le c_2(x)$ in some stated domain for the variable x. Lemma 2.4 and Corollary 2.5 below are standard results, cf. Exercises 04 and 05 in [6]; we give the details for completeness.

Lemma 2.4. *Uniformly for real numbers x, z with $x \ge 2$ and $0 < z < 2$,*

$$\sum_{P(n)\le x} \frac{z^{\Omega(n)}}{n} \asymp \frac{1}{2-z}(\log x)^z.$$

Proof. We have

$$\begin{aligned}\sum_{P(n)\le x} \frac{z^{\Omega(n)}}{n} &= \prod_{p\le x}\left(1 + \frac{z}{p} + \frac{z^2}{p^2} + \cdots\right) = \prod_{p\le x}\left(1 - \frac{z}{p}\right)^{-1}\\ &= \prod_{p\le x}\left(1 - \frac{1}{p}\right)^{-z} \prod_{p\le x}\left(1 - \frac{1}{p}\right)^{z}\left(1 - \frac{z}{p}\right)^{-1}.\end{aligned}$$

By the theorem of Mertens we have $\prod_{p\le x}(1-1/p)^{-z} \sim \mathrm{e}^{\gamma z}(\log x)^z$ uniformly for z in the interval $(0, 2)$, as $x \to \infty$, where γ is the Euler–Mascheroni constant. Thus, it suffices to prove that the second product above is of magnitude $1/(2-z)$. Using the power series for $\log(1-t)$, we have

$$\log\left(\prod_{p\le x}\left(1-\frac{1}{p}\right)^{z}\left(1-\frac{z}{p}\right)^{-1}\right)=\sum_{p\le x}\left(z\log\left(1-\frac{1}{p}\right)-\log\left(1-\frac{z}{p}\right)\right)$$
$$=z\log\frac{1}{2}-\log\left(1-\frac{z}{2}\right)+O\left(\sum_{3\le p\le x}\frac{1}{p^2}\right)=-\log(2-z)+O(1).$$

This then completes the proof of the lemma. □

We will use the entropy-like function $Q(x)$ defined for $x>0$ by

$$Q(x)=x\log x-x+1.$$

Note that $Q(x)\ge 0$ for all $x>0$ with equality only at $x=1$.

Corollary 2.5. *Uniformly for real numbers α,β,x with $0<\alpha\le 1\le\beta<2$ and $x\ge 3$, we have*

$$\sum_{\substack{P(n)\le x\\ \Omega(n)\le\alpha\log\log x}}\frac{1}{n}\ll(\log x)^{1-Q(\alpha)},\qquad \sum_{\substack{P(n)\le x\\ \Omega(n)\ge\beta\log\log x}}\frac{1}{n}\ll\frac{1}{2-\beta}(\log x)^{1-Q(\beta)}.$$

Proof. We have

$$\sum_{\substack{P(n)\le x\\ \Omega(n)\le\alpha\log\log x}}\frac{1}{n}\le\sum_{P(n)\le x}\frac{\alpha^{\Omega(n)-\alpha\log\log x}}{n}$$
$$=\sum_{P(n)\le x}\frac{\alpha^{\Omega(n)}}{n}(\log x)^{-\alpha\log\alpha}\ll(\log x)^{\alpha-\alpha\log\alpha},$$

using $0<\alpha\le 1$ and Lemma 2.4 with $z=\alpha$. Similarly, Lemma 2.4 with $z=\beta$ gives

$$\sum_{\substack{P(n)\le x\\ \Omega(n)\ge\beta\log\log x}}\frac{1}{n}\le\sum_{P(n)\le x}\frac{\beta^{\Omega(n)-\beta\log\log x}}{n}\ll\frac{1}{2-\beta}(\log x)^{\beta-\beta\log\beta}.$$

This completes the proof of the corollary. □

Proof of Theorem 2.1. Let x be a large real number, let ℓ_x denote the least common multiple of the integers in $[1,x]$, and let $n_x=\ell_x^2$. Thus, by the prime number theorem, we have $n_x=\mathrm{e}^{(2+o(1))x}$ as $x\to\infty$, so that

$$\log\log n_x=\log x+O(1). \tag{2.1}$$

Let

$$\mathcal{D}_x=\left\{d\mid\ell_x\ :\ \frac{\mathrm{e}}{4}\log\log x<\Omega(d)<\frac{\mathrm{e}}{2}\log\log x\right\}.$$

We note that each $d \in \mathcal{D}_x$ divides $n_x/\mathrm{rad}(n_x)$ and that $\mathcal{D}_x$ is product-free. Thus, by Lemma 2.3 we find that

$$\mathcal{S}_{\mathcal{D}_x} := \{a \in \mathbb{Z}/n_x\mathbb{Z} : \gcd(a, n_x) \in \mathcal{D}_x\}$$

is a product-free subset of $\mathbb{Z}/n_x\mathbb{Z}$, with density $\mathcal{D}(\mathcal{S}) = \frac{\varphi(n_x)}{n_x}\sum_{d\in\mathcal{D}_x}\frac{1}{d}$. Using (2.1) it suffices to show that for some positive constant c and x sufficiently large,

$$\frac{\varphi(n_x)}{n_x}\sum_{d\in\mathcal{D}_x}\frac{1}{d} \geq 1 - \frac{c}{(\log x)^{1-\frac{1}{2}\mathrm{e}\log 2}}. \tag{2.2}$$

We have

$$\sum_{d\in\mathcal{D}_x}\frac{1}{d} \geq \sum_{d\mid\ell_x}\frac{1}{d} - \sum_{\substack{P(d)\leq x\\ \Omega(d)\leq\frac{\mathrm{e}}{4}\log\log x}}\frac{1}{d} - \sum_{\substack{P(d)\leq x\\ \Omega(d)\geq\frac{\mathrm{e}}{2}\log\log x}}\frac{1}{d}.$$

Since $1 - Q(\frac{\mathrm{e}}{4}) = 1 - Q(\frac{\mathrm{e}}{2}) = \frac{1}{2}\mathrm{e}\log 2$, Corollary 2.5 implies there is some absolute constant $c' > 0$ with

$$\sum_{d\in\mathcal{D}_x}\frac{1}{d} \geq \sum_{d\mid\ell_x}\frac{1}{d} - c'(\log x)^{\frac{1}{2}\mathrm{e}\log 2}.$$

Now, letting σ denote the sum-of-divisors function,

$$\sum_{d\mid\ell_x}\frac{1}{d} = \frac{\sigma(\ell_x)}{\ell_x} = \prod_{p^a\|\ell_x}\frac{p^{a+1}-1}{p^a(p-1)} = \prod_{p\leq x}\frac{p}{p-1}\prod_{p^a\|\ell_x}\left(1-\frac{1}{p^{a+1}}\right)$$
$$\geq \prod_{p\leq x}\frac{p}{p-1}\cdot\left(1-\frac{1}{x}\right)^{\pi(x)} \geq \prod_{p\leq x}\frac{p}{p-1}\cdot\left(1-\frac{\pi(x)}{x}\right),$$

where $\pi(x)$ denotes the prime-counting function. Thus, since $\varphi(n_x)/n_x = \prod_{p\leq x}(p-1)/p$,

$$\frac{\varphi(n_x)}{n_x}\sum_{d\in\mathcal{D}_x}\frac{1}{d} \geq 1 - \frac{\pi(x)}{x} - c'(\log x)^{\frac{1}{2}\mathrm{e}\log 2}\prod_{p\leq x}\frac{p-1}{p}.$$

Using the theorem of Mertens for the product and the Chebyshev estimate $\pi(x) \ll x/\log x$, we obtain (2.2), completing the proof of Theorem 2.1. □

Remark 2.6. It is possible to uniformly save a factor $\sqrt{\log\log x}$ in Corollary 2.5 under the strengthened hypothesis that $\alpha \in [\epsilon, 1-\epsilon]$ and $\beta \in [1+\epsilon, 2-\epsilon]$, where $\epsilon > 0$ is fixed but arbitrary. This gives a slightly stronger version of Theorem 2.1: There is a positive constant C such that

$$D(n) > 1 - \frac{C}{(\log\log n)^{1-\frac{1}{2}\mathrm{e}\log 2}\sqrt{\log\log\log n}} \text{ for infinitely many } n. \tag{2.3}$$

The details are presented in a sequel paper [7], where the principal result is that (2.3), apart from the constant C, is best possible.

Remark 2.7. For a set $\mathcal{S}$ of positive integers, let $\mathcal{S}(x) = \mathcal{S} \cap [1, x]$. If $\mathcal{S}$ is product-free with least member a, then its upper asymptotic density, defined as

$$\overline{d}(\mathcal{S}) := \limsup_{x \to \infty} \frac{1}{x} |\mathcal{S}(x)|,$$

satisfies $\overline{d}(\mathcal{S}) \le 1 - \frac{1}{2a}$. To see this, suppose $x \ge a$ is arbitrary. Since $\mathcal{S}(x) \setminus \mathcal{S}(x/a)$ lies in $(x/a, x]$, we have $|\mathcal{S}(x)| - |\mathcal{S}(x/a)| \le x - \lfloor x/a \rfloor$. Also, multiplying each member of $\mathcal{S}(x/a)$ by a creates products in $[1, x]$ which cannot lie in $\mathcal{S}$, so we have $|\mathcal{S}(x)| \le x - |\mathcal{S}(x/a)|$. Adding these two inequalities leads to $|\mathcal{S}(x)| \le x - \frac{1}{2}\lfloor x/a \rfloor$, which proves the assertion.

3. Generalizations

If k, j are positive integers, we say a set of integers (or residue classes in $\mathbb{Z}/n\mathbb{Z}$) is (k, j)-product-free if there is no solution to $a_1 a_2 \ldots a_k = b_1 b_2 \ldots b_j$ with all $k + j$ letters being elements of the set. If $k = j$ then only the empty set is (k, j)-product-free. Indeed, if a is an element of the set, the equation $a^k = a^k$ shows that we cannot avoid $a_1 a_2 \ldots a_k = b_1 b_2 \ldots b_j$. Thus we restrict to cases where $k \ne j$, and we may as well assume that $k > j$. The case of $k = 2, j = 1$ is exactly the definition of product-free that was considered in the last section. In this section we record the following simple generalization.

Theorem 3.1. *For each real number $\epsilon > 0$ and integer $m \ge 3$ there is a positive integer n and a subset $\mathcal{S}$ of $\mathbb{Z}/n\mathbb{Z}$ of cardinality at least $(1 - \epsilon)n$ that is simultaneously (k, j)-product-free for all positive integers $k > j$ with $k + j \le m$.*

Proof. As in the proof of Theorem 2.1, let ℓ_x denote the least common multiple of the integers in $[1, x]$, but now we set $n_x = \ell_x^m$, and

$$\mathcal{D}_x = \left\{ d \mid \ell_x \,:\, \left(1 - \frac{1}{m}\right) \log\log x < \Omega(d) < \left(1 + \frac{1}{m}\right) \log\log x \right\}.$$

Let $k > j$ be positive integers with $k + j \le m$. If $d_1, \ldots, d_k \in \mathcal{D}_x$ and also $d_1', \ldots, d_j' \in \mathcal{D}_x$, it is easy to see that $d = d_1 \ldots d_k$ and $d' = d_1' \ldots d_j'$ are divisors of n_x. In addition, $d \ne d'$, since $\Omega(d) > k(1 - \frac{1}{m}) \log\log x \ge j(1 + \frac{1}{m}) \log\log x > \Omega(d')$. Thus, $\mathcal{D}_x$ is (k, j)-product-free as is the set $\mathcal{S}_{\mathcal{D}_x}$ (cf. Lemma 2.3). As in the proof of Theorem 2.1 it suffices to show that for each $\epsilon > 0$,

$$\frac{\varphi(n_x)}{n_x} \sum_{d \in \mathcal{D}_x} \frac{1}{d} \ge 1 - \epsilon$$

for all sufficiently large x depending on ϵ. Already from the proof of Theorem 2.1, we have

$$\frac{\varphi(n_x)}{n_x}\sum_{d|\ell_x}\frac{1}{d} \geq 1-\frac{\pi(x)}{x} \sim 1$$

as $x\to\infty$. Since $\varphi(n_x)/n_x \sim 1/(e^\gamma \log x)$ as $x\to\infty$, it suffices to show that

$$\sum_{\substack{d|\ell_x\\ d\notin\mathcal{D}_x}}\frac{1}{d} = o(\log x) \quad \text{as } x\to\infty. \tag{3.1}$$

Letting $\delta_1 = Q(1-1/m)$ and $\delta_2 = Q(1+1/m)$, we have $\delta_1,\delta_2 > 0$. Using Corollary 2.5,

$$\sum_{\substack{d|\ell_x\\ \Omega(d)\leq\left(1-\frac{1}{m}\right)\log\log x}}\frac{1}{d} \leq (\log x)^{1-\delta_1/2}, \qquad \sum_{\substack{d|\ell_x\\ \Omega(d)\geq\left(1+\frac{1}{m}\right)\log\log x}}\frac{1}{d} \leq (\log x)^{1-\delta_2/2}$$

for all large x. Thus, we have (3.1), which completes the proof of the theorem. ☐

Returning to the case when $k=j$, we can redefine the notion of (k,k)-product-free to mean that the equation $a_1a_2\dots a_k = b_1b_2\dots b_k$ implies that $\{a_1,a_2,\dots,a_k\} = \{b_1,b_2,\dots,b_k\}$ as multisets. For example, the primes are (k,k)-product-free for every k. This is essentially a best-possible result, for as shown by Erdős [3] in 1938, if $\mathcal{S}$ is a subset of the positive integers which is $(2,2)$-product-free, then the number of members of $\mathcal{S}$ in $[1,x]$ is $\pi(x)+O(x^{3/4})$.

The equation $abc=d^2$ was recently considered in [5], where it was shown (see Corollary 1) that if $\mathcal{S}$ is a set of integers such that

$$abc = d^2 \quad \text{has no solution with } a,b,c\in\mathcal{S}, \quad d \text{ arbitrary},$$

then the lower asymptotic density of $\mathcal{S}$ is at most 1/2. This result was inadvertently misquoted in [9], where it was asserted that such a result holds with all of $a,b,c,d\in\mathcal{S}$. In fact, this is false since Theorem 3.1 applied with $(k,j)=(3,2)$ implies the complementary result that for any $\epsilon>0$ there exists a set $\mathcal{S}$ of density exceeding $1-\epsilon$ such that

$$abc = d^2 \quad \text{has no solution with } \ a,b,c,d\in\mathcal{S}. \tag{3.2}$$

More precisely, it gives:

Corollary 3.2. *For each real number $\epsilon>0$, there is a positive integer n and a subset $\mathcal{S}$ of $\mathbb{Z}/n\mathbb{Z}$ of cardinality at least $(1-\epsilon)n$ such that $abc=d^2$ has no solution with $a,b,c,d\in\mathcal{S}$.*

4. A numerical example

In this section we give the details for a number N for which there exists a product-free subset of $\mathbb{Z}/N\mathbb{Z}$ of size larger than $N/2$. Our example is very large; it would be of interest to see if a substantially smaller number could be found.

Let $\mathcal{P}$ denote the set of the first 10,000,000 primes and let Q be their product. For each positive integer j, let

$$\sigma_j = \sum_{p\in\mathcal{P}} \frac{1}{p^j}, \quad S_j = \sum_{\substack{\mathrm{rad}(m)|Q \\ \Omega(m)=j}} \frac{1}{m}.$$

We have computed these sums for j up to 13, finding that to 6 decimal places,

$$\begin{array}{llll} \sigma_1 = 3.206219, & \sigma_2 = 0.452247, & \sigma_3 = 0.174763, & \sigma_4 = 0.076993, \\ \sigma_5 = 0.035755, & \sigma_6 = 0.017070, & \sigma_7 = 0.008284, & \sigma_8 = 0.004061, \\ \sigma_9 = 0.002004, & \sigma_{10} = 0.000994, & \sigma_{11} = 0.000494, & \sigma_{12} = 0.000246, \\ \sigma_{13} = 0.000123 & & & \end{array}$$

and

$$\begin{array}{llll} S_1 = 3.206219, & S_2 = 5.366043, & S_3 = 6.276492, & S_4 = 5.796977, \\ S_5 = 4.529060, & S_6 = 3.130763, & S_7 = 1.976769, & S_8 = 1.167289, \\ S_9 = 0.656256, & S_{10} = 0.356061, & S_{11} = 0.188345, & S_{12} = 0.097866, \\ S_{13} = 0.050226. & & & \end{array}$$

Concerning these calculations, we note that the computation for $\sigma_1 = S_1$ is the most time consuming. The other values of σ_j represent the starts of rapidly converging series, and in fact these values can be found on the web as values of the "prime zeta function." The remaining values of S_j are easily computed by a hand calculator using the identity

$$S_k = \frac{1}{k}\sum_{j=1}^{k} \sigma_j S_{k-j},$$

where by convention we take $S_0 = 1$ (see [8, page 23, (2.11)]).

Let

$$N = Q^{14} = \prod_{p\in\mathcal{P}} p^{14}$$

and let

$$\mathcal{D} = \{d \mid N \,:\, 3 \le \Omega(d) \le 5 \ \text{ or } \ 11 \le \Omega(d) \le 13\}.$$

A moment's reflection shows that $\mathcal{D}$ is product-free and that each member of $\mathcal{D}$ divides $N/\mathrm{rad}(N)$, and so from Lemma 2.3,

$$\mathcal{S}_{\mathcal{D}} = \{m \bmod N : \gcd(m, N) \in \mathcal{D}\}$$

is also product-free. Further,

$$\frac{|\mathcal{S}_{\mathcal{D}}|}{N} = \frac{\varphi(N)}{N}\sum_{d\in\mathcal{D}}\frac{1}{d}. \tag{4.1}$$

We may compute $\varphi(N)/N$ using σ_1 and σ_2 as follows:

$$\log\frac{\varphi(N)}{N} = \sum_{p\in\mathcal{P}}\log\left(1-\frac{1}{p}\right) = -\sigma_1 - \frac{1}{2}\sigma_2 + \sum_{p\in\mathcal{P}}\left(\frac{1}{p}+\frac{1}{2p^2}+\log\left(1-\frac{1}{p}\right)\right).$$

The last sum sum above is the start of a rapidly converging series, so we easily find that

$$\frac{\varphi(N)}{N} > 0.029542. \tag{4.2}$$

The sum in (4.1) is

$$\sum_{d\in\mathcal{D}}\frac{1}{d} = S_3+S_4+S_5+S_{11}+S_{12}+S_{13} = 16.938967.$$

Thus, with (4.1) and (4.2), we have

$$\frac{|\mathcal{S}_{\mathcal{D}}|}{N} > (0.029542)(16.9389) > 0.5004.$$

This number N is very large, it is about $10^{1.09\times10^9}$. However, it is possible to reduce the exponents somewhat for the larger primes in N. Let N' be N divided by the 12th power of each prime dividing N that is above 10^6. Then $D(N') > 0.5003N'$ and N' is about $10^{1.61\times10^8}$. We have made some effort at finding a smaller example, say below 10^{10^8}, but we were not successful.

5. Densities and further problems

Let $u\in[0,1)$ be a real number and, as in the introduction, let $\mathcal{N}(u)$ denote the set of natural numbers n with $D(n)>u$. Since $D(mn)\ge D(n)$, it follows that if $n\in\mathcal{N}(u)$, so too is every multiple of n. Consequently $\mathcal{N}(u)$ has a logarithmic density $\delta(\mathcal{N}(u)) := \lim_{x\to\infty}\frac{1}{\log x}\sum_{k\in\mathcal{N}(u),\,k\le x}\frac{1}{k}$, see [1, 2], denote this by $\delta(u)$. We have by Corollary 2.2 that $\delta(u)>0$ for all $u\in[0,1)$. We can say a bit more.

Proposition 5.1. *We have* $\liminf_{n\to\infty} D(n) = 1/2$. *Consequently for* $0\le u<\frac{1}{2}$ *the set* $\mathcal{N}(u)$ *has both a logarithmic density* $\delta(u)$ *and a natural density* $d(u)$ *satisfying* $d(u)=\delta(u)=1$.

Proof. Let p be an odd prime and let a be a positive integer. The set of nonzero residues mod p^a which are the product of a power of p and a quadratic nonresidue mod p is product-free, and this shows that $D(p^a)\to\frac{1}{2}$ as $a\to\infty$ (recall that $D(n)<1/2$ if $n/\mathrm{rad}(n)$ does not have at least 6 distinct prime factors). In addition, the set of nonzero residues mod 2^a which are the product of a power of 2 and an integer that is 3 mod 4 is product-free, so that $D(2^a) \to \frac{1}{2}$ as $a \to \infty$. Since $D(p) \to \frac{1}{2}$ as $p \to \infty$ through the primes, it follows that $D(q) \to \frac{1}{2}$ as $q \to \infty$ through the prime powers (which include the primes). Hence for each real number $\epsilon > 0$, there are at most finitely many prime powers q with $D(q) \le \frac{1}{2} - \epsilon$. Thus, if $D(n) \le \frac{1}{2} - \epsilon$, it follows that each prime power dividing n must come from this set, forcing the set of such n to be finite as well. This proves the first statement in the proposition. Let $u \in [0, 1/2)$. By what we just proved, the set $\mathcal{N}(u)$ consists of all but finitely many natural numbers. This establishes the second statement in the proposition. □

It follows from the principal results of [9] that $\delta(1/2) \le 1.56 \times 10^{-8}$, and so with Proposition 5.1 it follows that $\delta(u)$ is not continuous in the variable u at 1/2. From the numerical example in the last section, we have $\delta(1/2) > 10^{-1.62 \times 10^8}$. There is of course an enormous (multiplicative) gap between these two bounds for $\delta(1/2)$.

More generally Theorem 2.1 yields a lower bound for $\delta(u)$ as $u \to 1^-$. Setting $\alpha_0 := (1 - \frac{1}{2}\mathrm{e}\log 2)^{-1} \approx 17.26659$, we have

$$\delta(u) > 1/\exp\exp\left((C/(1-u))^{\alpha_0}\right). \tag{5.1}$$

Note that (2.3) allows a slight improvement in this estimate.

It seems likely that for each u, the set $\mathcal{N}(u)$ has an asymptotic density $d(\mathcal{N}(u))$. General facts about asymptotic densities give $\underline{d}(\mathcal{N}(u)) \le \delta(u) \le \overline{d}(\mathcal{N}(u))$, and a natural density $d(u) = \delta(u)$ exists for those values with $\underline{d}(\mathcal{N}(u)) = \overline{d}(\mathcal{N}(u))$. Our proofs show that $\underline{d}(\mathcal{N}(u)) > 0$ for $0 < u < 1$ and $\overline{d}(\mathcal{N}(u)) < 1$ for $u \ge \frac{1}{2}$.

As asked in [9], is it true that for $u \ge 1/2$, the "primitive" members of $\mathcal{N}(u)$ (namely, they are not divisible by any other member of $\mathcal{N}(u)$) are all squarefull? If so, then it would follow that the asymptotic density of $\mathcal{N}(u)$ exists for each value of u.

Acknowledgments. We thank Rosa Orellana for a helpful discussion concerning [8]. Part of this work was done while the three authors visited MSRI, as part of the semester program Arithmetic Statistics. They thank MSRI for support, funded through the NSF. The first author was supported in part by grants from the Göran Gustafsson Foundation, the Knut and Alice Wallenberg foundation, and the Swedish Research Council. The second author was supported in part by NSF grant DMS-0801029. The third author was supported in part by NSF grant DMS-1001180.

References

[1] H. Davenport and P. Erdős, On sequences of positive integers, *Acta Arith.* **2** (1936), 147–151.

[2] H. Davenport and P. Erdős, On sequences of positive integers, *J. Indian Math. Soc. (NS)* **15** (1951), 19–24.

[3] P. Erdős, On sequences of integers no one of which divides the product of two others and on some related problems, *Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk* **2** (1938), 74–82.

[4] B. Green and I. Z. Ruzsa, Sum-free sets in abelian groups, *Israel J. Math.* **147** (2005), 157–188.

[5] L. Hajdu, A. Schinzel, and M. Skalba, Multiplicative properties of sets of positive integers, *Arch. Math. (Basel)* **93** (2009), 269–276.

[6] R. R. Hall and G. Tenenbaum, Divisors, Cambridge University Press, 1988.

[7] P. Kurlberg, J. C. Lagarias and C. Pomerance, On the maximal density of product-free sets in $\mathbb{Z}/n\mathbb{Z}$, *Int. Math. Res. Not. IMRN*, to appear.

[8] I. G. Macdonald, Symmetric functions and Hall polynomials, 2nd ed., Oxford University Press, Oxford, 1995.

[9] C. Pomerance and A. Schinzel, Multiplicative properties of sets of residues, *Moscow J. Combinatorics and Number Theory* **1** (2011), 52–66.

Department of Mathematics, KTH, SE-10044, Stockholm, Sweden
E-mail address: kurlberg@math.kth.se

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
E-mail address: lagarias@umich.edu

Mathematics Department, Dartmouth College, Hanover, NH 03755, USA
E-mail address: carl.pomerance@dartmouth.edu