Sums and products

Carl Pomerance, Dartmouth College
Hanover, New Hampshire, USA

ASU Mathematics Colloquium
February 27, 2014

Let's begin with products. Take the $N \times N$ multiplication table. It has N^2 entries. It is a symmetric matrix, so most entries appear at least twice. How many distinct entries does it have?

Let $M(N)$ be the number of distinct entries in the $N \times N$ multiplication table.

$\times$	1	2	3	4	5
1	**1**	**2**	**3**	**4**	**5**
2	2	4	**6**	**8**	**10**
3	3	6	**9**	**12**	**15**
4	4	8	12	**16**	**20**
5	5	10	15	20	**25**

So, $M(5) = 14$.

×	1	2	3	4	5	6	7	8	9	10
1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	2	4	6	8	10	**12**	**14**	**16**	**18**	**20**
3	3	6	9	12	**15**	18	**21**	**24**	**27**	**30**
4	4	8	12	16	20	24	**28**	**32**	**36**	**40**
5	5	10	15	20	**25**	30	**35**	40	**45**	**50**
6	6	12	18	24	30	36	**42**	**48**	**54**	**60**
7	7	14	21	28	35	42	**49**	**56**	**63**	**70**
8	8	16	24	32	40	48	56	**64**	**72**	**80**
9	9	18	27	36	45	54	63	72	**81**	**90**
10	10	20	30	40	50	60	70	80	90	**100**

So, $M(10) = 42$.

What would you conjecture about $M(N)$ asymptotically?

Maybe

$$\lim_{N\to\infty} \frac{M(N)}{N^2} = \frac{1}{3}?$$

Maybe

$$\lim_{N\to\infty} \frac{M(N)}{N^2} = c > 0?$$

Maybe

$$\lim_{N\to\infty} \frac{M(N)}{N^2} = 0?$$

Here are some values of $M(N)/N^2$ (Brent & Kung 1981):

N	$M(N)$	$M(N)/N^2$
1	1	1.0000
3	6	0.6667
7	25	0.5102
15	89	0.3956
31	339	0.3528
63	1237	0.3117
127	4646	0.2881
255	17577	0.2703
511	67591	0.2588
1023	258767	0.2473
2047	1004347	0.2397
4095	3902356	0.2327
8191	15202049	0.2266

And some more values (Brent & Kung 1981, Brent 2012):

N	$M(N)$	$M(N)/N^2$
$2^{14}-1$	59410556	0.2213
$2^{15}-1$	232483839	0.2165
$2^{16}-1$	911689011	0.2123
$2^{17}-1$	3581049039	0.2084
$2^{18}-1$	14081089287	0.2049
$2^{19}-1$	55439171530	0.2017
$2^{20}-1$	218457593222	0.1987
$2^{21}-1$	861617935050	0.1959
$2^{22}-1$	3400917861267	0.1933
$2^{23}-1$	13433148229638	0.1909
$2^{24}-1$	53092686926154	0.1886
$2^{25}-1$	209962593513291	0.1865

And some statistically sampled values (Brent & P 2012):

N	$M(N)/N^2$	N	$M(N)/N^2$
2^{30}	0.1774	2^{100000}	0.0348
2^{40}	0.1644	2^{200000}	0.0312
2^{50}	0.1552	2^{500000}	0.0269
2^{100}	0.1311	$2^{1000000}$	0.0240
2^{200}	0.1119	$2^{2000000}$	0.0216
2^{500}	0.0919	$2^{5000000}$	0.0186
2^{1000}	0.0798	$2^{10000000}$	0.0171
2^{2000}	0.0697	$2^{20000000}$	0.0153
2^{5000}	0.0586	$2^{50000000}$	0.0133
2^{10000}	0.0517	$2^{100000000}$	0.0122
2^{20000}	0.0457	$2^{200000000}$	0.0115
2^{50000}	0.0390	$2^{500000000}$	0.0095

It's fairly "clear" that $M(N) = o(N^2)$ as $N \to \infty$.

Do we have $M(N)$ of the shape N^{2-c_1}?
Of the ahape $N^2/(\log N)^{c_2}$?
Of the ahape $N^2/(\log\log N)^{c_3}$?

N	$M(N)/N^2$	c_1
2^{10}	0.2473	2.02×10^{-1}
2^{10^2}	0.1311	2.93×10^{-2}
2^{10^3}	0.0798	3.65×10^{-3}
2^{10^4}	0.0517	4.27×10^{-4}
2^{10^5}	0.0348	4.84×10^{-5}
2^{10^6}	0.0240	5.38×10^{-6}
2^{10^7}	0.0171	5.87×10^{-7}
2^{10^8}	0.0122	6.36×10^{-8}

Do we have $M(N)$ of the shape N^{2-c_1}?
Of the ahape $N^2/(\log N)^{c_2}$?
Of the ahape $N^2/(\log\log N)^{c_3}$?

N	$M(N)/N^2$	c_1	c_2
2^{10}	0.2473	2.02×10^{-1}	.887
2^{10^2}	0.1311	2.93×10^{-2}	.479
2^{10^3}	0.0798	3.65×10^{-3}	.387
2^{10^4}	0.0517	4.27×10^{-4}	.335
2^{10^5}	0.0348	4.84×10^{-5}	.301
2^{10^6}	0.0240	5.38×10^{-6}	.277
2^{10^7}	0.0171	5.87×10^{-7}	.258
2^{10^8}	0.0122	6.36×10^{-8}	.244

Do we have $M(N)$ of the shape N^{2-c_1}?
Of the ahape $N^2/(\log N)^{c_2}$?
Of the ahape $N^2/(\log\log N)^{c_3}$?

N	$M(N)/N^2$	c_1	c_2	c_3
2^{10}	0.2473	2.02×10^{-1}	.887	2.12
2^{10^2}	0.1311	2.93×10^{-2}	.479	1.41
2^{10^3}	0.0798	3.65×10^{-3}	.387	1.35
2^{10^4}	0.0517	4.27×10^{-4}	.335	1.36
2^{10^5}	0.0348	4.84×10^{-5}	.301	1.39
2^{10^6}	0.0240	5.38×10^{-6}	.277	1.44
2^{10^7}	0.0171	5.87×10^{-7}	.258	1.48
2^{10^8}	0.0122	6.36×10^{-8}	.244	1.52

Paul Erdős studied this problem in two papers, one in 1955, the other in 1960.

Paul Erdős, 1913–1996

In 1955, Erdős proved (in Hebrew) that $M(N)/N^2 \to 0$ as $N \to \infty$ and indicated that it was likely that $M(N)$ is of the shape $N^2/(\log N)^c$.

In 1960, at the prodding of Linnik and Vinogradov, Erdős identified (in Russian) the value of "c". Let

$$c = 1 - \frac{1 + \log\log 2}{\log 2} = 0.08607\ldots.$$

Then $M(N^2) = N^2/(\log N)^{c+o(1)}$ as $N \to \infty$.

In work of Tenenbaum progress was made (in French) in nailing down the "$o(1)$".

In 2008, Ford showed (in English) that $M(N)$ is of order of magnitude

$$\frac{N^2}{(\log N)^c (\log\log N)^{3/2}}.$$

No matter the language, we still don't know an asymptotic estimate for $M(N)$, despite this just being about the multiplication table!

So how can the fact that $M(N)$ is small compared to N^2 be explained?

It all comes down to the function $\Omega(n)$, the total number of prime factors of n, counted with multiplicity. For example,

$$\Omega(8) = 3,\ \Omega(9) = 2,\ \Omega(10) = 2,\ \Omega(11) = 1,\ \Omega(12) = 3.$$

Some higher values: $\Omega(1024) = 10$, $\Omega(1009) = 1$, and $\Omega(2^{17} - 1) = 1,\ \Omega(2^{17}) = 17$.

But what is $\Omega(n)$ *usually*? That is, can $\Omega(n)$ be approximately predicted from the size of n if we throw out thin sets like primes and powers of 2?

Indeed it can.

In 1917, Hardy and Ramanujan proved that the normal order of $\Omega(n)$ is $\log\log n$. That is, for each $\epsilon > 0$, the set of integers n with

$$\Big|\Omega(n) - \log\log n\Big| < \epsilon \log\log n$$

has asymptotic density 1.

So, this explains the multiplication table. Most products $n_1 n_2$ have both $n_1 > N^{1/2}$ and $n_2 > N^{1/2}$, and most of these have $\Omega(n_1)$ and $\Omega(n_2)$ fairly close to $\log\log N$ (note that $\log\log(N^{1/2})$ differs from $\log\log N$ by less than 1). So most of the products formed have about $2\log\log N$ prime factors, which is an unusual value to have for a number below N^2.

G. H. Hardy

S. Ramanujan

So, $\log\log N$ for integers below N is the center of the distribution. To quantify $M(N)$ one needs to know about estimates for the tail, and that's where the constant c arises.

I should take a small diversion from our progress here and mention one of the most beautfiful theorems in number theory, the Erdős–Kac theorem. It says that the "standard deviation" for $\Omega(n)$ for integers up to N is $(\log\log N)^{1/2}$ and that the distribution is Gaussian. Namely, for each real number u, the set

$$\left\{n : \Omega(n) \le \log\log n + u(\log\log n)^{1/2}\right\}$$

has asymptotic density equal to $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{u} e^{-t^2/2}\,dt$.

Einstein: “God does not play dice with the universe.”

Einstein: "God does not play dice with the universe."

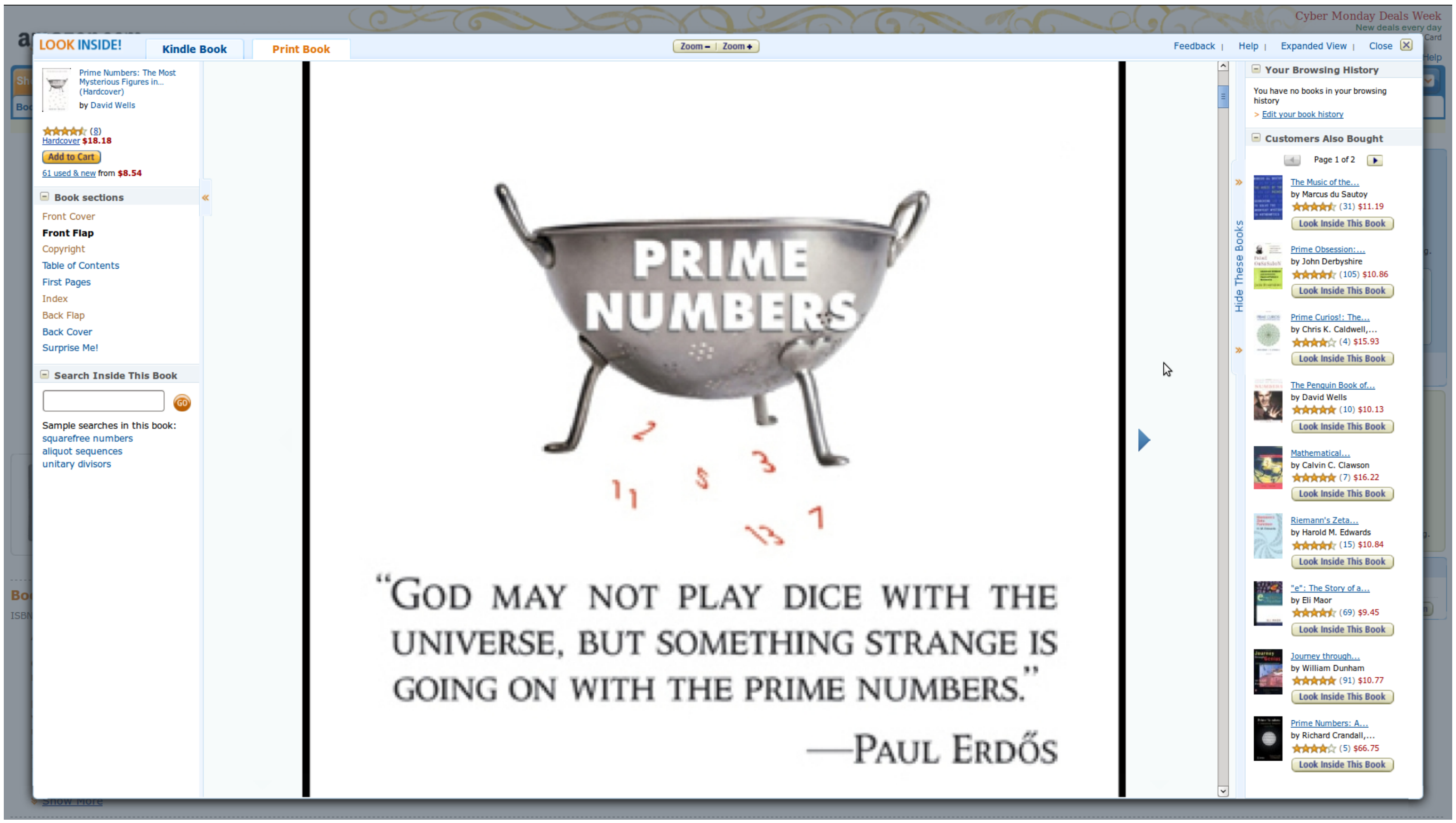
Erdős & Kac: Maybe so but something's going on with the primes.

Einstein: "God does not play dice with the universe."

Erdős & Kac: Maybe so but something's going on with the primes.

(Note: I made this up, it was a joke . . .)

Prime numbers, the most mysterious figures in math, D. Wells



Keeping with the theme of multiplication, what can be said about sets of positive integers that are *product-free*? This means that for any two members of the set, their product is not in the set. It is as far as you can get from being closed under multiplication.

It is easy to find such sets, for example the set of primes. But how dense can such a set be?

Consider the set

$$\Big\{n : \Omega(n) \text{ is odd}\Big\}.$$

This set is product-free and has asymptotic density $\frac{1}{2}$.

It's not clear if this is the best one can do, but at least there's a fairly simple proof that any product-free set must have upper asymptotic density strictly smaller than 1.

To make further progress, and perhaps to make things a little simpler, lets consider a periodic version of the problem.

Let $D(n)$ denote the maximal density of a product-free set that consists of residue classes modulo n.

For example, take the integers that are 2 (mod 3). The product of any two of them is 1 (mod 3), so is not in the set. And this set has asymptotic density $\frac{1}{3}$.

We have $D(3) = \frac{1}{3}$.

Can we do better with higher moduli?

Well, the set of integers that are 2 or 3 (mod 5) is product-free and has density $\frac{2}{5}$. That is $D(5) = \frac{2}{5}$.

For $n = 7$, we have $D(7) = \frac{3}{7}$. Namely, consider the classes 3, 5, or 6 (mod 7).

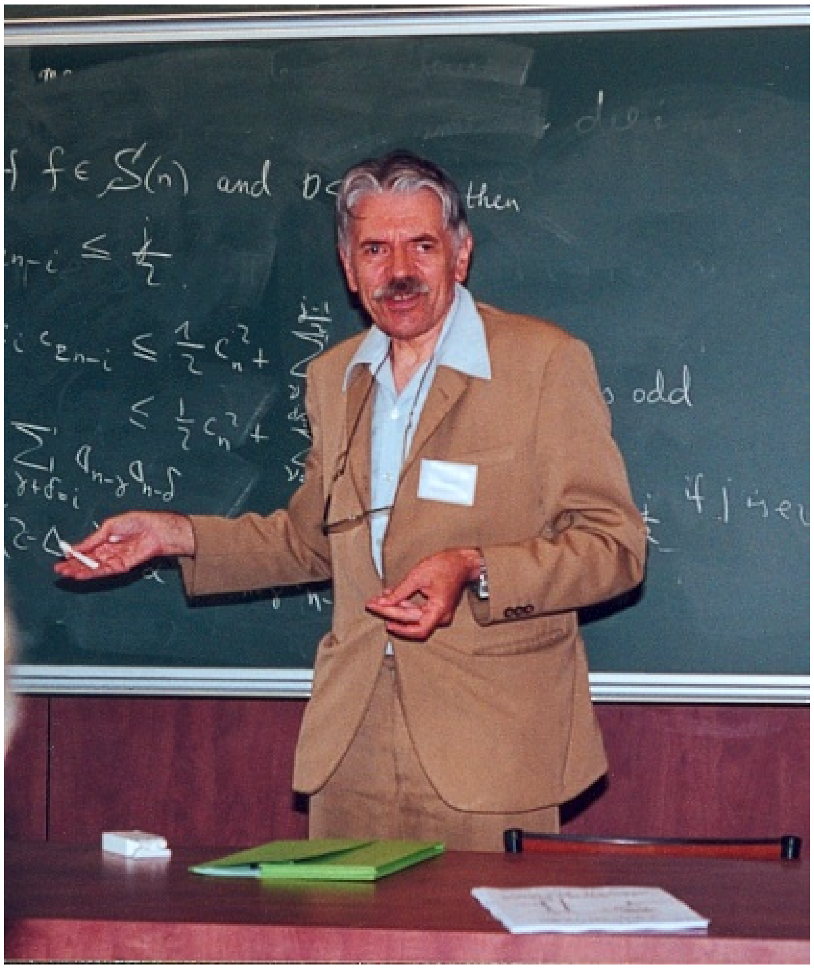
It is not hard to prove that $\liminf_{n \to \infty} D(n) = \frac{1}{2}$.

So, again we have met what seems to be some sort of boundary: $\frac{1}{2}$.

Do we have $D(n) < \frac{1}{2}$ for all n?

P, Schinzel (2011): *We have* $D(n) < \frac{1}{2}$ *for all* n *except possibly those* n *divisible by the square of a number with at least* 6 *distinct prime factors. Further, the asymptotic density of those* n *divisible by such a square is about* 1.56×10^{-8}.

Moscow Journal of Combinatorics and Number Theory,
1 (2011), 52–66.



Andrzej Schinzel

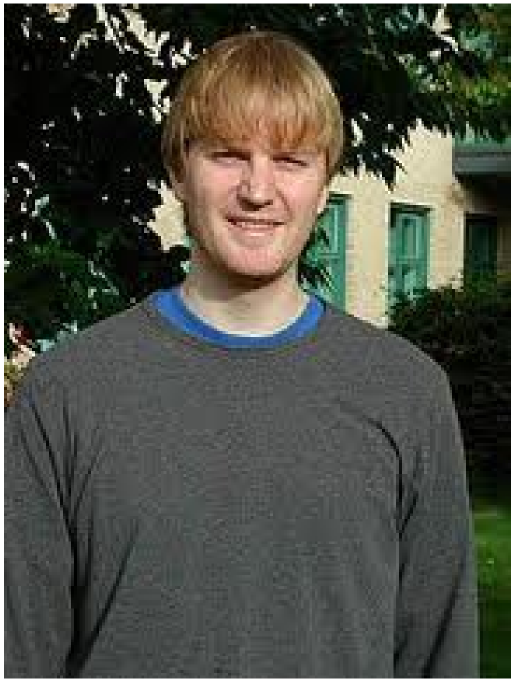
Surely that cements it, and $D(n) < \frac{1}{2}$ for all n, right?

Surely that cements it, and $D(n) < \frac{1}{2}$ for all n, right?

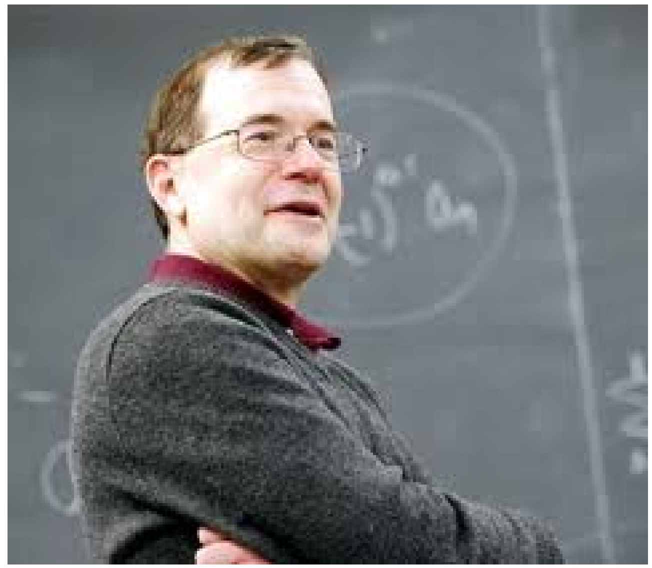
Well, no.

Kurlberg, Lagarias, P (2012): *There are infinitely many values of n with $D(n)$ arbitrarily close to 1. In particular, there are infinitely many values of n where all of the pairwise products of a subset of 99% of the residues $\pmod{n}$ all fall into the remaining 1% of the residue classes.*

Acta Arithmetica, in a special issue in honor of Andrzej Schinzel's 75th birthday.



Pär Kurlberg



Jeffrey C. Lagarias

Let's be more modest, just show me one n where $D(n) \geq \frac{1}{2}$.

It's not so easy!

Here's a number. Take the first 10,000,000 primes. For those primes below 1,000,000, take their 14th power, and for those that are larger, take their square, and then multiply these powers together to form N. Then $D(N) > 0.5003$. Further, $N \approx 10^{1.61 \times 10^8}$.

Can you find an example with fewer than 100,000,000 decimal digits?

What is behind this construction and proof?

It is actually very similar to the proof of the multiplication table theorem.

Suppose n is a high power of the product of all of the primes up to x, say the exponent is $\lfloor \log x \rfloor$. Then consider all residues $r \pmod{n}$ with

$$\frac{2}{3} \log \log x < \Omega(\gcd(r, n)) < \frac{4}{3} \log \log x.$$

Then these residues $r \pmod{n}$ form a product-free set, and in fact most residues $\pmod{n}$ satisfy this inequality.

Actually the numbers $\frac{2}{3}$ and $\frac{4}{3}$ are not optimal, but $\frac{\mathrm{e}}{4}$ and $\frac{\mathrm{e}}{2}$ are. Being especially careful with the estimates leads to the following result:

Kurlberg, Lagarias, P (2013): *There is a positive constant* c_1 *such that for infinitely many* n *we have*

$$D(n) > 1 - \frac{c_1}{(\log\log n)^{1-\frac{\mathrm{e}}{2}\log 2}(\log\log\log n)^{\frac{1}{2}}}.$$

Note that $1 - \frac{\mathrm{e}}{2}\log 2 = 0.0579153\ldots$.

This is optimal for our method of proof, but is this the optimal result? It turns out that yes, apart from the constant c_1, it is optimal:

Kurlberg, Lagarias, P (2013): *There is a positive constant* c_2 *such that for all* n *we have*

$$D(n) < 1 - \frac{c_2}{(\log\log n)^{1-\frac{e}{2}\log 2}(\log\log\log n)^{\frac{1}{2}}}.$$

The idea for this upper bound: use linear programming!

For a product-free set S in $\mathbb{Z}/n\mathbb{Z}$ and for $d \mid n$, let α_d be the proportion of those $s \in S$ with $\gcd(s, n) = d$ among all residues $r \pmod{n}$ with $\gcd(r, n) = d$.

Then each α_d is in $[0, 1]$.

Further, if $|S| \geq n/2$, then $\alpha_1 = 0$ and for all u, v with $uv \mid n$, we have

$$\alpha_u + \alpha_v + \alpha_{uv} \leq 2.$$

In some sense, $|S|/n$ is closely modeled by $\sum_{d|n} \alpha_d/d$.

So, the LP is to maximize $\sum_{d|n} \alpha_d/d$ given the above constraints.

Since we already know that $D(n)$ can be fairly large, we need not prove we have found the maximum of the LP, just some upper bound for it. It is known that any feasible solution to the *dual* LP gives an upper bound for the primary LP. Thus, we write down the dual LP, find a fairly trivial feasible solution, and then "shift mass" to make it better.

And, voilà, our upper bound for all n's tightly matches our constructed lower bound for champion n's.

Sated now with products, lets move on to sums . . .

No, we're not going to start with addition tables. The analogous problem is trivial, in the addition table for the integers from 1 to N there are precisely $2N - 1$ distinct sums.

But what about *sum-free* sets? Here we have a set of positive integers that contains none of the pairwise sums of its elements. How dense can such a set be?

This too is easy. The odd numbers form a sum-free set of asymptotic density $\frac{1}{2}$. And one cannot do better.

Here's the proof. Say $\mathcal{A}$ is a sum-free set of positive integers and $a \in \mathcal{A}$. Then the set $a + \mathcal{A}$ is disjoint from $\mathcal{A}$. If $\mathcal{A}$ has $N = N(x)$ members in $[1, x]$, then $a + \mathcal{A}$ has $N + O(1)$ numbers here, so $x \geq 2N + O(1)$. Hence for all x, we have $N(x) \leq \frac{1}{2}x + O(1)$. We conclude that the upper density of a sum-free set $\mathcal{A}$ of positive integers is at most $\frac{1}{2}$.

Let us look at a somewhat more subtle problem. How dense can a sum-free subset of $\mathbb{Z}/n\mathbb{Z}$ be?

If n is even, then take the odd residues, and this is best possible.

But what if n is odd?

Diananda & Yap (1969), **Green & Ruzsa** (2005):
If n is solely divisible by primes that are $1 \pmod 3$*, then the maximal density of a sum-free set in* $\mathbb{Z}/n\mathbb{Z}$ *is* $\frac{1}{3} - \frac{1}{3n}$*. If n is divisible by some prime that is* $2 \pmod 3$*, then the maximal density of a sum-free set in* $\mathbb{Z}/n\mathbb{Z}$ *is* $\frac{1}{3} + \frac{1}{3p}$*, where p is the least such prime. Otherwise, the maximal density of a sum-free set in* $\mathbb{Z}/n\mathbb{Z}$ *is* $\frac{1}{3}$.

This problem has been considered in general finite abelian groups and also for non-abelian groups. A survey article by recent Jeopardy contestant Kiran Kedlaya:
Product-free subsets of groups, then and now, Communicating mathematics, 169–177, Contemp. Math., **479**, Amer. Math. Soc., Providence, RI, 2009.

After hearing a shorter version of this talk a couple of years ago, several graduate students asked me the following question: What if you consider both sums *and* products?

Well, there is a famous and seminal problem here in which the Erdős multiplication-table theorem plays a role:

Among all sets $\mathcal{A}$ of N positive integers what is the minimum value of $|\mathcal{A}+\mathcal{A}|+|\mathcal{A}\cdot\mathcal{A}|$?

If one takes $\mathcal{A}=\{1,2,\ldots,N\}$, then $|\mathcal{A}+\mathcal{A}|=2N-1$ and $|\mathcal{A}\cdot\mathcal{A}|=N^2/(\log N)^{c+o(1)}$, so for large N,

$$|\mathcal{A}+\mathcal{A}|+|\mathcal{A}\cdot\mathcal{A}|>N^{2-\epsilon}.$$

If on the other hand we take $\mathcal{A} = \{1, 2, \ldots, 2^{N-1}\}$, then $|\mathcal{A} \cdot \mathcal{A}| = 2N - 1$ and $|\mathcal{A} + \mathcal{A}| = \frac{1}{2}N^2 + \frac{1}{2}N$, so that again

$$|\mathcal{A} + \mathcal{A}| + |\mathcal{A} \cdot \mathcal{A}| > N^{2-\epsilon}. \qquad (1)$$

Erdős & Szemerédi asked in 1983: Is (1) true for any set $\mathcal{A}$ of N positive integers?

There has been a parade of results getting better and better lower bounds, with game players being the posers Erdős & Szemerédi, then Nathanson, Chen, Elekes, Bourgain, Chang, Konyagin, Green, Tao, Solymosi, ...

Seeing a couple of Fields medalists in this list, with the problem still not solved, is a bit daunting!

But what the grad students asked was about dense sets $\mathcal{A}$ that are simultaneously sum-free and product-free.

For example, take the numbers that are 2 or 3 (mod 5). It is a set of asymptotic density $\frac{2}{5}$ and is both sum-free and product-free. We cannot do better than $\frac{1}{2}$ for the density (considering only the sum-free property), but can we beat $\frac{2}{5}$ for both sum-free and product-free?

Kurlberg, Lagarias, P (2012): *Say $\mathcal{A}$ is sum-free and product-free with upper density $D(\mathcal{A})$.*

1. *If $\mathcal{A} \subset \mathbb{Z}_{>0}$ with least element a, then $D(\mathcal{A}) \le \frac{1}{2}\left(1 - \frac{1}{5a}\right)$.*

2. *There is a constant $\kappa_1 > 0$, such that if $\mathcal{A} \subset \mathbb{Z}/n\mathbb{Z}$, then*

$$D(\mathcal{A}) \le \frac{1}{2} - \frac{\kappa_1}{(\log\log n)^{1-\frac{e}{2}\log 2}(\log\log\log n)^{\frac{1}{2}}}.$$

3. *There is a constant κ_2 and infinitely many n such that for some $\mathcal{A} \subset \mathbb{Z}/n\mathbb{Z}$,*

$$D(\mathcal{A}) \ge \frac{1}{2} - \frac{\kappa_2}{(\log\log n)^{1-\frac{e}{2}\log 2}(\log\log\log n)^{\frac{1}{2}}}.$$

There remains a numerical problem: find an example of a number n and a sum-free, product-free subset $\mathcal{A}$ of $\mathbb{Z}/n\mathbb{Z}$ with $\frac{|\mathcal{A}|}{n} > \frac{2}{5}$.

For $\frac{|\mathcal{A}|}{n} = \frac{2}{5}$, we have $n = 5$. Back-of-the-envelope calculations suggest that there is some n that beats $n = 5$ around

$$10^{10^{500,000}},$$

a number so large that not only can't we write the number in decimal notation, we can't even write the number of its digits in decimal notation.

But we haven't looked at this problem too closely and there may be a much more modest example.

Thank You!