Statistics in elementary number theory

Carl Pomerance, Dartmouth College
Hanover, New Hampshire, USA

CRM Workshop on
Statistics and Number Theory
15–19 September, 2014

In this talk we will consider 4 statistical problems in elementary number theory. With all of these problems, there remain substantial open questions.

The problems:

- Let $M(N)$ denote the number of distinct entries in the $N \times N$ multiplication table. How does $M(N)$ grow as $N \to \infty$?

- For random numbers j, n with $1 \leq j \leq \sqrt{n}$, how likely is it for the number of divisors of n lying in $[1, j]$ to be even?

- Let $s(n) = \sigma(n) - n$, be the sum of the divisors of n other than n. How likely is it for a random number to be a value of s, and how does this go for evens and for odds or other residue classes?

- Two numbers m, n are said to form an amicable pair if $s(m) = n$ and $s(n) = m$. What can one say about the distribution of numbers in an amicable pair, and in particular, what about their sum of reciprocals?

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, $M(5) = 14$.

$\times$	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 shape $N^2/(\log N)^{c_2}$?
Of the shape $N^2/(\log\log N)^{c_3}$?

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 shape $N^2/(\log N)^{c_2}$?
Of the shape $N^2/(\log\log N)^{c_3}$?

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 shape $N^2/(\log N)^{c_2}$?
Of the shape $N^2/(\log\log N)^{c_3}$?

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

Both the c_2 and c_3 columns seem hopeful, maybe it is a blend of the two? Here we take the statistically sampled data and try to fit $M(N)$ as $N^2/(\log N)^{c_2}(\log\log N)^{c_3}$.

N	$M(N)/N^2$	c_2	c_3
2^{10}	0.2473		
2^{10^2}	0.1311	-162.7	479.0
2^{10^3}	0.0798	$-.1107$	1.7318
2^{10^4}	0.0517	.02242	1.2680
2^{10^5}	0.0348	.05966	1.1170
2^{10^6}	0.0240	.07670	1.0382
2^{10^7}	0.0171	.07555	1.0441
2^{10^8}	0.0122	.08948	.9645

Long before such extensive calculations existed, 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 beautiful 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

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

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

We now move on to the second problem:

For random numbers j, n with $1 \leq j \leq \sqrt{n}$, how likely is it for the number of divisors of n lying in $[1, j]$ to be even?

Doing a thought experiment, one might guess it's 50-50. List the divisors of n that are at most $\sqrt{n}$:

$$1 = d_1 < d_2 < \cdots < d_k \leq \sqrt{n} < d_{k+1}.$$

(Note that for n not a square, $\tau(n)$, the total number of divisors of n, is even and, for such n, we have $k = \frac{1}{2}\tau(n)$.)

Surely picking a random number j in the interval $[1, \sqrt{n}]$ and finding $j \in [d_i, d_{i+1})$, wouldn't skew i towards being more often odd than even, or vice versa?

Actually, it is skewed in favor of evens, and this is fairly easy to see from the multiplication table theorem.

Suppose that $d_k < \frac{1}{10}\sqrt{n}$. For almost all n, we will have k even, since we may assume that n is not of the form pm^2 with p prime (such numbers n have asymptotic density 0). Thus, at least 90% of the values of $j \leq \sqrt{n}$ will fall in an even interval.

Good, but how likely is it for $d_k < \frac{1}{10}\sqrt{n}$? If not, then $n = d_k d_{k+1}$, with $d_{k+1} \leq 10\sqrt{n}$, so with $N = \lceil 10\sqrt{n} \rceil$, we have n as an entry in the $N \times N$ multiplication table. So n must come from a sparse set.

And we can repeat this with "10" replaced with any fixed number.

Let's take a different angle with this problem. *Fix* j and let $\tau_j(n)$ be the number of divisors of n in $[1, j]$. Let Δ_j be the asymptotic density of the set of n with $\tau_j(n)$ odd.

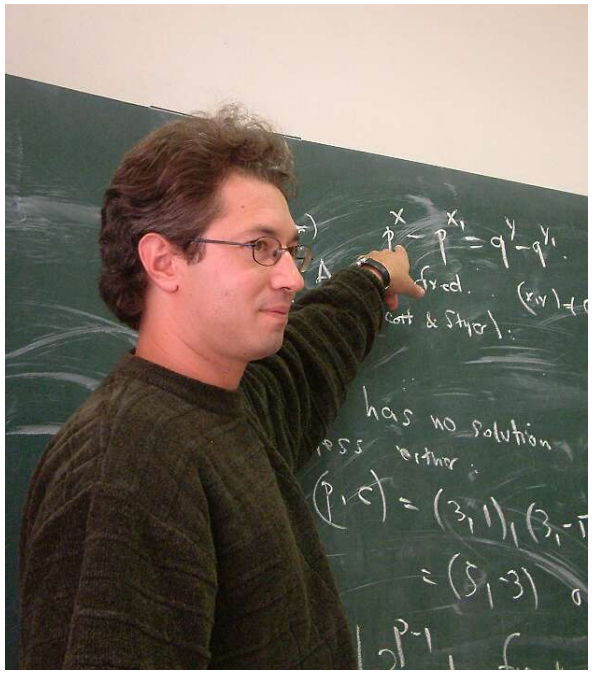
Note that $\tau_j(n)$ depends only on $\gcd(n, L_j)$, where L_j is the least common multiple of $1, 2, \ldots, j$. So, the density Δ_j exists.

How does Δ_j behave asymptotically?

We could begin by looking at some values of Δ_j. It's easy to compute

$$\Delta_1 = 1, \quad \Delta_2 = \frac{1}{2}, \quad \Delta_3 = \frac{1}{2}, \quad \Delta_4 = \frac{7}{12}.$$

j	L_j	$\#\{n \le L_j : \tau_j(n) \text{ odd}\}$	Δ_j
$5, 6$	60	33	0.55
7	420	225	0.5357142857
8	840	405	0.4821428571
9	2520	1305	0.5178571429
10	2520	1235	0.4900793651
$11, 12$	27720	13635	0.4918831169
13	360360	177705	0.4931318681
14	360360	170775	0.4739010989
15	360360	170181	0.4722527473
16	720720	359073	0.4982142857
17	12252240	6106815	0.4984243697
18	12252240	5919705	0.4831528765
19	232792560	112887225	0.4849262580
20	232792560	109706355	0.4712622903
21	232792560	110362725	0.4740818392



Florian Luca

Jeffrey Shallit

Recently (July, 2014):

Luca, P, & Shallit — *As* $j \to \infty$, $\Delta_j \to 0$. *In fact,*

$$\Delta_j = O\left(\frac{1}{(\log j)^{c/(1+c)}(\log\log j)^{1.5/(1+c)}}\right),$$

where $c = 1 - \dfrac{1+\log\log 2}{\log 2}$ *is the Erdős–Ford–Tenenbaum constant.*

It is easy to see that $\Delta_j > K/\log j$ for some positive constant K, but we have no idea what the "correct" order of magnitude of Δ_j is.

Odd counts for the 10,000 numbers following the $k \times 10^5$-th prime.

j	10^5	2×10^5	3×10^5	4×10^5	5×10^5	6×10^5
100	4131	4121	4077	4099	4123	4109
200	4061	4107	4174	4181	4231	4050
300	3800	3850	3954	3980	4002	3969
400	3630	3703	3800	3744	3877	3875
500	3466	3587	3673	3710	3793	3772
600	3351	3512	3526	3594	3722	3682
700	3294	3435	3502	3543	3627	3593
800	3213	3301	3431	3475	3577	3574
900	2822	3245	3337	3411	3522	3477
1000	2358	3197	3248	3334	3459	3439

The idea behind the proof: Suppose n has a small prime factor p, with $p^2 \nmid n$. (Think $n \equiv 2 \pmod 4$.) For j large, very few such numbers n will have a divisor in the interval $(j, pj]$. Let

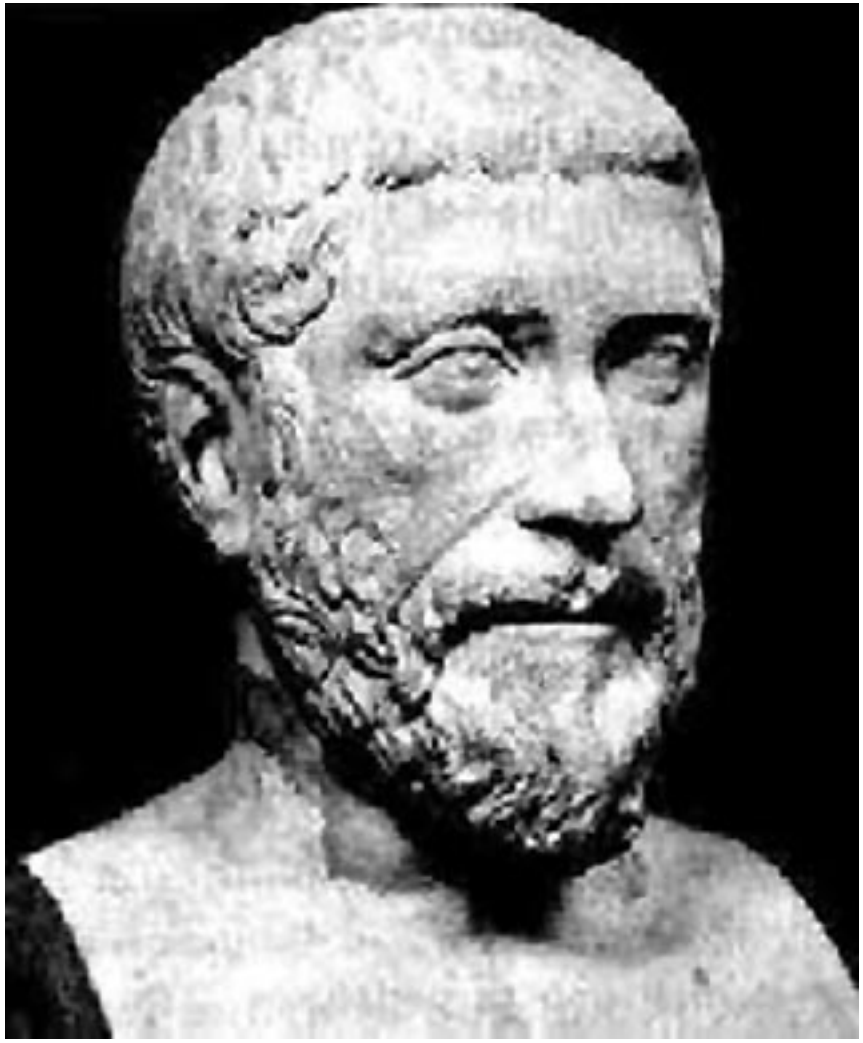
$$\mathcal{A} = \{d \mid n \,:\, d \le j,\, p \nmid d\}, \quad \mathcal{B} = \{d \mid n \,:\, d \le j,\, p \mid d\}.$$

Then $d \in \mathcal{A}$ if and only if $pd \in \mathcal{B}$, so $\#\mathcal{A} = \#\mathcal{B}$.

Hence $\tau_j(n) = \#\mathcal{A} + \#\mathcal{B}$ is even.

One might ask about other residue classes for $\tau_j(n)$. Our proof shows that for each fixed integer k, the set of n with $k \mid \tau(n)$ and $k \nmid \tau_j(n)$ has density $o(1)$ as $j \to \infty$.

For k not a power of 2, the density of the set of n with $k \nmid \tau(n)$ is positive. Within this set we don't know how likely it is for $k \mid \tau_j(n)$ as j grows. Perhaps some numerical experiments would help here.



Pythagoras, ca. 500 B.C.E.

Beyond his triangles, Pythagoras defined what is perhaps the first function of mathematics: $s(n)$, the sum of the proper divisors of n. We have

$$s(n) = \sigma(n) - n = \sum_{\substack{d \,|\, n \\ d<n}} d.$$

He discovered that $s(6) = 6$ and that

$$s(220) = 284, \quad s(284) = 220.$$

So, as soon as he had the first function, he had the first dynamical system!

The **Catalan–Dickson** conjecture: *There are no unbounded orbits* $n \to s(n) \to s(s(n)) \to \ldots$.

The **Guy–Selfridge** counter conjecture: *For asymptotically all even numbers* n*, the orbit starting with* n *is unbounded.*

The first number in doubt is $n = 276$, there are 4 others below 1000, known as the "Lehmer 5" .

The sociable-numbers conjecture: *Numbers involved in a cycle (sociable numbers) have asymptotic density* 0.

Kobayashi–Pollack–P (2009): The upper density is at most about 0.002. (For cycles of bounded length, Erdős proved the conjecture.)

These are hard questions.

So let's consider an easy (?) question: Which odd numbers are in the range of s, and which even numbers?

Note that if p, q are different primes, then $s(pq) = p + q + 1$. Since presumably all even numbers starting with 8 are the sum of two different primes (a slightly stronger form of Goldbach's conjecture), it would follow that all odd numbers starting with 9 are in the range of s.

Also, $s(2^k) = 2^k - 1$, so $1, 3, 7$ are also in the range of s.

Conjecture: *All odd numbers except for* 5 *are in the range of* s.

Theorem: *Asymptotically all odd numbers are in the range of* s.

(van der Corput, Chudakov, Estermann, Vaughan, Montgomery–Vaughan, Chen–Pan, Li, Pintz, Lu,. . .)

What about even values of s?

Erdős (1973): *A positive proportion of even numbers are not of the form* $s(n)$.

Chen & Zhao (2011) *This proportion is at least* 0.06.

P & Yang (2014): Let $U(x)$ denote the number of integers to x not of the form $s(n)$.

x	$U(x)/x$
10^5	0.138630
10^6	0.150232
10^7	0.157497
10^8	0.162469

A synopsis of the Erdős proof that $s(n)$ misses a positive proportion of the even numbers:

But for a negligible set of even numbers m, we may assume that if $s(n) = m$ for some n, then n is even. Indeed, otherwise n is an odd square, say k^2. If k is composite, then $s(k^2) \geq k^{3/2}$, so if $s(k^2) = m \leq x$, then $k \leq x^{2/3}$. On the other hand, if k is prime, then $s(k^2) \geq k$, and so $k \leq x$. Either way, there are only $o(x)$ possibilities.

Now assume that $s(n) = m$ and n is even. Then $s(n) \geq \frac{1}{2}n$, so that $n \leq 2x$. For any fixed integer u, almost all $n \leq 2x$ have $u \mid n$ if and only if $u \mid s(n)$. So, consider numbers $m \leq x$ divisible by 12. If $12 \mid n$, then $s(n) \geq \frac{4}{3}n$, so $n \leq \frac{3}{4}x$. But the number of $n \leq \frac{3}{4}x$ divisible by 12 is about $\frac{1}{16}x$, and so at least $(\frac{1}{12} - \frac{1}{16} + o(1))x = (\frac{1}{48} + o(1))x$ multiples of 12 up to x go untouched by the function $s(n)$.

But what of even numbers that *are* of the form $s(n)$?

Luca & P (June, 2014): *There is a positive proportion of even numbers in the range of* s.

This mostly would follow from:

The **Erdős, Granville, P, & Spiro** conjecture (1990): *If* $\mathcal{A}$ *is a set of asymptotic density* 0, *then* $s^{-1}(\mathcal{A})$ *has asymptotic density* 0.

Maybe our proof will shed some light on the conjecture. This conjecture would imply that for each positive integer k, $s^{(k)}(\mathbb{N})$ does not have density 0. (Our proof gives this for $k = 2$.)

We can also prove rigorously another consequence of the conjecture:

Luca & P (2014): *In any residue class* $a \pmod b$, *there is a positive proportion of numbers of the form* $s(n)$.

We have discussed 3 of the 4 problems. The 4th: What can we say about the distribution of the amicable numbers?

Recall that n is amicable if there is some $m \neq n$ with $s(m) = n,\ s(n) = m$.

About 12 million pairs are known, so about 24 million amicable numbers.

Erdős (1955): *The amicable numbers have asymptotic density zero.*

P (1981): *The number of amicable numbers in* $[1, x]$ *is at most* $x/\exp((\log x)^{1/3})$.

P (February 2014): *It's at most* $x/\exp((\log x)^{1/2})$.

From these latter results, it follows that

$$\sum_{n \text{ amicable}} \frac{1}{n} < \infty.$$

Can we numerically approximate this sum?

Using the known amicable numbers, the sum is at least 0.011984 . . . .

What about an upper bound?

Bayless & Klyve (2011): *The reciprocal sum is* $< 6.56 \times 10^8$

Hanh Nguyen (June, 2014): *The reciprocal sum is* < 4084.

One of the innovations in Nguyen's proof was not to follow so closely the distribution results. For example, the set of n with at least $4 \log\log n$ distinct prime factors has a fairly easily upper-bounded reciprocal sum, but the distribution of such numbers to x is way bigger than $x/\exp((\log x)^{1/2})$. The latter doesn't matter, it's the small reciprocal sum that's important. So, after this is done, many other parts of the argument simplify nicely.

MERCI, THANK YOU