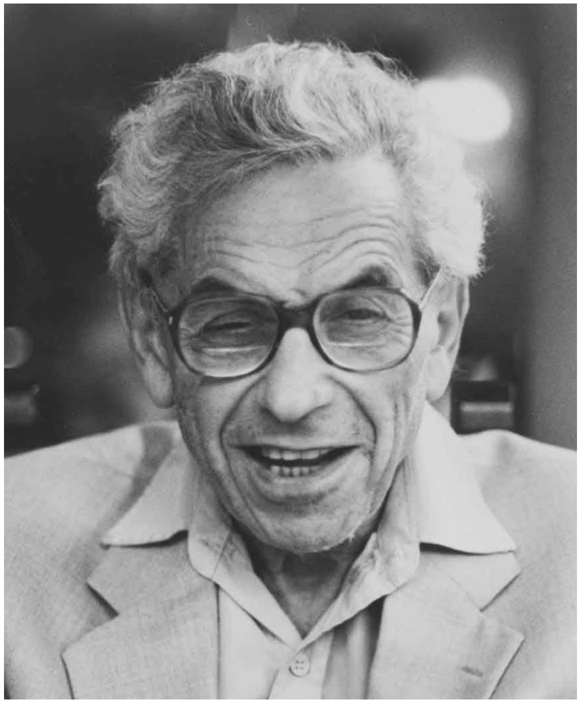
Some statistical problems concerning the arithmetic functions σ and φ

Carl Pomerance, Dartmouth College

based on joint work with

A. Anavi, F. Luca, P. Pollack, E. Treviño, and H.-S. Yang

Being the centennial of **Paul Erdős**, this talk is dedicated to him. He more than anyone else taught me how to think statistically about elementary number theory.



Paul Erdős, 1913–1996

The field of statistics concerns itself with intelligently gathering data, and making sense of it in ways that might lead us to valid inferences about underlying truths.

In number theory, our data might be ranks, orders, zeros, discriminants, etc., but it seems this statistical way of thinking all began centuries ago by studying how the prime numbers are distributed within the natural numbers.

Euclid proved there are infinitely many primes, and **Euler** showed that their reciprocal sum to x grows like $\log\log x$, which indicates that their counting function to x should grow like $x/\log x$. **Gauss** came to an even finer version of this conjecture after poring over tables of primes. Though proved at one level as the prime number theorem, we are still struggling to prove this law at a deeper level; this is the Riemann Hypothesis.

Though fundamental of course, prime numbers are not the only fish in the sea of integers. In fact, number theory began millennia ago with the study of other special numbers.

Perhaps the very first function ever described in mathematics, done so by **Pythagoras**, is $s(n)$. This takes a positive integer n and sends it to the sum of the positive divisors of n that are smaller than n. It is not clear why Pythagoras was interested in this function, but he noted that 220 and 284 form an *amicable pair* in that

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

Centuries after Pythagoras, but still 2300 years ago, Euclid went on to discuss solutions to $s(n) = n$, the *perfect numbers*. Others in antiquity discussed the *deficient numbers* ($s(n) < n$) and the *abundant numbers* ($s(n) > n$).

What is a modern number theorist to make of these problems? A fruitful avenue is to think statistically.

For example, **Davenport** proved in 1934 that not only do the deficient numbers and abundant numbers each possess a positive asymptotic density, but more generally, for each fixed real $u \geq 0$, the set

$$\{n : s(n) < un\}$$

has an asymptotic density that varies continuously with u, and is strictly increasing. In particular, the perfect numbers have asymptotic density 0.

Davenport's theorem, which was based on an earlier result of **Schoenberg** on Euler's function φ, was later given an ultimate generalization by **Erdős & Wintner**.

This talk will not survey the entire field of statistical problems in elementary number theory, but rather a small handful of papers in the last year or so in which I have participated. For a broader, albeit Erdős-centric view, see

P. Pollack and C. Pomerance, Paul Erdős and the rise of statistical thinking in elementary number theory, submitted.

All of the new papers mentioned in this talk are available on my home page; these slides are there too.

In 1932, **D. H. Lehmer** asked if there are any composite numbers n with $\varphi(n) \mid n-1$. Here φ is Euler's function, and it is clear that $\varphi(n) \mid n-1$ for $n=1$ and n prime. We still don't know if there are any such n that are composite.

How then can we do statistics if there are no data?

We can try and prove that there are not very many such numbers.

The main idea that has been used in this regard is that if d is a divisor of such a number n, then n is in a residue class modulo $d\varphi(d)$. Further, some d exists near $\sqrt{n}$. Using these and some other Erdős-type tricks, I showed in 1977 that the number of such numbers in $[1, x]$ is at most $x^{1/2}(\log x)^c$.

Since then, there have been a series of papers lowering the exponent c on $\log x$, by Z. Shan (1985), W. D. Banks & F. Luca (2007), and W. D. Banks, A. D. Güloglu, & C. W. Nevans (2008).

The record result so far is in

F. Luca and C. Pomerance, On composite integers n for which $\varphi(n) \mid n-1$, *Bol. Soc. Mat. Mex.*, **17** (2011), 13–21,

where it is proved that the number of such n in $[1, x]$ is at most $x^{1/2}(\log x)^{-1/2+o(1)}$ as $x \to \infty$.

I wrote in 1977 that it would seem to take a new idea to reduce the exponent $1/2$ on x; we're still waiting for this! Conjecturally the count should be bounded by $(\log x)^{O(1)}$, if not something smaller.

The problem of counting those n with $\varphi(n) \mid n-1$ can be generalized to those n with $\varphi(n) \mid n-a$, with a a fixed integer. Sometimes there is a "trivial" infinite family of solutions, such as the primes with $a=1$, or the numbers $6p$ with $p>3$ prime and $a=6$, but these are easily classified.

The above papers all hold for this more general problem of counting "sporadic" solutions to $\varphi(n) \mid n-a$.

One can also generalize to other arithmetic functions, perhaps the most prominent being $\sigma(n)$, the sum-of-divisors function. We have $\sigma(n) = s(n) + n$, where s is the function of Pythagoras mentioned at the start of the talk. We like σ better because it is multiplicative.

Here we are counting numbers n in $[1, x]$ with

$$\sigma(n) \equiv a \pmod{n}.$$

After removing "trivial" solutions, such as $6p$ when $a = 12$, the number of sporadic solutions is also bounded as with the φ version of the theorem, namely by $x^{1/2+o(1)}$. The proof uses the same overall scheme as in the 1977 paper, but with some new elements, mostly dealing with the *radical* of a number (the largest squarefree divisor).

This is in

A. Anavi, P. Pollack, and C. Pomerance, On congruences of the form $\sigma(n) \equiv a \pmod{n}$, *Int. J. Number Th.* **9** (2012), 115–124.

One can also ask about solutions n to $\sigma(n) \equiv a \pmod{n}$ with $\sigma(n)$ odd. For example, **Chowla** suggested the function $s'(n) = \sigma(n) - n - 1$ could be of interest, and **Cattaneo** defined a number to be *quasiperfect* if $s'(n) = n$ (so that $\sigma(n) = 2n + 1$). It is trivial to see that for any such n, we have $\sigma(n)$ odd.

Since $\sigma(n)$ is odd if and only if n is either a square or twice a square, we automatically have the count up to x being $O(x^{1/2})$. In the very recent paper

P. Pollack and C. Pomerance, On the distribution of some integers related to perfect and amicable numbers, submitted,

it is shown that the number of such n up to x is at most $x^{1/4+o(1)}$ as $x \to \infty$.

In 1973, Erdős showed that a positive proportion of numbers are not of the form $s(n)$, where s is the sum-of-proper-divisors function that was considered by Pythagoras. It was known already from the 1930s that almost all odd numbers are in the image of s. It is still not known if the image of s has an asymptotic density or if it contains a positive proportion of even numbers.

One can ask this type of question for other arithmetic functions; that is, what are the statistics for the image?

Erdős showed in 1935 that the number of values of φ or of σ in $[1, x]$ is $x/(\log x)^{1+o(1)}$ as $x \to \infty$. Much work has been done in trying to pin down the "$o(1)$" here, the current champion being **K. Ford**, but we still don't have an asymptotic formula for the count.

In his thesis from 1976, **H. J. J. te Riele** asked if the result on $s(n)$ could be extended to $s^*(n)$. This is the sum of those divisors d of n with $1 \leq d < n$ and $(d, n/d) = 1$. In the recent paper

C. Pomerance and H.-S. Yang, Variant of a theorem of Erdős on the sum-of-proper-divisors function, *Math. Comp.*, to appear,

we show that a positive proportion of even numbers are missing from the image of $s^*(n)$ and that a positive proportion of even numbers are present. (It is easy to see as with $s(n)$ that almost all odd numbers are in the image.) This paper will be presented in a session tomorrow afternoon by the second-named author.

This paper with Yang actually has some real data, and is not just an asymptotic estimate for the distribution of an interesting set of numbers. Based on our data, we conjecture that the image of s has asymptotic density close to $\frac{5}{6}$ and that the image of s^* has asymptotic density close to $\frac{99}{100}$.

For our last problem, consider the function $\varphi(n)$ of Euler. This function is far from being monotone. But here are two questions:

How long can an interval be where φ is monotone?

How large may a subset of $[1, x]$ be on which φ is monotone?

These questions are addressed in the recent paper

P. Pollack, C. Pomerance, and E. Treviño, Sets of monotonicity for Euler's totient function, *Ramanujan J.*, to appear.

We give a fairly precise answer to the first question.

Let γ be the Euler–Mascheroni constant and let

$$\alpha = \sum_{p \text{ prime}} \frac{1}{p} \log \frac{p}{p-1}.$$

Then the longest interval in $[1, x]$ on which φ is increasing has length

$$\frac{\log_3 x}{\log_6 x} + (\alpha - \gamma + o(1)) \frac{\log_3 x}{(\log_6 x)^2}$$

as $x \to \infty$. Here $\log_k$ indicates iteration. The number $\alpha - \gamma$ is .0028428289 .... This result also holds for φ decreasing on an interval. The proof is modeled after a paper of Erdős from 1958 that considers intervals on which φ does not vary appreciably.

It is easy to find a large subset of $[1, x]$ on which φ is nondecreasing, namely the primes. After some numerical experiments it seems that the size of the champion set is always at most $\pi(x) + 64$, where π is the prime-counting function. We cannot prove that the excess above $\pi(x)$ is bounded, but we did prove that the count is at most $x/(\log x)^{1+o(1)}$ as $x \to \infty$. Is the sum of reciprocals of the numbers in a champion such set $\log\log x + O(1)$?

For nonincreasing, we showed the size of the largest set in $[1, x]$ is greater than $x^{0.7}$ and smaller than $x/\exp((\frac{1}{2} + o(1))\sqrt{\log x \log\log x})$. The exponent 0.7 comes from a recent paper of **Baker & Harman** from which it can be shown that there are $x^{0.7}$ integers in $[1, x]$ at which φ is constant. Is it true that the size for nonincreasing is at most twice, say, the size for constant? We don't know.

THANK YOU