The Sieve of Eratosthenes and Rough Numbers

Carl Pomerance, Dartmouth College

This edition of the West Coast Number Theory Conference is dedicated to the memory of long-time attendee,

John Brillhart, 1930–2022.

I have many memories of John, some good, some difficult, but I am sad that he is no longer with us. In checking the Internet for photos of John, it seems strange that there are not very many of them, given his prominence. I did find this one.

Oxford Number Theory Conference, 1969

Most numbers n have a fairly large prime factor, say $> n^{\epsilon}$, and most have a fairly small prime factor, say $< 1/\epsilon$.

At one extreme we have the *smooth* (or *friable*) numbers with no large prime factors. And at the other extreme we have the *rough* (maybe *anti-friable*?) numbers with no small prime factors.

Smooth numbers are useful in the analysis of various factoring and discrete log algorithms, such as the **Brillhart–Morrison** continued fraction factoring algorithm.

Rough numbers have been viewed algorithmically as the uncanceled numbers in the sieve of **Eratosthenes**.

This talk discusses the distribution of rough numbers. In particular, let $\Phi(x,y)$ denote the number of integers $n \le x$ with no prime factors $\le y$.

For example, $\Phi(10,2) = 5$, since it is precisely the odd numbers that have no prime factors ≤ 2. And

$$\Phi(100,10) = 22,$$

since after removing numbers with a prime factor ≤ 10 from $[1,100]$, what is left is 1 and the 21 primes in $(10,100]$.

For y fixed and $x \to \infty$, inclusion–exclusion allows us to get an asymptotic formula:

$$\Phi(x,y) = \sum_{P(d)\le y} \mu(d)\left\lfloor \frac{x}{d}\right\rfloor \sim x\prod_{p\le y}\left(1-\frac{1}{p}\right), \quad x\to\infty,$$

where $P(d)$ is the largest prime factor of d and p runs over primes. But what if y is not fixed?

Using a well-known theorem of **Mertens** we have

$$\prod_{p\le y}\left(1-\frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log y}, \quad y\to\infty,$$

where γ is **Euler**'s constant ($e^{-\gamma}\approx .561$).

Can we put both asymptotics together?

Let's try it when $y = \sqrt{x}$. Might we have

$$\Phi(x, \sqrt{x}) \sim x \prod_{p \le \sqrt{x}} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma} x}{\log(\sqrt{x})}, \qquad x \to \infty?$$

Well, the left side is $\pi(x) - \pi(\sqrt{x}) + 1 \sim x/\log x$ by the Prime Number Theorem. And the right side is $(2e^{-\gamma})x/\log x$, and $2e^{-\gamma} \approx 1.123$, so no, this is **not** correct.

There is a bit of a mystery here. When does the simple sieve given by inclusion-exclusion fade away to something deeper with the distribution of primes?

Prominent here is the theorem of **Buchstab**:

$$\Phi(x,y) \sim \omega(u)x/\log y, \qquad x \geq y \to \infty.$$

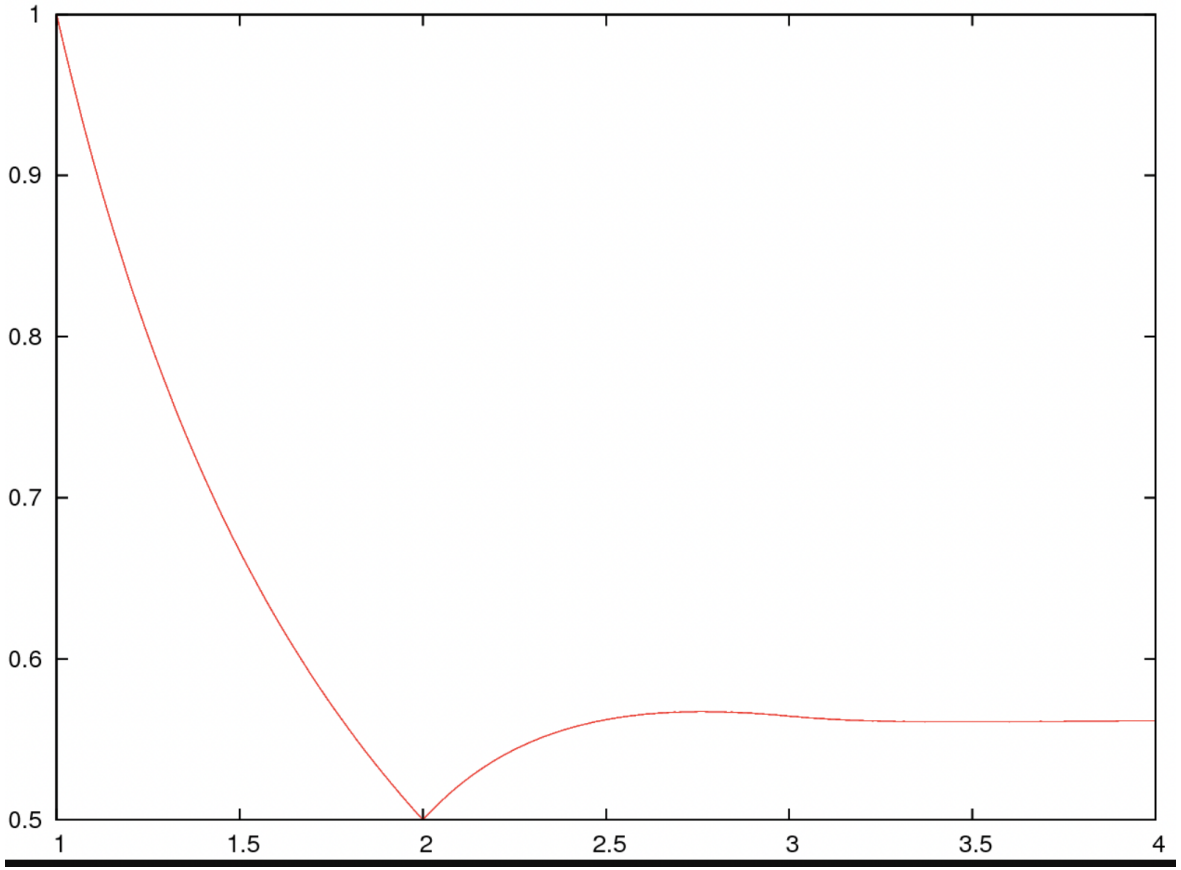
Here $u = \log x/\log y$ and $\omega(u)$ is the **Buchstab** function. Like the **Dickman–de Bruijn** ρ-function, ω is defined via a differential delay equation:

$$\begin{aligned} u\omega(u) &= 1, \quad 1 \leq u \leq 2, \\ (u\omega(u))' &= \omega(u-1), \quad u > 2, \end{aligned}$$

with ω continuous.

One can show that $1/2 \leq \omega(u) \leq 1$ for all $u \geq 1$ and that $\omega(u) \to e^{-\gamma} \approx .561$ as $u \to \infty$.

The **Buchstab** function, $1 \leq u \leq 4$, (from Wikipedia)



It's difficult to see since $\omega(u)$ approaches its horizontal asymptote of $e^{-\gamma}$ so rapidly, but it actually oscillates above and below this number infinitely often with a crossing in each interval framed by consecutive integers.

In one of the great theorems in analytic number theory in my lifetime, **Maier** used the oscillations of the **Buchstab** function to show that for each fixed $\kappa > 1$, there is a number $c_\kappa > 0$ such that the sets

$$\{x : \pi(x + (\log x)^\kappa) - \pi(x) > (1 + c_\kappa)(\log x)^{\kappa-1}\},$$
$$\{x : \pi(x + (\log x)^\kappa) - \pi(x) < (1 - c_\kappa)(\log x)^{\kappa-1}\}$$

are both unbounded.

While a great theorem, what I particularly like are results that are numerically explicit. For example, while we know from the Prime Number Theorem that

$$p_n \sim n \log n, \quad n \to \infty,$$

where p_n is the n-th prime, we have the wonderful inequality of **Rosser**:

$$p_n > n \log n, \quad n \geq 1.$$

Recently **Ford** asked me if we had a good numerically explicit upper bound for $\Phi(x, y)$. I passed this problem on to **Steve Fan**, a grad student here, and he proved:

$$\Phi(x, y) \leq \frac{x}{\log y}, \quad x \geq y \geq 2.$$

It is clear that **Fan**'s inequality $\Phi(x,y) \le x/\log y$ is best possible, since if $y = x^{1-\epsilon}$ then $\Phi(x,y) \sim x/\log x = (1-\epsilon)x/\log y$.

His principal tool in proving this was the large sieve of **Montgomery** and **Vaughan**.

When $\sqrt{x} \le y \le x$, $\Phi(x,y)$ counts precisely 1 and the primes in $(y,x]$. It might be nice to have an inequality when y is smaller. So we teamed up to prove the following.

Fan & P (2022): For $3 \le y \le \sqrt{x}$, we have $\Phi(x,y) < 0.6x/\log y$.

To prove this we consider various ranges for y. First, for $y < 71$ we do a complete inclusion-exclusion, necessarily checking some small cases. And for $71 \le y < 241$ we do a modified inclusion-exclusion using the Bonferroni inequalities and a pre-sieve with the primes 2, 3, 5. Each new prime involves checking values of x in an increasingly long interval. Once this gets larger than 3×10^7, we move to a new plan.

For larger values of y we use a numerically explicit version of **Selberg**'s sieve. Helping were not only the usual calculations of **Rosser & Schoenfeld** and **Dusart**, but some newer work of **Büthe**. From his work we have

$$\pi(x) < \mathrm{li}(x), \quad x \le 10^{19}.$$

Using these tools we establish our inequality when $u \geq 7.5$ ($u = \log x / \log y$) using separate arguments when $y \leq 500{,}000$ and $y > 500{,}000$.

Next, we deal with the case $2 \leq u < 3$, with separate arguments for y below and above 1100. We show a somewhat stronger inequality in these ranges, which is useful, since we bootstrap from these small values of u to $u = 7.5$ losing a little in each step. But we nevertheless complete the proof getting our result that $\Phi(x, y) \leq 0.6x / \log y$.

Back to smooth numbers: Let $\Psi(x,y)$ denote the number of integers $n \le x$ with all prime factors $\le y$. We know that $\Psi(x,y) \sim \rho(u)x$ in a large range, where ρ is the **Dickman–de Bruijn** function: $\rho(u)' = -\rho(u-1)/u$ for $u > 1$, with initial condition $\rho(u) = 1$ when $u \le 1$. Some years ago I conjectured that

$$\Psi(x,y) > \rho(u)x, \quad u = \log x/\log y, \quad x/2 \ge y \ge 2.$$

Very recent work of **Gorodetsky** proves essentially this under assumption of an error term in the Prime Number Theorem that is a bit stronger than the Riemann Hypothesis.

Thank you