SIEVE METHODS

BAILEY WHITBREAD

Contents

1. A Prototypical Sieve

The earliest example of a sieve is called the *Sieve of Eratosthene's*. [1] Let $z > 1$ be an integer and consider the integers on the interval $I = [z, z^2)$. That is, consider $S = I \cap \mathbb{Z}$. We strike out all of the multiples of 2 from S. Then we strike out all of the multiples of 3 from S, then all the multiples of 5, and so on. We do this for all of the primes less than z.

We can conclude that the integers that have survived this process must be prime. To see this, if n is a composite integer that has not been striked out, it must have at least two prime factors, p_1 and p_2. Since n survives, these prime factors satisfy $p_1, p_2 \geq z$. Thus $n \geq p_1 p_2 \geq z^2$. Thus n does not lie in the interval $I = [z, z^2)$.

We illustrate this idea with an example. Consider $z = 11$ so that $[z, z^2) = [11, 121)$. Then $[11, 121) \cap \mathbb{Z} = \{11, 12, ..., 120\}$. We first strike out all of the multiples of 2 by colouring them red. We colour the multiples of 3 orange, the multiples of 5 yellow and the multiples of 7 green.

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120

Then we can conclude that the numbers not striked out (i.e. the uncoloured numbers) are prime. This leaves us with the list of primes

$$\{11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113\}.$$

There's still one question unanswered: why can we stop at the prime 7? If we considered the prime 11, then we would strike off $11 \times 2, 11 \times 3, ..., 11 \times 11$. Notice that 11×11 will be the first new number striken off the list when we consider the prime 11. However $11 \times 11 = 121 \notin [z, z^2) \cap \mathbb{Z}$. Thus considering the prime 11, or any larger prime, is of no use.

2. The Framework of Sieve Theory

Fix a finite set $\mathcal{A} \subset \mathbb{Z}$, a set of primes $\mathcal{P}$, and an integer $z > 1$. We call $\mathcal{A}$ the *sieving set*, the set $\mathcal{P}$ the *sieve range* and z the *sieve level*. Then define

$$P(z) := \prod_{\substack{p \in \mathcal{P} \\ p < z}} p.$$

Then sieve theory is concerned with computing, or estimating, the value

$$S(\mathcal{A}, \mathcal{P}, z) := |\{a \in \mathcal{A} : (a, P(z)) = 1\}| = \sum_{(a, P(z))=1} 1.$$

where (a, b) is shorthand for $\gcd(a, b)$. Then $S(\mathcal{A}, \mathcal{P}, z)$ counts the elements in the sieving set, $\mathcal{A}$, that are indivisible by all of the primes in the sieve range, $\mathcal{P}$, up to the sieve level, z. If $(a, P(z)) > 1$, we can say that our sieve *sifts* a from $\mathcal{A}$. This allows us to interpret the value $S(\mathcal{A}, \mathcal{P}, z)$ as the number of elements that are *unsifted* by our sieve.

The Sieve of Eratosthenes is obtained when we fix some $N \in \mathbb{N}$, then take $\mathcal{A} = \{n \in \mathbb{N} : n \leq N\}$, $\mathcal{P}$ to be the set of all primes, and $z = \lfloor N^{1/2} + 1 \rfloor$.

3. A Weak Prime Number Theorem

This value $S(\mathcal{A}, \mathcal{P}, z)$ can be used to write a rudimentary proof of a weak form of the prime number theorem (PNT), which we recall now.

Theorem 1 (PNT)**.** *Recall the prime counting function $\pi(x) := |\{p \text{ prime} : p \leq x\}|$. Then*

$$\pi(x) \sim \frac{x}{\log x}.$$

We formulate a *weak prime number theorem* from [2] as follows.

Theorem 2 (Weak PNT)**.**

$$\pi(x) < \frac{x}{\log \log x} + \mathcal{O}(N^{\log 2}),$$

where we take $f(x) = \mathcal{O}(g(x))$ to mean that there exists constants $C > 0$ and x_0 such that

$$|f(x)| \leq Cg(x),$$

for all $x > x_0$.

To prove the weak PNT, we require a lemma.

Lemma 3. *If $f : \mathbb{N} \to \mathbb{R}$ is a completely multiplicative function (i.e. $f(ab) = f(a)f(b)$ and $f(1) = 1$), then there holds*

$$\sum_{d|n} \mu(d) f(d) = \prod_{p|n} (1 - f(p)).$$

Proof of Lemma 3. If $n = 1$ then the lemma is clear. Let $n > 1$ and write $n = p_1^{e_1} p_2^{e_2} ... p_r^{e_r}$ to be the unique prime decomposition of n. Also let $N = p_1 p_2 ... p_r$ be the product of the prime divisors of n. Then we compute

$$\sum_{d|n} \mu(d) f(d) = \sum_{d|N} \mu(d) f(d) = \sum_{I \subseteq [r]} (-1)^{|I|} f\Big(\prod_{i \in I} p_i\Big) = \sum_{I \subseteq [r]} \Big(\prod_{i \in I} -f(p_i)\Big).$$

Then notice that

$$\sum_{I \subseteq [r]} \Big(\prod_{i \in I} -f(p_i)\Big) = (1 - f(p_1))(1 - f(p_2)) \cdot ... \cdot (1 - f(p_r)) = \prod_{p|N} (1 - f(p)) = \prod_{p|n} (1 - f(p)).$$

□

Corollary 4. *In particular,*

$$\sum_{d|n} \frac{\mu(d)}{d} = \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Proof of Corollary 4. Take $f(x) = \frac{1}{x}$ in Lemma 3. □

Proof (of weak PNT). Fix $N \in \mathbb{N}$. Consider the sieving set $\mathcal{A} = \{n \in \mathbb{N} : n \leq N\}$, the sieve range $\mathcal{P} = \{p : p \text{ prime}\}$, and the sieve level $z = \log N$. Let d be a square-free positive integer. Then write $\mathcal{A}_d := \{a \in \mathcal{A} : d|a\}$ which allows us to compute

$$S(\mathcal{A}, \mathcal{P}, z) = \sum_{\substack{a \in \mathcal{A} \\ (a, P(z))=1}} 1 = \sum_{a \in \mathcal{A}} \sum_{\substack{d|a \\ d|P(z)}} \mu(d) = \sum_{d|P(z)} \mu(d) \sum_{\substack{a \in \mathcal{A} \\ d|a}} 1 = \sum_{d|P(z)} \mu(d)|\mathcal{A}_d|.$$

In particular, we have

$$|\mathcal{A}_d| = \sum_{\substack{n \leq N \\ d|n}} 1 = \left\lfloor \frac{N}{d} \right\rfloor = \frac{N}{d} - \left\{ \frac{N}{d} \right\}.$$

Thus

$$S(\mathcal{A}, \mathcal{P}, z) = \sum_{d|P(z)} \mu(d)|\mathcal{A}_d| = N \sum_{d|P(z)} \frac{\mu(d)}{d} - \sum_{d|P(z)} \mu(d) \left\{ \frac{N}{d} \right\}.$$

The prior corollary with $n = P(z)$ tells us that

$$\sum_{d|P(z)} \frac{\mu(d)}{d} = \prod_{p|P(z)} \left(1 - \frac{1}{p}\right) = \prod_{p<z} \left(1 - \frac{1}{p}\right).$$

Then we can write

$$S(\mathcal{A}, \mathcal{P}, z) = N \prod_{p<z} \left(1 - \frac{1}{p}\right) - \sum_{d|P(z)} \mu(d) \left\{ \frac{N}{d} \right\} = N \prod_{p<z} \left(1 - \frac{1}{p}\right) + R,$$

where $R := -\sum_{d|P(z)} \mu(d)\{\frac{N}{d}\}$ serves as a remainder term that we will estimate the size of now.

Notice that $\mu(d) \in \{-1, 0, 1\}$ and $\{\frac{N}{d}\} \in [0, 1)$. This tells us that $\mu(d)\{\frac{N}{d}\} = \mathcal{O}(1)$. Also observe that if $\{2, ..., z\}$ is the set of primes less than or equal to z then this set has $2^{\pi(z)}$ subsets. This tells us that there are $2^{\pi(z)}$ square-free divisors of $P(z)$. This is because each divisor is uniquely associated to a subset of $\{2, ..., z\}$, where we associate a subset with the product of its elements. This lets us write

$$R = -\sum_{d|P(z)} \mu(d) \left\{ \frac{N}{d} \right\} = \sum_{d|P(z)} \mathcal{O}(1) = \mathcal{O}(2^{\pi(z)}).$$

Now notice that, since $z = \log N$, we have

$$2^{\pi(z)} = 2^{\pi(\log N)} \leq 2^{\log N} = e^{\log N \log 2} = N^{\log 2},$$

where we observe that $\pi(x) \leq x$. Thus $R = \mathcal{O}(2^{\pi(z)}) = \mathcal{O}(N^{\log 2})$.

We can also estimate $S(\mathcal{A}, \mathcal{P}, z) \geq 1 + \pi(N) - \pi(z) \geq \pi(N) - z$. The first inequality comes from noticing that $S(\mathcal{A}, \mathcal{P}, z)$ will definitely count 1 and at least all of the primes between z and N, which is $\pi(N) - \pi(z)$. Then we can write

$$\pi(N) \leq z + S(\mathcal{A}, \mathcal{P}, z) = \log N + N \prod_{p<z} \left(1 - \frac{1}{p}\right) + \mathcal{O}(N^{\log 2}) = N \prod_{p<z} \left(1 - \frac{1}{p}\right) + \mathcal{O}(N^{\log 2}),$$

where the $\log N$ term is absorbed into $\mathcal{O}(N^{\log 2})$ since the function $N \mapsto \log N$ grows slower than the function $N \mapsto N^{\log 2}$ (see '*big-$\mathcal{O}$ notation*' for more details).

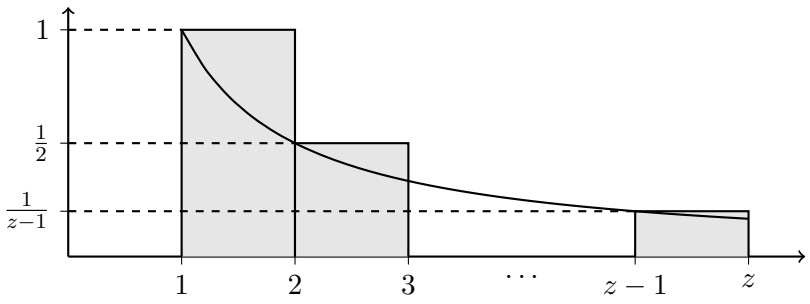
Now we'll estimate $\prod_{p<z}(1-\frac{1}{p})$. Notice that

$$(\star) \qquad \prod_{p<z}\left(1-\frac{1}{p}\right)^{-1} = \prod_{p<z}\sum_{m\geq 0}\frac{1}{p^m} \geq \sum_{n<z}\frac{1}{n} > \int_1^z \frac{1}{x}\,\mathrm{d}x = \log z.$$

To see the first equality of $(\star)$, notice that $(1-1/p)^{-1} = \frac{1}{1-1/p} = \sum_{m\geq 0}(1/p)^m$. To see the first inequality of $(\star)$, let p_z be the largest prime satisfying $p_z < z$. Notice that

$$\begin{aligned}\prod_{p<z}\sum_{m\geq 0}\frac{1}{p^m} &= \left(\frac{1}{2^0}+\frac{1}{2^1}+...\right)\left(\frac{1}{3^0}+\frac{1}{3^1}+...\right)\cdot...\cdot\left(\frac{1}{p_z^0}+\frac{1}{p_z^1}+...\right)\\ &\geq \sum_{n=p_1^{e_1}p_2^{e_2}...p_z^{e_z}}\frac{1}{n}\\ &\geq \sum_{n<z}\frac{1}{n},\end{aligned}$$

where the second last sum is over all n with prime decomposition only containing primes less than or equal to p_z. To see the second inequality of $(\star)$, notice that $\sum_{n<z}\frac{1}{n}$ corresponds to the upper Riemann sums of the function $f(x)=\frac{1}{x}$ on the interval $[1,z]$ with partition $P=\{1,2,3,...,z\}$. Then $U(f,P)=\sum_{n<z}\frac{1}{n}$. The inequality can be visualised below.



This lets us conclude that

$$\prod_{p<z}\left(1-\frac{1}{p}\right) < \frac{1}{\log z} = \frac{1}{\log\log N}.$$

We put this altogether and see that

$$\pi(N) \leq N\prod_{p<z}\left(1-\frac{1}{p}\right) + \mathcal{O}(N^{\log 2}) < \frac{N}{\log\log N} + \mathcal{O}(N^{\log 2}).$$

□

4. Prime Gaps and the Twin Prime Conjecture

Define p_n to be n^{th} prime number. Then we say the n^{th} *prime gap* is the number $p_{n+1}-p_n$. The first prime gap is $3-2=1$. However, after this, since all primes greater than 2 are odd, the prime gaps are at least 2. In other words, $p_{n+1}-p_n \geq 2$ for $n>1$.

It is natural to ask the question: "Is $p_{n+1}-p_n = 2$ infinitely often?" This question is equivalent to asking: "Is $\liminf_{n\to\infty} p_{n+1}-p_n = 2$?" This is the well-known *twin prime conjecture*. We detail the progress made towards proving this conjecture and the tools used in doing so.

In May, 2013, the mathematician Yitang Zhang published his paper *Bounded gaps between primes*. [3] This established the first finite bound on the value $\liminf_{n\to\infty} p_{n+1} - p_n$. He managed to prove that $\liminf_{n\to\infty} p_{n+1} - p_n < 7 \times 10^7$. This tells us that there are infinitely many prime pairs, p_{n+1} and p_n, with $p_{n+1} - p_n < 7 \times 10^7$.

Zhang's result is of particular importance due to his background. [4] After completing his PhD, he struggled to find an academic position. This led to him spending several years working outside of academia. He worked as a delivery worker and an accountant for some years. In 1999, he was hired by the University of New Hampshire. There he worked on his prime gaps result, which was published in the Annals of Mathematics.

Following this paper, a large amount of progress was made on the twin prime conjecture. A group of mathematicians formed the online collaborative group, the *Polymath project*. Together, they refined Zhang's result week by week. They publish the result that $\liminf_{n\to\infty} p_{n+1} - p_n \leq 4680$ in July 2013. [5]

Independently of the Polymath project, a post-doctoral researcher in the United Kingdom, James Maynard, was also looking at Zhang's paper. He noticed a large amount of Zhang's argument could be skipped, allowing him to publish the result that $\liminf_{n\to\infty} p_{n+1} - p_n \leq 600$. Together, James Maynard and the Polymath project refined this to $\liminf_{n\to\infty} p_{n+1} - p_n \leq 246$. This is where the result stands today.

5. The Elliot–Halberstam Conjecture

Progress on the twin prime conjecture has been made using the *Elliot–Halberstam conjecture*. However, the twin prime conjecture has not been proven in full. To state the Elliot–Halberstam conjecture, we must understand some results closely related to the prime number theorem. [6]

Theorem 5 (PNT (von Mangoldt))**.** *If $\Lambda : \mathbb{N} \to [0, \infty)$ is the von Mangoldt function, defined by*

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^k \text{ for some } k \geq 1, \\ 0, & \text{else.} \end{cases}$$

Then

$$\sum_{n\leq x} \Lambda(n) \sim x.$$

This statement is equivalent to the prime number theorem. To see this, we observe the inequality

$$(1-\varepsilon)(\pi(x) + \mathcal{O}(x^{1-\varepsilon})) \log x \leq \sum_{n\leq x} \Lambda(n) \leq \pi(x) \log x,$$

for any $0 < \varepsilon < 1$. We fix ε and divide through by x. Then send $x \to \infty$ and $\varepsilon \to 0$. Then $\lim_{x\to\infty} \frac{1}{x} \sum_{n\leq x} \Lambda(n) = 1$ and $\sum_{n\leq x} \Lambda(n) \sim x$.

We can also write a prime number theorem for arithmetic progressions.

Theorem 6 (PNT for Arithmetic Progressions)**.** *If $(a, q) = 1$ then let $\pi(x; q, a)$ be the number of primes less than or equal to x that are congruent to a modulo q. Then*

$$\pi(x; q, a) \sim \frac{1}{\phi(q)} \frac{x}{\log x}.$$

Notice the similarity between the regular PNT and the PNT for arithmetic progressions. Similarly, we can write a PNT for arithmetic progressions using the von Mangoldt function.

Theorem 7 (PNT for Arithmetic Progressions (von Mangoldt))**.** *If $(a, q) = 1$ then*

$$\sum_{\substack{n\leq x \\ n\equiv a \bmod q}} \Lambda(n) \sim \frac{1}{\phi(q)} \sum_{n\leq x} \Lambda(n).$$

Now we're ready to state the Elliot–Halberstam conjecture. [7]

Conjecture 8 (Elliot–Halberstam)**.** *Fix $0 < \theta < 1$. We say the Elliot–Halberstam conjecture holds at level θ, and write $EH[\theta]$, if*

$$\sum_{q \leq x^\theta} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \left| \sum_{\substack{n \leq x \\ n \equiv a \bmod q}} \Lambda(n) - \frac{1}{\phi(q)} \sum_{n \leq x} \Lambda(n) \right| << \frac{x}{\log^A x},$$

for all $A > 0$.

To understand this conjecture, we first look at the innermost term,

$$\left| \sum_{\substack{n \leq x \\ n \equiv a \bmod q}} \Lambda(n) - \frac{1}{\phi(q)} \sum_{n \leq x} \Lambda(n) \right|.$$

Recall that the PNT for arithmetic progressions written in terms of the von Mangoldt function is only an approximation. The term above is the error in this approximation. Then we consider the worst approximation by varying a and obtaining

$$\sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \left| \sum_{\substack{n \leq x \\ n \equiv a \bmod q}} \Lambda(n) - \frac{1}{\phi(q)} \sum_{n \leq x} \Lambda(n) \right|.$$

We recall that $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ if and only if $(a, q) = 1$. We must obey the condition $(a, q) = 1$ and we could have written $\sup_{(a,q)=1}$ but instead we write $\sup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times}$ to emphasise that a is varying.

Then we consider the sum over all $q \leq x^\theta$, where x^θ is a fixed value. This is performing a type of averaging for us. Overall, we're looking at the error in the prime number theorem for arithmetic progressions, then we're taking the worst error amoung the equivalence classes modulo q, and then we're averaging these worst errors amoung all choices of $q \leq x^\theta$.

Finally we notice the term $\frac{x}{\log^A x}$ is similar to the term $\frac{x}{\log x}$ in the prime number theorem. The Elliot–Halberstam conjecture is concerned with when we can make an improvement on the regular prime number theorem.

The Elliot–Halberstam conjecture has been proven for $0 < \theta < \frac{1}{2}$, a result known as the *Bombieri–Vinogradov theorem.* [8] It is known that if the Elliot–Halberstam conjecture is true for $0 < \theta < 1$ then $\liminf_{n \to \infty} p_{n+1} - p_n \leq 12$, a substantial improvement on the current known estimates. [9]

References

[1] H. Halberstam, H.-E. Richert, *Sieve Methods*, London Mathematical Society Monographs, Academic Press, 1974.

[2] Mario Sariols Quiles, *Sieve Theory and Applications*, University of Barcelona, 2017.

[3] Zhang, Yitang, *Bounded gaps between primes*, Department of Mathematics and Statistics, University of New Hampshire, 2013.

[4] Lin, Thomas, 2015, *After Prime Proof, an Unlikely Star Rises*, *Quanta Magazine*, April, 2. `www.quantamagazine.org/yitang-zhang-and-the-mystery-of-numbers-20150402`

[5] D. H. J. Polymath *The "bounded gaps between primes" Polymath project - a retrospective*, 2014, `https://arxiv.org/pdf/1409.8361.pdf`

[6] Tao, Terry. *The prime number theorem in arithmetic progressions, and dueling conspiracies*, 2009, `https://terrytao.wordpress.com/2009/09/24/the-prime-number-theorem-in-arithmetic-progressions-and-dueling-conspiracies/`

[7] Elliott, Peter D. T. A.; Halberstam, Heini, 1970, *A conjecture in prime number theory*, Symposia Mathematica, Vol. IV, London: Academic Press. pp. 5972.

[8] Bombieri, E. 1987, *Le Grand Crible dans la Thorie Analytique des Nombres*, Astrisque, 18 (Seconde ed.), Paris.

[9] Maynard, James, *Small gaps between primes*, 2019, `https://arxiv.org/pdf/1311.4600.pdf`