SURIM 2017 Project Write Up

Chen Lu

Stanford University

Abstract. This write up introduces some elementary results in analytic number theory, including Mertens' theorem, Chebychev's bound on $\pi(x)$, asymptotes for other arithmetic functions; it also outlines the ideas behind some more involved results, namely the prime number theorem, using complex analysis, and upper bounds for the number of twin primes, using sieve methods.

1 Introduction

We know, from the time of Euclid, that there are infinitely many primes. A natural question that follows is how are the primes distributed; somewhat surprisingly, we can find out a lot about this question, despite not having a good way of determining if a specific number is prime. Let $\pi(x)$ be the prime-counting function, which denotes the number of primes less than or equal to x. It was first conjectured, from numerical data, that $\pi(x) \sim x/\log x$, which means that $\pi(x)$ is almost equal to $x/\log x$ when x is large. Chebyshev then proved that, when x become large, $\pi(x)/(x/\log x)$ lies in the range $[\log 2, 2\log 2]$, and if $\lim_{x\to\infty} \pi(x)/(x/\log x)$ exists, the limit must be 1 (which would then implies $\pi(x) \sim x/\log x$). In 1859, Riemann introduced the connection between the zeros of the Riemann zeta function and the distribution of prime numbers; subsequently, in 1896, Hadamard and de la Valée Poussin independently proved $\pi(x) \sim x/\log x$, which is now known as the prime number theorem. Many different proofs of the prime number theorem have subsequently been found, including an elementary proof, by Selberg and Erdös in 1949, that does not involve the zeros of the zeta function. Moreover, if the Riemann hypothesis is true, we would know even more about $\pi(x)$ (that the error, when approximating $\pi(x)$ by $li(x)$, is of the order $O(\sqrt{x}\log x)$).

We can also try to investigate the distribution of twin primes, which are primes p where $p+2$ is also a prime, or of primes of the form p and $p+r$, which come in pairs. It has long been conjectured that there are infinitely many primes that come in pairs p and $p+r$, for some even number $r \geq 2$, although this result is only proven for $r \geq 246$, due to the efforts of Yitang Zhang, James Maynard and Terrence Tao. There are also conjectures on how these pairs of primes are distributed. Let $\pi_r(x)$ denote the number of pairs of primes p and $p+r$ less than or equal to x. Hardy and Littlewood conjectured that $\pi_r(x) \sim c(r)x/(\log x)^2$, where $c(r)$ is a constant depending on r. A proof of the conjecture is not known, but an upper bound on $\pi_r(x)$, which is of the same order of magnitude as the conjectured value, can be obtained by sieve methods.

In this write up, we aim to give an outline of the prime number theorem, and that of an upper bound for the number of primes of the form p and $p + r$. We will include most of the relevant background knowledge for these two results, and give references to some more technical arguments. We begin in section 2 on the technique of partial summation. The technique allows us to obtain Stirling's formula and three results due to Mertens, which will be useful in the sieve methods later on. We will also outline Chebyshev's theorem, which tells us something about the function $\pi(x)$.

Sections 3 and 4 demonstrate how to calculate the average values of important arithmetic functions. Section 3 deals with $\omega(n)$, the number of distinct prime factors of n, and $\Omega(n)$, the number of prime factors of n counted with multiplicity; section 4 deals with $d(n)$, the number of divisors of n, and other multiplicative functions. The work on $\omega(n)$ and $\Omega(n)$ is relevant for the sieve methods later on, and it can be applied to the multiplication table problem, which is presented in section 3. The work on multiplicative functions, combined with partial summation from section 2, will allow us to evaluate a wide range of expressions. Moreover, the use of Perron's formula, which we introduce in section 4, is relevant to the prime number theorem.

Section 5 contains the outline of the prime number theorem. It begins with a modification of Perron's formula, which simplifies the problem. Then we discuss properties of the Riemann zeta function, in particular its growth rate and its zeros. We finish by evaluating a contour integral and obtaining the desired main term.

Sections 6 and 7 are about sieve methods. Section 6 discusses Brun's pure sieve, and section 7 Selberg's sieve. The form of Brun's sieve that we present will give us upper bounds on the number of primes and twin primes, which are not on the order of the conjectured values, but will still enable us to show that the sum of the reciprocals of the twin primes converges (by using partial summation). Selberg's sieve will give us bounds for $\pi_r(x)$ on the order of Hardy and Littlewood's conjecture.

Before we dive into the various results, we will first introduce a few notations:

1. For functions f and g, $f = O(g)$ means that there exists a constant C such that $|f(x)| < Cg(x)$ for all large enough x.
2. $f = o(g)$ means that for all $\epsilon > 0$, there exists some x_0 such that $x > x_0$ implies $f(x) < \epsilon g(x)$.
3. $f \ll g$ means the same thing as $f = O(g)$, and $f \gg g$ means $g = O(f)$.
4. $f \sim g$ means that $\lim_{x \to infty} \frac{f(x)}{g(x)} = 1$, which is equivalent to $f(x) = g(x) + o(g(x))$

Acknowledgements. I would like to thank Professor Soundararajan for his guidance, and Pranav Nuti for many helpful discussions. In addition, I would like to thank George Schaeffer and the SURIM research program at Stanford University for providing facilities and support.

2 Estimation of Sums by Partial Summation

In this section, we will introduce the method of partial summation, which allows us to obtain asymptotic formulae such as $\log N! = N \log N - N + \frac{1}{2} \log N + C_0 + O\left(\frac{1}{N}\right)$ (Stirling's formula). This formula leads to the following result:

Theorem 1 (Chebyshev). *Let*

$$a = \liminf_{x \to \infty} \frac{\pi(x)}{x/\log x}$$

and

$$A = \limsup_{x \to \infty} \frac{\pi(x)}{x/\log x}$$

then

$$\log 2 \leq a \leq A \leq 2 \log 2$$

Moreover, if $a = A$, that is $\lim_{x \to \infty} \pi(x)/(x \log x)$ exists, then $a = A = 1$.

which provides a bound on $\pi(x)$ when x is large. We can also obtain the following three formulae, due to Mertens,

Theorem 2 (Mertens). *As $x \to \infty$, we have*

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x}$$

where γ is Euler's constant. Secondly, for some constant B, we have

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + B + O\left(\frac{1}{\log x}\right)$$

Finally, we have

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$$

The sums evaluated above will come up frequently, hence are very useful. The first sum comes up in the sieve of Eratosthenes and other sieve methods that we will discuss later on (more detailed discussion of the materials in this section can be found at [1]).

2.1 Partial summation

We begin by trying to estimate the sum $\sum_{n \leq x} \frac{1}{n}$. By considering the integral of $\frac{1}{t}$ from 1 to x, we conclude that this sum is roughly $\log x$. But we can achieve a more precise result than that. We will introduce a technique called partial summation. Suppose we have some integer sequence $a(n)$, where $A(x) = \sum_{n \leq x} a(n)$

denotes the partial sums. If we know something about $A(n)$, we want to use that knowledge to help us estimate the sum $\sum_{n\leq x} a(n)f(n)$, where f is a function on the reals. For example, in the sum $\sum_{n\leq x} \frac{1}{n}$, we have $a(n) = 1$ for all n (so $A(n)$ is simply n), and $f(x) = \frac{1}{x}$. Given this setting, we have the following result:

Lemma 1 (Partial Summation).

$$\sum_{n\leq x} a(n)f(n) = A(x)f(x) - \int_1^x f'(t)A(t)dt$$

Proof. Assume first that x is an integer. Note that

$$\begin{aligned}
\sum_{n\leq x} a(n)f(n) &= \sum_{n\leq x} (A(n) - A(n-1))f(n) \\
&= A(x)f(x) - \sum_{n\leq x-1} A(n)(f(n+1) - f(n)) \\
&= A(x)f(x) - \sum_{n\leq x-1} \int_n^{n+1} f'(t)A(t)dt \\
&= A(x)f(x) - \int_1^x f'(t)A(t)dt
\end{aligned}$$

As desired. The case when x is not an integer is very similar.

Using this technique on $\sum_{n\leq x} \frac{1}{n}$, we can get a better handle of the error, which comes from estimating the sum by an integral:

$$\begin{aligned}
\sum_{n\leq x} \frac{1}{n} &= 1 - \int_1^x -\frac{1}{t^2}\lfloor t \rfloor dt \\
&= 1 + \int_1^x \frac{1}{t^2}(t - \{t\})dt \\
&= 1 + \int_1^x \frac{1}{t}dt - \int_1^x \frac{\{t\}}{t^2}dt \\
&= \log x + \left(1 - \int_1^\infty \frac{\{t\}}{t^2}dt\right) + \int_x^\infty \frac{\{t\}}{t^2}dt
\end{aligned}$$

Where $\{t\} = t - \lfloor t \rfloor$ is the fractional part of t, which is bounded by 1, hence both the second and third term above are bounded by $\int_1^\infty \frac{1}{t^2}dt$, which is bounded. The second term, $1 - \int_1^\infty \frac{\{t\}}{t^2}dt$, is Euler's constant, denoted as γ. The third term is bounded by $\int_x^\infty \frac{1}{t^2}dt = \frac{1}{x}$. Thus we have the following result:

$$\sum_{n\leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right)$$

Partial summation will allow us to compute asymptotes for many other important sums, one of which is $\log N! = \sum_{n \leq N} \log n$. Following the earlier notation, we let $a(n) = 1$, $f(x) = \log x$, then we have:

$$\begin{aligned}\sum_{n \leq N} \log n &= N \log N - \int_1^N \frac{\lfloor t \rfloor}{t} dt \\ &= N \log N - N + 1 + \int_1^N \frac{\{t\}}{t} dt\end{aligned}$$

then we let

$$B(x) = \int_1^x \frac{\{t\}}{t} dt = \frac{x}{2} + C(x)$$

where $C(x)$ is bounded, then integrating by parts gives us:

$$\begin{aligned}\int_1^N \frac{\{t\}}{t} dt &= \frac{B(N)}{N} + \int_1^N \frac{B(t)}{t^2} dt \\ &= \frac{1}{2} + O\left(\frac{1}{N}\right) + \frac{1}{2} \log N + \int_1^N \frac{C(t)}{t^2} dt \\ &= \frac{1}{2} \log N + C + O\left(\frac{1}{N}\right)\end{aligned}$$

where C is some constant, so

$$\log N! = \sum_{n \leq N} \log n = N \log N - N + \frac{1}{2} \log N + C_0 + O\left(\frac{1}{N}\right)$$

which is actually Stirling's formula.

2.2 Chebyshev's bound on $\pi(x)$

We can say more about $\log N!$. In particular, we can express it as such:

$$\log N! = \sum_{n \leq N} \log n = \sum_{n \leq N} \sum_{p^\alpha || n} \log p^\alpha$$

where $p^\alpha || n$ means that α is the largest power of p that divides n. For convenience, we introduce the von Mangoldt function, $\Lambda(n)$, which is equal to $\log p$, when n is a prime power, p^k, and equal to 0 otherwise. Thus, going back to the sum and exchanging the order of summation, we have:

$$\begin{aligned}\sum_{n\leq N}\log n &= \sum_{n\leq N}\sum_{d|n}\Lambda(d)\\ &= \sum_{d\leq N}\Lambda(d)\sum_{\substack{d|n\\ n\leq N}}1\\ &= \sum_{d\leq N}\Lambda(d)\left\lfloor\frac{N}{d}\right\rfloor\\ &= N\sum_{d\leq N}\frac{\Lambda(d)}{d}+O\left(\sum_{d\leq N}\Lambda(d)\right)\end{aligned}\tag{1}$$

In the sum $\sum_{d\leq N}\frac{\Lambda(d)}{d}$, $\Lambda(d)$ serves to pick out the reciprocals of primes with weights of $\log p$; it also picks out reciprocals of prime powers, but these only make a negligible contribution to the sum, as we will see later when we discuss Mertens' theorem.

If we can get a handle on the error term $O\left(\sum_{d\leq N}\Lambda(d)\right)$, then we would be able to derive an asymptotic formula for $\sum_{d\leq N}\frac{\Lambda(d)}{d}$, from our earlier work on $\log N!$. The quantity $\sum_{d\leq N}\Lambda(d)$, which we will denote as $\psi(N)$, is very important and will come up in many of our later work. Notice that:

$$\psi(N)=\sum_{p\leq N}\sum_{\substack{k\geq 1\\ p^k\leq N}}\log p=\sum_{p\leq N}\log p\left\lfloor\frac{\log N}{\log p}\right\rfloor\leq\log N\pi(N)$$

where $\pi(N)$ denotes the number of primes less than or equal to N, which we will study in much greater depth later. Further manipulations with $\psi(N)$ can actually give us bounds on $\pi(N)$. First note that:

$$\begin{aligned}\log N! &= \sum_{n\leq N}\log n\\ &= \sum_{n\leq N}\sum_{d|n}\Lambda(d)\\ &= \sum_{d\leq N}\Lambda(d)\sum_{k\leq N/d}1\\ &= \sum_{k\leq N}\sum_{d\leq N/k}\Lambda(d)\\ &= \psi(N)+\psi(N/2)+\psi(N/3)+...\end{aligned}$$

so using the above identity on $2N$ and N, we have that:

$$\log\frac{2N!}{N!^2}=\psi(2N)-\psi(2N/2)+\psi(2N/3)-\psi(2N/4)+...$$

since $\psi(n)$ is increasing, we have that:

$$\psi(2N) - \psi(N) \leq \log \frac{2N!}{N!^2} \leq \psi(2N)$$

We know that $\frac{2N!}{N!^2} = \binom{2N}{N}$, and, because $\binom{2N}{N}$ is the largest of the coefficients $\binom{2N}{k}$, we have:

$$\frac{4^N}{2N+1} \leq \binom{2N}{N} \leq 4^N$$

Thus we are able to extract upper and lower bounds for $\psi(x)$ and $\pi(x)$. The lower bound for $\psi(x)$ comes easily:

$$\psi(2N) \geq \log \frac{4^N}{2N+1} = 2N \log 2 + O(\log N)$$

Since $\psi(2N) \leq \log 2N\pi(2N)$, we have, after replacing $2N$ with x,

$$\pi(x) \geq \frac{x}{\log x} \log 2 + O(1)$$

The upper bound for $\psi(x)$ requires a little more work (recall that we are interested in this upper bound because it would give us the error term in (1)). We have that:

$$\psi(2N) - \psi(N) \leq N \log 4$$

Replacing the integer N with a real number x, we have:

$$\psi(2x) - \psi(x) \leq x \log 4 + O(\log x)$$

Then replacing x by $x/2$, $x/4$, and so on, we have that:

$$\psi(2x) \leq (x + \frac{x}{2} + \frac{x}{4} + ...) \log 4 + O(\log x^2) = 2x \log 4 + O(\log x^2)$$

Thus we see that $\psi(x) = O(x)$. In fact, we will see later that $\psi(x) \sim x$, and this statement is equivalent to the prime number theorem, that $\pi(x) \sim \frac{x}{\log x}$. For now let us see how we can get an upper bound for $\pi(x)$. Note that:

$$\psi(2x) - \psi(x) = \sum_{x \leq n \leq 2x} \Lambda(n) \geq \sum_{x \leq p \leq 2x} \log p \geq \log x(\pi(2x) - \pi(x))$$

The first inequality holds because we are removing contributions from the prime powers, and the second one holds because $\log p \leq \log x$. $\sum_{p \leq N} \log p$ will also turn out to be a quantity of interest, and we will denote it as $\vartheta(N)$. Notice that $\vartheta(N)$ is very similar to $\psi(N)$, except without the contributions from the prime powers. In particular, note that

$$\psi(N) = \vartheta(N) + \vartheta(N^{\frac{1}{2}}) + \vartheta(N^{\frac{1}{3}}) + ... = \vartheta(N) + O\left(\sqrt{N}\log N\right)$$

and since $\psi(x) \sim x$, $\vartheta(x) \sim x$ as well, so again we see that contributions from prime powers are small when x gets large.

Going back to the inequality, we can follow a similar argument that we sketched out for $\psi(x)$ to get an upper bound for $\pi(x)$ (details see [1]), and so we have shown how to obtain the two bounds in Chebyshev's theorem (the statement in the theorem about the case when the limit exists is discussed in [1]).

2.3 Mertens' theorem

We now have an upper bound on $\psi(x)$, so we can go back to (1), and we can write out:

$$\log N! = N \sum_{d \leq N} \frac{\Lambda(d)}{d} + O(N)$$

combined with the result $\log N! = N\log N - N + \frac{1}{2}\log N + C_0 + O(\frac{1}{N})$, the formula above gives us:

$$\sum_{d \leq N} \frac{\Lambda(d)}{d} = \log N + O(1)$$

The sum above, as mentioned before, picks out the reciprocals of primes and prime powers with weight $\log p$:

$$\sum_{d \leq N} \frac{\Lambda(d)}{d} = \sum_{p \leq N} \frac{\log p}{p} + \sum_{p \leq \sqrt{N}} \log p \sum_{\substack{k \geq 2 \\ p^k \leq N}} \frac{1}{p^k}$$

the contribution from prime powers turns out to be just a constant, because the geometric series is bounded:

$$\sum_{\substack{k \geq 2 \\ p^k \leq N}} \frac{1}{p^k} < \sum_{k=2}^{\infty} \frac{1}{p^k} = O\left(\frac{1}{p^2}\right)$$

and since $\log p = O\left(\sqrt{p}\right)$, we have:

$$\sum_{p \leq \sqrt{(N)}} \frac{\log p}{p^2} \ll \sum_{p} \frac{\sqrt{p}}{p^2} < \sum_{n=1}^{\infty} \frac{1}{n^{\frac{3}{2}}} = O\left(1\right)$$

Where the order of the last sum comes from integration. Our calculation shows that the higher prime powers only contribute a constant's worth, so we have:

$$\sum_{d \leq N} \frac{\Lambda(d)}{d} = \sum_{p \leq N} \frac{\log p}{p} + O(1)$$

and

$$\sum_{p \leq N} \frac{\log p}{p} = \log N + O(1)$$

We have thus proved the third statement from Mertens' theorem, and the other two follow readily. The estimate for $\sum_{p \leq x} \frac{1}{p}$ can be achieved by partial summation, letting $a(n) = \frac{\log n}{n}$ when n is prime and 0 otherwise, and $f(x) = \frac{1}{\log x}$. The estimate for $\prod_{p \leq x} \left(1 - \frac{1}{p}\right)$ follows by taking log on the expression, considering its Taylor expansion, and using the result about $\sum_{p \leq x} \frac{1}{p}$.

2.4 Sieve of Eratosthenes

The quantity $\prod_{p \leq x} \left(1 - \frac{1}{p}\right)$ will appear in many places, and one such place is the estimation of $\pi(x)$. A direct way to estimate $\pi(x)$ is via the sieve of Eratosthenes: if we take away the numbers less than x which are multiples of primes less that $\sqrt{x}$, then the remaining numbers will be all the primes less than x. In general, starting with some $y < x$, if we take away the multiples of primes less than y, we will be left with the set

$$\{n \leq x : p | n \implies p \leq y\}$$

The way to count the numbers that are not multiples of a set of primes p_1, p_2, ..., p_k is by inclusion-exclusion: first, for each prime p_m, subtract away the number of multiples of p_k below x, which will be $\left\lfloor \frac{x}{p_i} \right\rfloor$; then, for each pair of primes p_i and p_j, add back the number of multiples of $p_i p_j$, $\left\lfloor \frac{x}{p_i p_j} \right\rfloor$, which would have been subtracted twice earlier; then subtract the number of multiples of three primes, $\left\lfloor \frac{x}{p_i p_j p_k} \right\rfloor$, and continue until we are left with the right amount.

To help with notation, we will introduce the Möbius function, $\mu(n)$, which is defined on the integers as follows:

$$\mu(n) = \begin{cases} 1 & \text{if } n \text{ is square-free and } n \text{ has an even number of prime factors;} \\ -1 & \text{if } n \text{ is square-free and } n \text{ has an off number of prime factors;} \\ 0 & \text{if } n \text{ is not square-free.} \end{cases}$$

where a number is not square-free means that it has a square that is greater than 1 as a divisor. An extremely important property of $\mu(n)$ is that it is a multiplicative function, which means that for any integers a and b, where $(a, b) = 1$, we have that $\mu(ab) = \mu(a)\mu(b)$. This property implies that the value that $\mu(n)$

takes is entirely determined by its values on the prime powers, where $\mu(p) = -1$, and $\mu(p^k) = 0$, for $k > 1$.

Going back to the inclusion-exclusion process, we see that the quantity we are looking for can be expressed as such:

$$\begin{aligned}\{n \leq x : p|n \implies p \leq y\} &= \sum_{\substack{d \\ p|d \implies p \leq y}} \mu(d) \#\{n \leq x : d|n\} \\ &= \sum_{\substack{d \\ p|d \implies p \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \\ &= x \sum_{\substack{d \\ p|d \implies p \leq y}} \frac{\mu(d)}{d} + O\left(\sum_{\substack{d \\ p|d \implies p \leq y}} |\mu(d)| \right)\end{aligned}$$

The main term above is

$$x \sum_{\substack{d \\ p|d \implies p \leq y}} \frac{\mu(d)}{d} = x \prod_{p \leq y} \left(1 - \frac{1}{p}\right)$$

and the error term is

$$O\left(\sum_{\substack{d \\ p|d \implies p \leq y}} |\mu(d)| \right) = O(2^{\pi(y)})$$

Then, letting $y = \sqrt{x}$ for example, and noting that $\pi(x) < y + \{n \leq x : p|n \implies p \leq y\}$, the main term above gives us that $\pi(x)$ is bounded by $\frac{2x}{\log x}$ (which is about a factor of 2 out). However, this bound is not valid, because when $y = \sqrt{x}$ the error term is much larger than the main term. In fact, in order to make the error term smaller than the main term, we have to choose $y = \log x$, giving us the bound of $\frac{x}{\log \log x}$, which is much worse. One way to proceed from the current result, and we will see this later, is to adapt the sieve method so that we can get a smaller error term. We will also present an overview of the proof of the prime number theorem (that $\pi(x) \sim \frac{x}{\log x}$) later on. Before moving onto these results, we will devote some time to presenting important functions on the integers.

3 $\omega(n)$ and $\Omega(n)$

We begin the exposition of integer functions with $\omega(n) = \sum_{p|n} 1$, the number of distinct prime factors of n, and $\Omega(n) = \sum_{p^\alpha || n} \alpha$, the number of prime factors of n counted with multiplicity. In particular, $\omega(n)$ is additive, meaning that

$f(ab) = f(a) + f(b)$, for $(a, b) = 1$, and $\Omega(n)$ is completely additive, meaning that $f(ab) = f(a) + f(b)$ always. These two functions are useful in many areas. $\omega(n)$ will be used to define a function that approximates $\mu(n)$ in our discussion on Brun's sieve later on. $\Omega(n)$ can be nicely applied to give a result due to Erdös:

Theorem 3 (Erdös). *The number of distinct integers in the $N \times N$ multiplication table is $o(N^2)$.*

which we will see at the end of this section (more details can again be found at [1]).

For such integer functions, we might like to investigate the maximum and minimum values that they can take, as well as their average value and variance. In the case for $\Omega(n)$, the minimum value is 1 (when n is a prime), and the maximum value $\max_{n \leq N} \Omega(n) = \left\lfloor \frac{\log N}{\log 2} \right\rfloor$, because each prime factor is less than or equal to 2. As for $\omega(n)$, the minimum value is also 1, and the maximum value we would expect to be obtained by a product of distinct primes, which turns out to be on the order of $\log N / \log \log N$, which is smaller than that of $\Omega(n)$.

We are more interested in the average values, that is looking for asymptotic formulae for $\frac{1}{N} \sum_{n \leq N} \omega(n)$ and $\frac{1}{N} \sum_{n \leq N} \Omega(n)$ as N gets large. Note that:

$$\bar{\omega}(n) = \frac{1}{N} \sum_{n \leq N} \omega(n) = \frac{1}{N} \sum_{n \leq N} \sum_{p|n} 1 = \frac{1}{N} \sum_{p \leq N} \sum_{\substack{n \leq N \\ p|n}} 1 = \frac{1}{N} \sum_{p \leq N} \left(\frac{N}{p} + O(1) \right)$$

by using Merten's theorem on the sum of reciprocal of primes, we have that the average of $\omega(n)$ is of order $\log \log N$. It turns out that the average of $\Omega(n)$ is also of order $\log \log N$, again because the contribution of prime powers is small.

As for the variance, note that

$$\begin{aligned} \frac{1}{N} (\omega(n) - \bar{\omega}(n))^2 &= \frac{1}{N} \sum_{n \leq N} \omega(n)^2 - \bar{\omega}(n)^2 \\ &= \frac{1}{N} \sum_{p_1, p_2 \leq N} \sum_{\substack{n \leq N \\ p_1|n \\ p_2|n}} 1 - \bar{\omega}(n)^2 \end{aligned}$$

After carefully investigating the first sum, we can show that it is of order $\log \log N^2 + O(\log \log N)$, and so the variance is $O(\log \log N)$, because $\bar{\omega}(n)^2$ is of order $\log \log N^2$. Thus we have the following result:

Theorem 4 (Hardy-Ramanujan; Turan). *The average value of $\omega(n)$ with $n \leq N$ is $\log \log N + B + O(1/\log N)$, where B is some constant, and the variance is*

$$\frac{1}{N} (\omega(n) - \bar{\omega}(n))^2 = O(\log \log N)$$

It turns out that the two asymptotic results above are also true for $\Omega(n)$, which can be deduced from the theorem explicitly by showing some quantity such as $\frac{1}{N}\sum_{n\leq N}(\Omega(n)-\omega(n))^2 \ll 1$. An important point to note is that the variance is small compared to the square of the mean, which means that the the typical values of $\omega(n)$, and similarly of $\Omega(n)$, are concentrated near the mean value. More explicitly, we have:

Corollary 1. *Let $f(N)$ be some function that tends to infinity with N. Then the set*

$$E = \{n \leq N : |\omega(n) - \log\log n| \geq f(N)\sqrt{\log\log N}\}$$

satisfies

$$|E| \ll \frac{N}{f(N)^2}$$

So almost all $n \leq N$ satisfies $\omega(n) \sim \log\log N$. The corollary comes easily from the theorem, because the variance is at least

$$\frac{1}{N}|E|(f(N)\sqrt{\log\log N})^2$$

and then we can use the result of the theorem. We know much more about the variances: Erdös and Kac showed that the quantity

$$\frac{\omega(n) - \log\log N}{\sqrt{\log\log N}}$$

has a normal distribution with mean 0 and variance 1.

With the Hardy and Ramanujan's theorem, we can achieve a neat result of Erdös known as the multiplication problem, which was stated in Theorem 3.

To prove the theorem, we will use Theorem 4 for $\Omega(n)$ instead of $\omega(n)$. Suppose $a, b \leq N$, so $n = ab$ appears in the multiplication table. If a and b are typical numbers below N, then they should have around $\log\log N$ prime factors, so n should have around $\Omega(n) = \Omega(a) + \Omega(b) = 2\log\log N$ prime factors. But a typical number below N^2 only has $\log\log N^2 \sim \log\log N$ factors, so the numbers in the multiplication table must be unusual (full proof is this idea presented with more details, which can be found in [1]).

What is known about $\Omega(n)$ and $\omega(n)$ is also related to permutations, because the cycle decomposition of a permutation can be thought of as a factorization, so we can also ask questions such as how many cycles does a typical element of S_n decompose into. More details can also be found in Sound's notes.

4 Multiplicative Functions

Multiplicative functions are functions on the integers that satisfy $f(ab) = f(a)f(b)$ for $(a, b) = 1$. We have already seen an example of a multiplicative function,

namely $\mu(n)$. Another example of such functions is $d(n) = \sum_{d|n} 1$, the divisor function, which counts the number of divisors of n. $d(n)$ is multiplicative because, if $n = p_1^{a_1} p_2^{a_2} ... p_k^{a_k}$ is written in its prime factorization, then $d(n) = (a_1 + 1)(a_2 + 1)...(a_k + 1)$, so indeed $d(ab) = d(a)d(b)$ for coprime a and b. We are particularly interested in how to evaluate averages of multiplicative functions, because an asymptotic formula for the average value, combined with the method of partial summation, will allow us to estimate a large variety of expressions. Moreover, in calculating the averages, we will introduce the use of Perron's formula, which will give a connection between the distribution of prime numbers and the zeros of the Riemann zeta function.

Let us consider $d(n)$. We can find its maximum value: we claim that for any $\epsilon > 0$, there exists a constant $C(\epsilon)$ such that

$$d(n) \leq C(\epsilon) n^{\epsilon}$$

To see this, note that

$$\frac{d(n)}{n^{\epsilon}} = \prod_{p^a || n} \frac{a+1}{p^{a\epsilon}} \leq \prod_{p^a || n} \max_{a \geq 0} \frac{a+1}{p^{a\epsilon}}$$

we can check by differentiation that $\frac{a+1}{p^{a\epsilon}}$ does attain a maximum, and for p large enough the maximum is simply 1, and thus the upper bound follows.

As for the average value $\bar{d}(n) = \frac{1}{N} \sum_{n \leq N} d(n)$, note that

$$\begin{aligned} \sum_{n \leq N} d(n) &= \sum_{n \leq N} \sum_{d|n} 1 \\ &= \sum_{d \leq N} \sum_{\substack{n \leq N \\ d|n}} 1 \\ &= \sum_{d \leq N} \frac{N}{d} + O(N) \\ &= N \log N + O(N) \end{aligned}$$

If we evaluate the sum by the hyperbola method, we can actually get that $\sum_{n \leq N} d(n) = N \log N + (2\gamma - 1)N + O(\sqrt{N})$ (details see [1]). In any case, we have $\bar{d}(n) = \log N + O(1)$.

4.1 Method of convolution

Before we proceed to more general cases of multiplicative functions, we will introduce an important operator on the multiplicative functions, and that is the Dirichlet convolution. For functions f and g, the Dirichlet convolution, $f * g$, is given by:

$$(f * g)(n) = \sum_{d|n} f(d)g(\frac{n}{d}) = \sum_{ab=n} f(a)g(b)$$

It is not hard to check that if f and g are both multiplicative, then $f * g$ will also be multiplicative, because for coprime a and b, the factors of ab can be uniquely written as the product of a factor of a times a factor of b. It is also not hard to check the Möbius inversion formula, that if $g = 1 * f$, then $f = \mu * g$ (the key is that $(1 * \mu)(n)$ is equal to 1 for $n = 1$ and 0 for all other n, which can be easily verified).

Going back to the example with $d(n)$, we note that $d = 1 * 1$. Moreover, the method for evaluating $\sum_{n \leq N} d(n)$ suggests a general method for evaluating sums of multiplicative functions: if we want to evaluate $\sum_{n \leq N} f(n)$, with $f(n)$ being quite complicated, we can write $f = 1 * g$, with $g = \mu * f$ by the inversion formula; with some luck, g should be a simpler function that f, and then we can evaluate the sum involving g (with $d(n)$, the function 1 is certainly simpler).

As an example, we will evaluate the sum $\sum_{n \leq N} \mu(n)^2$. Note that $\mu(n)^2$ is 1 on square-free integers and 0 elsewhere. We will write $\mu^2 = 1 * a$ for some function $a = \mu * \mu^2$. Because a is a multiplicative function, its values are determined by its values on the prime powers, which turn out to be $a(p^2) = -1$, and $a(p^k) = 0$ for $k \neq 1$. Thus we have:

$$\begin{aligned}
\sum_{n \leq N} \mu^2(n) &= \sum_{n \leq N} \sum_{cd=n} a(d) \\
&= \sum_{d \leq N} a(d) \sum_{\substack{n \leq N \\ d|n}} 1 \\
&= \sum_{d \leq N} a(d) \left(\frac{N}{d} + O(1) \right) \\
&= N \sum_{d \leq N} \frac{a(d)}{d} + O(\sqrt{N})
\end{aligned}$$

The error is $O(\sqrt{N})$ because $a(n)$ is 0 unless n is a square. Looking at the main term we have:

$$N \sum_{d \leq N} \frac{a(d)}{d} = N \sum_{d=1}^{\infty} \frac{a(d)}{d} + O\left(N \sum_{d \geq N} \frac{|a(d)|}{d} \right)$$

Again $a(n)$ is only non zero on squares, so

$$O\left(N \sum_{d \geq N} \frac{|a(d)|}{d} \right) = O\left(N \sum_{e \geq \sqrt{N}} \frac{1}{e^2} \right) = O(\sqrt{N})$$

at the same time, because $a(n)$ is multiplicative, we can exchange the sum in the main term into an Euler product:

$$\sum_{d=1}^{\infty} \frac{a(d)}{d} = \prod_{p} \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$$

Thus we have

$$\sum_{n \le N} \mu^2(n) = \frac{6}{\pi^2} N + O(\sqrt{N})$$

Using the method of writing functions as convolutions of simpler functions, we can also calculate sums of functions such as $d_3(n) = \sum_{abc=n} 1 = (1 * d)(n) = (1 * 1 * 1)(n)$, and we have $\sum_{n \le N} d_3(n) = NP(\log N) + O(N^{2/3} \log N)$, where P is a polynomial of degree 2 with leading coefficient $\frac{1}{2}$. More generally, let $d_k(n)$ be the k-th divisor function, which denotes the number of ways n can be written as a product of k numbers, then we have $d_k(n) = (1 * d_{k-1})(n)$, and we can show that $\sum_{n \le N} d_k(n) = NP_k(\log N) + O(N^{(k-1/k)+\epsilon}$ using our method.

Incidentally, note that $d_k(n)$ are the coefficients of the Dirichlet series of $\zeta^k(s)$:

$$\zeta^k(s) = \left(\sum_{n=1}^{\infty} \frac{1}{n^s}\right)^k = \sum_{n=1}^{\infty} \frac{d_k(n)}{n^s}$$

If we consider the coefficients of the Dirichlet series of $\zeta^\pi(s)$, then we can define the function $d_\pi(n)$, which is also multiplicative. But $d_\pi(n)$ does not only take integer values, so the sum $\sum_{n \le N} d_\pi(n)$ cannot be evaluated using the convolution method easily. Instead, there is a complex analytic method which handles such sums.

4.2 Perron's formula

The key to the complex analytic method is Perron's formula, which is given below:

Proposition 1. *For $y > 0$ and $c > 0$,*

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{y^s}{s} ds = \begin{cases} 1 & y > 1 \\ \frac{1}{2} & y = 1 \\ 0 & y < 1 \end{cases}$$

When $y > 1$, we want to shift the line of integration to the left, so y^s becomes very small, and we pick up a pole at 0 with residue 1; when $y < 1$, we shift the line of integration to the right, so again y^s becomes small, and this time we pass through no poles, so the integral evaluates to 0 (the case for $y = 1$ is more complicated).

With this formula in hand, we can try to evaluate $\sum_{n\leq x} a(n)$ by analyzing $A(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s}$. In particular, note that

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} A(s)\frac{x^s}{s}ds = \sum_{n=1}^{\infty} a(n)\frac{1}{2\pi i}\int_{(c)} \left(\frac{x}{n}\right)^s \frac{1}{s}ds = \sum_{n\leq x} a(n)$$

with the caveat that x should not be chosen to be an integer, so the weight of $\frac{1}{2}$ from Perron's formula does not come up. Using this method, we can try to find the asymptotic formula for $\sum_{n\leq N} d_\pi(n)$. Note that:

$$\sum_{n=1}^{\infty} \frac{d_\pi(n)}{n^s} = \zeta(s)^\pi = \exp(\pi \log \zeta(s))$$

and we will want to shift contours from $c > 1$ to some $c > 0$. The term above gives us a logarithmic singularity at $s = 1$, which turns out to give the following main term (more details can be found at [1]):

$$\sum_{n\leq x} d_\pi(n) \sim \frac{x(\log x)^{\pi-1}}{\Gamma(\pi)}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the gamma function, a meromorphic function with simple poles at the non-positive integers, with the property that $\Gamma(n) = (n-1)!$.

It turns out that Perron's formula allows us to relate the zeros of the zeta function to $\pi(x)$, and this idea leads to the proof of the prime number theorem. To be more specific, notice that for Re $s > 1$,

$$\begin{aligned}
\zeta(s) &= \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \\
\log \zeta(s) &= \sum_p -\log\left(1 - \frac{1}{p^s}\right) = \sum_p \sum_{k=1}^{\infty} \frac{1}{kp^{ks}} \\
\frac{\zeta'}{\zeta}(s) &= \sum_p \sum_{k=1}^{\infty} \frac{-k\log p}{kp^{ks}} = -\sum_p \sum_{k=1}^{\infty} \frac{\Lambda(p^k)}{kp^{ks}} = -\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}
\end{aligned}$$

Using Perron's formula, and taking $c > 1$, we have:

$$\psi(x) = \sum_{n\leq x} \Lambda(n) = \frac{1}{2\pi i}\int_{(c)} \left(-\frac{\zeta'}{\zeta}(s)\right)\frac{x^s}{s}ds \tag{2}$$

Moving the line of integration to the left, and picking up poles at $s = 1$, $s = 0$, and $s = \rho$ where ρ will denote all the zeros of the zeta function, we have that:

$$\psi(x) = x + \sum_\rho \frac{x^\rho}{\rho} + \frac{\zeta'}{\zeta}(0)$$

We mentioned earlier that $\psi(x) \sim x$ is equivalent to $\pi(x) \sim \frac{x}{\log x}$, so to prove the prime number theorem, we want to show that x becomes the main term in the formula above. In fact, we will not need to shift the line of integration all the way to $c = -\infty$ and include all the zeros of the zeta function. Instead, we only need to shift the line slightly to the left of 1, so we pick up the residue of x, and get an asymptotic formula from there. We will give an outline of the proof of the prime number theorem in the next section.

5 The Prime Number Theorem

We will now present one of the main results of this write up, the prime number theorem:

Theorem 5 (Prime Number Theorem). *As $x \to \infty$, we have*

$$\psi(x) \sim x$$

and

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

We present the proof in three steps. First we discuss the set up of the proof. We will not be using Perron's formula, as we demonstrated at the end of the previous section, but a variant of it. Then we quote some results on the growth rate of the Riemann zeta function, and demonstrate that the function has a zero free region close the the line Re $s = 1$. We finish by considering the relevant contour integral, and show that the main term is what we desire.

5.1 Set up

We begin by discussing why $\psi(x) \sim x$ and $\pi(x) \sim \frac{x}{\log x}$ are equivalent. We already know that $\psi(x) \sim x$ and $\vartheta(x) \sim x$ are equivalent. Because $\vartheta(x) = \sum_{p \leq x} \log p$ and $\pi(x) = \sum_{p \leq x} 1$, we can simply go from the asymptote of one to the other via partial summation (in fact we can get $\pi(x) \sim li(x)$, with $li(x)$ being the logarithmic integral). We will actually prove the prime number theorem by showing that $\psi_1(x) \sim \frac{x^2}{2}$, where

$$\psi_1(x) = \sum_{n \leq x} \Lambda(n)(x - n) = \int_0^x \psi(t)dt$$

Again we can check this statement is equivalent to the other statements by partial summation. The reason why we will work with $\psi_1(x)$ is that we can use a smoother variant of Perron's formula:

$$\frac{1}{2\pi i} \int_{(c)} \frac{y^s}{s(s+1)} ds = \begin{cases} 1 - \frac{1}{y} & y > 1 \\ 0 & y \leq 1 \end{cases}$$

from which we get the identity:

$$\psi_1(x) = \sum_{n \leq x} \Lambda(n)(x-n) = \left(\frac{1}{2\pi i} \int_{(c)} -\frac{\zeta'}{\zeta}(s) \frac{x^s}{s(s+1)} ds \right) x \tag{3}$$

This identity is easier to work with, compared to (2), because of the extra factor of $(s+1)$ in the denominator, which makes the integral converge more easily when we shift contours. However, it is also possible to prove the prime number theorem just by working with $\psi(x)$.

From (3), we will consider the integral

$$\frac{1}{2\pi i} \int_{(c)} F(s) ds = \frac{1}{2\pi i} \int_{(c)} -\frac{\zeta'}{\zeta}(s) \frac{x^{s+1}}{s(s+1)} ds \tag{4}$$

We will show how to shift the line of integration to get the desired main term of $\frac{x^2}{2}$ later on. Before we do so, we need to establish some properties of the zeta function.

5.2 Growth rate and zero free region of the zeta function

Recall that the Riemann zeta function is defined for Re $s > 1$ as:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

and it has an analytic continuation with a simple pole at $s = 1$ (more details about the zeta function can be found at [5]). Note that, for $s = \sigma + it$, if $\sigma > 1$ then $\zeta(s)$ is bounded by $\zeta(\sigma)$, so ζ is bounded uniformly to the right of the line Re $s = 1$. Moreover, the zeta function does not grow quickly on the line Re $s = 1$ away from the pole, which is exhibited in the following proposition:

Proposition 2. *Suppose $s = \sigma + it$, then for each $0 \leq \sigma_0 \leq 1$ and $\epsilon > 0$, there exists a constant c_ϵ such that:*

1. $|\zeta(s)| \leq c_\epsilon |t|^{1-\sigma_0+\epsilon}$, *if* $\sigma_0 \leq \sigma$ *and* $|t| \geq 1$.
2. $|\zeta'(s)| \leq c_\epsilon |t|^{\epsilon}$, *if* $\sigma \geq 1$ *and* $|t| \geq 1$.
3. $1/|\zeta(s)| \leq c_\epsilon$, *if* $\sigma \geq 1$ *and* $|t| \geq 1$.

The proof is not very difficult, and can be found in [4]. Another important property of the zeta function is that it has no zeros on the line Re $s = 1$, and we discuss the proof now. First, we have the following trig identity:

$$3 + 4\cos\theta + \cos 2\theta = 2(1 + \cos\theta)^2 \geq 0$$

We want to show that the zeta function has no zeros on the line Re $s = 1$ (when the real part is larger than 1, we can write the zeta function into an Euler product, which then shows that it cannot be zero). Now note that, for Re $s > 1$, we have:

$$\begin{aligned}\operatorname{Re}(\log\zeta(s)) &= \operatorname{Re}\left(\sum_p \sum_{k=1}^{\infty} \frac{1}{kp^{ks}}\right) \\ &= \sum_p \sum_{k=1}^{\infty} \frac{1}{k} p^{-k\sigma}\cos(kt\log p)\end{aligned}$$

Applying the formula above to $\zeta(1+\epsilon)$, $\zeta(1+\epsilon+it)$, and $\zeta(1+\epsilon+2it)$, and taking a clever linear combination, we have:

$$\begin{aligned}&3\log|\zeta(1+\epsilon)| + 4\log|\zeta(1+\epsilon+it)| + \log|\zeta(1+\epsilon+2it)| \\ &\quad= 3\operatorname{Re}(\log\zeta(1+\epsilon)) + 4\operatorname{Re}(\log\zeta(1+\epsilon+it)) + \operatorname{Re}(\log\zeta(1+\epsilon+2it)) \\ &\quad= \sum_p \sum_{k=1}^{\infty} \frac{1}{k} p^{-k(1+\epsilon)}\left(3 + 4\cos(kt\log p) + \cos(2kt\log p)\right) \\ &\quad\geq 0\end{aligned}$$

Hence we have

$$\left|\zeta(1+\epsilon)^3\zeta(1+\epsilon+it)^4\zeta(1+\epsilon+2it)\right| \geq 1$$

If the zeta function is 0 at some $1+it$, then taking $\epsilon \to 0$, we have

$$\left|\zeta(1+\epsilon)^3\zeta(1+\epsilon+it)^4\zeta(1+\epsilon+2it)\right| \to 0$$

which is a contradiction, so the zeta function cannot have any zeros on the line $\operatorname{Re} s = 1$. The intuition for considering the quantity $\left|\zeta(1+\epsilon)^3\zeta(1+\epsilon+it)^4\zeta(1+\epsilon+2it)\right|$ is that, if $1+it$ is a zero, then we are multiplying together three poles, four zeros, and some regular holomorphic component, which should be small.

Because the zeta function is holomorphic, it cannot have a limiting sequence of zeros. Hence there exists some δ such that for $\operatorname{Re} s \in (1-\delta, 1]$, the zeta function has no zeros. We now know enough about the zeta function to complete the proof of the prime number theorem.

5.3 Completing the proof

Going back to the integral in (4), we will shift the line of integration from $c - i\infty \to c + i\infty$, with $c > 1$, to the following (shown in Figure 1):

$$1 - i\infty \to 1 - iT \to 1 - \delta - iT \to 1 - \delta + iT \to 1 + iT \to 1 + i\infty$$

where T is some parameter, and δ is such that the zeta function has no zeros in $[1-\delta, 1]$, which means that shifting the contour as above will only pick up

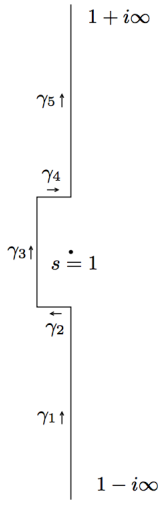

Fig. 1. Line of integration that we shift to; image taken from [4]

the residue at $s = 1$, which is $\frac{x^2}{2}$, and it will be the main term. We now want to show that the contributions from the other integrals are small compared to the main term, for which we will use the bounds from Proposition 2. We examine each line segment of the integral along the specified path. Firstly, for a large enough choice of T, we have:

$$\left|\int_{1-i\infty}^{1-iT} F(s)ds\right| \leq \frac{\epsilon x^2}{2} \text{ and } \left|\int_{1+iT}^{1+i\infty} F(s)ds\right| \leq \frac{\epsilon x^2}{2}$$

because $|x^{s+1}| = x^2$ when the real part of s is 1, and $|\zeta'(s)/\zeta(s)| \leq A|t|^\epsilon$ by the proposition, so the integral is bounded by

$$\left|\int_{1-i\infty}^{1-iT} F(s)ds\right| \leq Cx^2 \int_T^\infty \frac{|t|^\epsilon}{t^2} dt$$

The integral is bounded and can be made arbitrarily small, so we have our earlier bound. As for the other vertical segment, we have:

$$\left|\int_{1-\delta-iT}^{1-\delta+iT} F(s)ds\right| \leq C_T x^{2-\delta}$$

where the constant C_T depends on T. This bound is simply because that on the segment $1 - \delta - iT \to 1 - \delta + iT$ the component $-\frac{\zeta'}{\zeta}(s)\frac{1}{s(s+1)}$ is bounded above. Finally, for the two horizontal segments, we have:

$$\left| \int_{1-iT}^{1-\delta-iT} F(s)ds \right| \leq C'_T \int_{1-\delta}^{1} x^{1+\sigma} d\sigma \leq C'_T \frac{x^2}{\log x}$$

Thus the contribution from the integrals are all small compared to the main term, and so $\psi_1(x) \sim \frac{x^2}{2}$, which implies the prime number theorem.

6 Brun's Pure Sieve

In the final two sections, we will present the other main result of this write up: the upper bound for the number of twin primes (and of pairs of primes of the form p and $p+r$). The method we use is the sieve method, motivated by the sieve of Eratosthenes. In this section, we will develop Brun's pure sieve, which will give us the bounds $\pi(x) \ll (x \log\log x)/\log x$ (a factor of $\log\log x$ out from the true value, given by the prime number theorem) and $\pi_2(x) \ll (x \log\log x)^2/(\log x)^2$ (a factor of $(\log\log x)^2$ out from Hardy and Littlewood's conjectured value). Even though the bound for the number of twin primes is not optimal, we can still deduce from it, by using partial summation, that the sum of reciprocals of twin primes is bounded, which is a nice corollary. We achieve better bounds in the next section, using Selberg's sieve.

Let us define $S(x, y; P)$ to be the number of n such that $x < n \leq x + y$, and n is coprime with P. Recall our discussion of the sieve of Eratosthenes, which we can use to get an expression for $S(x, y; P)$. In particular, we need to use the fact that $\sum_{d|n} \mu(d)$ is 1 for $n = 1$ and 0 for all other n, from which we get the following:

$$\begin{aligned} S(x, y; P) &= \sum_{x<n\leq x+y} \sum_{\substack{d|n \\ d|P}} \mu(d) \\ &= \sum_{d|P} \mu(d) \sum_{\substack{d|n \\ x<n\leq x+y}} 1 \\ &= y \sum_{d|P} \frac{\mu(d)}{d} + O\left(\sum_{d|P} |\mu(d)| \right) \end{aligned}$$

The second line in the above calculation shows the inclusion-exclusion process from the sieve of Eratosthenes more explicitly. The issue with this formula, as we have previously discussed, is that the error term is very large. If we let $P = \prod_{p\leq z} p$ be a product of distinct primes, then, in order to get a meaningful bound, we must choose $z = \log y$, and then we have

$$S(x, y; P) \sim e^{-\gamma} \frac{y}{\log\log y}$$

Since we also have $\pi(x + y) - \pi(x) \leq \omega(P) + S(x, y; P)$, we have that

$$\pi(x+y) - \pi(x) \leq e^{-\gamma} \frac{y}{\log\log y}$$

which is a very weak bound (we might expect the actual bound to be on the order of $y/\log y$, from the prime number theorem, and this speculation turns out to be correct). In order to get a better upper bound, we might try to find some function $\mu_2(n)$ such that $\sum_{d|n} \mu_2(d) \geq \sum_{d|n} \mu_2(d)$ for all n, and that the error term $O\left(\sum_{d|P} \mu_2(d)\right)$ is small. Brun found a $\mu_2(n)$ that satisfies these conditions:

$$\mu_2(n) = \begin{cases} \mu(n) & \text{if } \omega(n) \leq 2h \text{ for some integer } h \\ 0 & \text{otherwise} \end{cases}$$

So we have

$$\sum_{d|n} \mu_2(d) = \sum_{j=0}^{2h} \sum_{\substack{d|n \\ \omega(d)=j}} \mu(d) = \sum_{j=0}^{2h} (-1)^j \binom{\omega(n)}{j} = (-1)^{2h} \binom{\omega(n)-1}{2h}$$

The last identity can be shown by comparing coefficients in the power series expansion of $(1-x)^{-1}(1-x)^{2h}$ and $(1-x)^{2h-1}$. Thus we see that $\mu_2(n)$ indeed satisfies our earlier requirements. Again letting $P = \prod_{p \leq z} p$, we have:

$$\begin{aligned} \pi(x+y) - \pi(x) &\leq z + S(x, y; P) \\ &= y \sum_{d|P} \frac{\mu_2(d)}{d} + O\left(z + \sum_{\substack{d|P \\ \omega(d) \leq 2h}} |\mu(d)| \right) \\ &= y \prod_{p \leq z} \left(1 - \frac{1}{p}\right) + O\left(z + \sum_{\substack{d|P \\ \omega(d) \leq 2h}} 1 + y \sum_{\substack{d|P \\ \omega(d) > 2h}} \frac{1}{d} \right) \end{aligned}$$

We know that the main term is $e^{-\gamma} y/\log z$ by Mertens. The second error term does not exceed z^{2h}. It turns out we can choose h so that the third term is bounded by $y/(\log z)^v$, where v is some integer larger than 3. Then, choosing $z = y^{1/(10 \log\log y)}$ (details see [3]), we have that:

$$\pi(x+y) - \pi(x) \ll \frac{y \log\log y}{\log y}$$

This result does improve upon our previous bound, although it still has an extra factor of $\log\log y$. The power of the sieve method is that its argument can be generalized to many problems. For example, we could also investigate $J(y)$, the number of twin primes below y. Again we let $P = \prod_{\leq z} p$, and we will bound $J(y)$ by the number of $n \leq y$ such that $n(n+2)$ is coprime to P, so we have

$$J(y) \leq z + \sum_{n \leq y} \sum_{\substack{d | n(n+2) \\ d | P}} \mu_2(d)$$
$$= z + \sum_{d|P} \mu_2(d) \sum_{\substack{n \leq y \\ d | n(n+2)}} 1$$

We want to answer the question of how many n modulo d are there such that $n(n+2) \equiv 0$. If d is prime, then the answer is 1 if $d = 2$ and 2 otherwise. The number of solutions to $n(n+2) \equiv 0 \bmod d$ can thus be written as $\rho(d)$, where $\rho(n)$ is completely multiplicative, and takes $\rho(2) = 1$, $\rho(p) = 2$ (this is because of the Chinese remainder theorem: if $d|n(n+2)$, then for each $p|d$, $p|n(n+2)$, so the number of solutions modulo d is the product of the number of solutions modulo p). Thus we have:

$$J(y) \leq y \sum_{d|P} \frac{\mu(d)\rho(d)}{d} + O\left(z + \sum_{\substack{d|P \\ \omega(d) \leq 2h}} \rho(d) + y \sum_{\substack{d|P \\ \omega(d) > 2h}} \frac{\rho(d)}{d} \right)$$

where the main term is given by:

$$\frac{1}{2} y \prod_{2 < p \leq z} \left(1 - \frac{2}{p}\right) \leq 2x \prod_{p \leq z} \left(1 - \frac{1}{p}\right)^2 \sim 2e^{-2\gamma} \frac{y}{(\log z)^2}$$

making a similar choice of h and z as we did before (more details discussed in [3]), we get the result:

Theorem 6 (Brun). *As $y \to \infty$, we have:*

$$J(y) \ll \frac{y(\log \log y)^2}{(\log y)^2}$$

and then, using partial summation, and setting $a(n)$ as the indicator function for twin primes, and $f(x) = \frac{1}{x}$, we get the following:

Corollary 2. *Let J denote the set of twin primes, then:*

$$\sum_{p \in J} \frac{1}{p} < \infty$$

7 Selberg's Sieve

We will now introduce Selberg's sieve, with the aim of getting a better bound for the number of pairs of primes of the form p and $p + r$.

We return to the identity

$$S(x,y;P) = \sum_{x<n\le x+y} \sum_{\substack{d|p\\ d|n}} \mu(n) = y \sum_{d|P} \frac{\mu(d)}{d} + O\left(\sum_{d|P} |\mu(d)|\right)$$

The method to find an upper bound that we discussed in the last section was to find some $\sum_{d|n} \mu_2(n) \ge \sum_{d|n} \mu(n)$ that makes the the error small. Selberg proposed a slightly different method: to find some function Λ_n such that $\left(\sum_{d|n} \Lambda_d\right)^2 \ge \sum_{d|n} \mu(n)$. We will require two other constraints be satisfied: that $\Lambda_1 = 1$, and $\Lambda_n = 0$ for $n > z$ for some z. Then we have:

$$\begin{aligned} S(x,y;P) &\le \sum_{x<n\le x+y} \left(\sum_{\substack{d|n\\ d|P}} \Lambda_d\right)^2 \\ &= \sum_{\substack{d|p\\ e|P}} \Lambda_d \Lambda_e \sum_{\substack{x<n\le x+y\\ d|n\\ e|n}} 1 \\ &= y \sum_{\substack{d|p\\ e|P}} \frac{\Lambda_d \Lambda_e}{[d,e]} + O\left(\left(\sum_{d|P} |\Lambda_d|\right)^2\right) \end{aligned}$$

where $[d,e] = \operatorname{lcm}(d,e)$. We now have a quadratic form in Λ_d, so we wish to diagonalize it. Using the identity $\sum_{d|n} \phi(d) = n$, and that $de = (d,e)[d,e]$, we have:

$$\begin{aligned} \sum_{\substack{d|p\\ e|P}} \frac{\Lambda_d \Lambda_e}{[d,e]} &= \sum_{\substack{d|p\\ e|P}} \frac{\Lambda_d}{d} \frac{\Lambda_e}{e} (d,e) \\ &= \sum_{\substack{d|p\\ e|P}} \frac{\Lambda_d}{d} \frac{\Lambda_e}{e} \sum_{\substack{f|d\\ f|e}} \phi(f) \\ &= \sum_{f|P} \phi(f) \sum_{\substack{d\\ f|d|P}} \frac{\Lambda_d}{d} \sum_{\substack{e\\ f|e|P}} \frac{\Lambda_e}{e} \\ &= \sum_{f|P} \phi(f) {y_f}^2 \end{aligned}$$

where

$$y_f = \sum_{\substack{d \\ f|d|P}} \frac{\Lambda(d)}{d}$$

by a variant of the Möbius inversion formula, we also have that

$$\frac{\Lambda_d}{d} = \sum_{\substack{f \\ d|f|P}} y_f \mu(f/d)$$

So our constraints on Λ_n, that $\Lambda_1 = 1$ and $\Lambda_n = 0$ for $n > z$, can be translated to the constraints $\sum_{f|P} y_f \mu(f) = 1$ and $y_f = 0$ for $f > z$. We have not yet specified the values of y_f and Λ_d, so what we want is to choose values for these two functions to minimize the quadratic form, subject to our linear constraints, which can be done using Lagrange multipliers. Because of the following factorization:

$$\sum_{f|P} \phi(f) {y_f}^2 = \sum_{\substack{f|p \\ f \leq z}} \phi(f) \left(y_f - \frac{\mu(f)}{\phi(f) L_P(z)} \right)^2 + \frac{1}{L_P(z)}$$

where

$$L_P(z) = \sum_{\substack{n \leq z \\ n|P}} \frac{\mu(n)^2}{\phi(n)}$$

we see that the optimal choice of y_f that minimizes our quadratic form is

$$y_f = \frac{\mu(f)}{\phi(f) L_P(z)}$$

Going back to

$$S(x, y; P) = y \sum_{\substack{d|p \\ e|P}} \frac{\Lambda_d \Lambda_e}{[d, e]} + O\left(\left(\sum_{d|P} |\Lambda_d| \right)^2 \right)$$

with our choice of y_f, and hence of Λ_d, we have (details found in [2]):

$$S(x, y; P) \leq \frac{y}{L_P(z)} + O\left(\frac{z^2}{L_P(z)^2} \right)$$

we can also show that $L_P(z) \leq \log z$, if we let $P = \prod_{p \leq z} p$. Letting $z = \sqrt{y}$, we have that

$$S(x, y; P) \leq \frac{2y}{\log y} + O\left(\frac{y}{(\log y)^2} \right)$$

and hence

$$\pi(x+y) - \pi(x) \leq \frac{2y}{\log y} + O\left(\frac{y}{(\log y)^2}\right)$$

And now we are only out by a factor of 2.

Notice that $S(x, y; P)$ is the set of numbers n in $(x, x+y]$ with $n \equiv 0$ modulo some $p|P$ sieved out. It is very natural to ask what the bound for $S'(x, y; P)$ would be, where we sieve out numbers that are c_p modulo each $p|P$, instead of sieving out the 0 residue class. By a simple argument in Montgomery and Vaughan, $S'(x, y; P)$ can be bounded by the same bound as the bound for $S(x, y; P)$ that we have above.

We can also considering sieving out multiple residue classes for each $p|P$ instead of just one (which is exactly what we did when we tried to bound twin primes in the previous section: sieving out $n(n+2) \equiv 0$ modulo p is simply sieving out the 0 and -2 residue classes). Let $\mathcal{B}(p)$ be a set of 'bad' residue classes that we want to sieve out, and let $b(p)$ be the size of $\mathcal{B}(p)$. Consider the quantity:

$$a(n) = \prod_{\substack{p|P \\ n \equiv \mathcal{B}(p) \bmod p}} p$$

so the values of n that remain after sieving are exactly the ones for which $a(n) = 1$. Now $p|a(n)$ if and only if $n \equiv \mathcal{B}(p)$ modulo p. By the Chinese remainder theorem, for some m and all $p|m$, there are exactly $\prod_{p|m} b(p)$ residue classes of m that lie in bad residue classes of p. Thus $b(m)$ is a completely multiplicative function that represents the number of bad residues modulo m. After sieving out all $\mathcal{B}(p)$ for $p|P$, we are left with:

$$\begin{aligned}
\sum_{x<n\leq x+y} \sum_{\substack{d|P \\ d|a(n)}} \mu(d) &\leq \sum_{x<n\leq x+y} \left(\sum_{\substack{d|P \\ d|a(n)}} \Lambda(d) \right)^2 \\
&= \sum_{\substack{d|P \\ e|P}} \Lambda_d \Lambda_e \sum_{\substack{x<n\leq x+y \\ [d,e]|a(n)}} 1 \\
&= y \sum_{\substack{d|P \\ e|P}} \frac{b([d,e])}{[d,e]} \Lambda_d \Lambda_e + O\left(\sum_{\substack{d|P \\ e|P}} |\Lambda_d \Lambda_e| b([d,e]) \right)
\end{aligned}$$

Again we diagonalize the main term, and we get

$$\sum_{\substack{d|P \\ e|P}} \frac{b([d,e])}{[d,e]} \Lambda_d \Lambda_e = \sum_{f|P} \frac{1}{g(f)} {y_f}^2$$

where

$$g(f) = \prod_{p|f} \frac{b(p)}{p - b(p)}$$

for square-free f, and

$$y_f = \sum_{\substack{d \\ f|d|P}} \frac{b(d)}{d} \Lambda_d$$

Again the linear constraints $\Lambda_1 = 1$ and $\Lambda_d = 0$ for $d > z$ translate to $\sum_{f|P} y_f \mu(f) = 1$ and $y_f = 0$ for $f > z$, and again we have a factorization

$$\sum_{f|P} \frac{1}{g(f)} {y_f}^2 = \sum_{\substack{f|p \\ f \le z}} \frac{1}{g(f)} \left(y_f - \frac{\mu(f) g(f)}{L} \right)^2 + \frac{1}{L}$$

where

$$L = \sum_{\substack{f|P \\ f \le z}} \mu(f)^2 g(f)$$

Thus the main term is minimized by taking

$$y_f = \frac{\mu(f) g(f)}{L}$$

for $f \le z$.

We can apply the work above to bound the number of twin primes, with $b(2) = 1$ and $b(p) = 2$ for $p > 2$. Letting $P = \prod_{p \le z} p$ and choosing $z = \sqrt{y}$, we can show that (details see [2]):

Theorem 7. *The number of integers n in $(x, x+y]$ such that $(n(n+2), P) = 1$ does not exceed*

$$\frac{8cy}{(\log y)^2} + O\left(\frac{y \log\log y}{(\log y)^3} \right)$$

where

$$c = 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2} \right)$$

Thus the number of twin primes in the interval $(x, x+y]$ is bounded by the above formula. We can also consider the number of primes p such that $p + r$ is also prime, for some even number r, in the interval $(x, x+y]$. In this case, we have $b(p) = 1$ for all $p|r$, and $b(p) = 2$ for $p \nmid r$. Going through the argument as before, we arrive at the following:

Theorem 8. *For an even number r, the number of primes $p \in (x, x+y]$ such that $p+r$ is also prime is bounded by*

$$\frac{8c(r)y}{(\log y)^2} + O\left(\frac{y \log\log y}{(\log y)^3}\right)$$

where

$$c(r) = \left(\prod_{\substack{p|r \\ p \neq 2}} \frac{p-1}{p-2}\right) c$$

where c is the constant from before.

Hardy and Littlewood's first conjecture states that the number of primes $p \leq y$ such that $p+r$ is also prime is asymptotic to $\frac{c(r)y}{(\log y)^2}$, so our bound from Selberg's sieve is off by a factor of 8.

References

1. Soundararajan, K.: *Notes for Math 155*. Available upon request.
2. Montgomery, H. L., Vaughan, R. C.: *Multiplicative Number Theory: I. Classical Theory*. Cambridge University Press, 2006.
3. Tenenbaum, G.: *Introduction to Analytic and Probabilistic Number Theory*. Cambridge University Press, 1995
4. Stein, E. M., Shakarchi, R.: *Complex Analysis*. Princeton University Press, 2003.
5. Stopple, J.: *A Primer of Analytic Number Theory*. Cambridge University Press, 2003.