MATH 155: ANALYTIC NUMBER THEORY

JESSE THORNER

1. April 2

(1) Syllabus / LaTeX
(2) Analytic number theory
   (a) Present analytic techniques and their applications in the setting of NT
   (b) Develop analytic intuition by making educated guesses and adjusting guesses in light of new information
(3) Tentative list of ourse topics
   (a) The distribution of prime numbers. In particular Chebyshev's bounds for $\pi(x)$, the number of primes up to x. Bertrand's postulate.
   (b) The Hardy-Ramanujan theorem on the number of prime factors of a typical integer. Statistics and mean values of arithmetic functions.
   (c) Introduction to sieve methods, twin primes or Goldbach.
   (d) Asymptotics/bounds for various arithmetic functions. Proof of the prime number theorem.
   (e) Dirichlet's theorem on primes in arithmetic progressions
(4) Notation
   (a) Big theme: approximating discrete phenomena with continuous phenomena
   (b) How would one approximate $\lfloor x \rfloor$ (the integer part of x)? Write $\{x\} = x - \lfloor x \rfloor$ (the fractional part of x).
   (c) Instead of working with $\{x\}$ directly, we note that $0 \leq \{x\} < 1$
   (d) Thus $\lfloor x \rfloor = x + O(1)$, where $O(1)$ means that the error term in the approximation of $\lfloor x \rfloor$ by x belongs to the class of bounded functions.

   Definition 1.1. We write $f(x) = O(g(x))$ for $x \in I$, or equivalently $f(x) \ll g(x)$ for $x \in I$, if there exists a constant (depending at most on f, g, and I) such that $|f(x)| \leq c \cdot |g(x)|$ for each $x \in I$. In this case, we say that the order of magnitude of f in I is small than that of g. Usually (but not always), $I = \mathbb{R}$.

   (e) Since the sum of 2 bounded functions is also bounded, $O(1) + O(1) = O(1)$
   (f) Similarly, $O(g(x)) \pm O(g(x)) = O(g(x))$
   (g) The point: Allows us to write complicated inequalities compactly
   (h) Also, it allows us to write inequalities as *asymptotic equalities*

   Definition 1.2. If $|f(x) - g(x)| \leq c \cdot h(x)$ for $x \in I$, where c is a constant, then we write $f(x) = g(x) + O(h(x))$ for $x \in I$. We say that $f(x) = o(g(x))$ (as $x \to x_0$) when $\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$. We say that $f(x) \sim g(x)$ (as $x \to x_0$) when $\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1$.

2. April 4

Conjecture 2.1 (Gauss). *If* $\pi(x) := \#\{p \leq x\}$*, then*

$$\pi(x) \sim \mathrm{Li}(x), \qquad \mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}.$$

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

Why would we guess this as the right asymptotic for $\pi(x)$? We begin by recalling the Sieve of Eratosthenes. Take the numbers $n \leq x$ and cross out multiples of the primes which lie below $\sqrt{x}$, and the numbers left behind will be 1 or prime.

More generally, if we take $y \leq x$ and cross out multiples of primes below y, then we will be left with the set

$$\{n \leq x \colon \text{each prime } p \mid n \text{ satisfies } p > y\}.$$

Thus we may expect that if $1 \leq y \leq \sqrt{x}$ is chosen appropriately, then

$$\begin{aligned} &\#\{n \leq x \colon \text{each prime } p \mid n \text{ satisfies } p > y\} - 1 + \pi(y) \\ &= \#\{n \leq x \colon \text{each prime } p \mid n \text{ satisfies } p > y\} + O(y) \end{aligned}$$

should be a reasonable approximation for $\pi(x)$.

We now appeal to the principle of inclusion-exclusion to understand this set. First, we introduce some simplifying notation.

Definition 2.2. We define $\pi(x, y)$ to be $\#\{n \leq x \colon \text{each prime } p \mid n \text{ satisfies } p > y\}$.

With this definition,

$$\begin{aligned} \pi(x,y) &= \lfloor x \rfloor - \sum_{\substack{p \text{ prime} \\ p \leq y}} \left\lfloor \frac{x}{p} \right\rfloor + \sum_{\substack{p_1, p_2 \text{ prime} \\ p_1 \neq p_2 \\ p_1, p_2 \leq y}} \left\lfloor \frac{x}{p_1 p_2} \right\rfloor - \sum_{\substack{p_1, p_2, p_3 \text{ prime} \\ p_1 \neq p_2 \neq p_3 \\ p_1, p_2, p_3 \leq y}} \left\lfloor \frac{x}{p_1 p_2 p_3} \right\rfloor + \sum_{\substack{p_1, p_2, p_3, p_4 \text{ prime} \\ p_1 \neq p_2 \neq p_3 \neq p_4 \\ p_1, p_2, p_3, p_4 \leq y}} \left\lfloor \frac{x}{p_1 p_2 p_3 p_4} \right\rfloor - \dots \\ &= \lfloor x \rfloor + \sum_{j=1}^{\infty} (-1)^j \sum_{\substack{p_1, \dots, p_j \text{ prime} \\ p_1 \neq \dots \neq p_j \\ p_1, \dots, p_j \leq y}} \left\lfloor \frac{x}{p_1 \cdots p_j} \right\rfloor. \end{aligned}$$

(Note that this is finite because $\lfloor t \rfloor = 0$ for all $0 \leq t < 1$.) This motivates the introduction of the Möbius function.

Definition 2.3. A function $f : \mathbb{N} \to \mathbb{N}$ is called **multiplicative** if $f(1) = 1$ and for every pair of coprime integers m and n, we have that $f(mn) = f(m)f(n)$. We say that f is **totally multiplicative** (or **completely multiplicative**) if $f(mn) = f(m)f(n)$ for *all* m and n.

Definition 2.4. The Möbius function $\mu(n)$ is the multiplicative function defined on prime powers by $\mu(p) = -1$ and $\mu(p^k) = 0$ for $k \geq 2$.

Definition 2.5. We say that a positive integer $n \geq 1$ is **squarefree** if for each prime $p \mid n$, one has that $p^2 \nmid n$.

Note that $\mu(n) = 0$ unless n is squarefree, in which case

$$\mu(n) = (-1)^{\#\{p \colon p \text{ prime}, \, p \mid n\}}.$$

With this notation, and observing that 1 is the only squarefree d with no prime factors,

$$\lfloor x \rfloor = \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ d \text{ has precisely zero prime factors}}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

and

$$-\sum_{\substack{p \text{ prime} \\ p \leq y}} \left\lfloor \frac{x}{p} \right\rfloor = \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ d \text{ has precisely one prime factor, and it is} \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

and

$$\sum_{\substack{p_1, p_2 \text{ prime} \\ p_1 \neq p_2 \\ p_1, p_2 \leq y}} \left\lfloor \frac{x}{p_1 p_2} \right\rfloor = \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ d \text{ has precisely two distinct prime factors, each being} \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

and for general $j \geq 1$,

$$(-1)^j \sum_{\substack{p_1, \ldots, p_j \text{ prime} \\ p_1 \neq \cdots \neq p_j \\ p_1, \ldots p_j \leq y}} \left\lfloor \frac{x}{p_1 \cdots p_j} \right\rfloor = \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ d \text{ has precisely } j \text{ distinct prime factors, each being} \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor.$$

We now combine the contributions from the squarefree d with no prime diviors, 1 prime divisor, 2 prime divisors, etc. to see that

$$\pi(x, y) = \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ \text{each prime } p \text{ which divides } d \text{ is} \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor.$$

But since $\lfloor \frac{x}{d} \rfloor = \frac{x}{d} + O(1)$, we find that

$$\begin{aligned} \pi(x, y) &= \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ \text{each prime } p \text{ which divides } d \text{ is} \leq y}} \mu(d) \Big(\frac{x}{d} + O(1) \Big) \\ &= x \sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ \text{each prime } p \text{ which divides } d \text{ is} \leq y}} \frac{\mu(d)}{d} + O\left(\sum_{\substack{d \geq 1 \\ d \text{ squarefree} \\ \text{each prime } p \text{ which divides } d \text{ is} \leq y}} |\mu(d)| \right) \end{aligned}$$

Note that $\{d \geq 1 \colon d \text{ squarefree, each } p \mid n \text{ is} \leq y\} = \{d \geq 1 \colon d \text{ divides } \prod_{p \leq y} p\}$. Thus

$$\pi(x, y) = x \sum_{d \mid \prod_{p \leq y} p} \frac{\mu(d)}{d} + O\left(\sum_{d \mid \prod_{p \leq y} p} |\mu(d)| \right)$$

As part of the first homework assignment, you will walk through the proof that

$$\pi(x, y) = x \prod_{p \leq y} \left(1 - \frac{1}{p} \right) + O(2^{\pi(y)}). \tag{2.1}$$

3. April 6

Consider (2.1). Once y is a bit larger than $\log x$, then the $2^{\pi(y)}$ error term dominates the main term, and we don't get any interesting information. Nevertheless, (2.1) gives us a potential candidate for a (hopefully) manageable quantity which might be asymptotic to $\pi(x)$. This suggests a preliminary problem for us to investigate: What is the size of

$$\prod_{p \leq y} \Big(1 - \frac{1}{p}\Big)?$$

Theorem 3.1 (Mertens)**.** *As $y \to \infty$*

$$\prod_{p \leq y} \Big(1 - \frac{1}{p}\Big) \sim \frac{e^{-\gamma}}{\log y}, \qquad \gamma = \lim_{N \to \infty} \Big(\sum_{n \leq N} \frac{1}{n} - \log N \Big) = 0.5772\ldots$$

The proposition and Mertens's theorem should suggest that $x/\log x$ is a natural size for the number of $p \leq x$. However, if we use (2.1) with $y = \sqrt{x}$ with Mertens's theorem, then together they suggest that

$$\pi(x) \sim x \prod_{p \leq \sqrt{x}} \Big(1 - \frac{1}{p}\Big) \sim x \frac{e^{-\gamma}}{\log \sqrt{x}} = 2e^{-\gamma} \frac{x}{\log x}.$$

But $2e^{-\gamma} = 1.229\ldots \neq 1$.

What went wrong? The problem is that while divisibility by small primes does behave independently, this fails when the primes are large. For example, a number below x cannot be divisible by 3 primes all larger than $x^{1/3}$. Our prediction did not take such facts into account.

Over the next several lectures, we will establish the following results.

Theorem 3.2 (Mertens)**.** *(1) As $x \to \infty$, we have*

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

(2) There exists a constant B such that as $x \to \infty$,

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

(3) As $x \to \infty$, we have

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

(I will not always be so precise with the behavior or x, unless the range of x is not clear from context.)

Additionally, we will establish the following elementary estimates for $\pi(x)$.

Theorem 3.3 (Chebyshev)**.** *Let*

$$a = \liminf_{x \to \infty} \frac{\pi(x)}{x/\log x}, \qquad A = \limsup_{x \to \infty} \frac{\pi(x)}{x/\log x}.$$

Then $\log 2 \leq a \leq A \leq \log 4$. Moreover, if $a = A$ (in which case the limit $\lim_{x \to \infty} \frac{\pi(x)}{x/\log x}$ exists), then $a = A = 1$.

To prove these results, we start with the unique factorization of integers into products of prime powers, and hope to understand the averages over primes by relating them to averages over integers.

Definition 3.4. We define the von Mangoldt function $\Lambda(n)$ by

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

For a prime p dividing n, write $p^\alpha \,||\, n$ to mean p^α is the largest power of p which divides n. We encode the factorization of n with the identity

$$\log n = \sum_{p^\alpha || n} \log p^\alpha = \sum_{p^\alpha || n} \alpha \log p = \sum_{p|n} \sum_{\substack{1 \leq j \leq \alpha \\ p^\alpha || n}} \log p = \sum_{d|n} \Lambda(n).$$

Now, for any integer $N \geq 1$, we have the identity

$$\log(N!) = \sum_{n \leq N} \log n = \sum_{n \leq N} \sum_{d|n} \Lambda(d) = \sum_{n \leq N} \sum_{\substack{d \leq N \\ d|n}} \Lambda(d).$$

We swap the order of summation to achieve

$$\log(N!) = \sum_{d \leq N} \Lambda(d) \sum_{\substack{n \leq N \\ d|n}} 1 = \sum_{d \leq N} \Lambda(d) \left\lfloor \frac{N}{d} \right\rfloor = N \sum_{d \leq N} \frac{\Lambda(d)}{d} + O\Big(\sum_{d \leq N} \Lambda(d) \Big).$$

A crucial observation which will be useful later is that $\Lambda(d)$ essentially picks out primes with weight $\log p$. While we are also summing over squares of primes, cubes of primes, etc., their contribution is small because there aren't too many prime powers. For instance,

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

Note that for any fixed prime p,

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

Therefore, since $\log t = O(\sqrt{t})$,

$$\sum_{p \leq \sqrt{N}} \log p \sum_{\substack{k \geq 2 \\ p^k \leq N}} \frac{1}{p^k} \ll \sum_{p \leq \sqrt{N}} \frac{\log p}{p^2} \ll \sum_{p \leq \sqrt{N}} \frac{\sqrt{p}}{p^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} = O(1).$$

Thus we may conclude

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

We are well on our way to part of Theorem 3.2, but we still need to understand how $\log(N!)$ grows as $N \to \infty$, and we still need to bound $\sum_{d \leq N} \Lambda(d)$. We will resume this next week.

4. April 9

Last time:

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

Note that

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

Once we establish the Chebyshev bound $\pi(N) \ll N/\log N$, we would have that $\psi(N) \ll N$, in which case

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

For intuition on how large $\log(N!)$, we can roughly approximate the sum with an integral and guess that

$$\log(N!) = \sum_{n \leq N} \log n \approx \int_1^N (\log t) dt = N \log N + O(N).$$

(We will make this rigorous and more accurate later.) Then

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

In other words,

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

We proved last time that

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

in which case we recover part of Theorem 3.2:

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

We have two tasks remaining (at least for the third claim). First, we must prove the Chebyshev bound $\pi(N) \ll N/\log N$, which was crucial to determine that $\psi(N) \leq \pi(N) \log N \ll N$. But how closely related are $\psi(N)$ and $\pi(N) \log N$? It is convenient to consider the close relative

$$\vartheta(N) = \sum_{p \leq N} \log p = \log \Big(\prod_{p \leq N} p \Big).$$

Note that

$$\psi(N) = \vartheta(N) + \vartheta(N^{1/2}) + \vartheta(N^{1/3}) + \cdots,$$

which establishes a close connection between $\psi(N)$ and $\vartheta(N)$.

Exercise 4.1. Prove that $\psi(N) = \vartheta(N) + O(\sqrt{N} \log N)$. Assuming Chebyshev's bound, improve the error term from $O(\sqrt{N} \log N)$ to $O(\sqrt{N})$.

Thus the question of comparing $\psi(N)$ with $\pi(N)\log N$ is the same as comparing $\vartheta(N)$ with $\pi(N)\log N$. Since most $n \le N$ are large (say $\ge 10^{-6}N$), we have for most n that $\log n \approx \log N$. We might expect that something similar holds for primes, that $\log p \approx \log N$ for most $p \le N$. Then we would have $\vartheta(N) \approx \pi(N)\log N$.

Second, we need a strong approximation for $\log(N!) = \sum_{n\le N} \log n$, which also is trying to quantify precisely that $\log n$ is usually close to $\log N$. Thus we have now several problems where we would like to transfer information for one kind of sum to a related kind of sum: for example from understanding $\sum_{n\le N} 1$ (easy), we'd like to understand $\sum_{n\le N} \log n$. Or from understanding $\pi(N) = \sum_{p\le N} 1$, we'd like to understand $\vartheta(N) = \sum_{p\le N} \log p$. Or from understanding $\sum_{p\le N} (\log p)/p$ (which we nearly have), we might want to understand $\sum_{p\le N} 1/p$ (another part of Mertens). These problems (and many others) can be tackled using a technique called **partial summation**, which we will develop next lecture.

We now begin our work towards Theorem 3.3. Our starting point is again

$$\log(N!) = \sum_{d\le N} \Lambda(d) \sum_{k\le N/d} 1 = \sum_{k\le K} \sum_{d\le N/k} \Lambda(d) = \psi(N) + \psi(N/2) + \psi(N/3) + \dots.$$

Take this identity at $2N$ and subtract off 2 copies of the identity with N, which yields

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

RHS = alternating series with monotonically decreasing terms. Thus

$$\psi(2N) - \psi(N) \le \log \frac{(2N)!}{(N!)^2} = \log \binom{2N}{N} \le \psi(2N).$$

Once we have a good approximation for $N!$, we will extract the Chebyshev bounds (along with part 3 of Mertens).

5. April 11

From last time:

$$\psi(2N) - \psi(N) \leq \log \binom{2N}{N} \leq \psi(2N).$$

Note that the middle binomial coefficient $\binom{2N}{N}$ is the largest of the $\binom{2N}{k}$ with $0 \leq k \leq 2N$. Also,

$$4^N = (1+1)^{2N} = \sum_{k=0}^{2N} \binom{2N}{k},$$

so

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

Thus

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

Since $\psi(2N) \leq \pi(2N) \log(2N)$, we obtain

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

or (by replacing x with the largest even number below x) that

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

This is a more precise form of one of Chebyshev's bounds:

$$\liminf_{x \to \infty} \frac{\pi(x)}{x/\log x} \geq \log 2.$$

Now to the upper bounds. From before, we have that

$$\psi(2N) - \psi(N) \leq \log \binom{2N}{N} \leq N \log 4.$$

By considering the integer part of x (as we did in the case of Chebyshev's lower bound), we find that

$$\psi(2x) - \psi(x) \leq x \log 4 + \begin{cases} \Lambda(2\lfloor x \rfloor) & \text{if } x \text{ is an integer,} \\ 0 & \text{otherwise} \end{cases} = x \log 4 + O(\log x).$$

Now we use the above bound with x replaced by $x/2$, $x/4$, $x/8$, etc. ($O(\log x)$ many times) to obtain

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

Thus

$$\psi(x) \leq x \log 4 + O(\log x).$$

Now,

$$\psi(2x) - \psi(x) = \sum_{x < n \leq 2x} \Lambda(n) \geq \sum_{x < p \leq 2x} \log p \geq (\log x) \sum_{x < p \leq 2x} 1 = (\pi(2x) - \pi(x)) \log x.$$

Thus

$$\pi(2x) - \pi(x) \le \frac{x}{\log x} \log 4 + O(\log x).$$

Using the geometric sum approach from our experience with ψ, we find that for any integer $2 \le K \le \log x$,

$$\pi(2x) - \pi(x/2^K) \le \log 4 \sum_{j=0}^{K} \frac{x/2^j}{\log(x/2^j)} + O(K \log x).$$

The tricky part here is how to deal with $\log(x/2^j)$. Intuitively, only the small j should be relevant, in which case $\log(x/2^j) \approx \log x$. To make this precise, let

$$K = \left\lfloor \frac{1}{2} \log\log x \right\rfloor.$$

Then $2^K \ll \sqrt{x}$, in which case

$$\pi(2x) - \pi(x/2^K) = \pi(2x) + O(\sqrt{x}).$$

If $0 \le j \le K$, then

$$\frac{1}{\log(x/2^j)} = \frac{1}{\log x - j \log 2} = \frac{1}{\log x} \Big(1 - \frac{j \log 2}{\log x}\Big)^{-1} = \frac{1}{\log x} \Big(1 + O\Big(\frac{j}{\log x}\Big)\Big).$$

(Here, we have used the fact that $1/(1-t) = 1 + O(t)$ for $|t| \le 1/2$.) Thus

$$\pi(2x) - \pi(x/2^K) = \pi(2x) + O(\sqrt{x}) \le \log 4 \sum_{j=0}^{K} \frac{x}{\log x} \frac{1}{2^j} \Big(1 + O\Big(\frac{j}{\log x}\Big)\Big) + O(\sqrt{x} + K \log x).$$

With our choice of K, the right-hand side of the above expression equals

$$\frac{2x}{\log x} \log 4 + \Big(\frac{x}{(\log x)^2}\Big).$$

(You should be able to fill in the details in this step.) We now conclude that

$$\pi(x) \le \frac{x}{\log x} \log 4 + O\Big(\frac{x}{(\log x)^2}\Big).$$

This completes the proof of Chebyshev's bounds.

Exercise 5.1. Prove that as $x \to \infty$,

$$(\log 2 - o(1)) \le \frac{\psi(x)}{x} \le (\log 4 + o(1)), \qquad (\log 2 - o(1)) \le \frac{\vartheta(x)}{x} \le (\log 4 + o(1)).$$

In order to proceed toward Mertens's theorem, it remains to estimate $\log(N!)$. To handle this, we will develop the method of partial summation, which also will enable us to prove the last part of Chebyshev's theorem.

6. April 13

We introduce partial summation, which will be a ubiquitous tool for the rest of the course. The idea is that if $a(n)$ is a function on the integers and we have some information about the partial sums

$$A(x) = \sum_{n \leq x} a(n),$$

then we should be able to extract some information about the sum $\sum_{n \leq x} a(n)f(n)$, where f is a "suitably nice" function on the positive reals. Think of f as being twice continuously differentiable. In applications, we will often have $f(t)$ equal to t, $1/t$, $\log t$, $1/\log t$, or something else of an explicit nature which is relatively well-behaved.

Partial summation can be thought of as a discretized version of integration by parts. Recall that if $u(t)$ and $v(t)$ are continuously differentiable, then

$$\int u(t)v'(t)dt = u(t)v(t) - \int u'(t)v(t)dt.$$

The idea is that in the discrete version, $f(n)$ behaves like $u(t)$ and $A(x)$ behaves like $v(t)$. To make this precise, suppose that f is twice continuously differentiable. Then

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

Upon rearranging, we find:

Proposition 6.1. *Let $a(n)$ be a function on the integers, and let $A(x) = \sum_{n \leq x} a(n)$. If f is a twice continuously differentiable function on the positive reals, then*

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

This simple idea is incredibly powerful. Let's work out an important example.

Example 6.2. Let $a(n) = 1$ for all n, in which case $A(x) = \lfloor x \rfloor = x - \{x\}$. Let $f(t) = \log t$. Partial summation yields

$$\log(N!) = \sum_{n \leq N} \log n = N \log N - \int_1^N \frac{\lfloor t \rfloor}{t} dt = N \log N - \int_1^N \frac{t - \{t\}}{t} dt = N \log N - N + 1 + \int_1^N \frac{\{t\}}{t} dt.$$

Using the simple bound $0 \leq \{t\} < 1$, we already see that

$$N \log N - N + 1 \leq \log(N!) \leq N \log N - N + 1 + \log N,$$

which is already much more precise than our original guess indicated. But we can be more precise! By plotting $\{t\}$ on any interval of length 1, it looks like the average value of $\{t\}$ is $1/2$. In fact, by working with a piece-wise integral, one can compute directly that

$$B(x) = \int_1^x \{t\}dt = \frac{x}{2} + C(x), \qquad C(x) = O(1).$$

(Each of the roughly x intervals of length 1 gives a contribution of $1/2$, and a single interval may be left over.) Now, integrating by parts gives us

$$\int_1^N \frac{\{t\}}{t} dt = \frac{B(N)}{N} + \int_1^N \frac{B(t)}{t^2} dt = \frac{1}{2} + O(N^{-1}) + \frac{\log N}{2} + \int_1^N \frac{C(t)}{t^2} dt.$$

Since $C(t) = O(1)$, the last integral is $O(1)$, but we can say more: Note

$$\int_1^N \frac{C(t)}{t^2}dt = \int_1^\infty \frac{C(t)}{t^2} - \int_N^\infty \frac{C(t)}{t^2}dt = c_0 + O(N^{-1})$$

for some constant c_0. The point is that $\int_1^\infty \frac{C(t)}{t^2}dt$ converges, and the tail is only of size $O(1/N)$. We may now conclude that

$$\log(N!) = N\log N - N + \frac{1}{2}\log N + c_0 + O(1/N).$$

This gives us a precise version of Stirling's formula, up to the computation of c_0.

Exercise 6.3. Show that $c_0 = \frac{1}{2}\log(2\pi)$.

Exercise 6.4. Prove that

$$\sum_{n\le N}\frac{1}{n} = \log N + \gamma + O(N^{-1}), \qquad \text{where } \gamma = 1 - \int_1^\infty \frac{\{t\}}{t^2}dt$$

is Euler's constant.

This completes the proof of the limsup and liminf bounds in Chebyshev's theorem, as well as the asymptotic formula for $\sum_{p\le x}(\log p)/p$ in Theorem 3.2.

Now that we have established the 3rd part of Theorem 3.2

$$A(x) = \sum_{p\le x}\frac{\log p}{p} = \log x + E(x), \qquad E(x) = O(1),$$

we can use partial summation to estimate the sum $\sum_{p\le x}\frac{1}{p}$, which is the content of the 2nd part of Mertens's theorem. Let $a(n) = (\log n)/n$ if n is prime and $a(n) = 0$ otherwise, and let $f(t) = 1/\log t$. (How do we handle the fact that $f(t)$ is badly behaved near $t = 1$?) Thus by partial summation,

$$\sum_{p\le x}\frac{1}{p} = \sum_{n\le x}\frac{a(n)}{\log n} = \frac{\log x + E(x)}{\log x} - \int_2^x (\log t + E(t))\Big(\frac{1}{\log t}\Big)' = 1 + O\Big(\frac{1}{\log x}\Big) + \int_2^x \frac{\log t + E(t)}{t(\log t)^2}dt.$$

Then

$$\int_2^x \frac{\log t + E(t)}{t(\log t)^2}dt = \log\log x - \log\log 2 + \int_2^x \frac{E(t)}{t(\log t)^2}dt.$$

Because of the convergence of the integral, we may rewrite it as

$$\int_2^\infty \frac{E(t)}{t(\log t)^2}dt - \int_x^\infty \frac{E(t)}{t(\log t)^2}dt = c_1 + O\Big(\int_x^\infty \frac{1}{t(\log t)^2}dt\Big) = c_1 + O\Big(\frac{1}{\log x}\Big).$$

Putting everything together, we recover part of Theorem 3.2:

$$\sum_{p\le x}\frac{1}{p} = \log\log x + B + O\Big(\frac{1}{\log x}\Big), \qquad B = c_1 - \log\log 2.$$

7. April 16

Final part of Mertens's theorem:

$$\prod_{p \leq x} \Big(1 - \frac{1}{p}\Big) \sim \frac{C}{\log x}.$$

(HW Problem: Determine C)

Take logarithms:

$$\sum_{p \leq x} \log \Big(1 - \frac{1}{p}\Big) = -\log\log x + \log C + o(1).$$

Taylor's theorem: If $|t| \leq 1/2$, then

$$\log(1-t) = -t - \frac{t^2}{2} - \frac{t^3}{3} + \cdots.$$

so for such t

$$\log(1-t) + t = O(t^2).$$

This sets us up to rearrange the sum:

$$\sum_{p \leq x} \log \Big(1 - \frac{1}{p}\Big) = -\sum_{p \leq x} \frac{1}{p} + \sum_{p \leq x} \Big(\log \Big(1 - \frac{1}{p}\Big) + \frac{1}{p}\Big).$$

By Mertens,

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

while the second term equals

$$\begin{aligned}
\sum_{p \leq x} \Big(\log \Big(1 - \frac{1}{p}\Big) + \frac{1}{p}\Big) &= \sum_{p} \Big(\log \Big(1 - \frac{1}{p}\Big) + \frac{1}{p}\Big) - \sum_{p > x} \Big(\log \Big(1 - \frac{1}{p}\Big) + \frac{1}{p}\Big) \\
&= \sum_{p} \Big(\log \Big(1 - \frac{1}{p}\Big) + \frac{1}{p}\Big) + O\Big(\sum_{p > x} \frac{1}{p^2}\Big) \\
&= const + O\Big(\sum_{n > x} \frac{1}{n^2}\Big) \\
&= const + O(1/\log x).
\end{aligned}$$

Taking exponents, this shows that

$$\prod_{p \leq x} \Big(1 - \frac{1}{p}\Big) = \frac{C}{\log x}\Big(1 + O\Big(\frac{1}{\log x}\Big)\Big),$$

which is a more precise version of Theorem 3.2.

Let's now finish Theorem 3.3: If $\lim_{x \to \infty} \pi(x)/(x/\log x)$ exists, then the limit must equal 1. With A the limsup and a the liminf, we will show that $A \geq 1$ and $a \leq 1$ using Theorem 3.2. Thus if the limit exists, then $A = a = 1$.

Let $\epsilon > 0$, and let $T = T(\epsilon) > 0$ be a constant which is large (depending on ϵ). From the definition of limsup, we have for any $\epsilon > 0$ that if $t \geq T$, then

$$\pi(t) \leq (A + \epsilon)\frac{t}{\log t}.$$

By partial summation,

$$\sum_{p\le x}\frac{\log p}{p}=\pi(x)\frac{\log x}{x}-\int_2^x\pi(t)\Big(\frac{\log t}{t}\Big)'dt.$$

Using the Chebyshev bound $\pi(x)\ll x/\log x$, the above expression equals

$$\begin{aligned}O(1)+\int_2^x\pi(t)\Big(\frac{\log t}{t^2}-\frac{1}{t^2}\Big)dt&\le O_\epsilon(1)+\int_T^x(A+\epsilon)\frac{t}{\log t}\Big(\frac{\log t}{t^2}+\frac{1}{t^2}\Big)dt\\&=(A+\epsilon)\log x+O_\epsilon(\log\log x).\end{aligned}$$

(If $f=O_\epsilon(g)$, then $|f|\le c_\epsilon|g|$, where c_ϵ is a positive constant depending at most on ϵ. You could compute what this dependency is, but it is not important for our arguments.) Thus

$$\sum_{p\le x}\frac{\log p}{p}\le(A+\epsilon)\log x+O_\epsilon(\log\log x).$$

Taking x sufficiently large, we must have that $A+\epsilon\ge 1$. Since $\epsilon>0$ can be made arbitrarily small, we must have that $A\ge 1$. By a symmetric argument using the infimum, we also see that $a\le 1$, which concludes the proof of final part of Theorem 3.3.

We have already shown that $\psi(x)$ and $\vartheta(x)$ are roughly $\pi(x)\log x$, but this can be made more precise using partial summation. In particular, if $\vartheta(x)=x+E(x)$, then

$$\pi(x)=\mathrm{Li}(x)+\frac{E(x)}{\log x}+O(1)+\int_2^x\frac{E(t)}{t(\log t)^2}.$$

where $\mathrm{Li}(x)=\int_2^x dt/(\log t)$. Recall from the HW that as $x\to\infty$,

$$\mathrm{Li}(x)=\frac{x}{\log x}+\frac{x}{(\log x)^2}+O\Big(\frac{x}{(\log x)^3}\Big).$$

This establishes that $\mathrm{Li}(x)$ is a much more accurate approximation for $\pi(x)$ than $x/\log x$. Historically speaking, $\mathrm{Li}(x)$ was the function that Gauss conjectured should approximate $\pi(x)$; this was based on extensive (for his time) numerical data he gathered.

Thus we see that $\pi(x)\sim\mathrm{Li}(x)\iff\vartheta(x)\sim x$ (or equivalently if $E(x):=\vartheta(x)-x=o(x)$). We have that $|\vartheta(x)-\psi(x)|\ll\sqrt{x}$ (from the homework). Thus $\vartheta(x)\sim x\iff\psi(x)\sim x$.

8. April 18

In keeping with our desire for precision, we might ask how large

$$|\pi(x) - \mathrm{Li}(x)|, \qquad |\vartheta(x) - x|, \qquad |\psi(x) - x|$$

might be if we knew that the prime number theorem were true.

Conjecture 8.1 (The Riemann Hypothesis)**.** $\psi(x) = x + O(x^{1/2}(\log x)^2)$.

Exercise 8.2. Show that

$$\psi(x) = x + O(x^{1/2}(\log x)^2) \iff \vartheta(x) = x + O(x^{1/2}(\log x)^2) \iff \pi(x) = \mathrm{Li}(x) + O(x^{1/2}\log x).$$

The Riemann hypothesis predicts that $|\pi(x) - \mathrm{Li}(x)|$ is quite small. Indeed, $\sqrt{x}\log x$ has only about half as many digits as $\mathrm{Li}(x)$. This concludes our preliminary discussion on the distribution of primes.

We now begin our inquiry into the mean value and distribution of an arithmetic function. We begin by studying how many prime numbers divide a typical integer $n \geq 2$.

Define

$$\omega(n) = \sum_{p|n} 1, \qquad \Omega(n) = \sum_{p^a || n} a.$$

Thus $\omega(n)$ counts the distinct primes dividing n while $\Omega(n)$ counts the primes dividing n with multiplicity. Some questions for general arithmetic functions $a(n)$:

(1) What is the maximal size of $a(n)$?
(2) What is the minimal size of $a(n)$?
(3) What is the average size of $a(n)$?
(4) Does $a(n)$ stay near its average, or does it fluctuate wildly?

Let's consider these questions for $\omega(n)$. Clearly $\omega(p) = 1$ on the primes. Now, how large is $\max_{n \leq N} \omega(n)$? For Ω, this is easy: $\max_{n \leq N} \Omega(n) = \lfloor \frac{\log N}{\log 2} \rfloor$, which is attained at the largest power of 2 below N. Instead of proceeding in this fashion for ω, we may ask, given k, what is the smallest n with $\omega(n) = k$. Clearly the answer is $n = p_1 p_2 \cdots p_k$, where $p_1 < p_2 < \dots$ denotes the sequence of primes. Thus

$$n = \exp\Big(\sum_{j \leq k} \log p_j\Big) = \exp(\vartheta(p_k)).$$

By the Chebyshev bounds, there exist constants $0 < c \leq 1 \leq C$ such that if k is large, then

$$\exp(cp_k) \leq n \leq \exp(Cp_k).$$

From the HW, there exist constants $0 < c_1 \leq 1 \leq C_1$ such that $c_1 k \log k \leq p_k \leq C_1 k \log k$ for k sufficiently large, so

$$\exp(cc_1 k \log k) \leq n \leq \exp(CC_1 k \log k).$$

Exercise 8.3. Show that the above equation implies that

$$k \asymp \frac{\log n}{\log \log n},$$

in which case the maximal order of $\omega(n)$ for $n \leq N$ is $(\log N)/\log\log N$ (a bit smaller than the story for $\Omega(n)$).

Now for the average order of $\omega(n)$. We now proceed to compute the asymptotics for

$$\frac{1}{N}\sum_{n\le N}\omega(n).$$

From the definition of $\omega(n)$,

$$\frac{1}{N}\sum_{n\le N}\sum_{p|n}1=\frac{1}{N}\sum_{p\le N}\sum_{\substack{n\le N\\ p|N}}=\frac{1}{N}\sum_{p\le N}\lfloor\frac{N}{p}\rfloor=\frac{1}{N}\sum_{p\le N}\Big(\frac{N}{p}+O(1)\Big).$$

By Mertens and Chebyshev, this equals

$$\sum_{p\le N}\frac{1}{p}+O\Big(\frac{\pi(N)}{N}\Big)=\log\log N+B+O\Big(\frac{1}{\log N}\Big),$$

so on average a number below N has about $\log\log N$ prime factors.

Let us now call $\bar\omega$ the average value of $\omega(n)$ for $n\le N$, so that $\bar\omega=\log\log N+B+O(1/\log N)$. To understand the distribution of ω better, we now compute the variance:

$$\frac{1}{N}\sum_{n\le N}(\omega(n)-\bar\omega)^2=\frac{1}{N}\sum_{n\le N}\omega(n)^2-2\bar\omega\frac{1}{N}\sum_{n\le N}1+\bar\omega^2=\frac{1}{N}\sum_{n\le N}\omega(n)^2-\bar\omega^2.$$

Unraveling the definition of ω, we see that

$$\frac{1}{N}\sum_{n\le N}\omega(n)^2=\sum_{n\le N}\Big(\sum_{p|n}1\Big)^2=\frac{1}{N}\sum_{n\le N}\sum_{p_1|n}\sum_{p_2|n}1=\frac{1}{N}\sum_{p_1,p_2\le N}\sum_{\substack{n\le N\\ p_1|n\\ p_2|n}}1.$$

If $p_1\neq p_2$, then the inner sum over n counts the multiples of p_1p_2. Hence

$$\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2}}\sum_{\substack{n\le N\\ p_1|n\\ p_2|n}}1=\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2}}\sum_{\substack{n\le N\\ p_1p_2|n}}1=\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2\\ p_1p_2\le N}}\Big(\frac{N}{p_1p_2}+O(1)\Big)=\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2\\ p_1p_2\le N}}\frac{1}{p_1p_2}+O\Big(\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2\\ p_1p_2\le N}}1\Big).$$

Note that

$$\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2\\ p_1p_2\le N}}1\le\frac{1}{N}\sum_{n\le N}1=1$$

so

$$\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2}}\sum_{\substack{n\le N\\ p_1|n\\ p_2|n}}1=\sum_{\substack{p_1,p_2\le N\\ p_1\neq p_2\\ p_1p_2\le N}}\frac{1}{p_1p_2}+O(1).$$

But if $p_1=p_2$, then the inner sum only counts multiples of p_1. Thus

$$\frac{1}{N}\sum_{\substack{p_1,p_2\le N\\ p_1=p_2}}\sum_{\substack{n\le N\\ p_1|n\\ p_2|n}}1=\frac{1}{N}\sum_{p\le N}\sum_{\substack{n\le N\\ p|n}}1=\frac{1}{N}\sum_{n\le N}\sum_{p|n}1=\bar\omega.$$

9. April 20

Putting together everything from before, we see that

$$\frac{1}{N}\sum_{n\leq N}\omega(n)^2=\sum_{\substack{p_1\neq p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}+\bar{\omega}+O(1).$$

Note (perhaps by partial summation) that $\sum_{p_1^2\leq N}\frac{1}{p_1^2}=O(1)$, so

$$\sum_{\substack{p_1\neq p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}=\sum_{\substack{p_1,p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}+O(1).$$

Now, observe that

$$\Big(\sum_{p\leq\sqrt{N}}\frac{1}{p}\Big)^2\leq\sum_{\substack{p_1,p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}\leq\Big(\sum_{p\leq N}\frac{1}{p}\Big)^2.$$

We can asymptotically evaluate the upper and lower bounds using Mertens. The lower bound is

$$(\log\log\sqrt{N}+B+O(1/\log N))^2=(\log\log N)^2+O(\log\log N),$$

and the upper bound is

$$(\log\log N+B+O(1/\log N))^2=(\log\log N)^2+O(\log\log N).$$

Collecting our estimates, we see that

$$\sum_{\substack{p_1\neq p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}=\sum_{\substack{p_1,p_2\\ p_1p_2\leq N}}\frac{1}{p_1p_2}+O(1)=(\log\log N)^2+O(\log\log N).$$

Collecting our estimates, we find that

$$\frac{1}{N}\sum_{n\leq N}\omega(n)^2=(\log\log N)^2+O(\log\log N).$$

But

$$\bar{\omega}^2=(\log\log N)^2+O(\log\log N).$$

Therefore...

Theorem 9.1 (Hardy-Ramanujan; Turán)**.** *Over $n\leq N$, the function $\omega(n)$ has mean $\bar{\omega}=\log\log N+B+O(1/\log N)$ and variance*

$$\frac{1}{N}\sum_{n\leq N}(\omega(n)-\bar{\omega})^2\ll\log\log N.$$

Key point: The variance is rather small compared to the mean. Thus for $n\leq N$, $\omega(n)$ has a tendency to stay very close to its mean.

Corollary 9.2. *Let $f(N)$ be any function such that $\lim_{N\to\infty}f(N)=\infty$. The set*

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

satisfies

$$\#\mathcal{E}(N)\ll\frac{N}{f(N)^2}.$$

In other words, for $n \leq N$, the **normal order** *of $\omega(n)$ is $\log\log N$.*

Proof. If $n \in \mathcal{E}(N)$, then

$$|\omega(n) - \bar{\omega}| \geq f(N)\sqrt{\log\log N}.$$

Thus

$$\frac{1}{N}\#\mathcal{E}(N)(f(N)\sqrt{\log\log N})^2 = \frac{1}{N}\sum_{n\in\mathcal{E}(N)}(f(N)\sqrt{\log\log N})^2 \leq \frac{1}{N}\sum_{n\leq N}(\omega(n)-\bar{\omega})^2.$$

The right hand side is $O(\log\log N)$ by the theorem of Hardy-Ramanujan. The conclusion follows. □

Much more is true! A famous theorem of Erdős and Kac shows that for any fixed interval (α, β),

$$\lim_{N\to\infty}\frac{1}{N}\#\Big\{n \leq N\colon \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}} \in (\alpha,\beta)\Big\} = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-x^2/2}dx.$$

In other words, for $n \leq N$, the quantity

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

is normally distributed with mean 0 and variance 1. This follows from computing **all** integral moments of $\omega(n)$ (whereas the normal order of $\omega(n)$ follows from the first two moments, which we have computed).

Using these ideas, Erdős proved the following beautiful theorem.

Theorem 9.3 (Erdős)**.** *The number of distinct integers in the $N \times N$ multiplication table is $o(N^2)$.*

Proof. By reasoning which is essentially similar to the case with $\omega(n)$, Corollary 9.2 also applies when ω is switched out with Ω. In particular, for all except $o(N)$ of the integers $n \leq N$, we have that

$$|\Omega(n) - \log\log N| < (\log\log N)^{2/3}.$$

We see from the definition of Ω that $\Omega(ab) = \Omega(a) + \Omega(b)$ for all integers $1 \leq a, b \leq N$. Therefore, for all but at most $o(N^2)$ entries ab in the $N \times N$ table, we have

$$|\Omega(ab) - 2\log\log N| < 2(\log\log N)^{2/3}.$$

But note that $\log\log(N^2) \sim \log\log N$. Therefore, by the normal order of Ω, only $o(N^2)$ integers $n \leq N^2$ have the property that $\Omega(n)$ is near $2\log\log N$. Therefore, most pairs (a, b) lead to numbers ab with an atypically large number of prime power divisors compared to most integers $n \leq N^2$. Therefore, there can be only $o(N^2)$ numbers in the table, as desired. □

Next week, we will develop techniques to study the distribution of the divisor function

$$d(n) = \sum_{d|n} 1.$$

10. April 23

We now apply the ideas in the study of $\omega(n)$ to the study of

$$d(n) = \sum_{d|n} 1 = \sum_{ab=n} 1.$$

This is a multiplicative function (by HW1), but it is not completely multiplicative: $d(p^k) = k+1$.

As with $\omega(n)$, the minimal order of $d(n)$ is easy since $d(p) = 2$ for every prime p. Now on to the maximal order

$$\max_{n \leq N} d(n).$$

First, let us try to construct large values of $d(n)$. By HW3,

$$n = p_1 p_2 \cdots p_k = e^{\vartheta(p_k)} \leq e^{(\log 4 + o(1))p_k} \leq e^{\frac{\log 4 + o(1)}{\log 2 - o(1)} k \log k},$$

where p_n is the n-th prime number. For such an n, $d(n) = k$. Inverting this relationship, we see that

$$\max_{n \leq N} d(n) \gg \exp\Big(c \frac{\log N}{\log \log N}\Big).$$

In order to relate $d(n)$ to the functions $\omega(n)$ and $\Omega(n)$, we note the simple identities

$$2^{\omega(n)} \leq d(n) \leq 2^{\Omega(n)}.$$

But these don't give us good upper bounds on the maximal order of $\Omega(n)$ (recall our upper bound for the maximal order of $\Omega(n)$).

We will prove that for every $\epsilon > 0$, there exists a positive constant $C(\epsilon)$ such that

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

To see this, note that for any $\epsilon > 0$,

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

Note that there exists a constant $r(\epsilon)$ such that if $p > r(\epsilon)$, then the maximum value for $(a+1)/p^{a\epsilon}$ is 1 (attained when $a = 0$). Therefore,

$$\prod_{p} \max_{a \geq 0} \frac{a+1}{p^{a\epsilon}} = \prod_{p \leq r(\epsilon)} \max_{a \geq 0} \frac{a+1}{p^{a\epsilon}}.$$

This depends only on ϵ; we may take $C(\epsilon)$ to be this truncated product.

Exercise 10.1. Prove that there exists an absolute constant $c > 0$ such that $d(n) \ll \exp(c(\log n)/\log \log n)$.

Finally, we consider the average value of $d(n)$:

$$\begin{aligned}\frac{1}{N}\sum_{n\le N} d(n) = \frac{1}{N}\sum_{n\le N}\sum_{d|n} 1 &= \frac{1}{N}\sum_{n\le N}\sum_{\substack{d\le N\\ d|n}} 1\\ &= \frac{1}{N}\sum_{d\le N}\sum_{\substack{n\le N\\ d|n}} 1\\ &= \frac{1}{N}\sum_{d\le N}\lfloor\frac{N}{d}\rfloor = \frac{1}{N}\sum_{d\le N}\Big(\frac{N}{d}+O(1)\Big) = \log N + O(1).\end{aligned}$$

We will work on improving the error term (this will be important for our proof of the prime number theorem), but for now, let us observe some curious behavior. Recall that for a typical $n \le N$, $\omega(n)$ and $\Omega(n)$ are typically of size $\log\log N$. From our bound

$$2^{\omega(n)} \le d(n) \le 2^{\Omega(n)},$$

we see that $d(n)$ is typically of size $2^{\log\log N} = (\log N)^{\log 2}$.

However, we just showed that the mean value of $d(n)$ for $n \le N$ is $\log N$. So the average value of $d(n)$ does not capture its true size; instead, the average value is dominated by the large values of $d(n)$.

Exercise 10.2. Compute the variance of $d(n)$ for $n \le N$. How do the mean and variance of $d(n)$ compare?

11. April 25

Our goal is to refine the asymptotic for the mean value of

$$d(n) = \sum_{d|n} 1.$$

Using the definition of $d(n)$, we showed that

$$\frac{1}{N} \sum_{n \leq N} d(n) = \frac{1}{N} \sum_{d \leq n} \left\lfloor \frac{N}{d} \right\rfloor.$$

But we approximated $\lfloor \frac{N}{d} \rfloor$ with $N/d + O(1)$, which resulted in a fairly large error term in the mean value. We will remedy this by using geometric considerations.

Consider once again the sum

$$\sum_{n \leq N} d(n) = \sum_{n \leq N} \sum_{\substack{a,b \leq N \\ ab = n}} 1 = \sum_{\substack{a,b \leq N \\ ab \leq N}} 1.$$

Let $1 \leq A \leq N$ and $1 \leq B \leq N$ be parameters such that $AB = N$; we will choose A and B optimally later. Our goal is to count the number of pairs (a, b) with $ab \leq N$. We now break this sum into pieces: we could have

(1) $a \leq A$ and $b \leq N$,
(2) $a \leq N$ and $b \leq B$,
(3) $a \leq A$ and $b \leq B$.

But in counting the pairs (a, b) in cases (1) and (2), we count the pairs (a, b) in case (3) twice, so we need to subtract off that contribution. Thus

$$\sum_{\substack{a \leq N \\ b \leq N \\ ab \leq N}} 1 = \sum_{\substack{a \leq A \\ b \leq N \\ ab \leq N}} 1 + \sum_{\substack{a \leq N \\ b \leq B \\ ab \leq N}} 1 - \sum_{\substack{a \leq A \\ b \leq B \\ ab \leq N}} 1 = \sum_{\substack{a \leq A \\ b \leq N \\ ab \leq N}} 1 + \sum_{\substack{a \leq N \\ b \leq B \\ ab \leq N}} 1 - \sum_{\substack{a \leq A \\ b \leq B}} 1.$$

(For the last equality, notice that the conditions $a \leq A$ and $b \leq B$ imply that $ab \leq N$ since $A \leq N$ and $B \leq N$.

For the sum corresponding to case (1), if $a \leq N$, $b \leq N$, and $ab \leq N$, then $b \leq N/a$ as $a \leq N$ varies. Thus

$$\begin{aligned} \sum_{\substack{a \leq A \\ b \leq N \\ ab \leq N}} 1 = \sum_{a \leq A} \sum_{b \leq N/a} 1 &= \sum_{a \leq A} \left(\frac{N}{a} + O(1) \right) \\ &= N(\log A + \gamma + O(A^{-1}) + O(A) \\ &= N(\log A + \gamma) + O(N/A + A) \\ &= N(\log A + \gamma) + O(A + B). \end{aligned}$$

We study the sum for case (2) similarly (in fact, we can read it off by a symmetry argument):

$$\sum_{\substack{a \leq N \\ b \leq B \\ ab \leq N}} 1 = \sum_{b \leq B} \sum_{a \leq N/b} 1 = N(\log B + \gamma) + O(A + B).$$

For case (3) (which must subtracted off!), we observe that

$$\sum_{\substack{a\le A\\ b\le B\\ ab\le N}} 1 = \Big(\sum_{a\le A} 1\Big)\Big(\sum_{b\le B} 1\Big) = (A+O(1))(B+O(1)) = AB + O(A+B) = N + O(A+B).$$

Summing the results for cases (1) and (2) and subtracting the result for case (3) yields

$$\sum_{n\le N} d(n) = N\log(AB) + N(2\gamma - 1) + O(A+B).$$

We now choose A and B to minimize the error, which is minimized when $A = B$, hence $A = B = \sqrt{N}$. Therefore, we conclude that

$$\sum_{n\le N} d(n) = N\log N + (2\gamma-1)N + O(\sqrt{N}),$$

or alternatively

$$\frac{1}{N}\sum_{n\le N} d(n) = \log N + 2\gamma - 1 + O(N^{-1/2}).$$

This is much more precise than our first attempt; this will be crucial in our proof of the prime number theorem. This method of summing the divisor function is called Dirichlet's hyperbola method.

In the homework, the idea of a convolution of two functions is introduced. In many instances, this allows us to express certain functions as being built from "simpler" functions. For example, $d(n) = (1*1)(n)$, where $1(n)$ is the function which returns 1 for every integer. If we have a function $f(n)$ which is a convolution of two functions $g(n)$ and $h(n)$, in which case

$$f(n) = \sum_{ab=n} g(n)h(n),$$

then we can apply the hyperbola method again. If $1 \le A \le N$ and $1 \le B \le N$ with $AB = N$, then

$$\begin{aligned}\sum_{n\le N} f(n) &= \sum_{ab\le N} g(a)h(b)\\ &= \sum_{\substack{ab\le N\\ a\le A}} g(a)h(b) + \sum_{\substack{ab\le N\\ b\le B}} g(a)h(b) - \sum_{\substack{a\le A\\ b\le B}} g(a)h(b)\\ &= \sum_{a\le A} g(a)\sum_{b\le N/a} h(b) + \sum_{b\le B} h(b)\sum_{a\le N/b} g(a) - \Big(\sum_{a\le A} g(a)\Big)\Big(\sum_{b\le B} h(b)\Big).\end{aligned}$$

As an example, recall that $\mu(n) = 0$ is n is not squarefree and $\mu(n) = (-1)^{\omega(n)}$ otherwise. Thus $\mu(n)^2 = 1$ for n squarefree and $\mu(n)^2 = 0$ otherwise. In particular, $\mu(p)^2 = 1$. This suggests that we may write $\mu^2 = 1 * h$ for some multiplicative function h, Indeed, we may define h implicitly in this manner.

If $\mu^2 = 1*h$, then

$$\mu^2(n) = \sum_{d|n} h(d).$$

By HW3, this is equivalent to saying that

$$h(n) = \sum_{d|n} \mu(d)^2 \mu(n/d).$$

From this, we can see that h is multiplicative (thus $h(1) = 1$) and

$$h(p^k) = \begin{cases} -1 & \text{if } k = 2, \\ 0 & \text{if } k = 1 \text{ or } k \geq 3. \end{cases}$$

(In particular, $h(n) = 0$ unless n is the square of a squarefree number.) Thus

$$\sum_{n \leq N} \mu(n)^2 = \sum_{ab \leq N} h(b) = \sum_{a \leq A} \sum_{b \leq N/a} h(b) + \sum_{b \leq B} h(b) \sum_{a \leq N/b} 1 - \Big(\sum_{a \leq A} 1 \Big) \Big(\sum_{b \leq B} h(b) \Big)$$

with $AB = N$. Thus

$$\sum_{n \leq N} \mu(n)^2 = \sum_{ab \leq N} h(b) = \sum_{a \leq A} \sum_{b \leq N/a} h(b) + \sum_{b \leq B} h(b) \sum_{a \leq N/b} 1 + O(A\sqrt{B}).$$

Notice in the subtracted term that the sum of 1 is much larger than the sum of $h(b)$. This suggests that we should make $B = X$ and $A = 1$, in which case the above analysis reduces to

$$\sum_{n \leq N} \mu(n)^2 = \sum_{b \leq B} h(b) \Big(\frac{N}{b} + O(1) \Big) + O(\sqrt{N}) = N \sum_{b \leq B} \frac{h(b)}{b} + O(\sqrt{N}).$$

Now, recall that $h(b)$ is nonzero only when b is the square of a squarefree number, and in this case the magnitude of $h(b)$ is 1. Thus

$$\sum_{b \leq B} \frac{h(b)}{b} = \sum_{b=1}^{\infty} \frac{h(b)}{b} + O\Big(\sum_{d^2 > N} \frac{1}{n^2} \Big) = \sum_{b=1}^{\infty} \frac{h(b)}{b} + O\Big(\frac{1}{\sqrt{N}} \Big).$$

By the multiplicativity of $h(b)$ and the work in HW3,

$$\sum_{b=1}^{\infty} \frac{h(b)}{b} = \prod_{p} \Big(1 - \frac{1}{p^2} \Big) = \Big(\sum_{n=1}^{\infty} \frac{1}{n^2} \Big)^{-1} = \frac{6}{\pi^2}.$$

Thus

$$\#\{n \leq x \colon n \text{ squarefree}\} = \sum_{n \leq N} \mu(n)^2 = \frac{6}{\pi^2} N + O(\sqrt{N}).$$

12. April 27

Let us recall a few of our results:

- From our work on the Chebyshev bounds, $\log n = \sum_{d|n} \Lambda(d) = (1 * \Lambda)(n)$.
- We have the chain of implications

$$\pi(x) = \mathrm{Li}(x) + o(\mathrm{Li}(x)) \iff \vartheta(x) = x + o(x) \iff \psi(x) = x + o(x).$$

- From last class,

$$\sum_{n \le x} d(n) = x \log x + (2\gamma - 1)x + O(\sqrt{x}).$$

- From our work on the Chebyshev bounds,

$$\sum_{n \le x} \log n = x \log x - x + O(\log x).$$

- We have the convolution identities

$$\mu * 1(n) = E(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1, \end{cases} \qquad f * E = f, \qquad d(n) = (1 * 1)(n).$$

- The hyperbola method allows us to estimate partial sums of functions of the form $f = g * h$ once we have a handle on the partial sums of $g(n)$ and $h(n)$.

With these in mind, we will prove:

Theorem 12.1. *Let*

$$M(x) = \sum_{n \le x} \mu(n).$$

If $M(x) = o(x)$, then $\psi(x) = x + o(x)$.

Since $\log = 1 * \Lambda$, we have that $\mu * \log = (1 * \mu) * \Lambda = E * \Lambda = \Lambda$. Now, notice that by combining our work for the partial sums of $d(n)$ and $\log n$, we have that

$$\begin{aligned} \sum_{n \le x} (\log n - d(n) + 2\gamma) &= x \log x - x + O(\log x) - x \log x - (2\gamma - 1)x + O(\sqrt{x}) + 2\gamma x \\ &= O(\sqrt{x}). \end{aligned}$$

Write $\alpha(n) = \log n - d(n) + 2\gamma$. Thus

$$\sum_{n \le x} \alpha(n) = O(\sqrt{x}).$$

Now, we can rewrite $\log(n)$ as $\log(n) = \alpha(n) + (1 * 1)(n) - 2\gamma$. Thus

$$\begin{aligned} \Lambda(n) = (\mu * \log)(n) &= \mu(n) * \alpha(n) + \mu * (1 * 1)(n) - 2\gamma(\mu * 1)(n) \\ &= \mu * \alpha(n) + (E * 1)(n) - 2\gamma E(n) \\ &= \mu * \alpha(n) + 1 - 2\gamma E(n). \end{aligned}$$

Now,

$$\sum_{n \le x} \Lambda(n) = x + O(1) + \sum_{n \le x} \alpha * \mu(n)$$

Thus the prime number theorem would follow if we know that

$$\sum_{n \leq x} \alpha * \mu(n) = o(x).$$

By the hyperbola method from last time, these partial sums equal

$$\sum_{a \leq A} \mu(a) \sum_{b \leq x/a} \alpha(b) + \sum_{b \leq B} \alpha(b) \sum_{a \leq N/b} \mu(a) - \Big(\sum_{a \leq A} \mu(a) \Big) \Big(\sum_{b \leq B} \alpha(b) \Big),$$

where $1 \leq A \leq x$, $1 \leq B \leq x$, and $AB = x$.

Now, we know that

$$\sum_{b \leq N} \alpha(b) = O(\sqrt{N}),$$

so

$$\sum_{a \leq A} \mu(a) \sum_{b \leq x/a} \alpha(b) = O\Big(\sum_{a \leq A} \frac{\sqrt{x}}{\sqrt{a}} \Big) = O(\sqrt{Ax}).$$

In order to proceed, we need to carefully spell out our hypothesis regarding the partial sums of $\mu(n)$. Our hypothesis reads that for any $\epsilon > 0$, there exists a constant $K(\epsilon) > 0$ such that if $K > K(\epsilon)$, then

$$\Big| \sum_{n \leq K} \mu(n) \Big| \leq \epsilon K.$$

Now, choose $A \geq K(\epsilon)$. Thus

$$\sum_{b \leq B} \alpha(b) \sum_{a \leq x/b} \mu(a) = O\Big(\epsilon x \sum_{n \leq B} \frac{|\alpha(b)|}{b} \Big) = O\Big(\epsilon x \sum_{b \leq B} |\alpha(b)| \Big).$$

Now, we observe that

$$\sum_{b \leq B} |\alpha(b)| \leq \sum_{b \leq B} (\log b + d(b) + 2\gamma) = B \log B + O(B),$$

so

$$\epsilon x \sum_{b \leq B} |\alpha(b)| \ll \epsilon x B \log B.$$

Finally,

$$\Big(\sum_{a \leq A} \mu(a) \Big) \Big(\sum_{b \leq B} \alpha(b) \Big) = O(\epsilon AB \log B) = O(\epsilon x \log B).$$

Collecting our estimates, we find that under our hypothesis for the partial sums of $\mu(n)$, for any $\epsilon > 0$ and any choice of parameters $A, B \in [1, x]$ which satsify $AB = x$,

$$\sum_{n \leq x} \Lambda(n) = x + O(1) + \sum_{n \leq x} \alpha * \mu(n) = x + O(\sqrt{Ax} + \epsilon x B \log B) = x + O\Big(\frac{x}{\sqrt{B}} + \epsilon x B \log B \Big).$$

We choose $B = (1/\epsilon)^{2/3}$, in which case $A = x/B$ is larger than $K(\epsilon)$ once x is sufficiently large. With this choice of A and B, we find that

$$\sum_{n \leq x} \Lambda(n) = x + O(\epsilon^{1/3} \log(1/\epsilon) x),$$

and $\epsilon^{1/3} \log(1/\epsilon) \to 0$ as $\epsilon \to 0$. This completes the proof of the theorem.

If you have a more precise upper bound for the partial sums of $\mu(n)$, then you can deduce a more precise version of the prime number theorem. This will be part of your homework assignment.

13. April 30

We have spent lots of time developing tools to understand the partial sums of various arithmetic functions. But the tools we have developed to date are not powerful enough on their own to bound partial sums like

$$\sum_{n \leq x} \mu(n), \qquad \sum_{n \leq x} \Lambda(n)$$

with the sort of asymptotic precision required for the Prime Number Theorem (PNT). One way to proceed is as in the original proof of PNT by Hadamard and de la Valleé Poussin (independently in 1896) which relies on results from complex analysis which are taught in a course like Math 116. Another way to proceed is to take the route of Selberg and Erdős (independently in 1948) which avoids the use of complex variables altogether, but it is fairly complicated (and we only half of the quarter left). As a compromise, we will use the results we have developed to date to the generating function

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}.$$

Our conclusions will not be as strong as one could prove using complex analysis, but for the PNT and related problems, we will come reasonably close.

Series of the shape given above are called **Dirichlet series**. We have already seen the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}, \qquad s > 1.$$

Without discussing convergence, $F(s)$ is simply a formal sum. But from HW2, if

$$\sum_{n \leq x} f(n) \ll x^{\delta}$$

for some $\delta \geq 0$, then in the region $s > \delta$,

$$F(s) = s \int_1^{\infty} \frac{\sum_{n \leq v} f(n)}{v^{s+1}} dv.$$

The proof proceeds just the same when $s = \sigma + it$ is a complex variable (with $\sigma, t \in \mathbb{R}$), except now the result is valid provided that $\sigma > \delta$.

Making the change of variables $x \mapsto e^u$, we see that

$$\frac{F(\sigma + it)}{\sigma + it} = \int_0^{\infty} \frac{\sum_{n \leq e^u} f(n)}{e^{\sigma u}} e^{-itu} du.$$

Definition 13.1. Let $f : \mathbb{R} \to \mathbb{C}$ be a function such that

$$\int_{-\infty}^{\infty} |f(v)| dv < \infty.$$

The function $\hat{f} : \mathbb{R} \to \mathbb{C}$ given by

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(v) e^{-2\pi i v \xi} dv$$

is called the **Fourier transform** of f.

In this way, we see that if $\sigma > \delta$, then we see that the map

$$t \mapsto \frac{F(\sigma + 2\pi i t)}{\sigma + 2\pi i t}$$

is the Fourier transform of the map

$$u \mapsto \alpha_\sigma(u) := \frac{1}{e^{\sigma u}} \sum_{n \leq e^u} f(n).$$

We would like to be able to "invert" this relationship so that we can relate the partial sums

$$\sum_{n \leq x} f(n)$$

to an integral involving $F(\sigma + it)$. We can do this whenever $\sigma > \delta$.

Theorem 13.2 (Fourier inversion formula)**.** *Suppose that f is "sufficiently nice" (in a way we will make precise next lecture). Then*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x \xi} d\xi.$$

We will prove this theorem next class.

So if $\sigma > \delta$ is given, then Fourier inversion implies that

$$\frac{\lim_{v \to u^+} \alpha_\sigma(v) + \lim_{v \to u^-} \alpha_\sigma(v)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F(\sigma + it)}{\sigma + it} e^{itu} dt,$$

where

$$\alpha(v) = \frac{1}{e^{\sigma v}} \sum_{n \leq e^u} f(n).$$

Now if we return back to x from e^u, we see that

$$\sum_{n < x} f(n) + \frac{f(x)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F(\sigma + it)}{\sigma + it} x^{\sigma + it} dt = \frac{1}{2\pi i} \int_{\Re(s) = \sigma_0} F(s) \frac{x^s}{s} ds, \tag{13.1}$$

where $\sigma_0 > \delta$ and $f(x) = 0$ if x is not an integer. This connects the partial sums of $f(n)$ with the analytic properties of $F(s)$.

14. May 2

In this lecture, we will prove the Fourier inversion formula. Let $f : \mathbb{R} \to \mathbb{C}$ be a continuous function such that

$$|f(x)| \ll_f s(1+|x|^2)^{-1} \qquad \text{for all } x.$$

We call the space of such functions $\mathcal{M} = \mathcal{M}(\mathbb{R})$.

Proposition 14.1. *If $f, g \in \mathcal{M}$, and $a, b \in \mathbb{C}$ and $\delta \in \mathbb{R}$, then*

(1) $\int_{-\infty}^{\infty} f(x-\delta)dx = \int_{-\infty}^{\infty} f(x)dx$.
(2) $\delta \int_{-\infty}^{\infty} f(\delta x)dx = \int_{-\infty}^{\infty} f(x)dx$, *provided* $\delta > 0$.
(3) $\lim_{\delta \to 0} \int_{-\infty}^{\infty} |f(x-\delta) - f(x)|dx = 0$.

Proof. (1) $\int_{[-N,N]} f(x-\delta)dx - \int_{[-N,N]} f(x)dx \to 0$
(2) $\delta \int_{[-N,N]} f(\delta x)dx = \int_{[-\delta N, \delta N]} f(x)dx$
(3) Use the formal definition of continuity.

□

Example 14.2. If $f(x) = e^{-\pi x^2}$, then f and all of its derivatives lie in $\mathcal{M}$.

Definition 14.3. For $f \in \mathcal{M}$, the **Fourier transform** of f is

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i \xi}dx, \qquad \xi \in \mathbb{R}.$$

We write $f(x) \to \hat{f}(\xi)$ to mean that $\hat{f}$ denotes the Fourier transform of ξ.

Proposition 14.4. *If $f \in \mathcal{M}$ and $\delta \in \mathbb{R}$, then*

(1) $f(x+\delta) \to \hat{f}(\xi)e^{2\pi i \delta \xi}$
(2) $f(x)e^{-2\pi i x \delta} \to \hat{f}(\xi + \delta)$
(3) $f(\delta x) \to \delta^{-1}\hat{f}(\delta^{-1}\xi)$, *provided* $\delta > 0$
(4) If $x^k|f^{(\ell)}(x)| \in \mathcal{M}$ *for all* $k, \ell \geq 0$, *then* $f'(x) \to 2\pi i \xi \hat{f}(\xi)$
(5) If $x^k|f^{(\ell)}(x)| \in \mathcal{M}$ *for all* $k, \ell \geq 0$, *then* $-2\pi i x f(x) \to \frac{d}{d\xi}\hat{f}(\xi)$.

Proof. (1) Follows from Proposition 14.1(1).
(2) Follows from definition of the Fourier transform.
(3) Follows from Proposition 14.1(2).
(4) Use integration by parts.
(5) Use the formal definition of the derivative.

□

We will take special note that

$$\Big(\int_{-\infty}^{\infty} e^{-\pi x^2}dx\Big)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\pi(x^2+y^2)}dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-\pi r^2} r dr d\theta = 1. \tag{14.1}$$

From this, we note that

Theorem 14.5. *If $f(x) = e^{-\pi x^2}$, then $\hat{f}(\xi) = f(\xi)$.*

Proof. Uses Proposition 14.4(4,5). □

Corollary 14.6. *If $\delta > 0$ and $K_\delta(x) = \delta^{-1/2}e^{-\pi x^2/\delta}$, then $\hat{K}_\delta(\xi) = e^{-\pi\delta\xi^2}$.*

Proof. Uses Proposition 14.4(3). □

Note that

(1) $\int_{-\infty}^{\infty} K_\delta(x)dx = \int_{-\infty}^{\infty} |K_\delta(x)|dx = 1$ (perform a change of variables and use (14.1))
(2) For every $\eta > 0$, we have $\int_{|x|>\eta} |K_\delta(x)|dx \to 0$ as $\delta \to 0$. (change of variables)

Definition 14.7. If $f, g \in \mathcal{M}$, their **convolution** is defined by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-t)g(t)dt.$$

For fixed x and varying t, the function $f(x-t)g(t) \in \mathcal{M}$.

Corollary 14.8. *If $f \in \mathcal{M}$, then $(f * K_\delta)(x) \to f(x)$ uniformly in x as $\delta \to 0$. In other words, for all $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that for all $0 < \delta < \delta(\epsilon)$, we have*

$$\sup_{x\in\mathbb{R}} |(f * K_\delta)(x) - f(x)| < \epsilon.$$

Proof. Note that for any $\eta > 0$, one has

$$|(f*K_\delta)(x)-f(x)| = \int_{-\infty}^{\infty} K_\delta(t)[f(x-t)-f(x)]dt \le \Big(\int_{|t|>\eta} + \int_{|t|\le\eta}\Big) K_\delta(t)|f(x-t)-f(x)|dt.$$

The proof now follows from the properties of K_δ and f, including Proposition 14.1(3). □

Proposition 14.9. *If $f, g \in \mathcal{M}$, then $\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx = \int_{-\infty}^{\infty} \hat{f}(y)g(y)dy$.*

Proof. If $F(x, y) = f(x)g(y)e^{-2\pi ixy}$, then

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)dxdy = \int_{-\infty}^{\infty} \hat{f}(y)g(y)dy.$$

□

Finally, we prove:

Theorem 14.10 (Fourier inversion)**.** *If $f \in \mathcal{M}$, then*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi ix\xi}d\xi.$$

Proof. By the transformation in Proposition 14.4(1), it suffices to prove that

$$f(0) = \int_{-\infty}^{\infty} \hat{f}(\xi)d\xi.$$

Let $G_\delta(x) = e^{-\pi\delta x^2}$, so that $\widehat{G_\delta}(\xi) = K_\delta(\xi)$. By Proposition 14.9,

$$\int_{-\infty}^{\infty} f(x)K_\delta(x)dx = \int_{-\infty}^{\infty} \hat{f}(\xi)G_\delta(\xi)d\xi.$$

By the aforementioned properties of K_δ, the first integral tends to $f(0)$ as $\delta \to 0$. The second integral clearly converges to $\int_{-\infty}^{\infty} \hat{f}(\xi)d\xi$, which finishes the proof. □

15. May 4

We now reap the benefits of the Fourier inversion formula. Recall that our original motivation was the identity

$$\sum_{n<x} f(n) + \frac{f(x)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F(\sigma+it)}{\sigma+it} x^{\sigma+it} dt, \qquad \sigma > \delta.$$

For our proof of the prime number theorem, a related formula will be more convenient.

Exercise 15.1. Let $s = \sigma + it$ be a complex number. Let $f : \mathbb{N} \to \mathbb{C}$ have the associated Dirichlet series $F(s)$ which converges absolutely in the region $\sigma \geq \sigma_0$. Prove that

$$\sum_{n \leq x} f(n) \log \frac{x}{n} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\sigma_0 + it) \frac{x^{\sigma_0+it}}{(\sigma_0+it)^2} dt$$

Another striking consequence is the following:

Theorem 15.2 (Plancherel)**.** *If $f \in \mathcal{M}$, then*

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi.$$

Proof. Let $f \in \mathcal{M}$, and define $f^\flat(x) = \overline{f(-x)}$. Then $\widehat{f^\flat}(\xi) = \overline{\hat{f}(\xi)}$. Now, let $h = f * f^\flat$. Then by the Fourier inversion formula,

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = h(0) = \int_{-\infty}^{\infty} \hat{h}(\xi) d\xi = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi.$$

□

These ideas extend results on other spaces of functions, such as the space of square-integrable functions on $[-\pi, \pi]$. In this setting, we can write Fourier inversion as

$$f(x) = \sum_{n \in \mathbb{Z}} c_n e^{inx}, \qquad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx.$$

(A different normalization for the Fourier transform is used in this formulation, but that is fine.) In this setting, the analogue of Plancherel's formula is Parseval's identity:

$$\sum_{n \in \mathbb{Z}} |c_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx,$$

Exercise 15.3. Use Parseval's identity to prove that $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, and $\zeta(6) = \pi^6/945$. Determine a recursive relationship that allows you to compute $\zeta(2n)$ for all integers $n \geq 1$.

We now discuss how one might determine $F(s)$ from better-known Dirichlet series using convolutions. Suppose $f, g : \mathbb{N} \to \mathbb{C}$ are functions, and let

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}, \qquad G(s) = \sum_{n=1}^{\infty} \frac{g(n)}{n^s},$$

where s is in the region where both F and G converge absolutely. A straightforward computation shows that

$$F(s)G(s) = \sum_{n=1}^{\infty} \frac{(f*g)(n)}{n^s}.$$

Moreover, if $E(1) = 1$ and $E(n) = 0$ for all $n > 1$, then the Dirichlet series for $E(n)$ is simply the 1 function. Therefore, if $f(1) = 1$ (which is the case when f is multiplicative), then $1/F(s)$ gives the Dirichlet series for the convolution inverse of f.

It will be important for us to be able to consider derivatives of $F(s)$, where the derivative is now with respect to a complex variable s instead of a real variable. In order to define the derivative of a function of a complex variable, we appeal to the limit definition of the derivative from multivariable calculus. For this discussion, we consider only points in the region for which $F(s)$ is absolutely convergent. (It is straightforward to check that if the sum defining $F(\sigma_0 + it_0)$ converges absolutely, then $F(s)$ converges absolutely for all $\Re(s) \geq \sigma_0$.)

If $s \neq s_0$ are two points in the region of absolute convergence, then we can consider

$$\frac{F(s) - F(s_0)}{s - s_0} = \sum_{n=1}^{\infty} f(n) \frac{1}{s - s_0}\Big(\frac{1}{n^s} - \frac{1}{n^{s_0}}\Big).$$

One can justify (since we are in the region of absolute convergence!) that

$$\frac{d}{ds}F(s) = \lim_{s\to s_0} \frac{F(s) - F(s_0)}{s - s_0} = \sum_{n=1}^{\infty} f(n) \lim_{s\to s_0} \frac{1}{s - s_0}\Big(\frac{1}{n^s} - \frac{1}{n^{s_0}}\Big).$$

The limit here is along all paths in the complex plane that lead to s_0, much like the way limits are defined for functions of two variables. In fact, we can evaluate this limit in this exact manner, in which case we conclude that

$$\frac{d}{ds}F(s) = -\sum_{n=1}^{\infty} \frac{f(n)\log n}{n^s}.$$

Exercise 15.4. Let $\sigma > 1$.

(1) Find the Dirichlet series for $\mu(n)$, $d_k(n)$, $|\mu(n)|$, $\varphi(n)$, and $2^{\omega(n)}$.

(2) Prove that if f is totally multiplicative, then

$$\sum_{n=1}^{\infty} \frac{f(n)\Lambda(n)}{n^s} = -\frac{F'(s)}{F(s)}.$$

(3) Define $\zeta^*(s) = (s-1)\zeta(s)$. Prove that for any integer $k \geq 1$,

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}(-\log m)^k = \Big(\frac{1}{\zeta(s)}\Big)^{(k)} = (s-1)\Big(\frac{1}{\zeta^*(s)}\Big)^{(k)} + k\Big(\frac{1}{\zeta^*(s)}\Big)^{(k-1)}.$$

(Here, $f^{(k)}$ denotes the k-th derivative of f.)

In the region of absolute convergence, we can also obtain a complex-variable version of a previous homework problem: If s is in the region of absolute convergence, then for f a multiplicative function,

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \Big(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \cdots\Big).$$

Moreover, if f is completely multiplicative, then

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_{p} \left(1 - \frac{f(p)}{p^s}\right)^{-1}, \qquad \frac{1}{F(s)} = \sum_{n=1}^{\infty} \frac{f(n)\mu(n)}{n^s} = \prod_{p} \left(1 - \frac{f(p)}{p^s}\right)$$

These results are already apparent for the Riemann zeta function.

16. May 7

We are now in a position to outline our approach to PNT.

Theorem 16.1. *Let $k \geq 4$ be an integer, and let*

$$F(x) = \sum_{n \leq x} \mu(n)(\log n)^k \log \frac{x}{n}.$$

We have the bound $F(x) \ll_k x(\log x)^{3(k+1)/4}$.

Assuming Theorem 16.1, we now prove the PNT. We will then concentrate our efforts on proving Theorem 16.1.

Proof of PNT assuming Theorem 16.1. Let x be large, and let $2 \leq y \leq x$. Define the function

$$H(x) = \sum_{n \leq x} \mu(n)(\log n)^k.$$

Consider the difference

$$\begin{aligned} F(x+y) - F(x) &= \sum_{n \leq x} \mu(n)(\log n)^k \Big(\log \frac{x+y}{n} - \log \frac{x}{n} \Big) + \sum_{x < n \leq x+y} \mu(n)(\log n)^k \frac{x+y}{n} \\ &= H(x) \log \frac{x+y}{x} + \sum_{x < n \leq x+y} \mu(n)(\log n)^k \log \frac{x+y}{n}. \end{aligned}$$

We bound the sum over $x < n \leq x+y$ trivially:

$$\sum_{x < n \leq x+y} \mu(n)(\log n)^k \log \frac{x+y}{m} \leq \log \frac{x+y}{x} \sum_{x < n \leq x+y} (\log n)^k \ll_k y(\log x)^k \log \frac{x+y}{x}.$$

Thus

$$F(x+y) - F(x) = \Big(H(x) + O_k(y(\log x)^k) \Big) \log \frac{x+y}{x},$$

so

$$H(x) = \frac{1}{\log \frac{x+y}{x}} (F(x+y) - F(x)) + O_k(y(\log x)^k).$$

By Taylor's theorem, if $y = o(x)$ and x is sufficiently large, then

$$\frac{1}{\log \frac{x+y}{x}} \ll \frac{x}{y}.$$

Given our range of y relative to x, we now use Theorem 16.1 to bound $F(x+y) - F(x)$ trivially:

$$\begin{aligned} F(x+y) - F(x) &\leq |F(x+y)| + |F(x)| \\ &\ll_k (x+y)(\log(x+y))^{3(k+1)/4} + x(\log x)^{3(k+1)/4} \\ &\ll_k x(\log x)^{3(k+1)/4}. \end{aligned}$$

Putting together our estimates, we find that

$$H(x) \ll_k y(\log x)^k + \frac{x^2}{y} (\log x)^{3(k+1)/4}.$$

Now, choose

$$y = \frac{x}{(\log x)^{(k-3)/8}}.$$

Substituting these choices into our bound for $H(x)$, we find that

$$H(x) \ll_k x(\log x)^{(7k+3)/8}.$$

Observe that the trivial bound for $H(x)$ is $O(x(\log x)^k)$. Thus our result is nontrivial once we select $k \geq 4$.

By partial summation, we see that this bound for $H(x)$ implies the bound

$$\sum_{n \leq x} \mu(n) = \sum_{n \leq x} \frac{\mu(n)(\log n)^k}{(\log n)^k} \ll_k \frac{x}{(\log x)^{k-(7k+3)/8}} \ll_k \frac{x}{(\log x)^{k/10}}.$$

for $k \geq 10$, say. It now follows from HW4 that

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = x + O_k\Big(\frac{x}{(\log x)^{k/40}}\Big).$$

From HW2, we have that $|\psi(x) - \vartheta(x)| \ll \sqrt{x}$, which is dominated by $x/(\log x)^A$ for any fixed $A > 0$ by HW1. Thus

$$\vartheta(x) = x + O_k\Big(\frac{x}{(\log x)^{k/40}}\Big).$$

By partial summation, we have

$$\pi(x) = \mathrm{Li}(x) + O_k\Big(\frac{x}{(\log x)^{k/40+1}}\Big).$$

Now, we may take k as large as we wish. In other words, if $A > 0$ and x is sufficiently large with respect to A, then we conclude that

$$\pi(x) = \mathrm{Li}(x) + O_A\Big(\frac{x}{(\log x)^A}\Big).$$

□

17. May 9

1792/1793 - Gauss predicts that $\pi(x) \sim \mathrm{Li}(x)$
1851 - Chebyshev proved that $\liminf \leq 1 \leq \limsup$ (and the Chebyshev bounds soon after)
1874 - Mertens (also proved a version for APs)
1859 - Riemann publishes his one paper in number theory. Consider

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

Theorem 17.1 (Riemann, 1859)**.** *Define the function*

$$\xi(s) = s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s), \qquad \Gamma(s/2) = \int_0^\infty e^{-t}t^{s/2-1}dt \text{ for } \Re(s) > 1.$$

Then ξ is complex differentiable everywhere in the complex plane, and $\xi(s) = \xi(1-s)$ for all $s \in \mathbb{C}$.

Proof. From the definition of Γ and the change of variables $t \mapsto n^2\pi x$, we obtain

$$\pi^{-s/2}\Gamma(s/2)n^{-s} = \int_0^\infty x^{s/2-1}e^{-n^2\pi x}dx.$$

Define

$$\omega(x) = \sum_{n=1}^{\infty} e^{-n^2\pi x}, \qquad \theta(x) = \sum_{n\in\mathbb{Z}} e^{-n^2\pi x}, \qquad 2\omega(x) = \theta(x) - 1.$$

Thus for $\sigma > 1$,

$$\begin{aligned}\pi^{-s/2}\Gamma(s/2)\zeta(s) &= \int_0^\infty x^{s/2-1}\omega(x)dx \\ &= \int_1^\infty x^{s/2-1}\omega(x)dx + \int_1^\infty x^{-s/2-1}\omega(1/x)dx\end{aligned}$$

Exercise 17.2. Prove that for $f \in \mathcal{M}$, we have that

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \hat{f}(m).$$

From this, deduce that $\theta(x^{-1}) = x^{1/2}\theta(x)$ for $x > 0$. From this, deduce that

$$\omega(x^{-1}) = -\frac{1}{2} + \frac{1}{2}x^{1/2} + x^{1/2}\omega(x).$$

With this exercise, it follows that

$$s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s) = 1 + s(s-1)\int_1^\infty (x^{s/2-1} + x^{-(s-1)/2})\omega(x)dx.$$

This holds for $\sigma > 1$, but the integral on the right converges rapidly for any s because $\omega(x) = O(e^{-\pi x})$ as $x \to \infty$. Note that even though $\zeta(s) \sim 1/(s-1)$ as $s \to 1^+$, it was shown in an earlier HW that $\lim_{s\to 1^+}(s-1)\zeta(s) = 1$. □

Theorem 17.3 (Hadamard)**.** *Suppose $f(s)$ is complex-differentiable everywhere in $\mathbb{C}$ and $|f(s)| \ll_\epsilon e^{|s|^{1+\epsilon}}$ as $|s| \to \infty$. Then there exist constants A and B such that*

$$f(s) = e^{A+Bs} \prod_{\substack{\rho \\ f(\rho)=0}} \Big(1 - \frac{s}{\rho}\Big) e^{s/\rho}.$$

Moreover, $\xi(s)$ satisfies the hypotheses.

By the Euler product representation of $\zeta(s)$, we find that $\zeta(s) \neq 0$ for $\Re(s) > 1$. By the functional equation, $\zeta(s) \neq 0$ for $\Re(s) < 0$ except when $s = -2, -4, -6, -8, \ldots$. These zeros exist to get counteract the poles of $s\Gamma(s/2)$; this must be the case because $\xi(s)$ is complex-differentiable *everywhere*; it can have not bad points. This implies that the zeros of $\xi(s)$ (which are also zeros of $\zeta(s)$) must lie in the **critical strip** $0 \leq \Re(s) \leq 1$. We call the zeros $-2, -4, -6, \ldots$ the **trivial zeros** of $\zeta(s)$, and we call the zeros ρ of $\xi(s)$ the **nontrivial zeros** of $\zeta(s)$.

From one of the current HW problems, we know that

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)}.$$

We have already reasoned out that if $\sigma \geq \sigma_0 > 1$, then

$$\sum_{n<x} \Lambda(n) + \frac{\Lambda(x)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} -\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} \frac{x^{\sigma+it}}{\sigma+it} dt = \frac{1}{2\pi i} \int_{\Re(s)=\sigma_0} -\frac{\zeta'(s)}{\zeta(s)} \frac{x^s}{s} ds.$$

The residue theorem (and more precisely, the argument principle) from complex analysis tells us that since $\zeta(s)$ can be continued to a complex-differentiable function outside of $s = 1$, then we can "push the integral to the left" and choose values of $\sigma < 1$. In doing so, we encounter singularities at each zero of $\zeta(s)$, at the pole of $\zeta(s)$ at $s = 1$, and at the pole of x^s/s at $s = 0$. Each one of those singularities contributes to the size of the integral. Thus one can prove the so-called **explicit formula**

$$\sum_{n<x} \Lambda(n) + \frac{\Lambda(x)}{2} = x - \sum_{\rho} \frac{x^\rho}{\rho} - \log(2\pi) - \frac{1}{2}\log(1 - x^{-2}).$$

(The $\log(2\pi)$ comes from the fact that x^s/s has a singularity at $s = 0$, and the $-\frac{1}{2}\log(1-x^{-2})$ comes from the contribution of the trivial zeros.)

Conjecture 17.4 (The Riemann Hypothesis)**.** *Each nontrivial zero ρ satisfies $\Re(\rho) = 1/2$.*

This gives the Prime Number Theorem with its optimal error term. Also, writing each zero ρ as $1/2 + it$, this the explicit formula can be seen as a Fourier expansion for $\psi(x)$.

Theorem 17.5 (Hadamard, de la Vallée Poussin, 1896)**.** *$\Re(\rho) < 1$ for each nontrivial ρ.*

This is the bare minimum one can get away with in order to prove the Prime Number Theorem. de la Vallée Poussin later proved that there exists a constant $c > 0$ such that

$$\Re(s) \geq 1 - \frac{c}{\log(2 + |\Im(s)|)} \implies \zeta(s) \neq 0$$

This results in the existence of some computable constant $c' > 0$ such that

$$\sum_{n \leq x} \Lambda(n) = x + O\Big(x \exp(-c'\sqrt{\log x})\Big).$$

18. May 14

We now have all of the tools we need to prove Theorem 16.1. Recall that the goal is to bound the sum

$$\sum_{n\le x}\mu(n)(\log n)^k\log\frac{x}{n}.$$

We apply Exercise 15.1 to see that this sum equals

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big(\sum_{n=1}^{\infty}\frac{\mu(n)(\log n)^k}{n^{\sigma_0+it}}\Big)\frac{x^{\sigma_0+it}}{(\sigma_0+it)^2}dt,$$

provided that σ_0 is large enough so that the series in the integral converges absolutely in the region $\Re(s)\ge\sigma_0$. For future convenience, we rewrite this as an integral over the complex variable $s=\sigma+it$ instead of an integral over the real variable t (like we did in (13.1)). Thus the above equals

$$\frac{1}{2\pi i}\int_{\Re(s)=\sigma_0}\Big(\sum_{n=1}^{\infty}\frac{\mu(n)(\log n)^k}{n^s}\Big)\frac{x^s}{s^2}ds. \tag{18.1}$$

But what Dirichlet series are we actually working with? By a HW assignment,

$$\sum_{n=1}^{\infty}\frac{\mu(n)(\log n)^k}{n^s}=(-1)^k\Big(\frac{d^k}{ds^k}\sum_{n=1}^{\infty}\frac{\mu(n)}{n^s}\Big)=(-1)^k\Big((s-1)\Big(\frac{1}{\zeta^*(s)}\Big)^{(k)}+k\Big(\frac{1}{\zeta^*(s)}\Big)^{(k-1)}\Big),$$

where $\zeta^*(s)=(s-1)\zeta(s)$. Thus (18.1) equals

$$\frac{1}{2\pi i}\int_{\Re(s)=\sigma_0}\Big((s-1)\Big(\frac{1}{\zeta^*(s)}\Big)^{(k)}+k\Big(\frac{1}{\zeta^*(s)}\Big)^{(k-1)}\Big)\frac{x^s}{s^2}ds.$$

This expression may seem complicated, but an identity from calculus will simplify our task for us.

Exercise 18.1. Let $k\ge1$ be an integer, and let f be a function of a complex variable. Suppose that the first k derivatives of f exist and are continuous at s. Prove that

$$\Big(\frac{1}{f(s)}\Big)^{(k)}=\frac{1}{f(s)}\sum_{a_1+2a_2+3a_3+\cdots+ka_k=k}\frac{(a_1+a_2+\cdots+a_k)!}{(a_1!)(a_2!)\cdots(a_k!)}\Big(\frac{-f'(s)}{1!f(s)}\Big)^{a_1}\Big(\frac{-f''(s)}{2!f(s)}\Big)^{a_2}\cdots\Big(\frac{-f^{(k)}(s)}{k!f(s)}\Big)^{a_k},$$

where $a_1,a_2,\dots,a_k$ run over the nonnegative integers.

While this identity may seem unsightly, it shifts the burden of the proof to two tasks:

(1) Obtain upper bounds for the derivatives of $\zeta(s)$ (hence $\zeta^*(s)$). We can do this using partial summation!
(2) Obtain a lower bound for $\zeta(s)$. This is the crux of the reason why it took so long for a proof of PNT to emerge.

We let $s=\sigma+it$, where $\sigma\in(1,2)$. We first obtain the upper bounds for the derivatives of $\zeta^*(s)=(s-1)\zeta(s)$. We know that

$$(-1)^k\zeta(s)^{(k)}=\sum_{n=1}^{\infty}\frac{(\log n)^k}{n^s}.$$

To simplify matters, we will separately estimate the head and the tail of the sum. Let us truncate this sum at some threshold X to be determined later. Then for $\Re(s) > 1$,

$$(-1)^k \zeta(s)^{(k)} = \sum_{1 \leq n \leq X} \frac{(\log n)^k}{n^s} + \sum_{n > X} \frac{(\log n)^k}{n^s} = \sum_{n > X} \frac{(\log n)^k}{n^s} + O((\log X)^{k+1}).$$

By partial summation, if

$$\sum_{n=1}^{\infty} |a(n) f(n)| < \infty, \qquad \sum_{n \leq x} a(n) = g(x) + O(|h(x)|),$$

then

$$\sum_{n > x} a(n) f(n) = -f(x) g(x) + O(|f(x) h(x)|) + \int_x^{\infty} (g(t) f'(t) + O(|f'(t) h(t))|) dt. \tag{18.2}$$

Integrating by parts, we see that

$$\sum_{n > x} a(n) f(n) = \int_x^{\infty} f(t) g'(t) dt + O\Big(|f(x) h(x)| + \int_x^{\infty} |h(t) f'(t)| dt\Big).$$

Now, since

$$\sum_{n \leq x} \frac{1}{n^{it}} = \lfloor x \rfloor x^{-it} + it \int_1^x \frac{\lfloor v \rfloor}{v^{1+it}} dv = \frac{x^{1-it}}{1-it} + O(1 + |t| \log x),$$

we can apply (18.2) with $a(n) = n^{-it}$, $f(n) = (\log n)^k n^{-\sigma}$, $g(x) = \frac{x^{1-it}}{1-it}$, and $h(x) = 1 + |t| \log x$ to find that

$$\begin{aligned} \sum_{n > X} \frac{(\log n)^k}{n^s} &= \int_X^{\infty} \frac{(\log v)^k}{v^s} dv + O_k\Big(\frac{|s|}{X} (\log X)^{k+1} + |s| \int_X^{\infty} \frac{(\log v)^{k+1}}{v^{1+\sigma}} dv\Big) \\ &= \int_X^{\infty} \frac{(\log v)^k}{v^s} dv + O_k\Big(\frac{|s|}{X} (\log X)^{k+1}\Big). \end{aligned}$$

(We have used the fact that $\Re(s) \in (1, 2)$, so the size of $|s|$ is governed by t.)

Since $\Re(s) > 1$, we have

$$\int_X^{\infty} \frac{(\log v)^k}{v^s} dv = \int_1^{\infty} \frac{(\log v)^k}{v^s} dv - \int_1^X \frac{(\log v)^k}{v^s} dv = \frac{k!}{(s-1)^{k+1}} + O_k((\log X)^{k+1}).$$

We now collect our estimates with to find that

$$\zeta^{(k)}(s) = \frac{(-1)^k k!}{(s-1)^{k+1}} + O_k\Big(\Big(1 + \frac{|s|}{X}\Big)(\log X)^{k+1}\Big). \tag{18.3}$$

Upon choosing $X = 2|s|$, this gives us an upper bound on $(\zeta^*(s))^{(k)}$ since

$$\begin{aligned} (\zeta^*(s))^{(k)} &= (s-1)\zeta^{(k)}(s) + k\zeta^{(k-1)}(s) \\ &= (s-1)\frac{(-1)^k k!}{(s-1)^{k+1}} + k\frac{(-1)^{k-1}(k-1)!}{(s-1)^{(k-1)+1}} + O_k(\max\{|s-1|, 1\}(\log 2|s|)^{k+1}) \\ &\ll_k |s|(\log 2|s|)^{k+1}. \end{aligned} \tag{18.4}$$

19. May 16

In order to prove the prime number theorem, it remains to establish lower bounds on $\zeta(s)$ with $\Re(s)$ close to 1. By continuity, establishing a lower bound for $\zeta(s)$ along the line $\sigma + it$ with $\sigma \approx 1$ is the same as establishing a zero-free region near the line $\Re(s) = 1$ for the analytic continuation of $\zeta(s)$. Historically, this was the most difficult part of the proof after Riemann introduced his program. The original ideas were due to Hadamard and de la Vallée Poussin, but we will present the ideas as Mertens reformulated them.

The idea of Hadamard and de la Vallée Poussin was to show that $\zeta(1 + it) = 0$ implies $\zeta(1 + 2it) = \infty$, but this contradicts (18.3). Thus $|\zeta(1 + it)|$ must be bounded from below. Their original idea is roughly as follows. Suppose to the contrary that $\zeta(s)$ has a zero at $1 + it$ of order r. If $s = 1 + it + 1/\log x$, with x large, then by Taylor's theorem, $\zeta(1 + it + 1/\log x) \sim c(\log x)^{-r}$ for some positive integer r (the order of the putative zero) and some constant $c \neq 0$ as $x \to \infty$. We can truncate the Euler product of $\zeta(s)$ up to x and obtain

$$\zeta(1 + 1/\log x + it) \sim \prod_{p \leq x} \Big(1 - \frac{1}{p^{1+it}}\Big)^{-1} \qquad \text{as } x \to \infty.$$

Note that

$$\Big|1 - \frac{1}{p^{1+it}}\Big|^{-1} \geq \Big(1 + \frac{1}{p}\Big)^{-1},$$

and by Mertens

$$\prod_{p \leq x} \Big(1 + \frac{1}{p}\Big)^{-1} \sim \frac{c}{\log x} \implies |\zeta(1 + it + 1/\log x)| \gg \frac{1}{\log x}.$$

This forces r to be 1 (otherwise the asymptotic upper bound would not match the asymptotic lower bound in the above equation). This suggests that $-1/p^{1+it} \approx 1/p$, i.e. $p^{it} \approx -1 = \mu(p)$, for "most primes $p \leq x$". Squaring both sides, this asserts that $p^{2it} \approx 1$ for "most primes $p \leq x$". But since

$$\zeta(1 + 2it + 1/\log x) = \prod_{p} \Big(1 - \frac{1}{p^{1+2it+1/\log x}}\Big)^{-1},$$

it would follow that $\zeta(1 + 2it) = \infty$ (upon taking $x \to \infty$). But this cannot happen, since it would violate (18.3) from the last lecture. Thus we achieve a contradiction and conclude that $\zeta(1 + it) \neq 0$ for all t. Thus $|\zeta(1 + it)| \gg_t 1$ for all t.

For our proof, we will use a (perhaps less enlightening) reformulation due to Mertens. Since $2(\cos\theta)^2 - 1 = \cos 2\theta$, we may observe the innocuous identity

$$3 + 4\cos\theta + \cos 2\theta = 2(1 + \cos\theta)^2 \geq 0. \tag{19.1}$$

By taking the logarithm of the Euler product for $\zeta(\sigma + it)$ with $\sigma > 1$, we find that

$$\log \zeta(\sigma + it)) = -\sum_{p} \log\Big(1 - \frac{1}{p^{\sigma+it}}\Big) = \sum_{p} \sum_{j=1}^{\infty} j^{-1} p^{-j\sigma} e^{-it \log p^j}.$$

Thus

$$\Re(\log \zeta(\sigma + it)) = \sum_{p} \sum_{j=1}^{\infty} j^{-1} p^{-j\sigma} \cos(t \log p^j),$$

and (19.1) implies

$$3\log\zeta(\sigma) + 4\Re(\log\zeta(\sigma+it)) + \Re(\log\zeta(\sigma+2it)) \geq 0.$$

Since $|e^{\sigma+it}| = e^{\sigma}$, it follows from exponentiation that

$$\zeta(\sigma)^3 \cdot |\zeta(\sigma+it)|^4 \cdot |\zeta(\sigma+2it)| \geq 1, \quad \text{hence} \quad |\zeta(\sigma+it)| \geq \frac{1}{\zeta(\sigma)^{\frac{3}{4}} \cdot |\zeta(\sigma+2it)|^{\frac{1}{4}}}.$$

Using (18.3) from last lecture, we conclude that for $s = \sigma + it$ with $\sigma > 1$,

$$\begin{aligned} |\zeta^*(s)| = |(s-1)\zeta(s)| \geq \frac{|s-1|}{\zeta(\sigma)^{3/4}|\zeta(\sigma+2it)|^{1/4}} &\gg \frac{|s-1|(1-\sigma)^{3/4}}{|\frac{1}{1-\sigma+2it} + O(\log 2|\sigma+2it|)|^{1/4}} \\ &\gg (\sigma-1)^{3/4}\frac{|s|}{(\log 2|s|)^{1/4}}. \end{aligned} \tag{19.2}$$

For s close to 1, it follows from (18.3) that $\lim_{s\to 1}(s-1)\zeta(s) = 1$, provided that the s approaches 1 from within the region of absolute convergence. So we obtain the same lower bound even in this seemingly-bad scenario. Using Exercise 18.1, we find that for some constant $\kappa > 0$ depending only on k,

$$\Big(\frac{1}{\zeta^*(s)}\Big)^{(k)} \ll_k \frac{(\log 2|s|)^{\kappa}}{(\sigma-1)^{3(k+1)/4}|s|}.$$

Putting together our upper bound (18.4) and our lower bound (19.2), we find

$$\Big(\frac{1}{\zeta(s)}\Big)^{(k)} = (-1)^k\Big((s-1)\Big(\frac{1}{\zeta^*(s)}\Big)^{(k)} + k\Big(\frac{1}{\zeta^*(s)}\Big)^{(k-1)}\Big) \ll_k \frac{(\log 2|s|)^{\kappa}}{(\sigma-1)^{3(k+1)/4}}.$$

Therefore,

$$\begin{aligned} \sum_{n\leq x}\mu(n)(\log n)^k \log\frac{x}{n} &= \frac{(-1)^k}{2\pi i}\int_{\Re(s)=\sigma_0}\Big(\frac{1}{\zeta(s)}\Big)^{(k)}\frac{x^s}{s^2}ds \\ &\ll_k \int_{\Re(s)=\sigma_0}\frac{(\log 2|s|)^{\kappa}}{(\sigma_0-1)^{3(k+1)/4}}\frac{|x^s|}{|s^2|}|ds| \\ &\ll_k \frac{x^{\sigma_0}}{(\sigma_0-1)^{3(k+1)/4}}\int_{\Re(s)=\sigma_0}\frac{(\log 2|s|)^{\kappa}}{|s|^2}|ds|. \end{aligned}$$

We can now translate back to integrating over t instead of $s = \sigma + it$ (since σ is fixed, namely $\sigma = \sigma_0$, and only t varies). Since the line of integration σ_0 is a number which we will choose close to 1, say $\sigma_0 \in (1,2)$, this leads to the above quantity being bounded by

$$\frac{x^{\sigma_0}}{(\sigma_0-1)^{3(k+1)/4}}\int_{-\infty}^{\infty}\frac{(\log 2|t|)^{\kappa}}{t^2+1}dt.$$

We may now choose any value of $\sigma_0 \in (1,2)$ we want. A convenient choice is $\sigma_0 = 1 + \frac{1}{\log x}$. Then $x^{\sigma_0} = ex$ and $(\sigma_0-1)^{-1} = \log x$. Thus our integral is now

$$\ll_k x(\log x)^{3(k+1)/4}\int_{-\infty}^{\infty}\frac{(\log 2|t|)^{\kappa}}{t^2+1}dt \ll_k x(\log x)^{3(k+1)/4}.$$

This concludes the proof of Theorem 16.1, and hence our proof of the prime number theorem!

20. May 18

We just proved the prime number theorem! While this problem has attracted much attention over the last few centuries, it turns out that the techniques we have developed are also well-suited for proving refinements to the prime number theorem. In particular, one may ask: Which *subsets* of the primes can we count "well"? What constitutes "well"? Perhaps we might be able to prove nothing more than variants of Mertens's theorem or Chebyshev's bounds. Perhaps we might be fortunate enough to count certain subsets of the primes asymptotically.

One natural starting point for this question is to count primes in residue classes $a \pmod q$; such primes form lie in the arithmetic progression $\{nq+a\}_{n=1}^{\infty}$. If $\gcd(a,q) > 1$, then such a subset can contain at most finitely many primes. Thus we consider arithmetic progressions $a \pmod q$ with $\gcd(a,q) = 1$. For a given $q \geq 1$, there are $\varphi(q)$ choices of residue classes a. It turns out that the study of primes in arithmetic progressions has some beautiful connections to algebraic number theory, which we will describe later (if we have enough time).

When we introduced convolutions, we used them to help us decompose sums of complicated arithmetic functions in terms sums of simpler functions. We will decompose $\mathbf{1}_{n\equiv a \pmod q}(n)$ in terms of totally multiplicative functions called Dirichlet characters.

Definition 20.1. A **Dirichlet character modulo** q is a completely multiplicative function $\chi : \mathbb{N} \to \mathbb{C}$ such that $\chi(n+q) = \chi(n)$ for all $n \in \mathbb{N}$ (thus χ is q-periodic) and $\chi(n) \neq 0$ if and only if $\gcd(n,q) = 1$.

If you are familiar with group theory, then an equivalent way to think of χ is as a homomorphism $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$, which is extended to the integers in such a way that $\chi(n) = 0$ if $\gcd(n,q) > 1$. We can see from the definition of χ that if $\chi(n) \neq 0$, then $\chi(n)$ is a q-th root of unity, in which case $|\chi(n)| = 1$. There are $\varphi(q)$ Dirichlet characters modulo q.

Apart from being fascinating examples of a completely multiplicative function, Dirichlet characters can be used to construct an indicator function for arithmetic progressions $a \pmod q$ when $\gcd(a,q) = 1$. As shown in HW5,

$$\mathbf{1}_{a,q}(n) := \begin{cases} 1 & \text{if } n \equiv a \pmod q, \\ 0 & \text{otherwise} \end{cases} = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \bar{\chi}(a)\chi(n). \tag{20.1}$$

Thus in order to detect primes in an arithmetic progression, we could consider analogues of the prime counting functions $\pi(x)$, $\vartheta(x)$, and $\psi(x)$ given by

$$\pi(x;q,a) := \sum_{p\leq x} \mathbf{1}_{a,q}(p), \quad \vartheta(x;q,a) := \sum_{p\leq x} \mathbf{1}_{a,q}(p)\log p, \quad \psi(x;q,a) := \sum_{n\leq x} \Lambda(n)\mathbf{1}_{a,q}(n).$$

We will prove the following generalization of Mertens's theorem.

Theorem 20.2 (Mertens)**.** *If* $\gcd(a,q) = 1$*, then there are infinitely many primes* $p \equiv a \pmod q$*. In particular,*

$$\sum_{p\leq x} \frac{\mathbf{1}_{a,q}(p)}{p} = \frac{\log\log x}{\varphi(q)} + O_q(1).$$

Let χ_0 denote the Dirichlet character modulo q which corresponds to the identity element of $(\mathbb{Z}/q\mathbb{Z})^\times$; we call this character the **trivial character**. Then $\chi_0(n) \in \{0,1\}$ for all n, and

$\chi(n) = 1$ precisely when $\gcd(n, q) = 1$. By total multiplicativity,

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_{p} \Big(1 - \frac{\chi(p)}{p^s}\Big)^{-1}, \qquad L(s,\chi_0) = \zeta(s) \prod_{p \mid q} \Big(1 - \frac{1}{p^s}\Big). \tag{20.2}$$

Much like $\zeta(s)$, we can see from the Euler product representation of $L(s,\chi)$ that $L(s,\chi) \neq 0$ for $\Re(s) > 1$. So we may freely take logarithms to see via Taylor's theorem that

$$\log L(s,\chi) = -\sum_{p} \log(1 - \chi(p)p^{-s}) = \sum_{p} \chi(p)p^{-s} + O(1).$$

Now,

$$\sum_{p} p^{-s}\mathbf{1}_{a,q}(p) = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \bar{\chi}(a) \log L(s,\chi) + O_q(1).$$

We have already seen from partial summation that as $s \to 1^+$ that

$$\zeta(s) = \frac{1}{s-1} + O(1).$$

Since $L(s,\chi_0)$ is $\zeta(s)$ with the Euler factors at $p \mid q$ missing, we see that as $s \to 1^+$,

$$\sum_{p} p^{-s}\mathbf{1}_{a,q}(p) = \frac{1}{\varphi(q)} \log \frac{1}{s-1} + \frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} \bar{\chi}(a) \log L(s,\chi) + O(\varphi(q)). \tag{20.3}$$

Theorem 20.3. *If $\chi \neq \chi_0$ is a Dirichlet character modulo q, then $L(1,\chi) \neq 0$.*

Proof of Theorem 20.2 assuming Theorem 20.3. If $L(1,\chi) \neq 0$, then $\lim_{s \to 1^+} \log L(s,\chi)$ is a finite number depending only on q. Then we may take $s = 1 + (\log x)^{-1}$ in (20.3), and the result follows from partial summation. □

Our first step is a quick bound for character sums. Suppose $\chi \neq \chi_0$. Note that by periodicity,

$$\Big| \sum_{n \leq x} \chi(n) - \Big\lfloor \frac{x}{q} \Big\rfloor \sum_{a=1}^{q} \chi(a) \Big| \leq q.$$

There exists a number a_1 with $(a_1, q) = 1$ and $\chi(a_1) \neq 1$. (This is analogous the proof of HW5, Problem 5, Part 1 that I provided.) Note that since $\chi(a) = 0$ if $(a,q) > 1$, then

$$\chi(a_1) \sum_{a=1}^{q} \chi(a) = \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \chi(a_1)\chi(a) = \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \chi(a_1 a) = \sum_{\substack{1 \leq b \leq q \\ (b,q)=1}} \chi(b) = \sum_{b=1}^{q} \chi(b).$$

Since $\chi(a_1) \neq 1$, this implies that

$$\sum_{a=1}^{q} \chi(a) = 0, \qquad \text{hence} \qquad \Big| \sum_{n \leq x} \chi(n) \Big| \leq \min\{q, x\}. \tag{20.4}$$

Consequently, a minor variation in the arguments in HW2 gives us the estimate

$$L(\sigma + it, \chi) = \sum_{n \leq x} \frac{\chi(n)}{n^{\sigma + it}} + O_{q,\sigma,t}(x^{-\sigma}), \qquad x \geq 1, \qquad \sigma > 0. \tag{20.5}$$

21. May 21

We now proceed to show that $L(1,\chi) \neq 0$ for each Dirichlet character $\chi \neq \chi_0$ modulo q. Our second step is to borrow an idea from our proof of the prime number theorem. Recall

$$3 + 4\cos\theta + \cos 2\theta \geq 0,$$

and that we applied this to show that

$$3\log\zeta(\sigma) + 4\Re(\log\zeta(\sigma+it)) + \Re(\log\zeta(\sigma+2it)) \geq 0, \qquad \sigma > 1.$$

which immediately implies

$$\zeta(\sigma)^3 \cdot |\zeta(\sigma+it)|^4 \cdot |\zeta(\sigma+2it)| \geq 1, \qquad \sigma > 1.$$

We will play the same game here. By a nearly identical calculation (carried out in HW6), we find that

$$3\log L(\sigma,\chi_0) + 4\Re(\log L(\sigma+it,\chi)) + \Re(\log L(\sigma+2it,\chi^2)) \geq 0, \qquad \sigma > 1, \tag{21.1}$$

which immediately implies (by $\Re(e^{a+bi}) = e^a$)

$$L(\sigma,\chi_0)^3 \cdot |L(\sigma+it,\chi)|^4 \cdot |L(\sigma+it,\chi^2)| \geq 1 \qquad \sigma > 1.$$

Now, as $\sigma \to 1^+$, we have

$$L(\sigma,\chi_0) \sim \frac{1}{\sigma-1}\prod_{p|q}(1-p^{-\sigma})^{-1}$$

Now, if $L(1,\chi) = 0$, then as $\sigma \to 1^+$, there exists a constant A such that $L(\sigma,\chi) \sim A(\sigma-1)$. Thus as $\sigma \to 1^+$,

$$L(\sigma,\chi_0)^3 \cdot |L(\sigma,\chi)|^4 \cdot |L(\sigma,\chi^2)| \sim A(\sigma-1) \cdot |L(\sigma,\chi^2)| \cdot \prod_{p|q}(1-p^{-1})^3.$$

Suppose first that $\chi^2 \neq \chi_0$. Then by (20.4) and HW2, we know that $|L(\sigma,\chi^2)| < q$, in which case we find that as $\sigma \to 1^+$, $L(\sigma,\chi_0)^3 \cdot |L(\sigma,\chi)|^4 \cdot |L(\sigma,\chi^2)| \to 0$, a contradiction. This proves Theorem 20.3, provided that $\chi^2 \neq \chi_0$. Suppose now that $\chi^2 = \chi_0$, in which case χ is real-valued. If $L(1,\chi) = 0$, then as $\sigma \to 1^+$,

$$L(\sigma,\chi_0)^3 \cdot |L(\sigma,\chi)|^4 \cdot |L(\sigma,\chi^2)| \sim A\prod_{p|q}(1-p^{-1})^4.$$

This does not lead to the desired contradiction (we believe that the asymptotic $L(\sigma,\chi) \sim A(\sigma-1)$ is false). We must now resort to other means.

From now on, we let $\chi \neq \chi_0$ be a real character. In order to prove that $L(1,\chi) \neq 0$, we first observe via partial summation that

$$\sum_{n\leq x}\log\frac{x}{n} = x - 1 - \int_1^x \frac{\{t\}}{t}dt.$$

Now, we consider the sum

$$T(x) = \sum_{n\leq x}(\chi * 1)(n)\log\frac{x}{n}.$$

Since χ is real and totally multiplicative, $\chi(n) \in \{-1, 0, 1\}$ and $\chi(n^b) = \chi(n)^b$. Therefore, by HW1,

$$(\chi * 1)(n) = \sum_{d|n} \chi(d) = \prod_{p^a || n} \Big(1 + \sum_{r=1}^{a} \chi(p)^a\Big) \geq 0, \qquad (\chi * 1)(n^2) \geq 1.$$

Thus

$$T(x) \geq \sum_{m^2 \leq x} \log \frac{x}{m^2} = 2 \sum_{m \leq \sqrt{x}} \log \frac{\sqrt{x}}{m} \geq 2\sqrt{x} - O(\log x).$$

In order to estimate $T(x)$ more precisely, we use (20.5) to compute

$$\begin{aligned} T(x) = \sum_{a \leq x} \chi(a) \sum_{b \leq x/a} \log \frac{x/a}{b} &= \sum_{a \leq x} \chi(a) \Big(\frac{x}{a} - 1 - \int_1^{x/a} \frac{\{t\}}{t} dt \Big) \\ &= x \sum_{a \leq x} \frac{\chi(a)}{a} - \sum_{a \leq x} \chi(a) - \int_1^x \Big(\sum_{a \leq x/t} \chi(a) \Big) \frac{\{t\}}{t} dt \\ &= xL(1, \chi) - x \sum_{a > x} \frac{\chi(a)}{a} - \sum_{a \leq x} \chi(a) - \int_1^x \Big(\sum_{a \leq x/t} \chi(a) \Big) \frac{\{t\}}{t} dt. \end{aligned}$$

By partial summation and (20.4), we find that if $x > q$, then

$$\begin{aligned} &x \Big| \sum_{a > x} \frac{\chi(a)}{a} \Big| = x \Big| - \frac{\sum_{a \leq x} \chi(a)}{x} + \int_x^{\infty} \frac{\sum_{a \leq t} \chi(a)}{t^2} dt \Big| \leq 2q, \\ &\Big| - \int_1^x \Big(\sum_{a \leq x/t} \chi(a) \Big) \frac{\{t\}}{t} dt \Big| \leq \int_1^x \frac{\min\{q, x/t\}}{t} dt \leq q \log x, \quad \Big| \sum_{a \leq x} \chi(a) \Big| \leq q. \end{aligned}$$

Putting together our upper and lower bounds for $T(x)$, we find that for all $x \geq 1$,

$$\frac{2\sqrt{x} - 3q - q \log x - O(\log x)}{x} \leq L(1, \chi).$$

If we choose $x = cq^2 (\log q)^4$ with c a sufficiently large constant independent of q, then we obtain the lower bound

$$\frac{1}{q(\log q)^2} \ll L(1, \chi).$$

This finishes the proof of Theorem 20.3, hence also the proof of Theorem 20.2.

Much more is true. One can extend Riemann's program for $\zeta(s)$ to produce similar properties (analytic continuation, functional equation, etc.) for $L(s, \chi)$. When $x \notin \mathbb{Z}$, we can also use (20.1) to express

$$\psi(x; q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \Lambda(n)$$

as

$$\frac{1}{\varphi(q)} \sum_{\chi \pmod q} \bar{\chi}(a) \sum_{n \leq x} \Lambda(n) \chi(n) = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \frac{\bar{\chi}(a)}{2\pi} \int_{-\infty}^{\infty} -\frac{L'(\sigma + it, \chi)}{L(\sigma + it, \chi)} \frac{x^{\sigma + it}}{\sigma + it} d\sigma, \quad \sigma > 1.$$

As was the case in Mertens's theorem above, the main contribution will come from the trivial character (whose Dirichlet series is nearly $\zeta(s)$). Either by working using complex analysis

to study these integrals or by generalizing our approach to the prime number theorem (and incorporating some ideas from topics yet to come), we could prove:

Theorem 21.1 (PNT for arithmetic progressions)**.** *For all $A > 0$, there exists a constant $c(A) > 0$ such that following is true. Let $q \leq (\log x)^A$, and let $\gcd(a, q) = 1$. We have that*

$$\psi(x; q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \Lambda(n) = \frac{x}{\varphi(q)} + O\Big(\frac{x}{\exp(c(A)\sqrt{\log x})}\Big).$$

The constant $c(A)$ cannot be computed. Equivalently (by passing from prime powers to primes as in HW2 and partial summation),

$$\pi(x; q, a) := \#\{p \leq x \colon p \equiv a \pmod q\} = \frac{\mathrm{Li}(x)}{\varphi(q)} + O\Big(\frac{x}{\exp(c(A)\sqrt{\log x})}\Big).$$

Note that the range of q relative to x is **very** restrictive. Also, note that the constant $c(A)$ cannot be computed. Both of these shortcomings are tied to the fact that bounding $L(1, \chi)$ from below when χ is a real character is incredibly hard! This is intimately connected to the problem of understanding the behavior of the ideal class group of quadratic number fields, a problem that Gauss was rather fond of (in addition to counting primes!).

Of course, we believe that a generalization of the Riemann hypothesis (GRH) holds for $L(s, \chi)$ as well. There is a functional equation relating $L(s, \chi)$ to $L(1 - s, \bar{\chi})$, and we expect that all of the nontrivial zeros of $L(s, \chi)$ lie on the line $\Re(s) = 1/2$. Equivalently,

Conjecture 21.2 (GRH for Dirichlet characters)**.** *Let $\epsilon > 0$. If $\gcd(a, q) = 1$ and $x \gg_\epsilon q^{2+\epsilon}$,*

$$\pi(x; q, a) = \frac{\mathrm{Li}(x)}{\varphi(q)} + O(\sqrt{x} \log x).$$

One fascinating development of twentieth-century mathematics was the following result, which was proved independently by Enrico Bombieri and A. I. Vinogradov in the 1965.

Theorem 21.3 (Bombieri–Vinogardov)**.** *Fix $\theta \in (0, 1/2)$. For all $A > 0$,*

$$\sum_{q \leq x^\theta} \max_{\gcd(a,q)=1} \Big|\psi(x; q, a) - \frac{x}{\varphi(q)}\Big| \ll_A \frac{x}{(\log x)^A}.$$

Equivalently (by passing from prime powers to primes as in HW2 and partial summation)

$$\sum_{q \leq x^\theta} \max_{\gcd(a,q)=1} \Big|\pi(x; q, a) - \frac{\mathrm{Li}(x)}{\varphi(q)}\Big| \ll_A \frac{x}{(\log x)^A}.$$

Thus the mean value of the error term in the prime number theorem for arithmetic progressions is as about as small as the Generalized Riemann Hypothesis predicts for all except a density zero subset of the moduli $q \leq x^{1/2-\epsilon}$ for any $\epsilon > 0$. This has served as a powerful substitute for the Generalized Riemann Hypothesis in many intriguing settings, such as the recent work on the twin prime conjecture. We believe much more to be true:

Conjecture 21.4 (Elliot–Halberstam)**.** *We can replace the $\theta \in (0, 1/2)$ with the condition $\theta \in (0, 1)$ in the Bombieri-Vinogradov theorem.*

22. May 23

We have spent a good bit of time describing how to obtain asymptotic results, but in many situations, it is also important to be able to obtain upper and lower bounds on various sums. We saw one idea at the beginning of the quarter with the Sieve of Eratosthenes. To facilitate our discussion, let

$$P^-(n) = \min\{p\colon p \mid n\}, \qquad P^+(n) = \max\{p\colon p \mid n\},$$

with the convention that $P^-(1) = \infty$, $P^+(1) = 0$. We considered the sum

$$\pi(x,z) = \sum_{\substack{n \leq x \\ P^-(n) > z}} 1$$

and reasoned that $\pi(x) = \pi(x, \sqrt{x}) + \pi(\sqrt{x}) - 1$. For values of $z < \sqrt{x}$, we saw that

$$\pi(x) \leq \pi(x,z) + \pi(y).$$

Our investigation of $\pi(x,z)$ was inefficient because, in our new notation, an inclusion-exclusion argument led to

$$\pi(x,z) = \sum_{P^+(n) \leq z} \mu(n) \left\lfloor \frac{x}{n} \right\rfloor,$$

and we approximated with $\lfloor x/n \rfloor$ by $x/n + O(1)$. This leads to understanding partial sums of $\mu(n)$, which as we have seen, are quite difficult to manage.

We revisit this problem using a simple, yet incredibly powerful, observation due to Selberg. In the language of convolutions, it was proved in HW3 that

$$\sum_{d \mid n} \mu(d) = (\mu * 1)(n) = E(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

But if $\lambda(d)$ is **any** real-valued function such that $\lambda(1) = 1$, then

$$\sum_{d \mid n} \mu(d) \leq \Big(\sum_{d \mid n} \lambda(d) \Big)^2.$$

Indeed, the left- and right-hand sides are equal if $n = 1$ (since $\lambda(1) = 1$). If $n \geq 2$, the left-hand side is zero, while the right hand side is nonnegative (because $\lambda(n) \in \mathbb{R}$).

For convenience, write

$$P(z) = \prod_{p \leq z} p.$$

Note that $P^-(n) > z$ if and only if $\gcd(n, P(z)) = 1$. But this can be expressed in terms of μ:

$$\gcd(n, P(z)) = 1 \iff \sum_{d \mid \gcd(n, P(z))} \mu(d) = 1.$$

Now, let us proceed without specifying λ for the time being; we can choose it optimally later. Then

$$\pi(x,y) = \sum_{\substack{n \leq x \\ P^-(n) > z}} 1 = \sum_{n \leq x} \sum_{d \mid \gcd(n, P(z))} \mu(d) \leq \sum_{n \leq x} \Big(\sum_{d \mid \gcd(n, P(z)} \lambda(d) \Big)^2.$$

For compact notation, we introduce $(a,b)=\gcd(a,b)$ and $[a,b]=\operatorname{lcm}(a,b)$. Then the above becomes

$$\sum_{n\le x}\Big(\sum_{d|(n,P(z)}\lambda(d)\Big)^2=\sum_{n\le x}\Big(\sum_{d_1,d_2|(n,P(z))}\lambda(d_1)\lambda(d_2)\Big)=\sum_{d_1,d_2|P(z)}\lambda(d_1)\lambda(d_2)\sum_{\substack{n\le x\\ [d_1,d_2]|n}}1.$$

We now proceed with our usual floor-function approximation, which leads us to

$$\pi(x,y)\le\sum_{d_1,d_2|P(z)}\lambda(d_1)\lambda(d_2)\Big(\frac{x}{[d_1,d_2]}+O(1)\Big)=x\sum_{d_1,d_2|P(z)}\frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}+O\Big(\sum_{d_1,d_2|P(z)}|\lambda(d_1)|\cdot|\lambda(d_2)|\Big).$$

We now start to impose restrictions on λ. For convenience, we assume that

$$\lambda(d)=0 \text{ for all } d>z \text{ and all } d\nmid P(z).$$

Then

$$\pi(x,y)\le\sum_{d_1,d_2\le z}\lambda(d_1)\lambda(d_2)\Big(\frac{x}{[d_1,d_2]}+O(1)\Big)=x\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}+O\Big(\Big(\sum_{d\le z}|\lambda(d)|\Big)^2\Big).$$

If we also restrict λ so that $|\lambda(d)|\le 1$ for all d, then this becomes

$$\pi(x,y)\le x\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}+O(z^2).$$

It remains to choose λ so that our main term becomes something useful.

Using the identity $[d_1,d_2](d_1,d_2)=d_1d_2$ and the identity (from HW4) $\sum_{\delta|d}\varphi(\delta)=d$, we transform the main term by

$$\begin{aligned}\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}=\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}(d_1,d_2)&=\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}\sum_{\delta|(d_1,d_2)}\varphi(\delta)\\&=\sum_{\delta\le z}\varphi(\delta)\sum_{\substack{d_1,d_2\le z\\ \delta|(d_1,d_2)}}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}=\sum_{\delta\le z}\varphi(\delta)\Big(\sum_{\substack{d\le z\\ \delta|d}}\frac{\lambda(d)}{d}\Big)^2.\end{aligned}$$

Under the transformation

$$u_\delta=\sum_{\substack{d\le z\\ \delta|d}}\frac{\lambda(d)}{d},\tag{22.1}$$

this becomes

$$\sum_{d_1,d_2\le z}\frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}=\sum_{\delta\le z}\varphi(\delta)u_\delta^2.$$

By proceeding in a manner similar to HW3 (Exercise 3), we see (22.1) implies

$$\frac{\lambda(d)}{d}=\sum_{\substack{\delta\le z\\ d|\delta}}\mu(\delta/d)u_\delta.\tag{22.2}$$

23. May 25

We continue from last time. Our hypothesis that $\lambda(1) = 1$ combined with (22.2) (with $d = 1$) implies that $\sum_{\delta \leq z} \mu(\delta) u_\delta = 1$. From (22.1), we see that the condition $\lambda(d) = 0$ if $d > z$ yields the condition that $u_\delta = 0$ for $\delta > z$. This leaves us with the task of minimizing

$$\sum_{\delta \leq z} \varphi(\delta) u_\delta^2 \qquad \text{subject to} \qquad u_\delta = 0 \text{ for } \delta > z, \qquad \sum_{\delta \leq z} \mu(\delta) u_\delta = 1.$$

This can be accomplished with Lagrange multipliers!

The solution to the Lagrange multiplier problem is

$$u_\delta = \frac{\mu(\delta)}{\varphi(\delta) V(z)}, \qquad V(z) = \sum_{\substack{d \mid P(z) \\ d \leq z}} \frac{\mu(d)^2}{\varphi(d)}.$$

This translates into

$$\lambda(d) = d \sum_{\substack{\delta \leq z \\ d \mid \delta}} \frac{\mu(\delta/d)\mu(\delta)}{\varphi(\delta) V(z)} = \frac{\mu(d) d}{V(z)} \sum_{\substack{\delta \leq z \\ d \mid \delta}} \frac{\mu(\delta)^2}{\varphi(\delta)}.$$

(We have used the fact that $\mu(\delta/d) = \mu(\delta)/\mu(d) = \mu(\delta)\mu(d)$ when δ is squarefree.)

Exercise 23.1. We required that $\lambda(1) = 1$, $\lambda(d) = 0$ for all $d > z$ and all $d \nmid P(z)$, and $|\lambda(d)| \leq 1$ for all d. Verify these conditions.

By considering our choice of u (which induced a choice of λ), we find that

$$\sum_{d_1, d_2 \leq z} \frac{\lambda(d_1)\lambda(d_2)}{[d_1, d_2]} = \sum_{\delta \leq z} \varphi(\delta) u_\delta^2 = \sum_{\delta \leq z} \varphi(\delta) \Big(\frac{\mu(\delta)}{\varphi(\delta) V(z)} \Big)^2 = \frac{1}{V(z)^2} \sum_{\delta \leq z} \frac{\mu(\delta)^2}{\varphi(\delta)} = \frac{1}{V(z)}.$$

Putting everything together, we conclude that

$$\pi(x, z) \leq \frac{x}{V(z)} + O(z^2).$$

It remains to find a lower bound for $V(z)$.

Exercise 23.2. Prove that $V(z) \geq \sum_{n \leq z} \frac{1}{n}$.

This gives us the conclusion of Selberg's sieve:

$$\pi(x, z) = \#\{n \leq x \colon P^-(n) > z\} \leq \frac{x}{\log z} + O(z^2).$$

The main terms

$$x \prod_{p < z} \Big(1 - \frac{1}{p}\Big), \qquad x \Big(\sum_{d \leq z} \frac{\mu(d)^2}{\varphi(d)} \Big)^{-1}$$

provided by the sieve of Eratosthenes and Selberg's sieve (respectively) have the same $O(x/\log z)$ order of magnitude. However, the error term in the sieve of Eratosthenes was $O(2^{\pi(z)})$, whereas our error term here is $O(z^2)$. This is an amazing difference in quality! In fact, by choosing $z = (x/\log x)^{1/2}$, Selberg's sieve recovers the Chebyshev upper bound! As we will see next lecture, we can generalize Selberg's sieve into an incredibly flexible setup that allows us to a plethora of upper-bound problems. Note that

$$\lambda(d) \approx \mu(d) \frac{\log(z/d)}{\log z}.$$

In other words, $\lambda(d)$ is like a "smoothed" version of $\mu(d)$.

We now state one (of many!) formal generalization of the results from last time. We omit the proof because of time considerations and also because the proof largely follows what we did last class (but more care is required). Let $\mathcal{A}$ be a finite set of integers, let $z \geq 1$ be a real number, and let $\mathcal{P}$ denote a set of primes. Set

$$P(z) = \prod_{\substack{p<z \\ p\in\mathcal{P}}} p, \qquad S(\mathcal{A}, z) = \#\{n \in \mathcal{A} \colon (n, P(z)) = 1\}.$$

For each $d \mid P(z)$, let $\mathcal{A}_d = \{n \in \mathcal{A} \colon d \mid n\}$. We make the tacit assumption that

$$\#\mathcal{A}_d = \frac{\#\mathcal{A}}{f(d)} + r_d.$$

- $f : \mathbb{N} \to [1, \infty)$ is a multiplicative function such that $f(p) > 1$ for each $p \mid P(z)$. The function $1/f(d)$ represents the "probability" that a member of $\mathcal{A}$ is divisible by d. (In our estimate of $\pi(x, z)$, $1/f(d) = 1/d$.)
- r_d is some real number, which we can think of as an error term. (In our estimate of $\pi(x, z)$, $r_d = O(1)$.)

We expect that $\mathcal{A}_{d_1}$ and $\mathcal{A}_{d_2}$ are roughly independent if $(d_1, d_2) = 1$; for most sieve-friendly problems, this is a close approximation to reality. The analogue of the sieve of Eratosthenes is the **exact** formula

$$S(\mathcal{A}, z) = \#\mathcal{A} \sum_{d|P(z)} \frac{\mu(d)}{f(d)} + \sum_{p|P(z)} \mu(d) r_d = \#\mathcal{A} \prod_{p<z} \Big(1 - \frac{1}{f(p)}\Big) + \sum_{d|P(z)} \mu(d) r_d.$$

The product

$$W(z) = \prod_{p<z} \Big(1 - \frac{1}{f(p)}\Big)$$

represents the "probability" that $n \in \mathcal{A}$ has no prime factors less than z. Our work at the beginning of the quarter suggested that $S(\mathcal{A}, z) \approx \#\mathcal{A} \cdot W(z)$ if we can safely ignore the error terms, which happen to accumulate rapidly in most situations (even if $r_d = O(1)$ as was the case with Eratosthenes!). The Selberg sieve, in a sense, is a grand gambit: At the sacrifice of an exact formula, we obtain an upper bound of the same order as $\#\mathcal{A} \cdot W(z)$, but the error terms accumulate at a manageable rate. This turns out to be incredibly effective!

Theorem 23.3 (Selberg, 1947)**.** *With the notation and conventions set above,*

$$\#S(\mathcal{A}, z) \leq \frac{\#\mathcal{A}}{V(z)} + R(z),$$

where

$$V(z) = \sum_{\substack{d\leq z \\ d|P(z)}} \frac{\mu(d)^2}{(\mu * f)(d)}, \qquad R(z) = \sum_{\substack{d_1,d_2\leq z \\ d_1,d_2|P(z)}} |r_{[d_1,d_2]}|.$$

Observe that

$$\frac{1}{V(z)} \geq \prod_{p\leq z} \Big(1 + \frac{1}{f(p) - 1}\Big)^{-1} = \prod_{p<z} \Big(1 - \frac{1}{f(p)}\Big) = W(z),$$

which illustrates that this bound is reasonably close to optimal (although in some cases it is not!). The sum $V(z)$ can be unwieldy sometimes, but the following lemma helps often.

Lemma 23.4. *Let $\widetilde{f}$ be the completely multiplicative function defined on the primes by $\widetilde{f}(p) = f(p)$. Then*

$$V(z) \geq \sum_{\substack{n \leq z \\ \text{each prime dividing } n \text{ is in } \mathcal{P}}} \frac{1}{\widetilde{f}(n)}.$$

Proof. Let $\mathcal{N}(d) = \{n \colon p \mid d \iff p \mid n\}$. Notice for d squarefree that

$$(f * \mu)(d) = \prod_{p \mid d} (f(p) - 1).$$

Thus

$$\frac{1}{(f * \mu)(d)} = \prod_{p \mid d} \frac{1/f(p)}{1 - 1/f(p)} = \prod_{p \mid d} \sum_{j=1}^{\infty} \frac{1}{f(p)^k} = \sum_{n \in \mathcal{N}(d)} \frac{1}{\widetilde{f}(n)}.$$

Notice that if $d_1 \neq d_2$ are squarefree, then $\mathcal{N}(d_1) \cap \mathcal{N}(d_2)$ is empty, and every integer s lies in $\mathcal{N}(\prod_{p \mid s} p)$. Thus

$$V(z) = \sum_{\substack{d \leq z \\ d \mid P(z)}} \frac{1}{(f * \mu)(d)} = \sum_{\substack{d \leq z \\ d \mid P(z)}} \sum_{n \in \mathcal{N}(d)} \frac{1}{\widetilde{f}(n)}.$$

Since $\widetilde{f}(n) > 0$, we can omit all $n > z$, which gives us a lower bound. The resulting sum is over $n \leq z$ which are divisible only by primes which lie in $\mathcal{P}$. □

24. May 30

In 1912, Edmund Landau (a number theorist and complex analyst) listed four basic problems regarding the distribution of primes which he characterized as "unattackable at the present state of mathematics". They were:

(1) Goldbach's conjecture: Can every even integer $n \geq 2$ be written as a sum of two primes?
(2) Twin prime conjecture: Are there infinitely many primes p such that $p+2$ is also prime?
(3) Legendre's conjecture: Does there exist a prime in the interval $(n^2, (n+1)^2)$ for every integer $n \geq 1$?
(4) Are there infinitely integers n such that n^2+1 is prime?

We will discuss the first two questions in some detail. These are rather interesting questions because the primes are defined multiplicatively, whereas these questions are additive in nature. The relationship between multiplicative and additive structure is rather delicate, and techniques related to the proof of the prime number theorem (or its extension to arithmetic progressions) cannot address these questions on their own. That being said, one can make some nontrivial progress toward closely related questions via sieve methods. We will discuss these problems using the Selberg sieve.

Before we begin, we will describe a probabilistic model which, while imperfect, provides a mechanism by which one can make reasonable conjectures. In 1936, Cramér gave an interesting approach to statistical questions about prime numbers based on a random model (the Cramér model). The idea is that if $\mathbb{P}$ denotes the set of all primes, then $(\mathbf{1}_{\mathbb{P}}(n))_{n=1}^{\infty}$ behaves like a sequence of independent Bernoulli random variables with parameter $1/\log n$ ($n \geq 3$). In other words, for $n \geq 3$,

$$\text{Prob}[\mathbf{1}_{\mathbb{P}}(n) = 1] = \frac{1}{\log n} \qquad \text{Prob}[\mathbf{1}_{\mathbb{P}}(n) = 0] = 1 - \frac{1}{\log n},$$

and all such events are independent. This must be taken with a hefty dose of salt. A number is either prime or composite; probability does not enter the picture! Indeed, this gives positive probability to the number $n(n+1)$ being prime, but this is even for every $n \geq 1$. However, the prime number theorem can be thought of as giving the cumulative distribution function of such a random variable, so this should be a decent tool for giving reasonable conjectures.

Cramér model predicts that if $2N$ is a large even integer, then

$$r(N) := \#\{p_1 + p_2 \colon p_1 + p_2 = 2N,\, p_1, p_2 \in \mathbb{P}\} = \#\{p < 2N \colon 2N - p \in \mathbb{P}\}$$
$$\sim \sum_{m=3}^{N} \frac{1}{\log m} \cdot \frac{1}{\log(2N-m)} \sim c\frac{2N}{(\log N)^2}.$$

We have more refined guesses that lead to a good guess as two what c should be. We will discuss this next lecture; it turns out c depends mildly on N. Our work on Goldbach's conjecture will yield an upper bound for $r(2N)$ of the conjectured order of magnitude.

Theorem 24.1. *If $2N$ is a large even integer, then*

$$r(2N) \ll \frac{N}{\varphi(N)} \cdot \frac{N}{(\log N)^2}.$$

We begin with a lemma that will be useful both here and in future sieving problems.

Lemma 24.2. *Let $f(x)$ be a polynomial with integer coefficients, and let*

$$\rho_f(d) = \#\{c \pmod d \colon f(c) \equiv 0 \pmod d\}.$$

Then

$$\#\{n \leq x \colon d \mid f(n)\} = \frac{\rho(d)}{d}x + O(\rho(d)).$$

Proof. The quantity we wish to estimate equals

$$\sum_{\substack{c \pmod d \\ f(c) \equiv 0 \pmod d}} \#\{n \leq x \colon n \equiv c \pmod d\}.$$

The result follows since $\#\{n \leq x \colon n \equiv c \pmod d\} = \frac{x}{d} + O(1)$. □

Let $\mathcal{P} = \mathbb{P}$. Since $r(2N) = \#\{p < 2N \colon 2N - p \in \mathbb{P}\}$, we find that

$$r(2N) \leq 2\pi(z) + S(\mathcal{A}, z), \qquad \mathcal{A} = \{n(2N-n) \colon n \leq 2N\}.$$

To estimate $S(\mathcal{A}, z)$, we consider for each $d \mid P(z)$ the set $\mathcal{A}_d = \{n \in \mathcal{A} \colon d \mid n\}$. First, we consider $\mathcal{A}_p$ for primes $p \mid P(z)$. By Lemma 24.2,

$$\#\mathcal{A}_p = \frac{\rho(p)}{p}\#\mathcal{A} + O(\rho(p)),$$

where $\rho(p) = \#\{c \pmod p \colon c(2N-c) \equiv 0 \pmod p\}$. If $p \mid 2N$, then this has a single solution ($c = 0$). If $p \nmid 2N$, then this has two solutions, namely $c = 0$ and the residue class in which $2N$ resides. Thus

$$\rho(p) = \begin{cases} 1 & \text{if } p \mid 2N, \\ 2 & \text{if } p \nmid 2N. \end{cases}$$

The multiplicativity of $\rho(d)$ follows from the the Chinese remainder theorem. Thus for any $d \mid P(z)$, we have

$$\#\mathcal{A}_d = \frac{\rho(d)}{d}\#\mathcal{A} + O(\rho(d)).$$

We now apply Theorem 23.3 to find that if $f(d) = d/\rho(d)$, then

$$r(2N) \leq \frac{\#\mathcal{A}}{V(z)} + O\Big(\sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 \mid P(z)}} \rho([d_1, d_2])\Big), \qquad V(z) = \sum_{\substack{d \leq z \\ d \mid P(z)}} \frac{\mu(d)^2}{(\mu * f)(d)}$$

Note that $\rho(d) \leq 2^{\omega(d)}$. Since $\omega([d_1, d_2]) \leq \omega(d_1 d_2) \leq \omega(d_1) + \omega(d_2)$, we find that

$$\sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 \mid P(z)}} \rho([d_1, d_2]) \leq \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 \mid P(z)}} 2^{\omega(d_1) + \omega(d_2)} = \Big(\sum_{\substack{d \leq z \\ d \mid P(z)}} 2^{\omega(d)}\Big)^2 \leq \Big(\sum_{\substack{n \leq z \\ n \mid P(z)}} d(n)\Big)^2 \ll z^2 (\log z)^2.$$

Since $\#\mathcal{A} = 2N$, it remains to bound $V(z)$ from below. To give a lower bound for $V(z)$, we use Lemma 23.4:

$$V(z) \geq \sum_{n \leq z} \frac{\widetilde{\rho}(n)}{n},$$

where if $n = \prod_{p^a||n} p^a$, then

$$\widetilde{\rho}(n) = \prod_{p^a||n} \rho(p)^a = \prod_{\substack{p^a||n \\ p\nmid 2N}} 2^a = 2^{\Omega_{2N}(n)},$$

where $\Omega_{2N}(n)$ is the number of primes (with multiplicity) dividing n which are coprime to $2N$.

Exercise 24.3. (1) Prove that

$$2^{\Omega_{2N}(n)} \geq \sum_{\substack{d|n \\ \gcd(d,2N)=1}} 1.$$

(2) Use this to prove that $V(z) \gg \frac{\varphi(N)}{N}(\log z)^2$.

We now conclude that

$$r(2N) \ll \frac{N}{\frac{\varphi(N)}{N}(\log z)^2} + z^2(\log z)^2.$$

We obtain the theorem by choosing $z = N^{1/4}$.

25. June 1

We will Selberg's sieve to study the distribution of primes p such that $p+2$ is also prime. By the Cramér model,

$$\mathrm{Prob}(\mathbf{1}_{\mathbb{P}}(n)=\mathbf{1}_{\mathbb{P}}(n+2)=1)=\frac{1}{\log n}\cdot\frac{1}{\log(n+2)}.$$

Summing this over all $3\leq n\leq x$, we make the prediction that

$$\pi_2(x)\sim c\frac{x}{(\log x)^2}$$

for some appropriate constant $c>0$.

A more precise conjecture was made by Hardy and Littlewood in 1922. Instead of treating the primality of n and $n+2$ as being independent, we know that if n is large, then in order for both n and $n+2$ to be prime, neither n nor $n+2$ can be divisible by 2, 3, 5, and so on. If n is chosen randomly, then the probability that both n and $n+2$ are odd is $1/2$, whereas the probability that two random numbers are both odd is $1/4$. For each prime $\ell\geq 3$, the probability that both n and $n+2$ are not divisible by ℓ is

$$1-2/\ell,$$

whereas the probability that two random numbers are both not divisible by ℓ is

$$(1-1/\ell)^2.$$

For the prime 2, we must correct the probability $1/4$ by multiplying by

$$2=(1-1/2)(1-1/2)^{-2}.$$

For each prime $\ell\geq 3$, we must correct the probability $(1-1/\ell)^2$ by multiplying by

$$(1-2/\ell)(1-1/\ell)^{-2}.$$

The idea is that if we multiply all of these corretion factors together, then we have accounted for "all the ways" in which n and $n+2$ are dependent, producing the right value of c from our discussion in the Cramér model. Thus we conjecture that

$$\pi_2(x)\sim\mathfrak{S}\int_2^x\frac{dt}{(\log t)^2}\sim\mathfrak{S}\frac{x}{(\log x)^2},\quad \mathfrak{S}=\Big(1-\frac{1}{2}\Big)\Big(1-\frac{1}{2}\Big)^{-2}\prod_{\ell\geq 3}\Big(1-\frac{2}{\ell}\Big)\Big(1-\frac{1}{\ell}\Big)^{-2}=1.3203\ldots$$

By similar reasoning, it is reasonable to conjecture that with $r(2N)$ as given last lecture,

$$r(2N)\sim\mathfrak{S}\Big(\prod_{\substack{p\mid 2N\\ p\geq 3}}\frac{p-1}{p-2}\Big)\int_2^{2N}\frac{dt}{(\log t)^2}.$$

Theorem 25.1. *If x is large, then $\pi_2(x)\ll x(\log x)^{-2}$.*

To begin, let

$$\mathcal{A}=\{n(n+2)\colon n\leq x\},\qquad \mathcal{P}=\text{all primes}$$

Then $\pi_2(x)\leq S(\mathcal{A},z)+\pi(z)$, so it suffices to give an upper bound for $S(\mathcal{A},z)$. For $d\mid P(z)$, define

$$\mathcal{A}_d=\{n(n+2)\colon n\leq x,\ d\mid n(n+2)\}.$$

If we let

$$\rho(d)=\#\{c\pmod d\colon c(c+2)\equiv 0\pmod d\},$$

then by Lemma 24.2,

$$\#\mathcal{A}_p = \frac{\rho(p)}{p}\#\mathcal{A} + O(\rho(p)).$$

If $p = 2$, then $c = 0$ is the only solution to $c(c+2) \equiv 0 \pmod{p}$. Thus $\rho(2) = 1$. If $p \neq 2$, then both $c = 0$ and $c = p-2$ are solutions, so $\rho(p) = 2$. Thus

$$\rho(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p \neq 2. \end{cases}$$

Now, $\rho(\prod_{p|d} p) = \prod_{p|d} \rho(p)$ by the Chinese remainder theorem, giving $\rho(d)$ for squarefree d.

By Selberg's sieve,

$$\pi_2(x) - \pi_2(z) \leq \frac{x}{V(z)} + R(z).$$

We will show that $V(z) \gg (\log z)^2$ and $R(z) \ll (z \log z)^2$. Then we could choose $z = x^{1/4}$, for example, to conclude the theorem.

We now set out to prove $R(z) \ll (z \log z)^2$. Since $r_d = \rho(d) \leq 2^{\omega(d)}$ for d squarefree, the bound for $R(z)$ proceeds just as in the case of our upper bound for $r(2N)$. Hence $R(z) \ll (z \log z)^2$, as claimed.

For the lower bound on $V(z)$, we have by Lemma 23.4 that

$$V(z) \geq \sum_{n \leq z} \frac{\widetilde{\rho}(n)}{n}.$$

If $n = \prod_{p^a || n} p^a$, then

$$\widetilde{\rho}(n) = \prod_{\substack{p^a || n \\ p \neq 2}} 2^a = 2^{\Omega_2(n)},$$

where $\Omega_2(n)$ is the number of odd prime divisors of n with multiplicity (think $\Omega(n)$ instead of $\omega(n)$). By checking on prime powers, we find that

$$2^{\Omega_2(n)} \geq d_{odd}(n), \qquad d_{odd}(n) = \sum_{d|n,\ 2\nmid d} 1.$$

Thus

$$V(z) \geq \sum_{n \leq z} \frac{d_{odd}(n)}{n}.$$

Note that

$$\sum_{n \leq z} d_{odd}(n) = z \sum_{\substack{n \leq z \\ 2\nmid n}} \frac{1}{d} + O(z) = z\Big(\sum_{n \leq z} \frac{1}{n} - \sum_{n \leq z/2} \frac{1}{2n}\Big) + O(z) = \frac{z}{2} \log z + O(z)$$

Hence by partial summation, $V(z) \gg (\log z)^2$, as desired.

With a lot more work, the ideas we have developed in this course can be wielded to prove

$$\pi_2(x) \leq (8 + o(1))\mathfrak{S} \frac{x}{(\log x)^2},$$

and Theorem 21.3 (which we did not prove) further implies that

$$\pi_2(x) \leq (4 + o(1))\mathfrak{S} \frac{x}{(\log x)^2}.$$

26. June 4

27. June 6