MATH 8440 / FALL 2026

Topics in Analytic and Combinatorial Number Theory

Elementary methods in the distribution of prime numbers

Paul Pollack

University of Georgia

The course at a glance

Infinitely many primes. Euclid, Euler, squarefree integers, and two proofs from algebra.

Prime-producing sequences. Euler's quadratic and an example of Erdős.

Counting primes. Chebyshev's bounds and Bertrand's postulate.

Mertens' estimates. Prime sums, products, and guesses about primes.

Primes in progressions. Schur's lemma and the classes 1 and -1.

Dirichlet's theorem. Characters and the values $L(1, \chi)$.

The prime number theorem. Möbius sums, smooth numbers, and the hyperbola method.
Bonus: $\sum \mu(n)/n = 0$.

Infinitely many primes (1 of 3)

Contents	Pages
Euclid's proof	6
The proofs of Stieltjes and Braun	7–8
Euclid's lemma	9
Furstenberg's proof	10–11
Goldbach's proof and the Fermat numbers	12–13
Prime factors of $2^p - 1$	14
Multiplicative functions	15
Euler's factorization theorem	16–17

Course overview

Infinitely many primes (2 of 3)

Contents	Pages
Euler's proof that $\sum_p 1/p$ diverges	18
Upper and lower bounds for $\sum_{p \leq x} 1/p$	19–20
Hacks's proof using $\zeta(2)$	21
Infinitude from $\zeta(2)^2/\zeta(4)$	22
Exercises on Euler products	23
Perott's proof using squarefree integers	24
Erdős's lower bound for $\pi(N)$	25
Erdős's proof that $\sum_p 1/p$ diverges	26

Course overview

Infinitely many primes (3 of 3)

Contents	Pages
The identity $\sum_{d\mid n}\mu(d)=\mathbf{1}_{n=1}$	27
Infinitude from the support of μ	28
Möbius inversion	29–32
Counting smooth integers	33–35
The maximal ideals of $\mathbb{Z}[X]$	36–37
Washington's proof	38–39

Course overview

Euclid's proof

A prime is an integer greater than 1 whose only positive divisors are 1 and itself. Here are the first few:

$$2, \quad 3, \quad 5, \quad 7, \quad 11, \quad 13, \quad 17.$$

Let $p_1, \ldots, p_k$ be any finite list of distinct primes, and put $P = p_1 \cdots p_k$. Look at $P + 1$. We know $P + 1 > 1$, so $P + 1$ has some prime factor q. But $P + 1 \equiv 1 \pmod{p_i}$ for all i, and so $q \notin \{p_1, \ldots, p_k\}$. So no finite list of primes is complete.

The proof of Stieltjes

Let $p_1, \ldots, p_k$ be a finite list of distinct primes, with product P. Factor $P = AB$ with $A, B > 0$, and consider $A + B$.

Each p_i divides exactly one of A, B: if it divided both, then p_i^2 would divide P. Hence $p_i \nmid A + B$.

But $A + B \geq 2$, so it has a prime factor. Again we have discovered a prime not on our original list.

Euclid's proof is the case $B = 1$.

Braun. Suppose $k \geq 2$. Put the fractions over a common denominator:

$$\frac{1}{p_1} + \cdots + \frac{1}{p_k} = \frac{N}{P}, \qquad N = \frac{P}{p_1} + \cdots + \frac{P}{p_k}.$$

For each i, p_i divides P/p_j for all $j \neq i$. Thus $N \equiv P/p_i \pmod{p_i}$. If $p_i \mid N$, then $p_i \mid P/p_i$ and $p_i^2 \mid P$, a contradiction. Since $k \geq 2$, we have $N > 1$. Take any prime $q \mid N$; it is not on our list. This is why the proof needs $k \geq 2$.

Another variant is to look at $m! + 1$. Every prime factor exceeds m.

Homework 1, Problem 2. Here is a slightly unusual proof of Euclid's lemma. Fix a prime p and an integer a, and consider $I = \{b \in \mathbb{Z} : p \mid ab\}$. The difference of two members of I again belongs to I, as does every integer multiple of a member of I. Thus I is an ideal of $\mathbb{Z}$.

Write $I = d\mathbb{Z}$ with $d > 0$ (notice that $p \in I$). Then $d \mid p$, so either $d = 1$ or $d = p$. If $d = 1$, then $1 \in I$, whence $p \mid a$. If $d = p$, then every b for which $p \mid ab$ is itself divisible by p. These are precisely the two alternatives in Euclid's lemma.

For $m > 0$ let

$$U_{a,m} = \{n \in \mathbb{Z} : n \equiv a \pmod{m}\}.$$

The *profinite topology* is the topology on $\mathbb{Z}$ with basic open sets $U_{a,m}$. To check that this makes sense, observe that these sets cover $\mathbb{Z}$. If n belongs to two of them, the progression through n modulo the product of their moduli lies in both (Homework 1, Problem 1).

Each set $U_{a,m}$ is closed. Indeed, its complement is the union of the other classes modulo m. Also, every nonempty open set is infinite, since it contains a full arithmetic progression.

Furstenberg's proof

Suppose $p_1, \ldots, p_k$ are all the primes. The set of integers divisible by at least one of them is

$$U_{0,p_1} \cup \cdots \cup U_{0,p_k}.$$

This set is closed (a finite union of closed sets). Hence its complement is open. But the complement consists of the integers divisible by no prime, namely $\{-1, 1\}$. This is a contradiction, since nonempty open sets are infinite.

Suppose $a_1, a_2, \ldots$ is an infinite sequence of integers greater than 1, with $(a_i, a_j) = 1$ whenever $j < i$. Let p_i be a prime factor of a_i. Then $p_1, p_2, \ldots$ is an infinite sequence of distinct primes.

How do we get such a sequence? Goldbach takes $a_1 = 3$, and for $i > 1$ sets

$$a_i = 2 + \prod_{1 \le j < i} a_j.$$

First, notice that each a_i is odd, by induction. If $d \mid a_i$ and $d \mid a_j$, where $j < i$, then $d \mid a_j \mid a_i - 2$. Thus $d \mid 2$; since a_i is odd, $d = 1$.

The sequence begins $3, 5, 17, 257, 65537$. This suggests $a_i = 2^{2^{i-1}} + 1$.

Goldbach's proof and the Fermat numbers

Write $F_n = 2^{2^n} + 1$ for the nth Fermat number. Factoring a difference of squares repeatedly gives

$$F_n - 2 = 2^{2^n} - 1 = (2^{2^{n-1}} + 1)(2^{2^{n-1}} - 1) = F_{n-1} \cdots F_0 \quad (n \geq 1).$$

Thus $a_i = F_{i-1}$. We do not need the Fermat numbers themselves to be prime. Indeed, $F_5 = 641 \cdot 6700417$.

If $p \mid F_n$, then $2^{2^n} \equiv -1 \pmod{p}$. The order of 2 modulo p divides 2^{n+1} but does not divide 2^n, so it equals 2^{n+1}.

In particular $2^{n+1} \mid p - 1$. Distinct n give distinct prime divisors, and for each fixed k infinitely many of them are $1 \bmod 2^k$.

Homework 1, Problem 3. Let $n \geq 2$ and let p be a prime divisor of F_n. Prove the stronger divisibility $2^{n+2} \mid p - 1$.

Infinitude of primes from the factors of $2^p - 1$

Suppose there is a largest prime p. Let q be any prime factor of $2^p - 1$. Then

$$2^p \equiv 1 \pmod{q}.$$

Hence the order of 2 modulo q divides p. The order is not 1, since $q \nmid 2 - 1$. It follows that the order is p. By Lagrange's theorem, $p \mid q - 1$. Hence $q > p$, contrary to the choice of p.

An arithmetic function is a function on the positive integers. We call f *multiplicative* if $f(1) = 1$ and $f(ab) = f(a)f(b)$ whenever $(a, b) = 1$. Hence

$$f(p_1^{e_1} \cdots p_k^{e_k}) = \prod_{i=1}^{k} f(p_i^{e_i})$$

whenever the p_i are distinct. Conversely, this identity implies multiplicativity (Homework 1, Problem 8).

The functions 1 and n^s are examples, as is Euler's $\phi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times$. For ϕ, apply the Chinese remainder theorem and count units on both sides. For $\sigma(n) = \sum_{d|n} d$, observe that specifying a divisor amounts to choosing, independently for each p_i, an exponent between 0 and e_i.

Euler's factorization theorem

Theorem

For a multiplicative function f,

$$\sum_{n\geq 1} f(n) = \prod_{p} (1 + f(p) + f(p^2) + \cdots)$$

if $\sum_n |f(n)|$ converges, or if the product with $|f|$ in place of f is finite.

Suppose first that $\sum_n |f(n)| < \infty$. On multiplying out just the factors corresponding to $p \leq x$, unique factorization gives

$$\prod_{p\leq x} (1 + f(p) + f(p^2) + \cdots) = \sum_{\substack{n\geq 1 \\ p|n \Rightarrow p\leq x}} f(n).$$

What have we left out? Only integers having a prime factor $> x$, all of which are themselves $> x$. Thus the error is bounded by $\sum_{n>x} |f(n)|$, which tends to zero.

Suppose now that we are given convergence of the product for $|f|$. Then

$$\sum_{n \le x} |f(n)| \le \prod_{p \le x} (1 + |f(p)| + |f(p^2)| + \cdots) \le \prod_{p} (1 + |f(p)| + |f(p^2)| + \cdots).$$

The partial sums on the left increase and are bounded. Thus the series converges absolutely, and the first argument applies.

Euler's proof that $\sum_p 1/p$ diverges

What happens if we try $f(n) = 1/n$? For any one prime p we have

$$1 + \frac{1}{p} + \frac{1}{p^2} + \cdots = (1 - 1/p)^{-1}.$$

If there were only finitely many primes, the product would be finite. Euler's theorem would then give convergence of $\sum_{n \geq 1} 1/n$!

In fact $\sum_p 1/p$ diverges. Otherwise the partial products would again be bounded, since

$$(1 - 1/p)^{-1} = 1 + \frac{1}{p-1} \leq 1 + \frac{2}{p} \leq e^{2/p}.$$

The products would be bounded by $\exp(2 \sum_p 1/p)$, so the harmonic series would converge. This proves Euler's assertion.

Upper and lower bounds for $\sum_{p \le x} 1/p$

For $x > 1$ the integral test gives $\log x < \sum_{n \le x} 1/n \le 1 + \log x$. Every term on the left occurs on multiplying out the product below, so

$$\log x < \prod_{p \le x} (1 - 1/p)^{-1} \le \exp\left(\sum_{p \le x} \frac{1}{p-1} \right).$$

Since $1/(p-1) = 1/p + 1/(p(p-1))$ and

$$\sum_{p \le x} \frac{1}{p(p-1)} \le \sum_{n \ge 2} \left(\frac{1}{n-1} - \frac{1}{n} \right) = 1,$$

taking logarithms gives $\sum_{p \le x} 1/p > \log\log x - 1$.

Upper and lower bounds for $\sum_{p \le x} 1/p$

Upper bounds are more difficult! Can we show that $\sum_{p \le x} 1/p$ is significantly smaller than $\log x$? Square the sum. How often can a given integer occur as a denominator? At most twice, since pq determines the unordered pair $\{p, q\}$. Hence

$$\left(\sum_{p \le x} \frac{1}{p} \right)^2 = \sum_{p,q \le x} \frac{1}{pq} \le 2 \sum_{n \le x^2} \frac{1}{n} \le 2 + 4 \log x.$$

Thus $\sum_{p \le x} 1/p = O(\sqrt{\log x}) = o(\log x)$. So we win! Higher powers give still better bounds.

Homework 1, Problem 11. Take higher powers of $\sum_{p \le x} 1/p$. By letting the exponent depend on x, prove that this sum is at most $\log \log x + O(\log \log \log x)$.

Hacks's proof using $\zeta(2)$

For $s > 1$, put $\zeta(s) = \sum_{n \geq 1} n^{-s}$. Euler's factorization gives

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1}.$$

At $s = 2$ the series equals $\pi^2/6$. Recall Euler's calculation: in the sine product

$$\frac{\sin t}{t} = \prod_{n \geq 1} \left(1 - \frac{t^2}{n^2 \pi^2}\right)$$

the coefficient of t^2 is $-\zeta(2)/\pi^2$. Comparing this with $-1/3!$ gives Euler's evaluation. If there were only finitely many primes, then $\pi^2/6$ would be a finite product of rational numbers. But $\pi^2/6$ is irrational (Homework 1, Problem 13).

Infinitude of primes from $\zeta(2)^2/\zeta(4)$

Comparing the coefficients of t^4 gives $(\zeta(2)^2 - \zeta(4))/(2\pi^4) = 1/5!$. Thus $\zeta(4) = \pi^4/90$, and

$$\frac{5}{2} = \frac{\zeta(2)^2}{\zeta(4)} = \prod_p \frac{1-p^{-4}}{(1-p^{-2})^2} = \prod_p \frac{p^2+1}{p^2-1}.$$

The quotient $5/2$ is rational. But there is something suspicious here: how does the 3 in the denominator of the first factor, $5/3$, go away? If there were only finitely many factors, some numerator p^2+1 would have to be divisible by 3. But -1 is not a square modulo 3, since the only squares are 0 and 1. This contradiction proves infinitude.

Homework 1, Problem 10. Let χ be zero on even integers, 1 on $1 \bmod 4$, and -1 on $3 \bmod 4$. Show that χ is completely multiplicative. Assuming there are only finitely many primes, show that

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \prod_p (1 - \chi(p)/p)^{-1}.$$

Use the value of the alternating series to reach a contradiction.

Homework 2, Problem 3. Let $\mathcal{P}$ be the set of primes dividing at least one Fermat number. Prove that $\sum_{p \in \mathcal{P}} 1/p$ converges. Use the order of 2 modulo p to count divisors of the higher-index Fermat numbers, and their sizes for the lower indices.

Perott's proof using squarefree integers

A positive integer is *squarefree* if it is divisible by no square greater than 1. It is determined by its set of prime divisors. With only k primes there would be only 2^k squarefree integers.

How many $n \leq N$ fail to be squarefree? At most

$$\sum_{m=2}^{\lfloor\sqrt{N}\rfloor} \left\lfloor \frac{N}{m^2} \right\rfloor \leq N \sum_{m\geq 2} \frac{1}{m^2} = cN.$$

Since $c < \int_1^\infty t^{-2}\,\mathrm{d}t = 1$, at least $(1-c)N$ integers up to N are squarefree. For N large this is more than 2^k, however large we took k.

Erdős's lower bound for $\pi(N)$

Every positive integer n can be written uniquely as uv^2, with u squarefree. Throw into u those primes appearing to an odd power in n.

Let $\pi(N) = \#\{p \le N\}$. If $n \le N$, there are at most $2^{\pi(N)}$ choices for u, since u is determined by its prime factors. There are at most $\sqrt{N}$ choices for v. Hence

$$N \le 2^{\pi(N)}\sqrt{N}, \qquad \pi(N) \ge \frac{\log N}{\log 4}.$$

The bound is crude: at $N = 10^6$, it gives $\pi(N) \ge 9.96\ldots$, whereas $\pi(N) = 78\,498$. But it tends to infinity, which is enough for our purpose.

Erdős's proof that $\sum_p 1/p$ diverges

Suppose $\sum_p 1/p$ converges. Choose y so that $\sum_{p>y} 1/p < 1/2$.

How many $n \le N$ have all their prime factors at most y? Writing $n = uv^2$ as before gives at most $2^{\pi(y)}\sqrt{N}$ possibilities.

Every remaining n is a multiple of a prime $p > y$. There are at most

$$\sum_{y<p\le N} \left\lfloor \frac{N}{p} \right\rfloor \le N \sum_{p>y} \frac{1}{p} < N/2$$

such multiples. We conclude that $N/2 \le 2^{\pi(y)}\sqrt{N}$. But y has been fixed all along, and this fails for all sufficiently large N.

Put $\mu(1) = 1$, put $\mu(n) = 0$ if n is not squarefree, and put $\mu(n) = (-1)^k$ if n is a product of k distinct primes.

For $n > 1$ with k distinct prime factors, only products of subsets of those primes contribute. Thus

$$\sum_{d|n} \mu(d) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} = (1-1)^k = 0.$$

For $n = 1$ we get 1, the contribution of the empty set of primes. Thus the single formula $\sum_{d|n} \mu(d) = \mathbf{1}_{n=1}$ covers both cases.

Infinitude of primes from the support of μ

Suppose $g(n) = \sum_{d|n} f(d)$ and both f and g vanish beyond some point. We claim f is identically zero.

Otherwise let D be the largest index with $f(D) \neq 0$. For $|z| < 1$,

$$\sum_{n \geq 1} g(n) z^n = \sum_{d \leq D} f(d) \sum_{j \geq 1} z^{dj} = \sum_{d \leq D} f(d) \frac{z^d}{1 - z^d}.$$

What's the problem? The left side is a polynomial, but the right side has a pole! Let z tend radially to a primitive Dth root ζ. For $d < D$, $\zeta^d \neq 1$, so those terms have finite limits. The $d = D$ term is unbounded, since $f(D) \neq 0$. Contradiction.

Take $f = \mu$ and $g = \mathbf{1}_{n=1}$. Finitely many primes would mean finitely many squarefree numbers, giving $\mu = 0$, although $\mu(1) = 1$.

We can regard $\sum_{n\geq 1} a_n n^{-s}$ as a formal series. What matters is the coefficient of each n^{-s}. Since $d^{-s}e^{-s} = (de)^{-s}$,

$$\left(\sum_{d\geq 1} \frac{a_d}{d^s}\right)\left(\sum_{e\geq 1} \frac{b_e}{e^s}\right) = \sum_{n\geq 1} \frac{1}{n^s} \sum_{de=n} a_d b_e.$$

Each coefficient is a finite sum. Here s is a formal variable; we are not asserting convergence for any value of s.

For power series the indices *add*; here they *multiply*. Formal power series are useful for additive problems, Dirichlet series for multiplicative ones.

Put $\zeta(s) = \sum_{n \geq 1} n^{-s}$. Then

$$\zeta(s)^2 = \sum_{n \geq 1} \frac{1}{n^s} \sum_{de=n} 1 = \sum_{n \geq 1} \frac{\tau(n)}{n^s}, \qquad \tau(n) = \#\{d : d \mid n\}.$$

More generally, the coefficient of n^{-s} in $\zeta(s)^k$ counts ordered k-tuples of positive integers with product n.

Multiplying the series for an arithmetic function f by $\zeta(s)$ replaces $f(n)$ by $\sum_{d|n} f(d)$. This is exactly the operation we want to undo in Möbius inversion.

Homework 1, Problem 12. Consider formal Dirichlet series with complex coefficients, with termwise addition and the multiplication above. Prove that this ring is an integral domain and determine its units.

How might we get f in terms of g? The key is to invert $\zeta(s)$.

$$\zeta(s)\sum_{n\geq 1}\frac{\mu(n)}{n^s}=\sum_{n\geq 1}\frac{1}{n^s}\sum_{d|n}\mu(d)=1.$$

Thus, if $g(n)=\sum_{d|n}f(d)$, then

$$\sum_{n\geq 1}\frac{g(n)}{n^s}=\zeta(s)\sum_{n\geq 1}\frac{f(n)}{n^s},\qquad \sum_{n\geq 1}\frac{f(n)}{n^s}=\left(\sum_{d\geq 1}\frac{\mu(d)}{d^s}\right)\left(\sum_{e\geq 1}\frac{g(e)}{e^s}\right).$$

Comparing coefficients gives $f(n)=\sum_{d|n}\mu(d)g(n/d)$. Multiplying back by $\zeta(s)$ proves the converse.

Why did we define μ this way?

There is still something of a mystery here. We gave a direct proof, so there is no logical gap. But there is still a psychological gap: why this definition of μ?

Start from the formal Euler product and invert:

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}, \qquad \zeta(s)^{-1} = \prod_p (1 - p^{-s}) = \sum_{n \geq 1} \frac{\mu(n)}{n^s}.$$

To obtain n^{-s}, select $-p^{-s}$ for each prime dividing n. Each prime can be selected at most once, so n must be squarefree; its coefficient is $(-1)^{\#\{p : p \mid n\}}$.

What does convergence mean here? Each coefficient is eventually constant. The product over $p \leq N$ already has the right coefficient for every $n \leq N$.

Infinitude of primes by counting smooth integers

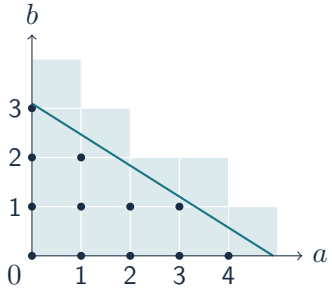
Call n y-*smooth* if every prime factor is at most y. Write $\Psi(x, y)$ for their count up to x, including 1.

For $y = 2$ we get $1 + \lfloor \log x / \log 2 \rfloor$. For $y = 3$ count nonnegative integer pairs (a, b) with

$$a \log 2 + b \log 3 \leq \log x.$$

Denote by $\Delta(t)$ the triangle $a, b \geq 0$, $a \log 2 + b \log 3 \leq t$. Our guess is that the number of lattice points is about the area. Draw a unit square with lower left corner at each lattice point counted. These squares cover $\Delta(\log x)$: round both coordinates down to find a square containing any given point. How far can they protrude? Increasing both coordinates by 1 adds only $\log 6$, so every square lies in $\Delta(\log x + \log 6)$.

Infinitude of primes by counting smooth integers



Each dot represents one integer $2^a 3^b \leq 30$. For instance, $(3, 1)$ represents 24. The next dot in that column would represent 72.

The shaded squares cover the triangle $a \log 2 + b \log 3 \leq \log 30$ in the first quadrant. Each has area 1, so their combined area is $\Psi(30, 3)$.

Infinitude of primes by counting smooth integers

The area of $\Delta(t)$ is $t^2/(2\log 2\log 3)$. The two areas differ by $O(\log x)$, so

$$\Psi(x,3) = \frac{(\log x)^2}{2\log 2\log 3} + O(\log x).$$

For the first k primes, draw unit cubes in k dimensions. Their union contains the simplex with $t = \log x$ and lies in the one with $t = \log x + \sum_i \log p_i$. Changing u_i to $u_i \log p_i$ gives volume $t^k/(k!\prod_i \log p_i)$. Thus

$$\Psi(x,p_k) = \frac{(\log x)^k}{k!\prod_{i=1}^k \log p_i} + O_k((\log x)^{k-1}).$$

If p_k were the last prime, $\Psi(x,p_k) = \lfloor x \rfloor$. But $\Psi(x,p_k)/x \to 0$, whereas $\lfloor x \rfloor / x \to 1$. This proves infinitude.

Suppose (f) is maximal. Then $\mathbb{Z}[X]/(f)$ is a field. A constant f will not do: $f = 0, \pm 1$ gives $\mathbb{Z}[X]$ or the zero ring; for $|f| > 1$ the class of X is a nonzero nonunit.

If $\deg f > 0$, choose an integer n with $|f(n)| > 1$. The constant $f(n)$ has nonzero class modulo $f(X)$. If it had an inverse, we could write

$$f(n)g(X) - f(X)h(X) = 1, \qquad g, h \in \mathbb{Z}[X].$$

Set $X = n$. The left side is divisible by $f(n)$, while the right side is 1. Thus $\mathbb{Z}[X]$ has no principal maximal ideal.

Infinitude of primes from the maximal ideals of $\mathbb{Z}[X]$

Suppose the complete list of primes is $p_1, \ldots, p_k$, and put $P = p_1 \cdots p_k$. What numbers do we get by adjoining $1/P$ to $\mathbb{Z}$? Every positive integer b divides some P^j. Thus $a/b = (aP^j/b)/P^j$ belongs to $\mathbb{Z}[1/P]$, and we get all of $\mathbb{Q}$. In particular,

$$\mathbb{Q} = \mathbb{Z}[1/P] \cong \mathbb{Z}[X]/(PX - 1).$$

Send X to $1/P$. Conversely send a/P^j to the class of aX^j; this is well defined since $PX = 1$ in the quotient.

But $\mathbb{Q}$ is a field. This would make $(PX - 1)$ a principal maximal ideal of $\mathbb{Z}[X]$, and we have just seen that there are none.

Washington's proof

Recall that the ring of integers $\mathcal{O}_K$ consists of the elements of K satisfying monic polynomials over $\mathbb{Z}$.

A *Dedekind domain* is a Noetherian, integrally closed domain whose nonzero prime ideals are maximal. We use two facts:

- Every nonzero prime ideal lies over a rational prime p: its intersection with $\mathbb{Z}$ is $p\mathbb{Z}$. At most $[K : \mathbb{Q}]$ prime ideals lie over any one p.
- $\mathcal{O}_K$ is a Dedekind domain. A Dedekind domain with only finitely many nonzero prime ideals is a principal ideal domain, hence has unique factorization of elements.

Thus finitely many primes would make $\mathcal{O}_K$ a principal ideal domain, hence a unique factorization domain, for every K. But this is false for some K.

In $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}]$ we have

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

Take norms, where $N(a + b\sqrt{-5}) = a^2 + 5b^2$. Norms multiply, and the only units are ± 1. There is no element of norm 2 or 3. Consequently elements of norms $4, 9, 6$ are irreducible (how could one factor their norms?). So all four factors are irreducible. Neither factor on the right can be associate to one on the left, since their norms differ. So unique factorization does fail.

Prime-producing sequences

Contents	Pages
Composite values of polynomials	41
Euler's prime-producing polynomial	42–43
Primes as polynomial values in several variables	44–46
Bunyakovsky's conjecture	47
Erdős's theorem on $2^n + k$	48–51
Pseudoprimes and Carmichael numbers	52–53

Course overview

Composite values of polynomials

Goldbach called Euler's attention to $f(X) = X^2 + 19X - 19$. At $X = 2, 4, 8, 16, 32$ one gets a prime each time. At 64 one gets $5293 = 67 \cdot 79$. Could a more fortunate choice of f succeed?

Let $f \in \mathbb{Z}[X]$ be nonconstant. Choose $n_0 > 0$ with $|f(n_0)| > 1$ and a prime $p \mid f(n_0)$. Then $p \mid f(n)$ whenever $n \equiv n_0 \pmod{p}$. Taking n large gives $|f(n)| > p$.

Even restriction to powers of 2 does not help. Otherwise, for large n_0 we would have $|f(2^{n_0})| = p$ with p odd. For every $n \equiv n_0 \pmod{p-1}$, Fermat's theorem gives $p \mid f(2^n)$. Now take n large.

Euler's prime-producing polynomial

Euler's $f(X) = X^2 - X + 41$ takes prime values for $1 \le n \le 40$. At $n = 41$, it gives 41^2. Why such a long run? The explanation comes from unique factorization in $\mathbb{Z}[(1+\sqrt{-163})/2]$.

Put $\theta = (1+\sqrt{-163})/2$. Then $\theta^2 = \theta - 41$, $\bar{\theta} = 1 - \theta$, and

$$\mathcal{O} = \mathbb{Z}[\theta] = \{a + b\theta : a, b \in \mathbb{Z}\}, \qquad f(n) = (n - \theta)(n - \bar{\theta}).$$

We use the fact that $\mathcal{O}$ is a unique factorization domain. Its norm is

$$N(a + b\theta) = (a + b\theta)(a + b\bar{\theta}) = a^2 + ab + 41b^2.$$

Why is $\mathcal{O} = \mathbb{Z} + \mathbb{Z}\theta$? The relation $\theta^2 = \theta - 41$ reduces every higher power to this form. Also $\bar{\theta} = 1 - \theta \in \mathcal{O}$, so $\mathcal{O}$ is closed under conjugation.

An element of norm 1 has its conjugate for an inverse. Conversely, a unit has norm 1: take norms in $\alpha\beta = 1$.

Euler's prime-producing polynomial

Suppose $1 \leq n < 41$ and $f(n)$ is composite. Since $1 < f(n) < 41^2$, some prime $p < 41$ divides $f(n)$.

If p were irreducible in $\mathcal{O}$, unique factorization would force $p \mid n - \theta$ or $p \mid n - 1 + \theta$. Neither is possible: dividing by p leaves coefficient $\pm 1/p$ on θ.

Thus $p = \alpha\beta$ for nonunits. Taking norms gives $p^2 = N\alpha N\beta$. Both norms are positive integers greater than 1, so both must equal p. Writing $\alpha = a + b\theta$ gives

$$p = (a + b/2)^2 + \frac{163}{4}b^2.$$

If $b = 0$, then $p = a^2$. Otherwise $p \geq 163/4$, hence $p \geq 41$. Both are impossible.

The primes as polynomial values in several variables

What if we allow polynomials in several variables? There is a polynomial in 26 variables whose positive values, as the variables range over the nonnegative integers, are precisely the primes. The word *positive* is important! Zero and negative values are allowed.

A prime-producing polynomial in 26 variables

The positive values of the following polynomial, for nonnegative integer values of $a, b, \ldots, z$, are precisely the primes.

$$
\begin{aligned}
P(a,b,\ldots,z) = (k+2)\Big\{ & 1-(wz+h+j-q)^2 \\
& -\Big((gk+2g+k+1)(h+j)+h-z\Big)^2 \\
& -(2n+p+q+z-e)^2 \\
& -\Big(16(k+1)^3(k+2)(n+1)^2+1-f^2\Big)^2 \\
& -\Big(e^3(e+2)(a+1)^2+1-o^2\Big)^2 \\
& -\Big((a^2-1)y^2+1-x^2\Big)^2 \\
& -\Big(16r^2y^4(a^2-1)+1-u^2\Big)^2 \\
& -\Big(((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2\Big)^2 \\
& -(n+\ell+v-y)^2-\Big((a^2-1)\ell^2+1-m^2\Big)^2-(ai+k+1-\ell-i)^2 \\
& -\Big(p+\ell(a-n-1)+b(2an+2a-n^2-2n-2)-m\Big)^2 \\
& -\Big(q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x\Big)^2 \\
& -\Big(z+p\ell(a-p)+t(2ap-p^2-1)-pm\Big)^2\Big\}.
\end{aligned}
$$

Degree 25. Here ℓ denotes the variable l.

Jones, Sato, Wada, and Wiens, *American Mathematical Monthly* 83 (1976), 449–464.

Returning to one variable

Which functions assume infinitely many prime values?

For a linear polynomial $a + mX$, with $m > 0$ and $(a, m) = 1$, Dirichlet's theorem gives the answer. What about higher degree polynomials?

Bunyakovsky's conjecture

For a nonconstant $f \in \mathbb{Z}[X]$ with positive leading coefficient we certainly want f irreducible. We must also exclude a *fixed prime divisor*, a prime p dividing $f(n)$ for every integer n.

For example, $X^2 + X + 2$ has relatively prime coefficients but is even at every integer.

Bunyakovsky's conjecture

Under those two conditions, $f(n)$ is prime for infinitely many positive integers n.

For a linear polynomial this is Dirichlet's theorem. Already for $X^2 + 1$ we do not know the answer. Iwaniec proved that $n^2 + 1$ has at most two prime factors (counted with multiplicity) for infinitely many n.

Erdős's theorem on the numbers $2^n + k$

Erdős proved that there is an odd positive integer k such that $2^n + k$ is composite for all positive integers n. We work piecemeal, imposing more conditions on k to cover all n.

Baby problem: find a k for which $2^n + k$ is never prime for an *odd* n. Then $2^n \equiv -1 \pmod 3$, so choose $k \equiv 1 \pmod 3$, $k > 3$. If $n \equiv 2 \pmod 4$, then $2^n \equiv -1 \pmod 5$, so choose $k \equiv 1 \pmod 5$.

Recall that if $p \mid F_j$, then $2^{2^j} \equiv -1 \pmod p$ and the order of 2 modulo p is 2^{j+1}. Use prime factors of the Fermat numbers to cover the remaining classes.

Erdős's theorem on the numbers $2^n + k$

n lies in	a divisor of $2^n + k$	impose on k
$1 \bmod 2$	3	$1 \bmod 3$
$2 \bmod 4$	5	$1 \bmod 5$
$4 \bmod 8$	17	$1 \bmod 17$
$8 \bmod 16$	257	$1 \bmod 257$
$16 \bmod 32$	65537	$1 \bmod 65537$
$32 \bmod 64$	641	$1 \bmod 641$
$0 \bmod 64$	6700417	$-1 \bmod 6700417$

Why does the table cover every n? Either n is odd, or $2 \mid n$. In the latter case, either $n \equiv 2 \pmod 4$, or $4 \mid n$. Keep going: after the first six rows, only $64 \mid n$ remains.

Finishing the covering argument

The last two rows use the two prime factors of F_5:

$$F_5 = 2^{32} + 1 = 641 \cdot 6700417.$$

If $n \equiv 32 \pmod{64}$, then $2^n \equiv -1 \pmod{641}$. If $64 \mid n$, then $2^n \equiv 1 \pmod{6700417}$.

This is why two factors of the same Fermat number are useful: one deals with $32 \pmod{64}$, the other with $0 \pmod{64}$.

Our seven moduli, together with 2, are pairwise coprime. The Chinese remainder theorem supplies an odd $k > 6700417$ satisfying every condition. Whichever row contains n, the indicated prime is a proper divisor of $2^n + k$.

Which k do we get? The six congruences with residue 1 say $k = 1 + Mt$, where

$$M = 3 \cdot 5 \cdot 17 \cdot 257 \cdot 65537 \cdot 641 = 641(2^{32} - 1).$$

Put $q = 6\,700\,417$. Since $2^{32} + 1 = 641q$, $M \equiv -2 \cdot 641 \pmod{q}$. Thus $1 + Mt \equiv -1 \pmod{q}$ precisely when $641t \equiv 1 \pmod{q}$.

Now $641 \cdot 5\,634\,204 - 539q = 1$. Taking $t = 5\,634\,204$ gives

$$k = 1 + 641(2^{32} - 1) \cdot 5\,634\,204 = 15\,511\,380\,746\,462\,593\,381.$$

This k is odd and exceeds all seven primes, so each indicated divisor is proper.

Homework 2, Problem 4. Prove that for every integer $n \geq 1$, the number $2^n + 78557$ is divisible by at least one of $3, 5, 7, 13, 19, 37, 73$.

Pseudoprimes to base 2

If N is an odd prime, then $2^{N-1} \equiv 1 \pmod{N}$ by Fermat's little theorem. Does the converse hold?

Consider $341 = 11 \cdot 31$. Since $2^{10} = 1024 = 3 \cdot 341 + 1$, we have $2^{340} \equiv 1 \pmod{341}$. A composite that passes this test is called a *pseudoprime to base* 2.

Carmichael numbers

Some composites pass for every base coprime to them. For example, $561 = 3 \cdot 11 \cdot 17$, and $2, 10, 16$ all divide 560. If $(a, 561) = 1$, Fermat's little theorem gives $a^{560} \equiv 1$ modulo each of $3, 11, 17$, hence modulo 561. Such numbers are called *Carmichael numbers*.

Contents	Pages
A heuristic for the prime number theorem	55–56
The product formula for rational numbers	57
Infinitude from the product formula	58–59
Chebyshev's estimates for $\pi(x)$	60–66
A quantitative form of Bertrand's postulate	67–68

Course overview

A heuristic for the prime number theorem

We know $\pi(x) \to \infty$. How fast? Let's guess!

Our lower bound, together with Homework 1, Problem 11, gives $\sum_{p \leq x} 1/p \sim \log\log x$.

Partial summation (Homework 2, Problem 7) writes the sum as

$$\frac{\pi(x)}{x} + \int_2^x \frac{\pi(t)}{t^2}\,\mathrm{d}t.$$

Since $\pi(x)/x \leq 1$, the integral alone is asymptotic to $\log\log x$.

A heuristic for the prime number theorem

Suppose we replace $\pi(t)$ by a good smooth approximation $P(t)$. We expect

$$\int_2^x \frac{P(t)}{t^2}\,\mathrm{d}t \sim \log\log x.$$

Differentiating suggests

$$\frac{P(x)}{x^2} \sim \frac{1}{x\log x}, \qquad P(x) \sim \frac{x}{\log x}.$$

"Hence" $\pi(x) \sim x/\log x$.

This is a guess, not a proof: an asymptotic relation need not survive differentiation. The prime number theorem says that the guess is right!

The product formula for rational numbers

For a prime p, let $v_p(n)$ be its exponent in a nonzero integer n. Extend by $v_p(a/b) = v_p(a) - v_p(b)$, and put $|r|_p = p^{-v_p(r)}$ for $r \neq 0$, $|0|_p = 0$.

As in Homework 2, Problem 6, $|rs|_p = |r|_p|s|_p$ and $|r+s|_p \leq \max(|r|_p, |s|_p)$: after taking a common denominator, the smaller power of p dividing both summands still divides their sum.

For a nonzero rational r,

$$|r| \prod_p |r|_p = 1.$$

Write $r = \pm \prod p^{e_p}$. Then $|r| = \prod p^{e_p}$, while the factor at p is p^{-e_p}. All but finitely many factors are 1.

Infinitude of primes from the product formula

Given a finite list S of primes, let $N = 1 + \prod_{p \in S} p$. For $p \in S$, $|N|_p = 1$. The product formula reads

$$\prod_{p \notin S} |N|_p = \frac{1}{N} < 1.$$

Since the product is less than 1, at least one prime outside S divides N. This is Euclid's proof again.

From infinitude to counting primes

If we understand both the size and the prime factorization of an integer N, we get information about primes.

For counting primes up to x, the candidate is

$$N_x = 1 \cdot 2 \cdot 3 \cdots \lfloor x \rfloor = \lfloor x \rfloor !.$$

Every prime up to x divides it. We can estimate its size by summing $\log n$, and count its prime factors by counting multiples of $p, p^2, p^3, \ldots$.

Theorem

Let $0 < c_1 < \log 2$ and $c_2 > 2\log 2$. For all sufficiently large x,

$$c_1 \frac{x}{\log x} \le \pi(x) \le c_2 \frac{x}{\log x}.$$

We first bound $\psi(x) = \sum_{p^k \le x} \log p$ and compare it with $\pi(x)\log x$. We compute $\log \lfloor x \rfloor !$ in two ways: summing $\log n$ and counting prime factors. Subtracting twice the identity at $x/2$ from the identity at x gives an alternating sum with decreasing terms.

Chebyshev's estimates for $\pi(x)$

Put $T(x) = \log \lfloor x \rfloor ! = \sum_{n \le x} \log n$, with natural logarithms. Put $N = \lfloor x \rfloor$ and compare areas under $\log t$. For $x \ge 2$,

$$\int_1^N \log t \, \mathrm{d}t \le T(x) \le \int_1^{N+1} \log t \, \mathrm{d}t.$$

The two integrals differ by at most $\log(N+1)$, and $\int_1^x \log t \, \mathrm{d}t = x \log x - x + 1$ lies between them. Consequently

$$T(x) = x \log x - x + O(\log x).$$

The lower bound $T(x) \ge x \log x - x - K \log x$ is Homework 2, Problem 8. Here $O(\log x)$ means a quantity bounded in absolute value by a constant times $\log x$ for $x \ge 2$.

How many copies of p are there in $\lfloor x \rfloor !$? Each multiple of p contributes one, each multiple of p^2 a second, and so on:

$$v_p(\lfloor x \rfloor !) = \sum_{n \le x} \sum_{\substack{k \ge 1 \\ p^k | n}} 1 = \sum_{k \ge 1} \left\lfloor \frac{x}{p^k} \right\rfloor .$$

Put $\theta(x) = \sum_{p \le x} \log p$ and $\psi(x) = \sum_{p^k \le x} \log p$, where $k \ge 1$. Taking logarithms and reversing summation gives

$$T(x) = \sum_{p^k \le x} \log p \left\lfloor \frac{x}{p^k} \right\rfloor = \sum_{j \le x} \sum_{p^k \le x/j} \log p = \sum_{j \le x} \psi(x/j).$$

In $T(x) = \sum_j \psi(x/j)$ the even-indexed terms sum to $T(x/2)$. Subtracting $2T(x/2)$ changes precisely their signs:

$$\psi(x) - \psi(x/2) + \psi(x/3) - \psi(x/4) + \cdots = T(x) - 2T(x/2).$$

The sum stops once $x/j < 2$. For $x \geq 4$,

$$\begin{aligned} T(x) - 2T(x/2) &= x\log x - x - x\log(x/2) + x + O(\log x) \\ &= x\log 2 + O(\log x). \end{aligned}$$

We have an alternating sum with decreasing terms! Its first term gives an upper bound for the sum, the first two terms a lower bound, and so on.

Keeping the first term gives $\psi(x) \geq x \log 2 - O(\log x)$. Powers of any single prime contribute at most $\log x$, since

$$\psi(x) = \sum_{p \leq x} \log p \left\lfloor \frac{\log x}{\log p} \right\rfloor \leq \pi(x) \log x.$$

It follows that

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

Thus any $c_1 < \log 2$ works in the lower bound once x is large enough.

Stopping after two terms gives

$$\psi(x) - \psi(x/2) \le x \log 2 + C \log x \qquad (x \ge 4).$$

Replace x by $x/2$, then $x/4$, and keep going. Choose k as large as possible with $x/2^{k-1} \ge 4$. On adding, the left side is $\psi(x) - \psi(x/2^k)$. Since $2 \le x/2^k < 4$, the last term is bounded. The main terms sum to at most $x \log 2 \sum_{j \ge 0} 2^{-j} = 2x \log 2$. There are $O(\log x)$ error terms, each $O(\log x)$. Thus

$$\psi(x) \le (2 \log 2)x + O((\log x)^2).$$

Fix $\epsilon \in (0,1)$ and let $y = x^{1-\epsilon}$. Every prime in $(y, x]$ contributes at least $\log y$ to $\psi(x)$. Thus

$$\pi(x) \le \pi(y) + \frac{\psi(x)}{\log y} \le x^{1-\epsilon} + \frac{1}{1-\epsilon}\frac{\psi(x)}{\log x}.$$

Divide by $x/\log x$. The first term tends to zero, and the limsup of the second is at most $(2\log 2)/(1-\epsilon)$. Now let $\epsilon \downarrow 0$.

Chebyshev's estimates

For $0 < c_1 < \log 2$ and $c_2 > 2\log 2$, $c_1 x/\log x \le \pi(x) \le c_2 x/\log x$ for all sufficiently large x.

Changing constants covers every $x \ge 2$: the ratio is bounded above and away from zero on the remaining bounded interval.

Fix $K > 2$. Choose $c_1 < \log 2$ and $c_2 > 2\log 2$ close enough to the endpoints that $Kc_1/c_2 > 1$. For large x,

$$\frac{\pi(Kx)}{\pi(x)} \geq K\frac{c_1}{c_2}\frac{\log x}{\log(Kx)} > 1.$$

This gives a prime in $(x, Kx]$. But the ratio argument does not reach $K = 2$: we need $Kc_1/c_2 > 1$, whereas $c_1/c_2 < 1/2$.

For $K = 2$, go back to the alternating sum and keep one more term:

$$\begin{aligned}\psi(x) - \psi(x/2) &\geq x\log 2 - \psi(x/3) - O(\log x)\\ &\geq \tfrac{1}{3}(\log 2)x - O((\log x)^2).\end{aligned}$$

A quantitative form of Bertrand's postulate

If $p^k \le x$ with $k \ge 2$, then $p \le \sqrt{x}$. All powers of any one p contribute at most $\log x$. Hence

$$0 \le \psi(x) - \theta(x) \le \sqrt{x} \log x.$$

In particular,

$$\psi(x) - \psi(x/2) \le (\pi(x) - \pi(x/2)) \log x + \sqrt{x} \log x.$$

Comparing with the lower bound gives

$$\pi(x) - \pi(x/2) \ge \frac{\log 2}{3} \frac{x}{\log x} - O(\sqrt{x}).$$

For each $0 < c_3 < \frac{1}{3} \log 2$ and all sufficiently large x, $(x/2, x]$ contains more than $c_3 x / \log x$ primes. In particular it contains at least one.

Mertens' estimates (1 of 2)

Contents	Pages
Mertens' estimate for $\sum_{p \leq x} (\log p)/p$	71
Partial summation	72
Mertens' estimate for $\sum_{p \leq x} 1/p$	73–74
Mertens' product up to a constant factor	75
The harmonious truth	76
Mertens' product theorem	77
Reading supplement: the constant in the product	78–80

Course overview

Mertens' estimates (2 of 2)

Contents	Pages
The theorem of Montgomery and Wagon	81–83
How (not) to prove the PNT	84
Gauss, Legendre, and Littlewood's theorem	85–90
A heuristic for primes ending in 3	91
A heuristic for twin primes	92
A heuristic for prime values of $n^2 + 1$	93–94
Numerical examples	95

Course overview

Mertens' estimate for $\sum_{p\le x}(\log p)/p$

Recall $\log\lfloor x\rfloor! = \sum_{p^k\le x}\log p\lfloor x/p^k\rfloor$. Dropping the floor incurs error at most $\psi(x) = O(x)$. Thus

$$x\sum_{p^k\le x}\frac{\log p}{p^k} = \log\lfloor x\rfloor! + O(x) = x\log x + O(x).$$

Divide by x. If it were not for the terms with $k \ge 2$, we would already have the estimate we want. But these nuisance terms contribute a bounded amount:

$$\sum_p\sum_{k\ge 2}\frac{\log p}{p^k} = \sum_p\frac{\log p}{p(p-1)} \le \sum_{n\ge 2}\frac{\log n}{n(n-1)} < \infty.$$

Consequently $\sum_{p\le x}(\log p)/p = \log x + O(1)$ for $x \ge 2$.

The partial summation formula

From Homework 2, Problem 7: put $A(t) = \sum_{n \le t} a_n$. For continuously differentiable h and $x \ge 1$,

$$\sum_{n \le x} a_n h(n) = A(x)h(x) - \int_1^x A(t)h'(t)\,\mathrm{d}t.$$

Substitute $h(n) = h(x) - \int_n^x h'(t)\,\mathrm{d}t$ on the left. The coefficient of $h'(t)$ is $-\sum_{n \le t} a_n = -A(t)$.

For example, take a_n to be 1 on primes and 0 elsewhere, and $h(t) = 1/t$:

$$\sum_{p \le x} \frac{1}{p} = \frac{\pi(x)}{x} + \int_2^x \frac{\pi(t)}{t} \frac{\mathrm{d}t}{t}.$$

Mertens' estimate for $\sum_{p \le x} 1/p$

Write $A(t) = \sum_{p \le t} (\log p)/p = \log t + R(t)$ for $t \ge 2$, where $R(t) = O(1)$. Partial summation with $h(t) = 1/\log t$ gives

$$\sum_{p \le x} \frac{1}{p} = \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2}\,\mathrm{d}t.$$

The $\log t$ term contributes $1 + \log\log x - \log\log 2$. Since R is bounded,

$$\int_2^\infty \frac{R(t)}{t(\log t)^2}\,\mathrm{d}t$$

converges absolutely. Its tail has absolute value at most

$$C \int_x^\infty \frac{\mathrm{d}t}{t(\log t)^2} = \frac{C}{\log x}.$$

Thus we can put

$$B = 1 - \log\log 2 + \int_2^\infty \frac{R(t)}{t(\log t)^2}\,\mathrm{d}t.$$

Also $R(x)/\log x = O(1/\log x)$. Collecting estimates gives

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

The point is that the error in $A(t) = \log t + O(1)$ contributes a convergent integral. Replacing that integral by its limit leaves only a tail of size $O(1/\log x)$.

Mertens' product up to a constant factor

Take logarithms and use $-\log(1-u) = \sum_{k\geq 1} u^k/k$:

$$\log \prod_{p\leq x} (1-1/p)^{-1} = \sum_{p\leq x} \frac{1}{p} + \sum_{p\leq x} \sum_{k\geq 2} \frac{1}{kp^k}.$$

The double sum converges over all primes; its $p > x$ tail is at most $\sum_{n>x} 1/(n(n-1)) = O(1/x)$. For some constant b, the logarithm is $\log\log x + b + O(1/\log x)$. Taking exponentials and using $e^{O(1/\log x)} = 1 + O(1/\log x)$ gives

$$\prod_{p\leq x} (1-1/p) = \frac{e^{-b}}{\log x} + O\left(\frac{1}{(\log x)^2}\right).$$

What is b?

The harmonious truth

We know the harmonic sum is within 1 of $\log x$. In fact, the difference tends to a limit.

Put $\gamma_n = 1/n - \int_n^{n+1} \mathrm{d}t/t$. By monotonicity,

$$0 \le \gamma_n \le \frac{1}{n} - \frac{1}{n+1}, \qquad \gamma = \sum_{n \ge 1} \gamma_n$$

exists, and $\sum_{n>N} \gamma_n \le 1/(N+1)$. Thus

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

Take $N = \lfloor x \rfloor$. Since $0 \le \log((N+1)/x) \le 1/x$,

$$\sum_{n \le x} \frac{1}{n} = \log x + \gamma + O(1/x).$$

Mertens' product theorem

Mertens' product theorem says that b is precisely this γ. For $x \geq 3$,

$$\prod_{p \leq x} (1 - 1/p) = \frac{e^{-\gamma}}{\log x} + O\left(\frac{1}{(\log x)^2}\right).$$

It remains to identify the constant. We do this by evaluating the same limit in two different ways.

Reading supplement: the constant in Mertens' product

Put $B_1 = \lim_{x\to\infty}(\sum_{p\le x} 1/p - \log\log x)$ and $B_2 = \sum_p \sum_{k\ge 2} 1/(kp^k)$. We must show $B_1 + B_2 = \gamma$.

For real $s > 1$, removing prime terms from the logarithm of the Euler product gives

$$F(s) := \log\zeta(s) - \sum_p p^{-s} = \sum_p \sum_{k\ge 2} \frac{1}{kp^{ks}}.$$

The summands are bounded by the convergent series for B_2. Hence $F(s) \to B_2$ as $s \downarrow 1$. We now derive another expression for $F(s)$ which will make visible that the limit is $\gamma - B_1$.

P. Pollack, *Not Always Buried Deep*, Chapter 3, Theorem 3.17, pp. 99–100.

Reading supplement: the constant in Mertens' product

Put $h = s - 1$. The integral test gives $1/h \le \zeta(1+h) \le 1 + 1/h$. Since $1 - e^{-h} = h(1 + O(h))$,

$$\log \zeta(1+h) = -\log(1 - e^{-h}) + O(h) = \sum_{n \ge 1} \frac{e^{-hn}}{n} + O(h).$$

Put $H(t) = \sum_{n \le t} 1/n$, $P(u) = \sum_{p \le u} 1/p$. Using $e^{-hn} = h \int_n^\infty e^{-ht}\,\mathrm{d}t$ gives

$$\sum_{n \ge 1} \frac{e^{-hn}}{n} = h \int_0^\infty H(t) e^{-ht}\,\mathrm{d}t.$$

Likewise $p^{-h} = h \int_{\log p}^\infty e^{-ht}\,\mathrm{d}t$, so $\sum_p p^{-1-h} = h \int_0^\infty P(e^t) e^{-ht}\,\mathrm{d}t$.

Subtracting gives

$$F(1+h) = h\int_0^\infty (H(t) - P(e^t))e^{-ht}\,\mathrm{d}t + O(h).$$

The logarithms cancel: $H(t) - P(e^t) = \gamma - B_1 + O(1/(1+t))$ for $t \geq 1$. Enlarging the constant extends the estimate to $0 \leq t \leq 1$.

The constant integrates to $\gamma - B_1$. Split the error at $t = 1/h$:

$$h\int_0^\infty \frac{e^{-ht}}{1+t}\,\mathrm{d}t \leq h\log(1+1/h) + h\int_{1/h}^\infty he^{-ht}\,\mathrm{d}t = o(1).$$

Thus $B_2 = \gamma - B_1$, proving Mertens' product theorem.

Montgomery and Wagon

Suppose $\pi(x) \sim xW(x)$, with W positive and eventually decreasing. Then $W(x) \sim 1/\log x$. In particular, if $\pi(x) \sim cx/\log x$ with $c > 0$, then $c = 1$.

In other words: if $\pi(x)/x$ is "steadily decreasing to zero", then the PNT holds. To see what Mertens' estimate says about W, first express it in terms of $\pi(t)$. Partial summation gives

$$\sum_{p \le x} \frac{\log p}{p} = \frac{\pi(x)\log x}{x} + \int_2^x \frac{\pi(t)(\log t - 1)}{t^2}\,\mathrm{d}t.$$

By Chebyshev, the first term is $O(1)$ and $\int_2^x \pi(t)\,\mathrm{d}t/t^2 = O(\log\log x)$. Mertens now gives $\int_2^x (\pi(t)/t)\log t\,\mathrm{d}t/t \sim \log x$.

By Homework 3, Problem 7(a), replace $\pi(t)/t$ by $W(t)$ in the integral: their ratio is eventually between $1-\delta$ and $1+\delta$, and a fixed initial interval contributes $O(1)$. Divide by $\log x$ and let $\delta \to 0$. More explicitly, put

$$A(x) = \int_2^x W(t) \frac{\log t}{t}\, \mathrm{d}t.$$

For each fixed $0 < \delta < 1$ and all sufficiently large x,

$$(1-\delta)A(x) + O_\delta(1) \le \int_2^x \frac{\pi(t)}{t}\frac{\log t}{t}\, \mathrm{d}t \le (1+\delta)A(x) + O_\delta(1).$$

The middle expression is $\sim \log x$. Thus

$$\frac{1}{1+\delta} \le \liminf \frac{A(x)}{\log x} \le \limsup \frac{A(x)}{\log x} \le \frac{1}{1-\delta},$$

and so $A(x) \sim \log x$.

Where does monotonicity enter?

We have $\int_2^x W(t)(\log t)\,\mathrm{d}t/t \sim \log x$. We must show $W(x)\log x \to 1$. This is a Tauberian step: recover the function from a weighted average, or "unscramble an egg".

Fix $0 < \epsilon < 1$. Subtract the integrals up to x and $x^{1+\epsilon}$. Their difference is $\epsilon \log x + o(\log x)$, and is at most

$$W(x) \int_x^{x^{1+\epsilon}} \frac{\log t}{t}\,\mathrm{d}t = W(x)\epsilon(1+\epsilon/2)(\log x)^2.$$

Hence $\liminf W(x)\log x \geq (1+\epsilon/2)^{-1}$. On $[x^{1-\epsilon}, x]$, the integral is at least $W(x)\epsilon(1-\epsilon/2)(\log x)^2$, so $\limsup W(x)\log x \leq (1-\epsilon/2)^{-1}$. Let $\epsilon \downarrow 0$.

We used

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

to bound $\psi(x)$. Could we recover $\psi(x) \sim x$ by unscrambling this average?

Suppose f is nonnegative and nondecreasing and

$$\sum_{n \le x} f(x/n) = x \log x + Bx + o(x).$$

Optimistic Tauberian Theorem: $f(x) \sim x$.

Ingham proved that this is true! But the simplest proof uses the PNT.

Legendre proposed $x/(\log x - 1.08366)$. Gauss counted primes in blocks of a thousand. Density $1/\log t$ near t suggested

$$\mathrm{Li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}.$$

Lesson 1: Legendre and Gauss are both right (to first order). For Legendre this is immediate. For Gauss, integrate by parts:

$$\mathrm{Li}(x) = \frac{x}{\log x} - \frac{2}{\log 2} + \int_2^x \frac{\mathrm{d}t}{(\log t)^2}.$$

Split the last integral at $\sqrt{x}$: the pieces are $O(\sqrt{x})$ and $O(x/(\log x)^2)$. This works for any fixed power of $1/\log t$.

Lesson 2: Gauss was more right. Repeated integration by parts gives, for fixed $k \geq 1$,

$$\mathrm{Li}(x) = \sum_{j=1}^{k} \frac{(j-1)!x}{(\log x)^j} + O_k\left(\frac{x}{(\log x)^{k+1}}\right).$$

The classical PNT with error term gives $\pi(x) = \mathrm{Li}(x) + O(xe^{-c\sqrt{\log x}})$ for some $c > 0$. We quote this stronger result. It gives the same expansion for $\pi(x)$; in particular,

$$\frac{x}{\pi(x)} = \log x \left(1 + \frac{1}{\log x} + O((\log x)^{-2})\right)^{-1} = \log x - 1 + O(1/\log x).$$

Thus if $\pi(x) = x/(\log x - c(x))$, the limit of $c(x)$ is 1. Wherefore 1.08366?

Keeping three terms gives

$$\pi(x) = \frac{x}{\log x} + \frac{x}{(\log x)^2} + \frac{2x}{(\log x)^3} + O\left(\frac{x}{(\log x)^4}\right),$$
$$\frac{x}{\log x - 1} = \frac{x}{\log x} + \frac{x}{(\log x)^2} + \frac{x}{(\log x)^3} + O\left(\frac{x}{(\log x)^4}\right).$$

Thus $x/(\log x - 1)$ underestimates by about $x/(\log x)^3$. To compensate, use $x/(\log x - (1+\delta))$! Paasche and Hiary found that 1.08366 is close to the value minimizing the average error for those data.

Comparing $\pi(x)$ with $\mathrm{Li}(x)$

Recall our convention $\mathrm{Li}(x) = \int_2^x \mathrm{d}t/\log t$. The last two columns are rounded to one decimal place.

x	$\pi(x)$	$\mathrm{Li}(x)$	$\mathrm{Li}(x) - \pi(x)$
10^2	25	29.1	4.1
10^3	168	176.6	8.6
10^4	1 229	1 245.1	16.1
10^6	78 498	78 626.5	128.5
10^8	5 761 455	5 762 208.3	753.3
10^{10}	455 052 511	455 055 613.5	3 102.5
10^{12}	37 607 912 018	37 607 950 279.8	38 261.8

The error is getting larger, but it is a smaller and smaller fraction of $\pi(x)$. Notice also: every entry in the last column is positive!

Prime counts: R. C. Vaughan's table; integral values computed with the convention above.

How much is $\mathrm{Li}(x)$ off by? At $x = 10^{20}$, rounded to the nearest integer:

$x/\log x$	$2\,171\,472\,409\,516\,259\,138$
$x/(\log x - 1.08366)$	$2\,223\,801\,523\,570\,829\,204$
$\mathrm{Li}(x)$	$2\,220\,819\,602\,783\,663\,483$
$\pi(x)$	$2\,220\,819\,602\,560\,918\,840$

Gauss's guess is too large by about $222\,744\,643$, or one part in 10^{10}. The Riemann hypothesis predicts $|\pi(x) - \mathrm{Li}(x)| = O_\epsilon(x^{1/2+\epsilon})$ for every $\epsilon > 0$.

Again $\mathrm{Li}(x) > \pi(x)$. Is this true for all large x?

Prime count from R. C. Vaughan's table. Here $\mathrm{Li}(x) = \int_2^x \mathrm{d}t/\log t$.

Littlewood's theorem

Our table might lead one to conjecture that $\mathrm{Li}(x) - \pi(x) \to \infty$ as $x \to \infty$. Littlewood's theorem has a surprising consequence: this difference changes sign infinitely often.

Littlewood (1914)

The difference $\pi(x) - \mathrm{Li}(x)$ takes both positive and negative values for arbitrarily large x. In particular, it changes sign infinitely often.

In particular, there are arbitrarily large x for which the inequality suggested by our table is reversed.

A heuristic for primes ending in 3

Adding $1/\log(10k+3)$ over $10k+3 \le x$ gives about $\mathrm{Li}(x)/10$. Fishy! The same prediction would apply to numbers ending in 2, among which precisely one is prime.

The issue is that $3+10k$ is biased, arithmetically: it is always odd and never divisible by 5. We must compare its chance of passing each divisibility test with that of a random integer.

The correction factors are

$$\frac{1}{1-1/2} = 2, \qquad \frac{1}{1-1/5} = \frac{5}{4}.$$

At every other prime the fraction passing the divisibility test is the usual one. The revised guess is $2 \cdot (5/4) \cdot \mathrm{Li}(x)/10 = \mathrm{Li}(x)/4$.

A heuristic for twin primes

How do we think the count of twin primes behaves? The naive guess is $\int_2^x \mathrm{d}t/(\log t)^2$. Again we need correction factors.

For $n, n+2$ to be prime, both must avoid each small prime p. For independent integers the fraction passing is $(1-1/p)^2$.

At $p=2$ the pair passes half the time, giving correction 2. At odd p, forbidden classes $0, -2$ are distinct, giving correction

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

Our guess for $\#\{p \le x : p+2 \text{ prime}\}$ is

$$2\prod_{p>2}\left(1-\frac{1}{(p-1)^2}\right)\int_2^x \frac{\mathrm{d}t}{(\log t)^2}.$$

The product is positive since $\sum_{p>2} 1/(p-1)^2$ converges. We cannot even show that the actual count tends to infinity.

The uncorrected guess is $\mathrm{Li}(x)/2$, since $\log(n^2+1) \sim 2\log n$. Put $\rho(p) = \#\{a \bmod p : a^2+1 \equiv 0 \pmod p\}$. The correction is $c_p = (1-\rho(p)/p)/(1-1/p)$.

At 2 it is 1. For odd p, a root has order 4 in $\mathbb{F}_p^\times$. Conversely, an element of order 4 squares to -1. Since $\mathbb{F}_p^\times$ is cyclic, such an element exists precisely when $4 \mid p-1$. A root a comes with the distinct root $-a$. Thus

$$c_p = \begin{cases} (p-2)/(p-1), & p \equiv 1 \pmod 4, \\ p/(p-1), & p \equiv 3 \pmod 4. \end{cases}$$

The revised guess for $n \le x$ with n^2+1 prime is $A\,\mathrm{Li}(x)/2$, where $A = \prod_p c_p$, taking primes in increasing order.

Why does this product converge?

For odd primes put $\chi(p) = 1$ if $p \equiv 1 \pmod 4$ and $\chi(p) = -1$ if $p \equiv 3 \pmod 4$. Then

$$c_p = 1 - \frac{\chi(p)}{p-1}, \qquad \log c_p = -\frac{\chi(p)}{p} + O(1/p^2).$$

The error terms have a convergent sum. So convergence of the product comes down to convergence of

$$\sum_{p \text{ odd}} \frac{\chi(p)}{p} = -\frac{1}{3} + \frac{1}{5} - \frac{1}{7} - \frac{1}{11} + \frac{1}{13} + \cdots .$$

This says that the primes are balanced between $1 \pmod 4$ and $3 \pmod 4$, from the point of view of reciprocal sums. Our proof of Dirichlet's theorem will give $\sum_{p \le x} \chi(p)(\log p)/p = O(1)$; partial summation then gives the required convergence.

In the last row, x bounds n; the primes can be as large as $x^2 + 1$.

	x	Guess	Actual count
$p \le x,\ p \equiv 3 \pmod{10}$	10^{12}	9 401 987 570	9 401 979 904
$p \le x,\ p + 2$ prime	10^{12}	1 870 559 867	1 870 585 220
$n \le x,\ n^2 + 1$ prime	10^{14}	2 199 893 807 666	2 199 894 223 892

The rounded guesses are $\mathrm{Li}(x)/4$, $2C_2 \int_2^x \mathrm{d}t/(\log t)^2$, and $A\,\mathrm{Li}(x)/2$. Here $C_2 = 0.6601618158\ldots$ is the twin-prime product and $A = 1.3728134628\ldots$.

Counts: OEIS A073506, Nicely's twin-prime table (MathWorld), and OEIS A206709.

Contents	Pages
The odd classes modulo 4	97–98
Schur's theorem on prime divisors of polynomial values	99
The four odd classes modulo 8	100–103
Primes congruent to 1 modulo m	104–106
Primes congruent to -1 modulo m	107–114
The criterion of Schur and Murty	115

Course overview

The odd classes modulo 4

Suppose $p_1, \ldots, p_r$ are all the primes congruent to $3 \pmod 4$. Put $N = 4p_1 \cdots p_r - 1$. Since $N \equiv 3 \pmod 4$, some prime divisor is $3 \pmod 4$ (a product of primes $1 \pmod 4$ would still be $1 \pmod 4$); but none of the p_i divides N. Contradiction.

Replacing 4 by 3 proves that infinitely many primes are $2 \pmod 3$.

For $1 \pmod 4$, use X^2+1 somehow! If an odd prime p divides n^2+1, then $n^2 \equiv -1 \not\equiv 1 \pmod p$ and $n^4 \equiv 1 \pmod p$. Thus n has order 4 in $\mathbb{F}_p^\times$, and $4 \mid p-1$.

It remains to show that infinitely many primes divide values of n^2+1. The next theorem gives much more.

Homework 4, Problem 1. Let H be a proper subgroup of $(\mathbb{Z}/m\mathbb{Z})^\times$. Adapt Euclid's proof to show that infinitely many primes $p \nmid m$ have residue class outside H.

Theorem

A nonconstant polynomial $f \in \mathbb{Z}[X]$ has a root modulo infinitely many primes.

If $f(0) = 0$, every prime works. Otherwise put $a = f(0)$ and $g(X) = f(aX)/a$. Why does g have integer coefficients? Write $f(X) = a + a_1X + \cdots + a_dX^d$. Then

$$g(X) = 1 + a_1X + a_2aX^2 + \cdots + a_da^{d-1}X^d.$$

Thus $g \in \mathbb{Z}[X]$, $g(0) = 1$, and $\deg g = \deg f > 0$.

Suppose only $p_1, \ldots, p_r$ divide nonzero values of f, and put $P = p_1 \cdots p_r$. Choose j with $|g(jP)| > 1$. Each prime divisor of $g(jP)$ divides $f(ajP)$ and hence is among the p_i. But $g(jP) \equiv 1 \pmod{p_i}$ for each i, a contradiction.

The class 1 modulo 8

If an odd prime p divides $n^4 + 1$, then $n^8 \equiv 1 \pmod{p}$ but $n^4 \not\equiv 1 \pmod{p}$. Thus n has order 8 in $\mathbb{F}_p^\times$, and $p \equiv 1 \pmod{8}$.

Schur's theorem applied to $X^4 + 1$ now gives infinitely many primes in this class.

The class 3 modulo 8

Suppose the primes in this class form a finite list. Choose an odd positive integer n divisible by every prime on the list.

Now $n^2 + 2 \equiv 3 \pmod 8$. Each prime divisor p satisfies $(\frac{-2}{p}) = 1$. The supplementary laws of quadratic reciprocity give $p \equiv 1$ or $3 \pmod 8$. They cannot all be $1 \pmod 8$.

Thus some prime divisor is $3 \pmod 8$. It cannot divide n, since it is odd and divides $n^2 + 2$. This is a new prime, a contradiction.

The class 5 modulo 8

Suppose the primes in this class form a finite list. By the Chinese remainder theorem, choose $n \equiv 2 \pmod 4$ divisible by every prime on the list.

Then $n^2 + 1 \equiv 5 \pmod 8$. Every prime divisor is $1 \pmod 4$, hence 1 or $5 \pmod 8$. They cannot all be $1 \pmod 8$, so some prime divisor is $5 \pmod 8$.

This prime does not divide n, contradicting the choice of n.

The class 7 modulo 8

Suppose the primes in this class form a finite list. Choose an odd $n \geq 3$ divisible by every prime on the list.

The positive integer $n^2 - 2$ is $7 \pmod 8$. Each prime divisor p satisfies $(\frac{2}{p}) = 1$. By the supplementary law for 2, $p \equiv 1$ or $7 \pmod 8$. Some divisor must be $7 \pmod 8$.

Again this odd prime cannot divide n, giving the contradiction.

Primes congruent to 1 modulo m

Let $\Phi_m(X)$ be the monic polynomial whose roots are the primitive mth roots of unity. Grouping the roots of $X^m - 1$ by their exact orders gives the first identity; multiplicative Möbius inversion gives the second:

$$X^m - 1 = \prod_{d \mid m} \Phi_d(X), \qquad \Phi_m(X) = \prod_{d \mid m} (X^d - 1)^{\mu(m/d)}.$$

Indeed the exponent of Φ_e on the right is $\sum_{e \mid d \mid m} \mu(m/d) = \mathbf{1}_{e=m}$. For example,

$$\Phi_3 = X^2 + X + 1, \quad \Phi_4 = X^2 + 1, \quad \Phi_8 = X^4 + 1.$$

Homework 4, Problem 3. Let G be an abelian group, written multiplicatively, and let $f, g : \mathbb{Z}_{>0} \to G$. Prove that if $g(n) = \prod_{d \mid n} f(d)$ for every n, then $f(n) = \prod_{d \mid n} g(d)^{\mu(n/d)}$.

Inductively, $\Phi_m \in \mathbb{Z}[X]$: divide $X^m - 1$ by the monic integer polynomial $\prod_{d \mid m,\ d<m} \Phi_d$. Long division has integer coefficients, and the remainder is zero over $\mathbb{C}$.

The following argument works over *any* field of characteristic $p \nmid m$. Suppose $\Phi_m(a) = 0$. Since $\Phi_m \mid X^m - 1$, we have $a^m = 1$, so $a \neq 0$ and its order divides m.

If the order were $d < m$, then $0 = a^d - 1 = \prod_{e \mid d} \Phi_e(a)$. Thus a is also a root of some Φ_e with $e \mid d$, $e < m$. Hence $(X - a)^2 \mid \Phi_e(X)\Phi_m(X) \mid X^m - 1$.

But $(X^m - 1)' = mX^{m-1}$ is nonzero at each root when $p \nmid m$. Therefore a has order exactly m.

Infinitely many primes congruent to 1 modulo m

Schur's theorem supplies infinitely many primes for which Φ_m has a root. Discard the finitely many primes dividing m.

For each remaining prime p, the root has order m in $\mathbb{F}_p^\times$. Lagrange's theorem gives $m \mid p-1$.

Thus there are infinitely many primes $p \equiv 1 \pmod{m}$.

Toward the class -1 modulo m

For the class $-1 \pmod m$, we look for a polynomial whose prime divisors lie in the two classes 1 and -1.

We shall then force a prime divisor in the second class. The idea is to pair each primitive root of unity with its inverse.

Let $m \geq 3$ and $h = \varphi(m)/2$. Pairing each primitive root ζ with ζ^{-1} shows that Φ_m has even degree and is *palindromic*: its coefficients read the same in reverse order. We can write

$$\Phi_m(X) = X^h \Psi_m(X + X^{-1})$$

for a monic $\Psi_m \in \mathbb{Z}[Y]$ of degree h. Indeed $X^{-h}\Phi_m(X)$ is an integer linear combination of 1 and $X^j + X^{-j}$. Put $S_j = X^j + X^{-j}$. These are polynomials in $Y = X + X^{-1}$, since

$$S_0 = 2, \quad S_1 = Y, \quad S_{j+1} = Y S_j - S_{j-1}.$$

For $m = 3, 4, 5, 8$ the polynomials Ψ_m are, respectively,

$$Y + 1, \quad Y, \quad Y^2 + Y - 1, \quad Y^2 - 2.$$

Homework 4, Problem 4–5. Prove the reciprocal symmetry of Φ_m and the existence and uniqueness of Ψ_m in the displayed identity.

A negative value of Ψ_m

Put $\zeta = e^{2\pi i a/m}$, where $(a, m) = 1$. Since $\Phi_m(\zeta) = 0$ and $\zeta \neq 0$,

$$0 = \zeta^h \Psi_m(\zeta + \zeta^{-1}), \qquad \zeta + \zeta^{-1} = 2\cos(2\pi a/m).$$

Pair a with $m - a$, and take $0 < a < m/2$. There are $\varphi(m)/2$ such a, and the cosines are distinct: cosine is strictly decreasing on $(0, \pi)$. This accounts for the full degree of Ψ_m.

In particular $r = 2\cos(2\pi/m)$ is a simple root. Write $\Psi_m(Y) = (Y - r)Q(Y)$ with $Q(r) \neq 0$. Near r, Q keeps its sign while $Y - r$ changes sign. Hence there is an open interval I on which $\Psi_m < 0$.

Suppose $p \nmid m$ and $\Psi_m(r) = 0$ in $\mathbb{F}_p$. We want a z with $z + z^{-1} = r$. Choose a root of $X^2 - rX + 1$ in an algebraic closure of $\mathbb{F}_p$.

The constant term is 1, so $z \neq 0$. Dividing $z^2 - rz + 1 = 0$ by z gives

$$z + z^{-1} = r.$$

Now substitute in the defining identity:

$$\Phi_m(z) = z^h \Psi_m(z + z^{-1}) = z^h \Psi_m(r) = 0.$$

Since $p \nmid m$, our argument for cyclotomic polynomials shows that z has order exactly m.

Raise $z + z^{-1} = r$ to the pth power. In characteristic p, $(a+b)^p = a^p + b^p$; and $r^p = r$ since $r \in \mathbb{F}_p$. Thus

$$z^p + (z^p)^{-1} = z + z^{-1}.$$

Consider more generally $u + u^{-1} = v + v^{-1}$, with $u, v \neq 0$. Rearranging,

$$u - v = \frac{1}{v} - \frac{1}{u} = \frac{u - v}{uv}.$$

So either $u = v$ or $uv = 1$. Taking $u = z^p$, $v = z$, we get

$$z^{p-1} = 1 \quad \text{or} \quad z^{p+1} = 1.$$

Since z has order m, $m \mid p - 1$ or $m \mid p + 1$. Hence $p \equiv \pm 1 \pmod{m}$.

Forcing a new prime in the class -1

Can we just apply Schur? Not yet: it could give only primes $1 \pmod m$. We must force the minus sign to occur.

Suppose $p_1, \dots, p_r$ are all the primes $-1 \pmod m$, and put $D = mp_1 \cdots p_r$. Choose

$$v = D^\ell, \qquad u = 1 + Dj, \qquad u/v \in I.$$

For fixed ℓ, these fractions are spaced $D^{1-\ell}$ apart. Choose ℓ so that this spacing is less than the length of I, then choose j. Notice that $(u, v) = 1$.

Write $\Psi_m(Y) = Y^h + a_{h-1}Y^{h-1} + \cdots + a_0$. Then

$$v^h \Psi_m(u/v) = u^h + a_{h-1}u^{h-1}v + \cdots + a_0 v^h.$$

Every term is an integer, and all but the first are divisible by v.

The integer

$$N = v^h \Psi_m(u/v)$$

is negative. Since Ψ_m is monic, $N \equiv u^h \pmod{v}$. Consequently $(N, v) = 1$ and $N \equiv 1 \pmod{m}$.

Finishing the class -1

If $q \mid N$, then $q \nmid v$ and $\Psi_m(u/v) = 0$ modulo q. Hence $q \equiv \pm 1 \pmod{m}$.

All the supposed primes $-1 \pmod{m}$ divide v, so every prime divisor of N is $1 \pmod{m}$. But $N < 0$. Writing $N = -q_1 \cdots q_j$ with prime factors repeated according to multiplicity gives $N \equiv -1 \pmod{m}$.

This contradicts $N \equiv 1 \pmod{m}$ when $m \geq 3$. For $m = 1, 2$, the classes 1 and -1 coincide, and the result is already known.

How far can this method go?

A theorem of Schur and Murty says that a polynomial whose prime divisors, with finitely many exceptions, lie in the classes 1 and $a \pmod m$, with infinitely many in the latter class, exists precisely when

$$a^2 \equiv 1 \pmod m.$$

When $a^2 \equiv 1 \pmod m$, $H = \{1, a\}$ is a subgroup of $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$. Write its fixed field as $\mathbb{Q}(\alpha)$ and take the minimal polynomial of α, scaled to have integer coefficients. This gives the polynomial in Schur's direction. Murty proved the converse.

Even the class $2 \pmod 5$ lies beyond this approach. For this we need Dirichlet's theorem.

Dirichlet's theorem (1 of 2)

Contents	Pages
Statement and plan of proof	118–119
The modulus 4	120–122
Orthogonality relations for characters	123–127
Convergence of $L(s, \chi)$ for nonprincipal χ	128–129
Reduction to nonvanishing	130

Course overview

Dirichlet's theorem (2 of 2)

Contents	Pages
Nonvanishing for nonreal characters	131–133
Nonvanishing for real nonprincipal characters	134–137
Conclusion and a counting consequence	138–139
Siegel zeros and Linnik's theorem	140–142

Course overview

Dirichlet's theorem: statement and plan

Theorem

Fix $m \geq 1$ and $(a, m) = 1$. For $x \geq 3$,

$$\sum_{\substack{p \leq x \\ p \equiv a \bmod m}} \frac{\log p}{p} = \frac{\log x}{\varphi(m)} + O_m(1).$$

Thus the primes equidistribute themselves in the coprime residue classes, if we count each prime with weight $(\log p)/p$. In particular each such class contains infinitely many primes.

It suffices to use prime powers p^k in place of primes: the extra contribution is at most $\sum_p (\log p)/(p(p-1)) = O(1)$.

Dirichlet's theorem: plan of proof

For $m = 4$ we detect a congruence with a linear combination of two completely multiplicative functions. In general we use Dirichlet characters modulo m.

Mertens' theorem handles the principal character. For each nonprincipal character χ, we shall show

$$L(1, \chi) \sum_{p^k \leq x} \frac{\chi(p^k) \log p}{p^k} = O_m(1).$$

It suffices to prove $L(1, \chi) \neq 0$. First we show that at most one nonprincipal character can have $L(1, \chi) = 0$. Nonreal characters occur in conjugate pairs, so this leaves the real characters to consider.

Dirichlet's theorem for the modulus 4

Let χ_0 be 1 on odd integers and 0 on even integers. Let χ be 1 on $1 \pmod 4$, -1 on $3 \pmod 4$, and 0 on even integers. Both are completely multiplicative, and

$$\mathbf{1}_{n\equiv 1 \bmod 4} = \tfrac{1}{2}(\chi_0(n) + \chi(n)), \qquad \mathbf{1}_{n\equiv 3 \bmod 4} = \tfrac{1}{2}(\chi_0(n) - \chi(n)).$$

Mertens' theorem gives

$$\sum_{p^k \le x} \chi_0(p^k) \frac{\log p}{p^k} = \log x + O(1),$$

since the powers of 2 contribute at most $\log 2$. Substituting gives $\frac{1}{2}\log x + O(1)$, which is what we wanted to end up with! So we should be trying to show that the remaining character sum is $O(1)$.

A positive value of $L(1, \chi)$

The alternating series test gives

$$L = \sum_{n \geq 1} \frac{\chi(n)}{n} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots > 1 - \frac{1}{3} > 0,$$

$$\sum_{n \leq t} \frac{\chi(n)}{n} = L + O(1/t).$$

The series $\sum_n \chi(n)(\log n)/n$ also converges: $(\log t)/t$ eventually decreases to zero.

On the other hand, prime factorization gives $\log n = \sum_{p^k | n} \log p$. Substitute this into the partial sum.

Write $n = p^k j$ and use complete multiplicativity:

$$\sum_{n \le x} \frac{\chi(n) \log n}{n} = \sum_{p^k \le x} \frac{\chi(p^k) \log p}{p^k} \sum_{j \le x/p^k} \frac{\chi(j)}{j}$$

$$= L \sum_{p^k \le x} \frac{\chi(p^k) \log p}{p^k} + O\left(\frac{1}{x} \sum_{p^k \le x} \log p \right).$$

The error is $O(\psi(x)/x) = O(1)$ by Chebyshev's estimate. The left side is bounded, and $L > 0$, so the character sum is $O(1)$.

Substitution gives $\frac{1}{2} \log x + O(1)$ for each odd class.

Characters of a finite abelian group

How can we write $\mathbf{1}_{n \equiv a \bmod m}$ as a linear combination of completely multiplicative functions? We first work with a finite abelian group G.

A character of G is a homomorphism to the complex numbers of absolute value 1. The characters form a group $\widehat{G}$ under pointwise multiplication; the identity is the constant character 1, and $\chi^{-1} = \overline{\chi}$.

Easy case: $G = \langle g \rangle$ has order h. A character is determined by $\chi(g)$, and $\chi(g)^h = \chi(g^h) = 1$. Conversely, for any hth root ω, put $\chi(g^j) = \omega^j$. This is well defined because changing j by a multiple of h does not change ω^j.

General case: write $G = \langle g_1 \rangle \times \cdots \times \langle g_r \rangle$, with g_i of order h_i.

A character is determined by $\chi(g_1), \ldots, \chi(g_r)$. Each $\chi(g_i)$ must be an h_ith root of unity, so there are at most $h_1 \cdots h_r$ characters.

Conversely, choose any such roots $\omega_1, \ldots, \omega_r$, independently, and put

$$\chi(g_1^{e_1} \cdots g_r^{e_r}) = \omega_1^{e_1} \cdots \omega_r^{e_r}.$$

The direct product decomposition makes this well defined. It is a character, and every character arises this way. Hence

$$|\widehat{G}| = h_1 \cdots h_r = |G|.$$

For nontrivial χ, choose $h \in G$ with $\chi(h) \neq 1$. Multiplication by h permutes G, so

$$\chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G} \chi(hg) = \sum_{g \in G} \chi(g).$$

Thus $\sum_g \chi(g) = 0$.

If $g \neq 1$, some character η has $\eta(g) \neq 1$: use a nonidentity cyclic coordinate, send its generator to a primitive root, and send the other generators to 1.

Multiplication by η permutes $\widehat{G}$, so the same argument gives $\sum_{\chi \in \widehat{G}} \chi(g) = 0$.

Averaging over group elements tells characters apart; averaging over characters tells group elements apart. If $\chi = \eta$, every term in the first average is 1. Otherwise $\chi\overline{\eta}$ is nontrivial, so its sum is 0. For the second average use $\chi(g)\overline{\chi(h)} = \chi(gh^{-1})$:

$$\frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\eta(g)} = \mathbf{1}_{\chi=\eta}, \qquad \frac{1}{|G|}\sum_{\chi\in\widehat{G}}\chi(g)\overline{\chi(h)} = \mathbf{1}_{g=h}.$$

The units modulo 8 are generated by 3 and 5, each of order 2. Their signs can be chosen independently:

	1	3	5	7
χ_0	1	1	1	1
χ_1	1	-1	1	-1
χ_2	1	1	-1	-1
χ_3	1	-1	-1	1

Homework 4, Problem 6. For a finite abelian group G, let $T = (\chi(g))_{\chi\in\widehat{G},\, g\in G}$. Compute TT^*, where T^* is the conjugate transpose.

Take $G = (\mathbb{Z}/m\mathbb{Z})^\times$. Extend a group character to all integers by periodicity on the coprime classes and by zero elsewhere. These are the $\varphi(m)$ Dirichlet characters modulo m.

Each is periodic, completely multiplicative, and vanishes precisely when $(n, m) > 1$. The principal character χ_0 is 1 on the coprime classes. For $(a, m) = 1$,

$$\mathbf{1}_{n \equiv a \bmod m} = \frac{1}{\varphi(m)} \sum_{\chi \bmod m} \overline{\chi(a)} \chi(n).$$

This is group orthogonality if $(n, m) = 1$; otherwise both sides vanish.

Homework 4, Problem 7. Show that the three properties above characterize Dirichlet characters.

We use that the values of χ average to zero. Suppose $\chi \neq \chi_0$, and put $S(t) = \sum_{n \le t} \chi(n)$. Each complete block of m consecutive integers sums to 0; only an incomplete block remains. Thus $|S(t)| \le \varphi(m)$. Partial summation gives

$$\sum_{n \le x} \frac{\chi(n)}{n^s} = S(x)x^{-s} + s\int_1^x S(t)t^{-s-1}\,\mathrm{d}t.$$

For real $s > 0$, the first term tends to zero and the integral converges absolutely. Thus $L(s,\chi) = \sum_{n \ge 1} \chi(n)n^{-s}$ converges.

For χ_0 the series converges for $s > 1$ and diverges at $s = 1$, by comparison on the progression $1 \pmod m$.

Homework 4, Problem 8(a). Derive the displayed formula from partial summation.

Subtract the finite sum from its limit and use $s\int_x^\infty t^{-s-1}\,\mathrm{d}t = x^{-s}$:

$$\sum_{n>x}\frac{\chi(n)}{n^s} = s\int_x^\infty \big(S(t)-S(x)\big)t^{-s-1}\,\mathrm{d}t.$$

The sum on $(x,t]$ consists of full periods and one incomplete period, so $|S(t)-S(x)|\le\varphi(m)$. The tail has absolute value at most $\varphi(m)x^{-s}$.

In particular, $\sum_{n\le x}\chi(n)/n = L(1,\chi)+O_m(1/x)$. Also $\sum_n \chi(n)(\log n)/n$ converges by partial summation, since $(\log t)/t$ decreases to zero for $t>e$.

Homework 4, Problem 8(b,c). For nonprincipal $\chi \bmod m$ and real $s>0$, prove the tail bound above for $x\ge 1$, and prove convergence of $\sum_{n\ge1}\chi(n)(\log n)/n$.

Orthogonality gives

$$\sum_{\substack{p^k \le x \\ p^k \equiv a \bmod m}} \frac{\log p}{p^k} = \frac{1}{\varphi(m)} \sum_{\chi \bmod m} \overline{\chi(a)} \sum_{p^k \le x} \frac{\chi(p^k) \log p}{p^k}.$$

For χ_0, the inner sum is $\log x + O_m(1)$: primes dividing m contribute at most $\sum_{p|m} \log p/(p-1)$.

For nonprincipal χ, the calculation for modulus 4 gives

$$\sum_{n \le x} \frac{\chi(n) \log n}{n} = L(1,\chi) \sum_{p^k \le x} \frac{\chi(p^k) \log p}{p^k} + O_m(\psi(x)/x).$$

Both the left side and $\psi(x)/x$ are bounded. New goal: show that $L(1,\chi) \neq 0$. Then we can divide through by it, just as for $m = 4$.

Write $\Lambda(n) = \log p$ if $n = p^k$, $k \geq 1$, and 0 otherwise. Since $\log n = \sum_{d|n} \Lambda(d)$, Möbius inversion gives

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) = -\sum_{d|n} \mu(d) \log d.$$

The term $\log n \sum_{d|n} \mu(d)$ vanishes even at $n = 1$. Consequently,

$$\sum_{d|n} \mu(d) \log(x/d) = \Lambda(n) + \mathbf{1}_{n=1} \log x.$$

On multiplying by $\chi(n)/n$ and summing, the $n = 1$ term contributes precisely $\log x$.

Multiply by $\chi(n)/n$, sum over $n \le x$, and put $n = de$:

$$\log x + \sum_{n \le x} \frac{\chi(n)\Lambda(n)}{n} = \sum_{d \le x} \frac{\mu(d)\chi(d)}{d} \log(x/d) \sum_{e \le x/d} \frac{\chi(e)}{e}.$$

If $L(1, \chi) = 0$, the inner sum is $O_m(d/x)$. The right side is then

$$O_m\left(\frac{1}{x} \sum_{d \le x} \log(x/d)\right) = O_m(1),$$

since the sum is $\lfloor x \rfloor \log x - \log \lfloor x \rfloor ! = O(x)$. Thus a zero implies $\sum_{n \le x} \chi(n)\Lambda(n)/n = -\log x + O_m(1)$.

There is at most one exceptional character

We compute the sum over all characters in two different ways. Let N count the nonprincipal χ with $L(1,\chi)=0$. First add the prime-power sums character by character. The principal character contributes $\log x + O_m(1)$, the N zeros each contribute $-\log x + O_m(1)$, and all others contribute $O_m(1)$.

The total is $(1-N)\log x + O_m(1)$. In the other order, orthogonality gives

$$\varphi(m) \sum_{\substack{p^k \le x \\ p^k \equiv 1 \bmod m}} \frac{\log p}{p^k} \ge 0.$$

Hence $N \le 1$. But a nonreal zero would give two distinct characters $\chi, \overline{\chi}$ with zero value, since $L(1,\overline{\chi}) = \overline{L(1,\chi)}$. So nonreal characters do not vanish.

Nonvanishing for real nonprincipal characters

Put $r_\chi(n) = \sum_{d|n} \chi(d)$. We will show that a weighted average of $r_\chi(n)$ tends to infinity, and that this forces $L(1,\chi) \neq 0$.

Notice that r_χ is multiplicative: for $(a,b) = 1$, every divisor of ab is uniquely uv with $u \mid a$, $v \mid b$, and $\chi(uv) = \chi(u)\chi(v)$. On prime powers,

$$r_\chi(p^k) = \sum_{j=0}^{k} \chi(p)^j = \begin{cases} k+1, & \chi(p) = 1, \\ 1, & \chi(p) = 0, \\ 1 \text{ or } 0, & \chi(p) = -1, \ k \text{ even or odd.} \end{cases}$$

Thus $r_\chi(n) \geq 0$ and $r_\chi(j^2) \geq 1$. In particular,

$$\sum_{n \leq x} \frac{r_\chi(n)}{\sqrt{n}} \geq \sum_{j \leq \sqrt{x}} \frac{r_\chi(j^2)}{j} \geq \sum_{j \leq \sqrt{x}} \frac{1}{j} \longrightarrow \infty.$$

Now estimate the same sum another way.

Splitting the sum at the square root

Put $y = \sqrt{x}$ and split the pairs $de \le x$ according as $d \le y$ or $d > y$:

$$\sum_{n \le x} \frac{r_\chi(n)}{\sqrt{n}} = \sum_{d \le y} \frac{\chi(d)}{\sqrt{d}} \sum_{e \le x/d} \frac{1}{\sqrt{e}} + \sum_{e \le y} \frac{1}{\sqrt{e}} \sum_{y < d \le x/e} \frac{\chi(d)}{\sqrt{d}}.$$

Why is $e \le y$ in the second region? Since $d > y$ and $de \le x = y^2$, we have $e < y$. There is no overlap, since the first region has $d \le y$.

The second inner sum is a difference of two tails of $L(1/2, \chi)$, each $O_m(y^{-1/2})$.

Thus the second double sum is $O_m(y^{-1/2} \sum_{e \le y} e^{-1/2}) = O_m(1)$ by the integral test.

Keeping the constant term

For $u \geq 1$ we need

$$\sum_{e \leq u} \frac{1}{\sqrt{e}} = 2\sqrt{u} + C + O(u^{-1/2}).$$

An $O(1)$ error here would give only $O(\sqrt{y})$ after summing absolute values, so the constant matters.

Subtract $\int_e^{e+1} t^{-1/2}\,\mathrm{d}t$ from $e^{-1/2}$. The difference lies between 0 and $e^{-1/2} - (e+1)^{-1/2}$, so its sum converges with tail $O(u^{-1/2})$.

Finally $\int_1^{\lfloor u \rfloor + 1} t^{-1/2}\,\mathrm{d}t = 2\sqrt{u} - 2 + O(u^{-1/2})$, proving the formula.

The first double sum is

$$2\sqrt{x}\sum_{d\le y}\frac{\chi(d)}{d}+C\sum_{d\le y}\frac{\chi(d)}{\sqrt{d}}+O\left(\frac{1}{\sqrt{x}}\sum_{d\le y}1\right).$$

The first term is $2\sqrt{x}L(1,\chi)+O_m(\sqrt{x}/y)$; the middle sum is bounded by convergence of $L(1/2,\chi)$. Since $y=\sqrt{x}$,

$$\sum_{n\le x}\frac{r_\chi(n)}{\sqrt{n}}=2\sqrt{x}L(1,\chi)+O_m(1).$$

If $L(1,\chi)=0$, this is bounded, contradicting the lower bound $\sum_{j\le\sqrt{x}}1/j\to\infty$.

Conclusion of Dirichlet's theorem

We can now finish the proof. For every nonprincipal χ, we have $L(1,\chi) \neq 0$, and hence

$$\sum_{p^k \leq x} \frac{\chi(p^k)\log p}{p^k} = O_m(1).$$

Substitute into orthogonality and remove the terms with $k \geq 2$:

$$\boxed{\sum_{\substack{p \leq x \\ p \equiv a \bmod m}} \frac{\log p}{p} = \frac{\log x}{\varphi(m)} + O_m(1).}$$

In particular every coprime class contains infinitely many primes.

Subtract the estimate at x/K from the estimate at x. The difference is $\log K/\varphi(m) + O_m(1)$, at least 1 if we fix K large enough.

Each prime counted contributes at most $K \log x/x$. Thus at least $x/(K \log x)$ primes in the class lie in $(x/K, x]$, for all large x.

Homework 3, Problem 8. Let A be a set of positive integers satisfying

$$\sum_{\substack{a \in A \\ a \le x}} \frac{\log a}{a} = c \log x + O(1), \qquad c > 0.$$

Prove that $\#\{a \in A : a \le x\}$ lies between two positive constant multiples of $x/\log x$ for all sufficiently large x.

Let χ be a real nonprincipal character modulo m. If $L(1,\chi)=0$, the argument above would give

$$\sum_{p\le x}\chi(p)\frac{\log p}{p} = -\log x + O_m(1), \quad \text{hence} \quad \sum_{\substack{p\le x\\ \chi(p)=-1}}\frac{\log p}{p} = \log x + O_m(1).$$

We ruled this out. In fact, our proof of nonvanishing gives

$$\sum_{\substack{p\le x\\ \chi(p)=-1}}\frac{\log p}{p} = \frac{1}{2}\log x + O_m(1).$$

For each fixed m, this sum is therefore at most $\frac{3}{4}\log x$ once x is sufficiently large. But how large must x be, in terms of m?

OpenAI (October 2026). There is an absolute constant $A > 0$ such that, for every $m \geq 2$ and every real nonprincipal character χ modulo m,

$$\sum_{\substack{p \leq x \\ \chi(p) = -1}} \frac{\log p}{p} \leq \frac{3}{4} \log x \qquad (x \geq m^A).$$

The point is that the *same* A works for all m and χ.

This resolved the longstanding problem of *Siegel zeros*: hypothetical real zeros of $L(s, \chi)$ extremely close to 1. The bound implies that, for some absolute $c > 0$,

$$L(s, \chi) \neq 0 \qquad \left(1 - \frac{c}{\log m} < s \leq 1, \quad s \text{ real}\right).$$

OpenAI, *Uniform exclusion of Landau–Siegel zeros* (2026).

Linnik's theorem

Another question in which the dependence on m matters is how far we must go to find the first prime in a given coprime residue class.

Linnik's theorem. There is an absolute constant L such that, for every $m \geq 2$ and every a coprime to m, there is a prime

$$p \equiv a \pmod{m}, \qquad p \leq m^L.$$

The bound on the preceding slide combines with Selberg's symmetry formulas to give an elementary proof of this theorem. See

The missing Lin(ni)k (Hexagon, 2026).

The prime number theorem (1 of 2)

Contents	Pages
Statement and plan of proof	145
Squarefree integers and Möbius cancellation	146–150
The implication from $M(x) = o(x)$ to the PNT	151
A logarithmic average of $M(t)/t$	152–153
The Fundamental Lemma	154
Deduction of cancellation from the lemma	155–159
The integral comparison	160–169
Excess and deficit	170–173

Course overview

The prime number theorem (2 of 2)

Contents	Pages
The excess below y	174–175
The inequality for E_y	176–182
The integrated inequality	183–188
The excess above y	189–190
Conclusion: $M(x) = o(x)$	191
The hyperbola method and $\sum_{n \leq x} \tau(n)$	192–194
Deduction of $\psi(x) \sim x$	195–198
BONUS: $\sum_{n \geq 1} \mu(n)/n = 0$	199–202

Course overview

The prime number theorem: statement and plan

We will prove the Prime Number Theorem in the form

$$\psi(x) = \sum_{n \le x} \Lambda(n) \sim x.$$

Our first goal is cancellation in the Möbius sum:

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

We give a proof using smooth numbers, following Daboussi. First we prove $M(x) = o(x)$; then we deduce $\psi(x) \sim x$.

The number of squarefree integers up to x

Recall that $\mu(1) = 1$ and

$$\mu(n) = \begin{cases} 0, & p^2 \mid n \text{ for some prime } p, \\ (-1)^k, & n = p_1 \cdots p_k \text{ with distinct primes } p_i. \end{cases}$$

There are two natural sums:

$$A(x) = \sum_{n \le x} |\mu(n)|, \qquad M(x) = \sum_{n \le x} \mu(n).$$

The first counts squarefree integers. The second counts with a sign: $+1$ for an even number of prime factors and -1 for an odd number.

Theorem

For $x \geq 1$, $A(x) = \dfrac{6}{\pi^2}x + O(\sqrt{x})$.

We use the identity

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

The sum expands $\prod_{p^2|n}(1-1)$. If n is squarefree this is an empty product, equal to 1. Otherwise some factor is zero.

Reversing the squarefree sum

Summing and reversing order,

$$A(x) = \sum_{n \le x} \sum_{d^2 \mid n} \mu(d) = \sum_{d \le \sqrt{x}} \mu(d) \left\lfloor \frac{x}{d^2} \right\rfloor$$
$$= x \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} + O(\sqrt{x}).$$

Extend the sum to infinity, since

$$\left| \sum_{d > \sqrt{x}} \frac{\mu(d)}{d^2} \right| \le \sum_{d > \sqrt{x}} \frac{1}{d^2} = O(x^{-1/2}).$$

By absolute convergence,

$$\left(\sum_{d\geq 1}\frac{\mu(d)}{d^2}\right)\left(\sum_{e\geq 1}\frac{1}{e^2}\right)=\sum_{n\geq 1}\frac{1}{n^2}\sum_{d\mid n}\mu(d)=1.$$

Since $\sum_e e^{-2}=\pi^2/6$, the leading constant is $6/\pi^2$. We have proved

$$A(x)=\frac{6}{\pi^2}x+O(\sqrt{x}).$$

Cancellation in the Möbius sum

What about the signed sum? We want to prove

Theorem (von Mangoldt, 1897)

$$M(x) = o(x).$$

In other words, squarefree integers split equally between the signs: each of $\mu(n) = 1$ and $\mu(n) = -1$ occurs $3x/\pi^2 + o(x)$ times up to x.

If $M(x) = o(x)$, asymptotically half of the squarefree integers have each sign. We will use this to prove the PNT.

Theorem

$$M(x) = o(x) \quad \Longrightarrow \quad \psi(x) \sim x.$$

We prove this implication after establishing $M(x) = o(x)$.

A bounded Möbius sum with weight $1/n$

The Möbius function gives a number-theoretically friendly answer to the question of "who is number 1":

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

Let $x \geq 1$ and sum over $n \leq x$:

$$1 = \sum_{d \leq x} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = x \sum_{d \leq x} \frac{\mu(d)}{d} - \sum_{d \leq x} \mu(d) \left\{ \frac{x}{d} \right\}.$$

Since $|\mu(d)| \leq 1$ and $0 \leq \{x/d\} < 1$,

$$\left| \sum_{d \leq x} \frac{\mu(d)}{d} \right| < \frac{1 + \lfloor x \rfloor}{x} \leq 2 \qquad (x \geq 1).$$

A logarithmic average of $M(t)/t$

Expressing the same sum by partial summation, we have

$$\sum_{n \le x} \frac{\mu(n)}{n} = \frac{M(x)}{x} + \int_1^x \frac{M(t)}{t} \frac{\mathrm{d}t}{t}.$$

Since $|M(x)| \le x$, we obtain

$$\left| \int_1^x \frac{M(t)}{t} \frac{\mathrm{d}t}{t} \right| < 3.$$

Thus $M(t)/t$ tends to zero on logarithmic average:

$$\frac{1}{\log x} \int_1^x \frac{M(t)}{t} \frac{\mathrm{d}t}{t} \longrightarrow 0.$$

We want $M(x)/x$ itself to tend to zero.

Put $a = \limsup_{t\to\infty} |M(t)|/t$, so $0 \le a \le 1$. Write $v^+ = \max\{v, 0\}$.

Fundamental Lemma

As $x \to \infty$,

$$\int_1^x \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} = o(\log x).$$

Roughly speaking, $|M(t)|/t$ spends most of its time close to its limit superior a, with time measured logarithmically (with respect to $\mathrm{d}t/t$).

First we deduce $a = 0$ from this lemma; then we prove it.

Suppose $a > 0$. In every interval $[X, BX]$, we will find a u with $|M(u)|/u \leq a/2$. Otherwise M would keep one sign and its logarithmic integral would be too large.

Then $|M(t)|/t \leq 3a/4$ throughout a short interval starting at u. Each such interval contributes a fixed positive amount to

$$\int_1^x \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t},$$

contradicting the Fundamental Lemma.

Suppose for a contradiction that $a > 0$, and fix $B = \exp(12/a)$. We first show that every interval $[X, BX]$, with $X \geq 1$, contains a u for which $|M(u)| \leq (a/2)u$.

Indeed, if there is no such u, then $M(t)$ is nonzero throughout $[X, BX]$. Since M takes integer values and jumps by at most 1, it has constant sign there. This gives

$$\left| \int_X^{BX} \frac{M(t)}{t} \frac{\mathrm{d}t}{t} \right| = \int_X^{BX} \frac{|M(t)|}{t} \frac{\mathrm{d}t}{t} \geq \frac{a}{2} \log B = 6.$$

But the two integrals from 1 to BX and 1 to X each have absolute value less than 3. Their difference is less than 6, a contradiction.

A whole interval of small values

Now let $X \geq X_0 := 8/a$, and choose $u \in [X, BX]$ with $|M(u)| \leq (a/2)u$. To obtain a lower bound for the integral in the Fundamental Lemma, we extend this estimate to nearby values of t.

For $u \leq t \leq u(1 + a/8)$,

$$\begin{aligned} |M(t)| &\leq |M(u)| + |M(t) - M(u)| \\ &\leq \frac{a}{2}u + (t - u) + 1 \\ &\leq \frac{a}{2}u + \frac{a}{8}u + 1 \leq \frac{3a}{4}t. \end{aligned}$$

We used $t \geq u \geq 8/a$. Since $a \leq 1$, this whole interval lies in $[X, 2BX]$.

A fixed positive contribution

For every $X \geq X_0$,

$$\int_X^{2BX} \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} \geq \int_u^{u(1+a/8)} \frac{a}{4} \frac{\mathrm{d}t}{t}$$
$$= \frac{a}{4} \log(1 + a/8) > 0.$$

This lower bound is independent of X: $\int_u^{u(1+a/8)} \mathrm{d}t/t = \log(1 + a/8)$ for every u.

Use the last estimate at $X = X_0, (2B)X_0, \dots, (2B)^{N-1}X_0$. The intervals have disjoint interiors, so

$$\int_1^{(2B)^N X_0} \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} \geq N\frac{a}{4}\log(1 + a/8).$$

The Fundamental Lemma makes the left side $o(N)$, since

$$\log((2B)^N X_0) = N\log(2B) + \log X_0.$$

This contradiction forces $a = 0$.

It remains to show

$$\int_1^y \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} = o(\log y).$$

We compare $|M(t,y)|/t$ with $a\Psi(t,y)/t$, where both sums are restricted to y-smooth integers.

For $t \le y$, these are $|M(t)|/t$ and $a\lfloor t \rfloor/t$, nearly the quantities in the lemma.

Let $P^+(n)$ be the largest prime factor of n, with $P^+(1) = 1$. An integer is y-smooth if $P^+(n) \le y$. Put

$$\Psi(x,y) = \sum_{\substack{n \le x \\ P^+(n) \le y}} 1, \qquad M(x,y) = \sum_{\substack{n \le x \\ P^+(n) \le y}} \mu(n).$$

Throughout take $y \ge 3$ and $P = \prod_{p \le y} p$.

For $t \le y$, $M(t,y) = M(t)$. For fixed y and large enough t, the smooth Möbius sum vanishes.

First, $|M(t,y)| \leq \Psi(t,y)$ for all $t \geq 1$.

For $1 \leq t \leq y$, every integer counted is smooth, so

$$M(t,y) = M(t), \qquad \Psi(t,y) = \lfloor t \rfloor.$$

For $t \geq P$, every squarefree smooth integer is a divisor of P, and all divisors have appeared. Hence

$$M(t,y) = \sum_{d|P} \mu(d) = 0.$$

The integral comparison we need

Following Daboussi, we compare $|M(x,y)|$ with $a\Psi(x,y)$. The following estimate gets us started: for every $y \geq 3$,

$$\int_1^\infty \left(\frac{|M(t,y)|}{t} - a\frac{\Psi(t,y)}{t} \right) \frac{\mathrm{d}t}{t} \geq 0.$$

Below y the integrand is $|M(t)|/t - a\lfloor t \rfloor/t$, differing from $|M(t)|/t - a$ by at most a/t.

We first establish the inequality with y fixed, then let y tend to infinity.

Each positive integer admits a unique decomposition in the form

$$n = bc, \qquad P^+(b) \le y, \qquad (c, P) = 1.$$

Put the full prime powers for $p \le y$ into b, and the remaining prime powers into c. Since $(b, c) = 1$, $\mu(bc) = \mu(b)\mu(c)$. Thus

$$M(x) = \sum_{\substack{c \le x \\ (c,P)=1}} \mu(c) \sum_{\substack{b \le x/c \\ P^+(b) \le y}} \mu(b) = \sum_{\substack{c \le x \\ (c,P)=1}} \mu(c) M(x/c, y).$$

Since $M(t, y) = 0$ for $t \geq P$,

$$|M(x)| \leq \sum_{\substack{x/P < c \leq x \\ (c,P)=1}} |M(x/c, y)|.$$

Keep y fixed and group by $j = \lfloor x/c \rfloor$, where $1 \leq j < P$. Then $x/(j+1) < c \leq x/j$ and $M(x/c, y) = M(j, y)$, giving

$$|M(x)| \leq \sum_{j=1}^{P-1} |M(j, y)| \sum_{\substack{x/(j+1) < c \leq x/j \\ (c,P)=1}} 1.$$

Counting integers coprime to P

View y as fixed. Every block of P consecutive integers contains precisely $\varphi(P)$ integers coprime to P, so

$$\sum_{\substack{x/(j+1)<c\le x/j\\(c,P)=1}} 1 = x\frac{\varphi(P)}{P}\left(\frac{1}{j}-\frac{1}{j+1}\right)+O_y(1).$$

There are only $P-1$ terms, so the total error is still $O_y(1)$:

$$|M(x)| \le x\frac{\varphi(P)}{P}\sum_{j=1}^{P-1}|M(j,y)|\left(\frac{1}{j}-\frac{1}{j+1}\right)+O_y(1).$$

Divide by x and take the limsup:

$$a \le \frac{\varphi(P)}{P}\sum_{j=1}^{P-1}|M(j,y)|\left(\frac{1}{j}-\frac{1}{j+1}\right).$$

Since $M(t,y) = M(j,y)$ on $[j, j+1)$,

$$\begin{aligned} a &\le \frac{\varphi(P)}{P} \sum_{j=1}^{P-1} |M(j,y)| \int_j^{j+1} \frac{\mathrm{d}t}{t^2} \\ &= \frac{\varphi(P)}{P} \int_1^{\infty} \frac{|M(t,y)|}{t} \frac{\mathrm{d}t}{t}. \end{aligned}$$

Consequently

$$\int_1^{\infty} \frac{|M(t,y)|}{t} \frac{\mathrm{d}t}{t} \ge a \frac{P}{\varphi(P)} = a \prod_{p \le y} (1 - 1/p)^{-1}.$$

The corresponding integral for Ψ

The product just obtained also occurs in the integral of $\Psi(t,y)/t$. To see this, first observe that, for fixed y,

$$\Psi(x,y) \le \left(1+\frac{\log x}{\log 2}\right)^{\pi(y)} = o(x),$$

since each prime exponent is at most $\log x/\log 2$. Partial summation gives

$$\sum_{\substack{n\le x\\ P^+(n)\le y}} \frac{1}{n} = \frac{\Psi(x,y)}{x} + \int_1^x \frac{\Psi(t,y)}{t}\frac{\mathrm{d}t}{t}.$$

Let $x\to\infty$:

$$\int_1^\infty \frac{\Psi(t,y)}{t}\frac{\mathrm{d}t}{t} = \sum_{P^+(n)\le y}\frac{1}{n} = \prod_{p\le y}(1-1/p)^{-1}.$$

The integral comparison is proved

We have shown

$$\int_1^\infty \frac{|M(t,y)|}{t}\frac{\mathrm{d}t}{t} \geq a\prod_{p\leq y}(1-1/p)^{-1} = a\int_1^\infty \frac{\Psi(t,y)}{t}\frac{\mathrm{d}t}{t}.$$

Equivalently,

$$\int_1^\infty \left(\frac{|M(t,y)|}{t} - a\frac{\Psi(t,y)}{t}\right)\frac{\mathrm{d}t}{t} \geq 0.$$

We now consider $y \to \infty$.

To apply the integral comparison to the Fundamental Lemma, we separate its integrand into positive and negative parts. Define excess and deficit functions by

$$E_y(t) = \left(\frac{|M(t,y)|}{t} - a\frac{\Psi(t,y)}{t}\right)^+, \quad D_y(t) = \left(a\frac{\Psi(t,y)}{t} - \frac{|M(t,y)|}{t}\right)^+.$$

Their difference is $|M(t,y)|/t - a\Psi(t,y)/t$. The total excess is at least the total deficit, both measured with respect to $\mathrm{d}t/t$:

$$\int_1^\infty D_y(t)\frac{\mathrm{d}t}{t} \le \int_1^\infty E_y(t)\frac{\mathrm{d}t}{t}.$$

Both integrals converge: $E_y(t) \le \Psi(t,y)/t$ and $D_y(t) \le a\Psi(t,y)/t$.

When $1 \le t \le y$, the deficit $D_y(t)$ is closely related to the integrand in the Fundamental Lemma:

$$\begin{aligned} D_y(t) &= \left(a\frac{\lfloor t \rfloor}{t} - \frac{|M(t)|}{t}\right)^+ \\ &= \left(a - \frac{|M(t)|}{t} - a\frac{\{t\}}{t}\right)^+ \\ &\ge \left(a - \frac{|M(t)|}{t}\right)^+ - \frac{a}{t}. \end{aligned}$$

We used $(v-h)^+ \ge v^+ - h$ for $h \ge 0$. Thus $(a - |M(t)|/t)^+ \le D_y(t) + a/t$.

Integrating and comparing deficit with excess,

$$\begin{aligned}\int_1^y \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} &\leq \int_1^y D_y(t)\frac{\mathrm{d}t}{t} + a\int_1^y \frac{\mathrm{d}t}{t^2} \\ &\leq \int_1^\infty E_y(t)\frac{\mathrm{d}t}{t} + a.\end{aligned}$$

We have therefore reduced the Fundamental Lemma to proving that

$$\int_1^\infty E_y(t)\frac{\mathrm{d}t}{t} = o(\log y).$$

We have $0 \leq E_y(t) \leq |M(t,y)|/t \leq \Psi(t,y)/t \leq 1$. Hence

$$\begin{aligned}\int_1^\infty E_y(t)\frac{\mathrm{d}t}{t} &\leq \int_1^\infty \frac{\Psi(t,y)}{t}\frac{\mathrm{d}t}{t} \\ &= \prod_{p\leq y}(1-1/p)^{-1} = O(\log y),\end{aligned}$$

by Mertens' product estimate.

So the remaining fight is over the case of a single letter: We must improve $O(\log y)$ to $o(\log y)$.

We split the integral at y. The required estimate below y follows from the definition of a:

$$\int_1^y E_y(t)\frac{\mathrm{d}t}{t} = o(\log y).$$

Fix $\epsilon > 0$. By the definition of a, choose $T \geq 1$ so that

$$\frac{|M(t)|}{t} \leq a + \epsilon \qquad (t \geq T).$$

Here $M(t, y) = M(t)$ and $\Psi(t, y) = \lfloor t \rfloor$ when $t \leq y$.

Finishing the estimate below y

For $T \le t \le y$,

$$E_y(t) = \left(\frac{|M(t)|}{t} - a + a\frac{\{t\}}{t} \right)^+ \le \epsilon + \frac{a}{t}.$$

Using $E_y(t) \le 1$ on the initial interval,

$$\begin{aligned} \int_1^y E_y(t) \frac{\mathrm{d}t}{t} &\le \log T + \int_T^y \left(\epsilon + \frac{a}{t} \right) \frac{\mathrm{d}t}{t} \\ &\le \log T + \epsilon \log(y/T) + a. \end{aligned}$$

Divide by $\log y$ and let $y \to \infty$. The limsup is at most every $\epsilon > 0$, so it is zero.

An identity for the smooth Möbius sum

To handle the remaining range (y, ∞), we need two logarithmic relations. We begin their derivation with some simple consequences of Möbius inversion. Recall

$$\Lambda(n) = -\sum_{d|n} \mu(d) \log d.$$

Möbius inversion gives

$$-\mu(n) \log n = \sum_{d|n} \Lambda(d)\mu(n/d) = \sum_{p^k|n} \log p \, \mu(n/p^k).$$

We sum over smooth integers.

$$
\begin{aligned}
-\sum_{\substack{n\le x\\ P^+(n)\le y}}\mu(n)\log n &= \sum_{\substack{n\le x\\ P^+(n)\le y}}\sum_{p^k\mid n}\log p\,\mu(n/p^k)\\
&= \sum_{\substack{p\le y\\ p^k\le x}}\log p\sum_{\substack{m\le x/p^k\\ P^+(m)\le y}}\mu(m)\\
&= \sum_{\substack{p\le y\\ p^k\le x}}\log p\,M(x/p^k,y).
\end{aligned}
$$

We put $n=p^km$. Since $p\le y$, smoothness of n is equivalent to smoothness of m.

On the other hand, partial summation gives

$$\sum_{\substack{n\le x\\ P^+(n)\le y}} \mu(n)\log n = M(x,y)\log x - \int_1^x M(t,y)\frac{\mathrm{d}t}{t}.$$

Substituting into the preceding identity and rearranging,

$$M(x,y)\log x = -\sum_{\substack{p\le y\\ p^k\le x}} \log p\, M(x/p^k,y) + \int_1^x M(t,y)\frac{\mathrm{d}t}{t}.$$

Taking absolute values

Taking absolute values and appealing to the triangle inequality,

$$|M(x,y)|\log x \leq \sum_{\substack{p\leq y\\ p^k\leq x}} \log p\,|M(x/p^k,y)| + \int_1^x |M(t,y)|\frac{\mathrm{d}t}{t}.$$

We will compare this with an exact identity for $\Psi(x,y)$.

Start instead with $\log n = \sum_{p^k \mid n} \log p$. Summing over smooth integers gives

$$\sum_{\substack{n \le x \\ P^+(n) \le y}} \log n = \sum_{\substack{p \le y \\ p^k \le x}} \log p \, \Psi(x/p^k, y).$$

Partial summation yields

$$\Psi(x,y) \log x = \sum_{\substack{p \le y \\ p^k \le x}} \log p \, \Psi(x/p^k, y) + \int_1^x \Psi(t,y) \frac{\mathrm{d}t}{t}.$$

We subtract a times this identity from the inequality for $|M|$.

We obtain

$$\big(|M(x,y)| - a\Psi(x,y)\big)\log x \le \sum_{\substack{p\le y\\ p^k\le x}} \log p\,\big(|M(x/p^k,y)| - a\Psi(x/p^k,y)\big) + \int_1^x \big(|M(t,y)| - a\Psi(t,y)\big)\frac{\mathrm{d}t}{t}.$$

Replacing each difference on the right by its positive part only increases the right-hand side and so preserves the inequality. Since the resulting right-hand side is nonnegative, we may also replace the left-hand side by its positive part.

Since $(|M(t,y)| - a\Psi(t,y))^+ = tE_y(t)$,

$$xE_y(x)\log x \le \sum_{\substack{p\le y\\ p^k\le x}} \log p\, \frac{x}{p^k} E_y(x/p^k) + \int_1^x tE_y(t)\frac{\mathrm{d}t}{t}.$$

Divide by x:

$$E_y(x)\log x \le \sum_{\substack{p\le y\\ p^k\le x}} \frac{\log p}{p^k} E_y(x/p^k) + \int_1^x \frac{t}{x} E_y(t)\frac{\mathrm{d}t}{t}.$$

Next we integrate with respect to $\mathrm{d}x/x$.

Integrating the excess inequality

Recall

$$E_y(x)\log x \le \sum_{\substack{p\le y\\ p^k\le x}} \frac{\log p}{p^k} E_y(x/p^k) + \int_1^x \frac{t}{x} E_y(t)\frac{\mathrm{d}t}{t}.$$

For fixed y, $M(t,y)=0$ for $t\ge P$, and so $E_y(t)=0$ there. The integral on the left is finite.

Here all manipulations of sums and integrals are justified by nonnegativity.

Integrating the prime-power terms

For the term indexed by p^k, substitute $t = x/p^k$:

$$\int_{p^k}^{\infty} \frac{\log p}{p^k} E_y(x/p^k) \frac{\mathrm{d}x}{x} = \frac{\log p}{p^k} \int_1^{\infty} E_y(t) \frac{\mathrm{d}t}{t}.$$

Therefore the first contribution is

$$\left(\sum_{p \le y} \sum_{k \ge 1} \frac{\log p}{p^k} \right) \int_1^{\infty} E_y(t) \frac{\mathrm{d}t}{t}.$$

For the remaining term, we reverse the order of integration:

$$\int_1^\infty \left(\int_1^x \frac{t}{x} E_y(t) \frac{\mathrm{d}t}{t} \right) \frac{\mathrm{d}x}{x} = \int_1^\infty E_y(t) \left(\int_t^\infty \frac{t}{x} \frac{\mathrm{d}x}{x} \right) \frac{\mathrm{d}t}{t}.$$

The inner integral is

$$\int_t^\infty \frac{t}{x^2} \, \mathrm{d}x = \left[-\frac{t}{x} \right]_{x=t}^{x=\infty} = 1.$$

Thus this contribution is simply $\int_1^\infty E_y(t) \, \mathrm{d}t/t$.

Combining the contributions,

$$\int_1^\infty E_y(t) \log t \frac{\mathrm{d}t}{t} \leq \left(1 + \sum_{p \leq y} \sum_{k \geq 1} \frac{\log p}{p^k}\right) \int_1^\infty E_y(t) \frac{\mathrm{d}t}{t}.$$

The coefficient on the right is estimated using Mertens' theorem.

Sum the geometric series in k:

$$\begin{aligned}\sum_{p\le y}\sum_{k\ge 1}\frac{\log p}{p^k} &= \sum_{p\le y}\frac{\log p}{p-1}\\ &= \sum_{p\le y}\frac{\log p}{p}+\sum_{p\le y}\frac{\log p}{p(p-1)}\\ &= \log y + O(1).\end{aligned}$$

The first sum is Mertens' estimate. The second is bounded by the convergent series $\sum_{n\ge 2}(\log n)/(n(n-1))$.

A weighted bound for the excess

For an absolute constant $C > 0$,

$$\int_1^\infty E_y(t) \log t \frac{\mathrm{d}t}{t} \leq (\log y + C) \int_1^\infty E_y(t) \frac{\mathrm{d}t}{t}.$$

Subtracting $\log y$ times the excess integral gives

$$\int_1^\infty E_y(t) \log(t/y) \frac{\mathrm{d}t}{t} \leq C \int_1^\infty E_y(t) \frac{\mathrm{d}t}{t} = O(\log y).$$

With the factor $\log(t/y)$ omitted, we already had an $O(\log y)$ bound; the advantage here is that the new factor is sensitive to where the excess occurs.

Let's mop up

Recall that $\int_1^y E_y(t)\,\mathrm{d}t/t = o(\log y)$. Slightly larger values of t, say $y \le t \le y^{1+\epsilon}$ with $\epsilon > 0$ fixed, pose no trouble. Since $E_y(t) \le 1$,

$$\int_y^{y^{1+\epsilon}} E_y(t)\frac{\mathrm{d}t}{t} \le \epsilon \log y,$$

which will be acceptable to us upon taking ϵ small.

It remains to consider t that are substantially larger than y, meaning $t \ge y^{1+\epsilon}$. Moving the contribution from $[1, y]$ in the weighted inequality to the right,

$$\begin{aligned}\int_y^\infty E_y(t)\log(t/y)\frac{\mathrm{d}t}{t} &\le \int_1^y E_y(t)\log(y/t)\frac{\mathrm{d}t}{t} + O(\log y)\\ &\le (\log y)\int_1^y E_y(t)\frac{\mathrm{d}t}{t} + O(\log y)\\ &= o((\log y)^2).\end{aligned}$$

Combining the three ranges

On the other hand, $\log(t/y) \geq \epsilon \log y$ when $t \geq y^{1+\epsilon}$, so

$$\int_y^\infty E_y(t) \log(t/y) \frac{\mathrm{d}t}{t} \geq \epsilon \log y \int_{y^{1+\epsilon}}^\infty E_y(t) \frac{\mathrm{d}t}{t}.$$

Comparing with the preceding estimate, we find that

$$\int_{y^{1+\epsilon}}^\infty E_y(t) \frac{\mathrm{d}t}{t} = o(\log y).$$

Combining the estimates for the three ranges gives

$$\int_1^\infty E_y(t) \frac{\mathrm{d}t}{t} \leq \epsilon \log y + o(\log y).$$

Dividing by $\log y$ and letting $y \to \infty$, the limsup is at most ϵ. But $\epsilon > 0$ is arbitrary, and so $\int_1^\infty E_y(t)\, \mathrm{d}t/t = o(\log y)$, as desired.

Conclusion: $M(x) = o(x)$

The deficit estimate now gives

$$\int_1^y \left(a - \frac{|M(t)|}{t}\right)^+ \frac{\mathrm{d}t}{t} \leq a + \int_1^\infty E_y(t)\frac{\mathrm{d}t}{t} = o(\log y).$$

This proves the Fundamental Lemma. We already showed that the lemma implies $a = 0$. Hence

$$\boxed{M(x) = \sum_{n \leq x} \mu(n) = o(x).}$$

We want to prove $\psi(x) \sim x$.

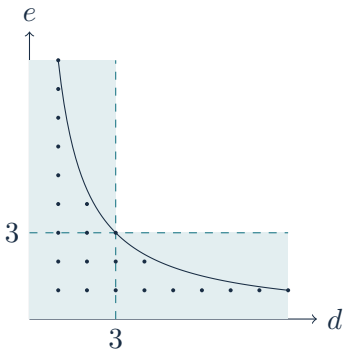
Strategy. Write

$$\Lambda(n) - 1 + (\text{stuff with small partial sums}) = \sum_{de=n} \mu(d)g(e),$$

where g has small partial sums.

Sum both sides on $n \leq x$. The left side is $\psi(x) - x + o(x)$. We show that the right side is $o(x)$, using the hyperbola method.

The hyperbola method



Here $x = 9$. Each dot is a pair of positive integers d, e with $de \leq 9$.
Count the pairs with $d \leq 3$, then those with $e \leq 3$. Every pair is counted, and the nine pairs with both coordinates at most 3 are counted twice.
For general x, split at $\sqrt{x}$.

The average order of $\tau(n)$

Since $\tau(n)$ counts positive divisors,

$$\sum_{n \le x} \tau(n) = \sum_{de \le x} 1 = 2 \sum_{d \le \sqrt{x}} \left\lfloor \frac{x}{d} \right\rfloor - \lfloor \sqrt{x} \rfloor^2.$$

Every pair has $d \le \sqrt{x}$ or $e \le \sqrt{x}$; the square is counted twice.

Dropping the floors in the sum costs $O(\sqrt{x})$, and $\lfloor \sqrt{x} \rfloor^2 = x + O(\sqrt{x})$. The harmonic-sum estimate therefore gives

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

Subtracting our estimate for $\sum_{n\le x}\tau(n)$ from the one for $\sum_{n\le x}\log n$ gives

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

Too large! So put

$$g(n) = \log n - \tau(n) + 2\gamma, \qquad G(x) = \sum_{n\le x} g(n).$$

Since $2\gamma\lfloor x\rfloor = 2\gamma x + O(1)$, we have $G(x) = O(\sqrt{x})$.

Recall $\sum_{d|n}\mu(d)\log(n/d) = \Lambda(n)$. What happens with τ in place of the logarithm?

Expand $\tau(n/d)$ and sum on e first:

$$\sum_{d|n} \mu(d)\tau(n/d) = \sum_{de|n} \mu(d) = \sum_{e|n} \sum_{d|n/e} \mu(d) = \sum_{e|n} \mathbf{1}_{n/e=1} = 1.$$

Consequently

$$\sum_{dj=n} \mu(d)g(j) = \Lambda(n) - 1 + 2\gamma \mathbf{1}_{n=1}.$$

Summing, it remains to prove

$$\psi(x) - \lfloor x \rfloor + 2\gamma = \sum_{dj \leq x} \mu(d)g(j) = o(x).$$

Using the hyperbola method again

Fix a positive integer K. For $x \geq K$, split according as $j \leq K$ or $j > K$:

$$\sum_{dj \leq x} \mu(d)g(j) = \sum_{j \leq K} g(j)M(x/j) + \sum_{d \leq x/K} \mu(d)\big(G(x/d) - G(K)\big).$$

For each fixed j, $M(x/j) = o(x)$, so the first sum is $o_K(x)$.

In the second, $x/d \geq K$, and

$$|G(x/d) - G(K)| \ll \sqrt{x/d} + \sqrt{K} \ll \sqrt{x/d},$$

with constants independent of K. Its absolute value is bounded by a constant times

$$\sqrt{x} \sum_{d \leq x/K} \frac{1}{\sqrt{d}} \leq 2\sqrt{x}\sqrt{x/K} = \frac{2x}{\sqrt{K}}.$$

For every fixed $K \geq 1$,

$$\psi(x) - x = o_K(x) + O(x/\sqrt{K}).$$

Divide by x and take the limsup:

$$\limsup_{x \to \infty} \left| \frac{\psi(x)}{x} - 1 \right| \leq \frac{C}{\sqrt{K}}.$$

Since K can be taken arbitrarily large, the limsup is zero. Thus

$$\boxed{\psi(x) \sim x.}$$

We know $M(x) = o(x)$. First we prove a lemma:

$$R(x) := \sum_{d \le x} \mu(d) \left\{ \frac{x}{d} \right\} = o(x).$$

Fix an integer $K > 2$. We break the sum at x/K. Then

$$\left| \sum_{d \le x/K} \mu(d) \left\{ \frac{x}{d} \right\} \right| \le \frac{x}{K}.$$

Now consider $x/K < d \le x$. I claim this part of the sum is $o(x)$!

BONUS: partial summation

Put $a = x/K$. Partial summation gives

$$x \sum_{a<d\le x} \frac{\mu(d)}{d} = M(x) - KM(a) + x\int_a^x \frac{M(t)}{t^2}\,\mathrm{d}t.$$

For $d > a$, the integer $\lfloor x/d \rfloor$ counts $1 \le j \le K-1$ with $d \le x/j$. Thus

$$\sum_{a<d\le x} \mu(d)\left\lfloor \frac{x}{d}\right\rfloor = \sum_{j=1}^{K-1} \big(M(x/j) - M(a)\big).$$

Subtracting the two identities gives

$$\sum_{x/K<d\le x} \mu(d)\left\{\frac{x}{d}\right\} = x\int_{x/K}^x \frac{M(t)}{t^2}\,\mathrm{d}t - \sum_{j=2}^{K} M(x/j).$$

Divide the last identity by x. Each term $M(x/j)/x$ goes to zero, for $j = 2, \dots, K$. Also,

$$\left| \int_{x/K}^{x} \frac{M(t)}{t^2}\, \mathrm{d}t \right| \leq \sup_{t \geq x/K} \frac{|M(t)|}{t} \int_{x/K}^{x} \frac{\mathrm{d}t}{t} = \log K \sup_{t \geq x/K} \frac{|M(t)|}{t} \longrightarrow 0.$$

Therefore

$$|R(x)| \leq \frac{x}{K} + o_K(x), \qquad \limsup_{x \to \infty} \frac{|R(x)|}{x} \leq \frac{1}{K}.$$

Take K large. This proves $R(x) = o(x)$.

We showed that

$$1 = \sum_{n \le x} \sum_{d|n} \mu(d) = \sum_{d \le x} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor$$
$$= x \sum_{d \le x} \frac{\mu(d)}{d} - R(x).$$

Rearranging,

$$\sum_{d \le x} \frac{\mu(d)}{d} = \frac{1}{x} + \frac{R(x)}{x} \longrightarrow 0.$$

We win!

$$\boxed{\sum_{n=1}^{\infty} \frac{\mu(n)}{n} = 0.}$$

Course materials

MATH 8440, Fall 2026
University of Georgia. The course page has the dated lecture record and four assignments.

Homework 1 Homework 2 Homework 3 Homework 4

Not Always Buried Deep
Paul Pollack, *A Second Course in Elementary Number Theory*, AMS, 2009. Chapters 1, 3, and 4 accompany the lectures on infinitude, prime estimates, and progressions. The constant in Mertens' product is Theorem 3.17.

From Möbius sums to the PNT
The course handout for the final sequence of lectures.

Further reading

The missing Lin(ni)k

Paul Pollack, Hexagon, 2026. An elementary proof of Linnik's theorem using Selberg's symmetry formulas and the exclusion of Siegel zeros.

Daboussi's elementary proof of the prime number theorem, reconsidered

Paul Pollack, Hexagon, 2026. A simplification of Daboussi's smooth-number argument for $M(x) = o(x)$, underlying the final part of these notes.

Uniform exclusion of Landau–Siegel zeros

OpenAI, October 2026. The breakthrough excluding real zeros uniformly close to 1 on the scale $1/\log m$.