From Möbius sums to the PNT

An elementary proof of the Prime Number Theorem

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

Our goal

We will prove the Prime Number Theorem in the form

$$\psi(x) = \sum_{n \le x} \Lambda(n) \sim x, \qquad \text{and hence} \qquad \pi(x) \sim \frac{x}{\log x}.$$

Our first goal is cancellation in the Möbius sum:

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

The main step is a Fundamental Lemma, which we reduce to an estimate involving smooth numbers.

We then finish this estimate and deduce the asymptotic formula for $\psi(x)$.

Partial sums of the Möbius function

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 to consider:

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

The first counts squarefree integers. The second counts those having an even number of prime factors, minus those having an odd number.

Counting squarefree integers

Theorem

For $x \geq 1$,

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

For the proof, we use the identity

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

To see this, expand $\prod_{p^2 | n} (1 - 1)$. The product is 1 when n is squarefree (an empty product), and 0 otherwise.

Counting squarefree integers, continued

Summing on $n \le x$ and reversing the order of summation, we find

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

We can extend the sum to infinity at little cost, since

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

Thus $A(x) = x \sum_{d \ge 1} \mu(d)/d^2 + O(\sqrt{x})$.

From $A(x)$ to $M(x)$

It remains to evaluate the constant. By absolute convergence,

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

Hence $\sum_{d\geq 1}\mu(d)/d^2=6/\pi^2$, proving the formula for $A(x)$.

Understanding the behavior of $M(x)$ as $x\to\infty$ is much more difficult!

Theorem (to be proved)

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

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

Why care about $M(x)$?

The assertion $M(x) = o(x)$ is intrinsically interesting: among squarefree integers, an even number of prime factors and an odd number of prime factors occur equally often in the limit.

But there is another reason to care. This cancellation is exactly what we need to understand the asymptotic distribution of primes.

$$M(x) = o(x) \quad \Longrightarrow \quad \pi(x) \sim \frac{x}{\log x}.$$

We will prove this implication at the end. For now, we concentrate on $M(x)$.

A first bit of cancellation

The sum $\sum_{n\le x}\mu(n)/n$ is much easier to handle. Summing $\sum_{d|n}\mu(d)=\mathbf{1}_{n=1}$ on $n\le x$ gives

$$\begin{aligned}1&=\sum_{d\le x}\mu(d)\lfloor\frac{x}{d}\rfloor\\&=x\sum_{d\le x}\frac{\mu(d)}{d}-\sum_{d\le x}\mu(d)\{\frac{x}{d}\}.\end{aligned}$$

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

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

An averaged version

Let us put this in a form we can use later. Partial summation gives

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

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

$$\left| \int_1^x \frac{M(t)}{t} \frac{dt}{t} \right| < 3 \qquad (x \ge 1).$$

So $M(t)/t$ tends to 0 on logarithmic average:

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

There is still a gap between this and $M(x)/x \to 0$.

The Fundamental Lemma

Put

$$a = \limsup_{t\to\infty} \frac{|M(t)|}{t}, \qquad 0 \le a \le 1.$$

We want to prove that $a = 0$. Write $v^+ = \max\{v, 0\}$.

Fundamental Lemma

As $x \to \infty$,

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

Roughly speaking, $|M(t)|/t$ spends most of its time near its limsup, where we measure time using dt/t.

Why this is enough

Before proving the lemma, let us see how it gives $a = 0$. Suppose, for a contradiction, that $a > 0$.

If $|M(t)|/t$ stayed above $a/2$ for too long, then $M(t)$ would keep the same sign, and its signed integral would become too large.

Thus small values must occur. Moreover, $M(t)$ changes slowly enough that each small value gives a whole interval on which $|M(t)|/t < a$.

These intervals will contribute too much to

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

contradicting the Fundamental Lemma.

Finding a small value

Assume $a > 0$, and take $B = \exp(12/a)$.

Claim

For every $X \geq 1$, some $u \in [X, BX]$ satisfies $|M(u)|/u \leq a/2$.

Otherwise $M(t)$ never vanishes on $[X, BX]$. Since it takes integer values and jumps by at most 1, it cannot change sign there. Hence

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

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

One small value gives many

Let $X \geq X_0 := 8/a$, and choose $u \in [X, BX]$ with $|M(u)| \leq au/2$.

We claim that $|M(t)|/t \leq 3a/4$ throughout $[u, u(1 + a/8)]$. Indeed,

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

In the last step we used $t \geq u \geq 8/a$. Since $a \leq 1$, the whole interval lies in $[X, 2BX]$.

The resulting saving

Consequently, for every $X \geq X_0$,

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

This is a fixed positive saving, independent of X.

Notice the role of dt/t: moving $[u, u(1 + a/8)]$ farther out does not change its logarithmic length.

Adding the savings

Now use the preceding estimate with

$$X = X_0,\ (2B)X_0,\ \ldots,\ (2B)^{N-1}X_0.$$

The intervals have disjoint interiors. Adding their contributions gives

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

The Fundamental Lemma says that the left side is $o(N)$, since

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

This is a contradiction, since $\frac{a}{4}\log(1 + a/8) > 0$.

It remains to prove the Fundamental Lemma.

Restricting the prime factors

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

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

Throughout, take $y \ge 3$ and put

$$P = \prod_{p \le y} p.$$

For $t \le y$, we still have $M(t,y) = M(t)$. The advantage is that, with y fixed, we can say much more about $M(t,y)$ for large t.

Some useful observations

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

For $1 \leq t \leq y$, every integer up to t is y-smooth, so

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

For $t \geq P$, every squarefree y-smooth integer is a divisor of P, and every divisor of P has already appeared. Thus

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

Thus, for fixed y, the sum $M(t,y)$ eventually vanishes altogether.

An integral comparison

Our first task is to show that, for every $y \geq 3$,

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

Below y, the integrand is

$$\frac{|M(t)|}{t} - a\frac{\lfloor t \rfloor}{t},$$

which differs from $|M(t)|/t - a$ by at most a/t.

This will let us bound the deficit in the Fundamental Lemma by the total excess, apart from a harmless error coming from $\lfloor t \rfloor$.

Splitting off the small prime factors

Every positive integer has a unique factorization

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

Simply put all prime-power factors belonging to primes $\le y$ into b, and the remaining ones into c. Then $(b, c) = 1$, so that

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

Grouping by $\lfloor x/c \rfloor$

Taking absolute values, and recalling that $M(t, y) = 0$ for $t \geq P$,

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

To estimate the right side, group terms according to $j = \lfloor x/c \rfloor$:

$$\frac{x}{j+1} < c \leq \frac{x}{j}, \qquad M(x/c, y) = M(j, y).$$

Thus

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

For now, y and P are fixed, and $x \to \infty$.

Counting integers coprime to P

Each block of P consecutive integers contains exactly $\varphi(P)$ integers coprime to P. Thus

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

Since y is fixed, we may absorb all the errors into $O_y(1)$. Thus

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

Dividing by x and taking the limsup gives

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

Writing the sum as an integral

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

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

Therefore

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

The same calculation for Ψ

This time we get an exact answer. First note that, for fixed y,

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

Indeed, each of the $\pi(y)$ prime exponents is at most $\log x/\log 2$. Now partial summation gives

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

Letting $x\to\infty$, we obtain

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

Putting the two estimates together

We have shown that

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

Equivalently,

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

We now let $y \to \infty$.

Excess and deficit

To keep track of the excess and the deficit, put

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

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

Then

$$\frac{|M(t,y)|}{t} - a\frac{\Psi(t,y)}{t} = E_y(t) - D_y(t).$$

The comparison just proved says precisely that

$$\int_1^\infty D_y(t)\,\frac{dt}{t} \le \int_1^\infty E_y(t)\,\frac{dt}{t}.$$

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

Returning to the Fundamental Lemma

For $1 \le t \le y$, we have

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

Here we used $(v-h)^+ \ge v^+ - h$ for $h \ge 0$.

Consequently,

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

What remains to be shown

Integrating, and using that total deficit is at most total excess,

$$\begin{aligned}\int_1^y \left(a - \frac{|M(t)|}{t}\right)^+ \frac{dt}{t} &\le \int_1^y D_y(t)\,\frac{dt}{t} + a\int_1^y \frac{1}{t}\,\frac{dt}{t} \\ &\le \int_1^\infty D_y(t)\,\frac{dt}{t} + a \\ &\le \int_1^\infty E_y(t)\,\frac{dt}{t} + a.\end{aligned}$$

We will be finished if we can show

$$\int_1^\infty E_y(t)\,\frac{dt}{t} = o(\log y) \qquad (y \to \infty).$$

A first estimate for the excess

There is an easy upper bound. Since

$$0 \le E_y(t) \le \frac{|M(t,y)|}{t} \le \frac{\Psi(t,y)}{t} \le 1.$$

we have

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

by Mertens' product estimate.

We need to change $O(\log y)$ to $o(\log y)$.

The excess below y

Recall that

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

We split the integral at $t = y$. The first part is easy, essentially because of how we defined a.

Claim

$$\int_1^y E_y(t)\,\frac{dt}{t} = o(\log y) \qquad (y \to \infty).$$

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

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

Using the definition of a

For $T \le t \le y$, we have $M(t,y) = M(t)$ and $\Psi(t,y) = \lfloor t \rfloor$. Thus

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

Since $E_y(t) \le 1$, for $y \ge T$ we obtain

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

Divide by $\log y$ and let $y \to \infty$. The limsup is at most ϵ. Since $\epsilon > 0$ is arbitrary, the claim follows.

A logarithmic identity

The range $t > y$ is much trickier!

We will use a logarithmic identity to control how far out the excess can lie. Recall from our proof of Dirichlet's theorem that

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

By Möbius inversion,

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

We sum this identity over y-smooth integers.

Summing over smooth integers

Reversing the order of summation gives

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

In the third line, put $n = p^k m$; smoothness is preserved since $p \le y$.

Partial summation

On the other hand,

$$\sum_{\substack{n\le x\\ P^+(n)\le y}} \mu(n)\log n = \sum_{n\le x} \log n\big(M(n,y)-M(n-1,y)\big)$$
$$= M(x,y)\log x - \int_1^x M(t,y)\,\frac{dt}{t}.$$

Combining the two expressions and rearranging, we get

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

Taking absolute values

We have just shown that

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

Taking absolute values therefore gives

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

The matching identity for Ψ

Start instead from

$$\log n = \sum_{p^k \mid n} \log p.$$

Summing over y-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 now yields the exact identity

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

The two formulas have the same shape. Subtracting will give an inequality for the excess.

Subtracting the positive model

Subtract a times the identity for Ψ from the inequality for $|M|$:

$$\begin{aligned}&(|M(x,y)| - a\Psi(x,y)) \log x \\ &\quad \le \sum_{\substack{p \le y \\ p^k \le x}} \log p \big(|M(x/p^k, y)| - a\Psi(x/p^k, y)\big) \\ &\qquad + \int_1^x \left(|M(t,y)| - a\Psi(t,y)\right) \frac{dt}{t}.\end{aligned}$$

Replace every difference on the right by its positive part. This can only increase the right side.

The right side is now nonnegative, so we may also replace the left side by its positive part.

An inequality for the excess

Since

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

we obtain

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

Divide by x:

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

We will integrate this inequality with respect to dx/x.

Why integrate?

Recall the inequality

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

Integrating over $1 \le x < \infty$ will turn each term on the right into a multiple of the total excess mass.

For fixed y, recall that $M(t, y) = 0$ for $t \ge P$. Since $a \ge 0$, we also have $E_y(t) = 0$ there. Thus the integral on the left is finite.

All terms on the right are nonnegative, so the interchanges of sums and integrals below are valid.

Integrating the prime-power sum

For each $p \le y$ and $k \ge 1$, the range of integration is $x \ge p^k$. Put $t = x/p^k$, so that $dx/x = dt/t$. Then

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

Consequently,

$$\int_1^{\infty} \sum_{\substack{p \le y \\ p^k \le x}} \frac{\log p}{p^k} E_y(x/p^k)\, \frac{dx}{x} = \left(\sum_{p \le y} \sum_{k \ge 1} \frac{\log p}{p^k} \right) \int_1^{\infty} E_y(t)\, \frac{dt}{t}.$$

Integrating the remaining term

Reversing the order of integration gives

$$\int_1^\infty \left(\int_1^x \frac{t}{x} E_y(t) \, \frac{dt}{t} \right) \frac{dx}{x} = \int_1^\infty E_y(t) \left(\int_t^\infty \frac{t}{x} \, \frac{dx}{x} \right) \frac{dt}{t}.$$

But

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

So this contribution is just

$$\int_1^\infty E_y(t) \, \frac{dt}{t}.$$

The integrated inequality

Putting the two contributions together, we obtain

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

The factor $\log t$ on the left records how far out the excess lies.

On the right, we have the total excess mass multiplied by a prime-power sum. Mertens' estimate will show that this multiplier is $\log y + O(1)$.

Evaluating the prime-power sum

Summing the geometric series, we find

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

Here we used Mertens' estimate for the first sum. The second sum is bounded, since

$$\sum_{n\ge 2}\frac{\log n}{n(n-1)} < \infty.$$

Measuring the excess relative to y

For some absolute constant $C > 0$, we have proved

$$\int_1^\infty E_y(t) \log t \, \frac{dt}{t} \le (\log y + C) \int_1^\infty E_y(t) \, \frac{dt}{t}.$$

Subtract $\log y$ times the excess mass. This gives

$$\int_1^\infty E_y(t) \log(t/y) \, \frac{dt}{t} \le C \int_1^\infty E_y(t) \, \frac{dt}{t} = O(\log y).$$

The weight $\log(t/y)$ is negative below y and positive above y. We now have an upper bound that takes the location of the excess into account.

The three ranges

We can now finish proving

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

Recall that $0 \le E_y(t) \le 1$, and that

$$\int_1^y E_y(t)\,\frac{\mathrm{d}t}{t} = o(\log y), \qquad \int_1^\infty E_y(t)\log(t/y)\,\frac{\mathrm{d}t}{t} \le C\log y.$$

Fix $\epsilon > 0$. Split the integral into

$$[1, y], \qquad (y, y^{1+\epsilon}], \qquad (y^{1+\epsilon}, \infty).$$

The first range is already handled. The middle range is short on the logarithmic scale. The weighted estimate will control the last range.

Slightly larger values of t

For the middle range, the trivial bound is enough:

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

This will be fine, because we can take ϵ arbitrarily small.

It remains to handle $t > y^{1+\epsilon}$.

Moving the negative part to the other side

The weight $\log(t/y)$ is negative below y. Thus

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

Here we used the estimate below y, and $\log y = o((\log y)^2)$.

The integral on the left is nonnegative.

The far tail

If $t \geq y^{1+\epsilon}$, then $\log(t/y) \geq \epsilon \log y$. Hence

$$\epsilon \log y \int_{y^{1+\epsilon}}^{\infty} E_y(t) \frac{\mathrm{d}t}{t} \leq \int_{y}^{\infty} E_y(t) \log(t/y) \frac{\mathrm{d}t}{t} = o\big((\log y)^2\big).$$

For each fixed $\epsilon > 0$, this gives

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

Notice that the lower limit here is $y^{1+\epsilon}$, and the upper limit is ∞.

Finishing the excess estimate

Putting the three pieces together,

$$0 \le \int_1^\infty E_y(t)\,\frac{\mathrm{d}t}{t} \le o(\log y) + \epsilon \log y + o_\epsilon(\log y).$$

Divide by $\log y$ and let $y \to \infty$, with ϵ fixed:

$$0 \le \limsup_{y\to\infty} \frac{1}{\log y} \int_1^\infty E_y(t)\,\frac{\mathrm{d}t}{t} \le \epsilon.$$

Since $\epsilon > 0$ is arbitrary,

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

The Fundamental Lemma, and $M(x) = o(x)$

Recall the comparison between deficit and excess:

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

This proves the Fundamental Lemma. As we already showed, this implies

$$\boxed{M(x) = o(x).}$$

From $M(x) = o(x)$ to the PNT

Our end-game is the Prime Number Theorem in the form

$$\psi(x) = \sum_{p^k \le x} \log p = \sum_{n \le x} \Lambda(n) \sim x.$$

We will use

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

and

$$\sum_{n \le x} \log n = x \log x - x + O(\log x) \qquad (x \ge 2).$$

Here $\tau(n) = \sum_{d|n} 1$, and γ is Euler's constant.

Dirichlet's hyperbola method

We have

$$\sum_{n \le x} \tau(n) = \sum_{n \le x} \sum_{de = n} 1 = \sum_{de \le x} 1.$$

This counts points below the hyperbola $de = x$, with positive integer coordinates.

Every such point has $d \le \sqrt{x}$ or $e \le \sqrt{x}$. Counting these two regions and subtracting their overlap gives

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

The two regions contribute equally, by symmetry.

Evaluating the divisor sum

Using the harmonic-sum estimate,

$$\begin{aligned}\sum_{d\le\sqrt{x}}\left\lfloor\frac{x}{d}\right\rfloor &= x\sum_{d\le\sqrt{x}}\frac{1}{d}+O(\sqrt{x})\\ &= x\left(\log\sqrt{x}+\gamma+O(x^{-1/2})\right)+O(\sqrt{x})\\ &= \tfrac{1}{2}x\log x+\gamma x+O(\sqrt{x}).\end{aligned}$$

Also, $\lfloor\sqrt{x}\rfloor^2 = x+O(\sqrt{x})$. Thus

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

The sum of the logarithms

Since $\log t$ is increasing, comparison with its integral gives

$$\sum_{n \le x} \log n = \int_1^x \log t \, \mathrm{d}t + O(\log x) = x \log x - x + O(\log x).$$

Subtracting the divisor-sum estimate,

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

So define

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

Then $F(x) = O(\sqrt{x})$, since $\lfloor x \rfloor = x + O(1)$.

Two consequences of Möbius inversion

From $\log n = \sum_{d|n} \Lambda(d)$, we obtain

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

Similarly, $\tau(n) = \sum_{d|n} 1$ gives

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

Finally, recall that

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

These three identities explain our choice of f.

A convolution with a small summatory function

We have

$$\sum_{de=n} \mu(d) f(e) = \sum_{de=n} \mu(d) \big(\log e - \tau(e) + 2\gamma \big)$$
$$= \Lambda(n) - 1 + 2\gamma \mathbf{1}_{n=1}.$$

Summing on $n \leq x$, for $x \geq 1$, gives

$$\psi(x) - \lfloor x \rfloor + 2\gamma = \sum_{de \leq x} \mu(d) f(e).$$

The left side is $\psi(x) - x + O(1)$. It is enough to show that the double sum is $o(x)$.

Splitting at a fixed cutoff

Fix a positive integer K, and suppose $x \geq K$. Split according to $j \leq K$ or $j > K$:

$$\begin{aligned}\sum_{dj \leq x} \mu(d) f(j) &= \sum_{j \leq K} f(j) \sum_{d \leq x/j} \mu(d) + \sum_{d \leq x/K} \mu(d) \sum_{K < j \leq x/d} f(j) \\ &= \sum_{j \leq K} f(j) M(x/j) + \sum_{d \leq x/K} \mu(d) \big(F(x/d) - F(K)\big).\end{aligned}$$

In the second sum, $j > K$ and $dj \leq x$ imply $d < x/K$. Including $d = x/K$ adds an empty inner sum.

The finitely many small values of j

For each fixed j,

$$M(x/j) = o(x),$$

because $M(t) = o(t)$.

There are only K terms, so

$$\sum_{j \le K} f(j) M(x/j) = o_K(x).$$

The subscript is a reminder: we keep K fixed while letting $x \to \infty$.

For the other sum, we will use $F(t) = O(\sqrt{t})$.

Bounding the remaining sum

For $d \le x/K$, we have $\sqrt{K} \le \sqrt{x/d}$. Hence

$$|F(x/d) - F(K)| \ll \sqrt{x/d} + \sqrt{K} \ll \sqrt{x/d}.$$

It follows that

$$\begin{aligned}\left| \sum_{d \le x/K} \mu(d)\big(F(x/d) - F(K)\big) \right| &\ll \sqrt{x} \sum_{d \le x/K} \frac{1}{\sqrt{d}} \\ &\ll \sqrt{x}\sqrt{x/K} = \frac{x}{\sqrt{K}}.\end{aligned}$$

The implied constant does not depend on K. We used $\sum_{d \le z} d^{-1/2} \le 2\sqrt{z}$ for $z \ge 1$.

Letting the cutoff grow

Combining the two estimates,

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

Divide by x, then let $x \to \infty$, keeping K fixed:

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

Now let $K \to \infty$. The right side tends to 0, so

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