9.1. **The Prime Number Theorem.** How frequent are the primes? At the end of the eighteenth century, Gauss and Legendre suggested giving up looking for a formula for the nth prime, and proposed instead estimating the number of primes up to x. So, define the prime-counting function $\pi(x)$ by

$$\pi(x) = \sum_{\substack{p \le x \\ p \text{ prime}}} 1.$$

Gauss conjectured on computational evidence that $\pi(x) \sim \frac{x}{\log x}$. This was proved by independently by Hadamard and de la Vallée Poussin in 1896, and became known as

Theorem 9.1 (The Prime Number Theorem)**.** *We have* $\pi(x) \sim \frac{x}{\log x}$ *as* $x \to \infty$.

It turns out to be more convenient to work with

$$\theta(x) = \sum_{\substack{p \le x \\ p \text{ prime}}} \log p,$$

which is called *Chebyshev's θ-function.* In terms of this function it can be shown (not difficult) that the Prime Number Theorem is equivalent to the statement $\theta(x) \sim x \quad (x \to \infty)$.

We won't prove PNT here, but instead a weaker version, and in terms of $\theta(x)$:

Theorem 9.2. *As* $x \to \infty$ *we have*

$$(\log 2)x + o(x) < \theta(x) < (2\log 2)x + o(x),$$

so that

$$0.6931x + o(x) < \theta(x) < 1.3863x + o(x).$$

9.2. **Proof of Theorem 9.2.**

9.2.1. *The upper bound.*

Proposition 9.3. *We have* $\theta(x) < (2\log 2)x + O(\log^2 x)$.

Proof. Consider $\binom{2n}{n}$. By the Binomial Theorem, it is less than $(1+1)^{2n} = 4^n$. Also, it is divisible by all primes p with $n < p \le 2n$, so

$$4^n > \binom{2n}{n} \ge \prod_{n<p\le 2n} p = e^{\theta(2n)-\theta(n)}.$$

Hence $\theta(2n) - \theta(n) \le 2n \log 2$.

Now if $2n \le x < 2n+2$ (i.e., $n \le x/2 < n+1$) then $\theta(x/2) = \theta(n)$ and

$$\theta(x) \le \theta(2n) + \log(2n+1) \le \theta(2n) + \log(x+1),$$

so that, for each x,

$$\begin{aligned}\theta(x)-\theta(x/2) &\le \theta(2n)+\log(x+1)-\theta(n)\\ &\le 2n\log 2+\log(x+1)\\ &\le x\log 2+\log(x+1).\end{aligned}$$

So (standard telescoping argument for $x, x/2, x/2^2, \dots, x/2^k$ where $x/2^{k-1}\ge 2$, $x/2^k<2$, $\theta(x/2^k)=0$):

$$\begin{aligned}\theta(x) &= \left(\theta(x)-\theta\left(\frac{x}{2}\right)\right)+\left(\theta\left(\frac{x}{2}\right)-\theta\left(\frac{x}{2^2}\right)\right)+\left(\theta\left(\frac{x}{2^2}\right)-\theta\left(\frac{x}{2^3}\right)\right)+\dots\left(\theta\left(\frac{x}{2^{k-1}}\right)-\theta\left(\frac{x}{2^k}\right)\right)\\ &\le \log 2\left(x+\frac{x}{2}+\cdots+\frac{x}{2^{k-1}}\right)+k\log(x+1)\\ &\le 2x\log 2+\lfloor\log_2 x\rfloor\log(x+1)\\ &\le 2x\log 2+O(\log^2 x).\end{aligned}$$

□

9.2.2. *The lower bound.* To obtain an inequality in the other direction, we look at

$$d_n=\operatorname{lcm}(1,2,\dots,n)=e^{\sum_{p^m\le x}\log p}.$$

Define

$$\psi(x)=\sum_{\substack{p^m\le x\\ p\text{ prime}}}\log p;$$

(i.e., $\log p$ to be counted m times if p^m is the highest power of p that is $\le x$). So $d_n=e^{\psi(n)}$.

Lemma 9.4. *We have* $\psi(x)<\theta(x)+2x^{1/2}\log x+O(\log^2 x)$.

Proof. Now

$$\begin{aligned}\psi(x) &= \sum_{p\le x}\log x+\sum_{p^2\le x}\log x+\sum_{p^3\le x}\log x+\dots\\ &= \theta(x)+\theta(x^{1/2})+\theta(x^{1/3})+\cdots+\theta(x^{1/k}),\end{aligned}$$

where k is greatest such that $x^{1/k}\ge 2$, i.e., $k=\lfloor\log_2 x\rfloor$

$$\begin{aligned}&< \theta(x)+\log_2 x\,\theta(x^{1/2})\\ &< \theta(x)+2x^{1/2}\log x+O(\log^2 x), \qquad \text{using Prop. 9.3.}\end{aligned}$$

□

Curious note: this k is the same one as in the proof of Prop. 9.3, though they have apparently different definitions.

We can now prove

Proposition 9.5. *We have* $\theta(x)\ge x\log 2+O(x^{1/2}\log x)$.

Proof. Consider the polynomial $p(t) = (t(1-t))^n$ on the interval $[0, 1]$. As $t(1-t) \le \frac{1}{4}$ on that interval (calculus!), we have

$$0 \le p(t) \le \frac{1}{4^n} \qquad \text{on } [0,1].$$

Writing $p(t) = \sum_{k=0}^{2n} a_k t^k \in \mathbb{Z}[t]$, then

$$\frac{1}{4^n} \ge \int_0^1 p(t)dt = \sum_{k=0}^{2n} \frac{a_k}{k+1} = \frac{N}{d_{2n+1}} \ge \frac{1}{d_{2n+1}},$$

for some $N \in \mathbb{N}$, on putting the fractions over a common denominator. Hence we have successively

$$\begin{aligned} d_{2n+1} &\ge 4^n \\ \psi(2n+1) &\ge 2n \log 2 && \text{on taking logs} \\ \theta(2n+1) &\ge 2n\log 2 - 2\log(2n+1)\sqrt{2n+1} && \text{by Lemma 9.4} \\ \theta(x) &\ge x \log 2 + O(x^{1/2}\log x). \end{aligned}$$

□

Combining Propositions 9.3 and 9.5, we certainly obtain Theorem 9.2.

9.3. **Some standard estimates.**

Lemma 9.6. *For $t > -1$ we have $\log(1+t) \le t$, with equality iff $t = 0$.*
For $n \in \mathbb{N}$ we have $n \log(1 + \frac{1}{n}) < 1$.

Proof. The first inequality comes from observing that the tangent $y = t$ to the graph of $y = \log(1+t)$ at $t = 0$ lies above the graph, touching it only at $t = 0$. The second inequality comes from putting $t = 1/n$ in the first inequality. □

Lemma 9.7 (Weak Stirling Formula)**.** *For $n \in \mathbb{N}$ we have*

$$n \log n - n < \log(n!) \le n \log n.$$

Proof. Now for $j \ge 2$ we have

$$\begin{aligned} \log j &= j\log j - (j-1)\log(j-1) - (j-1)\log\left(1 + \frac{1}{j-1}\right) \\ &= j \log j - (j-1)\log(j-1) - \delta_j, \end{aligned}$$

where $0 < \delta_j < 1$, using Lemma 9.6 for $n = j - 1$. So, on summing for $j = 2, \ldots, n$ we get

$$\begin{aligned}\log(n!) &= \sum_{j=2}^{n} \log j \\ &= \sum_{j=2}^{n} j \log j - (j-1)\log(j-1) - \delta_j \\ &= n \log n - \sum_{j=2}^{n} \delta_j \\ &= n \log n - \Delta,\end{aligned}$$

where $0 < \Delta < n$, since $1 \log 1 = 0$ and all the other $j \log j$ terms apart from $n \log n$ telescope. □

Proposition 9.8. *We have*

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

where $\gamma = 0.577\ldots$, the Euler-Mascheroni constant.

Proof. Draw the graph of $y = 1/t$ for t from $0+$ to $N+1$, where $N = \lfloor x \rfloor$. On each interval $[n, n+1]$ draw a rectangle of height $1/n$, so that these rectangles for $n = 1, 2, \ldots, N$ completely cover the area under the curve from $t = 1$ to $t = N+1$. The pie-shaped pieces of the rectangles above the curve, when moved to the left to lie above the interval $[0, 1]$, are non-intersecting, and more than half-fill the 1×1 square on that interval. Say their total area is γ_n. Then, as $n \to \infty$, γ_n clearly tends to a limit γ, the Euler-Mascheroni constant.

The sum of the areas of the rectangles above $[n, n+1]$ for $n = 1, 2, \ldots, N$ is clearly $\sum_{n=1}^{N} 1/n$ (the total area of the parts of the rectangles below the curve). On the other hand, it is $\int_1^{N+1} \frac{dx}{x} = \log(N+1)$ (the total area of the parts of the rectangles below the curve), plus γ_n (the total area of the parts of the rectangles above the curve). Hence

$$\sum_{n \le x} \frac{1}{n} = \sum_{n=1}^{N} 1/n = \log(N+1) + \gamma_n.$$

Since $\log(N+1) - \log x = O\left(\frac{1}{x}\right)$ and $\gamma - \gamma_n = O\left(\frac{1}{x}\right)$ (check!), we have the result. □

9.4. **More estimates of sums of functions over primes.** Let us put $\mathcal{P}_x = \prod_{p \le x} \frac{1}{1-p^{-1}}$. Then

Proposition 9.9. *We have $\mathcal{P}_x > \log x$.*

Proof. We have

$$\mathcal{P}_x = \prod_{p \le x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^n} + \ldots\right).$$

On multiplying these series together, we obtain a sum of terms that includes all fractions $\frac{1}{n}$, where $n \le x$. This is simply because all prime factors of such n are at most x. Hence

$$\mathcal{P}_x > \sum_{n \le x} \frac{1}{n} > \log x,$$

by Prop. 9.8. □

Corollary 9.10. *There are infinitely many primes.*

Proposition 9.11. *We have*

$$\sum_{p \le x} \frac{1}{p} > \log\log x - 1.$$

Proof. We have

$$\begin{aligned}\log \mathcal{P}_x &= \sum_{p \le x} \log\left(1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots + \frac{1}{p^k} + \dots\right)\\ &< \sum_{p \le x} \frac{1}{p} + \sum_{p \le x} \frac{1}{p(p-1)},\end{aligned}$$

on applying Lemma 9.6 with $t = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots + \frac{1}{p^k} + \dots$, and summing the GP, starting with the $1/p^2$ term,

$$\begin{aligned}&< \sum_{p \le x} \frac{1}{p} + \sum_{n=1}^{\infty} \frac{1}{(n+1)n}\\ &= \sum_{p \le x} \frac{1}{p} + \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right)\\ &= \sum_{p \le x} \frac{1}{p} + 1,\end{aligned}$$

because of the telescoping of $\sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right)$. Hence

$$\sum_{p \le x} \frac{1}{p} > \log \mathcal{P}_x - 1 > \log\log x - 1,$$

using Prop. 9.9. □

Proposition 9.12. *We have*

$$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1) \qquad \textit{as } x \to \infty.$$

Proof. Now from Problem Sheet 1, Q8, we have

$$n! = \prod_{p \le n} p^{\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \cdots},$$

so that (taking logs)

$$\begin{aligned} \log(n!) &= \sum_{p \le n} \left(\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \dots \right) \log p \\ &= \sum_{p \le n} \left\lfloor \frac{n}{p} \right\rfloor \log p + S_n, \end{aligned}$$

where

$$\begin{aligned} S_n :&= \sum_{p \le n} \left(\left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \dots \right) \log p \\ &\le \sum_{p \le n} \left(\frac{n}{p^2} + \frac{n}{p^3} + \dots \right) \log p \\ &= n \sum_{p \le n} \frac{\log p}{p(p-1)} \\ &< n \sum_{k=1}^{\infty} \frac{\log(k+1)}{(k+1)k} \\ &= nc, \end{aligned}$$

for some positive constant c, since the last sum is convergent. Hence $nc > S_n > 0$. Also, for $n = \lfloor x \rfloor$ we have

$$\begin{aligned} n \sum_{p \le x} \frac{\log p}{p} &\ge \sum_{p \le x} \left\lfloor \frac{n}{p} \right\rfloor \log p \\ &> \sum_{p \le x} \left(\frac{n}{p} - 1 \right) \log p \\ &= n \sum_{p \le x} \frac{\log p}{p} - \theta(x). \end{aligned}$$

Hence

$$n \sum_{p \le x} \frac{\log p}{p} \ge \sum_{p \le x} \left\lfloor \frac{n}{p} \right\rfloor \log p > n \sum_{p \le x} \frac{\log p}{p} - O(x),$$

since $\theta(x) = O(x)$, by Theorem 9.2. Now add the inequality $nc > S_n > 0$ to the above inequality, to obtain

$$n \sum_{p \le x} \frac{\log p}{p} + nc > \log(n!) > n \sum_{p \le x} \frac{\log p}{p} - O(x).$$

Dividing by n, and using the fact that $\frac{\log(n!)}{n} = \log n - O(1)$ from Prop. 9.7, we have

$$\sum_{p \le x} \frac{\log p}{p} + O\left(1\right) > \log n - O\left(1\right) > \sum_{p \le x} \frac{\log p}{p} - O\left(1\right).$$

Hence

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

□

9.5. **The average size of the divisor function $\tau(n)$.** The following result is a way of saying that an integer n has $\log n + 2\gamma - 1$ divisors, on average. Recall that $\tau(n)$ is the number of (positive) divisors of n.

Proposition 9.13. *We have, as $x \to \infty$, that*

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

Proof. Now

$$\begin{aligned}
\sum_{n \le x} \tau(n) &= \sum_{n \le x} \sum_{\ell | n} 1 \\
&= \sum_{\ell \le x} \sum_{\substack{n = k\ell \\ k \le \frac{x}{\ell}}} 1 \\
&= \sum_{\ell \le x} \left\lfloor \frac{x}{\ell} \right\rfloor,
\end{aligned}$$

on recalling that $\lfloor y \rfloor$ is the number of positive integers $\le y$,

$$\begin{aligned}
&= 2 \sum_{\ell \le \sqrt{x}} \left\lfloor \frac{x}{\ell} \right\rfloor - \left\lfloor \sqrt{x} \right\rfloor^2 && \text{by Q10, Problem Sheet 1} \\
&= 2 \sum_{\ell \le \sqrt{x}} \frac{x}{\ell} - x + O(\sqrt{x}) \\
&= 2x \left(\log \sqrt{x} + \gamma + O\left(\frac{1}{\sqrt{x}} \right) \right) - x + O(\sqrt{x}) && \text{using Prop. 9.8} \\
&= x \log x + (2\gamma - 1)x + O(\sqrt{x}).
\end{aligned}$$

□