LECTURE 7: SUMMARY

We started by recalling the notation $\mathrm{ord}_p(n)$, where p is prime and $n \in \mathbb{N}$. We defined it to be the largest integer k such that $p^k \mid n$. In other words, $\mathrm{ord}_p(n)$ is the unique integer satisfying

$$p^{\mathrm{ord}_p(n)} \mid n \qquad \text{and} \qquad p^{\mathrm{ord}_p(n)+1} \nmid n.$$

We also noted that this gives an upper bound on $\mathrm{ord}_p(n)$: we have

$$p^{\mathrm{ord}_p(n)} \mid n \Longrightarrow p^{\mathrm{ord}_p(n)} \leq n$$

whence $\mathrm{ord}_p(n) \leq \frac{\log n}{\log p}$. (As will always be the case in this course, $\log x$ denotes the natural logarithm of x.)

Next, we reconsidered Euclid's proof of the infinitude of primes. What does it tell us about the number of primes below x? For convenience, we gave this quantity a name:

$$\pi(x) := \{p \leq x : p \text{ is prime}\}.$$

Note that this function is well-defined even when x is not an integer. For example, $\pi(4.7) = 2$, since there are precisely two primes below 4.7 (namely, 2 and 3).

For the time being, let $p_1, p_2, p_3, \ldots$ denote the sequence of all primes in increasing order (e.g. $p_1 = 2, p_2 = 3$, etc). From Euclid's proof, we know that there exists a prime $p \mid (p_1 p_2 \cdots p_{n-1} + 1)$, and that this prime isn't p_1, p_2, ..., or p_{n-1}. It follows that $p_n \leq p$ (since p_n is the smallest prime which is different from $p_1, \ldots, p_{n-1}$). And $p \leq p_1 p_2 \cdots p_{n-1} + 1$. Thus, we conclude that

$$p_n \leq p_1 p_2 \cdots p_{n-1} + 1.$$

This is a pretty terrible bound, as a few examples demonstrate. Nonetheless, it's good enough to prove something about the growth of primes. First, we have the following result.

Theorem 1. *Let p_n denote the nth largest prime, where $p_1 = 2$. Then*

$$p_n \leq 2^{2^{n-1}} \quad \forall n \in \mathbb{N}.$$

Proof. We proceed by (strong) induction. The theorem clearly holds for $n = 1$, which will serve as our base case. Next, suppose the bound holds for $p_1, p_2, \ldots, p_{n-1}$. Then

$$\begin{aligned} p_n &\leq p_1 p_2 \cdots p_{n-1} + 1 \\ &\leq 2^{2^0 + 2^1 + \cdots + 2^{n-2}} + 1 \\ &= 2^{2^{n-1}-1} + 1 \\ &\leq 2^{2^{n-1}}. \end{aligned}$$

□

Corollary 2. $\pi(x) \geq \log\log x$ *for all* $x \geq 2$.

Date: January 29th, 2013.

Proof. Given $x \geq 2$, there exists $n \in \mathbb{N}$ such that $p_n \leq x < p_{n+1}$. It follows that $\pi(x) = n$. By the preceding theorem, we have

$$x < p_{n+1} \leq 2^{2^n} = 2^{2^{\pi(x)}}.$$

Taking logs twice and simplifying gives

$$\pi(x) > \frac{\log\log x - \log\log 2}{\log 2}.$$

Note that $\log 2 < 1$, whence $\log\log 2 < 0$. It follows immediately from above that

$$\pi(x) > \log\log x$$

as claimed. □

One immediate question is: why the lower bound $\log\log x$ when we've actually proved something stronger? Well, first of all, both this and the sharper lower bound we discovered during the course of the proof ($\frac{\log\log x - \log\log 2}{\log 2}$) are extremely weak, so quibbling about which one to use is irrelevant. Moreover, $\log\log x$ is more aesthetically pleasing.

The point of the above corollary is to quantify Euclid's theorem; now we know not only that there are infinitely many primes, but also something about how they're distributed. However, much more is known. When he was a teenager, Gauss conjectured that

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

Here the notation $\sim$ (read: 'is asymptotic to') has a precise meaning:

$$f(x) \sim g(x) \iff \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

Another way to think about this:

$$f(x) \sim g(x) \iff f(x) = g(x) + \mathrm{Err(x)}, \text{ where } \lim_{\mathrm{x}\to\infty} \frac{\mathrm{Err(x)}}{\mathrm{g(x)}} = 0.$$

Later on, Gauss made a more precise conjecture:[1]

$$\pi(x) \sim \int_2^x \frac{dt}{\log t}.$$

This relation has played a key role in number theoretic investigations ever since. One milestone occurred a century after Gauss' original conjecture, when Hadamard and de la Vallée Poussin (independently) verified its validity:

Theorem 3 (Prime Number Theorem). $\pi(x) \sim \displaystyle\int_2^x \frac{dt}{\log t}$.

To do this, they completed an outline set down by Riemann four decades earlier. This paper, Riemann's only one on number theory, contained a number of conjectures. All of these have been proved but one: this is the notorious Riemann Hypothesis, which is a more precise version of the Prime Number Theorem:

[1] Since $\int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}$, it is not immediately clear why one conjecture is more precise than the other. The difference lies in the error term, which is much smaller in Gauss' second approximation than in the first.

Conjecture 4 (Riemann Hypothesis). *For every $\epsilon > 0$, there exists a positive constant C_ϵ such that*

$$\left|\pi(x) - \int_2^x \frac{dt}{\log t}\right| < C_\epsilon\, x^{1/2+\epsilon}$$

for all $x \geq 2$.

This is widely considered the most outstanding problem in all of mathematics today. Its resolution would have enormous consequences. Unfortunately, any such result is quite far from being known. In fact, it is not currently known whether there exists *any* $\alpha < 1$ such that

$$\left|\pi(x) - \int_2^x \frac{dt}{\log t}\right| < x^\alpha.$$

We will return to this subject next lecture.

We finished the lecture with a different approach to proving the infinitude of primes. The approach begins with the observation that starting with a few primes and multiplying them together in all possible ways generates very few integers, even if we allow reusing each prime an arbitrary number of times. We first treat an easy case of this. Suppose we have two primes p_1 and p_2. (This notation no longer indicates that these are the smallest two primes; the p_i are simply arbitrary primes.) Fix a huge number x. How many natural numbers smaller than x are generated by the p_i's? In other words, how many numbers up to x are of the form $p_1^a p_2^b$? We have the following calculation:

$$\begin{aligned}\left|\left\{(a,b) \in \mathbb{Z}^2 : a, b \geq 0 \text{ and } p_1^a p_2^b \leq x\right\}\right| &= \sum_{\substack{a\geq 0 \\ p_1^a p_2^b \leq x}} \sum_{b \geq 0} 1 \leq \sum_{\substack{a \geq 0 \\ p_1^a \leq x}} \sum_{\substack{b \geq 0 \\ p_2^b \leq x}} 1 \\ &= \left(\sum_{\substack{a \geq 0 \\ p_1^a \leq x}} 1\right)\left(\sum_{\substack{b \geq 0 \\ p_2^b \leq x}} 1\right) \\ &\leq \left(\frac{\log x}{\log p_1} + 1\right)\left(\frac{\log x}{\log p_2} + 1\right)\end{aligned}$$

Since $p_1 \geq 2$ and $p_2 \geq 3$, we conclude that

$$\begin{aligned}\left|\left\{(a,b) \in \mathbb{Z}^2 : a, b \geq 0 \text{ and } p_1^a p_2^b \leq x\right\}\right| &\leq \left(\frac{\log x}{\log 2} + 1\right)\left(\frac{\log x}{\log 3} + 1\right) \\ &\leq (2\log x)(\log x) \qquad \text{for all sufficiently large } x \\ &= 2(\log x)^2.\end{aligned}$$

This is a very small proportion of the number less the x: it is a calculus exercise to prove that $\frac{2(\log x)^2}{x} \to 0$ as $x \to \infty$, so we might even say that 0% of all integers are generated by any two fixed primes.

With a bit more notation, one can generalize the above procedure to prove the following:

Theorem 5. *Any finite set of primes generates 0% of all integers. More precisely, for all sufficiently large x, the set of primes $\{p_1, p_2, \ldots, p_k\}$ generate at most $2(\log x)^k$ of the natural numbers $n \leq x$.*

It follows that there must be infinitely many primes; the Fundamental Theorem of Arithmetic guarantees that the set of all primes generates every integer, and the above theorem shows that any finite set of primes generates very few integers.