Last time: We proved that the n^{th} prime number is less than $2^{2^{n-1}}$ for all $n \geq 1$. What does this tell us about the rate at which primes tend to infinity?

Let $\pi(x) := \#\{p \leq x\}$. Examples: $\pi(10) = 4$, $\pi(10.2) = 4$. If we graph $\pi(x)$ for small numbers, it looks really jagged. But if you zoom out it looks super smooth. What smooth function does it look like? The Prime Number Theorem answers this. The first, simpler answer is: it looks like $x/\log x$. A more refined answer: it looks like a certain integral, called the Logarithmic Integral (or $\mathrm{Li}(x)$ for short). Here are some pictures, courtesy of Robert Lemke Oliver.

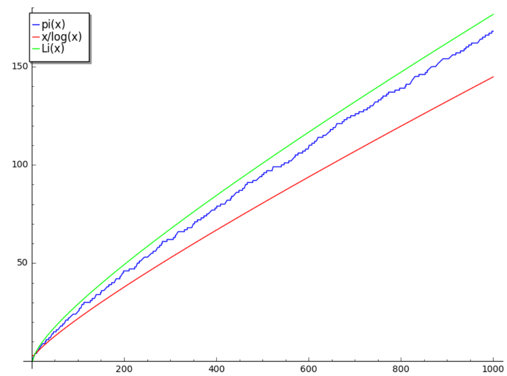

Figure 1: $\pi(x)$ is bumpy

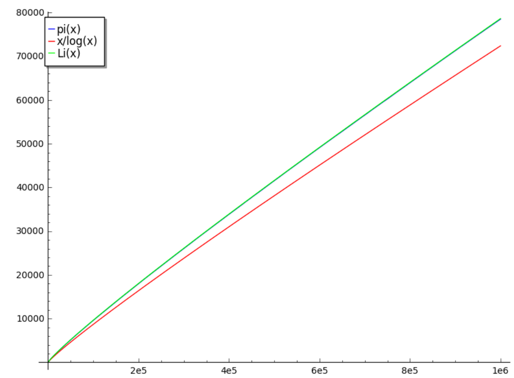

Figure 2: I swear the blue line is there

We'll come back to these ideas soon. First, we observe that we can leverage our result from last class to say something about the growth of $\pi(x)$. Given $x \geq 2$, there exists some prime $p_n \leq x < p_{n+1}$ (i.e. p_n is the largest prime less than or equal to x). Then $\pi(x) = n$. Using our bound, $x < p_{n+1} \leq 2^{2^n} = 2^{2^{\pi(x)}}$. Thus $x < 2^{2^{\pi(x)}}$, from which it follows that $\pi(x) > \log_2(\log_2(x))$.

Note: no mathematician uses $\log_2$ or even $\log_{10}$. The natural log to use is $\log_e = \ln$, which is well behaved because its derivative is $\frac{1}{x}$. From here on out, we'll write log to denote the natural log. How do we rewrite our above bound in terms of log rather than $\log_2$? It turns out that, with the proper notation, this is a simple task.

<u>Definition:</u> We write $f(x) \ll g(x)$ iff there exists some constant $C \geq 0$ such that $|f(x)| \leq C \cdot g(x)$ for all sufficiently large x.

Thus, for example,

$$\pi(x) > \log_2(\log_2(x)) = \frac{\log(\log_2(x))}{\log(2)} = \frac{\log(\frac{\log(x)}{\log(2)})}{\log(2)} \gg \log\log x.$$

More examples using $\ll$:

$$\begin{aligned} x &\ll 10x + 2 \\ 10x + 2 &\ll x \\ \sin x &\ll x \\ \sin x &\ll 1 \end{aligned}$$

Note in particular (from the last example) that $f(x) \ll g(x)$ does *not* mean that $f(x)/g(x)$ tends to a limit; it merely means that the quantity $f(x)/g(x)$ is bounded for all large x.

Note that $\pi(x) \gg \log\log x$ is a pretty terrible bound; $\log\log x$ grows *super* slowly. (e.g. $\log\log 10^{70} \approx 5$). In fact even just $\log x$ grows pretty slowly:

Claim: $\log x \leq x$ for all $x \geq 1$.
Proof: Observe that $\log x$ grows more slowly than x. To see this consider their derivatives, $\frac{1}{x}$ and 1 respectively. So when $x \geq 1$, $\log x$ has a small derivative than x. At $x = 1$, $\log 1 = 0 \leq 1 = x$. So $\log x$ starts smaller than x *and* it grows more slowly, so $\log x \leq x$ for all $x \geq 1$. □

Claim: $\log(x) \leq \sqrt{x}$ for all $x \geq 4$.
Proof (Ben): The derivative of $\log x$ is $\frac{1}{x}$ and the derivative of $\sqrt{x}$ is $\frac{1}{2\sqrt{x}}$. When does $\log x$ grow slower than $\sqrt{x}$?

$$\frac{1}{x} \leq \frac{1}{2\sqrt{x}} \quad\iff\quad 2 \leq \frac{x}{\sqrt{x}} = \sqrt{x} \quad\iff\quad 4 \leq x.$$

Thus for all $x \geq 4$, we see that $\log x$ grows more slowly than $\sqrt{x}$. Moreover, *at* $x = 4$ we have

$$\log(4) = \log 2^2 = 2\log 2 < 2 = \sqrt{4}.$$

Thus $\log x$ starts smaller than $\sqrt{x}$ and grows more slowly, too, so it can never catch up. This proves the claim. □

Note that our new notation allows us to be lazier: we can just write $\log x \ll \sqrt{x}$, and note worry about where the precise inequality begins. Similarly, we can prove that $\log x \ll x^{1/1000}$. An even stronger statement is that $\log x = o(x^{1/1000})$, where the 'little-oh' notation is defined as follows:

Definition: $f(x) = o(g(x))$ if and only if $\lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 0$.
Examples: $x = o(x^2)$ and $\frac{1}{x} = o(1)$.

We now return to our question from before: how does $\pi(x)$ grow?

Prime Number Theorem: $\pi(x) \sim \frac{x}{\log(x)}$.

Definition: We say $f(x) \sim g(x)$ ("$f(x)$ is asymptotic to $g(x)$") if and only if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Gauss guessed that there's a better version of this:

Prime Number Theorem v2.0: $\pi(x) \sim \int_2^x \frac{dt}{\log(t)}$.

(See the pictures above to see why this is a better estimate.)

Proving the Prime Number Theorem is quite hard, as evidenced by the fact that Gauss never managed to do it. (In fact, it took another century after Gauss' conjecture before it was proved (independently) by Hadamard and de la Vallée Poussin in 1896.) Among other things, the proof requires complex analysis (think 'calculus with complex numbers'). Although we won't prove it in this class, we'll be able to prove something almost as good:

Theorem(Chebyshev): $\frac{x}{\log(x)} \ll \pi(x) \ll \frac{x}{\log(x)}$

This seems similar to prime number theorem (it is!), but it's weaker: it just says $\dfrac{\pi(x)}{x/\log x}$ is bounded between two constants, whereas the prime number theorem says that the quotient actually tends to 1. Today we'll make good progress towards the upper bound:

Claim: $\pi(x) \ll \frac{x}{\log(x)}$.
Proof: (Chebyshev's proof, with a flourish by Ramanujan)
We will analyze $\binom{n}{k}$ = the number of ways to choose k objects from n objects. Recall the formula $\binom{n}{k} = \frac{n!}{(n-k)!k!}$, and its connection to the Binomial Theorem:

$$(a+b)^n = (a+b)(a+b)\cdots(a+b) = \sum_{j=0}^{n} \binom{n}{k} a^k b^{n-k}$$

because we're picking k brackets to select a from, and then selecting b's from all the rest.

Let's look at $\binom{2n}{n}$. What can you say about the prime factorization of $\binom{2n}{n}$? We stared at the formula for a bit before having an idea:

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2}$$

Max and Miranda conjectured that 2 will divide $\binom{2n}{n}$, because there should be a lot more factors of 2 in the numerator than in the denominator. We were unable to prove this, but it inspired Oliver to notice that all the primes between n and $2n$ will definitely not get cancelled, so they will divide $\binom{2n}{n}$. More formally, for all primes p such that $n < p \leq 2n$ we have $p \mid \binom{2n}{n}$, i.e. $\nu_p(\binom{2n}{n}) \geq 1$ whenever $n < p \leq 2n$. Thus we deduce

$$\prod_{n<p\leq 2n} p \,\Big|\, \binom{2n}{n}$$

The LHS is about primes, while the RHS isn't, which is good! To make the situation even nicer, we observe that

$$\binom{2n}{n} \leq 2^{2n}.$$

Indeed, we have

$$(1+1)^{2n} = \binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{2n} \geq \binom{2n}{n}.$$

Putting everything together, we have

$$\prod_{n<p\leq 2n} p \leq 2^{2n}.$$

Now products are hard to manipulate but sums are easier, so we apply a log:

$$\prod_{n<p\leq 2n} p \leq 2^{2n} \implies \log(\prod_{n<p\leq 2n} p) \leq \log(2^{2n}) \implies \sum_{n<p\leq 2n} \log p \leq 2n\log 2 \ll n.$$

Next we derive a lower bound on the LHS of this inequality. Note that for every prime $p \in (n, 2n]$ we have $\log p \geq \log n$. It follows that

$$n \gg \sum_{n<p\leq 2n} \log p \geq \sum_{n<p\leq 2n} \log n = \big(\pi(2n) - \pi(n)\big)\log n.$$

Thus, we deduce the bound

$$\pi(2n) - \pi(n) \ll \frac{n}{\log n}.$$

Looking back at the claim, we see that this looks pretty good! Next time we'll quickly deduce the claim from this bound, and then tackle the lower bound in Chebyshev's theorem.