CHEBYSHEV'S THEOREM AND BERTRAND'S POSTULATE

LEO GOLDMAKHER

ABSTRACT. In 1845, Joseph Bertrand conjectured that there's always a prime between n and $2n$ for any integer $n > 1$. This was proved less than a decade later by Chebyshev; much more importantly, Chebyshev was led to prove the first good approximation to the prime number theorem. My goal in this essay is to give a motivated (albeit anachronistic) account of Chebyshev's ideas.

1. INTRODUCTION AND CONTEXT [1 PAGE]

One of the foundational results of number theory is the Prime Number Theorem, conjectured privately by Gauss in the 1790s and publicly by Dirichlet in the 1830s:

Theorem 1.1 (Prime Number Theorem). *Let $\pi(x)$ denote the number of primes $\leq x$, and set* $\mathrm{Li}(x) := \displaystyle\int_2^x \frac{dt}{\log t}$ *(this is called the* logarithmic integral*). Then*

$$\lim_{x\to\infty} \frac{\pi(x)}{\mathrm{Li}(x)} = 1.$$

Remark. *Written in standard analytic number theory notation, this result would be expressed $\pi(x) \sim \mathrm{Li}(x)$. It's also worth noting that $\mathrm{Li}(x) \sim \frac{x}{\log x}$, so the PNT could be written in the simpler form $\pi(x) \sim \frac{x}{\log x}$. The disadvantage of this is that $\frac{\pi(x)}{x/\log x}$ tends to 1 much more slowly than $\frac{\pi(x)}{\mathrm{Li}(x)}$ does. In other words, both $\frac{x}{\log x}$ and $\mathrm{Li}(x)$ are eventually good approximations to $\pi(x)$, but the latter is much better.*[1]

The proof of the Prime Number Theorem turned out to be unexpectedly difficult, and resisted all attempts until the 1890s when Hadamard and de la Vallée Poussin (independently building on a fundamental 1859 paper of Riemann's) succeeded in proving PNT.

These days, many proofs of the prime number theorem are known. The original approach is conceptually the simplest and yields quite a bit more information than is stated in the theorem (namely, the rate of the convergence to 1), but involves complex analysis and some serious technical work. Wiener and Ikehara, building on earlier ideas of Landau, were able to shorten the proof and replace most of the complex analysis by Fourier analysis, but at the cost of losing control of the rate of convergence. This was later massively simplified by D. J. Newman, whose proof has been described very succinctly by Zagier (it rests on some very clever bounds combined with a result of Landau's in complex analysis). In the 1940s Erdős and Selberg independently came up with an approach that uses only elementary tools, but is also qualitative (no rate of convergence). Since then there have been some elementary approaches developed that are quantitative, but all of these are quite technical.

In short, none of the proofs of the prime number theorem are easy; there are short proofs that are fairly opaque and don't yield good quantitative results, and conceptually beautiful proofs that involve a lot of difficult estimates. However, half a century before the prime number theorem was first proved, Chebyshev was able to obtain some results that are almost as good – and whose proofs are conceptual and not too technical.

2. CHEBYSHEV'S THEOREM [2 PAGES]

Chebyshev's first result was that if $\frac{\pi(x)}{x/\log x}$ tends to a limit, then that limit *must* be 1. The only problem was, he couldn't prove that the limit exists! Still, he was able to prove that $\pi(x)$ has the correct order of growth:

Date: September 25, 2019.

[1]How good is the approximation? Riemann conjectured that $|\pi(x) - \mathrm{Li}(x)| = O(x^{1/2+\epsilon})$. This guess, now called the *Riemann Hypothesis*, is one of the most famous open problems in mathematics.

Theorem 2.1 (Chebyshev, 1850)**.** *There exist constants $a, b > 0$ such that*

$$\frac{ax}{\log x} \leq \pi(x) \leq \frac{bx}{\log x} \qquad \forall x \geq 2.$$

To prove something like this, Chebyshev needed to find some quantity that we understand independently of knowing about primes, but which also can be nicely expressed in terms of primes. Chebyshev's main insight was that the central binomial coefficients fit the bill. Before stating his results, it's important to build up some intuition with examples.

Exercise 1. *The prime factorization of $\binom{8}{4}$ is*

$$\binom{8}{4} = \frac{8 \times 7 \times 6 \times 5}{4!} = 2 \times 5 \times 7.$$

Determine, by hand, the prime factorization of $\binom{2n}{n}$ for $n = 5, 6, 7, 8, 9, 10$ and 16. What patterns do you observe in the factorization?

Our main goal below will be to quantify the patterns you observed, and to prove them. To fully appreciate the results, it is essential that you complete the exercise above before reading further.

Using central binomial coefficients to deduce something about primes is doomed to fail unless we can say something about $\binom{2n}{n}$ that has nothing to do with its prime factorization. At a minimum, it would be nice to know roughly how large it is.

Exercise 2. *Prove that* $\dfrac{4^n}{2n} \leq \dbinom{2n}{n} \leq 4^n$ *for all positive integers n.* [Hint. Consider $(1+1)^{2n}$.]

Remark. *It turns out the truth is roughly the geometric mean of these two bounds: Stirling's formula implies the asymptotic $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$. Working harder, one can prove the explicit bounds*

$$\frac{2\sqrt{\pi}}{e^2} \cdot \frac{4^n}{\sqrt{n}} \leq \binom{2n}{n} \leq \frac{e}{\pi\sqrt{2}} \cdot \frac{4^n}{\sqrt{n}}$$

but for our purposes, the weaker bounds of exercise 2 will suffice.

Now that we have some rudimentary understanding of how $\binom{2n}{n}$ behaves, we can study its prime factorization. As a first step, it will be helpful to develop some notation about prime factorizations in general. Any positive integer k has a unique factorization into prime powers, say,

$$k = \prod_p p^{\nu_p(k)}$$

In words, $\nu_p(k)$ is the exponent of the largest power of p dividing k.

Example 1. *The prime factorization of 50 is $50 = 2 \times 5^2$, so $\nu_2(50) = 1$, $\nu_3(50) = 0$, $\nu_5(50) = 2$, and $\nu_p(50) = 0$ for all primes $p \geq 7$.*

Thus understanding the prime factorization of $\binom{2n}{n}$ is equivalent to understanding $\nu_p\left(\binom{2n}{n}\right)$.

Exercise 3. *Prove that $\nu_p\left(\binom{2n}{n}\right) = 0$ for all $p > 2n$. (Qualitatively: $\binom{2n}{n}$ has no huge prime factors.)*

Exercise 4. *Prove that $\nu_p\left(\binom{2n}{n}\right) = 1$ for all $p \in (n, 2n]$. (Qualitatively: every large prime appears exactly once in the factorization of $\binom{2n}{n}$.)*

Since $\binom{2n}{n} = \frac{(2n)!}{(n!)^2}$, studying the prime factorization of the central binomial coefficient leads naturally to studying the prime factorization of factorials. Remarkably, there is an exact formula one can write down for $\nu_p(k!)$; this is the goal of the next exercise, and plays a fundamental role in the proof of Chebyshev's lower bound.

Exercise 5. *Prove that*

$$\nu_p(k!) = \sum_{j\geq 1} \left\lfloor \frac{k}{p^j} \right\rfloor .$$

Hint. Consider the following demonstration that $\nu_2(19!) = 16$:

1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***
	✓		✓		✓		✓		✓		✓		✓		✓		✓	
			✓				✓				✓				✓			
							✓								✓			
															✓			

Exercise 6. *Here we'll show that every prime power appearing in the factorization of $\binom{2n}{n}$ is small.*

(a) Consider the function $f(x) := \lfloor 2x \rfloor - 2\lfloor x \rfloor$. Describe the behavior of f as explicitly and simply as possible. (You might wish to graph it first.)

(b) Prove that every prime power appearing in the factorization of $\binom{2n}{n}$ is no larger than $2n$. In other words, prove that whenever $p^k \mid \binom{2n}{n}$, it must follow that $p^k \leq 2n$. [This is not intended to be obvious!]

The exercises above make precise your earlier empirical observations about the prime factorization of $\binom{2n}{n}$: small primes (less than n) appear with small exponent by exercise 6(b); larger primes (between n and $2n$) each appears with exponent 1; and huge primes (larger than $2n$) never appear.

It turns out that this information is enough to count primes with remarkable precision. Indeed, we deduce

$$\frac{2^{2n}}{2n} \leq \binom{2n}{n} = \prod_{p\leq 2n} p^{\nu_p} \leq \prod_{p \leq 2n} 2n = (2n)^{\pi(2n)},$$

and applying the natural logarithm to both sides yields

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

Not only does this look like the lower bound in Chebyshev's theorem 2.1, it implies it:

Exercise 7. *Prove that there exists a constant $a > 0$ such that $\pi(x) \geq a\frac{x}{\log x}$ for all real numbers $x \geq 2$.*

Our work on the factorization of $\binom{2n}{n}$ also yields the upper bound in Chebyshev's theorem. First, observe

$$4^n \geq \binom{2n}{n} \geq \prod_{p\in(n,2n]} p \geq \prod_{p \in (n,2n]} n = n^{\pi(2n)-\pi(n)}, \tag{2.1}$$

whence

$$\pi(2n) - \pi(n) \leq \frac{n \log 4}{\log n}$$

for all integers $n \geq 2$. This looks a lot like the upper bound in Chebyshev's theorem 2.1, in that it's bounding the number of primes in an interval of length n. Trouble is, it's not the interval we want: we've counted primes in $(n, 2n]$, whereas Chebyshev's theorem counts primes in $(0, n]$. This problem is surmountable:

Exercise 8. *The goal of this exercise is to deduce the upper bound in Chebyshev's theorem.*

(a) Prove that there exists a constant c such that $\pi(2x) - \pi(x) \leq c \cdot \frac{x}{\log x}$ for all real numbers $x \geq 2$.

(b) Given a real number $x \geq 2$, consider the sequence ${}^x/_2, {}^x/_4, {}^x/_8, \ldots$ Let ℓ be the largest positive integer such that ${}^x/_{2^\ell} > \sqrt{x}$. Prove that

$$\pi({}^x/_{2^{\ell+1}}) \leq \sqrt{x} \qquad \textit{and} \qquad \frac{1}{2} < \frac{\log({}^x/_{2^k})}{\log x} \leq 1 \quad \forall k \leq \ell.$$

(c) Prove that there exists a constant b such that $\pi(x) \leq b \cdot \frac{x}{\log x}$ for all real numbers $x \geq 2$. [Hint: Use part (a) a bunch of times, keeping part (b) in mind.]

This exercise concludes the proof of Chebyshev's theorem.

Exercise 9. *The goal of this exercise is to make Chebyshev's theorem 2.1 completely explicit, by determining admissible choices for the constants a and b.*

(a) Prove that $\pi(x) \geq \frac{\log 2}{2} \cdot \frac{x}{\log x}$ for all $x \geq 2$.
(b) Prove that $\pi(2^k) \leq 3 \cdot \frac{2^k}{k}$ for all positive integers k. [Hint: Induction!]
(c) Deduce that $\pi(x) \leq (6 \log 2) \frac{x}{\log x}$ for all $x \geq 2$.

3. Bertrand's Postulate [1.5 pages]

Chebyshev's theorem 2.1 tells us quite a bit about how primes are distributed in the interval $(0, x]$. What can we deduce about the distribution of primes in other intervals of length x, e.g. the dyadic interval $(x, 2x]$?

Exercise 10. *Prove that if $\frac{a}{b} > 0.6$ in Chebyshev's theorem 2.1, then $(x, 2x]$ would contain at least one prime for every $x \geq 32$.*

Since one can manually verify that there's a prime in $(x, 2x]$ for every $x \in [1, 32]$, this exercise would imply the existence of a prime in every dyadic interval $(x, 2x]$ with $x \geq 1$.

Unfortunately, our proof of Chebyshev's theorem doesn't yield constants that are strong enough to satisfy the hypothesis of exercise 10 (see exercise 9). Still, some playing around will convince you that there really does some to be a prime in every dyadic interval. This was formally conjectured by Bertrand in 1845:

Theorem 3.1 (Bertrand's Postulate)**.** *For every positive integer n, there exists a prime in the interval $(n, 2n]$.*

This statement is sometimes also referred to as Chebyshev's theorem, since he proved it in 1852. The key idea is to refine our work above on the prime factorization of $\binom{2n}{n}$. For example, you might have noticed empirically that there's always a string of consecutive medium-sized primes missing from the prime factorization of $\binom{2n}{n}$. The following exercise makes this precise:

Exercise 11. *Our goal is to show that none of the primes in $(2n/3, n]$ appear in the factorization of $\binom{2n}{n}$.*

(a) Prove that $p^2 \geq 2n$ for all primes $p \in (2n/3, n]$.
(b) Prove that $\nu_p\left(\binom{2n}{n}\right) = 0$ for all $p \in (2n/3, n]$.

Summing up all our knowledge on the factorization of $\binom{2n}{n}$ thus far, we've shown that none of the primes larger than $2n$ appear, every prime in $(n, 2n]$ appears precisely once, and none of the primes in the interval $(2n/3, n]$ appear at all. What about primes smaller than $2n/3$? Turns out that, with minimal effort, we can say a bit more:

Exercise 12. *Prove that every prime in $(\sqrt{2n}, 2n/3]$ appears at most once in the prime factorization of $\binom{2n}{n}$.*

To prove Bertrand's postulate, we require one more refinement of our work above. Recall that in (2.1) we produced an upper bound on the product of all primes in a dyadic interval of length n:

$$\prod_{p \in (n, 2n]} p \leq 4^n.$$

We claim the same bound holds for the product over all primes in a different interval of length n:

Lemma 3.2. *For any positive integer n we have* $\prod_{p \in (0, n]} p \leq 4^n$.

False Proof. From (2.1) we have

$$\prod_{\frac{n}{2} < p \leq n} p \leq 4^{n/2} \quad , \quad \prod_{\frac{n}{4} < p \leq \frac{n}{2}} p \leq 4^{n/4} \quad , \quad \prod_{\frac{n}{8} < p \leq \frac{n}{4}} p \leq 4^{n/8} \quad , \quad \text{etc.}$$

Multiplying all these together yields the claim. □

Exercise 13. *Our goal is to prove the lemma. The key idea is to sharpen the estimate from (2.1).*

(a) What's the problem with the 'false proof' given above?

(b) Prove $\prod_{m<p\le 2m-1} p \le 4^{m-1}$ *for any integer* $m \ge 1$. [Hint: Consider the prime factorization of $\binom{2m-1}{m}$.]
(c) Prove Lemma 3.2. [Hint: Induction!]
(d) Prove that $\prod_{p\in(0,x]} p \le 4^x$ *for any real number* $x \ge 1$. [Hint: Your proof should be very short.]

We have now assembled all the tools required to prove Bertrand's postulate.

Theorem 3.3 (Bertrand's postulate)**.** *There's a prime in the interval* $(n, 2n]$ *for every integer* $n \ge 1$.

Proof. Given a positive integer n, suppose there were no primes in $(n, 2n]$. From our work above it would follow that all prime factors of $\binom{2n}{n}$ would have to be smaller than $2n/3$. Moreover, we've proved above that the primes larger than $\sqrt{2n}$ all have exponent at most 1, and that all prime powers appearing in the factorization of $\binom{2n}{n}$ are bounded by $2n$. Combining these observations with Exercise 13(d), we deduce

$$\frac{4^n}{2n} \le \binom{2n}{n} \le \prod_{p\le\sqrt{2n}} (2n) \times \prod_{\sqrt{2n}<p\le 2n/3} p \le (2n)^{\sqrt{2n}} \times 4^{2n/3}.$$

Taking logarithms yields

$$\frac{n}{3}\log 4 \le (1+\sqrt{2n})\log(2n),$$

which is clearly false for all sufficiently large n; a bit of calculus shows that this inequality fails for all $n \ge 468$. In particular, this proves that Bertrand's postulate holds for all integers $n \ge 468$. It's now a straightforward matter to verify that it also holds for all small positive integers; the most expedient way to do so is to note that each prime in the sequence

$$2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 631$$

is less than twice the previous entry. □

Exercise 14. *The goal of this exercise is to fill in the details above.*
(a) Carefully justify that Bertrand's postulate holds for all $n \ge 468$.
(b) Carefully justify that Bertrand's postulate holds for all $n < 468$.

Remark. *In exercise 9 we saw that* $\frac{\log 2}{2}\cdot\frac{x}{\log x} \le \pi(x) \le (6\log 2)\frac{x}{\log x}$. *It turns out that Chebyshev obtained much stronger constants: he proved that* $0.92\frac{x}{\log x} < \pi(x) < 1.11\frac{x}{\log x}$, *which immediately yields Bertrand's postulate via exercise 10.*

4. Final Remarks [0.5 pages]

What makes Chebyshev's approach successful? This type of question is always difficult to answer—it's successful because it works!—but one key feature is that the prime factorization of $\binom{2n}{n}$ essentially looks like the product of all the primes between n and $2n$, with some additional noise coming from small primes.[2] Can one find another quantity that, on one hand, looks roughly like a product of a string of primes, but on the other hand can be studied without any *a priori* knowledge about primes? We leave this open-ended question as a challenge to the reader, and conclude with one natural candidate: the general binomial coefficient. What can be said about the prime factorization of $\binom{n}{k}$? Following the general outline we used to prove Bertrand's postulate 3.1, one can show:

Theorem 4.1 (Sylvester, 1892)**.** *Whenever* $n \ge 2k$, *the binomial coefficient* $\binom{n}{k}$ *has a prime factor strictly larger than* k.

Proof. Exercise! □

Note that this subsumes Bertrand's postulate (take $n = 2k$).

Dept of Mathematics and Statistics, Williams College, Williamstown, MA, USA 01267
Email address: Leo.Goldmakher@williams.edu

[2] In view of our work above, this can be made precise: $\log\binom{2n}{n} = \sum_{n<p\le 2n}\log p + \sum_{\sqrt{2n}<p\le 2n/3}(0 \text{ or } 1)\log p + O(\sqrt{2n}\log 2n)$.