1 Chebyshev's Theorem

Today we prove some asymptotic results about the distribution of prime numbers. Specifically, we derive estimates for the *prime-counting functions*

$$\vartheta(x) = \sum_{p \le x} \ln(p)$$
$$\psi(x) = \sum_{p^k \le x} \ln(p)$$
$$\pi(x) = \sum_{p \le x} 1$$

Note that we will always use p to denote a prime.

Lacking the tools of complex analysis, it is difficult to find the exact asymptotic formulas; however, our elementary methods suffice to determine the asymptotics up to a constant multiple. Our main result is Chebyshev's Theorem:

Theorem 1.1: [2, Theorem 6.3] There exist positive constants c_1 and c_2 such that

$$c_1 x \le \vartheta(x) \le \psi(x) \le \pi(x)\ln(x) \le c_2 x. \tag{1}$$

for all $x \ge 2$. Moreover,

$$\liminf_{x\to\infty} \frac{\vartheta(x)}{x} = \liminf_{x\to\infty} \frac{\psi(x)}{x} = \liminf_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \ge \ln(2) \tag{2}$$
$$\limsup_{x\to\infty} \frac{\vartheta(x)}{x} = \limsup_{x\to\infty} \frac{\psi(x)}{x} = \limsup_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \le 2\ln(2) \tag{3}$$

We will prove this in three steps.

2 Comparing the three functions

Since all terms in the sum defining $\vartheta(x)$ are included in the sum defining $\psi(x)$, $\vartheta(x) \le \psi(x)$. For a given p there are $\left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor$ choices for k so that $p^k \le x$, so

$$\psi(x) = \sum_{p^k \le x} \ln(p) = \sum_{p \le x} \left\lfloor \frac{\ln(x)}{\ln(p)} \right\rfloor \ln(p) \le \sum_{p \le x} \ln(p) = \pi(x)\ln(x).$$

This shows the middle two inequalities in (1).

Given $\vartheta(x) \le \psi(x) \le \pi(x)\ln(x)$, to show that the three quantities in (2) and (3) are equal it suffices to show that

$$\liminf_{x\to\infty} \frac{\vartheta(x)}{x} \ge \liminf_{x\to\infty} \frac{\pi(x)\ln(x)}{x}, \qquad \limsup_{x\to\infty} \frac{\vartheta(x)}{x} \ge \limsup_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \tag{4}$$

To compare $\vartheta(x) = \sum_{p \le x} \ln(p)$ and $\pi(x)\ln(x) = \sum_{p \le x} \ln(x)$, note that for p "close" to x, we have $\ln(p)$ "close" to $\ln(x)$ and relatively large, while the terms for small p will not contribute much to either sum. Thus we can just consider the terms with $p > x^{1-\delta}$, where $\delta \in (0,1)$.

$$\begin{aligned}
\vartheta(x) &\geq \sum_{x^{1-\delta}<p\leq x} \ln(p) \\
&\geq \sum_{x^{1-\delta}<p\leq x} \ln(x^{1-\delta}) \\
&= \ln(x^{1-\delta})(\pi(x) - \pi(x^{1-\delta})) \\
&= (1-\delta)\ln(x)(\pi(x) - \pi(x^{1-\delta})) \\
&\geq (1-\delta)\ln(x)(\pi(x) - x^{1-\delta})
\end{aligned}$$

Hence

$$\frac{\vartheta(x)}{x} \geq \frac{(1-\delta)\pi(x)\ln(x)}{x} - \frac{(1-\delta)\ln(x)}{x^{\delta}}.$$

Letting $\delta \to 0$ gives (4).

3 Upper Bound

We show that $\vartheta(x) \leq 2x\ln(x)$. Instead of thinking about bounding $\vartheta(x)$, it is easier to think about bounding $e^{\vartheta(x)} = \prod_{p\leq x} p$.

Lemma 3.1: [1, 3.?] For any $x \in \mathbb{N}$,

$$\prod_{p\leq x} p \leq 4^{x-1} \tag{5}$$

Proof. Use strong induction on x. For $x = 1, 2$ the statement holds. The induction step from odd $x > 1$ to $x+1$ is obvious, since $x+1$ is not a prime.

Consider the induction step from even x to $x+1$. Let $x = 2n$. The key idea is that there cannot be "too many" primes between $n+2$ and $2n+1$, because...

1. These primes all divide $\binom{2n+1}{n} = \frac{(2n+1)!}{n!(n+1)!}$.

2. $\binom{2n+1}{n}$ can easily be bounded from above:

$$\binom{2n+1}{n} = \frac{1}{2}\left(\binom{2n+1}{n} + \binom{2n+1}{n+1}\right) \leq \frac{1}{2}\sum_{i=0}^{n}\binom{2n+1}{i} = 4^n.$$

Then

$$\prod_{p\leq x+1} p = \prod_{p\leq n+1} p \prod_{n+2\leq p\leq 2n+1} p \leq 4^n \cdot \binom{2n+1}{n} \leq 4^{2n}.$$

$\square$

Taking the logarithm of both sides of (5) gives $\vartheta(x) \leq (x-1)\ln(4) \leq 2x\ln(x)$.

4 Lower Bound

We show that $\liminf_{x\to\infty} \frac{\pi(x)\ln(x)}{x} \geq \ln(2)$. First consider when x is even, say equal to $2n$. Like in Section 3, we consider a binomial coefficient, this time $\binom{2n}{n}$. We show that each prime cannot appear as a factor in $\binom{2n}{n}$ "too many" times, so it can be bounded above by $(2n)^{\pi(2n)}$. We can easily bound $\binom{2n}{n}$ below:

$$\binom{2n}{n} \geq \frac{2^{2n}}{2n}$$

since it is the largest among $2, \binom{2n}{1}, \dots, \binom{2n}{2n-1}$. Putting these two bounds together will give the desired bound for $\pi(2n)$.

We need the following to count the highest prime powers dividing $\binom{2n}{n}$:

Lemma 4.1: [1, Lemma 6.3] For every positive integer n,

$$v_p(n!) = \sum_{k=1}^{\left\lfloor \frac{\ln(n)}{\ln(p)} \right\rfloor} \left\lfloor \frac{n}{p^k} \right\rfloor,$$

where $v_p(m)$ denotes the largest integer i such that $p^i | m$.

Proof. There are $\left\lfloor \frac{n}{p^k} \right\rfloor$ multiples of p^k less than or equal to n. In the sum $\sum_{k\geq 1} \left\lfloor \frac{n}{p^k} \right\rfloor$, each multiple of p^k less than n is counted k times, once each as a multiple of $p, p^2, \dots, p^k$. $\square$

From Lemma 4.1, we get

$$v_p\left(\binom{2n}{n}\right) = v_p\left(\frac{(2n)!}{n!^2}\right) = v_p((2n)!) - 2v_p(n!) = \sum_{k=1}^{\left\lfloor \frac{\ln(2n)}{\ln(p)} \right\rfloor} \left\lfloor \frac{2n}{p^k} \right\rfloor - 2\left\lfloor \frac{n}{p^k} \right\rfloor$$

Since each term of the sum is at most 1,

$$v_p\left(\binom{2n}{n}\right) \leq \left\lfloor \frac{\ln(2n)}{\ln(p)} \right\rfloor \leq \frac{\ln(2n)}{\ln(p)}.$$

Thus

$$\frac{2^{2n}}{2n} \leq \binom{2n}{n} = \prod_{p\leq 2n} p^{v_p\left(\binom{2n}{n}\right)} \leq (2n)^{\pi(2n)}.$$

Taking logs and remembering $x = 2n$ gives $x\ln(2) - \ln(x) \leq \pi(x)\ln(x)$, which gives the desired bound. For odd x, the value of $\frac{\pi(x)\ln(x)}{x}$ can be compared to the value for $x - 1$.

Finally, (2) and (3), and the fact that all the prime-counting functions are positive for $x \geq 2$, show the existence of c_1 and c_2 in (1). This finishes the proof of Theorem 1.1.

5 The nth prime

We found an estimate for the number of primes less than or equal to a given number; we can use this bound to find an estimate for the nth prime number.

Theorem 5.1: [2, Theorem 6.4] Let p_n denote the nth prime number. Then there exist constants c_3, c_4 such that

$$c_3 n \ln(n) \leq p_n \leq c_4 n \ln(n)$$

for all $n \geq 2$.

Proof. From Theorem 1.1,

$$\frac{c_1 p_n}{\ln(p_n)} \leq \pi(p_n) = n \leq \frac{c_2 p_n}{\ln(p_n)}, \tag{6}$$

so

$$\frac{n \ln(p_n)}{c_2} \leq p_n \leq \frac{n \ln(p_n)}{c_1}.$$

The LHS is at least $c_3 n \ln(n)$ by the trivial bound $n \leq p_n$. On the RHS, use the LHS of (6) again to get $\ln(p_n) \leq \ln\left(\frac{n \ln(p_n)}{c_1}\right)$, giving $\ln(p_n) \leq c \ln(n)$ for some c. $\square$

References

[1] Andreescu, T., Dospinescu, G.: "Problems from the Book" XYZ Press, TX, 2008.

[2] Nathanson, M.: "Additive Number Theory: The Classical Bases," Springer, NY, 1996, p. 153-158.