7. Distribution of primes

7.1 Elementary estimates for the function $\pi(x)$

Definition 7.1. *For $x > 0$, we denote by $\pi(x)$ the number of primes p such that $p \leq x$.*

The fundamental result on the distribution of prime numbers is the *prime number theorem* (PNT), which states that

$$\pi(x) \sim \frac{x}{\ln x}\,,$$

i.e. $\lim_{x\to\infty} \frac{\pi(x)}{x/\ln(x)} = 1$. This result was first conjectured by Gauss, and it was proved independently by Hadamard and de la Vallée Poussin in 1896.

An even better approximation for function $\pi(x)$ is the so-called *logarithmic integral function*

$$\mathrm{li}(x) = \int_2^x \frac{1}{\ln t}\, dt.$$

By L'Hospital's rule, we obtain

$$\lim_{x\to\infty} \frac{\mathrm{li}(x)}{x/\ln(x)} = 1.$$

Hence, the PNT is equivalent to $\pi(x) \sim \mathrm{li}(x)$.

We will first prove a weaker statement of the prime number theorem, which was proved by the Russian mathematician Pafnuti Lvovich Chebyshev (1821 – 1894), that there are positive real numbers a and b such that

$$a\frac{x}{\ln x} < \pi(x) < b\frac{x}{\ln x} \tag{7.1}$$

for x large enough.

The proof (following [354, Chapter 12.1]) will use the information on the prime factorization of binomial coefficients of the form $\binom{2n}{n}$.

Let us recall the definition of the binomial coefficient and the binomial theorem 1.2:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k},$$

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + y^n.$$

We will now see that Theorem 6.1 can help us in obtaining information on the distribution of prime numbers.

Lemma 7.1. *Let n be a positive integer.*

(i) $2^n \leq \binom{2n}{n} < 2^{2n}$

(ii) $\prod_{n<p\leq 2n} p$ *divides* $\binom{2n}{n}$
(the product runs through all prime numbers from the interval $(n, 2n]$).

(iii) Let $r(p) = \lfloor \log_p 2n \rfloor$. Then $\binom{2n}{n}$ divides $\prod_{p\leq 2n} p^{r(p)}$.

(iv) If $n > 2$ and $\frac{2n}{3} < p \leq n$, then p does not divide $\binom{2n}{n}$.

(v) $\prod_{p\leq n} p < 4^n$

Proof:

(i) Due to $2n - k \geq 2(n-k)$, we have

$$\binom{2n}{n} = \frac{2n}{n}\cdot\frac{2n-1}{n-1}\cdot\ldots\cdot\frac{n+1}{1} \geq 2^n.$$

Furthermore,

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

(ii) Let $p \in (n, 2n]$ be a prime number. Then p divides $(2n)!$, but it does not divide $n!$. Hence, p divides $\binom{2n}{n} = \frac{(2n)!}{n!\cdot n!}$.

(iii) The exponent of p in the prime factorization of $(2n)!$ is $\sum_{j=1}^{\infty}\left\lfloor\frac{2n}{p^j}\right\rfloor = \sum_{j=1}^{r(p)}\left\lfloor\frac{2n}{p^j}\right\rfloor$, and in the prime factorization of $n!$ is $\sum_{j=1}^{r(p)}\left\lfloor\frac{n}{p^j}\right\rfloor$. Therefore, the exponent of p in the prime factorization of $\binom{2n}{n}$ is

$$\sum_{j=1}^{r(p)}\left(\left\lfloor\frac{2n}{p^j}\right\rfloor - 2\left\lfloor\frac{n}{p^j}\right\rfloor\right) = \sum_{j=1}^{r(p)}\left(\left\lfloor\frac{n}{p^j}+\frac{1}{2}\right\rfloor - \left\lfloor\frac{n}{p^j}\right\rfloor\right) \leq \sum_{j=1}^{r(p)} 1 = r(p).$$

(Here, we used formula $\lfloor 2x\rfloor = \lfloor x+\frac{1}{2}\rfloor + \lfloor x\rfloor$ from Example 6.1.)

(iv) Let $\frac{2n}{3} < p \leq n$. Then $2p > n$ and $3p > 2n$, so p appears in the prime factorization of $n!$ with exponent $\left\lfloor\frac{n}{p}\right\rfloor = 1$, and in the factorization of $(2n)!$ with the exponent $\left\lfloor\frac{2n}{p}\right\rfloor = 2$. Thus, p appears in the factorization of $\binom{2n}{n}$ with the exponent $2-1-1=0$, i.e. p does not divide $\binom{2n}{n}$.

(v) We prove the statement by induction over n. For $n = 1, 2, 3$, by direct inspection, we can check that the statement holds. Assume now that $n \geq 4$ and that the statement holds for all positive integers less than n.

If n is even, say $n = 2m$, then n is not prime, so we have

$$\prod_{p\leq 2m} p = \prod_{p\leq 2m-1} p < 4^{2m-1} < 4^n.$$

Let n be odd, say $n = 2m+1$ with $m \geq 2$. Every prime number $p \in (m+1, 2m+1]$ divides $\binom{2m+1}{m+1} = \binom{2m+1}{m} = \frac{(2m+1)!}{m!\cdot(m+1)!}$, so we have

$$\prod_{p\leq 2m+1} p \leq \binom{2m+1}{m}\prod_{p\leq m+1} p < \binom{2m+1}{m}\cdot 4^{m+1}.$$

Furthermore, we have

$$\begin{aligned} 2^{2m+1} = (1+1)^{2m+1} &= 1 + \cdots + \binom{2m+1}{m} + \binom{2m+1}{m+1} + \cdots + 1 \\ &> 2\cdot\binom{2m+1}{m}, \end{aligned}$$

so we obtain

$$\prod_{p\leq 2m+1} p < 4^m\cdot 4^{m+1} = 4^{2m+1}. \qquad \square$$

Theorem 7.2. *For $n \geq 2$,*

$$\frac{n}{8 \ln n} < \pi(n) < \frac{6n}{\ln n}.$$

Proof: From Lemma 7.1, parts (ii) and (iii), it follows that

$$n^{\pi(2n)-\pi(n)} < \prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq \prod_{p\leq 2n} p^{r(p)} \leq (2n)^{\pi(2n)},$$

so from Lemma 7.1, part (i), we obtain

$$n^{\pi(2n)-\pi(n)} < 2^{2n} \quad \text{and} \quad 2^n \leq (2n)^{\pi(2n)}. \tag{7.2}$$

Let us now insert $n = 2^k$ in (7.2). We have

$$k(\pi(2^{k+1}) - \pi(2^k)) < 2^{k+1} \quad \text{and} \quad 2^k \leq (k+1)\pi(2^{k+1}).$$

It is clear that $\pi(2^{k+1}) \leq 2^k$ (even numbers greater than 2 are not prime), so we have

$$(k+1)\pi(2^{k+1}) - k\pi(2^k) < \pi(2^{k+1}) + 2^{k+1} \leq 3 \cdot 2^k. \tag{7.3}$$

Let us sum up relations (7.3) for $k = m, m-1, \ldots, 1, 0$. After cancelling ("telescoping"), we obtain

$$(m+1)\pi(2^{m+1}) \leq 3(2^m + 2^{m-1} + \cdots + 2^1 + 2^0) < 3 \cdot 2^{m+1}.$$

From this and (7.2), we conclude that

$$\frac{2^m}{m+1} \leq \pi(2^{m+1}) < \frac{3 \cdot 2^{m+1}}{m+1}. \tag{7.4}$$

Now, for the given integer $n \geq 2$, let $m = \lfloor \log_2 n \rfloor - 1$. Then $2^{m+1} \leq n < 2^{m+2}$. Let us also notice that for any $x > 0$, we have $\ln 2^x = x \ln 2 < x$ and $\ln 2^x > \frac{x}{2}$. Furthermore, from (7.4), we obtain

$$\pi(n) \leq \pi(2^{m+2}) < \frac{3 \cdot 2^{m+2}}{m+2} < \frac{6 \cdot 2^{m+1}}{\ln(2^{m+2})} < \frac{6n}{\ln n};$$

$$\pi(n) \geq \pi(2^{m+1}) > \frac{2^m}{m+1} = \frac{2^{m+2}}{8 \cdot \frac{m+1}{2}} > \frac{2^{m+2}}{8 \ln(2^{m+1})} > \frac{n}{8 \ln n}. \qquad \square$$

The following result, the so-called Bertrand's postulate, was conjectured by Bertrand (1845) and first proved by Chebyshev (1852). The proof given below follows the proof which was given by the Hungarian mathematician Paul Erdős (1913 – 1996) in [175].

Theorem 7.3 (Bertrand, Chebyshev)**.** *For every positive integer n, there is a prime number p such that $n < p \leq 2n$.*

Proof: For $n = 1, 2, 3$, the statement is evidently true: $1 < 2 \leq 2$, $2 < 3 \leq 4$, $3 < 5 \leq 6$. Assume that the statement does not hold for an integer $n > 3$. From Lemma 7.1.(iv), it follows that all prime factors of $\binom{2n}{n}$ satisfy $p \leq \frac{2n}{3}$. Let $p^{s(p)}$ be the largest power of p which divides $\binom{2n}{n}$. By Lemma 7.1.(iii), we have $p^{s(p)} \leq p^{r(p)} \leq 2n$. If $s(p) \geq 2$, then $p \leq \sqrt{2n}$, so at most $\lfloor \sqrt{2n} \rfloor$ prime numbers appear in the factorization of $\binom{2n}{n}$ with an exponent ≥ 2. Therefore,

$$\binom{2n}{n} \leq (2n)^{\lfloor \sqrt{2n} \rfloor} \cdot \prod_{p \leq \frac{2n}{3}} p.$$

Among all binomial coefficients $\binom{2n}{k}$, the largest is the middle one, i.e. $\binom{2n}{n}$. Now, from $2^{2n} = (1+1)^{2n} = 1 + \cdots + \binom{2n}{n} + \cdots + 1 < (2n+1)\binom{2n}{n}$, it follows that $\binom{2n}{n} > \frac{4^n}{2n+1}$. Therefore, by Lemma 7.1.(v),

$$\frac{4^n}{2n+1} < (2n)^{\lfloor \sqrt{2n} \rfloor} \cdot \prod_{p \leq \frac{2n}{3}} p < 4^{2n/3} \cdot (2n)^{\sqrt{2n}}.$$

From $2n + 1 < (2n)^2$, we obtain $4^{n/3} < (2n)^{2+\sqrt{2n}}$, i.e.

$$\frac{n \ln 4}{3} < (2 + \sqrt{2n}) \ln 2n.$$

However, the function $f(x) = \frac{x \ln 4}{3} - (2 + \sqrt{2x}) \ln 2x$ is increasing and positive, for x large enough ($f'(x) > 0$ for $x \geq 200$ and $f(507) > 0$ implies that $f(x) > 0$ for $x \geq 507$). By the obtained contradiction, we proved the theorem for $n \geq 507$. The statement of the theorem for $n < 507$, follows from the fact that in the following sequence of prime numbers

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

each member is less than the double of the previous member. □

Example 7.1. *Find all solutions of the equation $m! = x^2$ in positive integers.*

Solution: We claim that the only solution is $(m, x) = (1, 1)$. We have $1! = 1^2$, while $2! = 2$ and $3! = 6$ are not squares. Let $m \geq 4$. By Theorem 7.3, there is a prime number p such that

$$\frac{m}{2} < p \leq m. \tag{7.5}$$

Indeed, if m is even, say $m = 2n$, then there is a prime number p, such that $n < p \leq 2n$, so (7.5) holds, and if m is odd, say $m = 2n + 1$, then there is p such that $n < p \leq 2n$ and $n + 1 \leq p \leq 2n$, so due to $\frac{m}{2} < n + 1$, (7.5) holds again. However, then $m!$ is divisible by p, but it is not divisible by p^2 since $2p > m$, so $m!$ cannot be the square of an integer. □

The result from Example 7.1 is a special case of Erdős-Selfridge's theorem from 1975, which states that the product of consecutive integers is never a power of an integer, i.e. the equation

$$(n + 1)(n + 2) \cdot \ldots \cdot (n + k) = x^l$$

does not have solutions in integers for $k \geq 2$, $l \geq 2$ and $n \geq 0$. A proof of this theorem can be found in [323, Chapter 3.2].

7.2 Chebyshev functions

In this section, we will study some functions connected with the distribution of prime numbers and give an alternative proof of inequality (7.1), with better constants a, b than those in Theorem 7.2 (following [328, Chapter 8.1]).

Definition 7.2. *The* von Mangoldt function $\Lambda(n)$, $n \in \mathbb{N}$ *is defined by* $\Lambda(n) = \ln p$ *if* $n = p^k$, *and* $\Lambda(n) = 0$ *otherwise. Let us also define*

$$\psi(x) = \sum_{n \leq x} \Lambda(n), \quad \vartheta(x) = \sum_{p \leq x} \ln p, \quad T(x) = \sum_{n \leq x} \ln n.$$

The functions ψ *and* ϑ *are called* Chebyshev functions.

Theorem 7.4.

$$\sum_{d|n} \Lambda(d) = \ln n$$

Proof: Let $n = \prod_{i=1}^{k} p_i^{\alpha_i}$. Then $\ln n = \sum_{i=1}^{k} \alpha_i \ln p_i$. However, $p_i^{\alpha_i} \parallel n$, so $p_i^e \mid n$ if and only if e is one of the numbers $1, 2, \ldots, \alpha_i$. Therefore,

$$\sum_{i=1}^{k} \alpha_i \ln p_i = \sum_{i=1}^{k} \sum_{p_i^e | n} \ln p_i = \sum_{d|n} \Lambda(d). \qquad \square$$

Proposition 7.5. *For any real number* $x \geq 1$, *there is a real number* α, $|\alpha| \leq 1$, *such that* $T(x) = x \ln x - x + \alpha \ln ex$.