DISTRIBUTION OF PRIME NUMBERS

W W L CHEN

Chapter 2

ELEMENTARY PRIME NUMBER THEORY

2.1. Euclid's Theorem Revisited

We have already seen the elegant and simple proof of Euclid's theorem, that there are infinitely many primes. Here we shall begin by proving a slightly stronger result.

THEOREM 2A. *The series*

$$\sum_p \frac{1}{p}$$

is divergent.

PROOF. For every real number $X \geq 2$, write

$$P_X = \prod_{p \leq X} \left(1 - \frac{1}{p}\right)^{-1}.$$

Then

$$\log P_X = -\sum_{p \leq X} \log\left(1 - \frac{1}{p}\right) = S_1 + S_2,$$

where

$$S_1 = \sum_{p \leq X} \frac{1}{p} \qquad \text{and} \qquad S_2 = \sum_{p \leq X} \sum_{h=2}^{\infty} \frac{1}{hp^h}.$$

Since

$$0 \le \sum_{h=2}^{\infty} \frac{1}{hp^h} \le \sum_{h=2}^{\infty} \frac{1}{p^h} = \frac{1}{p(p-1)},$$

we have

$$0 \le S_2 \le \sum_{p} \frac{1}{p(p-1)} \le \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = 1,$$

so that $0 \le S_2 \le 1$. On the other hand, we have

$$P_X = \prod_{p \le X} \left(\sum_{h=0}^{\infty} \frac{1}{p^h} \right) \ge \sum_{n \le X} \frac{1}{n} \to \infty \quad \text{as } X \to \infty.$$

The result follows. ◯

For every real number $X \ge 2$, we write

$$\pi(X) = \sum_{p \le X} 1,$$

so that $\pi(X)$ denotes the number of primes in the interval $[2, X]$. This function has been studied extensively by number theorists, and attempts to study it in depth have led to major developments in other important branches of mathematics.

As can be expected, many conjectures concerning the distribution of primes were made based purely on numerical evidence, including the celebrated Prime number theorem, proved in 1896 by Hadamard and de la Vallée Poussin, that

$$\lim_{X \to \infty} \frac{\pi(X) \log X}{X} = 1.$$

We shall prove this in Chapter 5, and give another proof in Chapter 6. Here we shall be concerned with the weaker result of Tchebycheff, that there exist positive absolute constants c_1 and c_2 such that for every real number $X \ge 2$, we have

$$c_1 \frac{X}{\log X} < \pi(X) < c_2 \frac{X}{\log X}.$$

2.2. The Von Mangoldt Function

The study of the function $\pi(X)$ usually involves, instead of the characteristic function of the primes, a function which counts not only primes, but prime powers as well, and with weights. Accordingly, we introduce the von Mangoldt function $\Lambda : \mathbb{N} \to \mathbb{C}$, defined for every $n \in \mathbb{N}$ by writing

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^r \text{, with } p \text{ prime and } r \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

THEOREM 2B. *For every $n \in \mathbb{N}$, we have*

$$\sum_{m \mid n} \Lambda(m) = \log n.$$

PROOF. The result is clearly true for $n = 1$, so it remains to consider the case $n \geq 2$. Suppose that $n = p_1^{u_1} \ldots p_r^{u_r}$ is the canonical decomposition of n. Then the only non-zero contribution to the sum on the left hand side comes from those natural numbers m of the form $m = p_j^{v_j}$ with $j = 1, \ldots, r$ and $1 \leq v_j \leq u_j$. It follows that

$$\sum_{m|n} \Lambda(m) = \sum_{j=1}^{r} \sum_{v_j=1}^{u_j} \log p_j = \sum_{j=1}^{r} \log p_j^{u_j} = \log n.$$

This completes the proof. ◯

THEOREM 2C. *As $X \to \infty$, we have*

$$\sum_{m \leq X} \Lambda(m) \left[\frac{X}{m}\right] = X \log X - X + O(\log X).$$

PROOF. It follows from Theorem 2B that

$$\sum_{n \leq X} \log n = \sum_{n \leq X} \sum_{m|n} \Lambda(m) = \sum_{m \leq X} \Lambda(m) \sum_{\substack{n \leq X \\ m|n}} 1 = \sum_{m \leq X} \Lambda(m) \left[\frac{X}{m}\right].$$

It therefore suffices to prove that

$$\sum_{n \leq X} \log n = X \log X - X + O(\log X) \quad \text{as } X \to \infty. \tag{1}$$

To prove (1), note that $\log X$ is an increasing function of X. In particular, for every $n \in \mathbb{N}$, we have

$$\log n \leq \int_n^{n+1} \log u \, \mathrm{d}u,$$

so that

$$\sum_{n \leq X} \log n - \log(X+1) \leq \int_1^X \log u \, \mathrm{d}u.$$

On the other hand, for every $n \in \mathbb{N}$, we have

$$\log n \geq \int_{n-1}^{n} \log u \, \mathrm{d}u,$$

so that

$$\sum_{n \leq X} \log n = \sum_{2 \leq n \leq X} \log n \geq \int_1^{[X]} \log u \, \mathrm{d}u = \int_1^X \log u \, \mathrm{d}u - \int_{[X]}^X \log u \, \mathrm{d}u \geq \int_1^X \log u \, \mathrm{d}u - \log X.$$

The inequality (1) now follows on noting that

$$\int_1^X \log u \, \mathrm{d}u = X \log X - X + 1.$$

This completes the proof. ◯

2.3. Tchebycheff's Theorem

The crucial step in the proof of Tchebycheff's theorem concerns obtaining bounds on sums involving the von Mangoldt function. More precisely, we prove the following result.

THEOREM 2D. *There exist positive absolute constants c_3 and c_4 such that*

$$\sum_{m\le X}\Lambda(m)\ge\frac{1}{2}X\log 2\quad\textit{if } X\ge c_3, \tag{2}$$

and

$$\sum_{\frac{X}{2}<m\le X}\Lambda(m)\le c_4X\quad\textit{if } X\ge 0. \tag{3}$$

PROOF. If $m\in\mathbb{N}$ satisfies $X/2<m\le X$, then clearly $[X/2m]=0$. It follows from this and Theorem 2C that as $X\to\infty$, we have

$$\begin{aligned}\sum_{m\le X}\Lambda(m)\left(\left[\frac{X}{m}\right]-2\left[\frac{X}{2m}\right]\right)&=\sum_{m\le X}\Lambda(m)\left[\frac{X}{m}\right]-2\sum_{m\le\frac{X}{2}}\Lambda(m)\left[\frac{X}{2m}\right]\\&=(X\log X-X+O(\log X))-2\left(\frac{X}{2}\log\frac{X}{2}-\frac{X}{2}+O(\log X)\right)=X\log 2+O(\log X).\end{aligned}$$

Hence there exists a positive absolute constant c_5 such that for all sufficiently large X, we have

$$\frac{1}{2}X\log 2<\sum_{m\le X}\Lambda(m)\left(\left[\frac{X}{m}\right]-2\left[\frac{X}{2m}\right]\right)<c_5X.$$

We now consider the function $[\alpha]-2[\alpha/2]$. Clearly $[\alpha]-2[\alpha/2]<\alpha-2(\alpha/2-1)=2$. Note that the left hand side is an integer, so we must have $[\alpha]-2[\alpha/2]\le 1$. It follows that for all sufficiently large X, we have

$$\frac{1}{2}X\log 2<\sum_{m\le X}\Lambda(m).$$

The inequality (2) follows. On the other hand, if $X/2<m\le X$, then $[X/m]=1$ and $[X/2m]=0$, so that for all sufficiently large X, we have

$$\sum_{\frac{X}{2}<m\le X}\Lambda(m)\le c_5X.$$

The inequality (3) follows easily. ◯

We now state and prove Tchebycheff's theorem.

THEOREM 2E. (TCHEBYCHEFF) *There exist positive absolute constants c_1 and c_2 such that for every real number $X\ge 2$, we have*

$$c_1\frac{X}{\log X}<\pi(X)<c_2\frac{X}{\log X}.$$

PROOF. To prove the lower bound, note that

$$\sum_{m\le X}\Lambda(m)=\sum_{\substack{p,n\\p^n\le X}}\log p=\sum_{p\le X}(\log p)\sum_{1\le n\le\left[\frac{\log X}{\log p}\right]}1=\sum_{p\le X}(\log p)\left[\frac{\log X}{\log p}\right]\le\pi(X)\log X.$$

It follows from (2) that

$$\pi(X) \geq \frac{X \log 2}{2 \log X} \quad \text{if } X \geq c_3.$$

Since $\pi(2) = 1$, we get the lower bound for a suitable choice of c_1.

To prove the upper bound, note that in view of (3) and the definition of the von Mangoldt function, the inequality

$$\sum_{\frac{X}{2^{j+1}} < p \leq \frac{X}{2^j}} \log p \leq c_4 \frac{X}{2^j}$$

holds for every integer $j \geq 0$ and every real number $X \geq 0$. Suppose that $X \geq 2$. Let the integer $k \geq 0$ be defined such that $2^k < X^{1/2} \leq 2^{k+1}$. Then

$$\sum_{X^{1/2} < p \leq X} \log p \leq \sum_{j=0}^{k} \sum_{\frac{X}{2^{j+1}} < p \leq \frac{X}{2^j}} \log p \leq c_4 X \sum_{j=0}^{k} 2^{-j} < 2c_4 X,$$

so that

$$\sum_{X^{1/2} < p \leq X} 1 \leq \sum_{X^{1/2} < p \leq X} \frac{\log p}{\log X^{1/2}} < \frac{4c_4 X}{\log X},$$

whence

$$\pi(X) \leq X^{1/2} + \frac{4c_4 X}{\log X} < \frac{c_2 X}{\log X}$$

for a suitable c_2. ○

2.4. Some Results of Mertens

We conclude this chapter by obtaining an improvement of Theorem 2A.

THEOREM 2F. (MERTENS) *As $X \to \infty$, we have*

(4)
$$\sum_{m \leq X} \frac{\Lambda(m)}{m} = \log X + O(1),$$

(5)
$$\sum_{p \leq X} \frac{\log p}{p} = \log X + O(1),$$

and

(6)
$$\sum_{p \leq X} \frac{1}{p} = \log \log X + O(1).$$

PROOF. Recall Theorem 2C. As $X \to \infty$, we have

$$\sum_{m \leq X} \Lambda(m) \left[\frac{X}{m} \right] = X \log X - X + O(\log X).$$

Clearly $[X/m] = X/m + O(1)$, so that as $X \to \infty$, we have

$$\sum_{m\le X} \Lambda(m)\left[\frac{X}{m}\right] = X\sum_{m\le X}\frac{\Lambda(m)}{m} + O\left(\sum_{m\le X}\Lambda(m)\right).$$

It follows from (3) that

$$\sum_{m\le X}\Lambda(m) \le \sum_{j=0}^{\infty}\sum_{\frac{X}{2^{j+1}}<m\le\frac{X}{2^j}}\Lambda(m) \le 2c_4X,$$

so that as $X\to\infty$, we have

$$X\sum_{m\le X}\frac{\Lambda(m)}{m} = X\log X + O(X).$$

The inequality (4) follows. Next, note that

$$\sum_{m\le X}\frac{\Lambda(m)}{m} = \sum_{\substack{p,k\\ p^k\le X}}\frac{\log p}{p^k} = \sum_{p\le X}\frac{\log p}{p} + \sum_{p\le X}(\log p)\sum_{2\le k\le\frac{\log X}{\log p}}\frac{1}{p^k}.$$

As $X\to\infty$, we have

$$\sum_{p\le X}(\log p)\sum_{2\le k\le\frac{\log X}{\log p}}\frac{1}{p^k} \le \sum_{p\le X}(\log p)\sum_{k=2}^{\infty}\frac{1}{p^k} = \sum_{p\le X}\frac{\log p}{p(p-1)} \le \sum_{n=2}^{\infty}\frac{\log n}{n(n-1)} = O(1).$$

The inequality (5) follows. Finally, for every real number $X\ge 2$, let

$$T(X) = \sum_{p\le X}\frac{\log p}{p}.$$

Then it follows from (5) that there exists a positive absolute constant c_6 such that $|T(X)-\log X| < c_6$ whenever $X\ge 2$. On the other hand,

$$\begin{aligned}\sum_{p\le X}\frac{1}{p} &= \sum_{p\le X}\frac{\log p}{p}\left(\frac{1}{\log X}+\int_p^X\frac{\mathrm{d}y}{y\log^2 y}\right) = \frac{T(X)}{\log X}+\int_2^X\frac{T(y)\,\mathrm{d}y}{y\log^2 y}\\ &= \frac{T(X)-\log X}{\log X}+\int_2^X\frac{(T(y)-\log y)\,\mathrm{d}y}{y\log^2 y}+1+\int_2^X\frac{\mathrm{d}y}{y\log y}.\end{aligned}$$

It follows that as $X\to\infty$, we have

$$\left|\sum_{p\le X}\frac{1}{p}-\log\log X\right| < \frac{c_6}{\log X}+\int_2^X\frac{c_6\,\mathrm{d}y}{y\log^2 y}+1-\log\log 2 = O(1).$$

The inequality (6) follows. ○

Problems for Chapter 2

1. Prove that $\Lambda(n)+\sum_{m\mid n}\mu(m)\log m = 0$ for every $n\in\mathbb{N}$.

2. For any arithmetic function f, we define f' to be the arithmetic function given by $f'(n) = f(n)\log n$ for every $n \in \mathbb{N}$. Then for the arithmetic function U defined by $U(n) = 1$ for every $n \in \mathbb{N}$, we have $U'(n) = \log n$ and $U''(n) = \log^2 n$ for every $n \in \mathbb{N}$.
   (i) Suppose that f and g are arithmetic functions.
      (I) Prove that $(f+g)' = f' + g'$ and $(f * g)' = (f' * g) + (f * g')$.
      (II) Suppose that $f(1) \neq 0$. By noting that $(f * f^{-1})'(n) = 0$ for every $n \in \mathbb{N}$, prove that $(f^{-1})' = -f' * (f * f)^{-1}$.
   (ii) Explain why $\Lambda * U = U'$. Then establish Selberg's identity $\Lambda' + (\Lambda * \Lambda) = U'' * \mu$.

3. Prove that for every real number $X \geq 2$, we have $\displaystyle\prod_{p \leq X}\left(1 - \frac{1}{p}\right)^{-1} > \log X$.

4. Use the well-known inequality

$$\frac{t}{1+t} < \log(1+t) < t, \qquad \text{where } t > -1 \text{ and } t \neq 0,$$

to show that

$$\sum_{p \leq X} \frac{1}{p-1} > \log\log X \qquad \text{and} \qquad \sum_{p \leq X} \frac{1}{p} > \log\log X - 1.$$

5. Suppose that
   - λ_n is an increasing sequence of real numbers with limit infinity;
   - c_n is an arbitrary sequence of real or complex numbers; and
   - f has continuous derivative for $X \geq \lambda_1$.

   For every $X \geq \lambda_1$, let

$$C(X) = \sum_{\lambda_n \leq X} c_n.$$

   Establish the partial summation formula, that for every $X \geq \lambda_1$, we have

$$\sum_{\lambda_n \leq X} c_n f(\lambda_n) = C(X)f(X) - \int_{\lambda_1}^{X} C(y) f'(y)\,\mathrm{d}y.$$

6. Use Theorem 2F and partial summation to show that as $X \to \infty$, we have

$$\int_2^X \frac{\pi(y)}{y^2}\,\mathrm{d}y = \sum_{p \leq X} \frac{1}{p} + o(1) \sim \log\log X.$$

7. Derive the Prime number theorem, that

$$\pi(X) \sim \frac{X}{\log X} \quad \text{as } X \to \infty,$$

   from the hypothetical relation

$$\sum_{p \leq X} \log p \sim X \quad \text{as } X \to \infty,$$

   and the information

$$\int_2^X \frac{\mathrm{d}y}{\log y} = \frac{X}{\log X} + o\left(\frac{X}{\log X}\right) \qquad \text{as } X \to \infty.$$

8. Show that the series $\displaystyle\sum_{p \leq X} \frac{1}{p\log p}$ converges as $X \to \infty$.