Introduction to Analytic Number Theory

Math 531 Lecture Notes, Fall 2005

A.J. Hildebrand
Department of Mathematics
University of Illinois

http://www.math.uiuc.edu/~hildebr/ant

Version 2013.01.07

Chapter 3

Distribution of primes I: Elementary results

The Prime Number Theorem (PNT), in its most basic form, is the asymptotic relation $\pi(x) \sim x/\log x$ for the prime counting function $\pi(x)$, the number $\pi(x)$ of primes $\leq x$. This result had been conjectured by Legendre and (in a more precise form) by Gauss, based on examining tables of primes. However, neither succeeded in proving the PNT (and it certainly wasn't for lack of trying!). It was only much later, near the end of the 19th century, that a proof of the PNT was given, independently by J. Hadamard and C. de la Vallee Poussin, via a new, analytic, approach that was not available to Gauss and his contemporaries. We will give a proof of the PNT, in a strong form with an explicit error term, in a later chapter.

In this chapter we establish a number of elementary results on the distribution of primes that are much easier to prove than the PNT and which, for the most part, have been known long before the PNT was proved. These results are of interest in their own right, and they have many applications.

3.1 Chebyshev type estimates

Getting upper and lower bounds for the prime counting function $\pi(x)$ is surprisingly difficult. Euclid's result that there are infinitely many primes shows that $\pi(x)$ tends to infinity, but the standard proofs of the infinitude of prime are indirect and do not give an explicit lower bound for $\pi(x)$, or give only a very weak bound. For example, Euclid's argument shows that the n-th prime p_n satisfies the bound $p_n \leq p_1 \dots p_{n-1} + 1$. By induction, this implies that $p_n \leq e^{e^{n-1}}$ for all n, from which one can deduce the bound

$\pi(x) \geq \log\log x$ for sufficiently large x. This bound is far from the true order of $\pi(x)$, but it is essentially the best one can derive from Euclid's argument.

Euler's proof of the infinitude of primes proceeds by showing that $\sum_{p\leq x} 1/p \geq \log\log x - c$ for some constant c and sufficiently large x. Although this gives the correct order for the partials sum of the reciprocals of primes (as we will see below, the estimate is accurate to within an error $O(1)$), one cannot deduce from this a lower bound for $\pi(x)$ of comparable quality. In fact, one can show (see the exercises) that the most one can deduce from the above bound for $\sum_{p\leq x} 1/p$ is a lower bound of the form $\pi(x) \gg \log x$. While this is better than the bound obtained from Euclid's argument, it is still far from the true order of magnitude.

In the other direction, getting non-trivial upper bounds for $\pi(x)$ is not easy either. Even showing that $\pi(x) = o(x)$, i.e., that the primes have density zero among all integers, is by no means easy, when proceeding "from scratch". (Try to prove this bound without resorting to any of the results, techniques, and tricks you have learned so far.)

In light of these difficulties in getting even relatively weak nontrivial bounds for $\pi(x)$ it is remarkable that, in the middle of the 19th century, the Russian mathematician P.L. Chebyshev was able to determine the precise order of magnitude of the prime counting function $\pi(x)$, by showing that there exist positive constants c_1 and c_2 such that

$$c_1\frac{x}{\log x} \leq \pi(x) \leq c_2\frac{x}{\log x}$$

for all sufficiently large x. In fact, Chebyshev proved such an inequality with constants $c_1 = 0.92\ldots$ and $c_2 = 1.10\ldots$. This enabled him to conclude that, for sufficiently large x (and, in fact, for all $x \geq 1$) there exists a prime p with $x < p \leq 2x$, an assertion known as **Bertrand's postulate**.

In establishing these bounds, Chebyshev introduced the auxiliary functions

$$\theta(x) = \sum_{p\leq x} \log p, \quad \psi(x) = \sum_{n\leq x} \Lambda(n),$$

which proved to be extremely useful in subsequent work. Converting results on $\pi(x)$ to results on $\psi(x)$ or $\theta(x)$, or vice versa, is easy (see Theorem 3.2 below), and we will state most of our results for all three of these functions and use whichever version is most convenient for the proof.

Theorem 3.1 (Chebyshev estimates). *For $x \geq 2$ we have*

(i) $$\psi(x) \asymp x,$$

(ii) $$\theta(x) \asymp x,$$

(iii) $$\pi(x) \asymp \frac{x}{\log x}.$$

Proof. We will establish (i), and then deduce (ii) and (iii) from (i).

To prove (i), we need to show that there exist positive constants c_1 and c_2 such that

$$c_1 x \leq \psi(x) \leq c_2 x \tag{3.1}$$

holds for all $x \geq 2$.

We begin by noting that it suffices to establish (3.1) for $x \geq x_0$, for a suitable $x_0 \geq 2$. Indeed, suppose there exists a constant $x_0 \geq 2$ such that (3.1) holds for $x \geq x_0$. Since for $2 \leq x \leq x_0$ we have, trivially, $\psi(x)/x \leq \psi(x_0)/2$ and $\psi(x)/x \geq \psi(2)/x_0 = \log 2/x_0$, it then follows that (3.1) holds for all $x \geq 2$ with constants $c_1' = \min(c_1, \log 2/x_0)$ and $c_2' = \max(c_2, \psi(x_0)/2)$ in place of c_1 and c_2.

In what follows, we may therefore assume that x is sufficiently large. (Recall our convention that O-estimates without explicit range are understood to hold for $x \geq x_0$ with a sufficiently large x_0.)

Define

$$S(x) = \sum_{n \leq x} \log n, \quad D(x) = S(x) - 2S(x/2).$$

To prove (3.1) we will evaluate $D(x)$ in two different ways. On the one hand, using the asymptotic estimate for $S(x)$ established earlier (see Corollary 2.8), we have

$$\begin{aligned} D(x) &= x(\log x - 1) + O(\log x) - 2(x/2)(\log(x/2) - 1) + O(\log(x/2)) \\ &= (\log 2)x + O(\log x). \end{aligned}$$

Since $1/2 < \log 2 < 1$, this implies

$$x/2 \leq D(x) \leq x \quad (x \geq x_0) \tag{3.2}$$

with a suitable x_0.

On the other hand, using the identity $\log n = (\Lambda * 1)(n) = \sum_{d|n} \Lambda(d)$ and interchanging summations, we have

$$S(x) = \sum_{d \leq x} \Lambda(d) \sum_{\substack{n \leq x \\ d|n}} 1 = \sum_{d \leq x} \Lambda(d)[x/d] \quad (x \geq 1), \tag{3.3}$$

where $[t]$ denotes the greatest integer function. Applying (3.3) to $S(x)$ and $S(x/2)$, we get

$$D(x) = S(x) - 2S(x/2) = \sum_{d\le x} \Lambda(d) f(x/d) \quad (x \ge 2), \tag{3.4}$$

where $f(t) = [t] - 2[t/2]$. (Note that in the evaluation of $S(x/2)$ the summation range $d \le x/2$ can be extended to $d \le x$, since the terms with $x/2 < d \le x$ do not contribute to the sum due to the factor $[x/2d]$.) By the elementary inequalities $[s] \le s$ and $[s] > s - 1$, valid for any real number s, we have

$$f(t) \begin{cases} < t - 2(t/2 - 1) = 2, \\ > t - 1 - 2(t/2) = -1. \end{cases}$$

Since the function $f(t) = [t] - 2[t/2]$ is integer-valued, it follows that

$$f(t) \begin{cases} = 1 & \text{if } 1 \le t < 2, \\ \in \{0, 1\} & \text{if } t \ge 2. \end{cases}$$

Hence (3.4) implies

$$D(x) \begin{cases} \le \sum_{d\le x} \Lambda(d) = \psi(x) \\ \ge \sum_{x/2<d\le x} \Lambda(d) = \psi(x) - \psi(x/2) \end{cases} \qquad (x \ge 2). \tag{3.5}$$

Combining (3.2) and (3.5), we obtain

$$\psi(x) \ge D(x) \ge x/2 \quad (x \ge x_0), \tag{3.6}$$

and

$$\psi(x) \le D(x) + \psi(x/2) \le x + \psi(x/2) \quad (x \ge x_0). \tag{3.7}$$

The first of these inequalities immediately gives the lower bound in (3.1) (with $c_1 = 1/2$). To obtain a corresponding upper bound, we note that iteration of (3.7) yields

$$\psi(x) \le \sum_{i=0}^{k-1} x2^{-i} + \psi(x2^{-k}),$$

for any positive integer k such that $x2^{-k+1} \ge x_0$. Choosing k as the maximal such integer, we have $2^{-k+1}x \ge x_0 > 2^{-k}x$ and thus $\psi(x2^{-k}) \le \psi(x_0)$, and hence obtain

$$\psi(x) \le \sum_{i=0}^{k-1} x2^{-i} + \psi(x_0) \le 2x + \psi(x_0),$$

which gives the upper bound in (3.1) for $x \geq x_0$ with a sufficiently large constant c_2 (in fact, we could take $c_2 = 2 + \epsilon$, for $x \geq x_0(\epsilon)$, for any fixed $\epsilon > 0$ with a suitable $x_0(\epsilon)$).

This completes the proof of (i).

To deduce (ii), we note that

$$\begin{aligned} \psi(x) - \theta(x) &= \sum_{p^m \leq x} \log p - \sum_{p \leq x} \log p \\ &= \sum_{p \leq \sqrt{x}} \log p \sum_{2 \leq m \leq \log x/\log p} 1 \\ &\leq \sum_{p \leq \sqrt{x}} (\log p) \left[\frac{\log x}{\log p}\right] \leq \sqrt{x} \log x, \end{aligned} \tag{3.8}$$

so that

$$\theta(x) \begin{cases} \leq \psi(x), \\ \geq \psi(x) - \sqrt{x} \log x. \end{cases}$$

Hence the upper bound in (3.1) remains valid for $\theta(x)$, with the same values of c_2 and x_0, and the lower bound holds for $\theta(x)$ with constant $c_1/2$ (for example) instead of c_1, upon increasing the value of x_0 if necessary.

The lower bound in (iii) follows immediately from that in (ii), since $\pi(x) \geq (1/\log x) \sum_{p \leq x} \log p = \theta(x)/\log x$. The upper bound follows from the inequality

$$\pi(x) \leq \pi(\sqrt{x}) + \frac{1}{\sqrt{\log x}} \sum_{\sqrt{x} < p \leq x} \log p \leq \sqrt{x} + \frac{2}{\log x} \theta(x)$$

and the upper bound in (ii). □

Alternate proofs of Theorem 3.1. The proof given here rests on the convolution identity $\Lambda * 1 = \log$, which relates the "unknown" function Λ to two extremely well-behaved functions, namely 1 and log. Given this relation, it is natural to try to use it to derive information on the average behavior of the function $\Lambda(n)$ from the very precise information that is available on the behavior of the functions 1 and $\log n$. The particular way this identity is used in the proof of the theorem may seem contrived. Unfortunately, more natural approaches don't work, and one has to resort to some sort of "trickery" to get any useful information out of the above identity. For example, it is tempting to try to simply invert the relation $\Lambda * 1 = \log$ to express Λ as $\Lambda = \mu * \log$, interpret Λ as a "perturbation" as the function log and proceed as in the convolution method described in Section 2.4. Unfortunately, the error terms in this approach are too large to be of any use.

There exist alternate proofs of Theorem 3.1, but none is particularly motivated or natural, and all involve some sort of "trick". For example, a commonly seen argument, which may be a bit shorter than the one given here, but has more the character of pulling something out of the air, is based on an analysis of the middle binomial coefficient $\binom{2n}{n}$: On the one hand, writing this coefficient as a fraction $(n+1)(n+2)\cdots(2n)/n!$ and noting that every prime p with $n < p \le 2n$ divides the numerator, but not the denominator, we see that $\binom{2n}{n}$ is divisible by the product $\prod_{n<p\le 2n} p$. On the other hand, the binomial theorem gives the bound

$$\binom{2n}{n} \le \sum_{k=0}^{2n} \binom{2n}{k} = 2^{2n}.$$

Hence $\prod_{n<p\le 2n} p \le 2^{2n}$, and taking logarithms, we conclude

$$\theta(2n) - \theta(n) = \sum_{n<p\le 2n} \log p \le (2\log 2)n,$$

for any positive integer n. By iterating this inequality, one gets $\theta(2^k) \le (2\log 2)2^k$ for any positive integer k, and then $\theta(x) \le (4\log 2)x$ for any real number $x \ge 2$. This proves the upper bound in (ii) with constant $4\log 2$. The lower bound in (ii) can be proved by a similar argument, based on an analysis of the prime factorization of $\binom{2n}{n}$, and the lower bound

$$\binom{2n}{n} \ge \frac{1}{2n+1} \sum_{k=0}^{2n} \binom{2n}{k} = \frac{2^{2n}}{2n+1}.$$

The constants in Chebyshev's estimates. An inspection of the above argument shows that it yields (3.1) with any constants c_1 and c_2 satisfying $c_1 < \log 2 = 0.69\ldots$ and $c_2 > 2\log 2 = 1.38\ldots$, for sufficiently large x. Chebyshev used a more complicated version of this argument, in which the linear combination $S(x) - 2S(x/2)$ is replaced by $S(x) - S(x/2) - S(x/3) - S(x/5) + S(x/30)$, to obtain $c_1 = 0.92\ldots$ and $c_2 = 1.10\ldots$ as constants in (3.1). For most applications, the values of these constants are not important. However, since the PNT had not been proved at the time Chebyshev proved his estimates, there was a strong motivation to obtain constants as close to 1 as possible. It is natural to ask if, by considering more general linear combinations of the functions $S(x/k)$, one can further improve these constants. This is indeed the case; in fact, Diamond and Erdös showed that it is possible to obtain constants c_1 and c_2 arbitrarily close to 1, by using Chebyshev's approach with a suitable linear combination of the function $S(x)$. Now, the assertion that (3.1) holds with constants c_1 and c_2 arbitrarily close to 1, clearly implies the PNT in the form $\psi(x) \sim x$, so it would seem that Chebyshev's method in fact yields a proof of the PNT. However, this is not the case, since in proving that c_1 and c_2 can be taken arbitrarily close to 1, Diamond and Erdös had to use the PNT.

Theorem 3.2 (Relation between π, ψ, and θ)**.** *For $x \geq 2$ we have*

(i) $$\theta(x) = \psi(x) + O(\sqrt{x}),$$

(ii) $$\pi(x) = \frac{\psi(x)}{\log x} + O\left(\frac{x}{\log^2 x}\right).$$

Proof. A slightly weaker version of (i), with error term $O(\sqrt{x}\log x)$ instead of $O(\sqrt{x})$, was established in (3.8) above. To obtain (i) as stated, we use again the identity (3.8) for $\psi(x)-\theta(x)$, but instead of estimating the right-hand side trivially, we apply Chebyshev's bound (which we couldn't use while proving Theorem 3.1). This gives for $\psi(x) - \theta(x)$ the bound $\leq \pi(\sqrt{x})\log x \ll (\sqrt{x}/\log\sqrt{x})\log x \ll \sqrt{x}$.

To obtain (ii), we write

$$\pi(x) = \sum_{p\leq x} 1 = \sum_{p\leq x}(\log p)(1/\log p)$$

and "eliminate" the factor $1/\log p$ by partial summation:

$$\pi(x) = \frac{\theta(x)}{x} - \int_2^x \theta(t)\left(-\frac{1}{t(\log t)^2}\right)dt = \frac{\theta(x)}{x} + \int_2^x \frac{\theta(t)}{t(\log t)^2}dt.$$

Since $\theta(t) \ll t$ by Chebyshev's estimate, the last integral is of order

$$\begin{aligned} &\ll \int_2^x \frac{1}{(\log t)^2}dt \leq \int_2^{\sqrt{x}} \frac{1}{(\log 2)^2} + \int_{\sqrt{x}}^x \frac{1}{(\log\sqrt{x})^2}dt \\ &\ll \sqrt{x} + \frac{x}{(\log\sqrt{x})^2} \ll \frac{x}{(\log x)^2}, \end{aligned}$$

so we have

$$\pi(x) = \frac{\theta(x)}{\log x} + O\left(\frac{x}{\log^2 x}\right).$$

In view of (i), we may replace $\theta(x)$ by $\psi(x)$ on the right-hand side, and thus obtain (ii). □

Corollary 3.3 (Equivalent formulations of PNT)**.** *The following relations are equivalent:*

(i) $$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty),$$

(ii) $$\theta(x) \sim x \quad (x \to \infty),$$

(iii) $$\psi(x) \sim x \quad (x \to \infty).$$

Proof. By the previous theorem, the functions $\psi(x)$, $\theta(x)$, and $\pi(x)\log x$ differ by an error term that is of order $O(x/\log x)$ (at worst), and hence are of smaller order (by a factor $1/\log x$) than the main terms in the asserted relations. □

3.2 Mertens type estimates

A second class of estimates below the level of the PNT are estimates for certain weighted sums over primes, such as the sum of reciprocals of primes up to x. These estimates seem surprisingly strong as the error terms involved are by at least a logarithmic factor smaller than the main term, yet they are not strong enough to imply the PNT.

Theorem 3.4 (Mertens' estimates)**.** *We have*

(i) $$\sum_{n\le x} \frac{\Lambda(n)}{n} = \log x + O(1),$$

(ii) $$\sum_{p\le x} \frac{\log p}{p} = \log x + O(1),$$

(iii) $$\sum_{p\le x} \frac{1}{p} = \log\log x + A + O\left(\frac{1}{\log x}\right),$$

(iv) $$\prod_{p\le x}\left(1-\frac{1}{p}\right) = \frac{e^{-\gamma}}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right),$$

where A is a constant and γ is Euler's constant.

Before proving this result, we make some remarks and derive two corollaries.

Estimate (iv) is usually referred to as **Mertens' formula**. We will prove this result here with an unspecified constant in place of $e^{-\gamma}$; the proof that this constant is equal to $e^{-\gamma}$ requires additional tools and will be deferred until the next chapter. It is easy to show that the product on the left-hand side of (iv) is equal to the density of positive integers that have no prime factor $\le x$, i.e., $\lim_{y\to\infty}(1/y)\#\{n \le y : p|n \Rightarrow p > x\}$. (For example, this follows by applying Wintner's mean value theorem (Theorem 2.11) to the characteristic function of the integers with no prime factor $\le x$.) Mertens' formula shows that this density, i.e., the "probability" that an integer has no prime factors $\le x$, tends to zero as $x \to \infty$ at a rate proportional to $1/\log x$.

An equivalent formulation of (iv), obtained by taking the reciprocal on each side, is

$$\text{(iv')}\qquad \prod_{p\le x}\left(1-\frac{1}{p}\right)^{-1} = e^{\gamma}(\log x)\left(1+O\left(\frac{1}{\log x}\right)\right).$$

This version has the following interesting interpretation:

Let $P(x)$ denote the product on the left of (iv'). Expanding each of the factors $(1-1/p)^{-1}$ into a geometric series and multiplying out all terms in this product, one obtains a sum over terms $(*)$ $\prod_{p\le x} p^{-\alpha_p}$, where the exponents α_p run over all non-negative integers. Now, $(*)$ is the reciprocal of a positive integer n all of whose prime factors are $\le x$, and by the fundamental theorem of arithmetic each such reciprocal $1/n$ has exactly one representation in the form $(*)$. Hence, letting $A_x = \{n \in \mathbb{N} : p|n \Rightarrow p \le x\}$ denote the set of such integers n, we have $P(x) = \sum_{n\in A_x} 1/n$. Now A_x clearly contains every positive integer $n \le x$, so we have the lower bound $P(x) \ge \sum_{n\le x} 1/n$, which is asymptotic to $\log x$. The estimate (iv') shows that $P(x)$ is (asymptotically) by a factor e^{γ} larger than this trivial lower bound. The difference between the actual estimate and the trivial bound, $(e^{\gamma}-1)\log x$, can be interpreted as a measure of how many integers from A_x were missed by only counting integers $n \le x$.

The estimates (i)–(iii) can be viewed as average versions of the PNT, expressed in terms of $\psi(x)$, $\theta(x)$, and $\pi(x)$, respectively. For example, (i) implies that the *logarithmic mean value* of the von Mangoldt function $\Lambda(n)$ exists and is equal to 1. The existence of the *ordinary (asymptotic) mean value* of $\Lambda(n)$ would imply (in fact, is equivalent to) the PNT. However, as we have seen in Theorems 2.13 and 2.14, the existence of an asymptotic mean value is a strictly stronger assertion than the existence of a logarithmic mean value, so the PNT does not follow from these estimates.

The following corollary makes the interpretation of Mertens' estimates as average versions of the PNT more explicit.

Corollary 3.5. *We have*

$$\int_1^x \frac{\psi(t)/t}{t}dt = \log x + O(1). \tag{3.9}$$

Proof. By partial summation and Chebyshev's estimate (Theorem 3.1), the left-hand side of (i) in Theorem 3.4 equals

$$\frac{\psi(x)}{x} + \int_1^x \frac{\psi(t)}{t^2}dt = O(1) + \int_1^x \frac{\psi(t)/t}{t}dt,$$

so (i) implies the estimate of the corollary. □

A simple consequence of this estimate is the following result, which says that the proportionality constant in the PNT, if it exists (i.e., if $\psi(x) \sim cx$ for *some* constant c), must be equal to 1.

Corollary 3.6. *Let A_* and A^* denote, respectively, the* lim inf*, and the* lim sup*, of $\psi(x)/x$. Then $A_* \leq 1 \leq A^*$. Moreover, if the limit $A = \lim_{x\to\infty} \psi(x)/x$ exists, then $A = 1$.*

Proof. The second assertion clearly follows from the first. To prove the first assertion, suppose, for example, that A^* is strictly less than 1. Then there exist $\epsilon > 0$ and $x_0 \geq 2$ such that $\psi(x) \leq (1-\epsilon)x$ for $x \geq x_0$. Hence, for $x \geq x_0$, the left-hand side of (3.9) is

$$\leq \int_1^{x_0} \frac{\psi(t)/t}{t}dt + (1-\epsilon)\int_{x_0}^x \frac{1}{t}dt \leq \psi(x_0)\int_1^\infty \frac{1}{t^2}dt + (1-\epsilon)\log x,$$

which contradicts (3.9) if x is sufficiently large. Hence $A^* \geq 1$, and a similar argument shows $A_* \leq 1$. □

Proof of Theorem 3.4. To prove (i) we begin, as in the proof of Chebyshev's estimate for $\psi(x)$, with two evaluations for $S(x) = \sum_{n\leq x} \log n$. On the one hand, by Corollary 2.7, we have $S(x) = x\log x + O(x)$. On the other hand, (3.3) and Chebyshev's estimate imply

$$\begin{aligned} S(x) &= \sum_{d\leq x} \Lambda(d)[x/d] = x\sum_{n\leq x} \frac{\Lambda(d)}{d} + O\left(\sum_{d\leq x} \Lambda(d)\right) \\ &= x\sum_{n\leq x} \frac{\Lambda(d)}{d} + O(x). \end{aligned}$$

Setting the last expression equal to $x \log x + O(x)$ and dividing by x, we obtain (i).

The estimate (ii) follows from (i) on noting that the difference between the sums in (i) and (ii) equals

$$\sum_{\substack{p^m \le x \\ m \ge 2}} \frac{\log p}{p^m},$$

which can be bounded by

$$\le \sum_p \log p \sum_{m=2}^{\infty} \frac{1}{p^m} \le 2 \sum_p \frac{\log p}{p^2} < \infty.$$

Hence the sums in (i) and (ii) differ by a term of order $O(1)$, and so (ii) follows from (i).

We now deduce (iii) from (ii). To this end we write the summand $1/p$ as $((\log p)/p)(1/\log p)$ and apply partial summation to "remove" the factor $1/\log p$. Defining $L(t)$ and $R(t)$ by

$$L(t) = \sum_{p \le t} \frac{\log p}{p} = \log t + R(t)$$

(so that $R(t) = O(1)$ by (ii)), we obtain

$$\begin{aligned} \sum_{p \le x} \frac{1}{p} &= \frac{L(x)}{\log x} - \int_2^x L(t) \frac{-1}{t(\log t)^2} dt \\ &= 1 + \frac{R(x)}{\log x} + \int_2^x \frac{1}{t \log t} dt + \int_2^x \frac{R(t)}{t(\log t)^2} dt \\ &= 1 + O\left(\frac{1}{\log x}\right) + \log\log x - \log\log 2 + I(x), \end{aligned}$$

where $I(x) = \int_2^x R(t)/(t \log^2 t) dt$. To obtain the desired estimate (iii), it suffices to show that, for some constant C,

$$I(x) = C + O\left(\frac{1}{\log x}\right). \tag{3.10}$$

To prove this, note that, since $R(t) = O(1)$ and the integral $\int_2^\infty (t \log^2 t)^{-1} dt$ converges, the infinite integral $I(\infty) = \int_2^\infty R(t)/(t \log^2 t) dt$ converges. Setting $C = I(\infty)$, we have

$$I(x) = C - \int_x^\infty \frac{R(t)}{t(\log t)^2} dt = C + O\left(\int_x^\infty \frac{1}{t(\log t)^2} dt\right) = C + O\left(\frac{1}{\log x}\right),$$

which proves (3.10).

It remains to prove (iv). We will establish (iv) with *some* positive constant B in place of $e^{-\gamma}$, but defer the proof that this constant equals $e^{-\gamma}$ (which is, in fact, the most difficult part of the proof of Theorem 3.4), to a later chapter.

Taking logarithms, (iv) becomes

$$\sum_{p\le x} \log\left(1-\frac{1}{p}\right) = -\gamma - \log\log x + \log\left(1 + O\left(\frac{1}{\log x}\right)\right).$$

Since, for $|y| \le 1/2$, $|\log(1+y)| \asymp |y|$, this estimate is equivalent to

$$-\sum_{p\le x} \log\left(1-\frac{1}{p}\right) = C + \log\log x + O\left(\frac{1}{\log x}\right), \tag{3.11}$$

with $C = -\gamma$. We will show that the latter estimate holds, with a suitable constant C.

Using the expansion $-\log(1-x) = \sum_{n\ge 1} x^n/n$ ($|x|<1$), we have

$$-\sum_{p\le x} \log\left(1-\frac{1}{p}\right) = \sum_{p\le x}\frac{1}{p} + \sum_{p\le x} r_p,$$

with $r_p = \sum_{m=2}^{\infty} 1/(mp^m)$. Since $|r_p| \le (1/2)\sum_{m=2}^{\infty} p^{-m} \le p^{-2}$, the series $\sum_p r_p$ is absolutely convergent, with sum R, say, and we have

$$\sum_{p\le x} r_p = R - \sum_{p>x} r_p = R + O\left(\sum_{p>x}\frac{1}{p^2}\right) = R + O\left(\frac{1}{x}\right).$$

Hence the difference between the left-hand sides of (iii) and (3.11) is $R + O(1/x)$. Therefore (3.11) follows from (iii).

□

3.3 Elementary consequences of the PNT

The following result gives some elementary consequences of the PNT.

Theorem 3.7 (Elementary consequences of the PNT)**.** *The PNT implies:*

(i) The nth prime p_n satisfies $p_n \sim n\log n$ as $n\to\infty$.

(ii) The function $\omega(n)$, the number of distinct prime factors of n, has maximal order $(\log n)/(\log\log n)$, i.e., it satisfies

$$\limsup_{n\to\infty} \frac{\omega(n)}{(\log n)/(\log\log n)} = 1.$$

(iii) For every $\epsilon > 0$ there exists $x_0 = x_0(\epsilon) \geq 2$ such that for all $x \geq x_0$ there exists a prime p with $x < p \leq (1+\epsilon)x$.

(iv) The set of rational numbers p/q with p and q prime is dense on the positive real axis.

(v) Given any finite string $a_1 \dots a_n$ of digits $\{0, 1, \dots, 9\}$ with $a_1 \neq 0$, there exists a prime number whose decimal expansion begins with this string.

Proof. (i) Since $p_n \to \infty$ as $n \to \infty$, the PNT gives

$$n = \pi(p_n) \sim \frac{p_n}{\log p_n} \quad (n \to \infty). \tag{3.12}$$

This implies that, for any fixed $\epsilon > 0$ and all sufficiently large n, we have $p_n^{1-\epsilon} \leq n \leq p_n$, and hence $(1-\epsilon)\log p_n \leq \log n \leq \log p_n$. The latter relation shows that $\log p_n \sim \log n$ as $n \to \infty$, and substituting this asymptotic formula into (3.12) yields $n \sim p_n/\log n$, which is equivalent to the desired relation $p_n \sim n \log n$.

(ii) First note that, given any positive integer k, the least positive integer n with $\omega(n) = k$ is $n_k = p_1 \dots p_k$, where p_i denotes the i-th prime. Since $\log n/\log\log n$ is a monotone increasing function for sufficiently large n, it suffices to consider integers n from the sequence $\{n_k\}$ in the limsup in (ii). We then need to show that

$$\limsup_{k\to\infty} \frac{k}{(\log n_k)/(\log\log n_k)} = 1. \tag{3.13}$$

The PNT and the asymptotic formula for p_k proved in part (i) implies

$$\log n_k = \sum_{i=1}^{k} \log p_i = \theta(p_k) \sim p_k \sim k \log k \quad (k \to \infty)$$

and

$$\begin{aligned}\log\log n_k &= \log((1+o(1))k\log k) = \log k + \log\log k + \log(1+o(1)) \\ &= (1+o(1))\log k.\end{aligned}$$

Substituting these estimates on the left side of (3.13) gives the desired relation.

(iii) By the PNT we have, for any fixed $\epsilon > 0$,

$$\frac{\pi((1+\epsilon)x)}{\pi(x)} \sim \frac{(1+\epsilon)x/\log((1+\epsilon)x)}{x/\log x} = (1+\epsilon)\frac{\log x}{\log(1+\epsilon)+\log x},$$

and thus $\lim_{x\to\infty} \pi((1+\epsilon)x)/\pi(x) = 1+\epsilon$. This implies $\pi((1+\epsilon)x) > \pi(x)$ for any $\epsilon > 0$ and $x \geq x_0(\epsilon)$, which is equivalent to the assertion in (iii).

(iv) Given a positive real number α and $\epsilon > 0$, we need to show that there exist primes p and q with $|p/q - \alpha| \leq \epsilon$, or equivalently $(*)$ $\alpha q - \epsilon q \leq p \leq \alpha q + \epsilon q$. To this end, set $\epsilon' = \epsilon/\alpha$, and let q be any prime such that $\alpha q \geq x_0$, where $x_0 = x_0(\epsilon')$ is defined as in the previous proof relative to ϵ'. Thus, for $x \geq x_0$, there exists a prime p with $x < p \leq (1+\epsilon')x$. Taking $x = \alpha q$, we conclude that there exists a prime p with $\alpha q < p \leq (1+\epsilon')\alpha q = \alpha q + \epsilon q$, as desired.

(v) Given a string of digits $a_1 \dots a_r$, with $a_1 \neq 0$, let $A = a_1 \dots a_r$ denote the integer formed by these digits. Since $a_1 \neq 0$, A is a positive integer. Now observe that a positive integer n begins with the string $a_1 \dots a_r$ if and only if, for some integer $k \geq 0$, $(*)$ $10^k A \leq n < 10^k(A+1)$. Applying the result of (iii) with some $\epsilon < 1/A$ (say, $\epsilon = 1/(A+1)$), we see that the interval $(*)$ contains a prime for sufficiently large k. □

3.4 The PNT and averages of the Moebius function

As we have seen above, the PNT is "equivalent" to each of the relations $\psi(x) \sim x$ and $\theta(x) \sim x$, in the sense that deducing one statement from the other, and vice versa, is substantially easier than proving either of these statements. (One should be aware that this type of "equivalence" is an imprecise, and to some extent subjective, notion, but in the context of the prime number theorem this informal usage of the term "equivalent" has become standard. Of course, from a purely logical point of view, all true statements are equivalent to each other.)

There exist many other prime number sums or products for which an asymptotic estimation is equivalent, in the same sense, to the PNT. These equivalences are usually neither particularly deep or unexpected, and are easily established.

In this section we prove the equivalence of the PNT to a rather different type of result, namely that the Moebius function has mean value zero. In contrast to the above-mentioned equivalences, the connection between the PNT and the mean value of the Moebius function lies much deeper and is more difficult to establish (though still easier than a proof of the PNT). The precise statement is the following.

Theorem 3.8 (Relation between PNT and the Moebius function)**.** *The PNT is equivalent to the relation*

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n\le x}\mu(n)=0, \tag{3.14}$$

i.e., the assertion that the mean value $M(\mu)$ of the Moebius function exists and is equal to 0*.*

The proof of this result will be given in the second part of this section. We first make some remarks and establish several auxiliary results.

Primes and the Moebius function. What is so surprising about this result is that there does not seem to be any obvious connection between the distribution of primes (which is described by the PNT), and the distribution of the values of the Moebius function (which is described by the result that $M(\mu)=0$). If one restricts to squarefree numbers, then the Moebius function encodes the *parity* of the number of prime factors of an integer. The assertion that $M(\mu)=0$ can then be interpreted as saying that the two parities, even and odd, occur with the same asymptotic frequency. More precisely, this may be formulated as follows: Let $Q(x)$ denote the number of squarefree positive integers $\le x$, and $Q_+(x)$, resp. $Q_-(x)$, the number of squarefree positive integers $\le x$ with an even, resp. odd, number of prime factors. Then $\sum_{n\le x}\mu(n)=Q_+(x)-Q_-(x)$, so the relation $M(\mu)=0$ is equivalent to

$$Q_-(x)=Q_+(x)+o(x)=(1/2)Q(x)+o(x)\sim\frac{3}{\pi^2}x\quad(x\to\infty),$$

in view of the asymptotic relation $Q(x)\sim(6/\pi^2)x$. The equivalence between the PNT and the relation $M(\mu)=0$ therefore means that an asymptotic formula for the function $\pi(x)$, which counts positive integers $\le x$ with *exactly one* prime factor, is equivalent to an asymptotic formula for the function $Q_-(x)$, which counts positive integers $\le x$ with *an odd number* of prime factors. That those two counting functions should be so closely related is anything but obvious.

Next, we prove a simple, and surprisingly easy-to-prove, bound for "logarithmic" averages of the Moebius function. This result may be regarded as a Moebius function analogue of Mertens' estimates for "logarithmic" prime number sums given in Theorem 3.4.

Lemma 3.9 (Mertens' type estimate for the Moebius function)**.** *For any* $x \geq 1$,

$$\left| \sum_{n \leq x} \frac{\mu(n)}{n} \right| \leq 1.$$

Proof. Note first that, without loss of generality, we can assume that $x = N$, where N is a positive integer. We then evaluate the sum $S(N) = \sum_{n \leq N} e(n)$, where e is the convolution identity, defined by $e(n) = 1$ if $n = 1$ and $e(n) = 0$ otherwise, in two different ways. On the one hand, by the definition of $e(n)$, we have $S(N) = 1$; on the other hand, writing $e(n) = \sum_{d|n} \mu(d)$ and interchanging summations, we obtain

$$S(N) = \sum_{d \leq N} \mu(d)[N/d] = N \sum_{d \leq N} \frac{\mu(d)}{d} - \sum_{d \leq N} \mu(d)\{N/d\},$$

where $\{t\}$ denotes the fractional part of t. We now bound the latter sum. Since N is an integer, we have $\{N/d\} = 0$ when $d = N$. Thus we can restrict the summation to those terms for which $1 \leq d \leq N-1$, and using the trivial bound $|\mu(d)\{N/d\}| \leq 1$ for these terms, we see that this sum is bounded by $N - 1$. Hence,

$$\left| N \sum_{d \leq N} \frac{\mu(d)}{d} \right| \leq (N-1) + |S(N)| = (N-1) + 1 = N,$$

which gives the asserted bound for $x = N$. □

Corollary 3.10 (Logarithmic mean value of the Moebius function)**.** *The Moebius function has logarithmic mean value* $L(\mu) = 0$*. Moreover, if the ordinary mean value* $M(\mu)$ *exists, it must be equal to* 0.

Proof. The first statement follows immediately from the definition of the logarithmic mean value and Theorem 3.9. The second statement follows from the first and the general result (Theorem 2.13) that if the ordinary mean value $M(f)$ of an arithmetic function f exists, then the logarithmic mean value $L(f)$ exists as well, and the two mean values are equal. □

Our second auxiliary result is a general result relating the ordinary mean value of an arithmetic function to a mean value involving logarithmic weights. Its proof is a simple exercise in partial summation and is omitted here.

Lemma 3.11. *Given an arithmetic function f, define a mean value $H(f)$ by*

$$H(f) = \lim_{x\to\infty} \frac{1}{x\log x} \sum_{n\le x} f(n)\log n,$$

if the limit exists. Then $H(f)$ exists if and only if the ordinary mean value $M(f)$ exits.

We are now ready to prove the main result of this section.

Proof of 3.8. The proof of this result is longer and more complex than any of the proofs we have encountered so far. Yet it is still easier than a proof of the PNT itself. Its proof requires much of the arsenal of tools and tricks we have assembled so far: convolution identities between arithmetic functions, partial summation, convolution arguments, the Dirichlet hyperbola method, and an estimate for sums of the divisor functions.

We will use the fact that the PNT is equivalent to the relation

$$\psi(x) \sim x \quad (x\to\infty). \tag{3.15}$$

We will also use the result of Lemma 3.11 above, according to which (3.14) is equivalent to

$$\lim_{x\to\infty} \frac{1}{x\log x} \sum_{n\le x} \mu(n)\log n = 0. \tag{3.16}$$

To prove Theorem 3.8 it is therefore enough to show the implications (i) (3.15) $\Rightarrow$ (3.16) and (ii) (3.14) $\Rightarrow$ (3.15).

(i) Proof of (3.15) $\Rightarrow$ (3.16): This is the easier direction. The proof rests on the following identity which is a variant of the identity $\log = 1*\Lambda$.

Lemma 3.12. *We have*

$$\mu(n)\log n = -(\mu*\Lambda)(n) \quad (n\in\mathbb{N}). \tag{3.17}$$

Proof. Suppose first that n is squarefree. In this case we have, for any divisor d of n, $\mu(n) = \mu(d)\mu(n/d)$ (since any such divisor d must be squarefree and relatively prime to its complementary divisor n/d) and $\mu(d)\Lambda(d) = -\Lambda(d)$

(since for squarefree d, $\Lambda(d)$ is zero unless d is a prime, in which case $\mu(d) = -1$). Thus, multiplying the identity $\log n = \sum_{d|n} \Lambda(n/d)$ by $\mu(n)$, we obtain

$$\mu(n)\log n = \sum_{d|n} \mu(d)\mu(n/d)\Lambda(n/d) = -\sum_{d|n} \mu(d)\Lambda(n/d),$$

which proves (3.17) for squarefree n. If n is not squarefree, the left-hand side of (3.17) is zero, so it suffices to show that $(\mu * \Lambda)(n) = 0$ for non-squarefree n. Now $(\mu * \Lambda)(n) = \sum_{p^m|n}(\log p)\mu(n/p^m)$. If n is divisible by the squares of at least two primes, then none of the numbers n/p^m occurring in this sum is squarefree, so the sum vanishes. On the other hand, if n is divisible by exactly one square of a prime, then n is of the form $n = p_0^{m_0} n_0$ with $m_0 \geq 2$ and n_0 squarefree, $(n_0, p_0) = 1$, and the above sum reduces to $(\log p_0)\mu(n_0) + (\log p_0)\mu(n_0 p_0)$, which is again zero since $\mu(n_0 p_0) = \mu(n_0)\mu(p_0) = -\mu(n_0)$. This completes the proof of the lemma. □

Now, suppose (3.15) holds, and set

$$H(x) = \sum_{n \leq x} \mu(n) \log n.$$

We need to show that, given $\epsilon > 0$, we have $|H(x)| \leq \epsilon x \log x$ for all sufficiently large x. From the identity (3.17) we have

$$\begin{aligned} H(x) &= -\sum_{n \leq x} \sum_{d|n} \mu(d)\Lambda(n/d) \\ &= -\sum_{d \leq x} \mu(d) \sum_{m \leq x/d} \Lambda(m) = -\sum_{d \leq x} \mu(d)\psi(x/d). \end{aligned}$$

Let $\epsilon > 0$ be given. By the hypothesis (3.15) there exists $x_0 = x_0(\epsilon) \geq 1$ such that, for $x \geq x_0$, $|\psi(x) - x| \leq \epsilon x$. Moreover, we have trivially $|\Psi(x)| \leq \sum_{n \leq x} \log n \leq x \log x$ for all $x \geq 1$. Applying the first bound with x/d in place of x for $d \leq x/x_0$, and the second (trivial) bound for $x/x_0 < d \leq x$, we obtain, for $x \ge x_0$,

$$
\begin{aligned}
(3.18)\qquad |H(x)| &\le \left| \sum_{d \le x/x_0} \mu(d) \frac{x}{d} \right| + \sum_{d \le x/x_0} |\mu(d)| \left| \psi\left(\frac{x}{d}\right) - \frac{x}{d} \right| \\
&\quad + \sum_{x/x_0 < d \le x} |\mu(d)| \left| \psi\left(\frac{x}{d}\right) \right| \\
&\le \left| \sum_{d \le x/x_0} \mu(d) \frac{x}{d} \right| + \sum_{d \le x/x_0} |\mu(d)| \epsilon \frac{x}{d} \\
&\quad + \sum_{x/x_0 < d \le x} |\mu(d)| \frac{x}{d} \log(x/d) \\
&= \left| \sum\nolimits_1 \right| + \sum\nolimits_2 + \sum\nolimits_3,
\end{aligned}
$$

say. Of the three sums here, the first is bounded by

$$\left| \sum\nolimits_1 \right| = x \left| \sum_{d \le x/x_0} \frac{\mu(d)}{d} \right| \le x$$

by Lemma 3.9. The second sum is bounded by

$$\sum\nolimits_2 \le \epsilon x \sum_{d \le x/x_0} \frac{1}{d} = \epsilon x(\log(x/x_0) + O(1)) \le \epsilon x(\log x + O(1)),$$

by Theorem 2.5. The third sum satisfies

$$\sum\nolimits_3 \le (\log x_0) x_0 \sum_{d \le x} 1 \le (\log x_0) x_0 x,$$

and hence is of order $O_\epsilon(x)$ (with the O-constant depending on ϵ via x_0). Collecting these estimates, we obtain

$$|H(x)| \le \epsilon x \log x + O_\epsilon(x) \quad (x \ge x_0),$$

which implies

$$\limsup_{x \to \infty} \frac{|H(x)|}{x \log x} \le \epsilon.$$

Since ϵ was arbitrary, the limsup must be zero, i.e., (3.16) holds.

(ii) Proof of (3.14) $\Rightarrow$ (3.15): The proof rests on the identity $\Lambda = \log * \mu$, which follows from $\log = 1 * \Lambda$ by Moebius inversion. However, a

direct application of this identity is not successful: Namely, writing $\Lambda(n) = \sum_{d|n}(\log d)\mu(n/d)$ and inverting the order of summation as usual yields

$$\psi(x) = \sum_{n\le x}\Lambda(n) = \sum_{d\le x}(\log d)\sum_{n\le x/d}\mu(n).$$

Assuming a bound of the form $\epsilon x/d$ for the (absolute value of) the inner sum would yield a bound $\le \epsilon\sum_{d\le x}(\log d)(x/d)$, which has order of magnitude $\epsilon x(\log x)^2$ and thus would not even yield a Chebyshev type bound for $\psi(x)$. Even assuming stronger bounds on $\sum_{n\le x}\mu(n)$ (e.g., bounds of the form $O(x/\log^A x)$ for some constant A) would at best yield Chebyshev's bound $\psi(x) \ll x$ in this approach.

To get around these difficulties, we take the unusual (and surprising) step of approximating a smooth function, namely $\log n$, by an arithmetic function that is anything but smooth, but which has nice arithmetic properties. The approximation we choose is the function $f(n) = d(n) - 2\gamma$, where $d = 1*1$ is the divisor function and γ is Euler's constant. This choice is motivated by the fact that the summatory function of $f(n)$ approximates the summatory function of $\log n$ very well. Indeed, on the one hand, Theorem 2.20 gives

$$\begin{aligned}\sum_{n\le x} f(n) &= \sum_{n\le x} d(n) - 2\gamma\sum_{n\le x}1\\ &= x(\log x + 2\gamma - 1) + O(\sqrt{x}) - 2\gamma x + O(1)\\ &= x(\log x - 1) + O(\sqrt{x}),\end{aligned}$$

while, on the other hand, by Corollary 2.8,

$$\sum_{n\le x}\log n = x(\log x - 1) + O(\log x).$$

Thus, if we define the "remainder function" $r(n)$ by

$$\log = f + r = (1*1) - 2\gamma + r,$$

we have

$$\sum_{n\le x} r(n) = \sum_{n\le x}\log n - \sum_{n\le x} f(n) = O(\sqrt{x}) \quad (x\ge 1). \tag{3.19}$$

Replacing the function log by $f+r = (1*1)-2\gamma\cdot 1+r$ in the identity $\Lambda = \mu*\log$, we obtain, using the algebraic properties of the Dirichlet convolution,

$$\begin{aligned}\Lambda &= \mu*(1*1-2\gamma\cdot 1+r)\\ &= (\mu*1)*1 - 2\gamma(\mu*1) + \mu*r = 1 - 2\gamma e + \mu*r,\end{aligned}$$

where e is the usual convolution identity. It follows that

$$\psi(x) = \sum_{n\le x} 1 - 2\gamma + \sum_{n\le x} (\mu * r)(n) = x + O(1) + E(x), \tag{3.20}$$

where

$$E(x) = \sum_{n\le x} (\mu * r)(n).$$

Thus, in order to obtain (3.15), it remains to show that the term $E(x)$ is of order $o(x)$ as $x \to \infty$.

It is instructive to compare the latter sum $E(x)$ with the sum $\psi(x) = \sum_{n\le x} (\mu * \log)(n)$ we started out with. Both of these sums are convolution sums involving the Moebius function. The difference is that, in the sum $E(x)$, the function $\log n$ has been replaced by the function $r(n)$, which, by (3.19), is much smaller on average than the function $\log n$. This makes a crucial difference in our ability to successfully estimate the sum. Indeed, writing

$$E(x) = \sum_{d\le x} \mu(d) \sum_{n\le x/d} r(n)$$

and bounding the inner sum by (3.19) would give the bound

$$E(x) \ll \sum_{d\le x} \sqrt{x/d} = \sqrt{x} \sum_{d\le x} \frac{1}{\sqrt{d}} \ll x,$$

which is by a factor $(\log x)^2$ better than what a similar argument with the function $\log n$ instead of $r(n)$ would have given and strong enough to yield Chebyshev's bound for $\psi(x)$.

Thus, it remains to improve the above bound from $O(x)$ to $o(x)$, by exploiting our assumption (3.14) (which was not used in deriving the above bound). To this end, we use a general version of the Dirichlet hyperbola method: We fix y with $1 \le y \le x$ and split the sum $E(x)$ into

$$E(x) = \sum_{n\le x} \sum_{d|n} r(d)\mu(n/d) = \sum_{dm\le x} r(d)\mu(m) = \sum\nolimits_1 + \sum\nolimits_2 - \sum\nolimits_3, \tag{3.21}$$

where

$$\sum\nolimits_1 = \sum_{d\le y} r(d)M(x/d), \quad \sum\nolimits_2 = \sum_{m\le x/y} \mu(m)R(x/m), \quad \sum\nolimits_3 = R(y)M(x/y),$$

with

$$M(x) = \sum_{n\le x} \mu(n), \quad R(x) = \sum_{n\le x} r(n).$$

We proceed to estimate the three sums arising in the decomposition (3.21). Let $\epsilon > 0$ be given. Then, by our assumption (3.14), there exists $x_0 = x_0(\epsilon)$ such that $|M(x)| \le \epsilon x$ for $x \ge x_0$. Moreover, by (3.19), there exists a constant c such that $|R(x)| \le c\sqrt{x}$ for all $x \ge 1$. Applying these bounds we obtain, for $x \ge x_0 y$,

$$\left|\sum\nolimits_3\right| \le c\sqrt{y}\epsilon(x/y) \le c\epsilon x,$$

and

$$\left|\sum\nolimits_2\right| \le \sum_{m\le x/y} |\mu(m)| c\sqrt{x/m} \le c\sqrt{x} \sum_{m\le x/y} \frac{1}{\sqrt{m}} \le \frac{2cx}{\sqrt{y}},$$

where the latter estimate follows (for example) from Euler's summation formula in the form

$$\begin{aligned} \sum_{n\le t} \frac{1}{\sqrt{n}} &= 1 + \int_1^t \frac{1}{\sqrt{s}} ds - \frac{\{t\}}{\sqrt{t}} - \int_1^t \frac{\{s\}}{s^2} ds \\ &\le 1 + \int_1^t t^{-1/2} dt \le 2\sqrt{t} \quad (t \ge 2). \end{aligned}$$

Finally, we have

$$\left|\sum\nolimits_1\right| \le \sum_{d\le y} |r(d)|\epsilon(x/d) \le C(y)\epsilon x,$$

where

$$C(y) = \sum_{d\le y} \frac{|r(d)|}{d}$$

is a constant depending only on y.

Substituting the above bounds into (3.21), we obtain

$$|E(x)| \le x\left(c\epsilon + C(y)\epsilon + \frac{2c}{\sqrt{y}}\right).$$

It follows that

$$\limsup_{x\to\infty} \frac{|E(x)|}{x} \le \epsilon(c + C(y)) + \frac{2c}{\sqrt{y}},$$

for any fixed $\epsilon > 0$ and $y \ge 1$. Since $\epsilon > 0$ was arbitrary, the above limsup is bounded by $\le 2c/\sqrt{y}$, and since y can be chosen arbitrarily large, it must be equal to 0. Hence $\lim_{x\to\infty} E(x)/x = 0$, which is what we set out to prove. (Note that, for this argument to work it was essential that the constant c did not depend on ϵ or y, and that the constant $C(y)$ did not depend on ϵ.) □

3.5 Exercises

3.1 Let $P(x)=\prod_{p\le x} p$. Show that the PNT is equivalent to the relation $P(x)^{1/x}\to e$ as $x\to\infty$.

3.2 Let $L(n)=[1,2,\dots,n]$, where $[\dots]$ denotes the least common multiple. Show that the limit $\lim_{n\to\infty} L(n)^{1/n}$ exists if and only if the PNT holds.

3.3 Let a_n be a nonincreasing sequence of positive numbers. Show that $\sum_p a_p$ converges if and only if $\sum_{n=2}^{\infty} a_n/\log n$ converges.

3.4 For positive integers k define the generalized von Mangoldt functions Λ_k by the identity $\sum_{d|n}\Lambda_k(d)=(\log n)^k$ (which for $k=1$ reduces to the familiar identity for the ordinary von Mangoldt function $\Lambda(n)$). Show that $\Lambda_k(n)=0$ if n has more than k distinct prime factors.

3.5 Call a positive integer n *round* if it has no prime factors greater than $\sqrt{n}$. Let $R(x)$ denote the number of round integers $\le x$. Estimate $R(x)$ to within an error $O(x/\log x)$. (Hint: Estimate first the slightly different counting function

$$R_0(x)=\#\{n\le x : p|n\Rightarrow p\le\sqrt{x}\},$$

and then show that the difference between $R(x)$ and $R_0(x)$ is of order $O(x/\log x)$ and thus negligible.)

3.6 Let α be a fixed non-zero real number, and let $S_\alpha(x)=\sum_{p\le x} p^{-1-i\alpha}$. Use the prime number theorem in the form $\pi(x)=x/\log x+O(x/\log^2 x)$ to derive an estimate for $S_\alpha(x)$ with error term $O_\alpha(1/\log x)$.

3.7 Let

$$Q(x)=\prod_{p\le x}\left(1+\frac{1}{p}\right).$$

Obtain an estimate for $Q(x)$ with relative error $O(1/\log x)$. Express the constant arising in this estimate in terms of well-known mathematical constants. (Hint: Relate $Q(x)$ to the product $P(x)=\prod_{p\le x}(1-1/p)$ estimated by Mertens' formula.)

3.8 Let

$$R(x) = \prod_{p \le x} \left(1 + \frac{2}{p}\right).$$

Obtain an estimate for $R(x)$ with relative error $O(1/\log x)$. (Constants arising in the estimate need not be evaluated explicitly.

3.9 Using only Mertens' type estimates (but not the PNT), obtain an asymptotic estimate for the partial sums

$$S(x) = \sum_{p \le x} \frac{1}{p \log p}$$

with as good an error term as you can get using only results at the level of Mertens.

3.10 (i) Show that the estimate (a stronger version of one of Mertens' estimates)

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \log x + C + o(1), \tag{1}$$

where C is a constant, implies the PNT.

(ii) (Harder) Show that the converse also holds, i.e., the PNT implies (1).

3.11 Show that

$$\sum_{n \le x} \mu^2(n) = \frac{6}{\pi^2} x + o(\sqrt{x}) \quad (x \to \infty).$$

(With error term $O(\sqrt{x})$ this was proved in Theorem 2.18, quite easily and without using the PNT. To improve this error term to $o(\sqrt{x})$ requires an appeal to the PNT and a more careful treatment of the term that gave rise to the $O(\sqrt{x})$ error.)

3.12 Euler's proof of the infinitude of primes shows that $(*)$ $\sum_{p \le x} 1/p \ge \log\log x - C$, for some constant C and all sufficiently large x. This is a remarkably good lower bound for the sum of reciprocals of primes (it is off by only a term $O(1)$), so it is of some interest to see what this bound implies for $\pi(x)$. The answer is, surprisingly little, as the following problems show.

(i) Deduce from $(*)$, *without using any other information about the primes*, that there exists $\delta > 0$ such that $\pi(x) > \delta \log x$ for all sufficiently large x. In other words, show that if A is any sequence of positive integers satisfying

$$\sum_{a \le x, a \in A} \frac{1}{a} \ge \log\log x - C \tag{1}$$

for some constant C and all sufficiently large x, then there exists a constant $\delta > 0$ such that the counting function $A(x) = \#\{a \in A, a \le x\}$ satisfies

$$A(x) \ge \delta \log x \tag{2}$$

for all sufficiently large x.

(ii) (Harder) Show that this result is nearly best possible, in the sense that it becomes false if the function $\log x$ on the right-hand side of (2) is replaced by a power $(\log x)^\alpha$ with an exponent α greater than 1. In other words, given $\epsilon > 0$, construct a sequence A of positive integers, satisfying (1) above, but for which the counting function $A(x) = \#\{a \in A, a \le x\}$ satisfies

$$\liminf_{x \to \infty} A(x)(\log x)^{-1-\epsilon} = 0. \tag{3}$$