Since $|x^{it}| = 1$ and $|a + id| \geq |a + ic| \geq (a + c)/\sqrt{2}$, the first two terms are each less then $\sqrt{2}\, x^a/(a + c) \log x$ and the integral in the third term is less than

$$\begin{aligned}\int_c^\infty \frac{dt}{a^2 + t^2} &= \int_0^\infty \frac{dt}{a^2 + (c + t)^2} \leq \int_0^\infty \frac{dt}{a^2 + c^2 + t^2} \\ &= \int_0^\infty \frac{(a^2 + c^2)^{1/2}\, du}{a^2 + c^2 + [u(a^2 + c^2)^{1/2}]^2} = \frac{1}{(a^2 + c^2)^{1/2}} \int_0^\infty \frac{du}{1 + u^2} \\ &\leq \frac{\sqrt{2}}{a + c} \cdot \frac{\pi}{2}.\end{aligned}$$

Thus

$$\left|\frac{1}{2\pi i} \int_{a+ic}^{a+id} \frac{x^s\, ds}{s}\right| \leq \frac{1}{2\pi}\left(2\sqrt{2} + \frac{\pi}{\sqrt{2}}\right) \frac{x^a}{(a + c) \log x}$$

which is the desired result.

3.4 THE DENSITY OF THE ROOTS

This section is devoted to the proof of the following theorem:

Theorem The vertical density of the roots ρ of $\xi(\rho) = 0$ is less than 2 $\log T$ for large T. More specifically, there is an H such that for $T \geq H$ the number of roots ρ with imaginary parts in the range $T \leq \operatorname{Im} \rho \leq T + 1$ is less than $2 \log T$.

Proof Von Mangoldt's proof of this fact is based on Hadamard's proof of the product formula and on a strong version of Stirling's formula which was published by Stieltjes in 1889.

Hadamard's theorem that the series $\sum |\rho - \frac{1}{2}|^{-2}$ converges implies (see Sect. 3.2) that the termwise integration of the series $\xi'(s)/\xi(s) = \sum (s - \rho)^{-1}$ is valid over any finite segment. Hence, in particular,

$$\int_{2+iT}^{2+i(T+1)} \frac{\xi'(s)}{\xi(s)}\, ds = \sum_\rho \int_{2+iT}^{2+i(T+1)} \frac{ds}{s - \rho}$$

for any T. Now for any fixed ρ the imaginary part of the integral on the right

$$\operatorname{Im}\left(\int_{2+iT}^{2+i(T+1)} \frac{ds}{s - \rho}\right)$$

is equal (because $dz/z = d \log r + i\, d\theta$) to the angle subtended by the segment $[2 + iT, 2 + i(T + 1)]$ at the point ρ. Thus it is always positive—because the roots ρ are to the left of $\operatorname{Re} s = 2$—and if ρ lies in the range $T \leq \operatorname{Im} \rho \leq T + 1$, then it is at least the angle subtended by the segment $[2 + iT, 2 + i(T + 1)]$ at the point iT, which is $\operatorname{Arctan} \frac{1}{2}$. Thus if n denotes the number of roots ρ in $T \leq \operatorname{Im} \rho \leq T+1$, it follows that

$$n \cdot \operatorname{Arctan} \frac{1}{2} \leq \operatorname{Im} \int_{2+iT}^{2+i(T+1)} \frac{\xi'(s)}{\xi(s)}\, ds. \tag{1}$$

Thus an upper bound on n will result from an upper bound on the integral on the right.

On the other hand, the integral of $\xi'(s)\, ds/\xi(s)$ over the interval from $2 + iT$ to $2 + i(T+1)$ is, by the fundamental theorem of calculus, equal to the amount by which the function

$$\log \xi(s) = \log \Pi(s/2) - (s/2) \log \pi + \log (s-1) + \log \zeta(s)$$

changes between these two points. It is in the estimation of $\log \Pi(s/2)$ that von Mangoldt uses Stieltjes' version of Stirling's formula. Specifically, he uses the fact that the modulus of the error in the approximation

$$\log \Pi(z) \sim (z + \tfrac{1}{2}) \log z - z + \tfrac{1}{2} \log 2\pi$$

is at most $(6|z|)^{-1}$ for z in the halfplane $\operatorname{Re} z \geq 0$. (For a complete statement and proof of Stieltjes' result see Section 6.3.) Thus

$$\begin{aligned}
\log \xi(s) &\sim \left(\frac{s}{2} + \frac{1}{2}\right) \log s - \left(\frac{s}{2} + \frac{1}{2}\right) \log 2 - \frac{s}{2} + \frac{1}{2} \log 2\pi \\
&\quad - \frac{s}{2} \log \pi + \log(s-1) + \log \zeta(s) \\
&= \frac{s+1}{2} \log s - \frac{s}{2} \log 2\pi - \frac{s}{2} \\
&\quad + \log (s-1) + \log \zeta(s) + \text{const}
\end{aligned}$$

with an error of at most $(6|s|)^{-1}$. As was noted in Section 1.13, formula (3), the modulus of $\log \zeta(s)$ is at most $\log \zeta(2)$ on the line $\operatorname{Re} s = 2$, so neglecting this term introduces an error of at most $\log \zeta(2) = \log (\pi^2/6) < 1$. Thus the change in $\log \xi(s)$ between $2 + i(T+1)$ and $2 + iT$ is approximately

$$\begin{aligned}
&\frac{3 + i(T+1)}{2} \log[2 + i(T+1)] - \frac{3 + iT}{2} \log(2 + iT) \\
&\quad - \frac{i}{2} \log 2\pi - \frac{i}{2} + \log \frac{1 + i(T+1)}{1 + iT} \\
&= \frac{i}{2} \log[2 + i(T+1)] + \frac{3 + iT}{2} \log\left(1 + \frac{i}{2 + iT}\right) \\
&\quad + \text{const} + \log\left(1 + \frac{1}{1 + iT}\right)
\end{aligned}$$

and the error has modulus at most $2(6T)^{-1} + 2$. Neglecting terms which are bounded for large T puts this estimate in the form

$$\frac{i}{2} \log(iT) + \frac{3 + iT}{2} \cdot \frac{i}{2 + iT} \sim \frac{i}{2} \log T.$$

Thus the error in the approximation

$$\int_{2+iT}^{2+i(T+1)} \frac{\xi'(s)}{\xi(s)}\,ds \sim \frac{i}{2}\log T$$

remains bounded in modulus as $T \longrightarrow \infty$—say by K—hence (1) gives

$$n \cdot \operatorname{Arctan} \tfrac{1}{2} \le \tfrac{1}{2}\log T + K,$$

$$n \le \frac{\frac{1}{2}\log T + K}{\pi/8} < 2\log T$$

for all sufficiently large T, as was to be shown.

3.5 PROOF OF VON MANGOLDT'S FORMULA FOR $\psi(x)$

The derivation of von Mangoldt's formula

(1) $$\psi(x) = x - \sum_\rho \frac{x^\rho}{\rho} + \sum_n \frac{x^{-2n}}{2n} - \frac{\zeta'(0)}{\zeta(0)} \qquad (x > 1)$$

which was given in Section 3.2 depended on the integral formula which was proved in Section 3.3 and on three termwise integrations. This section is devoted to proving that these termwise integrations are valid. In all three cases, the series converge uniformly on any finite segment $[a - ih, a + ih]$ $(a > 1)$, so the integral can be computed termwise on finite segments and the problem is to show that the limit of their sum as $h \longrightarrow \infty$ is equal to the sum of their limits.

Consider first the limit

(2) $$\lim_{h\to\infty} \sum_{n=1}^{\infty} \Lambda(n) \frac{1}{2\pi i} \int_{a-ih}^{a+ih} \left(\frac{x}{n}\right)^s \frac{ds}{s}.$$

The limit of a finite sum is the sum of the limits; hence one can disregard the finite number of terms in which $n \le x$ and consider only the terms $n > x$. For these, the estimate (1) of Section 3.3 gives

$$\left|\Lambda(n) \frac{1}{2\pi i} \int_{a-ih}^{a+ih} \left(\frac{x}{n}\right)^s \frac{ds}{s}\right| \le \log n \frac{x^a}{n^a \pi h(\log n - \log x)}$$
$$\le \text{const}\frac{1}{n^a h}.$$

Hence their sum over $n > x$ is at most a constant times h^{-1} and therefore approaches zero as $h \longrightarrow \infty$. This shows that the limit (2) can be evaluated termwise.

Consider next the limit

(3) $$\lim_{h\to\infty} \sum_{n=1}^{\infty} \frac{x^{-2n}}{2n} \frac{1}{2\pi i} \int_{a-ih}^{a+ih} \frac{x^{s+2n}\,ds}{s + 2n} \qquad (x > 1).$$

The limit of the nth term of this series is $x^{-2n}/2n$ by (3) of Section 3.3; hence the sum of the limits converges. Now by (4) of Section 3.3 the nth term is at most

$$\frac{x^{-2n}}{2n}\cdot 2\left|\frac{1}{2\pi i}\int_{a+2n}^{a+2n+ih}\frac{x^t\,dt}{t}\right| \leq \frac{x^{-2n}}{n}\cdot K\frac{x^{a+2n}}{(a+2n)\log x} \leq \frac{\text{const}}{n^2}$$

for all h. Thus the series (3) converges uniformly in h and one can pass to the limit $h \longrightarrow \infty$ termwise, as was to be shown. [Given $\epsilon > 0$, choose N large enough that the sum (3) differs by at most ϵ from the sum of the first N terms for all h. By enlarging N if necessary, one can also assume that the sum of the limits $x^{-2n}/2n$ differs by at most ϵ from the sum of the first N of them. Now choose H large enough that each of the first N terms of (3) differs by at most ϵ/N from its limit when $h \geq H$. Then the sum (3) differs by at most 3ϵ from the sum of the limits provided only that $h \geq H$. Since ϵ is arbitrary this proves the desired result.]

Consider finally the limit

$$\lim_{h\to\infty}\sum_{\rho}\frac{x^\rho}{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^{s-\rho}}{s-\rho}\,ds. \tag{4}$$

This limit has now been shown to *exist* because it has been shown that the limit (9) of Section 3.2 exists [and is equal to $\psi(x)$] and it has been shown that the limit of the sum over n in (9) also exists. It has also been shown that the individual terms of (4) approach limits x^ρ/ρ as $h \longrightarrow \infty$, but it has *not* been shown that the sum of these limits converges; in fact, the proof that $\sum x^\rho/\rho$ converges when summed in the order of increasing $|\operatorname{Im}\rho|$ is the major difficulty in the proof of von Mangoldt's formula. Broadly speaking, von Mangoldt overcomes this difficulty by approaching the limit (4) "diagonally," that is, by considering the limit

$$\lim_{h\to\infty}\sum_{|\operatorname{Im}\rho|\leq h}\frac{x^\rho}{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^{s-\rho}}{s-\rho}\,ds \tag{5}$$

as an intermediate step between the limit (4) which is known to exist and the sum of the limits

$$\lim_{h\to\infty}\sum_{|\operatorname{Im}\rho|\leq h}\frac{x^\rho}{\rho} \tag{6}$$

which is to be shown to exist and be equal to (4).

Specifically, consider for each h the differences

$$\sum_{\rho}\frac{x^\rho}{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^{s-\rho}}{s-\rho}\,ds - \sum_{|\operatorname{Im}\rho|\leq h}\frac{x^\rho}{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^{s-\rho}}{s-\rho}\,ds \tag{7}$$

and

$$(8) \qquad \sum_{|\text{Im}\,\rho| \leq h} \frac{x^\rho}{\rho} \frac{1}{2\pi i} \int_{a-ih}^{a+ih} \frac{x^{s-\rho}}{s-\rho}\, ds - \sum_{|\text{Im}\,\rho| \leq h} \frac{x^\rho}{\rho}.$$

It will be shown that both these differences approach zero as $h \to \infty$. Then, since the limit (4) exists, it follows first that the "diagonal" limit (5) exists and is equal to it [because (7) goes to zero] and hence that the limit (6) exists and is equal to it [because (8) goes to zero hence (6) equals (5)] as desired.

Consider first the estimate of (7). Let $\rho = \beta + i\gamma$ denote a typical root. Then by (4) of Section 3.3 the modulus of (7) is at most

$$\begin{aligned}
&\sum_{|\gamma| > h} \left|\frac{x^\rho}{\rho}\right| \left|\frac{1}{2\pi i} \int_{a-ih}^{a+ih} \frac{x^{s-\rho}}{s-\rho}\, ds\right| \\
&\leq 2 \sum_{\gamma > h} \frac{x^\beta}{\gamma} \left|\frac{1}{2\pi i} \int_{a-\beta+i(\gamma-h)}^{a-\beta+i(\gamma+h)} \frac{x^t\, dt}{t}\right| \\
&\leq 2 \sum_{\gamma > h} \frac{x^\beta}{\gamma} \cdot K \frac{x^{a-\beta}}{(a-\beta+\gamma-h)\log x} \\
&\leq 2K \frac{x^a}{\log x} \sum_{\gamma > h} \frac{1}{\gamma(\gamma - h + c)}
\end{aligned}$$

where $c = a - 1 > 0$ so that $c \leq a - \beta$ for all roots ρ. Now if the roots γ beyond h are grouped in intervals $h < \gamma \leq h + 1$, $h + 1 < \gamma \leq h + 2$, $h + 2 < \gamma \leq h + 3, \ldots$, then (assuming h is large enough that the estimate of Section 3.4 applies beyond h) the interval $h + j < \gamma \leq h + j + 1$ contains at most $2 \log (h + j)$ of the γ's and the modulus of (7) is at most a constant times

$$\sum_{j=0}^{\infty} \frac{\log(h+j)}{(h+j)(j+c)},$$

and it remains only to show that this sum approaches zero as $h \to \infty$. One can do this by choosing h large enough that $\log (h + j) < (h + j)^{1/2}$ for all $j \geq 0$; then the summand is at most one over $(h + j)^{1/4}(c + j)^{1/4} \cdot (c + j)$, so the sum is at most a constant times $h^{-1/4} \to 0$ as was to be shown.

Consider now the estimate of (8). The fact that the terms corresponding to ρ and $\bar{\rho}$ give equal contributions and the estimates (3) and (4) of Section 3.3 show that the modulus of (8) is at most

$$\begin{aligned}
&2 \sum_{0 < \gamma \leq h} \left|\frac{x^\rho}{\rho}\right| \left|\frac{1}{2\pi i} \int_{a-\beta-i\gamma-ih}^{a-\beta-i\gamma+ih} \frac{x^t\, dt}{t} - 1\right| \\
&\leq 2 \sum_{0 < \gamma \leq h} \frac{x^\beta}{\gamma} \left|\frac{1}{2\pi i} \int_{a-\beta-i(h+\gamma)}^{a-\beta+i(h+\gamma)} \frac{x^t\, dt}{t} - 1\right| \\
&\quad + 2 \sum_{0 < \gamma \leq h} \frac{x^\beta}{\gamma} \left|\frac{1}{2\pi i} \int_{a-\beta+i(h-\gamma)}^{a-\beta+i(h+\gamma)} \frac{x^t\, dt}{t}\right|
\end{aligned}$$

$$
\begin{aligned}
&\le 2 \sum_{0<\gamma\le h} \frac{x^\beta}{\gamma} \frac{x^{a-\beta}}{\pi(h+\gamma)\log x} \\
&\quad + 2 \sum_{0<\gamma\le h} \frac{x^\beta}{\gamma} K \frac{x^{a-\beta}}{(a-\beta+h-\gamma)\log x} \\
&\le \frac{2x^a}{\pi \log x} \sum_{0<\gamma\le h} \frac{1}{\gamma(h+\gamma)} \\
&\quad + \frac{2Kx^a}{\log x} \sum_{0<\gamma\le h} \frac{1}{\gamma(c+h-\gamma)},
\end{aligned}
$$

where $c = a - 1 > 0$ as before. Thus it suffices to prove that these two sums over γ approach zero as $h \longrightarrow \infty$. Let H be a large integer such that the estimate of Section 3.4 applies beyond H and let the roots be grouped in intervals $H \le \gamma < H + 1, H + 1 \le \gamma < H + 2, \ldots$. Then the interval $H + j \le \gamma \le H + j + 1$ contains at most $2 \log (H + j)$ of the γ's and

$$
\sum_{0<\gamma\le h} \frac{1}{\gamma(h+\gamma)} \le \sum_{0<\gamma\le H} \frac{1}{\gamma(h+\gamma)} + \sum_{0\le j\le h-H} \frac{2\log(H+j)}{(H+j)(h+H+j)}.
$$

The first sum has a fixed finite number of terms and therefore clearly has the limit zero as $h \longrightarrow \infty$. The second sum is at most

$$
\begin{aligned}
&2 \sum_{0\le j\le h-H} (\log h)\left[\frac{1}{h}\left(\frac{1}{H+j} - \frac{1}{h+H+j}\right)\right] \\
&\le \frac{\log h}{h} 2 \sum_{0\le j\le h-H} \frac{1}{H+j} \\
&\le 2\frac{\log h}{h} \int_{H-1}^{h} \frac{dt}{t} \le 2\frac{(\log h)^2}{h}
\end{aligned}
$$

which approaches zero as $h \longrightarrow \infty$. A similar estimate shows that

$$
\sum_{H\le\gamma\le h} \frac{1}{\gamma(c+h-\gamma)}
$$

approaches zero as $h \longrightarrow \infty$ and completes the proof that (6) equals (4).

Thus the limits (2), (3), and (4) can be evaluated termwise—provided the sum of the limits of (4) is defined in the sense of (6)—and the derivation of Section 3.2 proves von Mangoldt's formula (1).

3.6 RIEMANN'S MAIN FORMULA

There are at least two reasons why Riemann's formula for $J(x)$ [Section 3.1, formula (1)] is generally neglected today. First, it contains essentially the same information as von Mangoldt's formula for $\psi(x)$ but is less "natural" than this formula in the sense that it is harder to prove and harder to generalize. Secondly, Riemann's reason for establishing the formula in the first place—to show an explicit analytic connection between the arithmetic function $\pi(x)$ and the empirically derived approximation $\mathrm{Li}(x)$—was rendered superfluous by Chebyshev's observation that the prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ can be deduced from the more natural theorem $\psi(x) \sim x$. Thus, in all respects, the formula for $\psi(x)$ is preferable to that for $J(x)$.

Nonetheless, it is the formula for $J(x)$ that was stated† by Riemann and for this reason, if for no other, it is of great interest to know whether or not the formula is valid. Von Mangoldt proved that it is. Von Mangoldt did not, however, follow Riemann's method of proving the formula for $J(x)$; once the product formula for $\xi(s)$ was established by Hadamard, the only real difficulty which remained in Riemann's derivation of the formula for $J(x)$ was the proof that the termwise integration of the sum over ρ (Section 1.15) is valid, but von Mangoldt does not justify this termwise integration directly and proves the formula for $J(x)$ by a quite different method. It may be that he did this simply as a matter of convenience or it may be that he was in fact unable to derive estimates which would justify the termwise integration of Section 1.15 directly. In any event, Landau [L2] in 1908 proved in a more or less direct manner that termwise integration is valid.

3.7 VON MANGOLDT'S PROOF OF RIEMANN'S MAIN FORMULA

For $r > 0$ consider the definite integral

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left[-\frac{\zeta'(s+r)}{\zeta(s+r)}\right]x^s\,ds \qquad (a > 1, \quad x > 1). \tag{1}$$

When the formula $-\zeta'(s+r)/\zeta(s+r) = \sum \Lambda(n)n^{-s-r}$ is substituted in this equation and integration is carried out termwise, one obtains the value $\sum_{n<x} \Lambda(n)n^{-r}$ where, as usual, at any point $x = p^n$ where the value jumps, it is defined to split the difference. Von Mangoldt denotes this function by $\psi(x, r)$. An equivalent definition of $\psi(x, r)$ is

$$\psi(x, r) = \int_0^x x^{-r}\,d\psi(x). \tag{2}$$

On the other hand, the derivation of the formula

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{s}{s-1} - \sum_\rho \frac{s}{\rho(s-\rho)} + \sum_n \frac{s}{2n(s+2n)} - \frac{\zeta'(0)}{\zeta(0)}$$

†Landau [L3] began a tradition of referring to this and other statements of Riemann as "conjectures," which gives the very mistaken impression that Riemann had some doubts about them. The only conjecture which Riemann makes in the paper is that Re $\rho = \frac{1}{2}$.

of Section 3.2 is easily modified to give the more general formula

$$-\frac{\zeta'(s+r)}{\zeta(s+r)} = -\frac{s}{(r-1)(s+r-1)} + \sum_\rho \frac{s}{(r-\rho)(s+r-\rho)} + \sum_n \frac{s}{(r+2n)(s+r+2n)} - \frac{\zeta'(r)}{\zeta(r)}$$

which is valid for $r > 0$ except at $r = 1$. When this expression is substituted in (1) and the integration is carried out termwise, the result is

$$\frac{x^{1-r}}{1-r} - \sum_\rho \frac{x^{\rho-r}}{\rho-r} + \sum_n \frac{x^{-2n-r}}{2n+r} - \frac{\zeta'(r)}{\zeta(r)}. \tag{3}$$

The estimates of Section 3.5 can be modified to show that for fixed $r > 0$, $x > 1$, $r \neq 1$ the termwise integrations are valid and hence that

$$\int_0^x x^{-r}\, d\psi(x) = \frac{x^{1-r}}{1-r} - \sum_\rho \frac{x^{\rho-r}}{\rho-r} + \sum_n \frac{x^{-2n-r}}{2n+r} - \frac{\zeta'(r)}{\zeta(r)}.$$

The first and last terms on the right have poles at $r = 1$ which cancel each other, so the entire right side defines a continuous function of r as can be seen by rewriting it in the form

$$\int_0^x x^{-r}\, d\psi(x) = \left[\frac{x^{1-r}}{1-r} - \frac{\zeta'(r)}{\zeta(r)}\right] - x^{-r} \sum_\rho \frac{x^\rho}{\rho} - x^{-r} \sum_\rho \frac{r x^\rho}{\rho(\rho-r)} + \sum_n \frac{x^{-2n-r}}{2n+r} \qquad (x > 1) \tag{4}$$

and noting that near any value of r these series converge uniformly in r and hence define continuous functions of r. Thus the formula is valid for $r = 1$ as well.

Now integrate both sides dr from $r = 0$ to $r = \infty$. On the left one obtains

$$\int_0^\infty \int_0^x x^{-r}\, d\psi(x)\, dr = \int_0^x \left(\int_0^\infty x^{-r}\, dr\right) d\psi(x) = \int_0^x \frac{1}{\log x}\, d\psi(x) = \int_0^x dJ(x) = J(x).$$

(Or, less elegantly,

$$\int_0^\infty \sum_{n<x} \Lambda(n) n^{-r}\, dr = \sum_{n<x} \frac{\Lambda(n)}{\log n} = J(x)$$

with the usual adjustment if x is a prime power.) The sum over n on the right can be integrated termwise by the Lebesgue dominated convergence theorem because it is dominated by $x^{-r} \sum_n x^{-2n}$. The result is

$$\sum_n \int_0^\infty \frac{x^{-2n-r}\, dr}{2n+r} = \sum_n \int_{2n}^\infty \frac{x^{-u}\, du}{u} = \sum_n \int_{2n}^\infty \int_x^\infty t^{-u-1}\, dt\, du = \sum_n \int_x^\infty \frac{t^{-2n-1}}{\log t}\, dt = \int_x^\infty \frac{dt}{t(t^2-1)\log t}.$$

The first sum over ρ does not involve r and can therefore be integrated termwise. The second sum over ρ can be integrated termwise by the Lebesgue dominated convergence theorem because it is dominated by $rx^{1-r}\sum|\rho|^{-2}$. The result of these two termwise integrations is

$$-\sum_{\rho}\left[\int_0^\infty x^{-r}\frac{x^\rho}{\rho}\,dr + \int_0^\infty x^{-r}\frac{rx^\rho}{\rho(\rho-r)}\,dr\right]$$
$$= -\sum_{\rho}\int_0^\infty \frac{x^{\rho-r}}{\rho-r}\,dr = -\sum_{\operatorname{Im}\rho>0}[\operatorname{Li}(x^\rho)+\operatorname{Li}(x^{1-\rho})]$$

as will now be shown. [Here, as in (4), the sum over ρ is to be taken in the order of increasing $|\operatorname{Im}\rho|$.]

The formula needed above is

$$\int_0^\infty \frac{x^{\rho-r}}{\rho-r}\,dr = \operatorname{Li}(x^\rho) \mp i\pi, \tag{5}$$

where $x > 1$, $\operatorname{Re}\rho > 0$, and† where the sign of $i\pi$ is opposite to that of $\operatorname{Im}\rho$. This can be proved by setting $t = (\rho - r)\log x$, $dt = -(\log x)\,dr$ to put the integral in the form

$$-\int_\infty^0 \frac{x^{\rho-r}(\log x)\,dr}{(\rho-r)\log x} = \int_{\rho\log x-\infty\log x}^{\rho\log x}\frac{e^t\,dt}{t}.$$

Since the integrand vanishes very rapidly near $-\infty$, the lower limit of integration can be taken as $-\infty$; hence

$$\int_0^\infty \frac{x^{\rho-r}\,dr}{\rho-r} = \int_{-\infty}^{\rho\log x}\frac{e^t\,dt}{t},$$

where the path of integration passes above the singularity at $t = 0$ if $\operatorname{Im}\rho > 0$, below if $\operatorname{Im}\rho < 0$. Now, if it is stipulated that the path of integration in the integral

$$\int_{-\infty}^{\beta\log x}\frac{e^t\,dt}{t}$$

must enter the halfplane $\operatorname{Re} t > 0$ by crossing the *positive* imaginary axis, then this integral defines an analytic function of β in $\operatorname{Re}\beta > 0$ which for real β is equal to

$$\int_0^{x^\beta}\frac{du}{\log u} = \operatorname{Li}(x^\beta) - i\pi$$

(see Section 1.15). Hence the same is true by analytic continuation for all

†Thus, as with Riemann, the possibility $\operatorname{Re}\rho = 0$ is not included. The extension of the formula to cover this case is trivial, but, as Hadamard showed (see Section 4.2), none of the ρ's lies on the imaginary axis.

β in Re $\beta > 0$, and

$$\int_0^\infty \frac{x^{\rho-r}}{\rho - r}\,dr = \operatorname{Li}(x^\rho) - i\pi$$

for Im $\rho > 0$ follows. The case Im $\rho < 0$ is analogous and (5) is proved.

It remains only to show that the remaining pair of terms in (4) when integrated dr from 0 to ∞ give the remaining terms in Riemann's formula, that is, to show that

$$\int_0^\infty \left[\frac{x^{1-r}}{1-r} - \frac{\zeta'(r)}{\zeta(r)}\right] dr = \operatorname{Li}(x) + \log \xi(0).$$

This can be proved as follows.

The two terms can be integrated separately

$$\int_0^\infty \left[\frac{x^{1-r}}{1-r} - \frac{\zeta'(r)}{\zeta(r)}\right] dr = \int_0^\infty \frac{x^{1-r}}{1-r}\,dr - \int_0^\infty \frac{\zeta'(r)}{\zeta(r)}\,dr$$

provided the path from 0 to ∞ is perturbed slightly to avoid the singularity at $r = 1$, say by passing slightly above it. The first integral is then the limiting case $\rho = 1 - i\epsilon$ of formula (5) and is therefore $\operatorname{Li}(x) + i\pi$. The second integral can be evaluated by integrating

$$\frac{d}{dr}\log \xi(r) = \frac{d}{dr}\log\left[\pi^{-r/2}\,\Pi\left(\frac{r}{2}\right)\right] + \frac{1}{r-1} + \frac{\zeta'(r)}{\zeta(r)}$$

from 0 to K passing above $r = 1$ to obtain

$$\log \xi(K) - \log \xi(0) = \log \pi^{-K/2}\,\Pi\left(\frac{K}{2}\right) + \log(K-1) - i\pi$$
$$+ \int_0^K \frac{\zeta'(r)}{\zeta(r)}\,dr,$$
$$-\int_0^K \frac{\zeta'(r)}{\zeta(r)}\,dr = \log \xi(0) - i\pi + \log \zeta(K).$$

The limit as $K \longrightarrow \infty$ is thus $\log \xi(0) - i\pi$ and the desired formula follows.

In summary, then, when (4) is integrated dr from 0 to ∞, the result is Riemann's main formula

$$J(x) = \operatorname{Li}(x) - \sum_\rho \operatorname{Li}(x^\rho) + \int_x^\infty \frac{dt}{t(t^2-1)\log t} + \log \xi(0) \qquad (x > 1)$$

which is thereby proved.

3.8 NUMERICAL EVALUATION OF THE CONSTANT

Von Mangoldt found that the numerical value of the constant in the formula for $\psi(x)$ is

$$\zeta'(0)/\zeta(0) = \log 2\pi. \tag{1}$$

The series $\Sigma(x^{-2n}/2n)$ can also be summed using the series expansion

$$\log [1/(1 - x)] = x + \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 + \cdots$$

to put the formula for $\psi(x)$ in the form

$$\psi(x) = x - \sum_\rho \frac{x^\rho}{\rho} + \frac{1}{2}\log\left(\frac{x^2}{x^2 - 1}\right) - \log 2\pi \qquad (x > 1),$$

where the sum over ρ is in the order of increasing $|\operatorname{Im} \rho|$.

The value (1) of the constant can be obtained as follows.† Consider the function

$$\frac{\zeta(s)}{\Pi(-s)} = \frac{1}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)^s}{e^x - 1} \cdot \frac{dx}{x}, \tag{2}$$

where the path of integration is as in Section 1.4. It will first be shown that the derivative of this function is zero at $s = 1$, that is, it will be shown that

$$\frac{1}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)\log(-x)}{e^x - 1}\frac{dx}{x} = 0. \tag{3}$$

Let the path of integration be written as a sum of three parts as in Section 1.4 so that this integral becomes

$$\begin{aligned}
&\frac{1}{2\pi i}\int_\infty^\epsilon \frac{(-x)(\log x - i\pi)}{e^x - 1}\frac{dx}{x} \\
&\qquad + \frac{1}{2\pi i}\int_{|x|=\epsilon} \frac{(-x)(\log \epsilon + i\theta - i\pi)}{e^x - 1}\frac{dx}{x} \\
&\qquad + \frac{1}{2\pi i}\int_\epsilon^\infty \frac{(-x)(\log x + i\pi)}{e^x - 1}\frac{dx}{x} \\
&= -\int_\epsilon^\infty \frac{dx}{e^x - 1} - \frac{\log \epsilon}{2\pi i}\int_{|x|=\epsilon} \frac{x}{e^x - 1} \cdot \frac{dx}{x} \\
&\qquad - \frac{1}{2\pi i}\int_{-\pi}^{\pi} \frac{x}{e^x - 1}\phi \, d\phi,
\end{aligned}$$

where $x = \epsilon e^{i(\phi+\pi)}$ in the last integral. Since $x(e^x - 1)^{-1}$ is 1 at $x = 0$, the middle integral is $(-\log \epsilon)$ by the Cauchy integral formula and the last integral approaches zero as $\epsilon \downarrow 0$.

†For an alternative proof see Section 6.8.

The first integral can be evaluated directly

$$-\int_\epsilon^\infty \frac{dx}{e^x - 1} = -\int_\epsilon^\infty \left(\sum_{n=1}^\infty e^{-nx}\right) dx = -\left[\sum_{n=1}^\infty \frac{e^{-nx}}{-n}\right]\Bigg|_\epsilon^\infty = -\sum \frac{(e^{-\epsilon})^n}{n}$$

$$= \log(1 - e^{-\epsilon}) = \log\left(\epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{6} - \cdots\right)$$

$$= \log \epsilon + \log\left(1 - \frac{\epsilon}{2} + \cdots\right).$$

Thus the $\log \epsilon$'s cancel and the limit as $\epsilon \downarrow 0$ of the remaining terms is zero, which proves (3). Now by the functional equation the function (2) can also be written in the form

$$\zeta(s)/\Pi(-s) = (2\pi)^{s-1}\zeta(1 - s)2 \sin(\pi s/2).$$

Since its derivative is zero at $s = 1$, its logarithmic derivative

$$\log(2\pi) - \frac{\zeta'(1 - s)}{\zeta(1 - s)} + \frac{\pi}{2}\frac{\cos(\pi s/2)}{\sin(\pi s/2)}$$

must also be zero at $s = 1$, which gives (1).

Since the logarithmic derivative of $\xi(s)$ is on the one hand

$$\sum_\rho \frac{d}{ds}\log\left(1 - \frac{s}{\rho}\right) = \sum_\rho \frac{1}{s - \rho}$$

and on the other hand

$$\frac{d}{ds}\log \Pi\left(\frac{s}{2}\right) - \frac{1}{2}\log \pi + \frac{1}{s - 1} + \frac{\zeta'(s)}{\zeta(s)},$$

the sum of the series $\sum (1/\rho)$ is the value of

$$-\frac{1}{2}\frac{\Pi'(s/2)}{\Pi(s/2)} + \frac{1}{2}\log \pi + \frac{1}{1 - s} - \frac{\zeta'(s)}{\zeta(s)}$$

at $s = 0$. Now logarithmic differentiation of the product formula for $\Pi(x)$ [(4) of Section 1.3] gives

$$\frac{\Pi'(s)}{\Pi(s)} = \sum_{n=1}^\infty \left[-\frac{1}{s + n} - \log n + \log(n + 1)\right]$$

$$-\frac{\Pi'(0)}{\Pi(0)} = \lim_{n\to\infty}\left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log(n + 1)\right].$$

The number on the right side of this equation is by definition *Euler's constant*, and is traditionally denoted γ. Thus

(4) $$\sum_\rho \frac{1}{\rho} = \frac{1}{2}\gamma + \frac{1}{2}\log \pi + 1 - \log 2\pi.$$

This formula was known to Riemann, who used it in his computations of the roots ρ (see Section 7.6 below). From this it is clear that the formula (1) was also known to Riemann.

Chapter 4

The Prime Number Theorem

4.1 INTRODUCTION

The prime number theorem is the statement that the relative error in the approximation $\pi(x) \sim \text{Li}(x)$ approaches zero as $x \longrightarrow \infty$. Following the work of Chebyshev it was well known that the prime number theorem could be deduced from the theorem that the relative error in the approximation $\psi(x) \sim x$ approaches zero as $x \longrightarrow \infty$. But von Mangoldt's formula for $\psi(x)$ shows that $\psi(x) \sim x$ has this property if and only if

$$\lim_{x\to\infty} \frac{-\sum_\rho (x^\rho/\rho) + \sum_n (x^{-2n}/2n) + \text{const}}{x} = 0$$

or, what is the same, if and only if

$$\lim_{x\to\infty} \sum_\rho \frac{x^{\rho-1}}{\rho} = 0. \tag{1}$$

If the limit of this sum could be taken termwise, then it would suffice to prove that $x^{\rho-1} \longrightarrow 0$ for all ρ or, what is the same, that $\text{Re}\,\rho < 1$ for all ρ. Since $\text{Re}\,\rho \leq 1$ for all ρ (by the Euler product formula—see Section 1.9), this amounts to proving that there are no roots ρ on the line $\text{Re}\,s = 1$. Thus, given von Mangoldt's 1894 formula for $\psi(x)$, the proof of the prime number theorem can be reduced to proving that there are no roots ρ on the line Re $s = 1$ and to proving that the above limit can be evaluated termwise.

Both Hadamard [H2] and de la Vallée Poussin [V1] succeeded in 1896 in filling in these remaining steps in the proof of the prime number theorem. They both circumvented proving that the limit (1) can be evaluated termwise and instead they each derived a variation of von Mangoldt's formula, namely,

(2) $$\int_0^x t^{-2}\psi(t)\,dt = \log x - \sum_\rho \frac{x^{\rho-1}}{\rho(\rho-1)} - \sum_n \frac{x^{-2n-1}}{2n(2n+1)} + \frac{1}{x}\frac{\zeta'(0)}{\zeta(0)} + \text{const} \qquad (x > 1)$$

in de la Vallée Poussin's case† and

(3) $$\int_0^x t^{-1}\psi(t)\,dt = x - \sum_\rho \frac{x^\rho}{\rho^2} - \sum_n \frac{x^{-2n}}{(2n)^2} - \frac{\zeta'(0)}{\zeta(0)}\log x + \text{const} \qquad (x > 1)$$

in Hadamard's case. Either of these formulas is quite easy to prove using von Mangoldt's methods—easier‡ in fact than von Mangoldt's formula for $\psi(x)$—and if it is known that $\operatorname{Re}\rho < 1$ for all roots ρ, then either of them can be used to conclude by straightforward estimates that $\psi(x) \sim x$. Thus, although it certainly required insight to see that a formula such as (2) or (3) could be used, the substantial step beyond von Mangoldt's work which was required for the proof of the prime number theorem was the proof that there are no roots ρ on the line $\operatorname{Re} s = 1$.

Hadamard's proof that there are no roots ρ on $\operatorname{Re} s = 1$ is given in Section 4.2. De la Vallée Poussin admitted that Hadamard's proof was the simpler of the two, and although simpler proofs have since been found (see Section 5.2), Hadamard's is perhaps still the most straightforward and natural proof of this fact. Section 4.3 is devoted to a proof that $\psi(x) \sim x$. This proof follows the same general line of argument as was followed by both Hadamard and de la Vallée Poussin, but it is somewhat simpler in that it is based on the formula

(4) $$\int_0^x \psi(t)\,dt = \frac{x^2}{2} - \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - \sum_n \frac{x^{-2n+1}}{2n(2n-1)} - x\frac{\zeta'(0)}{\zeta(0)} + \text{const} \qquad (x > 1)$$

rather than on the analogous, but somewhat more complicated, formulas (2) or (3). Finally, Section 4.4 gives the very simple deduction of $\pi(x) \sim \operatorname{Li}(x)$ from $\psi(x) \sim x$.

†For the value of the constant, which is $-1 - \gamma$, see footnote †, p. 94.

‡At the time, de la Vallee Poussin was under the mistaken impression that von Mangoldt's proof was fallacious, so of course he gave his own proof. Hadamard too preferred to avoid making appeal to von Mangoldt's more difficult estimates and produced his own proof.

4.2 HADAMARD'S PROOF THAT Re $\rho < 1$ FOR ALL ρ

The representation

$$\log \zeta(s) = \int_0^\infty x^{-s}\, dJ(x) \tag{1}$$

$$= \sum_p \frac{1}{p^s} + \frac{1}{2}\sum_p \frac{1}{p^{2s}} + \frac{1}{3}\sum_p \frac{1}{p^{3s}} + \cdots$$

is valid throughout the halfplane $\text{Re}\, s > 1$. The presence of zeros ρ of $\zeta(s)$ on the line $\text{Re}\, s = 1$ would imply the presence of points $s = \sigma + it$ slightly to the right of $\text{Re}\, s = 1$ where $\text{Re} \log \zeta(s)$ was near $-\infty$. The series in (1) has the property that the sum of the terms after the first is bounded by the number

$$B = \frac{1}{2}\sum_p \frac{1}{p^2} + \frac{1}{3}\sum_p \frac{1}{p^3} + \frac{1}{4}\sum_p \frac{1}{p^4} + \cdots$$

for the entire halfplane $\text{Re}\, s \geq 1$ including $\text{Re}\, s = 1$; hence

$$\text{Re} \log \zeta(\sigma + it) \geq \sum_p \frac{\cos(t \log p)}{p^\sigma} - B$$

can approach $-\infty$ as $\sigma \downarrow 1$ only if the first term approaches $-\infty$. In short, if $1 + it$ were a zero ρ of $\zeta(s)$, then it would follow that

$$\lim_{\sigma \downarrow 1} \sum_p \frac{\cos(t \log p)}{p^\sigma} = -\infty \tag{2}$$

for this value of t. The objective is to show that this is impossible.

Now the fact that

$$\lim_{s \to 1} (s-1)\zeta(s) = \frac{\xi(1)}{\Pi(\frac{1}{2})\pi^{-1/2}} = \frac{(\frac{1}{2})}{\frac{1}{2}\pi^{1/2}\pi^{-1/2}} = 1$$

implies that

$$\lim_{\sigma \downarrow 1} [\text{Re} \log \zeta(\sigma) + \log(\sigma - 1)] = 0,$$

$$\lim_{\sigma \downarrow 1} \left[\sum_p \frac{1}{p^\sigma} + \text{bounded} + \log(\sigma - 1)\right] = 0,$$

which means that for σ slightly larger than one

$$\sum_p \frac{1}{p^\sigma} \sim -\log(\sigma - 1) \sim +\infty. \tag{3}$$

On the other hand, the derivation of (2) is easily strengthened to give

$$\lim_{\sigma \downarrow 1} \left[\sum_p \frac{\cos(t \log p)}{p^\sigma} + \text{bounded} - \text{Re} \log(\sigma - 1)\right] = 0,$$

$$\sum_p \frac{\cos(t \log p)}{p^\sigma} \sim \log(\sigma - 1) \sim -\infty. \tag{4}$$

If this were true then, because of (3), one would expect that the number $\cos(t \log p)$ would have to be nearly -1 for the overwhelming majority of primes p. This would imply a surprising regularity in the distribution of the numbers $\log p$, namely, that most of them lie near the points of the arithmetic progression $(2n + 1)t^{-1}\pi$ for this particular t. However, such a regularity cannot exist because it would imply that $\cos(2t \log p)$ was nearly $+1$ for the overwhelming majority of primes p, hence

$$\sum_p \frac{\cos(2t \log p)}{p^\sigma} \sim +\infty,$$

$$\operatorname{Re} \log \zeta(\sigma + 2it) \sim +\infty$$

which would imply the existence of a pole of $\zeta(s)$ at $s = 1 + 2it$. Since $\zeta(s)$ has no poles other than $s = 1$, this line of argument might be expected to yield a proof of the impossibility of (4) as desired. The actual proof requires little more than a quantitative description of the "overwhelming majority" of the primes p for which $t \log p \sim (2n + 1)\pi$ would have to be true.

Assume that $1 + it$ is a zero of $\zeta(s)$. Then $\zeta(s)/(s - 1 - it)$ would be analytic near $s = 1 + it$ so the real part of its log would be bounded above, say by K; hence with $s = \sigma + it$

$$\sum_p \frac{\cos(t \log p)}{p^\sigma} + \frac{1}{2}\sum_p \frac{\cos(2t \log p)}{p^{2\sigma}} + \frac{1}{3}\sum_p \frac{\cos(3t \log p)}{p^{3\sigma}} + \cdots - \operatorname{Re} \log(\sigma - 1) < K,$$

(5)
$$\sum_p \frac{\cos(t \log p)}{p^\sigma} < \log(\sigma - 1) + K + B$$

for all $\sigma > 1$ near 1. This is a quantitative version of (4); the objective is to show that no t has this property.

Let δ be a small positive number (for the sake of definiteness, $\delta = \pi/8$ will work in the following proof) and let the terms of the sum on the left side of (5) be divided into those terms corresponding to primes p for which there is an n such that

(6)
$$|(2n + 1)\pi - t \log p| < \delta$$

and those terms for which p does not satisfy this condition. In the terms of the second group $\cos(t \log p) = \cos(\pi - \alpha) = -\cos \alpha$ where $\delta \le |\alpha| \le \pi$, hence $\cos(t \log p) > -\cos \delta$ for these terms and (5) implies

$$-S' - (\cos \delta)S'' < \log(\sigma - 1) + K + B$$

where S' denotes the sum of $p^{-\sigma}$ over all primes p which satisfy (6) and S'' denotes the sum of $p^{-\sigma}$ over all primes p which do not. Now (3) says that $S' + S'' \sim -\log(\sigma - 1)$, and the derivation of (3) shows easily that there is a K' such that $S' + S'' < -\log(\sigma - 1) + K'$ for all $\sigma > 1$ near 1. Thus (5) implies

$$-S' - (\cos\delta)S'' < -S' - S'' + K' + K + B,$$
$$(1 - \cos\delta)S'' < \text{const}$$

for all $\sigma > 1$ near 1. Since $S' + S''$ becomes infinite as $\sigma \downarrow 1$, this shows that (5) implies

(7) $$\lim_{\sigma\downarrow 1}\frac{S''}{S' + S''} = 0, \qquad \lim_{\sigma\downarrow 1}\frac{S'}{S' + S''} = 1$$

which is a specific sense in which the "overwhelming majority" of primes must satisfy (6) if (5) is true.

However, since $1 + 2it$ is not a pole of $\zeta(s)$, the real part of $\log\zeta(s)$ is bounded above near $s = 1 + 2it$, say by K'', hence

(8) $$\sum_p \frac{\cos(2t\log p)}{p^\sigma} < K'' + B.$$

When the terms in this sum are split into those terms for which p satisfies (6) and those terms for which p does not, the terms in the first group have cos $(2t\log p) = \cos(2\alpha)$, where $|\alpha| < 2\delta$; hence $\cos(2t\log p) > \cos 2\delta > 0$, so (8) implies

$$S'\cos 2\delta - S'' < K'' + B,$$
$$\frac{S''}{S' + S''} > \cos 2\delta\,\frac{S'}{S' + S''} - \frac{\text{const}}{S' + S''}$$

which contradicts (7). Thus (5) is impossible and the proof is complete.

4.3 PROOF THAT $\psi(x) \sim x$

Since von Mangoldt's formula

$$\psi(x) = x - \sum\frac{x^\rho}{\rho} + \sum\frac{x^{-2n}}{2n} - \frac{\zeta'(0)}{\zeta(0)} \qquad (x > 1)$$

is obtained by evaluating the definite integral

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left[-\frac{\zeta'(s)}{\zeta(s)}\right]\frac{x^s\,ds}{s} \qquad (x > 1, \quad a > 1)$$

in two different ways, it is natural to expect that the antiderivative of von Mangoldt's formula, namely,

$$\int_0^x \psi(t)\,dt = \frac{x^2}{2} - \sum\frac{x^{\rho+1}}{\rho(\rho+1)} - \sum\frac{x^{-2n+1}}{2n(2n-1)}$$
$$- \frac{\zeta'(0)}{\zeta(0)}x + \text{const} \qquad (x > 1)$$

could be obtained by evaluating the antiderivative of this definite integral, namely,

$$\text{(1)} \qquad \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left[-\frac{\zeta'(s)}{\zeta(s)}\right]\frac{x^{s+1}\,ds}{s(s+1)} \qquad (x>1,\quad a>1)$$

in two different ways.

The first way of evaluating (1) is to use the expansion $-\zeta'(s)/\zeta(s) = \sum \Lambda(n)n^{-s}$. If termwise integration of this expansion is valid, then (1) is

$$\text{(2)} \qquad \sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^{s+1}\,ds}{n^s s(s+1)}.$$

The partial fractions expansion

$$\text{(3)} \qquad \frac{1}{s(s+1)} = \frac{1}{s} - \frac{1}{s+1}$$

gives

$$\frac{x^{s+1}}{n^s s(s+1)} = \frac{x}{s}\left(\frac{x}{n}\right)^s - \frac{n}{s+1}\left(\frac{x}{n}\right)^{s+1}$$

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^{s+1}\,ds}{n^s s(s+1)} = \frac{x}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{n}\right)^s\frac{ds}{s} - \frac{n}{2\pi i}\int_{a+1-i\infty}^{a+1+i\infty}\left(\frac{x}{n}\right)^u\frac{du}{u}$$

$$= \begin{cases} x-n & \text{if } n \le x, \\ 0 & \text{if } n \ge x, \end{cases}$$

so that if termwise integration is valid, then (1) is equal to†

$$\sum_{n\le x}\Lambda(n)(x-n) = \int_0^x (x-t)\,d\psi(t) = \int_0^x \psi(t)\,dt.$$

The proof that termwise integration is valid, and hence that (1) is equal to $\int_0^x \psi(t)\,dt$, is easily accomplished by writing the integrand of (1) in the form

$$-\frac{\zeta'(s)}{\zeta(s)}\frac{x^{s+1}}{(s+1)s} = \left[x\sum_{n=1}^{\infty}\Lambda(n)\left(\frac{x}{n}\right)^s\frac{1}{s}\right] - \left[\sum_{n=1}^{\infty}\Lambda(n)\left(\frac{x}{n}\right)^{s+1}\frac{n}{s+1}\right]$$

to express (1) as the difference of two integrals of infinite sums over n, and by then observing that von Mangoldt's method shows immediately that each of these two integrals can be evaluated termwise.

The second way of evaluating (1) is to use the expansion

$$\text{(4)} \qquad -\frac{\zeta'(s)}{\zeta(s)} = \frac{s}{s-1} - \sum_{\rho}\frac{s}{\rho(s-\rho)} + \sum_{n=1}^{\infty}\frac{s}{2n(s+2n)} - \frac{\zeta'(0)}{\zeta(0)}$$

of Section 3.2, (7). When this is used in conjunction with (3) in expressing the integrand of (1), the term $1/s$ divides evenly, but the term $1/(s+1)$ does not.

†Note that the case $x = n$ presents no difficulties because then $x - n = 0$. This occurs because the antiderivative of a function with jump discontinuities has no jump discontinuities.

To simplify the resulting expression note that

$$-\frac{\zeta'(s)}{\zeta(s)}+\frac{\zeta'(-1)}{\zeta(-1)}=\frac{s}{s-1}-\frac{-1}{-1-1}-\sum_\rho\left[\frac{s}{\rho(s-\rho)}-\frac{-1}{\rho(-1-\rho)}\right]$$
$$+\sum_n\left[\frac{s}{2n(s+2n)}-\frac{-1}{2n(-1+2n)}\right]$$
$$-\frac{\zeta'(s)}{\zeta(s)}=\frac{s+1}{2(s-1)}-\sum_\rho\frac{s+1}{(\rho+1)(s-\rho)}$$
$$+\sum_n\frac{s+1}{(2n-1)(s+2n)}-\frac{\zeta'(-1)}{\zeta(-1)}$$

so that the integrand in (1) is equal to

$$-\frac{\zeta'(s)}{\zeta(s)}x^{s+1}\left(\frac{1}{s}-\frac{1}{s+1}\right)=\frac{x^{s+1}}{s-1}-\sum_\rho\frac{x^{s+1}}{\rho(s-\rho)}+\sum_n\frac{x^{s+1}}{2n(s+2n)}$$
$$-\frac{\zeta'(0)}{\zeta(0)}\frac{x^{s+1}}{s}-\frac{x^{s+1}}{2(s-1)}+\sum_\rho\frac{x^{s+1}}{(\rho+1)(s-\rho)}$$
$$-\sum_n\frac{x^{s+1}}{(2n-1)(s+2n)}+\frac{\zeta'(-1)}{\zeta(-1)}\frac{x^{s+1}}{s+1}.$$

Von Mangoldt's estimates (Section 3.5) with only slight modifications show that both of the sums over ρ and both of the sums over n can be integrated termwise. This gives for (1) the value

$$x^2-\sum_\rho\frac{x^{\rho+1}}{\rho}+\sum_n\frac{x^{1-2n}}{2n}-\frac{\zeta'(0)}{\zeta(0)}x-\frac{x^2}{2}+\sum_\rho\frac{x^{\rho+1}}{\rho+1}-\sum_n\frac{x^{1-2n}}{2n-1}+\frac{\zeta'(-1)}{\zeta(-1)}$$
$$=\frac{x^2}{2}-\sum_\rho\frac{x^{\rho+1}}{\rho(\rho+1)}-\sum_n\frac{x^{1-2n}}{2n(2n-1)}-\frac{\zeta'(0)}{\zeta(0)}x+\frac{\zeta'(-1)}{\zeta(-1)}$$

which by the above is also equal to $\int_0^x\psi(t)\,dt$. This completes the proof of the formula†

$$\int_0^x\psi(t)\,dt=\frac{x^2}{2}-\sum_\rho\frac{x^{\rho+1}}{\rho(\rho+1)}-\sum_n\frac{x^{-2n+1}}{2n(2n-1)}-\frac{\zeta'(0)}{\zeta(0)}x+\frac{\zeta'(-1)}{\zeta(-1)}$$

which holds for all $x>1$.

†This is a special case ($u=0$, $v=-1$) of the formula

$$y^u\int_0^y t^{-u}\,d\psi(t)-y^v\int_0^y t^{-v}\,d\psi(t)=-\left[y^u\frac{\zeta'(u)}{\zeta(u)}-y^v\frac{\zeta'(v)}{\zeta(v)}\right]+\frac{u-v}{(u-1)(v-1)}y$$
$$-(u-v)\sum_\rho\frac{y^\rho}{(u-\rho)(v-\rho)}$$
$$-(u-v)\sum_n\frac{y^{-2n}}{(u+2n)(v+2n)}\qquad(y>1)$$

which de la Vallée Poussin proved in 1896 by using the above method in conjunction with elementary arguments—not those of von Mangoldt—to justify the termwise integrations. However, he used the case $u=1$, $v=0$, which gives (2) of Section 4.1, in his proof of the prime number theorem. Note that the constant in formula (4) of Section 4.1 has been shown to be $\zeta'(-1)/\zeta(-1)$.

Using this formula for $\int_0^x \psi(t)\,dt$ it is quite easy to show that

$$\int_0^x \psi(t)\,dt \sim \frac{x^2}{2}, \tag{5}$$

where, as usual, the symbol $\sim$ means that the relative error in the approximation approaches zero as $x \to \infty$. One need only note that $\int_0^x \psi(t)\,dt - (x^2/2)$ divided by $x^2/2$ is, by the formula, equal to $2\Sigma x^{\rho-1}/\rho(\rho+1)$ plus terms which go to zero as $x \to \infty$. Since $\Sigma \rho^{-1}(\rho+1)^{-1}$ converges absolutely and since $|x^{\rho-1}| \leq 1$ (because $\operatorname{Re}\rho \leq 1$), it follows that the series $\Sigma x^{\rho-1}/\rho(\rho+1)$ converges uniformly in x and hence that the limit as $x \to \infty$ can be evaluated termwise; since each term goes to zero (because $\operatorname{Re}\rho < 1$), it follows that this limit is zero and hence that the relative error in (5) approaches zero as $x \to \infty$.

To deduce $\psi(x) \sim x$, let $\epsilon > 0$ be given and let X be such that

$$(1-\epsilon)\frac{x^2}{2} < \int_0^x \psi(t)\,dt < (1+\epsilon)\frac{x^2}{2}$$

for all $x \geq X$. Then for $y > x \geq X$ it follows that

$$\int_0^y \psi(t)\,dt - \int_0^x \psi(t)\,dt \tag{6}$$

is at least

$$(1-\epsilon)\frac{y^2}{2} - (1+\epsilon)\frac{x^2}{2} = (1-\epsilon)\frac{y^2-x^2}{2} - 2\epsilon\frac{x^2}{2}$$

and at most

$$(1+\epsilon)\frac{y^2}{2} - (1-\epsilon)\frac{x^2}{2} = (1+\epsilon)\frac{y^2-x^2}{2} + 2\epsilon\frac{x^2}{2}.$$

On the other hand, ψ is an increasing function, so (6) is at least

$$\int_x^y \psi(t)\,dt \geq (y-x)\psi(x)$$

and at most

$$\int_x^y \psi(t)\,dt \leq (y-x)\psi(y).$$

Combining these inequalities gives

$$(y-x)\psi(x) \leq (6) \leq (1+\epsilon)\frac{y^2-x^2}{2} + 2\epsilon\frac{x^2}{2},$$

$$(y-x)\psi(y) \geq (6) \geq (1-\epsilon)\frac{y^2-x^2}{2} - 2\epsilon\frac{x^2}{2},$$

for all $y > x \geq X$. With $y = \beta x$ these inequalities give

$$\frac{\psi(x)}{x} \leq (1+\epsilon)\frac{\beta+1}{2} + \frac{\epsilon}{\beta-1},$$

$$\frac{\psi(y)}{y} \geq (1-\epsilon)\frac{\beta+1}{2\beta} - \frac{\epsilon}{\beta(\beta-1)}.$$

Given any $\beta > 1, \epsilon > 0$, these inequalities are satisfied for all sufficiently large x, y. Now the quantity on the right side of the first inequality can be made less than any number greater than 1 by first choosing $\beta > 1$ near enough to 1 that the first term is very near 1 and by then choosing ϵ so small that the second term is still small; this shows that $\psi(x)/x$ is less than any number greater than 1 for x sufficiently large. Similarly the second inequality can be used to show that $\psi(y)/y$ is greater than any number less than 1 for y sufficiently large. This completes the proof that $\psi(x) \sim x$.

4.4 PROOF OF THE PRIME NUMBER THEOREM

Once the estimate $\psi(x) \sim x$ has been proved, the prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ is easily deduced. The technique used below is essentially the technique used by Chebyshev in 1850 [C3] to deduce his estimate of $\pi(x)$ from his estimate of $\psi(x)$.

Let $\theta(x)$ denote† the sum of the logarithms of all the primes p less than x, with the usual understanding that if x itself is a prime, then $\theta(x) = \frac{1}{2}[\theta(x+\epsilon) + \theta(x-\epsilon)]$. Then the relationship of θ and ψ is analogous to the relationship of π and J, and in analogy to the formula (1) of Section 1.17 relating π and J there is the formula

$$\psi(x) = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \theta(x^{1/4}) + \cdots$$

relating θ and ψ. The series on the left has the property that each term is larger than the following term and that all terms $\theta(x^{1/n})$ in which $x^{1/n} < 2$ are zero. Thus there are at most $\log x/\log 2$ nonzero terms and

$$\theta(x) < \psi(x) < \theta(x) + \theta(x^{1/2}) \log x/\log 2$$

which gives

$$\psi(x) - \theta(x^{1/2}) \log x/\log 2 < \theta(x) < \psi(x). \tag{1}$$

The inequality on the right together with $\psi(x) \sim x$ shows that $\theta(x)/x^{1+\epsilon} \longrightarrow 0$ as $x \longrightarrow \infty$. Hence $\theta(x^{1/2}) \log x/x = [\theta(x^{1/2})/(x^{1/2})^{1+\epsilon}][\log x/(x^{1/2})^{1-\epsilon}]$ goes to zero as $x \longrightarrow \infty$ and $\theta(x) \sim x$ follows from (1) and $\psi(x) \sim x$.

Now let $\epsilon > 0$ be given and let X be such that $(1-\epsilon)x \le \theta(x) \le (1+\epsilon)x$ whenever $x \ge X$. Then for $y > x \ge X$ it follows that

$$\begin{aligned}\pi(y) - \pi(x) &= \int_x^y \frac{d\theta(t)}{\log t} \\ &= \left[\frac{\theta(t)}{\log t}\right]_x^y + \int_x^y \frac{\theta(t)\,dt}{(\log t)^2 t}\end{aligned}$$

†This notation, introduced by Chebyshev, is now standard.

is at most

$$\begin{aligned}
&\frac{(1+\epsilon)y}{\log y} - \frac{(1-\epsilon)x}{\log x} + \int_x^y \frac{(1+\epsilon)t\,dt}{(\log t)^2 t} \\
&\quad = 2\epsilon\frac{x}{\log x} + (1+\epsilon)\left\{\left[\frac{t}{\log t}\right]_x^y + \int_x^y \frac{t\,dt}{(\log t)^2 t}\right\} \\
&\quad = 2\epsilon\frac{x}{\log x} + (1+\epsilon)\left\{\int_x^y \frac{dt}{\log t}\right\} \\
&\quad = 2\epsilon\frac{x}{\log x} + (1+\epsilon)[\mathrm{Li}(y) - \mathrm{Li}(x)]
\end{aligned}$$

and is at least

$$-2\epsilon\frac{x}{\log x} + (1-\epsilon)[\mathrm{Li}(y) - \mathrm{Li}(x)].$$

Thus for fixed x it follows that $\pi(y)/\mathrm{Li}(y)$ is at most

$$1 + \epsilon + \frac{\pi(x) - 2\epsilon x(\log x)^{-1} - (1+\epsilon)\,\mathrm{Li}(x)}{\mathrm{Li}(y)} \leq 1 + 2\epsilon$$

for all sufficiently large y and similarly that it is at least $1 - 2\epsilon$ for all sufficiently large y. Since ϵ was arbitrary, this proves the prime number theorem $\mathrm{Li}(y) \sim \pi(y)$.

Chapter 5

De la Vallée Poussin's Theorem

5.1 INTRODUCTION

After it was proved that the relative error in the approximation $\pi(x) \sim \mathrm{Li}(x)$ approaches zero as x approaches infinity, the next step was an estimate of the *rate* at which it approaches zero. De la Vallée Poussin [V2] proved in 1899 that there is a $c > 0$ such that the relative error approaches zero at least as fast as $\exp[-(c \log x)^{1/2}]$ does; that is,

$$\left|\frac{\pi(x) - \mathrm{Li}(x)}{\mathrm{Li}(x)}\right| < e^{-\sqrt{c \log x}}$$

for all sufficiently large x. The next two sections are devoted to the proof of this fact. Section 5.4 contains an application of this theorem to the question of comparing $\mathrm{Li}(x)$ to other possible approximations; it is shown, in essence, that Dirichlet, Gauss, Chebyshev, and Riemann were correct in preferring $\mathrm{Li}(x)$ to another approximation suggested earlier by Legendre. The next section, 5.5, shows that de la Vallée Poussin's theorem can be improved considerably if the Riemann hypothesis is true and that in fact the Riemann hypothesis is *equivalent* to the statement that the relative error in $\pi(x) \sim \mathrm{Li}(x)$ goes to zero faster than $x^{-(1/2)+\epsilon}$ as $x \to \infty$ (for all $\epsilon > 0$). Finally, the last section is devoted to the very simple proof of von Mangoldt's theorem that Euler's formula $\sum \mu(n)/n = 0$ [$\mu(n)$ is the Möbius function] is true in the strong sense that the series $\sum \mu(n)/n$ is *convergent* to the sum zero; this proof makes very effective use of de la Vallée Poussin's theorem that the relative error in the prime number theorem approaches zero at least as fast as $\exp[-(c \log x)^{1/2}]$.

5.2 AN IMPROVEMENT OF Re $\rho < 1$

De la Vallée Poussin's estimate of the error in the prime number theorem is based on a strengthened version of the theorem that the roots $\rho = \beta + i\gamma$ satisfy $\beta < 1$, namely, the theorem that there exist constants† $c > 0$, $K > 1$, such that

$$\beta < 1 - \frac{c}{\log \gamma} \tag{1}$$

for all roots $\rho = \beta + i\gamma$ in the range $\gamma > K$. Since $\log \gamma > \log K > 0$, the inequality (1) is stronger than $\beta < 1$, but the amount by which it is stronger decreases as γ increases and (1) does not preclude the possibility that there are roots ρ arbitrarily near to the line $\operatorname{Re} s = 1$; although (1) has been improved upon somewhat since 1899, no one has yet been able to prove that $\operatorname{Re} \rho$ has any upper bound less than 1.

De la Vallée Poussin's proof of (1) makes use of a technique by which Mertens [M6] simplified the proof that $\operatorname{Re} \rho < 1$. This technique is based on the elementary inequality

$$4 \geq 2(1 - \cos \theta),$$
$$4(1 + \cos \theta) \geq 2(1 - \cos^2 \theta) = 1 - \cos 2\theta,$$
$$3 + 4 \cos \theta + \cos 2\theta \geq 0,$$

which holds for all θ. Combining this with the formula‡ $-\zeta'(s)/\zeta(s) = \int_0^\infty x^{-s}\, d\psi(x)$ for $\operatorname{Re} s > 1$ gives

$$\operatorname{Re}\left\{-3\frac{\zeta'(\sigma)}{\zeta(\sigma)} - 4\frac{\zeta'(\sigma + it)}{\zeta(\sigma + it)} - \frac{\zeta'(\sigma + 2it)}{\zeta(\sigma + 2it)}\right\}$$
$$= \int_0^\infty x^{-\sigma}[3 + 4\cos(t \log x) + \cos(2t \log x)]\, d\psi(x) \geq 0.$$

Hence

$$\operatorname{Re}\left\{3\frac{\zeta'(\sigma)}{\zeta(\sigma)} + 4\frac{\zeta'(\sigma + it)}{\zeta(\sigma + it)} + \frac{\zeta'(\sigma + 2it)}{\zeta(\sigma + 2it)}\right\} \leq 0 \tag{2}$$

for all $\sigma > 1$ and for all real t.

This inequality can be used to prove $\beta < 1$ as follows. Assume there is a real value of t such that $\zeta(1 + it) = 0$ and let $f(s)$ be the function $[\zeta(s)]^3 \cdot [\zeta(s + it)]^4 \cdot [\zeta(s + 2it)]$. Then the first factor of $f(s)$ has a pole of order 3 at

†Specifically, de la Vallée Poussin proved that $\beta < 1 - c(\log \gamma - \log n)^{-1}$ for $\gamma \geq 705$, where $c = 0.034$ and $n = 47.8$. This requires, of course, much more careful estimates than those given here.

‡Mertens uses $\log \zeta(s) = \int_0^\infty x^{-s}\, dJ(x)$ instead.

$s = 1$, the second has a zero of order at least 4, and the third has no pole; hence $f(s)$ has a zero of order at least 1 at $s = 1$, say $f(s) = (s-1)^n g(s)$, where $n \geq 1$ and where $g(s)$ has no zero and no pole at $s = 1$. Then the logarithmic derivative of g approaches a limit as $s \to 1$, so the logarithmic derivative of $f(s)$ differs from $n(s-1)^{-1}$ by a bounded amount as $s \to 1$. Thus the logarithmic derivative of $f(s)$ has large positive real part when $s = \sigma$ is real and greater than 1. But (2) shows that this is impossible and hence that no t satisfies $\zeta(1 + it) = 0$.

De la Vallée Poussin used the inequality (2) to prove (1) by using the formula

$$\frac{\zeta'(s)}{\zeta(s)} = -\frac{1}{s-1} + \sum_\rho \frac{1}{s-\rho} - \frac{1}{2}\frac{\Pi'(s/2)}{\Pi(s/2)} + \frac{1}{2}\log \pi$$

[see (6) of Section 3.2] to estimate the terms of (2). Specifically, this formula and the formula $\Pi'(x)/\Pi(x) \sim \log x$ of Section 6.3 show that for $1 \leq \sigma \leq 2$ and for t large, the real part of $\zeta'(s)/\zeta(s)$ is approximately

$$\sum_\rho \operatorname{Re} \frac{1}{s-\rho} - \frac{1}{2}\log\left|\frac{t}{2}\right| + \frac{1}{2}\log \pi$$

[because $(s-1)^{-1}$ is small] and in particular that one can choose $K > 1$ such that

$$\operatorname{Re}\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} \geq \sum_\rho \operatorname{Re}\frac{1}{\sigma+it-\rho} - \log t + \frac{1}{2}\log \pi$$

holds throughout the region $1 \leq \sigma \leq 2$, $t \geq K$. The terms of the sum over ρ are positive (because $\sigma \geq 1 > \operatorname{Re}\rho$), so the same inequality holds *a fortiori* if some or all terms of the sum over ρ are omitted. Now $(\sigma - 1)\zeta(\sigma)$ is nonsingular at $\sigma = 1$, so the logarithmic derivative of $\zeta(\sigma)$ differs from $-(\sigma-1)^{-1}$ by a bounded amount and (2) gives

$$0 \geq -\frac{3}{\sigma-1} + \frac{4}{\sigma+it-\rho} - 4\log t - \log(2t) + \text{const}$$

when all but one term of the sums over ρ are omitted. Thus for all roots ρ and $1 < \sigma \leq 2$, $t \geq K$

$$\frac{3}{4}\frac{1}{\sigma-1} + C\log t \geq \frac{1}{\sigma+it-\rho},$$

where $C > 0$ is independent of σ, t, and ρ. If $\rho = \beta + i\gamma$, where $\gamma \geq K$, then one can set $t = \gamma$ to find

$$\frac{3}{4}\frac{1}{\sigma-1} + C\log\gamma \geq \frac{1}{\sigma-\beta}$$

for all σ in the range $1 < \sigma \leq 2$. Thus

$$\left[\frac{3}{4(\sigma - 1)} + C \log \gamma\right]^{-1} \leq \sigma - \beta = (\sigma - 1) - (\beta - 1),$$

$$\beta - 1 \leq (\sigma - 1) - \frac{4(\sigma - 1)}{3 + 4(\sigma - 1)C \log \gamma}$$

$$= \frac{-(\sigma - 1) + 4(\sigma - 1)^2 C \log \gamma}{3 + 4(\sigma - 1)C \log \gamma},$$

$$\beta \leq 1 - \frac{y - 4y^2}{3 + 4y} \cdot \frac{1}{C \log \gamma},$$

where $y = (\sigma - 1)C \log \gamma$. Thus y can have any value between 0 and $C \log \gamma \geq C \log K$. Since $y - 4y^2 > 0$ for small values of y, one need only fix a small positive value of y such that $y < C \log K$, $y - 4y^2 > 0$ in order to draw the desired conclusion (1) for all roots $\rho = \beta + i\gamma$ which satisfy $\gamma \geq K$.

5.3 DE LA VALLÉE POUSSIN'S ESTIMATE OF THE ERROR

Since the main step in the proof of the prime number theorem is to use the estimate $\beta < 1$ to prove that $\sum x^{\rho-1}/\rho(\rho + 1)$ approaches zero as $x \to \infty$, it is natural to try to use the improved estimate $\beta < 1 - c(\log \gamma)^{-1}$ (for $\gamma \geq K$) of the last section to prove an improved estimate of $\sum x^{\rho-1}/\rho(\rho + 1)$. De la Vallée Poussin accomplished this by the following very simple argument. Note first that

$$\left|\sum_\rho \frac{x^{\rho-1}}{\rho(\rho + 1)}\right| \leq \sum_\rho \frac{x^{\beta-1}}{\gamma^2} = \sum_{|\gamma|<K} \frac{x^{\beta-1}}{\gamma^2} + 2 \sum_{\gamma \geq K} \frac{x^{\beta-1}}{\gamma^{1-\delta}} \cdot \frac{1}{\gamma^{1+\delta}}.$$

The first term on the right is the sum of a finite number of terms each of which is a constant times a negative power of x (namely, $x^{\beta-1}$); hence there are positive constants C, ϵ such that this term is less than $Cx^{-\epsilon}$ for all $x > 1$. If δ is any positive constant, then $2 \sum \gamma^{-1-\delta}$ converges by the theorem of Section 2.5; so the second term on the right is less than a constant times the maximum value of $x^{\beta-1}/\gamma^{1-\delta}$ if this expression does have a maximum value for $\gamma \geq K$. But

$$\frac{x^{\beta-1}}{\gamma^{1-\delta}} \leq \frac{x^{-c/(\log \gamma)}}{\gamma^{1-\delta}}$$

and the logarithmic derivative of the quantity on the right with respect to γ (considered for the moment as a continuous variable) is

$$\frac{c \log x}{(\log \gamma)^2} \cdot \frac{1}{\gamma} - \frac{1 - \delta}{\gamma}$$

which is positive, negative, or zero according to whether $c \log x$ is greater than, less than, or equal to $(1-\delta)(\log \gamma)^2$. Thus if x is large enough that $c \log x > (1-\delta)(\log K)^2$ and if $1-\delta > 0$, then there is a maximum for $\gamma \geq K$ at the point where $c \log x = (1-\delta)(\log \gamma)^2$. At this point

$$\frac{c \log x}{\log \gamma} = (1-\delta) \log \gamma = [c(1-\delta) \log x]^{1/2};$$

hence

$$\frac{x^{\beta-1}}{\gamma^{1-\delta}} \leq \frac{x^{-c/(\log \gamma)}}{\gamma^{1-\delta}} \leq \frac{\exp\{-[c(1-\delta)\log x]^{1/2}\}}{\exp\{[c(1-\delta)\log x]^{1/2}\}} = \exp\{-2[c(1-\delta)\log x]^{1/2}\}.$$

Set $\delta = \frac{3}{4}$ and let C_1 denote $2\sum \gamma^{-1-\delta}$. Then the above estimates combine to give

$$\left|\sum \frac{x^{\rho-1}}{\rho(\rho+1)}\right| < Cx^{-\epsilon} + C_1 \exp[-(c \log x)^{1/2}]$$

for all sufficiently large x. Finally, since $x^{-\epsilon}$ goes to zero much faster than $\exp[-(c \log x)^{1/2}]$, since the constants C, C_1 can be absorbed by decreasing c slightly, and since $2\sum x^{\rho-1}/\rho(\rho+1)$ is the relative error in the approximation $\int_0^x \psi(t)\,dt \sim x^2/2$ except for terms which are like x^{-1} as $x \to \infty$ [see (4) of Section 4.1] this proves that *there is a constant* $c > 0$ *such that the relative error in the approximation* $\int_0^x \psi(t)\,dt \sim x^2/2$ *is less than* $\exp[-(c \log x)^{1/2}]$ *for all sufficiently large* x.

Now, by essentially the same arguments which were used to deduce the prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ from $\int_0^x \psi(t)\,dt \sim x^2/2$, one can deduce an estimate of the relative error in the prime number theorem from the above estimate of the relative error in $\int_0^x \psi(t)\,dt \sim x^2/2$.

The first step is to derive an estimate of the relative error in $\psi(x) \sim x$. Let $\varepsilon(x) = \exp[-(c \log x)^{1/2}]$, where c is as above so that

$$[1-\varepsilon(x)]\frac{x^2}{2} \leq \int_0^x \psi(t)\,dt \leq [1+\varepsilon(x)]\frac{x^2}{2}$$

for all sufficiently large x. Then, as before, $\int_0^y \psi(t)\,dt - \int_0^x \psi(t)\,dt$ for $y > x$ is on the one hand at most

$$[1+\varepsilon(y)]\frac{y^2}{2} - [1-\varepsilon(x)]\frac{x^2}{2} \leq [1+\varepsilon(x)]\left(\frac{y^2-x^2}{2}\right) + 2\varepsilon(x)\frac{x^2}{2}$$

and at least

$$[1-\varepsilon(x)]\left(\frac{y^2-x^2}{2}\right) - 2\varepsilon(x)\frac{x^2}{2}$$

and on the other hand is at most

$$\int_x^y \psi(y)\,dt = (y-x)\psi(y)$$

and at least

$$\int_x^y \psi(x)\,dt = (y-x)\psi(x).$$

Thus

$$(y-x)\psi(x) \le [1+\varepsilon(x)]\left(\frac{y^2-x^2}{2}\right) + 2\varepsilon(x)\frac{x^2}{2},$$

$$(y-x)\psi(y) \ge [1-\varepsilon(x)]\left(\frac{y^2-x^2}{2}\right) - 2\varepsilon(x)\frac{x^2}{2},$$

and, therefore,

$$\psi(x) - x \le \frac{y-x}{2} + \varepsilon(x)\frac{y+x}{2} + \varepsilon(x)\frac{x^2}{y-x},$$

$$\psi(y) - y \ge -\frac{y-x}{2} - \varepsilon(x)\frac{y+x}{2} - \varepsilon(x)\frac{x^2}{y-x}.$$

In the first inequality set $y = \{1 + [\varepsilon(x)]^{1/2}\}x$ to obtain

$$\begin{aligned}\frac{\psi(x)-x}{x} &\le \frac{[\varepsilon(x)]^{1/2}}{2} + \varepsilon(x)\left[1+\frac{1}{2}\varepsilon(x)\right] + [\varepsilon(x)]^{1/2}\\ &\le \text{const}\,[\varepsilon(x)]^{1/2},\end{aligned}$$

and in the second inequality set $x = \{1 - [\varepsilon(y)]^{1/2}\}y$ to obtain

$$\begin{aligned}\frac{\psi(y)-y}{y} &\ge -\frac{[\varepsilon(y)]^{1/2}}{2} - \varepsilon(x)\left\{1-\frac{1}{2}[\varepsilon(y)]^{1/2}\right\} - \varepsilon(x)\frac{1}{[\varepsilon(y)]^{1/2}}\\ &\ge -\frac{[\varepsilon(x)]^{1/2}}{2} - \varepsilon(x) - \frac{\varepsilon(x)}{[\varepsilon(y)]^{1/2}} \ge \frac{-2\varepsilon(x)}{[\varepsilon(y)]^{1/2}}\\ &= -2\exp\{-[c\log y + c\log[1-(\varepsilon(y))^{1/2}]]^{1/2} + \tfrac{1}{2}(c\log y)^{1/2}\}\\ &\ge -2\exp\{-(c\log y)^{1/2}[(1-\text{small})^{1/2} - \tfrac{1}{2}]\}.\end{aligned}$$

Thus *there is a constant* $c_0 > 0$ *such that the relative error in the approximation* $\psi(x) \sim x$ *is less than* $\exp[-(c_0\log x)^{1/2}]$ *for all sufficiently large* x.

The next step is to consider the approximation $\theta(x) \sim x$. But since, as before,

$$\psi(x) - \theta(x^{1/2})(\log x)/\log 2 < \theta(x) < \psi(x)$$

and since $\theta(x^{1/2})\log x \sim x^{1/2}\log x$ grows much more slowly than $x\exp[-(c_0\log x)^{1/2}]$, it follows immediately that the relative error in this approximation is also less than $\exp[-(c_0\log x)^{1/2}]$ for all sufficiently large x.

The final step is to consider the approximation $\pi(x) \sim \text{Li}(x)$. Here the formula

$$\pi(y) - \pi(x) = \int_x^y \frac{d\theta(t)}{\log t} = \frac{\theta(y)}{\log y} - \frac{\theta(x)}{\log x} + \int_x^y \frac{\theta(t)\,dt}{(\log t)^2 t}$$

shows that if $y > x$ are in the range where the relative error in $\theta(x) \sim x$ is less than $\varepsilon(x) = \exp[-(c_0\log x)^{1/2}]$, then $\pi(y)$ is at most

$$\begin{aligned}&\pi(x) + \frac{y[1+\varepsilon(y)]}{\log y} - \frac{x[1-\varepsilon(x)]}{\log x} + \int_x^y \frac{t[1+\varepsilon(t)]}{(\log t)^2 t}\,dt\\ &\quad = \pi(x) + \frac{y\varepsilon(y)}{\log y} + \frac{x\varepsilon(x)}{\log x} + \int_x^y \frac{dt}{\log t} + \int_x^y \frac{\varepsilon(t)\,dt}{(\log t)^2}.\end{aligned}$$

Let x be fixed, let y be larger than x^2, and let the final integral be divided into an integral from x to $y^{1/2}$ and an integral from $y^{1/2}$ to y. Then

$$\begin{aligned}\pi(y) &\le \text{const} + \frac{y\varepsilon(y)}{\log y} + \mathrm{Li}(y)\\ &\quad + (y^{1/2} - x)\frac{\varepsilon(x)}{(\log x)^2} + (y - y^{1/2})\frac{4\varepsilon(y^{1/2})}{(\log y)^2}\\ &\le \text{const} + \mathrm{Li}(y)\\ &\quad + \frac{y}{\log y}\left[\varepsilon(y) + \frac{\text{const}\log y}{y^{1/2}} + \frac{4\varepsilon(y^{1/2})}{\log y}\right]\end{aligned}$$

and similarly

$$\pi(y) \ge \text{const} + \mathrm{Li}(y) - \frac{y}{\log y}\left[\varepsilon(y) + \frac{\text{const}\log y}{y^{1/2}} + \frac{4\varepsilon(y^{1/2})}{\log y}\right].$$

Since the quantity in square brackets is less than $\varepsilon(y^{1/2})$ for all sufficiently large y, it will suffice to prove that $y/\log y$ divided by $\mathrm{Li}(y)$ is bounded as $y \to \infty$ in order to prove that the relative error in $\pi(y) \sim \mathrm{Li}(y)$ is less than $\varepsilon(y^{1/2})$ for y sufficiently large. It will be shown in the next section that $\mathrm{Li}(y) \sim y/\log y$, and this will complete the proof that *there is a constant* $c_1 > 0$ *such that the relative error in the approximation* $\pi(y) \sim \mathrm{Li}(y)$ *is less than* $\exp[-(c_1 \log y)^{1/2}]$ *for all sufficiently large values of* y. This is de la Vallée Poussin's estimate† of the error in the prime number theorem.

5.4 OTHER FORMULAS FOR $\pi(x)$

The approximate formula for $\pi(x)$ which appears in Legendre's *Theorie des Nombres* [L4] is

$$\pi(x) \sim \frac{x}{\log x - A}, \tag{1}$$

where A is a constant whose value Legendre gives as 1.08366, apparently on empirical grounds. Legendre does not specify the sense in which the approximation (1) is to be understood, but if it is interpreted in the usual sense of "the relative error approaches zero as $x \to \infty$," then the value of A is irrelevant because

$$\frac{x}{\log x - A} \sim \frac{x}{\log x - B}$$

†De la Vallée Poussin wrote the estimate in the form $(c_1 \log y)^{1/2} \exp[-(c_1 \log y)^{1/2}]$ and gave the explicit value 0.032 of c_1. He did not, however, give any explicit estimate of how large y must be in order for this estimate of the relative error in $\pi(y) \sim \mathrm{Li}(y)$ to be valid.

for any two numbers A, B (because the ratio is $\log x - A$ over $\log x - B$, which approaches 1) and hence if (1) is true for one value of A, it must be true for all values of A (because "$\sim$" is transitive). Therefore if Legendre's value $A = 1.08366$ has any significance, it must lie in some other interpretation of the approximation (1).

The prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ shows that (1) is true if and only if $\mathrm{Li}(x) \sim x/(\log x - A)$ for some—and hence all—values of A. But by integration by parts

$$\begin{aligned}\mathrm{Li}(x) &= \mathrm{Li}(2) + \int_2^x \frac{dt}{\log t}\\ &= \mathrm{Li}(2) + \frac{x}{\log x} - \frac{2}{\log 2} + \int_2^x \frac{dt}{(\log t)^2},\end{aligned}$$

$$\mathrm{Li}(x) - \frac{x}{\log x} = \mathrm{const} + \int_2^x \frac{dt}{(\log t)^2},$$

and it suffices to show that the integral on the right divided by $x/\log x$ approaches zero as $x \to \infty$ in order to conclude that (1) is true with $A = 0$. [Intuitively this is the obvious statement that the average value of $(\log t)^{-2}$ for $2 \le t \le x$ is much less than $(\log x)^{-1}$.] This is easily accomplished by dividing the interval of integration at $x^{1/2}$ to find

$$\begin{aligned}\frac{\log x}{x}\int_2^x \frac{dt}{(\log t)^2} &= \frac{\log x}{x}\int_2^{x^{1/2}} \frac{dt}{(\log t)^2} + \frac{\log x}{x}\int_{x^{1/2}}^x \frac{dt}{(\log t)^2}\\ &\le \frac{\log x}{x}\cdot\frac{x^{1/2}-2}{(\log 2)^2} + \frac{\log x}{x}\cdot\frac{x - x^{1/2}}{(\log x^{1/2})^2}\\ &\le \frac{\log x}{(\log 2)^2 x^{1/2}} + \frac{4}{\log x} \to 0.\end{aligned}$$

Thus the prime number theorem implies (1) but it implies no particular value of A.

Chebyshev [C2] was able to show that if any value of A is any better than any other, then this value must be $A = 1$. The special property of the value $A = 1$ which is needed is the fact that the approximation

$$\mathrm{Li}(x) \sim \frac{x}{\log x - A} \tag{2}$$

is best when $A = 1$. To see this, note that

$$\frac{x}{\log x - A} = \frac{x}{\log x}\left[1 + \left(\frac{A}{\log x}\right) + \left(\frac{A}{\log x}\right)^2 + \cdots\right],$$

$$\frac{x}{\log x - A} \sim \frac{x}{\log x} + \frac{Ax}{(\log x)^2}, \tag{3}$$

where $\sim$ means that the error is much smaller than the last term $Ax(\log x)^{-2}$

in the sense that their ratio approaches zero as $x \longrightarrow \infty$, while

$$\begin{aligned}\operatorname{Li}(x) &= \frac{x}{\log x} + \text{const} + \int_2^x \frac{dt}{(\log t)^2} \\ &= \frac{x}{\log x} + \text{const} + \frac{x}{(\log x)^2} - \frac{2}{(\log 2)^2} + 2\int_2^x \frac{dt}{(\log t)^3}.\end{aligned}$$

Hence

$$\operatorname{Li}(x) \sim \frac{x}{\log x} + \frac{x}{(\log x)^2}, \tag{4}$$

where the error is much smaller than the last term because

$$\frac{(\log x)^2}{x}\int_2^x \frac{dt}{(\log t)^3} \le \frac{(\log x)^2}{x}\frac{x^{1/2}-2}{(\log 2)^3} + \frac{(\log x)^2}{x}\frac{x - x^{1/2}}{(\log x^{1/2})^3} \longrightarrow 0.$$

The combination of (3) and (4) shows that the error in (2) divided by $x(\log x)^{-2}$ is $A - 1$ plus a quantity which approaches zero as $x \longrightarrow \infty$. This gives a precise sense in which $A = 1$ is the best value of A in (2). Chebyshev was able to prove enough about the approximation $\pi(x) \sim \operatorname{Li}(x)$ to prove that if A has a best value in (1), then that value must also be $A = 1$.

However, even the prime number theorem is not enough to prove that $A = 1$ is best in (1). To prove this, one would have to show that the approximation $\pi(x) \sim \operatorname{Li}(x)$, like the approximations (3) and (4), has the property that its error grows much less rapidly than $x(\log x)^{-2}$, and the prime number theorem says only that it grows much less rapidly than $\operatorname{Li}(x)$, which in turn grows like $x(\log x)^{-1}$ by (2). Thus a stronger estimate of the error in $\pi(x) \sim \operatorname{Li}(x)$ is needed.

Now since $\exp[(c\log x)^{1/2}]$ grows more rapidly than any power of $(c\log x)^{1/2}$, it follows from de la Vallée Poussin's estimate of the error in $\pi(x) \sim \operatorname{Li}(x)$ that

$$\left|\frac{\pi(x) - \operatorname{Li}(x)}{x(\log x)^{-2}}\right| \le \frac{\operatorname{Li}(x)}{x}\cdot\frac{(\log x)^2}{\exp[(c\log x)^{1/2}]} \longrightarrow 0$$

and hence that the error in (1) divided by $x(\log x)^{-2}$ is $A - 1$ plus a quantity which approaches zero as $x \longrightarrow \infty$. Thus de la Vallée Poussin's estimate proves that the value $A = 1$ in (1) is better than any other value.

More generally, successive integration by parts shows that (4) can be generalized to

$$\begin{aligned}\operatorname{Li}(x) \sim \frac{x}{\log x} &+ \frac{x}{(\log x)^2} + 2\frac{x}{(\log x)^3} \\ &+ 6\frac{x}{(\log x)^4} + \cdots + (n-1)!\,\frac{x}{(\log x)^n},\end{aligned}$$

where the error (for any fixed n) grows much less rapidly than the last term $x(\log x)^{-n}$ as $x \longrightarrow \infty$. De la Vallée Poussin's estimate shows that the error in

$\pi(x) \sim \mathrm{Li}(x)$ also grows less rapidly than $x(\log x)^{-n}$ and hence proves that the approximation

$$\pi(x) \sim \frac{x}{\log x} + \frac{x}{(\log x)^2} + \cdots + (n-1)!\frac{x}{(\log x)^n} \tag{5}$$

is valid in the sense that (for any fixed n) the error in the approximation divided by the last term on the right approaches zero as $x \longrightarrow \infty$. The case $n = 1$ is essentially the prime number theorem and the case $n = 2$ is essentially the theorem that $A = 1$ is best in (1).

Thus de la Vallée Poussin's estimate of the error proves that the approximation $\pi(x) \sim \mathrm{Li}(x)$ is better than the approximation (5) for any value of n. This is the principal consequence which de la Vallée Poussin himself drew from his estimate of the error.

Formula (5) is an example of an *asymptotic expansion*, which is an expansion such as $\pi(x) = \sum (n-1)!\, x(\log x)^{-n}$ in which the error resulting from taking a finite number of terms is of a lower order of magnitude than the last term used. Another more familiar example of an asymptotic expansion is the Taylor series of an infinitely differentiable function $f(x) = \sum f^{(n)}(a)(x-a)^n/n!$. The fact that this is an asymptotic expansion—that is, the fact that the error resulting from using just n terms decreases more rapidly than $(x-a)^n$ as $x \longrightarrow a$—is Taylor's theorem. This in no way implies, of course, that for fixed $x \neq a$ the error approaches zero as $n \longrightarrow \infty$. For any fixed x formula (5) for $\pi(x)$ becomes worthless as $n \longrightarrow \infty$ because $(n-1)!$ grows much faster than $(\log x)^n$. Another example of an asymptotic expansion is Stirling's series (3) of Section 6.3. Although pure mathematicians shun asymptotic expansions which do not converge as $n \longrightarrow \infty$, mathematicians who engage in computation are well aware that asymptotic expansions (for example, Stirling's series) are often more practical than convergent expansions [for example, the product formula (3) of Section 1.3 for Π].

Recall that Riemann's approximate formula for $\pi(x)$ was

$$\pi(x) \sim \mathrm{Li}(x) + \sum_{n=2}^{N} \frac{\mu(n)}{n} \mathrm{Li}(x^{1/n}), \tag{6}$$

where $N > \log x/\log 2$, and that on empirical grounds this formula appeared to be much better than $\pi(x) \sim \mathrm{Li}(x)$. The second term in this formula is $-\frac{1}{2}\mathrm{Li}(x^{1/2}) \sim -\frac{1}{2}x^{1/2}(\log x^{1/2})^{-1} = -x^{1/2}(\log x)^{-1}$, whereas de la Vallée Poussin's estimate shows that the error in $\pi(x) \sim \mathrm{Li}(x)$ is less than $\mathrm{Li}(x) \cdot\exp[-(c\log x)^{1/2}] \sim x(\log x)^{-1}\exp[-(c\log x)]^{1/2}$ so that $-\frac{1}{2}\mathrm{Li}(x^{1/2})$ divided by the error estimate is about $-x^{-1/2}\exp[+(c\log x)^{1/2}] = -\exp[-\frac{1}{2}\log x + (c\log x)^{1/2}] \longrightarrow 0$. Thus de la Vallée Poussin's estimate is not strong enough to prove that even the second term of (6) has any validity, much less the remaining terms. It was, in fact, proved by Littlewood [L13] that Riemann's formula (6) is *not* better as $x \longrightarrow \infty$ than the simpler formula $\pi(x) \sim \mathrm{Li}(x)$. In other words, Littlewood proved that in the long run as $x \longrightarrow \infty$ the "periodic" terms $\mathrm{Li}(x^\rho)$ in the formula for $\pi(x)$ (see Section 1.17) are as significant as the monotone increasing term $-\frac{1}{2}\mathrm{Li}(x^{1/2})$ and *a fortiori* as significant as the following terms $\mathrm{Li}(x^{1/3})$, $\mathrm{Li}(x^{1/4}), \ldots$ of (6).

5.5 ERROR ESTIMATES AND THE RIEMANN HYPOTHESIS

In view of the strong relationship between de la Vallée Poussin's estimate of the error in the prime number theorem and his estimate $\beta < 1 - c(\log \gamma)^{-1}$ of $\beta = \mathrm{Re}\,\rho$, it is not surprising that the Riemann hypothesis $\mathrm{Re}\,\rho = \frac{1}{2}$ should imply much stronger estimates of the error. The best such estimate that has been found up to now is the estimate proved by von Koch [K1] in 1901, namely, that if the Riemann hypothesis is true then the relative errors in $\psi(x) \sim x$ and $\pi(x) \sim \mathrm{Li}(x)$ are both less than a constant times $(\log x)^2 x^{-1/2}$ for all sufficiently large x. This estimate implies that the relative errors are eventually less than $x^{-(1/2)+\epsilon}$ for all $\epsilon > 0$, whereas de la Vallée Poussin's estimate $\exp[-(c \log x)^{1/2}]$ fails to show that they are less than $x^{-\epsilon}$ for any $\epsilon > 0$.

If the Riemann hypothesis is true, then the magnitude of the relative error in the approximation $\int_0^x \psi(t)\,dt \sim \frac{1}{2}x^2$ is less than

$$\begin{aligned} &2\left|\sum_\rho \frac{x^{\rho-1}}{\rho(\rho+1)}\right| + \text{const}\, x^{-2} + x^{-1}\zeta'(0)/\zeta(0) \\ &\quad \leq 2\cdot x^{-1/2}\cdot\sum_\rho \left|\frac{1}{\rho(\rho+1)}\right| + \text{const}\, x^{-1} \\ &\quad \leq \text{const}\, x^{-1/2} \end{aligned}$$

for all sufficiently large x. However, the previous method of deducing from this an estimate of the relative error in $\psi(x) \sim x$ involves taking a square root and hence yields only the estimate $x^{-1/4}$ and not von Koch's estimate $x^{-1/2}(\log x)^2$. Therefore some other method of estimating the error in $\psi(x) \sim x$ is necessary. The following method is due to Holmgren [H10]. In the estimate (see Section 4.3)

$$\begin{aligned} \psi(x) &\leq \int_x^{x+1} \psi(t)\,dt = \int_0^{x+1} \psi(t)\,dt - \int_0^x \psi(t)\,dt \\ &= \left|\frac{(x+1)^2}{2} - \frac{x^2}{2} + \sum_\rho \frac{(x+1)^{\rho+1} - x^{\rho+1}}{\rho(\rho+1)} + \text{bounded}\right| \\ &\leq x + \sum_\rho \left|\frac{(x+1)^{\rho+1} - x^{\rho+1}}{\rho(\rho+1)}\right| + \text{const} \end{aligned}$$

let the terms corresponding to roots $\rho = \frac{1}{2} + i\gamma$ (assuming the Riemann hypothesis) for which $|\gamma| \leq x$ be estimated using

$$\left|\frac{(x+1)^{\rho+1} - x^{\rho+1}}{\rho(\rho+1)}\right| = \frac{1}{|\rho|}\left|\int_x^{x+1} t^\rho\,dt\right| \leq \frac{|x+1|^{1/2}}{|\rho|} \leq \frac{(2x)^{1/2}}{|\gamma|}$$

and let those corresponding to roots for which $|\gamma| > x$ be estimated using

$$\left|\frac{(x+1)^{\rho+1} - x^{\rho+1}}{\rho(\rho+1)}\right| \leq \frac{2|x+1|^{\operatorname{Re}\rho+1}}{|\rho(\rho+1)|} \leq \frac{2(2x)^{3/2}}{\gamma^2}.$$

Let H be as in the theorem of Section 3.4 so that the number of roots $\rho = \frac{1}{2} + i\gamma$ in the interval $t \leq \gamma \leq t+1$ is less than $2 \log t$ for $t \geq H$. Then the above estimates give (when $x > H$)

$$\begin{aligned}
\psi(x) &\leq x + x^{1/2} \sum_{|\gamma|<H} \frac{\sqrt{2}}{|\gamma|} + 2x^{1/2} \sum_{H<\gamma<x} \frac{\sqrt{2}}{\gamma} + 2x^{3/2} \sum_{x<\gamma} \frac{2^{5/2}}{\gamma^2} \\
&\leq x + \text{const}\, x^{1/2} + \text{const}\, x^{1/2} \int_H^x \frac{\log t\,dt}{t} + \text{const}\, x^{3/2} \int_x^\infty \frac{\log t\,dt}{t^2} \\
&\leq x + \text{const}\, x^{1/2} + \text{const}\, x^{1/2} \frac{(\log t)^2}{2}\bigg|_H^x \\
&\qquad + \text{const}\, x^{3/2} \left\{-\frac{\log t}{t}\bigg|_x^\infty + \int_x^\infty \frac{dt}{t^2}\right\} \\
&\leq x + \text{const}\, x^{1/2} + \text{const}\, x^{1/2}(\log x)^2 + \text{const}\, x^{3/2}\left\{\frac{\log x}{x} + \frac{1}{x}\right\} \\
&\leq x + \text{const}\, x^{1/2}(\log x)^2.
\end{aligned}$$

The same technique applied to the estimate

$$\psi(x) \geq \int_{x-1}^x \psi(t)\,dt = \left|\frac{x^2}{2} - \frac{(x-1)^2}{2} + \sum_\rho \frac{x^{\rho+1} - (x-1)^{\rho+1}}{\rho(\rho+1)} + \text{bounded}\right|$$

gives

$$\begin{aligned}
\psi(x) &\geq x - \sum_{|\gamma|<H} \frac{x^{1/2}}{|\gamma|} - 2 \sum_{H<\gamma<x} \frac{x^{1/2}}{\gamma} - 2 \sum_{x>\gamma} \frac{2x^{3/2}}{\gamma^2} \\
&\geq x - \text{const}\, x^{1/2} - \text{const}\, x^{1/2} \int_H^x \frac{\log t\,dt}{t} - \text{const}\, x^{3/2} \int_x^\infty \frac{\log t\,dt}{t^2} \\
&\geq x - \text{const}\, x^{1/2}(\log x)^2
\end{aligned}$$

which completes the proof that the relative error in $\psi(x) \sim x$ is less than a constant times $(\log x)^2 x^{-1/2}$ (assuming the Riemann hypothesis). Since

$$\psi(x) - \theta(x^{1/2})\log x/\log 2 \leq \theta(x) \leq \psi(x)$$

and since $\theta(x^{1/2}) \log x \sim x^{1/2} \log x$ is much smaller than $x^{1/2}(\log x)^2$, the same is true of $\theta(x) \sim x$. Now

$$\begin{aligned}
\pi(x) - \operatorname{Li}(x) &= \int_c^x \frac{d[\theta(t) - t]}{\log t} + \text{const} \\
&= \frac{\theta(x) - x}{\log x} + \int_c^x \frac{\theta(t) - t}{t(\log t)^2}\,dt + \text{const}
\end{aligned}$$

for any constant $c > 1$; hence if $|\theta(t) - t| < Kt^{1/2}(\log t)^2$ for $t \geq c$, then $x > c$ implies

$$|\pi(x) - \mathrm{Li}(x)| \leq \frac{Kx^{1/2}(\log x)^2}{\log x} + \int_c^x \frac{K\,dt}{t^{1/2}} + \text{const}$$
$$\leq K\frac{x}{\log x}\frac{(\log x)^2}{x^{1/2}} + Kx^{1/2} + \text{const},$$

which proves, since $\mathrm{Li}(x) \sim (x/\log x)$ as $x \to \infty$, that the relative error in $\pi(x) - \mathrm{Li}(x)$ is eventually less than a constant times $(\log x)^2 x^{-1/2}$ if the Riemann hypothesis is true, as was to be shown.

On the other hand, if the Riemann hypothesis is false, then there is a root ρ with $\mathrm{Re}\,\rho > \frac{1}{2}$ and hence (see below) a "periodic" term in Riemann's formula for $\pi(x)$ which grows more rapidly in magnitude than $x^{1/2}$, so it is reasonable to assume that the error in the prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ would not in that case grow less rapidly than $x^{(1/2)+\epsilon}$.

The rate of growth of $\mathrm{Li}(x^\rho)$ for ρ in the first quadrant $\mathrm{Re}\,\rho > 0$, $\mathrm{Im}\,\rho > 0$ is easily estimated using the formula $\mathrm{Li}(x^\rho) = \int_{C^+} (\log t)^{-1} t^{\rho-1}\,dt + i\pi$ of Section 1.15 and integration by parts as in Section 5.4 to find

$$\begin{aligned}\mathrm{Li}(x^\rho) &= \int_{C^+} \frac{t^{\rho-1}\,dt}{\log t} + i\pi \\ &= \int_{C^+} \frac{d}{dt}\left\{\frac{t^\rho}{\rho \log t}\right\} dt - \int_{C^+} \frac{t^\rho}{\rho}\frac{(-1)}{(\log t)^2}\frac{dt}{t} + i\pi \\ &= \frac{x^\rho}{\rho \log x} + \int_{C^+} \frac{t^{\rho-1}\,dt}{\rho(\log t)^2} + i\pi \\ &= \frac{x^\rho}{\rho \log x} + \int_2^x \frac{t^{\rho-1}\,dt}{\rho(\log t)^2} + \text{const.}\end{aligned}$$

The first term, which has modulus $x^{\mathrm{Re}\,\rho}|\rho|^{-1}(\log x)^{-1}$, dominates as $x \to \infty$ because

$$\begin{aligned}&\int_2^{x^{1/2}} \frac{t^{\rho-1}\,dt}{(\log t)^2} + \int_{x^{1/2}}^x \frac{t^{\rho-1}\,dt}{(\log t)^2} \\ &\leq \int_2^{x^{1/2}} \frac{t^{\mathrm{Re}\,\rho-1}\,dt}{(\log 2)^2} + \int_{x^{1/2}}^x \frac{t^{\mathrm{Re}\,\rho-1}\,dt}{(\log x^{1/2})^2} \\ &\leq \frac{x^{(1/2)\,\mathrm{Re}\,\rho}}{\mathrm{Re}\,\rho(\log 2)^2} + \frac{4x^{\mathrm{Re}\,\rho}}{\mathrm{Re}\,\rho(\log x)^2}\end{aligned}$$

has modulus much less than a constant times $x^{\mathrm{Re}\,\rho}(\log x)^{-1}$.

Theorem The Riemann hypothesis is equivalent to the statement that for every $\epsilon > 0$ the relative error in the prime number theorem $\pi(x) \sim \mathrm{Li}(x)$ is less than $x^{-(1/2)+\epsilon}$ for all sufficiently large x. [If they are true, then the relative error in the prime number theorem is in fact less than a constant times $x^{-1/2}(\log x)^2$.]

Proof It remains only to show that if the relative error in $\pi(x) \sim \mathrm{Li}(x)$ is less than $x^{-(1/2)+\epsilon}$, then the Riemann hypothesis must be true. Assume therefore that for every $\epsilon > 0$ the relative error in $\pi(x) \sim \mathrm{Li}(x)$ is eventually