Math 229: Introduction to Analytic Number Theory

Proof of the Prime Number Theorem;
the Riemann Hypothesis

We finally have all the ingredients that we need to assemble a proof of the Prime Number Theorem with an explicit error bound. We shall give an upper bound on $|(\psi(x)/x) - 1|$ that decreases faster than any power of $1/\log x$ as $x \to \infty$, though slower than any positive power of $1/x$. Specifically, we show:

Theorem. *There exists an effective constant* $C > 0$ *such that*

$$\psi(x) = x + O(x \exp(-C\sqrt{\log x})) \tag{1}$$

for all $x \geq 1$.

Proof: There is no difficulty with small x, so we may and shall assume that $x \geq e$, so $\log x \geq 1$. We use our integral approximation

$$\psi(x) = \frac{1}{2\pi i}\int_{1+\frac{1}{\log x}-iT}^{1+\frac{1}{\log x}+iT} -\frac{\zeta'}{\zeta}(s)\, x^s \frac{ds}{s} + O\left(\frac{x\log^2 x}{T}\right) \qquad (T \in [1, x]) \tag{2}$$

to $\psi(x)$. Assume that $T \geq e$, and that T does not coincide with the imaginary part of any ρ. Shifting the line of integration leftwards, say to real part -1, yields

$$\psi(x) - \left(x - \sum_{|\mathrm{Im}(\rho)|<T} \frac{x^\rho}{\rho}\right) = I_1 + I_2 - \frac{\zeta'}{\zeta}(0) + O\left(\frac{x\log^2 x}{T}\right), \tag{3}$$

in which I_1, I_2 are the integrals of $-(\zeta'(s)/\zeta(s))x^s\, ds/s$ over the vertical line $\sigma = -1, |t| < T$ and the horizontal lines $\sigma \in [-1, 1 + (1/\log x)]$, $t = \pm T$ respectively. We next show that I_1 is small, and that I_2 can be made small by adding $O(1)$ to T. The vertical integral I_1 is clearly

$$\ll \frac{\log T}{x} \sup_{|t|<T} \left|\frac{\zeta'}{\zeta}(-1+it)\right| \ll \frac{\log^2 T}{x}.$$

The horizontal integrals in I_2 are

$$\ll \frac{1}{T}\int_{-1}^{1+\frac{1}{\log x}} x^\sigma\, d\sigma \cdot \sup_{\sigma\in[-1,2]} \left|\frac{\zeta'}{\zeta}(\sigma + iT)\right|.$$

The σ integral is $\ll x/\log x$. We have seen already that for $s = \sigma + iT$ and $-1 \leq \sigma \leq 2$ we have

$$\zeta'(s)/\zeta(s) = \sum_{|T-\mathrm{Im}\,\rho|<1} \frac{1}{s-\rho} + O(\log T),$$

in which the sum has $O(\log T)$ terms. Since the number of $\operatorname{Im}\rho$ in the interval $[T-1, T+1]$ is $\ll \log T$, some point in the middle half of that interval is at distance $\gg 1/\log T$ from all of them; choosing that as our new value of T, we see that each term is $\ll \log T$, and thus that the sum is $\ll \log^2 T$. In conclusion, then,

$$I_2 \ll x \log^2 T / T \log x.$$

Better estimates can be obtained (we could save a factor of $\log T$ by averaging over $[T - \frac{1}{2}, T + \frac{1}{2}]$), but are not necessary because $x \log^2 T/T \log x$ is already less than the error $(x \log^2 x)/T$ in (2).

Thus the RHS of (3) may be absorbed into the $O((x \log^2 x)/T)$ error. In the LHS, we use our zero-free region, that is, the lower bound

$$1 - \sigma > c/\log |t|, \tag{4}$$

to find that

$$|x^\rho| = x^{\operatorname{Re}(\rho)} \ll x^{1-\frac{c}{\log T}} = x \exp\Bigl(-c\,\frac{\log x}{\log T}\Bigr).$$

Since[1]

$$\sum_{|\operatorname{Im}(\rho)|<T} \frac{1}{|\rho|} < \sum_{|\operatorname{Im}(\rho)|<T} \frac{1}{|\operatorname{Im}\rho|}$$

$$= 2\int_1^T \frac{dN(t)}{t} = \frac{2N(T)}{T} + 2\int_1^T \frac{N(t)\,dt}{t^2} \ll \log T + \int_1^T \frac{\log t\,dt}{t} \ll \log^2 T,$$

we thus have

$$\sum_{|\rho|<T} \frac{x^\rho}{\rho} \ll x \log^2 T \exp\Bigl(-c\,\frac{\log x}{\log T}\Bigr).$$

Therefore

$$\left|\frac{\psi(x)}{x} - 1\right| \ll \left(\frac{1}{T} + \exp\bigl(-c\frac{\log x}{\log T}\bigr)\right) \log^2 x.$$

We choose T so that the logarithms $-\log T$, $-\log x/\log T$ of the two terms $1/T$, $\exp(-c \log x/\log T)$ are equal. That is, we take $T = \exp\sqrt{\log x}$. Then both terms are $O(\exp(-C_1 \log^{1/2} x))$ for some $C_1 > 0$. We then absorb the factor $\log^2 x$ into this estimate by changing C_1 to any positive $C < C_1$, and at last complete the proof of (1). □□

The equivalent result for $\pi(x)$ follows by partial summation:

Corollary. *There exists an effective constant $C > 0$ such that*

$$\pi(x) = \operatorname{li}(x) + O(x \exp(-C\sqrt{\log x})).$$

for all $x \geq 1$.

[1] We can use $\int_1^T$ because we have shown that there are no complex zeros ρ with $|\operatorname{Im}(\rho)| \leq 1$. If there were such zeros, we could absorb their terms x^ρ/ρ into the error estimate. We shall do this in the proof of the corresponding estimates on $\psi(x, \chi)$.

[Recall that $\mathrm{li}(x)$ is the principal value of $\int_0^x dy/\log y$, whence

$$\mathrm{li}(x) = \int_2^x dy/\log y + O(1) = x/\log x + O(x/\log^2 x).]$$

Proof: We have seen already that

$$\pi(x) = \frac{\psi(x)}{\log x} + \int_2^x \psi(y) \frac{dy}{y \log^2 y} + O(x^{1/2}). \tag{5}$$

On the other hand, integration by parts yields

$$\mathrm{li}(x) = \frac{x}{\log x} - \int_2^x y \, d(1/\log y) + O(1) = \frac{x}{\log x} + \int_2^x \frac{dy}{\log^2 y} + O(1).$$

The Corollary now follows from (1). □

The Riemann Hypothesis and some consequences

The error estimate in (1), while sufficient to prove the Prime Number Theorem, is not nearly as strong as one might wish. The growth rate of $|\psi(x) - x|$ and $|\pi(x) - \mathrm{li}(x)|$ hinges on the *Riemann Hypothesis (RH)*, which we introduce next.

The RH and its generalizations are arguably the most important open problems in mathematics. We shall see and explore some of these generalizations later. The original RH is Riemann's inspired guess that *all the nontrivial zeros of* $\zeta(s)$ *have real part equal to* $1/2$, i.e., lie on the critical line $\sigma = 1/2$ at the center of the critical strip. At the time there was scant evidence for the conjecture: the symmetry of the zeros with respect to the critical line, and also numerical computations of the first few zeros (not reported in Riemann's memoir but found among his papers after his death). The conjecture is now supported by a wealth of numerical evidence, as well as compelling analogies with "geometrical" zeta functions for which the conjecture has been proved — notably the zeta functions of varieties over finite fields, for which the RH was proved by Hasse [1936] (elliptic curves), Weil [1940, 1941, 1948] (arbitrary curves and abelian varieties), and Deligne (the general case). These analogies also suggest that proving the "arithmetical" RH and its generalizations will involve fundamental new insights in number theory, quite beyond the immediate applications to the distribution of primes and related arithmetical functions. For now we content ourselves with the most direct connections between the RH and the error estimate in the Prime Number Theorem.

If the RH holds then we may take $T = x$ in (3) to find $\psi(x) = x + O(x^{1/2} \log^2 x)$. More generally:

Proposition. *Suppose there exists* θ *with* $1/2 \le \theta < 1$ *such that* $\mathrm{Re}\,\rho \le \theta$ *for all zeros* ρ *of* ζ. *Then* $\psi(x) = x + O(x^\theta \log^2 x)$ *and* $\pi(x) = \mathrm{li}(x) + O(x^\theta \log x)$ *for large* x.

Proof: Take $T = x + O(1)$ in (3). By our bounds on I_1, I_2, the right-hand side is $O(\log^2 x)$. By hypothesis, each of the terms x^ρ/ρ has absolute value at most

$x^\theta/|\rho| < x^\theta/|\operatorname{Im}\rho|$. Hence

$$\left|\sum_{|\operatorname{Im}(\rho)|<T} \frac{x^\rho}{\rho}\right| < 2x^\theta \sum_{0<\operatorname{Im}(\rho)<T} \frac{1}{\operatorname{Im}\rho}.$$

We have seen already that the last sum is $O(\log^2 T)$; here $T = x + O(1)$, so we conclude that

$$\psi(x) - x = O(x^\theta \log^2 T) + O(\log^2 x) = O(x^\theta \log^2 x),$$

as claimed. The corresponding estimate on $\pi(x) - \operatorname{li}(x)$ then follows from (5), since $\theta \geq 1/2$. $\square$

A converse implication also holds:

Proposition. *Suppose there exists θ with $1/2 \leq \theta < 1$ such that $\psi(x) = x + O_\epsilon(x^{\theta+\epsilon})$ for all $\epsilon > 0$. Then $\zeta(s)$ has no zeros of real part $> \theta$. The same conclusion holds if $\pi(x) = \operatorname{li}(x) + O_\epsilon(x^{\theta+\epsilon})$.*

(So, for instance, RH is *equivalent* to the assertion that $\pi(x) = \operatorname{li} x + O(x^{1/2}\log x)$. The hypotheses on $\pi(x)$ and $\psi(x)$ are equivalent, again by (5).)

Proof: Write $-\zeta'(s)/\zeta(s) = \sum_n \Lambda(n) n^{-s}$ as a Stieltjes integral and integrate by parts to find

$$-\frac{\zeta'}{\zeta}(s) = s\int_1^\infty \psi(x) x^{-s-1}\,dx = \frac{s}{s-1} + s\int_1^\infty (\psi(x) - x)\, x^{-s-1}\,dx \quad (\sigma > 1).$$

If $\psi(x) - x \ll_\epsilon x^{\theta+\epsilon}$ then the resulting integral for $s/(s-1) + \zeta'(s)/\zeta(s)$ extends to an analytic function on $\sigma > \theta$, whence that half-plane contains no zeros of $\zeta(s)$. $\square$

Note the amusing consequence that an estimate $\psi(x) = x + O_\epsilon(x^{\theta+\epsilon})$ would automatically improve to $\psi(x) = x + O(x^\theta \log^2 x)$, and similarly for $\pi(x)$.

Remarks

One may naturally ask whether $\psi(x)$ tends to be larger or smaller than its approximation x, and likewise whether $\pi(x)$ tends to be larger or smaller than $\operatorname{li}(x)$. For the former question, our formula (3) suggests that $\psi(x)$ can as easily be larger or smaller than x: the terms x^ρ/ρ in the formula (3) for $x - \psi(x)$ oscillate as x increases, and if we choose $\log x$ uniformly from $[1, U]$ then the phase of each term tends to uniform distribution on the circle as $U \to \infty$. It may be surprising then that $\pi(x)$ behaves quite differently: it is very hard to find any x such that $\pi(x) > \operatorname{li}(x)$. This is because $\pi(x)$ is expressed as a Stieltjes integral involving not $\psi(x)$ but $\sum_{p<x} \log p$, and

$$\psi(x) - \sum_{p<x} \log p \sim \psi(x^{1/2}) \sim x^{1/2}.$$

Under the Riemann Hypothesis, $x^{1/2}$ is exactly of the same asymptotic order as each of the terms x^ρ/ρ in (3), and much larger than each single term because

$|\rho|^{-1} < 1/14$. For large x, we can imagine the terms x^ρ/ρ ($\operatorname{Im}\rho > 0$) as random complex numbers z_ρ drawn independently from the circle $|z| = x^{1/2}/\rho$.[2] Then $\sum_\rho x^\rho/\rho = 2\operatorname{Re}\sum_{(\operatorname{Im}\rho)>0} z_\rho$. Since $\sum_\rho 1/|\rho|^2 < \infty$, this heuristic suggests that for "random large x" the scaled error $x^{-1/2}(\psi(x) - x)$ is drawn from a distribution symmetric about the origin, and thus that $x^{-1/2}(\sum_{p<x}\log p - x)$ is drawn from a distribution symmetric about -1. Since $\sum_\rho 1/|\rho| = +\infty$, it is possible for $-2\operatorname{Re}\sum_{(\operatorname{Im}\rho)>0} z_\rho$ to exceed x, and thus for $\sum_{p<x}\log p$ to exceed x and likewise for $\pi(x)$ to exceed $\operatorname{li}(x)$. But this does not happen routinely, and indeed it was once thought that $\operatorname{li}(x)$ might always exceed $\pi(x)$.

Littlewood first showed that the difference changes sign infinitely often. In particular, there exist x such that $\pi(x) > \operatorname{li}(x)$. But none has been found yet. The earliest explicit upper bound on the smallest such x was the (in)famously astronomical "Skewes' number" [Skewes 1933]. That bound has since fallen, but still stands at several hundred digits, too large to reach directly even with the best algorithms known for computing $\pi(x)$ — algorithms that themselves depend on the analytical formulas such as (2); see [LO 1982].

Exercises

1. Use the partial-fraction decomposition of ζ'/ζ to get the following exact formula:

$$\psi(x) = x - \sum_\rho \frac{x^\rho}{\rho} - \frac{\zeta'}{\zeta}(0) - \frac{1}{2}\log(1 - x^{-2}).$$

Here $\sum_\rho$ is taken to mean $\lim_{T\to\infty}\sum_{|\rho|<T}$; and if $x = p^k$, so that $\psi(x)$ is discontinuous at x, then we interpret $\psi(x)$ as $(\psi(x-\epsilon)+\psi(x+\epsilon))/2$. Note that $-\frac{1}{2}\log(1 - x^{-2})$ is the sum of $-x^r/r$ over the trivial zeros $r = -2, -4, -6, \ldots$ See [Davenport 1967, Chapter 17].

2. Show that the improvement $1-\sigma > c_\epsilon/\log^{(2/3)+\epsilon}|t|$ on (4) yields an estimate $O\big(x\exp(-C_\epsilon \log^{(3/5)-\epsilon} x)\big)$ on the error in the Prime Number Theorem.

3. Prove that

$$\lim_{x\to\infty}\left(\log x - \sum_{n=1}^{x}\frac{\Lambda(n)}{n}\right) = \boldsymbol{\gamma},$$

and give an error bound both unconditionally and under the Riemann Hypothesis. Deduce that $\log x - \sum_{p<x}(\log p)/p$ and $\log\log x - \sum_{p<x} 1/p$ approach finite limits as $x\to\infty$. (The last of these refines Euler's theorem that $\sum_p 1/p$ diverges.)

4. [A theorem of Mertens; see for instance [Titchmarsh 1951], pages 38–39.] Prove that

$$\lim_{x\to\infty}\left(\log\log x - \sum_{n=1}^{x}\frac{\Lambda(n)}{n\log n}\right) = -\boldsymbol{\gamma},$$

[2] We shall later make this heuristic more precise, and show that it is equivalent to the conjecture that the numbers $\gamma > 0$ such that $\zeta(\frac{1}{2} + i\gamma) = 0$ are $\mathbf{Q}$-linearly independent. This conjecture is almost certainly true and extremely difficult to prove. See [RS 1994] and [BFHR 2001] for more information.

(Warning: this requires a contour integral involving $\log\big((s-1)\zeta(s)\big)$, which cannot be pushed past the zero-free region.) Deduce that

$$\lim_{x\to\infty}\left(\log x \prod_{p<x}\frac{p-1}{p}\right) = e^{-\gamma}.$$

As in the previous exercise, and give error bounds both unconditionally and under the Riemann Hypothesis.

In the last exercise, we illustrate the power of the method we used to prove the Prime Number Theorem by applying it to different kind of asymptotic averaging problem. We'll address a special case posed as an open problem in [Rawsthorne 1984]:

> Set $a_0 = 1$ and for $n \geq 1$, $a_n = a_{n'} + a_{n''} + a_{n'''}$ where $n' = \lfloor n/2 \rfloor$, $n'' = \lfloor n/3 \rfloor$, $n''' = \lfloor n/6 \rfloor$. Find $\lim_{n\to\infty} a_n/n$.

(It is not immediately obvious even that the limit exists.) The general problem can be solved in much the same way, though one usually gets somewhat less precise estimates on the vertical distribution of the zeros than are available for our special case. Only two solutions were received (see *Math. Magazine* **58**, 51–52): the solution outlined here, and a solution by Erdős, Hildebrand, Odlyzko, Pudaite, and Reznick, which they subsequentely generalized in [EHOPR 1987]. Their method corresponds to one of the "elementary proofs" of the Prime Number Theorem. The sequence $\{a(n)\}$ of Rawsthorne's problem is now #A007731 in Sloane's *On-Line Encyclopedia of Integer Sequences.*

5. i) Let $f(s) = 1 - 2^{-s} - 3^{-s} - 6^{-s}$. Note that f has a simple zero at $s = 1$. Prove that all its other zeros lie in the strip $|\sigma| < 1$, and that f has $\frac{\log 6}{2\pi}T + O(1)$ zeros ρ with $0 < \operatorname{Im}\rho < T$; more precisely, that each rectangle

$$\{\sigma + it : |\sigma| \leq 1, \ \left|\frac{\log 6}{2\pi}t - n\right| < 1/2\}$$

($n \in \mathbf{Z}$) contains a unique zero of f (so in particular the zeros are all simple). NB Unlike the case of $\zeta(s)$, here there is no functional equation, nor a "Riemann Hypothesis"; indeed, it can be shown that some complex zeros have real parts arbitrarily close to 1, as well as zeros whose real parts are arbitrarily close to -1.
ii) Let a_n be the coefficients of the Dirichlet series $\sum_{n=1}^{\infty} a_n/n^s = 1/f(s)$. Show that $a_n \geq 0$, with equality unless $n = 2^a 3^b$ for some integers a, b. Find a constant C such that $\sum_{n<x} a_n \sim Cx$ as $x\to\infty$. Can you give an explicit error bound?
iii) Solve Rawsthorne's problem above. How far can you generalize it? (Warning: for more general recursions of this kind you may have to contend with multiple poles, or simple poles that nearly coincide and have large residues.)

References

[BFHR 2001] Bays, C., Ford, K., Hudson, R.H., Rubinstein, M.: Zeros of Dirichlet L-functions near the Real Axis and Chebyshev's Bias. *J. Number Th.* **87** (2001) #1, 54–76.

[Hasse 1936] Hasse, H.: Zur Theorie der abstrakten elliptischen Funktionkörper I, II, III, *J. reine angew. Math.* **175** (1936), 55–62, 69–88, and 193–208.

[EHOPR 1987] Erdős, P., Hildebrand, A., Odlyzko, A., Pudaite, P., Reznick, B.: The asymptotic behavior of a family of sequences, *Pacific J. Math.* **126** (1987), 227–241.

[LO 1982] Lagarias, J.C., Odlyzko, A.M.: New algorithms for computing $\pi(x)$. Pages 176–193 in *Number Theory: New York 1982* (D.V. and G.V. Chudnovsky, H. Cohn, M.B. Nathanson, eds.; Berlin: Springer 1984, LNM **1052**).

[Rawsthorne 1984] Rawsthorne, D.A.: Problem 1185, *Math. Magazine* **57** (1984), p.42.

[RS 1994] Rubinstein, M. Sarnak, P.: Chebyshev's Bias. *Exp. Math.* **3** (1994) #3, 173–197.

[Skewes 1933] Skewes, S.: On the difference $\pi(x) - \mathrm{li}(x)$ (I). *J. London Math. Soc.* (1st ser.) **8** (1933), 277–283.

[Weil 1940] Weil, A.: Sur les fonctions algébriques à corps de constantes fini, *C.R. Acad. Sci. Paris* **210** (1940), 592–594.

[Weil 1941] Weil, A.: On the Riemann hypothesis in functionfields, *Proc. Nat. Acad. Sci. USA* **27** (1941), 345–347.

[Weil 1948] Weil, A.: *Variétés abéliennes et courbes algébriques*, Paris: Hermann 1948.