

# THE PRIME NUMBER THEOREM AND THE RIEMANN HYPOTHESIS

PETE L. CLARK

## 1. SOME HISTORY OF THE PRIME NUMBER THEOREM

Recall we have defined, for positive real  $x$ ,

$$\pi(x) = \# \{\text{primes } p \leq x\}.$$

The following is probably the single most important result in number theory.

**Theorem 1.** (*Prime Number Theorem*) *We have  $\pi(x) \sim \frac{x}{\log x}$ ; i.e.,*

$$\lim_{x \rightarrow \infty} \frac{\pi(x) \log x}{x} = 1.$$

1.1. **Gauss at 15.** The prime number theorem (affectionately called “PNT”) was apparently first conjectured in the late 18th century, by Legendre and Gauss (independently). In particular, Gauss conjectured an equivalent – but more appealing – form of the PNT in 1792, at the age of 15 (!!!).

Namely, he looked at the frequency of primes in intervals of lengths 1000:

$$\Delta(x) = \frac{\pi(x) - \pi(x - 1000)}{1000}.$$

Computing by hand, Gauss observed that  $\Delta(x)$  seemed to tend to 0, however very slowly. To see how slowly he computed the reciprocal, and found

$$\frac{1}{\Delta(x)} \approx \log x,$$

meaning that

$$\Delta(x) \approx \frac{1}{\log x}.$$

Evidently 15 year old Gauss knew both differential and integral calculus, because he realized that  $\Delta(x)$  was a slope of the secant line to the graph of  $y = \pi(x)$ . When  $x$  is large, this suggests that the slope of the tangent line to  $\pi(x)$  is close to  $\frac{1}{\log x}$ , and hence he guessed that the function

$$\text{Li}(x) := \int_2^x \frac{dt}{\log t}$$

was a good approximation to  $\pi(x)$ .

**Proposition 2.** *We have*

$$\text{Li}(x) \sim \frac{x}{\log x}.$$

Proof: A calculus exercise (L'Hôpital's rule!).

Thus PNT is equivalent to  $\pi(x) \sim \text{Li}(x)$ . The function  $\text{Li}(x)$  – called the **logarithmic integral** – is not elementary, but has a simple enough power series expansion (see for yourself). Nowadays we have lots of data, and one can see that the error  $|\pi(x) - \text{Li}(x)|$  is in general much smaller than  $|\pi(x) - \frac{x}{\log x}|$ , so the dilogarithm gives a “better” asymptotic expansion. (How good? Read on.)

**1.2. A partial result.** As far as I know, there was no real progress for more than fifty years, until the Russian mathematician Pafnuty Chebyshev proved the following two impressive results.

**Theorem 3.** (*Chebyshev, 1848, 1850*)

a) *There exist explicitly computable positive constants  $C_1, C_2$  such that for all  $x$ ,*

$$\frac{C_1 x}{\log x} < \pi(x) < \frac{C_2 x}{\log x}.$$

b) *If  $\lim_{x \rightarrow \infty} \frac{\pi(x)}{x/(\log x)}$  exists, it necessarily equals 1.*

Remarks:

(i) For instance, one version of the proof gives  $C_1 = 0.92$  and  $C_2 = 1.7$ . (But I don't know what values Chebyshev himself derived.)

(ii) The first part shows that  $\pi(x)$  is of “order of magnitude”  $\frac{x}{\log x}$ , and the second shows that if it is “regular enough” to have an asymptotic value at all, then it must be asymptotic to  $\frac{x}{\log x}$ . Thus the additional trouble in proving PNT is establishing this *regularity* in the distribution of the primes, a quite subtle matter. (We have seen that other arithmetical functions, like  $\varphi$  and  $d$  are far less regular than this – their upper and lower orders differ by more than a multiplicative constant, so the fact that this regularity should exist for  $\pi(x)$  is by no means assured.)

(iii) Chebyshev's proof is quite elementary: it uses less machinery than some of the other topics in this course. However we will not give the time to prove it here: blame it on your instructor's failure to “understand” the proof.

**1.3. A complex approach.** The next step was taken by Riemann in 1859. We have seen the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

and its relation to the primes (e.g. obtaining an instantaneous proof that  $\pi(x) \rightarrow \infty$  by the above product factorization). However, Riemann considered  $\zeta(s)$  as a function of a complex variable:  $s = \sigma + it$  (indeed he used these rather strange names for the real and imaginary parts in his 1859 paper, and we have kept them ever since), so

$$n^s = n^{\sigma+it} = n^{\sigma} n^{it}.$$

Here  $n^{\sigma}$  is a real number and  $n^{it} = e^{i(\log n)t}$  is a point on the unit circle, so in modulus we have  $|n^s| = n^{\sigma}$ . From this we get that  $\zeta(s)$  is *absolutely convergent* for  $\sigma = \Re(s) > 1$ . Using standard results from analysis, one sees that it indeed defines

an *analytic* function in the half-plane  $\sigma > 1$ . Riemann got the zeta function named after him by observing the following:

**Fact:**  $\zeta(s)$  extends (“meromorphically”) to the entire complex plane and is analytic everywhere except for a simple pole at  $s = 1$ .

We recall in passing, for those with some familiarity with complex variable theory, that the extension of an analytic function defined in one (connected) domain in the complex plane to a larger (connected) domain is *unique* if it exists at all: this is the *principle of analytic continuation*. So the zeta function is well-defined. The continuation can be shown to exist via an integral representation valid for  $\sigma > 0$  and a **functional equation** relating the values of  $\zeta(s)$  to that of  $\zeta(1-s)$ . (Note that the line  $\sigma = \frac{1}{2}$  is fixed under the  $s \mapsto 1-s$ .) Riemann conjectured, but could not prove, certain simple (to state!) analytic properties of  $\zeta(s)$ , which he saw had profound implications on the distribution of the primes.

**1.4. A nonvanishing theorem.** Nevertheless it is a testament to the difficulty of the subject that even after this epochal paper the proof of PNT did not come for almost another 40 years: in 1896, Jacques Hadamard and Charles de la Vallée-Poussin proved PNT, independently, but by rather similar methods. The key point in both of their proofs (which Riemann could not establish) was that  $\zeta(s) \neq 0$  for any  $s = 1 + it$ , i.e., along the line with  $\sigma = 1$ .

Their proof does come with an explicit error estimate, albeit an ugly one: they showed in fact

**Theorem 4.** *There exist positive constants  $C$  and  $a$  such that*

$$|\pi(x) - \text{Li}(x)| \leq Cxe^{-a\sqrt{\log x}}.$$

It is not completely obvious that this is indeed an error bound, i.e., that

$$\lim_{x \rightarrow \infty} \frac{e^{-a\sqrt{\log x}}}{\text{Li}(x)} = 0;$$

this is left as another calculus exercise.

**1.5. An elementary proof is prized.** Much was made of the fact that the proof of PNT, a theorem of number theory, used nontrivial results from complex analysis (which by the end of the 19th century had been developed to a large degree of sophistication). Many people speculated on the existence of an “elementary” proof, a yearning that to my knowledge was never formalized precisely. Roughly speaking it means a proof that uses no extraneous concepts from higher analysis (such as complex analytic functions) but only the notion of a limit and the definition of a prime. It thus caused quite a stir when Atle Selberg and Paul Erdős (not independently, but not quite collaboratively either – the story is a controversial one!) gave what all agreed to be an elementary proof of PNT in 1949. In 1950 Selberg (but not Erdős) received the Fields Medal.

It seems fair to say that in recent times the excitement about the elementary proof has dimmed: most experts agree that it is less illuminating and less natural than the proof via Riemann’s zeta function. Moreover the elementary proof remains quite intricate: ironically, more so than the analytic proof for those with

some familiarity with functions of a complex variable. For those who do not, the time taken to learn some complex analysis and then the proof of Hadamard - de la Vallée-Poussin will be somewhat longer but ultimately more profitable than the time spent digesting the elementary proof!<sup>1</sup>

**1.6. Equivalents of PNT.** It turns out that there are many statements which are “equivalent” to PNT: i.e., for which it is much easier to show that they imply and are implied by PNT than to prove them. A useful one is

**Theorem 5.** *Let  $p_n$  be the  $n$ th prime. Then*

$$p_n \sim n \log n.$$

Note that this result implies (by the integral test) that  $\sum_{p \leq n} \frac{1}{p} \sim \log \log n$ ; strangely this consequence is much easier to prove than PNT itself.

Far more intriguing is that that PNT is equivalent to an asymptotic formula for the average value of the Möbius function:

**Theorem 6.**

$$\lim_{N \rightarrow \infty} \frac{\sum_{n=1}^N \mu(n)}{N} = 0.$$

Recall that the Möbius function is 0 if  $n$  is not squarefree (which we know occurs with density  $1 - \frac{6}{\pi^2}$ ) and is  $(-1)^r$  if  $n$  is a product of  $r$  distinct primes. We also saw that the set of all positive integers divisible by only a bounded number, say  $k$ , of primes is equal to zero, so most integers  $1 \leq n \leq N$  are divisible by lots of primes, and by adding up the values of  $\mu$  we are recording  $+1$  if this large number is even and  $-1$  if this large number is odd. It is very tempting to view this parity as being essentially random, similar to what would happen if we flipped a coin for each (squarefree)  $n$  and gave ourselves  $+1$  if we got heads and  $-1$  if we got tails.

With this “randomness” idea planted in our mind, the above theorem seems to assert that if we flip a large number  $N$  of coins then (with large probability) the number of heads minus the number of tails is small compared to the total number of coin flips. But now it seems absolutely crazy that this result is equivalent to PNT since – under the (as yet completely unjustified) assumption of randomness – it is far too weak: doesn’t probability theory tell us that the running total of heads minus tails will be likely to be on the order of the **square root** of the number of coin flips? Almost, but not quite. And *is* this probabilistic model justified? Well, that is the \$ 1 million dollar question.

## 2. COIN-FLIPPING AND THE RIEMANN HYPOTHESIS

Let us define the **Mertens function**

$$M(N) = \sum_{n=1}^N \mu(n).$$

The goal of this lecture is to discuss the following seemingly innocuous question.

**Question 1.** *What is the upper order of  $M(N)$ ?*

---

<sup>1</sup>Nevertheless Selberg became one of the great analytic number theorists of the 20th century: it turned out that the elementary proof of PNT was among his more minor work.

Among other incentives for studying this question there is a large financial one: if the answer is close to what we think it is, then proving it will earn you \$ 1 million!

Recall  $\mu(n)$  takes on only the values  $\pm 1$  and 0, so the “trivial bound” is

$$M(N) \leq N.$$

In fact we can do better, since we know that  $\mu(n) = 0$  iff  $n$  is not squarefree, and we know, asymptotically, how often this happens. This leads to an asymptotic expression for the “absolute sum”:

$$\sum_{n=1}^N |\mu(n)| = \#\{\text{squarefree } n \leq N\} \sim \frac{6}{\pi^2} N.$$

However, in the last lecture we asserted that  $\frac{M(N)}{N} \rightarrow 0$ , which we interpreted as saying that the average order of  $\mu$  is asymptotically 0. Thus the problem is one of *cancellation* in a series whose terms are sometimes positive and sometimes negative. Stop for a second and recall how much more complicated the theory of “conditional” convergence of such series is than the theory of convergence of series with positive terms. It turns out that the problem of *how much cancellation to expect* in a series whose terms are sometimes positive and sometimes negative (or a complex series in which the arguments of the terms are spread around on the unit circle) is absolutely a fundamental one in analysis and number theory. Indeed in such matters we can draw fundamental inspiration (if not proofs, directly) from **probability theory**, and to do so – i.e., to make heuristic probabilistic reasoning even in apparently “deterministic” situations – is an important theme in modern mathematics ever since the work of Erdős and Kac in the mid 20th century.

But our story starts before the 20th century. In the 1890’s Mertens<sup>2</sup> conjectured:

$$(MC1) \quad M(N) \leq \sqrt{N} \text{ for all sufficiently large } N.$$

This is quite bold. As we have seen, in studying orders of magnitude, it is safer to hedge one’s bets by at least allowing a multiplicative constant, leading to the weaker

$$(MC2) \quad M(N) \leq C\sqrt{N} \text{ for all } N.$$

The noted Dutch mathematician Stieltjes claimed a proof of (MC2) in 1885. But his proof was never published and was not found among his papers after his death.

It would be very interesting to know exactly why Mertens believed in (MC1). He did check the inequality for all  $N$  up to  $N = 10^4$ , but this is an amusingly small search by contemporary standards. (In your homework you are asked to fire up your computer to compute many more values than this.) The problem is not as computationally tractable as one might wish, because computing the Möbius function requires factorization of  $n$ , which is famously rather hard. Nevertheless we now know that (MC1) holds for all  $N \leq 10^{14}$ .

---

<sup>2</sup>Franz Mertens, 1840–1927

It says something about the difficulty of such questions that, while the mathematical community has viewed (MC1) and (MC2) with dubiousness for some time, (MC1) was disproved only in 1985:

**Theorem 7.** (*de Riele, Odlyzko*): *There are explicit constants  $C_1 > 1$ ,  $C_2 < -1$  such that*

$$\limsup_N \frac{M(N)}{\sqrt{N}} \geq C_1,$$

$$\liminf_N \frac{M(N)}{\sqrt{N}} \leq C_2.$$

That is to say, each of the inequalities  $-N \leq M(N)$  and  $M(N) \leq N$  fails for infinitely many  $N$ . Their proof does not supply a concrete value of  $N$  for which  $M(N) > \sqrt{N}$ , but we know now that there is such an  $N < 10^{156}$ .

We still do not know whether (MC2) holds – so conceivably Stieltjes was right all along and the victim of some terrible mix up – although I am about to spin a tale to try to persuade you that (MC2) should be almost, but not quite, true.

But first, what about the million dollars?

In the last section we mentioned two interesting “equivalents” of PNT. The following theorem takes things to another level:

**Theorem 8.** *The following statements (none of which are known!) are equivalent:*

a) *For all  $\epsilon > 0$ , there exists a constant  $C_\epsilon$  such that  $|M(N)| \leq C_\epsilon N^{\frac{1}{2} + \epsilon}$ .*

b)  *$|\pi(x) - \text{Li}(x)| \leq \frac{1}{8\pi} \sqrt{x} \log x$  for all  $x \geq 2657$ .*

c) *Suppose  $\zeta(s_0) = 0$  for some  $s_0$  with real part  $0 < \Re(s_0) < 1$ . Then  $\Re(s_0) = \frac{1}{2}$ .*

We note that the rather esoteric-sounding part c) – which refers to the behavior of the zeta function in a region which it is not obvious how it is defined (but was first shown by Riemann to be defined there) – is the illustrious **Riemann hypothesis**. One can see immediately why we care about this weird hypothesis: it is equivalent to a wonderful error bound in the prime number theorem (which one can show to be essentially “best possible” – it is known that the error *cannot* be taken to be  $C\sqrt{x}$  for all  $x$ ). In 2000 the Clay Math Institute set the Riemann Hypothesis as one of seven \$ 1 million prize problems. If you don’t know complex analysis, no problem: just prove part a), that the cancellation in the partial sums of the Möbius function is enough to make the sum less than or equal to a constant times any power of  $N$  greater than  $\frac{1}{2}$ .

Note that (MC1) (which is false!)  $\implies$  (MC2)  $\implies$  condition a) of the theorem, so in announcing a proof of (MC2) Stieltjes was announcing a *stronger* result than the Riemann hypothesis, which did not have a million dollar purse in his day but was no less a mathematical holy grail then than now. (So you can decide how likely it is that Stieltjes’s paper got lost in the mail and never found.)

But why should we believe in the Riemann hypothesis anyway? There is some experimental evidence for it – in any rectangle  $|t| \leq N$ ,  $0 < \sigma < 1$  the zeta function

can have only finitely many zeros (this holds for any function meromorphic on  $\mathbb{C}$ ), so one can “find all the zeros” up to a certain imaginary part, and the fact that all of these zeros lie on the critical line – i.e., have real part  $\frac{1}{2}$  – has been experimentally confirmed in a certain range of  $t$ . It is also known that there are infinitely many zeros lying on the critical line (Hardy) and that even a positive proportion of them as we go up lie on the critical line (Selberg – as I said, a great mathematician). For various reasons this evidence is rather less than completely convincing.

So let us go back to randomness – suppose  $\mu$  really were a random variable. What would it do, in all probability?

We can consider instead the random walk on the integers, where we start at 0 and at time  $i$ , step to the right with probability  $\frac{1}{2}$  and step to the left with probability  $\frac{1}{2}$ . Formally speaking, our walk is given by an infinite sequence  $\{\epsilon_i\}_{i=1}^{\infty}$ , each  $\epsilon_i = \pm 1$ . The set of all such sign sequences,  $\{\pm 1\}^{\infty}$  forms in a natural way a probability space (meaning it has a natural measure – but don’t worry about the details; just hold on for the ride). Then we define a random variable

$$S(N) = \epsilon_1 + \dots + \epsilon_n,$$

meaning a function that we can evaluate on any sign sequence, and it tells us where we end up on the integers after  $N$  steps. Now the miracle of modern probability theory is that it makes perfect sense to ask what the lim sup of  $S_N$  is.

If you happened to catch an undergraduate course in probability theory (good for you. . .) you will probably remember that  $S_N$  should be no larger than  $\sqrt{N}$ , more or less. But this seems disappointing, because that is (MC1) (or maybe (MC2)), which feels quite dubious for the partial sums of the Möbius function. But in between Mertens’ day and ours probability theory grew up, and we now know that  $\sqrt{N}$  is not *exactly* the correct upper bound. Rather, it is given by the following spectacular theorem:

**Theorem 9.** (Kolmogorov) *With probability 1, we have*

$$\limsup_{N \rightarrow \infty} \frac{S_N}{\sqrt{2N \log \log N}} = 1.$$

That is to say, if you randomly flip a fair coin  $N$  times, then in all probability there will be infinitely many moments in time when your running tally of heads minus tails is larger than any constant times the square root of the number of flips. (Similarly, and symmetrically, the limit infimum is  $-1$ .) So true randomness predicts that (MC2) is false. On the other hand, it predicts that the Riemann Hypothesis is true, since indeed for all  $\epsilon > 0$  there exists a constant  $C_\epsilon$  such that  $\sqrt{2 \log \log N} < C_\epsilon N^\epsilon$ .

So if we believed in the “true randomness” of  $\mu$ , we would believe the following

**Conjecture 10.**

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{M(N)}{\sqrt{N \log \log N}} &< \infty. \\ \liminf_{N \rightarrow \infty} \frac{M(N)}{\sqrt{N \log \log N}} &> -\infty. \end{aligned}$$

Just to make sure, this conjecture is still significantly more precise than the  $|M(N)| \leq C_\epsilon N^{\frac{1}{2}+\epsilon}$  which is equivalent to the Riemann Hypothesis, making it unclear exactly how much we should pay the person who can prove it: \$ 2 million? Or more??

Kolmogorov’s “law of the iterated logarithm,” and hence Conjecture 10, does not seem to be very well-known outside of probabilistic circles.<sup>3</sup> In searching the literature I found a paper from the 1960’s predicting such a “logarithm law” for  $M(N)$ . More recently I have seen another paper suggesting that perhaps it should be  $\sqrt{\log \log \log N}$  instead of  $\sqrt{\log \log N}$ . To be sure, the Möbius function is clearly *not* random, so one should certainly be provisional in one’s beliefs about the precise form of the upper bounds on  $M(N)$ . The game is really to decide whether the Möbius function is “random enough” to make the Riemann hypothesis true.

Nevertheless the philosophy expressed here is a surprisingly broad and deep one: whenever one meets a sum  $S_N$  of  $N$  things, each of absolute value 1, and varying in sign (or in argument in the case of complex numbers), one wants to know how much cancellation there is, i.e., how far one can improve upon the trivial bound of  $|S_N| \leq N$ . The mantra here is that if there is really no extra structure in the summands – i.e., “randomness” – then one should expect  $S_N \approx \sqrt{N}$ , more or less! More accurately the philosophy has two parts, and the part that expresses that  $|S_N|$  should be *no smaller* than  $\sqrt{N}$  unless there is hidden structure is an extremely reliable one. An example of hidden structure is  $a_n = e^{\frac{2\pi i}{N}}$ , when in fact

$$\sum_{n=1}^n a_n = 0.$$

But here we have chosen to sum over all of the  $N$ th roots of unity in the complex plane, a special situation. The second part of the philosophy allows us to hope that  $S_N$  is not *too much* larger than  $\sqrt{N}$ . In various contexts, any of  $C\sqrt{N}$ ,  $\sqrt{N} \log N$ ,  $N^{\frac{1}{2}+\epsilon}$ , or even  $N^{1-\delta}$  for some  $\delta > 0$ , may count as being “not too much larger.” So in truth our **philosophy of almost squareroot error** is a little bit vague. But it can be, and has been, a shining light in a dark place,<sup>4</sup> and we will see further instances of such illumination.

Acknowledgement: The two lectures on these topics were delivered without formal lecture notes but only a small cheat sheet. I would have had difficulty recovering what was said were it not for the very clear class notes of Ms.<sup>5</sup> Diana May.

---

<sup>3</sup>I learned about Kolmogorov’s theorem from a talk at Harvard given by W. Russell Mann.

<sup>4</sup>When all other lights go out?

<sup>5</sup>Now Dr.