university of groningen

faculty of science and engineering

mathematics and applied mathematics

An application of Tauber theory: proving the Prime Number Theorem

Bachelor's Project Mathematics

February 2018

Student: J. Koolstra

Supervisors: Prof.dr. J. Top and Dr. A. Sterk

Abstract

We look at the origins of Tauber theory, and apply it to prove the prime number theorem (PNT). Specifically, we prove a weak version of the Wiener-Ikehara Tauberian theorem due to Newman. Its application requires us to establish some properties of the Riemann zeta function. Most notably with regard to its meromorphic continuation, and the distribution of its zeros.

Contents

Even before I had begun my more detailed investigations into higher arithmetic, one of my first projects was to turn my attention to the decreasing frequency of primes, to which end I counted primes in several chiliads. I soon recognized that behind all of its fluctuations, this frequency is on average inversely proportional to the logarithm.

Gauss to Encke, 1849

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again. The never-satisfied man is so strange; if he has completed a structure, then it is not in order to dwell in it peacefully, but in order to begin another. I imagine the world conqueror must feel thus, who, after one kingdom is scarcely conquered, stretches out his arms for others.

Gauss to Bolyai, 1808

ὑμεῖς τε γὰρ οἱ λέγοντες μάλιστ᾽ ἂνοὕτως ἐν ἡμῖν τοῖς ἀκούουσιν εὐδοκιμοῖτε καὶ οὐκ ἐπαινοῖσθε—εὐδοκιμεῖνμὲν γὰρ ἔστιν παρὰ ταῖς ψυχαῖς τῶν ἀκουόντων ἄνευ ἀπάτης, ἐπαινεῖσθαιδὲ ἐν λόγῳ πολλάκις παρὰ δόξαν ψευδομένων.

Prodicus in the Protagoras

Introduction

Tauber theory grew out of a single observation of A. Tauber (1866-1942) in 1897. He proved a nontrivial condition under which Abel summability implies ordinary summability. The necessity of such a condition had always been obvious, because Abel summability is easily seen to be stronger than ordinary summability, but a neat explicit formulation of it was new. Now the quest was on to establish more and stronger similar results.

The famous collaborators G. H. Hardy (1877–1947) and J. E. Littlewood (1885–1977) were the first to pick up on the result and realize its potential as a representative of a general theory. They were intrigued by the idea and together set out to derive many related but much more intricate Tauber type theorems. In particular, in 1911, Littlewood significantly weakened the assumption of Tauber's original result, and later in 1914, together with Hardy, proved a nontrivial condition for moving from Abel summability to the more restrictive Cesàro summability. They dubbed these results "Tauberian theorems", antonymic to the well known Abelian theorems.

In 2004, J. Korevaar published an article "A simple proof of the prime number theorem" [1] that surveys the modification of Newman's 1980 simple proof of the prime number theorem (PNT) that we shall primarily study.[1] PNT states that the number of primes under x is asymptotically distributed as $x/\log x$. In the article, Newman's method is adapted to prove a weak version of the Wiener-Ikehara Tauberian theorem. With the knowledge of the Riemann zeta function and Tchebychef's ψ-function that we develop, this theorem is strong enough to imply PNT.

Korevaar explains that in particular number theory, and especially the search for simpler proofs of the prime number theorem, have formed a major impetus for the development of a more general Tauber theory. The direction we take in this thesis to highlight Tauber theory therefore presents a natural approach. Tauber theory has developed in the past century into a well established field of research, with many more deep results and techniques that are far beyond our scope. Indeed, the prime number theorem is only the start.

We proceed as follows. In the next section we introduce the concept of summability and then focus on the summability methods of Cesàro and Abel. We also prove how these methods can be related using the Tauberian theorems of Tauber and Hardy-Littlewood. In the subsequent section our

[1] The article is in Dutch. It appeared on the occasion of the publication of Korevaar's survey book on Tauber theory, "Tauberian theory, a century of developments".

attention shifts from the discrete to the continuous, and we derive a weak version of the Wiener-Ikehara theorem, dubbed the "poor man's" Wiener-Ikehara theorem. It forms the main body of the thesis, and most of the hard work. Then we are ready to prove the prime number theorem. We do so in two parts. First we derive all the essential properties of the Riemann zeta function that we need, featuring most prominently its meromorphic continuation to the complex plane, and its non-vanishing on the boundary line of its natural domain of definition. Finally we combine all the ingredients of the preceding sections, and finish the proof of the prime number theorem by relating it to the Tchebychef ψ-function.[2]

[2] With regard to prerequisites, it suffices to read, for example, Stein and Shakarchi [2], chapters I through III. Most importantly, one should know a bit about holomorphic functions and their basic properties, like analyticity, and having a unique analytic continuation. Assumed is Cauchy's theorem, and Cauchy's integral formula.

Discrete Tauber theory

We begin with a brief introduction of the concept of summability, and then quickly focus on the summability methods of Cesàro and Abel (Definition 1 and 3 respectively). Most importantly we prove that in a precise sense Abel summability is stronger than Cesàro summability (Theorem 1). We also prove how these methods can be related the other way around, and to ordinary summability, using the (discrete) Tauberian theorems of Tauber, Littlewood and Hardy-Littlewood (resp. Theorem 2, 4 and 3).

Summability methods

Techniques for assigning to (divergent) series reasonable sums are called summability methods. Taken together they allow us to form a notion of summability that can function as an object of study in itself. The most natural and common such summability method is to assign to a series the limit of its partial sums, as in calculus. Generalized concepts of summability, and older attempts thereon, grew out of an interest in divergent series with an appealingly simple structure or natural occurrence. Such series arise for example as formal solutions to certain differential equations, or by pondering over the meaning of sums like $1 + 1 + 1 + \cdots$. Interestingly, specialized summability methods have found use beyond the pure mathematical in theoretical physics.

Before Cauchy, Bolzano and Weierstrass introduced modern rigor in analysis, and so for the first time offered precise definitions of intuitive concepts like convergence and divergence, divergent series had been the subject of various intense debates. Indeed, the treatment of divergent series prior to the formalization of these foundations early in the nineteenth century, had mostly relied on a kind of heuristic reasoning that was notoriously intractable. It led Abel to conclude in 1826 that "divergent series are an invention of the devil". Afterwards, as a result of this widespread sentiment, that was promulgated by the certainty of the new rigor, it took a surprisingly long time before someone dared to get involved with divergent series again, and therefore also for the modern concept of summability to appear. The most notable such new involvement, in which this concept is made explicit, is undoubtedly Hardy's 1949 book "Divergent series". In its preface, his student Littlewood remarks that "in the early years of the century the subject, while in no way mystical or unrigorous, was regarded as sensational, and about the present title, now colorless, there hung an aroma of paradox and audacity".

A prototypical example of an appealing divergent series is the Grandi series $1-1+1-1+\cdots$.[3] It diverges; but that is not a very mathematically satisfying conclusion. Would it not be more natural to make sense of the sum as the average position of the partial sums, jumping between 0 and 1, and *assign* it the sum 1/2? And if we agree on this assignment, can we then fit it in a more general framework, that handles more such diverging series? In fact, this is precisely Cesàro summation, the first important summability method we look at.

Definition 1 (Cesàro summability)**.** *Given is a sequence of numbers* $\{a_n\}_{1\leq n<\infty}$ *with partial sums* $s_n = \sum_{k=1}^n a_k$*. We define the new sequence*

$$\sigma_n = \frac{1}{n}\sum_{k=1}^n s_k$$

that averages over these partial sums. If the limit $\lim \sigma_n = A$ *exists, we say that the formal series* $\sum a_n$ *is* Cesàro summable*, and* assign *it the limit* A*, called the* Cesàro sum *of the series.*

In contrast, we refer to the usual interpretation of series as according to the *ordinary summability method*. It is easily verified that the Grandi series is indeed Cesàro summable with sum 1/2.

Without proof let us now make a simple yet prototypical observation.

Observation 1. *Ordinary summability implies Cesàro summability. Moreover, if a series can be ordinarily summed, then both methods assign the same value to that series.*

Some terminology is in place. We will not need all of it, but it is useful to know some in order to get acquainted with how we would like to interact mathematically with summability methods. How to think about them effectively, and how to organize them by their characteristics.

A *summability method* Σ, sometimes referred to as Σ-*summation*, is defined as a (partial) function that maps sequences of numbers (from $\mathbb{C}^{\mathbb{N}}$) into the complex plane. In the context of Σ we identify these sequences with the series they formally define. For example we may reference the series $\sum a_n$ as S, and then identify $\Sigma(S)$ with $\Sigma(\{a_n\})$. In this case S does not refer to the outcome of the series, which might not even exist, but rather to the

[3] Wikipedia has a fairly good coverage of the interesting history of the Grandi series, maintained on the page "History of the Grandi series". It is quite representative of the attitude towards divergent series during the various historical periods discussed.

series as an object itself. So even $\Sigma(\sum a_n)$ has an unambiguous and intuitive meaning.

We introduce the following convenient jargon, in line with the terminology presented for Cesàro summability.

Definition 2 (A dictionary of summability)**.** *If a series S is in the domain of Σ, we say that it is* Σ-summable *and call $\Sigma(S)$ the* Σ-sum *of the series. We sometimes also say that Σ sums the series S to the sum $\Sigma(S)$. Furthermore, we isolate and name the following useful general properties that a summability method might have.*

1. Regularity*. If Σ sums all ordinarily convergent series to their ordinary sum, it is called* regular*.*

2. Linearity*. Let S and T be series summed by Σ and c some constant. We say that Σ is* linear *if $cS + T$ is Σ-summable and $\Sigma(cS + T) = c\Sigma(S) + \Sigma(T)$.*

3. Stability*. The method Σ is called* stable *if we can shift the initial terms of a series like $\Sigma(\sum_{n=0}^{\infty} a_n) = \Sigma(\sum_{n=1}^{\infty} a_n) + a_0$.*

In addition, suppose Π is another summability method, then Σ-summation is said to *conserve* Π if it sums all series summed by Π to their Π-sum. So regularity is nothing else than the conservation of ordinary summation. Also, if in addition Σ sums any other series not summed by Π, then Σ is called *stronger* than Π.

The notion of conservation induces a partial ordering $\preceq$ on the set of summability methods. This gives rise to the shorthands $\Pi \preceq \Sigma$ and $\Pi \prec \Sigma$ for the above concepts.

We can now succinctly state an improved observation about Cesàro summation.

Observation 2. *Cesàro summability is stronger than ordinary summability. Moreover, it is linear and stable.*

Next we turn to Abel summation, the other important summability method we discuss.

Definition 3 (Abel summability)**.** *Given is a sequence of numbers $\{a_n\}_{0 \leq n < \infty}$. We define for $0 \leq r < 1$ the family of* Abel means *$A(r)$ as the power series*

$$A(r) = \sum_{n=0}^{\infty} a_n r^n \,.$$

If all Abel means exist and converge to some value A as $r \to 1^-$, then $\sum a_n$ is said to be Abel summable *to A.*

The goal of the remainder of this subsection is to prove the next theorem, and simultaneously relate Abel summation to the idea behind Cesàro summability.

Theorem 1. *Abel summability is stronger than Cesàro summability.*

Proof. Let s_n denote the partial sums of the a_n sequence. Recall Hadamard's formula for the radius of convergence R of the power series $\sum a_n r^n$, that is

$$\frac{1}{R} = L = \limsup \sqrt[n]{|a_n|}\,.$$

A quick consequence is that the related power series $S = \sum_{n=0}^{\infty} s_n r^n$ has the same radius of convergence $R_S = R$. So we may write

$$\begin{aligned} A(r) &= a_0 + \sum_{n=1}^{\infty}(s_n - s_{n-1})r^n \\ &= \sum_{n=0}^{\infty} s_n r^n - \sum_{n=0}^{\infty} s_n r^{n+1} \\ &= \sum_{n=0}^{\infty} s_n (1-r) r^n\,. \end{aligned} \tag{1}$$

Notice that $\sum_{n=0}^{\infty}(1-r)r^n = 1$. This tells us that the Abel means form a parameterized family of summability methods that are constructed by weighted averaging of the partial sums. Abel summability is then the limit of the sums assigned by these methods as the family parameter is taken to $r \to 1^-$. Thus the summability methods discussed so far can be compared and summarized by saying that ordinary summation puts the full weight on s_n, Cesàro summation weights everything equally, and Abel summation assigns a family of weights $(1-r)r^n$.

Here peeks a connection to Tauber theory. Indeed, in the lecture notes [3] that we will mostly follow in the next subsection, Yum-Tong Siu explains that "nowadays a Tauberian theorem means a statement which uses an appropriate Tauberian condition to guarantee that a given way of taking weighted average (or weighted integral) gives the usual limit when the parameter in the given family of weighted average (or weighted integral) goes to an appropriate limit value".

Another consequence of equation (1) is the regularity of Abel summability. From Hadamard's formula it is immediate that the Abel means $A(r)$ exist when we assume the ordinary summability of $\sum a_n$, and hence the first condition of Abel summability is satisfied. Now without loss of generality also assume that $s_n \to 0$, so that it suffices to prove $\lim_{r\to 1^-} A(r) = 0$. Regularity then follows by ϵ-squeezing when we split the series in the right-hand side of (1) such that $|s_n| < \epsilon$ for all $n > k$, and bound

$$\begin{aligned} |A(r)| &\leq \sum_{n=0}^{\infty} |s_n|(1-r)r^n \\ &\leq (1-r)(|s_0| + \cdots + |s_k|) + (1-r) \sum_{n=k+1}^{\infty} \epsilon r^n \\ &= (1-r)(|s_0| + \cdots + |s_k|) + \epsilon r^k \,. \end{aligned}$$

In fact, this proves the theorem completely. Indeed, by simply applying equation (1) again we obtain

$$A(r) = (1-r)^2 \sum_{n=0}^{\infty} n\sigma_n r^n \,,$$

which can be treated as above.

The only remaining task is to exhibit a sum that is Abel summable but not Cesàro summable. The standard example is $\sum(-1)^{n+1}n$. $\square$

Where does this leave us? Figure 1 provides an overview of our current hierarchy. Such figures provide good aid if we quickly want to formulate some interesting questions. For example, we see that certain series are still out of reach of even Abel summability. Which series are those? What sort of methods might sum them? And how can we relate these methods to our current hierarchy? Are they also stable and linear?

One of the series currently beyond our bounds is $\sum n$, which plays a role in theoretical physics (string theory). It turns out that we can reasonably sum it to $-1/12$. A rather counter-intuitive sum if we consider that we are adding only positive integers. Interestingly, the same summation can be achieved by means of several very different techniques.[4] We will look

[4]One of these techniques, that is of particular interest, is due to Ramanujan (1887–1920). In 1913, he writes to Hardy: "I was expecting a reply from you similar to the one which a Mathematics Professor at London wrote asking me to study carefully Bromwich's Infinite Series and not fall pitfalls of divergent series. [...] I told him that the

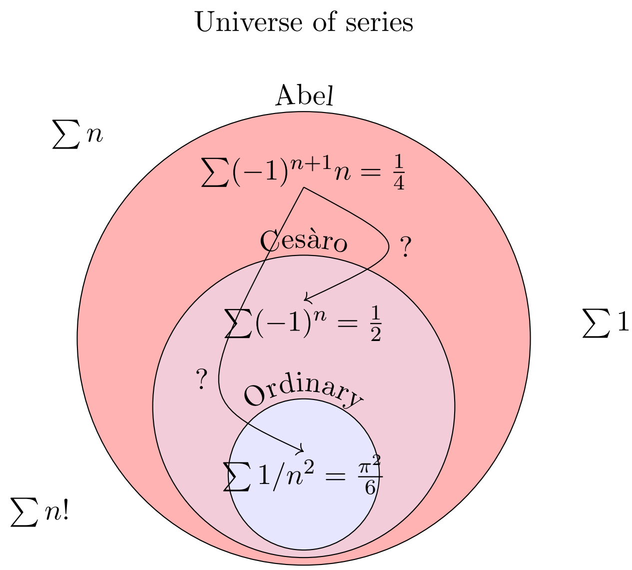

Figure 1: The $\Sigma_{\text{Ordinary}} \prec \Sigma_{\text{Cesàro}} \prec \Sigma_{\text{Abel}}$ hierarchy of summability methods. The reverse-inclusion conditions, listed next to the arrows, are not yet known to us, and are therefore indicated by a question mark. Note that inclusion here means not only summing the same series, but also doing so to the same sum.

at one particular method, that uses the analytic continuation of the zeta function. In fact, we shall see the zeta function and its extension a lot when we get to the prime number theorem, so that the result will just be a corollary (see Corollary 1). Unfortunately we will have to leave open how zeta function regularization, and other summability methods based on analytic continuation, fit in our simple hierarchy.

There are many more interesting questions that we cannot answer here, as they would go beyond the scope of this thesis. What we will answer however, in the next subsection, is which conditions can replace the question marks in the figure.

sum of an infinite number of terms of the series: $1 + 2 + 3 + 4 + \cdots = -\frac{1}{12}$ under my theory. If I tell you this you will at once point out to me the lunatic asylum as my goal. ".

The theorems of Tauber and Hardy-Littlewood

Theorems that prove one summability method conserving another are called Abelian theorems. A prime example is Theorem 1, which is indeed often referred to simply as Abel's theorem.[5]

Alfred Tauber proved in 1897 the first so-called Tauberian theorem. These theorems are the antonyms of Abelian theorems. They establish nontrivial conditions under which a weak summability method (usually the ordinary) conserves a strong one, so that they are of equal strength when the conditions are satisfied.

It is not at all obvious that such *Tauberian conditions* can be formulated for any meaningful hierarchy of summability methods. Remarkably, Tauber established a simple sufficient condition for the Cesàro-Abel hierarchy of Figure 1. For its proof, and the proofs of the other two (discrete) Tauberian theorems we discuss, we follow [3], which in turn follows the elegant proofs of Karamata of 1930 [7].

To maintain a clear relation to the previous subsection, we will here consider only *discrete* Tauberian theorems, roughly those theorems that involve summation instead of integration. The continuous Wiener-Ikehara theorem is presented in the next section, where we also present a more detailed comparison between the two types of Tauberian theorems, discrete and continuous.

Theorem 2 (Tauber, 1897)**.** *Under the Tauberian condition $na_n \to 0$, Abel summability implies ordinary summability.*

Proof. The proof is another example of the sort of series splitting arguments we have seen earlier.

Define the number $N(r) = \lfloor \frac{1}{1-r} \rfloor$, so that $N(1-r) \leq 1$ and $(N+1)(1-r) > 1$. Because $N \to \infty$ as $r \to 1^-$, and the limit $\lim_{r \to 1^-} \sum_{n=0}^{\infty} a_n r^n$ is guaranteed to exist by assumption of Abel summability, it suffices to prove

$$\lim_{r \to 1^-} \left(\sum_{n=0}^{\infty} a_n r^n - \sum_{n=0}^{N} a_n \right) = 0 \,.$$

[5] Abel's theorem might be familiar from real analysis. A common equivalent formulation is that the pointwise convergence of a power series on a set A implies the uniform convergence of that power series on any compact subset $K \subset A$. (See for example theorem 6.5.5 of [4].)

We split the series into

$$\sum_{n=N+1}^{\infty} a_n r^n - \sum_{n=0}^{N} a_n (1 - r^n) \, .$$

By the Tauberian condition there exists for arbitrary $\epsilon > 0$ a number N_0 such that whenever $n \geq N_0$ we have the bound $|na_n| < \epsilon$. Pick $\delta > 0$ sufficiently small to satisfy $N(r) \geq N_0$ whenever $1 - \delta < r < 1$. With $r(\delta)$ close enough to 1, we can then bound the first term in the split as

$$\begin{aligned} \left| \sum_{n=N+1}^{\infty} a_n r^n \right| &= \left| \sum_{n=N+1}^{\infty} n a_n \frac{r^n}{n} \right| \\ &\leq \epsilon \sum_{n=N+1}^{\infty} \frac{r^n}{n} \\ &\leq \frac{\epsilon}{N+1} \sum_{n=0}^{\infty} r^n \\ &= \frac{\epsilon}{(N+1)(1-r)} < \epsilon \, . \end{aligned}$$

For the second term, use again the Tauberian condition $na_n \to 0$, and recall observation 1, to justify the bound $\frac{1}{N} \sum_{n=0}^{N} |na_n| < \epsilon$ for large enough N, and therefore

$$\begin{aligned} \left| \sum_{n=0}^{N} a_n (1 - r^n) \right| &= \left| \sum_{n=0}^{N} a_n (1 - r)(1 + r + \cdots + r^{n-1}) \right| \\ &\leq \sum_{n=0}^{N} |na_n| (1 - r) \\ &\leq N\epsilon(1 - r) < \epsilon \, . \end{aligned}$$

$\square$

In 1914, Hardy and Littlewood established a similar minded result for going from Abel summability to Cesàro summability.

Theorem 3 (Hardy-Littlewood, 1914)**.** *Under the Tauberian condition $s_n \geq 0$, Abel summability implies Cesàro summability. In fact, it is sufficient to assume that $s_n \geq -C$ for some positive constant C.*

Proof. Assume the Abel sum $\sum a_n r^n \to A$ as $r \to 1^-$. By equation (1), and substituting r with r^{k+1}, then

$$\sum_{n=0}^{\infty} s_n(1-r)r^n \to A \Rightarrow \sum_{n=0}^{\infty} s_n(1-r^{k+1})r^n(r^n)^k \to A\,, \tag{2}$$

for all $k \geq 0$, as $r \to 1^-$. Noting that

$$\lim_{r\to 1^-} \frac{1-r^{k+1}}{1-r} = k+1 = 1\Big/\int_0^1 t^k dt\ ,$$

we can therefore write

$$(1-r)\sum_{n=0}^{\infty} s_n r^n (r^n)^k \to A\int_0^1 t^k dt\,.$$

More general, by taking linear combinations, we see that for arbitrary polynomial $P(t)$ similarly

$$(1-r)\sum_{n=0}^{\infty} s_n r^n P(r^n) \to A\int_0^1 P(t)dt\,. \tag{3}$$

The Weierstrass approximation theorem tells us that for any piecewise continuous function g on $[0,1]$ and arbitrarily small $\epsilon > 0$, there exist polynomials $P_\epsilon(t)$ and $Q_\epsilon(t)$ such that $P_\epsilon \leq g \leq Q_\epsilon$ and $\|Q_\epsilon - P_\epsilon\| < \epsilon$. By (3) therefore

$$(1-r)\sum_{n=0}^{\infty} s_n r^n P_\epsilon(r^n) \geq -\epsilon + A\int_0^1 P_\epsilon(t)dt \geq -2\epsilon + A\int_0^1 g(t)dt$$

and

$$(1-r)\sum_{n=0}^{\infty} s_n r^n Q_\epsilon(r^n) \leq \epsilon + A\int_0^1 Q_\epsilon(t)dt \leq 2\epsilon + A\int_0^1 g(t)dt\,,$$

if we take r close enough to 1. Now we use the Tauberian condition $s_n \geq 0$ to obtain the sandwich

$$(1-r)\sum_{n=0}^{\infty} s_n r^n P_\epsilon(r^n) \leq (1-r)\sum_{n=0}^{\infty} s_n r^n g(r^n) \leq (1-r)\sum_{n=0}^{\infty} s_n r^n Q_\epsilon(r^n)\,,$$

and consequently

$$(1-r)\sum_{n=0}^{\infty} s_n r^n g(r^n) \to A\int_0^1 g(t)dt\,.$$

The left-hand side provides enough freedom to finish the proof, since the choice of g is only subject to the fairly weak constraint of piecewise continuity. In particular we need to look for piecewise continuous function g and numbers r_N such that $r_N \to 1$ as $N \to \infty$, and

$$r_N^n g(r_N^n) = \begin{cases} 1 & \text{if } n \leq N \\ 0 & \text{if } n > N \end{cases} .$$

We also need g to be normalized like $\int_0^1 g(t)dt = 1$. The above condition suggests choosing $g(t) = (1/t)\chi_{[(r_N)^N,\,1]}(t)$. By the normalization condition we must then pick $r_N = e^{-1/N}$, which indeed matches the requirement that $r_N \to 1$ as $N \to \infty$. Furthermore, this choice implies $g = (1/t)\chi_{[1/e,\,1]}$, so that we have g independent of N.

Given these choices we obtain

$$\lim_{N\to\infty}(1-r_N)\sum_{n=0}^{\infty} s_n r_N^n g(r_N^n) = \lim_{N\to\infty}(1-r_N)\sum_{n=0}^{N} s_n = A\,.$$

But

$$\lim_{N\to\infty} N(1-r_N) = \lim_{N\to\infty}\frac{1-e^{-1/N}}{1/N} = 1\,,$$

so we can introduce the partial Cesàro sums σ_N (see Definition 1) like

$$\lim_{N\to\infty} N(1-r_N)\frac{1}{N}\sum_{n=0}^{N} s_n = \lim_{N\to\infty}\sigma_N = A\,,$$

which is what we needed to prove.

By the linearity of Abel and Cesàro summation, it suffices to assume $s_n \geq -C$. Indeed, simply replace a_0 with $a_0 + C$ and apply the theorem. □

In 1911 Littlewood significantly weakened the Tauberian condition $a_n = o(1/n)$ of the Tauber's original theorem, demonstrating that it was by no means a necessity.[6] The proof is very similar to that of Hardy-Littlewood, but requires some more technical detours.

Theorem 4 (Littlewood, 1911)**.** *Under the Tauberian condition $a_n = \mathcal{O}(1/n)$, Abel summability implies ordinary summability. In fact, it is sufficient to assume that $na_n > -C$ for some positive constant C.*

[6] Weaker Tauberian conditions make for *stronger* Tauberian theorems. We therefore also say, a bit counterintuitive, that we have *improved* the Tauberian conditions when we have weakened them.

Proof. Assume the Abel sum $\sum a_n r^n \to A$ as $r \to 1^-$. The proof is again by a sandwiching argument, but we need different polynomials to match the changed objective. In particular we must get rid of the loose r^n in (2). This has as a consequence that we are restricted to polynomials without constant term.

By taking linear combinations, we have for any polynomial $P(t)$ with $P(0) = 0$ that

$$\sum_{n=0}^{\infty} a_n P(r^n) \to AP(1)\,,$$

as $r \to 1^-$. We also impose the normalizing constraint $P(1) = 1$.

To circumvent these restrictions, we define $P(t)$ in terms of a freely chosen polynomial $Q(t)$ by setting

$$P(t) = t + t(1-t)Q(t)\,.$$

No matter what polynomial Q is chosen now, P has the imposed properties $P(0) = 0$ and $P(1) = 1$.

Similar as in the proof of Hardy-Littlewood, we take $g(t) = \chi_{[(r_N)^N,\,1]}(t)$ and $r_N = e^{-1/N}$; so $g = \chi_{[1/e,\,1]}$. To match $Q(t)$, we define on $[0,1]$ the piecewise continuous function

$$h(t) = \frac{g(t) - t}{t(1-t)}\,.$$

For arbitrary $\epsilon > 0$ we then find $Q_\epsilon(t)$ such that $h \le Q_\epsilon$ and $\|Q_\epsilon - h\| < \epsilon$. Hence $P_\epsilon(t)$ is such that $g \le P_\epsilon$ and

$$\int_0^1 \left(\frac{P_\epsilon(t) - t}{t(1-t)} - \frac{g(t) - t}{t(1-t)} \right) dt = \int_0^1 \left(\frac{P_\epsilon(t) - g(t)}{t(1-t)} \right) dt < \epsilon\,, \tag{4}$$

where we observe that any boundary issues are resolved by the integrability of $\int_0^\delta 1/t^2\,dt$.

As a result of the assumed normalization, we already have

$$\lim_{r \to 1^-} \sum_{n=0}^{\infty} a_n P_\epsilon(r^n) = A\,.$$

From this and the above bound, we would like to prepare the top slice of the sandwich, and show

$$\limsup_{r \to 1^-} \sum_{n=0}^{\infty} a_n g(r^n) \le A\,. \tag{5}$$

Without loss of generality assume $a_0 = 0$. By the Tauberian condition and $g \le P_\epsilon$, we can bound

$$\begin{aligned}\sum_{n=0}^{\infty} a_n g(r^n) - \sum_{n=0}^{\infty} a_n P_\epsilon(r^n) &\le C \sum_{n=1}^{\infty} \frac{1}{n} \left(P_\epsilon(r^n) - g(r^n)\right) \\ &\le C \sum_{n=1}^{\infty} \frac{1-r}{1-r^n} \left(P_\epsilon(r^n) - g(r^n)\right) \\ &= C \sum_{n=1}^{\infty} (r^n - r^{n+1}) \frac{P_\epsilon(r^n) - g(r^n)}{r^n(1-r^n)} \,.\end{aligned}$$

The key insight to Karamata's proof of Littlewood's theorem, is that the last bound can be interpreted as a Riemann sum of (4), with mesh size going to 0 as $r \to 1^-$.[7] So it follows that

$$\limsup_{r\to 1^-} \sum_{n=0}^{\infty} a_n g(r^n) \le \lim_{r\to 1^-} \sum_{n=0}^{\infty} a_n P_\epsilon(r^n) + C \int_0^1 \left(\frac{P_\epsilon(t) - g(t)}{t(1-t)} \right) dt \le A + C\epsilon \,,$$

which gives (5) after squeezing the ϵ.

Analogously we can show the other side of the sandwich

$$\liminf_{r\to 1^-} \sum_{n=0}^{\infty} a_n g(r^n) \ge A \,.$$

Together with (5) this proves the theorem by the choice of g and r_N, since

$$\lim_{N\to\infty} s_N = \lim_{N\to\infty} \sum_{n=0}^{\infty} a_n g(r^n) = \lim_{r\to 1^-} \sum_{n=0}^{\infty} a_n g(r^n) \,.$$

$\square$

[7] Alternatively, sacrificing brevity, the nicer Darboux integral can be used. Begin by writing

$$\begin{aligned}&\left| \sum_{n=1}^{\infty} (r^n - r^{n+1}) \frac{P_\epsilon(r^n) - g(r^n)}{r^n(1-r^n)} - \int_0^1 \frac{P_\epsilon(t) - g(t)}{t(1-t)} dt \right| \\ &\quad \le \sum_{n=1}^{\infty} \left| \frac{P_\epsilon(r^n) - g(r^n)}{r^n(1-r^n)} (r^n - r^{n+1}) - \int_{r^{n+1}}^{r^n} \frac{P_\epsilon(t) - g(t)}{t(1-t)} dt \right| \\ &\quad \le \sum_{n=1}^{\infty} \left(\sup_{s,\, t \in [r^{n+1}, r^n]} \left| \frac{P_\epsilon(s) - g(s)}{s(1-s)} - \frac{P_\epsilon(t) - g(t)}{t(1-t)} \right| \right) (r^n - r^{n+1}) \,,\end{aligned}$$

then use uniform continuity, and the fact that the integrated function has only one jump discontinuity.

The Wiener-Ikehara theorem

For the sake of completeness, we begin by stating some technical complex analytic results (Theorem 5 and 6). Then we are ready to prove the central result of the thesis: the (weak) Wiener-Ikehara Tauberian theorem (Theorem 7). We do so in two steps. First we prove a Tauberian theorem for the Laplace transform (Theorem 9), and then use that to prove a simple reformulation of Wiener-Ikehara (Theorem 8). For the proofs we follow the lecture notes of Siu [3].

Some complex analytic technicalities

The following basic results are used (tacitly) throughout the next subsection. They are technicalities, but form fundamental witnesses to the power and success of complex analysis. Because we are interested in the prime number theorem, not a development of complex analysis, they are stated without proof.[8]

Theorem 5. *The limit function f of a sequence of holomorphic functions $\{f_n\}$, is holomorphic in Ω, if the convergence is uniform in every compact subset of Ω.*

Theorem 6. *Let $f(z)$ be defined on the open set $\Omega \subset \mathbb{C}$ in terms of a Riemann integral,*

$$f(z) = \int_0^1 F(z,t)\,dt\,.$$

Suppose that:

1. *$F(z,t)$ is holomorphic in z for each t.*
2. *F is continuous on $\Omega \times [0,1]$.*

Then $f(z)$ is holomorphic on Ω.

A proof of the Wiener-Ikehara theorem

The Tauberian result that will ultimately entail the prime number theorem (PNT) is given in Korevaar's article [1] as

[8] The Cauchy-Goursat theorem and Cauchy's integral formula are also assumed known. All mentioned statements can be found, with their proofs, in for example [2] (chapter 2).

Theorem 7 (The weak Wiener-Ikehara theorem)**.** *Suppose that the Dirichlet series*

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n}{n^z},$$

with coefficients $a_n \geq 0$*, converges in the half plane* $\{z : \operatorname{Re} z > 1\}$*; so that the summation function* $f(z)$ *is automatically holomorphic in that half plane. Now also suppose that there exists a constant* A *such that the difference*

$$g(z) = f(z) - \frac{A}{z-1}$$

can be analytically continued to include the closure $\{z : \operatorname{Re} z \geq 1\}$ *of the domain of* $f(z)$*. And finally assume that* $s_n = \sum_{k=1}^{n} a_n$ *is in* $\mathcal{O}(n)$*. Then* $s_n/n \to A$ *as* $n \to \infty$*, that is* $s_n \sim An$*.*

In proving PNT, we will set $a_n = \Lambda(n)$, where $\Lambda(n)$ is the von Mangoldt function (Definition 5). The partial sum of this sequence is $s_n = \psi(n)$, the second Tchebychef function (Definition 6). It is straightforward to establish the required bound $\psi(n) \leq Cn$ (Theorem 14).

With respect to the Tauberian condition, we show in the next section that the Wiener-Ikehara theorem links $\psi(n)$, through the choice of a_n, to the Riemann zeta function $\zeta(z)$ (Definition 4) via the logarithmic derivative as $f(z) = -\zeta'(z)/\zeta(z)$ (see Theorem 12). Crucially, we prove that $\zeta(z)$ has a meromorphic continuation to the open right half plane $\{z : \operatorname{Re} z > 0\}$ (Theorem 10), and is nonzero on the critical line $\operatorname{Re} z = 1$ (Theorem 11). The behavior of $\zeta(z)$ around the simple pole $z = 1$ will then allow us to conclude that $g(z)$ can be analytically continued to the required half plane closure, by setting $A = 1$.

Satisfying all conditions, we can apply the weak Wiener-Ikehara theorem to obtain $\psi(n) \sim n$. Finally, a simple argument that is proved in the last section shows that PNT is equivalent to $\psi(n) \sim n$ (Theorem 15).

The difference with the original Wiener-Ikehara theorem, and the weakness in the above formulation, is that from the other assumptions alone, one can in fact deduce the supposition $s_n = \mathcal{O}(n)$. Newman's insight was that the full strength of the Wiener-Ikehara theorem is not needed to derive the prime number theorem from it. Moreover, he recognized that a proof of this simplification could be accomplished with considerably less sophistication. In particular, Newman was able to replace the used Wiener theory with some clever contour integration, requiring nothing more than Cauchy's theorem.

We now set out to prove Theorem 7. Recall Abel's partial summation formula

$$\sum_{1\le n\le \lfloor x\rfloor} a_n a(n) = s(x)a(x) - \int_1^x s(t)a'(t)dt\,, \tag{6}$$

where $s(x) = \sum_{1\le n\le\lfloor x\rfloor} a_n$, and $a(x)$ is assumed continuously differentiable. Set $a(x) = 1/x^z$, then $a'(x) = -zx^{-z-1}$, and we see that it suffices to prove

Theorem 8. *Let $s(x)$, $1 \le x < \infty$, be a nonnegative, nondecreasing, piecewise continuous function, such that $s(x) \le Cx$ for some constant C. Define*

$$f(z) = z\int_1^\infty s(x)x^{-z-1}dx\,,$$

which is automatically holomorphic in half plane $\{z : \operatorname{Re} z > 1\}$ because $s(x) = \mathcal{O}(x)$. If $g(z) = f(z) - \frac{A}{z-1}$ can be analytically continued to an open neighborhood of the line $\operatorname{Re} z = 1$, then $s(x) \sim Ax$.

The advantage of this reformulation of the Wiener-Ikehara theorem, is that its relation to the previously discussed *discrete* Tauber theory becomes more explicit. That is, it is helpful in order to understand why we consider it a Tauberian theorem in the first place.

Recall from the previous section that "nowadays a Tauberian theorem means a statement which uses an appropriate Tauberian condition to guarantee that a given way of taking weighted average (or weighted integral) gives the usual limit when the parameter in the given family of weighted average (or weighted integral) goes to an appropriate limit value".

The proof of Theorem 8 will be a consequence of the following Laplace transform Tauberian theorem.

Theorem 9 (Laplace transform Tauberian theorem)**.** *Let $F(t)$, $0 \le t < \infty$, be a bounded, piecewise continuous function. If the Laplace transform of F,*

$$\mathcal{L}\{F\}(z) = \int_0^\infty F(t)e^{-zt}dt\,,$$

can be analytically continued to an open neighborhood U of the line $\operatorname{Re} z = 0$, then $\lim_{z\to 0}\mathcal{L}\{F\}(z) = \mathcal{L}\{F\}(0) = \int_0^\infty F(t)dt$.

Here, the family of weights used to average the content of the function $F(t)$ is e^{-zt}, with a complex-valued family parameter $z \in \{w : \operatorname{Re} w > 0\}$. In contrast with the discussed discrete Tauber theory, we now weight "slices" of a continuous function, not elements of a sequence, and use an integral, not a sum, to take the (function) content averages. Indeed, the function's content is indexed by the continuous interval $0 \le t < \infty$, and not the discrete interval $0 \le n < \infty$.[9]

The deliberately general term "content" can be understood similarly as the concept of norm. That is, a definition of size for a class of mathematical objects. The difference is that we do not insist on any (norm) axioms, and that we are specifically concerned with families of weighted averages, either discretely or continuously indexed. Moreover, we have a particular interest in content definitions that involve sums or integrals. The goal of Tauber theory is to link these back to ordinary (common) assignments of content, such as the sum of elements for sequences (discrete case), or integration over an interval for piecewise continuous functions (continuous case).

In the above Tauberian theorem, the Tauberian condition is expressed as an assumption on the analytic horizon of the family of content assignments $\mathcal{L}\{F\}(z)$, namely that it can be analytically continued to an open neighborhood of the line $\operatorname{Re} z = 0$. The conclusion is, as in the discrete case, that if we take the family parameter to an appropriate limit, here $z \to 0$, the means converge to the ordinary definition of content, here $\int_0^\infty F(t)dt$. That is, $\lim_{z\to 0} \mathcal{L}\{F\}(z) = \int_0^\infty F(t)dt$.

Proof of Theorem 9. Let $G(z)$ be the analytic continuation of $\mathcal{L}\{F\}(z)$ to the open set $U \supset \{z : \operatorname{Re} z = 0\}$, and define

$$G_\lambda(z) = \int_0^\lambda F(t)e^{-zt}dt\,.$$

By Theorem 6, $G_\lambda(z)$ is entire. Also, by Theorem 5 and 6, and the boundedness assumption on $F(t)$, $G(z)$ is holomorphic in the closed right half plane. These are very strong properties, and allow us to proceed in the proof with relative ease.

It suffices to prove that $G(0) - G_\lambda(0) \to 0$ as $\lambda \to \infty$. Because of the Tauberian condition, we can use Cauchy's integral formula to conveniently rephrase this problem in terms of a contour integral. Generally we prefer to work with such integrals, because we have a lot of nice standard tools from complex analysis to handle them. Hence we write

$$G(0) - G_\lambda(0) = \frac{1}{2\pi i}\int_C \left(\frac{G(z) - G_\lambda(z)}{z}\right) dz\,.$$

[9] Compare the definition of Abel summability in Definition 3. Siu in [3] gives a tabulated comparison to Tauber's original theorem, which is quite useful.

The theorem will follow from an appropriate choice of contour C, and partial estimation of the Cauchy integral.

Pick $\epsilon > 0$ arbitrarily. For $x = \operatorname{Re} z > 0$, we have

$$G(z) - G_\lambda(z) = \int_\lambda^\infty F(t)e^{-zt}dt\,, \tag{7}$$

and bound

$$|G(z) - G_\lambda(z)| \le \frac{e^{-\lambda x}}{x} = \frac{\left|e^{-\lambda z}\right|}{\operatorname{Re} z}\,. \tag{8}$$

To neutralize the problematic denominator $\operatorname{Re} z$, which might blow up on C, we note for $|z| = R$ that

$$\frac{1}{z} + \frac{z}{R^2} = \frac{2\operatorname{Re} z}{R^2}\,, \tag{9}$$

Now the crux of the proof. We replace the usual $1/z$ kernel of Cauchy's integral formula with $e^{\lambda z}(1/z + z/R^2)$. This new kernel is likewise meromorphic on $\mathbb{C}$, and has a simple pole at $z = 0$ of residue 1.[10] The additional terms are key in obtaining the desired estimates. The reason is that for our choice of contour C in Cauchy's formula, both $\operatorname{Re} z > 0$ and $\operatorname{Re} z < 0$ occur. In the latter case, (8) might be problematic.

So we must show that

$$\left|\int_C (G(z) - G_\lambda(z))e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right)dz\right| < \epsilon\,, \tag{10}$$

for some valid contour C.[11] Let $\delta_R > 0$ be a function of R, chosen so that $\{|z| \le R : \operatorname{Re} z \ge -\delta_R\} \subset U$. We cut $C_{R,\delta_R} := C$ up into three parts, and show (10) for each part individually. As depicted in Figure 2, the chosen segments are: the right half circle

$$C_R^+ = \{|z| = R : \operatorname{Re} z > 0\}\,,$$

the union of two (small) circle arcs in the left half plane

$$A_{R,\delta_R} = \{|z| = R : -\delta_R < \operatorname{Re} z < 0\}\,,$$

and the vertical line segment connecting those arcs

$$L_{R,\delta_R} = \{|z| < R : \operatorname{Re} z = -\delta_R\}\,.$$

[10] A quick review of the proof of Cauchy's integral formula shows that this replacement of the kernel is indeed allowed here. Note that $((G(z) - G_\lambda(z)) - (G(0) - G_\lambda(0)))/z$ is bounded because $G(z) - G_\lambda(z)$ is holomorphic, and that $z/R^2 = (z^2/R^2)/z$.

[11] Piecewise smooth is good enough. See [2], appendix B.

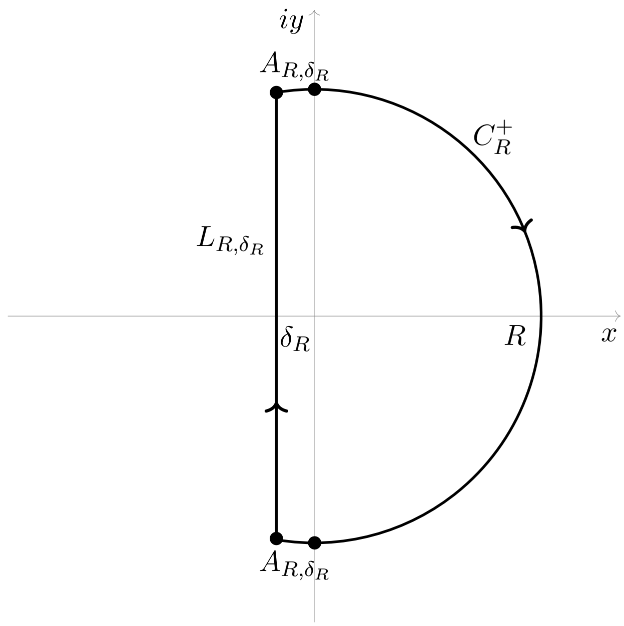

Figure 2: The integration contour for the Laplace transform Tauberian theorem. The contour is split into the three marked segments C_R^+, A_{R,δ_R}, and L_{R,δ_R}.

We start with C_R^+. From (8) and (9) it immediately follows that

$$\left|(G(z) - G_\lambda(z))e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right)\right| \leq \frac{\left|e^{-\lambda z}\right|}{\operatorname{Re} z}\left|e^{\lambda z}\right| \frac{2\operatorname{Re} z}{R^2} = \frac{2}{R^2}\,,$$

and therefore

$$\left|\int_{C_R^+} (G(z) - G_\lambda(z))e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right) dz\right| \leq \pi R \frac{2}{R^2} < \epsilon/3\,,$$

if R is chosen large enough. The somewhat mysterious choice for the non-standard Cauchy kernel $e^{\lambda z}(1/z + z/R^2)$ should begin to appear less opaque now.

In the left half plane we can no longer make use of (7), so for $A_{R,\delta_R} \cup L_{R,\delta_R}$ we treat $G(z)$ and $G_\lambda(z)$ separately. Their (assumed) analytical properties will provide the required bounds.

Since $G_\lambda(z)$ is entire, by the Cauchy–Goursat theorem, we may replace the contour $A_{R,\delta_R} \cup L_{R,\delta_R}$ with $\{|z| = R : \operatorname{Re} z < 0\}$. Similarly to the situation above, but now for $x = \operatorname{Re} z < 0$, we have the bound

$$|G_\lambda(z)| \leq \frac{e^{-\lambda x}}{-x} = \frac{\left|e^{-\lambda z}\right|}{-\operatorname{Re} z}\,,$$

and therefore

$$\left|\int_{\substack{\{|z|=R:\\ \operatorname{Re} z<0\}}} G_\lambda(z)e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right) dz\right| \leq \pi R \frac{2}{R^2} < \epsilon/3\,,$$

if R is chosen large enough. Notice that the change of contours is necessitated by the condition $|z| = R$ required to apply (7).

For $G(z)$ we treat A_{R,δ_R} and L_{R,δ_R} separately. Fix R large enough, then picking $\delta_R > 0$ small, we immediately establish the bound $< \epsilon/6$ for A_{R,δ_R}. Finally, remember that even if R and δ_R are fixed, we still have the freedom to choose λ as big as we wish, and so for the L_{R,δ_R} contour

$$\left|\int_{L_{R,\delta_R}} G(z)e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right) dz\right| \leq Ce^{-\lambda\delta_R} < \epsilon/6\,.$$

□

We now finish the proof of the Wiener-Ikehara theorem by proving Theorem 8 (repeated here for convenience).

Theorem. *Let $s(x)$, $1 \le x < \infty$, be a nonnegative, nondecreasing, piecewise continuous function, such that $s(x) \le Cx$, for some constant C. Define*

$$f(z) = z \int_1^\infty s(x) x^{-z-1} dx \,,$$

which is automatically holomorphic in half plane $\{z : \operatorname{Re} z > 1\}$, because $s(x) = \mathcal{O}(x)$. If $g(z) = f(z) - \frac{A}{z-1}$ can be analytically continued to an open neighborhood of the line $\operatorname{Re} z = 1$, then $s(x) \sim Ax$.

Proof. Let $F(t) = e^{-t}s(e^t) - A$. Under the theorem's conditions, F is bounded and piecewise continuous on $0 \le t < \infty$. We may therefore take its Laplace transform $G(z) = \mathcal{L}\{F\}(z)$. For G we have

$$\begin{aligned}
G(z) &= \int_0^\infty F(t) e^{-zt}\, dt \\
&= \int_0^\infty \left(e^{-t}s(e^t) - A\right) e^{-zt}\, dt \\
&= \int_1^\infty \left(\frac{1}{x}s(x) - A\right) x^{-z} \frac{1}{x}\, dx \\
&= \frac{1}{z+1}(z+1) \int_1^\infty s(x) x^{-(z+1)-1}\, dx - \frac{A}{z} \\
&= \frac{1}{z+1}\left(f(z+1) - \frac{A}{z} - A\right) \,,
\end{aligned}$$

and so by the assumptions on $g(z) = f(z) - \frac{A}{z-1}$, we can apply the Laplace transform Tauberian theorem (Theorem 9) to it. In fact, this is the whole point of our curious choice of $F(z)$.

Now $\int_0^\infty F(t)\, dt$ exists, and therefore

$$\int_1^\infty \left(\frac{s(x)}{x} - A\right) \frac{1}{x}\, dx = \int_0^\infty (e^{-t}s(e^t) - A)\, dt = \int_0^\infty F(t)\, dt \qquad (11)$$

also exists, and by definition is finite. This strongly suggests using the following proof technique.

Suppose we could show that for any $\epsilon > 0$, there exists a constant X such that whenever $x_0 \ge X$, we have $s(x_0)/x_0 - A \le \epsilon$ and $s(x_0)/x_0 - A \ge -\epsilon$. That would establish the theorem. Using this observation, we proceed with a proof by contradiction; so assume that for some $\epsilon > 0$, there exists a sequence $\{x_n\}$, $x_n \to \infty$, such that $s(x_n)/x_n - A > \epsilon$. We would like to contradict this with the finiteness of (11). The method for proving $s(x_0)/x_0 - A \ge -\epsilon$ when $x_0 \ge X$ is entirely analogous.

Recall that $s(x)$ was assumed to be nonnegative and nondecreasing. Hence, if $s(x_n)/x_n - A > \epsilon$, it should hold that $s(x_n)/x_n - A > \epsilon/2$ for some interval up ahead. We may pick this interval conveniently as

$$I_{x_n,\epsilon} = \left[x_n, \frac{A+\epsilon}{A+\frac{\epsilon}{2}}x_n\right] \subset [x_n, \infty)\,.$$

Indeed, for $x \in I_{x_n,\epsilon}$ we see that per assumption $s(x) \geq s(x_n) > x_n(A+\epsilon)$, and so

$$s(x)/x - A > \frac{x_n(A+\epsilon)}{x_n\left(\frac{A+\epsilon}{A+\epsilon/2}\right)} - A = \frac{\epsilon}{2}\,.$$

We then compute

$$\int_{I_{x_n,\epsilon}} \left(\frac{s(x)}{x} - A\right)\frac{1}{x}\,dx > \frac{\epsilon}{2}\int_{x_n}^{\frac{A+\epsilon}{A+\epsilon/2}x_n} \frac{1}{x}\,dx = \frac{\epsilon}{2}\log\left(\frac{A+\epsilon}{A+\frac{\epsilon}{2}}\right) > 0\,,$$

which is independent of x_n. But by the proof by contradiction assumption, there are infinitely many such x_n. Moreover, their respective intervals $I_{x_n,\epsilon}$ need not necessarily overlap, as we choose $x_n \to \infty$. So we have arrived at a statement contradicting the finiteness of (11). $\square$

The Riemann zeta function

We need two results on the Riemann zeta function $\zeta(z)$ (Definition 4) to use it in applying (Korevaar's version of) the Wiener-Ikehara Tauberian theorem to the proof of PNT. Additionally, we must show that $f(z) = \sum_n \Lambda(n)/n^z = -\zeta'(z)/\zeta(z)$ (Theorem 12), to establish the link between the zeta function and the Tauberian condition needed to derive $\psi(x) \sim x$, and so PNT. Here $\Lambda(n)$ is the von Mangoldt function (Definition 5); $\psi(x)$ is the second Tchebychef function (Definition 6).

The first result we need on the zeta function is its meromorphic continuation to the open right half plane (Theorem 10). The second required result builds on the first, and says that $\zeta(z)$ is nonzero on the line $\operatorname{Re} z = 1$ (Theorem 11). Through the above linking formula, these theorems together then quickly prove the necessary meromorphic continuation of $f(z)$ to an open neighborhood of the line $\operatorname{Re} z = 1$ (Theorem 12), and so with $\psi(x) = \mathcal{O}(x)$ prove $\psi(x) \sim x$. The proofs are taken mainly from [2] and [3], but are entirely standard.

The meromorphic continuation of the zeta function

Already from the discussion in the previous section, the importance of the zeta function for number theory is manifest. Euler (1707-1783) first used it to show that the sum $\sum 1/p$ diverges; presenting the first quantitative statement on the number of primes since Euclid. Later, in 1859, Riemann introduced the idea of applying complex analytic techniques, via Euler's zeta function, to the analysis of the prime counting function $\pi(x)$ (Definition 7); he also initiated the still ongoing investigation into the distribution of the zeros of the zeta function.[12] Knowledge of this distribution has proven pivotal to the application of the Riemann zeta function to number theory. In particular, to prove the prime number theorem, we shall need that $\zeta(z)$ is nonzero on the line $\operatorname{Re} z = 1$. In the previous section we observed the critical nature of this line for the Tauberian condition of the Wiener-Ikehara theorem.

Definition 4. The (Euler) zeta function *is initially defined on the half plane*

[12]The classical reference is Riemann's 1859 paper "Über die Anzahl der Primzahlen unter einer gegebenen Grösse". Here Riemann stated arguably the most famous open problem in of all of mathematics, the Riemann hypothesis; and introduced the now standard "zeta notation" $\zeta(z) := \sum 1/n^z$.

$\{z : \operatorname{Re} z > 1\}$ *as*

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}\,,$$

where it is automatically holomorphic.

Riemann's major contribution was in his proof of the existence of a meromorphic continuation of Euler's zeta function to the entire complex plane. Identifying a single simple pole at $z = 1$ of residue 1. This extended zeta function we call the Riemann zeta function.

To apply Theorem 7, we only need the meromorphic continuation of $-\zeta'(z)/\zeta(z)$ to some open set containing the line $\operatorname{Re} z = 1$; so we can get away with a bit less than Riemann's result.[13] Namely a continuation up to the line $\operatorname{Re} z = 0$, which easily follows from Abel's partial summation formula (see next theorem).

Remember however that no matter what methods we choose to pursue the meromorphic continuation of the zeta function, we cannot arrive at different definitions of $\zeta(z)$ in the domain so extended (the identity theorem). Hence the use of the definite article "the" in "the meromorphic continuation" is warranted. In particular, we shall look at two different extension approaches (with a third given in the appendixes).

Theorem 10. *The zeta function can be meromorphically continued to the open right half plane, with a single simple pole at $z = 1$ of residue 1. For the extended function we have the explicit formula*

$$\zeta(z) = \frac{1}{z-1} + 1 - z\int_1^{\infty} \{t\} t^{-z-1}\, dt\,,$$

where $\{t\} = t - \lfloor t \rfloor$ is the fractional part function.

Proof. By Abel's partial summation formula (6), taking $a_n = 1$ and $a(x) = 1/x^z$, for $\operatorname{Re} z > 1$ we may write

$$\zeta(z) = z\int_1^{\infty} \lfloor t \rfloor\, t^{-z-1}\, dt\,.$$

Now notice that the problematic extra order of growth in the integrand, contributed by the integral part function $\lfloor t \rfloor$, can be canceled by rewriting

$$z\int_1^{\infty} \lfloor t \rfloor\, t^{-z-1}\, dt = z\int_1^{\infty} (\lfloor t \rfloor - t) t^{-z-1}\, dt + z\int_1^{\infty} t^{-z}\, dt\,,$$

[13] Recall that complex differentiable functions are automatically infinitely complex differentiable. See [2], chapter 2 for reference.

so that

$$\zeta(z) = \frac{1}{z-1} + 1 - z\int_1^\infty \{t\}t^{-z-1}\,dt\,.$$

$\square$

This method can be extended to yield meromorphic continuations to the sets $\{z : \operatorname{Re} z > -m\}$, for any $m \in \mathbb{N}$; and hence to the entire complex plane. Define $Q_0(x) = \{x\} - 1/2$, then

$$\zeta(z) = \frac{z}{z-1} - \frac{1}{2} - z\int_1^\infty \frac{Q_0(x)}{x^{z+1}}\,dx\,.$$

We continue to recursively define $Q_k(x)$ by imposing the three properties

1. $\frac{d}{dx}Q_{k+1} = Q_k$
2. $Q_k(x+1) = Q_k(x)$
3. $\int_0^1 Q_k(x)\,dx = 0\,.$

These polynomials are related to the *Bernoulli polynomials* on $0 \le x \le 1$ by the equation

$$Q_k(x) = \frac{B_{k+1}(x)}{(k+1)!}\,.$$

In turn, *Bernoulli numbers* B_k arise as special values of these polynomials, namely $B_k = B_k(0)$. The first few Bernoulli numbers are $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, $B_4 = -1/30$.

With property 1, rewrite

$$\zeta(z) = \frac{z}{z-1} - \frac{1}{2} - z\int_1^\infty \left(\frac{d^k}{dx^k}Q_k(x)\right)\frac{1}{x^{z+1}}\,dx\,.$$

Integration by parts then extends the meromorphic continuation of $\zeta(z)$ into $\{z : \operatorname{Re} z > -k-1\}$.

Specifically, we have

$$\zeta(z) = \frac{z}{z-1} - \frac{1}{2} - z\frac{Q_1(x)}{x^{z+1}}\bigg|_1^\infty - z(z+1)\int_1^\infty \frac{Q_1(x)}{x^{z+2}}\,dx$$
$$\zeta(z) = \frac{z}{z-1} - \frac{1}{2} + z\frac{B_2}{2} - z(z+1)E(z)\,.$$

Now using the above properties 2 and 3, we can apply Dirichlet's test to the integral $E(z)$. Therefore, by Theorem 5 and 6, we see that $E(z)$ is holomorphic in the half plane $\operatorname{Re} z > -2$. Repeated application of these steps yields the general extension.

We are now also able to sum $\sum n = -1/12$. If we symbolically extend Euler's zeta function, and associate it with Riemann's analytic continuation, we can imagine that $\zeta(-1) = \sum \frac{1}{n^{-1}}$. In this spirit, zeta function regularization acts as a summability method. In particular, we have

Corollary 1. *Zeta function regularization assigns the sums*

$$\sum_{n=1}^{\infty} n^k = \zeta(-k) = -\frac{B_{k+1}}{k+1}\,,$$

for nonnegative integers k. Specifically, $\sum 1 = -1/2$ and $\sum n = -1/12$.

Another approach to the meromorphic continuation of $\zeta(z)$ is to start with the gamma function $\Gamma(z)$, initially defined for $s > 0$ as

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1}\, dt\,.$$

It can be shown that $1/\Gamma(z)$ admits an analytic continuation to the entire complex plane[14].

By Fubini-Tonelli theorem, since $1/(e^x - 1) = \sum e^{-nx}$, we may swap the integral and sum, and write

$$\zeta(z) = \frac{1}{\Gamma(z)} \int_0^{\infty} \frac{x^{z-1}}{e^x - 1}\, dx\,,$$

for $\operatorname{Re} z > 1$. Splitting the integral, we therefore obtain

$$\zeta(z) = \frac{1}{\Gamma(z)} \int_0^{1} \frac{x^{z-1}}{e^x - 1}\, dx + E(z)\,,$$

with $E(z)$ entire.

Recall the generating function of the Bernoulli numbers (commonly taken as definition) as

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n\,.$$

Using Fubini-Tonelli again then yields

$$\int_0^{\infty} \frac{x^{z-1}}{e^x - 1}\, dx = \sum_{n=0}^{\infty} \frac{B_n}{n!(z+n-1)}\,.$$

[14]See for example [2], chapter 6, pp. 160-168; and in particular, theorem 1.6, p. 165.

The right hand side is entire, except for poles at $z = 1, 0, -1, \cdots$. However, leaving $z = 1$, these are all canceled by the zeros of $1/\Gamma(z)$. So by the identity theorem, we are done.

The zeta function is nonzero on the line $\mathrm{Re}\, z = 1$

The following truly inspired proof is due to F. Mertens (1840–1927). It tremendously simplifies a considerable hurdle in the early proofs of the prime number theorem. The details are taken from [2], chapter 7.

Theorem 11. *The Riemann zeta function does not vanish on the line* $\mathrm{Re}\, z = 1$.

Proof. At the heart of Mertens's proof is the trigonometric identity

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

and the auxiliary function $h(x)$, defined for $x > 1$ as

$$h(x) = \zeta(x)^3 \zeta(x+iy)^4 \zeta(x+2iy)\,.$$

Recall Euler's product formula for his zeta function

$$\zeta(z) = \prod_{p \text{ prime}} \frac{1}{1-p^{-z}}\,,$$

which is valid in the plane $U = \{z : \mathrm{Re}\, z > 1\}$. An important corollary is that $\zeta(z)$ has no zeros in U, and so $\log\zeta(z)$ is holomorphic there (see [2], chapter 3, Theorems 5.2 and 6.2; chapter 5, proposition 3.1; chapter 7, pp. 182-184).

We can conveniently use product formulas to absorb logarithms. For the zeta function

$$\log\zeta(z) = \log \prod_{p \text{ prime}} \frac{1}{1-p^{-z}}$$
$$= \sum_{p \text{ prime}} \log\left(\frac{1}{1-p^{-z}}\right)$$
$$= \sum_{p \text{ prime}} \sum_{n=1}^{\infty} \frac{p^{-nz}}{n} \tag{12}$$
$$= \sum_{p,m} \frac{p^{-mz}}{m}$$
$$= \sum_{n=1}^{\infty} \frac{c_n}{n^z},$$

where $c_n = 1/m \geq 0$ if $n = p^m$, and zero otherwise. Recall here the analytic continuation of the power series expansion of the real logarithm

$$\log\left(\frac{1}{1-x}\right) = \sum_{n=1}^{\infty} \frac{x^n}{n}.$$

Also, in the last steps we may indeed safely ignore the order of summation (up to a bijection), because the double series converges absolutely (see [2], chapter 7, pp. 197-199).

Set $z = x+iy$; verify that $\log|z| = \operatorname{Re}\log z$, and $\operatorname{Re}(n^{-z}) = n^{-x}\cos(y\log n)$. As a result of the trigonometric identity above, we can then bound the logarithm of Mertens's auxiliary function from below by

$$\begin{aligned}\log|h(x)| &= 3\log|\zeta(x)| + 4\log|\zeta(x+iy)| + \log|\zeta(x+2iy)| \\ &= \sum_{n=1}^{\infty} c_n n^{-x}(3 + 4\cos\theta_n + \cos 2\theta_n) \\ &\geq 0,\end{aligned}$$

where $\theta_n = y\log n$. Hence $|h(x)| \geq 1$. Using this bound we give a proof by contradiction.

Suppose contrary to the theorem that $\zeta(z_0) = 0$ at some point $z_0 = 1 + iy_0 \neq 1$ on the line $\operatorname{Re} z = 1$. We know that the extended zeta function is per definition holomorphic at this z_0; therefore if we approach along the horizontal line $x + iy_0$, we have the bound

$$|\zeta(x+iy_0) - \zeta(z_0)| = |\zeta'(z_0)||x-1| + o(x-1),$$

and hence for $C > 0$ constant

$$\begin{aligned} & |\zeta(x+iy_0)| \leq C|x-1| \\ \Rightarrow \qquad & |\zeta(x+iy_0)|^4 \leq C|x-1|^4 \,. \end{aligned}$$

Similarly, at the pole we find

$$|\zeta(x)|^3 \leq C'|x-1|^{-3} \,.$$

We now observe that in bounding $h(x)$, the small terms $|x-1|$ overpower the large terms $1/|x-1|$. So we have

$$|h(x)| = |\zeta(x)|^3 \, |\zeta(x+iy)|^4 \, |\zeta(x+2iy)| \leq C''(x-1) \,.$$

Taking then $x \to 1^+$ (recall that $h(x)$ is only defined for $x > 1$), we arrive at the sought after contradiction. □

Combining Theorem 10 and 11, it is straight forward to prove the result we need to apply the Wiener-Ikehara theorem to the proof of PNT via the zeta function (Theorem 12). But before we continue, we require one more technical lemma. Also, we now explicitly need the two auxiliary functions $\Lambda(n)$ and $\psi(x)$ we mentioned several times earlier. First the lemma.[15]

Lemma 1. *Suppose $\{F_n\}$ is a sequence of functions, holomorphic on the open set $\Omega \subset \mathbb{C}$. If there exist constants $c_n > 0$ such that $\sum c_n < \infty$, and $|F_n(z) - 1| < c_n$ for all $z \in \Omega$, then*

1. *The product $\prod F_n(z)$ converges uniformly in Ω to a holomorphic function $F(z)$.*
2. *If $F_n(z)$ does not vanish for any n,*

$$\frac{F'(z)}{F(z)} = \sum_{n=1}^{\infty} \frac{F_n'(z)}{F_n(z)} \,.$$

The following two number theoretic auxiliary functions pop up frequently when analyzing the prime counting function $\pi(x)$ (Definition 7). Studying the relations between such arithmetical functions is key in number theory. The associated Greek letters are standard notation (similarly to the use of zeta for Riemann's function).

[15]For a proof refer to [2], chapter 5, proposition 3.2 (pp. 141-142).

Definition 5. The von Mangoldt function $\Lambda(n)$ *is defined on the natural numbers as*

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \text{ for prime } p \\ 0 & \text{otherwise} \end{cases} .$$

Definition 6. The (second) Tchebychef function $\psi(x)$ *is the summation of the von Mangoldt function up to x (as in Abel's formula). That is,*

$$\psi(x) = \sum_{1 \leq n \leq \lfloor x \rfloor} \Lambda(n) .$$

We now return to the discussion about the role of the Wiener-Ikehara Tauberian theorem in proving the prime number theorem (refer to directly below Theorem 7). Recall that we there set $f(z) = \sum_n \frac{\Lambda(n)}{n^z}$. We establish the link between the zeta function and the Tauberian condition needed to derive $\psi(x) \sim x$ (and so PNT) through the formula $f(z) = -\zeta'(z)/\zeta(z)$; and subsequently can prove the meromorphic continuation of $f(z)$ to an open neighborhood of the line $\operatorname{Re} z = 1$, as required by the condition. This should clarify the interplay between the Tchebychef ψ-function, the Riemann zeta function, the Wiener-Ikehara Tauberian theorem, and the prime number theorem.

In the next section we then finally show $\psi(x) = \mathcal{O}(x)$, and prove the equivalence of PNT and $\psi(x) \sim x$. Thereby finishing the argument.

Theorem 12. *Define*

$$f(z) = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^z} ,$$

if $\operatorname{Re} z > 1$*, then*

$$f(z) = -\frac{\zeta'(z)}{\zeta(z)} . \tag{13}$$

Furthermore, $f(z)$ admits a meromorphic continuation to an open neighborhood of the line $\operatorname{Re} z = 1$*. Moreover, like $\zeta(z)$, $f(z)$ has as only singularity a simple pole at $z = 1$ of residue* 1*.*

Proof. Applying the above Lemma 1 to the Euler product formula, we obtain

$$-\frac{\zeta'(z)}{\zeta(z)} = \sum_{p \text{ prime}} \frac{p^{-z} \log p}{1 - p^{-z}} = \sum_{p \text{ prime}} \sum_{n=1}^{\infty} \frac{\log p}{(p^n)^z} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^z} .$$

Here the series rearrangement in the last step is allowed, because the double sum converges absolutely (compare the derivation of (12) in Theorem 11).

As a result of this formula, the meromorphic continuation of $f(z)$ follows from extending the logarithmic derivative $-\zeta'(z)/\zeta(z)$. The problem is however, that $1/\zeta(z)$ might not be defined. So we need to use our knowledge of the distribution of the zeros of the Riemann zeta function.

We observed earlier (in proof of Theorem 11) that the Euler product formula implies that $\zeta(z)$ is nonzero on the half plane $\{z : \operatorname{Re} z > 1\}$. The demonstrated properties of the Riemann zeta function (Theorem 10 and 11) extend this non-vanishing result to an open set U containing the critical line $\operatorname{Re} z = 1$. On U the reciprocal $1/\zeta(z)$ exists, and is holomorphic. Consequently, $f(z)$ admits a meromorphic continuation to U, and therefore to a neighborhood of the line $\operatorname{Re} z = 1$.

Finally, from (13) and the product rule for differentiation, we immediately see that $f(z)$, like $\zeta(z)$, has as only singularity a simple pole of residue 1 at $z = 1$. $\square$

The prime number theorem

We define yet another auxiliary arithmetical function.

Definition 7. The prime counting function $\pi(x)$ *denotes the number of primes up to and including* x.

The prime number theorem (PNT) now reads

Theorem 13 (Hadamard, de La Vallée Poussin, 1896)**.** *As* $x \to \infty$*, primes distribute like*

$$\pi(x) \sim \frac{x}{\log x}\,.$$

We are left to show the linear bound $\psi(x) \leq Cx$, for some constant C (Theorem 14), and relate Tchebychef's ψ-function to PNT (Theorem 15). The proofs are entirely standard.

The linear bound on the Tchebychef ψ-function

Theorem 14. *The second Tchebychef function* $\psi(x)$ *is linearly bounded. That is,* $\psi(x) = \mathcal{O}(x)$*.*

Proof. We can relate the product of primes in the interval $(n, 2n]$ to the binomial coefficient $\binom{2n}{n}$ through the inequality

$$\prod_{\substack{n<p\leq 2n \\ p \text{ prime}}} p \leq \frac{(n+1)(n+2)\cdots(2n)}{1\cdot 2\cdots n} = \binom{2n}{n}\,.$$

Binomial coefficient are easier to manipulate and estimate than products of primes. In particular, we can immediately see that $\binom{2n}{n} < 2^{2n}$. Therefore by taking logarithms, we establish the crude bound

$$\sum_{\substack{n<p\leq 2n \\ p \text{ prime}}} \log p < 2n \log 2\,.$$

Now assume $n = 2^m$. We can then divide $(1, n]$ into subintervals $(2^l, 2^{l+1}]$, for $l \in [0, m-1]$, to obtain

$$\sum_{\substack{p\leq 2^m \\ p \text{ prime}}} \log p < 2^{m+1} \log 2\,. \tag{14}$$

Recall the definition of the Tchebychef's ψ-function (Definition 6). We see immediately that we can rewrite it to

$$\psi(x) = \sum_{p \text{ prime}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p\,.$$

The key idea of the proof is to differentiate relative to x between "small" and "big" primes, and accordingly split the above formula. We can then linearly bound both parts separately.

We say that p is a small prime relative to x if $p^2 \leq x$. Otherwise we say it is a big prime (relative to x). These cases correspond to $\lfloor \log x / \log p \rfloor > 1$ and $\lfloor \log x / \log p \rfloor = 1$ respectively. For the small primes $p \leq \sqrt{x}$, and therefore

$$\begin{aligned}
\sum_{p \text{ small}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p &\leq \sum_{p \leq \sqrt{x}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p \\
&\leq \sum_{p \leq \sqrt{x}} \log x \\
&= \pi(\sqrt{x}) \log x\,.
\end{aligned}$$

Where we take to our advantage, the ability to introduce the square root in the term $\pi(\sqrt{x})$. For the big primes, take m such that $2^m \leq x \leq 2^{m+1}$; then by equation (14) we may bound

$$\begin{aligned}
\sum_{p \text{ big}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p &= \sum_{p \text{ big}} \log p \\
&\leq 2^{m+2} \log 2 \leq 4x \log 2\,.
\end{aligned}$$

Here the advantage is the "cheap" disposal of the term $\lfloor \log x / \log p \rfloor$.

Now notice that $\pi(x) \leq x$. The two bounded pieces of $\psi(x)$ taken together then yield the result

$$\begin{aligned}
\psi(x) &\leq \pi(\sqrt{x}) \log x + 4x \log 2 \\
&\leq \sqrt{x} \log x + 4x \log 2 \\
&= \left(\frac{\log x}{\sqrt{x}} + 4 \log 2 \right) x \\
&= \mathcal{O}(x)\,.
\end{aligned}$$

$\square$

Equivalent formulations of PNT

By Theorem 12 and 14, we can apply the Wiener-Ikehara theorem with $A = 1$ to $a_n = \Lambda(n)$, and conclude that $\psi(x) \sim x$. Hence the prime number theorem is a corollary of

Theorem 15. *If $\psi(x) \sim x$, then $\pi(x) \sim x/\log x$, for $x \to \infty$.*

Proof. Under the assumption $\psi(x) \sim x$, it suffices to prove $\psi(x) \sim \pi(x)\log x$. We do so by the standard sandwiching technique of bounding with the same number, the limsup from above, and the liminf from below (as in the proof of Littlewood's 1911 Theorem 4). We see the limsup bound at once, but the liminf bound is more difficult. We again need to split the problem.

Note that $1 \leq \liminf \psi(x)/(\pi(x)\log x)$ iff $\limsup\,(\pi(x)\log x)/\psi(x) \leq 1$, so that we may also prove the latter inequality. Choose $y_x < x$, and separately consider the primes $p \leq y_x$, and the primes $y_x < p \leq x$, to obtain

$$\begin{aligned}\pi(x) &= \sum_{p \leq y_x} 1 + \sum_{y_x < p \leq x} 1 \\ &= \pi(y_x) + \frac{1}{\log y_x} \sum_{y_x < p \leq x} \log p \\ &\leq y_x + \frac{\psi(x)}{\log y_x}\,.\end{aligned}$$

Rewriting yields

$$\frac{\pi(x)\log x}{\psi(x)} \leq \frac{y_x \log x}{\psi(x)} + \frac{\log x}{\log y_x}\,.$$

So since $\psi(x) \sim x$, we can conveniently swap $\psi(x)$ for x, and it suffices to pick y_x such that

$$\limsup_{x\to\infty} \frac{\log x}{\log y_x} \leq 1\,,$$

and

$$\limsup_{x\to\infty} \frac{y_x \log x}{x} \leq 0\,.$$

A good choice is

$$y_x = \frac{x}{(\log x)^2}\,.$$

□

One reason to be interested in PNT is its use in understanding the structure of the prime number sequence p_n. Indeed, p_n appears to behave quite randomly; and it is not even easy to provide an approximating sequence. Using the prime number theorem however, we can establish such an (asymptotic) estimate. In fact, the estimate and PNT are equivalent.[16]

Theorem 16. *The prime number theorem is equivalent to the asymptotic estimate*

$$p_n \sim n \log n \qquad \text{as } n \to \infty \,.$$

Proof. ($\Rightarrow$) First note that $\pi(p_n) = n$; then in the PNT estimate set $x = p_n$ to get $p_n \sim n \log p_n$. Thus we need to show $\log p_n \sim \log n$. Take logarithms and use continuity to obtain

$$\frac{\log n}{\log p_n} + \frac{\log \log p_n}{\log p_n} - 1 \to 0 \,.$$

The middle term vanishes, hence by definition $p_n \sim n \log n$.

($\Leftarrow$) Let $p_n \leq x < p_{n+1}$, then

$$\frac{p_n}{n \log n} \leq \frac{x}{\pi(x) \log \pi(x)} < \frac{p_{n+1}}{n \log n} \,,$$

and therefore $x \sim \pi(x) \log \pi(x)$. Analogous to how we established $\log p_n \sim \log n$, we can now prove $\log \pi(x) \sim \log x$. Hence $\pi(x) \sim x / \log x$. $\square$

As a final observation we note that PNT can be understood "probabilistically", despite that being prime is not a probabilistic concept. The asymptotic estimate suggests using the heuristic $\text{Prob}(n \text{ prime}) = 1/\log n$; the expected number of primes under x is then given by

$$\sum_{2 \leq n \leq x} \mathcal{P}(n \text{ prime}) = \sum_{2 \leq n \leq x} \frac{1}{\log n} \,.$$

This approximation is numerically far superior to $x/\log x$ (see also [5]). In fact, for the closely related (offset) logarithmic integral function

$$\text{Li}(x) = \int_2^x \frac{1}{\log t} \, dt \,,$$

we have the following precise theorem due to H. von Koch (1870-1924).

[16] In the sense that there are quick proofs of PNT $\Rightarrow$ estimate, and estimate $\Rightarrow$ PNT. Similarly we can say that $\psi(x) \sim x$ and PNT are equivalent. Theorem 15 proves one side of this statement.

Theorem 17 (von Koch, 1901)**.** *The Riemann zeta function has no zeros in the strip* $\alpha < \operatorname{Re} z < 1$ *if and only if*

$$\pi(x) - \mathrm{Li}(x) = \mathcal{O}(x^{\alpha+\epsilon}) \qquad \text{as } x \to \infty$$

for every $\epsilon > 0$ *(and fixed* $1/2 \leq \alpha < 1$*).*

The theorem links the distribution of the zeros of $\zeta(z)$, through the Riemann hypothesis, to the estimation of the prime counting function. This illustrates once again the importance of the locations of the zeta function zeros for number theory.[17]

It should not surprise the reader that

Theorem 18. *The (offset) logarithmic integral function has the asymptotic estimate*

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

So PNT is equivalent to $\pi(x) \sim \mathrm{Li}(x)$*.*

Proof. Partial integration yields

$$\int_2^x \frac{1}{\log t} \, dt = \frac{x}{\log x} - \frac{2}{\log 2} + \int_2^x \frac{1}{(\log t)^2} \, dt \, .$$

So we are done if we can prove that the last integral is of order $o(x/\log x)$. A clever split of the integration domain suffices:

$$\begin{aligned} \int_2^x \frac{1}{(\log t)^2} \, dt &= \left(\int_2^{\sqrt{x}} + \int_{\sqrt{x}}^x \right) \frac{1}{(\log t)^2} \, dt \\ &\leq C\sqrt{x} + C' \frac{x - \sqrt{x}}{(\log \sqrt{x})^2} \\ &\leq \tilde{C} \frac{x}{(\log x)^2} \, . \end{aligned}$$

□

[17]The Riemann hypothesis proposes that $\zeta(z) = 0 \Rightarrow \operatorname{Re} z = 1/2$. I am not sure how much credit von Koch deserves for the theorem. Whether he simply first stated it, proved a weaker version, or showed just one implication, I don't know.

Appendix A: the prime polynomial theorem

It is instructive to see how the zeta function, along with the various familiar auxiliary functions, come up in the proof of the polynomial version of PNT.[18] Interestingly, the bounds provided by this prime polynomial theorem (PPT) satisfy a von Koch type condition (refer to Theorem 17), or a "Riemann hypothesis" for polynomials.

We first revisit the Riemann zeta function and the auxiliary arithmetical functions of the PNT proof, and define polynomial variants. We use the standard notation $\mathbb{F}_q$ to denote a finite field of order q. Similarly we use $\mathbb{F}_q[x]$ to mean the ring of polynomials over $\mathbb{F}_q$.

Take the degree of a polynomial as norm. The ring $\mathbb{F}_q[x]$ is then Euclidean in the sense that Euclid's algorithm is valid. Consequently, we can establish an analogue of Bézout's lemma, and therefore it makes sense to talk about irreducibles (polynomials without nontrivial factors) as primes. Being prime is understood here in the sense of Euclid's lemma. That is, P is prime if and only if $P|AB \Rightarrow P|A$ or $P|B$. From now on we treat the terms "irreducible" and "prime" as synonyms.

In fact, like the natural numbers, $\mathbb{F}_q[x]$ is a unique factorization domain. Meaning that a version of the fundamental theorem of arithmetic holds. In particular, every polynomial in $\mathbb{F}_q[x]$ has a unique representation as a product of irreducibles, up to order and multiplication by scalars.

Definition 8. The polynomial prime counting function $\pi_q(n)$ *gives the number of monic irreducible polynomials in* $\mathbb{F}_q[x]$ *of degree* n*.*

Definition 9. The polynomial von Mangoldt function *is given by*

$$\Lambda(f) = \begin{cases} \deg P & \text{if } f = P^k \text{ a power of a prime } P \\ 0 & \text{otherwise} \end{cases} .$$

Definition 10. The polynomial Tchebychef ψ-function *is defined using the von Mangoldt function as the finite sum*

$$\psi(n) = \sum_{\substack{\deg f = n \\ f \text{ monic}}} \Lambda(f) \, . \tag{15}$$

[18] More proofs of polynomial equivalents of (famous) number theoretic theorems can be found in the bachelor thesis [8]. In particular, the thesis discusses several other (elementary) proofs of the polynomial version of PNT, as well as equivalents for Dirichlet's theorem and the notion of Tchebychef bias.

Definition 11. *The polynomial zeta function $\zeta_q(z)$ is initially defined on $\{z : \operatorname{Re} z > 1\}$ as*

$$\zeta_q(z) = \sum_{\substack{0 \neq f \in \mathbb{F}_q[x] \\ f \text{ monic}}} \frac{1}{|f|^z} \,. \tag{16}$$

Here $|f| = q^{\deg f}$. That is, the norm of f is the number of distinct polynomials in $\mathbb{F}_q[x]$ of degree less than $\deg f$.

Note that in the definition of the zeta function $\zeta_q(s)$, we rely on the absolute convergence of the series to be able to ignore the precise order of summation (recall Riemann's rearrangement theorem). We will continue to associate $|f| = q^{\deg f}$.

We formulate PPT in probabilistic terms, as we did earlier for PNT. This has as advantage that it makes the statement appear more intriguing, and therefore easier to remember.

Theorem 19 (PPT)**.** *A random polynomial in $\mathbb{F}_q[x]$ of degree n has the asymptotic probability $1/n$ to be irreducible. More precisely*

$$\pi_q(n) = \frac{q^n}{n} + \mathcal{O}\left(\frac{\sqrt{q}^n}{n}\right) \,. \tag{17}$$

Writing $x = q^n$, this resembles the prime number theorem:

$$\pi_q(n) \sim x / \log_q(x) \,.$$

We can also formulate an Euler product for $\zeta_q(z)$. Its proof is analogous to proof of the usual Euler product, except for some technicalities. It is likewise a codification of the fundamental theorem of arithmetic (for $\mathbb{F}_q[x]$).

Note that for monic polynomials P and Q, conveniently $|PQ| = |P||Q|$.

Theorem 20 (Euler product)**.** *For $\operatorname{Re} z > 1$, we have*

$$\zeta_q(z) = \prod_{\substack{P \text{ prime} \\ \text{and monic}}} \frac{1}{1 - |P|^{-z}} \,,$$

where the product is taken from small to large polynomial degree.

A nice corollary is the polynomial variant of Euler's classical zeta function application.

Corollary 2. *There are infinitely many monic irreducible polynomials. In fact, we have the stronger divergence result*

$$\sum_{\substack{P \text{ irreducible} \\ \text{and monic}}} \frac{1}{|P|} \to \infty \,. \tag{18}$$

Very different from the regular zeta function however, is the simplicity of the meromorphic continuation of $\zeta_q(z)$.

Theorem 21. *The zeta function has a meromorphic continuation to the entire complex plane given by*

$$\zeta_q(s) = \frac{1}{1 - q^{1-z}} \,.$$

Proof. Assume $\operatorname{Re} z > 1$. A simple series rearrangement gives

$$\begin{aligned}
\sum_{\substack{0 \neq f \in \mathbb{F}_q[x] \\ f \text{ monic}}} \frac{1}{|f|^z} &= \sum_{n=0}^{\infty} \left(\sum_{\substack{\deg f = n \\ f \text{ monic}}} \frac{1}{|f|^z} \right) \\
&= \sum_{n=0}^{\infty} \frac{1}{q^{nz}} q^n \\
&= \sum_{n=0}^{\infty} (q^{1-z})^n \,.
\end{aligned}$$

By working backwards, this also immediately provides the required absolute convergence. Note here the convenient choice of $|\cdot|$.

The geometric series formula yields the stated meromorphic continuation, with poles at $s = 1 + i\frac{2\pi}{\log q} n$, for $n \in \mathbb{N}$. $\square$

We now prove an *explicit* formula for the (polynomial) Tchebychef ψ-function, reminiscent of the *asymptotic* formula $\psi(x) \sim x$.[19] Using some

[19]Explicit formulas also exist for the usual Tchebychef ψ-function (and for other arithmetical functions), but they are more involved, and harder to prove. Such number theoretic formulas were initially investigated by Riemann in relation to the distribution of the primes, beginning with his seminal 1859 paper. In fact, the explicit formula

$$\int_1^x \psi(x)\, dx = \frac{x^2}{2} - \sum_\rho \frac{x^\rho}{\rho(\rho+1)} + E(x) \,,$$

can be developed into a PNT proof, where it plays a role similar to the Wiener-Ikehara theorem in our proof. Here the sum is taken over all zeros ρ of the Riemann zeta function in the critical strip ($0 \leq \operatorname{Re} z \leq 1$); and $E(x)$ is an error term in $\mathcal{O}(x)$. This clarifies the relation of PNT to Mertens's theorem (see Theorem 11).

simple bounding, we can then directly deduce PPT from it. As a first step, we clarify the connection between the ψ-function and prime counting function $\pi_q(x)$ in the next lemma.

Lemma 2. *We have the identity*

$$\psi(n) = \sum_{d|n} d\pi_q(d)\,.$$

Proof. Consider the definition of the Tchebychef ψ-function. If the n-th degree polynomial f is a power of a prime P of degree k, as in the definition of the von Mangoldt function, then obviously $k|n$; and since f is monic, P must be monic as well. This monic prime is counted uniquely by $\pi_q(k)$, and is correctly weighted (in the above identity) with $k = \deg P$.

Conversely, any monic irreducible P of degree $d|n$, produces a unique monic polynomial of degree n when raised to the integral power n/d; and again, the weights $\deg P = d$ match. Hence the two series are rearrangements of each other. So they must be equal, because they contain only finitely many terms.

$\square$

For the proof of the explicit formula, and subsequent PPT proof, we follow [6].

Theorem 22. *The Tchebychef ψ-function has the explicit formula*

$$\psi(n) = q^n\,.$$

Proof. Define the function $Z(u)$ in the punctured neighborhood $D = \{u \in \mathbb{C} : 0 < |u| < q^{-1}\}$ as

$$Z(u) = \sum_{\substack{0 \neq f \in \mathbb{F}_q[x] \\ f \text{ monic}}} u^{\deg f}\,, \tag{19}$$

and set s such that $u = q^{-s}$. Because $u^{\deg f} = \left(q^{\deg f}\right)^s = 1/|f|^s$, we see that $Z(u)$ converges absolutely, and therefore is well defined; and from Theorem 20 we have the Euler product

$$Z(u) = \prod_{\substack{P \text{ prime} \\ \text{and monic}}} \frac{1}{1 - u^{\deg P}}\,.$$

Furthermore, by Theorem 21, we have the meromorphic continuation

$$Z(u) = \frac{1}{1 - qu}\,.$$

The idea of the proof is to compare the different forms of the logarithmic derivative $Z'(u)/Z(u)$ obtained from the two formulas, by manipulating them into power series. The coefficients of these power series are uniquely determined; and their equality immediately yields the explicit formula.

For the first formula we use Lemma 1 and (double) series rearrangement to get

$$\begin{aligned} u\frac{Z'(u)}{Z(u)} &= \sum_{\substack{P \text{ prime} \\ \text{and monic}}} \frac{\deg(P)u^{\deg P}}{1-u^{\deg P}} \\ &= \sum_{\substack{P \text{ prime} \\ \text{and monic}}} \deg(P)\sum_{k=1}^{\infty} u^{k\deg P} \\ &= \sum_{f \text{ monic}} \Lambda(f)u^{\deg f} \\ &= \sum_{n=1}^{\infty} \psi(n)u^n \end{aligned}$$

The second $Z(u)$ formula gives

$$u\frac{Z'(u)}{Z(u)} = u\frac{q}{1-qu} = \sum_{n=1}^{\infty} q^n u^n ,$$

and therefore $\psi(n) = q^n$. □

We now have all the ingredients required to quickly derive PPT.

Proof. Fix n. By the explicit formula and Lemma 2, we can establish for any $m \in \mathbb{N}$ the crude bound

$$m\pi_q(m) \le \sum_{d|m} d\pi_q(d) = q^m .$$

From this we obtain the squeeze

$$0 \le \psi(n) - n\pi_q(n) = \sum_{\substack{d|n \\ d<n}} d\pi_q(d) \le \sum_{\substack{d|n \\ d<n}} q^d .$$

Now, in order to provide a more workable bound, we use another crude observation, namely that there are no more than $\lfloor n/2 \rfloor$ proper divisors $d < n$

of n; and that these divisors cannot be larger than $\lfloor n/2 \rfloor$. As a result we obtain from the geometric series formula

$$\sum_{\substack{d|n \\ d<n}} q^d \leq \sum_{d=1}^{\lfloor n/2 \rfloor} q^d = \frac{q^{\lfloor n/2 \rfloor + 1} - q}{q-1} \leq \frac{q^{\lfloor n/2 \rfloor}}{1 - 1/q} \leq 2q^{n/2}\,,$$

and therefore

$$0 \leq \psi(n) - n\pi_q(n) \leq 2q^{n/2}\,.$$

Using the explicit formula again we get

$$\frac{q^n}{n} - 2\frac{q^{n/2}}{n} \leq \pi_q(n) \leq \frac{q^n}{n}\,.$$

Hence

$$\pi_q(n) = \frac{q^n}{n} + \mathcal{O}\left(\frac{q^{n/2}}{n}\right)\,.$$

Finally, note that we have established a slightly stronger version of the theorem we set out to prove. Specifically, we have $\pi_q(n) \leq q^n/n$. $\square$

References

[1] J. Korevaar. *Een eenvoudig bewijs van de priemgetalstelling*. Nieuw Archief voor Wiskunde, pp. 284-291, 5/5 nr. 4, december 2004.

[2] E. M. Stein and R. Shakarchi. *Complex Analysis*. Princeton University Press, 2003.

[3] Y. Siu. *Tauberian Theorems*.
https://web.archive.org/web/20161020160706/http://www.math.harvard.edu/~siu/math212b/tauberian_theorem.pdf.

[4] S. Abbott. *Understanding Analysis*. Springer, 2001.

[5] K. Conrad. *Primality statistics*.
http://www.math.uconn.edu/~kconrad/ross2016/primestats.pdf

[6] Z. Rudnick. *Notes on the prime polynomial theorem — course notes, 2015*.
http://www.math.tau.ac.il/~rudnick/courses/sieves2015/PPT.pdf

[7] J. Karamata. *Über die Hardy-Littlewoodschen Umkehrungen des Abelschen Stetigkeitssatzes*. Math. Z., pp. 319–320, no. 1, 1930.

[8] S. Taams. *Prime Polynomials over Finite Fields and Chebyshev's Bias* (bachelor thesis).
http://fse.studenttheses.ub.rug.nl/15351/1/BSc_Math_2017_Taams_S.pdf.