Tauberian Theorems

For a series $\sum_{n=0}^{\infty} a_n$ of complex numbers, the convergence statement $\sum_{n=0}^{\infty} a_n = A$ means that the n-th partial sum $s_n = \sum_{k=0}^{n} a_k$ as a sequence of complex numbers converges to the complex number A. There are more general ways to define the convergence of a series $\sum_{n=0}^{\infty} a_n$ to A.

For example, Cesàro convergence (named after Ernesto Cesàro) which is defined as the convergence of the Cesàro sum

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_{n-1}}{n}$$

as a sequence to A. Cesàro convergence implies the usual convergence but not vice versa.

Another even more general way to define convergence is Abel convergence, which is defined as the convergence of

$$\sum_{n=0}^{\infty} a_n x^n$$

to A as $x \to 1^-$. Abel convergence implies Cesàro convergence but not vice versa.

By Abel's summation by parts the power series $\sum_{n=0}^{\infty} a_n x^n$ can be rewritten as

$$\begin{aligned}
\sum_{n=0}^{\infty} a_n x^n &= a_0 + \sum_{n=1}^{\infty} \left(s_n - s_{n-1}\right) x^n \\
&= a_0 + \sum_{n=1}^{\infty} s_n x^n - \sum_{n=1}^{\infty} s_{n-1} x^n \\
&= a_0 + \sum_{n=1}^{\infty} s_n x^n - \sum_{n=0}^{\infty} s_n x^{n+1} \\
&= \sum_{n=0}^{\infty} s_n x^n - \sum_{n=0}^{\infty} s_n x^{n+1} \\
&= \sum_{n=0}^{\infty} s_n (1-x) x^n .
\end{aligned}$$

We can interpret the construction of the Cesàro sum σ_n and the Abel sum

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} s_n (1-x) x^n$$

as different ways of taking some weighted average of the sequence $(s_\nu)_{\nu \in \mathbb{N} \cup \{0\}}$. The original s_n can be regarded as the (n-dependent) weighted average of the sequence $(s_\nu)_{\nu \in \mathbb{N} \cup \{0\}}$ by assigning all the weight to s_n and just zero weight to the other s_ν with $\nu \neq n$. The Cesàro sum

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_{n-1}}{n}$$

can be regarded as the (n-dependent) weighted average of the sequence $(s_\nu)_{\nu \in \mathbb{N} \cup \{0\}}$ by assigning the weight $\frac{1}{n}$ to each of $s_0, s_1, \cdots, s_n$ and the zero weight to the other s_ν with $\nu > n$.

The Abel summation assigns the weight $(1-x)x^n$ to s_n for $n \in \mathbb{N} \cup \{0\}$. This weight is x-dependent for $0 < x < 1$. The sum

$$\sum_{n=0}^{\infty} (1-x) x^n$$

of the weights $(1-x)x^n$ for $n \in \mathbb{N} \cup \{0\}$ is 1 from the summing of the geometric series $\sum_{n=0} x^n$. We have a family of weights $((1-x)x^n)_{n=0}^{\infty}$ indexed by the parameter $0 < x < 1$. Abel convergence means that the weighted average

$$\sum_{n=0}^{\infty} s_n (1-x) x^n$$

of the sequence $(s_n)_{n=0}^{\infty}$ approaches A as the parameter x in the family of weights $((1-x)x^n)_{n=0}^{\infty}$ approaches 1 from below.

The original theorem of Tauber of 1897 gives the sufficient *Tauberian condition* of $a_n = o\left(\frac{1}{n}\right)$ for the more general Abel convergence to be reduced back to the stronger usual convergence. 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. A way of taking weighted average can also be referred to as a *kernel*.

In this set of lecture notes on Tauberian theorems, we will do the following.

(1) Prove the original theorem of Tauber of 1897. The proof is completely analogous to the proof of its boundedness version presented as Lemma 2.3 on pp.84-85 in the book on *Fourier Analysis* by Stein and Shakarchi.

(2) Prove the theorem of Littlewood of 1911 which weakens the condition in Tauber's original theorem of 1897 from $a_n = o\left(\frac{1}{n}\right)$ to $a_n = O\left(\frac{1}{n}\right)$ and also prove the related theorem of Hardy-Littlewood of 1914 stating that Abel convergence implies Cesàro convergence if $s_n \geq 0$. We will use the very elegant simple method of Karamata of 1930 to prove both.

(3) Introduce the three related families of weighted averages:

(i) $\frac{1}{\zeta(z)}\frac{1}{n^z}$ from Riemann's zeta function, with $n \in \mathbb{N}$ as the summing variable and z as the parameter variable.

(ii) ze^{-zt} from Laplace transform, with $0 \leq t < \infty$ as the integrating variable and z as the parameter variable.

(iii) zx^{-z-1} from Mellin transform, with $1 \leq x < \infty$ as the integrating variable and z as the parameter variable.

We will then use the Tauberian theorem for the family ze^{-zt} from Laplace transform to present Newman's 1980 simple proof of the Prime Number Theorem.

At the end a remark will be given concerning Wiener's approach to Tauberian theory and its interpretation in terms of Gelfand representation. However, we will not go into the details, because we have not yet introduced Lebesgue's theory of integration and Fourier transforms for integrable and square integrable Lebesgue measurable functions.

First, let us state and prove the original theorem of Tauber of 1897.

Theorem of Tauber of 1897. Let $a_n \in \mathbb{C}$ for $n \in \mathbb{N} \cup \{0\}$ and $A \in \mathbb{C}$ such that the Abel sum $\sum_{n=0}^{\infty} a_n x^n \to A$ as $x \to 1^-$. Assume in addition the Tauberian condition $a_n = o\left(\frac{1}{n}\right)$. Then $s_n \to A$ as $n \to \infty$.

Proof. For $0 < x < 1$ let N be the integral part of $\frac{1}{1-x}$ so that $N \leq \frac{1}{1-x}$ and $N + 1 > \frac{1}{1-x}$. Since $N \to \infty$ as $x \to 1^-$, it suffices to prove that

$$\sum_{n=0}^{\infty} a_n x^n - \sum_{n=0}^{N} a_n \to 0$$

as $x \to 1^-$. Rewrite

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

Given $\varepsilon > 0$, there exists N_0 such that $|na_n| < \varepsilon$ for $n \geq N_0$ and there exists $\delta > 0$ such that $N \geq N_0$ if $1 - \delta < x < 1$. Then for $1 - \delta < x < 1$,

$$\begin{aligned}\left|\sum_{n=N+1}^{\infty} a_n x^n\right| &= \left|\sum_{n=N+1}^{\infty} na_n \frac{x^n}{n}\right| \\ &< \varepsilon \sum_{n=N+1}^{\infty} \frac{x^n}{n} \leq \frac{\varepsilon}{N+1} \sum_{n=0}^{\infty} x^n \\ &= \frac{\varepsilon}{(N+1)(1-x)} < \varepsilon\end{aligned}$$

and

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

Thus

$$\left|\sum_{n=0}^{\infty} a_n x^n - \sum_{n=0}^{N} a_n\right| < 2\varepsilon$$

for $1 - \delta < x < 1$. Q.E.D.

Now we introduce the method of Karamata of 1930 to prove the theorem of Hardy-Littlewood of 1914 and the theorem of Littlewood of 1911 in that order.

Theorem of Hardy-Littlewood of 1914. Let $a_n \in \mathbb{R}$ for $n \in \mathbb{N} \cup \{0\}$ and $A \in \mathbb{R}$ such that the Abel sum $\sum_{n=0}^{\infty} a_n x^n \to A$ as $x \to 1^-$. Assume in addition the Tauberian condition $s_n \geq 0$ for all $n \in \mathbb{N} \cup \{0\}$ (where $s_n = a_0 + \cdots + a_n$). Then $\sigma_n \to A$ as $n \to \infty$ (where σ_n is the Cesàro sum $\frac{s_0 + \cdots + s_{n-1}}{n}$).

Proof. We follow Karamata's very elegant simple proof of 1930 whose idea consists of the following three steps.

(i) Replace x by x^k with $k \geq 1$ in

$$\sum_{n=0}^{\infty} s_n (1-x) x^n = \sum_{n=0}^{\infty} a_n x^n \to A \quad \text{as } x \to 1^-.$$

Take an $\mathbb{R}$-linear combination of the result over a finite number of k's to replace x^n by a polynomial of x^n.

(ii) Sandwich a piecewise continuous function g by two polynomials with the L^1 norm of the difference of the two polynomials approaching 0.

(iii) Choose g to achieve the effect of a characteristic function so that the infinite sum $\sum_{n=0}^{\infty} s_n (1-x) x^n$ essentially becomes σ_N (with some factor of normalization).

The Tauberian condition $s_n \geq 0$ is needed when the sandwiching of g by the two polynomials is multiplied by s_n.

The precise argument is as follows. From

$$\sum_{n=0}^{\infty} s_n (1-x) x^n = \sum_{n=0}^{\infty} a_n x^n \to A \quad \text{as } x \to 1^-.$$

it follows with the replacement of x^{k+1} (with $k \in \mathbb{N} \cup \{0\}$) that

$$\sum_{n=0}^{\infty} s_n (1 - x^{k+1}) x^n (x^n)^k \to A \quad \text{as } x \to 1^-.$$

Using

$$\lim_{x \to 1^-} \frac{1 - x^{k+1}}{1 - x} = \lim_{x \to 1^-} \left(1 + x + \cdots + x^k\right) = k + 1 = \int_{t=0}^{1} t^k dt,$$

we get

$$(1-x)\sum_{n=0}^{\infty} s_n x^n (x^n)^k \to A\int_{t=0}^{1} t^k dt \quad \text{as } x \to 1^-,$$

from which we get, by taking a linear combination over a finite number of k's,

$$(1-x)\sum_{n=0}^{\infty} s_n x^n P(x^n) \to A\int_{t=0}^{1} P(t) dt \quad \text{as } x \to 1^-$$

for any polynomial $P(x)$. This is the same as saying that

$$(1-x)\sum_{n=0}^{\infty} s_n T(x^n) \to A\int_{t=0}^{1} T(t) dt \quad \text{as } x \to 1^-$$

for any polynomial $T(x)$ *without constant term.* Note that we cannot allow any constant term in $T(x)$, otherwise $T(x) \equiv 1$ would have required the condition

$$(1-x)\sum_{n=0}^{\infty} s_n \to A \quad \text{as } x \to 1^-.$$

Our formulation of using $P(t)$ instead of $T(t)$ simply means factoring $T(t) = tP(t)$ so that we remove the restriction of no constant term from $P(t)$ and make its use in the sandwiching process easier.

We now come to the sandwiching process. For any piecewise continuous function $g(t)$ on $[0,1]$ and any $\varepsilon > 0$, we can find polynomials $P_\varepsilon(t)$ and $Q_\varepsilon(t)$ with $P_\varepsilon \leq g \leq Q_\varepsilon$ on $[0,1]$ such that

$$\int_{t=0}^{1} \left(Q_\varepsilon(t) - P_\varepsilon(t)\right) dt < \varepsilon.$$

From

$$(1-x)\sum_{n=0}^{\infty} s_n x^n P_\varepsilon(x^n) \to A\int_{t=0}^{1} P_\varepsilon(t) dt \quad \text{as } x \to 1^-$$

it follows that there exists $\delta_1 > 0$ such that

$$(1-x)\sum_{n=0}^{\infty} s_n x^n P_\varepsilon(x^n) \geq -\varepsilon + A\int_{t=0}^{1} P_\varepsilon(t) dt \geq -2\varepsilon + A\int_{t=0}^{1} g(t) dt$$

for $1-\delta_1 < x < 1$. From

$$(1-x)\sum_{n=0}^{\infty} s_n x^n Q_\varepsilon(x^n) \to A\int_{t=0}^{1} Q_\varepsilon(t)dt \quad \text{as } x \to 1^-$$

it follows that there exists $\delta_2 > 0$ such that

$$(1-x)\sum_{n=0}^{\infty} s_n x^n Q_\varepsilon(x^n) \leq \varepsilon + A\int_{t=0}^{1} Q_\varepsilon(t)dt \leq 2\varepsilon + A\int_{t=0}^{1} g(t)dt$$

for $\delta_2 < x < 1$. Since $s_n \geq 0$ for $n \in \mathbb{N}\cup\{0\}$, it follows that

$$(1-x)\sum_{n=0}^{\infty} s_n x^n P_\varepsilon(x^n) \leq (1-x)\sum_{n=0}^{\infty} s_n x^n g(x^n) \leq (1-x)\sum_{n=0}^{\infty} s_n x^n Q_\varepsilon(x^n)$$

and

$$-2\varepsilon + A\int_{t=0}^{1} g(t)dt \leq (1-x)\sum_{n=0}^{\infty} s_n x^n g(x^n) \leq 2\varepsilon + A\int_{t=0}^{1} g(t)dt$$

for $1-\max(\delta_1,\delta_2) < x < 1$. Thus,

$$(1-x)\sum_{n=0}^{\infty} s_n x^n g(x^n) \to A\int_{t=0}^{1} g(t)dt \quad \text{as } x \to 1^-$$

for any piecewise continuous function $g(t)$ on $[0,1]$.

Let us explore a good way of choosing $g(t)$. For any given $N \in \mathbb{N}$ we would like to choose a piecewise continuous function $g(t)$ on $[0,1]$ and $0 < x_N < 1$ with $x_N \to 1$ as $N \to \infty$ such that

$$\sum_{n=0}^{\infty} s_n x_N^n g(x_N^n) = \sum_{n=0}^{N} s_n.$$

For this purpose, we want $x_N^n g(x_N^n) = 1$ for $n \leq N$ and $x_N^n g(x_N^n) = 0$ for $n > N$. Since $x_N^n < x_N^N$ if and only if $n > N$, this means that we would like to have $tg(t) = 1$ for $t \geq (x_N)^N$ and $g(t) = 0$ for $t < (x_N)^N$, which means $g(t) = \frac{1}{t}\chi_{[(x_N)^N,1]}$, where $\chi_{[(x_N)^N,1]}$ is the characteristic function of the interval $[(x_N)^N, 1]$. We need also to use

$$\int_{t=0}^{1} g(t)dt = \int_{t=(x_N)^N}^{t=1} \frac{1}{t}\,dt = -N\log x_N.$$

A good way is to normalize $\int_{t=0}^{1} g(t)dt$ to be 1 by setting $-N \log x_N = 1$ so that $x_N = e^{-\frac{1}{N}}$ and $(x_N)^N = \frac{1}{e}$. We end up with the choice of $g(t) = \frac{1}{t}\chi_{[\frac{1}{e},1]}$ so that $\int_{t=0}^{1} g(t)dt = 1$ and

$$\sum_{n=0}^{\infty} s_n x_N^n g(x_N^n) = \sum_{n=0}^{N} s_n.$$

We apply

$$(1-x)\sum_{n=0}^{\infty} s_n x^n g(x^n) \to A \int_{t=0}^{1} g(t)dt \quad \text{as } x \to 1^-$$

to $x = x_N$ as $N \to \infty$ to get

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

Now

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

because

$$\lim_{u\to 0} \frac{1 - e^{-u}}{u} = -\left.\frac{d}{du} e^{-u}\right|_{u=0} = 1.$$

Hence $\lim_{N\to\infty} \sigma_N = A$. Q.E.D.

Remark. In the Theorem of Hardy-Littlewood of 1914, the condition $s_n \geq 0$ for $n \in \mathbb{N}$ can be weakened to $s_n \geq -C$ for some $C > 0$ and for $n \in \mathbb{N}$, because the condition $s_n \geq -C$ can be strengthened to $s_n \geq 0$ by simply replacing a_0 by $a_0 + C$.

Now we prove the theorem of Littlewood of 1911 again by the method of Karamata.

Theorem of Littlewood of 1911. Let $a_n \in \mathbb{R}$ for $n \in \mathbb{N} \cup \{0\}$ and $A \in \mathbb{R}$ such that the Abel sum $\sum_{n=0}^{\infty} a_n x^n \to A$ as $x \to 1^-$. Assume in addition the Tauberian condition $a_n = O\left(\frac{1}{n}\right)$ (or even the weaker condition $na_n \geq -C$ for some $C > 0$ and for $n \in \mathbb{N}$. Then $\sum_{n=0}^{N} a_n \to A$ as $N \to \infty$.

Proof. Again we follow Karamata's very elegant simple proof of 1930 whose idea consists of the following three steps.

(i) Replace x by x^k with $k \geq 1$ in

$$\sum_{n=0}^{\infty} a_n x^n \to A \quad \text{as } x \to 1^-.$$

Take an $\mathbb{R}$-linear combination of the result over a finite number of k's to replace x^n by a polynomial of x^n.

(ii) Sandwich a piecewise continuous function g by two polynomials with the L^1 norm of the difference of the two polynomials approaching 0.

(iii) Choose g to achieve the effect of a characteristic function so that the infinite sum $\sum_{n=0}^{\infty} a_m x^n$ essentially becomes σ_N (with some factor of normalization).

The Tauberian condition $a_n \geq -C$ is needed in the inequality from the sandwiching of g by the two polynomials is multiplied. We now implement these three steps.

By replacing x by x^k for some $k \geq 1$ in

$$\sum_{n=0}^{\infty} a_n x^n \to A \quad \text{as } x \to 1^-,$$

we obtain

$$\sum_{n=0}^{\infty} a_n (x^k)^n \to A \quad \text{as } x \to 1^-,$$

because $x^k \to 1^-$ if and only if $x \to 1^-$. By taking an $\mathbb{R}$-linear combination for $1 \leq k \leq m$ with coefficients b_k, we get

$$\sum_{n=0}^{\infty} a_n P(x^n) \to AP(1) \quad \text{as } x \to 1^-$$

for any polynomial $P(t) = \sum_{k=1}^{m} b_k t^k$ without constant term, because $P(1) = \sum_{k=1}^{m} b_k$. Note that unlike the case of Theorem of Hardy-Littlewood of 1914 where the limit is $A \int_0^1 \frac{P(t)}{t} dt$, in our present notation of $P(t)$ without constant term, involving the integral $\int_0^1 \frac{P(t)}{t} dt$ instead of the limit $AP(1)$ here which involves only the value of P at one point 1. Since the sandwiching the piecewise continuous function $g(t)$ by polynomials with approximation only in the L^1 norm, integrals have to enter picture. Here we have only the pointwise value $P(1)$ and integrals would have to come in sometime later.

Unlike in the proof of the theorem of Hardy-Littlewood of 1914, where the polynomial without constant term is put in the form $tP(t)$ so that there is no constraint of no constant term anymore for $P(t)$, here we keep the constraint of no constant term for $P(t)$. So when we do the sandwiching and approximation of $g(t)$ by $P(t)$ in the proof of the theorem of Hardy-Littlewood of 1914, the function $tg(t)$, which is approximated by the polynomial $tP(t)$, should yield the characteristic function to give the partial sum s_N. Thus there we choose $g(t) = \frac{1}{t}\chi_{[\frac{1}{e},1]}$, but here we use $P(t)$ instead of $tP(t)$ and as a result we are going to choose $g(t) = \chi_{[\frac{1}{e},1]}$ instead of $g(t) = \frac{1}{t}\chi_{[\frac{1}{e},1]}$. Again we choose $x_N = e^{-\frac{1}{N}}$ as the link between the two variables x and N so that

$$\sum_{n=0}^{N} a_n = \sum_{n=0}^{\infty} a_n g\left(x_N^n\right),$$

because $x_N^n = e^{-\frac{n}{N}} \geq \frac{1}{e}$ if and only if $n \geq N$.

In addition to the constraint $P(0) = 0$ (that is, no constant term), we impose the addition constraint $P(1)$ so that in the limit statement we have

$$\sum_{n=0}^{\infty} a_n P(x^n) \to A \quad \text{as } x \to 1^-$$

instead of

$$\sum_{n=0}^{\infty} a_n P(x^n) \to AP(1) \quad \text{as } x \to 1^-.$$

When we approximate a piecewise continuous function in L^1 norm by a polynomial, we would like to remove all the constraints on the polynomials so that we have the approximation we want. In the theorem of Hardy-Littlewood of 1914, the only constraint is $P(0) = 0$ and we can get rid of it by using $tP(t)$ so that $P(t)$ is without constraint, that is, we divide $P(t)$ by t to remove the constraint. Now we are going to remove the two constraints by dividing by $t(1-t)$ instead of just by t, but $P(1) = 1$ does not mean that $P(t)$ has 1 as root. If we use $P(t)-1$ so that $t=1$ is a root, the polynomial $P(t)-1$ does not have $t=0$ as a root. One way to make both $t=0$ and $t=1$ roots is to use $P(t)-t$ so that with the condition $P(0)=0$ and $P(1)=1$, we can divide $P(t)-t$ by $t(1-t)$ to define the polynomial $Q(t)=\frac{P(t)-t}{t(1-t)}$. In other words, $P(t)=t+t(1-t)Q(t)$.

We now define the piecewise continuous function $h(t)=\frac{g(t)-t}{t(1-t)}$ so that we can approximate it by polynomials $Q(t)$ without constraints and get $P(t)=t+t(1-t)Q(t)$ as a polynomial with constraints $P(0)=0$ and $P(1)=1$ to approximate $g(t)=t+t(1-t)h(t)$.

Given $\varepsilon>0$. We can find a polynomial $Q_\varepsilon(t)$ such that $h(t)\leq Q_\varepsilon(t)$ for $0\leq t\leq 1$ and $\int_{t=0}^1 (Q_\varepsilon(t)-h(t))\,dt<\varepsilon$. Let $P_\varepsilon(t)=t+t(1-t)Q_\varepsilon(t)$. Since both t and $1-t$ are nonnegative for $0\leq t\leq 1$, we have $g(t)\leq P_\varepsilon(t)$ for $0\leq t\leq 1$. The L^1 estimate $\int_{t=0}^1 (Q_\varepsilon(t)-h(t))\,dt<\varepsilon$ can now be rewritten as

$$\int_{t=0}^1\left(\frac{P_\varepsilon(t)-t}{t(1-t)}-\frac{g(t)-t}{t(1-t)}\right)dt<\varepsilon,$$

which is the same as

$$\int_{t=0}^1\frac{P_\varepsilon(t)-g(t)}{t(1-t)}\,dt<\varepsilon.$$

The polynomial $P_\varepsilon(t)$ gives only one side $g(t)\leq P_\varepsilon(t)$ of the sandwiching and we are going to use this side of the sandwiching to prove that

$$\limsup_{x\to 1^-}\sum_{n=0}^\infty a_n g(x^n)\leq A.$$

The other side of the sandwiching will analogously give

$$\liminf_{x\to 1^-}\sum_{n=0}^\infty a_n g(x^n)\geq A$$

so that both sides of the sandwiching together will give

$$\limsup_{x\to 1^-}\sum_{n=0}^\infty a_n g(x^n)=A.$$

To verify

$$\limsup_{x\to 1^-}\sum_{n=0}^\infty a_n g(x^n)\leq A,$$

since

$$\limsup_{x\to 1^-}\sum_{n=0}^{\infty} a_n P_\varepsilon(x^n) = A,$$

we consider

$$\sum_{n=0}^{\infty} a_n g(x^n) - \sum_{n=0}^{\infty} a_n P_\varepsilon(x^n) = \sum_{n=0}^{\infty} a_n \left(g(x^n) - P_\varepsilon(x^n)\right).$$

We have to link this back to ε through

$$\int_{t=0}^{1} \frac{P_\varepsilon(t) - g(t)}{t(1-t)}\, dt < \varepsilon.$$

For this purpose, we make use of $-na_n\left(P_\varepsilon(x^n) - g(x^n)\right) \geq C\left(P_\varepsilon(x^n) - g(x^n)\right)$ from $na_n \geq -C$ and $P_\varepsilon(x^n) - g(x^n) \geq 0$ and write

$$\begin{aligned}
\sum_{n=0}^{\infty} a_n g(x^n) - \sum_{n=0}^{\infty} a_n P_\varepsilon(x^n) &= -\sum_{n=0}^{\infty} a_n \left(P_\varepsilon(x^n) - g(x^n)\right) \\
&\leq C \sum_{n=0}^{\infty} \frac{1}{n}\left(P_\varepsilon(x^n) - g(x^n)\right) \\
&\leq C \sum_{n=0}^{\infty} \frac{1-x}{1-x^n}\left(P_\varepsilon(x^n) - g(x^n)\right) \\
&= C \sum_{n=0}^{\infty} (1-x)x^n \frac{P_\varepsilon(x^n) - g(x^n)}{x^n(1-x^n)} \\
&= C \sum_{n=0}^{\infty} \left(x^n - x^{n+1}\right) \frac{P_\varepsilon(x^n) - g(x^n)}{x^n(1-x^n)}.
\end{aligned}$$

At this point, one key ingenious observation is that by interpretation in terms of limits of Riemann sums,

$$(\natural) \qquad \lim_{t\to 1^-}\sum_{n=0}^{\infty} \left(x^n - x^{n+1}\right) \frac{P_\varepsilon(x^n) - g(x^n)}{x^n(1-x^n)} = \int_{t=0}^{1} \frac{P_\varepsilon(t) - g(t)}{t(1-t)}\, dt.$$

We now assume $(\natural)$ first and continue with the verification of

$$\limsup_{x\to 1^-}\sum_{n=0}^{\infty} a_n g(x^n) \leq A$$

and then return to the verification of ($\natural$) afterward. From

$$\sum_{n=0}^{\infty} a_n g(x^n) \leq \sum_{n=0}^{\infty} a_n P_\varepsilon(x^n) + C \sum_{n=0}^{\infty} \left(x^n - x^{n+1}\right) \frac{P_\varepsilon(x^n) - g(x^n)}{x^n(1-x^n)}$$

it follows that

$$\begin{aligned}\limsup_{x \to 1^-} \sum_{n=0}^{\infty} a_n g(x^n) &\leq \lim_{x \to 1^-} \sum_{n=0}^{\infty} a_n P_\varepsilon(x^n) + C \lim_{x \to 1^-} \sum_{n=0}^{\infty} \left(x^n - x^{n+1}\right) \frac{P_\varepsilon(x^n) - g(x^n)}{x^n(1-x^n)} \\ &= A + \int_{t=0}^{1} \frac{P_\varepsilon(t) - g(t)}{t(1-t)}\, dt \leq A + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\limsup_{x \to 1^-} \sum_{n=0}^{\infty} a_n g(x^n) \leq A.$$

For the proof of the other direction

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

we choose $\hat{Q}_\varepsilon(t) \leq h(t)$ for $0 \leq t \leq 1$ such that $\int_{t=0}^{1} \left(h(t) - \hat{Q}_\varepsilon(t)\right) dt < \varepsilon$. Let $\hat{P}_\varepsilon(t) = t + t(1-t)\hat{Q}_\varepsilon(t)$. Then $\hat{P}_\varepsilon(t) \leq g(t)$ for $0 \leq t \leq 1$ and $\int_{t=0}^{1} \left(h(t) - \hat{Q}_\varepsilon(t)\right) dt < \varepsilon$ can now be rewritten as

$$\int_{t=0}^{1} \left(\frac{g(t) - t}{t(1-t)} - \frac{\hat{P}_\varepsilon(t) - t}{t(1-t)} \right) dt < \varepsilon,$$

which is the same as

$$\int_{t=0}^{1} \frac{g(t) - \hat{P}_\varepsilon(t)}{t(1-t)}\, dt < \varepsilon.$$

To verify

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

we consider

$$\sum_{n=0}^{\infty} a_n \hat{P}_\varepsilon(x^n) - \sum_{n=0}^{\infty} a_n g(x^n) = \sum_{n=0}^{\infty} a_n \left(\hat{P}_\varepsilon(x^n) - g(x^n) \right).$$

We have to link this back to ε through

$$\int_{t=0}^{1} \frac{g(t) - \hat{P}_\varepsilon(t)}{t(1-t)}\, dt < \varepsilon.$$

For this purpose, we make use of $-na_n \left(g(x^n) - \hat{P}_\varepsilon(x^n) \right) \geq C \left(g(x^n) - \hat{P}_\varepsilon(x^n) \right)$ from $na_n \geq -C$ and $g(x^n) - \hat{P}_\varepsilon(x^n) \geq 0$ and write

$$\begin{aligned}
\sum_{n=0}^{\infty} \hat{P}_\varepsilon(x^n) a_n - \sum_{n=0}^{\infty} a_n g(x^n) &= -\sum_{n=0}^{\infty} a_n \left(g(x^n) - \hat{P}_\varepsilon(x^n) \right) \\
&\leq C \sum_{n=0}^{\infty} \frac{1}{n} \left(g(x^n) - \hat{P}_\varepsilon(x^n) \right) \\
&\leq C \sum_{n=0}^{\infty} \frac{1-x}{1-x^n} \left(g(x^n) - \hat{P}_\varepsilon(x^n) \right) \\
&= C \sum_{n=0}^{\infty} (1-x) x^n \frac{g(x^n) - \hat{P}_\varepsilon(x^n)}{x^n(1-x^n)} \\
&= C \sum_{n=0}^{\infty} \left(x^n - x^{n+1} \right) \frac{g(x^n) - \hat{P}_\varepsilon(x^n)}{x^n(1-x^n)}.
\end{aligned}$$

Again we leave to handle later the statement that

$$(\hat{\natural}) \qquad \lim_{t \to 1^-} \sum_{n=0}^{\infty} \left(x^n - x^{n+1} \right) \frac{g(x^n) - \hat{P}_\varepsilon(x^n)}{x^n(1-x^n)} = \int_{t=0}^{1} \frac{g(t) - \hat{P}_\varepsilon(t)}{t(1-t)}\, dt.$$

and continue with the verification of

$$\liminf_{x \to 1^-} \sum_{n=0}^{\infty} a_n g(x^n) \geq A$$

and then return to the verification of $(\hat{\natural})$ afterward. From

$$\sum_{n=0}^{\infty} a_n g(x^n) \geq \sum_{n=0}^{\infty} a_n \hat{P}_\varepsilon(x^n) - C \sum_{n=0}^{\infty} \left(x^n - x^{n+1} \right) \frac{g(x^n) - \hat{P}_\varepsilon(x^n)}{x^n(1-x^n)}$$

it follows that

$$\liminf_{x\to 1^-}\sum_{n=0}^\infty a_n g(x^n) \geq \lim_{x\to 1^-}\sum_{n=0}^\infty a_n \hat{P}_\varepsilon(x^n) - C\lim_{x\to 1^-}\sum_{n=0}^\infty \left(x^n - x^{n+1}\right)\frac{g(x^n)-\hat{P}_\varepsilon(x^n)}{x^n(1-x^n)}$$
$$= A - \int_{t=0}^1 \frac{g(t)-\hat{P}_\varepsilon(t)}{t(1-t)}\,dt \leq A-\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

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

We now handle ($\natural$) and ($\hat{\natural}$). Since both ($\natural$) and ($\hat{\natural}$) are completely analogous, we will only do ($\natural$). Take an arbitrary $\eta > 0$. Let B_ε be the supremum of the absolute value of the piecewise continuous function

$$\frac{P_\varepsilon(t)-g(t)}{t(1-t)}$$

for $0 \leq t \leq 1$. Then by using the partition of $[0,1]$ by an infinite number of points x^n for $n \in \mathbb{N}$ we get

$$\left|\sum_{n=0}^\infty \left(x^n - x^{n+1}\right)\frac{P_\varepsilon(x^n)-g(x^n)}{x^n(1-x^n)} - \int_{t=0}^1 \frac{P_\varepsilon(t)-g(t)}{t(1-t)}\,dt\right|$$
$$\leq \sum_{n=0}^\infty \left|\frac{P_\varepsilon(x^n)-g(x^n)}{x^n(1-x^n)}\left(x^n - x^{n+1}\right) - \int_{t=x^{n+1}}^{t=x^n}\frac{P_\varepsilon(t)-g(t)}{t(1-t)}\,dt\right|$$
$$\leq \sum_{n=0}^\infty \left(\sup_{s,t\in[x^{n+1},x^n]}\left|\frac{P_\varepsilon(s)-g(s)}{s(1-s)} - \frac{P_\varepsilon(t)-g(t)}{t(1-t)}\right|\right)\left(x^n - x^{n+1}\right).$$

Since there is only one jump discontinuity of the function

$$\frac{P_\varepsilon(t)-g(t)}{t(1-t)}$$

on $[0,t]$ (which is at $t=\frac{1}{e}$), there exists $\delta_{\eta,\varepsilon} > 0$ such that if $1-\delta_{\eta,\varepsilon} < x < 1$, then

$$\sup_{s,t\in[x^{n+1},x^n]}\left|\frac{P_\varepsilon(s)-g(s)}{s(1-s)} - \frac{P_\varepsilon(t)-g(t)}{t(1-t)}\right| < \eta$$

for $n \in \mathbb{N}$ except when $\frac{1}{e} \in [x^n, x^{n+1}]$, which can only occur for at most 2 values of n. Hence

$$\sum_{n=0}^{\infty}\left(\sup_{s,t\in[x^{n+1},x^n]}\left|\frac{P_\varepsilon(s)-g(s)}{s(1-s)}-\frac{P_\varepsilon(t)-g(t)}{t(1-t)}\right|\right)\left(x^n-x^{n+1}\right)$$
$$\leq \eta\sum_{n=0}^{\infty}\left(x^n-x^{n+1}\right)+4B_\varepsilon\left(1-\delta_{\eta,\varepsilon}\right).$$

Since $\eta > 0$ is arbitrary, we conclude that

$$\lim_{x\to 1^-}\sum_{n=0}^{\infty}\left(x^n-x^{n+1}\right)\frac{P_\varepsilon(x^n)-g(x^n)}{x^n(1-x^n)}=\int_{t=0}^{1}\frac{P_\varepsilon(t)-g(t)}{t(1-t)}\,dt.$$

This finishes the verification of ($\natural$) and the proof of the theorem of Littlewood of 1911.

We now discuss Newman's 1980 simple proof of the Prime Number Theory which uses a Tauberian theorem. The Tauberian theorem used is formuated in terms the Laplace (or equivalent the Mellin) transform. The family of weights used to take the weighted average is ze^{-zt} for the Laplace transform version (with $0 \leq t < \infty$) and is zx^{-z-1} in for the Mellin transform (with $1 \leq x < \infty$). The two are related by $x = e^t$. The precise statement of the Laplace transform version of the Tauberian theorem is the following.

Theorem (Laplace Transform Version of Tauberian Theorem). Let $F(t)$ for $0 \leq t < \infty$ be a bounded, piecewise continuous function. Let

$$G(z)=\int_{t=0}^{\infty}F(t)e^{-zt}dt,$$

which is automatically holomorphic on $\{\operatorname{Re} z > 0\}$. If $G(z)$ can be extended to a holomorphic function on an open neighborhood U of $\{\operatorname{Re} z = 0\}$ in $\mathbb{C}$ (as the Tauberian condition), then $\int_{t=0}^{\infty} F(t)dt$ as the limit of $\int_{t=0}^{\lambda} F(t)dt$ is equal to $G(0)$.

Remark. Let us put this Laplace transform version of the Tauberian theorem in the context of (the series version of the Tauberian theorem as given in) Tauber's theorem in 1897 (and its later refinement by Littlewood in 1911) by the following tabulated analogy.

Laplace Transform Version	Series Version
$F(t)$	a_n
$0 \leq t < \infty$	$0 \leq n < \infty$
$G(z) = \int_{t=0}^{\infty} F(t) e^{-zt} dt$	$\sum_{n=0}^{\infty} a_n x^n$
$\operatorname{Re} z > 0$	$0 \leq x < 1$
Tauberian Condition: $G(z)$ holomorphic in neighborhood of $\{\operatorname{Re} z > 0\}$ in $\mathbb{C}$	*Tauberian Condition:* $\sum_{n=0}^{\infty} a_n x^n$ continuous on $x \in [0,1]$ and $a_n = o(\frac{1}{n})$ $\big($or Littlewood's $a_n = O(\frac{1}{n})\big)$
Conclusion: $\lim_{\lambda \to \infty} \int_{t=0}^{\lambda} F(t) dt = \lim_{\substack{z \to 0, \\ \operatorname{Re} z > 0}} G(z)$	*Conclusion:* $\lim_{N \to \infty} \sum_{n=0}^{N} a_n = \lim_{x \to 1^-} \sum_{n=0}^{\infty} a_n x^n$
$\lambda \to \infty$	$N \to \infty$

Before its application to the proof of the Prime Number Theorem the Laplace transform version of Tauberian theorem will be changed to the following Mellin transform version of Tauberian theorem.

Theorem (Mellin Transform Version of Tauberian Theorem). Let $f(x)$ be a nonnegative, piecewise continuous, nondecreasing function for $1 \leq x < \infty$ such that $f(x) = O(x)$ as $x \to \infty$. Denote by $g(z)$ the Mellin transform of $f(x)$ so that

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

which is automatically holomorphic on $\{\operatorname{Re} z > 1\}$. If for some complex number c the function $g(z) - \frac{c}{z-1}$ can be extended to a holomorphic function on an open neighborhood U of $\{\operatorname{Re} z = 1\}$ in $\mathbb{C}$ (as the Tauberian conition), then $\frac{f(x)}{x} \to c$ as $x \to \infty$.

The Laplace transform version of Tauberian theorem and the Mellin transform version of Tauberian theorem are related by $F(t) = e^{-t} f(e^t) - c$. In its application to the proof of the Prime Number Theorem, the second Chebyshev function $\psi(x) = \sum_{n \leq x} \Lambda(n)$ will be chosen as $f(x)$, where $\Lambda(n)$ is the von Mangoldt function whose value is 0 unless $n = p^k$ for some prime number when its value is $\log p$.

We now prove the Laplace transform version of Tuaberian theorem by applying Cauchy's integral formula to verify that $G(0) - G_\lambda(0) \to 0$ as $\lambda \to \infty$, where

$$G_\lambda(z) = \int_{t=0}^{\lambda} F(t) e^{-zt} dt$$

and the kernel $\frac{1}{z}$ in Cauchy's integral formula is changed to

$$e^{\lambda z}\left(\frac{1}{z} + \frac{z}{R^2}\right)$$

while the contour for the Cauchy integral formula is chosen to be the boundary of $\{\,\mathrm{Re}\, z > -\delta,\ |z| < R\,\}$ for some appropriately chosen $R > 0$ and $\delta > 0$.

Proof of Laplace Transform Version of Tauberian Theorem. Since $F(t)$ is bounded on $\{\,0 \le t < \infty\,\}$, without loss of generality we can assume that $\sup_{0 \le t < \infty} |F(t)| = 1$. For $\lambda > 0$ let

$$G_\lambda(z) = \int_{t=0}^{\lambda} F(t) e^{-zt} dt,$$

which is automatically holomorphic on all of $\mathbb{C}$. We are going to prove the conclusion

$$\lim_{\lambda \to \infty} \int_{t=0}^{\lambda} F(t) dt = G(0)$$

by applying Cauchy's integral formula to verify that $G(0) - G_\lambda(0) \to 0$ as $\lambda \to \infty$, where

$$G_\lambda(z) = \int_{t=0}^{\lambda} F(t) e^{-zt} dt.$$

The straightforward application of the original form of Cauchy's integral formula would read

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

where C is a simple closed curve in U enclosing a neighborhood of 0 in U. In order to estimate the contour integral

$$\int_C \left(G(z) - G_\lambda(z)\right) \frac{1}{z}\, dz$$

as $\lambda \to \infty$, we need to estimate $G(z) - G_\lambda(z)$ for $z \in C$. Let $z = x + iy$. By definition

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

so that

$$|G(z) - G_\lambda(z)| \leq \int_{t=\lambda}^{\infty} e^{-xt} dt = \left[-\frac{e^{-xt}}{x} \right]_{t=\lambda}^{t=\infty} = \frac{e^{-\lambda x}}{x} = \frac{\left|e^{-\lambda z}\right|}{\operatorname{Re} z}$$

if $x = \operatorname{Re} z > 0$. For $x = \operatorname{Re} z > 0$ the factor $e^{-\lambda x} = \left|e^{-\lambda z}\right|$ is decaying exponentially as $\lambda \to \infty$. In order not to waste such an exponential decay, we modify the kernel $\frac{1}{z}$ in the Cauchy integral formula to $\frac{e^{\lambda z}}{z}$ to offset this wasted exponential decay so that the factor $e^{\lambda x} = \left|e^{-\lambda z}\right|$ coming from the new kernel $\frac{e^{\lambda z}}{z}$ can be used to our advantage in the part of the contour C where $x = \operatorname{Re} z < 0$ when $\lambda \to \infty$.

Though the domain U where $G(z)$ is holomorphic contains $\{ \operatorname{Re} z = 0 \}$, yet we have no control how close the boundary of U is to the origin $z = 0$. In the above estimate

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

the denominator $\operatorname{Re} z$ would make things difficult when z is close to 0. An important technique which Newman introduced is that since on $|z| = R$,

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

the disadvantage of having $\operatorname{Re} z$ in the denominator of $\frac{\left|e^{-\lambda z}\right|}{\operatorname{Re} z}$ can be offset by the advantage of having $\operatorname{Re} z$ in the numerator of $\frac{2\operatorname{Re} z}{R^2}$ at least on $\{ |z| = R \}$ if we modify the Cauchy kernel $\frac{1}{z}$ to $\frac{1}{z} + \frac{z}{R^2}$. Combining the two considerations together, Newman ended up with replacing the usual Cauchy kernel $\frac{1}{z}$ by the new Cauchy kernel

$$e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right)$$

which like $\frac{1}{z}$ is holomorphic on $\mathbb{C}$ as a function of z except a simple pole at $z = 0$ with residue 1. Now we use the modified Cauchy integral formula

$$G(0) - G_\lambda(0) = \int_C \left(G(z) - G_\lambda(z) \right) e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz$$

to verify that $G(0) - G_\lambda(0) \to 0$ as $\lambda \to \infty$.

Now we come to the choice of the contour C. With the new modified Cauchy kernel we can offset the disadvantage of $\mathrm{Re}\, z$ in the denominator by the advantage of $\mathrm{Re}\, z$ only on the circle $\{\, |z| = R \,\}$. For that reason we want the contour C to be as close to the circle $\{\, |z| = R \,\}$ as possible. However, though the domain U where $G(z)$ is holomorphic contains the right half-plane $\{\, \mathrm{Re}\, z > 0 \,\}$, we have no addition information about it other than its being a neighborhood of the imaginary axis $\{\, \mathrm{Re}\, z = 0 \,\}$ in $\mathbb{C}$.

The best we can do to make the contour C as close to the circle $\{\, |z| = R \,\}$ as possible is to set it to be the boundary of $\{\, |z| < R, \mathrm{Re}\, z > -\delta_R \,\}$ with $\delta_R > 0$ chosen (as a function of R) so that $\{\, |z| \leq R, \mathrm{Re}\, z \geq -\delta_R \,\}$ is contained in U. Because of the dependence of C on R and δ_R in such a choice, we denote C by C_{R,δ_R}. To do our estimate of the new modified Cauchy integral formula, we need to differentiate the different cases according to which part of C_{R,δ_R} the variable z lies. For that reason we break up the contour C_{R,δ_R} into three parts. The first part is the right half-circle of radius R which we denote by

$$C_R^+ = \{\, |z| = R, \mathrm{Re}\, z > 0 \,\} .$$

The second part is the union of two circular arcs on the left half-plane which we denote by

$$A_{R,\delta_R} = \{\, |z| = R, -\delta_R < \mathrm{Re}\, z < 0 \,\} .$$

The third part is the vertical line segment on the line $\{\, \mathrm{Re}\, z = -\delta_R \,\}$ in the left half-plane which we denote by

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

To do the estimate, we start out with an arbitrary $\varepsilon > 0$. We will first choose $R \geq R_\varepsilon$. Then we will choose $0 < \delta_R \leq \delta_{R,\varepsilon}$. Then we will choose $\lambda \geq \lambda_{R,\delta_R,\varepsilon}$ to make $|G(0) - G_\lambda(0)|$ less than some universal constant times ε when $\lambda \geq \lambda_{R,\delta_R,\varepsilon}$.

On the first part C_R^+ of the contour C_R, from

$$|G(z) - G_\lambda(z)| \leq \frac{\left|e^{-\lambda z}\right|}{\mathrm{Re}\, z}$$

and

$$\frac{1}{z} + \frac{z}{R^2} = \frac{2\mathrm{Re}\, z}{R^2}$$

it follows that

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

Thus

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

for $R\ge R_\varepsilon$ if R_ε is set to be $\frac{2\pi}{\varepsilon}$.

Now we fix an $R\ge R_\varepsilon$ and choose $\delta_R>0$ small enough so that $C_{R,\delta}$ together with the domain it encloses is contained in the domain U where $G(z)$ is assumed to be holomorphic. On the second and third parts A_{R,δ_R} and L_{R,δ_R} of the contour C_{R,δ_R}, we are going to do the estimate separately for $G(z)$ and $G_\lambda(z)$ instead of doing it at the same time for $G(z)-G_\lambda(z)$. The reason is that on the right half-plane we have the explicit integral formula

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

which we use to get our estimate on the first part C_R^+ of the contour $C_{R,\delta}$, but on the left half-plane, while we still have the explicit integral formula

$$G_\lambda(z)=\int_{t=0}^\lambda F(t)e^{-zt}dt$$

for $G_\lambda(z)$, we have no integral formula for $G(z)$ except that we know that $G(z)$ is independent of λ. In our separate estimates for $G_\lambda(z)$ and $G(z)$, in the case of $G_\lambda(z)$, we use its explicit integral formula and in the case of $G(z)$ we use the fact that $G(z)$ is independent of λ.

Since $G_\lambda(z)$ is holomorphic on all of $\mathbb{C}$, by Cauchy's theorem

$$\int_{A_{R,\delta_R}+L_{R,\delta_R}}G_\lambda(z)e^{\lambda z}\left(\frac{1}{z}+\frac{z}{R^2}\right)dz=\int_{\substack{|z|=R,\\ \operatorname{Re} z<0}}G_\lambda(z)e^{\lambda z}\left(\frac{1}{z}+\frac{z}{R^2}\right)dz.$$

From the explicit integral formula

$$G_\lambda(z)=\int_{t=0}^\lambda F(t)e^{-zt}dt$$

for $x = \operatorname{Re} z < 0$ we have the estimate

$$|G_\lambda(z)| \leq \int_{t=0}^{\lambda} e^{-xt} dt = \left[-\frac{e^{-xt}}{x} \right]_{t=0}^{t=\lambda} = \frac{e^{-\lambda x} - 1}{-x} \leq \frac{e^{-\lambda x}}{-x} = \frac{\left| e^{-\lambda z} \right|}{-\operatorname{Re} z}$$

and on $\{ |z| = R, \operatorname{Re} z < 0 \}$,

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

Thus

$$\left| \int_{A_{R,\delta_R} + L_{R,\delta_R}} G_\lambda(z)\, e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz \right| \leq \pi R\, \frac{2}{R^2} < \varepsilon$$

for $R \geq R_\varepsilon = \frac{2\pi}{\varepsilon}$.

We now handle the estimate for $G(z)$ on $A_{R,\delta_R} + L_{R,\delta_R}$. We do this for A_{R,δ_R} first. Since $\left| e^{\lambda z} \right| \leq 1$ on the left half-plane $\{ \operatorname{Re} z < 0 \}$, for fixed R, clearly there exists sompositive number $\delta_{R,\varepsilon}$ such that

$$\left| \int_{A_{R,\delta_R}} G(z)\, e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz \right| < \varepsilon$$

for $0 < \delta_R \leq \delta_{R,\varepsilon}$.

We now fix $R \geq R_\varepsilon$ and then fix $0 < \delta \leq \delta_{R,\varepsilon}$ and let λ vary in the estimate of

$$\left| \int_{L_{R,\delta_R}} G(z)\, e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz \right|.$$

Here the only dependence on λ is through $e^{\lambda z}$. On $\{ \operatorname{Re} z = -\delta_R \}$ we have

$$\left| e^{\lambda z} \right| \leq e^{-\lambda \delta_R}$$

which goes to 0 as $\lambda \to \infty$. Hence there exists some $\lambda_{R,\delta_R,\varepsilon} > 0$ such that

$$\left| \int_{L_{R,\delta_R}} G(z)\, e^{\lambda z} \left(\frac{1}{z} + \frac{z}{R^2} \right) dz \right| < \varepsilon$$

for $\lambda \geq \lambda_{R,\delta_R,\varepsilon}$. This concludes the proof of the Laplace transform version of the Tauberian theorem. Q.E.D.

We now prove the following Mellin transform version of the Tauberian theorem.

Theorem (Mellin Transform Version of Tauberian Theorem). Let $f(x)$ be a nonnegative, piecewise continuous, nondecreasing function for $1 \leq x < \infty$ such that $f(x) = O(x)$ as $x \to \infty$. Denote by $g(z)$ the Mellin transform of $f(x)$ so that

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

which is automatically holomorphic on $\{\operatorname{Re} z > 1\}$. If for some complex number c the function $g(z) - \frac{c}{z-1}$ can be extended to a holomorphic function on an open neighborhood U of $\{\operatorname{Re} z = 1\}$ in $\mathbb{C}$ (as the Tauberian conition), then $\frac{f(x)}{x} \to c$ as $x \to \infty$.

Proof. First of all we would like to remark that the growth condition $f(x) = O(x)$ as $x \to \infty$ is needed to obtain the complex-analyticity of

$$g(z) = z \int_{x=1}^{\infty} f(x) x^{-z-1} dx$$

on $\{\operatorname{Re} z > 1\}$ because of the need to handle the improperness of the integral as $x \to \infty$.

Secondly we would like to point out that though the Laplace transform version of the Tauberian theorem is related to the Mellin transform version of the Tauberian theorem by the change of variables $x = e^t$, in the Mellin transform version of the Tauberian theorem the Tauberian condition is that for some complex number c the function $g(z) - \frac{c}{z-1}$ can be extended to a holomorphic function on an open neighborhood U of $\{\operatorname{Re} z = 1\}$ in $\mathbb{C}$, whereas the Tauberian condition for the Laplace transform version involves only the holomorphic extension of $G(z)$ to an open neighborhood U of $\{\operatorname{Re} z = 0\}$ in $\mathbb{C}$ without any subtraction of a principal part.

The introduction of the principal part $\frac{c}{z-1}$ to the meromorphic extension of $g(z)$ to an open neighborhood U of $\{\operatorname{Re} z = 1\}$ in $\mathbb{C}$ is necessitated by its application to the proof of the Prime Number Theorem. Because of this principal part $\frac{c}{z-1}$ in the meromorphic extension of $g(z)$, the relation of $f(x)$ and $F(t)$ needs something more complicated than the change of variables $x = e^t$ to reduce the the Mellin transform version of the Tauberian theorem to the Laplace transform version of the Tauberian theorem. Since

$$\int_0^\infty e^{-zt}dt = \frac{1}{z} \quad \text{for } \operatorname{Re} z > 0,$$

in order to handle the complication from the principal part $\frac{c}{z-1}$ in the meromorphic extension of $g(z)$ it suffices to introduce

$$F(t) = e^{-t} f(e^t) - c$$

in the use of the change of variables, where the factor e^{-t} is to take care of the Jacobian in the integral from the change of variables and the constant $-c$ is to take care of the principal part $\frac{c}{z-1}$ in the meromorphic extension of $g(z)$. Note that the growth condition $f(x) = O(x)$ implies that $F(t)$ is uniformly bounded for $0 \leq t < \infty$.

Let $G(z)$ be the Laplace transform of $F(t)$. To relate $G(z)$ to $g(z)$, with the change of variables $x = e^t$ we have

$$\begin{aligned} G(z) &= \int_{t=0}^\infty F(t) e^{-zt} dt \\ &= \int_{t=0}^\infty \left(e^{-t} f(e^t) - c\right) e^{-zt} dt \\ &= \int_{x=1}^\infty \left(\frac{1}{x} f(x) - c\right) x^{-z} \frac{dx}{x} \\ &= \int_{x=1}^\infty f(x) x^{-z-2} dx - \frac{c}{z} \\ &= \frac{1}{z+1} \left(g(z+1) - \frac{c}{z} - c\right). \end{aligned}$$

Since the Tauberian condition in the Mellin transform version of the Tauberian theorem gives the holomorphic extension of $g(z+1) - \frac{c}{z}$ to an open neighborhood of $\{ \operatorname{Re} z = 0 \}$ in $\mathbb{C}$, it follows that $G(z)$ can be holomorphically extended to an open neighborhood of $\{ \operatorname{Re} z = 0 \}$ in $\mathbb{C}$.

By the Laplace transform version of the Tauberian theorem, the integral

$$\int_{t=0}^\infty F(t) dt = \int_{t=0}^\infty \left(e^{-t} f(e^t) - c\right) dt = \int_{x=1}^\infty \left(\frac{f(x)}{x} - c\right) \frac{dx}{x}$$

is finite. We are going to use the nondecreasing property of the function $f(x)$ to conclude from the finiteness of the integral

$$\int_{x=1}^{\infty}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x}$$

that $\frac{f(x)}{x}\to c$ as $x\to\infty$. Of course, if $\frac{f(x)}{x}-c$ is bounded from below by a positive number (or bounded from above by a negative number) on some subset of infinite length in $[1,\infty)$, there is a contradiction of the finiteness of the integral

$$\int_{x=1}^{\infty}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x},$$

but we need to get a contradiction for the stronger statement that the subset of $[1,\infty)$ where $\frac{f(x)}{x}-c$ is bounded from below by any positive number is bounded (and also the subset of $[1,\infty)$ where $\frac{f(x)}{x}-c$ is bounded from above by any negative number is bounded). The nondecreasing property of $f(x)$ is used to conclude from $\frac{f(x_0)}{x_0}-c\geq\varepsilon$ at some $1\leq x_0<\infty$ that there is some interval of length $I_{x_0,\varepsilon}$ in $[x_0,\infty)$ where $\frac{f(x)}{x}-c\geq\frac{\varepsilon}{2}$ so that

$$\frac{\varepsilon}{2}\int_{I_{x_0,\varepsilon}}\frac{dx}{x}$$

has a positive lower bound as $x_0\to\infty$, yielding a contradiction to the finiteness of the integral

$$\int_{x=1}^{\infty}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x}.$$

In order to get an interval $I_{x_0,\varepsilon}$ from a point x_0, the key point is that the positive lower bound ε in the inequality $\frac{f(x_0)}{x_0}-c\geq\varepsilon$ is replaced by the positive lower bound $\frac{\varepsilon}{2}$ in the inequality $\frac{f(x)}{x}-c\geq\frac{\varepsilon}{2}$. Here is the precise argument.

Suppose $\varepsilon>0$ and $\frac{f(x_0)}{x_0}-c\geq\varepsilon$ for some $1\leq x_0<\infty$. Define the interval

$$I_{x_0,\varepsilon}=\left[x_0,\,\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\,x_0\right].$$

By the nondecreasing property of $f(x)$, for $x_0\leq x\leq\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\,x_0$ we have

$$f(x)\geq f(x_0)\geq x_0(c+\varepsilon)\geq x\left(c+\frac{\varepsilon}{2}\right)$$

so that $\frac{f(x)}{x} - c \geq \frac{\varepsilon}{2}$ on the interval $I_{x_0,\varepsilon}$. We compute

$$\int_{I_{x_0,\varepsilon}} \frac{dx}{x} = \Big[\log x\Big]_{x=x_0}^{x=\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\,x_0} = \log\left(\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\right).$$

From

$$\int_{x_0}^{\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}x_0}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x} \geq \frac{\varepsilon}{2}\int_{I_{x_0,\varepsilon}}\frac{dx}{x} = \frac{\varepsilon}{2}\log\left(\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\right)$$

we get a contradiction from

$$\lim_{x_0\to\infty}\int_{x_0}^{\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}x_0}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x} \geq \frac{\varepsilon}{2}\log\left(\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}\right) > 0$$

and the finiteness of

$$\int_{x=1}^{\infty}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x},$$

if for some $\varepsilon > 0$ there exists a sequence $1 \leq x_0 < \infty$ going to ∞ with $\frac{f(x_0)}{x_0} - c \geq \varepsilon$.

Likewise, if for some $\varepsilon > 0$ there exists a sequence $1 \leq x_0 < \infty$ going to ∞ with $\frac{f(x_0)}{x_0} - c \leq -\varepsilon$, then we can define the interval

$$J_{x_0,\varepsilon} = \left[\frac{c-\varepsilon}{c-\frac{\varepsilon}{2}}\,x_0,\, x_0\right].$$

By the nondecreasing property of $f(x)$, for $\frac{c-\varepsilon}{c-\frac{\varepsilon}{2}}\,x_0 \leq x \leq x_0$ we have

$$f(x) \leq f(x_0) \leq x_0(c-\varepsilon) \leq x\left(c-\frac{\varepsilon}{2}\right)$$

so that $\frac{f(x)}{x} - c \leq -\frac{\varepsilon}{2}$ on the interval $J_{x_0,\varepsilon}$. From

$$\int_{J_{x_0,\varepsilon}}\frac{dx}{x} = \log\left(\frac{c-\frac{\varepsilon}{2}}{c-\varepsilon}\right).$$

and

$$\int_{\frac{c-\frac{\varepsilon}{2}}{c-\varepsilon}\,x_0}^{x_0}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x} \geq \frac{\varepsilon}{2}\int_{J_{x_0,\varepsilon}}\frac{dx}{x} = \frac{\varepsilon}{2}\log\left(\frac{c-\frac{\varepsilon}{2}}{c-\varepsilon}\right)$$

we get a contradiction from

$$\lim_{x_0\to\infty}\int_{x_0}^{\frac{c+\varepsilon}{c+\frac{\varepsilon}{2}}x_0}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x}\geq\frac{\varepsilon}{2}\log\left(\frac{c-\frac{\varepsilon}{2}}{c-\varepsilon}\right)>0$$

and the finiteness of

$$\int_{x=1}^{\infty}\left(\frac{f(x)}{x}-c\right)\frac{dx}{x}.$$

This finishes the proof that $\frac{f(x)}{x}\to c$ as $x\to\infty$. Q.E.D.

Riemann Zeta Function as Mellin Transform of Integral Part Function. For $\operatorname{Re} z>1$, the Riemann zeta function

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

can be written as the Riemann-Stieltjes integral

$$\zeta(z)=\int_{t=\frac{1}{2}}^{\infty}t^{-z}d\lfloor t\rfloor$$

(where t^{-z} means $e^{-z\log t}$ and $\lfloor t\rfloor$ means the largest integer not exceeding t) which after integration by parts becomes

$$\zeta(z)=\left[t^{-z}\lfloor t\rfloor\right]_{t=\frac{1}{2}}^{\infty}+z\int_{t=\frac{1}{2}}^{\infty}\lfloor t\rfloor t^{-(z+1)}dt=z\int_{t=1}^{\infty}\lfloor t\rfloor t^{-(z+1)}dt.$$

This means that on $\{\operatorname{Re} z>1\}$ the Riemann zeta function is representable in the following formula as the Mellin transform of the integral part function $t\mapsto\lfloor t\rfloor$.

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

Instead of using the Riemann-Stieltjes integral, the formula

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

which expresses the Riemann zeta function as the Mellin transform of the integral part function $t\mapsto\lfloor t\rfloor$, can be directly derived by summation by parts as follows. On $\{\operatorname{Re} z > 1\}$ we have

$$\begin{aligned}
z\int_{t=1}^{\infty} \lfloor t \rfloor t^{-(z+1)} dt &= z\sum_{n=1}^{\infty} n \int_{t=n}^{n+1} t^{-(z+1)} dt \\
&= z\sum_{n=1}^{\infty} n \left[-\frac{t^{-z}}{z}\right]_{t=n}^{t=n+1} \\
&= \sum_{n=1}^{\infty} n \left(\frac{1}{n^z} - \frac{1}{(n+1)^z}\right) \\
&= \sum_{n=1}^{\infty} \frac{n}{n^z} - \sum_{n=1}^{\infty} \frac{n}{(n+1)^z} \\
&= \sum_{n=1}^{\infty} \frac{n}{n^z} - \sum_{n=2}^{\infty} \frac{n-1}{n^z} \\
&= \frac{1}{n^z} + \sum_{n=2}^{\infty} \frac{n}{n^z} - \sum_{n=2}^{\infty} \frac{n-1}{n^z} \\
&= \sum_{n=1}^{\infty} \frac{1}{n^z} \\
&= \zeta(z).
\end{aligned}$$

Meromorphic extension of Riemann Zeta Function by Comparison with Mellin Transform of Identity Function. The formula

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

which expresses the Riemann zeta function as the Mellin transform of the integral part function $t \mapsto \lfloor t \rfloor$, holds only on $\{\operatorname{Re} z > 1\}$, because the order of growth of the integral part function $t \mapsto \lfloor t \rfloor$ is like t, which cancels one order in the denominator of $t^{-(z+1)}$. If we can somehow remove one order from the integral part function $t \mapsto \lfloor t \rfloor$, then we can get a formula for the Riemann zeta function $\zeta(z)$ on $\{\operatorname{Re} z > 1\}$. For this purpose we can consider

$$z\int_{t=1}^{\infty} \left(\lfloor t \rfloor - t\right) t^{-(z+1)} dt$$

on $\{\operatorname{Re} z > 1\}$ which is equal to

$$\begin{aligned} z \int_{t=1}^{\infty} \lfloor t \rfloor t^{-(z+1)} dt - z \int_{t=1}^{\infty} t^{-z} dt \\ = z\zeta(z) - z \left[\frac{t^{-z+1}}{-z+1} \right]_{t=1}^{t=\infty} \\ = z\zeta(z) - \frac{z}{z-1}. \end{aligned}$$

That is,

$$\zeta(z) = 1 + \frac{1}{z-1} + z \int_{t=1}^{\infty} \left(\lfloor t \rfloor - t \right) t^{-(z+1)} dt.$$

Since the integral

$$\int_{t=1}^{\infty} \left(\lfloor t \rfloor - t \right) t^{-(z+1)} dt$$

defines a holomorphic function on $\{\operatorname{Re} z > 0\}$, it follows that

$$\zeta(z) = 1 + \frac{1}{z-1} + z \int_{t=1}^{\infty} \left(\lfloor t \rfloor - t \right) t^{-(z+1)} dt$$

defines the meromorphic extension of the Riemann zeta function $\zeta(z)$ to $\{\operatorname{Re} z > 0\}$, which is holomorphic except at $z = 1$ where it has a simple pole with residue 1.

Nonvanishing of Riemann Zeta Function on Line of Abscissa 1 *by Mertens Auxiliary Function and Completion of Squares.* We are going to prove the nonvanishing of the Riemann zeta function $\zeta(z)$ on the line $\{\operatorname{Re} z = 1\}$ of abscissa 1 by using

(i) the completion of squares

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

(ii) the Mertens auxiliary function

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

From

$$\begin{aligned}\log|\zeta(z)| &= \mathrm{Re}\,\log\zeta(z)\\ &= \mathrm{Re}\,\log\prod_{p\ \mathrm{prime}}\frac{1}{1-p^{-z}}\\ &= \mathrm{Re}\sum_{p\ \mathrm{prime}}\log\frac{1}{1-p^{-z}}\\ &= \mathrm{Re}\sum_{p\ \mathrm{prime}}\sum_{n=1}^{\infty}\frac{p^{-nz}}{n}\end{aligned}$$

it follows that

$$\begin{aligned}\log|h(x)| &= 3\log|\zeta(x)| + 4\log|\zeta(x+iy)| + \log|\zeta(x+2iy)|\\ &= \sum_{p\ \mathrm{prime}}\sum_{n=1}^{\infty}\frac{p^{-nx}}{n}\mathrm{Re}\left(3+4p^{-iy}+p^{-2iy}\right)\\ &= \sum_{p\ \mathrm{prime}}\sum_{n=1}^{\infty}\frac{p^{-nx}}{n}\left(3+4\cos(y\log p)+\cos(2y\log p)\right)\\ &= \sum_{p\ \mathrm{prime}}\sum_{n=1}^{\infty}\frac{p^{-nx}}{n}2\left(1+\cos(y\log p)\right)^2 \geq 0.\end{aligned}$$

Thus

$$|h(x)| = |\zeta(x)|^3|\zeta(x+iy)|^4|\zeta(x+2iy)| \geq 1.$$

Suppose $\zeta(1+iy)$ vanishes for some $y\in\mathbb{R}$. Then $y\neq 0$, because $z=1$ is a simple pole with residue 1 for $\zeta(z)$. We get a contradiction from

$$|h(x)| = |\zeta(x)|^3|\zeta(x+iy)|^4|\zeta(x+2iy)| \geq 1$$

as $x\to 1^+$, because $|\zeta(x)|^3$ blows up of order 3 and $|\zeta(x+iy)|^4$ vanishes to order 4 whereas $|\zeta(x+2iy)|$ remains bounded. More precise description of the argument leading to the contradiction is as follows. From the vanishing of $\zeta(1+iy)$ we have

$$\lim_{x\to 1^+}\left|\frac{\zeta(x+iy)}{x-1}\right| < \infty,$$

which contradicts

$$\frac{1}{x-1} \leq \frac{|h(x)|}{x-1} = |(x-1)\zeta(x)|^3\left|\frac{\zeta(x+iy)}{x-1}\right|^4|\zeta(x+2iy)|$$

as $x \to 1^+$, because the left-hand side goes to ∞ while the right-hand side is bounded by some finite number. Here

$$\lim_{x\to 1^+} |(x-1)\zeta(x)| < \infty$$

is used, which is a consequence of the fact that $z = 1$ is a simple pole of $\zeta(z)$ with residue 1.

Meromorphic Extension of Negative of Logarithmic Derivative of Riemann Zeta Function to Neighborhood of Line of Abscissa 1. By putting together the following statements, we can conclude that $-\frac{\zeta'(z)}{\zeta(z)}$ admits a meromorphic extension to an open neighborhood of $\{\operatorname{Re} z \geq 1\}$ in $\mathbb{C}$ with 1 as the only pole and the principal part at 1 is $\frac{1}{z-1}$.

(i) The product formula

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

for the Riemann zeta function for $\operatorname{Re} z > 1$ (which comes from the unique factorization of any positive integer into a product of prime numbers) shows that $\zeta(z)$ is holomorphic nowhere zero on $\operatorname{Re} z > 1$.

(ii) The formula

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

on $\{\operatorname{Re} z > 1\}$ shows that $\zeta(z)$ can be meromorphically extended to to $\{\operatorname{Re} z > 0\}$, which is holomorphic except at $z = 1$ where it has a simple pole with residue 1.

(iii) The nonvanishing of $\zeta(z)$ on $\{\operatorname{Re} z = 1\}$ from the argument of the Mertens auxiliary function and the completion of squares shows that $-\frac{\zeta'(z)}{\zeta(x)}$ is holomorphic at every point of $\{\operatorname{Re} z = 1\} - \{1\}$.

Negative of Logarithmic Derivative of Riemann Zeta Function as Mellin Transform of Second Chebyshev Function. From the product formula

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

for the Riemann zeta function for $\operatorname{Re} z > 1$ (which comes from the unique factorization of any positive integer into a product of prime numbers), by taking the logarithmic derivative of both sides, we get

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

where $\psi(x) = \sum_{n \leq x} \Lambda(n)$ and $\Lambda(n)$ is the von Mangoldt function whose value is 0 unless $n = p^k$ for some prime number when its value is $\log p$.

We now use summation by parts to replace the use of the von Mangoldt function $\Lambda(k)$ by the use of the second Chebyshev function

$$\psi(x) = \sum_{k \leq x} \Lambda(k).$$

Recall the formula for summation by parts is

$$a_1 b_1 + \cdots + a_n b_n = s_1(b_1 - b_2) + s_2(b_2 - b_3) + \cdots + s_{n-1}(b_{n-1} - b_n) + s_n b_n,$$

where $s_k = a_1 + \cdots + a_k$. By application of summation by parts, we get

$$\sum_{k=1}^{N+1} \frac{1}{k^z} \Lambda(k) = \sum_{k=1}^{N} \psi(k) \left(\frac{1}{k^z} - \frac{1}{(k+1)^z} \right) + \psi(N+1) \frac{1}{(N+1)^z}$$

Since clearly $\psi(x) \leq x \log x$, it follows that for $\operatorname{Re} z > 1$,

$$\lim_{N \to \infty} \psi(N+1) \frac{1}{(N+1)^z} = 0$$

and

$$\sum_{k=1}^{\infty} \frac{1}{k^z} \Lambda(k) = \sum_{k=1}^{\infty} \psi(k) \left(\frac{1}{k^z} - \frac{1}{(k+1)^z} \right)$$

Thus from

$$\int_{t=k}^{k+1} t^{-(z+1)}dt = \left[-\frac{t^{-z}}{z}\right]_{t=n}^{t=n+1} = \frac{1}{z}\left(\frac{1}{k^z} - \frac{1}{(k+1)^z}\right)$$

we conclude that

$$\begin{aligned}
-\frac{\zeta'(z)}{\zeta(z)} &= \sum_{k=1}^{\infty} \frac{1}{k^z}\Lambda(k) \\
&= \sum_{k=1}^{\infty} \psi(k)\left(\frac{1}{k^z} - \frac{1}{(k+1)^z}\right) \\
&= \sum_{k=1}^{\infty} \psi(k) z \int_{t=k}^{k+1} t^{-(z+1)}dt = z\int_{t=1}^{\infty} \psi(t)t^{-(z+1)}dt.
\end{aligned}$$

In other words, the negative of the logarithmic derivative of the Riemann zeta Function is the Mellin transform of second Chebyshev function.

Linear Bound of Second Chebyshev Function from Argument of Powers of 2. We are going to verify the linear growth bound for the second Chebyshev function, that is, $\psi(x) = O(x)$. The verification is done in the following three steps.

(i) Bound of sum of log prime between n and $2n$, with the use of binomial coefficient $\binom{2n}{n}$.

(ii) Bound of sum of log prime not exceeding a power of 2, by breaking up primes not exceeding a power of 2 into intervals bounded by consecutive powers of 2.

(iii) For a number x greater than a prime p, differentiate between the case of x greater than the square of p and the contrary.

Here are the details for the three steps.

Step One. The statement for this step is

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

The key argument is that

$$\prod_{\substack{n<p\leq 2n,\\ p \text{ pime}}} p \leq (n+1)(n+2)\cdots(2n)$$

but $p \geq n$ cannot divide any of the numbers $2, 3, \cdots, n$ so that

$$\prod_{\substack{n<p\leq 2n,\\ p \text{ pime}}} p \leq \frac{(n+1)(n+2)\cdots(2n)}{2\cdot 3\cdots n} = \binom{2n}{n} < (1+1)^{2n} = 2^{2n},$$

because $\binom{2n}{n}$ is an integer and $(n+1)(n+2)\cdots(2n) = n!\binom{2n}{n}$. We get our statement

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

by taking the logarithm of

$$\prod_{\substack{n<p\leq 2n,\\ p \text{ pime}}} p < 2^{2n}.$$

A special case of this statement is that when $n = 2^m$, we have

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

In words, this statement says that the sum of the logarithm of primes between two consecutive powers of 2 is less than the larger power of 2 times $\log 2$.

Step Two. The statement for this step is

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

In words, this statement says that the sum of the logarithm of primes not exceeding a power of 2 is less than that power of 2 times $2\log 2$. This statement just follows from the special case in Step One by dividing up $[1, 2^m]$ into the disjoint union of intervals whose end-points are consecutive power of 2.

$$\sum_{\substack{p\leq 2^m,\\ p \text{ prime}}} \log p \leq \sum_{\ell=0}^{m} \sum_{\substack{2^\ell<p\leq 2^\ell,\\ p \text{ prime}}} \log p < \sum_{\ell=0}^{m} 2^\ell \log 2 = 2^{m+1}\log 2.$$

Step Three. The second Chebyshev function $\psi(x)$ is defined as the sum of $\log p$ with p prime and $p^k \leq x$ for some $k \in \mathbb{N}$. Thus for a given prime p the number of $\log p$ occurring in the sum for $\psi(x)$ is precisely the in integral part $\left\lfloor \frac{\log x}{\log p} \right\rfloor$ of $\frac{\log x}{\log p}$. The formula for $\psi(x)$ is

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

For any prime $p \leq x$, we differentiate between two cases $\left\lfloor \frac{\log x}{\log p} \right\rfloor = 1$ and $\left\lfloor \frac{\log x}{\log p} \right\rfloor > 1$. When $\left\lfloor \frac{\log x}{\log p} \right\rfloor > 1$, we have $\left\lfloor \frac{\log x}{\log p} \right\rfloor \geq 2$ and $\frac{\log x}{\log p} \geq 2$, which means that $x \geq p^2$ and $\sqrt{x} \geq p$. When we consider only those primes p with $\left\lfloor \frac{\log x}{\log p} \right\rfloor > 1$ in the sum for $\psi(x)$, we get

$$\begin{aligned}
\sum_{\substack{p \text{ prime}, \\ \left\lfloor \frac{\log x}{\log p} \right\rfloor > 1}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p &\leq \sum_{\substack{p \text{ prime}, \\ p \leq \sqrt{x}}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p \\
&\leq \sum_{\substack{p \text{ prime}, \\ p \leq \sqrt{x}}} \frac{\log x}{\log p} \log p \\
&= \sum_{\substack{p \text{ prime}, \\ p \leq \sqrt{x}}} \log x \\
&= \pi(\sqrt{x}) \log x.
\end{aligned}$$

On the other hand, When we consider only those primes p with $\left\lfloor \frac{\log x}{\log p} \right\rfloor = 1$ in the sum for $\psi(x)$, we get

$$\sum_{\substack{p \text{ prime}, \\ \left\lfloor \frac{\log x}{\log p} \right\rfloor = 1}} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p = \sum_{\substack{p \text{ prime}, \\ p \leq x}} \log p.$$

For any given $x \geq 2$ there exists a unique nonnegative integer m such that $2^m < x \leq 2^{m+1}$. From Step Two we have

$$\sum_{p \leq 2^{m+1}} \log p \leq 2^{m+2} \log 2 \leq 4x \log 2.$$

Hence

$$\sum_{\substack{p\ \text{prime},\\ \left\lfloor\frac{\log x}{\log p}\right\rfloor=1}} \left\lfloor\frac{\log x}{\log p}\right\rfloor \log p \leq 4x\log 2.$$

Putting the two cases together, we get

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

Growth Order of Second Chebyshev Function as Prime Number Function Times Logarithmic Function. The statement is that $\psi(x) \approx \pi(x)\log x$, under the assumption that $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$, so that the proof of the Prime Number Theorem $\lim_{x\to\infty}\frac{\pi(x)}{\frac{x}{\log x}} = 1$ is reduced to $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$. One direction of the comparison

$$\limsup_{x\to\infty}\frac{\psi(x)}{\pi(x)\log x} \leq 1$$

is straightforward and does not require the assumption $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$, because

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

For the other direction of the comparison

$$\limsup_{x\to\infty}\frac{\pi(x)\log x}{\psi(x)} \leq 1,$$

the trick is to cut the number of primes $p \leq x$ by breaking the counting into the following two parts for an appropriate $y < x$ and then relate the second part to $\psi(x)$ by using $\log y < \log p$ for the primes $y < p \leq x$.

(i) The number of primes $p \leq y$.

(i) The number of primes $y < p \leq x$.

We have

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

Let us investigate how we should choose y in the inequality

$$\pi(x) \leq y + \frac{\psi(x)}{\log y}$$

in order to conclude that

$$\limsup_{x \to \infty} \frac{\pi(x) \log x}{\psi(x)} \leq 1.$$

Rewrite the inequality

$$\pi(x) \leq y + \frac{\psi(x)}{\log y}$$

as

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

Since we assume as known $\lim_{x \to \infty} \frac{\psi(x)}{x} = 1$, to get our conclusion an obvious way is to have

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

and

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

We can achieve both by setting $y = \frac{x}{(\log x)^\gamma}$ for any $\gamma > 1$, for example, $\gamma = 2$.

Proof of Prime Number Theorem from Applying to Second Chebyshev Function Mellin Transform Version of Tauberian Theorem. The proof of the Prime Number Theorem now is a consequence of the following steps which we have obtained above.

Step One. The second Chebyshev function $\psi(x) = \sum_{n \leq x} \Lambda(n)$ is clearly nondecreasing and piecewise continuous. We have checked that its growth order is at most linear, *i.e.*, $\psi(x) = O(x)$ as $x \to \infty$.

Step Two. We have also checked that the Mellin transform of $\psi(x)$ is $-\frac{\zeta'(z)}{\zeta(z)}$ which is meromorphic on an open neighborhood of $\{\operatorname{Re} z \geq 1\}$ with 1 as the only pole and $\frac{1}{z-1}$ as its principal part.

Step Three. We can now apply the Mellin Transform Version of Tauberian Theorem to

$$f(x) = \psi(x), \quad g(z) = -\frac{\zeta'(z)}{\zeta(z)}, \quad \text{and} \quad c = 1$$

to conclude that

$$\lim_{x \to \infty} \frac{\psi(x)}{x} = 1.$$

Step Four. We have also checked that

$$\lim_{x \to \infty} \frac{\psi(x)}{\pi(x) \log x} = 1$$

so that we can obtain the conclusion of the Prime Number Theorem which is

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

Wiener's Approach to Tauberian Theory. On $\mathbb{R}$ we consider a special kind of family weights for taking average. One starts with a single function $h(t)$ on $\mathbb{R}$ which is absolutely integrable, say piecewise continuous (actually after the introduction of Lebesgue theory, just Lebesgue measurable). The family indexed by $x \in \mathbb{R}$ is just the family of translates $h(t-x)$ of $h(t)$ to the right by x. For a bounded, piecewise continuous function $f(x)$ on $\mathbb{R}$, we can form the function

$$x \mapsto \int_{t\in\mathbb{R}} f(t)h(x-t)dt,$$

from the weights indexed by the variable x, which of course is just the convolution $f * h$ of f and h. If we would like to talk about weighted average, we can require that $\int_{t=0}^{\infty} h(t)dt = 1$, but it is not important.

The principle of Tauberian theorems is the relation between convergence in using two different families of weights when an appropriate Tauberian condition is fulfilled. Instead of just one single $h(t)$, we take another function $g(t)$ so that we have two different families, one defined by $h(t)$ and the other defined by $g(t)$. We ask under what Tauberian condition is the convergence of the function

$$x \mapsto \int f(t)h(x-t)dt$$

to A as $x \to x_0$ implies the convergence of the function

$$x \mapsto \int f(t)g(x-t)dt$$

to A as $x \to x_0$. Since we want to forego the condition $\int h(t)dt = 1$, as explained above, we will always choose $A = 0$. We use $x \to \infty$ as $x \to x_0$. The Tauberian condition which Wiener obtained in 1932 is that the Fourier transform of h has no real zeroes. That is,

$$\int_{x\in\mathbb{R}} h(x)e^{i\xi x}dx \neq 0 \quad \text{for all } \xi \in \mathbb{R}.$$

The conclusion of Wiener's Tauberian theorem is that under such Tauberian condition, if the limit of the function

$$x \mapsto \int f(t)h(x-t)dt$$

as $x \to \infty$ is equal to

$$A\int_{\mathbb{R}} h(x)dx$$

and if $g(x)$ is another piecewise continuous function absolutely integrable function on $\mathbb{R}$, then the limit of the function

$$x \mapsto \int f(t)g(x-t)dt$$

as $x \to \infty$ is equal to

$$A \int_{\mathbb{R}} g(x) dx.$$

Again after the introduction of Lebesgue's theory of integration, the condition of piecewise continuity for f, g, h (which is used to guarantee the local integrability of $f(t)h(x-t)$ and $f(t)g(x-t)$ as a function of t on any interval of finite length in $\mathbb{R}$) can be replaced by f, g, h being Lebesgue measurable.

The idea of proof is replace h by its translates and take linear combination and then approximate any given g by linear combinations of translates of h. As a matter of fact the proof is close to Karamata's method which first replaces x^n by x^{kn} (for fixed k and variable n) and then take a linear combination and approximate some function defined by characteristic functions by such linear combinations. The replacement of x^n by x^{kn} (for fixed k and variable n) is the same as performing a translation $n \to kn$ with respect to the group law of multiplication for integers.

Wiener's Tauberian theorem	Littlewood's Tauberian theorem
$f(t)$	a_n
$t \in \mathbb{R}$	$0 \leq n < \infty$
$\int_{t\in\mathbb{R}} f(t)h(x-t)dt$ (h uniformly bounded on $\mathbb{R}$)	$\sum_{n=0}^{\infty} a_n x^n$
$x \in \mathbb{R}$	$0 \leq x < 1$
Tauberian Condition: $\int_{t\in\mathbb{R}} f(t)h(x-t)dt \to 0$ as $x \to \infty$	*Tauberian Condition:* $a_n = O(\frac{1}{n})$ (or $\text{Karamata's } \inf_{n\in\mathbb{N}} na_n > -\infty$)
Conclusion: $\lim_{x\to\infty} \int_{t\in\mathbb{R}} f(t)g(x-t)dt \to 0$ for all g in $L^1(\mathbb{R})$	*Conclusion:* $\lim_{N\to\infty} \sum_{n=0}^{N} a_n$ exists and equals $\lim_{x\to 1^-} \sum_{n=0}^{\infty} a_n x^n$
$x \to \infty$	$N \to \infty$
Wiener's method of proof: (i) Replace h by translate $x \mapsto h(x-y)$ by y (ii) Take linear combination $x \mapsto \sum_j b_j h(x-y_j)$ (iii) Approximate $g(x)$ by $x \mapsto \sum_j b_j h(x-y_j)$ with approximate choice of b_j's	*Karamata's method of proof:* (i) Replace x^n by translate x^{kn} by k in multiplicative group law (ii) Take linear combination $x^n \mapsto \sum_j b_j (x^n)^k$ (iii) Approximate some function from characteristic function by $x^n \mapsto \sum_j b_j (x^n)^k$ with approximate choice of b_j's

There is a Fourier series version of Wiener's Tauberian theorem instead of the Fourier transform version. There are three equivalent formulations for it.

In order to state these formulations, we introduce the vector space $\ell_1(\mathbb{Z})$ over $\mathbb{C}$ which consists of all $(a_n)_{n\in\mathbb{Z}}$ with $\sum_{n\in\mathbb{Z}} |a_n| < \infty$. We introduce multiplication into the $\mathbb{C}$-vector space $\ell_1(\mathbb{Z})$ by using convolution so that the product of $\mathbf{a} = (a_n)_{n\in\mathbb{Z}}$ and $\mathbf{b} = (b_n)_{n\in\mathbb{Z}}$ is $\mathbf{c} = (c_n)_{n\in\mathbb{Z}}$ with $c_n = \sum_{k\in\mathbb{Z}} a_k b_{n-k}$. We use the notation $\mathbf{c} = \mathbf{a} * \mathbf{b}$ to denote the convolution $\mathbf{c}$ of $\mathbf{a}$ and $\mathbf{b}$. Note that when $\mathbf{a} \in \ell_1(\mathbb{Z})$ and $\mathbf{t} = (t_n)_{n\in\mathbb{Z}}$ is only a bounded sequence and not an element of $\ell_1(\mathbb{Z})$, we can still form the convolution $\mathbf{a} * \mathbf{t}$ and get a sequence, but the sequence $\mathbf{a} * \mathbf{t}$ is in general not an element of $\ell_1(\mathbb{Z})$ and is only a bounded sequence.

The ring $\ell_1(\mathbb{Z})$ is naturally isomorphic to the ring of all absolutely convergent Fourier series $\sum_{n\in\mathbb{Z}} a_n e^{inx}$ on $\mathbb{R}$ where addition and multiplication are defined as the usual addition and multiplication for functions. The absolutely convergent Fourier series $\sum_{n\in\mathbb{Z}} a_n e^{inx}$ corresponding to the element $(a_n)_{n\in\mathbb{Z}}$ of $\ell_1(\mathbb{Z})$ is called its *Fourier transform.*

The three equivalent formulations of Fourier series version of Wiener's Tauberian theorem are as follows.

(1) For an element $\mathbf{a} = (a_n)_{n\in\mathbb{Z}}$ of $\ell_1(\mathbb{Z})$, the span of its translates (by multiplication) is dense in $\ell_1(\mathbb{Z})$ if and only if its Fourier transform $\hat{\mathbf{a}}$ has no real zeroes.

(2) If the Fourier transform of an element $\mathbf{a} = (a_n)_{n\in\mathbb{Z}}$ of $\ell_1(\mathbb{Z})$ does not have any real zeroes and if for some bounded sequence $\mathbf{t}$ the convolution $\mathbf{a} * \mathbf{t}$ as a bounded sequence approaches zero at infinity, then the convolution $\mathbf{b}*\mathbf{t}$ as a bounded sequence also approaches zero at infinity for any element $\mathbf{b} = (b_n)_{n\in\mathbb{Z}}$ of $\ell_1(\mathbb{Z})$.

(3) If a function f on $\mathbb{R}$ with period 2π whose Fourier series is absolutely convergent has not zeroes on $\mathbb{R}$, then its reciprocal $\frac{1}{f}$ has absolutely convergent Fourier series.

The first formulation is related to Karamata's method of proof, which uses approximation by linear combinations of translates of the kernel (or the weighted average). The second formulation is along the lines of the historic Tauberian theorems involving taking limits for two different kernels (or ways of taking weighted averages). The third formulation is assigned as a homework problem.

Israel Moiseevich Gelfand in 1941 introduced the language of commutative Banach algebras, multiplicative linear functionals, maximal ideal space, and the Gelfand representation to interpret the Tauberian theory of Wiener. Gelfand's work led to fundamental developments in the theory of harmonic analysis and Fourier analysis on locally compact abelian groups.