2016

A Measure Theoretic Approach to Problems of Number Theory with Applications to the Proof of the Prime Number Theorem

Russell Lee Jahn
Minnesota State University Mankato

MINNESOTA STATE UNIVERSITY

MASTER'S THESIS

A Measure Theoretic Approach to Problems of Number Theory with Applications to the Proof of the Prime Number Theorem

Author:
Russell Jahn

Supervisor:
Dr. Wook Kim

A thesis submitted in fulfilment of the requirements
for the degree of Master of Arts

in the

Department of Mathematics
Minnesota State University
Mankato, Minnesota

May, 2016

Endorsement

March 30, 2016

A Measure Theoretic Approach to Problems of Number Theory with Applications to the Proof of the Prime Number Theorem

Russell Jahn

This thesis has been examined and approved by the following members of the student's committee.

Wook Kim, Ph.D.

Brian F. Martensen, Ph.D.

Dan W. Singer, Ph.D.

MINNESOTA STATE UNIVERSITY

Abstract

Department of Mathematics

Master of Arts

A Measure Theoretic Approach to Problems of Number Theory with Applications to the Proof of the Prime Number Theorem

by Russell Jahn

In this paper we demonstrate how the principles of measure theory can be applied effectively to problems of number theory. Initially, necessary concepts from number theory will be presented. Next, we state standard concepts and results from measure theory to which we will need to refer. We then develop our repertoire of measure theoretic machinery by constructing the needed measures and defining a generalized version of the multiplicative convolution of measures. A suitable integration by parts formula, one that is general enough to handle various combinations of measures, will then be derived. At this juncture we will be ready to demonstrate the effectiveness of measure theory tactics on number theory problems. Specifically, due to its highly receptive nature to measure theoretic techniques, the prime number theorem will be proved. First, we prove the theorem by what is termed an "elementary" method. Secondly, the Riemann zeta function is employed to enable us to give a much shorter proof. In both cases, we borrow the ideas from several sources and apply them to our proofs. The approach taken in this paper, however, is distinctive in the sense that the driving force of the proofs is measure theory. Although there is greater overhead in learning the appropriate material used in this approach, we argue that once this material is understood it can be beneficially applied to problems of suitable complexity.

Acknowledgements

First and foremost, I would like to thank my thesis advisor, Dr. Wook Kim, for his invaluable insight, guidance and suggestions and his devotion to this project. I would also like to thank my instructors at Minnesota State University: Dr. Brian Martensen, Dr. Ruijun Zhao, Dr. Mezbahur Rahman, and Dr. Chia Tung. A special thanks goes to Dr. Dan Singer for the inspiration and motivation I gained from him during my years at Minnesota State.

Most of all, I would like to thank my wonderful family – Reeta, Ryan, Misty, Robert, and Matthew – for their support throughout my academic pursuits.

Contents

List of Symbols

$\mathbb{C}$	set of complex numbers
$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}^+$	set of positive real numbers
$\mathbb{R}^*$	set of extended real numbers
$\mathcal{P}$	set of primes
$\Re(s)$	real part of complex number s
$\Im(s)$	imaginary part of complex number s
(a, b)	greatest common divisor of a and b
$[x]$	greatest integer of x
$\mathscr{A}$	set of arithmetic functions
$\mathscr{A}_{\mathscr{M}}$	set of multiplicative functions
$o(g)$	little-o of g
$o_y(g)$	little-o of g with parameter y
$O(g)$	Big-O of g
$O_y(g)$	Big-O of g with parameter y
$0(n)$	0 for all n
$e(n)$	$\begin{cases} 1, & n = 1; \\ 0, & n > 1. \end{cases}$
$1_S(n)$	$\begin{cases} 1, & n \in S; \\ 0, & n \notin S \end{cases}$ for $S \subset \mathbb{N}$.
$1_{\mathcal{P}}(n)$	$\begin{cases} 1, & n \text{ prime}; \\ 0, & n \text{ not prime}. \end{cases}$

$1(n)$	$1_{\mathbb{N}}(n) = 1$ for all n.
$\mu(n)$	$\begin{cases} 1, & n = 1 \text{ or } n \text{ is the product of an even number of distinct primes;} \\ -1, & n \text{ is the product of an odd number of distinct primes;} \\ 0, & \text{otherwise.} \end{cases}$ μ is called the Möbius function.
$\lvert\mu\rvert(n)$	$\begin{cases} 1, & n = 1 \text{ or } n \text{ is the product of distinct primes;} \\ 0, & \text{otherwise.} \end{cases}$
$\Lambda(n)$	$\begin{cases} \log p, & n = p^\alpha \text{ for } p \text{ prime and } \alpha \in \mathbb{N}; \\ 0, & \text{otherwise.} \end{cases}$
$f * g$	multiplicative convolution of arithmetic functions f and g
$\in$	is a member of
$\notin$	is not a member of
$\emptyset$	empty set
$\subset$	is a subset of
$\supset$	is a superset of
$\cup$	union
$\cap$	intersection
A^c	complement of set A
$A \setminus B$	set A minus set B
$f(A)$	image of A under f
$f^{-1}(A)$	inverse image of A under f
tA	$\{ta : a \in A\}$ for set A
$\{a_i\}_{i=1}^n$	finite sequence of a_i
$\{a_i\}_i$	countable sequence of a_i
$\bigcup_{i=1}^n A_i$	finite union of sets A_i
$\bigcup_i A_i$	countable union of sets A_i

$\bigcap_{i=1}^{n} A_i$	finite intersection of sets A_i
$\bigcap_{i} A_i$	countable intersection of sets A_i
$\sum_{d\mid n} f(n)$	the sum of addends $f(d)$ for all d that divide n
$\sum_{i=1}^{n} a_i$	finite sum of a_i
$\sum_{i} a_i$	countable sum of a_i
$\Rightarrow$	implies
$\iff$	if and only if
$\to$	to or approaches
$\downarrow$	approaches from the right
$\uparrow$	approaches from the left
$:=$	is defined as
$=:$	is the definition of
$\lvert a\rvert$	absolute value if $a \in \mathbb{R}$; modulus if $a \in \mathbb{C}$
a^-	$a - \epsilon$ for sufficiently small $\epsilon > 0$
a^+	$a + \epsilon$ for sufficiently small $\epsilon > 0$
$\sup$	least upper bound
$\inf$	greatest lower bound
$\lim$	limit
$\limsup$	limit superior
$\liminf$	limit inferior
$\sim$	is asymptotic to
$f\colon \Omega \to \Pi$	mapping of f with domain Ω and image in Π
$f \circ g$	composition of f of g
Ω	arbitrary set
$\mathcal{P}(\Omega)$	power set of Ω
$\mathcal{A}$	arbitrary σ-algebra
$\sigma(\mathcal{C})$	σ-algebra generated by set $\mathcal{C}$
χ_A	characteristic function with respect to set A
$\mathcal{B}$	Borel sets of $\mathbb{R}$
$\mathcal{B}^+$	Borel sets of $\mathbb{R}^+$

$\mathcal{B}^*$	Borel sets of $\mathbb{R}^*$
$\mu,\ \nu,\ \rho$	arbitrary measures (μ is also Möbius function)
$\mu(A)$	value of the measure of set A with respect to measure μ
$(\Omega, \mathcal{A}, \mu)$	arbitrary measure space
$(\Omega, \mathcal{A})$	arbitrary measurable space
$(\mathbb{R}^+, \mathcal{B}^+)$	measurable space of $\mathbb{R}^+, \mathcal{B}^+$
a.e.	almost everywhere
μ-a.e.	μ almost everywhere for arbitrary measure μ
$\int_a^b f(t)dt$	Riemann integral of f from a to b
$\int_A f(t)d\mu(t)$ or $\int_A f d\mu$	abstract Lebesgue integral of arbitrary f over A with respect to μ
$f^+(x)$	$\max\{f(x), 0\}$ for arbitrary f
$f^-(x)$	$-\min\{f(x), 0\}$ for arbitrary f
$\mathcal{L}^1(A, \mu)$	set of all extended real valued Lebesgue integrable functions over A with respect to μ on arbitrary measure space $(\Omega, \mathcal{A}, \mu)$
$\mathcal{L}^1(\Omega, \mu)$	set of all extended real valued Lebesgue integrable functions over Ω with respect to μ on arbitrary measure space $(\Omega, \mathcal{A}, \mu)$
$\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$	set of all extended real valued Lebesgue integrable functions over Ω with respect to μ on arbitrary measure space $(\Omega, \mathcal{A}, \mu)$
$\mathcal{L}^1(\mu)$	set of all extended real valued Lebesgue integrable functions over Ω with respect to μ on arbitrary measure space $(\Omega, \mathcal{A}, \mu)$
$f(n),\ g(n),\ h(n),\ r(n)$	arbitrary arithmetic functions
$F(x),\ G(x),\ H(x),\ R(x)$	arbitrary Type 1 summatory functions induced by $f,\ g,\ h,\ r$, respectively
$\lambda_F,\ \lambda_G,\ \lambda_H,\ \lambda_R$	arbitrary $\mathcal{B}^+$ measures induced by Type 1 summatory functions F, G, H, R, respectively

$N(x)$	$\sum_{n\leq x} 1(n)$
$Q(x)$	$\sum_{n\leq x} \|\mu\|(n)$
$(Q-M)(x)$	$\sum_{n\leq x} (\|\mu\|-\mu)(n)$
$\psi(x)$	$\sum_{n\leq x} \Lambda(n)$
$\pi(x)$	$\sum_{n\leq x} 1_{\mathcal{P}}(n)$
$\vartheta(x)$	$\sum_{n\leq x} \log n \cdot 1_{\mathcal{P}}(n)$
$\chi(x)$	$\sum_{n\leq x} e(n)$
$M(x)$	$\sum_{n\leq x} \mu(n)$
λ	Lebesgue measure restricted to $\mathcal{B}^+$
$\lambda_{\|_A}$	Lebesgue measure restricted to A
λ_0	the null measure ($\lambda_0(E)=0$ for all $E\in\mathcal{B}^+$)
λ_n	$\int \chi_{[n,\infty)}(t)d\lambda(t)$ for $n\in\mathbb{N}$
λ_1	$\int \chi_{[1,\infty)}(t)d\lambda(t) = \lambda_{\|_{[1,\infty)}}$
λ_2	$\int \chi_{[2,\infty)}(t)d\lambda(t)$
λ_N	$\mathcal{B}^+$ measure induced by $N(x)$
λ_Q	$\mathcal{B}^+$ measure induced by $Q(x)$
λ_{Q-M}	$\mathcal{B}^+$ measure induced by $(Q-M)(x)$
λ_ψ	$\mathcal{B}^+$ measure induced by $\psi(x)$
λ_π	$\mathcal{B}^+$ measure induced by $\pi(x)$
λ_ϑ	$\mathcal{B}^+$ measure induced by $\vartheta(x)$
λ_χ	$\mathcal{B}^+$ measure induced by $\chi(x)$
δ_1	$\mathcal{B}^+$ measure induced by $\chi(x)$
λ_M	set function on $\mathcal{B}^+$ induced by $M(x)$
$L^n\mu$	$\int \chi_{[1,\infty)}(t)\log^n t\, d\mu(t)$ for $n\in\mathbb{N}$
$L\lambda_N$	$\int \chi_{[1,\infty)}(t)\log t\, d\lambda_N(t)$
$L\lambda_\psi$	$\int \chi_{[1,\infty)}(t)\log t\, d\lambda_\psi(t)$
$L\lambda_1$	$\int \chi_{[1,\infty)}(t)\log t\, d\lambda_1(t)$
$T^n\mu$	$\int \chi_{[1,\infty)}(t)t^n d\mu(t)$ for $n\in\mathbb{Z}$
$T^{-1}\lambda_1$	$\int t^{-1}d\lambda_1(t)$

$\nu \ll \mu$	arbitrary measure ν is absolutely continuous with respect to arbitrary measure μ
$d\nu/d\mu$	Radon-Nikodym derivative of arbitrary measure ν with respect to arbitrary measure μ
$\times$	Cartesian product
$\Gamma \times \Lambda$	arbitrary product space of Γ with Λ
$\mathcal{S} \times \mathcal{T}$	arbitrary product σ-algebra of $\mathcal{S}$ with $\mathcal{T}$
$(\Gamma \times \Lambda, \mathcal{S} \times \mathcal{T})$	arbitrary product measurable space
$\mathbb{R} \times \mathbb{R}$	product space of $\mathbb{R}$ with $\mathbb{R}$
$\mathbb{R}^+ \times \mathbb{R}^+$	product space of $\mathbb{R}^+$ with $\mathbb{R}^+$
$\mathbb{R}^* \times \mathbb{R}^*$	product space of $\mathbb{R}^*$ with $\mathbb{R}^*$
$\mathcal{B} \times \mathcal{B}$	Borel sets of $\mathbb{R} \times \mathbb{R}$
$\mathcal{B}^+ \times \mathcal{B}^+$	Borel sets of $\mathbb{R}^+ \times \mathbb{R}^+$
$\mathcal{B}^* \times \mathcal{B}^*$	Borel sets of $\mathbb{R}^* \times \mathbb{R}^*$
$(\mathbb{R}^+ \times \mathbb{R}^+, \mathcal{B}^+ \times \mathcal{B}^+)$	product measurable space of $\mathbb{R}^+ \times \mathbb{R}^+, \mathcal{B}^+ \times \mathcal{B}^+$
$\mu \times \nu$	arbitrary product measure of μ with ν
$\mu * \nu$	multiplicative convolution of arbitrary measures μ and ν
$\mathcal{E}$	$\{E \in \mathcal{B}^+ : \sup E < \infty\}$
$\mathscr{M}$	$\{\sigma$-finite $\mathcal{B}^+$ measures $\mu : \mu((0,1)) = 0$ and $\mu(E) < \infty$ for $E \in \mathcal{E}\}$
$\mathscr{L}(\mathcal{E})$	$\{$linear combinations over $\mathbb{R}$ of measures in $\mathscr{M}$ restricted to $\mathcal{E}\}$
$\widetilde{f}(t)$	$\begin{cases} f(t), & t \in \mathbb{N}; \\ 0, & t \notin \mathbb{N} \end{cases}$ for arbitrary arithmetic f
$p,\ q,\ p_i,\ q_i$	primes
UPF	unique prime factorization
PNT	prime number theorem
ξ	$(\lambda_\psi - \lambda_1 - \delta_1) * T^{-1}\lambda_1$
$\Xi(x)$	$\xi((0,x]) = \int_{[1,x]} \frac{\psi(t) - t}{t} d\lambda_1(t)$

$\lvert\Xi\rvert(x)$	$\lvert\Xi(x)\rvert$
α	$\limsup_{x\to\infty} \frac{\lvert\Xi\rvert(x)}{x}$
S	$\{p^e : p \text{ prime}, e \geq 1, e \in \mathbb{N}\}$
R	$\{x : \Xi(x) \text{ changes sign}\}$
$\widehat{\nu}$	Mellin transform associated with ν
$\sigma_c(\widehat{\nu})$	abscissa of convergence of the Mellin transform associated with ν
e	base of natural logarithm
i	the imaginary number i
π	the number π
γ	Euler's constant

Note that some symbols are used for more than one application.

Chapter 1

Introduction

The purpose of this paper is to demonstrate how measure theory can be effectively applied to problems of number theory and to show that in certain situations this application of measure theory can make methodical the analysis of complicated expressions. An example would be, from Chapter 3, the evaluation of what essentially reduces to

$$\begin{aligned}
&-\int_1^x \frac{1}{u}\left(\int_1^u \frac{1}{t}\left(\sum_{n\le u/t} \log n\cdot\Lambda(n) - \frac{u}{t}\log\left(\frac{u}{t}\right) + \frac{u}{t} - 1\right) dt\right.\\
&\qquad\qquad\left. + \sum_{n\le u}\left(\psi\left(\frac{u}{n}\right) - \frac{u}{n}\right)\Lambda(n)\right) du\\
&+ (\log x)\left(\int_1^x \frac{1}{t}\left(\sum_{n\le x/t} \log n\cdot\Lambda(n) - \frac{x}{t}\log\left(\frac{x}{t}\right) + \frac{x}{t} - 1\right) dt\right.\\
&\qquad\qquad\left. + \sum_{n\le x}\left(\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right)\Lambda(n)\right)\\
&- \sum_{m\le x}\left(\int_1^{\frac{x}{m}} \frac{1}{t}\left(\sum_{n\le x/(mt)} \log n\cdot\Lambda(n) - \frac{x}{mt}\log\left(\frac{x}{mt}\right) + \frac{x}{mt} - 1\right) dt\right.\\
&\qquad\qquad\left. + \sum_{n\le x/m}\left(\psi\left(\frac{x}{mn}\right) - \frac{x}{mn}\right)\Lambda(n)\right)\Lambda(m).
\end{aligned}$$

where

$$\Lambda \text{ is the von Mangoldt function } \Lambda(n) = \begin{cases} \log p, & n = p^\alpha,\ \ p \text{ prime}, \alpha\in\mathbb{N}; \\ 0, & \text{otherwise.} \end{cases}$$

and

$$\psi(x) = \sum_{n\le x} \Lambda(n).$$

This is possible to do with standard methods, of course, but it would be a daunting task to be sure and something that would strike fear in even the most seasoned mathematician. However, by recasting the above expression in the context of measures and applying the techniques we will develop in Chapter 2, evaluating this will become methodical and straightforward. Make no mistake, though. This ease in analysis does not come without a price. There is the upfront cost of learning the measure theoretic techniques to the point where they can be applied instinctively. Once this is done, we claim, the analysis of such expressions will become easily manageable.

Initially, measure was the generalization to arbitrary sets on real spaces of familiar concepts such as length, area, volume, etc. During the last years of the 19th century and early part of the 20th century, É. Borel (1871-1956) and H. Lebesgue (1875-1941) were largely responsible for the development of what has become to be known as measure theory. Among his many advancements to the theory, Lebesgue defined the "Lebesgue integral", of which we will make great use, and derived its properties. Gradually, measure theory was generalized from real spaces to abstract spaces and, today, measure theory has applications to branches of analysis, probability, geometry, physics, and more.

Chapter 2 is dedicated to the development of the measure theoretic instruments that we will require in our demonstration. The first two sections will simply be a listing of notation, number theory concepts, and measure theory concepts that we will need to access in order to develop the necessary tools. The remainder of the chapter will focus on constructing these tools, among them being: (1) the needed Borel measures, (2) a generalized multiplicative convolution of measures, (3) a commutative algebra with identity over the reals of set functions with the binary operations of addition and generalized multiplicative convolution, and (4) an accommodating integration by parts formula. Numerous examples will be given to reinforce the understanding of the application of these concepts. These examples will be referenced frequently in the subsequent chapters.

Although not an exhaustive test of the potential benefits that measure theory can impart, we choose for our demonstration the proof of the prime number theorem, or PNT, because of its receptiveness to measure theory tactics. This theorem states that the number of primes less than or equal to a given number is asymptotically equal to that number divided by the natural logarithm of that number. Mathematically, we write this as

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

or

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

where $\pi(x)$ is the number of primes less than or equal to x. This relationship between $\pi(x)$ and $\frac{x}{\log x}$ was first hypothesized by C. Gauss (1777-1855) in 1792 and A. Legendre (1752-1833) in 1798 [3]. A century passed before the first proofs occurred, however. These were accomplished by J. Hadamard (1865-1963) and C. de la Vallée Poussin (1866-1962) in 1896. Quite lengthy, their proofs required the use of complex analysis and the properties of Riemann's zeta function. During the succeeding half century, this type of proof was considerably improved and shortened, most notably by the discovery of the Wiener-Ikehara theorem in 1931. Not until 1949 was a way found to prove the theorem in what is coined an "elementary" manner, meaning that complex analysis is not utilized. A. Selberg (1917-2007) and P. Erdős (1913-1996) accomplished this at roughly the same time; in fact, there was a mildly contentious disagreement as to whom the credit should be given for being the first [3].

We will prove the theorem while borrowing elements and making modifications from several proofs. In Chapter 3 we will do an "elementary" proof. As we shall see, the proof is far from being elementary in the literal sense of the word. This proof will provide numerous opportunities to employ the measure theoretic techniques developed in Chapter 2. A zeta function dependent proof will be the substance of Chapter 4. Although not nearly as fertile in places to interject these techniques as the "elementary" proof, it nonetheless is a vehicle for demonstrating the effective deployment of measure theory on a number theory problem. We feel the use of measure theory and the modifications we make to previous proofs make our topic original enough to warrant the writing of this thesis.

Lastly, we will make some concluding remarks to close out the paper. For the reader's convenience, an index and a list of symbols has been provided to assist with navigation.

Chapter 2

Development of Number and Measure Theoretic Concepts

We shall require concepts from both number thory and measure theory in the upcoming proofs of the PNT. Chapter 2 is dedicated to the development of these concepts. This is a key component of the paper and is vital to the understanding of our ultimate goal: proofs of the PNT. Thus, this development is presented here rather than relegated to the back pages of an appendix.

In Section 2.1, we list the relevant material from number theory. It should be noted that much of this material is only displayed in terms of its results. That is, the proofs are not necessarily given in this text. Proofs are standard and may be found in a quality number theory text such as [2]. More standard number theory results will be presented as we require them in Chapters 3 and 4.

In Sections 2.2 through 2.5, much of the relevant measure theory material is also simply listed. However, a large portion of the material is fully fleshed out as it is specific to the paper and is less likely to be found, or unable to be found due to originality, in the literature. In other words, if a concept has a proof omitted, it is considered a standard result and the proof can be found in a quality text such as [5]. If a concept has a proof attached, it is considered to be a result that is specific enough to this paper to warrant the inclusion of a proof. We take this avenue rather than using phrases such as, "It can be shown that..."; or simply citing references that prove a result that is similar, but not exactly the same, as the one in question. We feel this approach will make our arguments the most thorough and clearest possible. Of course, all original ideas requiring justification will have the proofs included.

2.1 Basic Material from Number Theory

For easy reference, we here state the notation, functions, definitions, theorems, examples and identities from number theory that will be relevant in our discussion. It is assumed the reader has a working knowledge of this subject. Proofs may be found in a quality number theory text such as [2].

As is customary, the set of natural numbers will be denoted as $\mathbb{N}$; the set of integers as $\mathbb{Z}$; the set of real numbers as $\mathbb{R}$; the set of positive real numbers as $\mathbb{R}^+$; the set of complex numbers as $\mathbb{C}$; and the set of primes as $\mathcal{P}$.

The real part of a complex number s will be denoted as $\Re(s)$. The imaginary part of a complex number s will be denoted as $\Im(s)$.

The greatest common divisor of two natural numbers a and b will be denoted as (a, b).

A function f is called ***arithmetic*** if $f\colon \mathbb{N} \to \mathbb{C}$. The set of all arithmetic functions will be denoted as $\mathscr{A}$. An arithmetic function g is called ***multiplicative*** if g is not identically 0 and $g(ab) = g(a)g(b)$ for $a, b \in \mathbb{N}$ and whenever $(a, b) = 1$. The set of all multiplicative functions will be denoted as $\mathscr{A}_{\mathscr{M}}$. An arithmetic function h is called ***completely multiplicative*** if h is not identically 0 and $h(ab) = h(a)h(b)$ for any $a, b \in \mathbb{N}$.

Consider f and $g \in \mathscr{A}$. If $f(n) = g(n)$ for all $n \in \mathbb{N}$ we write $f = g$. Similary for $f < g$, $f \leq g$, etc. We define $(f + g)(n) = f(n) + g(n)$ and $(\alpha f)(n) = \alpha f(n)$ for $\alpha \in \mathbb{C}$. It is easy to see that $\mathscr{A}$ is a vector space over $\mathbb{C}$ with respect to addition and scalar multiplication.

The following is a list of arithmetic functions we will encounter. These functions will be the basis for the measures we will construct in Section 2.3.

Example 2.1.

$0(n) = 0$ for all n.

$$e(n) = \begin{cases} 1, & n = 1; \\ 0, & n > 1. \end{cases}$$

$$1_S(n) = \begin{cases} 1, & n \in S; \\ 0, & n \notin S \end{cases} \text{ for arbitrary set } S \subset \mathbb{N}.$$

$$1_{\mathcal{P}}(n) = \begin{cases} 1, & n \text{ prime;} \\ 0, & n \text{ not prime.} \end{cases}$$

$1(n) = 1_{\mathbb{N}}(n) = 1$ for all n.

$$\mu(n) = \begin{cases} 1, & n = 1 \text{ or } n \text{ is the product of an even number of distinct primes;} \\ -1, & n \text{ is the product of an odd number of distinct primes;} \\ 0, & \text{otherwise.} \end{cases}$$

μ is called the Möbius function.

$$|\mu|(n) = \begin{cases} 1, & n = 1 \text{ or } n \text{ is the product of distinct primes;} \\ 0, & \text{otherwise.} \end{cases}$$

$$\Lambda(n) = \begin{cases} \log p, & n = p^{\alpha} \text{ for } p \text{ prime and } \alpha \in \mathbb{N}; \\ 0, & \text{otherwise.} \end{cases}$$ □

We will see in Section 2.4 that the function defined next is simply a particular case of a more general concept.

Definition 2.2. Define $f * g(n) := \sum_{ij=n} f(i)g(j) = \sum_{d|n} f(d)g(\frac{n}{d}) = \sum_{d|n} f(\frac{n}{d})g(d)$ the ***multiplicative convolution*** of two arithmetic functions f and g for $n \in \mathbb{N}$.

We notice that the multiplicative convolution of two arithmetic functions is another arithmetic function. In fact, $\mathscr{A}$ is not only a vector space over $\mathbb{C}$ with respect to addition and scalar multiplication, it is much more.

Theorem 2.3. *$\mathscr{A}$ is a commutative algebra with identity over $\mathbb{C}$ with respect to addition and multiplicative convolution. In particular, for f, g and $h \in \mathscr{A}$ and $\alpha \in \mathbb{C}$,*

$f + g \in \mathscr{A}$.

$\alpha f \in \mathscr{A}$.

$f * g \in \mathscr{A}$.

$f * e = f$.

$f * g = g * f$.

$f * (g * h) = (f * g) * h$.

$\alpha(f * g) = (\alpha f) * g = f * (\alpha g)$.

$f * (g + h) = f * g + f * h$.

Proof. Omitted. ■

The last theorem in this section has far reaching consequences in number theory. Its result will be a vital component throughout this paper.

Theorem 2.4. $\mu * 1 = e$.

Proof. Omitted. ■

We close this section with a definition on the expression of estimates.

Definition 2.5. Let f be a real or complex valued function on a set $A \subset \mathbb{R}$.

a) If there exists $M > 0$ such that $|f| \leq M|g|$ on A, then we write $f = O(g)$ on A. If, in addition, f depends on a parameter y, we write $f = O_y(g)$ on A.

b) If $A = [a, \infty)$ for some suitable a and $\lim_{x \to \infty} f(x)/g(x) = 0$, then we write $f = o(g)$. If, in addition, f depends on a parameter y, we write $f = o_y(g)$.

c) We define $O(\log^k x) = O(1)$ when $x = 1$ to avoid the cumbersome $O(\log^k ex)$ notation.

2.2 Basic Material from Measure Theory

We now commence the task of developing the necessary measure theoretic instruments to prove the PNT. This section is largely devoted to simply listing standard measure theory concepts and results to which we will refer in the subsequent sections. Consequently, very minimal narrative is included. Much of the notation we will be using will be defined in this section, as well. These standard results and notation are retrieved almost entirely from [5]. Readers with a working knowledge of measure theory may wish to proceed directly to Section 2.3.

In Section 2.1 we defined the symbols we will use for certain sets. We now add to that list by denoting the set of extended real numbers as $\mathbb{R}^*$. That is, $\mathbb{R}^* = \mathbb{R} \cup \{\pm\infty\}$. Furthermore, we extend the ordering and operations of $\mathbb{R}$ to $\mathbb{R}^*$ by the following definitions:

$-\infty < \infty$.

$-\infty < x < \infty$ for all $x \in \mathbb{R}$.

$x - \infty = x + (-\infty)$ for all $x \in \mathbb{R}^*$ if either side is defined (e.g. not $\infty - \infty$).

$x - (-\infty) = x + \infty$ for all $x \in \mathbb{R}^*$ if either side is defined (e.g. not $\infty - \infty$).

$x + \infty = \infty + x = \infty$ and $x - \infty = -\infty + x = -\infty$ for all $x \in \mathbb{R}$.

$x \cdot \infty = \infty \cdot x = \infty$ and $x \cdot (-\infty) = (-\infty) \cdot x = -\infty$ for $x > 0$.

$x \cdot \infty = \infty \cdot x = -\infty$ and $x \cdot (-\infty) = (-\infty) \cdot x = \infty$ for $x < 0$.

$x \cdot \infty = \infty \cdot x = 0$ and $x \cdot (-\infty) = (-\infty) \cdot x = 0$ for $x = 0$.

$\infty + \infty = \infty$.

$-\infty - \infty = -\infty$.

$\infty \cdot \infty = (-\infty)(-\infty) = \infty$.

$\infty(-\infty) = (-\infty)\infty = -\infty$.

$\infty - \infty$ and $-\infty + \infty$ are undefined.

If $x = \infty$ and $y = \infty$, then we say $x = y$.

If $x = -\infty$ and $y = -\infty$, then we say $x = y$.

As is customary, the complement of a set A with respect to a set Ω will be denoted A^c. The set Ω should be clear from the context. The set A minus B will be denoted $A \setminus B$. The power set of a set Ω will be denoted $\mathcal{P}(\Omega)$. The symbol $f(A)$ will denote $\{f(x) : x \in A\}$ for a function $f\colon \Omega \to \Pi$ where $A \subset \Omega$. The symbol $f^{-1}(A)$ will denote $\{x \in \Omega : f(x) \in A\}$ for a function $f\colon \Omega \to \Pi$ where $A \subset \Pi$.

The symbol $\{a_n\}_{n=1}^m$ will denote a finite sequence of elements. The symbol $\{a_n\}_n$ will denote a countable sequence of elements.

If two functions f and g defined on a set A are such that $f(x) = g(x)$ for all $x \in A$, then we will write $f = g$ on A. Similarly for $f < g,\ f \leq g$, etc. For $f\colon A \to \mathbb{C}$, we define $(f + g)(x) = f(x) + g(x),\ (fg)(x) = f(x)g(x),\ (\alpha f)(x) = \alpha f(x)$ for $\alpha \in \mathbb{C}$, etc.

The following function will be ubiquitous in the remainder of the paper.

Definition 2.6 (Characteristic Function)**.** Let Ω be a set. Let $A \subset \Omega$. Then the function $\chi_A\colon \Omega \to \mathbb{R}$ such that

$$\chi_A(t) = \begin{cases} 1, & t \in A; \\ 0, & t \notin A, \end{cases}$$

is called the ***characteristic function*** of A.

We will have cause to consider extended real valued functions on a set Ω. That is, $f\colon \Omega \to \mathbb{R}^*$. The next definition allows us to include such functions in the upcoming definition of measurable functions.

Definition 2.7 (Open Subsets of $\mathbb{R}^*$)**.** A subset of $\mathbb{R}^*$ is said to be ***open*** if it can be expressed as a union of sets of the form $(a, b),\ [-\infty, b),$ and $(a, \infty]$, where $a, b \in \mathbb{R}$.

The concept of a σ-algebra is central to the workings of measure theory. Here is the definition.

Definition 2.8 (σ-algebra)**.** Let Ω be a set. Let $\mathcal{A} \subset \mathcal{P}(\Omega)$ and $\mathcal{A} \neq \emptyset$. Then $\mathcal{A}$ is called a σ-algebra if

a) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, and

b) $\{A_n\}_n \subset \mathcal{A} \Rightarrow \bigcup\limits_n A_n \in \mathcal{A}$.

Proposition 2.9. *Let Ω be a set. Let $\mathcal{C} \subset \mathcal{P}(\Omega)$ and $\mathcal{C} \neq \emptyset$. Then there exists a smallest σ-algebra of subsets of Ω containing $\mathcal{C}$. It is denoted $\sigma(\mathcal{C})$.*

Proof. Omitted. ■

The σ-algebra in part b) of the following will be of utmost interest to us.

Definition 2.10.

a) $\mathcal{B}$ denotes the smallest σ-algebra of subsets of $\mathbb{R}$ that contains all the open sets. $\mathcal{B}$ is called the collection of ***Borel sets of*** $\mathbb{R}$.

b) $\mathcal{B}^+$ denotes the smallest σ-algebra of subsets of $\mathbb{R}^+$ that contains all the open sets. $\mathcal{B}^+$ is called the collection of ***Borel sets of*** $\mathbb{R}^+$.

c) $\mathcal{B}^*$ denotes the smallest σ-algebra of subsets of $\mathbb{R}^*$ that contains all the open sets. $\mathcal{B}^*$ is called the collection of ***Borel sets of*** $\mathbb{R}^*$.

The basis of all measure theory is, of course, the measure. We now make the definition.

Definition 2.11 (Measure, Measurable Space, Measure Space)**.** Let Ω be a set and $\mathcal{A}$ be a σ-algebra of subsets of Ω. Suppose $\mu\colon \mathcal{A} \to \mathbb{R}^*$. Then μ is a ***measure*** on $\mathcal{A}$ if all of the following are satisfied:

a) $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.

b) $\mu(\emptyset) = 0$.

c) If $\{A_n\}_n \subset \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n). \text{ This is called } \textbf{\textit{countable additivity}}.$$

$(\Omega, \mathcal{A})$ denotes a ***measurable space*** and $(\Omega, \mathcal{A}, \mu)$ denotes a ***measure space***.

In particular, we denote ***Lebesgue measure*** by λ. See [5] for a complete discussion on the construction of this measure.

The following theorem lists some of the properties a measure possesses.

Theorem 2.12. *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $A, B \in \mathcal{A}$. Then*

a) $\mu(A) < \infty$ and $A \subset B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$.

b) $A \subset B \Rightarrow \mu(A) \leq \mu(B)$. This is called ***monotonicity****.*

c) If $\{A_n\}_n \subset \mathcal{A}$ such that $\mu(A_1) < \infty$ and $A_n \supset A_{n+1}$ for all $n \in \mathbb{N}$, then

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

d) If $\{A_n\}_n \subset \mathcal{A}$ such that $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

e) If $\{A_n\}_n \subset \mathcal{A}$, then

$$\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n). \text{ This is called } \textbf{\textit{countable subadditivity}}.$$

Proof. Omitted. ∎

Measure theory is very flexible in that we may be able to exploit properties even if they do not hold universally. The next definition will make specific these types of properties.

Definition 2.13 (Almost Everywhere, Almost Always)**.** A property is called ***μ-almost everywhere***, or ***μ-a.e.***, if it holds except on a set N such that $\mu(N) = 0$. The term ***almost always*** is also used.

We will need to restrict ourselves to a particular class of functions. These functions are defined in the following.

Definition 2.14 (Measurable Transformation, Measurable Function)**.** Let $(\Gamma, \mathcal{S})$ and $(\Lambda, \mathcal{T})$ be measurable spaces. A mapping $\varphi\colon \Gamma \to \Lambda$ is called a ***measurable transformation*** if $\varphi^{-1}(T) \in \mathcal{S}$ for each $T \in \mathcal{T}$.

In particular, let $(\Omega, \mathcal{A})$ be a measurable space and consider the measurable spaces $(\mathbb{R}, \mathcal{B}),\ (\mathbb{R}^+, \mathcal{B}^+),$ and $(\mathbb{R}^*, \mathcal{B}^*)$. Then

a) a function $f\colon \Omega \to \mathbb{R}$ is called $\mathcal{A}$-measurable if $f^{-1}(B) \in \mathcal{A}$ for each $B \in \mathcal{B}$;

b) a function $f\colon \Omega \to \mathbb{R}^+$ is called $\mathcal{A}$-measurable if $f^{-1}(B) \in \mathcal{A}$ for each $B \in \mathcal{B}^+$;

c) a function $f\colon \Omega \to \mathbb{R}^*$ is called $\mathcal{A}$-measurable if $f^{-1}(B) \in \mathcal{A}$ for each $B \in \mathcal{B}^*$.

Measurable functions possess closure properties.

Theorem 2.15. *Let $(\Omega, \mathcal{A})$ be a measurable space. The set of all real valued $\mathcal{A}$-measurable functions is an algebra. That is, for $\mathcal{A}$-measurable functions f and g and $\alpha \in \mathbb{R}$,*

a) $f + g$ is $\mathcal{A}$-measurable,

b) αf is $\mathcal{A}$-measurable, and

c) $f \cdot g$ is $\mathcal{A}$-measurable.

Proof. Omitted. ∎

The measurable functions in which we will ultimately be interested are those in part b) of the next theorem.

Theorem 2.16.

a) The set of $\mathcal{B}$-measurable functions is the smallest algebra of real valued functions on $\mathbb{R}$ that contains the continuous functions and is closed under pointwise limits.

b) The set of $\mathcal{B}^+$-measurable functions is the smallest algebra of real valued functions on $\mathbb{R}^+$ that contains the continuous functions and is closed under pointwise limits.

Proof. Omitted. ∎

This next function is the basis of the definition of the Lebesgue integral to be defined shortly.

Definition 2.17 (Simple Function, Canonical Representation)**.** Consider the measurable space $(\Omega, \mathcal{A})$. A ***simple function*** is an $\mathcal{A}$-measurable function S whose range is a finite set. Let $\{a_i\}_{i=1}^{n}$ be the nonzero distinct values of S. Let $A_i = \{x : S(x) = a_i\}$, $1 \leq i \leq n$. Then

$$S = \sum_{i=1}^{n} a_i \chi_{A_i}$$

is the ***canonical representation*** of S.

Proposition 2.18. *Consider the measurable space $(\Omega, \mathcal{A})$. Let S be an $\mathcal{A}$-measurable simple function with canonical representation $\sum_{i=1}^{n} a_i \chi_{A_i}$. Then $A_i \in \mathcal{A}$ for all i and the A_i are pairwise disjoint.*

Proof. Omitted. ■

Here is a proposition that relates measurable functions with measurable sets. We will be calling on this proposition in later sections.

Proposition 2.19. *Let $(\Omega, \mathcal{A})$ be a measurable space. Let $f : \Omega \to \mathbb{R}$. The following statements are equivalent:*

a) f is $\mathcal{A}$-measurable.

b) $f^{-1}((-\infty, a)) \in \mathcal{A}$ for each $a \in \mathbb{R}$.

c) $f^{-1}((-\infty, a]) \in \mathcal{A}$ for each $a \in \mathbb{R}$.

d) $f^{-1}((a, \infty)) \in \mathcal{A}$ for each $a \in \mathbb{R}$.

e) $f^{-1}([a, \infty)) \in \mathcal{A}$ for each $a \in \mathbb{R}$.

Proof. Omitted. ■

Besides the measures themselves, the main measure theoretic instrument we will employ is the abstract Lebesgue integral. The Lebesgue integral will allow easy passage between sums and integrals. Here now is the definition of the abstract Lebesgue integral of a nonnegative function.

Definition 2.20 (Lebesgue Integral of a Nonnegative Function)**.** Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let S be a nonnegative $\mathcal{A}$-measurable simple function on Ω with canonical representation $S = \sum_{i=1}^{n} a_i \chi_{A_i}$. Then the ***abstract Lebesgue integral of*** S ***over*** Ω ***with***

respect to μ is defined by

$$\int_{\Omega} S(x)d\mu(x) = \sum_{i=1}^{n} a_i \mu(A_i).$$

If $E \in \mathcal{A}$, then the ***abstract Lebesgue integral of*** S ***over*** E ***with respect to*** μ is defined by

$$\int_{E} S(x)d\mu(x) = \int_{\Omega} \chi_E(x)S(x)d\mu(x).$$

Let f be a nonnegative extended real-valued $\mathcal{A}$-measurable function on Ω. Then the ***abstract Lebesgue integral of*** f ***over*** Ω ***with respect to*** μ is defined by

$$\int_{\Omega} f(x)d\mu(x) = \sup_{S} \int_{\Omega} S(x)d\mu(x),$$

where the supremum is taken over all nonnegative $\mathcal{A}$-measurable simple functions S such that $S \leq f$. If $E \in \mathcal{A}$, then the ***abstract Lebesgue integral of*** f ***over*** E ***with respect to*** μ is defined by

$$\int_{E} f(x)d\mu(x) = \int_{\Omega} \chi_E(x)f(x)d\mu(x).$$

Some properties of the abstract Lebesgue integral for nonnegative extended real valued measurable functions are specified in the following proposition.

Proposition 2.21. *Consider the measure space* $(\Omega, \mathcal{A}, \mu)$. *Let* $f\colon \Omega \to \mathbb{R}^+ \cup \{0, \infty\}$ *and* $g\colon \Omega \to \mathbb{R}^+ \cup \{0, \infty\}$ *be A-measurable functions. Let* $\alpha \geq 0$. *Let* $A, B \in \mathcal{A}$ *such that* $B \subset A$. *Then*

a) $f \leq g$, μ*-a.e.*$\Rightarrow \int_A f d\mu \leq \int_A g d\mu$.

b) $\int_B f d\mu \leq \int_A f d\mu$.

c) $f = 0$, μ*-a.e.*$\Rightarrow \int_A f d\mu = 0$.

d) $\mu(A) = 0 \Rightarrow \int_A f d\mu = 0$.

e) $\int_A \alpha f d\mu = \alpha \int_A f d\mu$.

Proof. Omitted. ■

We now state a major convergence theorem and corollary for the abstract Lebesgue integral for a nonnegative extended real valued measurable function.

Theorem 2.22 (Monotone Convergence Theorem or MCT)**.** *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $\{f_n\}_n$ be a monotone nondecreasing sequence of nonnegative extended real valued $\mathcal{A}$-measurable functions. Then*

$$\int_A \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_A f_n d\mu \text{ for each } A \text{ in } \mathcal{A}.$$

Proof. Omitted. ■

Corollary 2.23. *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let f, g be nonnegative extended real valued $\mathcal{A}$-measurable functions. In addition, let $\{f_n\}_{n=1}^{\infty}$ be a sequence of nonnegative extended real valued $\mathcal{A}$-measurable functions. Let $A \in \mathcal{A}$. Then*

a) $\int_A (f+g) d\mu = \int_A f d\mu + \int_A g d\mu$.

b) $\int_A \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_A f_n d\mu$.

c) If $\{A_n\}_n \subset \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\int_{\cup_n A_n} f d\mu = \sum_n \int_{A_n} f d\mu.$$

Proof. Omitted. ■

We need the following definition and proposition in order to broaden the concept of the abstract Lebesgue integral of a nonnegative extended real valued measurable function: the integral of an extended real valued measurable function.

Definition 2.24. Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $f : \Omega \to \mathbb{R}^*$. Define

$$f^+(x) = \max\{f(x), 0\} \text{ and } f^-(x) = -\min\{f(x), 0\} \text{ for each } x \in \Omega.$$

Proposition 2.25. *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $f : \Omega \to \mathbb{R}^*$ be $\mathcal{A}$-measurable. Then f^+ and f^- are $\mathcal{A}$-measurable.*

Proof. Omitted. ■

We are now able to define the integral of an extended real valued measurable function.

Definition 2.26 (Integral of an Extended Real Valued Measurable Function)**.** Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $f\colon \Omega \to \mathbb{R}^*$ be $\mathcal{A}$-measurable and let $A \in \mathcal{A}$. Then the abstract **Lebesgue integral of** f **over** A **with respect to** μ is defined

$$\int_A f d\mu = \int_A f^+ d\mu - \int_A f^- d\mu$$

provided at least one of the integrals on the right side is finite. Furthermore, if

$$\int_A |f| d\mu < \infty$$

then we say f **is Lebesgue integrable over** A. We denote this as $f \in \mathcal{L}^1(A, \mu)$ where $\mathcal{L}^1(A, \mu)$ is the collection of all extended real valued Lebesgue integrable functions over A. In a like manner, the collection of all extended real valued Lebesgue integrable functions over Ω is denoted $\mathcal{L}^1(\Omega, \mu)$ or $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ or simply $\mathcal{L}^1(\mu)$. Note that f is Lebesgue integrable implies f is $\mathcal{A}$-measurable.

Lebesgue integrable functions have the following properties.

Theorem 2.27. *Consider the measure space* $(\Omega, \mathcal{A}, \mu)$. *Let* $f, g \in \mathcal{L}^1(\Omega, \mu)$ *and* $\alpha \in \mathbb{R}$. *Then*
a) $f + g \in \mathcal{L}^1(\Omega, \mu)$ *and*

$$\int_\Omega (f + g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.$$

b) $\alpha f \in \mathcal{L}^1(\Omega, \mu)$ *and*

$$\int_\Omega \alpha f d\mu = \alpha \int_\Omega f d\mu.$$

c) $f,\ g \in \mathbb{R}$ *and* $f \leq g$ *on* $\Omega \Rightarrow \int_\Omega f d\mu \leq \int_\Omega g d\mu$.
d) $\left| \int_\Omega f d\mu \right| \leq \int_\Omega |f| d\mu$.
e) $\mu(A) = 0 \Rightarrow \int_A f d\mu = 0$.
f) $\{A_n\} \subset \mathcal{A}$ *and* $A_i \cap A_j = \emptyset$ *for* $i \neq j \Rightarrow \displaystyle\int_{\bigcup\limits_n A_n} f d\mu = \sum_n \int_{A_n} f d\mu$.

Proof. Omitted. ■

Here is an important result that compares Riemann integral functions with Lebesgue integral functions.

Theorem 2.28. *Let f be Riemann integrable on $[a,b]$. Then $f \in \mathcal{L}^1([a,b],\lambda)$ and*

$$\int_{[a,b]} f(t)d\lambda(t) = \int_a^b f(t)dt.$$

Proof. Omitted. ■

Just about all of our work will require measures that possess certain properties. One of these properties is the substance of the following definition.

Definition 2.29 (σ-finite Measure Space)**.** Let $(\Gamma, \mathcal{S}, \mu)$ be a measure space. Suppose there exists a sequence $\{S_n\} \subset \mathcal{S}$ such that $\bigcup\limits_n S_n = \Gamma$ and $\mu(S_n) < \infty$ for each n. Then $(\Gamma, \mathcal{S}, \mu)$ is called a ***σ-finite measure space*** and μ is called a ***σ-finite measure on $\mathcal{S}$***.

In order to define what will be known as the multiplicative convolution of measures, we will require the concept of a product measure space.

Definition 2.30 (Product Measure Space)**.** Let $(\Gamma, \mathcal{S}, \mu)$ and $(\Lambda, \mathcal{T}, \nu)$ be σ-finite measure spaces. The Cartesian product of Γ with Λ is defined to be $\{(x,y) : x \in \Gamma,\ y \in \Lambda\}$. The σ-algebra generated by the set $\{S \times T : S \in \mathcal{S},\ T \in \mathcal{T}\}$ is called the ***product σ-algebra of $\mathcal{S}$ with $\mathcal{T}$*** and is denoted $\mathcal{S} \times \mathcal{T}$. The measure on $\mathcal{S} \times \mathcal{T}$ is called the ***product measure of μ with ν*** and is denoted $\mu \times \nu$ where $\mu \times \nu(S \times T) = \mu(S)\nu(T)$ for $S \in \mathcal{S},\ T \in \mathcal{T}$. The measure space $(\Gamma \times \Lambda, \mathcal{S} \times \mathcal{T}, \mu \times \nu)$ is the ***product measure space*** of $(\Gamma, \mathcal{S}, \mu)$ with $(\Lambda, \mathcal{T}, \nu)$.

The product σ-algebra in which we will be the most interested is the subject of part b) of the following.

Definition 2.31.

a) $\mathcal{B} \times \mathcal{B}$ denotes the smallest σ-algebra of subsets of $\mathbb{R} \times \mathbb{R}$ that contains all the open sets. $\mathcal{B} \times \mathcal{B}$ is called the collection of ***Borel sets of*** $\mathbb{R} \times \mathbb{R}$.

b) $\mathcal{B}^+ \times \mathcal{B}^+$ denotes the smallest σ-algebra of subsets of $\mathbb{R}^+ \times \mathbb{R}^+$ that contains all the open sets. $\mathcal{B}^+ \times \mathcal{B}^+$ is called the collection of ***Borel sets of*** $\mathbb{R}^+ \times \mathbb{R}^+$.

c) $\mathcal{B}^* \times \mathcal{B}^*$ denotes the smallest σ-algebra of subsets of $\mathbb{R}^* \times \mathbb{R}^*$ that contains all the open sets. $\mathcal{B}^* \times \mathcal{B}^*$ is called the collection of ***Borel sets of*** $\mathbb{R}^* \times \mathbb{R}^*$.

Lastly in this section we define two types of measures and state with proof an important proposition.

Definition 2.32. Let $(\Omega, \mathcal{A})$ be a measurable space. A measure μ on $\mathcal{A}$ is called ***discrete*** if $\{x\} \in \mathcal{A}$ for all $x \in \Omega$ and there exists a countable set K such that $\mu(K^c) = 0$. A measure ν on $\mathcal{A}$ is called ***continuous*** if $\{x\} \in \mathcal{A}$ for all $x \in \Omega$ and $\nu(\{x\}) = 0$ for all $x \in \Omega$.

Proposition 2.33. *Let $(\Omega, \mathcal{A})$ be a measurable space. Let μ be a discrete measure on $\mathcal{A}$ with corresponding countable set K. Let $A \in \mathcal{A}$ and let $g\colon \Omega \to \mathbb{R}^+ \cup \{0\}$ be $\mathcal{A}$-measurable. Then*

$$\int_A g(t)d\mu(t) = \sum_{x \in A \cap K} g(x)\mu(\{x\}).$$

Proof. Using Proposition 2.21 and Corollary 2.23, we have

$$\begin{aligned}
\int_A g(t)d\mu(t) &= \int_{A \cap K} g(t)d\mu(t) + \int_{A \cap K^c} g(t)d\mu(t) \\
&= \int_{A \cap K} g(t)d\mu(t) = \int_{\bigcup\limits_{x \in A \cap K} \{x\}} g(t)d\mu(t) \\
&= \sum_{x \in A \cap K} \int_{\{x\}} g(t)d\mu(t) = \sum_{x \in A \cap K} \int_{\{x\}} g(x)d\mu(t) \\
&= \sum_{x \in A \cap K} g(x)\mu(\{x\}).
\end{aligned}$$

■

2.3 Construction of Measures

This section is dedicated to the construction of the measures we will require to prove the PNT, and to the derivation of some of their properties. We commence with a definition.

Definition 2.34. Let $f\colon \mathbb{N} \to \mathbb{R}^+ \cup \{0\}$. Define $F\colon \mathbb{R}^+ \to \mathbb{R}^+ \cup \{0\}$ by $F(x) = \sum\limits_{n \leq x} f(n)$. Then F is called the ***Type 1 summatory function*** induced by f.
Let $g\colon \mathbb{N} \to \mathbb{R}$ such that $g(n) < 0$ for at least one n. Define $G\colon \mathbb{R}^+ \to \mathbb{R}$ by $G(x) = \sum\limits_{n \leq x} g(n)$. Then G is called the ***Type 2 summatory function*** induced by g.

It is clear that a Type 1 or 2 summatory function is right continuous and finite for any finite x, and its limit as $x \downarrow 0$ is 0. In addition, a Type 1 summatory function is monotone nondecreasing and a Type 2 summatory function is not monotone nondecreasing.

Define $\lambda_F(E) := \sum_{n \in E} f(n)$ where F and f are as in Definition 2.34 and where $E \in \mathcal{B}^+$. Verification that λ_F is a σ-finite measure on $(\mathbb{R}^+, \mathcal{B}^+)$ is straightforward and we omit it. We officially make the definition.

Definition 2.35. Consider the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$. Let F be the Type 1 summatory function induced by f. Let $\lambda_F(E) := \sum_{n \in E} f(n)$ for $E \in \mathcal{B}^+$. Then the σ-finite measure λ_F on $\mathcal{B}^+$ is called the ***Borel measure induced by the Type 1 summatory function*** F.

Remark 2.36. For $0 < a \leq b < \infty$, note that $\lambda_F((a, b]) = F(b) - F(a)$. □

There are other ways to create measures. One very direct way is the substance of the following proposition.

Proposition 2.37. *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let $f\colon \Omega \to \mathbb{R}^+ \cup \{0\}$ be $\mathcal{A}$-measurable. Let $\nu(A) = \int_A f d\mu$ for $A \in \mathcal{A}$. Then ν is a measure on $\mathcal{A}$ and we will write $\nu = \int f d\mu$ (i.e. with an indefinite integral sign).*

Proof. Verification of the three properties that constitute a measure is straightforward and we omit it. ■

Certain linear combinations over $\mathbb{R}$ of measures is also a measure. Formally, we have

Proposition 2.38. *Let $(\Omega, \mathcal{A})$ be a measurable space. Let $\{\mu_i\}_{i=1}^n$ be measures on $\mathcal{A}$. Let $\{\alpha_i\}_{i=1}^n \in \mathbb{R}^+ \cup \{0\}$. Define $\left(\sum_{i=1}^n \alpha_i \mu_i\right)(A) = \sum_{i=1}^n \alpha_i \mu_i(A)$ for $A \in \mathcal{A}$. This definition is consistent since all terms in the sum are nonnegative. Then $\sum_{i=1}^n \alpha_i \mu_i$ is a measure on $\mathcal{A}$.*

Proof. Demonstrating that $\sum_{i=1}^n \alpha_i \mu_i$ satifies the three properties required to be a measure is straightforward and we omit the details. ■

Listed in the following example are some of the measures we will encounter in the proofs of the PNT.

Example 2.39.

All of the following measures on $(\mathbb{R}^+, \mathcal{B}^+)$ are clearly σ-finite. See Example 2.1 for the definition of the arithmetic functions in the following.

λ denotes the Lebesgue measure restricted to $\mathcal{B}^+$.

$\lambda_0 =$ the null measure ($\lambda_0(E) = 0$ for all $E \in \mathcal{B}^+$).

$\lambda_n = \int \chi_{[n,\infty)}(t)d\lambda(t)$ for $n \in \mathbb{N}$.

λ_N denotes the Borel measure induced by $N(x) := \sum_{n \le x} 1(n)$.

λ_Q denotes the Borel measure induced by $Q(x) := \sum_{n \le x} |\mu|(n)$.

λ_{Q-M} denotes the Borel measure induced by $(Q - M)(x) := \sum_{n \le x} (|\mu| - \mu)(n)$.

λ_ψ denotes the Borel measure induced by $\psi(x) := \sum_{n \le x} \Lambda(n)$.

λ_π denotes the Borel measure induced by $\pi(x) := \sum_{n \le x} 1_{\mathcal{P}}(n)$.

λ_ϑ denotes the Borel measure induced by $\vartheta(x) := \sum_{n \le x} \log n \cdot 1_{\mathcal{P}}(n)$.

λ_χ denotes the Borel measure induced by $\chi(x) := \sum_{n \le x} e(n)$. This will most often be referred to as δ_1.

A special case: Although $M(x) := \sum_{n \le x} \mu(n)$ is not a Type 1 summatory function (it is a Type 2) since $\mu(n)$ is not always nonnegative, we set $\lambda_M = \lambda_Q - \lambda_{Q-M}$ and define $\lambda_M(E) = \lambda_Q(E) - \lambda_{Q-M}(E)$ for $E \in \mathcal{B}^+$ provided the right side exists. Note that λ_M is not a measure. □

Measures induced by Type 1 summatory functions possess special properties. The derivation of these properties is the substance of the following proposition and corollary.

Proposition 2.40. *Consider the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$. Let R be the Type 1 summatory function induced by r (see Definition 2.34). Let λ_R be the measure on $\mathcal{B}^+$ induced by R. Then*

a) $\lambda_R(\{n\}) = r(n)$ for $n \in \mathbb{N}$,

b) $\lambda_R((n-1, n)) = 0$ for $n \in \mathbb{N}$, and

c) $\lambda_R(\mathbb{N}^c) = 0$.

Proof. a) $\lambda_R(\{n\}) = \sum_{k \in \{n\}} r(k) = r(n)$.

b) $\lambda_R((n-1,n)) = \sum_{k\in(n-1,n)} r(k) = 0.$

c)

$$\lambda_R(\mathbb{N}^c) = \lambda_R\left(\bigcup_{n=1}^{\infty}(n-1,n)\right) = \sum_{n=1}^{\infty}\lambda_R((n-1,n)) = \sum_{n=1}^{\infty} 0 = 0.$$

■

Corollary 2.41. *Let the hypotheses be as in Proposition 2.40. In addition, let $B \in \mathcal{B}^+$ and let $g\colon \mathbb{R}^+ \to \mathbb{R}^+ \cup \{0\}$ be $\mathcal{B}^+$-measurable. Then*

$$\int_B g(t)d\lambda_R(t) = \sum_{n\in B} g(n)r(n).$$

Proof. By Proposition 2.40, λ_R is a discrete measure on $\mathcal{B}^+$ with corresponding countable set $\mathbb{N}$. The result now follows from Proposition 2.33 and Proposition 2.40. ■

In order to present our last examples of measures, we first require a definition, theorem, and proposition that introduce us to the Radon-Nikodym derivative.

Definition 2.42 (Absolutely Continuous Measures)**.** Let μ and ν be measures on the measurable space $(\Omega, \mathcal{A})$. We say ν is ***absolutely continuous*** with respect to μ if $\mu(A) = 0 \Rightarrow \nu(A) = 0$ for $A \in \mathcal{A}$. We denote this as $\nu \ll \mu$.

Theorem 2.43 (Radon-Nikodym)**.** *Let μ and ν be σ-finite measures on the measurable space $(\Omega, \mathcal{A})$. Suppose $\nu \ll \mu$. Then there exists $f\colon \Omega \to \mathbb{R}^+ \cup \{0\} \cup \{\infty\}$ that is $\mathcal{A}$-measurable such that*

$$\nu(A) = \int_A f(t)d\mu(t) \text{ for all } A \in \mathcal{A}.$$

Furthermore, if also $g\colon \Omega \to \mathbb{R}^+ \cup \{0\} \cup \{\infty\}$ is $\mathcal{A}$-measurable and $\nu(A) = \int_A g(t)d\mu(t)$ for all $A \in \mathcal{A}$, then $g = f$, μ-a.e. The function f (or g) is called the Radon-Nikodym derivative of ν with respect to μ and is denoted $d\nu/d\mu$.

Proof. Omitted. ■

Proposition 2.44. *Let μ and ν be σ-finite measures on the measurable space $(\Omega, \mathcal{A})$. Let $E \in \mathcal{A}$. Suppose $\nu \ll \mu$ and $h\colon \Omega \to \mathbb{R}^+ \cup \{0\}$ is $\mathcal{A}$-measurable. Then*

$$h\ d\nu/d\mu \text{ is } \mathcal{A}\text{-measurable and } \int_E h(t)d\nu(t) = \int_E h(t)(d\nu/d\mu)(t)d\mu(t).$$

Proof. Omitted. ■

With the preceding results in mind, we now define two more measures.

Definition 2.45. Let μ be a measure on $(\mathbb{R}^+, \mathcal{B}^+)$. Then,

$$L^n\mu \text{ is defined to be the measure } \int \chi_{[1,\infty)}(t) \log^n t\ d\mu(t) \text{ for } n \in \mathbb{N} \text{ and}$$

$$T^n\mu \text{ is defined to be the measure } \int \chi_{[1,\infty)}(t) t^n d\mu(t) \text{ for } n \in \mathbb{Z}.$$

Remark 2.46. $L^n\mu$ and $T^n\mu$ in Definition 2.45 are measures on $(\mathbb{R}^+, \mathcal{B}^+)$ by Proposition 2.37. □

The measures in the preceding definition have the following important property.

Proposition 2.47. *Let $L^n\mu$ and $T^n\mu$ be as defined in Definition 2.45. In addition, let μ be σ-finite. Then $L^{n+1}\mu = L(L^n\mu)$ for all $n \in \mathbb{N}$ and $T^{n+1}\mu = T(T^n\mu)$ for all $n \in \mathbb{Z}$.*

Proof. It is clear μ σ-finite $\Rightarrow$ $L^n\mu$ σ-finite. Also note that $(dL^n\mu/d\mu)(t) = \chi_{[1,\infty)}(t) \log^n t$, μ-a.e. for all $n \in \mathbb{N}$ by the Radon-Nikodym theorem. Let $E \in \mathcal{B}^+$ and $n \in \mathbb{N}$.

$$\begin{aligned} L(L^n\mu)(E) &= \int_E \chi_{[1,\infty)}(t) \log t\ dL^n\mu(t) \\ &= \int_E \chi_{[1,\infty)}(t) \log t (dL^n\mu/d\mu)(t)d\mu(t) \text{ by Proposition 2.44} \\ &= \int_E \chi_{[1,\infty)}(t) \log t\ \chi_{[1,\infty)}(t) \log^n t\ d\mu(t) = \int_E \chi_{[1,\infty)}(t) \log^{n+1} t\ d\mu(t) \\ &= L^{n+1}\mu(E). \end{aligned}$$

Hence, $L(L^n\mu) = L^{n+1}\mu$ for all $n \in \mathbb{N}$. The proof that $T^{n+1}\mu = T(T^n\mu)$ for all $n \in \mathbb{Z}$ is done similarly. ■

A particular case of Proposition 2.47 is the matter of the following example.

Example 2.48. Let $f\colon \mathbb{N} \to \mathbb{R}^+ \cup \{0\}$. and F its corresponding Type 1 summatory function. Let $n \in \mathbb{N}$. Then

$$L\lambda_F(\{n\}) = \int_{\{n\}} \chi_{[1,\infty)}(t) \log t \; d\lambda_F(t) = \log n \; \lambda_F(\{n\}) = \log n \; f(n). \tag{2.1}$$

Similarly,

$$L^m \lambda_F(\{n\}) = \log^m n \; f(n) \text{ for } m \in \mathbb{N}. \tag{2.2}$$

□

The measure $L^n\mu$ in Definition 2.45 will play a large role in the upcoming proofs of the PNT. In particular, it will allow us to derive the Selberg formulae used in the elementary proof of the PNT. Also, we will discover in Section 2.4 that when $n = 1$ and μ is a convolution (yet to be defined) of measures, then L acts in an analogous manner as the derivative of a product of differentiable functions.

Continuing now with our construction of measures, we will have a desire to represent certain set functions on a measurable space as a difference of measures. Under certain conditions this is possible. We end this section with a proposition that contains the details.

Proposition 2.49. *Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space. Let $f\colon \Omega \to \mathbb{R}$ and $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. Let $\nu(A) = \int_A f(t) d\mu(t)$ for $A \in \mathcal{A}$. Let $S = \{x \in \Omega : f(x) \geq 0\}$. Then*

a) There exist σ-finite measures $\nu_1 = \int \chi_S(t) f(t) d\mu(t)$, $\nu_2 = \int -\chi_{S^c}(t) f(t) d\mu(t)$ on $(\Omega, \mathcal{A})$ such that $\nu(A) = \nu_1(A) - \nu_2(A)$ for each $A \in \mathcal{A}$,

b) $d\nu_1/d\mu = \chi_S f$, μ-a.e., and

c) $d\nu_2/d\mu = -\chi_{S^c} f$, μ-a.e.

Proof. $S = \{x \in \Omega : f(x) \geq 0\} = f^{-1}([0,\infty)) \in \mathcal{A}$ by Proposition 2.19. Then χ_S and χ_{S^c} are $\mathcal{A}$-measurable which implies $\chi_S f$ and $-\chi_{S^c} f$ are $\mathcal{A}$-measurable. Furthermore, $\int_\Omega |\chi_S(t) f(t)| d\mu(t) \leq \int_\Omega |f(t)| d\mu(t) < \infty$. Similarly for $-\chi_{S^c} f$. Hence, $\chi_S f, \; -\chi_{S^c} f \in \mathcal{L}^1(\Omega, \mu)$.

Now, define $\nu_1(A) = \int_A \chi_S(t) f(t) d\mu(t)$ and $\nu_2(A) = \int_A -\chi_{S^c}(t) f(t) d\mu(t)$ for $A \in \mathcal{A}$. By Proposition 2.37, ν_1 and ν_2 are measures on $(\Omega, \mathcal{A})$. Moreover they are σ-finite since

$0 \leq \nu_1(\Omega) = \int_\Omega \chi_S(t)f(t)d\mu(t) = \int_\Omega |\chi_S(t)f(t)|d\mu(t) < \infty$. Likewise for ν_2. Furthermore,

$$\begin{aligned}\nu_1(A) - \nu_2(A) &= \int_A \chi_S(t)f(t)d\mu(t) - \int_A -\chi_{S^c}(t)f(t)d\mu(t) \\ &= \int_A (\chi_S(t) + \chi_{S^c}(t))f(t)d\mu(t) \\ &= \int_A f(t)d\mu(t) = \nu(A).\end{aligned}$$

It is apparent that $\nu_1 \ll \mu$ and $\nu_2 \ll \mu$. Hence, by the Radon-Nikodym Theorem 2.43, $d\nu_1/d\mu = \chi_S f,\ \mu$-a.e. and $d\nu_2/d\mu = -\chi_{S^c} f,\ \mu$-a.e. ■

Now that we have constructed some of the measures we will use in the proofs of the PNT, we can continue the development of the measure theoretic concepts by defining a binary operation of measures. Next up: the multiplicative convolution of measures.

2.4 The Multiplicative Convolution of Measures

In order to prove the PNT in the most efficient way we need to develop further measure theoretic devices. The multiplicative convolution of measures, as it is called, will play an important part. It is a generalization of the familiar convolution of arithmetic functions as discussed in Section 2.1. However, to make the convolution usable in our proofs of the PNT, we will have to generalize it even further. Specifically, we want to create a convolution of certain set functions that is well defined and that is consistent with the usual convolution when the aforementioned set functions are measures.

We begin this process with the following standard results that will be instrumental in our ability to define and to manipulate the standard convolution. We start with a standard result that relates Borel sets with measurable functions in an arbitrary measurable space $(\Omega, \mathcal{A})$.

Proposition 2.50. *Consider the measurable spaces* $(\Omega, \mathcal{A})$, $(\mathbb{R}, \mathcal{B})$, $(\mathbb{R}^+, \mathcal{B}^+)$, *and* $(\mathbb{R}^*, \mathcal{B}^*)$. *If* $f\colon \Omega \to \mathbb{R}$, *then* f *is* $\mathcal{A}$*-measurable if and only if* $f^{-1}(O) \in \mathcal{A}$ *for all open* $O \in \mathbb{R}$. *If* $f\colon \Omega \to \mathbb{R}^+$, *then* f *is* $\mathcal{A}$*-measurable if and only if* $f^{-1}(O) \in \mathcal{A}$ *for all open* $O \in \mathbb{R}^+$.

If $f\colon \Omega \to \mathbb{R}^*$, then f is $\mathcal{A}$-measurable if and only if $f^{-1}(O) \in \mathcal{A}$ for all open $O \in \mathbb{R}^*$.

Proof. Omitted. ■

The following theorem will allow us to show the convolution of measures is another measure.

Theorem 2.51 (Change of Variable)**.** *Let $(\Gamma, \mathcal{S}, \mu)$ be a measure space. Let $(\Lambda, \mathcal{T})$ be a measurable space. Let φ be a measurable transformation from $(\Gamma, \mathcal{S})$ to $(\Lambda, \mathcal{T})$ (see Definition 2.14). Let*

$f\colon \Lambda \to \mathbb{R}^+ \cup \{0\} \cup \{\infty\}$ be $\mathcal{T}$-measurable on Λ. Then

a) $\mu \circ \varphi^{-1}$ is a measure on $\mathcal{T}$ where $\mu \circ \varphi^{-1}(T)$ is defined as $\mu(\varphi^{-1}(T))$ for $T \in \mathcal{T}$.

b) $f \circ \varphi$ is $\mathcal{S}$-measurable on Γ.

c) $\displaystyle\int_{\varphi^{-1}(T)} f \circ \varphi(s) d\mu(s) = \int_T f(t) d\mu \circ \varphi^{-1}(t)$ for each $T \in \mathcal{T}$.

Proof. Omitted. ■

This next result is known as Tonelli's Theorem. It is one of the most powerful results in measure theory. With certain hypotheses, it allows the computation of an integral on a product measure space by iteration of integrals on the factor measure spaces. We will be calling on this theorem time and again.

Theorem 2.52 (Tonelli's Theorem)**.** *Let $(\Gamma, \mathcal{S}, \mu)$ and $(\Lambda, \mathcal{T}, \nu)$ be σ-finite measure spaces. Let $f\colon \Gamma \times \Lambda \to \mathbb{R}^+ \cup \{0\} \cup \{\infty\}$ be $\mathcal{S} \times \mathcal{T}$-measurable. Then*

a) $f(x, y)$ is $\mathcal{T}$-measurable for each $x \in \Gamma$.

b) $f(x, y)$ is $\mathcal{S}$-measurable for each $y \in \Lambda$.

c) $\int_\Lambda f(x, y) d\nu(y)$ is $\mathcal{S}$-measurable.

d) $\int_\Gamma f(x, y) d\mu(x)$ is $\mathcal{T}$-measurable.

e) $\displaystyle\int_{\Gamma\times\Lambda} f(x, y) d(\mu\times\nu)(x, y) = \int_\Gamma \left(\int_\Lambda f(x, y) d\nu(y) \right) d\mu(x) = \int_\Lambda \left(\int_\Gamma f(x, y) d\mu(x) \right) d\nu(y)$.

Proof. Omitted. ■

We are at last ready to define the multiplicative convolution of measures - the measure theoretic instrument that will be at the heart of the proofs of the PNT.

Definition 2.53 (Multiplicative Convolution of Measures)**.** Consider the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$. Let μ and ν be measures on $\mathcal{B}^+$. Define

$$\mu * \nu(E) = \int_{\varphi_2^{-1}(E)} d(\mu \times \nu)(s,t),\ E \in \mathcal{B}^+,$$

where $\varphi_2 \colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is the multiplication map defined by $\varphi_2(s,t) = st$.

$\mu * \nu$ is called the ***multiplicative convolution*** - or more simply, the ***convolution*** - of the measures μ and ν.

Remark 2.54. In the interest of thoroughness, let us now show the above definition is well defined. Note that

$$\varphi_2 \text{ continuous } \Rightarrow \varphi_2^{-1}(O) \text{ is open in } \mathbb{R}^+ \times \mathbb{R}^+ \text{ for } O \text{ open in } \mathbb{R}^+ \Rightarrow \varphi_2^{-1}(O) \in \mathcal{B}^+ \times \mathcal{B}^+.$$

By Proposition 2.50, then, φ_2 is $\mathcal{B}^+ \times \mathcal{B}^+$-measurable. Therefore, by Definition 2.14, $\varphi_2^{-1}(E) \in \mathcal{B}^+ \times \mathcal{B}^+$. Thus, the above definition is well defined. □

We see from Definition 2.53 that $\mu * \nu$ is a set function on $\mathcal{B}^+$. In fact, it is a measure on $\mathcal{B}^+$, which we now state formally.

Theorem 2.55. *Consider the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$. Let μ and ν be measures on $(\mathbb{R}^+, \mathcal{B}^+)$. Then $\mu * \nu$ is a measure on $(\mathbb{R}^+, \mathcal{B}^+)$.*

Proof. From Remark 2.54, $\varphi_2^{-1}(E) \in \mathcal{B}^+ \times \mathcal{B}^+$ for all $E \in \mathcal{B}^+$. Therefore φ_2 is a measurable transformation from $\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ (see Definition 2.14). Consequently, $(\mu \times \nu) \circ \varphi_2^{-1}$ is a measure on $\mathcal{B}^+$ by Theorem 2.51 part a). Furthermore, by Theorem 2.51 part c) for $E \in \mathcal{B}^+$,

$$\begin{aligned}(\mu * \nu)(E) &= \int_{\varphi_2^{-1}(E)} d(\mu \times \nu)(s,t) \\ &= \int_E d(\mu \times \nu) \circ \varphi_2^{-1}(u) \\ &= (\mu \times \nu) \circ \varphi_2^{-1}(E).\end{aligned}$$

Thus, $\mu * \nu$ is a measure on $(\mathbb{R}^+, \mathcal{B}^+)$. ■

The following examples will demonstrate the convolution.

Example 2.56. Let μ and ν be σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$. Let $E \in \mathcal{B}^+$. Then

$$\begin{aligned}(\mu * \nu)(E) &= \int_{\varphi_2^{-1}(E)} d(\mu \times \nu)(s,t) \\ &= \int_{\mathbb{R}^+ \times \mathbb{R}^+} \chi_{\varphi_2^{-1}(E)}(s,t) d(\mu \times \nu)(s,t) \\ &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \chi_{\varphi_2^{-1}(E)}(s,t) d\nu(t) d\mu(s) \text{ by Tonelli's theorem.}\end{aligned}$$

One can easily verify that $\chi_{\varphi_2^{-1}(E)}(s,t) = \chi_{s^{-1}E}(t)$ and $\chi_{\varphi_2^{-1}(E)}(s,t) = \chi_{t^{-1}E}(s)$. Thus,

$$(\mu * \nu)(E) = \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} \chi_{s^{-1}E}(t) d\nu(t) \right) d\mu(s) = \int_{\mathbb{R}^+} \nu(s^{-1}E) d\mu(s). \tag{2.3}$$

Similarly, $(\mu * \nu)(E) = \int_{\mathbb{R}^+} \mu(t^{-1}E) d\nu(t).$ (2.4)

In particular, for $\mu = \lambda_F$ and $\nu = \lambda_G$ where λ_F and λ_G are the Borel measures induced by Type 1 summatory functions F and G, respectively, and for $x \geq 1$ and $E = (0, x]$, then

$$\begin{aligned}(\lambda_F * \lambda_G)((0,x]) &= \int_{\mathbb{R}^+} \lambda_G((0, x/s]) d\lambda_F(s) \\ &= \left(\int_{(0,1)} + \int_{[1,x]} + \int_{(x,\infty)} \right) \lambda_G((0, x/s]) d\lambda_F(s) \\ &= \int_{[1,x]} \lambda_G((0, x/s]) d\lambda_F(s). \end{aligned} \tag{2.5}$$

Similarly, $(\lambda_F * \lambda_G)((0,x]) = \int_{[1,x]} \lambda_F((0, x/t]) d\lambda_G(t).$

More generally, if μ and ν are σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$ such that $\mu((0,1)) = 0,\ \nu((0,1)) = 0,\ E \in \mathcal{B}^+$, and $x = \max\{\sup E, 1\} < \infty$, then

$$(\mu * \nu)(E) = \int_{[1,x]} \nu(s^{-1}E) d\mu(s) = \int_{[1,x]} \mu(t^{-1}E) d\nu(t). \tag{2.6}$$

□

Example 2.57. Let μ be a σ-finite measure on $(\mathbb{R}^+, \mathcal{B}^+)$ and $E \in \mathcal{B}^+$. Observe that δ_1 is a discrete measure with corresponding countable set $\{1\}$. Then, by Proposition 2.33,

$$\mu * \delta_1(E) = \int_{\mathbb{R}^+} \mu(t^{-1}E) d\delta_1(t) = \sum_{n \in \{1\}} \mu(n^{-1}E)\delta_1(\{n\}) = \mu(E).$$

Consequently, $\mu * \delta_1 = \mu$. □

The measures we will utilize in the proofs of the PNT will largely be those measures that are induced by Type 1 summatory functions. The following theorem and corollary provide valuable results which we will employ in the proofs of the PNT.

Theorem 2.58. *Consider the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$. Let F, G be Type 1 summatory functions induced by f, g respectively. Let λ_F, λ_G be the Borel measures induced by F, G respectively. Then*

a) $\lambda_F * \lambda_G(E) = \sum\limits_{n \in E} \sum\limits_{m|n} f(m^{-1}n)g(m)$,

b) $\lambda_F * \lambda_G(\{n\}) = \sum\limits_{m|n} f(m^{-1}n)g(m)$, $n \in \mathbb{N}$, *and*

c) $\lambda_F * \lambda_G((0, x]) = \sum\limits_{n \leq x} \sum\limits_{m|n} f(m^{-1}n)g(m)$.

Proof. For given $t \in \mathbb{N}$ and $E \in \mathcal{B}^+$, let $A = \{n \in \mathbb{R}^+ : n \in (t^{-1}E) \cap \mathbb{N}\}$ and $B = \{t^{-1}m \in \mathbb{N} : m \in E \cap \mathbb{N}\}$. We now show that $A = B$. Let $n \in A$.

$$\begin{aligned} n \in A &\Rightarrow n = t^{-1}e \text{ for some } e \in E \text{ and } n = N \text{ for some } N \in \mathbb{N} \\ &\Rightarrow e \in E,\ t^{-1}e \in \mathbb{N},\ e = nt \in \mathbb{N} \\ &\Rightarrow n \in B \Rightarrow A \subset B. \end{aligned}$$

Now, let $n \in B$.

$$\begin{aligned} n \in B &\Rightarrow n = t^{-1}m \text{ for some } m \in E \cap \mathbb{N} \text{ and } n \in \mathbb{N} \\ &\Rightarrow n \in (t^{-1}E) \cap \mathbb{N} \\ &\Rightarrow n \in A \Rightarrow B \subset A \Rightarrow A = B. \end{aligned}$$

Define $\widetilde{f}(y) = \begin{cases} f(y), & y \in \mathbb{N}; \\ 0, & y \in \mathbb{N}^c. \end{cases}$ Then, $A = B$ implies $\sum_{n \in t^{-1}E} f(n) = \sum_{m \in E} \widetilde{f}(t^{-1}m)$. It is not hard to show $\widetilde{f}(t^{-1}u)$ is $\mathcal{B}^+ \times \mathcal{B}^+$-measurable. Moreover, by Tonelli's Theorem 2.52, $\widetilde{f}(t^{-1}u)$ is $\mathcal{B}^+$-measurable with respect to t for each u and is $\mathcal{B}^+$-measurable with respect to u for each t. By combining (2.3) with multiple uses of Tonelli's Theorem 2.52 and recalling λ_N is the counting measure, we then have

a)

$$\begin{aligned}
\lambda_F * \lambda_G(E) &= \int_{\mathbb{R}^+} \int_{t^{-1}E} d\lambda_F(s) d\lambda_G(t) \\
&= \int_{\mathbb{R}^+} \left(\sum_{n \in t^{-1}E} f(n) \right) d\lambda_G(t) \\
&= \int_{\mathbb{R}^+} \left(\sum_{m \in E} \widetilde{f}(t^{-1}m) \right) d\lambda_G(t) \\
&= \int_{\mathbb{R}^+} \int_E \widetilde{f}(t^{-1}u) d\lambda_N(u) d\lambda_G(t) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \chi_E(u) \widetilde{f}(t^{-1}u) d\lambda_N(u) d\lambda_G(t) \\
&= \int_{\mathbb{R}^+} \chi_E(u) \int_{\mathbb{R}^+} \widetilde{f}(t^{-1}u) d\lambda_G(t) d\lambda_N(u) = \int_E \int_{\mathbb{R}^+} \widetilde{f}(t^{-1}u) d\lambda_G(t) d\lambda_N(u) \\
&= \sum_{n \in E} \sum_{m|n} f(m^{-1}n) g(m).
\end{aligned}$$

In particular,

b) $\lambda_F * \lambda_G(\{n\}) = \sum_{m|n} f(m^{-1}n)g(m)$, $n \in \mathbb{N}$.

c) $\lambda_F * \lambda_G((0, x]) = \sum_{n \leq x} \sum_{m|n} f(m^{-1}n)g(m)$.

In b), we see that $\lambda_F * \lambda_G(\{n\})$ is simply the traditional definition of the multiplicative convolution of the arithmetic functions f and g. In other words, $\lambda_F * \lambda_G(\{n\}) = f * g(n)$. In a), we note that by letting $H(x) := \sum_{n \leq x} \sum_{m|n} f(m^{-1}n)g(m)$ we see $\lambda_F * \lambda_G = \lambda_H$, the Borel measure induced by the Type 1 summatory function H induced by $f * g$. ∎

Corollary 2.59. *Let f, g, and h be nonnegative arithmetic functions such that $f * g(n) = h(n)$ for all $n \in \mathbb{N}$. Let λ_F, λ_G, and λ_H be the Borel measures induced by f, g, and h, respectively. Then $\lambda_F * \lambda_G = \lambda_H$.*

Proof. Let $E \in \mathcal{B}^+$. Then

$$\lambda_F * \lambda_G(E) = \sum_{n \in E} f * g(n) = \sum_{n \in E} h(n) = \lambda_H(E).$$

Therefore, since E is arbitrary, $\lambda_F * \lambda_G = \lambda_H$. ■

Now that we have looked at some examples of the convolution, let us derive one of its properties.

Proposition 2.60. *Let μ and ν be σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$. Then $\mu * \nu = \nu * \mu$.*

Proof. Let $E \in \mathcal{B}^+$. From (2.3) and (2.4) we get

$$\mu * \nu(E) = \int_{\mathbb{R}^+} \nu(s^{-1}E)d\mu(s) = \int_{\mathbb{R}^+} \mu(t^{-1}E)d\nu(t) = \nu * \mu(E).$$

Therefore, since E is arbitrary, $\mu * \nu = \nu * \mu$. ■

We are now prepared to take the first step in generalizing the convolution to a broader domain of set functions. With that in mind, we state the next proposition.

Proposition 2.61. *Let $\mu = \sum_{i=1}^{n} \alpha_i \mu_i$ and $\nu = \sum_{j=1}^{m} \beta_j \nu_j$ where $\mu_i,\ \nu_j$ are σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$ and $\alpha_i,\ \beta_j \in \mathbb{R}^+ \cup \{0\}$ for each $i,\ j$. Define $\mu(A) = \sum_{i=1}^{n} \alpha_i \mu_i(A)$ and $\nu(A) = \sum_{j=1}^{m} \beta_j \nu_j(A)$ for $A \in \mathcal{B}^+$. This definition is consistent since all terms in the sums are nonnegative. Then μ and ν are σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$ and*

$$\mu * \nu = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j$$

where

$$\left(\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j \right)(E) \text{ is defined as } \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j(E),\ E \in \mathcal{B}^+.$$

Proof. By Proposition 2.38, μ and ν are measures on $(\mathbb{R}^+, \mathcal{B}^+)$. For each i, the σ-finiteness of μ_i implies there exists $\{A_{i,k_i}\}_{k_i} \subset \mathcal{B}^+$ such that $\bigcup_{k_i} A_{i,k_i} = \mathbb{R}^+$ and

$\mu_i(A_{i,k_i}) < \infty$ for each k_i. Furthermore,

$$\bigcup_{k_i} A_{i,k_i} = \mathbb{R}^+ \text{ for each } i \Rightarrow \bigcap_{i=1}^{n} \left(\bigcup_{k_i} A_{i,k_i} \right) = \mathbb{R}^+$$
$$\Rightarrow \bigcup_{k_1,k_2,\ldots,k_n} (A_{1,k_1} \cap A_{2,k_2} \cap \ldots \cap A_{n,k_n}) = \mathbb{R}^+.$$

Now, $A_{1,k_1} \cap A_{2,k_2} \cap \ldots \cap A_{n,k_n} \subset A_{i,k_i}$ for each $\{i, k_i\}$. This implies $\mu_i(A_{1,k_1} \cap A_{2,k_2} \cap \ldots \cap A_{n,k_n}) < \infty$ for each i. It follows that $\mu(A_{1,k_1} \cap A_{2,k_2} \cap \ldots \cap A_{n,k_n}) < \infty$. Consequently, μ is σ-finite. Similarly, ν is σ-finite. So, we can now utilize Tonelli's theorem in the following. For $E \in \mathcal{B}^+$,

$$\begin{aligned}
\mu * \nu(E) &= \int_{\mathbb{R}^+} \mu(t^{-1}E) d\nu(t) = \int_{\mathbb{R}^+} \sum_{i=1}^{n} \alpha_i \mu_i(t^{-1}E) d\nu(t) \\
&= \sum_{i=1}^{n} \alpha_i \int_{\mathbb{R}^+} \mu_i(t^{-1}E) d\nu(t) = \sum_{i=1}^{n} \alpha_i \int_{\mathbb{R}^+} \nu(s^{-1}E) d\mu_i(s) \\
&= \sum_{i=1}^{n} \alpha_i \int_{\mathbb{R}^+} \sum_{j=1}^{m} \beta_j \nu_j(s^{-1}E) d\mu_i(s) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \int_{\mathbb{R}^+} \nu_j(s^{-1}E) d\mu_i(s) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j(E).
\end{aligned}$$

The result follows. ■

Still in the context of the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$, we now want to generalize the operation of convolution even further to set functions that are linear combinations over $\mathbb{R}$ of Borel measures. In doing so, we need to make sure the operation is well defined and that it is consistent with our prior results. Under certain circumstances this is possible, as we now show.

We begin by defining

$$\mathcal{E} = \{E \in \mathcal{B}^+ : \sup E < \infty\} \text{ and}$$
$$\mathscr{M} = \{\sigma\text{-finite } \mathcal{B}^+ \text{ measures } \mu : \mu((0,1)) = 0 \text{ and } \mu(E) < \infty \text{ for } E \in \mathcal{E}\}.$$

Remark 2.62. Actually, $\mu(E) < \infty$ for $E \in \mathcal{E}$ implies μ is σ-finite, but not vice versa. □

Now, let $\mu,\ \nu \in \mathscr{M}$ and $E \in \mathcal{E}$. Let $x = \max\{\sup E, 1\}$. We know from Theorem 2.55 that $\mu * \nu$ is a measure on $\mathcal{B}^+$. Moreover by (2.3),

$$\begin{aligned}\mu * \nu(E) &= \int_{\mathbb{R}^+} \nu(s^{-1}E)d\mu(s)\\ &= \int_{(0,1)} \nu(s^{-1}E)d\mu(s) + \int_{[1,x]} \nu(s^{-1}E)d\mu(s) + \int_{(x,\infty)} \nu(s^{-1}E)d\mu(s)\\ &= \int_{[1,x]} \nu(s^{-1}E)d\mu(s) < \infty \text{ since } [1,x] \in \mathcal{E} \text{ and } s^{-1}E \in \mathcal{E} \text{ for } s \geq 1.\end{aligned}$$

In particular, $\mu * \nu((n-1,n]) < \infty$ for all $n \in \mathbb{N}$. This implies, of course, that $\mu * \nu$ is σ-finite. We also have

$$\begin{aligned}\mu * \nu((0,1)) &= \int_{\mathbb{R}^+} \nu((0,s^{-1}))d\mu(s)\\ &= \int_{[1,\infty)} \nu((0,s^{-1}))d\mu(s) = 0 \text{ since } (0,s^{-1}) \subset (0,1) \text{ for } s \geq 1.\end{aligned}$$

Therefore, $\mu * \nu \in \mathscr{M}$. (2.7)

Not only do we have closure in $\mathscr{M}$ with respect to convolution, we have the following two properties.

Proposition 2.63. *Let $\mu,\ \nu$, and $\rho \in \mathscr{M}$. Then*

a) $\mu * \nu = \nu * \mu$.

b) $\mu * (\nu * \rho) = (\mu * \nu) * \rho$.

Proof. a) This follows from Proposition 2.60

b) Let $E \in \mathcal{B}^+$. Utilizing (2.3), (2.4), and Tonelli's theorem we have

$$\begin{aligned}\mu * (\nu * \rho)(E) &= \int_{\mathbb{R}^+} (\nu * \rho)(s^{-1}E)d\mu(s)\\ &= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} \nu(t^{-1}(s^{-1}E))d\rho(t)\right) d\mu(s)\\ &= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} \nu(s^{-1}(t^{-1}E))d\mu(s)\right) d\rho(t)\\ &= \int_{\mathbb{R}^+} (\nu * \mu)(t^{-1}E)d\rho(t)\\ &= \int_{\mathbb{R}^+} (\mu * \nu)(t^{-1}E)d\rho(t)\end{aligned}$$

$$= (\mu * \nu) * \rho(E).$$

Therefore, $\mu * (\nu * \rho) = (\mu * \nu) * \rho$ and we may define $\mu * \nu * \rho$ as the common value. ■

Continuing, let $\bar{\mu} = \sum_{i=1}^{n} \alpha_i \mu_i$ where $\alpha_i \in \mathbb{R}$ and $\mu_i \in \mathscr{M}$. Define $\bar{\mu}(A) = \sum_{i=1}^{n} \alpha_i \mu_i(A)$ for $A \in \mathcal{E}$. Observe the sum on the right side of the last equation exists since $\mu_i(A) < \infty$ for each i. Similary, let $\bar{\nu} = \sum_{j=1}^{m} \beta_j \nu_j$ where $\beta_j \in \mathbb{R}$ and $\nu_j \in \mathscr{M}$. Define $\bar{\nu}(A) = \sum_{j=1}^{m} \beta_j \nu_j(A)$ for $A \in \mathcal{E}$. Observe the sum on the right side of the last equation exists since $\nu_j(A) < \infty$ for each j. Note, also, that $\bar{\mu}$ and $\bar{\nu}$ are linear combinations over $\mathbb{R}$ of measures in $\mathscr{M}$ and that they are not necessarily measures, themselves.

Using Proposition 2.61 as a guide, we would like to define the convolution of $\bar{\mu}$ and $\bar{\nu}$ on $\mathcal{E}$ as

$$\bar{\mu} * \bar{\nu} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j,$$

which is also a linear combination over $\mathbb{R}$ of measures in $\mathscr{M}$. Observe, then, since each $\mu_i * \nu_j(E) < \infty$, we can define $\bar{\mu} * \bar{\nu}(E) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j(E)$ for $E \in \mathcal{E}$. Notice, also, if all $\alpha_i,\ \beta_j$ are nonnegative this definition is consistent with the result of Proposition 2.61.

However, before we can formally state this as a definition, we must show this generalized convolution is well defined. In other words, if in addition $\bar{\mu} = \sum_{i'=1}^{n'} \alpha'_{i'} \mu'_{i'}$ where $\alpha'_{i'} \in \mathbb{R}$ and $\mu'_{i'} \in \mathscr{M}$; and $\bar{\nu} = \sum_{j'=1}^{m'} \beta'_{j'} \nu'_{j'}$ where $\beta'_{j'} \in \mathbb{R}$ and $\nu'_{j'} \in \mathscr{M}$, then $\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j = \sum_{i'=1}^{n'} \sum_{j'=1}^{m'} \alpha'_{i'} \beta'_{j'} \mu'_{i'} * \nu'_{j'}$ on $\mathcal{E}$. We proceed with multiple uses of Tonelli's Theorem 2.52 and (2.6). Let $E \in \mathcal{E}$ and $x = \max\{\sup E, 1\}$.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j(E) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \int_{[1,x]} \mu_i(t^{-1}E) d\nu_j(t)$$
$$= \sum_{j=1}^{m} \beta_j \int_{[1,x]} \left(\sum_{i=1}^{n} \alpha_i \mu_i(t^{-1}E) \right) d\nu_j(t) = \sum_{j=1}^{m} \beta_j \int_{[1,x]} \left(\sum_{i'=1}^{n'} \alpha'_{i'} \mu'_{i'}(t^{-1}E) \right) d\nu_j(t)$$
$$= \sum_{i'=1}^{n'} \sum_{j=1}^{m} \alpha'_{i'} \beta_j \int_{[1,x]} \mu'_{i'}(t^{-1}E) d\nu_j(t) = \sum_{i'=1}^{n'} \sum_{j=1}^{m} \alpha'_{i'} \beta_j \int_{[1,x]} \nu_j(s^{-1}E) d\mu'_{i'}(s)$$
$$= \sum_{i'=1}^{n'} \alpha'_{i'} \int_{[1,x]} \left(\sum_{j=1}^{m} \beta_j \nu_j(s^{-1}E) \right) d\mu'_{i'}(s) = \sum_{i'=1}^{n'} \alpha'_{i'} \int_{[1,x]} \left(\sum_{j'=1}^{m'} \beta'_{j'} \nu'_{j'}(s^{-1}E) \right) d\mu'_{i'}(s)$$

$$= \sum_{i'=1}^{n'} \sum_{j'=1}^{m'} \alpha'_{i'} \beta'_{j'} \int_{[1,x]} \nu'_{j'}(s^{-1}E) d\mu'_{i'}(s) = \sum_{i'=1}^{n'} \sum_{j'=1}^{m'} \alpha'_{i'} \beta'_{j'} \mu'_{i'} * \nu'_{j'}(E)$$

and the convolution is well defined. We can now make the formal definition.

Definition 2.64 (Generalized Convolution of Measures)**.** Let $\bar{\mu} = \sum_{i=1}^{n} \alpha_i \mu_i$ where $\alpha_i \in \mathbb{R}$ and $\mu_i \in \mathscr{M}$. Define $\bar{\mu}(A) = \sum_{i=1}^{n} \alpha_i \mu_i(A)$ for $A \in \mathcal{E}$. Observe the sum on the right side of the last equation exists (e.g. not $\infty - \infty$) since $\mu_i(A) < \infty$ for each i. Similary, let $\bar{\nu} = \sum_{j=1}^{m} \beta_j \nu_j$ where $\beta_j \in \mathbb{R}$ and $\nu_j \in \mathscr{M}$. Define $\bar{\nu}(A) = \sum_{j=1}^{m} \beta_j \nu_j(A)$ for $A \in \mathcal{E}$. Observe the sum on the right side of the last equation exists since $\nu_j(A) < \infty$ for each j. Then we define the convolution

$$\bar{\mu} * \bar{\nu} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j \text{ on } \mathcal{E}$$

where

$$\left(\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j \right)(A) \text{ is defined as } \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \mu_i * \nu_j(A),\ A \in \mathcal{E}.$$

Now, let $\mathscr{L}(\mathcal{E}) = \{$linear combinations over $\mathbb{R}$ of measures in $\mathscr{M}$ restricted to $\mathcal{E}\}$. By the foregoing results, $\mathscr{L}(\mathcal{E})$ is closed under convolution. In fact, we have much more.

Proposition 2.65. *$\mathscr{L}(\mathcal{E})$ is a commutative algebra over $\mathbb{R}$ with identity under the operations of addition and convolution. That is, for $\bar{\mu},\ \bar{\nu},\ \bar{\rho} \in \mathscr{L}(\mathcal{E})$ and $\alpha \in \mathbb{R}$,*

a) $\bar{\mu} + \bar{\nu} \in \mathscr{L}(\mathcal{E})$,

b) $\alpha\bar{\mu} \in \mathscr{L}(\mathcal{E})$,

c) $\bar{\mu} * \bar{\nu} \in \mathscr{L}(\mathcal{E})$,

d) $\alpha(\bar{\mu} * \bar{\nu}) = (\alpha\bar{\mu}) * \bar{\nu} = \bar{\mu} * (\alpha\bar{\nu})$,

e) $\bar{\mu} * \bar{\nu} = \bar{\nu} * \bar{\mu}$,

f) $\bar{\mu} * (\bar{\nu} * \bar{\rho}) = (\bar{\mu} * \bar{\nu}) * \bar{\rho}$,

g) $\bar{\mu} + \bar{\nu} = \bar{\nu} + \bar{\mu}$,

h) $\bar{\mu} + (\bar{\nu} + \bar{\rho}) = (\bar{\mu} + \bar{\nu}) + \bar{\rho}$,

i) $\bar{\mu} * (\bar{\nu} + \bar{\rho}) = \bar{\mu} * \bar{\nu} + \bar{\mu} * \bar{\rho}$,

j) $\bar{\mu} * \delta_1 = \bar{\mu}$, *k)* $\bar{\mu} + \lambda_0 = \bar{\mu}$,

l) $\bar{\mu} + (-\bar{\mu}) = \lambda_0$.

Proof. Properties a), b), d), g), h), i), k), and l) follow easily from the definition. Property c) follows from (2.7) and the definition. Properties e) and f) follow from the corresponding properties of the individual measures in Proposition 2.63. We omit the details. We verify j). Let $\bar{\mu}$ as in Definition 2.64 and $E \in \mathcal{E}$. Note that $\delta_1 \in \mathscr{L}(\mathcal{E})$. We have

$$\bar{\mu} * \delta_1(E) = \sum_{i=1}^{n} \alpha_i \mu_i * \delta_1(E) = \sum_{i=1}^{n} \alpha_i \mu_i(E) = \bar{\mu}(E).$$

Therefore, $\bar{\mu} * \delta_1 = \bar{\mu}$ on $\mathcal{E}$. ■

Example 2.66. All the measures mentioned in Example 2.39 except Lebesgue measure λ are members of $\mathscr{M}$ (and so of $\mathscr{L}(\mathcal{E})$). λ_M is not a measure; but it is a member of $\mathscr{L}(\mathcal{E})$. □

Example 2.67. Let $\mu \in \mathscr{M}$. It is clear by Definition 2.45 and Proposition 2.47 that $L^n\mu \in \mathscr{M}$ (and so of $\mathscr{L}(\mathcal{E})$) for $n \in \mathbb{N}$ and that $T^n\mu \in \mathscr{M}$ (and so of $\mathscr{L}(\mathcal{E})$) for $n \in \mathbb{Z}$. □

The following are two examples which demonstrate the use of the generalized convolution.

Example 2.68. μ is the Möbius function in this example. Let $E \in \mathcal{E}$. Then

$$\begin{aligned}\lambda_M * \lambda_N(E) &= \lambda_Q * \lambda_N(E) - \lambda_{Q-M} * \lambda_N(E) \\ &= \sum_{n\in E} |\mu| * 1(n) - \sum_{n \in E} (|\mu| - \mu) * 1(n) \text{ by Theorem 2.58} \\ &= \sum_{n\in E}\sum_{m|n} |\mu|(m)1(m^{-1}n) - \sum_{n\in E}\sum_{m|n}(|\mu| - \mu)(m)1(m^{-1}n) \\ &= \sum_{n\in E}\sum_{m|n} \mu(m)1(m^{-1}n) = \sum_{n\in E} \mu * 1(n) = \sum_{n\in E} e(n) \text{ by Th. 2.4} \\ &= \delta_1(E).\end{aligned}$$

Therefore, $\lambda_M * \lambda_N = \delta_1$ on $\mathcal{E}$. □

Example 2.69. Let $\bar{\mu} = \sum_{i=1}^{n} \alpha_i \mu_i$ be in $\mathscr{L}(\mathcal{E})$ where $\alpha_i \in \mathbb{R}$ and $\mu_i \in \mathscr{M}$. Then, for $x \geq 1$,

$$\begin{aligned}
\bar{\mu} * T^{-1}\lambda_1((0,x]) &= \sum_{i=1}^{n} \alpha_i \mu_i * T^{-1}\lambda_1((0,x]) \\
&= \sum_{i=1}^{n} \alpha_i \int_{[1,x]} \mu_i((0,t^{-1}x])dT^{-1}\lambda_1(t) \\
&= \sum_{i=1}^{n} \alpha_i \int_{[1,x]} \mu_i((0,t^{-1}x])t^{-1}d\lambda_1(t) \\
&= \sum_{i=1}^{n} \alpha_i \int_{[1,x]} \mu_i((0,u])u^{-1}d\lambda_1(u) \text{ (after a change of variable } u = x/t) \\
&= \int_{[1,x]} \sum_{i=1}^{n} \alpha_i \mu_i((0,u])u^{-1}d\lambda_1(u).
\end{aligned}$$

□

Next, we state and prove a theorem and corollary regarding integration with respect to a convolution. We will need these results in Chapter 3. First, though, we need the following standard measure theory result.

Proposition 2.70. *Consider the measurable space* $(\Omega, \mathcal{A})$. *Then*

$f\colon \Omega \to \mathbb{R}^+ \cup \{0,\infty\}$ *is* $\mathcal{A}$*-measurable*

$\Rightarrow$

there exists a nondecreasing sequence of nonnegative $\mathcal{A}$*-measurable simple functions that converges pointwise to* f.

Proof. Omitted. ■

Theorem 2.71. *Consider the measurable space* $(\mathbb{R}^+, \mathcal{B}^+)$. *Let* μ, ν *be* σ*-finite measures on* $\mathcal{B}^+$. *Let* $f\colon \mathbb{R}^+ \to \mathbb{R}^+ \cup \{0,\infty\}$ *be* $\mathcal{B}^+$*-measurable. Then*

$$\int_E f(u)d\mu * \nu(u) = \int_{\mathbb{R}^+} \int_{s^{-1}E} f(st)d\nu(t)d\mu(s) \text{ for all } E \in \mathcal{B}^+.$$

Proof. We proceed by bootstrapping. First, let $f = \chi_A$ for $A \in \mathcal{B}^+$. Then

$$\int_E \chi_A(u)d\mu * \nu(u) = \int_{E\cap A} d\mu * \nu(u) = \mu * \nu(E \cap A).$$

On the other hand,

$$\begin{aligned}\int_{\mathbb{R}^+}\int_{s^{-1}E}\chi_A(st)d\nu(t)d\mu(s) &= \int_{\mathbb{R}^+}\int_{s^{-1}E}\chi_{s^{-1}A}(t)d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}E\cap s^{-1}A} d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}(E\cap A)} d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\nu(s^{-1}(E\cap A))d\mu(s)\\ &= \mu*\nu(E\cap A).\end{aligned}$$

Now let $f := S = \sum_{i=1}^{n} a_i\chi_{A_i}$ where $a_i \geq 0$ and $A_i \in \mathcal{B}^+$ for all i (i.e. a nonnegative, $\mathcal{B}^+$-measurable simple function). Then

$$\begin{aligned}\int_E S(u)d\mu*\nu(u) &= \int_E \sum_{i=1}^{n} a_i\chi_{A_i}(u)d\mu*\nu(u)\\ &= \sum_{i=1}^{n} a_i \int_E \chi_{A_i}(u)d\mu*\nu(u)\\ &= \sum_{i=1}^{n} a_i \int_{\mathbb{R}^+}\int_{s^{-1}E}\chi_{A_i}(st)d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}E}\sum_{i=1}^{n} a_i\chi_{A_i}(st)d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}E} S(st)d\nu(t)d\mu(s).\end{aligned}$$

Finally, let $f\colon \mathbb{R}^+ \to \mathbb{R}^+ \cup \{0,\infty\}$ be $\mathcal{B}^+$-measurable. By Proposition 2.70, there exists a nondecreasing sequence of nonnegative $\mathcal{B}^+$-measurable simple functions S_m such that $\lim_{m\to\infty} S_m(x) = f(x)$ for all $x \in \mathbb{R}^+$. Then

$$\begin{aligned}\int_E f(u)d\mu*\nu(u) &= \int_E \lim_{m\to\infty} S_m(u)d\mu*\nu(u)\\ &= \lim_{m\to\infty}\int_E S_m(u)d\mu*\nu(u) \text{ by MCT}\\ &= \lim_{m\to\infty}\int_{\mathbb{R}^+}\int_{s^{-1}E} S_m(st)d\nu(t)d\mu(s)\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}E}\lim_{m\to\infty} S_m(st)d\nu(t)d\mu(s) \text{ by MCT twice more}\\ &= \int_{\mathbb{R}^+}\int_{s^{-1}E} f(st)d\nu(t)d\mu(s).\end{aligned}$$

■

This next corollary will be vital in the proof of Selberg's formulae in Chapter 3.

Corollary 2.72. *Let $\mu,\ \nu \in \mathscr{M}$ and $n \in \mathbb{Z}$. Then*

$$L(\mu * \nu) = L\mu * \nu + \mu * L\nu \quad \textit{on } \mathcal{E} \textit{ and}$$
$$T^n(\mu * \nu) = (T^n\mu) * (T^n\nu) \quad \textit{on } \mathcal{E}.$$

Proof. Let $E \in \mathcal{E}$. Then

$$\begin{aligned} L(\mu * \nu)(E) &= \int_E \chi_{[1,\infty)}(u) \log u \, d\mu * \nu(u) \\ &= \int_E \log u \, d\mu * \nu(u) \\ &= \int_{\mathbb{R}^+} \int_{s^{-1}E} \log(st) d\nu(t) d\mu(s) \text{ by Theorem 2.71} \\ &= \int_{\mathbb{R}^+} \int_{s^{-1}E} (\log s + \log t) d\nu(t) d\mu(s) \\ &= \int_{\mathbb{R}^+} \nu(s^{-1}E) dL\mu(s) + \int_{\mathbb{R}^+} L\nu(s^{-1}E) d\mu(s) \\ &= L\mu * \nu(E) + \mu * L\nu(E) = (L\mu * \nu + \mu * L\nu)(E). \end{aligned}$$

Since E is arbitrary in $\mathcal{E}$, the first result is proved. The second result is proved in a similar manner. ∎

Although not pertaining to convolution, this would be an appropriate time to define what we mean by $L^n\bar{\mu}, n \in \mathbb{N}$ and $T^m\bar{\mu}, m \in \mathbb{Z}$ where $\bar{\mu} \in \mathscr{L}(\mathcal{E})$. We will need this in Section 3.5. Of course, our definition must be well defined and consistent with Definition 2.45 for the case where $\bar{\mu}$ is a measure.

So, let $\bar{\mu} = \sum_{i=1}^{k} \alpha_i \mu_i$ where $\alpha_i \in \mathbb{R}$ and $\mu_i \in \mathscr{M}$. For the moment, define $L^n\bar{\mu}(E) = \sum_{i=1}^{k} \alpha_i L^n \mu_i(E)$ for $E \in \mathcal{E}$. Then, clearly if $\alpha_1 = 1$ and $\alpha_i = 0,\ i > 1$ (i.e. $\bar{\mu}$ is a measure), then this is consistent wth Definition 2.45.

It remains to show well definedness. So, suppose $\bar{\mu}$ also equals $\sum_{j=1}^{l} \beta_j \nu_j$ where $\beta_j \in \mathbb{R}$ and $\nu_j \in \mathscr{M}$. Let $\rho = \int n \frac{\log^{n-1} t}{t} \, d\lambda_1(t)$. Then $\rho \in \mathscr{M}$; $\rho((0,t]) = \log^n t$ for $t \geq 1$; and $d\rho/d\lambda_1(t) = n\frac{\log^{n-1} t}{t}$, λ_1-a.e. Let $E \in \mathcal{E}$. Without loss of generality, we can assume $E \cap (0,1) = \emptyset$. Then

$$\begin{aligned}
&\sum_{i=1}^{k} \alpha_i L^n \mu_i(E) = \sum_{i=1}^{k} \alpha_i \int_E \chi_{[1,\infty)}(t) \log^n t \, d\mu_i(t) \\
&= \sum_{i=1}^{k} \alpha_i \int_E \log^n t \, d\mu_i(t) = \sum_{i=1}^{k} \alpha_i \int_E \rho((0,t]) d\mu_i(t) \\
&= \sum_{i=1}^{k} \alpha_i \int_E \int_{\mathbb{R}^+} \chi_{(0,t]}(s) d\rho(s) d\mu_i(t) = \sum_{i=1}^{k} \alpha_i \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \chi_E(t) \chi_{(0,t]}(s) d\rho(s) d\mu_i(t) \\
&= \sum_{i=1}^{k} \alpha_i \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \chi_E(t) \chi_{[s,\infty)}(t) d\mu_i(t) d\rho(s) \text{ (by Tonelli and that } \chi_{(0,t]}(s) = \chi_{[s,\infty)}(t)) \\
&= \sum_{i=1}^{k} \alpha_i \int_{\mathbb{R}^+} \mu_i(E \cap [s,\infty)) d\rho(s) = \int_{\mathbb{R}^+} \sum_{i=1}^{k} \alpha_i \mu_i(E \cap [s,\infty)) d\rho(s) \\
&= \int_{\mathbb{R}^+} \sum_{j=1}^{l} \beta_j \nu_j(E \cap [s,\infty)) d\rho(s) \\
&= \sum_{j=1}^{l} \beta_j L^n \nu_j(E) \text{ by reversing the previous steps}
\end{aligned}$$

and the definition is well defined. A similar argument holds for T^m. We can now make the formal definition.

Definition 2.73. Let $\bar{\mu} \in \mathscr{L}(\mathcal{E})$ such that $\bar{\mu} = \sum_{i=1}^{k} \alpha_i \mu_i$ where $\alpha_i \in \mathbb{R}$ and $\mu_i \in \mathscr{M}$. Then we define

$$L^n \bar{\mu} = \sum_{i=1}^{k} \alpha_i L^n \mu_i \text{ on } \mathcal{E} \text{ for } n \in \mathbb{N}$$

and

$$T^m \bar{\mu} = \sum_{i=1}^{k} \alpha_i T^m \mu_i \text{ on } \mathcal{E} \text{ for } m \in \mathbb{Z},$$

where

$$\left(\sum_{i=1}^{k} \alpha_i L^n \mu_i \right)(E) \text{ is defined as } \sum_{i=1}^{k} \alpha_i L^n \mu_i(E), \ E \in \mathcal{E}$$

and

$$\left(\sum_{i=1}^{k} \alpha_i T^m \mu_i\right)(E) \text{ is defined as } \sum_{i=1}^{k} \alpha_i T^m \mu_i(E),\ E \in \mathcal{E}.$$

2.5 Integration by Parts and Euler's Summation Formula

There will be many instances where we will require an integration by parts formula. The formula will have to be general enough to accommodate: (1) integration of a discrete measure with respect to a discrete measure, (2) integration of a discrete measure with respect to a continuous measure, (3) integration of a continuous measure with respect to a discrete measure, and (4) integration of a continuous measure with respect to a continuous measure. To that end, we begin this section with a derivation of a suitable integration by parts formula. Following that will be given several examples on the use of the formula. We then will end the section with an important application of integration by parts: Euler's summation formula.

Theorem 2.74 (Integration by Parts)**.** *Let μ and ν be σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$. Let $0 \leq a \leq b$. Then*

$$\begin{aligned} &\int_{(a,b]} \mu((0,y])d\nu(y) + \int_{(a,b]} \nu((0,x])d\mu(x) \\ &\quad = \int_{(a,b]} \mu(\{y\})d\nu(y) + \mu((0,a])\nu((a,b]) + \mu((a,b])\nu((0,a]) + \mu((a,b])\nu((a,b]). \end{aligned} \tag{2.8}$$

In particular, with $(0,0]$ defined as $\emptyset$,

$$\begin{aligned} &\int_{(0,b]} \mu((0,y])d\nu(y) + \int_{(0,b]} \nu((0,x])d\mu(x) \\ &\qquad = \int_{(0,b]} \mu(\{y\})d\nu(y) + \mu((0,b])\nu((0,b]). \end{aligned} \tag{2.9}$$

Proof. Let $H(x,y) = \chi_{(0,y]}(x)\chi_{(a,b]}(y) + \chi_{(a,b]}(x)\chi_{(0,x]}(y)$. Let $L(x,y) = \chi_{\{y\}}(x)\chi_{(a,b]}(y) + \chi_{(0,a]}(x)\chi_{(a,b]}(y) + \chi_{(a,b]}(x)\chi_{(0,a]}(y) + \chi_{(a,b]}(x)\chi_{(a,b]}(y)$. It can be easily verified that $H(x,y) = L(x,y)$ for all $(x,y) \in \mathbb{R}^+ \times \mathbb{R}^+$. Moreover, H and L are clearly $\mathcal{B}^+ \times \mathcal{B}^+$ measurable. Then, by Tonelli's Theorem 2.52,

$$
\begin{aligned}
&\int_{\mathbb{R}^+\times\mathbb{R}^+} H(x,y)d(\mu\times\nu)(x,y)\\
&=\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\left(\chi_{(0,y]}(x)\chi_{(a,b]}(y)+\chi_{(a,b]}(x)\chi_{(0,x]}(y)\right)d\mu(x)d\nu(y)\\
&=\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\int_{\mathbb{R}^+}\chi_{(0,y]}(x)d\mu(x)d\nu(y)+\int_{\mathbb{R}^+}\chi_{(a,b]}(x)\int_{\mathbb{R}^+}\chi_{(0,x]}(y)d\nu(y)d\mu(x)\\
&=\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\mu((0,y])d\nu(y)+\int_{\mathbb{R}^+}\chi_{(a,b]}(x)\nu((0,x])d\mu(x)\\
&=\int_{(a,b]}\mu((0,y])d\nu(y)+\int_{(a,b]}\nu((0,x])d\mu(x).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\int_{\mathbb{R}^+\times\mathbb{R}^+} L(x,y)d(\mu\times\nu)(x,y)\\
&=\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\big(\chi_{\{y\}}(x)\chi_{(a,b]}(y)+\chi_{(0,a]}(x)\chi_{(a,b]}(y)\\
&\qquad+\chi_{(a,b]}(x)\chi_{(0,a]}(y)+\chi_{(a,b]}(x)\chi_{(a,b]}(y)\big)d\mu(x)d\nu(y)\\
&=\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\int_{\mathbb{R}^+}\chi_{\{y\}}(x)d\mu(x)d\nu(y)+\int_{\mathbb{R}^+}\chi_{(0,a]}(x)\int_{\mathbb{R}^+}\chi_{(a,b]}(y)d\nu(y)d\mu(x)\\
&\qquad+\int_{\mathbb{R}^+}\chi_{(a,b]}(x)\int_{\mathbb{R}^+}\chi_{(0,a]}(y)d\nu(y)d\mu(x)+\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\int_{\mathbb{R}^+}\chi_{(a,b]}(x)d\mu(x)d\nu(y)\\
&=\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\mu(\{y\})d\nu(y)+\int_{\mathbb{R}^+}\chi_{(0,a]}(x)\nu((a,b])d\mu(x)\\
&\qquad+\int_{\mathbb{R}^+}\chi_{(a,b]}(x)\nu((0,a])d\mu(x)+\int_{\mathbb{R}^+}\chi_{(a,b]}(y)\mu((a,b])d\nu(y)\\
&=\int_{(a,b]}\mu(\{y\})d\nu(y)+\mu((0,a])\nu((a,b])+\mu((a,b])\nu((0,a])+\mu((a,b])\nu((a,b]).
\end{aligned}
$$

The result follows by the equality of H and L. The particular case follows from the fact that the measure of the empty set is 0. ∎

Here now are some examples of the use of integration by parts. We will come across examples such as these time and again in the proofs of the PNT.

Example 2.75. Let $0 \leq a < b$. Let F and G be Type 1 summatory functions with corresponding arithmetic functions f and g, respectively. Let λ_F and λ_G be the measures induced by F and G, respectively. Then application of Theorem 2.74 reveals

$$\sum_{a<n\leq b} \lambda_F((0,n])g(n) + \sum_{a<n\leq b} \lambda_G((0,n])f(n)$$
$$= \sum_{a<n\leq b} f(n)g(n) + \lambda_F((0,b])\lambda_G((0,b]) - \lambda_F((0,a])\lambda_G((0,a]).$$

□

Example 2.76. Let $0 \leq a < b$. Let F be the Type 1 summatory function with corresponding arithmetic function f. Let λ_F be the measure induced by F. Let λ be Lebesgue measure. Then application of Theorem 2.74 reveals

$$\sum_{a<n\leq b} nf(n) + \int_{(a,b]} \lambda_F((0,x])d\lambda(x) = b\lambda_F((0,b]) - a\lambda_F((0,a]).$$

□

We will require the following definition and propositions in order to furnish our next two examples on the use of integration by parts.

Definition 2.77. For $t \in \mathbb{R}^+$ and $f\colon \mathbb{N} \to \mathbb{R}$, define $\widetilde{f}(t) = \begin{cases} f(t), & t \in \mathbb{N}; \\ 0, & t \notin \mathbb{N}. \end{cases}$

Proposition 2.78. *Let $\widetilde{f}$ be as in Definition 2.77. Then $\widetilde{f}$ is $\mathcal{B}^+$-measurable.*

Proof. Let O be open in $\mathbb{R}$. We look at two cases.
(i) Suppose O does not contain 0. Then $\widetilde{f}^{-1}(O) = \{n \in \mathbb{N} : f(n) \in O\} \in \mathcal{B}^+$ since it is a closed set.
(ii) Suppose O contains 0. Then $\widetilde{f}^{-1}(O) = \{n \in \mathbb{N} : f(n) \in O\} \cup \mathbb{N}^c \in \mathcal{B}^+$ since it is the union of a closed set and an open set.
Hence, by Proposition 2.50, $\widetilde{f}$ is $\mathcal{B}^+$-measurable. ■

Proposition 2.79. *Let $f\colon \mathbb{N} \to \mathbb{R}^+ \cup \{0\}$ and F the Type 1 summatory function induced by f. Let $\widetilde{f}$ be as in Definition 2.77. Then*

$$\lambda_F = \int \widetilde{f}(t)d\lambda_N(t) \text{ on } (\mathbb{R}^+, \mathcal{B}^+) \text{ and } d\lambda_F/d\lambda_N = \widetilde{f}, \quad \lambda_N\text{-a.e.}$$

Proof. Let $E \in \mathcal{B}^+$. Then

$$\lambda_F(E) = \sum_{n \in E} f(n) = \int_E \widetilde{f}(t) d\lambda_N(t).$$

The first result follows. The second result follows by Theorem 2.43 since both measures are σ-finite and clearly $\lambda_F \ll \lambda_N$.

■

Proposition 2.80. *Let μ and ν be σ-finite measures on the measurable space $(\Omega, \mathcal{A})$. Let $E \in \mathcal{A}$. Suppose $\mu \ll \nu$ and $f \in \mathcal{L}^1(E, \mu)$. Then*

$$f\ d\mu/d\nu \in \mathcal{L}^1(E, \nu) \text{ and } \int_E f(t) d\mu(t) = \int_E f(t)(d\mu/d\nu)(t) d\nu(t).$$

Proof. Omitted. ■

Example 2.81. From Example 2.39, we defined $\lambda_M(E) = \lambda_Q(E) - \lambda_{Q-M}(E)$ for $E \in \mathcal{B}^+$ provided the right side exists. μ is the symbol for the Möbius function in this example. Now, by Proposition 2.79, $d\lambda_Q/d\lambda_N = \widetilde{|\mu|},\ \ \lambda_N$-a.e. and $d\lambda_{Q-M}/d\lambda_N = \widetilde{|\mu| - \mu} = \widetilde{|\mu|} - \widetilde{\mu},\ \ \lambda_N$-a.e.

Let ν be a σ-finite measure on $(\mathbb{R}^+, \mathcal{B}^+)$ such that $\nu((0, b]) < \infty$. Let $0 \leq a \leq b < \infty$. Then

$$\begin{aligned}
&\int_{(a,b]} \lambda_M((0,t]) d\nu(t) \\
&= \int_{(a,b]} \left(\lambda_Q((0,t]) - \lambda_{Q-M}((0,t])\right) d\nu(t) \\
&= \int_{(a,b]} \lambda_Q((0,t]) d\nu(t) - \int_{(a,b]} \lambda_{Q-M}((0,t]) d\nu(t) \\
&= \Bigg(- \int_{(a,b]} \nu((0,s]) d\lambda_Q(s) + \int_{(a,b]} \lambda_Q(\{t\}) d\nu(t) \\
&\qquad + \lambda_Q((0,a])\nu((a,b]) + \lambda_Q((a,b])\nu((0,a]) + \lambda_Q((a,b])\nu((a,b]) \Bigg) \\
&\quad - \Bigg(- \int_{(a,b]} \nu((0,s]) d\lambda_{Q-M}(s) + \int_{(a,b]} \lambda_{Q-M}(\{t\}) d\nu(t) \\
&\qquad + \lambda_{Q-M}((0,a])\nu((a,b]) + \lambda_{Q-M}((a,b])\nu((0,a]) + \lambda_{Q-M}((a,b])\nu((a,b]) \Bigg)
\end{aligned}$$

by integration by parts

$$= -\left(\int_{(a,b]} \nu((0,s])\widetilde{|\mu|}(s)d\lambda_N(s) - \int_{(a,b]} \nu((0,s])\left(\widetilde{|\mu|}(s) - \widetilde{\mu}(s)\right)d\lambda_N(s)\right)$$
$$+ \int_{(a,b]} \lambda_M(\{t\})d\nu(t) + \lambda_M((0,a])\nu((a,b]) + \lambda_M((a,b])\nu((0,a]) + \lambda_M((a,b])\nu((a,b])$$

by Definition 2.77 and Proposition 2.80

$$= -\int_{(a,b]} \nu((0,s])\widetilde{\mu}(s)d\lambda_N(s) + \int_{(a,b]} \lambda_M(\{t\})d\nu(t)$$
$$+ \lambda_M((0,a])\nu((a,b]) + \lambda_M((a,b])\nu((0,a]) + \lambda_M((a,b])\nu((a,b]).$$

□

Example 2.82.

a) Let $0 \leq a \leq b < \infty$.

b) Let ν and ρ be σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$ such that $\nu((0,b]) < \infty$.

c) Let $h\colon \mathbb{R}^+ \to \mathbb{R}$ such that $h \in \mathcal{L}^1([a,b],\rho)$.

d) Define $g\colon \mathbb{R}^+ \to \mathbb{R}$ by $g(t) = \chi_{[a,b]}(t)h(t)$.

Then $g(t) \in \mathcal{L}^1(\mathbb{R}^+,\rho)$. In addition,

e) let $\mu(A) = \int_A g(t)d\rho(t)$ for $A \in \mathcal{B}^+$; and,

f) let $S = \{x \in \mathbb{R}^+ : g(x) \geq 0\}$.

Note that μ is not necessarily a measure.

By Proposition 2.49,

1) $\chi_S g,\ -\chi_{S^c} g \in \mathcal{L}^1(\mathbb{R}^+,\rho)$,

2) $\mu_1 = \int \chi_S(t)g(t)d\rho(t)$ and $\mu_2 = \int -\chi_{S^c}(t)g(t)d\rho(t)$ are σ-finite measures on $(\mathbb{R}^+, \mathcal{B}^+)$,

3) $\mu_1(A) - \mu_2(A) = \mu(A)$ for all $A \in \mathcal{B}^+$, and

4) $d\mu_1/d\rho = \chi_S g,\ \ \rho$-a.e. and $d\mu_2/d\rho = -\chi_{S^c} g,\ \ \rho$-a.e. Using the same series of steps as in Example 2.81 reveals

$$\int_{(a,b]} \mu((0,t])d\nu(t) = -\int_{(a,b]} \nu((0,s])h(s)d\rho(s) + \int_{(a,b]} \mu(\{t\})d\nu(t)$$
$$+ \mu((0,a])\nu((a,b]) + \mu((a,b])\nu((0,a]) + \mu((a,b])\nu((a,b]).$$

□

Example 2.83. The following result will be used in Sections 3.3 and 3.5. Let $\bar{\mu}$ be as in Definition 2.73 and $x \geq 1$. Then

$$\begin{aligned}
L\bar{\mu}((0,x]) &= \sum_{i=1}^{k} \alpha_i L\mu_i((0,x]) = \sum_{i=1}^{k} \alpha_i \int_{(0,x]} \chi_{[1,\infty)}(t) \log t \, d\mu_i(t) \\
&= \sum_{i=1}^{k} \alpha_i \left(- \int_{(0,x]} \mu_i((0,t])t^{-1} d\lambda_1(t) + (\log x)\mu_i((0,x]) \right) \text{ by int. by parts} \\
&= - \int_{(0,x]} \sum_{i=1}^{k} \alpha_i \mu_i((0,t])t^{-1} d\lambda_1(t) + (\log x) \sum_{i=1}^{k} \alpha_i \mu_i((0,x]) \\
&= - \int_{(0,x]} \bar{\mu}((0,t])t^{-1} d\lambda_1(t) + (\log x)\bar{\mu}((0,x]) \\
&= - \int_{[1,x]} \bar{\mu}((0,t])t^{-1} d\lambda_1(t) + (\log x)\bar{\mu}((0,x]) \\
&= - \int_{[1,x]} \bar{\mu}((0,u^{-1}x])u^{-1} d\lambda_1(u) + (\log x)\bar{\mu}((0,x]) \text{ with change of var. } u = \frac{x}{t} \\
&= -\bar{\mu} * T^{-1}\lambda_1((0,x]) + (\log x)\bar{\mu}((0,x]).
\end{aligned}$$

□

Here is another important example to which we will refer in Section 3.5.

Example 2.84. Let $\bar{\mu}$ be as in Definition 2.73 and $x \geq 1$. In addition, let $\nu \in \mathscr{M}$. Then, for $0 \leq a \leq b < \infty$,

$$\begin{aligned}
&\int_{(a,b]} \bar{\mu}((0,t]) d\nu(t) = \int_{(a,b]} \sum_{i=1}^{k} \alpha_i \mu_i((0,t]) d\nu(t) = \sum_{i=1}^{k} \alpha_i \int_{(a,b]} \mu_i((0,t]) d\nu(t) \\
&= \sum_{i=1}^{k} \alpha_i \Bigg(- \int_{(a,b]} \nu((0,t]) d\mu_i(t) + \int_{(a,b]} \mu_i(\{t\}) d\nu(t) \\
&\qquad + \mu_i((0,a])\nu((a,b]) + \mu_i((a,b])\nu((0,a]) + \mu_i((a,b])\nu((a,b]) \Bigg) \\
&= - \sum_{i=1}^{k} \alpha_i \int_{(a,b]} \nu((0,t]) d\mu_i(t) + \int_{(a,b]} \sum_{i=1}^{k} \alpha_i \mu_i(\{t\}) d\nu(t) \\
&\qquad + \sum_{i=1}^{k} \alpha_i \mu_i((0,a])\nu((a,b]) + \sum_{i=1}^{k} \alpha_i \mu_i((a,b])\nu((0,a]) + \sum_{i=1}^{k} \alpha_i \mu_i((a,b])\nu((a,b]) \\
&= - \sum_{i=1}^{k} \alpha_i \int_{(a,b]} \nu((0,t]) d\mu_i(t) + \int_{(a,b]} \bar{\mu}(\{t\}) d\nu(t) \\
&\qquad + \bar{\mu}((0,a])\nu((a,b]) + \bar{\mu}((a,b])\nu((0,a]) + \bar{\mu}((a,b])\nu((a,b]).
\end{aligned}$$

□

Our final result demonstrating integration by parts will be Euler's summation formula. We will state it as a theorem; however, we require one further definition to set the scene.

Definition 2.85 (Absolutely Continuous Function)**.** Let $0 < a \leq b < \infty$. Let $f\colon \mathbb{R}^+ \to \mathbb{R}$; f' exist λ-a.e. on $[a, b]$; $f' \in \mathcal{L}^1([a, b], \lambda)$; and

$$f(x) = f(a) + \int_{[a,x]} f'(t)d\lambda(t), \;\; a \leq x \leq b.$$

Then f is called ***absolutely continuous*** on $[a, b]$.

Here now is Euler's summation formula set in the context of measure theory. We will appeal to it on multiple occasions during the proofs of the PNT.

Theorem 2.86 (Euler's Summation Formula)**.** *Let* $0 < a \leq b < \infty$*. Let* f *be absolutely continuous on* $[a, b]$*. Then*

$$\sum_{a<n\leq b} f(n) = \int_{(a,b]} f(t)d\lambda(t) - (t - [t])f(t)\Big|_a^b + \int_{(a,b]} f'(t)(t - [t])d\lambda(t).$$

Proof. Noting that $\lambda_N((0, s]) = [s]$ and substituting f' for h; λ_N for ν; and λ for ρ in Example 2.82 yields

$$\int_{(a,b]} (f(t) - f(a))d\lambda_N(t) = -\int_{(a,b]} [s]f'(s)d\lambda(s) + (f(b) - f(a))[b].$$

Again, this time substituting f' for h; λ for ν; and λ for ρ in Example 2.82 yields

$$\int_{(a,b]} (f(t) - f(a))d\lambda(t) = -\int_{(a,b]} sf'(s)d\lambda(s) + (f(b) - f(a))b.$$

Then, taking the difference between the two equalities and rearranging gives the desired result. ■

We now possess the relevant background material of number theory and measure theory. The concepts discussed in this section shall be instrumental in proving the PNT in the upcoming sections.

Chapter 3

An Elementary Proof of the Prime Number Theorem

The proof we present here of the PNT is by a method which has come to be known as "elementary". This simply means that mathematics of a higher level than calculus, such as complex analysis and Fourier analysis, are not used.

Alte Selberg was the first, along with Paul Erdős at about the same time, to succeed in proving the PNT in this manner. Indeed, his accomplishment contributed to his winning the Field's Medal in 1950. His proof, and just about all later "elementary" proofs, are based on his discovery of the so called "Selberg Formulae", which can be written as

$$(L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) = 2x\log x + O(x)$$

or

$$\psi(x)\log x + \int_{[1,x]} \lambda_\psi((0,t^{-1}x])d\lambda_\psi(t) = 2x\log x + O(x) \text{ for } x \geq 1.$$

Since that time, many improvements have been made to Selberg's original proof by several mathematicians. We borrow some of their ideas in the proof we now present.

In order to prove the PNT in the most efficient and clear fashion, we incorporate elements of Selberg's original proof [7] and the modifications of Selberg's proof by T. Tatuzawa and K. Iseki [8], N. Levinson [4], and P. Bateman and H. Diamond [2]. The similarities of our proof and the others ends there, however, as we incorporate our own modifications and, of course, invoke the elements of measure theory.

The proof consists of the following six sections:

(3.1) State and prove preliminary results.

(3.2) Prove

$$\psi(x) - x = o(x) \iff \text{PNT.}$$

(3.3) Derive Selberg's formulae.

(3.4) Define a continuous and piecewise differentiable function $\Xi(x)$ and show

$$\Xi(x) = o(x) \iff \psi(x) - x = o(x).$$

(3.5) Using Selberg's formulae, prove

$$\limsup_{x\to\infty} \frac{|\Xi(x)|}{x} = \limsup_{x\to\infty} \frac{1}{\log x}\int_{[1,x]} \frac{|\Xi(t)|}{t^2} d\lambda_1(t).$$

(3.6) Show

$$\limsup_{x\to\infty} \frac{1}{\log x}\int_{[1,x]} \frac{|\Xi(t)|}{t^2} d\lambda_1(t) = 0, \text{ thus proving the PNT.}$$

For convenience, we restate most of the measures we will use. See Section 2.1 for a description of the arithmetic functions. Also note that we will work exclusively in the measurable space $(\mathbb{R}^+, \mathcal{B}^+)$ for the remainder of the paper.

λ denotes the Lebesgue measure restricted to $\mathcal{B}^+$.

$\lambda_0 =$ the null measure ($\lambda_0(E) = 0$ for all $E \in \mathcal{B}^+$).

$\lambda_n = \int \chi_{[n,\infty)}(t) d\lambda(t)$ for $n \in \mathbb{N}$.

$\lambda_1 = \int \chi_{[1,\infty)}(t) d\lambda(t) = \lambda_{|_{[1,\infty)}}$.

$\lambda_2 = \int \chi_{[2,\infty)}(t) d\lambda(t)$.

$T^n\lambda_1 = \int t^n d\lambda_1(t)$. for $n \in \mathbb{Z}$.

$T^{-1}\lambda_1 = \int t^{-1} d\lambda_1(t)$.

λ_N denotes the Borel measure induced by $N(x) := \sum_{n \le x} 1(n)$.

λ_Q denotes the Borel measure induced by $Q(x) := \sum_{n \le x} |\mu|(n)$.

λ_{Q-M} denotes the Borel measure induced by $(Q - M)(x) := \sum_{n \le x} (|\mu| - \mu)(n)$.

λ_ψ denotes the Borel measure induced by $\psi(x) := \sum_{n \le x} \Lambda(n)$.

λ_π denotes the Borel measure induced by $\pi(x) := \sum_{n \le x} 1_{\mathcal{P}}(n)$ where $\mathcal{P} = \{\text{primes}\}$.

λ_ϑ denotes the Borel measure induced by $\vartheta(x) := \sum_{n \le x} \log n \cdot 1_{\mathcal{P}}(n)$.

δ_1 denotes the Borel measure induced by $\chi(x) := \sum_{n \le x} e(n)$.

$L^n \lambda_N = \int \chi_{[1,\infty)}(t) \log^n t \, d\lambda_N(t)$ for $n \in \mathbb{N}$.

$L^n \lambda_\psi = \int \chi_{[1,\infty)}(t) \log^n t \, d\lambda_\psi(t)$ for $n \in \mathbb{N}$.

$L\lambda_N = \int \chi_{[1,\infty)}(t) \log t \, d\lambda_N(t)$.

$L\lambda_\psi = \int \chi_{[1,\infty)}(t) \log t \, d\lambda_\psi(t)$.

$L\lambda_1 = \int \chi_{[1,\infty)}(t) \log t \, d\lambda_1(t)$.

A special case: Although $M(x) := \sum_{n \le x} \mu(n)$ is not a Type 1 summatory function (it is a Type 2) since $\mu(n)$ is not always nonnegative, we set $\lambda_M = \lambda_Q - \lambda_{Q-M}$ and define $\lambda_M(E) = \lambda_Q(E) - \lambda_{Q-M}(E)$ for $E \in \mathcal{B}^+$ provided the right side exists. Note that λ_M is not a measure.

All of the above are members of $\mathscr{L}(\mathcal{E})$ except for Lebesgue measure λ.

3.1 Preliminary Results

This first section is devoted to proving three lemmas that will be required in the proofs of the PNT. In so doing, this will be the first opportunity to utilize the measure theoretic machinery that was constructed and demonstrated in Chapter 2.

Lemma 3.1 (Chebyshev's Identity)**.**

$$L\lambda_N = \lambda_\psi * \lambda_N.$$

Proof. Let $E \in \mathcal{B}^+$ and let $n = p_1^{e_1} \cdots p_r^{e_r}$ be the unique prime factorization, or UPF, of n.

$$L\lambda_N(\{n\}) = (\log n)(1(n)) = \log n \text{ by (2.1)}.$$

By Theorem 2.58,

$$\begin{aligned}\lambda_\psi * \lambda_N(\{n\}) = \Lambda * 1(n) &= \sum_{ij=n} 1(i)\Lambda(j) = \sum_{ij=n} \Lambda(j) \\ &= \sum_{m=1}^{r} \sum_{l=1}^{e_m} \log p_m = \sum_{m=1}^{r} e_m \log p_m \\ &= \sum_{m=1}^{r} \log p_m^{e_m} = \log p_1^{e_1} \cdots p_r^{e_r} \\ &= \log n.\end{aligned}$$

Hence,

$$L\lambda_N(E) = \sum_{n\in E} L\lambda_N(\{n\}) = \sum_{n\in E} \lambda_\psi * \lambda_N(\{n\}) = \lambda_\psi * \lambda_N(E).$$

Since E is arbitrary, the result follows. ■

Lemma 3.2. *For $x \geq 1$,*

$$\psi(x) \leq x(1 + \log 2) + o(x) \tag{3.1}$$

and

$$\psi(x) = O(x). \tag{3.2}$$

Proof. We assume $x \geq 1$. Let $\nu = \int 2\chi_{(2,\infty)}(t)t^{-1}d\lambda(t)$. Then $\nu \in \mathscr{L}(\mathcal{E})$ and the Radon-Nikodym derivative $d\nu/d\lambda(t) = 2\chi_{(2,\infty)}(t)t^{-1}$, λ-a.e. Additionally, let $\rho = \lambda_N * (\delta_1 - \nu)$. Then $\rho \in \mathscr{L}(\mathcal{E})$.

Now, $\rho((0, x]) = 1$ for $x \in [1, 2)$. And for $x \geq 2$,

$$\begin{aligned}\rho((0,x]) &= \lambda_N * \delta_1((0,x]) - \lambda_N * \nu((0,x]) = \lambda_N((0,x]) - \int_{[1,x]} \lambda_N((0,t^{-1}x])d\nu(t) \\ &= \lambda_N((0,x]) - \int_{[1,x]} \lambda_N((0,t^{-1}x])2\chi_{(2,\infty)}(t)t^{-1}d\lambda(t) \\ &= [x] - 2\int_{[2,x]} \lambda_N((0,t^{-1}x])t^{-1}d\lambda(t) \\ &= [x] - 2\int_{[1,x/2]} \lambda_N((0,u])u^{-1}d\lambda(u) \text{ after a change of variable } u = x/t \\ &= \big([x] - (x-2)\big) + \left(2\int_{[1,x/2]} (u - [u])u^{-1}d\lambda(u)\right).\end{aligned}$$

Now, the first term has period 1 and is decreasing on [2,3) and the second term is nonnegative and increasing. Therefore,

$$\rho((0,x]) \geq \big([3^-] - (3^- - 2)\big) + 0 = 1 \text{ for all } x \geq 2$$

and so

$$\rho((0,x]) \geq 1 \text{ for all } x \geq 1.$$

Then, for $x \geq 1$,

$$\begin{aligned} \lambda_\psi * \lambda_N * (\delta_1 - \nu)((0,x]) = \lambda_\psi * \rho((0,x]) &= \int_{[1,x]} \rho((0,t^{-1}x])d\lambda_\psi(t) \\ &\geq \int_{[1,x]} 1 d\lambda_\psi(t) = \int_{(0,x]} 1 d\lambda_\psi(t) \\ &= \lambda_\psi((0,x]) = \psi(x). \end{aligned} \tag{3.3}$$

Moreover,

$$\begin{aligned} L\lambda_N * (\delta_1 - \nu)((0,x]) &= L\lambda_N((0,x]) - L\lambda_N * \nu((0,x]) \\ &= L\lambda_N((0,x]) - \int_{[1,x]} L\lambda_N((0,t^{-1}x])d\nu(t) \\ &= L\lambda_N((0,x]) - 2\int_{[1,x]} L\lambda_N((0,t^{-1}x])2\chi_{(2,\infty)}(t)t^{-1}d\lambda(t) \\ &= L\lambda_N((0,x]) - 2\int_{[2,x]} L\lambda_N((0,t^{-1}x])t^{-1}d\lambda(t) \\ &= L\lambda_N((0,x]) - 2\int_{[1,x/2]} L\lambda_N((0,u])u^{-1}d\lambda(u) \end{aligned} \tag{3.4}$$

after a change of variable $u = x/t$. Now, using Euler's summation formula Theorem 2.86, for $x \geq 1$,

$$\begin{aligned} L\lambda_N((0,x]) = L\lambda_N((1^-,x]) &= \sum_{n \leq x} \log n \\ &= \int_{(1^-,x]} \log s \, d\lambda(s) - (s - [s]) \log s \Big|_{1^-}^{x} + \int_{(1^-,x]} (s - [s]) \frac{1}{s} d\lambda(s) \\ &= x \log x - x + 1 - (x - [x]) \log x + \int_{(1^-,x]} (s - [s]) \frac{1}{s} d\lambda(s). \end{aligned}$$

Estimating the integral,

$$0 \leq \int_{(1^-,x]} (s - [s])\frac{1}{s} d\lambda(s) \leq \int_{(1^-,x]} \frac{1}{s} d\lambda(s) = \log x.$$

Therefore,

$$L\lambda_N((0,x]) = x \log x - x + O(\log x). \tag{3.5}$$

Bringing (3.4) and (3.5) together we then have

$$\begin{aligned} L\lambda_N * (\delta_1 - \nu)((0,x]) &= x \log x - x + O(\log x) - 2\int_{[1,x/2]} \big(u \log u - u + O(\log u)\big)\frac{1}{u} d\lambda(u) \\ &= x \log x - x + O(\log x) - 2\int_{[1,x/2]} \big(\log u - 1 + O(u^{-1}\log u)\big) d\lambda(u) \\ &= x \log x - x + O(\log x) - 2\left(\frac{x}{2}\log\frac{x}{2} - \frac{x}{2} + 1 - \frac{x}{2} + 1\right) + O(\log^2 x) \\ &= x \log 2 + x + O(\log^2 x). \end{aligned} \tag{3.6}$$

All together now using (3.3), (3.6), and Lemma 3.1,

$$\begin{aligned} \psi(x) &\leq \lambda_\psi * \lambda_N * (\delta_1 - \nu)((0,x]) \\ &= L\lambda_N * (\delta_1 - \nu)((0,x]) \\ &= x \log 2 + x + O(\log^2 x) \\ &= x(\log 2 + 1) + o(x). \end{aligned}$$

Furthermore, this implies

$$\psi(x) = O(x).$$

■

Lemma 3.3 (Mertens' Estimates)**.** *For* $x \geq 1$*,*

$$\int_{[1,x]} \frac{1}{t} d\lambda_\psi(t) = \log x + O(1) \tag{3.7}$$

and

$$\int_{[1,x]} \frac{\psi(t)}{t^2} d\lambda(t) = \log x + O(1). \tag{3.8}$$

Proof. Note all of the measures used in the following are in $\mathscr{L}(\mathcal{E})$.

$$\begin{aligned}
\int_{[1,x]} \frac{x}{t} d\lambda_\psi(t) &= (\lambda_1 + \delta_1) * \lambda_\psi((0,x]) \\
&= (\lambda_N + \lambda_1 - \lambda_N + \delta_1) * \lambda_\psi((0,x]) \\
&= \lambda_N * \lambda_\psi((0,x]) + (\lambda_1 - \lambda_N + \delta_1) * \lambda_\psi((0,x]) \\
&= L\lambda_N((0,x]) + \int_{[1,x]} (\lambda_1 - \lambda_N + \delta_1)((0,t^{-1}x])d\lambda_\psi(t) \text{ by Lem. 3.1} \\
&= x\log x - x + O(\log x) + \int_{[1,x]} (t^{-1}x - 1 - [t^{-1}x] + 1)d\lambda_\psi(t) \text{ by (3.5)} \\
&= x\log x + O(x) + \int_{[1,x]} O(1)d\lambda_\psi(t) \\
&= x\log x + O(x) + O(\lambda_\psi([1,x])) = x\log x + O(x) + O(\psi(x)) \\
&= x\log x + O(x) \text{ by (3.2).}
\end{aligned}$$

Therefore,

$$\int_{[1,x]} \frac{1}{t} d\lambda_\psi(t) = \log x + O(1)$$

and (3.7) is proved.

Now, define $\nu = \int t^{-2} d\lambda_1(t)$. Then $\nu \in \mathscr{L}(\mathcal{E})$ and $d\nu/d\lambda_1(t) = t^{-2},\ \lambda_1$-a.e. Utilizing the integration by parts formula reveals

$$\begin{aligned}
\int_{[1,x]} \frac{\psi(t)}{t^2} d\lambda(t) &= \int_{[1,x]} \frac{\psi(t)}{t^2} d\lambda_1(t) \\
&= \int_{[1,x]} \psi(t)d\nu(t) = \int_{(0,x]} \lambda_\psi((0,t])d\nu(t) \\
&= -\int_{(0,x]} \nu((0,t])d\lambda_\psi(t) + \lambda_\psi((0,x])\nu((0,x]) \text{ by Th. 2.74} \\
&= -\int_{(0,x]} (1 - t^{-1})d\lambda_\psi(t) + \psi(x)(1 - x^{-1}) \\
&= \int_{(0,x]} t^{-1} d\lambda_\psi(t) - \psi(x)x^{-1} = \int_{[1,x]} t^{-1} d\lambda_\psi(t) - \psi(x)x^{-1} \\
&= \log x + O(1) \text{ by (3.2) and (3.7).}
\end{aligned}$$

Thus, (3.8) is proved. ∎

3.2 The Equivalence of $\psi(x) - x = o(x)$ and PNT

As the title of this section indicates, we prove the following theorem whose result is a key component of both of the proofs we present of the PNT.

Theorem 3.4. *$\psi(x) - x = o(x) \iff$ PNT.*

Proof. We will find bounds for $\frac{\pi(x)}{\frac{x}{\log x}}$ and apply the pinching principle. Let $x \geq 2$.

$$\begin{aligned}
\psi(x) = \lambda_\psi((0,x]) &= \sum_{n \leq x} \Lambda(n) = \sum_{\substack{p^e \leq x \\ e \geq 1}} \log p \\
&= \sum_{p \leq x} \sum_{1 \leq e \leq \frac{\log x}{\log p}} \log p = \sum_{p \leq x} \log p \sum_{1 \leq e \leq \frac{\log x}{\log p}} 1 \\
&= \sum_{p \leq x} \log p \left[\frac{\log x}{\log p}\right] \\
&\leq \log x \sum_{p \leq x} 1 \\
&= \pi(x) \log x.
\end{aligned}$$

Therefore,

$$\frac{\psi(x)}{x} \leq \frac{\pi(x)}{\frac{x}{\log x}}. \tag{3.9}$$

Next,

$$\begin{aligned}
\pi(x) = \lambda_\pi((0,x]) &= \sum_{p \leq x} 1 \leq \sum_{p \leq x} 1 + \sum_{\substack{p^e \leq x \\ e \geq 2}} \frac{1}{e} = \sum_{\substack{p^e \leq x \\ e \geq 1}} \frac{1}{e} \\
&= \sum_{\substack{p^e \leq x \\ e \geq 1}} \frac{\log p}{e \log p} = \sum_{\substack{p^e \leq x \\ e \geq 1}} \frac{\log p}{\log p^e} = \sum_{2 \leq n \leq x} \frac{\Lambda(n)}{\log n} \\
&= \int_{[2,x]} \frac{1}{\log t} d\lambda_\psi(t).
\end{aligned} \tag{3.10}$$

We want to be able to use integration by parts on (3.10); but, we first need to get it in the right form. With that in mind, let $\nu = \int t^{-1} \log^{-2} t \, d\lambda_2(t)$. Then $\nu \in \mathscr{L}(\mathcal{E})$ and $d\nu/d\lambda_2(t) = t^{-1} \log^{-2} t$, λ_2-a.e. Furthermore, $\nu((0,t]) = -\log^{-1} t + \log^{-1} 2$ for $t \geq 2$.

Inequality (3.10) then becomes

$$\begin{aligned}
\pi(x) &\le \int_{[2,x]} -\nu((0,t])d\lambda_\psi(t) + \frac{1}{\log 2}\int_{[2,x]} d\lambda_\psi(t) \\
&= -\int_{(0,x]} \nu((0,t])d\lambda_\psi(t) + \frac{1}{\log 2}\lambda_\psi([2,x]) \\
&= -\left(-\int_{(0,x]} \lambda_\psi((0,t])d\nu(t) + \nu((0,x])\lambda_\psi((0,x])\right) + \frac{1}{\log 2}\psi(x) \text{ by (2.9)} \\
&= \int_{(0,x]} \psi(t)t^{-1}\log^{-2} t\, d\lambda_2(t) - (-\log^{-1} x + \log^{-1} 2)\psi(x) + \frac{1}{\log 2}\psi(x) \\
&= \int_{[2,x]} \psi(t)t^{-1}\log^{-2} t\, d\lambda(t) + \frac{\psi(x)}{\log x}. \qquad (3.11)
\end{aligned}$$

We now must get a bound on $I := \int_{[2,x]} \psi(t)t^{-1}\log^{-2} t\, d\lambda(t)$. By (3.2),

$$0 \le I \le \int_{[2,x]} kt \cdot t^{-1}\log^{-2} t\, d\lambda(t) \text{ for some } k > 0.$$

This implies

$$0 \le \frac{I}{\frac{x}{\log x}} \le \frac{\int_{[2,x]} k\log^{-2} t\, d\lambda(t)}{\frac{x}{\log x}}.$$

By L'hospital's Rule,

$$\lim_{x\to\infty} \frac{\int_{[2,x]} k\log^{-2} t\, d\lambda(t)}{\frac{x}{\log x}} = \lim_{x\to\infty} \frac{k\frac{1}{\log^2 x}}{\frac{1}{\log x} - \frac{1}{\log^2 x}} = 0.$$

Therefore,

$$I = o\left(\frac{x}{\log x}\right).$$

Combining this with (3.11) yields

$$\pi(x) \le \frac{\psi(x)}{\log x} + o\left(\frac{x}{\log x}\right).$$

This implies

$$\frac{\pi(x)}{\frac{x}{\log x}} \le \frac{\psi(x)}{x} + o(1). \qquad (3.12)$$

Bringing (3.9) and (3.12) together gives us

$$\frac{\psi(x)}{x} \leq \frac{\pi(x)}{\frac{x}{\log x}} \leq \frac{\psi(x)}{x} + o(1) \tag{3.13}$$

or

$$\psi(x) \leq \pi(x)\log x \leq \psi(x) + o(x). \tag{3.14}$$

Now we are ready to show the result of the theorem. First, assume $\psi(x) = x + o(x)$. Then, using (3.13),

$$1 + o(1) \leq \frac{\pi(x)}{\frac{x}{\log x}} \leq 1 + o(1).$$

This implies

$$\pi(x) \sim \frac{x}{\log x} \text{ (i.e. PNT).}$$

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

$$\pi(x) = \frac{x}{\log x} + o(x \log^{-1} x).$$

This implies, using (3.14),

$$\begin{aligned}\psi(x) - x &\leq \left(\frac{x}{\log x} + o(x\log^{-1} x)\right)\log x - x \\ &= o(x)\end{aligned}$$

and

$$\begin{aligned}\psi(x) - x &\geq \left(\frac{x}{\log x} + o(x\log^{-1} x)\right)\log x - x + o(x) \\ &= o(x).\end{aligned}$$

Therefore,

$$\psi(x) - x = o(x).$$

■

3.3 Selberg's Formulae

The Selberg Formula written in the measure theoretic sense,

$$(L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) = 2x\log x + O(x),$$

is at the core of the elementary proofs. Actually, Selberg [7] derived the formula in terms of $\vartheta(x)$ for use in his proof. In 1951, T. Tatuzawa and K. Iseki [8] elegently proved the formula in terms of $\psi(x)$.

Due to the intimate relationship of $\Lambda(n)$ and the counting measure $1(n)$ through Chebyshev's identity (Lemma 3.1), $\psi(x)$ became the preferred function to use in proofs that succeeded Selberg's. Choosing $\psi(x)$ over $\vartheta(x)$ also has the added benefit of enabling the exploitation of Mertens' estimates (Lemma 3.3).

In the same spirit of brevity as with Tatuzawa and Iseki [8], we now present a concise proof of the Selberg formulae; but, of course, we will employ measure theory to achieve the goal.

Theorem 3.5 (The Selberg Formulae). *For $x \geq 1$,*

$$(L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) = 2x\log x + O(x). \tag{3.15}$$

$$\psi(x)\log x + \int_{[1,x]} \lambda_\psi((0,t^{-1}x])d\lambda_\psi(t) = 2x\log x + O(x). \tag{3.16}$$

Proof. Assume we are working exclusively in $\mathcal{E}$. Let $\rho = \sum_{i=1}^{n} \alpha_i\rho_i$ where $\alpha_i \in \mathbb{R}$, $\rho_i \in \mathscr{M}$. Then $\rho \in \mathscr{L}(\mathcal{E})$. In addition, let $\nu = L(\rho * \lambda_N) + \rho * \lambda_N * T^{-1}\lambda_1$. Then $\nu \in \mathscr{L}(\mathcal{E})$. On the one hand,

$$\nu((0,x]) = (\log x)\rho * \lambda_N((0,x]) \text{ by Example 2.83.} \tag{3.17}$$

On the other hand,

$$\begin{aligned}
\nu &= L\left(\left(\sum_{i=1}^{n} \alpha_i\rho_i\right) * \lambda_N\right) + \rho * \lambda_N * T^{-1}\lambda_1 \\
&= L\left(\sum_{i=1}^{n} \alpha_i(\rho_i * \lambda_N)\right) + \rho * \lambda_N * T^{-1}\lambda_1
\end{aligned}$$

$$= \sum_{i=1}^{n} \alpha_i L(\rho_i * \lambda_N) + \rho * \lambda_N * T^{-1}\lambda_1 \text{ by Definition 2.73}$$
$$= \sum_{i=1}^{n} \alpha_i (L\rho_i * \lambda_N + \rho_i * L\lambda_N) + \rho * \lambda_N * T^{-1}\lambda_1 \text{ by Corollary 2.72}$$
$$= \sum_{i=1}^{n} \alpha_i (L\rho_i * \lambda_N) + \sum_{i=1}^{n} \alpha_i (\rho_i * \lambda_\psi * \lambda_N) + \rho * \lambda_N * T^{-1}\lambda_1 \text{ by Lemma 3.1}$$
$$= L\rho * \lambda_N + \rho * \lambda_\psi * \lambda_N + \rho * T^{-1}\lambda_1 * \lambda_N. \tag{3.18}$$

Thus, recalling $\lambda_N * \lambda_M = \delta_1$,

$$\nu * \lambda_M = L\rho + \rho * \lambda_\psi + \rho * T^{-1}\lambda_1. \tag{3.19}$$

In particular,

$$\nu * \lambda_M((0, x]) = (\log x)\rho((0, x]) + \rho * \lambda_\psi((0, x]) \text{ by Example 2.83.} \tag{3.20}$$

But, also,

$$\nu * \lambda_M((0, x]) = \int_{[1,x]} \nu((0, t^{-1}x])d\lambda_Q(t) - \int_{[1,x]} \nu((0, t^{-1}x])d\lambda_{Q-M}(t). \tag{3.21}$$

Let us now be specific and define $\rho = \lambda_\psi - \lambda_1 + \gamma\delta_1$ where γ = Euler's constant $= 1 - \int_{[1,\infty)} (t - [t])t^{-2} d\lambda_1(t)$. In order to simplify (3.21) we need get an estimate for $\nu((0, x])$. Exploiting Lemma 3.1 once again yields

$$\nu((0, x]) = (\log x)\rho * \lambda_N((0, x]) \text{ by (3.17)}$$
$$= (\log x)\big(\lambda_\psi * \lambda_N((0, x]) - \lambda_1 * \lambda_N((0, x]) + \gamma\lambda_N((0, x])\big)$$
$$= (\log x)\left(\sum_{n \le x} \log n - x\sum_{n \le x} \frac{1}{n} + (\gamma + 1)x + O(1)\right) \tag{3.22}$$

after simplification. Multiple uses of Euler's summation formula then reveals

$$\nu((0, x]) = O(\log^2 x) = O(\sqrt{x}). \tag{3.23}$$

Inserting this into (3.21) brings

$$\begin{aligned}
|\nu * \lambda_M((0,x])| &\le \int_{[1,x]} |\nu((0,t^{-1}x])| d\lambda_Q(t) + \int_{[1,x]} |\nu((0,t^{-1}x])| d\lambda_{Q-M}(t) \\
&\le \int_{[1,x]} |\nu((0,t^{-1}x])| d\lambda_N(t) + 2\int_{[1,x]} |\nu((0,t^{-1}x])| d\lambda_N(t) \\
&= 3\int_{[1,x]} O(\sqrt{x/t}) d\lambda_N(t) \text{ by (3.23)} \\
&= O\left(3\sqrt{x}\sum_{n\le x}\frac{1}{\sqrt{n}}\right) \\
&= O\big(3\sqrt{x}(2\sqrt{x}+O(1))\big) \text{ using Euler's summation formula} \\
&= O(x). \qquad (3.24)
\end{aligned}$$

Bringing together (3.20), (3.24), and the definition of ρ gives

$$\begin{aligned}
&(\log x)(\lambda_\psi - \lambda_1 + \gamma\delta_1)((0,x]) + (\lambda_\psi - \lambda_1 + \gamma\delta_1) * \lambda_\psi((0,x]) = O(x) \\
\Rightarrow\ &\psi(x)\log x - (x-1)\log x + \gamma\log x + \int_{[1,x]} \lambda_\psi((0,t^{-1}x]) d\lambda_\psi(t) \\
&\quad - \int_{[1,x]} \lambda_1((0,t^{-1}x]) d\lambda_\psi(t) + \gamma\psi(x) = O(x) \\
\Rightarrow\ &\psi(x)\log x - (x-1)\log x + \gamma\log x + \int_{[1,x]} \lambda_\psi((0,t^{-1}x]) d\lambda_\psi(t) \\
&\quad - \int_{[1,x]} \left(\frac{x}{t} - 1\right) d\lambda_\psi(t) + \gamma\psi(x) = O(x) \\
\Rightarrow\ &\psi(x)\log x + \int_{[1,x]} \lambda_\psi((0,t^{-1}x]) d\lambda_\psi(t) = 2x\log x + O(x) \text{ using Lemmas 3.3 and 3.2} \\
&\qquad (3.25)
\end{aligned}$$

and (3.16) is proved.

Moreover,

$$\begin{aligned}
L\lambda_\psi((0,x]) &= \int_{(0,x]} \chi_{[1,\infty)}(t)\log t\, d\lambda_\psi(t) \\
&= \int_{(1,x]} \log t\, d\lambda_\psi(t) \\
&= -\int_{(1,x]} \lambda_\psi((0,t])t^{-1} d\lambda_1(t) + (\log x)\lambda_\psi((1,x]) \text{ by int. by parts} \\
&= -\int_{(1,x]} \psi(t)t^{-1} d\lambda_1(t) + (\log x)\psi(x). \qquad (3.26)
\end{aligned}$$

Estimating the integral,

$$\begin{aligned} 0 &\leq \int_{(1,x]} \psi(t)t^{-1}d\lambda_1(t) \\ &\leq \int_{(1,x]} k't \cdot t^{-1}d\lambda_1(t) \text{ for some k'>0 by (3.2)} \\ &= O(x). \end{aligned}$$

Combining this estimate with (3.26) gives

$$L\lambda_\psi((0,x]) = (\log x)\psi(x) + O(x). \tag{3.27}$$

Bringing together (3.25) and (3.27) gives

$$(L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) = 2x\log x + O(x)$$

and (3.15) is proved. ■

3.4 The Transformation of $\psi(x) - x$, Part I

Our ultimate goal is to show $\psi(x) - x = o(x)$ which, by Section 3.2, implies the PNT. However, $\psi(x) - x$ is a discontinuous function thereby making difficult the use of the tools of analysis. To overcome that impediment, we begin a two stage process in this section to transform $\psi(x) - x$ into something that will be responsive to the techniques that will be employed in Section 3.6. We first will define an element ξ of $\mathscr{L}(\mathcal{E})$ such that $\xi((0,x])$ is equivalent to $\psi(x) - x$ in the sense that

$$\xi((0,x]) = o(x) \iff \psi(x) - x = o(x).$$

Take note that $\psi(x) - x$ can be written as $(\lambda_\psi - \lambda_1 - \delta_1)((0,x])$ for $x \geq 1$. With that in mind, we define $\xi = (\lambda_\psi - \lambda_1 - \delta_1) * T^{-1}\lambda_1$. Now let $x \geq 1$. Then by Example 2.69,

$$\begin{aligned} \xi((0,x]) &= (\lambda_\psi - \lambda_1 - \delta_1) * T^{-1}\lambda_1((0,x]) \\ &= \int_{[1,x]} (\lambda_\psi - \lambda_1 - \delta_1)((0,u])u^{-1}d\lambda_1(u) \end{aligned}$$

$$= \int_{[1,x]} \frac{\psi(u) - u}{u} d\lambda_1(u). \tag{3.28}$$

For convenience of notation, define $\Xi(x) = \xi((0, x])$. We see $\Xi(x)$ is a type of average of $\psi(x) - x$ and that $\Xi(x)$ is continuous and piecewise differentiable. In other words, $\Xi(x)$ is what is known as a "smoothing" of $\psi(x) - x$. By Lemma 3.2, we note

$$|\Xi(x)| \leq \int_{[1,x]} \frac{ku}{u} d\lambda_1(u) \text{ for some } k > 0.$$

Therefore,

$$\Xi(x) = O(x). \tag{3.29}$$

For simplicity of notation in the remaining analysis, let us use the symbols $|\Xi|(x)$ and $|\Xi(x)|$ interchangeably. Now, to prove $\Xi(x) = o(x) \iff \psi(x) - x = o(x)$ we will first need the following Lemma.

Lemma 3.6. $\liminf_{x\to\infty} \frac{\psi(x)}{x} + \limsup_{x\to\infty} \frac{\psi(x)}{x} = 2.$

Proof. By Lemma 3.2 and since $\frac{\psi(x)}{x} \geq 0$,

$$0 \leq \liminf_{x\to\infty} \frac{\psi(x)}{x} \leq \limsup_{x\to\infty} \frac{\psi(x)}{x} \leq 1 + \log 2 < \infty. \tag{3.30}$$

Let $a = \liminf_{x\to\infty} \frac{\psi(x)}{x}$ and $A = \limsup_{x\to\infty} \frac{\psi(x)}{x}$. For $\epsilon > 0$, there exists $N > 1$ such that $\frac{\psi(y)}{y} \leq A + \epsilon$ for all $y \geq N$. Let $M = \sup_{1\leq y<N} \{\frac{\psi(y)}{y}\}$. Let $x \geq N$. Then $\frac{x}{t} \geq N \iff t \leq \frac{x}{N}$. From Selberg's formula (3.16) and Lemma 3.3,

$$\begin{aligned}
2x\log x + O(x) &= \psi(x)\log x + \int_{[1,x]} \lambda_\psi((0, t^{-1}x])d\lambda\psi(t) \\
&= \psi(x)\log x + \int_{[1,\frac{x}{N}]} \frac{\psi(\frac{x}{t})}{\frac{x}{t}} \frac{x}{t} d\lambda_\psi(t) + \int_{(\frac{x}{N},x]} \frac{\psi(\frac{x}{t})}{\frac{x}{t}} \frac{x}{t} d\lambda_\psi(t) \\
&\leq \psi(x)\log x + x(A+\epsilon) \int_{[1,\frac{x}{N}]} \frac{1}{t} d\lambda_\psi(t) + xM \int_{(\frac{x}{N},x]} \frac{1}{t} d\lambda_\psi(t) \\
&= \psi(x)\log x + x(A+\epsilon)(\log \frac{x}{N} + O(1)) + xM(\log x - \log \frac{x}{N} + O(1)) \\
&= \psi(x)\log x + (A+\epsilon)x\log x + O(x).
\end{aligned}$$

Dividing through by $x \log x$ implies

$$2 + O(\log^{-1} x) \leq \frac{\psi(x)}{x} + (A + \epsilon) + O(\log^{-1} x) \text{ for all } x \geq N.$$

In particular,

$$2 \leq (A + \epsilon) + \liminf_{x \to \infty} \frac{\psi(x)}{x}.$$

Therefore, since ϵ is arbitrary,

$$2 \leq a + A. \tag{3.31}$$

Similarly, with the same type of argument

$$2 + O(\log^{-1} x) \geq \frac{\psi(x)}{x} + (a + \epsilon) + O(\log^{-1} x) \text{ for all } x \geq \text{ some } N_1.$$

In particular,

$$2 \geq (a + \epsilon) + \limsup_{x \to \infty} \frac{\psi(x)}{x}$$

which implies, since ϵ is arbitrary,

$$2 \geq a + A. \tag{3.32}$$

Finally, bringing together (3.31) and (3.32) yields

$$a + A = 2.$$

■

Now, we are in position to prove the main result.

Theorem 3.7. $\Xi(x) = o(x) \iff \psi(x) - x = o(x)$.

Proof. ($\Leftarrow$) Let $\psi(x) - x = o(x)$. Therefore, for $\epsilon > 0$, there exists $N_1 > 1$ such that $\frac{|\psi(x)-x|}{x} < \frac{\epsilon}{2}$ for all $x \geq N_1$. Then by (3.28)

$$\begin{aligned} |\Xi|(x) &\leq \int_{[1,x]} \frac{|\psi(t) - t|}{t} d\lambda_1(t) \\ &= \int_{[1,N_1)} \frac{|\psi(t) - t|}{t} d\lambda_1(t) + \int_{[N_1,x]} \frac{|\psi(t) - t|}{t} d\lambda_1(t). \end{aligned}$$

Now, let $M = \int_{[1,N_1)} \frac{|\psi(t)-t|}{t} d\lambda_1(t)$. There exists $N_2 > 1$ such that $\frac{M}{x} < \frac{\epsilon}{2}$ for all $x \geq N_2$. Let $N = \max\{N_1, N_2\}$. Then

$$\frac{|\Xi|(x)}{x} \leq \frac{M}{x} + \frac{\int_{[N_1,x]} \frac{|\psi(t)-t|}{t} d\lambda_1(t)}{x} < \frac{\epsilon}{2} + \frac{\epsilon}{2}\frac{x-N_1}{x} \leq \epsilon \text{ for all } x \geq N.$$

Since ϵ is arbitrary, $\Xi(x) = o(x)$.

($\Rightarrow$) Let $\Xi(x) = o(x)$. Here we borrow from Levinson [4] for the first half of the proof for this implication. We will find upper and lower bounds for $\frac{\psi(x)}{x}$. We have $\frac{\Xi(x)}{x} = o(1)$. Therefore, for $1 > \epsilon > 0$, there exists $N > 1$ such that $-\epsilon^2 < \frac{\Xi(x)}{x} < \epsilon^2$ for all $x \geq N$. This implies

$$-\epsilon^2 x < \Xi(x) < \epsilon^2 x \text{ for all } x \geq N.$$

Now,

$$\int_{(x,x+x\epsilon]} \frac{\psi(t)-t}{t} d\lambda_1(t) = \Xi(x+x\epsilon) - \Xi(x) < \epsilon^2(x+x\epsilon) + \epsilon^2 x < 3\epsilon^2 x.$$

But, also,

$$\int_{(x,x+x\epsilon]} \frac{\psi(t)-t}{t} d\lambda_1(t) \geq \left(\frac{\psi(x)}{x+x\epsilon} - 1\right) \int_{(x,x+x\epsilon]} d\lambda_1(t) = \left(\frac{\psi(x)}{x+x\epsilon} - 1\right) x\epsilon.$$

Combining these last two inequalities brings

$$\frac{\psi(x)}{x} < (1+3\epsilon)(1+\epsilon) \text{ for all } x \geq N.$$

And so

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq (1+3\epsilon)(1+\epsilon).$$

Since ϵ is arbitrary,

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq 1. \tag{3.33}$$

We now depart from Levinson [4] for the reverse inequality by noticing that we can simply employ Lemma 3.6 as follows.

$$2 = \liminf_{x\to\infty} \frac{\psi(x)}{x} + \limsup_{x\to\infty} \frac{\psi(x)}{x} \leq \liminf_{x\to\infty} \frac{\psi(x)}{x} + 1.$$

Therefore,

$$\liminf_{x\to\infty} \frac{\psi(x)}{x} \geq 2 - 1 = 1. \tag{3.34}$$

Bringing together (3.33) and (3.34) we can say

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

and

$$\psi(x) - x = o(x).$$

■

3.5 The Transformation of $\psi(x) - x$, Part II

This section is devoted to the second stage of the transformation of $\psi(x) - x$. The next step will be to represent $\alpha := \limsup_{x\to\infty} \frac{|\Xi|(x)}{x}$ in an alternate way. Here, again, Selberg's formulae will prove vital. We require the next two lemmas before we can accomplish this, however. Let us first get a preliminary bound on α.

Lemma 3.8. $0 \leq \alpha \leq \log 2$.

Proof. Clearly, $\alpha \geq 0$. We know from Lemma 3.2 that

$$\psi(x) \leq (1 + \log 2)x + o(x),$$

which implies

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq (1 + \log 2).$$

Furthermore, we have

$$\psi(x) - x \leq x \log 2 + o(x)$$

which implies

$$\frac{\psi(x) - x}{x} \leq \log 2 + o(1)$$

which further implies

$$\limsup_{x\to\infty} \frac{\psi(x) - x}{x} \leq \log 2.$$

Moreover, by Lemma 3.6,

$$2 = \liminf_{x\to\infty} \frac{\psi(x)}{x} + \limsup_{x\to\infty} \frac{\psi(x)}{x}.$$

Therefore,

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \geq \liminf_{x\to\infty} \frac{\psi(x)}{x} \geq 2 - 1 - \log 2 = 1 - \log 2$$

and then

$$\limsup_{x\to\infty} \frac{\psi(x) - x}{x} = -1 + \limsup_{x\to\infty} \frac{\psi(x)}{x} \geq -\log 2.$$

Therefore,

$$\limsup_{x\to\infty} \frac{|\psi(x) - x|}{x} \leq \log 2.$$

This implies

$$\frac{|\psi(x) - x|}{x} \leq \log 2 + o(1).$$

So,

$$\begin{aligned}
\frac{|\Xi|(x)}{x} &\leq \frac{1}{x} \int_{[1,x]} \frac{|\psi(t) - t|}{t} d\lambda_1(t) \\
&\leq \frac{1}{x} \int_{[1,x]} (\log 2 + o(1)) d\lambda_1(t) \\
&= \frac{1}{x}(x-1)(\log 2 + o(1)) \\
&< \log 2 + o(1).
\end{aligned}$$

Then, passing to the limit yields

$$\alpha \leq \log 2.$$

■

Lemma 3.9. $|\Xi|$ *is absolutely continuous on* $[1, x]$ *for* $1 \leq x < \infty$.

Proof. We know

$$\begin{cases} \frac{\psi(t)-t}{t} \text{ is piecewise continuous on } [1,\infty) \text{ which implies } \Xi(x) \text{ is continuous on } [1,\infty). \\ \psi(t) \text{ is discontinuous only at } p^e \text{ (prime power, } e \geq 1). \\ \psi(t) \text{ is constant between consecutive prime powers.} \end{cases}$$

Now, let x be given, x not a prime power and ${p_1}^{e_1} < x < {p_2}^{e_2}$ where ${p_1}^{e_1}, {p_2}^{e_2}$ are consecutive prime powers (e.g. $3^2, 11^1$). Let $S = \{p^e : p \text{ prime}, e \in \mathbb{N}\}$. Then

$$\Xi(x) = \int_{[1,{p_1}^{e_1}]} \frac{\psi(t)-t}{t} d\lambda_1(t) + \int_{({p_1}^{e_1},x]} \frac{\psi(t)-t}{t} d\lambda_1(t).$$

Clearly, then, $\Xi'(x)$ exists and

$$\Xi'(x) = \frac{\psi(x)-x}{x}.$$

Therefore, for $x \notin S$,

$$\Xi'(x) = \frac{\psi(x)-x}{x}.$$

Moreover, this implies $\Xi''(x)$ exists for $x \notin S$ and equals $-\frac{\psi(x)}{x^2} < 0$ (recalling $\psi(x)$ is constant between consecutive prime powers). These facts imply $\Xi(x)$ is concave down in each $({p_1}^{e_1}, {p_2}^{e_2})$ which implies

$$\#\{\text{zeros of } \Xi(x) \text{ for } x \in ({p_1}^{e_1}, {p_2}^{e_2})\} \leq 2.$$

Now, let

$$R = \{x : \Xi(x) \text{ changes sign}\}.$$

Clearly,

$$R \subset \{x : \Xi(x) = 0\} =: V.$$

Then

$$|\Xi|'(x) \text{ exists} \iff x \in R^c \cap S^c =: W.$$

Or, put another way,

$$|\Xi|'(x) \text{ does not exist} \iff x \in R \cup S \; (\subset V \cup S).$$

We see by the preceding statements that $(R\cup S)\cap[1,x]$ is a finite set for any $1\le x<\infty$ and consequently $\lambda_1((R\cup S)\cap[1,x])=0$. Therefore, $|\Xi|'$ exists λ_1-a.e. on $[1,x]$.

Now, let $1\le t\le x$. Note since $t\ge 1$ we may use λ_1 instead of λ in Definition 2.85 (recalling $\lambda_1=\lambda_{|[1,\infty)}$). In addition, let $(R\cup S)\cap[1,t]=\{1,t_1,t_2,...,t_n\}$ in order from least to greatest. Then

$$\begin{aligned}
&\int_{[1,t]}|\Xi|'(u)d\lambda_1(u)\\
&=\int_{(R\cup S)^c\cap[1,t]}|\Xi|'(u)d\lambda_1(u)+\int_{(R\cup S)\cap[1,t]}|\Xi|'(u)d\lambda_1(u)\\
&=\int_{(R\cup S)^c\cap[1,t]}|\Xi|'(u)d\lambda_1(u)=\int_{(R\cup S)^c\cap(1,t)}|\Xi|'(u)d\lambda_1(u)\\
&=\int_{(1,t_1)}|\Xi|'(u)d\lambda_1(u)+\sum_{i=1}^{n-1}\int_{(t_i,t_{i+1})}|\Xi|'(u)d\lambda_1(u)+\int_{(t_n,t)}|\Xi|'(u)d\lambda_1(u)\\
&=\big(|\Xi|(t_1)-|\Xi|(1)\big)+\big(|\Xi|(t_n)-|\Xi|(t_1)\big)+\big(|\Xi|(t)-|\Xi|(t_n)\big)\\
&=|\Xi|(t)-|\Xi|(1)=|\Xi|(t)<\infty.
\end{aligned}$$

Therefore, $|\Xi|'\in\mathcal{L}^1([1,x],\lambda_1)$ and $|\Xi|(t)=|\Xi|(1)+\int_{[1,t]}|\Xi|'(u)d\lambda_1(u)$, $1\le t\le x$. Having satisfied the hypotheses of Definition 2.85, we conclude $|\Xi|$ is absolutely continuous on $[1,x]$ for $1\le x<\infty$. ■

In order to show $\limsup_{x\to\infty}\frac{|\Xi|(x)}{x}=0$, we want to find a relationship between $\frac{|\Xi|(x)}{x}$ and the more manageable function $\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)$. Our derivation is primarily based on Bateman and Diamond's [2] and Levinson's [4] modifications of Selberg's [7] original proof. In particular, we will exploit Levinson's idea of first "smoothing" each term in Selberg's formula. In contrast, however, we will employ the measure theoretic techniques developed in Chapter 2 rather than the standard methods utilized by these authors. Specifically, we have the following result.

Lemma 3.10. $\alpha=\limsup_{x\to\infty}\frac{1}{\log x}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)$.

Proof. Consider the following two expressions. The reason will shortly become apparent. Let $x\ge 1$.

$$\Big[L\Big((L\lambda_\psi+\lambda_\psi*\lambda_\psi)*T^{-1}\lambda_1-(L\lambda_1+\lambda_\psi*(\lambda_1+\delta_1))*T^{-1}\lambda_1\Big)$$
$$-\Big((L\lambda_\psi+\lambda_\psi*\lambda_\psi)*T^{-1}\lambda_1-(L\lambda_1+\lambda_\psi*(\lambda_1+\delta_1))*T^{-1}\lambda_1\Big)*\lambda_\psi\Big]((0,x]). \tag{3.35}$$

In accordance with Definition 2.73 and recalling $\xi=(\lambda_\psi-\lambda_1-\delta_1)*T^{-1}\lambda_1$, expression (3.35) can be written as

$$\Big[L\Big((L\lambda_\psi-L\lambda_1)*T^{-1}\lambda_1+\xi*\lambda_\psi\Big)-\Big((L\lambda_\psi-L\lambda_1)*T^{-1}\lambda_1+\xi*\lambda_\psi\Big)*\lambda_\psi\Big]((0,x]). \tag{3.36}$$

Our idea is to evaluate each of these expressions separately and then equate them. We begin by calculating (3.35). First, we note that for $u\geq 1$,

$$\begin{aligned}\lambda_\psi*(\lambda_1+\delta_1)((0,u])&=\int_{[1,u]}(\lambda_1+\delta_1)((0,t^{-1}u])d\lambda_\psi(t)=\int_{[1,u]}t^{-1}u\,d\lambda_\psi(t)\\&=u(\log u+O(1))=u\log u+O(u)\text{ by Lemma 3.3.}\end{aligned} \tag{3.37}$$

Next, we perform the "smoothing" on Selberg's formula. For $x\geq 1$,

$$\begin{aligned}(L\lambda_\psi+\lambda_\psi*\lambda_\psi)*T^{-1}\lambda_1((0,x])&=\int_{[1,x]}(L\lambda_\psi+\lambda_\psi*\lambda_\psi)((0,t^{-1}x])dT^{-1}\lambda_1(t)\\&=\int_{[1,x]}(L\lambda_\psi+\lambda_\psi*\lambda_\psi)((0,u])u^{-1}d\lambda_1(u)\text{ by Ex. 2.69}\\&=\int_{[1,x]}\big(2u\log u+O(u)\big)u^{-1}d\lambda_1(u)\text{ by Selberg's (3.15);}\end{aligned} \tag{3.38}$$

and

$$\begin{aligned}&(L\lambda_1+\lambda_\psi*(\lambda_1+\delta_1))*T^{-1}\lambda_1((0,x])\\&=\int_{[1,x]}(L\lambda_1+\lambda_\psi*(\lambda_1+\delta_1))((0,t^{-1}x])dT^{-1}\lambda_1(t)\\&=\int_{[1,x]}(L\lambda_1+\lambda_\psi*(\lambda_1+\delta_1))((0,u])u^{-1}d\lambda_1(u)\end{aligned}$$

$$= \int_{[1,x]} \big((u \log u - u + 1) + (u \log u + O(u))\big) u^{-1} d\lambda_1(u) \text{ by (3.37)}$$
$$= \int_{[1,x]} \big(2u \log u + O(u)\big) u^{-1} d\lambda_1(u). \quad (3.39)$$

Subtracting (3.39) from (3.38) yields, for $x \geq 1$,

$$\Big((L\lambda_\psi + \lambda_\psi * \lambda_\psi) * T^{-1}\lambda_1 - (L\lambda_1 + \lambda_\psi * (\lambda_1 + \delta_1)) * T^{-1}\lambda_1\Big)((0, x]) = O(x). \quad (3.40)$$

Then, using Example 2.83 and (3.40),

$$L\Big((L\lambda_\psi + \lambda_\psi * \lambda_\psi) * T^{-1}\lambda_1 - (L\lambda_1 + \lambda_\psi * (\lambda_1 + \delta_1)) * T^{-1}\lambda_1\Big)((0, x])$$
$$= -\int_{[1,x]} O(t) t^{-1} d\lambda_1(t) + (\log x) O(x) = O(x \log x). \quad (3.41)$$

Moreover, using (3.40),

$$\Big((L\lambda_\psi + \lambda_\psi * \lambda_\psi) * T^{-1}\lambda_1 - (L\lambda_1 + \lambda_\psi * (\lambda_1 + \delta_1)) * T^{-1}\lambda_1\Big) * \lambda_\psi((0, x])$$
$$= \int_{[1,x]} O(t^{-1}x) d\lambda_\psi(t) = O\left(x \int_{[1,x]} t^{-1} d\lambda_\psi(t)\right)$$
$$= O(x \log x) \text{ by Mertens' estimate, once again.} \quad (3.42)$$

Finally, by subtracting (3.42) from (3.41), expression (3.35) evaluates to

$$\Big[L\Big((L\lambda_\psi + \lambda_\psi * \lambda_\psi) * T^{-1}\lambda_1 - (L\lambda_1 + \lambda_\psi * (\lambda_1 + \delta_1)) * T^{-1}\lambda_1\Big)$$
$$- \Big((L\lambda_\psi + \lambda_\psi * \lambda_\psi) * T^{-1}\lambda_1 - (L\lambda_1 + \lambda_\psi * (\lambda_1 + \delta_1)) * T^{-1}\lambda_1\Big) * \lambda_\psi\Big]((0, x])$$
$$= O(x \log x). \quad (3.43)$$

On the other hand, we now evaluate (3.36). First, note that for $u \geq 1$,

$$\lambda_\psi * \xi((0, u]) = \int_{[1,u]} \xi((0, t^{-1}u]) d\lambda_\psi(t) = \int_{[1,u]} O(t^{-1}u)\, d\lambda_\psi(t) \text{ by (3.29)}$$
$$= O\left(u \int_{[1,u]} t^{-1}\, d\lambda_\psi(t)\right) = O(u \log u) \text{ by Lemma 3.3.} \quad (3.44)$$

Next, for $x \geq 1$,

$$
\begin{aligned}
&(L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,x]) \\
&= \int_{[1,x]} (L\lambda_\psi - L\lambda_1)((0,t^{-1}x])dT^{-1}\lambda_1(t) \\
&= \int_{[1,x]} (L\lambda_\psi - L\lambda_1)((0,u])u^{-1}d\lambda_1(u) \text{ again by Ex. 2.69} \\
&= \int_{[1,x]} \big((\psi(u)\log u + O(u)) - (u\log u - u + 1)\big)u^{-1}d\lambda_1(u) \text{ by (3.27)} \qquad (3.45) \\
&= \int_{[1,x]} \left(\log u \frac{\psi(u)-u}{u} + O(1)\right) d\lambda_1(u) = \int_{(1,x]} \left(\log u \frac{\psi(u)-u}{u}\right) d\lambda_1(u) + O(x) \\
&= -\int_{(1,x]} \xi((0,u])u^{-1}d\lambda_1(u) + (\log x)\xi((1,x]) + O(x) \Big(\text{by int. by parts Ex.2.82} \\
&\qquad \text{with } \lambda_1 \text{ as } \rho, \int u^{-1}d\lambda_1(u) \text{ as } \nu, \frac{\psi(u)-u}{u} \text{ as } h(u),\ (1,x] \text{ as } (a,b]\Big) \\
&= -\int_{(1,x]} O(1)d\lambda_1(u) + (\log x)\xi((0,x]) + O(x) \text{ again by (3.29)} \\
&= (\log x)\xi((0,x]) + O(x). \qquad (3.46)
\end{aligned}
$$

Then, employing Example 2.83 produces

$$
\begin{aligned}
&L\Big((L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1\Big)((0,x]) \\
&= -\int_{(0,x]} (L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,u])u^{-1}d\lambda_1(u) + (\log x)(L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,x]) \\
&= -\int_{[1,x]} (L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,u])u^{-1}d\lambda_1(u) + (\log x)(L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,x]) \\
&= -\int_{[1,x]} \big((\log u)\xi((0,u]) + O(u)\big)u^{-1}d\lambda_1(u) + (\log x)\big((\log x)\xi((0,x]) + O(x)\big) \text{ by (3.46)} \\
&= -\int_{[1,x]} \big((\log u)O(u) + O(u)\big)u^{-1}d\lambda_1(u) + (\log^2 x)\xi((0,x]) + O(x\log x) \text{ by (3.29)} \\
&= (\log^2 x)\xi((0,x]) + O(x\log x). \qquad (3.47)
\end{aligned}
$$

Using the same steps, we also have

$$
\begin{aligned}
&L(\xi * \lambda_\psi)((0,x]) \\
&= -\int_{(0,x]} \xi * \lambda_\psi((0,u])u^{-1}d\lambda_1(u) + (\log x)\xi * \lambda_\psi((0,x]) \\
&= -\int_{[1,x]} \xi * \lambda_\psi((0,u])u^{-1}d\lambda_1(u) + (\log x)\xi * \lambda_\psi((0,x])
\end{aligned}
$$

$$= -\int_{[1,x]} O(u\log u)u^{-1}d\lambda_1(u) + (\log x)\xi * \lambda_\psi((0,x]) \text{ by (3.44)}$$
$$= (\log x)\xi * \lambda_\psi((0,x]) + O(x\log x). \tag{3.48}$$

Adding (3.47) and (3.48) gives, for $x \geq 1$,

$$L\Big((L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1 + \xi * \lambda_\psi\Big)((0,x]) = (\log^2 x)\xi((0,x]) + (\log x)\xi * \lambda_\psi((0,x]) + O(x\log x). \tag{3.49}$$

Moreover,

$$\Big((L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1 + \xi * \lambda_\psi\Big) * \lambda_\psi((0,x])$$
$$= \int_{[1,x]} (L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1((0,t^{-1}x])d\lambda_\psi(t) + \xi * \lambda_\psi * \lambda_\psi((0,x])$$
$$= \int_{[1,x]} \big((\log(t^{-1}x))\xi((0,t^{-1}x]) + O(t^{-1}x)\big)d\lambda_\psi(t) + \xi * \lambda_\psi * \lambda_\psi((0,x]) \text{ by (3.46)}$$
$$= (\log x)\xi * \lambda_\psi((0,x]) - \int_{[1,x]} \xi((0,t^{-1}x])\log t\, d\lambda_\psi(t) + O(x\log x) + \xi * \lambda_\psi * \lambda_\psi((0,x])$$
$$= (\log x)\xi * \lambda_\psi((0,x]) - \int_{[1,x]} \xi((0,t^{-1}x])dL\lambda_\psi(t) + O(x\log x) + \xi * \lambda_\psi * \lambda_\psi((0,x])$$
$$= (\log x)\xi * \lambda_\psi((0,x]) - \xi * L\lambda_\psi((0,x]) + \xi * \lambda_\psi * \lambda_\psi((0,x]) + O(x\log x). \tag{3.50}$$

Then, by subtracting (3.50) from (3.49), expression (3.36) evaluates to

$$\Big[L\Big((L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1 + \xi * \lambda_\psi\Big) - \Big((L\lambda_\psi - L\lambda_1) * T^{-1}\lambda_1 + \xi * \lambda_\psi\Big) * \lambda_\psi\Big]((0,x])$$
$$= (\log^2 x)\xi((0,x]) + \xi * \big(L\lambda_\psi - \lambda_\psi * \lambda_\psi\big)((0,x]) + O(x\log x). \tag{3.51}$$

Finally, equating (3.43) with (3.51) yields

$$(\log^2 x)\xi((0,x]) + \xi * \big(L\lambda_\psi - \lambda_\psi * \lambda_\psi\big)((0,x]) = O(x\log x), \quad x \geq 1. \tag{3.52}$$

Continuing, rewrite (3.52) as

$$-\log^2 x \cdot \xi((0,x]) = \xi * L\lambda_\psi((0,x]) - \xi * \lambda_\psi * \lambda_\psi((0,x]) + O(x\log x)$$
$$\Rightarrow -\log^2 x \cdot \xi((0,x]) = \int_{[1,x]} \xi((0,t^{-1}x])dL\lambda_\psi(t) - \int_{[1,x]} \xi((0,t^{-1}x])d\lambda_\psi * \lambda_\psi(t) + O(x\log x)$$

$$\Rightarrow -\log^2 x \cdot \Xi(x) = \int_{[1,x]} \Xi(t^{-1}x) dL\lambda_\psi(t) - \int_{[1,x]} \Xi(t^{-1}x) d\lambda_\psi * \lambda_\psi(t) + O(x \log x)$$

$$\Rightarrow \log^2 x \cdot |\Xi|(x) \leq \int_{[1,x]} |\Xi|(t^{-1}x) dL\lambda_\psi(t) + \int_{[1,x]} |\Xi|(t^{-1}x) d\lambda_\psi * \lambda_\psi(t) + O(x \log x)$$

$$\Rightarrow \log^2 x \cdot |\Xi|(x) \leq \int_{[1,x]} |\Xi|(t^{-1}x) d(L\lambda_\psi + \lambda_\psi * \lambda_\psi)(t) + O(x \log x). \tag{3.53}$$

Referring to the proof of Lemma 3.9, let

$$f(u) = \begin{cases} |\Xi|'(u), & u \in (S \cup R)^c; \\ 0, & u \in S \cup R. \end{cases}$$

Then f is $\mathcal{B}^+$-measurable and $f = |\Xi|'$, λ_1-a.e. Now, define $\nu = \int \chi_{[1,x]}(u) f(u) d\lambda_1(u)$. By Proposition 2.49, there exist σ-finite measures ν_1 and ν_2 on $\mathcal{B}^+$ such that $\nu(A) = \nu_1(A) - \nu_2(A)$ for all $A \in \mathcal{B}^+$. Consequently, ν_1 and $\nu_2 \in \mathscr{M}$ and $\nu \in \mathscr{L}(\mathcal{E})$. Also, for $1 \leq t \leq x$, notice

$$\begin{aligned} \nu((0, t^{-1}x]) &= \int_{(0,t^{-1}x]} \chi_{[1,x]}(u) f(u) d\lambda_1(u) = \int_{[1,t^{-1}x]} f(u) d\lambda_1(u) \\ &= \int_{[1,t^{-1}x]} |\Xi|'(u) d\lambda_1(u) = |\Xi|(t^{-1}x) \text{ by Lemma 3.9.} \end{aligned} \tag{3.54}$$

We then have

$$\begin{aligned} &\int_{[1,x]} |\Xi|(t^{-1}x) d(L\lambda_\psi + \lambda_\psi * \lambda_\psi)(t) \\ &= \int_{[1,x]} \nu((0, t^{-1}x]) d(L\lambda_\psi + \lambda_\psi * \lambda_\psi)(t) \\ &= \int_{[1,x]} \nu_1((0, t^{-1}x]) d(L\lambda_\psi + \lambda_\psi * \lambda_\psi)(t) - \int_{[1,x]} \nu_2((0, t^{-1}x]) d(L\lambda_\psi + \lambda_\psi * \lambda_\psi)(t) \\ &= \nu_1 * (L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) - \nu_2 * (L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0,x]) \\ &= \int_{[1,x]} (L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0, t^{-1}x]) d\nu_1(t) - \int_{[1,x]} (L\lambda_\psi + \lambda_\psi * \lambda_\psi)((0, t^{-1}x]) d\nu_2(t) \\ &= \int_{[1,x]} \left(2\frac{x}{t} \log \frac{x}{t} + O(x/t)\right) d\nu_1(t) - \int_{[1,x]} \left(2\frac{x}{t} \log \frac{x}{t} + O(x/t)\right) d\nu_2(t) \text{ by Selberg} \\ &= \int_{[1,x]} 2\frac{x}{t} \log \frac{x}{t} d\nu_1(t) - \int_{[1,x]} 2\frac{x}{t} \log \frac{x}{t} d\nu_2(t) + \int_{[1,x]} O(x/t) d\nu_1(t) - \int_{[1,x]} O(x/t) d\nu_2(t) \end{aligned}$$

$$=: I_1 - I_2 + J_1 - J_2. \tag{3.55}$$

Now,

$$\begin{aligned}
I_1 - I_2 &= \int_{[1,x]} 2\frac{x}{t}\log\frac{x}{t}d\nu_1(t) - \int_{[1,x]} 2\frac{x}{t}\log\frac{x}{t}d\nu_2(t) \\
&= 2\int_{[1,x]} (L\lambda_1 + \lambda_1)((0,t^{-1}x])d\nu_1(t) - 2\int_{[1,x]} (L\lambda_1 + \lambda_1)((0,t^{-1}x])d\nu_2(t) \\
&= 2(L\lambda_1 + \lambda_1) * \nu_1((0,x]) - 2(L\lambda_1 + \lambda_1) * \nu_2((0,x]) \\
&= 2(L\lambda_1 + \lambda_1) * \nu((0,x]) \\
&= 2\int_{[1,x]} \nu((0,t^{-1}x])dL\lambda_1(t) + 2\int_{[1,x]} \nu((0,t^{-1}x])d\lambda_1(t) \\
&= 2\int_{[1,x]} \nu((0,t^{-1}x])dL\lambda_1(t) + 2\int_{[1,x]} |\Xi|(t^{-1}x)d\lambda_1(t) \text{ by (3.54)} \\
&= 2\int_{[1,x]} \nu((0,t^{-1}x])dL\lambda_1(t) + 2\int_{[1,x]} O(t^{-1}x)d\lambda_1(t) \text{ by (3.29)} \\
&= 2\int_{[1,x]} \nu((0,t^{-1}x])dL\lambda_1(t) + O(x\log x).
\end{aligned} \tag{3.56}$$

It remains to calculate $J_1 - J_2$. Employing Proposition 2.49 once again yields

$$\begin{aligned}
J_1 - J_2 &= \int_{[1,x]} O(x/t)d\nu_1(t) - \int_{[1,x]} O(x/t)d\nu_2(t) \\
&= \int_{[1,x]} O(x/t)(d\nu_1/d\lambda_1)(t)d\lambda_1(t) - \int_{[1,x]} O(x/t)(d\nu_2/d\lambda_1)(t)d\lambda_1(t) \\
&= \int_{[1,x]} O(x/t)\chi_{[1,x]}(t)f(t)d\lambda_1(t) \\
&= \int_{[1,x]} O(x/t)|\Xi|'(t)d\lambda_1(t).
\end{aligned} \tag{3.57}$$

Moreover, again referring to the proof of Lemma 3.9, a simple analysis reveals

$$|\Xi|'(t) = \begin{cases} \Xi'(t) = \frac{\psi(t)-t}{t} & \text{if } \Xi(t) > 0; \\ -\Xi'(t) = -\frac{\psi(t)-t}{t} & \text{if } \Xi(t) < 0, \end{cases} \quad \text{for } t \in (R \cup S)^c. \tag{3.58}$$

Then

$$|J_1 - J_2| \leq \int_{[1,x]} k\frac{x}{t}||\Xi|'(t)|d\lambda_1(t) \text{ for some } k > 0$$

$$= kx \int_{[1,x]} \frac{|\psi(t) - t|}{t^2} d\lambda_1(t) \text{ by (3.58)}$$
$$\leq kx \int_{[1,x]} \frac{\psi(t) + t}{t^2} d\lambda_1(t)$$
$$\leq kx \int_{[1,x]} \frac{k't + t}{t^2} d\lambda_1(t) \text{ for some } k' > 0 \text{ by Lemma 3.2}$$
$$= k(k'+1)x \log x$$
$$= O(x \log x). \tag{3.59}$$

Combining (3.53), (3.55) ,(3.56) and (3.59) produces

$$\log^2 x \cdot |\Xi|(x) \leq 2 \int_{[1,x]} \nu((0, t^{-1}x])dL\lambda_1(t) + O(x \log x)$$
$$= 2 \int_{[1,x]} |\Xi|(t^{-1}x) \log t \, d\lambda_1(t) + O(x \log x). \tag{3.60}$$

Letting $u = \frac{x}{t}$ and then dividing by $x \log^2 x$ gives

$$\frac{|\Xi|(x)}{x} \leq \frac{2}{\log^2 x} \int_{[1,x]} \frac{|\Xi|(u)}{u^2} \log \frac{x}{u} d\lambda_1(u) + O(\log^{-1} x). \tag{3.61}$$

We are now in position to complete the proof. By (3.61) and integration by parts we obtain

$$\frac{|\Xi|(x)}{x}$$
$$\leq \frac{2}{\log^2 x} \left(\log \frac{x}{u} \int_{[1,u]} \frac{|\Xi|(s)}{s^2} d\lambda_1(s) \Big|_1^x + \int_{[1,x]} \frac{1}{u} \left(\int_{[1,u]} \frac{|\Xi|(s)}{s^2} d\lambda_1(s) \right) d\lambda_1(u) \right) + O(\log^{-1} x)$$
$$= \frac{2}{\log^2 x} \int_{[1,x]} \frac{1}{u} \left(\int_{[1,u]} \frac{|\Xi|(s)}{s^2} d\lambda_1(s) \right) d\lambda_1(u) + O(\log^{-1} x). \tag{3.62}$$

Now,

$$\frac{1}{\log u} \int_{[1,u]} \frac{|\Xi|(s)}{s^2} d\lambda_1(s) \leq \frac{1}{\log u} \int_{[1,u]} \frac{Ks}{s^2} d\lambda_1(s)$$
$$= K \text{ for some } K > 0 \text{ by (3.29)}.$$

Therefore, we can say

$$\frac{1}{\log u}\int_{[1,u]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s) \leq \left(\limsup_{v\to\infty}\frac{1}{\log v}\int_{[1,v]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\right)+o(1)$$
$$< \infty.$$

Combining this result with (3.62) gives

$$\begin{aligned}
&\frac{|\Xi|(x)}{x}\\
&\leq \frac{2}{\log^2 x}\int_{[1,x]}\left(\frac{1}{\log u}\int_{[1,u]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\right)\frac{\log u}{u}d\lambda_1(u)+O(\log^{-1}x)\\
&\leq \frac{2}{\log^2 x}\left(\limsup_{v\to\infty}\frac{1}{\log v}\int_{[1,v]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\right)\int_{[1,x]}\frac{\log u}{u}d\lambda_1(u)\\
&\qquad +\frac{2}{\log^2 x}\int_{[1,x]}o(1)\frac{\log u}{u}d\lambda_1(u)+O(\log^{-1}x).
\end{aligned} \tag{3.63}$$

Let us get an estimate for the second term. For $\epsilon > 0$, there exists $N_1 > 0$ such that $|o(1)| < \frac{\epsilon}{2}$ for all $u \geq N_1$ and

$$\begin{aligned}
\int_{[1,x]}|o(1)|\frac{\log u}{u}d\lambda_1(u) &< \int_{[1,N_1)}|o(1)|\frac{\log u}{u}d\lambda_1(u)+\int_{[N_1,x]}\frac{\epsilon}{2}\frac{\log u}{u}d\lambda_1(u)\\
&=: M+\int_{[N_1,x]}\frac{\epsilon}{2}\frac{\log u}{u}d\lambda_1(u) \text{ for some } M>0 \text{ and for } x\geq N_1.
\end{aligned}$$

Furthermore, there exists N_2 such that $\frac{2M}{\log^2 x} < \frac{\epsilon}{2}$ for all $x \geq N_2$. Let $N = \max\{N_1, N_2\}$. Then

$$\begin{aligned}
\left|\frac{2}{\log^2 x}\int_{[1,x]}o(1)\frac{\log u}{u}d\lambda_1(u)\right| &< \frac{2M}{\log^2 x}+\frac{\epsilon}{2}\left(1-\frac{\frac{1}{2}\log^2 N_1}{\frac{1}{2}\log^2 x}\right)\\
&< \frac{\epsilon}{2}+\frac{\epsilon}{2}\\
&= \epsilon \text{ for all } x\geq N.
\end{aligned}$$

Since ϵ is arbitrary,

$$\frac{2}{\log^2 x}\int_{[1,x]}o(1)\frac{\log u}{u}d\lambda_1(u)=o(1).$$

Combining this with (3.63) brings

$$\begin{aligned}\frac{|\Xi|(x)}{x} &\leq \frac{2}{\log^2 x}\left(\limsup_{v\to\infty}\frac{1}{\log v}\int_{[1,v]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\right)\int_{[1,x]}\frac{\log u}{u}d\lambda_1(u)+o(1)+O(\log^{-1}x)\\ &= \limsup_{v\to\infty}\frac{1}{\log v}\int_{[1,v]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)+o(1).\end{aligned}$$

Passing to the limit implies

$$\alpha \leq \limsup_{x\to\infty}\frac{1}{\log x}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s). \tag{3.64}$$

Now we show the reverse inequality. For $\epsilon>0$, there exists x' such that $\frac{|\Xi|(x)}{x}\leq\alpha+\epsilon$ for all $x\geq x'$. So, for $x\geq x'$,

$$\begin{aligned}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s) &\leq \int_{[1,x')}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)+\int_{[x',x]}\frac{\alpha+\epsilon}{s}d\lambda_1(s)\\ &= \int_{[1,x')}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)+(\alpha+\epsilon)(\log x-\log x')\\ &= \left(\int_{[1,x')}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)-(\alpha+\epsilon)\log x'\right)+(\alpha+\epsilon)\log x.\end{aligned}$$

By the finiteness of the first term, this implies

$$\limsup_{x\to\infty}\frac{1}{\log x}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\leq\alpha+\epsilon.$$

Since ϵ is arbitrary,

$$\limsup_{x\to\infty}\frac{1}{\log x}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s)\leq\alpha.$$

Therefore, combining with (3.64),

$$\alpha = \limsup_{x\to\infty}\frac{1}{\log x}\int_{[1,x]}\frac{|\Xi|(s)}{s^2}d\lambda_1(s).$$

■

3.6 The Prime Number Theorem

We now proceed to show $\alpha=0$. We will require the following two Lemmas.

Lemma 3.11. $\left|\int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t)\right| \leq K$ *for some* $K > 0$.

Proof.

$$\begin{aligned}
\int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) &= \int_{(1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) = \int_{(1,x]} \frac{\int_{[1,t]} \frac{\psi(u)-u}{u} d\lambda_1(u)}{t^2} d\lambda_1(t) \\
&= \int_{(1,x]} \frac{\int_{(0,t]} \frac{\psi(u)-u}{u} d\lambda_1(u)}{t^2} d\lambda_1(t) \\
&= -\int_{(1,x]} \left(-\frac{1}{t}+1\right) \frac{\psi(t)-t}{t} d\lambda_1(t) + \left(-\frac{1}{x}+1\right) \int_{(1,x]} \frac{\psi(t)-t}{t} d\lambda_1(t) \\
&= -\frac{\Xi(x)}{x} + \int_{[1,x]} \frac{\psi(t)}{t^2} d\lambda_1(t) - \log x
\end{aligned}$$

by judicious choices of ν, ρ, and h in integration by parts Example 2.82 and noting λ_1 is a continuous measure. By Lemma 3.3 and (3.29), this equals $O(1)$. Therefore,

$$\left|\int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t)\right| \leq K \text{ for some } K > 0.$$

■

Lemma 3.12. $|\Xi(x_2) - \Xi(x_1)| \leq k(x_2 - x_1)$ *for some* $k > 0$ *and for all* $1 \leq x_1 \leq x_2$.

Proof. We know $\Xi'(x) = \frac{\psi(x)-x}{x}$ on (a, b) if no prime power is contained in (a, b). This implies

$$|\Xi'(x)| \leq \frac{\psi(x)}{x} + 1 \leq \frac{cx}{x} + 1 \leq k \text{ on any such } (a, b) \text{ for some } c, k > 0.$$

Therefore, since $\Xi(x)$ is continuous on $[a, b]$ and differentiable on (a, b), we have, for $a \leq x_1 \leq x_2 \leq b$,

$$\Xi(x_2) - \Xi(x_1) = (x_2 - x_1)\Xi'(\kappa) \text{ for some } x_1 < \kappa < x_2$$

by the mean value theorem. Therefore,

$$|\Xi(x_2) - \Xi(x_1)| \leq k(x_2 - x_1) \text{ on all such intervals.}$$

Now, let $x_1 \downarrow p^e$ (prime power). Then

$$\lim_{x_1 \downarrow p^e} |\Xi(x_2) - \Xi(x_1)| \leq \lim_{x_1 \downarrow p^e} k(x_2 - x_1)$$

which implies

$$|\Xi(x_2) - \Xi(p^e)| \leq k(x_2 - p^e) \text{ by the continuity of } \Xi(x) \text{ and absolute value.}$$

Similarly,

$$|\Xi(p^e) - \Xi(x_1)| \leq k(p^e - x_1).$$

So, for the case $x_1 < p^e < x_2$ where no other prime power is in (x_1, x_2) we get

$$\begin{aligned} |\Xi(x_2) - \Xi(x_1)| &\leq |\Xi(x_2) - \Xi(p^e)| + |\Xi(p^e) - \Xi(x_1)| \\ &\leq k(x_2 - x_1). \end{aligned}$$

These last three inequalities imply

$$|\Xi(x_2) - \Xi(x_1)| \leq k(x_2 - x_1) \text{ for some } k > 0 \text{ and for all } 1 \leq x_1 \leq x_2.$$

■

Now, we are poised to complete the proof of the PNT.

Theorem 3.13 (Prime Number Theorem)**.** $\pi(x) \sim \frac{x}{\log x}$.

Proof. First, suppose $\Xi(x)$ has only a finite number of zeros in $[1, \infty)$ (we know $\Xi(1) = 0$). Let $[a, \infty)$ contain no zeros for some a. Then, for $x \geq a$,

$$\begin{aligned} \int_{[1,x]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) &= \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \int_{[a,x]} \frac{|\Xi|(t)}{t^2} d\lambda(t) \\ &= \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \left| \int_{[a,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \\ &= \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \left| \int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) - \int_{[1,a)} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \\ &= \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \left| \int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) - \int_{[1,a]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \end{aligned}$$

$$\leq \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \left| \int_{[1,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| + \left| \int_{[1,a]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right|$$
$$\leq \int_{[1,a)} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + 2K \text{ from Lemma 3.11.}$$

This implies

$$\alpha = \limsup_{x \to \infty} \frac{1}{\log x} \int_{[1,x]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) = 0.$$

So, now suppose there exists infinitely many zeros of $\Xi(x)$ in $[1, \infty)$. A common theme of the proofs that use the Selberg formulae is the analysis of integrals of functions, such as $\frac{|\Xi|(t)}{t^2}$, on non-overlapping intervals. Here we do the same, but in an original manner.

Let $a < b$ be consecutive zeros. We will examine $\int_{(a,b]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t)$ in three cases according to the size of $\log b - \log a$. Here we deviate from other proofs in that we demonstrate the form of the size of the intervals can be assumed to be more general than in other proofs. Furthermore, we take full advantage of the closeness of $\Xi(x)$ to 0 at both a and b in the intermediate interval.

Let us first define some needed variables. Let $\epsilon > 0$. Since $\limsup_{x \to \infty} \frac{|\Xi|(x)}{x} = \alpha$, there exists x' such that $\frac{|\Xi|(x)}{x} \leq \alpha + \epsilon$ for all $x \geq x'$. Define the following:

Let β, ϵ as satisfying $\alpha + \epsilon < \beta < \alpha + 2\epsilon < 1$ (recall $\alpha \leq \log 2 < 1$). (3.65)

Let $m > 1$. (3.66)

Choose k in Lemma 3.12 large enough so that $mk > \beta$. (3.67)

Choose l such that $l > \frac{\beta}{K} \log \frac{mk + \beta}{mk - \beta}$ for K in Lemma 3.11. (3.68)

In addition, let $I = \int_{(a,b]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t)$, throughout the following analysis.

(case i) $\log b - \log a \leq \log \frac{mk+\beta}{mk-\beta}, a \geq x'$.

From Lemma 3.12, $|\Xi|(t) = |\Xi(t) - \Xi(a)| \leq k(t-a)$. Here we have $b \leq a\frac{mk+\beta}{mk-\beta}$. So

$$a < t \leq b \leq a\frac{mk+\beta}{mk-\beta}$$

which implies

$$\begin{aligned} t - a &\leq t - t\frac{mk-\beta}{mk+\beta} \\ &= t\frac{2\beta}{mk+\beta} \\ &< t\frac{2\beta}{mk} \end{aligned}$$

which leads to

$$\begin{aligned} |\Xi|(t) &< kt\frac{2\beta}{mk} \\ &= \frac{2\beta}{m}t. \end{aligned}$$

Therefore

$$\begin{aligned} I &= \int_{(a,b]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) \\ &< \int_{(a,b]} \frac{\frac{2\beta}{m}t}{t^2} d\lambda_1(t) \\ &= \frac{2}{m}\beta(\log b - \log a). \end{aligned} \tag{3.69}$$

(case ii) $\log\frac{mk+\beta}{mk-\beta} < \log b - \log a \leq \frac{lK}{\beta}, a \geq x'.$ (3.70)

Since $l > \frac{\beta}{K}\log\frac{mk+\beta}{mk-\beta}$, we have $\frac{lK}{\beta} > \log\frac{mk+\beta}{mk-\beta}$. So, the interval is well defined and

$$\begin{aligned} I &= \int_{(a,b]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) \\ &= \int_{(a,a\frac{mk}{mk-\beta}]} + \int_{(a\frac{mk}{mk-\beta},b\frac{mk}{mk+\beta}]} + \int_{(b\frac{mk}{mk+\beta},b]} \\ &=: I_1 + I_2 + I_3. \end{aligned}$$

Let us verify that this division is consistent.

$$\log\left(b\frac{mk}{mk+\beta}\right) - \log\left(a\frac{mk}{mk-\beta}\right) = \log b - \log a - \log\frac{mk+\beta}{mk-\beta}$$
$$> 0.$$

Therefore

$$a\frac{mk}{mk-\beta} < b\frac{mk}{mk+\beta}.$$

and the division is well defined.

First, look at $I_1 : a < t \leq a\frac{mk}{mk-\beta}$. Then

$$\begin{aligned} t - a &\leq t - t\frac{mk-\beta}{mk} \\ &= t\frac{\beta}{mk}. \end{aligned}$$

Therefore,

$$\begin{aligned} |\Xi|(t) &\leq kt\frac{\beta}{mk} \\ &= t\frac{\beta}{m}. \end{aligned}$$

This implies,

$$\begin{aligned} I_1 &\leq \int_{(a, a\frac{mk}{mk-\beta}]} \frac{t\frac{\beta}{m}}{t^2} d\lambda_1(t) \\ &= \frac{\beta}{m}\log\frac{mk}{mk-\beta}. \end{aligned} \tag{3.71}$$

Next, look at $I_3 : b\frac{mk}{mk+\beta} < t \leq b$. Then

$$\begin{aligned} b - t &< t\frac{mk+\beta}{mk} - t \\ &= t\frac{\beta}{mk}. \end{aligned}$$

Therefore,

$$|\Xi|(t) < kt\frac{\beta}{mk}$$

$$= t\frac{\beta}{m}.$$

This implies

$$\begin{aligned} I_3 &< \int_{(b\frac{mk}{mk+\beta}, b]} \frac{t\frac{\beta}{m}}{t^2} d\lambda_1(t) \\ &= -\frac{\beta}{m} \log \frac{mk}{mk+\beta}. \end{aligned} \tag{3.72}$$

Lastly, look at $I_2 : a\frac{mk}{mk-\beta} < t \leq b\frac{mk}{mk+\beta}$. From (3.65), $|\Xi|(t) < \beta t$ and we get

$$\begin{aligned} I_2 &< \int_{(a\frac{mk}{mk-\beta}, b\frac{mk}{mk+\beta}]} \frac{\beta t}{t^2} d\lambda_1(t) \\ &= \beta \left(\log b - \log a - \log \frac{mk+\beta}{mk-\beta} \right). \end{aligned} \tag{3.73}$$

Therefore, combining (3.71), (3.72), and (3.73),

$$\begin{aligned} I &< \frac{\beta}{m} \log \frac{mk}{mk-\beta} - \frac{\beta}{m} \log \frac{mk}{mk+\beta} + \beta \left(\log b - \log a - \log \frac{mk+\beta}{mk-\beta} \right) \\ &= \beta \left(\log b - \log a + \left(1 - \frac{1}{m}\right) \log \frac{mk-\beta}{mk+\beta} \right). \end{aligned} \tag{3.74}$$

Now, $\frac{2\beta}{mk+\beta} < 1$ by (3.67). This implies

$$\begin{aligned} \log \frac{mk-\beta}{mk+\beta} &= \log \left(1 - \frac{2\beta}{mk+\beta}\right) \\ &= -\frac{2\beta}{mk+\beta} - \sum_{i=2}^{\infty} \frac{(\frac{2\beta}{mk+\beta})^i}{i} \\ &< -\frac{2\beta}{mk+\beta}. \end{aligned}$$

Combining this result with (3.74) gives

$$\begin{aligned} I &< \beta(\log b - \log a) \left(1 - \left(1 - \frac{1}{m}\right) \frac{2\beta}{(mk+\beta)(\log b - \log a)}\right) \\ &< \beta(\log b - \log a) \left(1 - \left(1 - \frac{1}{m}\right) \frac{2\beta}{2mk\frac{lK}{\beta}}\right) \text{ by (3.67) and (3.70)} \\ &= \beta(\log b - \log a) \left(1 - \left(1 - \frac{1}{m}\right) \frac{\beta^2}{lmkK}\right). \end{aligned} \tag{3.75}$$

(case iii) $\log b - \log a > \dfrac{lK}{\beta}, a \geq x'.$ (3.76)

$$\begin{aligned}
I &= \int_{(a,b]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) \\
&= \left| \int_{(a,b]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \\
&= \left| \int_{[1,b]} \frac{\Xi(t)}{t^2} d\lambda_1(t) - \int_{[1,a]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \\
&\leq \left| \int_{[1,b]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| + \left| \int_{[1,a]} \frac{\Xi(t)}{t^2} d\lambda_1(t) \right| \\
&\leq 2K \text{ by Lemma 3.11} \\
&< \frac{2\beta}{l}(\log b - \log a) \text{ by (3.76).}
\end{aligned} \tag{3.77}$$

Summarizing, by (3.69), (3.75) and (3.77),

$$\begin{aligned}
&\text{(case i) } I < \frac{2}{m}\beta(\log b - \log a). \\
&\text{(case ii) } I < \beta(\log b - \log a)\left(1 - \left(1 - \frac{1}{m}\right)\frac{\beta^2}{lmkK}\right). \\
&\text{(case iii) } I < \frac{2}{l}\beta(\log b - \log a).
\end{aligned}$$

We see that if $\left(1 - \left(1 - \frac{1}{m}\right)\frac{\beta^2}{lmkK}\right) \geq \max\{\frac{2}{m}, \frac{2}{l}\}$, then all three cases will satisfy

$$I < \beta(\log b - \log a)\left(1 - \left(1 - \frac{1}{m}\right)\frac{\beta^2}{lmkK}\right).$$

By combining this with our other restrictions: $\begin{cases} \left(1 - \left(1 - \frac{1}{m}\right)\frac{\beta^2}{lmkK}\right) \geq \max\{\frac{2}{m}, \frac{2}{l}\} \\ K \text{ satisfies requirement in Lemma 3.11} \\ k \text{ satisfies requirement in Lemma 3.12} \\ m > 1 \\ mk > \beta \\ l > \frac{\beta}{K}\log\frac{mk+\beta}{mk-\beta} \end{cases}$

we see that if m and l are any numbers greater than 2, then all interval sizes with respect to m and l will satisfy all conditions by choosing K and k large enough. This is the crucial element, as we now show we will be able to telescope the sum of all the integrals I.

Recalling from section 3.5 that there are a finite number of zeros of $\Xi(x)$ in $[1, x]$, let $\{x_1, x_2, \ldots, x_r\}$ be the zeros of $\Xi(x)$ in ascending order in $[x', x]$. Then

$$\begin{aligned}
&\frac{1}{\log x}\int_{(1,x]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) \\
&< \frac{1}{\log x}\left(\int_{(1,x_1]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t) + \beta\left(1-\left(1-\frac{1}{m}\right)\frac{\beta^2}{lmkK}\right)\sum_{i=1}^{r-1}\log\frac{x_{i+1}}{x_i}\right. \\
&\qquad\left. + \int_{(x_r,x]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t)\right) \\
&= \frac{\int_{(1,x_1]} \frac{|\Xi|(t)}{t^2} d\lambda_1(t)}{\log x} + \beta\left(1-\left(1-\frac{1}{m}\right)\frac{\beta^2}{lmkK}\right)\frac{\log\frac{x_r}{x_1}}{\log\frac{x}{1}} + \frac{\left|\int_{(x_r,x]} \frac{\Xi(t)}{t^2} d\lambda_1(t)\right|}{\log x}.
\end{aligned}$$

The numerator of the first term is an integral of a bounded function over a bounded set and the numerator of the third term is finite by Lemma 3.11. It then follows

$$\begin{aligned}
\alpha &= \limsup_{x\to\infty} \frac{1}{\log x}\int_{[1,x]} \frac{|\Xi|(t)}{t^2} d\lambda(t) \\
&= \limsup_{x\to\infty} \frac{1}{\log x}\int_{(1,x]} \frac{|\Xi|(t)}{t^2} d\lambda(t) \\
&\leq \beta\left(1-\left(1-\frac{1}{m}\right)\frac{\beta^2}{lmkK}\right) \\
&< (\alpha+2\epsilon)\left(1-\left(1-\frac{1}{m}\right)\frac{(\alpha+\epsilon)^2}{lmkK}\right) \\
&< (\alpha+2\epsilon) - \left(1-\frac{1}{m}\right)\frac{(\alpha+\epsilon)^3}{lmkK}.
\end{aligned}$$

This implies

$$\left(1-\frac{1}{m}\right)\frac{(\alpha+\epsilon)^3}{lmkK} < 2\epsilon$$

and so

$$\alpha < \left(\frac{2lmKk}{1-\frac{1}{m}}\epsilon\right)^{\frac{1}{3}}.$$

Since ϵ is arbitrary,

$$\alpha = 0.$$

Therefore, by sections 3.5, 3.4 and 3.2,

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

■

Chapter 4

A Prime Number Theorem Proof Based on Riemann's Zeta Function

The Riemann zeta function, defined as

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \text{ for } \Re(s) > 1,$$

and named after G. Riemann (1826-1866), can arguably called a misnomer in that L. Euler (1707-1783) was the first to analyze its properties. In fact, the definition of $\zeta(s)$ and the product representation

$$\zeta(s) = \prod_{p} \left(1 - p^{-s}\right)^{-1} \text{ for } \Re(s) > 1,$$

are due to Euler.

On the other hand, Riemann contributed great advances in the analysis such as the meromorphic continuation of $\zeta(s)$ to the entire complex plane and the functional equation. He was also the first to predict that the prime number conjecture could be proved by examining the zeros of $\zeta(s)$. Indeed, much of the development of complex analysis in the latter half of the 19th century was due to the search for a proof. However, Riemann himself was unsuccessful.

Not until 1896 did a proof emerge, this by both de la Vallée Poussin and by Hadamard. Their proofs, just as Riemann predicted, both relied on $\zeta(s)$ having no zeros on the line $\Re(s) = 1$. Essential boundedness properties of $\zeta(s)$ made their proofs quite elaborate and lengthy, however.

Then in 1931, with the discovery of the Wiener-Ikehara Theorem, the PNT was proved in a very brief manner although still relying on $\zeta(s)$ having no zeros on $\mathfrak{R}(s) = 1$. Below we present a proof utilizing the Wiener-Ikehara Theorem, but in a manner that departs from the standard way this theorem has been used.

The year 1980 brought another simplification of the Riemann zeta function dependent type proof. The mathematician D. J. Newman [6] demonstrated an elegant proof that did not require the Wiener-Ikehara Theorem, but, instead a much simpler Tauberian type theorem of his own creation. By virtue of this nonusage of Weiner-Ikehara, a strong argument can be made that Newman's proof, out of all the proofs of this type, is the simplest to date.

As mentioned, we now give a proof that employs the Wiener-Ikehara Theorem. Although the use of measure theory in this proof is not near to the extent as it is in the elementary proof, it nevertheless displays its utility, most notably in Section 4.3.

Note: Throughout, p is prime and $s \in \mathbb{C}$ where $s = \sigma + it$ and $\sigma, t \in \mathbb{R}$. As usual, $\mathfrak{R}(s)$ and $\mathfrak{I}(s)$ denote the real and imaginary parts of s, respectively. The proof consists of the following six sections:

(4.1) State required standard results from analytic number theory.

(4.2) State a particular case of the Wiener-Ikehara Theorem.

(4.3) Prove

$$\sum_{n \leq x} \mu(n) = o(x) \Rightarrow \psi(x) - x = o(x).$$

(4.4) Prove $\zeta(s)$ is analytic for $\mathfrak{R}(s) > 1$ and has a meromorphic continuation to $\mathfrak{R}(s) > 0$ with a simple pole at $s = 1$.

(4.5) Prove $\zeta(s)$ has no zeros for $\mathfrak{R}(s) \geq 1$.

(4.6) Find the Mellin transform $\widehat{\lambda_G}(s)$ associated with

$$\lambda_G, \text{ the Borel measure induced by } G(x) := \sum_{n \leq x} (1 + \mu)(n)$$

and apply the Wiener-Ikehara Theorem to $\widehat{\lambda_G}(s)$ to show $\sum_{n \leq x} \mu(n) = o(x)$, thus proving the PNT by section 4.3 and Theorem 3.4.

4.1 Preliminary Results

We will require certain standard definitions and results specific to this proof. We state them in a manner that is consistent with our notation and definitions. The proofs of the theorem and lemmas are standard and can be found in any quality text on analytic number theory. For other general concepts, see Chapter 2.

We begin the section with definitions and results concerning the Mellin transform associated wth a measure.

Definition 4.1. Let $\nu \in \mathscr{M}$. The ***Mellin transform*** associated with ν is defined by

$$\widehat{\nu}(s) = \int_{(0,\infty)} x^{-s} d\nu(x)$$

for the set of points $s \in \mathbb{C}$ for which the integral converges. In particular, if $\nu = \lambda_F$, the Borel measure induced by $F(x) := \sum_{n \leq x} f(n)$, then $\widehat{\lambda_F}(s) = \sum_{n=1}^{\infty} n^{-s} f(n)$. This last series is called the ***Dirichlet series of*** f.

Definition 4.2. The ***abscissa of convergence*** of $\widehat{\nu}(s)$ is defined to be

$$\sigma_c(\widehat{\nu}) = \inf\{\sigma : \widehat{\nu}(s) \text{ converges for some } s \text{ with } \Re(s) = \sigma\}.$$

If the integral defining the Mellin transform associated wth ν diverges for all s we set $\sigma_c(\widehat{\nu}) = +\infty$. If the integral defining the Mellin transform associated with ν converges for all s we set $\sigma_c(\widehat{\nu}) = -\infty$.

Lemma 4.3. *Let F be a Type 1 summatory function and λ_F be the Borel measure induced by F. Let $s_0 = \sigma_0 + it_0 \in \mathbb{C}$. If $\int_{(0,\infty)} x^{-s_0} d\lambda_F(x)$ converges, then the integral converges on the open half plane $\{s : \sigma > \sigma_0\}$.*

Proof. Omitted. ■

Theorem 4.4. *Let F and G be Type 1 summatory functions and λ_F and λ_G be the Borel measures induced by F and G, respectively. Suppose s is a point at which one of the Mellin* *transforms* $\widehat{F}(s)$ *and* $\widehat{G}(s)$ *converges absolutely and the other converges. Then*

$$\int_{(0,\infty)} x^{-s} d\lambda_F * \lambda_G(x) = \widehat{F}(s) \cdot \widehat{G}(s).$$

Proof. Omitted. ∎

We end this section with two results pertaining to the product representation of Dirichlet series.

Lemma 4.5. *Let* $f \in \mathscr{A}_{\mathscr{M}}$. *Assume* $\sum_{n=1}^{\infty} |f(n)|$ *converges. Then*

$$\sum_{n=1}^{\infty} f(n) = \prod_p (1 + f(p) + f(p^2) + \cdots).$$

Proof. Omitted. ∎

We will require Euler's product representation of $\zeta(s)$ in order to prove the nonvanishing of $\zeta(s)$ for $\Re(s) \geq 1$ in section 4.5. We state it as a corollary to the preceding lemma.

Corollary 4.6 (Euler's Product Representation of $\zeta(s)$)**.**

$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$ *for* $\Re(s) > 1$.

Proof. Let $\Re(s) > 1$. We recall the definition

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \text{ for } \Re(s) > 1.$$

Clearly, n^{-s} is multiplicative. Now,

$$\sum_{n=1}^{\infty} |n^{-s}| = \sum_{n=1}^{\infty} n^{-\sigma} \text{ which is convergent by the p-test.}$$

Therefore, by Lemma 4.5,

$$\begin{aligned} \zeta(s) &= \prod_p \sum_{k=0}^{\infty} (p^{-s})^k \\ &= \prod_p \left(1 - p^{-s}\right)^{-1} \text{ by geometric series.} \end{aligned}$$

∎

4.2 The Wiener-Ikehara Theorem

We now state without proof a particular case of the Wiener-Ikehara Theorem. A proof can be found in any quality analytic number theory text. This is a Tauberian theorem in that it takes known information of the Mellin transform associated with a measure and gives information about the measure itself. We reproduce the theorem here in a form that is compatible with the measures induced by Type 1 summatory functions.

Theorem 4.7 (Wiener-Ikehara). *Let F be a Type 1 summatory function and let λ_F be the Borel measure induced by F. Let the abscissa of convergence $\sigma_c(\widehat{\lambda_F}) = \alpha > 0$ and suppose that there exist a real number l and a function φ, continuous on the closed half plane $\{s : \sigma \geq \alpha\}$, such that*

$$\widehat{\lambda_F}(s) = \int_{(0,\infty)} x^{-s} d\lambda_F(x) = l(s-\alpha)^{-1} + \varphi(s)$$

holds on the corresponding open half plane. Then

$$\lambda_F((0,x]) = F(x) = \frac{lx^\alpha}{\alpha} + o(x^\alpha).$$

Proof. Omitted. ∎

4.3 $\sum_{n\leq x} \mu(n) = o(x)$ Implies $\psi(x) - x = o(x)$

We are going to depart from the typical way the zeta function dependent proof is carried out. Normally, the function $\psi(x)$ is the centerpiece of the proof. As a further demonstration of the usefulness of measure theory, however, we will utilize $M(x) = \sum_{n\leq x} \mu(n)$ as our function of choice. In order to accomplish this, we will need to prove $\sum_{n\leq x} \mu(n) = o(x) \Rightarrow \psi(x) - x = o(x)$. This will require the result of the following lemma.

This lemma is a measure theoretic adaptation of what is known as "Dirichlet's hyperbola method". We must state and prove the hyperbola method in such a way that will allow it to work in concert with the measures in $\mathscr{M}$.

Lemma 4.8 (Dirichlet's Hyperbola Method). *Let $\mu,\ \nu \in \mathscr{M}$. Let $x \geq 1$ and $x = yz$ where $y,\ z \geq 1$. Then*

$$\mu * \nu((0,x]) = \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,y]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,y])\nu((0,z]).$$

Proof. First, assume $z = 1$ and $x = y$. Then

$$\begin{aligned}
&\int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,y]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,y])\nu((0,z]) \\
&= \int_{\{1\}} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,x]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,x])\nu((0,1]) \\
&= \mu((0,x])\nu(\{1\}) + \int_{[1,x]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,x])\nu((0,1)) - \mu((0,x])\nu(\{1\}) \\
&= \int_{[1,x]} \nu((0,t^{-1}x])d\mu(t) = \mu * \nu((0,x]).
\end{aligned}$$

Similarly, we have equality when $y = 1$ and $x = z$.

Suppose, now, that $y,\ z > 1$. Let $\nu_1 = \int \chi_{(0,z]}(t)d\nu(t)$ and $\nu_2 = \int \chi_{(z,\infty)}(t)d\nu(t)$. Then ν_1 and $\nu_2 \in \mathscr{M}$. Furthermore, ν_1 and ν_2 have the following properties.

$$\nu = \nu_1 + \nu_2. \tag{4.1}$$

$$\nu_1((0,u]) = \begin{cases} \nu((0,u]), & u \leq z; \\ \nu((0,z]), & u > z. \end{cases} \tag{4.2}$$

$$\nu_2((0,u]) = \begin{cases} 0, & u \leq z; \\ \nu((0,u]) - \nu((0,z]), & u > z. \end{cases} \tag{4.3}$$

We then get

$$\begin{aligned}
&\mu * \nu((0,x]) = \mu * (\nu_1 + \nu_2)((0,x]) = \mu * \nu_1((0,x]) + \mu * \nu_2((0,x]) \text{ by(4.1)} \\
&= \int_{[1,x]} \mu((0,t^{-1}x])d\nu_1(t) + \int_{[1,x]} \nu_2((0,t^{-1}x])d\mu(t) \\
&= \left(\int_{[1,z]} + \int_{(z,x]}\right) \mu((0,t^{-1}x])d\nu_1(t) + \left(\int_{[1,x/z)} + \int_{[x/z,x]}\right) \nu_2((0,t^{-1}x])d\mu(t) \\
&= \int_{[1,z]} \mu((0,t^{-1}x])d\nu_1(t) + 0 + \int_{[1,x/z)} \nu_2((0,t^{-1}x])d\mu(t) + 0 \text{ by (4.2) and (4.3)}
\end{aligned}$$

$$= \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,x/z)} \big(\nu((0,t^{-1}x]) - \nu_1((0,t^{-1}x])\big)d\mu(t) \text{ by (4.1) and (4.2)}$$
$$= \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,x/z)} \big(\nu((0,t^{-1}x]) - \nu((0,z])\big)d\mu(t) \text{ by (4.2)}$$
$$= \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,x/z)} \nu((0,t^{-1}x])d\mu(t) - \mu([1,x/z))\nu((0,z])$$
$$= \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,x/z]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,x/z])\nu((0,z])$$
$$= \int_{[1,z]} \mu((0,t^{-1}x])d\nu(t) + \int_{[1,y]} \nu((0,t^{-1}x])d\mu(t) - \mu((0,y])\nu((0,z]).$$

■

We are now ready to prove the main result of this secton.

Lemma 4.9. $\sum_{n\leq x} \mu(n) = o(x) \Rightarrow \psi(x) - x = o(x)$.

Proof. Recall $M(x) = \sum_{n\leq x} \mu(n)$ and $N(x) = \sum_{n\leq x} 1(n)$. Let $x \geq 1$ and assume we will be working exclusively in the domain $\mathcal{E}$. From Lemma 3.1, $L\lambda_N = \lambda_\psi * \lambda_N$ and

$$L\lambda_N = \lambda_\psi * \lambda_N \Rightarrow L\lambda_N * \lambda_M = \lambda_\psi * \lambda_N * \lambda_M \Rightarrow L\lambda_N * \lambda_M = \lambda_\psi.$$

We then have

$$\begin{aligned} \psi(x) = \lambda_\psi((0,x]) &= L\lambda_N * \lambda_M((0,x]) \\ &= (L\lambda_N - \lambda_1 * \lambda_N) * \lambda_M((0,x]) + \lambda_1 * \lambda_N * \lambda_M((0,x]) \\ &= (L\lambda_N - \lambda_1 * \lambda_N) * \lambda_M((0,x]) + \lambda_1((0,x]) =: A + B. \end{aligned} \tag{4.4}$$

It is clear that

$$B = x - 1. \tag{4.5}$$

Let us now evaluate A. Consider

$$\begin{aligned} (L\lambda_N - \lambda_1 * \lambda_N)((0,x]) &= \int_{(0,x]} \chi_{[1,\infty)}(t) \log t \, d\lambda_N(t) - \int_{[1,x]} \lambda_1((0,t^{-1}x])d\lambda_N(t) \\ &= \int_{[1,x]} \log t \, d\lambda_N(t) - \int_{[1,x]} \left(\frac{x}{t} - 1\right) d\lambda_N(t) \\ &= \int_{[1,x]} \log t \, d\lambda_N(t) - \int_{[1,x]} \frac{x}{t} d\lambda_N(t) + \lambda_N([1,x]). \end{aligned} \tag{4.6}$$

Using Euler's summation formula gives

$$\int_{[1,x]} \log t \, d\lambda_N(t) = \sum_{n \le x} \log n = x \log x - x + O(\log x) \tag{4.7}$$

and

$$\int_{[1,x]} \frac{x}{t} d\lambda_N(t) = x \sum_{n \le x} \frac{1}{n} = x \log x + kx + O(1) \text{ for some } k > 0. \tag{4.8}$$

Combining (4.6), (4.7), and (4.8) yields

$$\begin{aligned}
&(L\lambda_N - \lambda_1 * \lambda_N)((0,x]) \\
&= \big(x \log x - x + O(\log x)\big) - \big(x \log x + kx + O(1)\big) + \big(x + O(1)\big) \\
&= -kx + O(\log x) = -k\lambda_N((0,x]) + O(\log x). \qquad (4.9)
\end{aligned}$$

Now, let $\rho = L\lambda_N - \lambda_1 * \lambda_N + k\lambda_N$. Then $\rho \in \mathscr{L}(\mathcal{E})$ and

$$\rho((0,x]) = O(\log x) \tag{4.10}$$

and

$$\begin{aligned}
\rho * \lambda_M((0,x]) = \Big(&L\lambda_N * \lambda_Q - \lambda_1 * \lambda_N * \lambda_Q + k\lambda_N * \lambda_Q \\
&- L\lambda_N * \lambda_{Q-M} + \lambda_1 * \lambda_N * \lambda_{Q-M} - k\lambda_N * \lambda_{Q-M}\Big)((0,x]). \qquad (4.11)
\end{aligned}$$

Now, let $x = yz$ where $y, z \ge 1$. Applying Lemma 4.8 to each term in (4.11) individually and grouping in an advantageous manner yields

$$\begin{aligned}
\rho * \lambda_M((0,x]) = &\left(\int_{[1,x/z]} \rho((0,t^{-1}x]) d\lambda_Q(t) - \int_{[1,x/z]} \rho((0,t^{-1}x]) d\lambda_{Q-M}(t) \right) \\
&+ \left(\int_{[1,z]} \lambda_M((0,t^{-1}x]) dL\lambda_N(t) - \int_{[1,z]} \lambda_M((0,t^{-1}x]) d\lambda_1 * \lambda_N(t) \right. \\
&\left. + \int_{[1,z]} \lambda_M((0,t^{-1}x]) dk\lambda_N(t) \right) - \rho((0,z])\lambda_M((0,x/z]) \\
&=: A_1 + A_2 + A_3. \qquad (4.12)
\end{aligned}$$

We now evaluate $|A_1|$, $|A_2|$, and $|A_3|$.

$|A_1|$:

$$\begin{aligned}
|A_1| &\leq \int_{[1,x/z]} |\rho((0,t^{-1}x])| d\lambda_Q(t) + \int_{[1,x/z]} |\rho((0,t^{-1}x])| d\lambda_{Q-M}(t) \\
&= \int_{[1,x/z]} |\rho((0,t^{-1}x])| d(\lambda_Q + \lambda_{Q-M})(t) \\
&= \int_{[1,x/z]} |\rho((0,t^{-1}x])| (\widetilde{|\mu|} + \widetilde{|\mu|} - \widetilde{\mu}) d\lambda_N(t) \text{ (see Def. 2.77 and Prop. 2.79)} \\
&\leq 3 \int_{[1,x/z]} |\rho((0,t^{-1}x])| d\lambda_N(t) \leq 3K \int_{[1,x/z]} \log \frac{x}{t} d\lambda_N(t) \text{ for some } K > 0 \text{ by (4.10)} \\
&= 3K\Big((\log x)(x/z + O(1)) - \big(x/z \log(x/z) - x/z + O(\log(x/z))\big)\Big) \\
&\qquad \text{by Euler's summation formula} \\
&= 3K\left(x\frac{\log z}{z} + \frac{x}{z}\right) + O(\log x) \leq 6Kx\frac{\log z}{z} + O(\log x) \text{ for } z \geq e.
\end{aligned}$$

Therefore,

$$\frac{|A_1|}{x} \leq 6K\frac{\log z}{z} + O(x^{-1}\log x) \text{ for } z \geq e. \tag{4.13}$$

$|A_3|$:

$$|A_3| = |\rho((0,z])| \cdot |\lambda_M((0,x/z])| \leq (K \log z) \cdot o(x/z) \text{ by (4.10) and hypothesis.}$$

So,

$$\frac{|A_3|}{x} \leq K\frac{\log z}{z}\frac{o(x/z)}{x/z} \leq K\frac{o(x/z)}{x/z}. \tag{4.14}$$

Now, let $\epsilon > 0$. There exists z_1 such that $\frac{\log z}{z} < \epsilon$ for all $z \geq z_1$. Fix $z = \max\{e, z_1\}$. Moreover, there exists x_3 such that $\frac{o(x/z)}{x/z} < \epsilon$ for all $x \geq x_3$. Then combining (4.13) and (4.14) gives

$$\frac{|A_1| + |A_3|}{x} < \epsilon(7K) + O(x^{-1}\log x) \text{ for all } x \geq x_3. \tag{4.15}$$

$|A_2|$:

By the hypothesis, $\lambda_M((0,x]) = o(x)$. Therefore, there exists x_2 such that $\frac{|\lambda_M((0,x])|}{x} < \epsilon$ for all $x \geq x_2$. Hence, $|\lambda_M((0,t^{-1}x])| < \epsilon\frac{x}{t}$ for $t \leq \frac{x}{x_2}$. Now, choose x such that $x \geq zx_2$.

Then

$$\begin{aligned}|A_2| &\leq \int_{[1,z]} |\lambda_M((0,t^{-1}x])| dL\lambda_N(t) + \int_{[1,z]} |\lambda_M((0,t^{-1}x])| d\lambda_1 * \lambda_N(t) \\ &\qquad + \int_{[1,z]} |\lambda_M((0,t^{-1}x])| dk\lambda_N(t) \\ &< \int_{[1,z]} \epsilon \frac{x}{t} d(L\lambda_N + \lambda_1 * \lambda_N + k\lambda_N)(t) \\ &\leq \epsilon x (L\lambda_N + \lambda_1 * \lambda_N + k\lambda_N)([1,z]).\end{aligned}$$

Or, put another way,

$$\frac{|A_2|}{x} < \epsilon(L\lambda_N + \lambda_1 * \lambda_N + k\lambda_N)([1,z]). \tag{4.16}$$

Finally, for $x \geq \max\{x_3, zx_2\}$, combining (4.15) and (4.16) yields

$$\frac{|A_1| + |A_2| + |A_3|}{x} < \epsilon\big(7K + (L\lambda_N + \lambda_1 * \lambda_N + k\lambda_N)([1,z])\big) + O(x^{-1}\log x).$$

Therefore, since ϵ is arbitrary,

$$\limsup_{x\to\infty} \frac{|\rho * \lambda_M((0,x])|}{x} \leq \limsup_{x\to\infty} \frac{|A_1| + |A_2| + |A_3|}{x} = 0.$$

Thus, $\rho * \lambda_M((0,x]) = o(x)$, which implies

$$\begin{aligned}&(L\lambda_N - \lambda_1 * \lambda_N + k\lambda_N) * \lambda_M((0,x]) = o(x) \\ \Rightarrow\ &(L\lambda_N - \lambda_1 * \lambda_N) * \lambda_M((0,x]) + k = o(x) \\ \Rightarrow\ &(L\lambda_N - \lambda_1 * \lambda_N) * \lambda_M((0,x]) = o(x).\end{aligned} \tag{4.17}$$

Bringing together (4.17) with (4.4) and (4.5) gives

$$\psi(x) = o(x) + x - 1 \quad \text{or} \quad \psi(x) - x = o(x).$$

■

4.4 The Meromorphicity of $\zeta(s)$ on the Half Plane $\Re(s) > 0$

The meromorphicity of $\zeta(s)$ for $\sigma > 0$ is the linchpin of the proof. Specifically, we have the following Lemma that will be needed in sections 4.5 and 4.6.

Lemma 4.10. *$\zeta(s)$ is analytic for $\Re(s) > 1$ and has a meromorphic continuation to $\Re(s) > 0$ with a simple pole at $s = 1$.*

Proof. Let $\sigma > 1$. We recall

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$

Let $\{S_k\}_{k=1}^{\infty}$ be an increasing sequence of positive integers such that $S_k \to \infty$. Clearly, $\sum_{n=1}^{S_k} n^{-s}$ is analytic on $\{s : \sigma > 1\}$. It is easily seen that $\sum_{n=1}^{S_k} n^{-s}$ converges uniformly to $\sum_{n=1}^{\infty} n^{-s}$ on every compact subset of $\{s : \sigma > 1\}$. By Weierstrass's theorem [1] on the uniform limit of analytic functions, then,

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \text{ is analytic on } \{s : \sigma > 1\}$$

and

$$\zeta'(s) = \sum_{n=1}^{\infty} -n^{-s} \log n \text{ on } \{s : \sigma > 1\}. \tag{4.18}$$

Also, for $\sigma > 1$, we may deduce from Theorem 2.86 (Euler's summation formula) that

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \frac{1}{s-1} + 1 - s \int_{[1,\infty)} t^{-s-1}(t - [t]) d\lambda(t).$$

Now, let $\{T_n\}_{n=1}^{\infty}$ be an increasing sequence of numbers such that $T_n \geq 1$ and $T_n \to \infty$. Clearly, $\int_{[1,T_n]} t^{-s-1}(t - [t]) d\lambda(t)$ is analytic on $\{s : \sigma > 0\}$. It is again quickly apparent that $\int_{[1,T_n]} t^{-s-1}(t - [t]) d\lambda(t)$ converges uniformly to $\int_{[1,\infty)} t^{-s-1}(t - [t]) d\lambda(t)$ on every compact subset of $\{s : \sigma > 0\}$. By Weierstrass's theorem [1] on the uniform limit of analytic functions, then,

$$\int_{[1,\infty)} t^{-s-1}(t - [t]) d\lambda(t) \text{ is analytic on } \{s : \sigma > 0\}.$$

This implies

$$H(s) := 1 - s\int_{[1,\infty)} t^{-s-1}(t-[t])d\lambda(t) \text{ is analytic on } \{s : \sigma > 0\}. \tag{4.19}$$

Therefore, $\zeta(s)$ can be extended to a meromorphic function to $\{s : \sigma > 0\}$ whose only singularity is a simple pole at $s = 1$ and we can write

$$\zeta(s) = \frac{1}{s-1} + H(s) \text{ for } \sigma > 0 \text{ and where } H(s) \text{ is analytic.} \tag{4.20}$$

■

4.5 The Nonvanishing of $\zeta(s)$ on the Half Plane $\Re(s) \geq 1$

A vital ingredient in our proof is the nonvanishing of $\zeta(s)$ for $\Re(s) \geq 1$. We state it in the following Lemma.

Lemma 4.11. *$\zeta(s) \neq 0$ for $\Re(s) \geq 1$.*

Proof. Let $\sigma > 1$. Consider $\prod_p (1 - p^{-s})$. Since $\sum_p p^{-s}$ and $\sum_p p^{-2\sigma}$ are convergent,

$$\prod_p (1 - p^{-s}) \text{ is convergent by the convergence criteria for infinite products.}$$

In particular, this implies $\prod_p (1 - p^{-s})$ is neither 0 nor ∞. Consequently, $\zeta(s) \neq 0$ for $\sigma > 1$.

Showing $\zeta(s) \neq 0$ on $\sigma = 1$ is much more difficult. There are various ways to do this. Due to its simplicity, we choose to adopt the approach originally due to Hadamard of utilizing a trigonometric identity. This method was subsequently improved by Poussin and Mertens. Here we borrow the further refinement by D. Zagier [9]. In addition, we go a step further and supply full detail not found elsewhere.

By (4.18) and Theorem 4.4, for $\sigma > 1$,

$$\begin{aligned}
\zeta'(s) &= -\sum_{n=1}^{\infty} n^{-s}\log n \text{ from (4.18)} \\
&= -\int_{(0,\infty)} t^{-s}\log t \, d\lambda_N(t)
\end{aligned}$$

$$= -\int_{(0,\infty)} t^{-s} dL\lambda_N(t)$$
$$= -\int_{(0,\infty)} t^{-s} d\lambda_\psi * \lambda_N(t) \text{ by Chebyshev's identity}$$
$$= -\left(\int_{(0,\infty)} t^{-s} d\lambda_\psi(t)\right)\left(\int_{(0,\infty)} t^{-s} d\lambda_N(t)\right) \text{ (integrals abs. convergent)}$$
$$= -\left(\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}\right)\zeta(s).$$

It follows that

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} \\ &= \sum_{\substack{p^e \\ e\geq 1}} \frac{\log p}{(p^e)^s} \\ &= \sum_p \log p \sum_{e=1}^{\infty} \frac{1}{(p^s)^e} \text{ by absolute convergence} \\ &= \sum_p \frac{\log p}{p^s - 1}. \end{aligned}$$

So, for $\sigma > 1$,

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= \sum_p \frac{\log p}{p^s} + \sum_p \frac{\log p}{p^s(p^s-1)} \text{ by the absolute convergence of both series} \\ &=: \Phi(s) + \sum_p \frac{\log p}{p^s(p^s-1)}. \end{aligned} \tag{4.21}$$

By the exact same argument we used to prove analyticity in section 4.4, we have

$$\Phi(s) \text{ is analytic for } \sigma > 1$$

and

$$\sum_p \frac{\log p}{p^s(p^s-1)} \text{ is analytic for } \sigma > \frac{1}{2}.$$

Hence, the meromorphicity of $\frac{\zeta'(s)}{\zeta(s)}$ for $\sigma > 0$ with poles only at $s \in \{1, \text{zeros of } \zeta\}$ imply the meromorphic continuation of $\Phi(s)$ to $\sigma > \frac{1}{2}$ with poles only at

$s \in \{1, \text{zeros of } \zeta\}$.

We proceed to show the nonvanishing of $\zeta(s)$ on $\sigma = 1$ by contradiction. So, suppose $\zeta(s)$ has a zero of order $\alpha \geq 1$ at $s = 1 + it$ for some $t \neq 0$. Further suppose $\zeta(s)$ has a zero of order $\beta \geq 0$ at $s = 1 + i2t$ (For $\sigma > 0$, recall the lone pole of $\zeta(s)$ is at $s = 1$, so β cannot be less than 0). The Schwarz reflection principle [1] then implies that $\zeta(s)$ has a zero of order α at $s = 1 - it$ and a zero of order β at $s = 1 - i2t$.

Now, since $\zeta(s)$ has a simple pole at $s = 1$, the residue of $\frac{\zeta'(s)}{\zeta(s)}$ at $s = 1$ is -1. Similarly, the residue of $\frac{\zeta'(s)}{\zeta(s)}$ at $s = 1 \pm it$ is α and at $s = 1 \pm i2t$ it is β.

The residues and the analyticity of $\sum_p \frac{\log p}{p^s(p^s-1)}$ imply

$$\lim_{\sigma \to 1} (\sigma - 1)\Phi(\sigma) = 1,$$

$$\lim_{\sigma \to 1} (\sigma - 1)\Phi(\sigma \pm it) = \lim_{\sigma \to 1} (\sigma \pm it - (1 \pm it))\Phi(\sigma \pm it) = -\alpha,$$

$$\lim_{\sigma \to 1} (\sigma - 1)\Phi(\sigma \pm i2t) = \lim_{\sigma \to 1} (\sigma \pm i2t - (1 \pm i2t))\Phi(\sigma \pm i2t) = -\beta.$$

The cleverness Hadamard is evident in recognizing the following identity and inequality. Since the series defining $\Phi(s)$ is absolutely convergent for $\sigma > 1$, we can write

$$\text{For } \sigma > 1, \quad \sum_{k=0}^{4} \binom{4}{k} \Phi(\sigma + i(k-2)t) = \sum_p \frac{\log p}{p^\sigma} \left(p^{i\frac{t}{2}} + p^{-i\frac{t}{2}}\right)^4$$
$$= \sum_p \frac{\log p}{p^\sigma} 2^4 \cos^4\left(\frac{t}{2} \log p\right)$$
$$\geq 0.$$

Multiplying both sides of the inequality by $\sigma - 1$ and letting $\sigma \downarrow 1$ yields

$$1(-\beta) + 4(-\alpha) + 6(1) + 4(-\alpha) + 1(-\beta) \geq 0$$
$$\Rightarrow 6 - 8\alpha - 2\beta \geq 0.$$

Since $\alpha \geq 1$ and $\beta \geq 0$, we have a contradiction. Therefore, $1 + it$ cannot be a zero of $\zeta(s)$. Since t is nonzero arbitrary and $s = 1$ is not a zero, we conclude

$$\zeta(s) \text{ has no zeros on the line } \sigma = 1.$$

■

4.6 The Prime Number Theorem

We now are prepared to prove the PNT.

Theorem 4.12 (Prime Number Theorem)**.** $\pi(x) \sim \frac{x}{\log x}$.

Proof. Recall that $M(x) = \sum_{n \le x} \mu(n)$ and $N(x) = \sum_{n \le x} 1(n)$. Define $(N + M)(x) = \sum_{n \le x}(1 + \mu)(n)$. Then $N + M$ is a Type 1 summatory function since $(1 + \mu)(n) \ge 0$ for all $n \in \mathbb{N}$. Note that $(N + M)(x) = N(x) + M(x)$. Let λ_{N+M} be the Borel measure induced by $N + M$.

For $\sigma > 1$, $\sum_{n=1}^{\infty} n^{-s}(1 + \mu)(n)$ is absolutely convergent. Therefore, $\sum_{n=1}^{\infty} n^{-s}(1 + \mu)(n)$ is convergent for $\sigma > 1$ and

$$\int_{(0,\infty)} t^{-s} d\lambda_{N+M}(t) = \sum_{n=1}^{\infty} n^{-s}(1 + \mu)(n) = \sum_{n=1}^{\infty} n^{-s}1(n) + \sum_{n=1}^{\infty} n^{-s}\mu(n), \; \sigma > 1. \tag{4.22}$$

For $\sigma > 1$, clearly

$$\sum_{n=1}^{\infty} n^{-s} = \zeta(s). \tag{4.23}$$

Let us evaluate $\sum_{n=1}^{\infty} \mu(n)n^{-s}$. Now, $\mu(n)n^{-s}$ is multiplicative and for p prime,

$$\mu(p^e) = \begin{cases} 1, & e = 0. \\ -1, & e = 1. \\ 0, & e \ge 2. \end{cases}$$

Furthermore, $\sum_{n=1}^{\infty} \mu(n)n^{-s}$ is clearly absolutely convergent for $\sigma > 1$. Therefore, by Lemma 4.5,

$$\sum_{n=1}^{\infty} \mu(n)n^{-s} = \prod_{p}(1 - p^{-s}) = \zeta(s)^{-1} \text{ (recall } \zeta(s) \ne 0 \text{ for } \sigma \ge 1).$$

Now, define $G \coloneqq N + M$ and λ_G the Borel measure induced by G. The Mellin transform associated with λ_G is then

$$\widehat{\lambda_G}(s) = \int_{(0,\infty)} t^{-s}\lambda_{N+M}(t) = \zeta(s)^{-1} + \zeta(s) \text{ for } \sigma > 1. \tag{4.24}$$

In order to apply Weiner-Ikehara to $\widehat{\lambda_G}(s)$, we first must determine the abscissa of convergence $\sigma_c(\widehat{\lambda_G})$. The next few lines are devoted to that purpose. Assume $x \geq 1$. We have

$$\begin{aligned}
&x\left(\int_{[1,x]} t^{-1} d\lambda_Q(t) - \int_{[1,x]} t^{-1} d\lambda_{Q-M}(t)\right) \\
&= \int_{[1,x]} xt^{-1} d\lambda_Q(t) - \int_{[1,x]} xt^{-1} d\lambda_{Q-M}(t) \\
&= \int_{[1,x]} \Big(\lambda_N((0,t^{-1}x]) - \big(\lambda_N((0,t^{-1}x]) - t^{-1}x\big)\Big) d\lambda_Q(t) \\
&\qquad\qquad - \int_{[1,x]} \Big(\lambda_N((0,t^{-1}x]) - \big(\lambda_N((0,t^{-1}x]) - t^{-1}x\big)\Big) d\lambda_{Q-M}(t) \\
&= \lambda_N * \lambda_Q((0,x]) - \lambda_N * \lambda_{Q-M}((0,x]) \\
&\qquad\qquad + \left(\int_{[1,x]} (t^{-1}x - [t^{-1}x]) d\lambda_Q(t) - \int_{[1,x]} (t^{-1}x - [t^{-1}x]) d\lambda_{Q-M}(t)\right) \\
&= \lambda_N * \lambda_M((0,x]) + \left(\int_{[1,x]} O(1) d\lambda_Q(t) - \int_{[1,x]} O(1) d\lambda_{Q-M}(t)\right) \\
&= 1 + O(Q(x)) + O((Q-M)(x)) = O(x). \qquad (4.25)
\end{aligned}$$

Thus,

$$\begin{aligned}
O(1) &= \int_{[1,x]} t^{-1} d\lambda_Q(t) - \int_{[1,x]} t^{-1} d\lambda_{Q-M}(t) \\
&= \sum_{n \leq x} n^{-1} |\mu|(n) - \sum_{n \leq x} n^{-1} (|\mu| - \mu)(n) \\
&= \sum_{n \leq x} n^{-1} \mu(n). \qquad (4.26)
\end{aligned}$$

Therefore, $\sum_{n=1}^{\infty} n^{-1}\mu(n)$ is finite. And, since $\sum_{n=1}^{\infty} n^{-1}1(n)$ is unbounded, we conclude from (4.22) and Lemma 4.3 the abscissa of convergence of $\widehat{\lambda_G}(s)$ is

$$\sigma_c(\widehat{\lambda_G}) = 1.$$

Now, recall that $\zeta(s)$ is meromorphic for $\sigma > 0$ and that $\zeta(s)$ does not vanish for $\sigma \geq 1$. These two results imply there exists an open set S that contains $\{s : \sigma \geq 1\}$ such that $\zeta(s)^{-1}$ is analytic on S. Consequently, referring to (4.20) and (4.24), $\widehat{\lambda_G}(s)$ can be written

$$\begin{aligned}\widehat{\lambda_G}(s) &= \frac{1}{s-1} + (H(s) + \zeta(s)^{-1}) \text{ for } \sigma > 1 \\ &=: \frac{1}{s-1} + \varphi(s)\end{aligned}$$

where $\varphi(s)$, in particular, is continuous on $\{s : \sigma \geq 1\}$.

Applying Weiner-Ikehara (Theorem 4.7) yields

$$\begin{aligned}N(x)+M(x) &= (N+M)(x) = G(x) = x + o(x) \\ \Rightarrow M(x) &= x - [x] + o(x) = o(x).\end{aligned}$$

Therefore, by Lemma 4.9 and Theorem 3.4,

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

■

Chapter 5

Conclusion

We have seen throughout the paper the means by which measure theory can play a role in problems of number theory. Let us now give specific examples of the usefulness of these applications.

Notice how Lebesgue integration and the integration by parts formula came into play in Section 3.3. A standard proof requires the evaluation of $\sum_{n\le x} \log n \cdot \Lambda(n)$. Written as a Lebesgue integral, this becomes $\int_{(1,x]} \log t \, d\lambda_\psi(t)$ which in turn is equal to $-\int_{(1,x]} \psi(t)t^{-1} d\lambda_1(t) + (\log x)\psi(x)$ using integration by parts. This, then, is easily seen to be $(\log x)\psi(x) + O(x)$. So, in three short steps, we are able to evaluate a sum that otherwise would not at all be obvious.

Further examples of the utility of incorporating measure theory into the proofs are evident in the derivation of Selberg's formulae in Chapter 3 and the proof that $M(x) = o(x) \Rightarrow \psi(x) - x = o(x)$ in Chapter 4. In both instances, we exploited to full advantage (1) the commutative, associative, and distributive properties of $\mathscr{L}(\mathcal{E})$ with respect to the generalized multiplicative convolution of measures, and (2) the fact that λ_N and λ_M are inverses of each other. Using this type of strategy allowed us to temporarily leave the untidy world of summations and integrations, enter the clean world of convolutions, do our manipulating and simplfying there, then return. In this way, this part of the derivations became algebraic in nature rather than relying on the evaluation of sums or integrals.

Lastly, let us return to the example in the introduction. We recall the unwieldy expression:

$$
\begin{aligned}
&-\int_1^x \frac{1}{u}\Bigg(\int_1^u \frac{1}{t}\left(\sum_{n\le u/t}\log n\cdot\Lambda(n)-\frac{u}{t}\log\left(\frac{u}{t}\right)+\frac{u}{t}-1\right)dt\\
&\qquad+\sum_{n\le u}\left(\psi\left(\frac{u}{n}\right)-\frac{u}{n}\right)\Lambda(n)\Bigg)du\\
&+(\log x)\Bigg(\int_1^x \frac{1}{t}\left(\sum_{n\le x/t}\log n\cdot\Lambda(n)-\frac{x}{t}\log\left(\frac{x}{t}\right)+\frac{x}{t}-1\right)dt\\
&\qquad+\sum_{n\le x}\left(\psi\left(\frac{x}{n}\right)-\frac{x}{n}\right)\Lambda(n)\Bigg)\\
&-\sum_{m\le x}\Bigg(\int_1^{\frac{x}{m}} \frac{1}{t}\left(\sum_{n\le x/(mt)}\log n\cdot\Lambda(n)-\frac{x}{mt}\log\left(\frac{x}{mt}\right)+\frac{x}{mt}-1\right)dt\\
&\qquad+\sum_{n\le x/m}\left(\psi\left(\frac{x}{mn}\right)-\frac{x}{mn}\right)\Lambda(n)\Bigg)\Lambda(m).
\end{aligned}
$$

In terms of measures, this becomes (3.36) from Section 3.5:

$$\Big[L\Big((L\lambda_\psi-L\lambda_1)*T^{-1}\lambda_1+\xi*\lambda_\psi\Big)-\Big((L\lambda_\psi-L\lambda_1)*T^{-1}\lambda_1+\xi*\lambda_\psi\Big)*\lambda_\psi\Big]((0,x]).$$

Using the above mentioned methods of convolution, Lebesgue integration, integration by parts, and earlier results; we were able to evaluate this expression in a methodical manner in just a few steps as

$$(\log^2 x)\xi((0,x])+\xi*\big(L\lambda_\psi-\lambda_\psi*\lambda_\psi\big)((0,x])+O(x\log x).$$

The operative word here is "methodical". Being methodical results in a minimum expenditure of thought required to attain the goal in question, which is always desirable. Not only does the notation place the expression in compact form, but notice how straightforward the evaluation process became by referring to results of prior convolutions, employing the integration by parts formula, and estimating remainders via antiderivatives. The proofs, most notably the "elementary" proof, are abundant with these types of examples.

We feel we have given a satisfactory demonstration on the potential usefulness of measure theory as applied to number theory. Nonetheless, there is room for further progress. Expanding on the idea of the algebra of set functions $\mathscr{L}(\mathcal{E})$, comes to mind. Notably, inverses (where they exist) of more of the elements could be found. For example, a short calculation shows that the inverse of λ_Q is defined by $\lambda_Q^{-1} := \lambda_M * \lambda_S$ where λ_S is the measure induced by $S(x) := \sum_{n \leq x} 1_s(n)$ where $s = \{\text{perfect squares}\}$. Finding such inverses would increase the simplification possibilities in multiply convoluted expressions. Another improvement would be to define the algebra over $\mathbb{C}$ rather than $\mathbb{R}$, thus permitting the analysis of complex valued problems. In addition, an interesting theoretical development would be to generalize the methods to spaces other than $\mathbb{R}^+$, such as arbitrary abelian topological groups. The convolution can be defined similarly on these spaces. Implementing these changes in turn would allow the inclusion of a greater number of problems that would benefit from the techniques.

As with any endeavor, however, the methods must be learned well enough so they can be applied with confidence. This is where the trade-off enters. For the author, at least, only a few weeks was necessary to gain the ability to apply these methods instinctively. The reader must decide for himself if the mastery of the material is worth the expenditure of time involved.

Granted, the proof of the PNT does not possess the complexity to demonstrate the full potential of these methods; it was chosen merely to stimulate the reader's curiosity of the possible benefits of applying measure theory to a number theory problem. However, it is not hard to imagine that application of the techniques might well be beneficial on a problem of a suitable complexity such as one that involves (1) repeatedly convoluted elements or (2) the evaluation of complicated integro-series expressions; particularly if the above mentioned further development of $\mathscr{L}(\mathcal{E})$ is undertaken. For the number theorist, especially, the dividends possibly gained might well be worth the investment of time in becoming proficient in the methods.

Bibliography

[1] Ahlfors, Lars V. *Complex analysis*. New York: McGraw-Hill, 1979. Print.

[2] Bateman, P. T.; Diamond, Harold G. *Analytic number theory: an introductory course*. New Jersey: World Scientific, 2004. Print.

[3] Goldfield, D. *The elementary proof of the prime number theorem: an historical perspective*. http://www.math.columbia.edu/ goldfeld/ErdosSelbergDispute.pdf.

[4] Levinson, Norman. *A motivated account of an elementary proof of the prime number theorem*. The American Mathematical Monthly, vol. 76 (1969), pp. 225-245.

[5] McDonald, John N.; Weiss, N. A. *A course in real analysis*. San Diego: Academic, 2013. Print.

[6] Newman, D. J. *Simple analytic proof of the prime number theorem*. The American Mathematical Monthly. Vol. 87, No. 9 (Nov., 1980), pp. 693-696.

[7] Selberg, Alte. *An elementary proof of the prime number theorem*. Annals of Mathematics, Second Series, Vol. 50, No. 2 (Apr., 1949), pp. 305-313.

[8] Tatuzawa, Tikao; Iseki, Kanesiroo. *On Selberg's elementary proof of the prime-number theorem*. Proc. Japan Acad. 27 (1951), no. 7, 340–342. doi:10.3792/pja/1195571353. http://projecteuclid.org/euclid.pja/1195571353.

[9] Zagier, D. *Newman's short proof of the prime number theorem*. The American Mathematical Monthly, Vol. 104, No. 8 (Oct., 1997), pp. 705-708.

Index