Pontificia Universidad Católica del Perú
Facultad de Ciencias e Ingeniería

Análisis, algoritmos y estimados de la identidad de Selberg

Trabajo de investigación para optar el grado académico de
Bachiller en Matemáticas

Autor
Manuel Alejandro Loaiza Vasquez
DNI 70452019

Asesor
Alfredo Bernardo Poirier Schmitz
ORCID 0000-0003-2789-3630
DNI 10803756

Lima, Perú
Julio 2023

Agradecimientos

Primero, quiero agradecer profundamente a mi asesor por su guía y comprensión durante mi carrera universitaria. Alfredo es un asesor fantástico: un oráculo de perspicaces observaciones, balanceado con humor y anécdotas sobre su experiencia.

Aún recuerdo hace unos años atrás, cuando cursaba el cuarto ciclo, haber aprendido un poco sobre teoría de números para los regionales de ICPC y divertirme con Alfredo leyendo elegantes artículos desarrollando aproximadamente el 90% de este trabajo, el cual lo presentamos en el 2020 y ganamos el primer concurso de iniciación a la investigación en estudios generales ciencias. Un compañero y yo comenzamos a trabajar con Alfredo tras tomar el curso Tópicos de Análisis al inicio de la pandemia luego de una charla conjunta en la que nos escogió como asesorados:

> La idea es salir de requisitos formales lo antes posible y seguir avanzando en sus carreras de investigadores. Las tesis deben tomarse como un juego, y como tal, las deben dar por finiquitadas lo antes posible.

Desafortunadamente, no seguí la línea de investigación que Alfredo visionó puesto que la pandemia hizo florecer en mí el entusiasmo de construir los bloques fundamentales de nuestras aplicaciones, dedicándome profesionalmente a la Ingeniería de Software en paralelo a la carga académica los dos años y medio posteriores, motivo por el cual el 10% restante de esta tesis está siendo concluido durante este último ciclo universitario.

Quiero agradecer a mis amistades a lo largo de los años: Marcelo Gallardo, Jemisson Coronel, Diego Hurtado de Mendoza y Hans Acha. Tengo la suerte de haber trabajado con estas personas supertalentosas.

Finalmente, quiero agradecer a mi familia por su apoyo constante e incondicional.

Abstract

One central theme of number theory is the distribution of the prime numbers over the positive integers. In one direction, from the works of Hadamard, de la Valleé Poussin and Newman, we know that the PNT (Prime Number Theorem) can be worked by complex analysis methods. On another direction, Selberg, Breusch, and Levinson proved the PNT using elementary techniques in the sense that it uses only real analysis. Less than a decade ago, Choudhary has strengthened Levinson's proof. All the elementary proofs mentioned above derive the PNT using Selberg's identity.

In this thesis, we establish another proof for the Selberg's identity simpler than Choudhary's in several respects refining the works discussed earlier. We also present a linear time algorithm for estimating a formula derived from Selberg's identity.

Keywords: Selberg's identity, analytic number theory, algorithmic number theory.

2020 Mathematics Subject Classification: 11A25, 11A41, 11Y16.

Resumen

Un tema central en la teoría de números es la distribución de los números primos sobre los enteros positivos. En una dirección, de los trabajos de Hadamard, de la Valleé Poussin and Newman, nosotros sabemos que el PNT (de su acrónimo en inglés *Prime Number Theorem*, Teorema del Número Primo) es cierto por métodos del análisis complejo. En otra dirección, Selberg, Breusch y Levinson probaron el PNT vía técnicas elementales, en el sentido de que solo usan análisis real. Hace menos de una década, Choudhary fortaleció la prueba de Levinson. Todas las pruebas elementales mencionadas derivan el PNT usando la identidad de Selberg.

En esta tesis, establecemos otra prueba para la identidad de Selberg más simple que la de Choudhary en muchos aspectos refinando los trabajos discutidos previamente. También presentamos un algoritmo de tiempo lineal para estimar una fórmula derivada de la identidad de Selberg.

Palabras clave: Identidad de Selberg, teoría analítica de números, teoría algorítmica de números.

Índice general

1 Introducción **1**

1.1 Nuestros resultados 1

1.2 Nuestras técnicas 2

1.3 Notación 2

2 Preliminares matemáticos **3**

3 La identidad de Selberg **5**

4 Experimentación numérica **17**

4.1 Algoritmos 17

4.2 Estimados 22

Bibliografía **22**

Capítulo 1

Introducción

El teorema del número primo ha sido extensivamente estudiado [Bre60, Cho17, Dia82, Erd49, Lev69, Liu22, New80, Ric21, Sel49, Sha59]. Este teorema afirma que la función $\pi(x)$ se aproxima asintóticamente a $x/\ln x$.

En 1852, Chebyshev [Che52] acota $\pi(x)/(x/\ln x)$ y concluye que el límite es igual a 1 en caso exista. En 1896, Hadamard [Had96] y de la Valleé Poussin probaron el PNT vía la función ζ de Riemann. En 1949, Selberg [Sel49] probó este teorema elegantemente sin uso de análisis complejo. Un par de meses luego, Erdős [Erd49] probó el teorema mediante el abuso de estimados tauberianos. En 1959, Shapiro [Sha59] prueba un par de teoremas tauberianos y equivalencias relacionadas a [Erd49], lo cual desembocó también en una nueva prueba del PNT. En 1969, Levinson [Lev69] se inspira de [Sel49, Bre60] para crear una demostración elemental profundizando el estudio de la función resto. En 2017, Choudhary [Cho17] prueba elementalmente el PNT reemplazando ciertos resultados de [Lev69] por corolarios de [Sha59].

El artículo de Selberg es vital para las pruebas elementales posteriores. Levinson y Choudhary invierten más de la mitad de sus trabajos en probar la fórmula asintótica de Selberg para luego rematar el PNT vía la función resto. Por otro lado, Breusch creativamente obtiene fórmulas asintóticas con la misma estructura que la de Selberg para funciones que él define en la primera parte de su artículo.

1.1 Nuestros resultados

El propósito de este trabajo es doble: en primer lugar presentaremos una nueva prueba elemental de la fórmula de Selberg en todo rigor; luego diseñaremos, analizaremos e implementaremos algoritmos para su verificación numérica.

Empezamos enunciando la fórmula asintótica de Selberg. En todo lo que sigue los símbolos p y q se referirán a números primos positivos.

Teorema 1.1 (Identidad de Selberg)**.** *Para todo número real x mayor o igual a 1 se cumple la fórmula*

$$\sum_{p \le x} \ln^2 p + \sum_{pq \le x} \ln p \ln q = 2x \ln x + O(x).$$

Ahora presentamos el aporte computacional en cuestión.

Teorema 1.2. *Sea x un número real mayor o igual a 1. Existe un algoritmo de tiempo $O(x)$ que computa*

$$K(x) = \frac{\left| \sum_{p \le x} \ln^2 p + \sum_{pq \le x} \ln p \ln q - 2x \ln x \right|}{x}.$$

1.2 Nuestras técnicas

Identidad de Selberg. [Sel49, sección 2] nos provee esta identidad como la herramienta más poderosa para probar el PNT. Para demostrarla, usaremos como hoja de ruta [Cho17] sin utilizar las fórmulas que derivan de [Cho17, teorema 3.17] ni los estimados tauberianos de Shapiro desarrollados en [Cho17, sección 4]. En su reemplazo, utilizaremos nuestra creatividad y ciertos resultados de [Apo76, Dia82, TI51].
Experimentación numérica. Nuestro algoritmo surge de la observación de que existen $O(x/\ln x)$ primos menores o iguales a x. Así, usando búsqueda binaria sobre números primos y sumatorias precomputados linealmente, realizamos la estimación en tiempo $O(x)$.

1.3 Notación

Emplearemos $f(x) = O(g(x))$ en vez de $f \in O(g)$ a pesar de que no se trate de una igualdad de conjuntos sino pertenencia de una función a una clase determinada de funciones. De la misma manera trataremos la aritmética entre familias de funciones con notación O. Los símbolos p y q, en caso de no especificarse, harán referencia a números primos positivos.

Capítulo 2

Preliminares matemáticos

En ruta a la identidad de Selberg tendremos que recordar algunas definiciones y estimados bastante conocidos.

Dada una función g denotamos $O(g(x))$ al conjunto de funciones

$$O(g(x)) = \{f : \text{existe una constante positiva } c \text{ y un momento } x_0 \text{ tal que}\\ 0 \leq f(x) \leq cg(x) \text{ para todo } x \geq x_0\}.$$

Teorema 2.1 (Fórmula de sumación de Euler). *Si f tiene una derivada continua f' en el intervalo $[a, b]$, con $0 < a < b$, entonces se satisface*

$$\sum_{a<n\leq x} f(n) = \int_a^b f(t)\, dt + \int_a^b (t - \lfloor t \rfloor) f'(t)\, dt + f(a)(\lfloor a \rfloor - a) - f(b)(\lfloor b \rfloor - b).$$

Teorema 2.2 (Fórmula de sumación de Abel). *Sea $a : \mathbb{N} \to \mathbb{C}$ una función aritmética y $A(x) = \sum_{n\leq x} a(n)$. Si f es una función con derivada continua en el intervalo $[y, x]$ con $0 < y < x$, entonces se cumple*

$$\sum_{y<n\leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)\, dt.$$

Para dos funciones aritméticas f y g definimos su **producto de Dirichlet** como la función aritmética h definida puntualmente por

$$h(n) = f * g(n) = \sum_{d\,|\,n} f(d) g\left(\frac{n}{d}\right).$$

Teorema 2.3 (Fórmula de Inversión de Möbius). *La ecuación*

$$f(n) = \sum_{d\,|\,n} g(d)$$

equivale a

$$g(n) = \sum_{d\,|\,n} f(d) \mu\left(\frac{n}{d}\right).$$

La función $\boldsymbol{\mu}$ **de Möbius** es definida como de costumbre. Primero ponemos $\mu(1) = 1$. Si $n > 1$, expresamos $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ como producto de primos distintos y definimos

$$\mu(n) = \begin{cases} (-1)^k & \text{si } \alpha_1 = \cdots = \alpha_k = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Para n entero positivo la función $\boldsymbol{\Lambda}$ **de Mangoldt** está dada por

$$\Lambda(n) = \begin{cases} \ln p & \text{si } n = p^m \text{ para algún } m \geq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Observe que Λ se define de modo que se satisfaga $\Lambda * \mathbf{1} = \ln$.

Para $x > 0$ definimos la función $\boldsymbol{\Psi}$ **de Chebyshev** con la fórmula

$$\Psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{\substack{m=1 \\ p^m \leq x}}^{\infty} \sum_{p} \Lambda(p^m) = \sum_{m=1}^{\infty} \sum_{p \leq x^{1/m}} \ln p.$$

Para todo $x > 0$ definimos la función $\boldsymbol{\vartheta}$ **de Chebyshev** mediante la ecuación

$$\vartheta(x) = \sum_{p \leq x} \ln p.$$

Capítulo 3

La identidad de Selberg

El objetivo de este capítulo es demostrar [Sel49, ecuación 2.8]. La prueba clásica de Selberg depende de la función número de divisores de n denotada por $\tau(n)$ y la función $\theta_n(x)$ definida como $\sum_{d\,|\,n}\mu(d)\ln^2(x/d)$. Nosotros no usamos estas funciones sino utilizamos las fórmulas asintóticas de [Cho17, corolario 4.2]. Choudhary enuncia estas como corolarios de los teoremas tauberianos de [Sha59] desarrollados en [Cho17, teorema 4.1]. En contraste, nuestras pruebas son trasparentes en uso de aquellos teoremas tauberianos y serán fruto de resultados clásicos de [Apo76] junto con brillantes ideas de artículos que gravitan en torno al tema.

Lema 3.1. *Para todo $x \geq 1$ tenemos*

$$\sum_{n\leq x}\frac{1}{n} = \ln x + \gamma + O\left(\frac{1}{x}\right),$$

aquí $\gamma \approx 0.5772\ldots$ es una constante conocida como ***la constante de Euler****.*

Demostración. La función $f : [1,\infty) \to \mathbb{R}$ con $f(x) = 1/x$ es continua y diferenciable en toda la recta, conque podemos aplicar el teorema 2.1 en cualquier intervalo $[2,k]$ y así obtener

$$\sum_{n=2}^{k}\frac{1}{n} = \int_1^k \frac{dt}{t} + \int_1^k (t-\lfloor t\rfloor)\left(\frac{1}{-t^2}\right)dt = \ln k - \int_1^k \frac{t-\lfloor t\rfloor}{t^2}\,dt,$$

lo cual conduce de inmediato a

$$\sum_{n=1}^{k}\frac{1}{n} - \ln k = 1 - \int_1^k \frac{t-\lfloor t\rfloor}{t^2}\,dt.$$

Para analizar qué ocurre cuando $k\to\infty$ escribimos

$$\gamma = \lim_{k\to\infty}\left(\sum_{n=1}^{k}\frac{1}{n} - \ln k\right) = 1 - \lim_{k\to\infty}\left(\int_1^k \frac{t-\lfloor t\rfloor}{t^2}\right)dt = 1 - \int_1^\infty \frac{t-\lfloor t\rfloor}{t^2}\,dt, \tag{3.1}$$

límite que existe, pues al tenerse

$$\int_1^k \frac{t-\lfloor t\rfloor}{t^2}\,dt \le \int_1^k \frac{1}{t^2}\,dt = 1-\frac{1}{k} \le 1$$

la convergencia queda garantizada por monotonicidad.

Finalmente, para establecer la fórmula anunciada reemplazamos (3.1) tras el uso del teorema 2.1 y obtenemos

$$\begin{aligned}
\sum_{n\le x}\frac{1}{n} &= \int_1^x \frac{1}{t}\,dt + \int_1^x (t-\lfloor t\rfloor)\left(-\frac{1}{t^2}\right)dt - (x-\lfloor x\rfloor)\left(\frac{1}{x}\right)+1\\
&= \ln x - \int_1^x \frac{t-\lfloor t\rfloor}{t^2}\,dt + 1 - \frac{x-\lfloor x\rfloor}{x}\\
&= \ln x - \int_1^x \frac{t-\lfloor t\rfloor}{t^2}\,dt + 1 - \int_1^\infty \frac{t-\lfloor t\rfloor}{t^2}\,dt + \int_1^\infty \frac{t-\lfloor t\rfloor}{t^2}\,dt - \frac{x-\lfloor x\rfloor}{x}\\
&= \ln x + \gamma + \int_x^\infty \frac{t-\lfloor t\rfloor}{t^2}\,dt - \frac{x-\lfloor x\rfloor}{x}\\
&\le \ln x + \gamma + \int_x^\infty \frac{1}{t^2}\,dt - \frac{x-\lfloor x\rfloor}{x}\\
&= \ln x + \gamma + \frac{1-(x-\lfloor x\rfloor)}{x},
\end{aligned}$$

alcanzando así nuestro objetivo. □

Lema 3.2. *Para todo $x \ge 1$ tenemos*

$$\sum_{n\le x}\ln n = x\ln x - x + O(\ln x).$$

Demostración. Esta vez utilizamos el teorema 2.1 en $f:[1,\infty)\to\mathbb{R}$ con $f(x)=\ln x$, función continua y diferenciable en toda la recta real positiva:

$$\begin{aligned}
\sum_{n\le x}\ln x &= \int_1^x \ln t\,dt + \int_1^x \frac{t-\lfloor t\rfloor}{t}\,dt + (\lfloor x\rfloor - x)\ln x\\
&= x\ln x - x + 1 + \int_1^x \frac{t-\lfloor t\rfloor}{t}\,dt + (\lfloor x\rfloor - x)\ln x\\
&\le x\ln x - x + 1 + \int_1^x \frac{1}{t}\,dt + \ln x\\
&= x\ln x - x + 1 + 2\ln x.
\end{aligned}$$

Como 1 está dominado por $\ln x$, se cumple $\sum_{n\le x}\ln n = x\ln x - x + O(\ln x)$, lo anunciado. □

Lema 3.3. *Para toda función aritmética f se cumple*

$$\sum_{n\le x}\sum_{d\,|\,n} f(d) = \sum_{n\le x} f(n)\left\lfloor\frac{x}{n}\right\rfloor.$$

Demostración. Sean f y g dos funciones aritméticas, F y G sus respectivas cumulativas; es decir, $F(x) = \sum_{n\leq x} f(x)$ y $G(x) = \sum_{n\leq x} g(x)$. La cumulativa del producto de Dirichlet de f y g está dada por

$$\sum_{n\leq x} f{*}g(n) = \sum_{n\leq x}\sum_{cd=n} f(c)g(d) = \sum_{c\leq x}\sum_{d\leq x/c} f(c)g(d) = \sum_{c\leq x} f(c) \sum_{d\leq x/c} g(d) = \sum_{c\leq x} f(c)G\left(\frac{x}{c}\right). \tag{3.2}$$

En particular, cuando $g = \mathbf{1}$, su cumulativa es

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

De este modo, al introducir $G(x) = \lfloor x \rfloor$ en (3.2) se logra

$$\sum_{n\leq x}\sum_{d|n} f(d) = \sum_{n\leq x} f * \mathbf{1}(n) = \sum_{n\leq x} f(n) \left\lfloor \frac{x}{n} \right\rfloor ,$$

lo buscado. □

Lema 3.4. *Para todo número real x tenemos*

$$\lfloor x \rfloor = x + O(1).$$

Demostración. Sea $x = n + r$ un número real no negativo, con n entero y $0 \leq r < 1$. De esta manera, por definición de máximo entero obtenemos $\lfloor x \rfloor = n = x - r = x + O(1)$ concluyendo trivialmente con la propiedad. □

Lema 3.5 ([Che52])**.** *Para todo $x \geq 1$ tenemos*

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

Demostración. Usaremos el desarrollo de los estimados de Chebyshev de [Dia82, sección 3] dado por

$$A \leq \liminf \frac{\Psi(x)}{x} \leq \limsup \frac{\Psi(x)}{x} \leq \frac{6A}{5},$$

con

$$A = -\frac{\ln 1}{1} + \frac{\ln 2}{2} + \frac{\ln 3}{3} + \frac{\ln 5}{5} - \frac{\ln 30}{30} \approx 0.921292...$$

Reescribimos la parte derecha de la desigualdad con valor absoluto ya que es una función positiva cual $\limsup |\Psi(x)|/x \leq 6A/5$. Por definición, existe n_0 a partir del cual se tiene $|\Psi(x)|/x \leq 6A/5$, para todo $x \geq n_0$. Como ello equivale a $|\Psi(x)| \leq (6A/5)x$, se consigue $\Psi(x) = O(x)$. □

Lema 3.6. *La función de Mangoldt se puede expresar como el producto de Dirichlet $\Lambda = \mu * \ln$.*

Demostración. Esta fórmula equivale a $\Lambda * \mathbf{1} = \ln$ vía el teorema 2.3. $\square$

Lema 3.7. *Para todo $x \geq 1$ tenemos*

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \ln x + O(1).$$

Demostración. El desarrollo de $\ln = \Lambda * \mathbf{1}$ cual $\ln n = \sum_{d \mid n} \Lambda(d)$ lleva a

$$\sum_{n \leq x} \ln n = \sum_{n \leq x} \sum_{d \mid n} \Lambda(d).$$

Una aplicación directa del lema 3.3 deriva en

$$\sum_{n \leq x} \ln n = \sum_{n \leq x} \Lambda(n) \left\lfloor \frac{x}{n} \right\rfloor.$$

De acá, en uso del lema 3.4 conseguimos

$$\begin{aligned} \sum_{n \leq x} \ln n = \sum_{n \leq x} \Lambda(n) \left(\frac{x}{n} + O(1) \right) &= x \sum_{n \leq x} \frac{\Lambda(n)}{n} + O\left(\sum_{n \leq x} \Lambda(x) \right) \\ = x \sum_{n \leq x} \frac{\Lambda(n)}{n} + O(\Psi(x)) &= x \sum_{n \leq x} \frac{\Lambda(n)}{n} + O(x) \end{aligned}$$

dado que, por el lema 3.5, $\Psi(x) = O(x)$ claramente implica $O(\Psi(x)) = O(x)$. Si aplicamos el lema 3.2 al lado izquierdo desembocamos en

$$x \ln x - x + O(\ln x) = x \sum_{n \leq x} \frac{\Lambda(n)}{n} + O(x).$$

Al despejar obtenemos

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \ln x - 1 + O(1) + O\left(\frac{\ln x}{x} \right) = \ln x + O(1),$$

pues $-1 + O(1) + O(\ln x / x)$ es acotado. $\square$

Lema 3.8. *Para $f, g : [1, \infty) \to \mathbb{R}$ sujetos a $g(x) = \sum_{n \leq x} f(x/n) \ln x$ tenemos*

$$\sum_{n \leq x} \mu(n) g\left(\frac{x}{n} \right) = f(x) \ln(x) + \sum_{n \leq x} f\left(\frac{x}{n} \right) \Lambda(n).$$

Demostración. Desarrollemos la sumatoria que queremos analizar

$$\begin{aligned}
\sum_{n\leq x}\mu(n)g\left(\frac{x}{n}\right) &= \sum_{n\leq x}\mu(n)\sum_{m\leq x/n} f\left(\frac{x}{nm}\right)\ln\left(\frac{x}{n}\right)\\
&= \sum_{nm\leq x}\mu(n)\ln\left(\frac{x}{n}\right) f\left(\frac{x}{nm}\right)\\
&= \sum_{c\leq x} f\left(\frac{x}{c}\right)\sum_{d\,|\,c}\mu(d)\ln\left(\frac{x}{d}\right)\\
&= \sum_{n\leq x} f\left(\frac{x}{n}\right)\sum_{d\,|\,n}\mu(d)\left[\ln\left(\frac{x}{n}\right)+\ln\left(\frac{n}{d}\right)\right]\\
&= \left[\sum_{n\leq x} f\left(\frac{x}{n}\right)\ln\left(\frac{x}{n}\right)\sum_{d\,|\,n}\mu(d)\right]+\left[\sum_{n\leq x} f\left(\frac{x}{n}\right)\sum_{d\,|\,n}\mu(d)\ln\left(\frac{n}{d}\right)\right].\\
&= f(x)+\ln x+\sum_{n\leq x} f\left(\frac{x}{n}\right)(\mu*\ln)(n).
\end{aligned}$$

Con ello, finalmente, utilizamos el lema 3.6 para concluir lo deseado. □

Lema 3.9. *Para todo $x\geq 1$ tenemos*

$$\ln^2 x = O(\sqrt{x}).$$

Demostración. Como sabemos que para todo $x\geq 1$ se cumple que $x>\ln x$ (vía análisis de la derivada de $x-\ln x$), se obtiene

$$\ln^2 x = \ln^2\left(\left(x^{1/4}\right)^4\right) = 16\ln^2\left(x^{1/4}\right) < 16(x^{1/4})^2 = 16\sqrt{x}.$$

□

Lema 3.10. *Para todo $x\geq 1$ tenemos*

$$\Psi(x)\ln x+\sum_{n\leq x}\Psi\left(\frac{x}{n}\right)\Lambda(n) = 2x\ln x+O(x).$$

Demostración. Para utilizar el lema 3.8, definimos convenientemente $f:[1,\infty)\to\mathbb{R}$ con

$$f(x) = \Psi(x)-x+\gamma+1.$$

Primero le brindaremos a $g(x)=\sum_{n\leq x} f\left(x/n\right)\ln x$ una expansión diferente cual es

$$\begin{aligned}
g(x) = \sum_{n\leq x} f\left(\frac{x}{n}\right)\ln x &= \sum_{n\leq x}\left(\Psi\left(\frac{x}{n}\right)-\frac{x}{n}+\gamma+1\right)\ln x\\
&= \sum_{n\leq x}\Psi\left(\frac{x}{n}\right)\ln x - x\ln x\sum_{n\leq x}\frac{1}{n}+(\gamma+1)\ln x\sum_{n\leq x}1. \qquad (3.3)
\end{aligned}$$

Analicemos por separado cada sumatoria de (3.3).

La primera resulta ser

$$\sum_{n\le x}\Psi\left(\frac{x}{n}\right)=\sum_{n\le x}\sum_{d\le x/n}\Lambda(d)=\sum_{n\le x}\sum_{d\,|\,n}\Lambda(d)=\sum_{n\le x}(\Lambda*\mathbf{1})(n)=\sum_{n\le x}\ln n,$$

de tipo $x\ln x-x+O(\ln x)$ por el lema 3.2. Al multiplicar el logaritmo obtenemos la expresión

$$\sum_{n\le x}\Psi\left(\frac{x}{n}\right)\ln x=x\ln^2 x-x\ln x+O(\ln^2 x).$$

Para la segunda recurrimos al lema 3.1 y logramos

$$-x\ln x\sum_{n\le x}\frac{1}{n}=-x\ln x\left(\ln x+\gamma+O\left(\frac{1}{x}\right)\right)=-x\ln^2 x-\gamma x\ln x+O(\ln x).$$

Para la tercera necesitamos el lema 3.4:

$$(\gamma+1)\ln x\sum_{n\le x}1=(\gamma+1)\ln x\lfloor x\rfloor=(\gamma+1)\ln x(x+O(1))=(\gamma+1)x\ln x+O(\ln x).$$

Finalmente, juntamos los tres resultados y obtenemos

$$\begin{aligned}g(x)&=x\ln^2 x-x\ln x+O(\ln^2 x)-x\ln^2 x-\gamma x\ln x+O(\ln x)+(\gamma+1)x\ln x+O(\ln x)\\&=O(\ln^2 x)+O(\ln x)=O(\ln^2 x).\end{aligned}$$

Del lema 3.8 obtenemos entonces

$$\sum_{n\le x}\mu(n)g\left(\frac{x}{n}\right)=(\Psi(x)-x+\gamma+1)\ln x+\sum_{n\le x}\left(\Psi\left(\frac{x}{n}\right)-\frac{x}{n}+\gamma+1\right)\Lambda(n).\tag{3.4}$$

El remate consiste en analizar ambos miembros de la desigualdad por separado. Tatuzawa e Iseki [TI51] trabajaron (3.4) mediante integrales. Nosotros hemos evitado aquello astutamente haciendo uso de la desigualdad triangular junto con el hecho de que se cumple $g(x)=O(\ln^2 x)$ para obtener a la izquierda

$$\left|\sum_{n\le x}\mu(n)g\left(\frac{x}{n}\right)\right|\le\sum_{n\le x}\left|g\left(\frac{x}{n}\right)\right|=O\left(\sum_{n\le x}g\left(\frac{x}{n}\right)\right)=O\left(\sum_{n\le x}\ln^2\left(\frac{x}{n}\right)\right).$$

Con esta expresión, el lema 3.9 nos conduce a

$$\sum_{n\le x}\mu(n)g\left(\frac{x}{n}\right)=O\left(\sum_{n\le x}\sqrt{\frac{x}{n}}\right)=O\left(\sqrt{x}\sum_{n\le x}\frac{1}{\sqrt{n}}\right)=O\left(\sqrt{x}\cdot\sqrt{x}\right)=O(x).\tag{3.5}$$

El término de la derecha en (3.4) lo reordenamos cual

$$\Psi(x)\ln x+\sum_{n\le x}\Psi\left(\frac{x}{n}\right)\Lambda(n)-x\ln x-x\sum_{n\le x}\frac{\Lambda(n)}{n}+(\gamma+1)\Psi(x).\tag{3.6}$$

Merced al lema 3.5 y el lema 3.7 reducimos (3.6) a

$$\begin{aligned}\Psi(x)\ln x + \sum_{n\le x}\Psi\left(\frac{x}{n}\right)\Lambda(n) - x\ln x - x(\ln x + O(1)) + O(x)\\ = \Psi(x)\ln x + \sum_{n\le x}\Psi\left(\frac{x}{n}\right)\Lambda(n) - 2x\ln x + O(x).\end{aligned} \tag{3.7}$$

Finalmente, igualamos (3.5) con (3.7) y obtenemos

$$\Psi(x)\ln x + \sum_{n\le x}\Psi\left(\frac{x}{n}\right)\Lambda(n) - 2x\ln x + O(x) = O(x),$$

equivalente a lo aseverado. □

La prueba del lema siguiente es remedo del argumento utilizado en [Lev69, lema 4] para liquidar

$$\Psi(x) = \pi(x)\ln x + O\left(\frac{x\ln\ln x}{\ln x}\right).$$

Lema 3.11. *Para todo $x \ge 1$ tenemos*

$$\Psi(x) = \vartheta(x) + O\left(\sqrt{x}\ln x\right).$$

Demostración. Directo de la definición observamos que se cumple

$$\Psi(x) = \sum_{n=1}^{\infty}\vartheta\left(x^{1/n}\right).$$

Notemos que al mismo tiempo esta sumatoria tiene apenas una cantidad finita de términos efectivos puesto que la función ϑ solo tiene sentido cuando es evaluada en valores mayores o iguales a 2. Para un x específico, hallamos ese momento m mediante la desigualdad $x^{1/m} \ge 2$, pues elevándola al cuadrado $x^{2/m} \ge 4 > e$ y aplicándole logaritmo $(2/m)\ln x > 1$ obtenemos $2\ln x > m$. Notamos que para valores mayores que $m = \lfloor 2\ln x\rfloor$ los constituyentes de la suma son nulos. Ahora podemos reescribir a Ψ como

$$\Psi(x) = \vartheta(x) + \vartheta(x^{1/2}) + \cdots + \vartheta(x^{1/m}) = \vartheta(x) + \sum_{n=2}^{m}\vartheta(x^{1/n}).$$

Para el análisis de $\sum_{n=2}^{m}\vartheta(x^{1/n})$ desdoblamos

$$\sum_{n=2}^{m}\vartheta(x^{1/n}) = \sum_{n=2}^{m}\sum_{p\le x^{1/n}}\ln p.$$

Trataremos de darle forma manipulativa sencilla. Si un número primo p será inmiscuido en la sumatoria, su logaritmo contribuirá a la sumatoria tantas veces como las raíces de x lo permitan: entre 1 y k, donde k es el máximo entero que obedece $p^k \le x$. Fácilmente hallamos que este máximo está dado por $k = \lfloor \ln x/\ln p\rfloor$.

Por su parte, para forzar por lo menos $p^2 \leq x$, se necesita $p \leq \sqrt{x}$, detalle importante que aprovecharemos.

Ahora analicemos la forma equivalente

$$\begin{aligned}\sum_{n=2}^{m} \sum_{p^n \leq x} \ln p &= \sum_{p \leq \sqrt{x}} \sum_{2 \leq n \leq \lfloor \ln x / \ln p \rfloor} \ln p = \sum_{p \leq \sqrt{x}} \ln p \sum_{2 \leq n \leq \lfloor \ln x / \ln p \rfloor} 1 \leq \sum_{p \leq \sqrt{x}} \ln p \left\lfloor \frac{\ln x}{\ln p} \right\rfloor \\ &\leq \sum_{p \leq \sqrt{x}} \ln p \left(\frac{\ln x}{\ln p} \right) = \sum_{p \leq \sqrt{x}} \ln x = \ln x \sum_{p \leq \sqrt{x}} 1 \leq \ln x \sum_{n \leq \sqrt{x}} 1 \leq \ln x \sqrt{x},\end{aligned}$$

lo que permite concluir $\sum_{n=2}^{m} \vartheta(x^{1/n}) = O(\sqrt{x} \ln x)$. Con lo anterior queda establecida la relación $\Psi(x) = \vartheta(x) + O(\sqrt{x} \ln x)$. □

Lema 3.12. *La serie*

$$\sum_{p=2}^{\infty} \frac{\ln p}{p(p-1)}$$

tomada sobre los primos converge.

Demostración. El primer paso es notar que el límite

$$\lim_{n \to \infty} \frac{\ln n}{n(n-1)} n^{3/2}$$

vale 0 como se deduce al descomponer

$$\frac{\ln n}{n(n-1)} n^{3/2} = \left(\frac{\ln n}{\sqrt{n}} \right) \left(\frac{n}{n-1} \right).$$

Analicemos el límite de lo que está dentro del paréntesis de la izquierda; es decir,

$$\lim_{n \to \infty} \frac{\ln n}{\sqrt{n}} = \frac{\infty}{\infty}.$$

Como este es de la forma ∞/∞, aplicamos la regla de L'Hospital

$$\lim_{n \to \infty} \frac{\ln n}{\sqrt{n}} = \lim_{n \to \infty} \frac{1/n}{1/(2\sqrt{n})} = \lim_{n \to \infty} \frac{2}{\sqrt{n}} = 0.$$

A la derecha tenemos

$$\lim_{n \to \infty} \frac{n}{n-1} = \lim_{n \to \infty} \frac{1}{1 - 1/n} = \frac{1}{1-0} = 1.$$

Como ambos límites existen, por aritmética de límites obtenemos

$$\lim_{n \to \infty} \frac{\ln n}{n(n-1)} n^{3/2} = \left(\lim_{n \to \infty} \frac{\ln n}{\sqrt{n}} \right) \left(\lim_{n \to \infty} \frac{n}{n-1} \right) = 0 \cdot 1 = 0.$$

Por supuesto, lo mismo es válido si crecemos a lo largo de primos, conque se tiene

$$\lim_{p \to \infty} \frac{\ln p}{p(p-1)} p^{3/2} = 0.$$

Por definición entonces, dado $\varepsilon > 0$, existe un n_0 a partir del cual se tiene

$$\frac{\ln p}{p(p-1)} p^{3/2} < \varepsilon$$

De este modo, al hacer $p_0 = \lfloor n_0 \rfloor + 1$, sabemos que para todo $p \geq p_0$ se tendrá

$$\sum_{p \geq p_0} \frac{\ln p}{p(p-1)} < \sum_{p \geq p_0} \frac{\varepsilon}{p^{3/2}}.$$

Como el menor primo es 2, logramos el estimado

$$\sum_{p \geq p_0} \frac{\ln p}{p(p-1)} < \sum_{p \geq p_0} \frac{\varepsilon}{p^{3/2}} \leq \int_1^\infty \frac{\varepsilon}{x^{3/2}}\, dx < 2\varepsilon.$$

Esto, por supuesto, lleva a

$$\sum_{p=2}^{\infty} \frac{\ln p}{p(p-1)} < \sum_{p < p_0} \frac{\ln p}{p(p-1)} + 2\varepsilon,$$

lo que equivale a la convergencia absoluta de la serie. □

Lema 3.13. *Para todo $x \geq 1$ tenemos*

$$\vartheta(x) \ln x + \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) \ln p = 2x \ln x + O(x).$$

Demostración. El primer paso es comparar la sumatoria con otra más a tono con nuestros intereses:

$$\begin{aligned}
\sum_{n \leq x} \Psi\left(\frac{x}{n}\right) \Lambda(n) - \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) \ln p &= \sum_{n \leq x} \sum_{m \leq x/n} \Lambda(m)\Lambda(n) - \sum_{p \leq x} \sum_{q \leq x/p} \ln q \ln p \\
&= \sum_{nm \leq x} \Lambda(n)\Lambda(m) - \sum_{pq \leq x} \ln p \ln q.
\end{aligned}$$

En el primero de los dos sumandos sobrevivientes, la función de Mangoldt solo actúa sobre las potencias de los primos. En particular, todas las combinaciones de primos con potencias iguales a uno van de la mano con la sumatoria que estamos restando a la derecha. De esta manera, apenas sobreviven aquellos términos con al menos uno de los exponentes mayor o igual a dos. De este modo, se consigue

$$\sum_{nm \leq x} \Lambda(n)\Lambda(m) - \sum_{pq \leq x} \ln p \ln q \leq \sum_{\substack{p^n q^m \leq x \\ n \geq 2 \\ m \geq 1}} \ln p \ln q + \sum_{\substack{p^n q^m \leq x \\ n \geq 1 \\ m \geq 2}} \ln p \ln q$$

$$= 2 \sum_{\substack{p^n q^m \leq x \\ n \geq 2 \\ m \geq 1}} \ln p \ln q = 2 \sum_{\substack{p^n \leq x \\ n \geq 2}} \ln p \sum_{\substack{q^m \leq x/p^n \\ m \geq 1}} \ln q = O\left(\sum_{\substack{p^n \leq x \\ n \geq 2}} \ln p\, \Psi\left(\frac{x}{p}\right) \right),$$

puesto que tras la desigualdad contamos por partida doble aquellos pares cuyos productos no son enteros libres de cuadrados. Para continuar, utilizamos el lema 3.5 en la última igualdad y logramos

$$\sum_{n\leq x}\Psi\left(\frac{x}{n}\right)\Lambda(n)-\sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p=O\left(\sum_{\substack{p^n\leq x\\ n\geq 2}}\ln p\,\frac{x}{p^n}\right)=O\left(x\sum_{\substack{p^n\leq x\\ n\geq 2}}\frac{\ln p}{p^n}\right).$$

Llegado este punto, nuevamente notamos que la contribución en la cola es exclusiva de los primos sujetos a $p^2\leq x$. Asimismo, los términos de la sumatoria están dominados por una serie geométrica, por lo que conseguimos

$$\begin{aligned}O\left(x\sum_{\substack{p^n\leq x\\ n\geq 2}}\frac{\ln p}{p^n}\right)&=O\left(x\sum_{p\leq\sqrt{x}}\ln p\sum_{n\geq 2}\frac{1}{p^n}\right)=O\left(x\sum_{p\leq\sqrt{x}}\ln p\sum_{m\geq 0}\frac{1}{p^{m+2}}\right)\\&=O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\sum_{m\geq 0}\frac{1}{p^m}\right)=O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\left(\frac{1}{1-1/p}\right)\right)\\&=O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\left(\frac{p}{p-1}\right)\right)=O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p(p-1)}\right).\end{aligned}$$

Como según el lema 3.12 la serie $\displaystyle\sum_{p=2}^{\infty}\frac{\ln p}{p(p-1)}$ coverge, las sumas parciales están acotadas, y reducimos a

$$\sum_{n\leq x}\Psi\left(\frac{x}{n}\right)\Lambda(n)-\sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p=O\left(x\sum_{p=2}^{\infty}\frac{\ln p}{p(p-1)}\right)=O(x).$$

Al inmiscuir el lema 3.10, esta relación se troca por

$$2x\ln x+O(x)-\Psi(x)\ln x-\sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p=O(x),$$

o lo que es lo mismo por

$$\Psi(x)\ln x+\sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p=2x\ln x+O(x).$$

De acá, el uso consecutivo del lema 3.11 (para Ψ) lleva a

$$2x\ln x+O(x)=(\vartheta(x)+O(\sqrt{x}\ln x))\ln x+\sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p$$

y del lema 3.9 (para $\ln^2 x$) a

$$\begin{aligned}\vartheta(x)\ln x + O(\sqrt{x}\ln^2 x) + \sum_{p\le x}\vartheta\left(\frac{x}{p}\right)\ln p &= \vartheta(x)\ln x + O(\sqrt{x}\sqrt{x}) + \sum_{p\le x}\vartheta\left(\frac{x}{p}\right)\ln p\\ 2x\ln x + O(x) &= \vartheta(x)\ln x + O(x) + \sum_{p\le x}\vartheta\left(\frac{x}{p}\right)\ln p.\end{aligned}$$

□

Lema 3.14. *La serie* $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^2}$ *converge a un número menor o igual a* 2.

Demostración. Esto es sencillo si utilizamos sumas telescópicas:

$$\sum_{n=1}^{k}\frac{1}{n^2} = 1+\sum_{n=2}^{k}\frac{1}{n^2} \le 1+\sum_{n=2}^{k}\frac{1}{n(n-1)} = 1+\sum_{n=2}^{k}\left(\frac{1}{n-1}-\frac{1}{n}\right) = 1+\left(\frac{1}{2-1}-\frac{1}{k}\right) = 2-\frac{1}{k}.$$

El resultado se sigue de inmediato. □

Finalmente el resultado teórico más importante de este capítulo.

Teorema 3.15 ([Sel49])**.** *Para todo* $x\ge 1$ *tenemos*

$$\sum_{p\le x}\ln^2 p + \sum_{pq\le x}\ln p\ln q = 2x\ln x + O(x).$$

Prueba de la fórmula de Selberg. Consolidemos la siguiente diferencia en una única suma

$$\begin{aligned}\vartheta(x)\ln x - \sum_{p\le x}\ln^2 p &= \sum_{p\le x}\ln p\ln x - \sum_{p\le x}\ln p\ln p\\ &= \sum_{p\le x}\ln p(\ln x - \ln p)\\ &= \sum_{p\le x}\ln p\ln\left(\frac{x}{p}\right).\end{aligned} \tag{3.8}$$

A continuación recurrimos a una versión gruesa del lema 3.1; al ser $1/x$ acotado para $x\ge 0$, obtenemos para la serie armónica

$$\sum_{n\le x}\frac{1}{n} = \ln x + \gamma + O\left(\frac{1}{x}\right) = \ln x + O(1),$$

o, lo que es lo mismo,

$$\ln x = \sum_{n\le x}\frac{1}{n} + O(1).$$

Reeplazamos este nuevo estimado en (3.8) para conseguir

$$\begin{aligned}
\vartheta(x)\ln x-\sum_{p\le x}\ln^2 p &= \sum_{p\le x}\ln p\ln\left(\frac{x}{p}\right)=\sum_{p\le x}\ln p\left(\sum_{n\le x/p}\frac{1}{n}+O(1)\right)\\
&=\sum_{p\le x}\ln p\sum_{n\le x/p}\frac{1}{n}+O\left(\sum_{p\le x}\ln p\right)=\sum_{p\le x}\sum_{n\le x/p}\frac{\ln p}{n}+O\left(\sum_{p\le x}\ln p\right)\\
&=\sum_{n\le x}\sum_{p\le x/n}\frac{\ln p}{n}+O\left(\sum_{p\le x}\ln p\right)=\sum_{n\le x}\sum_{p\le x/n}\frac{\ln p}{n}+O(\vartheta(x)). \qquad (3.9)
\end{aligned}$$

Pero una combinación del lema 3.5 y el lema 3.11 muestra que $\Psi(x)$ y $\vartheta(x)$ están en $O(x)$, propiedad que utilizaremos en la forma $O(\vartheta(x))=O(x)$.

A continuación desdoblamos uno de los sumandos en (3.9) para llegar a

$$\begin{aligned}
\vartheta(x)\ln x-\sum_{p\le x}\ln^2 p &= \sum_{n\le x}\sum_{p\le x/n}\frac{\ln p}{n}+O(x)\\
&=\sum_{n\le x}\frac{1}{n}\sum_{p\le x/n}\ln p+O(x)\\
&=\sum_{n\le x}\frac{1}{n}\cdot\vartheta\left(\frac{x}{n}\right)+O(x)\\
&=O\left(x\sum_{n\le x}\frac{1}{n^2}\right)+O(x)\\
&=O(2x)+O(x)\\
&=O(x),
\end{aligned}$$

pues la sumatoria de recíprocos al cuadrado está acotada por 2.

Para el remate es cuestión de reemplazar en el lema 3.13 y lograr

$$\begin{aligned}
2x\ln x+O(x) &= \vartheta(x)\ln x+\sum_{p\le x}\vartheta\left(\frac{x}{p}\right)\ln p\\
&=\sum_{p\le x}\ln^2 p+O(x)+\sum_{p\le x}\vartheta\left(\frac{x}{p}\right)\ln p\\
&=\sum_{p\le x}\ln^2 p+\sum_{p\le x}\sum_{q\le x/p}\ln q\ln p\\
&=\sum_{p\le x}\ln^2 p+\sum_{pq\le x}\ln p\ln q,
\end{aligned}$$

la fórmula de Selberg. $\square$

Capítulo 4

Experimentación numérica

4.1 Algoritmos

En esta sección introduciremos a detalle nuestro algoritmo para estimar la fórmula asintótica de Selberg del capítulo 3. Primero, realicemos ciertas manipulaciones a la fórmula demostrada

$$\frac{\sum_{p\leq x}\ln^2 p + \sum_{pq\leq x}\ln p\ln q - 2x\ln x}{x} = O(1).$$

Por definición, esto significa que existe un momento $n_0 > 0$ y una constante $c > 0$ tales que

$$K(x) = \left|\frac{\sum_{p\leq x}\ln^2 p + \sum_{pq\leq x}\ln p\ln q - 2x\ln x}{x}\right| \leq c$$

para todo $x \geq n_0$. Convenientemente hemos definido K para poder estimar asintóticamente la identidad de Selberg.

Ahora presentaremos como computar K.

Primero precomputaremos los números primos en el rango de interés para poder utilizarlos en ambas sumatorias.

Algorithm 1: Retorna un arreglo dinámico con todos los números primos menores o iguales a x.

```
1  Función ObtenerPrimos(x)
2      Sea M[0...x] un nuevo arreglo estático
3      Inicializamos M[2...x] con VERDADERO
4      M[0] = M[1] = FALSO
5      Sea P un nuevo arreglo dinámico
6      for n = 2 to x do
7          if M[n] then
8              INSERT(P, n)
9          end
10         for p ∈ P and np ≤ x do
11             M[np] = FALSO
12             if p | n then
13                 break
14             end
15         end
16     end
17     return P
18 end
```

Lema 4.1. *El algoritmo* `ObtenerPrimos` *computa correctamente los números primos menores o iguales a* x.

Demostración. Definamos el arreglo estático $M[0\ldots x]$ en la línea 2 donde $M[i]$ denota si i es un número primo y el arreglo dinámico P en la línea 5 el cual contendrá todos los números primos buscados. La línea 4 inicializa $M[0]$ y $M[1]$ en FALSO y el resto es inicializado como VERDADERO en la línea 3. Sea n el número que estamos analizando en las líneas $6-16$. Supongamos que n es un número primo. Luego, las líneas $7-9$ siempre insertarán a n en P puesto que la única manera que $M[n]$ sea FALSO es que algún primo $p \in P$ lo haya etiquetado de esta manera en la línea 11, lo cual es imposible pues $n < pn$. Por el teorema fundamental de la aritmética podemos expresar de manera única cualquier número natural mayor que 1 como producto de potencias de números primos. Sin pérdida de generalidad, expresemos $n = p_1^{\alpha_1}\ldots p_k^{\alpha_k}$ con los factores primos $p_1 \leq \cdots \leq p_k$ ordenados crecientemente. Ahora supongamos que n es compuesto y escribamos $n = p_1 m$ con $m = p_1^{\alpha_1 - 1}\ldots p_k^{\alpha_k}$. Afirmamos que $M[n]$ es FALSO al llegar a la línea 7 y, por ende, no es insertado en P. Necesariamente $1 < m < n$ pues n es compuesto. Como p_1 es menor o igual que todos los primos que dividen a m, la línea 11 asigna a $M[mp_1]$ el valor FALSO cuando el bucle de las líneas $6-16$ está analizando a m. □

Lema 4.2. *El tiempo de ejecución del algoritmo* `ObtenerPrimos` *para computar los números primos menores o iguales a x es $O(x)$.*

Demostración. En la prueba del lema 4.1 observamos que cada número compuesto es etiquetado como FALSO una única vez puesto que determinísticamente sabemos qué primo y qué número su producto hizo que cambie el valor de M en dicho compuesto. En la línea 12 nos aseguramos de que los múltiplos a analizar sean originados por el menor primo que divide a este número compuesto, pues ni bien deje de serlo, la línea 13 evita que sigamos analizando más múltiplos innecesariamente. □

Teorema 4.3. *Sea x un número entero positivo. El algoritmo* `ObtenerPrimos` *computa correctamente los números primos menores o iguales a x en tiempo y espacio $O(x)$.*

Demostración. Inmediato del lema 4.1 y el lema 4.2. □

Vamos preparando el clima para computar $\sum_{pq \leq x} \ln p \ln q$. Si fijamos p y conocemos todos los primos menores o iguales a x, me gustaría saber qué primos q mayores o iguales a p aportan a esta sumatoria. El siguiente algoritmo nos permitirá obtener la menor cota superior para este subconjunto.

Algorithm 2: Retorna el máximo primo q mayor o igual a p tal que $pq \leq x$.

```
Función ObtenerMáximoPrimo(p, x, P)
    l = 0
    r = P.length − 1
    if p · P[r] ≤ x then
        return P[r]
    end
    if p^2 > x then
        return −1
    end
    while r − l > 1 do
        m = ⌊(l + r)/2⌋
        if p · P[m] ≤ x then
            l = m
        else
            r = m
        end
    end
    return P[l]
end
```

Teorema 4.4. *Sea x un número entero positivo, p un número primo menor o igual a x y P un arreglo dinámico con los números primos menores o iguales a x ordenados crecientemente. El algoritmo `ObtenerMáximoPrimo` obtiene el máximo primo q mayor o igual a p tal que $pq \leq x$ o -1 en caso no exista en tiempo $O(\log_2 \pi(x))$.*

Demostración. Una precondición para la línea 4 es que $P[l]$ y $P[r]$ son el menor y mayor número primo menores o iguales a x, respectivamente. En las líneas $4-6$, si p multiplicado por el mayor número primo menor o igual a x sigue siendo menor o igual a x, este sería el primo buscado. Por otro lado, si $p^2 > x$ tenemos que $pq \geq p^2 > x$ para $q \geq p$ por lo que no existiría solución. Así, las líneas $7-9$ se aseguran de retornar -1 frente a esta situación. Caso contrario, notamos que si $p \cdot P[i] \leq x$ entonces $p \cdot P[i-1] \leq x$ para $i > 0$. Luego construimos la invariante $p \cdot P[l] \leq x$ y $p \cdot P[r] > x$ y la preservamos a lo largo del bucle de las líneas $10-17$, el cual se detiene cuando l y r son consecutivos. Como P tiene $\pi(x)$ elementos y en cada paso reducimos el espacio de búsqueda de $r - l + 1$ elementos a $\lfloor (r-l+1)/2 \rfloor + O(1)$ elementos, la complejidad en tiempo será $O(\log_2 \pi(x))$. □

Supongamos que tenemos un arreglo $L[0 \ldots x]$ donde $L[n] = \sum_{p \leq n} \ln p$. Luego podemos calcular $\sum_{l \leq p \leq r} \ln p = L[r] - L[l-1]$ para $l \geq 1$ en $O(1)$. La meta del siguiente algoritmo es computar $L[0 \ldots x]$ y de pasada a $\sum_{p \leq x} \ln^2 p$.

Algorithm 3: Retorna $\sum_{p \leq x} \ln^2 p$ y las sumas parciales de $\sum_{p \leq x} \ln p$.

Función `Sumar`$_1(x, P)$
    Sea $L[0 \ldots x]$ un nuevos arreglo estático
    $L[0] = 0$
    $s = 0$
    $i = 0$
    for $n = 1$ **to** x **do**
        $L[n] = L[n-1]$
        if $i < P.length$ *and* $P[i] == n$ **then**
            $L[n] = L[n] + \ln n$
            $s = s + \ln^2 n$
            $i = i + 1$
        end
    end
    return s *y* L
end

Teorema 4.5. *Sea x un número entero positivo y P un arreglo dinámico con los números primos menores o iguales a x ordenados crecientemente. El algoritmo `Sumar`$_1$ computa $\sum_{p \leq x} \ln^2 p$ y $L[0 \ldots x]$ en tiempo $O(x)$.*

Demostración. Trivialmente el tiempo de ejecución del algoritmo es $O(x)$ pues cada iteración del bucle en las líneas $6-13$ toma $O(1)$. Además, es tácito que el aporte de $\ln^2 n$ a s y de $\ln n$ a $L[n]$ se da cuando n es primo. □

Ya hemos preparado el clima y dispuesto los tecnicismos necesarios para computar el plato de fondo. Sea x un número entero positivo, expresamos

$$\sum_{pq\leq x}\ln p\ln q=\sum_{p\leq x}\sum_{p\leq q\leq r}\ln p\ln q=\sum_{p\leq x}\ln p\sum_{p\leq q\leq r}\ln q=\sum_{p\leq x}\ln p\cdot(L[r]-L[p-1])$$

donde r es el máximo primo mayor o igual a p tal que $pr\leq x$.

Algorithm 4: Retorna $\sum_{pq\leq x}\ln p\ln q$.

```
1  Función Sumar₂(x, L, P)
2      s = 0
3      for p ∈ P do
4          r = ObtenerMáximoPrimo(p, x, P)
5          if r == −1 then
6              break
7          end
8          s = s + ln p · (L[r] − L[p − 1])
9      end
10     return s
11 end
```

Teorema 4.6. *Sea x un número entero positivo, P un arreglo dinámico con los números primos menores o iguales a x ordenados crecientemente y $L[0\ldots x]$ un arreglo en donde $L[n]=\sum_{p\leq n}\ln p$. El algoritmo* `Sumar`$_2$ *computa $\sum_{pq\leq x}\ln p\ln q$ en $O(x)$.*

Demostración. El bucle de las líneas $3-9$ realizará a lo más $\pi(x)$ iteraciones. En cada iteración, la línea 4 toma $O(\log_2\pi(x))$ en tiempo y el resto de líneas toma $O(1)$ por lo que el tiempo de ejecución sería $O\left(\pi(x)\log_2\pi(x)\right)$. Rosser y Schoenfeld [RS62, corolario 1] probaron que $\pi(x)<cx/\ln x$ para $x>1$ y $c=1.25506$. Tras un mero manejo simbólico

$$\begin{aligned}\pi(x)\log_2\pi(x)&\leq\frac{cx}{\ln x}\log_2\frac{cx}{\ln x}\\&=\frac{cx}{\ln x}\left(\log_2 c+\log_2 x-\log_2\ln x\right)\\&=(c\log_2 c)\frac{x}{\ln x}+(c\log_2 e)\,x-c\left(\frac{x\log_2\ln x}{\ln x}\right)\\&=O(x)\end{aligned}$$

nuestra labor es recompensada inmediatamente. □

El siguiente algoritmo en cierto sentido resume nuestro trabajo hasta el momento.

Algorithm 5: Computa $K(x)$.

```
Función K(x)
    P = ObtenerPrimos(x)
    s , L = Sumar₁(x, P)
    t = Sumar₂(x, L, P)
    return |s + t − 2x ln x| / x
end
```

Teorema 4.7. *Sea x un número entero positivo. El algoritmo* `K` *computa $K(x)$ en $O(x)$.*

Demostración. Esto es fruto de el teorema 4.3, el teorema 4.5 y el teorema 4.6. □

4.2 Estimados

Realizamos nuestros experimentos en una MacBook Pro 15−inch 2017 corriendo macOS Ventura 13.4 con 16GB de RAM y un CPU Quad-Core Intel Core i7 con máxima frecuencia de reloj de 2.9GHz. El procesador contiene 4 núcleos y cada núcleo puede aprovechar 2 hilos con hyper-threading technology habilitada. Usamos el lenguaje de programación C++ para implementar nuestros algoritmos, utilizando el estándar C++17. Los números de 64 bits fueron emulados usando los tipos de datos `long long` y `long double` del compilador Apple clang version 14.0.3 (clang-1403.0.22.14.1).

Tabla 4.1: Resultados y tiempos de computación en la MacBook Pro.

x	$K(x)$	**Tiempo de computación (ms)**
10^1	3.4422963847	0
10^2	5.1926865457	0
10^3	6.3173110617	0
10^4	7.4463872101	2
10^5	8.5931825594	18
10^6	9.7471335703	156
10^7	10.8942869828	1406
10^8	12.0436377380	16150
10^9	13.1943961187	201913

El código fuente para replicar los resultados está disponible en GitHub en `https://github.com/ManuelLoaizaVasquez/selberg-identity/tree/main/code/`.

Bibliografía

[Apo76] Tom M. Apostol, *Introduction to Analytic Number Theory*, Springer New York, 1976.

[Bre60] Robert Breusch, *An elementary proof of the prime number theorem with remainder term*, Pacific Journal of Mathematics **10** (1960), no. 2, 487 – 497.

[Che52] Pafnuty Chebyshev, *Mémoire sur les nombres premiers*, Journal de mathématiques pures et appliquées **1** (1852), no. 17, 366 – 390.

[Cho17] Abhimanyu Choudhary, *An Elementary Proof of the Prime Number Theorem*, 2017.

[Dia82] Harold G. Diamond, *Elementary methods in the study of the distribution of prime numbers*, Bulletin (New Series) of the American Mathematical Society **7** (1982), no. 3, 553 – 589.

[Erd49] P. Erdős, *On a New Method in Elementary Number Theory Which Leads to An Elementary Proof of the Prime Number Theorem*, Proceedings of the National Academy of Sciences **35** (1949), no. 7, 374–384.

[Had96] J. Hadamard, *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques*, Bull. Soc. Math. France **24** (1896), 199–220. MR 1504264

[Lev69] Norman Levinson, *A Motivated Account of an Elementary Proof of the Prime Number Theorem*, The American Mathematical Monthly **76** (1969), no. 3, 225.

[Liu22] Zihao Liu, *A Direct Proof of the Prime Number Theorem using Riemann's Prime-counting Function*, Journal of Physics: Conference Series **2287** (2022), no. 1, 012008.

[New80] D. J. Newman, *Simple Analytic Proof of the Prime Number Theorem*, The American Mathematical Monthly **87** (1980), no. 9, 693.

[Ric21] Florian K. Richter, *A new elementary proof of the Prime Number Theorem*, Bulletin of the London Mathematical Society **53** (2021), no. 5, 1365–1375.

[RS62] J. Barkley Rosser and Lowell Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois Journal of Mathematics **6** (1962), no. 1.

[Sel49] Atle Selberg, *An Elementary Proof of the Prime-Number Theorem*, The Annals of Mathematics **50** (1949), no. 2, 305.

[Sha59] Harold N. Shapiro, *Tauberian Theorems and Elementary Prime Number Theory*, Communications on Pure and Applied Mathematics **12** (1959), no. 4, 579–610.

[TI51] Tikao Tatuzawa and Kanesiroo Iseki, *On Selberg's Elementary Proof of the Prime-Number Theorem*, Proceedings of the Japan Academy, Series A, Mathematical Sciences **27** (1951), no. 7.