Pontificia Universidad Católica del Perú

Facultad de Ciencias e Ingeniería

# Análisis, Algoritmos y Estimados de la Identidad de Selberg

Trabajo de tesis para optar el grado académico de
Bachiller en Ciencias con mención en Matemáticas


Autor

**Manuel Alejandro Loaiza Vasquez**

DNI 70452019

Asesor

**Alfredo Bernardo Poirier Schmitz**

ORCID 0000-0003-2789-3630

DNI 10803756


**Lima, Perú**

**Mes año**

Análisis, Algoritmos y Estimados de la Identidad de Selberg [1]

**Manuel Alejandro Loaiza Vasquez** [2]

Tesis presentada a consideración del cuerpo docente de la Facultad de Ciencias e Ingeniería de la Pontificia Universidad Católica del Perú como parte de los requisitos para obtener el grado académico de Bachiller en Matemáticas.

Miembros del Jurado:

Dr. nombre del presidente del jurado, xxxx

https://orcid.org/0000-0000-1111-2222

Presidente del jurado

Dr. Alfredo Bernardo Poirier Schmitz, PUCP

https://orcid.org/0000-0003-2789-3630

Asesor

Dr. nombre del jurado, zzzz

https://orcid.org/0000-0000-1111-2222

Tercer miembro

**Lima, Perú**
**Mes año**

[1] Version final con las correcciones del jurado

[2] Proyecto DGI, apoyo financiero, etc.

# Agradecimientos


Primero, quiero agradecer profundamente a mi asesor Alfredo Poirier 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 con su asesor John Milnor o su tiempo en el MSRI en UC Berkeley.

Aún recuerdo hace unos años atrás, cuando cursaba el cuarto ciclo, haber aprendido un poco sobre Teoría de Números para los regionales de ACM ICPC y divertirme con Alfredo leyendo elegantes artículos desarrollando aproximadamente el 80% de este trabajo, el cual lo presentamos en el 2020 y ganamos el Primer Concurso de Iniciación a la Investigación en Estudios Generales Ciencias. Jemisson 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 la adicción y 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 20% restante de esta tesis está siendo concluido en este décimo ciclo universitario.

Quiero agradecer a mis amistades a lo largo de los años: Marcelo Gallardo, Jemisson Coronel, Eduardo Llamoca, 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

In this work, we proof Selberg's Identity using elementary techniques, develop an algorithm in worst-case $O(x)$ time and implement it in C++ to estimate numeric results.

**Key Words:** Algorithmic Number Theory, Analytic Number Theory, Linear Sieve, Prime Number Theorem, Selberg's Identity.

**2010 Mathematics Subject Classification:**

# Resumen

Incluir el resumen de la tesis y tomar en cuenta el contexto de los tópicos considerados en el trabajo de tesis. Evite cualquier apreciación o juicio crítico. Es razonable que por medio del resumen se dé una descripción clara de los objetivos de la tesis, métodos utilizados y aportes del trabajo final.

**Palabras clave:** Palabra Clave 1, Palabra Clave 2, Palabra Clave 3.

# Índice general

# Índice de figuras

# Capítulo 1

# Introducción

Riemann, Erdos, Selberg, Newman, Tao y otros matemáticos han demostrado el teorema del número primo de distintas maneras así como Chebyshev y Euler lograron resultados parciales. Todos aportaron teoremas y técnicas que han logrado el desarrollo de nuevas teorías, así como la solución de conjeturas y propuestas de hipótesis aún sin una demostración. Además, desde hace miles de años Euclides y Eratóstenes como también en los últimos cincuenta años Miller, Rabin, Pollard, Gries y otros científicos han podido desarrollar algoritmos eficientes que permiten implementar y conseguir resultados combinando la Teoría de Números y las Ciencias de la Computación, lo cual tiene un alto impacto en el mundo contemporáneo tanto en la teoría como en la práctica, lo cual se ve reflejado en ramas como la Criptografía y la Matemática Computacional.

## 1.1 Nuestros resultados

El propósito de este trabajo es doble: en primer lugar probaremos la fórmula de Selberg en todo rigor (ver enunciado a continuación); luego diseñaremos y analizaremos la eficiencia de nuestros algoritmos e implementaremos programas para su verificación númerica.

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

**Teorema (La identidad de Selberg)** *Para todo número real $x$ mayor o igual a* 1 *se cumple la fórmula de Selberg*

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

## 1.2 Nuestras técnicas

Para obtener nuestros resultados, hemos utilizado herramientas básicas del análisis real, ejemplos de los cuales tenemos series, sucesiones, continuidad, límites, derivadas e integrales. Asimismo, haremos uso de funciones aritméticas y estimados de estas, las cuales serán controladas haciendo uso de la notación $O$, la cual también utilizaremos para realizar el análisis de complejidad asintótico de los algoritmos. El algoritmo principal dependerá de otros algoritmos que realizarán búsquedas binarias y una criba lineal para deteminar los números primos en un rango de modo eficiente. Finalmente, los estimados computacionales serán obtenidos tras realizar una implementación de los algoritmos propuestos en el lenguaje GNU C++ 17.

## 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 de funciones; de la misma manera trataremos la aritmética entre familias de funciones con notación *big* $O$. Los símbolos $p$ y $q$, en caso de no especificarse, harán referencia a números primos positivos.

## 1.4 Organización

En la sección 2 presentaremos los teoremas y definiciones que no probaremos pero son lugar común en área, utilizaremos como referencia [Apo76] y [CLRS09]. En la sección 3 realizaremos la demostración de la identidad de Selberg tras la prueba de ciertos lemas intermedios. En la sección 4 propondremos los algoritmos CribaLineal, BuscarÚltimaPosición y CalcularSuma para poder cumplir con el objetivo de realizar estimados con el algoritmo EstimarConstante, el mismo que emplea los tres algoritmos anteriores. Todos los algoritmos tendrán su respectivo análisis de complejidad asintótico. En la sección 5 implementaremos los algoritmos de la sección anterior en el lenguaje de programación C++ y presentaremos tablas con los estimados.

# 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*

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

Sean $f$ y $g$ dos funciones aritméticas, 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 2.2** (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 sigue. Primero definimos

$$\mu(1) = 1.$$

Si $n > 1$, expresamos $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ 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 definimos la función **Λ de Mangoldt** vía

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

Para $x > 0$ definimos la función **Ψ 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 **ϑ de Chebyshev** mediante la ecuación

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

# Capítulo 3

# La Identidad de Selberg

**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 $\gamma \approx 0.52$ 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 Theorem 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

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

la convergencia queda garantizada por monotonicidad.

Finalmente, para establecer la fórmula anunciada reemplazamos (3.1) tras el uso de

Theorem 2.1

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

alcanzando nuestro objetivo. □

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

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

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

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

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

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

$$\sum_{n\leq x}\sum_{d\mid n} f(d) = \sum_{n\leq x} f(n)\left\lfloor\frac{x}{n}\right\rfloor.$$

*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

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

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

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

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 $0 \le r < 1$. 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.** *Para todo $x \ge 1$ tenemos*

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

*Demostración.* Usaremos el desarrollo de los estimados de Chebyshev de [Dia82]

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

con

$$A = -\frac{\ln 1}{1} + \frac{\ln 2}{2} + \frac{\ln 3}{3} + \frac{\ln 5}{5} - \frac{\ln 30}{30} \approx 0.92129202293409078091 34...$$

Reescribimos la parte derecha de la desigualdad con valor absoluto ya que es una función positiva cual $\limsup |\Psi(x)|/x \le 6A/5$. Por definición, existe $n_0$ a partir del cual se tiene $|\Psi(x)|/x \le 6A/5$, para todo $x \ge n_0$. Como ello equivale a $|\Psi(x)| \le (6A/5)x$, 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 $\Lambda * 1 = \ln$ vía Theorem 2.2. □

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

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

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

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

Una aplicación directa de Lemma 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 de Lemma 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 Lemma 3.5, $\Psi(x) = O(x)$ claramente implica $O(\Psi(x)) = O(x)$. Si aplicamos Lemma 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* $g(x) = \sum_{n\leq x} f\left(\frac{x}{n}\right) \ln x$ *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 Lemma 3.6 para concluir lo deseado. $\square$

**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 Lemma 3.8, definimos convenientemente $f : [1, \infty) \to \mathbb{R}$ con

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

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

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

Analicemos por separado cada sumatoria de (3.3).

La primera resulta ser

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

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

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

Para la segunda recurrimos a Lemma 3.1 y logramos

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

Para la tercera necesitamos Lemma 3.4:

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

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 Lemma 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. Utilizamos la desigualdad triangular 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, Lemma 3.9 permite conseguir

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

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 a Lemma 3.5 y Lemma 3.7 reducimos (3.6)

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

Finalmente, igualamos (3.5) con (3.7)

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

equivalente a lo aseverado. □

**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 $\vartheta$ solo tiene sentido cuando es evaluada en valores mayores o iguales a 2. Para un $x$ específico, hallamos ese momento $m$ mediante la desigualdad $x^{1/m} \geq 2$, pues elevándola al cuadrado, $x^{2/m} \geq 4 > e$, y aplicándo logaritmo a ambos lados, $(2/m)\ln x > 1$, 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 escribir a $\Psi$ como

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

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

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

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 $p^k \leq x$. 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 de la doble sumatoria

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

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

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

Como este es de la forma $\infty/\infty$, aplicamos la regla de L'Hospital

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

Ahora analicemos el límite de lo que está dentro del paréntesis de la derecha

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

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. $\square$

**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:

$$\sum_{n\leq x} \Psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = \sum_{n\leq x}\sum_{m\leq \frac{x}{n}}\Lambda(m)\Lambda(n) - \sum_{p\leq x}\sum_{q\leq\frac{x}{p}}\ln q\ln p \tag{3.8}$$

$$= \sum_{nm\leq x}\Lambda(n)\Lambda(m) - \sum_{pq\leq x}\ln p\ln q. \tag{3.9}$$

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

$$\sum_{n\leq x} \Psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p \leq \sum_{\substack{p^n q^m\leq x\\ n\geq 2, m\geq 1}}\ln p\ln q + \sum_{\substack{p^n q^m\leq x\\ m\geq 2, n\geq 1}}\ln p\ln q \tag{3.10}$$

$$= 2\sum_{\substack{p^n q^m\leq x\\ n\geq 2, m\geq 1}}\ln p\ln q \tag{3.11}$$

$$= 2\sum_{\substack{p^n\leq x\\ n\geq 2}}\ln p\sum_{\substack{q^m\leq \frac{x}{p^n}\\ m\geq 1}}\ln q \tag{3.12}$$

$$= O\left(\sum_{\substack{p^n\leq x\\ n\geq 2}}\ln p\,\Psi\left(\frac{x}{p^n}\right)\right), \tag{3.13}$$

puesto que tras desigualdad contamos por partida doble aquellos pares con ambos exponentes al menos dos y ello contribuyen con valores positivos. Para continuar, utilizamos el **??** en la última igualdad y logramos

$$\sum_{n\leq x} \Psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = O\left(\sum_{\substack{p^n\leq x\\ n\geq 2}}\ln p\,\frac{x}{p^n}\right) \tag{3.14}$$

$$= O\left(x\sum_{\substack{p^n\leq x\\ n\geq 2}}\frac{\ln p}{p^n}\right). \tag{3.15}$$

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

$$\sum_{n\leq x} \Psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = O\left(x\sum_{p\leq\sqrt{x}}\ln p\sum_{n\geq 2}\frac{1}{p^n}\right) \tag{3.16}$$

$$= O\left(x\sum_{p\leq\sqrt{x}}\ln p\sum_{m\geq 0}\frac{1}{p^{m+2}}\right) \tag{3.17}$$

$$= O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\sum_{m\geq 0}\frac{1}{p^m}\right) \tag{3.18}$$

$$= O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\left(\frac{1}{1-\frac{1}{p}}\right)\right) \tag{3.19}$$

$$= O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p^2}\left(\frac{p}{p-1}\right)\right) \tag{3.20}$$

$$= O\left(x\sum_{p\leq\sqrt{x}}\frac{\ln p}{p(p-1)}\right). \tag{3.21}$$

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

$$\sum_{n\leq x} \Psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = O\left(x\sum_{p=2}^{\infty}\frac{\ln p}{p(p-1)}\right) \tag{3.22}$$

$$= O(x). \tag{3.23}$$

Al inmiscuir al **??**, esta relación se troca por

$$2x\ln x + O(x) - \Psi(x)\ln x - \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = O(x), \tag{3.24}$$

o lo que es lo mismo por

$$\Psi(x)\ln x + \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p = 2x\ln x + O(x). \tag{3.25}$$

De acá, el uso consecutivo del **??** (para $\Psi$) y el **??** (para $\ln^2 x$) deviene en la secuencia

$$(\vartheta(x) + O(\sqrt{x}\ln x))\ln x + \sum_{peqx} \vartheta\left(\frac{x}{p}\right)\ln p = 2x\ln x + O(x), \tag{3.26}$$

$$\vartheta(x)\ln x + O(\sqrt{x}\ln^2 x) + \sum_{p\le x} \vartheta\left(\frac{x}{p}\right)\ln p = 2x\ln x + O(x), \tag{3.27}$$

$$\vartheta(x)\ln x + O(\sqrt{x}\sqrt{x}) + \sum_{p\le x} \vartheta\left(\frac{x}{p}\right)\ln p = 2x\ln x + O(x), \tag{3.28}$$

$$\vartheta(x)\ln x + O(x) + \sum_{p\le x} \vartheta\left(\frac{x}{p}\right)\ln p = 2x\ln x + O(x). \tag{3.29}$$

□

**Lema 3.14.** *La serie*

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

*converge a un número menor o igual a* 2.

*Demostración.* Esto es sencillo si utilizamos sumas telescópicas:

$$\sum_{n=1}^{k} \frac{1}{n^2} = 1 + \sum_{n=2}^{k} \frac{1}{n^2} \tag{3.30}$$

$$\le 1 + \sum_{n=2}^{k} \frac{1}{n(n-1)} \tag{3.31}$$

$$= 1 + \sum_{n=2}^{k} \left(\frac{1}{n-1} - \frac{1}{n}\right) \tag{3.32}$$

$$= 1 + \left(\frac{1}{2-1} - \frac{1}{k}\right) \tag{3.33}$$

$$= 2 - \frac{1}{k}. \tag{3.34}$$

El resultado se sigue de inmediato. □

Finalmente el resultado teórico más importante de esta recopilación.

**Teorema 3.15** (Fórmula asintótica de Selberg)**.** *Para todo* $x \ge 1$ *tenemos*

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

*Nota.* En la fórmula dada arriba es indistinto si toman $p, q$ distintos o si se permite que sean iguales. En efecto, la diferencia entre una y otra alternativa es apenas

$$\sum_{p^2\le x} (\ln p)^2 \le \sum_{p^2\le x} (\ln x^{1/2})^2 \le \frac{\sqrt{x}\ln^2 x}{4} = O(x).$$

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

$$\vartheta(x)\ln x - \sum_{p\leq x}\ln^2 p = \sum_{p\leq x}\ln p\ln x - \sum_{p\leq x}\ln p\ln p \tag{3.35}$$
$$= \sum_{p\leq x}\ln p(\ln x - \ln p) \tag{3.36}$$
$$= \sum_{p\leq x}\ln p\ln\left(\frac{x}{p}\right). \tag{3.37}$$

A continuación recurrimos a una versión gruesa del **??**: Al ser $\frac{1}{x}$ acotado para $x \geq 0$, obtenemos para la serie armónica

$$\sum_{n\leq x}\frac{1}{n} = \ln x + \gamma + O\left(\frac{1}{x}\right) \tag{3.38}$$
$$= \ln x + O(1), \tag{3.39}$$

o, lo que es lo mismo,

$$\ln x = \sum_{n\leq x}\frac{1}{n} + O(1). \tag{3.40}$$

Reeplazamos este nuevo estimado en la Ecuación 3.37 para conseguir

$$\vartheta(x)\ln x - \sum_{p\leq x}\ln^2 p = \sum_{p\leq x}\ln p\ln\left(\frac{x}{p}\right) \tag{3.41}$$
$$= \sum_{p\leq x}\ln p\left(\sum_{n\leq\frac{x}{p}}\frac{1}{n} + O(1)\right) \tag{3.42}$$
$$= \sum_{p\leq x}\ln p\sum_{n\leq\frac{x}{p}}\frac{1}{n} + O\left(\sum_{p\leq x}\ln p\right) \tag{3.43}$$
$$= \sum_{p\leq x}\sum_{n\leq\frac{x}{p}}\frac{\ln p}{n} + O\left(\sum_{p\leq x}\ln p\right) \tag{3.44}$$
$$= \sum_{n\leq x}\sum_{p\leq\frac{x}{n}}\frac{\ln p}{n} + O\left(\sum_{p\leq x}\ln p\right) \tag{3.45}$$
$$= \sum_{n\leq x}\sum_{p\leq\frac{x}{n}}\frac{\ln p}{n} + O(\vartheta(x)). \tag{3.46}$$

Pero una combinación de **??** y **??** conduce a

$$\Psi(x) = O(x), \tag{3.47}$$
$$\vartheta(x) = O(x), \tag{3.48}$$

propiedad que utilizaremos en la forma $O(\vartheta(x)) = O(x)$.

A continuación desdoblamos uno de los sumandos en Ecuación 3.46 para llegar a

$$\vartheta(x)\ln x - \sum_{p\leq x}\ln^2 p = \sum_{n\leq x}\sum_{p\leq \frac{x}{n}}\frac{\ln p}{n} + O(x) \tag{3.49}$$

$$= \sum_{n\leq x}\frac{1}{n}\sum_{p\leq \frac{x}{n}}\ln p + O(x) \tag{3.50}$$

$$= \sum_{n\leq x}\frac{1}{n}\cdot\vartheta\left(\frac{x}{n}\right) + O(x) \tag{3.51}$$

$$= O\left(x\sum_{n\leq x}\frac{1}{n^2}\right) + O(x) \tag{3.52}$$

$$= O(2x) + O(x) = O(x), \tag{3.53}$$

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

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

$$2x\ln x + O(x) = \vartheta(x)\ln x + \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p \tag{3.54}$$

$$= \sum_{p\leq x}\ln^2 p + O(x) + \sum_{p\leq x}\vartheta\left(\frac{x}{p}\right)\ln p \tag{3.55}$$

$$= \sum_{p\leq x}\ln^2 p + \sum_{p\leq x}\sum_{q\leq \frac{x}{p}}\ln q\ln p \tag{3.56}$$

$$= \sum_{p\leq x}\ln^2 p + \sum_{pq\leq x}\ln p\ln q, \tag{3.57}$$

la fórmula de Selberg. $\square$

# Capítulo 4

# Algoritmos

## 4.1 Algoritmo Principal

Adicional a la prueba del Teorema 3.15, hemos diseñado unos algoritmos para poder obtener estimados con la fórmula asintótica de Selberg. Primero, realicemos ciertas manipulaciones a la fórmula demostrada

$$\sum_{p\leq x} \ln^2 p + \sum_{pq\leq x} \ln p \ln q = 2x\ln x + O(x) \tag{4.1}$$

$$\sum_{p\leq x} \ln^2 p + \sum_{pq\leq x} \ln p \ln q - 2x\ln x = O(x) \tag{4.2}$$

$$\frac{\sum_{p\leq x} \ln^2 p + \sum_{pq\leq x} \ln p \ln q - 2x\ln x}{x} = O(1). \tag{4.3}$$

Por definición, esto significa que existe un momento $n_0 > 0$ y una constante $c > 0$ a partir del cual

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

para todo $x \geq n_0$.

Asumamos que tenemos un número $x$ y queremos calcular

$$\left| \frac{\sum_{p\leq x} \ln^2 p + \sum_{pq\leq x} \ln p \ln q - 2x\ln x}{x} \right|.$$

La estrategia que utilizaremos consiste en lo siguiente: Primero precalcularemos los números primos en el rango $[1 \ldots x]$. Como solamente manipularemos logaritmos de números primos, aprovecharé de esto para poder precalcular las sumatorias y hallar sumas en rangos en $O(1)$ como diferencia de sumas. El contenedor log almacenará lo siguiente:

$$\log[p] = \begin{cases} \ln p & \text{si } p \text{ es primo,} \\ 0 & \text{en otro caso.} \end{cases}$$

Así, $\sum_{i=l}^{r} \log[i]$ solo tendrá la sumatoria de los logaritmos naturales de los números primos en el intervalo $[l, r]$.

El contenedor es_primo almacenará lo siguiente:

$$\text{es\_primo}[p] = \begin{cases} \text{verdadero} & \text{si p es primo,} \\ \text{falso} & \text{en otro caso.} \end{cases}$$

Los contenedores suma_log y suma_log$^2$ almacenarán

$$\text{suma\_log}[x] = \sum_{i=0}^{x} \log[i]$$

y

$$\text{suma\_log}^2[x] = \sum_{i=0}^{x} \log^2[i].$$

El primer preprocesamiento que realizaremos será llamar al método CribaLineal. Este método reibirá como parámetro al número n.

**Algorithm 1:** CribaLineal

```
Data: es_primo, primos, n.
Result: Números primos en el rango [1 ... n] guardados en primos.
begin
    es_primo[0] ← falso
    es_primo[1] ← falso
    for i ← 2 to n do
        es_primo[i] ← verdadero
    end
    primos ← ∅
    for i ← 2 to n do
        if es_primo[i] then
            primos ← primos ∪ {i}
        end
        for p ∈ primos and i · p ≤ n do
            es_primo[i · p] ← falso
            if i ≡ 0 mód p then
                break
            end
        end
    end
end
```

**Lema 4.1.** *Sea $n$ el número que representa el extremo derecho del intervalo $[1 \dots n]$ en el cual queremos hallar todos los números primos mediante la ejecución de CribaLineal. Luego el tiempo de ejecución es $O(n)$.*

*Demostración.* $\square$

**Algorithm 2:** BuscarÚltimaPosición

**Data:** $i, x$, primos

**Result:** Posición del mayor número primo $q$ tal que $pq \leq x$. En caso no exista, se retornará $-1$.

```
1  begin
2      p ← primos[i]
3      l ← i
4      r ← |primos| − 1
5      if p · primos[r] ≤ x then
6          return r
7      end
8      if p^2 > x then
9          return −1
10     end
11     while r − l > 1 do
12         m ← ⌊*⌋ (l+r)/2
13         if p · primos[m] ≤ x then
14             l ← m
15         else
16             r ← m
17         end
18     end
19     return l
20 end
```

**Lema 4.2.** *Sea $i$ el índice que representa al $i$-ésimo número primo y $p$ este $i$-ésimo número primo. Sea $x$ el número que representa la cota superior para el producto de $p$ con otro número primo $q$ tal que $pq \leq x$ y $q \geq p$. Obtendremos la posición del mayor número primo $q$ que cumpla con lo anterior o $-1$ en caso este número primo no exista mediante la ejecución de BuscarÚltimaPosición. Luego el tiempo de ejecución es $O(\log_2 n)$.*

*Demostración.* Obtener el valor del $i$-ésimo número primo e inicializar nuestros extremos de los intervalos para realizar la búsqueda binaria en las líneas $2-4$ toma $O(1)$ en tiempo. Nuestro objetivo es encontrar la última posición $j$ en la cual el $j$-ésimo número primo multiplicado por $p$ sea menor o igual a $x$ con $i \leq j$. Antes de analizar la invariante, quitémonos de encima los casos borde. El primer caso borde es en el cual el menor elemento en nuestro rango no cumple la condición, en este caso retornamos $-1$, pues para todo $q > p$ tenemos que $pq > p^2 > x$, por lo que no existirá par que satisfaga la condición. El segundo caso borde es cuando el número de la última posición cumple. En este caso retornamos de inmediato esta última posición, pues para cualquier $q < p_r$ tenemos $pq < p \cdot p_r \leq x$, por lo cual todos los demás primos en el rango también cumplirían. Analizar ambos casos nos tomaría $O(1)$ en tiempo. Al entrar al bucle en las líneas $11 - 18$ se satisface $i = l < r = \pi(x) - 1$ y además $p \cdot \text{primos}[l] \leq x$ y $p \cdot \text{primos}[r] > x$. Mientras que $l$ y $r$ disten al menos 2, computaremos $m = \lfloor * \rfloor \frac{l+r}{2}$ y podemos distinguir dos casos: el primer caso es cuando se cumple $p \cdot \text{primos}[m] \leq x$, $l$ cambiaría su valor a $m$ y, debido a que la diferencia entre $l$ y $r$ era mayor a 1, entonces $l' = m < r$ y la condición $p \cdot \text{primos}[l'] \leq x$ y $p \cdot \text{primos}[r] > x$ se sigue cumpliendo. En el segundo caso tenemos $p \cdot \text{primos}[m] > x$, aquí $r' = m$ y cuando $r' > l$ la condición se cumple de manera análoga al caso anterior, pero cuando $r' = l$, tenemos que en la siguiente iteración del bucle, al ya no distar 2, saldríamos del bucle y necesariamente la respuesta se encuentra en nuestro extremo izquierdo $l$. Ahora que ya sabemos que siempre terminamos obteniendo el último elemento que cumple la condición, analicemos el orden de complejidad de este algoritmo. Sea $T : \mathbb{N} \to \mathbb{N}$ la función que cuenta la cantidad de operaciones que realiza el bucle con $T(n) = 4 + T(\lceil \frac{n}{2} \rceil)$, puesto que en el peor de los casos nos quedamos con la mitad más grande; no obstante, esto lo podemos expresar como $T(n) = O(1) + T(\frac{n}{2})$ según [?] y utilizando inducción sobre esta última conseguimos $T(n) = O(\log_2 n)$. En el peor de los casos, $n = \pi(x)$, por lo que la complejidad total del algoritmo sería $O(\log_2 \pi(x))$, el cual es a su vez $O(\log_2 x)$ por definición. $\square$

**Algorithm 3:** CalcularSuma

**Data:** primos, suma_log, $x$.
**Result:** $\sum_{pq \leq x} \ln p \ln q$.

**begin**
  $suma \leftarrow 0$
  **for** $pos_p \leftarrow 0$ **to** $|primos| - 1$ **do**
    $pos_q \leftarrow$ BuscarÚltimaPosición($pos_p$, $x$)
    **if** $pos_q = -1$ **then**
      **break**
    **end**
    $p \leftarrow \text{primos}[pos_p]$
    $q \leftarrow \text{primos}[pos_q]$
    $suma \leftarrow suma + \ln p \cdot (\text{suma\_log}[q] - \text{suma\_log}[p-1])$
  **end**
  **return** *suma*
**end**

**Lema 4.3.** *Sea $x$ el número que representa la cota superior para el producto de dos números primos p y q en $\sum_{pq \leq x} \ln p \ln q$. Si obtenemos el valor de esta sumatoria mediante la ejecución de CalcularSuma, el tiempo de ejecución es $O(x)$.*

*Demostración.* Tenemos los números primos en el rango $[1 \ldots x]$ en el contenedor ordenado *primos*, entonces por definición $\pi(x) = |primos|$. Inicializar la suma en la línea 2 toma $O(1)$ en tiempo. El bucle en las líneas $3 - 11$ será ejecutado $O(\pi(x))$ veces. La función BuscarÚltimaPosición en la línea 4 es llamada exactamente una vez para cada primo $p \in primos$ y toma $O(\log_2 x)$ en tiempo de acuerdo con el lema 17, las demás operaciones dentro del bucle toman $O(1)$ en tiempo. De esta manera, el tiempo de ejecución del algoritmo es $O(\pi(x) \log_2 x)$. Asimismo, Rosser y Barkley [**?**, teorema 2 y corolario 1] probaron lo siguiente:

$$\pi(x) \leq 1.25506 \, \frac{x}{\ln x}. \tag{4.4}$$

Utilizamos esto para tener un mejor estimado cual

$$\begin{aligned} \pi(x) \log_2 x &\leq 1.25506 \, \frac{x}{\ln x} \log_2 x & (4.5)\\ &= 1.25506 \, \log_2 e \frac{x}{\log_2 e \cdot \log_e x} \log_2 x & (4.6)\\ &= 1.25506 \, \log_2 e \frac{x}{\log_2 x} \log_2 x & (4.7)\\ &= (1.25506 \, \log_2 e) x & (4.8)\\ & & (4.9) \end{aligned}$$

y concluimos que el tiempo de ejecución del algoritmo es $O(x)$. $\square$

**Algorithm 4:** EstimarConstante

**Data:** $x$.

**Result:** $\left|\frac{\sum_{p\leq x}\ln^2 p+\sum_{pq\leq x}\ln p\ln q-2x\ln x}{x}\right|$.

**begin**
    CribaLineal($x$)
    **for** $i \leftarrow 1$ **to** $x$ **do**
        suma_log[$i$] $\leftarrow 0$
        suma_log$^2$[$i$] $\leftarrow 0$
        **if** *es_primo*[$i$] **then**
            log[$i$] $\leftarrow \ln i$
        **else**
            log[$i$] $\leftarrow 0$
        **end**
    **end**
    **for** $i \leftarrow 2$ **to** $x$ **do**
        suma_log[$i$] $\leftarrow$ suma_log[$i-1$] + log[$i$]
        suma_log$^2$[$i$] $\leftarrow$ suma_log$^2$[$i-1$] + $(\log[i])^2$
    **end**
    **return** $\frac{|\text{suma\_log}^2[x]+\text{CalcularSuma}(x)-2x\ln x|}{x}$
**end**

**Teorema 4.4.** *Con $x$ el número natural para el cual se quiere determinar*

$$\frac{|\sum_{p\leq x}\ln^2 p+\sum_{pq\leq x}\ln p\ln q-2x\ln x|}{x}$$

*mediante la ejecución de EstimarConstante resulta que el tiempo de ejecución es $O(x)$.*

*Demostración.* □

## 4.2 Estimados

Sea $t$ la cantidad de casos de prueba y $n$ el máximo valor que puede tomar $x$ en la expresión que analizaremos. El siguiente programa en C++ determina el valor de

$$\frac{|\sum_{p\leq x}\ln^2 p+\sum_{pq\leq x}\ln p\ln q-2x\ln x|}{x}$$

para cada caso de prueba. En particular, usaremos $t = 28$ y $n = 3\cdot 10^7$. La complejidad asintótica en tiempo del programa es $O(tn)$, la cual se ve reflejada en un tiempo de ejecución de tan solo tres segundos.

```
real 0m3,123s
user 0m2,675s
sys 0m0,436s.
```

Tabla 4.1: Resultado del programa en cada uno de los 28 casos de prueba.

| $x$ | $\sum_{p\leq x}\ln^2 p+\sum_{pq\leq x}\ln p\ln q-2x\ln x$ | $\frac{\lvert\sum_{p\leq x}\ln^2 p+\sum_{pq\leq x}\ln p\ln q-2x\ln x\rvert}{x}$ |
|---|---:|---:|
| 10 | $-34.4229638465$ | 3.4422963847 |
| 100 | $-519.2686545658$ | 5.1926865457 |
| 1000 | $-6317.3110617078$ | 6.3173110617 |
| 10000 | $-74463.8721010727$ | 7.4463872101 |
| 100000 | $-859318.2559356594$ | 8.5931825594 |
| 1000000 | $-9747133.5703212193$ | 9.7471335703 |
| 1500000 | $-14918429.3651946987$ | 9.9456195768 |
| 2000000 | $-20177763.7803012519$ | 10.0888818902 |
| 2500000 | $-25505439.7770474999$ | 10.2021759108 |
| 3000000 | $-30875441.5873458827$ | 10.2918138624 |
| 3500000 | $-36291518.9878187147$ | 10.3690054251 |
| 4000000 | $-41746678.7374733434$ | 10.4366696844 |
| 4500000 | $-47221448.8244719136$ | 10.4936552943 |
| 5000000 | $-52725446.2863963772$ | 10.5450892573 |
| 5500000 | $-58263339.7575164592$ | 10.5933345014 |
| 6000000 | $-63830467.8915757891$ | 10.6384113153 |
| 6500000 | $-69414402.9962517989$ | 10.6791389225 |
| 7000000 | $-74998328.6200477667$ | 10.7140469457 |
| 7500000 | $-80619769.8168214446$ | 10.7493026422 |
| 8000000 | $-86252298.0080809856$ | 10.7815372510 |
| 8500000 | $-91907189.6865512790$ | 10.8126105514 |
| 9000000 | $-97554179.0078311940$ | 10.8393532231 |
| 9500000 | $-103244044.9516971762$ | 10.8677942054 |
| 10000000 | $-108942869.8283277971$ | 10.8942869828 |
| 15000000 | $-166436611.9307800680$ | 11.0957741287 |
| 20000000 | $-224778414.1259130784$ | 11.2389207063 |
| 25000000 | $-283766370.6499844969$ | 11.3506548260 |
| 30000000 | $-343248619.2823745428$ | 11.4416206427 |

# Bibliografía


[Apo76] Tom M. Apostol, *Introduction to analytic number theory*, Springer New York, 1976.

[CLRS09] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein, *Introduction to algorithms*, 3 ed., The MIT Press, MIT Press, London, England, July 2009.

[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.