Introducción a la Teoría de Números Probabílistica

Enrique Treviño

26 de marzo de 2020

1. Introducción

Considera la tabla de multiplicación de 10×10. Varios números entre 1 y 100 aparecen en la tabla, pero algunos se repiten. En la tabla hay 42 números distintos. ¿Que tal si consideramos una tabla de 100×100? Entonces tenemos 2906 números distintos. Y en la tabla de 1000×1000 tenemos 248083. Llamemos $N(n)$ al número de números distintos que aparecen en la tabla de multiplicación de $n \times n$. ¿Cuál es el límite de $\frac{N(n)}{n^2}$ cuando $n \to \infty$? Resulta que la respuesta es 0.

Teorema 1 (Teorema de la Tabla de Multiplicación)**.** *Sea* $N(n) = \#\{ij \mid 1 \leq i, j \leq n\}$. *Entonces*

$$\lim_{n\to\infty} \frac{N(n)}{n^2} = 0.$$

En este artículo daremos una introddución a la teoría de números probabílistica. Usaremos el problema de la tabla de multiplicación como motivación para aprender técnicas y teoremas de teoría de números analítica y de teoría de números probabílistica.

Uno de nuestros objetivos principales es demostrar un teorema clásico de Hardy y Ramanujan [6] usando técnicas probabílisticas basadas en ideas de Turán [11]. El teorema es el siguiente:

Teorema 2 (Hardy-Ramanujan)**.** *Sea* $\omega(n)$ *el número de divisores primos de* n. *Por ejemplo* $\omega(30) = 3$ *y* $\omega(8) = 1$. *Entonces casi todos los enteros satisfacen que* $\omega(n) \approx \log\log n$. *Para ser más precisos. Digamos que* $\varepsilon > 0$. *Consideremos la desigualdad*

$$|\omega(n) - \log\log n| < \varepsilon \log\log n. \tag{1}$$

Sea x *un número real positivo suficientemente grande. Consideremos el conjunto* $S(x)$ *de números* $n \leq x$ *que no satisfacen* (1), *entonces* $|S(x)| < \varepsilon x$.

Para poder demostrar este teorema, daremos un paseo por teoremas clásicos de teoría números analítica como el teorema de Mertens. Explicaremos algunas técnicas muy útiles que se han usado en el siglo veinte y terminaremos dando la demostración ingeniosa que hizo Erdős, en [3], para demostrar el Teorema de la Tabla Multiplicativa.

	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	12	14	16	18	20
3	3	6	9	12	15	18	21	24	27	30
4	4	8	12	16	20	24	28	32	36	40
5	5	10	15	20	25	30	35	40	45	50
6	6	12	18	24	30	36	42	48	54	60
7	7	14	21	28	35	42	49	56	63	70
8	8	16	24	32	40	48	56	64	72	80
9	9	18	27	36	45	54	63	72	81	90
10	10	20	30	40	50	60	70	80	90	100

Figura 1: Tabla de multiplicación de 10×10. Consiste de 42 números distintos.

2. Notación

En teoría de números analítica uno trabaja en estimar ciertas funciones. A diferencia de álgebra donde se acostumbra tener fórmulas exactas, en lo que estudiaremos aproximaremos funciones. Al aproximar, es importante tener una idea de que tan grande es el error. Para medir el tamaño de errores, tenemos dos tipos de notación muy comunes. El primero se llama notación "Big Oh" (o notación de Landau) y la segunda se llama notación de Vinogradov. La definición es la siguiente:

Definición 1 (Notación de Landau y de Vinogradov)**.** *Sean f y g dos funciones reales. Entonces decimos que $f(n) = \mathcal{O}(g(n))$ si existe un real positivo k tal que $|f(n)| \leq k|g(n)|$ para todo n. En lugar de $\mathcal{O}$, la notación de Vinogradov usa el símbolo $\ll$. Es decir, si $f(n) = \mathcal{O}(g(n))$ entonces podemos escribir $f(n) \ll g(n)$.*

En algunos casos en lugar de querer acotar con un múltiplo de $g(n)$, queremos demostrar que $f(n)$ es mucho más chica asintóticamente. Tenemos la siguiente definición para esos casos:

Definición 2. *Sean f y g funciones. Decimos que $f(n) = o(g(n))$ si*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

En el caso que el límite de la proporción es 1 (en lugar de 0), decimos que f es asintótico a g y escribimos $f(n) \sim g(n)$.

Algunos ejemplos para hacer la notación más clara.

Ejemplo 1. *Sabemos que $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$. Podemos reescribir esto como*

$$1 + 2 + \ldots + n = \frac{n^2}{2} + \mathcal{O}(n).$$

En tal notación estamos indicando que una buena aproximación es $\frac{n^2}{2}$. Esto sirve más cuando las fórmulas son más complicadas. Por ejemplo, la fórmula para $1^k + 2^k + 3^k + \ldots + n^k$ es muy complicada en términos de k (por ejemplo, véase [7]), pero podemos aproximarla muy bien. Primero, demos una cota superior usando integrales. Como x^k es creciente, tenemos

$$\sum_{j=1}^{n} j^k \leq n^k + \int_1^n t^k \,\mathrm{d}t = \frac{n^{k+1}}{k+1} - 1 + n^k = \frac{n^{k+1}}{k+1} + \mathcal{O}(n^k),$$

y

$$\sum_{j=1}^{n} j^k \geq 1 + \int_2^n t^k \,\mathrm{d}t = 1 + \frac{n^{k+1}}{k+1} - \frac{2^{k+1}}{k+1} = \frac{n^{k+1}}{k+1} + \mathcal{O}(1).$$

Entonces tenemos que la suma $1^k + 2^k + \ldots + n^k = \frac{n^{k+1}}{k+1} + \mathcal{O}(n^k)$. Es decir, la aproximamos por $\frac{n^{k+1}}{k+1}$ y tendríamos un error con orden de magnitud n^k. Si n es muy grande, el error es mucho más chico que el término principal.

Para el resto del artículo, las letras n y m se usaran sólo con enteros positivos y las letras p y q sólo para números primos.

Ejemplo 2. *La función $\omega(n)$ determina el número de divisores primos del número n. Con la convención de que p es primo, podemos escribir $\omega(n)$ de la siguiente manera:*

$$\omega(n) = \sum_{p|n} 1. \tag{2}$$

En el ejemplo anterior además de fijar la idea de que p es primo, podemos ver como lo usamos en una suma. Una suma se hace para todos los números que satisfagan cierta condición. En el caso de (2), la condición para ser un sumando es $p|n$.

Siguiendo la convención de teoría de números analítica, usaremos $\log(n)$ para el logaritmo natural de n. Es decir $\log(e^x) = x$. Otra función que usaremos es la función parte entera: $\lfloor x \rfloor$ es el entero menor o igual a x más cercano a x. Por ejemplo $\lfloor 3{,}1 \rfloor = 3$ y $\left\lfloor \frac{1}{7} \right\rfloor = 0$.

3. Resultados Preliminares

Un resultado que necesitaremos es una versión logarítmica de la fórmula de Stirling. La fórmula de Stirling dice que $n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$. Esta fórmula es un poco difícil de demostrar y no necesitamos algo tan fuerte para este artículo, pero podemos demostrar la siguiente versión más sencilla:

Lema 1 (Stirling Logarítmico)**.** *Sea n un entero positivo. Entonces*

$$\log(n!) = n\log n - n + \mathcal{O}(\log(n)). \tag{3}$$

Demostración. Primero notemos que $\log(n!) = \sum_{j\le n}\log(j)$. Como log es una función creciente entonces tenemos

$$\sum_{j\le n}\log(j) \le \int_1^n \log(t)\,\mathrm{d}t + \log(n) = n\log(n) - n + 1 + \log(n) = n\log(n) - n + \mathcal{O}(\log(n)).$$

Por otro lado, tenemos

$$\sum_{j\le n}\log(j) \ge 1 + \int_2^n \log(t)\,\mathrm{d}t = (n\log(n) - n) - (2\log(2) - 2) + 1 = n\log(n) - n + \mathcal{O}(1).$$

Entonces acotamos a $\log(n!)$ por arriba por $n\log(n) - n$ más un error de magnitud $\mathcal{O}(\log(n))$, y lo acotamos por abajo con $n\log(n) - n$ y un error de magnitud constante. Entonces $\log(n!)$ debe ser $n\log n - n$ con un error $\mathcal{O}(\log(n))$. $\square$

Una fórmula increíblemente útil que usaremos varias veces, y que es una de las herramientas más usadas en teoría de números analítica es la siguiente:

Teorema 3 (Fórmula de Sumación de Abel)**.** *Sea $\{a_n\}$ una secuencia y sea $f : \mathbb{R} \to \mathbb{R}$ una función diferenciable. Define $A(x)$ como*

$$A(x) = \sum_{n\le x} a_n = a_1 + a_2 + a_3 + \ldots + a_{\lfloor x\rfloor}.$$

La sumación de Abel o "Suma Parcial" es (para $b \ge a \ge 0$):

$$\sum_{a<n\le b} a_n f(n) = A(b)f(b) - A(a)f(a) - \int_a^b A(t)f^{'}(t)\,dt. \tag{4}$$

Demostración. Primero demostrémoslo para a y b enteros. Podemos suponer que $a \ge 0$. Dividamos $\int_a^b$ en pedazos:

$$\int_a^b A(t)f^{'}(t)\,\mathrm{d}t = \sum_{i=0}^{b-a-1}\int_{a+i}^{a+i+1} A(t)f'(t)\,\mathrm{d}t.$$

Como $A(t)$ es constante entre dos enteros, tenemos

$$\sum_{i=0}^{b-a-1}\int_{a+i}^{a+i+1} A(t)f'(t)\,\mathrm{d}t = \sum_{i=0}^{b-a-1} A(a+i)\int_{a+i}^{a+i+1} f'(t)\,\mathrm{d}t = \sum_{i=0}^{b-a-1} A(a+i)(f(a+i+1) - f(a+i)).$$

Pero ahora podemos cambiar el orden de la suma y cambiar las variables para obtener:

$$\begin{aligned}\sum_{i=0}^{b-a-1} A(a+i)(f(a+i+1) - f(a+i)) &= \sum_{j=a+1}^{b-1} f(j)(A(j-1) - A(j)) + A(b-1)f(b) - A(a)f(a)\\ &= A(b)f(b) - a_b f(b) - A(a)f(a) - \sum_{a+1}^{b-1} a_j f(j).\end{aligned}$$

Por lo tanto

$$\int_a^b A(t)f^{'}(t)\,\mathrm{d}t = A(b)f(b) - A(a)f(a) - \sum_{a<n\leq b} a_n f(n).$$

Queda demostrado para a y b enteros. Hagámoslo para a y b reales. Usando que lo demostramos para enteros tenemos

$$\sum_{a<n\leq b} a_n f(n) = \sum_{\lfloor a\rfloor<n\leq\lfloor b\rfloor} a_n f(n) = A(\lfloor b\rfloor)f(\lfloor b\rfloor) - A(\lfloor a\rfloor)f(\lfloor a\rfloor) - \int_{\lfloor a\rfloor}^{\lfloor b\rfloor} A(t)f^{'}(t)\,dt. \tag{5}$$

Queremos demostrar que el lado derecho de (5) es el mismo que el lado derecho de (4). Restemos las dos expresiones, tomando en cuenta que $A(x) = A(\lfloor x\rfloor)$,

$$A(b)(f(b) - f(\lfloor b\rfloor)) - A(a)(f(a) - f(\lfloor a\rfloor)) + \int_{\lfloor a\rfloor}^{\lfloor b\rfloor} A(t)f'(t)\,\mathrm{d}t - \int_a^b A(t)f'(t)\,\mathrm{d}t.$$

Esto nos da

$$A(b)(f(b) - f(\lfloor b\rfloor)) - A(a)(f(a) - f(\lfloor a\rfloor)) + \int_{\lfloor a\rfloor}^{a} A(t)f'(t)\,\mathrm{d}t - \int_{\lfloor b\rfloor}^b A(t)f'(t)\,\mathrm{d}t.$$

Pero como $A(t)$ es $A(\lfloor a\rfloor)$ para $t \in (\lfloor a\rfloor, a)$ y $A(t) = A(\lfloor b\rfloor)$ para $t \in (\lfloor b\rfloor, b)$, tenemos que esta diferencia es

$$A(b)(f(b) - f(\lfloor b\rfloor)) - A(a)(f(a) - f(\lfloor a\rfloor)) + A(\lfloor a\rfloor)\int_{\lfloor a\rfloor}^{a} f'(t)\,\mathrm{d}t - A(\lfloor b\rfloor)\int_{\lfloor b\rfloor}^b f'(t)\,\mathrm{d}t.$$

Usando que $\int_c^d f'(t)\,\mathrm{d}t = f(d) - f(c)$, terminamos demostrando que la diferencia es 0. Que es lo que queríamos demostrar. $\square$

4. Teorema de Mertens

El teorema de Mertens dice:

Teorema 4 (Mertens)**.** *Sea* $x \geq 3$. *Entonces*

$$\sum_{p\leq x} \frac{1}{p} = \log\log x + O(1).$$

El teorema nos da una muy buena aproximación de la suma de los recíprocos de los primos. Para poder demostrar esto, tendremos que demostrar algunos lemas primero. El primero, un lema demostrado por Erdős [2], nos da una cota superior para el producto de primos:

Lema 2.

$$\prod_{p\leq x} p \leq 4^x. \tag{6}$$

Demostración. Basta demostrarlo para x entero ya que $\prod_{p\leq x} p = \prod_{p\leq\lfloor x\rfloor} p$ y $4^{\lfloor x\rfloor} \leq 4^x$. Además podemos suponer que x es primo, ya que el producto entre dos primos es 1, pero 4^x sigue creciendo. Así que podemos suponer que x es primo. Para $x = 2$, el lado izquierdo de (6) es 2, mientras el lado derecho es $4^2 = 16$. Para $x = 3$, tenemos 6 contra 64. Así que podemos suponer que $x \geq 5$. Como x es primo mayor a 2, entonces $x = 2k+1$ para algún entero k. Ahora notemos que si $k+1 < p \leq 2k+1$, entonces $p|\binom{2k+1}{k}$. Esto es porque $p|(2k+1)!$, pero $p \not| k!$ y $p \not|(k+1)!$. Por lo tanto

$$\prod_{k+1<p\leq 2k+1} p \leq \binom{2k+1}{k+1}.$$

Ahora, el triángulo de Pascal nos dice que

$$\binom{2k+1}{0} + \binom{2k+1}{1} + \ldots + \binom{2k+1}{2k+1} = 2^{2k+1} = 2 \cdot 4^k.$$

Pero el lado izquierdo es mayor o igual a $\binom{2k+1}{k} + \binom{2k+1}{k+1} = 2\binom{2k+1}{k+1}$. Por lo tanto $\binom{2k+1}{k+1} \leq 4^k$. Usando inducción tenemos:

$$\prod_{p \leq x} p = \left(\prod_{p \leq k+1} p \right) \left(\prod_{k+1 < p \leq 2k+1} p \right) \leq 4^{k+1} \cdot 4^k = 4^{2k+1}.$$

Es lo que queríamos demostrar. $\square$

De este lema podemos probar el siguiente lema:

Lema 3. *Sea $\pi(x)$ el número de primos menores o iguales a x. Entonces $\pi(x) = \mathcal{O}\left(\frac{x}{\log(x)}\right)$.*

Demostración. Usando el Lema 2 y tomando logaritmos en (6) obtemos que

$$\sum_{p \leq x} \log(p) \leq \log(4)x. \tag{7}$$

Ahora, definamos la sucesión a_n definida por $\log(p)$ cuando $n = p$ es primo, y $a_n = 0$ cuando n no es primo. Definamos $f(n) = \frac{1}{\log(n)}$. Entonces, usando el Teorema de Abel (Teorema 3), obtenemos

$$\pi(x) = \sum_{p \leq x} 1 = \frac{\sum_{p \leq x} \log(p)}{\log(x)} + \int_2^x \frac{\sum_{p \leq t} \log(p)}{t \log^2(t)} \, \mathrm{d}t.$$

Usando que $\sum_{p \leq t} \log(p) \leq \log(4)t$, llegamos a que

$$\pi(x) \leq \mathcal{O}\left(\frac{x}{\log(x)}\right) + \log(4) \int_2^x \frac{1}{\log^2(t)} \, \mathrm{d}t. \tag{8}$$

Notemos que si $\sqrt{x} < t \leq x$, entonces $\log(t) \geq \frac{1}{2}\log(x)$. Por lo tanto

$$\begin{aligned} \int_2^x \frac{1}{\log^2(t)} \, \mathrm{d}t &\leq \int_2^{\sqrt{x}} \frac{1}{\log^2(t)} \, \mathrm{d}t + \int_{\sqrt{x}}^x \frac{4}{\log^2(x)} \, \mathrm{d}t \\ &\leq \sqrt{x} + \frac{4(x - \sqrt{x})}{\log^2 x} = \mathcal{O}\left(\frac{x}{\log^2(x)}\right). \end{aligned} \tag{9}$$

Usando (9) en (8) concluimos que $\pi(x) = \mathcal{O}\left(\frac{x}{\log(x)}\right)$. $\square$

El siguiente lema es casi Mertens:

Lema 4.

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + \mathcal{O}(1).$$

Demostración. Sea n un entero positivo. Factoricemos $n!$ como $n! = 2^{v_2} \cdot 3^{v_3} \cdot 5^{v_5} \cdots$. Entonces

$$\log(n!) = v_2 \log(2) + v_3 \log(3) + v_5 \log(5) + \cdots = \sum_{p \leq n} v_p \log(p). \tag{10}$$

Como $n!$ no tiene divisores primos mayores a n, por eso sólo necesitamos considerar los sumandos con $p \leq n$ (con $p > n, v_p = 0$). Ahora, es fácil calcular v_p. La idea es que el exponente más grande de p que divide a

$n!$ es el número de p's que aparecen en el producto $n(n-1)\cdots(1)$. Tenemos una p por cada múltiplo de p, luego otra por cada múltiplo de p^2 y así sucesivamente. Entonces

$$v_p = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor = \sum_{k=1}^{\log(n)/\log p} \left\lfloor \frac{n}{p^k} \right\rfloor. \tag{11}$$

La razón por la que cortamos la suma con el término $k = \log(n)/\log(p)$ es que si $k > \log(n)/\log(p)$ entonces $\frac{n}{p^k} < 1$ y, por lo tanto, $\left\lfloor \frac{n}{p^k} \right\rfloor = 0$. Combinando (10) con (11) llegamos a que

$$\log(n!) = \sum_{p \le n} \log(p) \sum_{k \le \frac{\log(n)}{\log(p)}} \left\lfloor \frac{n}{p^k} \right\rfloor. \tag{12}$$

Ahora, $\left\lfloor \frac{n}{p^k} \right\rfloor = \frac{n}{p^k} + \mathcal{O}(1)$. Usando el Lema 1 en (12), tenemos

$$n\log(n) - n + \mathcal{O}(\log(n)) = \sum_{p\le n} \log p \sum_{k \le \frac{\log(n)}{\log(p)}} \left(\frac{n}{p^k}\right) + \sum_{p \le n} \log(p) \sum_{k \le \frac{\log(n)}{\log(p)}} \mathcal{O}(1)$$

$$n\log(n) + \mathcal{O}(n) = n\sum_{p\le n} \frac{\log(p)}{p} + n\sum_{p \le n} \sum_{k=2}^{\frac{\log(n)}{\log(p)}} \frac{\log p}{p^k} + \sum_{p\le n} \mathcal{O}(\log(n)). \tag{13}$$

Tenemos que

$$\sum_{p\le n} \sum_{k=2}^{\frac{\log(n)}{\log(p)}} \frac{\log(p)}{p^k} \le \sum_{p} \sum_{k \ge 2} \frac{\log(p)}{p^k} = \sum_{p} \frac{\log(p)}{p(p-1)}. \tag{14}$$

Esta suma converge, así que es $\mathcal{O}(1)$. Usando el Lema 3 podemos ver que

$$\sum_{p \le n} \mathcal{O}(\log(n)) = \mathcal{O}(\pi(n)\log(n)) = \mathcal{O}(n). \tag{15}$$

Finalmente, aplicando (14) y (15) a (13) llegamos a

$$n\log(n) + \mathcal{O}(n) = n\sum_{p \le n} \frac{\log(p)}{p} + \mathcal{O}(n)$$

$$\log(n) + \mathcal{O}(1) = \sum_{p\le n} \frac{\log(p)}{p}.$$

Esto es lo que queríamos demostrar en el caso $x = n$, un número natural. Si x no es natural, tenemos

$$\sum_{p \le x} \frac{\log(p)}{p} = \sum_{p \le \lfloor x \rfloor} \frac{\log(p)}{p} = \log\log(\lfloor x \rfloor) + \mathcal{O}(1).$$

Como $\log\log(\lfloor x \rfloor) = \log\log(x) + \mathcal{O}(1)$, terminamos la demostración. $\square$

Finalmente estamos listos para demostrar Mertens:

Demostración del Teorema de Mertens. Usaremos el Teorema de Abel y el Lema 4:

$$\begin{aligned}
\sum_{p \le x} \frac{1}{p} = \sum_{p\le x} \frac{\log(p)}{p}\left(\frac{1}{\log(p)}\right) &= \frac{\sum_{p \le x} \frac{\log(p)}{p}}{\log(x)} + \int_2^x \frac{\sum_{p\le t} \frac{\log(p)}{p}}{t\log^2(t)}\,\mathrm{d}t \\
&= \frac{\log(x) + \mathcal{O}(1)}{\log(x)} + \int_2^x \frac{1}{t\log(t)}\,\mathrm{d}t + \mathcal{O}\left(\int_2^x \frac{1}{t\log^2(t)}\,\mathrm{d}t\right) \\
&= 1 + \mathcal{O}\left(\frac{1}{\log(x)}\right) + \log\log(x) - \log\log(2) + \mathcal{O}\left(\frac{1}{\log(x)}\right) \\
&= \log\log(x) + \mathcal{O}(1).
\end{aligned}$$

$\square$

4.1. Notas Históricas

Los resultados demostrados en este artículo han sido mejorados. El teorema de Mertens se puede escribir como $\sum_{p\le x}\frac{1}{p}=\log\log(x)+B_1+\mathcal{O}\left(\frac{1}{\log(x)}\right)$, donde B_1 es una constante (véase [9] para una demostración). Cabe notar, que el teorema de Mertens implica que hay infinitos números primos, ya que $\log\log(x)$ crece indefinidamente. La demostración de que cualesquiera progresión aritmética $a, a+b, a+2b, \ldots$, con a y b primos relativos, contiene infinitos primos (llamado el Teorema de Dirichlet) prueba algo más fuerte, prueba que la suma de los recíprocos de tales primos diverge (es decir, se va al infinito). Es sorprendente que no hay una demostración general que no pruebe ese resultado más fuerte.[1] Una de las razones por las que es difícil demostrar que hay infinitos primos gemelos (primos p tales que $p+2$ ó $p-2$ es primo) es que Brun (en [1]) demostró que la suma de los recíprocos de los primos gemelos converge.

El Lema 3 fue demostrado por primera vez por Chebyshev en 1848 con el objetivo de demostrar el postulado de Bertrand.[2] En 1859, Riemann en uno de los artículos más famosos de la historia de las matemáticas ([10]), explicó una estrategia para demostrar el teorema de los números primos que dice que $\pi(x)\sim\frac{x}{\log(x)}$. El teorema fue demostrado en 1896 por Hadamard y Vallée-Poussin.

Ejercicio 1. *Demuestra que existe una constante B_1 tal que*

$$\sum_{p\le x}\frac{1}{p}=\log\log(x)+B_1+\mathcal{O}\left(\frac{1}{\log(x)}\right).$$

De hecho, podemos calcular que

$$B_1=1-\log\log(2)+\int_2^\infty\frac{\sum_{p\le t}\frac{\log(p)}{p}-\log(t)}{t\log^2(t)}\,\mathrm{d}t.$$

Ejercicio 2 (Segundo Teorema de Mertens)**.** *Demuestra que existe una constante B_2 tal que*

$$\prod_{p\le x}\left(1-\frac{1}{p}\right)=\frac{e^{-B_1-B_2}}{\log(x)}+\mathcal{O}\left(\frac{1}{\log^2(x)}\right).$$

Podemos calcular que

$$B_2=\sum_p\sum_{k\ge 2}\frac{1}{k\cdot p^k}.$$

Ejercicio 3. *Demuestra que B_1+B_2 de los ejercicios anteriores es γ, la constante Euler-Mascheroni:*

$$\gamma=\lim_{n\to\infty}\left(\sum_{k\le n}\frac{1}{k}-\log(n)\right).$$

5. El promedio y la varianza de $\omega(n)$

El promedio de una función aritmética $f(n)$ es $\mu_f(x)=\frac{1}{x}\sum_{n\le x}f(n)$ y la varianza es $\sigma_f(x)=\frac{1}{x}\sum_{n\le x}\left(f(n)-\mu_f(x)\right)^2$.

Teorema 5. *El promedio de $\omega(n)$ es $\sim\log\log(x)$. Para ser más específicos:*

$$\frac{1}{x}\sum_{n\le x}\omega(n)=\log\log(x)+\mathcal{O}(1). \tag{16}$$

[1] La demostración clásica de que hay infinitos primos no implica que la suma de los recíprocos de los primos diverge.

[2] Para todo entero positivo n, existe un primo p que satisface $n<p\le 2n$.

Demostración. Empecemos escribiendo $\omega(n)$ como suma:

$$\sum_{n\leq x}\omega(n)=\sum_{n\leq x}\sum_{p|n}1.$$

Ahora, cambiaremos el orden de la suma. En lugar de fijar n y ver cuantos primos p dividen a n, lo podemos pensar como fijar un primo p y ver cuantos n son múltiplos de p, es decir $n=pm$ para alguna m. Tomando en cuenta que $\lfloor x\rfloor = x+\mathcal{O}(1)$, obtenemos

$$\begin{aligned}\sum_{n\leq x}\sum_{p|n}1=\sum_{p\leq x}\sum_{pm\leq x}1=\sum_{p\leq x}\sum_{m\leq \frac{x}{p}}1=\sum_{p\leq x}\left\lfloor\frac{x}{p}\right\rfloor&=\sum_{p\leq x}\frac{x}{p}+\sum_{p\leq x}\mathcal{O}(1)\\&=x\sum_{p\leq x}\frac{1}{p}+\mathcal{O}(\pi(x))=x\log\log(x)+\mathcal{O}(x)+\mathcal{O}\left(\frac{x}{\log(x)}\right)\\&=x\log\log(x)+\mathcal{O}(x).\end{aligned}$$

En las ecuaciones usamos el Teorema de Mertens y el Lema 3. Para concluir, basta dividir por x. $\square$

Teorema 6. *La varianza de $\omega(n)$ es $\mathcal{O}(\log\log(x))$. Para ser más específicos:*

$$\frac{1}{x}\sum_{n\leq x}(\omega(n)-\log\log x)^2=\mathcal{O}(\log\log(x)). \tag{17}$$

Demostración. Dividamos el problema en tres partes, separando la suma en tres sumas:

$$\sum_{n\leq x}(\omega(n)-\log\log(x))^2=\sum_{n\leq x}\omega^2(n)-2\log\log(x)\sum_{n\leq x}\omega(n)+(\log\log(x))^2\sum_{n\leq x}1=S_1-S_2+S_3, \tag{18}$$

donde $S_1=\sum_{n\leq x}\omega^2(n)$, $S_2=2\log\log(x)\sum_{n\leq x}\omega(n)$, y $S_3=(\log\log(x))^2\sum_{n\leq x}1$. Por el Teorema 5, tenemos

$$S_2=2\log\log(x)(x\log\log(x)+\mathcal{O}(x))=2x(\log\log(x))^2+\mathcal{O}(x\log\log(x)). \tag{19}$$

S_3 es fácil de calcular, ya que es simplemente $S_3=\lfloor x\rfloor(\log\log(x))^2$. Nos falta calcular S_1.

$$S_1=\sum_{n\leq x}\omega^2(n)=\sum_{n\leq x}\left(\sum_{p|n}1\right)\left(\sum_{q|n}1\right)=\sum_{n\leq x}\sum_{\substack{p|n\\q|n\\p\neq q}}1+\sum_{n\leq x}\sum_{p|n}1.$$

Esta última igualdad se obtiene separando en los casos donde tenemos primos iguales y en los que no. En la primera suma, cambiamos el orden de los sumandos, tomando en cuenta que podemos agarrar $p\leq x$, $q\leq x$ con $p\neq q$ y $pq|n$. En la segunda suma podemos usar el Teorema 5. Entonces tenemos:

$$\begin{aligned}S_1&=\sum_{p\leq x}\sum_{\substack{q\leq x\\q\neq p}}\sum_{n\leq\frac{x}{pq}}1+\sum_{n\leq x}\sum_{p|n}1\\&=\sum_{p\leq x}\sum_{\substack{q\leq x\\q\neq p}}\left\lfloor\frac{x}{pq}\right\rfloor+x\log\log(x)+\mathcal{O}(x)\\&=\sum_{p\leq x}\sum_{\substack{q\leq x\\q\neq p}}\frac{x}{pq}+\sum_{pq\leq x}\mathcal{O}(1)+x\log\log(x)+\mathcal{O}(x).\end{aligned}$$

Para obtener la última igualdad, usamos que $\left\lfloor\frac{x}{pq}\right\rfloor$ es 0 si $pq>x$. Entonces el error sucede sólo en los casos $pq\leq x$. Queremos contar todas las parejas (p,q) tales que $pq\leq x$. Pero entonces queremos los números menores a x que son producto de dos primos. Por cada uno de estos números, tenemos dos parejas (por el orden de p y q). Entonces, hay a los más $2x$ parejas. Por lo tanto $\sum_{pq\leq x}\mathcal{O}(1)=\mathcal{O}(x)$. Entonces tenemos

$$S_1=\sum_{p\leq x}\sum_{\substack{q\leq x\\ q\neq p}}\frac{x}{pq}+x\log\log(x)+\mathcal{O}(x)=x\sum_{p\leq x}\frac{1}{p}\sum_{\substack{q\leq x\\ q\neq p}}\frac{1}{q}+\mathcal{O}(x\log\log(x)) \tag{20}$$

$$=x\sum_{p\leq x}\frac{\log\log x+\mathcal{O}(1)}{p}+\mathcal{O}(x\log\log(x))=x\log\log(x)\sum_{p\leq x}\frac{1}{p}+\mathcal{O}(x\log\log(x)) \tag{21}$$

$$=x(\log\log(x))^2+\mathcal{O}(x\log\log(x)+\mathcal{O}(x\log\log(x))=x(\log\log(x))^2+\mathcal{O}(x\log\log(x). \tag{22}$$

Usamos el Teorema de Mertens varias veces para llegar a esa conclusión. Ahora, combinando (19) con (20) y el valor de S_3 tenemos:

$$\begin{aligned}S_1-S_2+S_3&=\left(x(\log\log(x))^2+\mathcal{O}(x\log\log(x)\right)-\left(2x(\log\log(x))^2+\mathcal{O}(x\log\log(x))\right)+\left(x(\log\log(x))^2\right)\\&=\mathcal{O}(x\log\log(x)).\end{aligned}$$

Al dividir por x, tenemos lo que queríamos demostrar. $\square$

Comentario 1. *Se puede demostrar que la varianza de $\omega(n)$ es asimptótica a $\log\log(n)$, pero requiere de más trabajo y de tener la versión de Mertens mencionada en el Ejercicio 1.*

Ejercicio 4. *Demuestra que*

$$\frac{1}{x}\sum_{n\leq x}\omega(n)=\log\log(x)+B_1+\mathcal{O}\left(\frac{1}{\log(x)}\right),$$

donde B_1 es la misma constante que en el Ejercicio 1.

Ejercicio 5. *Demuestra que*

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

Ejercicio 6. *Demuestra*

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

6. Casi todo n satisface que $\omega(n)\approx\log\log(n)$

Estamos listos para demostrar nuestro teorema principal.

Demostración del Teorema 2. Sea $\varepsilon>0$. Sea $m_\varepsilon(x)$ el número de enteros n que satisfacen $\sqrt{x}\leq n\leq x$ y que $|\omega(n)-\log\log(n)|\geq\varepsilon\log\log n$. Como $n\geq\sqrt{x}$, entonces

$$\log\log(n)\geq\log\log(\sqrt{x})=\log(1/2)+\log\log x=\log\log(x)-\log(2).$$

Por la desigualdad del triángulo

$$\begin{aligned}|\omega(n)-\log\log(x)|\geq||\omega(n)-\log\log(n)|-(\log\log(x)-\log\log(n))|&\geq|\,\varepsilon(\log\log(x)-\log(2))-\log(2)|\\&\geq\frac{\varepsilon}{2}\log\log x.\end{aligned}$$

La última desigualdad es cierta para x suficientemente grande, en particular para x que satisface $\log\log(x)>\log(2)\left(2+\frac{2}{\varepsilon}\right)$. Por lo tanto, si n es uno de los enteros que contamos en $m_\varepsilon(x)$, entonces $|\omega(n)-\log\log(x)|\geq\frac{\varepsilon}{2}\log\log(x)$. Pero esto implica que

$$\sum_{n\leq x}(\omega(n)-\log\log(x))^2\geq m_\varepsilon(x)\frac{\varepsilon^2}{4}(\log\log(x))^2. \tag{23}$$

Por otro lado, el lado izquierdo de (23) es $\mathcal{O}(x \log\log(x))$. Eso implica que

$$m_\varepsilon(x) = \mathcal{O}\left(\frac{x}{\varepsilon^2 \log\log x}\right).$$

Para x suficientemente grande, $m_\varepsilon(x) \leq \frac{\varepsilon}{2}x$ porque $\log\log(x) \to \infty$ cuando $x \to \infty$. Ahora los n que pueden ser excepción a (1), son los que están en $m_\varepsilon(x)$ o los números $n \leq \sqrt{x}$. Como $\sqrt{x} < \frac{\varepsilon}{2}x$ para x grande, entonces el número de cuates que no satisfacen (1) es menor a $\varepsilon\, x$. $\square$

Comentario 2. *La demostración se ve un poco fea por la parte técnica de trabajar con ε. Pero la filosofía detrás de la solución no es tan complicada. La idea es pensar como si fuera probabilidad. Tenemos una distribución ω cuyo promedio es $\log\log n$ ($\log\log n$ y $\log\log x$ son prácticamente lo mismo porque $\log$ crece muy lentamente) y cuya varianza es $\log\log n$. Entonces, por el teorema de Chebyshev*[3] *de probabilidad, tenemos que la probabilidad de que $|\omega(n) - \log\log n| \geq \varepsilon \log\log n$ es menor o igual a $\frac{K^2}{\varepsilon^2 \log\log(n)}$ para alguna constante K. Es decir se va a cero cuando $n \to \infty$.*

6.1. Erdős–Kac

Mencioné antes que la varianza de $\omega(n)$ se puede calcular como $\log\log(n)$ más un error pequeño, pero demostrar eso conlleva más trabajo. En particular, se pueden calcular varios momentos de la distribución y ver que siguen una distribución normal con promedio $\log\log n$ y varianza $\log\log n$ (véase [5], por ejemplo). Esta propiedad la conjeturó Mark Kac en una plática en los años treintas. Para fortuna de las matemáticas, Paul Erdős estaba en la plática y cuando Kac terminó la plática, Erdős le mencionó una idea para demostrar la propiedad. Este resultado se llama el Teorema Erdős–Kac y es donde nació la teoría de números probabílistica.

Teorema 7 (Erdős–Kac)**.** *Sean a y b reales tales que $a \leq b$. Sea $S_x(a,b)$ el número de $n \leq x$ que satisface*

$$\log\log(n) + a\sqrt{\log\log(n)} \leq \omega(n) \leq \log\log(n) + b\sqrt{\log\log(n)}.$$

Entonces

$$\lim_{x\to\infty} \frac{S_x(a,b)}{x} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, \mathrm{d}t.$$

Ejercicio 7. *Sea $\Omega(n)$ el número de divisores primos con multiplicidad de n. Por ejempo $\Omega(8) = 3$ porque el 2 aparece 3 veces, $\Omega(24) = 4$ porque $24 = 2^3 \cdot 3^1$. Demuestra que el Teorema 2 es cierto con Ω reemplazando a ω.*

7. Conclusión

Después de nuestro paseo por algunos resultados importantes de la teoría de números analítica, estamos listos para dar un bosquejo de la demostración del Teorema de la Tabla Multiplicativa de Erdős. No daremos la solución completa, para invitar al lector a usar lo que ha aprendido para completar los argumentos.

Bosquejo de Demostración del Teorema 1. Casi todos los números $j \leq n$ cumplen que $\Omega(j) \approx \log\log n$. Lo mismo para los números $i \leq n$. Entonces, casi todos los productos ij cumplen que

$$\Omega(ij) = \Omega(i) + \Omega(j) \approx 2\log\log(n).$$

Pero entonces, los números que forman parte de $N(n)$ son "raros". Para ser más específicos. Casi todos los números menores o iguales a n^2 cumplen que

$$\Omega(n) \approx \log\log(n^2) = \log(2\log(n) = \log(2) + \log\log(n) \approx \log\log(n).$$

[3] El teorema de Chebyshev dice que si μ es el promedio de una distribución X y σ^2 es la varianza, entonces la probabilidad de que $|X - \mu| \geq k\sigma^2$ es a lo más $\frac{1}{k^2}$.

Como casi todos tienen $\log\log(n)$ divisores primos (con multiplicidad), y los que están en la tabla de multiplicación tienen $2\log\log(n)$, entonces son muy raros, es decir,

$$\lim_{n\to}\frac{N(n)}{n^2}=0.$$

$\square$

Comentario 3. *Erdős demostró que* $N(n)=\mathcal{O}\left(\frac{n^2}{(\log(n))^{1+\varepsilon}}\right)$*. Luego Ford en [4] describió* ε *con más precisión calculando* $N(n)=\mathcal{O}\left(\frac{n^2}{(\log(n))^K(\log\log(n))^{3/2}}\right)$*, donde* K *es descrito explícitamente. Koukoulopoulos en [8] generalizó el problema a tablas multiplicativas en más dimensiones.*

Ejercicio 8. *Completa la demostración del Teorema de la Tabla de Multiplicación.*

Referencias

[1] Viggo Brun, *La série* $\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\frac{1}{19}+\frac{1}{29}+\frac{1}{31}+\frac{1}{41}+\frac{1}{43}+\frac{1}{59}+\frac{1}{61}+\ldots$ *où les dénominateurs sont "nombres premiers jumeaux" est convergente ou finite*, Bull. Sci. Math **43** (1919), 100–104, 124–128.

[2] Paul Erdős, *Beweis eines satzes von tschebyschef*, Acta Litt. Sci. Szeged **5** (1932), 194–198.

[3] Paul Erdös, *Some remarks on number theory*, Riveon Lematematika **9** (1955), 45–48. MR 0073619

[4] Kevin Ford, *The distribution of integers with a divisor in a given interval*, Ann. of Math. (2) **168** (2008), no. 2, 367–433. MR 2434882

[5] Andrew Granville and K. Soundararajan, *Sieving and the Erdos-Kac theorem*, Equidistribution in number theory, an introduction, NATO Sci. Ser. II Math. Phys. Chem., vol. 237, Springer, Dordrecht, 2007, pp. 15–27. MR 2290492

[6] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n [Quart. J. Math.* **48** *(1917), 76–92]*, Collected papers of Srinivasa Ramanujan, AMS Chelsea Publ., Providence, RI, 2000, pp. 262–275. MR 2280878

[7] Donald E. Knuth, *Johann Faulhaber and sums of powers*, Math. Comp. **61** (1993), no. 203, 277–294. MR 1197512

[8] Dimitris Koukoulopoulos, *On the number of integers in a generalized multiplication table*, J. Reine Angew. Math. **689** (2014), 33–99. MR 3187928

[9] Paul Pollack, *Not always buried deep*, American Mathematical Society, Providence, RI, 2009, A second course in elementary number theory. MR 2555430

[10] Bernhard Riemann, *Ueber die anzahl der primzahlen unter einer gegebenen grösse*, Monatsberichte der Berliner Akademie (1859), 671–680.

[11] Paul Turán, *On a Theorem of Hardy and Ramanujan*, J. London Math. Soc. **S1-9** (1934), no. 4, 274. MR 1574877