Probabilidades y distribución de los números primos

Francisco José Valladares Herrera

Probabilidades y distribución de los números primos

Francisco José Valladares Herrera

Trabajo de fin de grado que forma parte de los requisitos para la obtención del título de Grado en Matemáticas por la Universidad de Sevilla.

Dirigido por

Juan Arias de Reyna Martínez

A mi madre, siempre fuerte.

Summary

In this project, we will try to study the prime factors of a number taken at random. The main objective will be to introduce the probabilistic theory of numbers and, at the same time, several central themes of probability theory in general: expected value, variance and the central limit.

The story of probabilistic number theory begins with Hardy and Ramanujan, who were the ones who pushed the study of the central theme of this project: the distribution of $\omega(n)$, the number of prime divisors of a natural number n. Later generations, the area continued to develop remarkably with Erdös and Kac's theorem (which is the main theme we will try to demonstrate and understand) until this day, thanks to both number theory specialists and probability experts.

We will not assume any knowledge of complex analysis or measure theory. We only need to consider finite spaces of probability, so that we will start by studying expected value by treating ω as a random variable, using Chebyshev's inequality. Subsequently we will study the variance, and aply it in two ways: one using the inequality of Chebyshev, and another somewhat more useful (that is, it will give better theoretical results). Finally we will apply the central limit theorem, in order to try to understand and demonstrate the aforementioned Erdös - Kac theorem.

INTRODUCCIÓN

Los números primos no constituyen un concepto complicado que requiera años de estudios matemáticos; de hecho, se enseñan en los colegios, en los primeros cursos de matemáticas. Para saber lo que es un número primo basta con estar familiarizado con los primeros números naturales y las cuatro operaciones fundamentales. Sin embargo, han sido y siguen siendo uno de los retos más fabulosos de la historia de la ciencia. Su influencia no sólo está presente en el universo particular de las matemáticas, sino que, aunque no seamos conscientes de ello, los números primos desempeñan un papel decisivo en nuestra vida cotidiana: en la protección que requiere nuestro ordenador personal, en las transacciones bancarias o en la privacidad de nuestras conversaciones telefónicas, ya que son las piedras angulares de la seguridad informática.

En un sentido metafórico, los números primos son como un virus maléfico que, cuando ataca la mente de un matemático, es muy difícil de erradicar. Euclides, Fermat, Euler, Gauss, Riemann, Ramanujan y una larga lista de los matemáticos de mas renombre de la historia cayeron en sus redes. Algunos consiguieron zafarse de él de manera más o menos exitosa, tratando de entender las complejidades de su distribución y sorprendentes propiedades.

En este proyecto trataremos de dar unas nociones básicas de la teoría probabilística de números, centrándonos en el estudio de la distribución de los divisores primos de un número, y no en el estudio en sí de la distribución de los primos. La razón principal es que realmente se sabe poco con certeza acerca de la distribución de los números primos. Por ejemplo, se conjetura la probabilidad que una pareja de enteros $(n, n+2)$ esté formada por primos. En ocasiones como mucho, lo único que se posee es cotas superiores dadas por la teoría de cribas. Existen modelos tanto fructíferos como imperfectos - por ejemplo, el modelo de Cramér, que dice que el evento que un número n sea primo y el evento que un número m distinto sea primo se comportan muchas veces como si fueran eventos independientes. Está claro que

esto no debe ser creído completamente: por ejemplo, si $m = n + 1$ y $n > 2$, los números n y m no pueden ser ambos primos, lo cual seria una posibilidad si estuviéramos hablando de variables independientes.

A pesar de los muchos avances, la mayor parte de las cuestiones sobre los números primos quedan pendientes. Se podría decir que a lo largo de la historia, su estudio ha generado nuevas teorías, nuevos paradigmas, nuevos hitos que han marcado un antes y un después. En la actualidad, para el estudio de los números primos, se usa constantemente la teoría de la medida y el análisis complejo de una forma avanzada; de la misma manera cabría destacar el uso indispensable de la teoría de probabilidades. Fueron Ramanujan y Hardy los que realmente impulsaron esta forma de estudiar la distribución de los números primos. Sus trabajos ocupan actualmente a cientos de matemáticos en los departamentos de las universidades, y sus resultados se aplican en áreas tan dispares como la química de los polímeros, la arquitectura de los ordenadores o la investigación del cáncer.

De alguna forma, y aunque sonara paradójico decirlo, es una suerte que todavía no se hayan dejado dominar, pues eso quiere decir que a la comunidad matemática le queda mucho por avanzar y mucho más por descubrir.

Índice general

1 Resumen y objetivos

Sea $\omega(n)$ el número de divisores primos de un entero n. Tomaremos al azar un número entero n en un intervalo $(1,\ x)$ y nos planteamos preguntas como: ¿qué podríamos decir del valor que tomará $\omega(n)$?, ¿podríamos hablar de su esperanza probabilística, y cual sería?, ¿cuál es la probabilidad de que $\omega(n)$ tome valores que se alejen mucho de su esperanza? Estas preguntas irán enfocadas para al final entender y demostrar el *teorema de Erdös-Kac*, el cual relaciona de manera directa $\omega(n)$ con la distribución normal a pesar de no cumplirse las tres condiciones del famoso *teorema del límite central*.

Estudiaremos estas cuestiones como introducción a la *teoria de números probabilística*. Trataremos varios tópicos centrales de la teoria de probabilidades sin suponer demasiados conocimientos previos en el área. Dejaremos a un lado tanto la teoría de la medida como el análisis complejo. Entre los resultados que iremos reflejando, se desarrollarán las bases de la teoría de cribas como una aplicación de ideas probabilísticas.

2 | Notación y definiciones previas

A lo largo del proyecto, la letra p siempre designará a un número primo. Dados dos números enteros d y n, escribiremos que $d \mid n$ si d divide a n y diremos que $d \nmid n$ en caso contrario. Denotaremos $\lfloor x \rfloor$ el máximo entero n que no sea mayor que x. Cabe especificar también, que cuando decimos "logaritmo" o escribimos $\log x$, tenemos siempre en mente al logaritmo en base e, a menos que se especifique otra base explícitamente.

Utilizaremos la notación O y o de Landau de la siguiente manera: Dadas dos funciones f, g se escribe $f(x) = O(g(x))$ si existen unas constantes c_1, $c_2 > 0$ tales que $|f(x)/g(x)| < c_1$ para todo $x > c_2$, y se escribe $f(x) = o(g(x))$ cuando $\lim_{x\to\infty} f(x)/g(x) = 0$. Esta claro que $f(x) = o(g(x))$ implica $f(x) = O(g(x))$, pero no viceversa. En particular, $f(x) = O(1)$ quiere decir que f esta acotada por una constante, y $f(x) = o(1)$ quiere decir que f tiende a cero cuando x va al infinito. Escribimos $O_c(1)$,$o_{\delta,z}(1)$ si dichas constantes dependen de una c o δ y z concretas. A veces denotaremos por la misma letra C a diferentes constantes, incluso en la misma ecuación. Hacemos en esos casos un abuso de notación llamándolas a todas simbólicamente como C.

La expresión "$f(x) \ll g(x)$" es un sinónimo de "$f(x) = O(g(x))$" y la expresión "$f(x) \gg g(x)$" es un sinónimo de "$g(x) = O(f(x))$". Escribiremos además, $f(x) \sim g(x)$ cuando queremos decir que f es asintotica con respecto a g, i.e., $\lim_{x\to\infty} f(x)/g(x) = 1$.

Denotamos Prob(U) la probabilidad del evento aleatorio U, por $\mathbb{E}(X)$ la esperanza de la variable aleatoria X y por Var(X) la varianza de la variable X. Por último, definiremos también algunas funciones aritméticas:

- La función aritmética $\tau(d)$ representa la cantidad de divisores que tiene un cier-

to entero n:

$$\tau(n) = \sum_{d|n} 1$$

- La función aritmética $\omega(n)$ representa la cantidad de divisores primos que tiene un cierto entero n :

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

Observación 2.1. Cabe aclarar que por cada primo distinto se suma una vez por ejemplo: El número 12 tiene los tres divisores primos $\{2, 2, 3\}$ pero $\omega(12)$ = 2.

Observación 2.2. Aclaremos también que dado un número real x las sumas del tipo $\sum_{n\leq x}$ se refieren a las sumas $\sum_{n=1}^{N}$ con $N = \lfloor x \rfloor$. Ésto será conveniente sobre todo a la hora de acotar y usar la *fórmula de sumación de Abel.*

Observación 2.3. Observemos que $\sum_{p\leq x}$ se refiere a la suma de los términos primos p menores que x, por ejemplo: $\sum_{p\leq 10} \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7}$.

- La función $\Lambda(n)$ (*función de Von Mangoldt*) :

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^{\alpha} \text{ para algún primo } p \text{ y algún entero } \alpha > 0 \\ 0 & \text{en caso contrario} \end{cases}$$

- La función $X_m(n)$:

$$X_m(n) = \begin{cases} 1 & \text{si } m \mid n \\ 0 & \text{si } m \nmid n \end{cases} \tag{2.1}$$

- La *función de Möbius* $\mu(n)$:

$$\mu(n) = \begin{cases} 0 & \text{si } n \text{ es divisible por algún cuadrado} \\ (-1)^{\omega(n)} & \text{en caso contrario} \end{cases} \tag{2.2}$$

- La *función de Euler* $\varphi(n)$: denota el cardinal del conjunto de números primos con n y menores que n, es decir:

$$\varphi(n) = card\{k \perp n : 1 \leq k \leq n\}.$$

- La *función zeta de Riemann* $\zeta(s)$:

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

3 | Abel, Mertens y la esperanza

3.1 Nuestro espacio de probabilidad

Consideraremos un número real $x > 1$ y el espacio $\Omega_x = \{n \in \mathbb{N} : n \leq x\}$ es decir, el espacio formado por los números naturales $1, 2, \ldots, \lfloor x \rfloor = N$.

Dotaremos al espacio Ω_x de una probabilidad, la que asigna la misma probabilidad a cada punto del espacio, precisamente $\frac{1}{N}$. Podríamos poner un ejemplo:

Ejemplo 3.1. Si tenemos $\Omega_{5.23}$ entonces la probabilidad de que un número natural $n \in \Omega_{5.23}$ y sea impar es $\frac{3}{5}$.

Describiremos ahora la famosa *fórmula de sumación de Abel*, la cual usaremos muy a menudo en el trabajo (aquí encontramos la razón de usar la variable continua x en lugar de la discreta N).

3.2 Lema de Abel

Sea (a_n) una sucesión, estamos interesados en una suma del tipo $\sum_{n=1}^{N} a_n$ o incluso $\sum_{n=1}^{N} a_n f(n)$ donde f es una función con derivada continua. En estos casos es muy útil el lema de Abel.

Lema 3.1 (Lema de Abel). Sea un $x > 0$ real . Definiendo $S(x) = \sum_{n \leq x} a_n$ tenemos que

$$\sum_{n=1}^{N} a_n f(n) = S(x)f(x) - \int_1^x S(t)f'(t)dt \tag{3.1}$$

Observación 3.1. Notemos que en la integral aparece $S(t)$ siendo t un número real. es por esto que aun cuando nos interese estudiar $\sum_{n\leq x} a_n$ solo para $x \in \mathbb{N}$ debemos considerarlos para todo x real.

3.3 El concepto de Esperanza

A lo largo de esta sección, queremos dejar claro el concepto de esperanza probabilística, así como su íntima relación con las sumas frecuentes en el área del análisis matemático.

Definición 3.1 (Esperanza). *Dada una variable aleatoria X que toma los valores $x_1, x_2, \ldots, x_n$ con ciertas probabilidades $p_1, p_2, \ldots, p_n$, se define la esperanza de dicha variable, $\mathbb{E}(X)$, como la cantidad*

$$\sum_{i=1}^{n} p_i x_i.$$

Ejemplo 3.2. Podríamos poner el ejemplo clásico de un dado, si llamamos a X como la variable aleatoria que toma los valores:

$$X = \begin{cases} 1 & \text{con probabilidad } 1/6 \\ 2 & \text{con probabilidad } 1/6 \\ 3 & \text{con probabilidad } 1/6 \\ 4 & \text{con probabilidad } 1/6 \\ 5 & \text{con probabilidad } 1/6 \\ 6 & \text{con probabilidad } 1/6 \end{cases}$$

es fácil calcular su esperanza probabilísitca, que sería:

$$\mathbb{E}(X) = \sum_{i=1}^{n} p_i x_i = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \cdots + \frac{1}{6}\cdot 6 = \frac{7}{2}.$$

Una propiedad básica es que dadas $X_1, X_2, \ldots, X_n$ variables aleatorias, se cumple que $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$.

3.4 Promedio de $\tau(n)$

Nuestro objetivo es conocer los valores esperados que tomará la función aritmética $\omega(n)$ pero antes, estudiaremos los de $\tau(n)$ pues es un poco más sencillo de averiguar

y el método será idéntico.

Para que tenga sentido preguntarnos por el promedio de $\tau(n)$ fijamos un cierto x, es decir, tomaremos un entero n al azar en el intervalo $(1, x)$.

Usaremos la función X_m que definimos en (2.1) y la tomaremos formalmente como una variable aleatoria discreta. Así pués, por definición:

$$\tau = \sum_{d|n} 1 = \sum_{d|n} X_d(n) = \sum_{m\leq x} X_m(n).$$

Como consecuencia

$$\tau = \sum_{m\leq x} X_m.$$

Y por tanto preguntarnos por el promedio de τ no es más que preguntarnos por su esperanza, o sea :

$$\mathbb{E}(\tau) = \mathbb{E}\big(\sum_{m\leq x} X_m\big) = \sum_{m\leq x} \mathbb{E}(X_m).$$

Ahora calculemos $\mathbb{E}(X_m)$:

$$\mathbb{E}(X_m) = \text{Prob}(X_m(n) = 1) = \frac{1}{N}\sum_{\substack{n\leq x\\ m|n}} 1 = \frac{1}{N}\lfloor\frac{x}{m}\rfloor = \frac{1}{N}\big(\frac{x}{m} + O(1)\big) = \frac{1}{m} + O(\frac{1}{x}). \tag{3.2}$$

Observación 3.2. Recordemos que $N \leq x$ con $N = \lfloor x\rfloor$ y que estamos trabajando en Ω_x.

Observación 3.3. Queda apropiado aclarar de nuevo que, la notación $O(1)$ quiere decir "una cantidad y tal que $|y| \leq C$ para alguna constante C" y $O(1/x)$ quiere decir "una cantidad y tal que $|y| \leq C/x$". También que la ecuación $\lfloor x/m\rfloor = x/m + O(1)$ nos está diciendo simplemente que el valor absoluto de la diferencia entre $\lfloor x/m\rfloor$ y x/m es siempre menor que una pequeña constante, y dicha constante es uniforme e igual a 1, no depende de $m \in \Omega_x$, esto es usado a continuación.

Por lo tanto:

$$\mathbb{E}\big(\sum_{m=1}^{n} X_m\big) = \sum_{m=1}^{n} \mathbb{E}(X_m) = \sum_{m\leq x} \frac{1}{m} + \sum_{m\leq x} O\big(\frac{1}{x}\big) =^{*} \log x + O(1).$$

Concluimos que

$$\mathbb{E}(\tau) = \log x + O(1). \tag{3.3}$$

Expliquemos con un poco más de atención el por qué de la igualdad$(^*)$:

Aplicando la *técnica de sumación de Abel* (3.1) en $\sum_{m\leq x} 1/m$ (definiendo $S(x) = \sum_{m\leq x} 1$ y $f(t) = 1/t$) y luego acotando $\sum_{m\leq x} O(1/x)$ obtenemos que:

$$\sum_{m\leq x} \frac{1}{m} = \sum_{m\leq x} 1 \cdot \frac{1}{m} = \lfloor x \rfloor \frac{1}{x} - \int_1^x \lfloor t \rfloor \frac{(-1)}{t^2} dt = \lfloor x \rfloor \frac{1}{x} - \int_1^x (t - \{t\}) \frac{(-1)}{t^2} dt$$

$$= \log(x) + 1 + O(\frac{1}{x}). \tag{3.4}$$

Observación 3.4. En lo anterior, recordemos que $\lfloor t \rfloor = t - \{t\}$.

$$|\sum_{m\leq x} O(\frac{1}{x})| \leq \sum_{m\leq x} \frac{C}{x} = \frac{C}{x} \cdot \lfloor x \rfloor = 1 + O(\frac{1}{x}).$$

Observación 3.5. Es importante destacar que aquí C denota a la misma constante en cada sumando.

De lo cual, sumando ambas expresiones obtenemos

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

3.5 Promedio de $\omega(n)$

Ahora trataremos uno de los resultados importantes que nos preguntamos al principio: si $\omega(n)$ es el número de divisores primos de n, ¿cuánto sera $\omega(n)$ en promedio? Teniendo en cuenta que,

$$\omega(n) = \sum_{p\leq n} 1 = \sum_{p\leq x} X_p(n),$$

y que por tanto

$$\omega = \sum_{p \leq x} X_p$$

entonces, como hicimos con $\tau(n)$, calcularemos su esperanza:

$$\mathbb{E}(\omega) = \mathbb{E}\big(\sum_{p \leq x} X_p\big) = \sum_{p \leq x} \mathbb{E}(X_p) = \sum_{p \leq x} \frac{1}{p} + \sum_{p \leq x} O\big(\frac{1}{x}\big).$$

Ahora bien

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + O(1) \tag{3.5}$$

(como demostraremos en la sección 3.7, usando la esperanza y técnicas de sumación, este es el teorema de Mertens, 1875). Por lo tanto:

$$\mathbb{E}(\omega(n)) = \log\log x + O(1). \tag{3.6}$$

3.6 Cota superior de $\omega(n)$

En esta sección usaremos la desigualdad de Markov para dar una primera cota superior de $\omega(n)$ la cual seguidamente intentaremos de mejorar.

Teorema 3.1 (Desigualdad de Markov). *Sea X una variable aleatoria que toma siempre valores no negativos y sea $t \geq \mathbb{E}(X)$. Entonces se cumple que:*

$$\textit{Prob}(X \geq t) \leq \frac{\mathbb{E}(X)}{t} \quad (\textit{Desigualdad de Markov}). \tag{3.7}$$

Su demostración es sencilla:

Demostración. Por la definición de esperanza, siempre es cierto que

$$\mathbb{E}(X) \geq 0 \cdot \text{Prob}(X < t) + t \cdot \text{Prob}(X \geq t) = t \cdot \text{Prob}(X \geq t).$$

Luego

$$\text{Prob}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Con este resultado, obtenemos de manera inmediata que

$$\text{Prob}(\tau(n) \geq t) \leq \frac{\log x + O(1)}{t}, \tag{3.8}$$

$$\text{Prob}(\omega(n) \geq t) \leq \frac{\log\log x + O(1)}{t}. \tag{3.9}$$

De forma coloquial, esta cota nos viene a decir que, la probabilidad de que el número de factores primos de un cierto número n tomado en el intervalo $(1, x)$ sea mayor que un cierto t, sera menor que el valor de $(\log\log x)/t$ más un cierto error del orden $O(\frac{1}{t})$. No obstante, esta cota la podemos mejorar de la manera siguiente:

Es fácil ver que $\tau(n) \geq 2^{\omega(n)}$ luego es directo que

$$\text{Prob}(\omega(n) \geq t) \leq \text{Prob}(\tau(n) \geq 2^t) \leq \frac{\log x + O(1)}{2^t}. \tag{3.10}$$

Pero realmente, nos tenemos que preguntar en qué se diferencia la cota (3.9) a la nueva que hemos dado en (3.10), o sea, ¿cuánto es de mejor la nueva cota?

Si consideramos $t = (1+\epsilon)\log_2\log x$, entonces (3.9) nos da

$$\text{Prob}(\omega(n) \geq t) \leq \frac{\log 2 + o(1)}{1+\epsilon}$$

mientras que (3.10) nos da

$$\text{Prob}(\omega(n) \geq t) \leq \frac{1 + o(1)}{(\log x)^{\epsilon}}$$

lo cual es una cota mucho más fuerte (es decir, más baja).

Más adelante trataremos de estimar las distribuciones de $\omega(n)$ y $\tau(n)$ con mayor precisión.

3.7 Demostración del teorema de Mertens

Sencillamente en esta sección queremos demostrar un resultado famoso de teoria de números, que usamos directamente en la sección anterior y que usaremos más

adelante en el Teorema de Erdös-Kac. Lo que queremos probar es la igualdad siguiente:

$$\sum_{p\leq x}\frac{1}{p}=\log\log x+O(1). \tag{3.11}$$

Demostración. Dividiremos la demostracón en 6 pasos:

1. Para empezar, tendremos en cuenta el siguiente teorema:

   Teorema 3.2 (Teorema fundamental de la aritmética). *Todo número entero positivo puede ser expresado como un producto de primos de manera única. En otras palabras, para todo entero positivo n,*

   $$n=\prod_{p}p^{v_p(n)}$$

   donde $v_p(n)$ es el máximo entero no negativo k tal que $p^k \mid n$.

   Asi pues, tomando logaritmo en ambos lados de la igualdad, obtenemos que:

   $$n=p_1^{a_1}\cdot p_2^{a_2}\cdots p_k^{a_k}\Rightarrow \log n=a_1\cdot\log p_1+\cdots+a_k\cdot\log p_k=\sum_{i=1}^{k}a_i\log p_i=\sum_{p|n}v_p(n)\log p=\sum_{p^\alpha|n}\log p=\sum_{\substack{d|n\\ d=p^\alpha}}\log p=\sum_{d|n}\Lambda(d) \tag{3.12}$$

   donde $\Lambda(d)$ es la *función de Mangoldt* definida en los conceptos previos.
2. Si tomamos X_d como antes, es decir la variable aleatoria que toma el valor 1 cuando $d \mid n$ y el valor 0 cuando $d \nmid n$. Definiendo $Y=\sum_{d|n}\Lambda(d)X_d$. Entonces, por (3.12), para todo $n\in\Omega_x$, $Y(n)$ siempre toma valores de $\log n$, de donde es directo concluir que, teniendo en cuenta la definición de esperanza que dimos al principio,

   $$\mathbb{E}(Y)=\log x+O(1). \tag{3.13}$$

   Pues, aplicando el lema de Abel:

   $$\mathbb{E}(Y)=\frac{1}{N}\sum_{n\leq x}\log n=\frac{1}{N}\left(\lfloor x\rfloor\log x-\int_1^x\lfloor t\rfloor\frac{dt}{t}\right)=$$

   $$\log x+O\left(\frac{1}{N}(x-1)\right)=\log x+O(1)$$

3. A la vez, tenemos que:

   $$\mathbb{E}(Y)=\sum_{d\leq x}\Lambda(d)\mathbb{E}(X_d)=\sum_{d\leq x}\Lambda(d)\cdot\frac{1}{N}\left\lfloor\frac{x}{d}\right\rfloor,$$

asi que, igualándolo con (3.13),

$$\sum_{d\leq x}\Lambda(d)\frac{1}{N}\lfloor\frac{x}{d}\rfloor=\log x+O(1). \tag{3.14}$$

Como $\frac{1}{N}\lfloor\frac{N}{d}\rfloor=\frac{1}{d}-O(\frac{1}{x})$, estamos a un paso de obtener una estimación de $\sum_{d\leq x}\frac{\Lambda(d)}{d}$:

$$\sum_{d\leq x}\frac{\Lambda(d)}{d}=\log x+O(1)+\sum_{d\leq x}\Lambda(d)\cdot O\big(\frac{1}{x}\big)=\log x+O(1)+\frac{1}{x}\cdot O\big(\sum_{d\leq x}\Lambda(d)\big). \tag{3.15}$$

Sólo nos falta acotar $\sum_{d\leq x}\Lambda(d)$

4. Por (3.14), tenemos

$$\sum_{d\leq x}\Lambda(d)\cdot\frac{1}{N}\lfloor\frac{x}{d}\rfloor=\log x+O(1),$$

y para $x/2$ en lugar de x tenemos que

$$\sum_{d\leq x/2}\Lambda(d)\cdot\frac{1}{N/2}\lfloor\frac{x/2}{d}\rfloor=\log\frac{x}{2}+O(1)$$

y por lo tanto

$$\sum_{\frac{x}{2}\leq d\leq x}\Lambda(d)\leq\sum_{d\leq x}\Lambda(d)\cdot\big(\lfloor\frac{x}{d}\rfloor-2\lfloor\frac{x}{2d}\rfloor\big)=x(\log x+O(1))-x(\log\frac{x}{2}+O(1))=O(x)$$

para todo x. Por lo tanto, dividiendo una suma en *intervalos diádicos*, es decir, intervalos de la forma $M<d\leq 2M$; un procedimiento muy común en análisis, obtenemos que

$$\begin{aligned}\textstyle\sum_{d\leq x}\Lambda(d)&=\textstyle\sum_{\frac{x}{2}\leq d\leq x}\Lambda(d)+\sum_{\frac{x}{4}\leq d\leq\frac{x}{2}}\Lambda(d)+\sum_{\frac{x}{8}\leq d\leq\frac{x}{4}}\Lambda(d)+\dots\\&=O(x)+O(x/2)+O(x/4)+\dots=O(x).\end{aligned} \tag{3.16}$$

5. De (3.14) y (3.16) deducimos que, como

$$\sum_{d\leq x}\frac{\Lambda(d)}{d}=\log x+O(1)\frac{1}{x}\cdot O\big(\sum_{d\leq x}\Lambda(d)\big)$$

y como

$$\sum_{d\leq x}\Lambda(d)=O(x)$$

entonces

$$\sum_{d\leq x}\frac{\Lambda(d)}{d}=\log x+O(1)+\frac{1}{x}\cdot O(x)=\log x+O(1) \tag{3.17}$$

6. Ahora bien, los enteros d de la forma $d=p^{\alpha}$ tienen una contribución en la suma (3.17) prácticamente nula, en concreto, $O(1)$(porque $\sum_n \frac{\log n}{n^2}$ es convergente). Asi pués solo nos queda el caso $d=p$, todo lo demás es igual a $O(1)$ y por tanto tenemos que

$$S(x):=\sum_{p\leq x}\frac{\log p}{p}=\log x+O(1) \tag{3.18}$$

Para librarnos del factor $\log p$, usaremos la fórmula de sumación de Abel (3.1) en $\sum_{p\leq x}1/p$ de la siguiente forma:

$$\sum_{p\leq x}\frac{1}{p}=\sum_{p\leq x}\frac{\log p}{p}\cdot\frac{1}{\log p}=\sum_{n\leq x}a_n\cdot\frac{1}{\log p}=\sum_{n\leq x}a_n\cdot f(n)$$

definiendo

$$a_n=\begin{cases}\frac{\log p}{p} & \text{si } n=p \text{ primo}\\ 0 & \text{en caso contrario}\end{cases}$$

y

$$f(t)=\frac{1}{\log t}.$$

Por tanto, si $S(x)=\sum_{n\leq x}a_n$ usamos Abel (teniendo en cuenta (3.18)), como hicimos en (3.4):

$$\sum_{p\leq x}\frac{1}{p}=S(x)\frac{1}{\log x}+\int_1^x S(t)\frac{dt}{t\log^2 t}=\frac{\log x+O(1)}{\log x}+\int_2^x\frac{dt}{t\log t}+\int_2^x O(1)\frac{dt}{t\log^2 t}$$

Observación 3.6. Tengamos en cuenta que $S(t)=0$ para $1<t<2$ y podemos cambiar el límite inferior de 1 a 2 (observando también que $\log 1=0$).

resolviendo cada integral, y agrupando valores constantes con $O(1)$ obtenemos directamente que:

$$\sum_{p\leq x}\frac{1}{p}=\log\log x+O(1) \tag{3.19}$$

3.8 $\pi(x)$ y la técnica de sumación por partes

En esta sección, nos desviaremos levemente de nuestra preocupación sobre el comportamiento de $\omega(n)$ para dar una aplicación directa del *teorema de Mertens* que acabamos de demostrar, usándolo de nuevo a través de la *fórmula de sumación de Abel*.

Si denotamos $\pi(x)$ como el número de primos que hay entre 1 y N (recordemos que $\lfloor x \rfloor = N$) , podemos definirlo también de la siguiente manera:

$$\pi(x) = \sum_{p \leq x} 1 \leq \sum_{n \leq x} \Lambda(n) \cdot \frac{1}{\log n} = \sum_{n \leq x} \Lambda(n) \cdot f(n) \tag{3.20}$$

por lo que si definimos

$$\psi(x) = \sum_{n \leq x} \Lambda(n) \quad \text{y} \quad f(t) = \frac{1}{\log t}$$

y usamos Abel en (3.20)

$$\sum_{n \leq x} \Lambda(n) \cdot f(n) = \frac{\psi(x)}{\log x} - \int_2^x \psi(t) \frac{dt}{t \log t}.$$

Acotamos la integral, teniendo en cuenta que $|\psi(x)| \leq Ct$ (por (3.16)):

$$\left| \int_2^x \psi(t) \frac{dt}{t \log t} \right| \leq \int_2^x Ct \frac{dt}{t \log t} = C \int_2^x \frac{dt}{\log t} \leq C \int_2^{\sqrt{x}} \frac{dt}{\log t} + C \int_{\sqrt{x}}^x \frac{dt}{\log t} \leq C \frac{x}{\log x}.$$

Observación 3.7. Recordemos que cuando escribimos una constante C, a pesar de que en la misma ecuación haya diferentes constantes, hacemos un abuso de notación llamándolas a todas simbólicamente como C.

Luego concluimos que

$$\pi(x) \ll \frac{x}{\log x}. \tag{3.21}$$

Éste es un resultado de Chebyshev.

Cabría mencionar otro resultado más fuerte y preciso de Chebyshev, como por ejemplo:

Teorema 3.3 (Acotación inferior y superior de $\pi(x)$).

$$0.89\frac{x}{\log x} < \pi(x) < 1.11\frac{x}{\log x}$$

para todo $x \geq x_0$ siendo x_0 fijo.

Más tarde, en 1896, Hadamard y de la Vallée Poussin mostraron (independientemente el uno del otro) que

$$\pi(x) \sim \frac{x}{\log x}. \tag{3.22}$$

(*Teorema de los números primos*)

La mayoría de las demostraciones de (3.22) requieren iniciar el estudio de la *función zeta de Riemann*. Existen también pruebas sin usar el análisis complejo, estas demostraciones son elementales pero generalmente complicadas.

Después de pasear por la esperanza probabilística de $\omega(n)$, basándonos en resultados de Chebyshev-Mertens y la sumación de Abel, nos centraremos ahora en su varianza, para así acercarnos poco a poco a nuestro objetivo final.

4 Chebyshev y la varianza

4.1 El concepto de Varianza

Definición 4.1 (Varianza). *La varianza de una variable aleatoria X está definida como:*

$$Var(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{4.1}$$

Lema 4.1. Sean X, Y dos variable aleatorias independientes. Entonces se cumple que

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y), \tag{4.2}$$

y por tanto

$$Var(X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 = Var(X) + Var(Y). \tag{4.3}$$

En general, si $X_1, X_2, \ldots, X_n$ son variables independientes en pares (es decir, si X_i, X_j fuesen independientes para $i, j \in 1, 2, \ldots, n$ distintos cualesquiera),

$$Var(X_1 + X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n). \tag{4.4}$$

Usaremos a partir de ahora, una desigualdad importante en teoría de la probabilidad, la llamada *desigualdad de Chebyshev.*

Teorema 4.1 (Desigualdad de Chebyshev). *Para toda variable aleatoria X y todo $x > 0$ se tiene que*

$$Prob(|X - \mathbb{E}(X)| \geq x) \leq \frac{Var(X)}{x^2}. \tag{4.5}$$

Demostración.

$$Prob(|X - \mathbb{E}(X)) \geq x) = Prob(|X - \mathbb{E}(X)|^2 \geq x^2) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{x^2} = \frac{Var(X)}{x^2} \tag{4.6}$$

La aplicación que le queremos dar a la varianza y a esta desigualdad es la siguiente: sabemos ya que, en promedio, un número $n \leq x$ tiene $\log\log x + O(1)$ factores primos. ¿Pero cómo son de comunes los números que tienen menos o muchos más factores primos? Para ello primero tendríamos que calcular la varianza de nuestro amigo $\omega(n)$.

4.2 $\omega(n)$ y la Desigualdad de Chebyshev

Nuestro objetivo entonces, será calcular la varianza de $\omega(n)$. Recordamos que, $\omega(n) = \sum_{p \leq x} X_p$ donde X_p es lo que definimos en (2.1). Ahora bien, es obvio que $X_p^2 = X_p$, puesto que $0^2 = 0$ y $1^2 = 1$. Por lo tanto,

$$Var(X_p) = \mathbb{E}(X_p^2) - \mathbb{E}(X_p)^2 = \mathbb{E}(X_p) - \mathbb{E}(X_p)^2 = \mathbb{E}(X_p) - O(1/p^2)$$

y teniendo en cuenta (3.6), sabemos que $\sum_{p \leq x} \mathbb{E}(X_p) = \log\log x + O(1)$. Así:

$$\sum_{p \leq x} Var(X_p) = \sum_{p \leq x} \mathbb{E}(X_p) - \sum_{p \leq x} O\big(\frac{1}{p^2}\big) = \sum_{p \leq x} \mathbb{E}(X_p) - O(1) = \log\log x + O(1).$$

¿Entonces podríamos concluir que teniendo como variable aleatoria a $\omega(n)$, podríamos usar la desigualdad de Chebyshev con $\mathbb{E}(\omega(n)) = \log\log x + O(1)$? Para responder a ésto, no debemos apresurarnos, puesto que las variables X_{p_1}, X_{p_2} con $p_1 \neq p_2$ no son exactamente independientes, puesto que, por ejemplo, si p_1,$p_2 \geq \sqrt{x}$, entonces X_{p_1} y X_{p_2} no pueden ser 1 simultáneamente. No obstante, nos bastaría con que la igualdad $Var(XY) = Var(X) \cdot Var(Y)$ sea válida de manera aproximada.

Veamos: para $p_1 \neq p_2$, con $p_1, p_2 \leq x$,

$$\mathbb{E}(X_{p_1} X_{p_2}) = \mathbb{E}(X_{p_1 p_2}) = \frac{1}{p_1 p_2} + O\big(\frac{1}{x}\big)$$

(con constante absoluta)

$$\mathbb{E}(X_{p_1})\mathbb{E}(X_{p_2}) = (1/p_1 + O(1/x))(1/p_2 + O(1/x)) = \frac{1}{p_1 p_2} + O\big(\frac{1}{x}\big).$$

Luego para $1 \leq p_1, p_2 \leq x$

$$\mathbb{E}(X_{p_1} X_{p_2}) - \mathbb{E}(X_{p_1})\mathbb{E}(X_{p_2}) = O(\frac{1}{x}).$$

Por lo tanto,

$$Var(X_{p_1}+X_{p_2}) = \mathbb{E}((X_{p_1}+X_{p_2})^2) - \mathbb{E}(X_{p_1}+X_{p_2})^2 = Var(X_{p_1}) + Var(X_{p_2}) + O(1/x),$$

y, de la misma manera,

$$Var\big(\sum_{p\leq M} X_p\big) = \sum_{p\leq M} Var(X_p) + O(M^2/x)$$

para todo M. Ahora bien, podemos escoger M de tal manera que el término de error $O(M^2/x)$ sea pequeño, digamos $M = x^{1/3}$. Concluimos, por la desigualdad de Chebyshev, que

$$Prob(|X - \mathbb{E}(X)| \geq k) \leq \frac{\log\log x + O(1)}{k^2} \tag{4.7}$$

para $X = \sum_{p\leq x^{1/3}} X_p$.

Ahora bien, ¿cuál es la diferencia entre X y $\omega(n)$? Un número entero $n \leq N$ no puede tener más de dos divisores primos $> N^{1/3}$: más no caben. Por lo tanto $|X(n) - \omega(n)|$ nunca es más de 2. Obtenemos que

$$|\omega(n) - \log\log n| \leq |\omega(n) - X(n)| + |X(n) - \log\log x| + |\log\log x - \log\log n|$$

$$\leq 2 + |X(n) - \log\log x| + |\log\log x - \log\log n|$$

y por tanto

$$\{|\omega(n) - \log\log x| \geq k\} \subset \{|X(n) - \log\log x| > k-3\} \cup \{|\log\log x - \log\log n| > 1\}$$

en efecto, si $|\omega(n) - \log\log x| \geq k$ y $|\log\log x - \log\log n| \leq 1$ obtenemos

$$k \leq 2 + |X(n) - \log\log x| + 1.$$

Además

$$\log\log n < \log\log x - 1 \quad \Longrightarrow \quad \log n \leq \frac{1}{e}\cdot \log x \quad \Longrightarrow \quad n \leq x^{1/e}$$

hace que

$$P[|\log\log x - \log\log n| > 1] \leq x^{1/e-1} \leq \frac{1}{\sqrt{x}}.$$

Entonces

$$Prob(|\omega(n) - \log\log n| \geq k) \leq \frac{\log\log x + O(1)}{(k + O(1))^2}. \tag{4.8}$$

Dicho de otra manera,

$$Prob(|\omega(n) - \log\log n| \geq t\sqrt{\log\log x}) \leq \frac{1 + O(1/\sqrt{\log\log x})}{t^2} \tag{4.9}$$

para todo $t \geq 1$. Tanto el resultado (4.9) como la prueba que hemos presentado se deben a Erdös y Turán; Hardy y Ramanujan habían dado antes una prueba más complicada de un resultado ligeramente más debil.

Usemos esta cota para dar algún ejemplo a continuación:

Ejemplo 4.1. Escogemos $t = 10$, y obtenemos que

$$\text{Prob}(|\omega(n) - \log\log x| \geq 10\sqrt{\log\log x}) \leq \frac{1}{100} + o(1).$$

Ejemplo 4.2. Escogemos $t = \epsilon\sqrt{\log\log x}$, y obtenemos que

$$\text{Prob}(\omega(n) > (1+\epsilon)\log\log x) \leq \frac{1}{\epsilon^2 \log\log x} + o_\epsilon(1),$$

$$\text{Prob}(\omega(n) < (1-\epsilon)\log\log x) \leq \frac{1}{\epsilon^2 \log\log x} + o_\epsilon(1).$$

4.3 Una cota superior de $\omega(n)$ usando la varianza

Nuevamente nos preguntamos: ¿cuántos primos hay entre 1 y x? Tratemos de ver si podemos atacar el problema usando la varianza, en lugar de la esperanza. La idea central esta clara: los primos son algo que se desvían de la norma, y podemos usar la desigualdad de Chebyshev para obtener una cota sobre la probabilidad de eventos que se desvian de una norma. Podemos usar la desigualdad (4.8) (la cual hemos probado usando la misma desigualdad de Chebyshev) con $k = \log\log n$, y obtenemos

$$\begin{aligned} Prob(\omega(n) = 1) &\leq Prob(|\omega(n) - \log\log n| \geq \log\log x - 2) + o(x) \\ &\leq \frac{\log\log x + O(1)}{(\log\log x)^2 + 2\log\log x \cdot O(1) + O(1)^2} + o(x) \end{aligned}$$

Observación 4.1. La primera desigualdad se debe a que

$$\{n \leq x : \omega(n) = 1\} \subset \{n \leq x : |\omega(n) - \log\log n| \geq \log\log x - 2\} \cup \{n \leq x^{1/2}\} \cup \{n \leq e^e\}$$

pues si $n \leq x$, $\omega(n) = 1$ y $|\omega(n) - \log\log n| < \log\log x - 2$ entonces o bien $\omega(n) = 1 > \log\log n$ o bien $\log\log n - 1 < \log\log x - 2$. En el primer caso $e > \log n \implies e^e > n$. En el segundo, $\log\log n < \log\log x - 1$ luego $\log n < \frac{1}{e}\log x < \frac{1}{2}\log x$ luego $n \leq x^{1/2}$.

Donde

$$\frac{\log\log x + O(1)}{(\log\log x)^2 + 2\log\log x \cdot O(1) + O(1)^2} = \frac{\log\log x + O(1)}{(\log\log x)^2 + O(\log\log x)} = \frac{1 + O(1/\log\log x)}{\log\log x + O(1)}$$

y, agrupando errores,

$$\frac{1 + O(1/\log\log x)}{\log\log x + O(1)} = \frac{1}{\log\log x + O(1)}$$

por lo que, obtenemos

$$Prob(\omega(n) = 1) \leq \frac{1}{\log\log x + O(1)} \tag{4.10}$$

para cualquier n tomado al azar entre 1 y x. Por lo tanto hay a lo más $\frac{x}{\log\log x + O(1)}$ primos entre 1 y x. Esta cota es una cota sumamente débil: la cota que obtuvimos usando la esperanza era mucho mejor. Por tanto, tendríamos que ver si podríamos usar la desigualdad de Chebyshev de otra manera para obtener una cota más fuerte.

Veremos que es asi, y luego veremos que la gran ventaja de lo que haremos sobre lo que hicimos en las notas en la sección anterior es que el método que seguiremos ahora también sirve para obtener cotas sobre muchas cosas a parte del número de primos entre 1 y x.

4.4 Una cota superior de $\omega(n)$ usando un método más flexible

Lo que estamos haciendo es evaluar la varianza de $X = \sum_p X_p$. Al hacer tal cosa, utilizamos el hecho de que las variables X_p, $p \leq x^{1/3}$ (digamos) son casi independientes en pares. Hay todavia un hecho más general que no estamos usando: para

$p_1 < p_2 < \cdots < p_k$ cualesquiera tales que $p_1 p_2 \cdots p_k$ es bastante menor que x, las variables $X_{p_1}, X_{p_2}, \ldots, X_{p_k}$ son mutuamente independientes, o casi. ¿Qué podriamos hacer con esto?. Lo veremos a continuación.

Si definimos

$$Z_p(n) = \begin{cases} -(1-\frac{1}{p}) & \text{si } p \mid n \\ \frac{1}{p} & \text{si } p \nmid n \end{cases} \tag{4.11}$$

donde n es un entero aleatorio entre 1 y x. Es fácil ver que $\mathbb{E}(Z_p) = O(1/x)$:

$$\mathbb{E}(Z_p) = -(1-\frac{1}{p}) \cdot Prob(\{n \leq x : p \mid n\}) + \frac{1}{p} \cdot Prob(\{n \leq x : p \nmid n\}).$$

Tenemos que

$$\text{card}\{n \leq x : p \mid n\} = \left\lfloor \frac{x}{p} \right\rfloor.$$

El otro conjunto es su complementario, de manera que obtenemos

$$\mathbb{E}(Z_p) = -(1-\frac{1}{p})\frac{\lfloor x/p \rfloor}{N} + \frac{1}{p}\frac{N - \lfloor x/p \rfloor}{N} = \frac{1}{p} - \frac{1}{N}\left\lfloor \frac{x}{p} \right\rfloor.$$

Recordando que $\lfloor x/p \rfloor = \lfloor N/p \rfloor$ tendremos que

$$\mathbb{E}(Z_p) = \frac{1}{p} - \frac{1}{N}\left\lfloor \frac{N}{p} \right\rfloor = \frac{1}{p} - \frac{1}{N}\left(\frac{N}{p} - \left\{\frac{N}{p}\right\}\right) = \frac{1}{N}\left\{\frac{N}{p}\right\}$$

Esto prueba que $\mathbb{E}(Z_p) = O(1/N)$, pero para $x > 2$ tenemos $N \geq x/2$, así que finalmente

$$\mathbb{E}(Z_p) = O\left(\frac{1}{x}\right). \tag{4.12}$$

Si definimos para todo d sin divisores cuadrados (es decir, d no divisible por 4, ni por 9, ni por 16,...) Z_d tal que:

$$Z_d = \prod_{p \mid d} Z_p. \tag{4.13}$$

Verifiquemos ahora que $\mathbb{E}(Z_d) = O(\tau(d)/x)$ (aunque esta vez, no es de forma tan directa):

$$\mathbb{E}(Z_d) = \frac{1}{N}\sum_{n \leq x} Z_d(n) = \frac{1}{N}\Big(\sum_{n=1}^{dM} Z_d(n) + \sum_{n=dM+1}^{dM+r} Z_d(n)\Big) = \frac{M}{N}\sum_{n=1}^{d} Z_d(n) + \frac{1}{N}\sum_{n=1}^{r} Z_d(n).$$

Observación 4.2. Hemos tenido en cuenta, que todo entero $N = \lfloor x \rfloor = dM + r$, con $0 \leq r < d$.

Observación 4.3. En la tercera igualdad hemos tenido en cuenta que $Z_d(n) = Z_d(n+d)$ por su propia definición.

Como d es libre de cuadrados, es producto de primos distintos $d = p_1 \cdot p_2 \cdots p_k$. Sea $J_0 = \{p_1, p_2, \ldots, p_k\}$ el conjunto de sus factores primos. El valor de $Z_d(n) = Z_{p_1}(n) \cdots Z_{p_k}(n)$ depende de cuales de los primos $p_1, \ldots, p_k$ dividen a n. Al estudiar las sumas anteriores usaremos $J \subset J_0$ para denotar los primos $p \in J_0$ que dividen a n (J varía con n). Puede haber varios n que den el mismo J, en realidad $J = \{p \in J_0 : p \mid n\}$ luego $\prod_{p\in J} p = mcd(n, d)$, y $Z_d(n)$ se puede escribir como $Z_d(n) = \prod_{p\in J}(\frac{1}{p} - 1) \prod_{p\in J_0\setminus J} \frac{1}{p}$, entonces:

1. El primer sumando:

$$\sum_{n=1}^{d} Z_d(n) = \sum_{J\subset J_0} \prod_{p\in J}(\frac{1}{p} - 1) \prod_{p\in J_0\setminus J} \frac{1}{p} \cdot card\{n \leq d : mcd(n, d) = \prod_{p\in J} p\}$$

teniendo en cuenta que, fijado un J, el número de veces que se va a sumar $\sum_{p|n}$ es igual al cardinal:

$$card\{n \leq d : mcd(n, d) = \prod_{p\in J} p\} = card\{\frac{n}{\prod_{p\in J} p} \leq \frac{d}{\prod_{p\in J} p} : mcd(\frac{n}{\prod_{p\in J} p}, \prod_{p\in J_0\setminus J} p) = 1\}$$

$$= card\{m \leq \prod_{p\in J_0\setminus J} p = mcd(m, \prod_{p\in J_0\setminus J} p) = 1\} = \varphi(\prod_{p\in J_0\setminus J} p) = \prod_{p\in J_0\setminus J} (p-1).$$

Observación 4.4. Recordemos que $\varphi(n)$ es la función de Euler, la cual denota el cardinal del conjunto de números primos con n y menores que n.

tenemos que

$$= \sum_{J\subset J_0} \prod_{p\in J}(\frac{1}{p} - 1) \prod_{p\in J_0\setminus J} \frac{1}{p} \cdot \underbrace{card\{n \leq r : mcd(n, d) = \prod_{p\in J} p\}}_{\prod_{p\in J_0\setminus J}(p-1)}$$

por tanto

$$\sum_{n=1}^{d} Z_d(n) = \sum_{J\subset J_0} \prod_{p\in J}(\frac{1}{p} - 1) \prod_{p\in J_0\setminus J} (1 - \frac{1}{p}) = \prod_{p|d}(1 - \frac{1}{p}) \sum_{J\subset J_0} (-1)^{card(J)} = 0$$

puesto que

$$\sum_{J\subset J_0}(-1)^{card(J)} = \sum_{r=0}^{n}(-1)^r \begin{pmatrix} n \\ r \end{pmatrix} = (1-1)^n = 0.$$

2. El segundo sumando:

$$\left|\frac{1}{N}\sum_{n=1}^{r} Z_d(n)\right| \leq \frac{1}{N}\sum_{n=1}^{d}\left|Z_d(n)\right| \leq \frac{1}{N}\prod_{p|d}(1-\frac{1}{p})\cdot\sum_{J\subset J_0}|(-1)^{card(J)}|$$
$$\leq \frac{1}{N}2^{card(J_0)} = \frac{\tau(d)}{N} = O(\frac{\tau(d)}{x}).$$

Por tanto queda probado que $\mathbb{E}(Z_d) = O(\frac{\tau(d)}{x})$

Ahora bien, como hemos visto, si d_1,d_2 son distintos y carecen de divisores cuadrados, entonces

$$\mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2}) = x^{-2}\cdot O(\tau(d_1))O(\tau(d_2)) \tag{4.14}$$

y haciendo cuentas como hicimos en los dos puntos de arriba, podemos llegar también a que

$$\mathbb{E}(Z_{d_1}Z_{d_2}) = x^{-1}\cdot O(\tau(d_1d_2)) \leq x^{-1}\cdot O(\tau(d_1))O(\tau(d_2)). \tag{4.15}$$

Observación 4.5. No nos centraremos en calcular la igualdad anterior, pues realmente sería repetir cálculos.

Y por lo tanto

$$\mathbb{E}(Z_{d_1}Z_{d_2}) = \mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2}) + x^{-1}\cdot O(\tau(d_1))O(\tau(d_2)). \tag{4.16}$$

Ahora definiremos $Z = \sum^{*}_{d\leq M} Z_d$, donde $M = N^{0.49}$ (el asterisco $*$ en la suma $\sum^{*}_{d\leq M}$ quiere decir que d recorre sólo a los enteros sin divisores cuadrados). Calculemos ahora la esperanza y la varianza de Z ayudándonos de lo que ya sabemos, pero antes, enunciaremos un teorema que usaremos en los cálculos.

Teorema 4.2 (Dirichlet). *Si $\tau(n)$ es el número de divisores de un cierto entero n, entonces tenemos que, cuando x tiende a infinito:*

$$\sum_{n\leq x}\tau(n) = x\log x + (2\gamma-1)x + O(\sqrt{x}) \tag{4.17}$$

donde γ es la constante de Euler.

Observación 4.6. La técnica usada en la prueba de este teorema se debe a Dirichlet y se denomina **método de la hipérbola** (ver por ejemplo en el libro *An Introduction to the Theory of Numbrers*, de *Hardy y Wright* [3]).

Observación 4.7. El termino de error $O(\sqrt{x})$ puede mejorarse. La determinación del ínfimo de los θ tales que podamos sustituir $O(\sqrt{x})$ por $O(x^\theta)$ en el teorema anterior, constituye un problema abierto que se conoce con el nombre de **problema del divisor de Dirichlet**. En 1915 Hardy y Landau probaron que $\inf \theta \geq 1/4$.

- Esperanza:

$$\mathbb{E}(Z) = \mathbb{E}\big(\sum_{d\leq M}^{*} Z_d\big) = \sum_{d\leq M}^{*} \mathbb{E}(Z_d) = \sum_{d\leq M}^{*} O\big(\frac{\tau(d)}{x}\big) \leq \frac{C}{x}\sum_{d\leq M}^{*} \tau(d)$$

  ahora bien, usando el teorema anterior (4.17), tenemos que

$$\frac{C}{x}\sum_{d\leq M}^{*} \tau(d) \leq \frac{C}{x} M\log M \leq C\frac{\log x}{x^{0.51}} \leq \frac{C}{x^{1/2}} = O(x^{-1/2}). \tag{4.18}$$

- Varianza:

$$Var(Z) = \mathbb{E}(Z-\mathbb{E}(Z))^2 = \mathbb{E}(Z^2)-\mathbb{E}(Z)^2 = \mathbb{E}(\sum_{d\leq M}^{*} Z_d^2 + 2\sum_{d_1<d_2}^{*} Z_{d_1}Z_{d_2}) - \big(\mathbb{E}(Z)\big)^2$$

  luego teniendo en cuenta que

$$\big(\mathbb{E}(Z)\big)^2 = \sum_{d\leq M}^{*} \big(\mathbb{E}(Z_d)\big)^2 + 2\sum_{d_i<d_j} \mathbb{E}(Z_{d_j})\mathbb{E}(Z_{d_j})$$

  tenemos

$$Var(Z) = \sum_{d\leq M} Var(Z_d) + 2\sum_{d_1<d_2} \big(\mathbb{E}(Z_{d_1}Z_{d_2}) - \mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2})\big)$$

  y acotando el segundo sumando,

$$\Big|\sum_{d_1<d_2} \big(\mathbb{E}(Z_{d_1}Z_{d_2})-\mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2})\big)\Big| \leq \frac{C}{x}\sum_{d_1<d_2}^{*} \tau(d_1)\tau(d_2) \leq \frac{C}{x}\big(\sum_{d\leq M}^{*} \tau(d)\big)^2 \leq \frac{C}{x}(M\log M)^2 = \frac{C}{x}x^{1-0.02}(\log x)^2 = x^{-0.01}.$$

Por tanto llegamos a que

$$Var(Z) = \sum_{d\leq M}^{*} Var(Z_d) + O(x^{-0.01}) = \sum_{d\leq M}^{*} (\mathbb{E}(Z_d^2) - (\mathbb{E}(Z_d))^2 + O(x^{-0.01}).$$

Nos faltaría calcular explícitamente el valor de $\mathbb{E}(Z_d^2)$, así pues, mostraremos a continuación que precísamente $\mathbb{E}(Z_d^2) = \frac{\varphi(d)}{d^2} + O(\frac{\tau(d)}{x})$, donde $\varphi(d) = d \cdot \prod_{p|d}(1 - 1/p)$, es la función de Euler definida en las notaciones.

$$\mathbb{E}(Z_d^2) = \frac{1}{N}\sum_{n\leq x} Z_d^2(n) = \frac{1}{N}\big(\sum_{n=1}^{dM} Z_d^2(n) + \sum_{n=dM+1}^{dM+r} Z_d^2(n)\big) = \frac{M}{N}\sum_{n=1}^{d} Z_d^2(n) + \frac{1}{x}\sum_{n=1}^{r} Z_d^2(n).$$

Aquí hemos usado que todo número entero n se puede escribir como $n = dM + r$. Ahora, estudiemos el primer y el segundo sumando, definiendo de antemano los conjuntos $J_0 = \{p_1, \ldots, p_k, p_{k+1}, \ldots, p_{\omega(d)}\} \equiv$ primos que dividen a d y $J = \{p_1, \ldots, p_k\}$ los que dividen a a, con $J \subset J_0$ y siendo $a = mcd(n, d)$.

- $\frac{M}{N}\sum_{n=1}^{d} Z_d^2(n)$

$$\sum_{n=1}^{d} Z_d^2(n) = \sum_{a|d} \sum_{\substack{mcd(n,d)=a \\ 1\leq n\leq d}} Z_d^2(n).$$

Ahora bien:

$$\sum_{a|d} \sum_{\substack{mcd(n,d)=a \\ 1\leq n\leq d}} Z_d^2(n) = \sum_{a|d} \sum_{\substack{mcd(\frac{n}{a},\frac{d}{a})=1 \\ 1\leq n\leq d}} \prod_{p\in J}(1 - \frac{1}{p})^2 \prod_{p\in J_0\setminus J} \frac{1}{p^2} =$$

y como hay $\varphi(\frac{d}{a}) = \prod_{p\in J_0\setminus J}(p-1)$ sumandos en la segunda suma, entonces

$$= \sum_{J} \prod_{p\in J_0\setminus J} (p-1) \prod_{p\in J}(1 - \frac{1}{p})^2 \prod_{p\in J_0\setminus J} \frac{1}{p^2}$$

$$= \sum_{J} \prod_{p\in J}(1 - \frac{1}{p})^2 \prod_{p\in J_0\setminus J} \big((1 - \frac{1}{p})\frac{1}{p}\big) = \prod_{p\in J_0}(1 - \frac{1}{p})^2 \sum_{J} \prod_{p\in J_0\setminus J} \frac{1}{p}\frac{1}{(1 - \frac{1}{p})}$$

$$= \prod_{p\in J_0}(1-\frac{1}{p})^2 \prod_{p\in J_0}(1+\frac{1}{(p-1)}) = \prod_{p\in J_0} \frac{(1 - \frac{1}{p})^2}{(1 - \frac{1}{p})} = \frac{d}{d}\prod_{p\in J_0}(1-\frac{1}{p}) = \frac{\varphi(d)}{d}.$$

- $\sum_{n=1}^{r} Z_d^2(n)$

$$\sum_{n=1}^{r} Z_d^2(n) \leq \sum_{n=1}^{d} Z_d^2(n) = \frac{\varphi(d)}{d} \leq \tau(d)$$

puesto que, si d es libre de cuadrados y $0 \leq r \leq d$,

$$\frac{\varphi(d)}{d} = \prod_{p|d}(1 - \frac{1}{p}) \leq \prod_{p|d}(1+1) \leq 2^{\omega(d)} = \tau(d).$$

Por tanto, habiendo estudiado los dos sumandos, llegamos a que

$$\mathbb{E}(Z_d^2) = \frac{M}{N} \cdot \frac{\phi(d)}{d} + \frac{1}{N} O\big(\tau(d)\big)$$

y como $\frac{M}{N} = \frac{1}{d}$, puesto que $N = Md + r \Longrightarrow 1 = \frac{M}{N}d + \frac{r}{N} \Longrightarrow \frac{1}{d} = \frac{M}{N} + \frac{r}{Nd}$, y además $\frac{1}{N} \sim \frac{1}{x}$. Entonces considerando que $(\mathbb{E}(Z_d))^2$ es $O((\frac{\tau(d)}{x}))^2$ y que $\frac{\tau(d)}{x}$ es $O(1)$, concluimos que

$$\mathbb{E}(Z_d^2) = \frac{\phi(d)}{d^2} + O\big(\frac{\tau(d)}{x}\big). \tag{4.19}$$

Por consiguiente

$$Var(Z) = \sum_{d \leq M}^{*} \frac{\phi(d)}{d^2} + O(x^{-0.01}) \ll \log M \leq \log x \tag{4.20}$$

Observación 4.8. La primera desigualdad, se puede ver fácilmente. Basta con darse cuenta que $\phi(d) \leq d$.

Utilizaremos ahora la desigualdad de Chebyshev

$$Prob(|Z - \mathbb{E}(Z)| \geq k) \leq \frac{Var(Z)}{k^2}. \tag{4.21}$$

Si queremos usar la varianza para estimar el número de primos, debemos obtener una variable aleatoria tal que tenga esperanza pequeña y varianza grande en los elementos que queramos contar, en este caso los primos. Ahora bien, si el número n es primo y mayor que M, entonces, para cada d sin divisores cuadrados, entonces $Z_d(n)$

tomará el valor $\prod_{p|d} 1/p = 1/d$ (por la definición de Z_d, ver (4.11) y (4.13)). Por lo tanto, si n es primo y mayor que M,

$$Z(n) = \sum_{d\leq M}^{*} \frac{1}{d} \geq \frac{\sum_{d\leq M} \frac{1}{d}}{\sum_m \frac{1}{m^2}} \gg \sum_{d\leq M} \frac{1}{d} \gg \log M \gg \log x$$

(luego la variable es grande justamente en los primos mayores que M) donde utilizamos el hecho de que la suma $\sum_m \frac{1}{m^2}$ converge, aunque la primera desigualdad no es del todo directa, así que la veremos con más detenimiento:

Es decir, queremos probar con claridad que

$$\sum_{d\leq M}^{*} \frac{1}{d} \geq \frac{\sum_{d\leq M} \frac{1}{d}}{\sum_m \frac{1}{m^2}}.$$

Sabemos que, por definición que $\sum_{d\leq x}^{*} \frac{1}{d} = \sum_{d\leq x} \frac{|\mu(d)|}{d}$ (aquí hemos quitado el asterisco del sumatorio usado la definición de $\mu(d)$, ver en la sección de notaciones (2.2)) y usando el lema de Abel (definiendo $Q(x) = \sum_{d\leq x} |\mu(d)|$ y $f(t) = \frac{1}{t^2}$):

$$\sum_{d\leq x}^{*} \frac{1}{d} = \sum_{d\leq x} \frac{|\mu(d)|}{d} = \frac{Q(x)}{x} + \int_1^x Q(t)\cdot \frac{1}{t^2} dt$$

estudiemos la integral, sabiendo que $Q(x) = \frac{1}{\zeta(2)} \cdot x + O(\sqrt{x})$ por un resultado de teoría analítica de números, donde $\zeta(t)$ denota la *función zeta de Riemann* y usando como notación $R(t) = O(\sqrt{t})$:

$$\int_1^x Q(t)\cdot\frac{1}{t^2}dt = \int_1^x \frac{1}{\zeta(2)}\cdot t\cdot\frac{1}{t^2}dt + \int_1^x R(t)\cdot\frac{dt}{t^2} = \frac{1}{\zeta(2)}\log x + \int_1^\infty \frac{R(t)}{t^2}dt - \int_x^\infty \frac{R(t)}{t^2}dt.$$

y acotando superiormente a $\int_x^\infty \frac{R(t)}{t^2}dt$ con $O(\frac{1}{\sqrt{x}})$ llegamos a que:

$$\sum_{d\leq M}^{*} \frac{1}{d} = \frac{1}{\zeta(2)}\log x + H + O(\frac{1}{\sqrt{x}}) \geq C\log x = C\cdot\frac{\sum_{d\leq M}\frac{1}{d}}{\sum_{m=1}^{\infty}\frac{1}{m^2}}.$$

(donde H denota una constante)

Luego, efectivamente

$$\sum_{d\leq M}^{*} \frac{1}{d} \geq \frac{\sum_{d\leq M} \frac{1}{d}}{\sum_m \frac{1}{m^2}}.$$

Dejando atrás este inciso y continuando con nuestro razonamiento, dado que $\{p \leq x : p > M$, siendo p primo $\} \subset \{n \leq x : |Z - \mathbb{E}(Z)| > k\}$ con $k = \sum_{d\leq M}^{*} 1/d - \mathbb{E}(Z) \gg \log x - O(x^{-1/2}) \gg \log x$, se infiere de forma inmediata que

$$\text{Prob}(n \text{ es primo y mayor que } M) \leq \text{ Prob}(|Z - \mathbb{E}(Z)| \geq k)$$

Por (4.20) y (4.21), concluimos directamente que

$$\text{Prob}(n \text{ es primo y mayor que } M) \ll \frac{1}{\log x}$$

y por lo tanto

$$\text{Prob}(n \text{ es primo}) \ll \frac{1}{\log x} + \text{Prob}(n \leq M) = \frac{1}{\log x} + \frac{M}{x} \ll \frac{1}{\log x} \tag{4.22}$$

para n tomado al azar entre 1 y x. En otras palabras, el número de primos entre 1 y x es $\ll \frac{x}{\log x}$ (es trivial, pues hay x números posibles que tomar entre 1 y x con posibilidad de que sea primo $\frac{1}{\log x}$, es decir, $x \cdot \frac{1}{\log x}$).

Ésta es esencialmente la cota que ya obtuvimos usando la esperanza en la sección anterior, pero la diferencia con lo que ya hicimos está en el propio método que hemos usado ahora, puesto que el método presente obtiene una suma flexibilidad a la hora de usarlo para otras aplicaciones, como veremos a continuación (con menos detalle) de un problema similar.

4.4.1 Acotación del número de primos gemelos usando el método anterior

Procediendo como lo hicimos en el problema anterior, probaremos que

$$\text{Prob}(\text{tanto } n \text{ como } n+2 \text{ son primos}) \ll \frac{1}{(\log x)^2} \tag{4.23}$$

para n tomado al azar entre 1 y x.

1. La variable adecuada ahora es:

$$Z_p = \begin{cases} \frac{2}{p} & \text{si } p \nmid n \quad \text{y} \quad p \nmid (n+2) \\ -(1-\frac{2}{p}) & \text{si } p \mid n \quad \text{o} \quad p \mid (n+2) \end{cases} \tag{4.24}$$

donde como siempre, n es un entero aleatorio entre 1 y x, y

$$Z_d = \prod_{p|d} Z_p.$$

De la misma manera que antes, se puede ver que $\mathbb{E}(Z_{d_1} Z_{d_2})$ y $\mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2})$ son sumamente pequeños para d_1, d_2 distintos y sin divisores cuadrados.

2. Podríamos definir, como antes, $Z = \sum_{d\leq M}^{*} Z_d$. donde $M = x^{\frac{1}{2}-\epsilon}$. Ésto podría dar resultados. No obstante, tenemos el derecho de definir $Z = \sum_{d\leq M}^{*} c_d Z_d$, para unos arbitrarios c_d; hacemos ésto, y afrontamos la tarea de encontrar los c_d que nos den el mejor resultado, es decir, lo que haremos es optimizar el valor de c_d para tener una mejorar cuantitativa, dado que "por suerte", ciertos cálculos finales nos serán más simples de esta manera que si escogiéramos $c_d = 1$.
La idea es usar

$$\text{Prob}(|Z| \geq k) \leq \frac{\mathbb{E}(Z^2)}{k^2}$$

donde k es el valor que Z toma cuando n y $n+2$ son ambos primos.
Es fácil ver que, cuando n y $n+2$ son ambos primos, $Z = \sum_{d\leq M}^{*} c_d \frac{\tau(d)}{d}$, pues:

$$Z_d(n) = \prod_{p|d} Z_p(n) = \prod_{p|d} \frac{2}{p} = \frac{2^{\omega(d)}}{d} = \frac{\tau(d)}{d}.$$

Sin centrarnos mucho en este apartado (puesto que nos estamos centrando en el propio método más que en hacer todas las cuentas esta vez), de la misma forma que hicimos anteriormente se muestra que

$$\mathbb{E}(Z^2) = \sum_{d\leq M}^{*} c_d^2 \mathbb{E}(Z_d^2) + O\left(x^{-1}\left(\sum_{d\leq M}^{*} |c_d|\tau(d)^3\right)^2\right)$$

$$= \sum_{d\leq M}^{*} c_d^2 \prod_{p|d}(1-\frac{2}{p}) + O\left(x^{-1}\left(\sum_{d\leq M}^{*} |c_d|\tau(d)^3\right)^2\right). \tag{4.25}$$

3. Debemos, entonces, encontrar el mínimo de

$$\frac{\sum_{d\leq M}^{*} c_d^2 \prod_{p|d}(1-\frac{2}{p})}{\left(\sum_{d\leq M}^{*} c_d \frac{\tau(d)}{d}\right)} \tag{4.26}$$

la pregunta concretamente es: ¿cómo escogemos c_d de tal manera que (4.26) sea mínimo? o, más bien, ¿cuál es el mínimo valor tomado por (4.26)? para ello utilizaremos la *desigualdad de Cauchy*.

Lema 4.2 (Desigualdad de Cauchy). Para a_n, b_n cualesquiera, se tiene con total certeza que

$$\left(\sum_n a_n b_n\right)^2 \leq \sum_n a_n^2 \cdot \sum_n b_n^2 \tag{4.27}$$

con igualdad solo si hay algun r tal que $a_n = r b_n$ para todo n (o $a_n = 0$ para todo n), es decir, cuando son proporcionales.

La desigualdad de Cauchy no es sino la familiar afirmación de que el producto de dos vectores es menor o igual que el producto de sus normas. En verdad, no necesitaremos la desigualdad de Cauchy, sino simplemente el hecho (evidente) de que (4.27) se vuelve una igualdad cuando $a_n = b_n$ para todo n. La desigualdad de Cauchy solo cumple el rol de asegurarnos que estamos procediendo de la mejor manera posible (en este paso).
La expresión (4.26) es igual a $\frac{\sum^*_{d\leq M} a_d^2}{(\sum^*_{d\leq M} a_d b_d)^2}$ con

$$a_d = c_d \sqrt{\frac{\tau(d)}{d}\prod_{p|d}(1-\frac{2}{p})}, \quad b_d = \sqrt{\frac{\tau(d)}{d}}\left(\prod_{p|d}(1-\frac{2}{p})\right)^{-1/2}.$$

Por la desigualdad Cauchy, el mínimo de $\frac{\sum^*_{d\leq M} a_d^2}{(\sum^*_{d\leq M} a_d b_d)^2}$ es $\frac{1}{\sum^*_{d\leq M} b_d^2}$, es decir,

$$\frac{1}{\sum^*_{d\leq M} \frac{\tau(d)}{d}\prod_{p|d}(1-\frac{2}{p})^{-1}1}.$$

Este mínimo es alcanzado cuando $a_d = b_d$, i.e., cuando $c_d = \prod_{p|d}(1-\frac{2}{p})^{-1}$. Tenemos, entonces, utilizando (4.25):

$$\text{Prob}(n \text{ y } n{+}2 \text{ son primos}) \leq \frac{\mathbb{E}(Z^2)}{k^2} \leq \frac{1}{\sum^*_d \frac{\tau(d)}{d}\prod_{p|d}(1-\frac{2}{p})^{-1}} + O\left(\frac{M(\log x)^A}{x}\right).$$

El término $O(\frac{M(\log x)^A}{x})$ puede ser omitido pues es menor que el primero ($\ll x^{-1/2}$). Ahora bien,

$$\sum_d^* \frac{\tau(d)}{d}\prod_{p|d}(1-\frac{2}{p})^{-1} = \sum_d^* \frac{\tau(d)}{d}\prod_{p|d}(1+\frac{2}{p}+\frac{2^2}{p^2}+\dots) \geq \sum_{d\leq M}\frac{\tau(d)}{d} \geq \left(\sum_{d\leq M^{1/2}}\frac{1}{d}\right)^2$$

$$\sim (\log M^{1/2})^2 \gg (\log x)^2.$$

Por lo tanto

$$\text{Prob}(n \text{ y } n+2 \text{ son primos}) \ll \frac{1}{(\log x)^2}. \tag{4.28}$$

Lo que acabamos de hacer puede verse como una versión del método llamado *criba de Selberg* (1950).

La primera prueba del resultado (4.28) fue dada por V. Brun (1920). Como curiosidad, podemos incluir que actualmente se cree que

$$\text{Prob}(n \text{ y } n+2 \text{ son primos}) \sim \frac{c_2}{(\log x)^2} \tag{4.29}$$

donde

$$c_2 = 2\prod_{p\geq 3} \frac{p(p-2)}{(p-1)^2} \sim 1.32032\ldots$$

Sin embargo, esta conjetura sigue sin probarse; no se sabe ni quiera si es que hay un número infinito de primos n tales que $n+2$ sea también primo (*conjetura de los primos gemelos*). El enunciado(4.29) es parte de la *conjetura de Hardy-Littlewood*, la cual también especifica, por ejemplo, cuál debe ser la probabilidad que n, $n+2$ y $n+6$ sean todos primos. En el año 2013 Yitang Zhang probó que existe un número natural a tal que existen infinitos primos p tales que $p+a$ es también primo. Actualmente se conoce incluso que $a \leq 246$

A continuación estudiaremos el llamado *límite central*, de tal manera que podamos entenderlo y usarlo de alguna forma a los primos, y en especial a $\omega(n)$.

5 El límite central

Ya conocemos $\mathbb{E}(\omega(n))$ y $Var(\omega(n))$ para n un número tomado al azar entre 1 y x, donde x es grande. Ahora bien, lo que quisiéramos conocer de una vez por todas es la distribución de $\omega(n)$ en el límite $x \to \infty$.

Como antes, comenzaremos recordando que $\omega(n)$ es una suma de variables aleatorias, y enfocamos el problema de manera general. La siguiente observación se remonta en alguna forma a la de Moivre (1718): *si algo es la suma de muchas pequeñas cosas que nada o poco tienen que ver entre sí, este algo tendrá una distribución en forma de campana.*

5.1 Teorema del límite central

Teorema 5.1 (Teorema del límite central). *Sean $X_1, X_2, X_3, \ldots$ variables aleatorias mutuamente independientes. Si todas son de idéndica distribución; con esperanza μ y varianza σ, y con $\mathbb{E}(X_j^k)$ finita para todo $k \geq 0$. Entonces:*

$$\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mu)$$

tiende en distribución a

$$\frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{5.1}$$

La distribución dada por la función de densidad (5.1) es la conocida *distribución normal.*

La demostración de este teorema viene dada en las asignaturas básicas de Teoria de la Probabilidad, por lo que no nos centraremos en demostrarla. No obstante daremos

un pequeño esquema de la demostración, dado que la idea central del dicho esbozo, nos será de utilidad cuando examinemos $\omega(n)$. El método es el llamado *método de momentos*.

Demostración (Esquema de la prueba). Se trata de comparar los momentos $\mathbb{E}(S_n)$, $\mathbb{E}(S_n^2)$, $\mathbb{E}(S_n^3)$, ... de la variable $S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$ con los momentos $\mathbb{E}(S)$, $\mathbb{E}(S^2)$, $\mathbb{E}(S^3)$, ... de una variable S de distribución normal.

Por integración por partes, podemos ver que $\mathbb{E}(S^k) = (k-1)!!$ para k par; como S es simétrica con respecto al eje y, está claro que $\mathbb{E}(S^k) = 0$ para k impar. Podemos verificar $\mathbb{E}(S_n^k) = (k-1)!! + o_k(1)$ para k par, y $\mathbb{E}(S_n^k) = o_k(1)$ para k impar.

Como los momentos de S_n convergen a los momentos de S y la distribución normal satisface ciertas condiciones técnicas, podemos concluir que $S_n \to S$ utilizando un resultado auxiliar estándar del que hablaremos a continuación.

Teorema 5.2. *Sean $X_1, X_2, X_3, \ldots$ y X variables aleatorias tales que $\mathbb{E}(X_j^k)$ y $\mathbb{E}(X^k)$ son finitos para $j, k \geq 0$ cualesquiera. Supongamos que los momentos de X_j convergen a los momentos de X: $\lim_{j\to\infty} \mathbb{E}(X_j^k) = \mathbb{E}(X^k)$. Supongamos también que $\frac{1}{k!}\mathbb{E}(X^k) < C^k$ para algún $C > 0$ y todo $k > 0$. Entonces $X_1, X_2, X_3, \ldots$ convergen en distribución a X.*

(Daremos este último teorema como supuesto, pues se explica en cualquier curso básico de Teoria de la probabilidad, ver libro *Probability and Measure* de *Billingsley* [2])

5.2 Condiciones del teorema del límite central

Hemos asumido tres cosas acerca de las variables X_j para que podamos usar el teorema del límite central:

$$\begin{cases} (a) & \text{que son mutuamente independientes} \\ (b) & \text{que tienen la misma distribución} \\ (c) & \text{que para cada } k, \mathbb{E}(X_j^k) \text{ esta acotada independientemente de } j \end{cases}$$

Tanto (b) como (c) pueden relajarse; la *condición de Lindeberg* sustituye a las dos (lo veremos ahora). Es más difícil prescindir de (a); hay herramientas estándar para tal tarea pero ninguna cubre todos los casos que aparecen en la práctica.

Teorema 5.3 (Teorema del límite central - Lindeberg). *Sean $X_1, X_2, X_3, \ldots$ variables aleatorias mutuamente independientes. Sean*

$$S_n = \sum_{j=1}^{n} (X_j - \mathbb{E}(X_j)), \qquad s_n = \sqrt{Var(S_n)} = \sqrt{\sum_{j=1}^{n} Var(X_j)}.$$

Supongamos que, para todo $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{s_n^2} \int_{|t|\geq \epsilon s_n} t^2 f_j(t)dt < \infty, \qquad \textit{(condición de Lindeberg)} \tag{5.2}$$

donde f_j es la función de densidad de X_j. Entonces S_n/s_n tiende en distribución a la normal $\frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ cuando $n \to \infty$.

Observación 5.1. Si las variables X_j son discretas, la condición de Lindeberg sería equivalente a escribir

$$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{s_n^2} \sum_{x:|x|\geq \epsilon s_n} x^2 \text{Prob}(X_j = x) < \infty.$$

(tampoco nos desviaremos para demostrar este teorema; sigue siendo un resultado teórico de cursos intermedios de Teoría de la Probabilidad, se podría leer la demostración por ejemplo en el libro *All of Statistics* de autores *Wasserman y Larry*[7]).

Antes de ver como podemos arreglárnoslas sin (a), hagamos dos cosas: primero, verifiquemos que la falta de (b) en el caso que más nos interesa es inocua; para asuntos posteriores tenemos que hacer uso de los conceptos *transformada de fourier* y *función característica*:

Definición 5.1 (Transformada de Fourier). *La transformada de Fourier $\widehat{f} : \mathbb{R} \to \mathbb{C}$ de una función $f : \mathbb{R} \to \mathbb{C}$ se define como sigue:*

$$\widehat{f}(t) := \int_{-\infty}^{\infty} e^{itx} f(x)dx.$$

La función $f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ es un vector propio de la transformada de Fourier, es decir, para ese f, la transformada $\widehat{f}(t)$ resulta ser un múltiplo de $f(t)$: $\widehat{f}(t) = \sqrt{2\pi} f(t)$.

Definición 5.2 (Función característica). *Dada una variable aleatoria X, definimos la función característica $\widehat{X} : t \mapsto \mathbb{E}(e^{itX})$.*

Si X es continua,

$$\widehat{X}(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f(x) dx,$$

donde f es la función de densidad de X.

Si X es discreta,

$$\widehat{X}(t) = \mathbb{E}(e^{itX}) = \sum Prob(X = x) e^{itx}.$$

Lema 5.1. Cuando se tienen dos variables independientes X, Y, la variable $X + Y$ tiene como distribución la convolución de las distribuciones.

Lema 5.2. La transformada de Fourier $\widehat{f * g}$ de la convolución $f * g$ de dos funciones, es igual a $\widehat{f} \cdot \widehat{g}$. En consecuencia, $\widehat{X + Y} = \widehat{X} \cdot \widehat{Y}$.

Observación 5.2 (Como solventar que nuestras variables no tengan la misma distribución). Siendo $\omega = \sum_{p \leq x} X_p$ donde los X_p ni son independientes ni tienen la misma distribución. En primer lugar consideraremos unas variables X'_p que van a ser independientes pero tienen la misma distribución y vemos que $\omega' = \sum_{p \leq x} X'_p$ sí convergen a la distribución normal.

Ahora realizaremos unos cálculos previos, que nos servirán posteriormente para el límite central de $\omega(n)$:

Sean $X'_2, X'_3, X'_5, \ldots$ variables mutuamente independientes con la siguiente distribución:

$$X'_p = \begin{cases} 1 & \text{con probabilidad } 1/p \\ 0 & \text{con probabilidad } 1 - 1/p \end{cases} \tag{5.3}$$

(Escojemos los signos $X'_2, X'_3, X'_5, \ldots$ porque usaremos $X_2, X_3, X_5, \ldots$ más tarde)

Observación 5.3. Las variables aleatorias $X_p : \Omega_x \times [0, 1]$ están definidas por $X_p(n, t) = 1$ si $p \mid n$ y vale 0 en otro caso. Las variables X'_p son independientes y toman el valor 1 con probabilidad $1/p$ y 0 con probabilidad $1 - 1/p$.

Entonces $\mathbb{E}(X'_p) = \frac{1}{p}$, $\mathbb{E}((X'_p - \mathbb{E}(X'_p))^2) = \text{Var}(X'_p) = \frac{1}{p} - \frac{1}{p^2}$, $\mathbb{E}(|X'_p - \mathbb{E}(X'_p)|^3) \leq \frac{1}{p}$. Por consiguiente, la función característica de $X'_p - \mathbb{E}(X'_p)$ es

$$1 - \frac{t^2}{2}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \frac{O(t^3)}{p}.$$

Por el mismo razonamiento, la función característica de $\frac{1}{\sqrt{\log\log x}}(X'_p - \mathbb{E}(X'_p))$ es

$$1 - \frac{t^2}{2\log\log x}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \frac{O(t^3)}{p(\log\log x)^{3/2}}. \tag{5.4}$$

Definimos $S'_x = \frac{1}{\sqrt{\log\log x}} \sum_{p\leq x}(X'_p - \mathbb{E}(X'_p))$. Usando la regla $\widehat{X+Y} = \widehat{X}\cdot\widehat{Y}$ y (5.4), vemos que

$$\widehat{S'_x} = \prod_{p\leq x}\left(1 - \frac{t^2}{2\log\log x}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \frac{O(t^3)}{p(\log\log x)^{3/2}}\right)$$

$$= \prod_{p\leq x} e^{-\frac{t^2}{2\log\log x}\cdot(\frac{1}{p}-\frac{1}{p^2})\cdot(1+O(t/\sqrt{\log\log x})} = e^{-(1+o_t(1))\cdot\sum_{p\leq x}\frac{t^2}{2\log\log x}(\frac{1}{p}-\frac{1}{p^2})}$$

cuando $x \to \infty$. Por el teorema de Chebyshev-Mertens, sabemos que $\sum_{p\leq x}\frac{1}{p} = \log\log x + O(1)$, y por tanto:

$$\sum_{p\leq x}\frac{t^2}{2\log\log x}\left(\frac{1}{p} - \frac{1}{p^2}\right) = \frac{t^2}{2}(1 + O(1/\log\log x))$$

cuando $x \to \infty$, y en consecuencia

$$\widehat{S'_x} = e^{-t^2/2\cdot(1+o_t(1))} \to e^{-t^2/2}$$

cuando $x \to \infty$. Por el teorema de convergencia de Lévy (que enunciaremos a continuación), concluimos que S'_x converge en distribución a la normal.

Teorema 5.4 (Teorema de convergencia de P. Lévy). *Sean $X_1, X_2, X_3, \ldots$ variables aleatorias con funciones características $\widehat{X_1}, \widehat{X_2}, \widehat{X_3}, \ldots$ Asumiendo que para todo real t, la sucesión $\widehat{X_1}(t), \widehat{X_2}(t), \widehat{X_3}(t), \ldots$ tiene un límite $f(t)$. Si f es continua alrededor de $t = 0$, entonces f es la función característica $\widehat{X}$ de alguna variable aleatoria X, y las variables $X_1, X_2, X_3, \ldots$ convergen a X en distribución.*

5.3 El límite central de $\omega(n)$

Sean ahora $X_2, X_3, X_5, \ldots$ variables dadas por

$$X_p = \begin{cases} 1 & \text{si } p \mid n \\ 0 & \text{si } p \nmid n \end{cases} \tag{5.5}$$

donde n es un entero aleatorio entre 1 y x. Como ya sabemos, $\omega(n) = \sum_{p \leq x} X_p(n)$. Queremos probar que la distribución de $\omega(n)$ o, más bien dicho,

$$\frac{1}{\sqrt{\log \log x}}(\omega(n) - \log \log x)$$

tiende a la normal.

Cuando calculamos la varianza de $\omega(n)$, vimos que las variables X_p son casi independientes en pares: X_{p_1} y X_{p_2} son aproximadamente independientes para $p_1, p_2 < x^{1/2-\epsilon}$, $p_1 \neq p_2$ primos cualesquiera, y, en total, los términos de error son pequeños. No obstante, las variables X_p están muy lejos de ser mutuamente independientes. ¿Qué podemos hacer?.

Podemos probar el teorema del límite central para $\omega(n)$ por el método de momentos. Cuando calculamos el momento $\mathbb{E}(\omega(n)^k)$, sólo necesitamos el hecho que las variables X_p sean casi independientes "tomadas de k en k", esto es: k variables distintas cualesquiera entre $X_1, X_2, X_3, \ldots$ son aproximadamente independientes, por las mismas razones que ya vimos para $k = 2$. El término de error dependera de k, y por lo tanto la efectividad de convergencia de $\mathbb{E}(\omega(n)^k)$ a su límite dependerá de k; sin embargo, al método de momentos esto no le importa. Pasando todo esto a limpio, podemos probar *teorema de Erdös-Kac* en la siguiente sección.

6 El teorema de Erdös-Kac

Teorema 6.1 (Erdös-Kac). *Sean $Y_x : \Omega_x \to \mathbb{R}$ las variables aleatorias definidas en nuestro espacio de probabilidad Ω_x por*

$$Y_x(n) = \frac{1}{\sqrt{\log\log x}}(\omega(n) - \log\log x)$$

(n aquí representa un entero tomado al azar entre 1 y x). Entonces Y_x tiende en distribución a una normal

$$\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

cuando $x \to \infty$, es decir, que para cualesquiera que sean los números reales ω_1, ω_2, tales que $\omega_1 < \omega_2$,

$$\lim_{x\to\infty} P(\omega_1 < Y_x < \omega_2) = \lim_{x\to\infty} \frac{1}{x} \cdot card\{n \leq x, \omega_1 \leq \frac{\omega(n) - \log\log x}{\sqrt{\log\log x}} \leq \omega_2\}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\omega_1}^{\omega_2} e^{-t^2/2} dt.$$

La demostración que veremos se debe a Billingsley [2]; incluye varias ideas de la prueba de Erdös y Kac (1939) y de la prueba de Halberstam (1995).

Como antes, expresaremos $\omega(n)$ como una suma $\sum_{p\leq x} X_p$. Así como para utilizar la desigualdad de Chebyshev tuvimos que truncar la suma $\sum_{p\leq x} X_p$ (reemplazándola por $\sum_{p\leq x^{1/3}} X_p$), tendremos que truncarla ahora (aún más, ya que la reemplazaremos por $\sum_{p\leq g(x)} X_p$, donde $g(x)$ crece más lentamente que cualquier potencia de x). Primero mostraremos que esta truncación no nos sera dañina, es decir, que el total omitido $\sum_{g(x)<p\leq x} X_p$ es pequeño; luego procedermos a determinar la distribución de la suma truncada $\sum_{p\leq g(x)} X_p$ utilizando el método de momentos.

6.1 Demostración del teorema de Erdös-Kac

Demostración. Sea $g(x)$ una función tal que $g(x) = o_\epsilon(x^\epsilon)$ para todo $\epsilon > 0$ y $\log\log x - \log\log g(x) = o(\sqrt{\log\log x})$; podemos tomar, por ejemplo, $g(x) = x^{1/\log\log x}$. Vemos inmediatamente que:

$$Y_x(n) = (1 + o(1)) \cdot \frac{1}{\sqrt{\log\log g(x)}}(\omega(n) - \log\log x).$$

Ahora bien,

$$\sum_{p \le g(x)} 1/p = \log\log g(x) + O(1) = \log\log x + o(\sqrt{\log\log x}) = \log\log x + o(\sqrt{\log\log g(x)})$$

Concluimos que

$$Y_x(n) = (1 + o(1)) \cdot \frac{1}{\sqrt{\log\log g(x)}}\left(\omega(n) - \sum_{p \le g(x)} 1/p\right) + o(1). \tag{6.1}$$

Definamos ahora

$$S_m = \frac{1}{\sqrt{\log\log m}} \cdot \left(\sum_{p \le m}(X_p - 1/p)\right) \tag{6.2}$$

donde X_p es como en (5.5). Está claro que $\omega(n) = \sum_{p \le x} X_p(n)$ y por lo tanto $\omega(n) - \sum_{p \le g(x)} X_p(n)$ es igual a $\sum_{g(x) < p \le x} X_p(n)$. Calculamos la esperanza de esto último:

$$\mathbb{E}\left(\sum_{g(x) < p \le x} X_p\right) = \sum_{g(x) < p \le x} \frac{1}{p} = \log\log x - \log\log g(x) + O(1) \tag{6.3}$$

$$= o(\sqrt{\log\log x}) = o(\sqrt{\log\log g(x)}).$$

Como X_p toma solo valores positivos, la desigualdad de Markov nos permite deducir de (6.3) que

$$\frac{1}{\sqrt{\log\log g(x)}}\left(\sum_{g(x) < p \le x} X_p\right) = o(1)$$

con probabilidad $1 - o(1)$.

Veamos esto último con más detenimiento:

La desigualdad de Markov, nos decía que $\text{Prob}(X \ge t) \le \frac{\mathbb{E}(X)}{t}$, luego, tomando $X = \sum_{g(x) < p \le x} X_p$:

$$\text{Prob}\left(\sum_{g(x)<p\leq x} X_p \geq t\right) \leq \frac{o(\sqrt{\log\log g(x)})}{t}$$

definiendo $e(x) = o(\sqrt{\log\log g(x)})$, y $h(x) = \frac{e(x)}{\sqrt{\log\log g(x)}}$, que tiende a 0 cuando $x \to \infty$. Escogemos el valor de t como $t = \sqrt{h(x)} \cdot \sqrt{\log\log g(x)}$, entonces:

$$\text{Prob}\left(\sum_{g(x)<p\leq x} X_p \leq \sqrt{h(x)}\cdot\sqrt{\log\log g(x)}\right) > 1 - \frac{e(x)}{\sqrt{h(x)}\cdot\sqrt{\log\log g(x)}} = 1-\sqrt{h(x)} = 1-o(1)$$

por lo tanto

$$\text{Prob}\left(\frac{1}{\sqrt{\log\log g(x)}}\left(\sum_{g(x)<p\leq x} X_p\right) \leq o(1)\right) > 1 - o(1)$$

y de ahi, obtenemos lo que dijimos.

Concluimos ahora, por (6.2) y (6.1) que

$$Y_x = (1 + o(1))S_{g(x)} + o(1) \tag{6.4}$$

con probabilidad $1 - o(1)$ cuando $x \to \infty$. En consecuencia, si probamos que $S_{g(x)}$ tiende en distribución a la normal, habremos probado que Y_x tiende en distribución a la normal.

Hasta ahora, nuestra labor ha sido sólo la preparatoria: lo más que hemos hecho es truncar la suma $\sum_{p\leq x}$ y mostrar que el efecto de tal truncación es pequeño. A continuación, nuestra tarea es averiguar cuáles son los momentos de $S_{g(x)}$.

Sea $k \geq 0$. Sean X'_p como en (5.3) y $S'_m = \frac{1}{\sqrt{\log\log m}} \sum_{p\leq m}(X'_p - 1/p)$. Para k primos $p_1, p_2, \ldots, p_k$ cualesquiera (no necesariamente distintos),

$$\mathbb{E}(X_{p_1}X_{p_2}\cdots X_{p_k}) = \frac{1}{N}\sum_{n\leq x} X_{p_1}(n)X_{p_2}(n)\cdots X_{p_k}(n).$$

Teniendo en cuenta que, X_{p_i} vale 1 si $p_i \mid n$, tomamos el mínimo común múltiplo entre $\{p_1, p_2, \ldots, p_k\} \mid n$, (esto significa que cada uno divida a n) y lo denotamos por d. Luego, como el número de múltiplos de d menores o iguales que N es $\lfloor \frac{N}{d} \rfloor$ entonces:

$$\mathbb{E}(X_{p_1}X_{p_2}\cdots X_{p_k}) = \frac{1}{N}\left\lfloor \frac{N}{d} \right\rfloor.$$

Y de la misma manera, tenemos que, como $\mathbb{E}(X'_p) = \frac{1}{p}$ y como son independientes

$$\begin{aligned}\mathbb{E}(X'_{p_1}X'_{p_2}\cdots X'_{p_k}) &= \mathbb{E}(X'_{p_1})\mathbb{E}(X'_{p_2})\cdots\mathbb{E}(X'_{p_k}) \\ &= \frac{1}{p_1\cdot p_2\cdots p_k} = \frac{1}{d}.\end{aligned}$$

Observación 6.1. Tengamos en cuenta que $\mathbb{E}(X'_p) = 1/p$ y es obvio que, si hay primos iguales el resultado es el mismo pues $(X'_p)^k = X'_p$.

Bien pues, quedaría claro entonces que

$$|\mathbb{E}(X_{p_1}X_{p_2}\cdots X_{p_k}) - \mathbb{E}(X'_{p_1}X'_{p_2}\cdots X'_{p_k})| = \left|\frac{1}{N}\left\lfloor\frac{N}{d}\right\rfloor - \frac{1}{d}\right| \leq \frac{1}{N}, \tag{6.5}$$

donde d hemos dicho que es el mínimo común múltiplo de $p_1, p_2, \ldots, p_k$.

$g(x) = x^{1/\log\log x}$ es una constante (dado nuestro espacio $\Omega_x \times [0,1]$) . Consideramos ahora las variables

$$S = \frac{1}{\sqrt{\log\log g(x)}}\sum_{p\leq g(x)}(X_p - \tfrac{1}{p}), \quad S' = \frac{1}{\sqrt{\log\log g(x)}}\sum_{p\leq g(x)}(X'_p - \tfrac{1}{p}).$$

Queremos probar que

$$\mathbb{E}(S^k) = \mathbb{E}((S')^k) + O_k\left(\frac{g(x)^k}{x}\right)$$

Veamoslo detenidamente acotando la diferencia entre ambas esperanzas, y viendo que, efectivamente es $O_k(\frac{g(x)^k}{x})$:

Sea $U(x) = \log\log g(x)^{1/2}$. Tenemos entonces

$$U(x)^k|\mathbb{E}(S^k) - \mathbb{E}(S')^k| \leq \left|\mathbb{E}\left[\left(\sum_{p\leq g(x)}(X_p - \tfrac{1}{p})\right)^k - \left(\sum_{p\leq g(x)}(X'_p - \tfrac{1}{p})\right)^k\right]\right|.$$

Los primos $\leq g(x)$ son $\pi_0 = \pi(g(x))$ es decir, las sumas en p son el conjunto de primos $P := \{p_1, p_1, \ldots, p_{\pi_0}\}$. Desarrollando la potencia $k-$ésima se obtiene

$$\left(\sum_{p\leq g(x)}(X_p - \tfrac{1}{p})\right)^k = \sum_{(q_1,\ldots,q_k)\in P^k}\prod_{j=1}^{k}(X_{q_j} - \tfrac{1}{q_j}).$$

Por consiguiente

$$U(x)^k|\mathbb{E}(S^k) - \mathbb{E}(S')^k| \leq \sum_{(q_1,\ldots,q_k)\in P^k} \left|\mathbb{E}\left(\prod_{j=1}^{k}(X_{q_j} - \tfrac{1}{q_j}) - \prod_{j=1}^{k}(X'_{q_j} - \tfrac{1}{q_j})\right)\right|.$$

Ahora desarrollamos los productos, en cada uno de los k factores podemos escoger el primer sumando X_{q_j} o el segundo $-\frac{1}{q_j}$. Para cada $J \subset \{1, 2, \ldots, k\}$ tenemos uno de los términos

$$\prod_{j=1}^{k}(X_{q_j} - \tfrac{1}{q_j}) = \sum_{J\subset\{1,2,\ldots,k\}} \prod_{j\in\{1,2,\ldots,k\}\smallsetminus J} (-\tfrac{1}{q_j}) \prod_{j\in J} X_{q_j}.$$

Hacemos esto mismo con la segunda potencia $k-$ésima y asociamos los terminos correspondientes al mismo J. Al final obtenemos

$$U(x)^k|\mathbb{E}(S^k)-\mathbb{E}(S')^k| \leq \sum_{(q_1,\ldots,q_k)\in P^k} \sum_{J\subset\{1,2,\ldots,k\}} \prod_{j\in\{1,2,\ldots,k\}\smallsetminus J} \frac{1}{q_j} \left|\mathbb{E}(\prod_{j\in J} X_{q_j} - \prod_{j\in J} X'_{q_j})\right|.$$

Pero hemos visto antes en (6.5) que la diferencia de estas dos esperanzas es$\leq 1/\lfloor x\rfloor$. Por consiguiente lo que tenemos es

$$U(x)^k|\mathbb{E}(S^k) - \mathbb{E}(S')^k| \leq \sum_{(q_1,\ldots,q_k)\in P^k} \sum_{J\subset\{1,2,\ldots,k\}} \prod_{j\in\{1,2,\ldots,k\}\smallsetminus J} \frac{1}{q_j}\frac{1}{\lfloor x\rfloor}.$$

Como cada $q_j \geq 2$ tendremos $1/q_j \leq 1$ y entonces se simplifica a

$$U(x)^k|\mathbb{E}(S^k) - \mathbb{E}(S')^k| \leq \sum_{(q_1,\ldots,q_k)\in P^k} \sum_{J\subset\{1,2,\ldots,k\}} \frac{1}{\lfloor x\rfloor} \leq \frac{1}{\lfloor x\rfloor} 2^k \pi_0^k \leq 2^k \frac{g(x)^k}{\lfloor x\rfloor}.$$

Como $U(x) = \log\log g(x)^{1/2}$ tiende a infinito con x, esto implica también que

$$|\mathbb{E}(S^k) - \mathbb{E}((S')^k)| \leq 2^k \frac{g(x)^k}{\lfloor x\rfloor} = O\left(\frac{g(x)^k}{x}\right).$$

Como $g(x) = o_\epsilon(x^\epsilon)$, sabemos que $O(g(x)^k \cdot \frac{1}{x}) = o_k(1)$.

Ya vimos, que la distribución de S'_m tendía a la normal cuando $m \to \infty$; por lo tanto, los momentos $\mathbb{E}((S'_m)^k)$ de S_m tienden a los momentos $\mathbb{E}(W^k)$ de la normal W. Tenemos, entonces, que:

$$\lim_{x\to\infty} \mathbb{E}(S^k_{g(x)}) = \lim_{x\to\infty} \mathbb{E}((S'_{g(x)})^k) = \mathbb{E}(W^k)$$

para todo k. Concluimos entonces que $S^k_{g(x)}$ converge en distribución a la normal, y por lo tanto, Y_x converge en distribución a la normal.

Con lo anterior, quedaría probado el teorema de Erdös-Kac, y con ello, acabaría nuestra misión de mostrar algo más de los primos cuanto menos curioso e interesante. Queda plasmada en estas páginas como algo tan aparentemente sencillo como son los divisores primos de un número pueden llegar a dar tanto juego en una de las áreas de las matemáticas actualmente tan usadas como son la teoría de la probabilidad y el propio análisis. De la misma manera, problemas tan aparentemente sencillos, formulados hace siglos y combatidos por los mas grandes matemáticos, siguen aún vigentes y sin solución.

Bibliografía

[1] T. Apostol, *Introduction to Analytic Number Theory*, Springer, New York, 1976.

[2] P. Billingsley, *Probability and Measure*, Wiley, The University of Chicago, 1979.

[3] G.H. Hardy y E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1938.

[4] Harald Andrés Helfgott, Azar y aritmética, arXiv:0909.0922, (2009).

[5] M. Kac (1959), *Statistical Independence in Probability, Analysis and Number Theory*, Mathematical Association of America, New York, 1969. **9** (1934), 274–276.

[6] P. Turán, *On a theorem of Hardy and Ramanujan*, J. London Math. Soc. 51-9 (1934). 274.

[7] L. Wasserman, *All of statistics. A concise course in statistical inference.*, Springer, New York, 2004.