Azar y aritmética

Un capítulo de la teoría probabilística de números

Harald Andrés Helfgott

H. A. Helfgott, School of Mathematics, University of Bristol, Bristol, BS8 1TW, United Kingdom

E-mail address: harald.helfgott@gmail.com

2000 *Mathematics Subject Classification.* Primary 11K99

RESUMEN. Sea $\omega(n)$ el número de divisores primos de un entero n. Sea n un entero tomado al azar entre 1 y N. Qué se puede decir del valor que entonces tomará $\omega(n)$? Cuál es su esperanza? Cuál es su distribucion en el limite? Cuál es la probabilidad que $\omega(n)$ tome valores que se alejen mucho de su esperanza?

Estudiamos estas preguntas a guisa de introducción a la teoría de números probabilística. Trataremos varios tópicos centrales de la teoría de probabilidades sin suponer conocimientos previos en el área. No asumiremos ni teoría de la medida ni análisis complejo. En los ejercicios, entre otros tópicos, se desarrollarán las bases de la teoría de cribas como una aplicación de ideas probabilísticas.

Al miaj gepatroj

Índice general

Capítulo 1. Los divisores primos 1
1.1. La esperanza 1
1.2. La varianza 7
1.3. El límite central 18
1.4. Grandes desviaciones: cotas superiores. Valores críticos. 31
1.5. Grandes desviaciones: cotas inferiores. Entropía. 38

Apéndice A. Rudimentos de probabilidades 53

Apéndice B. Comentarios finales 59

Bibliografía 61

Índice alfabético 63

Prefacio

Este es un estudio de los factores primos de un número tomado al azar. El objeto principal es servir de introducción a la teoría probabilística de números y, al mismo tiempo, a varios temas centrales en la teoría de las probabilidades en general: la varianza, el límite central, las grandes desviaciones, la entropía.

La historia de la teoría de números probabilística comienza con Hardy y Ramanujan [**5**], quienes fueron los primeros en analizar el tema central de este libro: la distribución del número $\omega(n)$ de divisores primos de un número entero aleatorio n. En el curso de la generación siguiente – notablemente con el teorema de Erdös y Kac [**4**], el cual estudiaremos en la sección 1.3 – se fueron asimilando conceptos y técnicas de la teoría de probabilidades en general al estudio incipiente del tema. El área ha seguido desarrollándose hasta nuestros días, gracias tanto a especialistas en teoría de números como a probabilistas.

No asumiremos ningún conocimiento de análisis complejo ni de teoría de la medida. Al final de cada sección, se encontrará una serie de notas y problemas; esencialmente se trata de ejercicios guiados o esbozos de pruebas a seguir y completar con lápiz y papel. Entre otros tópicos, las notas de fin de sección desarrollan las bases de la teoría de cribas, tanto como una aplicación de conceptos probabilísticos, como para uso en el texto principal. Mi objetivo ha sido dar las pruebas que me parecen ser las más naturales, antes que las más conocidas.

El texto presente está basado en las notas de clase de un curso que dicté en Julio y Agosto de 2007 bajo los auspicios del IMCA (Instituto de Matemática y Ciencias Afines) en Lima, Perú. Agradezco tanto al IMCA como a la Universidad Mayor de San Marcos por su hospitalidad.

Notación

Sean d y n números enteros. Escribimos $d|n$ cuando queremos decir que d divide a n exactamente, es decir, sin dejar resto: $3|6$, $5|15$, $1|n$ para todo n. Escribimos $d \nmid n$ cuando d no divide a n, es decir, cuando la division de n por d deja resto: $4 \nmid 6$, $7 \nmid 15$, $(n+1) \nmid n$ para todo n.

La letra p siempre designará a un número primo. La función $\Lambda(n)$ (función de von Mangoldt) se define como sigue:

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^\alpha \text{ para algún primo } p \text{ y algún entero } \alpha > 0 \\ 0 & \text{si no es así.} \end{cases}$$

Denotamos por $\lfloor x \rfloor$ el máximo entero n que no sea mayor que x. Por ejemplo, $\lfloor 2{,}75 \rfloor = 2$, $\lfloor 7 \rfloor = 7$, $\lfloor \pi \rfloor = 3$.

Cuando decimos "logaritmo" o escribimos $\log x$, tenemos siempre en mente al logaritmo en base e, a menos que otra base se especifique explicitamente ("logaritmo en base 2", por ejemplo). Al contrario de los escritores franceses, utilizaremos la notación $\log_2 x$ para el logaritmo base 2 de x, y $\log \log x$ para el logaritmo (base e) del logaritmo (base e) de x.

Utilizaremos la notación O, o de Landau dadas dos funciones f, g, (a) se escribe $f(x) = O(g(x))$ cuando existen constantes $c_1, c_2 > 0$ tales que $|f(x)/g(x)| < c_1$ para todo $x > c_2$; (b) se escribe $f(x) = o(g(x))$ cuando $\lim_{x\to\infty} f(x)/g(x) = 0$. Está claro que $f(x) = o(g(x))$ implica $f(x) = O(g(x))$, pero no viceversa. Ejemplos: $x^2 = O(x^3)$, $x^2 = o(x^3)$, $f(x) = O(f(x))$ (para todo f), $\sin x = O(1)$, $x \sin x = O(x)$, $\sum_{n\leq x} 1/n = O(\log n)$, $\sum_{n\leq x}(-1)^n/n = O(1)$, $\prod_{n\leq x}(1-1/n) = o(1)$. En particular, $f(x) = O(1)$ quiere decir que f está 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 c o δ y z, por ejemplo.

La expresión "$f(x) \ll g(x)$" es un sinónimo de "$f(x) = O(g(x))$"; la expresión "$f(x) \gg g(x)$" es un sinónimo de "$g(x) = O(f(x))$".

Escribimos $f(x) \sim g(x)$ cuando queremos decir que f es asintótica con respecto a g, i.e.,

$$\lim_{x\to\infty} f(x)/g(x) = 1.$$

Si decimos que $f(x) = o(g(x))$ (o $f(x) \sim g(x)$) "cuando $x \to 0$", queremos decir que $\lim_{x\to 0} f(x)/g(x) = 0$ (o, respectivamente, $\lim_{x\to 0} f(x)/g(x) = 1$).

Denotamos por $\mathrm{Prob}(E)$ la probabilidad del evento aleatorio E, por $\mathbb{E}(X)$ la esperanza de la variable aleatoria X y por $\mathrm{Var}(X)$ la varianza de la variable X. Ver el apéndice A.

CAPíTULO 1

Los divisores primos

1.1. La esperanza

Definición de esperanza. Ejemplos. Recordemos que, si se tiene una variable aleatoria X que toma los valores $x_1, x_2, \ldots, x_n$ con probabilidades p_1, p_2, $\ldots$, p_n, donde $p_1 + p_2 + \ldots + p_n = 1$, entonces la esperanza se define como la cantidad

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

Así, por ejemplo, si X es el valor que da un dado arrojado al aire,

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

y por lo tanto

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

Si X es un dado trucado, muy bien podría tener la distribución

$$X = \begin{cases} 1 & \text{con probabilidad } 1/18 \\ 2, 3, \ldots, 5 & \text{con probabilidad } 1/9 \text{ en cada caso} \\ 6 & \text{con probabilidad } 1/2 \end{cases}$$

y entonces

$$\mathbb{E}(X) = \sum_{i} p_i x_i = \frac{1}{18} \cdot 1 + \frac{1}{9} \cdot 2 + \ldots + \frac{1}{9} \cdot 5 + \frac{1}{2} \cdot 6 = \frac{83}{18}.$$

Esperanza y sumas. Denotamos por $\mathbb{E}(X)$ la esperanza de una variable aleatoria X. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias. Es fácil ver que

$$\mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n). \tag{1.1.1}$$

Aplicación 1. Denotemos por $\tau(n)$ el número de divisores de un entero n. Cuánto es $\tau(n)$, en promedio?

Para que nuestra pregunta tenga sentido, debemos decir como estamos escogiendo n. Fijemos un entero N. Tomamos n al azar entre 1 y N, con la distribución uniforme. Estamos preguntando cuál es el valor de $\mathbb{E}(\tau(n))$.

Para todo entero m, definimos

$$X_m = \begin{cases} 0 & \text{si } m \nmid n, \\ 1 & \text{si } m|n. \end{cases} \tag{1.1.2}$$

(Formalmente, lo que tenemos es una variable aleatoria X que toma valores n entre 1 y N con la distribución uniforme, y varias variables aleatorias X_m que dependen de X.) Ahora bien,

$$\sum X_m = \tau(n).$$

Está claro que lo que queremos calcular es $\mathbb{E}(\sum X_m)$.

Ya sabemos (1.1.1) que

$$\mathbb{E}\left(\sum_m X_m\right) = \sum_m \mathbb{E}(X_m).$$

Ahora bien, cuál es el valor de $\mathbb{E}(X_m)$? Calculamos:

(1.1.3)

$$\mathbb{E}(X_m) = \mathrm{Prob}(X_m = 1) = \frac{1}{N}\sum_{\substack{n\leq N \\ m|n}} 1 = \frac{1}{N}\left\lfloor \frac{N}{m} \right\rfloor = \frac{1}{N}\left(\frac{N}{m} + O(1)\right) = \frac{1}{m} + O\left(\frac{1}{N}\right).$$

(Aquí, como de ahora en adelante, $O(1)$ quiere decir "una cantidad x tal que $|x| \leq C$ para alguna constante C" y $O(1/N)$ quiere decir "una cantidad x tal que $|x| \leq C/N$". La ecuación $\lfloor N/m \rfloor = N/m + O(1)$ nos está diciendo simplemente que el valor absoluto de la diferencia entre $\lfloor N/m \rfloor$ y N/m es siempre menor que una constante – en verdad, menor que 1.) Por lo tanto

$$\mathbb{E}\left(\sum_m X_m\right) = \sum_m \mathbb{E}(X_m) = \sum_{m\leq N} \frac{1}{m} + \sum_{m\leq N} O\left(\frac{1}{N}\right) = \log N + O(1).$$

concluimos que

$$\mathbb{E}(\tau(n)) = \log N + O(1). \tag{1.1.4}$$

Aplicación 2. Sea $\omega(n)$ el número de divisores primos de n. Cuánto es $\omega(n)$, en promedio?

Calculemos:

$$\mathbb{E}(\omega(n)) = \mathbb{E}\left(\sum_{p\leq N} X_p\right) = \sum_{p\leq N} \mathbb{E}(X_p) = \sum_{p\leq N} \frac{1}{p} + \sum_{p\leq N} O\left(\frac{1}{N}\right).$$

Ahora bien

$$\sum_{p\leq N} \frac{1}{p} = \log\log N + O(1) \tag{1.1.5}$$

(teorema de Chebyshev-Mertens, 1875; ver los ejercicios). Por lo tanto

$$\mathbb{E}(\omega(n)) = \log\log N + O(1). \tag{1.1.6}$$

Aplicación 3. Cuántos factores primos de un tamaño dado tiene un número tomado al azar?

Precisemos el rango. Sean dados δ_0, δ_1 tales que $0 \leq \delta_0 < \delta_1 < 1$. Tomemos n entre 1 y N bajo la distribución uniforme. Queremos saber la esperanza $\mathbb{E}(Y)$ del número Y de factores primos de n entre N^{δ_0} y N^{δ_1}.

Calculemos:

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}\left(\sum_{N^{\delta_0}<p\leq N^{\delta_1}} X_p\right) = \sum_{N^{\delta_0}<p\leq N^{\delta_1}} \left(\frac{1}{p} + O\left(\frac{1}{N}\right)\right)\\ &= \log\delta_1 - \log\delta_0 + o(1).\end{aligned}$$

Podemos plantearnos una segunda pregunta: cuál es la probabilidad que n tenga por lo menos un factor primo entre N^{δ_0} y N^{δ_1}? Un número $n \leq N$ no puede tener más de $1/\delta_0$ factores primos mayores que N^{δ_0}. Por lo tanto,

$$\mathrm{Prob}(Y > 0) \leq \frac{1}{1/\delta_0}\mathbb{E}(Y) \leq \delta_0 \cdot (\log\delta_1 - \log\delta_0 + o(1)).$$

Esto es sólo una cota superior. Veremos más tarde como estimar $\mathrm{Prob}(Y > 0)$ de manera más precisa.

Desigualdad de Markov. Sea X una variable aleatoria que toma siempre valores no negativos. Sea $t \geq \mathbb{E}(X)$. Entonces

$$\mathrm{Prob}(X \geq t) \leq \frac{\mathbb{E}(X)}{t} \qquad \text{(desigualdad de Markov).} \tag{1.1.7}$$

Esto tiene sentido: si, en promedio, cae 1 cm de lluvia al día, la probabilidad que caigan más de 10 cm no puede ser más de 0,1. (Por otra parte, dado que cae 1 cm de lluvia al día en promedio, la probabilidad que caigan 0 cm puede ser tan cercana a 1 como se quiera: muy bien podrían haber cien años de sequía y un día de diluvio. Esto nos muestra que una desigualdad tan general como la de Markov puede valer solo para la cola superior de la distribución, no para la cola inferior.)

La prueba es sencilla: por la definición de la esperanza, tenemos

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

y por lo tanto

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

Aplicaciones. Obtenemos de manera inmediata que

$$\begin{aligned}\mathrm{Prob}(\tau(n) \geq t) &\leq \frac{\log N + O(1)}{t},\\ \mathrm{Prob}(\omega(n) \geq t) &\leq \frac{\log\log N + O(1)}{t}.\end{aligned} \tag{1.1.8}$$

Podemos mejorar la segunda cota en (1.1.8) de la manera siguiente. Es fácil ver que $\tau(n) \geq 2^{\omega(n)}$. Luego

$$\text{Prob}(\omega(n) \geq t) \leq \text{Prob}(\tau(n) \geq 2^t) \leq \frac{\log N + O(1)}{2^t}. \tag{1.1.9}$$

Qué tan mejor es esto que la segunda línea de (1.1.8)? Consideremos $t = (1+\epsilon)\log_2 \log N$. Entonces (1.1.8) nos da $\text{Prob}(\omega(n) \geq t) \leq \frac{\log 2 + o(1)}{1+\epsilon}$, mientras que (1.1.9) nos da

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

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

Por otra parte, si t está entre $\log\log N$ y $\log_2 \log N$, la desigualdad (1.1.9) no nos da nada. Esto se debe al hecho que, si bien $\tau(n) = 2^{\omega(n)}$ gran parte del tiempo, $\mathbb{E}(\tau(n)) = \log N$, mientras que $\mathbb{E}(\omega(n)) = \log\log N$; en otras palabras, $\mathbb{E}(\tau(n)) \neq 2^{\mathbb{E}(\omega(n))}$. Las colas superiores de la distribución de $\omega(n)$ cobran gran efecto cuando $\omega(n)$ se pone en el exponente, al punto que afectan considerablemente la esperanza de $\mathbb{E}(\tau(n))$ (o la de $\mathbb{E}(2^{\omega(n)})$, la cual es del mismo orden de magnitud[1]).

Tendremos la oportunidad de estimar las distribuciones de $\omega(n)$ y $\tau(n)$ con mayor precisión más tarde.

Notas y problemas

1. *Sumas por partes.* La siguiente técnica es útil a menudo; la necesitaremos inmediatamente en la prueba de Chebyshev-Mertens y una y otra vez en el futuro. Digamos que tenemos que calcular

$$\sum_{n=1}^{N} h(n),$$

donde $h(n) = (f(n+1) - f(n)) \cdot g(n)$. Entonces

$$\begin{aligned}\sum_{n=1}^{N} h(n) &= \sum_{n=1}^{N} (f(n+1) - f(n)) \cdot g(n) = \sum_{n=1}^{N} f(n+1)g(n) - \sum_{n=1}^{N} f(n)g(n)\\ &= \sum_{n=2}^{N+1} f(n)g(n-1) - \sum_{n=1}^{N} f(n)g(n)\\ &= (f(N+1)g(N) - f(1)g(1)) - \sum_{n=2}^{N} f(n)(g(n) - g(n-1)).\end{aligned}$$

Esta técnica (*sumar por partes*) es útil cuando la suma $\sum_{n=1}^{N} f(n)(g(n)-g(n-1))$ es, por algún motivo, más facil de calcular que $\sum_{n=1}^{N}(f(n+1) - f(n)) \cdot g(n)$, o ya ha sido evaluada.

[1] Es decir, es de tamaño comparable, quítese o póngase un factor constante.

Se puede ver que el proceso es análogo a la integración por partes. (Uno puede, incluso, ver a la sumación por partes como un caso especial de la integración por partes, mediante el uso de una integral de Lebesgue.)

2. Probaremos el teorema de Chebyshev-Mertens (ecuación (1.1.5)).
   a) Todo número entero positivo puede ser expresado como un producto de primos de manera única[2]. 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 $p^k|n$.
   Tome logaritmos a ambos lados y muestre que
   $$\log n = \sum_{d|n} \Lambda(d), \tag{1.1.10}$$
   donde $\Lambda(d) = \log p$ si d es una potencia p^α de un primo p, y $\Lambda(d) = 0$ si no es así (*función de von Mangoldt*).
   b) Sea X_d como antes, es decir, la variable aleatoria que toma el valor 1 cuando $d|n$ y el valor 0 cuando $d \nmid n$. Sea $Y = \sum_{d|n} \Lambda(d) X_d$. Entonces, por (1.1.10), Y siempre toma el valor $\log n$. Concluya que
   $$\mathbb{E}(Y) = \log N + O(1). \tag{1.1.11}$$
   c) Al mismo tiempo, tenemos que
   $$\mathbb{E}(Y) = \sum_{d \leq N} \Lambda(d) \mathbb{E}(X_d) = \sum_{d \leq N} \Lambda(d) \cdot \frac{1}{N} \left\lfloor \frac{N}{d} \right\rfloor,$$
   así que
   $$\sum_{d \leq N} \Lambda(d) \cdot \frac{1}{N} \left\lfloor \frac{N}{d} \right\rfloor = \log N + O(1). \tag{1.1.12}$$
   Como $\frac{1}{N} \lfloor \frac{N}{d} \rfloor = \frac{1}{d} - O\left(\frac{1}{N}\right)$, estamos a un paso de obtener una estimación de $\sum_{d \leq N} \frac{\Lambda(d)}{d}$:
   $$\begin{aligned} \sum_{d \leq N} \frac{\Lambda(d)}{d} &= \log N + O(1) + \sum_{d \leq N} \Lambda(d) \cdot O\left(\frac{1}{N}\right) \\ &= \log N + O(1) + \frac{1}{N} \cdot O\left(\sum_{d \leq N} \Lambda(d)\right). \end{aligned} \tag{1.1.13}$$

[2] Este hecho es a veces llamado el *teorema fundamental de la aritmética*. Puede parecer extraño que un enunciado tan familiar tenga un nombre tan grandilocuente; empero, el hecho que un enunciado nos sea sumamente natural no quiere decir que no deba ser probado, o que sea cierto. Hay análogos del conjunto de enteros $\mathbb{Z}$, los así llamados *anillos de enteros* de los *campos algebraicos*; en la gran mayoría de ellos, el teorema fundamental de la aritmética deja de ser cierto. (Si bien todo elemento aún se factoriza en elementos irreducibles, ya no lo hace de manera única.) Tenemos, por ejemplo, las dos factorizaciones $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ en el anillo de enteros del campo algebraico $\mathbb{Q}(\sqrt{-5})$.

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

d) Por (1.1.12), tenemos

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

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

y por lo tanto

$$\begin{aligned}\sum_{\frac{N}{2}\leq d\leq N} \Lambda(d) &\leq \sum_{d\leq N} \Lambda(d) \cdot \left(\left\lfloor \frac{N}{d} \right\rfloor - 2 \left\lfloor \frac{N}{2d} \right\rfloor \right) \\ &= N(\log N + O(1)) - N \left(\log \frac{N}{2} + O(1) \right) = O(N)\end{aligned}$$

para todo N. Por lo tanto,

$$\begin{aligned}\sum_{d\leq N} \Lambda(d) &= \sum_{\frac{N}{2}<d\leq N} \Lambda(d) + \sum_{\frac{N}{4}<d\leq \frac{N}{2}} \Lambda(d) + \sum_{\frac{N}{8}<d\leq \frac{N}{4}} \Lambda(d) + \dots \\ &= O(N) + O(N/2) + O(N/4) + \dots = O(N).\end{aligned} \tag{1.1.14}$$

(Aquí lo que hemos hecho es dividir una suma en *intervalos diádicos*, es decir, intervalos de la forma $M < d \leq 2M$; este es un procedimiento muy común en el análisis.)

e) De (1.1.13) y (1.1.14), deducimos que

$$\sum_{d\leq N} \frac{\Lambda(d)}{d} = \log N + O(1). \tag{1.1.15}$$

f) Si bien (1.1.15) ya es un resultado útil, lo que queremos en verdad es estimar la suma $\sum_{p\leq N} \frac{1}{p}$. Ahora bien, la contribución de los enteros d de la forma $d = p^\alpha$, $\alpha \geq 2$, a la suma (1.1.15) es negligible, o, más precisamente, $O(1)$. (Por qué? Porque $\sum_n \frac{\log n}{n^2}$ es convergente.) Tenemos entonces que

$$\sum_{p\leq N} \frac{\log p}{p} = \log N + O(1). \tag{1.1.16}$$

Para liberarnos del factor $\log p$, podemos hacer una suma por partes (ver la nota 1 más arriba). Utilice tal técnica (cómo?) para concluir que

$$\sum_{p\leq N} \frac{1}{p} = \sum_{2\leq n\leq N} \frac{\log n + O(1)}{n(\log n)^2} + O(1).$$

Aproximando la suma mediante una integral, muestre que

$$\sum_{2\leq n\leq N} \frac{\log n + O(1)}{n(\log n)^2} = \log \log N + O(1)$$

y por lo tanto

$$\sum_{p \leq N} \frac{1}{p} = \log\log N + O(1) \qquad (\textit{teorema de Chebyshev-Mertens}). \tag{1.1.17}$$

g) Denotemos por $\pi(N)$ el número de primos entre 1 y N. (Aquí π es la letra griega que corresponde a p, la cual es la primera letra de la palabra *primo*; no hay otra connexión con el número $\pi = 3{,}14159\ldots$.) Utilice la técnica de la suma por partes para deducir de (1.1.14) que

$$\pi(N) \ll \frac{N}{\log N}. \tag{1.1.18}$$

Éste es un resultado de Chebyshev.

h) En verdad, siguiendo un procedimiento similar al que acabamos de poner en práctica, Chebyshev probó resultados más fuertes y más precisos; en particular, mostró que

$$(\log 2)\frac{N}{\log N}(1 + o(1)) \leq \pi(N) \leq (\log 4)\frac{N}{\log N}(1 + o(1)). \tag{1.1.19}$$

Nótese que Chebyshev dio una cota inferior, no sólo una cota inferior como (1.1.18). Probar (1.1.19) puede ser un problema interesante para el lector; alternativamente, se puede consultar [**8**, §2.2], por ejemplo. Aquí nos hemos querido concentrar en derivar Chebyshev-Mertens (1.1.17) de la manera más breve posible.

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

$$\pi(N) \sim \frac{N}{\log N} \qquad \text{(teorema de los números primos)} \tag{1.1.20}$$

La mayoría de las demostraciones de (1.1.20) requieren iniciar el estudio de la función zeta de Riemann. (Ver, e.g., [**8**, §5.6].) Existen también pruebas "elementales"[3], generalmente complicadas.

Utilizaremos el teorema de los números primos (1.1.20) muy poco en estas notas, ya que Chebyshev-Mertens nos será casi siempre suficiente.

1.2. La varianza

La *varianza* de una variable aleatoria está dada por

$$\mathrm{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{1.2.1}$$

Sean X, Y dos variables independientes. Entonces

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y), \tag{1.2.2}$$

y luego

$$\mathrm{Var}(X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{1.2.3}$$

[3]En el sentido de no utilizar el análisis complejo.

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

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n). \tag{1.2.4}$$

Teorema 1.1. *(Desigualdad de Chebyshev) Para toda variable aleatoria X y todo $x > 0$,*

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

En particular, si $X = X_1 + X_2 + \ldots + X_n$, donde $X_1, X_2, \ldots, X_n$ son variables independientes en pares,

$$\mathrm{Prob}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathrm{Var}(X)}{x^2} = \frac{\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)}{x^2}. \tag{1.2.5}$$

Demostración. Utilizamos la definición (1.2.1) y la desigualdad de Markov (1.1.7):

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

Si $X = X_1 + \ldots + X_n$, utilizamos (1.2.4) para evaluar $\mathrm{Var}(X)$. □

Aplicación. Sabemos ya que, en promedio, un número $n \leq N$ tiene $\log\log N + O(1)$ factores primos. Qué tan comunes son los números que tienen muchos menos o muchos más factores primos?

Sea X_p como en (1.1.2). Ahora bien, $X_p^2 = X_p$, puesto que $0^2 = 0$ y $1^2 = 1$. Por lo tanto,

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

La ecuación (1.1.6) nos dice que $\sum_{p\leq N} \mathbb{E}(X_p) = \log\log N + O(1)$. Así,

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

Podemos concluir que, para $\omega(n) = X = \sum_{p\leq N} X_p$, la desigualdad de Chebyshev es válida con $\mathbb{E}(X) = \log\log N + O(1)$? No nos apresuremos: las variables X_{p_1}, X_{p_2}, $p_1 \neq p_2$ no son exactamente independientes. (Por ejemplo, si $p_1, p_2 \geq \sqrt{N}$, entonces X_{p_1} y X_{p_2} no pueden ser 1 simultáneamente (por qué?); esto nunca pasaría con variables verdaderamente independientes.) Nos basta, empero, que la igualdad (1.2.2) sea válida de manera aproximada. Veamos: para $p_1 \neq p_2$,

$$\mathbb{E}(X_{p_1} X_{p_2}) = \mathbb{E}(X_{p_1 p_2}) = \frac{1}{p_1 p_2} + O\left(\frac{1}{N}\right),$$

$$\mathbb{E}(X_{p_1})\mathbb{E}(X_{p_2}) = (1/p_1 + O(1/N))(1/p_2 + O(1/N)) = \frac{1}{p_1 p_2} + O\left(\frac{1}{N}\right).$$

Por lo tanto,

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

y, de la misma manera,

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

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

$$\text{Prob}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\log\log N + O(1)}{x^2} \tag{1.2.6}$$

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

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

$$\text{Prob}(|\omega(n) - \log\log n| \geq x) \leq \frac{\log\log N + O(1)}{(x + O(1))^2}. \tag{1.2.7}$$

Dicho de otra manera,

$$\text{Prob}\left(|\omega(n) - \log\log N| \geq t\sqrt{\log\log N}\right) \leq \frac{1 + O(1/\sqrt{\log\log N})}{t^2} \tag{1.2.8}$$

para todo $t \geq 1$. (La constante implícita en $O(1/\sqrt{\log\log N})$ no depende de t.) Tanto el resultado (1.2.8) como la prueba que hemos presentado se deben a Turán [**9**]; Hardy y Ramanujan habían dado antes una prueba más complicada de un resultado ligeramente más débil [**5**].

Ejemplos. Escogemos $t = 10$, y obtenemos que

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

escogemos $t = \epsilon\sqrt{\log\log N}$, y obtenemos que

$$\begin{aligned} \text{Prob}(\omega(n) > (1+\epsilon)\log\log N) &\leq \frac{1}{\epsilon^2 \log\log N} + o_\epsilon(1), \\ \text{Prob}(\omega(n) < (1-\epsilon)\log\log N) &\leq \frac{1}{\epsilon^2 \log\log N} + o_\epsilon(1) \end{aligned} \tag{1.2.9}$$

para todo $\epsilon > 0$.

Notas y problemas

1. Nuevamente nos planteamos la pregunta: cuántos primos hay entre 1 y N? Tratemos de ver si podemos atacar el problema usando la varianza, antes que la esperanza. La idea central está clara: los primos son algo que se desvían de la norma, y podemos usar la desigualdad de Chebyshev (Teorema 1.1) para obtener una cota sobre la probabilidad de eventos que se desvían de una norma. Podemos usar la desigualdad (1.2.7) (la cual hemos probado utilizando la desigualdad de Chebyshev) con $x = \log\log n$, y obtenemos

$$\text{Prob}(\omega(n) = 1) \leq \frac{\log\log N + O(1)}{(\log\log N)^2 + O(\log\log N)} = \frac{1}{\log\log N + O(1)}.$$

Por lo tanto, hay a lo más $\frac{N}{\log\log N + O(1)}$ primos entre 1 y N. Esta es una cota sumamente débil: la cota (1.1.18) era mucho mejor. Podemos utilizar la desigualdad de Chebyshev de otra manera para obtener una cota más fuerte?

Veremos que es así, 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 que método que seguiremos ahora tambien sirve para obtener cotas sobre muchas cosas aparte del número de primos entre 1 y N.

Lo que estamos haciendo es evaluar la varianza de $X = \sum_p X_p$. Al hacer tal cosa, utilizamos el hecho que las variables X_p, $p \leq N^{1/3}$ (digamos) son casi independientes en pares. Hay todavía un hecho más general que no estamos usando: para $p_1 < p_2 < \ldots < p_k$ cualesquiera tales que $p_1 p_2 \cdots p_k$ es bastante menor que N, las variables $X_{p_1}, X_{p_2}, \ldots, X_{p_k}$ son mutuamente independientes, o casi. Qué podemos hacer con esto?

a) Defina

$$Z_p = \begin{cases} 1/p & \text{si } p \nmid n, \\ -(1 - 1/p) & \text{si } p | n, \end{cases} \tag{1.2.10}$$

donde n es un entero aleatorio entre 1 y N. Verifique que $\mathbb{E}(Z_p) = O(1/N)$.

b) Para todo d sin divisores cuadrados[4], defina

$$Z_d = \prod_{p|d} Z_p. \tag{1.2.11}$$

Verifique que $\mathbb{E}(Z_d) = O(\tau(d)/N)$. Ya sabemos que $\tau(d)$ es pequeño en promedio (ver (1.1.4)).

Concluya que, si d_1, d_2 son distintos y carecen de divisores cuadrados,

$$\begin{aligned} \mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2}) &= N^{-2} \cdot O(\tau(d_1))O(\tau(d_2)), \\ \mathbb{E}(Z_{d_1} Z_{d_2}) &= N^{-1} \cdot O(\tau(d_1 d_2)) \leq N^{-1} \cdot O(\tau(d_1))O(\tau(d_2)), \end{aligned} \tag{1.2.12}$$

y por lo tanto

$$\mathbb{E}(Z_{d_1} Z_{d_2}) = \mathbb{E}(Z_{d_1})\mathbb{E}(Z_{d_2}) + N^{-1} \cdot O(\tau(d_1))O(\tau(d_2)). \tag{1.2.13}$$

c) Defina $Z = \sum_{d \leq M}^{*} Z_d$, 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.)

Podemos ver que (1.2.13) es una versión aproximada de (1.2.2); es razonable tratar de obtener una versión aproximada de (1.2.4) en consecuencia. Muestre que

$$\begin{aligned} \mathbb{E}(Z) &= O(N^{-1/2}), \\ \mathrm{Var}(Z) &= \sum_{d \leq M}^{*} \mathrm{Var}(Z_d) + O(N^{-0,01}) = \sum_{d \leq M}^{*} \mathbb{E}(Z_d^2) + O(N^{-0,01}) \end{aligned} \tag{1.2.14}$$

Muestre también que $\mathbb{E}(Z_d^2) = \frac{\phi(d)}{d^2} + O\left(\frac{\tau(d)}{N}\right)$, donde $\phi(d) = d \cdot \prod_{p|d}(1 - 1/p)$ (*función de Euler*). Por consiguiente,

$$\mathrm{Var}(Z) = \sum_{d \leq M}^{*} \frac{\phi(d)}{d^2} + O\left(N^{-0,01}\right) \ll \log M \leq \log N. \tag{1.2.15}$$

[4] Es decir, d no divisible por 4, ni por 9, ni por 16, ni por 25, ...

d) Por la desigualdad de Chebyshev,

$$\mathrm{Prob}(|Z - \mathbb{E}(Z)| \geq x) \leq \frac{\mathrm{Var}(Z)}{x^2}. \tag{1.2.16}$$

Ahora bien, si el número n es primo y mayor que M, entonces, para cada d sin divisores cuadrados, la variable Z_d tomará el valor $\prod_{p|d} 1/p = 1/d$ (por la definición de Z_d; ver (1.2.10) y (1.2.11)). Por lo tanto, si n es primo y mayor que M,

$$Z = \sum_{d \leq M}^{*} \frac{1}{d} \gg \frac{\sum_{d \leq M} \frac{1}{d}}{\sum_m \frac{1}{m^2}} \gg \sum_{d \leq M} \frac{1}{d} \gg \log M \gg \log N,$$

donde utilizamos el hecho que la suma $\sum_m \frac{1}{m^2}$ converge.
Se infiere inmediatamente que

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

con $x = \sum_{d \leq M}^{*} 1/d - \mathbb{E}(Z) \gg \log N - O(N^{-1/2}) \gg \log N$. Por (1.2.15) y (1.2.16), concluimos que

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

y por lo tanto

$$\begin{aligned} \mathrm{Prob}(n \text{ es primo}) &\ll \frac{1}{\log N} + \mathrm{Prob}(n \leq M) \\ &= \frac{1}{\log N} + \frac{M}{N} \ll \frac{1}{\log N} \end{aligned} \tag{1.2.17}$$

para n tomado al azar entre 1 y N. En otras palabras, el número de primos entre 1 y N es $\ll \frac{N}{\log N}$.

Ésta es esencialmente la misma cota que ya obtuvimos en §1.1, Problema 2*d*. La ventaja del método presente reside en su suma flexibilidad: véase el problema siguiente.

2. Procediendo como lo hicimos en el problema anterior, probaremos que

$$\mathrm{Prob}(\text{tanto } n \text{ como } n+2 \text{ son primos}) \ll \frac{1}{(\log N)^2} \tag{1.2.18}$$

para n tomado al azar entre 1 y N.

a) Comenzamos definiendo

$$Z_p = \begin{cases} 2/p & \text{si } p \nmid n \text{ y } p \nmid n+2, \\ -(1 - 2/p) & \text{si } p|n \text{ o } p|n+2, \end{cases} \tag{1.2.19}$$

donde n es un entero aleatorio entre 1 y N, 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.

b) Podríamos definir, como antes, $Z = \sum_{d\leq M}^{*} Z_d$, donde $M = N^{\frac{1}{2}-\epsilon}$. (El asterisco $*$ en la suma $\sum_{d\leq M}$ denota que la suma recorre solo a los d sin divisores cuadrados.) Esto podría dar resultados. Empero, tenemos el derecho de definir $Z = \sum_{d\leq M}^{*} c_d Z_d$, para c_d's arbitrarios; hacemos esto, y afrontamos la tarea de encontrar los c_d que nos den el mejor resultado.
(Esta tarea de optimización nos dará una mejora cuantitativa, antes que cualitativa; podríamos obtener (1.2.18) sin este paso. Por suerte, ciertos cálculos finales nos serán más simples de esta manera que si escogieramos $c_d = 1$.)
Como antes, la idea es usar

$$\mathrm{Prob}(|Z| \geq x) \leq \frac{\mathbb{E}(Z^2)}{x^2},$$

donde x es el valor que Z toma cuando n y $n+2$ son ambos primos.
Muestre que, cuando n y $n+2$ son ambos primos,

$$Z = \sum_{d\leq M}^{*} c_d \frac{\tau(d)}{d}.$$

Muestre también que

$$\begin{aligned} \mathbb{E}(Z^2) &= \sum_{d\leq M}^{*} c_d^2 \mathbb{E}(Z_d^2) + O\left(N^{-1}\left(\sum_{d\leq M}^{*} |c_d|\tau(d)^3\right)^2\right) \\ &= \sum_{d\leq M}^{*} c_d^2 \frac{\tau(d)}{d} \prod_{p|d}\left(1-\frac{2}{p}\right) + O\left(N^{-1}\left(\sum_{d\leq M}^{*} |c_d|\tau(d)^3\right)^2\right). \end{aligned} \tag{1.2.20}$$

c) Debemos, entonces, encontrar el mínimo de

$$\frac{\sum_{d\leq M}^{*} c_d^2 \frac{\tau(d)}{d} \prod_{p|d}\left(1-\frac{2}{p}\right)}{\left(\sum_{d\leq M}^{*} c_d \frac{\tau(d)}{d}\right)^2} \tag{1.2.21}$$

La preguntas es: como escogemos c_d de tal manera que (1.2.21) sea mínimo? O, más bien: cuál es el mínimo valor tomado por (1.2.21)?
Para a_d, b_d cualesquiera,

$$\left(\sum_d a_d b_d\right)^2 \leq \sum_d a_d^2 \cdot \sum_d b_d^2 \qquad \text{(desigualdad de Cauchy)} \tag{1.2.22}$$

con igualdad sólo cuando $\vec{a}$ y $\vec{b}$ son proporcionales, i.e., cuando hay algún r tal que $a_d = rb_d$ para todo d (o $a_d = 0$ para todo d). (Prueba de la desigualdad de Cauchy: tenemos $\sum_{d<d'}(a_d b'_d - a_{d'} b_d)^2 \geq 0$ con igualdad sí y sólo sí $a_d b'_d - a_{d'} b_d = 0$ para todo par d, d', lo cual a su vez ocurre si y sólo si $\vec{a}$ y $\vec{b}$ son proporcionales. Expanda $\sum_{d<d'}(a_d b'_d - a_{d'} b_d)^2$, pase todos los términos

negativos al lado derecho de la desigualdad $\sum_{d<d'}(a_d b'_d - a_{d'} b_d)^2 \geq 0$, y sume $\sum a_d^2 b_d^2$ a cada lado.)

La desigualdad de Cauchy no es sino la familiar afirmación 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) que (1.2.22) se vuelve una igualdad cuando $a_d = b_d$ para todo d. La desigualdad de Cauchy sólo cumple el rol de asegurarnos que estamos procediendo de la mejor manera posible (en este paso).

La expresión (1.2.21) es igual a $\frac{\sum^*_{d\leq M} a_d^2}{\left(\sum^*_{d\leq M} a_d b_d\right)^2}$ con

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

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

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

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

Tenemos, entonces – utilizando (1.2.20) –

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

El término $O\left(\frac{M(\log N)^A}{N}\right)$ es negligible ($\ll N^{-1/2}$). Ahora bien,

$$\begin{aligned}\sum^*_{d\leq M}\frac{\tau(d)}{d}\prod_{p|d}\left(1-\frac{2}{p}\right)^{-1} &= \sum^*_{d\leq M}\frac{\tau(d)}{d}\prod_{p|d}\left(1+\frac{2}{p}+\frac{2^2}{p^2}+\frac{2^3}{p^3}+\dots\right)\\ &\geq \sum_{d\leq M}\frac{\tau(d)}{d} \geq \left(\sum_{d\leq M^{1/2}}\frac{1}{d}\right)^2\\ &\sim (\log M^{1/2})^2 \gg (\log N)^2.\end{aligned}$$

Por lo tanto,

$$\mathrm{Prob}(n \text{ y } n+2 \text{ son primos}) \ll \frac{1}{(\log N)^2}. \tag{1.2.23}$$

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

La primera prueba del resultado (1.2.18) fue dada por V. Brun (1920).

3. Se sabe que Prob(n es primo) $\sim \frac{1}{\log N}$; tal aseveración no es sino el teorema de los números primos (Hadamard – de la Vallée-Poussin, 1896), para el cual no se conoce una demostración simple. Se cree que

$$\text{Prob(tanto } n \text{ como } n+2 \text{ son primos)} \sim \frac{c_2}{(\log N)^2} \tag{1.2.24}$$

donde

$$c_2 = 2\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2} \sim 1{,}32032\ldots$$

Empero, esta conjetura sigue sin probarse; no se sabe siquiera si es que hay un número infinito de primos n tales que $n+2$ sea también primo (*conjetura de los primos gemelos*). (El enunciado (1.2.24) 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.)
4. Veamos ahora una bonita aplicación de los resultados de esta sección; tanto el resultado como la prueba se deben a Erdős [**3**]. El resultado es el siguiente: sólo una proporción $o(1)$ de los enteros $\leq N^2$ pueden expresarse como un producto $a \cdot b$ con $a, b \leq N$. (Cuando decimos una proporción $o(1)$ (o, coloquialmente, "proporción 0") de los enteros $\leq N^2$, queremos decir "$o(N^2)$ enteros $\leq N^2$".) Probemos este resultado.
   a) Sea ϵ un número pequeño (digamos $\epsilon = 1/10$). Entonces, por (1.2.9),

   $$\text{Prob}(\omega(a) < (1-\epsilon)\log\log N) = O\left(\frac{1}{\log\log N}\right) = o_\epsilon(1),$$
   $$\text{Prob}(\omega(b) < (1-\epsilon)\log\log N) = O\left(\frac{1}{\log\log N}\right) = o_\epsilon(1),$$

   donde a y b son enteros entre 1 y N tomados al azar.
   b) Muestre que $\mathbb{E}(\omega(\text{mcd}(a,b)))$ es $O(1)$.
   c) Concluya que

   $$\text{Prob}(\omega(a\cdot b) < (2-3\epsilon)\log\log N) = o_\epsilon(1).$$

   d) Nuevamente por (1.2.9),

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

   para n tomado al azar entre 1 y N. Tenemos, entonces, que hay a lo más $o(N^2)$ pares de enteros $a, b \leq N$ tales que $\omega(a\cdot b) < (2-3\epsilon)\log\log N$, y a lo más $o(N^2)$ enteros $n \leq N^2$ tal que $\omega(n) > (1+\epsilon)\log\log N^2$. Ahora bien, si $n = a\cdot b$ (y ϵ es pequeño y N es grande), se debe dar o lo uno o lo otro (puesto que $\log\log N^2 = \log\log N + \log 2$). Concluimos que hay a lo más $o(N^2)$ enteros n entre 1 y N^2 tales que

   $$n = a\cdot b$$

   para algún par de enteros $1 \leq a, b \leq N$. Esto es lo que queríamos demostrar.

Examinaremos este problema en más detalle cuando sepamos cómo obtener un término de error y darle un significado.

5. *Los números desmenuzables. Momentos variables.* Se dice que un número es *desmenuzable* si sólo tiene factores primos pequeños. Cuán comunes son los números desmenuzables? Esto es: cuántos números n de tamaño N tienen sólo factores primos $\leq N^{1/u}$, donde $u \to \infty$ cuando $N \to \infty$? (Por ejemplo: cuántos números de tamaño N tienen sólo factores primos $\leq N^{\frac{1}{\log\log N}}$?)

   a) Consideremos $N < n \leq 2N$, $z = N^{1/u}$ y $P(z) = \prod_{p \leq z} p$. Para simplificar las cosas, comenzaremos considerando sólo enteros n sin factores cuadrados; en otras palabras, nuestra meta inicial es acotar la probabilidad que un número n tomado al azar entre N y $2N$ no tenga factores cuadrados ni factores primos $> N^{1/u}$.
   Diremos que n es *u-desmenuzable* si no tiene factores primos $> N^{1/u}$. Si n no tiene factores cuadrados, n será u-desmenuzable si y sólo si

$$\gcd(n, P(z)) > N. \tag{1.2.25}$$

   Como $P(z) = \prod_{p \leq z} p$ y $n = \prod_{p|n} p$ son productos, y como preferimos trabajar con sumas, sacamos logaritmos en (1.2.25):

$$\log \gcd(n, P(z)) > \log N.$$

   Ahora bien, $\log \gcd(n, P(z)) = \sum_{p \leq z} (\log p) \cdot X_p$, donde $X_p = \begin{cases} 1 & \text{si } p|n, \\ 0 & \text{si } p \nmid n. \end{cases}$
   Por lo tanto, nuestra tarea es acotar

$$\mathrm{Prob}(X > \log N), \tag{1.2.26}$$

   donde $X = \sum_{p \leq z} (\log p) \cdot X_p$ y n es tomado al azar entre N y $2N$.

   b) Podemos acotar $\mathrm{Prob}(X > \log N)$ mediante Markov, o mediante Chebyshev, que no es sino Markov aplicado a X^2, o a $(X - \mathbb{E}(X))^2$; de la misma manera, podemos acotar $\mathrm{Prob}(X > \log N)$ mediante la desigualdad de Markov aplicada a X^k, para un k positivo de nuestra elección.
   Para cualquier variable aleatoria X y cualquier número par $k > 0$ (o para cualquier número $k > 0$ y cualquier variable aleatoria X que tome sólo valores no negativos),

$$\mathrm{Prob}(X > t) \leq \frac{\mathbb{E}(X^k)}{t^k}. \tag{1.2.27}$$

   Esto no es sino Markov aplicado a X^k. Las desigualdades de Markov y Chebyshev son los casos $k = 1$ y $k = 2$ de (1.2.27). A la utilización de (1.2.27) para k general se le llama *acotación por momentos.* (La expresión $\mathbb{E}(X^k)$ es llamada *un momento.*)

   c) Usaremos (1.2.27) para estimar (1.2.26). Escogeremos k al final; no será una constante, sino una función de N. Veamos:

$$\mathrm{Prob}(X > \log N) \leq \frac{\mathbb{E}(X^k)}{(\log N)^k},$$

donde $X = \sum_{p\leq z} (\log p) \cdot X_p$. Proseguimos:

$$\begin{aligned}\mathbb{E}(X^k) &= \mathbb{E}\left(\sum_{p_1,\ldots,p_k\leq z} (\log p_1)\cdots(\log p_k)\cdot X_{p_1}\cdots X_{p_k}\right)\\ &= \sum_{p_1,\ldots,p_k\leq z} (\log p_1)\cdots(\log p_k)\cdot \mathbb{E}\left(X_{p_1}\cdots X_{p_k}\right)\\ &= \sum_{p_1,\ldots,p_k\leq z} (\log p_1)\cdots(\log p_k)\cdot \mathbb{E}\left(X_{p_1'}\cdots X_{p_l'}\right),\end{aligned}$$

donde $p_1' < \ldots < p_l'$ son los primos distintos entre $p_1, \ldots, p_k$. (Por ejemplo, si $k = 4$ y $p_1 = 3$, $p_2 = 2$, $p_3 = 7$, $p_4 = 2$, entonces $l = 3$ y $p_1' = 2$, $p_2' = 3$, $p_3' = 7$.) Sabemos que

$$\mathbb{E}(X_{p_1}' \ldots X_{p_l}') = \frac{1}{N}\left(\left\lfloor\frac{2N}{m}\right\rfloor - \left\lfloor\frac{N}{m}\right\rfloor\right) \leq \frac{1}{N}\left\lfloor\frac{2N}{m}\right\rfloor \leq \frac{2}{m}, \tag{1.2.28}$$

donde $m = p_1'p_2'\ldots p_l'$. (Esta puede parecer una cota muy mala, pero, en este problema, es mejor para nuestra salud que $\leq \frac{1}{m} + \frac{1}{N}$.)
Concluimos que

$$\mathbb{E}(X^k) = \sum_{p_1,\ldots,p_k\leq z} \frac{2(\log p_1)(\log p_2)\cdots(\log p_k)}{p_1'p_2'\cdots p_l'} \tag{1.2.29}$$

d) Para estimar la suma (1.2.29), tenemos que estimar cuantas veces los primos distintos $p_1' < p_2' < \cdots < p_l'$ aparecen disfrazados de $p_1, p_2, \ldots, p_k$. Hay l^k maneras de colorear k objetos con l colores. Así,

$$\mathbb{E}(X^k) \leq 2\sum_{l=1}^{k} l^k \sum_{p_1'<\cdots<p_l'\leq z} (\log z)^{k-l}\frac{(\log p_1')\cdots(\log p_l')}{p_1'\cdots p_l'}.$$

Ahora bien,

$$\sum_{p_1'<\ldots<p_l'\leq z} \frac{(\log p_1')\cdots(\log p_l')}{p_1'\cdots p_l'} \leq \frac{1}{l!}\left(\sum_{p\leq z}\frac{\log p}{p}\right)^l$$

Sabemos que $\sum_{p\leq z} \frac{\log p}{p} = \log z + O(1)$ (ver (1.1.16)). Entonces

$$\mathbb{E}(X^k) \leq 2\cdot\sum_{l=1}^{k}\frac{l^k}{l!}\cdot(\log z)^k\cdot\left(1 + O\left(\frac{1}{\log z}\right)\right)^k.$$

e) Como $\lim_{n\to\infty}(1 + 1/n)^n = e$, tenemos que $\left(1 + O\left(\frac{1}{\log z}\right)\right)^k = e^{O\left(\frac{k}{\log z}\right)}$. Queda por estimar

$$\sum_{l=1}^{k}\frac{l^k}{l!} \leq k\cdot\max_{1\leq l\leq k}\frac{l^k}{l!}.$$

Como $\log l! = l \log l - l + O(\log l)$ (fórmula de Stirling), tenemos que

$$\frac{l^k}{l!} = O(l) \cdot \frac{l^k}{l^l e^{-l}} = O(l) e^{l + k \log l - l \log l}.$$

Verifique que, para k fijo, la función $l + k \log l - l \log l$ llega a su máximo cuando l es la solución a $k = l \log l$. Así,

$$\mathbb{E}(X^k) \ll (\log z)^k e^{O(k/\log z)} \cdot O(k^2) e^{l + k \log l - k},$$

donde l es la solución a $k = l \log l$.

f) Aplicamos (1.2.27):

$$\text{Prob}(X > \log N) \ll e^{O(k/\log z)} O(k^2) \cdot \left(\frac{\log z}{\log N} \right)^k e^{l + k \log l - k}. \tag{1.2.30}$$

Estimaremos $e^{O(k/\log z)} k^2$ al final; nuestra tarea ahora es encontrar el valor de k para el cual

$$\left(\frac{\log z}{\log N} \right)^k e^{l + k \log l - k} = \frac{e^{l + k \log l - k}}{u^k} \tag{1.2.31}$$

es mínimo (donde l es la solución a $k = l \log l$). Muestre que

$$\frac{d}{dk}(l + k \log l - k - k \log u) = \log l - \log u.$$

Por lo tanto, el mínimo se encuentra cuando $l = u$, es decir, cuando $k = u \log u$. Nótese que k no es una constante; por ello hablamos de *momentos variables*. Quizás resulte algo sorprendente que el valor óptimo de k es $u \log u$, puesto que esto es mayor que u (y, así, $z^k > N$; es por esto que escogimos la cota (1.2.28)).

g) Escogemos, entonces, $k = u \log u$. Obtenemos, por (1.2.30) y (1.2.31),

$$\text{Prob}(X > \log N) \ll e^{O\left(u \frac{\log u}{\log z}\right)} O(u \log u)^2 \cdot e^{u - u \log u}. \tag{1.2.32}$$

Concluimos que, si $z \to \infty$ y $u = \frac{\log N}{\log z} \to \infty$ cuando $N \to \infty$,

$$\text{Prob}(n \text{ es } u\text{-desmenuzable y carece de divisores cuadrados}) \ll u^{-u(1+o(1))}, \tag{1.2.33}$$

donde n es tomado al azar entre N y $2N$.

h) La desigualdad (1.2.33) es todo lo que necesitaremos en nuestra aplicación más importante. Empero, es válido preguntarse que pasa si se retira la restricción que n carezca de divisores cuadrados.

Usando (1.2.33), vemos que

$$\begin{aligned}
&\text{Prob}(n \text{ es } u\text{-desmenuzable})\\
&\quad = \text{Prob}\left(n \text{ es desmenuzable},\ k^2|n \text{ y } \frac{n}{k^2} \text{ es libre de factores cuadrados}\right)\\
&\quad = \sum_{1\le k\le K}\left(\frac{1}{k^2}+O\left(\frac{1}{N}\right)\right)\left(\frac{\log N/k^2}{\log z}\right)^{-\left(\frac{\log N/k^2}{\log z}\right)(1+o(1))} \qquad (1.2.34)\\
&\quad + O\left(\sum_{K<k\le\sqrt{N}}\left(\frac{1}{k^2}+\frac{1}{N}\right)\right)
\end{aligned}$$

para todo K. Ahora bien, para todo $\epsilon > 0$ y todo $k \le N^\epsilon$,

$$\begin{aligned}
\left(\frac{\log N/k^2}{\log z}\right)^{-\left(\frac{\log N/k^2}{\log z}\right)(1+o(1))} &= \left(\frac{\log N}{\log z}(1+O(\epsilon))\right)^{-\left(\frac{\log N}{\log z}\right)(1+O(\epsilon))(1+o(1))}\\
&= \left(\frac{\log N}{\log z}\right)^{-\left(\frac{\log N}{\log z}\right)(1+O(\epsilon))}
\end{aligned}$$

suponiendo que $z \le \frac{1}{2}N^{1-2\epsilon}$, digamos. Por lo tanto, haciendo que $K = N^\epsilon$, obtenemos de (1.2.34) que

$$\text{Prob}(n \text{ es } u\text{-desmenuzable}) = u^{-u(1+O(\epsilon))} + O\left(\frac{1}{N^\epsilon}\right).$$

Si (digamos) $u = (\log N)^{f(N)}$, donde $f(N) \to \infty$ cuando $N \to \infty$, entonces $N^{-\epsilon} = O(u^{-u})$. Haciendo que $\epsilon \to 0$, obtenemos

$$\text{Prob}(n \text{ es } u\text{-desmenuzable}) = u^{-u(1+O(\epsilon))} + O(u^{-u}) = u^{-u(1+o(1))} \qquad (1.2.35)$$

bajo la condición que $z \ge (\log N)^{f(N)}$ para alguna función f que satisfaga $f(N) \to \infty$ cuando $N \to \infty$.

Una condición de ese tipo, i.e., una cota inferior $z \ge$ para z, es en verdad necesaria para que (1.2.35) sea cierto. Por ejemplo, pruebe que, si z es una constante, Prob(n es u-desmenuzable) es mucho mayor que $u^{-u(1+o(1))}$.

1.3. El límite central

Ya conocemos $\mathbb{E}(\omega(n))$ y $\text{Var}(\omega(n))$ para n un número tomado al azar entre 1 y N, donde N es grande. Quisieramos saber, de una vez por todas, cuál es la distribución de $\omega(n)$, en el límite $N \to \infty$.

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 de Moivre (1718): si algo es la suma de muchas pequeñas cosas que nada o poco tienen que ver entre si, este algo tendrá una distribución en forma de campana. Antes de probar tal aseveración, debemos ponerla en forma precisa.

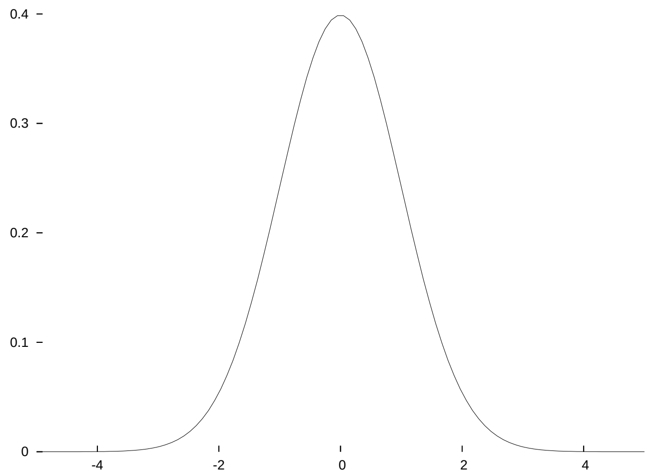


FIGURA 1. La distribución normal $f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. Este es el límite central de la distribución de $\omega(n)$: para n tomado al azar entre 1 y N, la probabilidad $\mathrm{Prob}(\omega(n) \leq \log\log N + t\sqrt{\log\log N})$ tiende a $\int_{-\infty}^{t} f(x)dx$.

Teorema 1.2 (Teorema del límite central)**.** *Sean $X_1, X_2, X_3, \ldots$ variables aleatorias mutuamente independientes. Asumamos que todas tienen la misma distribución; sea su esperanza E y su varianza V. Asumamos también que $\mathbb{E}(X_j^k)$ es finita para todo $k \geq 0$. Entonces $\frac{1}{\sqrt{nV}} \sum_{i=1}^{n} (X_i - E)$ tiende en distribución a*

$$\frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{1.3.1}$$

cuando $n \to \infty$.

La distribución dada por la función de densidad (1.3.1) es la afamada *distribución normal* (ver figura 1).

Antes de comenzar la demostración del teorema, recordemos que la *transformada de Fourier* $\hat{f} : \mathbb{R} \to \mathbb{C}$ de una función $f : \mathbb{R} \to \mathbb{C}$ se define como sigue:

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

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 $\hat{f}(t)$ resulta ser un múltiplo de $f(t)$: $\hat{f}(t) = \sqrt{2\pi} f(t)$. (La prueba está en las notas al final de esta sección.) Esta propiedad de la función $f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ será utilizada de manera crucial hacia el fin de la prueba siguiente.

DEMOSTRACIÓN. Podemos asumir sin pérdida de generalidad que $E = 0$ y $V = 1$.

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

$$\widehat{X}(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f(x)dx, \tag{1.3.2}$$

donde f es la función de densidad de X; si X es discreta,

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

En (1.3.2) vemos de manera especialmente clara que $\widehat{X}$ no es sino una transformada de Fourier.)

Tenemos

$$\widehat{X}(0) = \mathbb{E}(e^{i\cdot 0\cdot X}) = \mathbb{E}(1) = 1,$$

y, como

$$\widehat{X}'(t) = \mathbb{E}(iXe^{itX}),$$

se obtiene que $\widehat{X}'(0) = \mathbb{E}(iXe^0) = i\mathbb{E}(X)$. De la misma manera, $\widehat{X}''(0) = -\mathbb{E}(X^2)$, $\widehat{X}'''(0) = -i\mathbb{E}(X^3)$, etc.

Sea ahora X cualquiera de las variables X_i. Entonces $\widehat{X}(0) = 1$, $\widehat{X}'(0) = E = 0$, $\widehat{X}''(0) = -V = -1$, $|\widehat{X}'''(t)| = |\mathbb{E}(-iX^3e^{itX})| \leq |\mathbb{E}(|X|^3)| \leq 1 + |\mathbb{E}(X^4)| < \infty$. Ahora bien, si una función $f(t)$ es derivable $k+1$ veces alrededor del origen $t = 0$, y $f^{(k+1)}(t)$ está acotada por una constante c cuando t está cerca del origen, entonces

(1.3.3)

$$f(t) = f(0) + f'(0)\cdot t + \cdots + \frac{f^{(k)}(0)}{k!}t^k + O(c\cdot t^{k+1}) \qquad \text{(serie de Taylor truncada)}$$

cuando $t \to 0$. (Así como, cuando escribimos "$O(f(n))$ cuando $n \to \infty$" queremos decir "entre $-C\cdot f(n)$ y $C\cdot f(n)$ cuando n es mayor que una constante", escribimos, similarmente, "$O(f(t))$ cuando $t \to \infty$" cuando queremos decir "entre $-C\cdot f(t)$ y $C\cdot f(t)$ cuando $|t|$ es menor que una constante"; de la misma manera, así como "$o(f(n))$ cuando $n \to \infty$" quiere decir "entre $-g(n)\cdot f(n)$ y $g(n)\cdot f(n)$, donde g es alguna función con $\lim_{n\to\infty} g(n) = 0$", escribimos "$o(f(t))$ cuando $t \to 0$" cuando queremos decir "entre $-g(t)\cdot f(t)$ y $g(t)\cdot f(t)$, donde $\lim_{t\to 0} g(t) = 0$".) Como $c\cdot t \to 0$ cuando $t \to 0$, vemos por (1.3.3) que

$$f(t) = f(0) + f'(0)\cdot t + \cdots + \frac{f^{(k)}(0)}{k!}t^k + o(t^k).$$

Por lo tanto,

$$\widehat{X}(t) = 1 - \frac{t^2}{2} + o(t^2)$$

cuando $t \to 0$.

Para cualquier $r \neq 0$ y cualquier función f, la transformada de Fourier $\widehat{\frac{1}{r}f}$ de $\frac{1}{r}f$ satisface $\widehat{\frac{1}{r}f}(t) = \widehat{f}\left(\frac{t}{r}\right)$ (por qué?). Por lo tanto, para t fijo,

$$\widehat{\frac{1}{\sqrt{n}}X}(t) = \widehat{X}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right) \tag{1.3.4}$$

cuando $n \to \infty$. (Fíjense en "t fijo" y "$n \to \infty$"; hemos pasado a nuestro uso habitual de $o(\cdot)$.)

Cuando se tienen dos variables independientes X, Y, la variable $X+Y$ tiene como distribución la convolución de las distribuciones (por qué?). Ahora bien, la transformada de Fourier $\widehat{f*g}$ de la convolución $f*g$ de dos funciones es igual a $\widehat{f}\cdot\widehat{g}$ (por qué?). En consecuencia, $\widehat{X+Y}=\widehat{X}\cdot\widehat{Y}$. Repetiendo el proceso, obtenemos que la función característica de $X_1+X_2+\ldots+X_n$ es $\widehat{X_1}\cdot\widehat{X_2}\cdots\widehat{X_n}$.

Estamos considerando $S_n=\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j$. Vemos ahora que, por (1.3.4), su función característica debe ser

$$\widehat{S_n}=\left(\widehat{X}\left(\frac{t}{\sqrt{n}}\right)\right)^n=\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n \tag{1.3.5}$$

cuando $n\to\infty$.

Para ϵ pequeño, tenemos $(1+\epsilon)=e^{\epsilon+O(\epsilon^2)}$. Por lo tanto,

$$\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{2n}\right)\right)^n=\left(\left(1+\frac{1}{n}\right)^{-\frac{t^2}{2}(1+o(1))}\right)^n=\left(\left(1+\frac{1}{n}\right)^n\right)^{-\frac{t^2}{2}(1+o(1))}$$

cuando $n\to\infty$. Ahora bien,

$$\left(1+\frac{1}{n}\right)^n\to e$$

cuando $n\to\infty$. Así, concluimos que

$$\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{2n}\right)\right)^n\to e^{-t^2/2} \tag{1.3.6}$$

cuando $n\to\infty$. Por (1.3.5) y (1.3.6), obtenemos finalmente que

$$\widehat{S_n}\to e^{-t^2/2}$$

cuando $n\to\infty$. (Puede que la velocidad de convergencia dependa de t, pero esto no nos importa; el resultado que estamos por utilizar es robusto en ese sentido.)

La transformada $\widehat{W}$ de la normal $W=\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ es precisamente $e^{-t^2/2}$ (ver las notas). Tenemos, entonces, que, para cada t, $\widehat{S_n}(t)$ tiende a $\widehat{W}(t)$ cuando $n\to\infty$. Invocamos un resultado del análisis (*teorema de convergencia de Lévy*; ver notas) y concluimos que S_n tiende a W cuando $n\to\infty$. □

La idea central de la siguiente prueba alternativa nos será de utilidad cuando examinemos $\omega(n)$. El método es llamado *método de momentos*.

Esbozo de otra demostración. Compararemos los *momentos*

$$\mathbb{E}(S_n),\ \mathbb{E}(S_n^2),\ \mathbb{E}(S_n^3),\ldots$$

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 $\mathbb{E}(S^k)=(k-1)(k-3)(k-5)\cdots3\cdot1$ 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 que $\mathbb{E}(S_n^k)=(k-1)(k-3)(k-5)\cdots3\cdot1+o_k(1)$ para k par, y $\mathbb{E}(S_n^k)=o_k(1)$ para k impar (ver el problema 3).

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 (nota 4). $\square$

Condiciones del teorema del límite central. Hemos asumido tres cosas acerca de las variables X_j: (a) que son mutuamente independientes, (b) que tienen la misma distribución, (c) que, para cada k, $\mathbb{E}(X_j^k)$ está acotada independientemente de j (lo cual es lo mismo que $\mathbb{E}(X_j^k) < \infty$, si (b) se cumple).

Tanto (b) como (c) pueden relajarse; la *condición de Lindeberg* funge por las dos (nota 5). Es más difícil prescindir de (a); hay algunas herramientas estándar para tal tarea, pero ninguna cubre todos los casos que aparecen en la práctica.

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; luego, veamos como, en muchos otros casos, la falta de (b) (y de la condición de Lindeberg) hace que la conclusión sea falsa – es decir, que el límite no sea normal.

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

$$X'_p = \begin{cases} 1 & \text{con probabilidad } 1/p \\ 0 & \text{con probabilidad } 1 - 1/p. \end{cases} \tag{1.3.7}$$

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

Entonces $\mathbb{E}(X'_p) = \frac{1}{p}$, $\mathbb{E}((X'_p - \mathbb{E}(X'_p))^2) = \mathrm{Var}(X'_p) = \frac{1}{p} - \frac{1}{p^2}$, $\mathbb{E}(|X'_p - \mathbb{E}(X'_p)|^3) \leq \frac{1}{p}$. Por consiguiente (ver (1.3.3) y las líneas inmediatamente precedentes), 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 $\frac{1}{\sqrt{\log\log p}}(X'_p - \mathbb{E}(X'_p))$ es

$$1 - \frac{t^2}{2\log\log n}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \frac{O(t^3)}{p(\log\log n)^{3/2}}. \tag{1.3.8}$$

Definimos $S'_n = \frac{1}{\sqrt{\log\log n}} \sum_{p\leq n}(X'_p - \mathbb{E}(X'_p))$. Usando la regla $\widehat{X+Y} = \widehat{X}\cdot\widehat{Y}$ y (1.3.8), vemos que

$$\begin{aligned}
\widehat{S'_n} &= \prod_{p\leq n}\left(1 - \frac{t^2}{2\log\log n}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \frac{O(t^3)}{p(\log\log n)^{3/2}}\right) \\
&= \prod_{p\leq n} e^{-\frac{t^2}{2\log\log n}\cdot\left(\frac{1}{p}-\frac{1}{p^2}\right)\cdot(1+O(t/\sqrt{\log\log n}))} \\
&= e^{-(1+o_t(1))\cdot\sum_{p\leq n}\frac{t^2}{2\log\log n}\left(\frac{1}{p}-\frac{1}{p^2}\right)}
\end{aligned}$$

cuando $n \to \infty$. Por el teorema de Chebyshev-Mertens (1.1.17),

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

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

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

cuando $n \to \infty$. Por el teorema de convergencia de Levy, concluimos que S'_n converge en distribución a la normal (1.3.1).

* * *

Consideremos, en cambio, variables $X'_2, X'_3, X'_5, \ldots$ mutuamente independientes tales que

$$X'_p = \begin{cases} 1 & \text{con probabilidad } \rho_p \\ 0 & \text{con probabilidad } 1-\rho_p, \end{cases}$$

donde $\rho_2, \rho_3, \ldots$ son tales que $\sum_p \rho_p$ converge. Entonces $\sum_p \mathbb{E}(X'_p) < \infty$ y $\mathrm{Var}(\sum_p X'_p) = \sum_p \mathrm{Var}(X'_p) < \infty$. La función de distribución de $S'_n = \sum_{p\leq n} X'_p$ tenderá a un límite no normal $f(x)$. Ver el problema 6.

* * *

El límite central de $\omega(n)$. Sean, ahora, $X_2, X_3, X_5, \ldots$ variables dadas por

$$X_p = \begin{cases} 1 & \text{si } p|n, \\ 0 & \text{si } p\nmid n, \end{cases} \tag{1.3.9}$$

donde n es un entero aleatorio entre 1 y N. Como ya sabemos, $\omega(n) = \sum_{p\leq n} X_p$. Queremos probar que la distribución de $\omega(n)$ (o, más bien dicho, $\frac{1}{\sqrt{\log\log n}}(\omega(n) - \log\log n)$) 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 $p_1, p_2 < N^{1/2-\epsilon}$, $p_1 \neq p_2$ cualesquiera, y, en total, los términos de error son pequeños. Empero, 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 "de a k": k variables distintas cualesquiera entre $X_2, X_3, X_5, \ldots$ son aproximadamente independientes, por la mismas razones que ya vimos para $k = 2$. El término de error dependerá de k, y por lo tanto la tasa de convergencia de $\mathbb{E}(\omega(n)^k)$ a su límite dependerá de k; empero, al método de momentos esto no le importa (nota 4).

Pasemos esto en limpio.

Teorema 1.3 (Erdős-Kac [**4**]). *Sea n un entero tomado al azar entre* 1 *y N con la distribución uniforme. Entonces $Y_N = \frac{1}{\sqrt{\log\log N}}(\omega(n) - \log\log N)$ tiende en distribución a*

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

cuando $N \to \infty$.

La demostración que veremos se debe a Billingsley; incluye varias ideas de la prueba de Erdős y Kac (1939) y de la prueba de Halberstam (1955).

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

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

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

Ahora bien,

$$\begin{aligned}\sum_{p\leq g(N)} 1/p &= \log\log g(N) \;+\; O(1) = \log\log N + o(\sqrt{\log\log N})\\ &= \log\log N + o(\sqrt{\log\log g(N)}).\end{aligned}$$

Concluimos que

$$Y_N = (1+o(1))\cdot\frac{1}{\sqrt{\log\log g(N)}}\left(\omega(n) - \sum_{p\leq g(N)} 1/p\right) + o(1). \tag{1.3.10}$$

Definamos ahora

$$S_m = \frac{1}{\sqrt{\log\log m}}\cdot\left(\sum_{p\leq m}(X_p - 1/p)\right), \tag{1.3.11}$$

donde X_p es como en (1.3.9). Está claro que $\omega(n) = \sum_{p\leq N} X_p$, y por lo tanto $\omega(n) - \sum_{p\leq g(N)} X_p$ es igual a $\sum_{g(N)<p\leq N} X_p$. Calculamos la esperanza de esto último:

$$\begin{aligned}\mathbb{E}\left(\sum_{g(N)<p\leq N} X_p\right) &= \sum_{g(N)<p\leq N}\frac{1}{p} = \log\log N - \log\log g(N) + O(1)\\ &= o\left(\sqrt{\log\log N}\right) = o\left(\sqrt{\log\log g(N)}\right).\end{aligned} \tag{1.3.12}$$

Como X_p toma sólo valores positivos, la desigualdad de Markov nos permite deducir de (1.3.12) que

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

con probabilidad $1 - o(1)$. Concluimos (por (1.3.11) y (1.3.10)) que

$$Y_N = (1+o(1))S_{g(N)} + o(1) \tag{1.3.13}$$

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

Hasta ahora, nuestra labor ha sido sólo preparatoria: lo más que hemos hecho es truncar la suma $\sum_{p\leq N}$ y mostrar que el efecto de tal truncación es pequeño. A continuación, nuestra tarea es averiguar cuáles son los momentos de $S_{g(N)}$. Sea $k \geq 0$. Sean X'_p como en (1.3.7) 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),

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

donde d es el mínimo común múltiplo de $p_1, p_2, \ldots, p_k$. Por lo tanto,

$$\mathbb{E}(S^k_{g(N)}) = \mathbb{E}((S'_{g(N)})^k) + O_k\left(g(N)^k \cdot \frac{1}{N}\right),$$

puesto que $g(N)^k$ es el número de términos que aparecen cuando se expande $(S'_{g(N)})^k$. Como $g(N) = o_\epsilon(N^\epsilon)$, sabemos que $O\left(g(N)^k \cdot \frac{1}{N}\right) = o_k(1)$.

Ya vimos (después de (1.3.7)) que la distribución de S'_m tiende 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

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

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

Notas y problemas

1. Debemos probar que la transformada de Fourier

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

de $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ es igual a $e^{-t^2/2}$. Una de las maneras más simples es la siguiente. Tenemos que

$$\begin{aligned}\hat{f}(t) &= \int_{-\infty}^{\infty} e^{itx} \cdot \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2itx)}dx.\end{aligned}$$

Completando cuadrados,

$$\begin{aligned}\widehat{f}(t) &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{(-it)^2/2}e^{-(x-it)^2/2}dx \\ &= e^{-t^2/2} \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-it)^2/2}dx.\end{aligned}$$

Por el teorema de Cauchy en el análisis complejo[5], aplicado a la función $e^{-z^2/2}$ (analítica en todo el plano complejo), tenemos que

$$\int_{-\infty}^{\infty} e^{-(x-it)^2/2}dx = \int_{-\infty}^{\infty} e^{-x^2/2}dx \tag{1.3.14}$$

[5]El lector que no quiera utilizar el análisis complejo (el cual estamos evitando en general) puede saltarse este párrafo y ver la prueba alternativa al final de la nota presente.

Así tenemos que $\widehat{f}(t) = c \cdot e^{-t^2/2}$, donde c es la constante $c = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$. Sólo falta calcular c. Haremos esto en una de las formas más conocidas - evaluando la integral de $e^{-(x^2+y^2)}$ en el plano en dos maneras distintas.

Por una parte,

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2/2}\right) \\ &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\right)^2 \end{aligned} \tag{1.3.15}$$

Por otra parte, cambiando de coordenadas rectangulares (x, y) a coordenadas polares (r, θ), obtenemos

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy = \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2} r dr d\theta,$$

puesto que $dxdy = rdrd\theta$. Un breve cómputo muestra que

$$\int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2} r dr d\theta = 2\pi \int_{0}^{\infty} e^{-r^2/2} r dr = 2\pi \cdot 1 = 2\pi. \tag{1.3.16}$$

Comparando (1.3.15) y (1.3.16), vemos que

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Por lo tanto, $c = 1$, y, así, concluimos que $\widehat{f}(t) = e^{-t^2/2}$.

Como, en general, no estamos asumiendo ningún conocimiento del análisis complejo, es bueno indicar una prueba alternativa que no utilize el teorema de Cauchy (el cual usamos en el paso (1.3.14)). Una manera conocida es la que sigue. La función $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ puede ser descrita como la solución (necesariamente única) al problema de valor inicial dado por las siguientes condiciones: $f'(x) = -xf(x)$, $f(0) = \frac{1}{\sqrt{2\pi}}$. Tenemos

$$\begin{aligned} \widehat{f}'(t) &= \int_{-\infty}^{\infty} \left(\frac{d}{dt}(e^{itx} f(x))\right) dx = \int_{-\infty}^{\infty} ixe^{itx} f(x) dx = -i \int_{-\infty}^{\infty} (-xf(x)) e^{itx} dx \\ &= -i \int_{-\infty}^{\infty} f'(x) e^{itx} dx \end{aligned}$$

Hacemos una integración por partes, y obtenemos

$$\widehat{f}'(t) = i \int_{-\infty}^{\infty} f(x) \frac{d}{dx}(e^{itx}) dx = -t \int_{-\infty}^{\infty} f(x) e^{itx} dx = -t\widehat{f}(t).$$

También tenemos que

$$\widehat{f}(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

como probamos con anterioridad. Así, la función $g(t) = \frac{1}{\sqrt{2\pi}} \widehat{f}(t)$ satisface las condiciones $g'(t) = -tg(t)$, $g(0) = \frac{1}{\sqrt{2\pi}}$. Como f es la única función que satisface tales condiciones, concluimos que $g(t) = f(t)$. Por lo tanto,

$$\widehat{f}(t) = \sqrt{2\pi}g(t) = \sqrt{2\pi}f(t) = e^{-t^2/2},$$

como queríamos demostrar.

2. El siguiente es un resultado muy útil del análisis "suave"; la prueba involucra argumentos de compacticidad y convergencia, aparte de una transformación oportuna.

   Teorema 1.4 (Teorema de convergencia de P. Lévy)**.** *Sean $X_1, X_2, X_3, \dots$ variables aleatorias con funciones características $\widehat{X_1}, \widehat{X_2}, \widehat{X_3}, \dots$. Asumamos que, para todo real t, la sucesión $\widehat{X_1}(t), \widehat{X_2}(t), \dots$ 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 $X_1, X_2, X_3, \dots$ convergen a X en distribución.*

   La prueba puede encontrarse en [**6**, Vol. 2, §XV.3, Teorema 2].

3. Sean $X_1, X_2, X_3, \dots$ variables independientes con esperanza 0, varianza 1, y $\mathbb{E}(X_j^r)$ acotada para todo j y todo entero r entre 0 y k. Queremos estimar la esperanza de $(X_1 + X_2 + \dots + X_n)^k$. Comencemos expandiendo esta potencia en sus términos: $(X_1 + X_2 + \dots + X_n)^k = X_1^k + X_1^{k-1}X_2 + \dots + X_1X_2^2X_4^{k-4}X_n + \cdots$.
   a) Muestre que los términos donde algún X_j aparece a la potencia 1 tienen esperanza 0. (Decimos que X_j aparece a la potencia α si el término es de la forma $\dots X_j^\alpha \dots$; por ejemplo, X_j aparece a la potencia α en $X_1X_j^\alpha X_k^2$.)
   b) Muestre que hay a lo más $O_k(n^{(k-1)/2})$ términos (o $O_k(n^{k/2-1})$ si k es par) donde aparecen sólo potencias ≥ 2 y por lo menos una potencia ≥ 3.
   c) Nos quedan los términos donde toda variable que aparece, aparece a la potencia 2. Tal cosa puede ocurrir sólo cuando k es par; escribamos $k = 2l$. Muestre que cada término de la forma antedicha (es decir, cada término que contenga sólo cuadrados, e.g., $X_1^2X_3^2X_{10}^2$) ocurre exactamente $\frac{(2l)!}{2^l}$ veces.
   d) Si k es impar, concluimos que $\mathbb{E}((X_1 + X_2 + \dots + X_n)^k) = O_{k,c}(n^{(k-1)/2})$. Si k es par, concluimos que $\mathbb{E}((X_1 + X_2 + \dots + X_n)^k)$ es igual al número de términos distintos que contengan solo cuadrados, multiplicado por $\frac{(2l)!}{2^l}$, más $O_{k,c}(n^{k/2-1})$. A continuación, estimaremos este número de términos de manera indirecta.
   e) Expandamos la expresión $(x_1+x_2+\dots+x_n)^l$. Hay a lo más $O_k(n^{l-1})$ términos donde aparecen potencias > 1. Muestre que cada término donde no aparecen potencias > 1 ocurre $l!$ veces.
   f) Los términos de $(x_1 + x_2 + \dots + x_n)^l$ donde no aparecen potencias > 1 están en correspondencia uno a uno – sin contar el número de ocurrencias – con los términos de $(X_1 + X_2 + \dots + X_n)^{2l}$ que contienen sólo cuadrados.
   g) Definamos ahora $x_j = 1$ para todo j; entonces $(x_1 + x_2 + \dots + x_n)^l$ se vuelve n^l. Concluya que hay $\frac{1}{l!}n^l + O_k(n^{l-1})$ términos distintos donde no aparecen potencias > 1.
   h) Obtenemos inmediatamente que hay $\frac{1}{l!}n^l + O_k(n^{l-1})$ términos distintos en $(X_1 + X_2 + \dots + X_n)^k$ que contienen sólo cuadrados. Concluya que

$$\mathbb{E}((X_1 + X_2 + \dots + X_n)^k) = ((k-1)\cdot(k-3)\cdots 3\cdot 1)\cdot n^{k/2} + O_k(n^{k/2-1})$$

para k par, y $\mathbb{E}((X_1 + X_2 + \ldots + X_n)^k) = 0$ para k impar.

4. El método de momentos es válido gracias al resultado siguiente.

   Teorema 1.5. *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 $\mathbb{E}(X^k) < C^k$ para algún $C > 0$ y todo $k > 0$. Entonces $X_1, X_2, X_3, \ldots$ convergen en distribución a X.*

   La idea principal de la prueba es que los momentos de una distribución X determinan la serie de Taylor de $\widehat{X}(t)$ alrededor de $t = 0$. La condición $\mathbb{E}(X^k) < C^k$ asegura que la serie de Taylor alrededor de $t = 0$ tenga radio de convergencia infinito; así, la serie determina $\widehat{X}$, y, por ende, determina X. Para ver una prueba completa, consultar, por ejemplo, [**2**, §30, Teoremas 30.1 y 30.2].

   La condición $\frac{1}{k!}|\mathbb{E}(X^k)| < C^k$ se cumple para casi toda variable X "razonable". He aquí un ejemplo de un X para el cual la condición no se cumple, y, más aún, la conclusión del teorema no es cierta: sea $X = e^Y$, donde Y es una variable de distribución normal.
5. Las condiciones del teorema del límite central se pueden relajar de varias formas. La siguiente es una de las formas más comunes.

   Teorema 1.6 (Teorema del límite central – Lindeberg)**.** *Sean $X_1, X_2, X_3, \ldots$ variables aleatorias mutuamente independientes. Sean*

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

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

   $$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{s_n^2} \int_{|t|\geq \epsilon s_n} t^2 f_j(t)\,dt < \infty, \qquad \textit{(condición de Lindeberg)} \tag{1.3.17}$$

   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 $n \to \infty$.

   Si X_j es discreta, entonces, claro está, la condición de Lindeberg se escribe

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

   Esbozo de una prueba. Se procede como en la primera demostración que dimos del teorema del límite central. La condición de Lindeberg sirve para mostrar que

   $$\lim_{n\to\infty} \sum_{j=1}^{n} \left| \widehat{X_j/s_n}(t) - \left(1 - \frac{1}{2} t^2 \,\mathrm{Var}(X_j)/s_n^2\right) \right| = 0$$

para todo t. Esto nos permite mostrar que

$$\begin{aligned}\widehat{S_n/s_n} &= \prod_{j\leq n}\left(1-\frac{1}{2}t^2\operatorname{Var}(X_j)/s_n^2\right)^n + o(1)\\ &= \prod_{j\leq n} e^{-\frac{1}{2}t^2\operatorname{Var}(X_j)/s_n^2} + o(1) = e^{-t^2/2}+o(1),\end{aligned}$$

que es lo que deseamos. $\square$

Para una prueba completa, ver [**6**, XV.6, Teorema 1]. La condición de Lindeberg (1.3.17) es básicamente necesaria ([**6**, XV.6, Teorema 2]).

6. Sea

$$C_p = \begin{cases} 1 & \text{si } p^2|n \\ 0 & \text{si } p^2\nmid n,\end{cases}$$

donde n es tomado al azar entre 1 y N. Sea $C=\sum_{p\leq N} C_p$. (Como, para $p>\sqrt{N}$, no hay entero $n\leq N$ tal que $p^2|n$, tenemos que $C=\sum_p C_p = \sum_{p\leq\sqrt{N}} C_p$.) En otras palabras, C es la variable aleatoria que da el número de cuadrados de primos que dividen un entero tomado al azar entre 1 y N. Estudiemos la distribución de C.

a) Muestre que

$$\mathbb{E}(C) = \sum_{p\leq\sqrt{N}} C_p = \sum_{p\leq\sqrt{N}}\frac{1}{p^2} + O\left(N^{-1/2}\right) = \sum_p \frac{1}{p^2} + O\left(N^{-1/2}\right).$$

b) Muestre que

$$\operatorname{Prob}(C=0) = \prod_p\left(1-\frac{1}{p^2}\right) + O\left(N^{-1/2}\right).$$

Por consiguiente, $\lim_{N\to\infty}\operatorname{Prob}(C=0)$ existe y es igual a $\prod_p\left(1-\frac{1}{p^2}\right)$. El evento $C=0$ no es sino el evento que n carezca de divisores cuadrados (i.e. $d^2\nmid n$ para todo entero $d>1$).

c) Muestre que, para todo $k\geq 0$,

$$\begin{aligned}\operatorname{Prob}(C=k) &= \sideset{}{^*}\sum_{\substack{1\leq m\leq\sqrt{N}\\ \omega(m)=k}}\frac{1}{m^2}\cdot\left(\prod_{p\nmid m}\left(1-\frac{1}{p^2}\right) + O\left((N/m)^{-1/2}\right)\right)\\ &= \sideset{}{^*}\sum_{\substack{m\geq 1\\ \omega(m)=k}}\frac{1}{m^2}\cdot\prod_{p\nmid m}\left(1-\frac{1}{p^2}\right) + O(N^{-1/2}),\end{aligned} \tag{1.3.18}$$

donde $\sum^*$ denota una suma sólo sobre enteros sin divisores cuadrados. Por lo tanto, $\lim_{N\to\infty} \text{Prob}(C = k)$ existe y es igual a

$$\sum_{\substack{m\geq 1\\ \omega(m)=k}}^{*} \frac{1}{m^2}\cdot \prod_{p\nmid m}\left(1-\frac{1}{p^2}\right).$$

Tenemos, entonces, que C converge a una distribución discreta cuando $N \to \infty$. Esta distribución no es la normal (ya que es discreta) ni se le parece (no es simétrica alrededor de $\mathbb{E}(C)$: la probabilidad de $C < 0$ es cero, pero la probabilidad de $C > 2\mathbb{E}(C)$ tiende a un valor positivo). Lo crucial aquí es que $\lim_{N\to\infty} \mathbb{E}(C) < \infty$, es decir, el hecho que $\sum_p \mathbb{E}(X_p) = \sum_p \frac{1}{p^2}$ converge. En cambio, cuando examinábamos $X = \sum_p X_p$, teníamos que, como $\sum_p \frac{1}{p}$ diverge, la esperanza $\mathbb{E}(X)$ tendía a ∞ cuando $N \to \infty$ y la distribución límite era la normal.

d) Muestre que, para todo k,

$$\text{Prob}(C = k) \ll \frac{\lambda^k}{k!}$$

donde $\lambda = \sum_p \frac{1}{p^2}$. La distribución

(1.3.19) $$\frac{\lambda^k}{k!}e^{-\lambda}$$

es la famosa *distribución de Poisson*; se trata del límite $n \to \infty$ de la distribución de $Y_n = Y_{n,1} + \ldots + Y_{n,n}$, donde $\{Y_{i,j}\}_{i,j\geq 1}$ son variables mutuamente independientes con la distribución

$$Y_{n,j} = \begin{cases} 1 & \text{con probabilidad } \lambda/n \\ 0 & \text{con probabilidad } 1-\lambda/n. \end{cases}$$

(Demuestre que Y_n tiende, en efecto, a (1.3.19).)
Hemos, entonces, acotado la distribución de C por una distribución de Poisson de esperanza λ. Ésta no es una cota "ajustada": pruebe que

$$\text{Prob}(C = k) \ll_\epsilon \frac{\epsilon^k}{k!}$$

para todo $\epsilon > 0$, y, en consecuencia,

$$\text{Prob}(C = k) = o\left(\frac{\lambda^k}{k!}\right)$$

cuando $k \to \infty$, e incluso

$$\text{Prob}(C \geq k) = o\left(\frac{1}{k!}\right).$$

e) Sea

$$D_p = \begin{cases} 1 & \text{si } p|m \text{ y } p|n \\ 0 & \text{de lo contrario,} \end{cases}$$

donde (m, n) es un par de números entre 1 y N tomados al azar de acuerdo a la distribución uniforme. Sea $D = \sum_p D_p$. Entonces el evento $D = 0$ no es sino el evento que m y n sean primos entre si.
Muestre que

$$\mathrm{Prob}(D = 0) = \prod_p \left(1 - \frac{1}{p^2}\right) + O(N^{-1}),$$

$$\mathrm{Prob}(D = k) = \sideset{}{^*}\sum_{\substack{m\geq 1\\ \omega(m)=k}} \frac{1}{m^2} \cdot \prod_{p\nmid m} \left(1 - \frac{1}{p^2}\right) + O(N^{-1}).$$

Compare esto con (1.3.18). Como en $(6d)$, concluya que

$$\mathrm{Prob}(D \geq k) = o\left(\frac{1}{k!}\right).$$

1.4. Grandes desviaciones: cotas superiores. Valores críticos.

Sean $X_1, X_2, \ldots$ variables aleatorias mutuamente independientes con la distribución

$$X_j = \begin{cases} 0 & \text{con probabilidad } 1/2 \\ 1 & \text{con probabilidad } 1/2. \end{cases} \tag{1.4.1}$$

Sea $X = \sum_{j\leq n} X_j$. Sabemos que $\mathbb{E}(X) = \frac{1}{2}n$ y que la distribución de X será cercana a la normal alrededor de $\frac{1}{2}n$. Qué pasa lejos de $\frac{1}{2}n$?

En general, hablamos de pequeñas desviaciones cuando la distancia entre el valor de X y la esperanza $\mathbb{E}(X)$ es $O(\sqrt{\mathrm{Var}(X)})$, y de *grandes desviaciones* cuando la distancia es comparable a $\mathrm{Var}(X)$ o $\mathbb{E}(X)$. Podemos preguntarnos, por ejemplo, que tan a menudo se dan las grandes desviaciones $\omega(n) < \frac{1}{2}\log\log n$ o $\omega(n) > 6\log\log n$.

En el caso de las variables (1.4.1), podemos hacer los cálculos a mano. Tenemos $\mathrm{Prob}(X = m) = 2^{-n} \cdot \binom{n}{m}$ para todo m. Por lo tanto,

$$\mathrm{Prob}(Y > an) = 2^{-n} \sum_{m\geq an} \binom{n}{m}$$

para todo a, donde

$$\binom{n}{m} := \frac{n!}{(n-m)!m!} = \frac{1\cdot 2\cdot 3\cdot\cdots\cdot n}{(1\cdot 2\cdot\cdots\cdot(n-m))\cdot(1\cdot 2\cdot\cdots\cdot m)} = \frac{n(n-1)\cdots(n-m+1)}{1\cdot 2\cdot\cdots\cdot m}$$

es el número de maneras de escoger m cosas de entre n cosas. (Tenemos n posibilidades para la primera cosa elegida, $n-1$ posibilidades para la segunda, ..., $n-m+1$ posibilidades para la m–ésima, y no importa en qué orden de los $m!$ órdenes posibles hayamos elegido las m cosas. Por lo tanto, hay $\frac{n(n-1)\cdots(n-m+1)}{1\cdot 2\cdots m}$ maneras de escoger.)

Fijemos un $a \in [1/2, 1]$. Como $m \mapsto \binom{n}{m}$ es decreciente para $m \geq \frac{1}{2}n$, tenemos

$$2^{-n}\binom{n}{\lceil an\rceil} \leq \mathrm{Prob}(Y > an) \leq (n+1)2^{-n}\binom{n}{\lceil an\rceil}. \tag{1.4.2}$$

Ahora bien,

$$\log n! = \log 1 + \log 2 + \cdots + \log n = \int_1^n \log x\, dx + O(\log n) = n \log n - n + 1 + O(\log n),$$

así que

$$\begin{aligned}\log \binom{n}{m} &= n\log n - ((n-m)\log(n-m) + m\log m) + O(\log n) \\ &= -n \cdot \left(\left(1 - \frac{m}{n}\right) \cdot \log\left(1 - \frac{m}{n}\right) + \frac{m}{n}\log\frac{m}{n}\right) + O(\log n).\end{aligned} \tag{1.4.3}$$

Por (1.4.2) y (1.4.3), para $a \geq \frac{1}{2}$,

$$\begin{aligned}\log(\mathrm{Prob}(Y > an)) &\sim \log \frac{2^{-n}}{(a^a(1-a)^{(1-a)})^n} + O(\log n) \\ &= n(-a\log a - (1-a)\log(1-a) - \log 2) + O(\log n).\end{aligned}$$

En otras palabras, $\mathrm{Prob}(Y > an) = e^{(H+o(1))n}$, donde

$$H = -a \log \frac{a}{0{,}5} - (1-a)\log\frac{1-a}{0{,}5}. \tag{1.4.4}$$

La forma de (1.4.4) puede resultar familiar para los físicos (o químicos). Elaboraremos esta observación para más adelante; ahora, pasemos a estudiar $\omega(n)$.

* * *

Sean $X_2', X_3', X_5', \ldots$ variables mutuamente independientes con la distribución

$$X_p' = \begin{cases} 1 & \text{con probabilidad } \frac{1}{p} \\ 0 & \text{con probabilidad } 1 - \frac{1}{p}. \end{cases} \tag{1.4.5}$$

Sea $X' = \sum_{p \leq N} X_p'$. La esperanza $\mathbb{E}(X')$ es $\log\log N + O(1)$. Nos preguntaremos cuánto es la probabilidad $\mathrm{Prob}(X' < a\log\log N)$, $a < 1$, o $\mathrm{Prob}(X' > a \log\log N)$, $a > 1$.

Markov acota las cotas mediante $\mathbb{E}(X')$, Chebyshev mediante $\mathbb{E}(X'^2)$; muy bien podemos usar $\mathbb{E}(X'^k)$, o incluso $\mathbb{E}(e^{X'})$. En efecto, para a positivo y $\alpha \geq 0$,

$$\mathrm{Prob}(X' > a\log\log N) \leq \frac{\mathbb{E}\left(e^{\alpha X'}\right)}{e^{\alpha \cdot a \log\log N}}, \tag{1.4.6}$$

y, para a positivo y $\alpha \leq 0$,

$$\mathrm{Prob}(X' < a\log\log N) \leq \frac{\mathbb{E}\left(e^{\alpha X'}\right)}{e^{\alpha \cdot a \log\log N}}. \tag{1.4.7}$$

Así como la desigualdad de Chebyshev es simplemente la desigualdad de Markov aplicada a la variable X'^2, las desigualdades (1.4.6) y (1.4.7) son simplemente la desigualdad de Markov aplicada a la variable $e^{\alpha X'}$; tenemos la libertad de escoger el parámetro α como más nos convenga. (La aplicación de la desigualdad de Markov a una variable de la forma $e^{\alpha X'}$ es a veces denominada *el método de momentos exponenciales*; los momentos "usuales" son $\mathbb{E}(X')$, $\mathbb{E}(X'^2)$, $\mathbb{E}(X'^3)$,...)

Cómo evaluar $\mathbb{E}(e^{\alpha X'})$? Tenemos

$$e^{\alpha X'} = e^{\alpha \sum_{p \leq N} X'_p} = \prod_{p \leq N} e^{\alpha X'_p}.$$

Como $X'_2, X'_3, X'_5, \ldots$ son mutuamente independientes, las variables $e^{\alpha X'_2}, e^{\alpha X'_3}, e^{\alpha X'_5}, \ldots$ también lo son, y por lo tanto

$$\mathbb{E}(e^{\alpha X'}) = \mathbb{E}\left(\prod_{p \leq N} e^{\alpha X'_p}\right) = \prod_{p \leq N} \mathbb{E}\left(e^{\alpha X'_p}\right).$$

Es fácil calcular que

$$\mathbb{E}\left(e^{\alpha X'_p}\right) = 1 + \frac{1}{p}(e^\alpha - 1).$$

En consecuencia,

$$\mathbb{E}\left(e^{\alpha X'}\right) = \prod_{p \leq N} (1 + \frac{1}{p}(e^\alpha - 1)) \ll \prod_{p \leq N} \left(1 + \frac{1}{p}\right)^{e^\alpha - 1}. \tag{1.4.8}$$

(Esto es cierto aún si $\alpha < 0$, o incluso $\alpha = -\infty$. Demuestre la desigualdad en (1.4.8) si ésta no lo convence de inmediato; use el hecho que $\prod_{p \leq N}(1 + 1/p^2) \ll 1$, puesto que $\prod_p(1 + 1/p^2)$ converge (por qué?).) Ahora bien,

$$e^{\sum_{p \leq N} \frac{1}{p}} \ll \prod_{p \leq N} \left(1 + \frac{1}{p}\right) \ll e^{\sum_{p \leq N} \frac{1}{p}}, \tag{1.4.9}$$

y, como $\sum_{p \leq N} \frac{1}{p} = \log\log N + O(1)$ (Chebyshev-Mertens (1.1.17)), obtenemos que

$$\mathbb{E}\left(e^{\alpha X'}\right) \ll_\alpha \left(e^{\log\log N}\right)^{e^\alpha - 1} = (\log N)^{e^\alpha - 1}. \tag{1.4.10}$$

Podemos ahora sustituir (1.4.10) en (1.4.6) y (1.4.7). Nos queda escojer el valor óptimo de α. Para $a > 1$ y $\alpha > 0$,

$$\mathrm{Prob}(X' > a \log\log N) \ll_\alpha \frac{(\log N)^{e^\alpha - 1}}{e^{\alpha \cdot a \log\log N}} = (\log N)^{e^\alpha - 1 - \alpha a};$$

para $a < 1$ y $\alpha < 0$,

$$\mathrm{Prob}(X' < a \log\log N) \ll_\alpha \frac{(\log N)^{e^\alpha - 1}}{e^{\alpha \cdot a \log\log N}} = (\log N)^{e^\alpha - 1 - \alpha a}.$$

Debemos minimizar $e^\alpha - 1 - \alpha a$ para a positivo y fijo, cuidándonos que α sea negativo si $a > 1$, y positivo si $a < 1$. Sacando derivadas, vemos que $e^\alpha - 1 - \alpha a$ es mínimo para a dado cuando $\alpha = \log a$. Entonces

$$\begin{aligned} &\mathrm{Prob}(X' > a \log\log N) \ll_a (\log N)^{-(a \log a + 1 - a)} \quad \text{para } a > 1, \\ &\mathrm{Prob}(X' < a \log\log N) \ll_a (\log N)^{-(a \log a + 1 - a)} \quad \text{para } a < 1. \end{aligned} \tag{1.4.11}$$

Hemos obtenido las cotas superiores que deseábamos para las grandes desviaciones de $X' = \sum_{p \leq N} X'_p$, donde X'_p es como en (1.4.5). Las constantes implícitas en (1.4.11) dependen de a, pero de manera continua; por lo tanto, la misma constante puede servir para todo

$a \leq A$, A dado. (La constantes implícitas en (1.4.11) dependen de manera continua de $e^\alpha - 1 = e^{\log a} - 1 = a - 1$, por lo cual la dependencia en a es continua aún cerca de $a = 0$.)

* * *

Como de costumbre, pasamos a examinar las variables $X_2, X_3, X_5, \ldots$ dadas por

$$X_p = \begin{cases} 1 & \text{si } p|n, \\ 0 & \text{si } p \nmid n, \end{cases}$$

donde n es tomado al azar entre 1 y N con la distribución uniforme. Sea $X = \omega(n) = \sum_{p \leq N} X_p$. Las desigualdades (1.4.6) y (1.4.7) son aún válidas; partamos de ellas.

El problema es evaluar $\mathbb{E}(e^{\alpha X}) = \mathbb{E}(e^{\alpha\omega(n)})$. Si bien $e^{\alpha X} = e^{\alpha \sum_{p \leq N} X_p} = \prod_{p \leq N} e^{\alpha X_p}$, no podemos concluir que $\mathbb{E}(e^{\alpha X}) = \mathbb{E}\left(e^{\alpha \sum_{p \leq N} X_p}\right) = \prod_{p \leq N} \mathbb{E}\left(e^{\alpha X_p}\right)$, ya que las variables X_p no son mutuamente independientes. Empero, podemos mostrar sin mucha dificultad que

$$\sum_{n \leq N} \frac{e^{\alpha\omega(n)}}{n} \leq \prod_{p \leq N} \left(1 + e^\alpha \left(\frac{1}{p} + \frac{1}{p^2} + \ldots\right)\right) \ll \prod_{p \leq N} \left(1 + \frac{1}{p}\right)^{e^\alpha} \ll (\log N)^{e^\alpha}$$

(por (1.4.9)) y deducir con más trabajo de esto que

$$\sum_{n \leq N} e^{\alpha\omega(n)} \ll N(\log N)^{e^\alpha - 1}$$

(ver problema 1). Por lo tanto, $\mathbb{E}(e^{\alpha\omega(n)}) \ll (\log N)^{e^\alpha - 1}$. Fijamos $\alpha = \log a$ como antes, y concluimos (por (1.4.6) y (1.4.7)) que

$$\begin{aligned} &\text{Prob}(\omega(n) > a \log\log N) \ll_a (\log N)^{-(a \log a + 1 - a)} \quad \text{si } a > 1, \\ &\text{Prob}(\omega(n) < a \log\log N) \ll_a (\log N)^{-(a \log a + 1 - a)} \quad \text{si } a < 1. \end{aligned} \tag{1.4.12}$$

Las constantes implícitas en (1.4.12) dependen de a de manera continua – aún en la vecindad de $a = 0$.

* * *

Tanto (1.4.11) como (1.4.12) son cotas superiores. Encontraremos dentro de poco cotas inferiores muy cercanas a estas cotas superiores; mientras tanto, contentémonos con una aplicación de (1.4.12).

Cuántos enteros n entre 1 y N^2 pueden expresarse como el producto $n = a \cdot b$ de dos enteros a, b entre 1 y N? En otras palabras: si escribimos una tabla de multiplicación de 1000 por 1000 (digamos), habrá un millón de números en la tabla, y todos estos números estarán entre 1 y un millón; empero, como hay muchas repeticiones, podemos preguntarnos: cuántos de los números entre 1 y un millón están presentes en la tabla?

Este es el conocido *problema de la tabla de multiplicación*. Lo encontramos por primera vez en §1.2, ejercicio 4, donde probamos que el número de enteros $n \leq N^2$ que aparecen en la tabla (i.e., $n = a \cdot b$ para algún par $1 \leq a, b \leq N$) es $o(N^2)$. Ahora estableceremos una cota superior bastante más ajustada.

Comenzamos de la misma manera que antes: tomamos conciencia de que, si bien la esperanza de $\omega(n)$ (para n tomado al azar) es

$$\mathbb{E}(\omega(n)) = \log\log N^2 + O(1) = \log\log N + O(1),$$

la esperanza de $\omega(a \cdot b)$ (para a, b tomados al azar) es

$$\mathbb{E}(\omega(a\cdot b)) = \mathbb{E}(\omega(a)+\omega(b)-\omega(\mathrm{mcd}(a,b))) = \mathbb{E}(\omega(a))+\mathbb{E}(\omega(b))-O(1) = 2\log\log N+O(1).$$

Por lo tanto, el número de divisores $\omega(n)$ de todo número $n = a\cdot b$ ($n \leq N^2$, $a, b \leq N$) debe estar ya sea en la cola de la distribución de $\omega(n)$ o en la cola de la distribución de $\omega(a \cdot b)$. Ahora que sabemos como acotar las colas distantes (es decir, las grandes desviaciones) de $\omega(n) = X = \sum_p X_p$, podremos dar una buena cota superior para la (poca) probabilidad de $\omega(n)$ dentro de una u otra distribución.

Estamos en una situación que ya es propia no sólo de las probabilidades, sino de la estadística. La situación es así: hay una variable aleatoria observable U; su distribución se desconoce, pero se tienen dos conjeturas acerca de ésta. Hablemos, entonces, de la distribución 1 y la distribución 2. La esperanza $\mathbb{E}_1(U)$ de U según la distribución 1 es (digamos) u_1, mientras que la esperanza $\mathbb{E}_2(U)$ de U según la distribución 2 es $u_2 > u_1$. La tarea es determinar cuál de las dos distribuciones es más verosímil.

Debemos ponernos de acuerdo en un *valor crítico* t entre u_1 y u_2. Si, despues de n mediciones de U, vemos que el promedio $X = \frac{1}{n}(U_1 + U_2 + \ldots + U_n)$ es menor que t, decidiremos la disputa en favor de la distribución 1; si $X \geq t$, daremos la razón a 2. La pregunta es: que *valor crítico* t debemos escoger?

Si la distribución 1 es la verdadera distribución de U, la probabilidad de decidir la disputa erróneamente a favor de la distribución 2 es $\mathrm{Prob}_1(X \geq t)$ (donde $\mathrm{Prob}_1(\ldots)$ denota la probabilidad de un evento si se asume que la distribución 1 es la cierta). Si la distribución 2 es la cierta, la probabilidad de errar a favor de la distribución 1 es $\mathrm{Prob}_2(X < t)$ (donde $\mathrm{Prob}_2(\ldots)$ denota la probabilidad de un evento si se asume que la distribución 2 es la cierta). Resulta sensato, entonces, escoger el valor de t entre u_1 y u_2 para el cual

$$\mathrm{Prob}_1(X \geq t) + \mathrm{Prob}_2(X < t)$$

sea mínimo.

En el caso del problema de la tabla de multiplicación, estamos ante una situación parecida. Fijaremos un t entre 1 y 2. Para todo $n = a \cdot b$, $a, b \leq N$, tendremos ya sea $\omega(n) < t\log\log N$ o $\omega(n) \geq t\log\log N$. En el primer caso, cualquier par $a, b \leq N$ tal que $n = a \cdot b$ tendrá que estar en el conjunto $\{(a, b) : 1 \leq a, b \leq N,\ \omega(a \cdot b) < t\log\log N\}$, y el número de elementos de este conjunto es N^2 multiplicado por

$$\mathrm{Prob}(\omega(a \cdot b) < t\log\log N),$$

donde a y b son tomados al azar entre 1 y N con la distribución uniforme. En el segundo caso – es decir, $\omega(n) \geq t\log\log N$ – el número n estará en el conjunto $\{n : 1 \leq n \leq N, \omega(n) \geq t\log\log N\}$, cuyo número de elementos es N^2 multiplicado por

$$\mathrm{Prob}(\omega(n) \geq t\log\log N),$$

donde n es tomado al azar entre 1 y N^2 con la distribución uniforme. Por lo tanto, el conjunto de enteros $n \leq N^2$ tales que $n = a \cdot b$ para algún par $a, b \leq N$ tiene a lo más

$$N^2 \cdot (\mathrm{Prob}(\omega(n) \geq t \log\log N) + \mathrm{Prob}(\omega(a \cdot b) < t \log\log N)) \tag{1.4.13}$$

elementos. Queremos, entonces, escoger t tal que (1.4.13) sea tan pequeño como se pueda.

Para hacer esto, debemos primero estimar (1.4.13). Gracias a nuestra cota de grandes desviaciones (1.4.12),

$$\begin{aligned} \mathrm{Prob}(\omega(n) \geq t \log\log N) &= \mathrm{Prob}(\omega(n) \geq t \log\log N^2 - t\log 2) \\ &= \mathrm{Prob}(\omega(n) \geq t(1 - o(1)) \cdot \log\log N^2) \\ &\leq (\log N)^{-I(t)(1+o(1))}, \end{aligned} \tag{1.4.14}$$

donde $I(t) = t \log t + 1 - t$. Queda acotar $\mathrm{Prob}(\omega(a \cdot b) < t \log\log N)$. Está claro que

$$\omega(a \cdot b) = \omega(a) + \omega(b) - \omega(\mathrm{mcd}(a, b)).$$

Por §1.3, ejercicio $6e$, la probabilidad que $\mathrm{mcd}(a, b)$ sea $> \epsilon \log\log N$ es $\ll (\log N)^{-100}$ (o cualquier otra potencia que se quiera). (Hay varias maneras simples de evitar el uso del ejercicio $6e$; la ventaja de este último es que nos permite continuar sin salir nunca de un marco probabilístico.) Bastará, entonces, con estimar la probabilidad que $(\omega(a), \omega(b))$ quede en un cuadrado $\mathbf{C}_{t_a,t_b}$ de lado $\epsilon \log\log N$ que contenga el punto $(t_a \log\log N, t_b \log\log N)$, donde $t_a + t_b \leq t + \epsilon$. El máximo de tal probabilidad, multiplicado por el número de cuadrados a considerar (el cual es $O(1/\epsilon^2) = O_\epsilon(1)$), nos dará una cota superior para la probabilidad de $\omega(a \cdot b) \leq t$.

Como a y b son variables independientes, la cota de grandes desviaciones (1.4.12) nos dice que la probabilidad que $(\omega(a), \omega(b))$ esté dentro de $\mathbf{C}_{t_a,t_b}$ es

$$\ll (\log N)^{-(I(t_a)+I(t_b))+O(\epsilon)},$$

donde $I(t) = t \log t + 1 - t$. Nos estamos preguntando, entonces, cuál es el máximo de $I(t_a) + I(t_b)$, bajo la condición que $t_a + t_b \leq t$. Sacamos la segunda derivada de $I(x)$ y vemos que siempre es positiva. Por lo tanto, la gráfica de $x \mapsto I(x)$ se curva hacia arriba, y, así, $\frac{1}{2}(I(t_a) + I(t_b)) \geq I\left(\frac{t_a+t_b}{2}\right)$. Como t está entre 1 y 2, y $I(x)$ es decreciente en el intervalo entre $1/2$ y $2/2 = 1$, vemos que $I\left(\frac{t_a+t_b}{2}\right) \geq I(t/2+\epsilon) = I(t/2)+O(\epsilon)$. Concluimos que

$$(\log N)^{-(I(t_a)+I(t_b))+O(\epsilon)} \leq (\log N)^{-2I(t/2)+O(\epsilon)}.$$

Por lo tanto,

$$\mathrm{Prob}(\omega(a \cdot b) \leq t \log\log N) \leq (\log N)^{-2I(t/2)+O(\epsilon)}. \tag{1.4.15}$$

Comparando (1.4.14) y (1.4.15), vemos que tenemos que escoger $t \in [1, 2]$ de tal manera que

$$\mathrm{mín}(I(t), 2I(t/2))$$

sea máximo. Ahora bien, $I(t)$ es creciente y $2I(t/2)$ es decreciente en el intervalo $[1, 2]$. Por lo tanto, el máximo de $\mathrm{mín}(I(t), 2I(t/2))$ se alcanza cuando $I(t) = 2I(t/2)$. Resolvemos $I(t) = 2I(t/2)$, y encontramos $t = \frac{1}{\log 2}$. Hemos probado el siguiente resultado:

Teorema 1.7 (Erdős [**3**]). *El número de enteros n entre 1 y N^2 que pueden ser expresados como el producto $n = a \cdot b$ de dos enteros a, b entre 1 y N es*

$$O_\epsilon\left(N^2 \cdot (\log N)^{-I\left(\frac{1}{\log 2}\right)+\epsilon}\right)$$

para cualquier $\epsilon > 0$, donde $I(t) = t\log t + 1 - t$.

Numéricamente, $I\left(\frac{1}{\log 2}\right) = 0{,}08607\ldots$.

Notas y problemas

1. Debemos estimar $\sum_{n\leq N} e^{\alpha\omega(n)}$, donde $\alpha > 0$. Si se usan métodos analíticos – que no trataremos para evitar el análisis complejo – esto es rutina (dentro del método de Selberg–Delange – ver, e.g., [**9**, II.5]). Veremos como obtener de manera elemental una cota superior "del orden correcto" (es decir, $\ll$ la asintótica verdadera).
   a) Primero deduzca de (1.1.5) que
   $$\prod_{p\leq N}\left(1+\frac{1}{p}\right) \ll \log N,$$
   $$\prod_{p\leq N}\left(1+\frac{1}{p}+\frac{1}{p^2}+\ldots\right) \ll \log N. \tag{1.4.16}$$
   b) Acotemos $\sum_{n\leq N} e^{\alpha\omega(n)}/n$:
   $$\begin{aligned}\sum_{n\leq N}\frac{e^{\alpha\omega(n)}}{n} &\leq \prod_{p\leq N}\left(1+e^\alpha\left(\frac{1}{p}+\frac{1}{p^2}+\frac{1}{p^3}+\ldots\right)\right)\\ &\leq \prod_{p\leq N}\left(1+\frac{1}{p}+\frac{1}{p^2}+\ldots\right)^{e^\alpha} \ll_\alpha (\log N)^{e^\alpha}.\end{aligned} \tag{1.4.17}$$
   c) Podríamos concluir que $\sum_{n\leq N} e^{\alpha\omega(n)} \ll N\sum_{n\leq N}\frac{e^{\alpha\omega(n)}}{n} \ll N(\log N)^{e^\alpha}$, lo cual es correcto, pero esta cota se aleja de la realidad por un factor de $(\log N)$ (como veremos después). Cómo obtendremos una mejor cota?

   (Como dijimos, es rutina obtener la "suma parcial" $\sum_{n\leq N} e^{\alpha\omega(n)}$ a través del análisis (no elemental) de la función $L(s) = \sum_{n=1}^\infty e^{\alpha\omega(n)} n^{-s}$. Estamos tratando de obtener la suma parcial $\sum_{n\leq N} e^{\alpha\omega(n)}$ meramente de la suma parcial $\sum_{n\leq N} e^{\alpha\omega(n)} n^{-1}$, y quizás de un par de propiedades de $e^{\alpha\omega(n)}$.)

   La *convolución* $f * g$ de dos funciones $f, g : \mathbb{Z}^+ \to \mathbb{C}$ es $(f*g)(n) = \sum_{d|n} f(d)g(n/d)$. A menudo, en la teoría analítica de números, es conveniente expresar una función como la convolución de dos otras. Ya podemos obtener la suma parcial de $h(n) = e^{\alpha\omega(n)} * \mathbf{1}$, donde $\mathbf{1} : \mathbb{Z}^+ \to \mathbb{C}$ es la función

$\mathbf{1}(n) = 1$:

$$\begin{aligned}\sum_{n\leq N} h(n) &= \sum_{n\leq N}\sum_{d|n} e^{\alpha\omega(d)}\cdot 1 = \sum_{d\leq N} e^{\alpha\omega(d)} \sum_{\substack{m\leq N\\ d|m}} 1 \\ &\leq \sum_{d\leq N} e^{\alpha\omega(d)}\frac{N}{d} = N\sum_{d\leq N}\frac{e^{\alpha\omega(d)}}{d} \ll_\alpha N(\log N)^{e^\alpha}.\end{aligned} \tag{1.4.18}$$

Ahora bien, queremos conocer la suma parcial de $e^{\alpha\omega(n)}$, no la de $e^{\alpha\omega(n)} * \mathbf{1}$.

d) Utilicemos el hecho que $\log = \mathbf{1} * \Lambda$. (Esto no es sino (1.1.10).) Tenemos que

$$\sum_{n\leq N} e^{\alpha\omega(n)}\log(n) = \sum_{n\leq N}\sum_{d|n} e^{\alpha\omega(n)}\Lambda(n/d).$$

Ahora bien, $\omega(n) - 1 \leq \omega(d) \leq \omega(n)$ para todo n y todo $d|n$ tales que $\Lambda(n/d) \neq 0$. Así,

$$\sum_{n\leq N} e^{\alpha\omega(n)}\log(n) \ll \sum_{n\leq N}\sum_{d|n} e^{\alpha\omega(d)}\Lambda(n/d).$$

De (1.1.14) y (1.4.18), se deduce que

$$\begin{aligned}\sum_{n\leq N}\sum_{d|n} e^{\alpha\omega(d)}\Lambda(n/d) &= \sum_{d\leq N} e^{\alpha\omega(d)} \sum_{m\leq N/d}\Lambda(m) \\ &\ll \sum_{d\leq N} e^{\alpha\omega(d)}\cdot N/d \ll_\alpha N(\log N)^{e^\alpha}.\end{aligned}$$

Por lo tanto,

$$\sum_{n\leq N} e^{\alpha\omega(n)}\log(n) \ll_\alpha N(\log N)^{e^\alpha}.$$

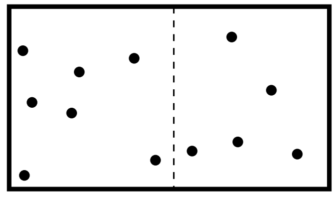
Usando la técnica de la suma por partes (§1.1, nota (1)), deduzca de esto que

$$\sum_{n\leq N} e^{\alpha\omega(n)} \ll_\alpha N(\log N)^{e^\alpha - 1}. \tag{1.4.19}$$

La constante implícita en $\ll$ depende de α, tanto aquí como antes. La dependencia es continua en α (o, si se prefiere, en $\beta = e^\alpha$; la dependencia es continua aún en la vecindad de $\beta = 0$).

1.5. Grandes desviaciones: cotas inferiores. Entropía.

Consideremos dos compartimientos separados por una membrana porosa. Llenémoslos con un gas:



De cuántas maneras puede ocurrir que haya m partículas en el compartimiento de la izquierda y $n-m$ en el de la derecha?

Hay $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ maneras de escoger m partículas de entre n; por lo tanto, hay $\binom{n}{m}$ maneras que haya m partículas en la izquierda y $n-m$ en la derecha.

Como $\log n! = \log 1 + \log 2 + \ldots + \log n = n \log n - n + O(\log n)$,

$$\begin{aligned}\log \binom{n}{m} &= \log n! - \log m! - \log(n-m)! \\ &= n \log n - m \log m - (n-m)\log(n-m) + O(\log n) \\ &= -n \cdot \left(\frac{m}{n} \log \frac{m}{n} + \frac{n-m}{n} \log \frac{n-m}{n} \right) + O(\log n).\end{aligned} \tag{1.5.1}$$

Qué pasa si cada una de las 2^n maneras de colocar n partículas en dos compartimientos es igualmente probable? (Hay 2^n maneras porque, dada cada partícula, podemos elegir en cual de dos compartimientos puede encontrarse.) Obtenemos de (1.5.1) que la probabilidad que la proporción de partículas en el comportamiento izquierdo sea $r = \frac{m}{n}$ es

$$\begin{aligned}\frac{\log \binom{n}{m}}{2^n} &= e^{-n(r \log r + (1-r)\log(1-r) + \log 2) + O(\log n)} \\ &= e^{n(H(r)+o(1))}.\end{aligned}$$

donde

$$\begin{aligned}H(r) &= -(r \log r + (1-r)\log(1-r) + \log 2) \\ &= -\left(r \log \frac{r}{0{,}5} + (1-r) \log \frac{1-r}{0{,}5} \right).\end{aligned}$$

La cantidad $H(r)$ no es sino la famosa *entropía*. Compárese con (1.4.4).

Claro está, no hay razón física por la cual podamos asumir que todas las disposiciones *iniciales* son igualmente probables; muy bien podemos comenzar llenando sólo uno de los compartimientos con gas. Empero, resulta poco sorprendente que, en la práctica y con el paso del tiempo, el sistema tienda a las proporciones $(r : 1-r)$ que resultarían más probables si las 2^n disposiciones fueran igualmente probables. Lo que hemos mostrado es que esto es lo mismo que decir que la entropía $H(r)$ del sistema tiende a crecer. (Esta tendencia es muy clara, ya que $H(r)$ está en el exponente de $e^{n(H(r)+o(1))}$; por eso se habla, en la termodinámica, de una *ley* según la cual la entropía siempre crece.)

* * *

Ahora que vemos que las expresiones del tipo $r \log r$ aparecen en un modelo de un fenómeno natural, resulta sensato esperar que el hasta ahora curioso exponente de $a \log a + 1 - a$ en (1.4.11) y (1.4.12) no sea simplemente una consecuencia de nuestra incapacidad. Nuestra tarea consiste ahora en dar cotas inferiores cercanas a las cotas superiores en (1.4.11) y (1.4.12), y de esta manera mostrar que el exponente $a \log a + 1 - a$ verdaderamente describe la probabilidad de las grandes desviaciones.

Comenzemos, como de costumbre, examinando variables mutuamente independientes $X_2', X_3', \ldots$ de distribución

$$X_p' = \begin{cases} 1 & \text{con probabilidad } \frac{1}{p} \\ 0 & \text{con probabilidad } 1 - \frac{1}{p}. \end{cases}$$

Sea $X' = \sum_{p\leq N} X'_p$. Queremos dar cotas inferiores para $\text{Prob}(X > a\log\log N)$, $a > 1$, y $\text{Prob}(X < a\log\log N)$, $a < 1$.

Utilizaremos el método del *ladeo exponencial*. Definimos nuevas variables $Y_2, Y_3, Y_5, \ldots$ mutuamente independientes y de distribución

$$Y_p = \begin{cases} 1 & \text{con probabilidad } \frac{a}{p} \cdot \left(1 + \frac{a-1}{p}\right)^{-1} \\ 0 & \text{con probabilidad } \left(1 - \frac{1}{p}\right) \cdot \left(1 + \frac{a-1}{p}\right)^{-1}. \end{cases} \tag{1.5.2}$$

(Está claro que estamos "ladeando" las variables hacia 1, pero donde está la "exponencial"? En Y_p, la probabilidad $1/p$ se ha vuelto $a/p = a^1 \cdot (1/p) \cdot \left(1 + \frac{a-1}{p}\right)^{-1}$; la probabilidad $1-1/p$ se ha vuelto $a^0 \cdot (1/p) \cdot \left(1 + \frac{a-1}{p}\right)^{-1}$. Si X_p tomara los valores $2, 3, \cdots$ con probabilidades no nulas, multiplicaríamos dichas probabilidades por $a^2, a^3, \cdots$, respectivamente. El factor de $\left(1 + \frac{a-1}{p}\right)^{-1}$ está allí simplemente para asegurar que la suma de las probabilidades sea 1.)

Sea $Y = \sum_{p\leq N} Y_p$. El evento $\text{Prob}(Y > a\log\log N)$ no es una gran desviación, sino un evento probable. Por el teorema del límite central,

$$\begin{aligned} \text{Prob}(a\log\log N < Y \leq a\log\log N + \sqrt{a\log\log N}) &= \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + o(1) \\ &\geq \frac{1}{5} + o(1), \end{aligned} \tag{1.5.3}$$

(digamos). Sea I el intervalo $(a\log\log N, a\log\log N + \sqrt{a\log\log N}]$. Comparemos

$$\begin{aligned} \text{Prob}(X' > a\log\log N) &= \sum_{\substack{\{x_p\}_{p\leq N}:\, x_p \in\{0,1\} \\ \sum_{p\leq N} x_p > a\log\log N}} \text{Prob}(X'_p = x_p \;\; \forall p \leq N) \\ &= \sum_{\substack{\{x_p\}_{p\leq N}:\, x_p \in\{0,1\} \\ \sum_{p\leq N} x_p > a\log\log N}} \prod_{p\leq N} \begin{cases} \frac{1}{p} & \text{si } x_p = 1 \\ 1 - \frac{1}{p} & \text{si } x_p = 0 \end{cases} \end{aligned} \tag{1.5.4}$$

y

(1.5.5)

$$\begin{aligned} \text{Prob}(Y \in I) = &\sum_{\substack{\{x_p\}_{p\leq N}:\, x_p \in\{0,1\} \\ \sum_{p\leq N} x_p \in I}} \text{Prob}(Y_p = x_p \;\; \forall p \leq N) \\ = &\sum_{\substack{\{x_p\}_{p\leq N}:\, x_p \in\{0,1\} \\ \sum_{p\leq N} x_p \in I}} \prod_{p<N} \begin{cases} \frac{a}{p} & \text{si } x_p = 1 \\ 1 - \frac{1}{p} & \text{si } x_p = 0 \end{cases} \cdot \prod_{p\leq N} \left(1 + \frac{a-1}{p}\right)^{-1} \\ = &\sum_{\substack{\{x_p\}_{p\leq N}:\, x_p \in\{0,1\} \\ \sum_{p\leq N} x_p \in I}} \text{Prob}(X'_p = x_p \;\; \forall p \leq N) \cdot \prod_{p\leq N} \left(1 + \frac{a-1}{p}\right)^{-1} \cdot \prod_{p\leq N: x_p = 1} a. \end{aligned}$$

Nuestro deseo es ir de la cota inferior (1.5.3) para (1.5.5) a una cota inferior para (1.5.4). Todo término $\mathrm{Prob}(X'_p = x_p \ \ \forall p \leq N)$ de la suma en (1.5.5) aparece en (1.5.4), puesto que $\sum_{p\leq N} x_p \in I = (a\log\log N, a\log\log N + \sqrt{a\log\log N}]$ implica $\sum_{p\leq N} x_p > a\log\log N$. Claro está, en (1.5.5), el término $\mathrm{Prob}(X'_p = x_p \ \ \forall p \leq N)$ aparece con dos factores; uno de ellos es

$$\prod_{p\leq N}\left(1+\frac{a-1}{p}\right)^{-1} \ll_a \prod_{p\leq N}\left(1+\frac{1}{p}\right)^{-(a-1)} \ll_a (\log N)^{-(a-1)},$$

y el otro es

$$\prod_{p\leq N:\, x_p=1} a = a^{\sum_{p\leq N} x_p}.$$

Como $\sum_{p\leq N} x_p \in I$, sabemos que $\sum_{p\leq N} x_p \leq a\log\log N + \sqrt{a\log\log N}$. Por ende,

$$\prod_{p\leq N:\, x_p=1} a \leq a^{a\log\log N + \sqrt{a\log\log N}}.$$

Concluimos que

$$\mathrm{Prob}(Y\in I) \ll_a \mathrm{Prob}(X > a\log\log N)\cdot a^{a\log\log N+\sqrt{a\log\log N}}\cdot(\log N)^{-(a-1)}.$$

Por lo tanto

$$\mathrm{Prob}(X' > a\log\log N) \gg_a \mathrm{Prob}(Y\in I)\cdot a^{-(a\log\log N+\sqrt{a\log\log N})}\cdot(\log N)^{a-1}.$$

Como sabemos que $\mathrm{Prob}(Y\in I) > \frac{1}{5}+o(1)$ (por (1.5.3)), obtenemos

$$\begin{aligned}\mathrm{Prob}(X' > a\log\log N) &\gg_a \frac{1}{5}a^{-(a\log\log N+\sqrt{a\log\log N})}\cdot(\log N)^{a-1}\\ &\gg (\log N)^{-(a\log a-a+1)-O((a^{1/2}\log a)\cdot(\log\log N)^{-1/2})}\end{aligned} \tag{1.5.6}$$

para $a>1$. Mediante exactamente el mismo método, podemos obtener

$$\mathrm{Prob}(X' < a\log\log N) \gg_a (\log N)^{-(a\log a-a+1)-O((a^{1/2}\log a)\cdot(\log\log N)^{-1/2})}. \tag{1.5.7}$$

para $a<1$. Hemos obtenido, entonces, cotas inferiores para complementar a (1.4.11). La constante implícita en $\gg$ es continua en a.

* * *

Aún nos falta derivar cotas inferiores similares para las variables

$$X_p = \begin{cases} 1 & \text{si } p|n \\ 0 & \text{si } p\nmid n.\end{cases}$$

Veremos que hay un método muy general para ir de resultados sobre X'_p a resultados sobre X_p. Ya pudimos haberlo utilizado para traducir las cotas superiores $\mathrm{Prob}(X' > a\log\log N) \ll \ldots$ en cotas superiores $\mathrm{Prob}(X > a\log\log N) \ll \ldots$. Hemos esperado hasta ahora en parte porque el método depende de un resultado técnico comúnmente considerado difícil: el *lema fundamental de la teoría de cribas*. (Ver el problema 1 al final de esta sección; en éste se desarrolla una prueba de una versión débil del lema fundamental – la única versión que necesitaremos.)

Como de costumbre, comenzaremos truncando las sumas. Definimos $S_m = \sum_{p\leq m} X_p$, y, en vez de trabajar con $X = S_N = \sum_{p\leq N} X_p$, trabajaremos con $S_{g(N)}$, donde $g(N)$ es un poco menor que N. Acotaremos la diferencia $X - S_{g(N)}$ al final, ya que a estas alturas esa parte es rutinaria. Escogeremos $g(N)$ en el último paso.

Sea $S'_m = \sum_{p\leq m} X'_p$. Debemos comparar $S_{g(N)}$ con $S'_{g(N)}$, o, más precisamente, mostrar que $\mathrm{Prob}(S_{g(N)} > a\log\log N)$ (o $\mathrm{Prob}(S_{g(N)} < a\log\log N)$) es aproximadamente igual a $\mathrm{Prob}(S'_{g(N)} > a\log\log N)$ (o con $\mathrm{Prob}(S'_{g(N)} < a\log\log N)$). Ya tenemos cotas para $S'_{g(N)}$, gracias a (1.5.6) y (1.5.7) (con $X = S'_{g(N)}$).

Tenemos, por una parte,

$$\mathrm{Prob}(S_{g(N)} > a\log\log N) = \sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N}} \mathrm{Prob}(X_p = x_p \ \ \forall p \leq g(N)), \tag{1.5.8}$$

y, por otra,

$$\begin{aligned}\mathrm{Prob}(S'_{g(N)} > a\log\log N) &= \sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N}} \mathrm{Prob}(X'_p = x_p \ \ \forall p \leq g(N)) \\ &= \sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N}} \prod_{p\leq g(N)} \begin{cases} \frac{1}{p} & \text{si } x_p = 1 \\ 1-\frac{1}{p} & \text{si } x_p = 0. \end{cases}\end{aligned} \tag{1.5.9}$$

Consideraremos primero los términos de (1.5.8) con $\prod_{p:x_p=1} p \leq N^{1-\epsilon}$. (Aquí $\epsilon > 0$ es un número fijo cualquiera entre 0 y 1.)

Sea $m = \prod_{p:x_p=1} p$. Supongamos que $m \leq N^{1-\epsilon}$. Sea $S = \{p \leq g(N) : x_p = 0\}$. Entonces

$$\mathrm{Prob}(X_p = x_p \ \ \forall p \leq g(N)) = \frac{1}{m}\,\mathrm{Prob}(X_p = 0 \ \ \forall p \in S),$$

donde en el lado derecho de la ecuación las variables X_p dependen de un número n tomado al azar entre 1 y N/m, no entre 1 y N. Por el lema fundamental de las cribas (problema 1, teorema 1.8),

$$\begin{aligned}\mathrm{Prob}(X_p = 0 \ \ \forall p \in S) &= \prod_{p\in S}\left(1-\frac{1}{p}\right)\cdot(1+O_A((\log N/m)^{-A})) \\ &= \prod_{p\in S}\left(1-\frac{1}{p}\right)\cdot(1+O_A((\log N)^{-A}))\end{aligned}$$

para cualquier A, siempre y cuando $\log g(N) \ll \frac{\log N/m}{\log\log(N/m)}$. Esto último ciertamente tiene lugar si $g(N) = o\left(\frac{\log N}{\log\log N}\right)$ (donde utilizamos el hecho que $m \leq N^{1-\epsilon}$).

Asumamos, entonces, $g(N) \ll \frac{\log N}{\log\log N}$. Podemos entonces concluir que

$$\sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N \\ \prod_{p:x_p=1} p \leq N^{1-\epsilon}}} \operatorname{Prob}(X_p = x_p \;\; \forall p \leq g(N))$$

es igual a

$$\sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N \\ \prod_{p:x_p=1} p \leq N^{1-\epsilon}}} \frac{1}{\prod_{p:x_p=1} p} \prod_{p:x_p=0} \left(1-\frac{1}{p}\right) \cdot (1+O_A((\log N)^{-A})),$$

lo cual no es sino

$$(1+O_A((\log N)^{-A}) \cdot \sum_{\substack{\{x_p\}_{p\leq g(N)}:\, x_p\in\{0,1\} \\ \sum_{p\leq g(N)} x_p > a\log\log N \\ \prod_{p:x_p=1} p \leq N^{1-\epsilon}}} \operatorname{Prob}(X_p' = x_p \;\; \forall p \leq g(N)),$$

es decir, la suma de los términos de (1.5.9) con $\prod_{p:x_p=1} p \leq N^{1-\epsilon}$, multiplicada por $(1+O_A((\log N)^{-A}))$. (Este es el paso crucial del método: hemos logrado ir de una suma que involucra a $\operatorname{Prob}(X_p = x_p \;\; \forall p \leq g(N))$ a una suma que involucra a $\operatorname{Prob}(X_p' = x_p \;\; \forall p \leq g(N))$, donde las variables X_p' son las variables mutuamente independientes que estudiamos al principio de la sección.)

Nos queda acotar los términos de (1.5.8) y (1.5.9) con $\prod_{p:x_p=1} p > N^{1-\epsilon}$. En un caso como el otro, el total de tales términos es a lo más

$$\sum_{\substack{Q\subset P \\ \prod_{p\in Q} p > N^{1-\epsilon}}} \frac{1}{\prod_{p\in Q} p},$$

donde P es el conjunto de los primos $p \leq g(N)$. Por un resultado intermedio (1.5.20) en la prueba del lema fundamental de las cribas,

$$\sum_{\substack{Q\subset P \\ \prod_{p\in Q} p > N^{1-\epsilon}}} \frac{1}{\prod_{p\in Q} p} \ll (\log g(N)) \cdot \left(\log\left(\frac{N^{1-\epsilon}}{g(N)}\right)\right)^{-\left(\log\left(\frac{N^{1-\epsilon}}{g(N)}\right)\right)\cdot(1+o(1))} \ll_A (\log N)^{-A}$$

para cualquier $A > 0$, donde utilizamos el hecho que $\log(N) = o\left(\frac{\log N}{\log\log N}\right)$.

Concluimos que

$$\operatorname{Prob}(S_{g(N)} > a\log\log N) = \operatorname{Prob}(S_{g(N)}' > a\log\log N) + O_A\left((\log N)^{-A}\right) \tag{1.5.10}$$

para cualquier $A > 0$, bajo la condición que $g(N) = o\left(\frac{\log N}{\log\log N}\right)$.

La desigualdad (1.5.6) puede aplicarse directamente a $S_{g(N)}$: como la definición de las variables X'_p no involucra a N, podemos simplemente utilizar $g(N)$ en vez de N. Obtenemos

$$\mathrm{Prob}(S'_{g(N)} > a \log\log N) \gg (\log g(N))^{-(a\log a - a + 1) + o(1)}.$$

Si $\log\log N - \log\log g(N) = o(\log\log N)$, entonces $\log g(N) > (\log N)^{1-o(1)}$. Por ende,

$$\mathrm{Prob}(S'_{g(N)} > a \log\log N) \gg (\log N)^{-(a\log a - a + 1) + o(1)}$$

bajo la condición que $\log\log N - \log\log g(N) = o(\log\log N)$. La ecuación (1.5.10) nos permite deducir de esto que

$$\mathrm{Prob}(S_{g(N)} > a \log\log N) \gg (\log N)^{-(a\log a - a + 1) + o(1)} \tag{1.5.11}$$

Empero, procederemos de manera distinta. Mostraremos que la probabilidad que $X - S_{g(N)} > \epsilon' \log\log N$ se satisfaga es muy pequeña – si es que $g(N)$ satisface una cierta condición fácil de cumplir.

Por Chebyshev-Mertens, $\sum_{g(N)<p\leq N} \frac{1}{p} = \log\log N - \log\log g(N) + O(1)$. Procediendo exactamente como en §1.4 (problemas $1a$, $1b$ y comienzo de $1c$), podemos mostrar que

$$\mathbb{E}(e^{\alpha(X - S_{g(N)})}) \leq \left(e^{\log\log N - \log\log g(N) + O(1)}\right)^{e^\alpha}$$

para todo α. (Podríamos continuar como en el resto de §1.4, problema 1, y reemplazar e^α por $e^\alpha - 1$, pero esto no será necesario.) Por Markov,

$$\mathrm{Prob}(X - S_{g(N)} > x) \leq \frac{\left(e^{\log\log N - \log\log g(N) + O(1)}\right)^{e^\alpha}}{e^{\alpha x}} \tag{1.5.12}$$

para cualquier α. Si $\log\log N - \log\log g(N) + O(1) = o(x)$, podemos escoger $\alpha = A + 1$ para A arbitariamente grande, y entonces (1.5.12) nos da que

$$\mathrm{Prob}(X - S_{g(N)} > x) \ll_A e^{-Ax} \tag{1.5.13}$$

para N suficientemente grande. Escogemos $x = \epsilon' \log\log N$ (con $\epsilon' > 0$ arbitrariamente pequeño) y $A = A'/\epsilon'$ (para A' arbitrariamente grande). Concluimos que

$$\mathrm{Prob}(X - S_{g(N)} > \epsilon' \log\log N) \ll_{A'} (\log N)^{-A'} \tag{1.5.14}$$

para N suficientemente grande, si se cumple la condición que $\log\log N - \log\log g(N) = o(\log\log N)$.

Atemos los cabos. Por (1.5.11),

$$\mathrm{Prob}(X > a\log\log N) \geq \mathrm{Prob}(S'_{g(N)} > a \log\log N) \gg (\log N)^{-(a\log a - a+1)+o(1)} \tag{1.5.15}$$

para $a > 1$. Pasemos al caso $a < 1$. Por (1.5.11) y (1.5.14),

$$\begin{aligned} \mathrm{Prob}(X < a\log\log N) &\geq \mathrm{Prob}(S'_{g(N)} < (a - \epsilon')\log\log N) \\ &\quad - \mathrm{Prob}(X - S_{g(N)} > \epsilon' \log\log N) \\ &\gg (\log N)^{-((a-\epsilon')\log(a-\epsilon') - (a-\epsilon') + 1) + o(1)} - O_{A',\epsilon'}((\log N)^{-A'}) \end{aligned}$$

para A' arbitrariamente grande y $\epsilon' > 0$ arbitrariamente pequeño. Como $t \to t\log t - t + 1$ es continua, esto implica que

$$\mathrm{Prob}(X > a \log\log N) \gg (\log N)^{-(a\log a - a+1)+o(1)}. \tag{1.5.16}$$

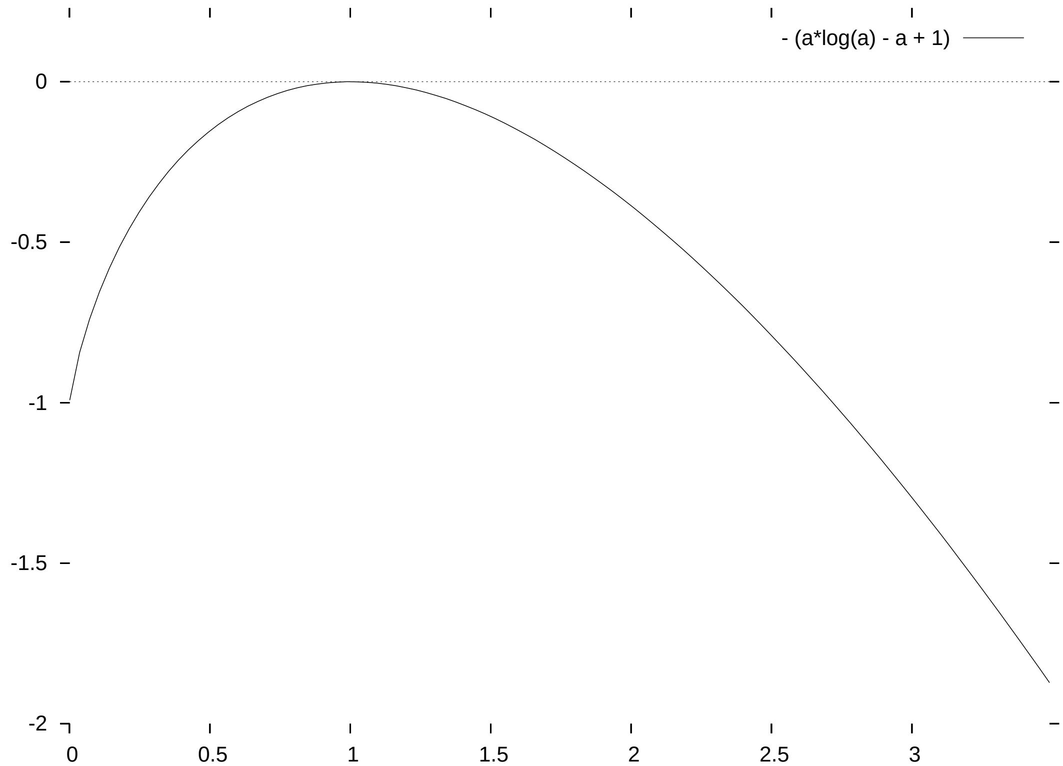


FIGURA 2. La distribución de $\omega(n)$, vista desde lejos, en escala logarítmica. Éste es el gráfico de $y = -I(x) = -(x \log x - x + 1)$. La probabilidad de $\omega(n) > t \log\log N$ (si $t > 1$) o $\omega(n) < t \log\log N$ (si $t < 1$) es igual a $(\log N)^{-I(t)+o(1)}$, lo cual es lo mismo que $(\log N)^{o(1)} \cdot \int_t^\infty (\log N)^{-I(x)} dx$ (si $t > 1$) o $(\log N)^{o(1)} \cdot \int_0^t (\log N)^{-I(x)} dx$ (si $t < 1$).

Queda solamente verificar que existe un $g(N)$ que satisface las condiciones impuestas: $\log g(N) = o\left(\frac{\log N}{\log\log N}\right)$ y $\log\log N - \log\log g(N) = o(\log\log N)$. La función $g(N) = \frac{\log N}{(\log\log N)^2}$ (por ejemplo) satisface ambas condiciones.

Hemos obtenido nuestro objetivo. Recordando (1.4.12) y $X = \omega(n)$, concluimos que

$$\begin{aligned}(\log N)^{-(a\log a+1-a)+o(1)} &\ll_a \mathrm{Prob}(\omega(n) > a\log\log N) &\ll_a (\log N)^{-(a\log a+1-a)} \quad \text{si } a > 1,\\ (\log N)^{-(a\log a+1-a)+o(1)} &\ll_a \mathrm{Prob}(\omega(n) < a\log\log N) &\ll_a (\log N)^{-(a\log a+1-a)} \quad \text{si } a < 1,\end{aligned}$$

donde n es tomado al azar entre 1 y N. Las constantes dependen de a de manera continua.

Notas y problemas

1. *Lema fundamental de las cribas (versión débil).*

   Sea $z = N^{1/s}$, donde $s \to \infty$ cuando $N \to \infty$. Sea P un conjunto de primos $\leq z$. Queremos determinar cuántos enteros $n \leq N$ son coprimos con todo $p \in P$.

a) Cuántos enteros $n \leq N$ son impares? La respuesta es el número de enteros $n \leq N$ (es decir, N) menos el número de enteros $n \leq N$ que son pares (es decir, $\lfloor N/2 \rfloor$). Muy bien – la respuesta es $N - \lfloor N/2 \rfloor = N/2 + O(1)$. Cuántos enteros $n \leq N$ son coprimos con 2 y 3? Tomamos los N enteros $n \leq N$, restamos los $\lfloor N/2 \rfloor$ enteros divisibles por 2 y los $\lfloor N/3 \rfloor$ enteros divisibles por 3, y nos damos cuenta que hemos sustraído los enteros divisibles tanto por 2 como por 3 (es decir, divisibles por 6) por partida doble; tendremos que contarlos una vez de vuelta. Vemos, entonces, que el número de enteros $n \leq N$ coprimos con 2 y 3 es

$$\begin{aligned} N - \left\lfloor \frac{N}{2} \right\rfloor - \left\lfloor \frac{N}{3} \right\rfloor + \left\lfloor \frac{N}{6} \right\rfloor &= N - \frac{N}{2} - \frac{N}{3} + \frac{N}{6} + 3 \cdot O(1) \\ &= \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right) N + O(1). \end{aligned}$$

Seguimos razonando de la misma manera (enfoque que tiene el nombre de *principio de inclusión-exclusión*) y concluimos que la probabilidad que un $n \leq N$ tomado al azar sea coprimo con todo $p \in P$ es igual a

(1.5.17)
$$\begin{aligned} \sum_{Q \subset P} (-1)^{|Q|} \operatorname{Prob}(n \text{ es divisible por todo } p \in Q) &= \sum_{Q \subset P} (-1)^{|Q|} \frac{1}{N} \left\lfloor \frac{N}{\prod_{p \in Q} p} \right\rfloor \\ = \sum_{\substack{Q \subset P \\ \prod_{p \in Q} p \leq N^{1-\epsilon}}} (-1)^{|Q|} \frac{1}{N} \left\lfloor \frac{N}{\prod_{p \in Q} p} \right\rfloor &+ \sum_{\substack{Q \subset P \\ \prod_{p \in Q} p > N^{1-\epsilon}}} (-1)^{|Q|} \frac{1}{N} \left\lfloor \frac{N}{\prod_{p \in Q} p} \right\rfloor \end{aligned}$$

para cualquier $\epsilon > 0$. Muestre que esto es

$$\prod_{p \in P} \left(1 - \frac{1}{p}\right) + O(\text{error}) + O(N^{-\epsilon}) + O(\text{error}),$$

donde

(1.5.18)
$$\text{error} = \sum_{\substack{Q \subset P \\ \prod_{p \in Q} p > N^{1-\epsilon}}} \frac{1}{\prod_{p \in Q} p}.$$

Habremos terminado una vez que acotemos este término de error,

b) Podemos escribir (1.5.18) de la siguiente manera:

(1.5.19)
$$\text{error} = \sum_{k=0}^{\infty} \sideset{}{^*}\sum_{\substack{N^{1-\epsilon} 2^k < n \leq N^{1-\epsilon} 2^{k+1} \\ p|n \Rightarrow p \leq z}} \frac{1}{n} \leq \sum_{k=0}^{\infty} \frac{1}{N^{1-\epsilon} 2^k} \sideset{}{^*}\sum_{\substack{N^{1-\epsilon} 2^k < n \leq N^{1-\epsilon} 2^{k+1} \\ p|n \Rightarrow p \leq z}} 1,$$

donde $\sum^*$ denota que la suma recorre sólo aquellos enteros n que no tienen divisores cuadrados. Lo que es verdaderamente crucial es que todo número n en la sumas $\sum_n$ en (1.5.19) es *desmenuzable*, i.e., no tiene factores primos grandes.

Podemos, entonces, utilizar (1.2.33), que nos da la (más bien pequeña) probabilidad que un número sea desmenuzable (y libre de divisores cuadrados). Obtenemos

$$\text{error} \leq \sum_{k=0}^{\infty} u_k^{-u_k(1+o(1))}, \tag{1.5.20}$$

donde $u_k = \frac{\log N^{1-\epsilon} 2^k}{\log z}$.

c) Como $u_k = a + kb$, donde $a = \frac{\log N^{1-\epsilon}}{\log z} = (1-\epsilon)s$ y $b = \frac{\log 2}{\log z}$, tenemos

$$\begin{aligned} u_k^{-u_k(1+o(1))} &= (a+kb)^{-(a+kb)(1+o(1))} \\ &\leq a^{-a(1+o(1))} e^{-kb(1+o(1))\log a}. \end{aligned}$$

Por lo tanto

$$\text{error} \leq a^{-a(1+o(1))} \sum_{k=0}^{\infty} e^{-k\delta},$$

donde $\delta = b(1+o(1))\log a$. Muestre que $\sum_{k=0}^{\infty} e^{-k\delta} \ll \frac{1}{\delta}$ para $\delta > 0$. Así,

$$\begin{aligned} \text{error} &\ll \frac{a^{-a(1+o(1))}}{\delta} \ll \frac{\log z}{\log s} \cdot ((1-\epsilon)s)^{-(1-\epsilon)s\cdot(1+o(1))} \\ &\ll (\log z) s^{-(1-2\epsilon)s(1+o(1))}. \end{aligned}$$

El error $(\log z) \cdot s^{-(1-2\epsilon)s\cdot(1+o(1))}$ será $O((\log N)^{-A})$ cuando $s > 3A\frac{\log\log N}{\log\log\log N}$. Si $s \gg \log\log N$, el error será $\ll_A (\log N)^{-A}$ para todo $A > 0$.

d) Hemos probado

Teorema 1.8 (Lema fundamental de las cribas, version débil)**.** *Sea* $z = N^{1/s}$*, donde* $s \gg \log\log N$*. Sea* P *un subconjunto de* $\{p \leq z : p \text{ primo}\}$*. Entonces, para* n *tomado al azar entre* 1 *y* N*,*

$$\text{Prob}(n \textit{ es coprimo con todo } p \in P) = (1+o(1)) \cdot \prod_{p\in P}\left(1 - \frac{1}{p}\right). \tag{1.5.21}$$

Explícitamente,

$$\begin{aligned} &\text{Prob}(n \textit{ es coprimo con todo } p \in P) \\ &\quad = \prod_{p\in P}\left(1-\frac{1}{p}\right) + O\left((\log z)\cdot s^{-(1-2\epsilon)s\cdot(1+o(1))}\right) + O(N^{1-\epsilon}) \\ &\quad = (1 + O_A((\log N)^{-A})) \cdot \prod_{p\in P}\left(1-\frac{1}{p}\right) \end{aligned} \tag{1.5.22}$$

para cualquier $\epsilon > 0$ *y cualquier* $A > 0$*.*

Nota. La versión fuerte del lema fundamental de las cribas consiste en la aseveración que (1.5.21) rige no sólo para $s \gg \log\log N$, sino para todo $s \to \infty$. No probaremos la versión fuerte aquí.

2. El lema fundamental es muy razonable: como la probabilidad que n sea coprimo con un $p \in P$ dado es $1 - 1/p$, la ecuación (1.5.21) nos dice (como muchas otras cosas que hemos probado) que los eventos $p|n$ se comportan hasta cierto punto como variables mutuamente independientes. Empero, si $z = N^{1/s}$ y s no va a ∞ cuando $N \to \infty$, la ley (1.5.21) no rige.

   Mostremos esto en el caso más simple. Si $z = N^{1/2}$ y P es el conjunto de todos los primos $p \leq P$, entonces n es coprimo con todo $p \in P$ sí y sólo sí n es primo. La probabilidad que n sea primo es

$$(1 + o(1)) \cdot \frac{1}{\log N} \tag{1.5.23}$$

   (este es el *teorema de los números primos*, el cual no probaremos). La aseveración (1.5.21) nos daría, de otro lado, $(1 + o(1)) \cdot \prod_{p \leq N^{1/2}} \left(1 - \frac{1}{p}\right)$. Se puede mostrar que

$$\prod_{p \leq N^{1/2}} \left(1 - \frac{1}{p}\right) = \frac{2e^{-\gamma}}{\log N}(1 + o(1)), \tag{1.5.24}$$

   donde γ es la constante de Euler $\gamma = 0{,}577\ldots$. Lo importante aquí es que $2e^{-\gamma} \neq 1$, por lo cual la predicción natural (1.5.24) no es compatible con la realidad (1.5.23).

3. *Desviaciones moderadas.*
   a) Si $X - \mathbb{E}(X)$ está en la escala de $\sqrt{\mathrm{Var}(X)}$ (es decir, entre $\frac{1}{1000}\sqrt{\mathrm{Var}(X)}$ y $1000\sqrt{\mathrm{Var}(X)}$, por ejemplo), hablamos del *límite central*, o de *pequeñas desviaciones*. Si $X - \mathbb{E}(X)$ está en la escala de $\mathrm{Var}(X)$ o $\mathbb{E}(X)$, hablamos de *grandes desviaciones*. Aún no hemos examinado el caso en el cual $X - \mathbb{E}(X)$ es bastante más grande que $\sqrt{\mathrm{Var}(X)}$ y bastante más pequeño que $\mathrm{Var}(X)$; como cabría esperar, esto se llama una *desviación moderada*.
   Lo que queremos examinar es

$$\mathrm{Prob}(X > (1 + \Delta(N)) \log\log N) \;\text{ y }\; \mathrm{Prob}(X < (1 - \Delta(N)) \log\log N),$$

   donde $\Delta(N) > 0$, $\Delta(N) = o(1)$ y $(\log\log N)^{-1/2} = o(\Delta(N))$. (Si $\Delta(N) = o(1)$ no se cumpliera, tendríamos una gran desviación; si $(\log\log N)^{-1/2} = o(\Delta(N))$ no se cumpliera, tendríamos una pequeña desviación.)
   b) Las cotas superiores de grandes desviaciones son aún válidas, como podemos ver repasando sus pruebas. Por lo tanto:

$$\mathrm{Prob}(X > (1 + \Delta(N)) \log\log N) \leq (\log N)^{-I(1+\Delta(N))},$$

   donde $I(a) = a \log a - a + 1$. Mediante una serie de Taylor, verifique que

$$I(1 + \Delta(N)) = \frac{1}{2}\Delta^2(N) + O(\Delta^3(N)) = \frac{1}{2}\Delta^2(N) \cdot (1 + o(1)).$$

   Concluimos que

$$\mathrm{Prob}(X > (1 + \Delta(N)) \log\log N) \leq (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))},$$

y, similarmente,

$$\mathrm{Prob}(X < (1-\Delta(N))\log\log N) \leq (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))}.$$

c) Las cotas inferiores tendrán que ser rehechas con más cuidado: el sumando de $o(1)$ en el exponente de (1.5.15) y (1.5.16) es ahora demasiado burdo, ya que $-(a\log a + 1 - a)$ será bastante más pequeño que $o(1)$.
Las cotas (1.5.6) y (1.5.7) son aún válidas. Muestre que, como estamos asumiendo $(\log\log N)^{-1/2} = o(\Delta(N))$, las cotas (1.5.6) y (1.5.7) toman la forma

$$\begin{aligned}\mathrm{Prob}(X' > (1+\Delta(N))\log\log N) &\gg (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))}\\ \mathrm{Prob}(X' < (1-\Delta(N))\log\log N) &\gg (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))}.\end{aligned} \tag{1.5.25}$$

d) La transición de X' a X se hace como antes. La única dificultad reside en el hecho que ya no basta probar que $\mathrm{Prob}(X - S_{g(N)} > \epsilon' \log\log N)$ es pequeña; debemos asegurarnos que $\mathrm{Prob}(X - S_{g(N)} > \epsilon'\Delta(N)\log\log N)$ sea pequeña. Para esto tendremos que modificar el valor de $g(N)$.
Sustituya $x = \epsilon'\Delta(N)\log\log N$ y $A = 1/\epsilon$ en (1.5.13) y obtenga

$$\mathrm{Prob}(X - S_{g(N)} > \epsilon'\Delta(N)\log\log N) \leq (\log N)^{-\Delta(N)}, \tag{1.5.26}$$

bajo la suposición que $\log\log N - \log\log g(N) + O(1) = o(\Delta(N)\log\log N)$. Como $\frac{1}{2}\Delta^2(N) = o(\Delta(N))$, el término de error (1.5.26) es mucho más chico que el término principal $(\log N)^{-\frac{1}{2}\Delta^2(N)(1+o(1))}$.

e) Muestre que se puede definir $g(N)$ de tal manera que

$$\log g(N) = o\left(\frac{\log N}{\log\log N}\right) \qquad \text{y} \qquad \log\log N - \log\log g(N) = o(\Delta(N)\log\log N).$$

Recuerde que $\Delta(N) \gg \sqrt{\log\log N}$.) Concluya que

$$\begin{aligned}\mathrm{Prob}(X' > (1+\Delta(N))\log\log N) &\ll (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))}\\ \mathrm{Prob}(X' < (1-\Delta(N))\log\log N) &\ll (\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))}.\end{aligned} \tag{1.5.27}$$

f) Podemos escribir $t = \Delta(N)\cdot(\log\log N)^{1/2}$. (La desviación es entonces $\Delta(N)\cdot(\log\log N) = t\sqrt{\log\log N}$.) Está claro que

$$(\log N)^{-\frac{1}{2}\Delta^2(N)\cdot(1+o(1))} = e^{-\frac{1}{2}t^2(1+o(1))}$$

Así, (1.5.25) y (1.5.27) nos dicen que la normal nos da, por lo menos, una idea aproximada de la escala de la probabilidad de las desviaciones moderadas.

4. Podemos utilizar el lema fundamental (aún en su versión debil) para dar una prueba alternativa del teorema de Erdős-Kac (teorema 1.3). (Lo siguiente está muy cercano del camino seguido originalmente por Erdős y Kac mismos.) Sea $g(x)$ una función tal que $g(x) = O\left(x^{1/\log\log x}\right)$ y $\log\log x - \log\log g(x) = o(\sqrt{\log\log N})$; podemos tomar $g(x) = x^{1/\log\log x}$, como en la primera prueba que dimos del teorema.

Procédase como en esa prueba hasta (1.3.13), i.e., muestre que podemos trabajar con la suma truncada

$$S_{g(N)} = \frac{1}{\sqrt{\log\log g(N)}} \cdot \sum_{p \leq g(N)} (X_p - 1/p).$$

Podemos luego utilizar el lema fundamental en la misma manera que lo usamos en esta sección, y así mostrar que

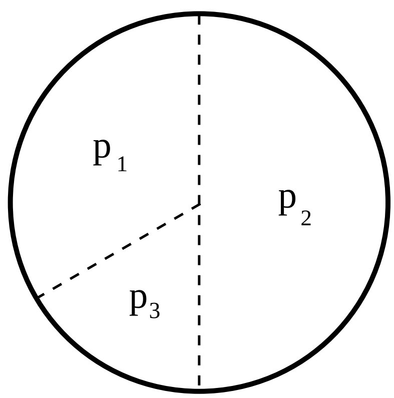
$$\mathrm{Prob}(S_{g(N)} \leq t) = \mathrm{Prob}(S'_{g(N)} \leq t) + o(1)$$

para todo t, donde $S'_{g(N)}$ es la suma de variables mutuamente independientes X'_p:

$$S'_{g(N)} = \frac{1}{\sqrt{\log\log g(N)}} \cdot \sum_{p \leq g(N)} (X'_p - 1/p).$$

En otras palabras, $S_{g(N)}$ tiende a la misma distribución que $S'_{g(N)}$ – es decir, tiende a la normal.

5. *Entropía relativa.* Consideremos tres compartimientos separados por membranas porosas. Llenémoslos de gas. Sea p_j la probabilidad que una partícula dada (todas son intercambiables) esté en el compartimiento j:



Cuál es la probabilidad que haya $n_1 = r_1 \cdot n$ partículas en el primer compartimiento, $n_2 = r_2 \cdot n$ en el segundo, y $n_3 = r_3 \cdot n$ en el tercero?

La probabilidad de una configuración específica – es decir, la probabilidad que n_j partículas específicas estén en el compartimiento j – es

$$p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3}.$$

El número de configuraciones con n_j partículas en la cámara j es

$$\frac{n!}{n_1! \cdot n_2! \cdot n_3!}.$$

Luego, la probabilidad que haya $n_j = r_j \cdot n$ partículas en el compartimiento j es

$$P = p_1^{n_1} p_2^{n_2} p_3^{n_3} \frac{n!}{n_1! \cdot n_2! \cdot n_3!}.$$

Cuánto es esto, aproximadamente? Como en el caso que vimos al principio de la sección, extraemos el logaritmo:

$$\begin{aligned}\log P &= n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n \log n - n + O(\log n)\\ &\quad - ((n_1 \log n_1 - n_1) + (n_2 \log n_2 - n_2) + (n_3 \log n_3 - n_3) + O(\log n))\\ &= -\left(n_1 \log \frac{n_1/n}{p_1} + n_2 \log \frac{n_2/n}{p_2} + n_3 \log \frac{n_3/n}{p_3}\right) + O(\log n)\\ &= -n\left(r_1 \log \frac{r_1}{p_1} + r_2 \log \frac{r_2}{p_2} + r_3 \log \frac{r_3}{p_3}\right) + O(\log n).\end{aligned}$$

Por lo tanto,

$$P = e^{-n(H+o(1))}$$

donde $H = r_1 \log \frac{r_1}{p_1} + r_2 \log \frac{r_2}{p_2} + r_3 \log \frac{r_3}{p_3}$. En general, definimos

$$H = \sum r_j \log \frac{r_j}{p_j} \qquad \text{(entropía relativa)}$$

y tenemos que la probabilidad que una proporción r_j de las partículas estén en el comportamiento j es

$$P = e^{-n(H+o(1))}.$$

6. *La entropía relativa y los primos.* Sea una partición

$$\{\text{los primos}\} = P_1 \cup P_2 \cup P_3 \qquad (P_1 \cap P_2 = P_1 \cap P_3 = P_2 \cap P_3 = \emptyset)$$

tal que, para cada $j = 1, 2, 3$,

$$\left(\frac{\log z_1}{\log z_0}\right)^{p_j} \ll \prod_{\substack{p \in P_j \\ z_0 < p \leq z_1}} \left(1 + \frac{1}{p}\right) \ll \left(\frac{\log z_1}{\log z_0}\right)^{p_j}$$

Sea

$$\begin{aligned}\omega_j(n) &= \text{número de divisores primos de } n \text{ en } P_j\\ &= |\{p \in P_j : p|n\}|\end{aligned}$$

Queremos estimar

$$\operatorname{Prob}((a_j - \epsilon) \log \log N < \omega_j(n) \leq (a_j + \epsilon) \log \log N \quad \forall j)$$

Por ejemplo: si $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{2}$, $p_3 = \frac{1}{3}$, cuál es la probabilidad que N tenga $\sim \frac{1}{2} \log \log N$, $\sim 5 \log \log N$ y $\sim \frac{1}{10} \log \log N$ divisores de cada tipo, respectivamente?

a) Como en §1.4, problema 1, pruebe que

$$\sum_{n \leq N} e^{\alpha_1 \omega_1(n) + \alpha_2 \omega_2(n) + \alpha_3 \omega_3(n)}/n \ll (\log N)^{p_1 e^{\alpha_1} + p_2 e^{\alpha_2} + p_3 e^{\alpha_3}}$$

para α_1, α_2, α_3 cualesquiera, y, prosiguiendo como en el mismo problema,

$$\mathbb{E}\left(e^{\sum_j \alpha_j \omega_j(n)}\right) \ll (\log N)^{\sum_j p_j e^{\alpha_j} - 1}. \tag{1.5.28}$$

b) Por Markov,

$$\mathrm{Prob}((a_j - \epsilon)\log\log N < \omega_j(n) \leq (a_j + \epsilon)\log\log N \quad \forall j)$$

es a lo más

$$\frac{\mathbb{E}\left(e^{\sum_j a_j \omega_j(n)}\right)}{e^{\sum \alpha_j a_j \log\log N + O_{\sum_j a_j}(\epsilon)}} \tag{1.5.29}$$

para α_j arbitrarios. Utilice (1.5.28) y encuentre los α_j para los cuales (1.5.29) es mínimo. Muestre, en conclusión, que

$$\mathrm{Prob}((a_j - \epsilon)\log\log N < \omega_j(n) \leq (a_j + \epsilon)\log\log N \quad \forall j) \ll (\log N)^{-I_{\vec{p}}(\vec{a}) + O(\epsilon)},$$

donde

$$I_{\vec{p}}(\vec{a}) = \left(\sum a_j \log \frac{a_j}{p_j}\right) + 1 - \sum_j a_j.$$

c) Proceda como en la sección presente para mostrar que

$$\mathrm{Prob}((a_j - \epsilon)\log\log N < \omega_j(n) \leq (a_j + \epsilon)\log\log N \quad \forall j) \gg (\log N)^{-I_{\vec{p}}(\vec{a}) + O(\epsilon) + o(1)}.$$

La cantidad $I_{\vec{p}}(\vec{a})$ es llamada la *entropía relativa* de $\vec{a}$ con respecto a $\vec{p}$.

APÉNDICE A

Rudimentos de probabilidades

Un *evento aleatorio* E es algo que puede ya sea ocurrir o no: digamos, la lluvia de mañana. Los ejemplos extremos son los eventos con probabilidad 0 – es decir, los que se sabe con certeza que no ocurrirán – y los eventos con probabilidad 1 – es decir, los que se sabe con certeza que ocurrirán. Todo otro evento tiene probabilidad entre 0 y 1. Denotamos la probabilidad del evento E mediante $\mathrm{Prob}(E)$.

Una *variable aleatoria* X puede tomar cualquier valor dentro de un conjunto. Los casos más comunes son las variables que toman valores dentro de un conjunto finito o infinito de enteros ("variables discretas") y las variables que toman valores dentro de los reales o algún otro espacio vectorial dado ("variables continuas"). La cantidad de lluvia que caerá mañana es un ejemplo del segundo tipo de variable; el número de días de lluvia del año próximo es un ejemplo del primero. Los eventos aleatorios son, claro está, un caso particular del primer tipo: pueden verse como las variables que toman los valores 0 y 1, o "no" y "sí". Tales variables son llamadas *variables de Bernoulli*:

$$X = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p. \end{cases}$$

La *función de cuantía* o *función de probabilidad* de una variable discreta X es la función f que asigna a cada valor posible x su probabilidad $f(x) = \mathrm{Prob}(X = x)$. La suma $\sum f(x)$ siempre es 1, ya que la probabilidad que la variable tome alguno de los posibles valores es 1. La función de cuantía de una variable de Bernoulli es

$$f(x) = \begin{cases} p & \text{para } x = 1 \\ 1-p & \text{para } x = 0 \\ 0 & \text{para todo otro } x. \end{cases}$$

La función de cuantía de la variable "cara de un dado" sería

$$f(x) = \begin{cases} 1/6 & \text{si } x \in \{1,2,3,4,5,6\} \\ 0 & \text{de otra manera,} \end{cases}$$

a menos, por supuesto, que el dado esté trucado.

Una variable continua generalmente toma cada uno de sus valores posibles con probabilidad 0: la probabilidad que caigan exactamente π centímetros de lluvia mañana es cero, o infinitesimal. Sin embargo, una tal variable aún puede ser descrita por una función de probabilidad, llamada, en este caso, *función de densidad*. Digamos que la variable en cuestión toma valores en $\mathbb{R}$. La función de densidad de una tal variable es una función $f : \mathbb{R} \to \mathbb{R}_0^+$ cuya integral es 1. La probabilidad que la variable tome su valor entre a y b

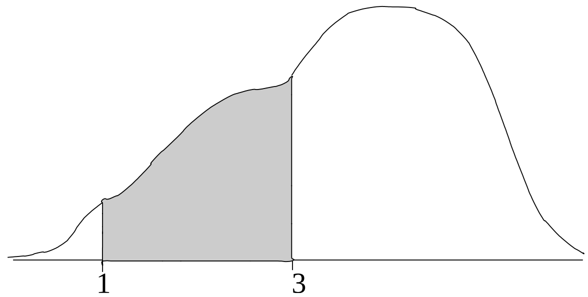


Figura 1. Distribución de una variable continua. La probabilidad que la variable tome un valor entre 1 y 3 es igual al área marcada.

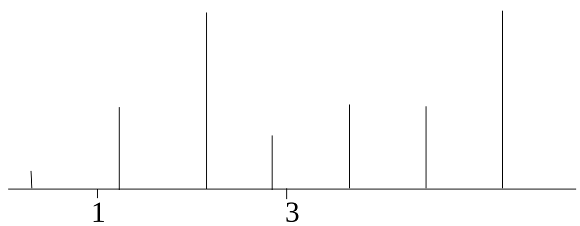


Figura 2. Distribución de una variable discreta. La probabilidad que la variable tome un valor entre 1 y 3 es igual a la suma de los valores $f(x)$ de la función de distribución para x entre 1 y 3, o, lo que es lo mismo, a la suma de las alturas de las barras entre 1 y 3.

está dada por la integral

$$\int_a^b f(x)dx.$$

Decimos que una variable X tiene la *distribución uniforme* si todos sus valores son igualmente probables. Tanto en el caso continuo como en el caso discreto, X tiene la distribución uniforme si y sólo si su función f es una función constante. Tanto un dado justo como una moneda justa tienen la distribución uniforme – si se define el dominio como $\{1, 2, \ldots, 6\}$ y $\{\text{cara}, \text{sello}\}$ en el otro, claro está.

Muy a menudo, los mismos enunciados y las mismas pruebas valen para las variables discretas y continuas si se utilizan sumas en un caso e integrales en el otro. También puede haber variables con un rango en parte continuo y en parte discreto. Por ello, lo correcto es tener un sólo marco para todas las distribuciones, de tal manera que la distinción entre las variables discretas y las variables continuas desaparezca en el plano formal. El lector puede adivinar que tal marco nos es dado por la integración de Lebesgue; en dicha perspectiva, las sumas son un caso particular de las integrales.

Tal es el formalismo aceptado en estos días, por excelentes razones. Empero, no nos preocuparemos, y hablaremos como si nuestras variables fueran discretas o continuas dependiendo de lo que haga que nuestra notación sea más conveniente.

Una alternativa elemental consiste en definir la *función de distribución acumulada* $P_X(x)$ de la siguiente manera:

$$P_X(x) = \text{Prob}(X \leq x). \tag{A.0.1}$$

En el caso continuo, $P(x) = \int_{-\infty}^{x} f(t)dt$; en el caso discreto, $P(x) = \sum_{t\leq x} f(t)$. Como la definición (A.0.1) es válida en ambos casos, uno puede utilizar $P_X(x)$ en vez de $f(t)$ y de esta manera hablar de ambos tipos de distribución a la vez.

No utilizaremos la función de distribución acumulada. Cuando decimos "la distribución $f(x) = e^{-x}$" o "la distribución $f(n) = \frac{1}{e\cdot n!}$", queremos decir "la distribución continua con función de densidad $f(x) = e^{-x}$" o "la distribución discreta dada por la función de cuantía $f(n) = \frac{1}{e\cdot n!}$", respectivamente; el caso que se tiene en mente estará claro en el contexto.

* * *

La *probabilidad condicional* $\text{Prob}(E_1|E_2)$ es la probabilidad de un evento E_1 dado que el evento E_2 ocurre. Tenemos

$$\text{Prob}(E_1|E_2) = \frac{\text{Prob}(E_1 \wedge E_2)}{\text{Prob}(E_2)},$$

donde $E_1 \wedge E_2$ se define como el evento que tanto E_1 como E_2 ocurran. (Si la probabilidad que llueva mañana es 0,1, y la probabilidad que llueva y enfríe es 0,07, la probabilidad que enfríe, dado que lloverá, es 0,7.)

Decimos que dos eventos E_1, E_2 son *independientes* si

$$\text{Prob}(E_1|E_2) = \text{Prob}(E_1) \quad \text{y} \quad \text{Prob}(E_2|E_1) = \text{Prob}(E_2).$$

Decimos que dos variables X, Y son independientes si

$$\text{Prob}(X = x|Y = y) = \text{Prob}(X = x) \quad \text{y} \quad \text{Prob}(Y = y|X = x) = \text{Prob}(Y = y).$$

para x, y cualesquiera. En otras palabras, X y Y son independientes si el valor tomado por una no nos dice nada acerca del valor de la otra. (Digamos: saber que mañana lloverá en Alaska no nos ayuda a saber si es que mañana lloverá en Iquitos.)

Si tenemos variables $X_1, X_2, \ldots, X_n$, decimos que son *independientes en pares* si X_i y X_j son independientes para $1 \leq i, j, \leq n$, $i \neq j$ cualesquiera. Decimos que $X_1, X_2, \ldots, X_n$ son *mutuamente independientes* si

$$\text{Prob}(X_i = x_i|X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots, X_n = x_n)$$

para $x_1, x_2, \ldots, x_n$ e i cualesquiera.

Ejercicio A.1. Muestre que, si $X_1, X_2, \ldots, X_n$ son mutuamente independientes, entonces son independientes en pares.

Sin embargo, si $X_1, X_2, \ldots, X_n$ son independientes en pares, ello no es de ninguna manera suficiente para que sean mutuamente independientes.

* * *

Sea X una variable que toma valores dentro de los reales (o los complejos). La *esperanza* $\mathbb{E}(X)$ de X es

$$\mathbb{E}(X) = \sum_x \text{Prob}(X = x) \cdot x$$

(o $\int_{-\infty}^{\infty} f(x)xdx$ en el caso continuo, donde f es la función de densidad). En otras palabras, se trata de una especie de promedio.

Ejercicio A.2. Muestre que, si X e Y son independientes, entonces

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Ésta es una condición necesaria, pero no suficiente, para que X e Y sean independientes.

Sean X una variable y E un evento. Definimos la *esperanza condicional* $\mathbb{E}(X|E)$ de X dado E de la siguiente manera:

$$\mathbb{E}(X|E) = \sum_x \mathrm{Prob}(X = x|E) \cdot x.$$

* * *

Existen diversas maneras de describir una variable, más allá de su distribución (que nos da una descripción completa) y su esperanza. La más común es la *varianza*:

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

La varianza será grande si los valores de X tienden a alejarse mucho de la esperanza de X, y pequeña si esto no sucede.

La *desviación estándar* $\sigma(X)$ no es sino la raíz cuadrada de la varianza:

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

La desviación estándar nos da una buena idea de la escala de las desviaciones – esto es, las distancias de los valores de X de la esperanza de X. Puede suceder que X esté a una o dos desviaciones estándar de su esperanza gran parte del tiempo, pero podrá estar a más de diez desviaciones estándar de distancia de su esperanza a lo más una de cada 100 veces (*desigualdad de Chebyshev*).

La varianza y la desviación estándar no distinguen entre las veces en que X toma valores más grandes y más pequeños que su esperanza. Hablamos de la *cola superior* de la distribución cuando queremos referirnos a aquellos valores posibles de X que son mucho más grandes que $\mathbb{E}(X)$; decimos *cola inferior* para referirnos a los valores que son mucho más pequeños que $\mathbb{E}(X)$.

* * *

Supongamos que tenemos dos variables X e Y con la misma función de densidad, excepto por un argumento; digamos, por ejemplo, que la función de densidad de X es $g(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{de otra manera,} \end{cases}$ y la función de densidad de Y es $h(x) = \begin{cases} 1 & \text{si } 0 < x \leq 1 \\ 0 & \text{de otra manera} \end{cases}$. Entonces tiene sentido decir que X e Y poseen la misma distribución: la probabilidad que X sea exactamente 1 es infinitesimal, de todas maneras, y la probabilidad que X esté entre a y b (para $a < b$ cualesquiera) es igual a la probabilidad que Y esté entre a y b.

Esto sugiere la definición siguiente. Sean dadas una variable Z y una sucesión de variables $Z_1, Z_2, Z_3, \ldots$. Decimos que las variables $Z_1, Z_2, Z_3, \ldots$ *convergen en distribución a* Z si, para $a < b$ cualesquiera,

$$\lim_{n\to\infty} \mathrm{Prob}(a < Z_n < b) = \mathrm{Prob}(a < Z < b). \tag{A.0.2}$$

Desde el punto de vista de la integración de Lebesgue, esta no es sino la "convergencia débil". Por qué? Ahora bien, si una sucesión de funciones de densidad (o cuantía) f_n converge en el sentido de la convergencia débil a una función de densidad (o cuantía) f, la sucesión de funciones de distribución acumulada $P_n(x) = \int_{-\infty}^{x} f_n(t)dt$ converge para cada x a la función de distribución acumulada $P(x) = \int_{-\infty}^{x} f_n(t)dt$. (Esto no es sino (A.0.2).) No es difícil probar que la convergencia de $P_n(x)$ a $P(x)$ es, incluso, uniforme en x. (Utilize $f_n(t), f(t) \geq 0$ y $\int_{-\infty}^{\infty} f_n(t)dt = \int_{-\infty}^{\infty} f(t)dt = 1$.)

Muy bien puede suceder que las variables Z_n sean discretas, y que su límite Z una variable continua.

APÉNDICE B

Comentarios finales

En esta introducción a la teoría probabilística de números, nos centramos en el estudio de los divisores primos de un número aleatorio, y no en el estudio de un primo aleatorio, o de los primos cercanos a un entero aleatorio. La razón principal es que aún se sabe poco con certeza acerca de estas otras preguntas. Por ejemplo, se conjetura, pero no se sabe, la probabilidad que un entero aleatorio n sea primo y que $n+2$ también sea primo (§1.2, nota 3). Las más de las veces, 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 (ver, por ejemplo, [**7**]), 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 sería una posibilidad si estuvieramos hablando de variables independientes.

En todo el texto, obedecimos a una restricción autoimpuesta: nos abstuvimos de usar teoría de la medida y análisis complejo. Para proseguir en el estudio de la teoría probabilística de números, es necesario usar los dos. El análisis complejo es sumamente útil para el estudio de los primos. Toda la teoría analítica de números depende del análisis; se trata de un caso clásico de cómo el estudio de lo continuo puede ayudar en el estudio de lo discreto. La idea principal es que, para estudiar sumas finitas, como $\sum_{n\leq N}\Lambda(n)$, N variable, debemos estudiar sumas infinitas $\sum_n \Lambda(n)n^{-s}$, s variable. Estas sumas infinitas se tratan como funciones complejas de s, generalmente analíticas o meromórficas.

La teoría de la medida se considera hoy en día como necesaria para desarrollar la teoría de probabilidades sobre una base rigurosa. Si se profundiza en el estudio de la teoría probabilística de números sobre una base puramente intuitiva, se llega fácilmente al punto donde el lenguaje mismo falta. Veamos, por ejemplo, el caso de las caminatas aleatorias. Tomemos un número n al azar. Consideremos los primos $p = 2, 3, 5, \ldots$ en orden. En cada paso, si p divide n, damos un paso a la derecha; si p no divide n, damos un paso mucho más corto a la izquierda. Billingsley [**1**] probó que la caminata que resulta tiende en distribución al mismo límite que una caminata aleatoria, es decir, el movimiento Browniano. Ahora bien, qué quiere decir que una caminata “tiende” a un cierto tipo de crecimiento en distribución? No se trata simplemente de una sucesión de números que convergen a otro número. Resulta difícil formular el resultado – ni que decir de su prueba – sin la teoría de la medida.

Hay dos desarrollos recientes notables. En primer lugar, la teoría de cribas ha probado ser más flexible y potente de lo que se creía hasta ahora, si se suplementa con otros métodos; están allí resultados inesperados en la teoría analítica de números obtenidos por Friedlander-Iwaniec, Goldston-Pintz-Yildirim, y otros. En segundo lugar, la teoría ergódica no sólo esta iluminando la teoría de números – incluído el estudio de los primos – sino que esta haciendo posibles pruebas de resultados que no se creían anteriormente accesibles. Un ejemplo muy reciente e impresionante es el teorema de Green y Tao sobre los números primos en progresiones aritméticas; no es que su prueba establezca propiedades sumamente delicadas de los números primos, sino más bien que muestra que algunas leyes muy precisas no son delicadas, al punto que deben regir para todo subconjunto de los enteros con ciertas propiedades generales – propiedades que los primos satisfacen. Toda la teoría ergódica esta basada sobre la teoría de la medida, y sería imposible comenzar el estudio de la primera sin utilizar la segunda.

La bibliografía tiene como fin ser útil antes que completa. El libro de Tenenbaum [**9**] es una introducción estándar y detallada al tema, con mucho más análisis que la presente monografía. El libro de Iwaniec y Kowalski [**8**] se ha vuelto la obra canónica de la teoría analítica de números para la época actual. Feller [**6**] es un texto clásico de probabilidades del cual generaciones se han beneficiado. Todos los artículos citados están entre los esenciales sobre el tema; la mayor parte de ellos son de lectura razonablemente accesible.

Bibliografía

[1] P. Billingsley, Prime numbers and Brownian motion, *Amer. Math. Monthly*, **80** (1973), 1099–1115.

[2] P. Billingsley, *Probability and measure*, 3rd ed., Wiley, 1995.

[3] P. Erdős, Una desigualdad asintótica en la teoría de números (en ruso), *Vestnik Leningrad Univ. Mat. Mekh. i Astr* **13** (1960), 41–49.

[4] P. Erdős and Mark Kac, The Gaussian law of errors in the theory of additive number theoretic functions, *Am. J. of Math.* **62** No. 1/4, (1940), pp. 738-742.

[5] G.H. Hardy y S. Ramanujan, The normal number of prime factors of a number, *Quart. J. Math.* **48** (1917), pp. 76-92.

[6] W. Feller, *An introduction to probability theory and its applications*, 2nd ed., Wiley, 1970, 2 vols.

[7] A. Granville, Harald Cramér and the distribution of prime numbers, *Scand. Actuarial J.* **1** (1995), pp. 12–28.

[8] H. Iwaniec and E. Kowalski, *Analytic number theory*, AMS Colloquium Publications, v. 53, American Mathematical Society, Providence, 2004.

[9] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.

[10] P. Turán, On a theorem of Hardy and Ramanujan, *J. London Math. Soc.* **9** (1934), 274–276.

Índice alfabético

$O(f), o(f)$, 2
$\Lambda(n)$, 5
$\omega(n)$, 2
$\tau(n)$, 1

teorema
  de convergencia de Lévy, 27

Billingsley, P., 23

Chebyshev, P., 7
condición de Lindeberg, 22, 28, 29
conjetura de Hardy-Littlewood, 14
conjetura de los primos gemelos, 14
convergencia en distribución, 56, 59
convolución, 37
criba
  de Selberg, 13
  lema fundamental, 41–43, 45, 47

de la Vallée Poussin, C., 7
desigualdad
  de Chebyshev, 8, 9, 23, 32, 56
  de Markov, 3, 8, 32
desviación estandar, 56
desviaciones
  grandes, 31, 38, 48
  moderadas, 48
distribución
  de Poisson, 30
  normal, 19, 28, 30
distribución uniforme, 54
divisores, 1
  divisores primos, 2

entropía, 38, 39
  relativa, 50
Erdős, P., 14, 23, 37
esperanza, 1, 55
esperanza condicional, 56

función característica, 20
función de densidad, 53
función de probabilidad, 53
función de von Mangoldt, 5
función zeta de Riemann, 7

Hadamard, J., 7
Halberstam, H., 23

independencia
  de variables, 55
  de eventos, 55
  en pares, 55
  mútua, 55
intervalos diádicos, 6

Kac, M., 23

límite central, 18, 23, 48
  teorema del, 19, 21, 22
ladeo exponencial, 40

modelo de Cramér, 59
momentos, 15, 21
  método de momentos, 21
  momentos variables, 15, 17
momentos exponenciales, 32

número de primos hasta N, 5, 9
números desmenuzables, 15, 17, 46

P. Lévy, 27
principio de inclusión-exclusión, 46
probabilidad condicional, 55
problema de la tabla de multiplicación, 14, 34, 37

sumas por partes, 4

teorema
  de Chebyshev-Mertens, 2, 5, 7
  de convergencia de Lévy, 21

de Erdős-Kac, 23, 49
de los números primos, 7, 14, 48
fundamental de la aritmética, 5
transformada de Fourier, 19, 21, 25

valor crítico, 35
variables de Bernoulli, 53
varianza, 7, 56