# Capítulo 2

# Métodos de criba

## 2.1. Inclusión-exclusión e ideas básicas

**Inclusión-exclusión. Acotación elemental de $\pi(x)$.**

Viajemos por la historia a la leyenda hasta los tiempos de Eratóstenes. Según se dice, para confeccionar su tabla de números primos tomaba el 2 y tachaba todos sus múltiplos (propios); tras el 2, el primero que se había salvado de la criba es el 3; repitiendo la operación con él se eliminarán todos los números coprimos con 6 distintos de 2 y 3. Si este proceso se continuase indefinidamente (y nos olvidamos del 1), se obtendría la lista de los números primos.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1ª pasada | | **2** | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | 9 | ~~10~~ | 11 | ~~12~~ | 13 | ~~14~~ | 15 |
| | ~~16~~ | 17 | ~~18~~ | 19 | ~~20~~ | 21 | ~~22~~ | 23 | ~~24~~ | 25 | ~~26~~ | 27 | ~~28~~ | 29 | ~~30~~ |
| 2ª pasada | | 2 | **3** | | 5 | | 7 | | ~~9~~ | | 11 | | 13 | | ~~15~~ |
| | | 17 | | 19 | | ~~21~~ | | 23 | | 25 | | ~~27~~ | | 29 | |
| 3ª pasada | | 2 | 3 | | **5** | | 7 | | | | 11 | | 13 | | |
| | | 17 | | 19 | | | | 23 | | ~~25~~ | | | | 29 | |

Como máquina de generar primos, el procedimiento de Eratóstenes no se muestra muy práctico, y más hoy en día cuando en la más oscura oficina se oyen tartamudeos de disco duro y murmullos de ventilador. Sin embargo pertenece a los prolegómenos teóricos de algunas técnicas combinatorias agrupadas bajo el nombre de *métodos de criba* que se han mostrado muy efectivas en el estudio de problemas variopintos que involucran los primos. En realidad es poco más que una cortesía histórica apelar al nombre de Eratóstenes, pues en nada es comparable la simpleza de poner cruces (a pesar de las quinielas) con el refinamiento técnico alcanzado por los métodos de criba. La contribución temprana de Legendre está más cercana del contenido de esta sección pero aun así no es vituperable afirmar que los métodos de criba comenzaron con V. Brun [Br] en 1915.

Tanto Eratóstenes como nosotros encontramos cansino tachar infinitos números y por tanto nos restringiremos a los primos en $[1, N]$. Todavía menos, en vez de una lista

de primos, para ilustrar los aspectos del método supongamos que sólo queremos aproximar $\pi(N)$, contar primos[1] en lugar de exhibirlos. Para ello representamos el proceso de Eratóstenes como restar el cardinal de la unión de algunos conjuntos y se vuelve especialmente útil el siguiente principio.

**Principio de inclusión-exclusión:** Sean $C_1, C_2, \ldots, C_M$ conjuntos finitos, entonces

$$\#(C_1 \cup C_2 \cup \cdots \cup C_M) = \sum_{l=1}^{M} (-1)^{l+1} \sum_{1 \le j_1 < j_2 < \ldots < j_l \le M} \#(C_{j_1} \cap C_{j_2} \cap \cdots \cap C_{j_l}).$$

Con la sencilla fórmula $\#(A \cup B) = \#A + \#B - \#(A \cap B)$ la prueba se reduce a inducción o sentido común.

Si se tomase $C_j = \{n \le N \ : \ p_j | n\}$ con $p_j$ el primo $j$-ésimo entonces $\#C_j = [N/p_j]$ y como todo número natural excepto el 1 es divisible por algún primo,

$$N - 1 = \left[\frac{N}{2}\right] + \left[\frac{N}{3}\right] + \cdots - \left[\frac{N}{2\cdot 3}\right] - \left[\frac{N}{2\cdot 5}\right] - \cdots + \left[\frac{N}{2\cdot 3\cdot 5}\right] + \left[\frac{N}{2\cdot 3\cdot 7}\right] + \ldots$$

que aislando el 1 se puede escribir como

$$1 = \sum_{n \le N} \mu(n) \left[\frac{N}{n}\right]$$

ya conocido por inversión de Möbius[2]. La razón de este valor tan escuálido es que en el proceso de Eratóstenes se ha cribado demasiado, hasta los propios primos se han tachado de la lista. Si se hubiera tratado de utilizar $C_j = \{n \le N \ : \ p_j | n, \ p_j > n\}$ las fórmulas habrían sido más complicadas y en cualquier caso dependerían de los primos, ¿tiene sentido dar una fórmula para $\pi(N)$ que depende a su vez de los primos?

El espíritu que anima y constriñe los métodos de criba llamados combinatorios es que hay que tratar de tachar lo menos posible para evitar una gran suma de errores y la dependencia de propiedades de los primos más fuertes que las buscadas. Por ejemplo, si sólo se criba con los primos 2 y 3

$$\pi(N) \le N - ([N/2] + [N/3] - [N/6]) + O(1) = N/3 + O(1).$$

Esto es, a lo más el 33'33 % de los números son primos, si se criba también con el 5, esta proporción pasa a ser del 26'66 %. Cuando se añaden más o menos primos para cribar el coeficiente de $N$ disminuye pero el error $O(1)$ aumenta. Con suerte, estudiando esta dependencia en el número de primos "cribadores" y relacionándolo con $N$, quizá aparezca un $N/\log N$, y como todo número no primo en $[1, N]$ tiene algún factor menor que $\sqrt{N}$, quizá la cota superior devenga en una igualdad con término de error, un flamante teorema

[1] Naturalmente, tras lo visto en el capítulo anterior y la relación con los ceros de la función $\zeta$, los razonamientos posteriores se muestran burdos pero es interesante y sorprendente notar cuán lejos se puede llegar con técnicas puramente combinatorias.

[2] El segundo miembro es $\sum_{n\le N} \mu(n) \sum_{m \le N/n} 1 = \sum_{k \le N} \sum_{n|k} \mu(n) = \sum_{k\le N} (1 * \mu)(k) = 1$ porque $\zeta(s) D_\mu(s) = 1$.

de los números primos sin emplear artillería pesada. Antes de seguir soñando es justo señalar que por el llamado *fenómeno de paridad* [He] hay impedimentos teóricos para demostrar el teorema de los números primos de esta forma o con cualquier técnica de criba clásica (las de este capítulo) sin emplear información adicional. Esto no es óbice para estudiar el procedimiento y comprobar dónde lleva.

Sea $\mathcal{A}_d = \{n \leq N \ : \ d|n\}$, está claro que $\mathcal{A}_{d_1} \cap \mathcal{A}_{d_2} = \mathcal{A}_{[d_1,d_2]}$ donde $[d_1, d_2]$ es el mínimo común múltiplo de $d_1$ y $d_2$. Para cada $z$ se cuentan los números que no han sido cribados por primos menores o iguales que $z$

$$\pi(N) - \pi(z) + 1 \leq N - \#\bigcup_{p<z} \mathcal{A}_p.$$

Por el principio de inclusión-exclusión, escribiendo $A_d = \#\mathcal{A}_d$

$$\pi(N) \leq \sum_{d|P(z)} \mu(d) A_d + \pi(z) - 1 \qquad \text{con} \qquad P(z) = \prod_{p<z} p. \tag{2.1}$$

Se tiene $A_d = [N/d] = N/d + O(1)$ y trivialmente $\pi(z) \leq z$ (una estimación menos burda no cambiaría el orden del resultado final) por tanto

$$\pi(N) \leq \sum_{d|P(z)} \mu(d)\frac{N}{d} + O\big(z + \sum_{d|P(z)} 1\big).$$

Como $f(d) = 1/d$ es una función multiplicativa es fácil escribir la primera suma como un producto, mientras que la segunda suma es $2^{\pi(z)}$ que, esta vez con una pérdida mayor pero no decisiva, estimamos por $O(2^z)$. Entonces

$$\pi(N) \leq N \sum_{p<z} (1 - p^{-1}) + O(2^z).$$

Utilizando la fórmula de Mertens (1.5) y sumando por partes

$$\log \prod_{p<z} (1 - p^{-1}) = -\sum_{p<z} \frac{1}{p} + O(1) = -\log\log x + O(1).$$

Por tanto

$$\pi(N) \ll \frac{N}{\log z} + 2^z.$$

Usando esta desigualdad está claro que ni siquiera se puede deducir $\pi(N) \ll N/\log N$, porque $z = N^\alpha$ causa que el último sumando sea exponencialmente grande. Eligiendo $z = (\log N - \log\log N)/\log 2$ se deduce

$$\pi(N) \ll \frac{N}{\log\log N}$$

que es lo mejor que se puede obtener con cualquier elección de $z$. Es un resultado bastante débil pero aun así contrasta con la simplicidad de las ideas empleadas.

Se puede dar al argumento visos de generalidad, al menos como pretexto para introducir alguna terminología. Para cualquier neófito que quiera introducirse en los métodos de criba, uno de los obstáculos primeros es una notación endiablada (el lector no experimentado puede contrastar esta opinión abriendo al azar el libro clásico de Halberstam y Richert [Ha-Ri]). Aquí, aun a riesgo de perder generalidad, utilizaremos una versión muy reducida de ella que extiende ligeramente la empleada en el ejemplo anterior.

**Notación:**

- $\mathcal{A}$ es un subconjunto finito de $\mathbb{N}$.
- $P(z) = \prod_{p<z} p$.
- $S(\mathcal{A}, z) = \#\{a \in \mathcal{A} \ : \ (a, P(z)) = 1\}$.
- $\mathcal{A}_d = \{a \in \mathcal{A} \ : \ d|a\}$ y $A_d = \#\mathcal{A}_d$.
- $X$, $g(d)$ y $r_d$ son tales que para $d$ libre de cuadrados $A_d = Xg(d)/d + r_d$ con $g$ multiplicativa, $0 \leq g(n) < n$ y $X$ dependiendo sólo de $\mathcal{A}$.

Respecto a la última definición, la idea es que $r_d$ es un término de error pequeño, así que $X$ es el cardinal de $\mathcal{A}$ o un número cercano a él y $g(d)/d$ es como la probabilidad de que un elemento de $\mathcal{A}$ escogido al azar sea divisible por $d$. Es conveniente notar que $g(d)$ y $r_d$ sólo están definidas para $d$ libre de cuadrados, si aparecieran en una suma sobre $d$ se sobreentiende que sólo se suma sobre dichos valores (típicamente basta con definirlas como cero en el resto de los casos).

**Proposición 2.1.1 (criba de Eratóstenes-Legendre)** *Con la notación anterior*

$$S(\mathcal{A}, z) = X \prod_{p<z} (1 - g(p)/p) + \sum_{d|P(z)} r_d \qquad con \quad P(z) = \prod_{p<z} p.$$

*Demostración:* Basta repetir el razonamiento que llevó a (2.1). □

Observación: Interpretando $X$ como $\#\mathcal{A}$ y $g(d)/d$ como la probabilidad de que un elemento de $\mathcal{A}$ sea divisible por $d$, la fórmula $S(\mathcal{A}, z)/X \approx \prod_{p<z}(1 - g(p)/p)$ que sugiere la proposición se muestra natural y cuantifica que la divisibilidad por diferentes primos son sucesos independientes[3] en cierta manera.

El siguiente lema auxiliar que probaremos a medias da el comportamiento de lo que pretende ser el término principal cuando $g$ es constante.

[3]Estos sucesos no pueden ser independientes "del todo", si fuera así entonces $x \prod_{p \leq x^{1/2}}(1 - 1/p)$ debería comportarse como $\pi(x) - \pi(x^{1/2})$ pero estas expresiones no son asintóticamente iguales, su cociente tiende a una constante distinta de uno.

**Lema 2.1.2** *Dado* $k \in \mathbb{Z} - \{0\}$

$$\prod_{k<p<x} \big(1 - \frac{k}{p}\big) = C_k (\log x)^{-k} + O((\log x)^{-k-1})$$

*donde* $C_k$ *es una constante positiva. De hecho* $C_1 = e^{-\gamma}$ *con* $\gamma$ *es la constante de Euler.*

*Demostración:* Tomando logaritmos y empleando $\log(1-h) = -h + O(h^2)$, basta probar

$$\sum_{p<x} \frac{1}{p} = \log\log x + \alpha_k + O((\log x)^{-1}).$$

Y esto es consecuencia de sumar por partes en la fórmula de Mertens $\sum_{p<x} p^{-1} \log p = \log x + O(1)$.

El cálculo de $C_1$ no es sencillo y se lleva acabo comparando $\sum p^{-s}$ y $\log \zeta(s)$. Para una prueba completa véase [Ha-Wr] §22.8. □

Ejemplo. Vamos a estimar la cantidad de enteros en $[1, N]$ tales que todos sus divisores (diferentes de 1) sean mayores que $\sqrt{\log N}$. Con la notación recién introducida esto es $S(\mathcal{A}, \sqrt{\log N})$ donde $\mathcal{A} = [1, N] \cap \mathbb{N}$. Como antes, $A_d = N/d + O(1)$ y se tiene

$$S(\mathcal{A}, \sqrt{\log N}) = N \prod_{p<\sqrt{\log N}} (1 - p^{-1}) + O\big(2^{\sqrt{\log N}}\big).$$

Por el lema anterior,

$$S(\mathcal{A}, \sqrt{\log N}) \sim \frac{2e^{-\gamma} N}{\log\log N}.$$

Como se ve en este ejemplo, la criba no sólo sirve para dar acotaciones, también se puede utilizar para obtener resultados asintóticamente correctos, aunque tiene sus limitaciones (véase [Bo]).

## 2.2. La criba de Brun

**Mejora de la acotación elemental de $\pi(x)$. La suma de los inversos de los primos gemelos converge.**

Si $\mathcal{A} = [1, N] \cap \mathbb{N}$, al cribar por ejemplo con los primos menores que 8 (que son nada más que 4), se obtiene una fórmula kilométrica:

$$\begin{aligned} S(\mathcal{A}, 8) \;=\; & N - \left[\frac{N}{2}\right] - \left[\frac{N}{3}\right] - \left[\frac{N}{5}\right] - \left[\frac{N}{7}\right] + \left[\frac{N}{2\cdot 3}\right] + \left[\frac{N}{2\cdot 5}\right] + \left[\frac{N}{2\cdot 7}\right] \\ & + \left[\frac{N}{3\cdot 5}\right] + \left[\frac{N}{3\cdot 7}\right] + \left[\frac{N}{5\cdot 7}\right] - \left[\frac{N}{2\cdot 3\cdot 5}\right] - \left[\frac{N}{2\cdot 3\cdot 7}\right] - \left[\frac{N}{2\cdot 5\cdot 7}\right] \\ & - \left[\frac{N}{3\cdot 5\cdot 7}\right] + \left[\frac{N}{2\cdot 3\cdot 5\cdot 7}\right] \end{aligned}$$

Esto explica por qué la criba de Eratóstenes no es muy buena, hay demasiados sumandos lo cual obliga a tomar $z$ excesivamente pequeño. Si sólo se aspira a cotas superiores o inferiores uno podría considerar la parte de la suma anterior correspondiente a los denominadores con menos de dos o tres primos:

$$\begin{aligned} S^- &= N - \left[\frac{N}{2}\right] - \left[\frac{N}{3}\right] - \left[\frac{N}{5}\right] - \left[\frac{N}{7}\right] \\ S^+ &= S^- + \left[\frac{N}{2\cdot 3}\right] + \left[\frac{N}{2\cdot 5}\right] + \left[\frac{N}{2\cdot 7}\right] + \left[\frac{N}{3\cdot 5}\right] + \left[\frac{N}{3\cdot 7}\right] + \left[\frac{N}{5\cdot 7}\right] \end{aligned}$$

que son sumas un poco menos extensas.

Esta claro que $S^- \leq S(\mathcal{A}, 8) \leq S^+$ porque los positivos ganan a los negativos o viceversa. Pues bien, éste es un hecho general: los que tienen menos de $2k+1$ factores primos contribuyen siempre más y los que tienen menos de $2k$ factores primos, menos. Ésta es la idea básica de Brun (aunque su aportación es mucho más compleja y poderosa que esto). Jugando con $k$ se puede disminuir a voluntad el número de sumandos y aun así conservar desigualdades. En general los métodos de criba basados en seleccionar sólo algunos de los sumandos que aparecerían en la criba de Eratóstenes-Legendre se dice que son metodos de *criba combinatoria*. A primera vista no parece que haya muchas más posibilidades que la apuntada inicialmente por Brun, pero muchos años de investigación sobre el tema han probado lo equivocado de esta impresión.

Un resultado simple que formaliza la idea anterior es

$$\sum_{\substack{d|n \\ \nu(d)<2k}} \mu(d) \leq \sum_{d|n} \mu(d) \leq \sum_{\substack{d|n \\ \nu(d)<2k+1}} \mu(d)$$

donde $\nu(d)$ representa el número de factores primos distintos de $d$ y $k, n \in \mathbb{N}$.

En vez de dar la demostración[4] y seguir la línea anterior encontrando $S^+$ y $S^-$, daremos una expresión exacta para $S(\mathcal{A}, z)$ que muestra claramente la dependencia en la paridad de los factores. Con este fin es conveniente enunciar una identidad elemental llamada *identidad de Buchstab* que desempeña un papel destacado en los métodos de criba. Se acompaña de otra identidad hermana similar.

**Lema 2.2.1 (identidad de Buchstab)** *Con la notación anterior*

$$S(\mathcal{A}, z) = \#\mathcal{A} - \sum_{p<z} S(\mathcal{A}_p, p).$$

*Además (para $q$ recorriendo los primos)*

$$\prod_{p<z}(1 - g(p)/p) = 1 - \sum_{p<z} \frac{g(p)}{p} \prod_{q<p}(1 - g(q)/q).$$

[4]Es bastante simple [Ci-Co] y se basa en la sencilla fórmula trivial $\sum \mu(d) = (-1)^r \binom{\nu(n)}{r}$ donde la suma recorre los divisores $d$ de $n$ con $\nu(d) = r$.

*Demostración:* La primera identidad dice que (!?) los elementos sin factores menores que $z$ son todos excepto los que tienen como menor factor a un primo menor que $z$. La segunda identidad es la misma (??) salvo que se indican probabilidades en vez de cardinales. □

La expresión antes anunciada es:

**Proposición 2.2.2** *Sea $r \in \mathbb{N}$, entonces*

$$S(\mathcal{A}, z) = \sum_{\substack{d|P(z) \\ \nu(d)<r}} \mu(d) A_d + (-1)^r \sum_{\substack{d|P(z) \\ \nu(d)=r}} S(\mathcal{A}_d, p_d)$$

*donde $p_d$ indica el menor factor primo de $d$. Además*

$$\prod_{p<z}(1 - g(p)/p) = \sum_{\substack{d|P(z) \\ \nu(d)<r}} \mu(d)\frac{g(d)}{d} + (-1)^r \sum_{\substack{d|P(z) \\ \nu(d)=r}} \frac{g(d)}{d} \prod_{q<p_d}(1 - g(q)/q)$$

*con $p_d$ como antes y $q$ recorriendo los primos.*

Observación: Nótese que inmediatamente se deducen de aquí las desigualdades:

$$\sum_{\substack{d|n \\ \nu(d)<2k}} \mu(d) A_d \leq S(\mathcal{A}, z) \leq \sum_{\substack{d|n \\ \nu(d)<2k+1}} \mu(d) A_d.$$

*Demostración:* En el primer caso basta aplicar inductivamente la identidad de Buchstab. Si se la emplea dos veces

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1<z} A_{p_1} + \sum_{p_2<p_1<z} S(A_{p_1p_2}, p_2),$$

y otra vez más

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1<z} A_{p_1} + \sum_{p_2<p_1<z} S(A_{p_1p_2}, p_2) - \sum_{p_3<p_2<p_1<z} S(A_{p_1p_2p_3}, p_3)$$

y así sucesivamente. La prueba de la otra identidad es similar. □

Recapitulemos la ventaja de las desigualdades de la observación: la fórmula exacta para $S(\mathcal{A}, z)$ (empleando la criba de Eratóstenes-Legendre) da lugar a una suma muy larga, exponencial en el número de factores primos en $P(z)$, donde se acumulan muchos términos de error. A través del parámetro $k$ se puede controlar la longitud de sumas mayorantes y minorantes de $S(\mathcal{A}, z)$ y muy similares a ella (lo único que se hace es seleccionar algunos sumandos de la suma inicial). Por un lado nos gustaría tomar $k$ pequeño para que hubiera pocos sumandos y por otra parte necesitaríamos que $k$ sea moderadamente grande para que las cotas no sean burdas.

A partir de aquí los ajustes dependen de cada problema y se pueden relacionar con ciertas variables y estructuras asignadas genéricamente al escenario de la criba que requieren pagar el precio de una notación críptica que aquí se intenta evitar. Antes de mostrar un resultado general, jugaremos con el método empleando el ejemplo de la sección anterior.

Sea $\mathcal{A} = [1, N] \cap \mathbb{N}$, ya sabemos que $A_d = N/d + O(1)$ y aplicando la proposición con $r = 2k$ y $r = 2k + 1$

$$N \sum_{\substack{d|P(z)\\ \nu(d)<2k}} \frac{\mu(d)}{d} - z^{2k} \leq S(\mathcal{A}, z) \leq N \sum_{\substack{d|P(z)\\ \nu(d)<2k+1}} \frac{\mu(d)}{d} + z^{2k}$$

donde se ha usado la cota $\#\{d|P(z) \ : \ 0 < \nu(d) \leq r\} \leq \sum_{t=1}^{r} \binom{z}{t} \leq z^r$ (si esto es dudoso, $\pm z^{2k}$ se puede reemplazar por $O(z^{2k})$ sin problemas).

Para que esta cota sea operativa necesitamos ver cuánto es el sumatorio. Sin la condición sobre $\nu(d)$ es trivialmente $\prod_{p<z}(1 - 1/p)$ pero como esta condición es esencial al método no es posible obviarla y se hace necesaria la ingrata tarea de compensar los términos que faltan. Ésta es la razón de ser de la segunda identidad de la Proposición 2.2.2 que para $g(d) = 1$ produce

$$\sum_{\substack{d|P(z)\\ \nu(d)<r}} \frac{\mu(d)}{d} = \prod_{p<z} \left(1 - \frac{1}{p}\right) - (-1)^r \sum_{\substack{d|P(z)\\ \nu(d)=r}} \frac{1}{d} \prod_{q<p_d} \left(1 - \frac{1}{q}\right). \tag{2.2}$$

Teniendo en cuenta que el último producto es menor que 1, la suma está mayorada por $\left(\sum_{p<z} 1/p\right)^r / r!$ que usando la fórmula de Mertens es $(\log\log z + O(1))^r / r!$. Escogiendo $r = [A \log\log N]$ se tiene (ejercicio) que la contribución de la suma final en (2.2) es $O((\log N)^{-B})$ donde $B > 0$ es arbitrariamente grande si $A$ lo es. En definitiva ese término no molesta.

Por otro lado, el primer producto es del mismo orden que $1/\log z$ y (2.2) implica

$$\sum_{\substack{d|P(z)\\ \nu(d)<r}} \frac{\mu(d)}{d} \asymp \frac{1}{\log z}.$$

Sustituyendo

$$S(\mathcal{A}, z) \asymp \frac{N}{\log z} \qquad \text{si} \qquad z^{2k} = o(N/\log z).$$

Escogiendo $z = \epsilon(N/\log N)^{1/2k}$ y recordando que $k \asymp \log\log N$, se obtiene $S(\mathcal{A}, z) \asymp N(\log\log N)/\log N$ y de aquí

$$\pi(N) \ll \frac{N}{\log N} \log\log N$$

que da el orden correcto de la función $\pi(N)$ salvo un factor $\log\log N$.

La mejora con respecto a la criba de Eratóstenes-Legendre se puede hacer general apurando bien las desigualdades.

**Teorema 2.2.3 (criba de Brun)** *Para cualquier $r \in \mathbb{N}$ se cumple*

$$\left|S(\mathcal{A},z) - X\prod_{p<z}(1-g(p)/p)\right| \leq X\sum_{\substack{d|P(z)\\ \nu(d)=r}}\frac{g(d)}{d} + \sum_{\substack{d|P(z)\\ \nu(d)\leq r}}|r_d|$$

*Demostración:* Por la Proposición 2.2.2, $S(\mathcal{A},z) - X\prod_{p<z}(1-g(p)/p)$ es

$$\sum_{\substack{d|P(z)\\ \nu(d)<r}}\mu(d)A_d - X\prod_{p<z}(1-g(p)/p) + (-1)^r\sum_{\substack{d|P(z)\\ \nu(d)=r}}S(\mathcal{A}_d,q_d)$$

$$= \sum_{\substack{d|P(z)\\ \nu(d)<r}}\mu(d)r_d + (-1)^r\sum_{\substack{d|P(z)\\ \nu(d)=r}}\left(S(\mathcal{A}_d,p_d) - X\frac{g(d)}{d}\prod_{p<q_d}(1-g(p)/p)\right)$$

Se cumple que $0 \leq S(\mathcal{A}_d,q_d) \leq A_d = Xg(d)/d + r_d$ y evidentemente el último producto está entre cero y uno. □

¿No parece todo esto combinatoria barata en la que uno quita y pone cosas irrelevantes? Las siguientes consecuencias hablan por sí solas.

**Corolario 2.2.4** *Sea $\pi_2(x) = \#\{p \leq x \ : \ p \text{ y } p+2 \text{ son primos}\}$. Se cumple*

$$\pi_2(x) \ll x\left(\frac{\log\log x}{\log x}\right)^2.$$

*Demostración:* Sea $\mathcal{A} = \{n(n+2) \ : \ n \leq x\}$. Si $q$ es un primo impar $A_q = 2X/q + O(1)$ y en general para $d$ libre de cuadrados impar $A_d = 2^{\nu(d)}X/d + O\big(2^{\nu(d)}\big)$ (ejercicio), que se completa fácilmente con $A_2 = X/2+O(1)$ y $A_{2d} = 2^{\nu(d)}X/2d+O\big(2^{\nu(d)}\big)$. Así pues se toma $X = x$ y $g(d) = 2^{\nu(d)+1}/(3+(-1)^d)$ para $d$ libre de cuadrados con $r_d = O(2^{\nu(d)})$. Para aplicar el teorema se necesitan las estimaciones:

$$\sum_{\substack{d|P(z)\\ \nu(d)=r}}\frac{2^{\nu(d)}}{d} = \sum_{\substack{d|P(z)\\ \nu(d)=r}}\frac{2^r}{d} \leq \frac{1}{r!}\left(\sum_{p<z}\frac{2}{p}\right)^r = \frac{(2\log\log z + O(1))^r}{r!}$$

y

$$\sum_{\substack{d|P(z)\\ \nu(d)\leq r}}2^{\nu(d)} \leq z^r\sum_{d|P(z)}\frac{2^{\nu(d)}}{d} = z^r\prod_{p<z}\big(1+\frac{2}{p}\big).$$

Como ya sabemos, eligiendo $r = [A\log\log x]$ la contribución de la suma con $\nu(d) = r$ está bajo control. Teniendo en cuenta estas acotaciones y $\prod_{p<z}(1+2/p) \asymp (\log z)^2$, el teorema asegura

$$\left|S(\mathcal{A},z) - x\prod_{p<z}(1-2/p)\right| \ll z^r(\log z)^2.$$

Así pues

$$\pi_2(x) \leq S(\mathcal{A}, z) + O(z) \ll \frac{x}{(\log z)^2} + z^r (\log z)^2$$

y tomando $z = x^{1/r}(\log x)^{-4/r}$ se termina la prueba. □

Más famosa y atractiva es la consecuencia de la consecuencia.

**Corolario 2.2.5** *La suma de los inversos de los primos gemelos converge.*

Observación: Como todas las constantes en estos resultados son efectivas, es posible acotar el error en las sumas parciales y con ello dejar trabajando al ordenador día y noche para obtener el valor del límite de la serie con unas cuantas cifras decimales (véase [Sh-Wr], actualmente hay resultados mucho más precisos). El número obtenido, usando primos hasta $2'55 \cdot 10^{15}$, es $1'90216058\ldots$. Se dice que ésta es la *constante de Brun*.

*Demostración:* Basta sumar por partes. Sea $a_n = 1$ si $n$ y $n+2$ son primos y $a_n = 0$ en otro caso,

$$\sum_{\substack{p,p+2 \text{ primos} \\ p \leq x}} \left( \frac{1}{p} + \frac{1}{p+2} \right) < 2 \sum_{n \leq x} \frac{a_n}{n} = 2\frac{\pi_2(x)}{x} + 2\int_1^x \frac{\pi_2(t)}{t^2}\, dt$$

y empleando el resultado anterior, la integral converge cuando $x \to \infty$. □

## 2.3. La criba de Selberg

**La criba cuadrática. Mejora del resultado de Brun. Otras aplicaciones.**

Ya habíamos visto que la base de la criba de Brun era seleccionar algunos términos en $S(\mathcal{A}, z) = \sum_{d|n} \mu(d) A_d$ pasando la igualdad a desigualdades. En esta sección veremos una criba no combinatoria, es decir, no se omiten términos, lo que se hará es reemplazar el coeficiente $\mu(d)$ por otra función. En principio esto no parece tener sentido porque las propiedades de $\mu$ son esenciales en el proceso de criba. Ver para creer, por ejemplo se cumple

$$S(\mathcal{A}, 4) \leq A_1 + \frac{1 - 2\sqrt{2}}{2} A_2 + \frac{1 - 2\sqrt{3}}{3} A_3 + \frac{1}{\sqrt{6}} A_6,$$

y en general, para cualquier $\alpha, \beta \in \mathbb{R}$

$$S(\mathcal{A}, 4) \leq A_1 + (\alpha^2 + 2\alpha) A_2 + (\beta^2 + 2\beta) A_3 + \alpha\beta A_6.$$

La prueba es elemental, simplemente desarrollar y agrupar en

$$S(\mathcal{A}, 4) = \sum_{\substack{a \in \mathcal{A} \\ (a,6)=1}} 1 \leq \sum_{\substack{a \in \mathcal{A} \\ (a,6)=1}} 1 + \sum_{\substack{a \in \mathcal{A} \\ (a,6)=2}} (1+\alpha)^2 + \sum_{\substack{a \in \mathcal{A} \\ (a,6)=3}} (1+\beta)^2 + \sum_{\substack{a \in \mathcal{A} \\ (a,6)=6}} (1+\alpha+\beta)^2$$

En 1947 A. Selberg desarrolló esta idea creando un método de criba consistente en buscar la "desigualdad óptima"que responde a demostraciones como la anterior con formas cuadráticas en los parámetros (véase [Se]). La optimización se adapta al problema, lo que permite en general superar los resultados de Brun. En el ejemplo anterior si se cumpliera $A_d \approx N/d$, $d = 1, 2, 3, 6$, entonces al minimizar en $\alpha$ y $\beta$ se obtiene más o menos $A_1 - 0'98A_2 - 0'95A_3 + 0'68A_6$. Sin embargo si todos los números de $A$ son pares con $A_1 \approx A_2 \approx N$, $A_3 \approx A_6 \approx N/3$, la desigualdad óptima sería $A_1 - 0'99A_2 - 0'79A_3 + 0'49A_6$.

El poder grandioso de la criba de Selberg contrasta enormemente con la simplicidad de su punto de partida (honor que comparten otras ideas geniales). Más que en el propio enunciado del siguiente resultado, el lector debería detenerse en la prueba.

**Lema 2.3.1** *Sea $\{\lambda_n\}_{n=1}^{\infty}$ una sucesión de números reales con $\lambda_1 = 1$ y $\lambda_n = 0$ si $n$ no es libre de cuadrados o $n \geq z$, entonces*

$$S(\mathcal{A}, z) \leq \sum_{d_1, d_2} \lambda_{d_1} \lambda_{d_2} A_{[d_1, d_2]}$$

*donde $[d_1, d_2]$ representa el mínimo común múltiplo.*

*Demostración:* Si $a \in \mathcal{A}$ no tiene divisores primos menores que $z$ se tiene que $\big(\sum_{d|a} \lambda_d\big)^2$ es uno, en cualquier caso esta cantidad es positiva y por consiguiente

$$S(\mathcal{A}, z) \leq \sum_{a \in \mathcal{A}} \big(\sum_{d|a} \lambda_d\big)^2 = \sum_{a \in \mathcal{A}} \sum_{d_1|a} \sum_{d_2|a} \lambda_{d_1} \lambda_{d_2}$$

y basta intercambiar el orden de sumación. □

Observación: La criba de Selberg en toda generalidad permite jugar con un parámetro llamado *nivel* que, al igual que $r$ en la criba de Brun, sirve para controlar el tamaño de las sumas. Para simplificar, aquí no se considerará tal diversión.

Con la confianza, o la hipótesis de que $A_d$ está bien aproximado por $Xg(d)/d$ (para $d$ libre de cuadrados) lo que se necesita es minimizar la siguiente forma cuadrática en los $\lambda_n$

$$Q = \sum_{d_1, d_2} \frac{g([d_1, d_2])}{[d_1, d_2]} \lambda_{d_1} \lambda_{d_2}$$

y esto parece muy difícil porque sus coeficientes son aritméticos. Una luz de esperanza es que si $d_1$ y $d_2$ fueran coprimos siempre, $g([d_1, d_2])/[d_1, d_2] = g(d_1)/d_1 \cdot g(d_2)/d_2$ y $Q$ sería simplemente $(\sum g(d)\lambda_d/d)^2$. Para salvar este obstáculo se puede tratar de separar el máximo común divisor porque $[d_1, d_2]$ coincide con $d_1 d_2$ salvo dividir por $(d_1, d_2)$. Antes de acabar de violentar las normas de los libros de estilo, anunciemos dónde queremos llegar.

**Proposición 2.3.2** *Se tiene*

$$\min Q = \left(\sum_{d<z} h(d)\right)^{-1} \qquad con \quad h(n) = \mu^2(n)\prod_{p|n}\frac{g(p)}{p-g(p)}$$

*donde el mínimo se toma sobre todas las sucesiones $\lambda_n$ como en el lema anterior. Además los valores para los que se alcanza el mínimo verifican $|\lambda_n| \le 1$.*

*Demostración:* Según lo dicho anteriormente,

$$Q = \sum_{d_1,d_2}\frac{g([d_1,d_2])}{[d_1,d_2]}\gamma((d_1,d_2))\lambda_{d_1}\lambda_{d_2} \qquad \text{donde} \quad \gamma(d) = \begin{cases} d/g(d) & \text{si } g(d)\neq 0 \\ 0 & \text{en otro caso}\end{cases}$$

Por inversión de Möbius, $\gamma(n) = \sum_{d|n} f(d)$ con $f = \mu * \gamma$ y se puede escribir

$$Q = \sum_{d_1,d_2}\frac{g([d_1,d_2])}{[d_1,d_2]}\lambda_{d_1}\lambda_{d_2}\sum_{m|(d_1,d_2)} f(m) = \sum_m f(m)\Big(\sum_{d\equiv 0\ (m)}\frac{g(d)}{d}\lambda_d\Big)^2.$$

Aquí y en el resto de la prueba se supondrá que $m$ es libre cuadrados y $m < z$.

Si se denota con $x_m$ la suma bajo el cuadrado, la forma cuadrática ya está diagonalizada. La restricción $\lambda_1 = 1$ equivale a $\sum \mu(m)x_m = 1$ porque

$$\sum_m \mu(m)\sum_{d\equiv 0\ (m)}\frac{g(d)}{d}\lambda_d = \sum_d \big(\sum_{m|d}\mu(m)\big)\frac{g(d)}{d}\lambda_d = \lambda_1.$$

En definitiva, el mínimo buscado es el de $\sum f(m)x_m^2$ sujeto a $\sum \mu(m)x_m = 1$. Éste es un problema de Cálculo II con letras raras pero sencillo. El mínimo se alcanza para

$$x_m = \frac{\mu(m)}{f(m)H} \qquad \text{con} \quad H = \sum\frac{1}{f(d)}$$

donde $d$ recorre los libres de cuadrados menores que $z$. Para asegurar que esto tiene sentido, nótese que $f(p) = (\mu * \gamma)(p) = p/g(p) - 1 \neq 0$ y por tanto $H = \sum_{d\le z} h(d)$. Evaluando $\sum f(m)x_m^2$ se tiene que el mínimo buscado es $H^{-1}$.

La comprobación de $|\lambda_n| \le 1$ es indirecta y más compleja de lo que cabría esperar. Fijado $n$ libre de cuadrados

$$H = \sum_{k|n} h(k)\sum_{\substack{d<z/k\\(d,n)=1}} h(d) \ge \big(\sum_{k|n} h(k)\big)\sum_{\substack{d<z/n\\(d,n)=1}} h(d). \tag{2.3}$$

No es difícil "despejar" los $\lambda_n$ a partir de $x_m$ (ejercicio), esto prueba de paso que el cambio de variable era lícito:

$$\lambda_n = \mu(n)\frac{n}{g(n)}\sum_{m\equiv 0\ (n)}\mu(m)x_m.$$

Utilizando la fórmula para los $x_m$ minimizantes se tiene

$$\sum_{\substack{d<z/n\\(d,n)=1}} h(d) = \frac{1}{h(n)} \sum_{m\equiv 0 \ (n)} h(m) = \frac{g(n)}{nh(n)} H\lambda_n = H\lambda_n \prod_{p|n} \frac{g(p)}{ph(p)},$$

mientras que por las propiedades multiplicativas

$$\sum_{k|n} h(k) = \prod_{p|n} (1+h(p)) = \prod_{p|n} \frac{ph(p)}{g(p)}.$$

Sustituyendo en (2.3) se llega a $|\lambda_n| \leq 1$. □

Con esto ya se tiene el cerebro de la criba de Selberg, sólo resta moldear unas formas bellas.

**Teorema 2.3.3 (criba de Selberg)** *Para cada $z > 1$*

$$S(\mathcal{A}, z) \leq X\big(\sum_{d<z} h(d)\big)^{-1} + \sum_{d<z^2} 3^{\nu(d)} |r_d|$$

*donde $h(n) = \mu^2(n) \prod_{p|n} g(p)/(p-g(p))$.*

*Demostración:* Por el Lema 2.3.1

$$S(\mathcal{A}, z) \leq \sum_{d_1,d_2} \frac{g[d_1,d_2]}{[d_1,d_2]} \lambda_{d_1}\lambda_{d_2} + \sum_{d_1,d_2} |r_{[d_1,d_2]}| \lambda_{d_1}\lambda_{d_2}.$$

con la elección óptima de la Proposición 2.3.2

$$S(\mathcal{A}, z) \leq X\big(\sum_{d<z} h(d)\big)^{-1} + \sum_{d<z^2} |r_d| \sum_{[d_1,d_2]=d} 1$$

y el último sumatorio es exactamente $3^{\nu(d)}$. □

Ejemplo. con $\mathcal{A} = [1, N] \cap \mathbb{N}$ se tiene $g(d) = 1$ y $|r_d| \leq 1$ por lo cual $h(p) = p^{-1} + p^{-2} + p^{-3} + \ldots$ y se cumple

$$S(\mathcal{A}, z) \leq N\big(\sum_{d<z} \frac{1}{d}\big)^{-1} + \sum_{d<z^2} \mu^2(d) 3^{\nu(d)}.$$

Con un "truquito" se puede estimar la última suma

$$\sum_{d<z^2} \mu^2(d) 3^{\nu(d)} < z^2 \sum_{d<z^2} \frac{\mu^2(d)}{d} 3^{\nu(d)} = z^2 \sum_{d_1d_2d_3<z^2} \frac{\mu^2(d_1d_2d_3)}{d_1d_2d_3} \leq z^2\big(\sum_{d<z^2} \frac{1}{d}\big)^3.$$

Por consiguiente, con las cotas para la serie armónica, $\log N < \sum_{n<N} 1/n < 1 + \log N$ se concluye

$$S(\mathcal{A}, z) < N/\log z + z^2(1 + 2\log z)^3. \tag{2.4}$$

Eligiendo $z = N^{1/2}/(\log N)^2$ se tiene para $N > 100$

$$S(\mathcal{A}, z) < 44\frac{N}{\log N}.$$

Con esto ¡por fin conseguimos una cota superiorpara $\pi(x)$ del orden correcto de magnitud! Hay otra cosa notoria y es que $z$ es "casi" como $N^{1/2}$ y por tanto $S(\mathcal{A}, z)$ es "casi" como $\pi(N) - \pi(z)$.

Con hipótesis adicionales es posible transformar el teorema anterior en un artículo de consumo en el que se sustituyen unos parámetros por un lado y se recoge el producto por otro.

**Teorema 2.3.4** *Con la notación del teorema anterior, si*

$$\sum_{p<z} h(p)\log p = \kappa \log z + O(1) \qquad y \qquad r_d = O(M^{\nu(d)})$$

*para alguna constante $M \in \mathbb{N}$, entonces*

$$S(\mathcal{A}, z) \leq C\frac{X}{(\log z)^\kappa} + O\big(\frac{X}{(\log z)^{\kappa+1}} + z^2(\log z)^{3M}\big)$$

*donde $C = \Gamma(\kappa+1)\prod(1 - g(p)/p)/(1-1/p)^\kappa$.*

*Demostración:* Por el teorema de Wirsing se tiene

$$\sum_{n<z} h(n) = \frac{(\log z)^\kappa}{\Gamma(\kappa+1)}\prod(1+h(p))(1-1/p)^\kappa + O((\log z)^{\kappa-1})$$

y por otra parte con el truco ya empleado para (2.4),

$$\sum_{d<z^2}\mu^2(d)(3M)^{\nu(d)} < z^2\sum_{d<z^2}\mu^2(d)\frac{(3M)^{\nu(d)}}{d} \leq z^2\bigg(\sum_{d<z^2}\frac{\mu^2(d)}{d}\bigg)^{3M} = O(z^2(\log z)^{3M}).$$

Basta sustituir estas dos estimaciones en el teorema anterior. □

Vayamos ahora con las aplicaciones.

**Corolario 2.3.5** *Sea $\pi_2(x) = \#\{p \leq x \ : \ p \text{ y } p+2 \text{ son primos}\}$, entonces*

$$\pi_2(x) \leq \frac{16x}{(\log x)^2}\prod_{p>2}\frac{p(p-2)}{(p-1)^2}\bigg(1 + O\big(\frac{\log\log x}{\log x}\big)\bigg)$$

Observación: Si se cumplen conjeturas profundas, debidas a Hardy y Littlewood, la desigualdad debería poder reemplazarse por una igualdad cambiando 16 por 4. En relación con estas conjeturas, la criba de Selberg también produce cotas superiores del orden de magnitud esperado para primos "trillizos" y otras familias numerosas.

*Demostración:* Tómese $\mathcal{A} = \{n(n+2) \,:\, n \le x\}$, entonces $A_d = xg(d)/d + r_d$ (para $d$ libre de cuadrados) con $g(d) = 2^{\nu(d)}$ si $d$ es impar y $g(d) = 2^{\nu(d)-1}$ si es par. En cualquier caso $g(d) \ge |r_d|$. De $h(p) = 2/(p-2)$ para $p > 2$ se deduce $\sum_{p<z} h(p)\log p = 2\log z + O(1)$ (fórmula de Mertens) y el teorema se aplica produciendo

$$S(\mathcal{A}, z) \le 4\prod_{p>2}(1-2/p)(1-1/p)^{-2}\frac{X}{(\log z)^2} + O\Big(\frac{X}{(\log z)^3} + z^2(\log z)^6\Big).$$

Escogiendo $z = x^{1/2}(\log x)^{-6}$ y notando que $S(\mathcal{A}, z) \ge \pi_2(x) + O(z)$ se termina la prueba. □

**Corolario 2.3.6** *Para $x, y > 1$*

$$\pi(x+y) - \pi(x) \le \frac{2y}{\log y} + O\left(\frac{y\log\log y}{(\log y)^2}\right).$$

*Demostración:* Evidentemente se puede aplicar el teorema con $\mathcal{A} = [x, x+y] \cap \mathbb{N}$ pero ya hemos hecho el trabajo porque todo funciona exactamente igual que en la deducción de (2.4) (allí $x = 1$, $y = N - 1$), por tanto

$$S(\mathcal{A}, z) < \frac{y}{\log z} + O(z^2(\log z)^3)$$

y basta tomar $z = y^{1/2}(\log y)^{-3}$ □

Nada impide considerar progresiones aritméticas en vez de intervalos trasladados o ¿por qué no? ambas cosas a la vez.

**Corolario 2.3.7 (desigualdad de Brun-Titchmarsh)** *Sea $\pi(x; q, a) = \#\{p \le z \,:\, p \equiv a \ (q)\}$ con $a, q \in \mathbb{N}$, $(a, q) = 1$. Para $x, y > 1$ se cumple*

$$\pi(x+y; q, a) - \pi(x; q, a) < \frac{2y}{\phi(q)\log(y/q)} + O\left(\frac{y}{\phi(q)(\log(y/q))^2}\right)$$

*Demostración:* Tómese $\mathcal{A} = \{n \in [x, x+y] \,:\, n \equiv a \ (q)\}$, $X = y/q$, $g(d) = 1$ si $(d, q) = 1$ y $g(d) = 0$ en otro caso. Entonces (para $d$ libre de cuadrados) $A_d = Xg(d)/d + r_d$ con $r_d = O(1)$ y todo lo que hay que hacer es meter estos ingredientes en la máquina del teorema. □

Después de todos estos ejemplos que alaban las glorias de la criba de Selberg es justo señalar el inconveniente de que no proporciona directamente cotas inferiores para $S(\mathcal{A}, z)$. Dando algún rodeo es posible paliar en parte esta deficiencia, por ejemplo la identidad de Buchstab permite pasar cotas superiores a inferiores. En otro contexto, esta idea será explotada en la próxima sección.

## 2.4. Nociones y aplicaciones de la criba lineal

**Una criba combinatoria. El teorema de Jurkat-Richert. Algunos ejemplos**

La criba de Selberg es la mejor criba entre aquellas cuyos coeficientes vienen determinados por ciertas formas cuadráticas. En esta sección volvemos a una criba de tipo combinatorio como la de Brun, es decir, se eligen sumandos $\sum_{d|P(z)} \mu(d)A_d$. En el caso que aquí se analiza resulta ser óptima (aunque no entraremos en ello) dejando un sabor agridulce porque por una parte permite mejorar algunas estimaciones y al tiempo muestra que sin hipótesis más restrictivas o particularidades específicas de algunos problemas no es posible ir más allá.

El teorema central de esta sección, el *teorema de Jurkat-Richert*, tiene una prueba que no es en absoluto sencilla y es difícil entrever las ideas principales siguiéndola paso a paso. Aferrándonos a las directrices del curso y favorecidos por la falta de diligencia, se omitirá dicha prueba aquí, reemplazándola por algunas ideas a veces etéreas, a veces tangibles.

Antes de nada volvamos a primeros principios para introducir algo más de notación. Por el principio de inclusión-exclusión (criba de Eratóstenes-Legendre) sabíamos que

$$S(\mathcal{A}, z) = \sum_{d|P(z)} \mu(d)A_d.$$

Una criba combinatoria consiste en hallar dos conjuntos $\mathcal{D}^-, \mathcal{D}^+ \subset \mathbb{N}$ tales que

$$\sum_{\substack{d|P(z)\\ d\in\mathcal{D}^-}} \mu(d)|A_d| \leq S(A, z) \leq \sum_{\substack{d|P(z)\\ d\in\mathcal{D}^+}} \mu(d)|A_d|.$$

El siguiente paso es sustituir $A_d = Xg(d)/d + r_d$ y crear un término principal a partir de la suma de $Xg(d)/d$ y un término de error a partir de la de $r_d$. Si $\mathcal{D}^-$ y $\mathcal{D}^+$ son "sustanciosos" este término principal debe ser comparable a $X\prod_{p<z}(1 - g(p)/p)$.

Digamos que $\mathcal{D}^-, \mathcal{D}^+ \subset [1, D]$. No tiene sentido $D < z$ porque esto significaría que se puede reducir $z$ teniendo el mismo resultado. Si $D = z^s$ con $s > 1$ estaremos vedando los valores que tengan más de $s$ factores primos próximos a $z$. De esta forma el parámetro $s$ se asemeja a $r$ en la criba de Brun, dando cierto control sobre el número de factores y si $s$ es suficientemente grande, se llegará a la criba de Eratóstenes-Legendre y el inconveniente es ya conocido: hay problemas para controlar el término de error. El tamaño admisible depende del término principal de que dispongamos, con las hipótesis que aparecerán más adelante este término será al menos comparable a $X/\log z$ y con un error $o(X/\log X)$ estamos a salvo porque en los ejemplos ya vistos no tiene sentido $z > X$.

Resumiendo, dados $z$ y $s > 1$ se define $D = z^s$ y se supone

$$\sum_{\substack{d|P(z)\\ d<D}} |r_d| \ll \frac{X}{(\log X)^C} \qquad \text{para algún} \quad C > 1. \tag{2.5}$$

Dependiendo del valor de $s$ el valor del término principal se puede ver amplificado o disminuido. Se esperan por tanto desigualdades del tipo:

$$f(s)\,\Pr(\mathcal{A},z)+O(E)\le S(\mathcal{A},z)\le F(s)\,\Pr(\mathcal{A},z)+O(E) \tag{2.6}$$

donde

$$\Pr(\mathcal{A},z)=X\prod_{p<z}(1-g(p)/p)\quad\text{y}\qquad E=\frac{X}{(\log X)^C}.$$

Dar a $f(s)$ y $F(s)$ nombre de función en vez de nombre de constante tiene su lógica por los razonamientos posteriores. De acuerdo con lo dicho anteriormente, si $\mathcal{D}^-$ y $\mathcal{D}^+$ se eligen "de la mejor manera", cuando $s\to\infty$ se tiene que cumplir $f(s),F(s)\to 1$ porque la criba de Eratóstenes da una igualdad con $\Pr(\mathcal{A},z)$ salvo el término de error. Por otro lado, si $s$ se empequeñece, las cotas se volverán burdas y parece creíble que para $s\in[1,\beta]$ sólo se pueda obtener la cota inferior trivial[5] $f(s)=0$.

Supongamos que tenemos alguna de las desigualdades de (2.6), digamos por ejemplo la inferior. En principio esta cota podría ser trivial o casi trivial (porque $s$ esté cerca de $\beta$). Lo que se mostrará es un proceso que transforma cotas inferiores en superiores mejorándolas cada vez. El objetivo es buscar el límite. Este proceso no es más que la identidad de Buchstab que se debería reflejar en los términos principales como

$$S(\mathcal{A},z)\lesssim X-\sum_{p<z}f(s_p)X_p\prod_{q<p}(1-g(q)/q)$$

donde ahora $X_p$ es la aproximación del cardinal de $\mathcal{A}_p$, así pues $X_p\approx Xg(p)/p$, y $s_p$ debería cumplir $D/p=p^{s_p}$ ya que éste es el análogo natural de $D=z^s$ al cambiar $S(\mathcal{A},z)$ por $S(\mathcal{A}_p,p)$. Teniendo todo esto en cuenta y restando a la fórmula anterior la siguiente forma de la segunda identidad del Lema 2.2.1

$$\Pr(\mathcal{A},z)=X-\sum_{p<z}\Pr(\mathcal{A}_p,p) \tag{2.7}$$

se obtiene

$$S(\mathcal{A},z)\lesssim\Pr(\mathcal{A},z)+\sum_{p<z}\left(1-f\big(\frac{\log(D/p)}{\log p}\big)\right)\Pr(\mathcal{A}_p,p).$$

Ahora ya estamos preparados para hacer la trampa mayor. Supondremos que se puede sustituir $\Pr(\mathcal{A},z)$ por $\mathrm{cte}X/\log z$ (lo cual es como decir de algún modo que $g(p)$ es igual a 1 en promedio). Si $\Pr(\mathcal{A},z)$ fuera realmente muy igual a $\mathrm{cte}X/\log z$, derivando en (2.7) se tendría (??) $X(\mathrm{cte}(\log z)^{-1})'=-\sum\Pr(\mathcal{A}_p,p)\delta(z-p)$ donde $\delta$ es la delta de Dirac. Por supuesto esto no tiene ningún sentido riguroso comenzando porque $\Pr(\mathcal{A},z)$ toma valores discretos y $\mathrm{cte}X/\log z$ es una función continua. Cerrando los ojos se sigue

$$\begin{aligned}S(\mathcal{A},z)&\lesssim\Pr(\mathcal{A},z)-X\int_1^z\left(1-f\big(\frac{\log(D/t)}{\log t}\big)\right)d(\mathrm{cte}(\log t)^{-1})\\&\approx\Pr(\mathcal{A},z)\left[1-\int_1^z\left(1-f\big(\frac{\log(D/t)}{\log t}\big)\right)\log z\;d((\log t)^{-1})\right].\end{aligned}$$

[5]Una analogía a través de un ejemplo es lo que ocurriría si en el proceso de Brun sólo se consideraran números con un factor primo ($r=1$), entonces $X-X/2-X/3-\cdots-X/p$ con $p\approx X$ es negativo y la cota es trivial.

Con el cambio de variable de $u = \log D/\log t$ y recordando que $D = z^s$, se llega a

$$S(\mathcal{A}, z) \lesssim \Pr(\mathcal{A}, z)\left(1 + \frac{1}{s}\int_s^\infty (1 - f(u-1))\, du\right).$$

Esto significa que dada una $f$ para la que se verifique la primera desigualdad de (2.6) se halla una $F$ válida para la segunda mediante

$$F(s) = 1 + \frac{1}{s}\int_s^\infty (1 - f(u-1))\, du$$

que se puede escribir en forma diferencial como $(sF(s))' = f(s-1)$. De la misma forma, dada una $F$ válida se obtiene una $f$ tal que $(sf(s))' = F(s-1)$. Si hasta $s = \beta$ sólo se tiene la cota inferior trivial, esto es, $f(s) = 0$ para $s \leq \beta$, entonces $F(s) = \text{cte}/s$ para $s \leq \beta+1$, ahora se podría sustituir en $(sf(s))' = F(s-1)$ y llegar a $f(s) = \text{cte}s^{-1}\log(s-1)$ para $\beta \leq s < \beta + 1$. No olvidemos que $\beta$ y la constante son desconocidas pero tenemos a nuestro favor que sabemos que $f$ y $F$ son 1 en el infinito, e iterando infinitas veces (??) se tendría un sistema $2 \times 2$ para estas incógnitas (??). En fórmulas, lo que se pretende es resolver

$$\begin{cases} (sF(s))' = f(s-1) & \text{si } s > \beta + 1 \\ (sf(s))' = F(s-1) & \text{si } s > \beta \\ 0 \leq f \leq f(\infty) = 1, & 1 = F(\infty) \leq F \\ f(s) = 0 \ \text{ si } s \leq \beta, & F(s) = \text{cte}/s \ \text{ si } \ s \leq \beta + 1 \end{cases}$$

Pues bien, se puede probar que las únicas posibilidades son $\beta = 2$ y que la constante tome el valor $2e^\gamma$ con $\gamma$ la constante de Euler. Entonces $f$ y $F$ son soluciones de la ecuación rara[6]

$$\begin{cases} (sF(s))' = f(s-1) & \text{si } s > 3 \\ (sf(s))' = F(s-1) & \text{si } s > 2 \\ f(s) = 0 \ \text{ si } s \leq 2, & F(s) = 2e^\gamma/s \ \text{ si } \ s \leq 3 \end{cases} \tag{2.8}$$

Dado un $s$ se puede hallar fácilmente $f(s)$ y $F(s)$ iterando a partir de las condiciones iniciales.

Después de toda esta historieta quedan dos cabos sueltos. En primer lugar el enunciado exacto en el que se materializan las ideas anteriores, y en segundo lugar explicar qué tiene que ver todo esto con la criba combinatoria mencionada al principio.

En la argumentación anterior se había empleado que $\Pr(\mathcal{A}, z)$ era, salvo constantes, como $X/\log z$ lo cual indica que de algún modo $g(p)$ es uno en promedio (por la fórmula de Mertens), incluso en intervalos pequeños porque por ejemplo la dudosa derivación en (2.7) parece requerir algún control en los incrementos.

[6]Esto es lo que se llama una *ecuación diferencial en diferencias*. Estas ecuaciones son naturales en modelos de población. Por ejemplo, que la tasa de variación de la población sea proporcional a la propia población no es muy creíble si los individuos tardan un tiempo $t_0$ grande en poder reproducirse tras su nacimiento. Esto se refleja en sustituir la ecuación $P'(t) = \alpha P(t)$ por $P'(t) = \alpha P(t - t_0)$.

La hipótesis se puede escribir como

$$\sum_{w\le p<z}\frac{g(p)\log p}{p}=\log z-\log w+O(1)\qquad\text{para}\quad 2\le w\le z. \tag{2.9}$$

Cuando se verifica esto se dice que la criba es *lineal* o que tiene *dimensión* uno. Se puede probar que es posible relajar la hipótesis cambiando la igualdad por menor o igual. En ese caso se dice que uno es la *dimensión débil*.

Con todo esto ya se puede enunciar el resultado principal.

**Teorema 2.4.1 (Jurkat-Richert)** *Sea un problema de criba con dimensión (débil) uno tal que para $D$ se cumple* (2.5)*. Entonces para $s=\log D/\log z>2$*

$$\big(f(s)+O(\Delta)\big)XV(z)+O(E)\le S(A,z)\le\big(F(s)+O(\Delta)\big)XV(z)+O(E)$$

*donde $E=X/(\log X)^C$, $\Delta=(\log\log\log D)^3/\log\log D$ y $f(s)$ y $F(s)$ son las soluciones de* (2.8)*.*

La segunda tarea pendiente es una explicación de la relación de este resultado con una criba combinatoria.

Supongamos que sabemos *a priori* que para $s\le 2$ sólo podemos obtener cotas inferiores triviales, entonces es una pérdida de términos de error aplicar la identidad de Buchstab con todos sus sumandos. Ya se mencionó que el $s$ correspondiente a $S(\mathcal{A}_p,p)$ es $s_p=\log(D/p)\big/\log p$, así que de nada sirven los términos con $2\ge\log(D/p)\big/\log p$. Despreciarlos pasa la igualdad a una desigualdad

$$S(\mathcal{A},z)\le\#\mathcal{A}-\sum_{\substack{p<z\\ p^3<D}}S(\mathcal{A}_p,p).$$

En una segunda iteración de Buchstab no se pueden descartar términos sin perder la desigualdad, por tanto

$$S(\mathcal{A},z)\le\#\mathcal{A}-\sum_{\substack{p_1<z\\ p_1^3<D}}A_{p_1}+\sum_{\substack{p_2<p_1<z\\ p_1^3<D}}S(\mathcal{A}_{p_1p_2},p_2).$$

Pero en la tercera iteración sí se pueden eliminar los términos con $2\ge s_{p_1p_2p_3}$, donde $s_{p_1p_2p_3}=\log(D/p_1p_2p_3)\big/\log p_3$, y se tiene

$$S(\mathcal{A},z)\le\#\mathcal{A}-\sum_{\substack{p_1<z\\ p_1^3<D}}A_{p_1}+\sum_{\substack{p_2<p_1<z\\ p_1^3<D}}A_{p_1p_2}-\sum_{\substack{p_3<p_2<p_1<z\\ p_1^3<D,\ p_3^3p_2p_1<D}}S(\mathcal{A}_{p_1p_2p_3},p_3).$$

Razonamientos análogos dan lugar a cotas inferiores. Con ello se tiene una criba combinatoria determinada por

$$\begin{aligned}\mathcal{D}^+ &= \{p_1p_2\cdots p_m\ :\ p_m<p_{m-1}<\ldots<p_1\text{ y }p_{2r+1}^3p_{2r}\cdots p_2p_1<D\text{ para }2r+1\le m\}\\ \mathcal{D}^- &= \{p_1p_2\cdots p_m\ :\ p_m<p_{m-1}<\ldots<p_1\text{ y }p_{2r}^3p_{2r-1}\cdots p_2p_1<D\text{ para }2r\le m\}\end{aligned}$$

Ésta es la *criba de Rosser*.

Al ser una criba combinatoria, se pueden controlar los términos de error bajo la condición (2.5) y se obtienen cotas del tipo

$$X \sum_{\substack{d|P(z)\\ d\in\mathcal{D}^-}} \mu(d)\frac{g(d)}{d} + \texttt{error} \le S(A,z) \le X \sum_{\substack{d|P(z)\\ d\in\mathcal{D}^+}} \mu(d)\frac{g(d)}{d} + \texttt{error}.$$

Ahora se espera extraer de estos sumandos un factor $V(z)$ y acumular las cantidades sobrantes en las funciones $f$ y $F$ pero eso es un trabajo duro y el torrente de palabras que precede al teorema la excusa para evitarlo.

Para festejar el fin del capítulo, veamos algunos ejemplos (esencialmente tomados de [He]). Cuando queramos subrayarlos los denominaremos corolarios.

Ejemplo. Volviendo al ejemplo de prueba de las secciones anteriores, para $\mathcal{A} = [1, N]$ se puede tomar $X = N$, $g(d) = 1$ y $r_d = O(1)$. La condición (2.5) está asegurada eligiendo $D \ll X/(\log X)^C$ y (2.9) es la fórmula de Mertens. Por otro lado, $s > 2$ requiere que $z$ no llegue a $N^{1/2}$ ¡justo el caso que se necesita para que $S(\mathcal{A}, z)$ sea como $\pi(z)$ con un error despreciable! Resignándonos a $z = N^{1/2-\epsilon}$ con $\epsilon$ pequeño, digamos $0 < \epsilon \le 1/6$, entonces $2 < s \le 3$ y según el teorema

$$\frac{2N}{\log N}\big(1 + o_\epsilon(1)\big)\log((1+2\epsilon)/(1-2\epsilon)) \le S(\mathcal{A}, N^{1/2-\epsilon}) \le \frac{2N}{\log N}\big(1 + o_\epsilon(1)\big)$$

donde se ha empleado que $\prod_{p<x}(1-1/p) \sim e^{-\gamma}/\log x$.

Ejemplo. Si se repite el ejemplo anterior pero ahora con $\mathcal{A} = [(N-1)^k, N^k]$ la diferencia es que $X = N^k - (N-1)^k = kN^{k-1} + O(N^{k-2})$ y $s > 2$ sugiere $z = N^{(k-1)/2-\epsilon}$ para lograr una cota inferior. Con ello se consigue $S(\mathcal{A}, N^{(k-1)/2-\epsilon}) > 0$ si $N$ es suficientemente grande. Por otro lado, si $n \in \mathcal{A}$ tienen todos sus factores primos mayores que $N^{(k-1)/2-\epsilon}$, necesariamente tiene a lo más $k/((k-1)/2-\epsilon)$ de ellos (pues $n \le N^k$) y este número es menor que 3 si $k \ge 4$ y $\epsilon$ es pequeño. Los casos $k = 2$ y $k = 3$ son peores.

**Corolario 2.4.2** *Entre dos cuadrados consecutivos (suficientemente grandes) siempre hay un número con a lo más cuatro factores primos y entre dos cubos consecutivos (suficientemente grandes) siempre hay un número con a lo más tres factores primos.*

Observación: Con los conocimientos actuales sobre diferencias de primos [Ba-Ha-Pi] la segunda parte de este corolario está muy anticuada pues se sabe que hay primos entre dos cubos (la primera también, pero sigue siendo un problema abierto saber si entre dos cuadrados hay siempre un primo).

Ejemplo. Sea $\mathcal{A} = \{n^2 + 1 \ : \ n \le N\}$. De la fórmula $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$, se tiene

$$A_p = |\{n \le N \ : \ p|n^2+1\}| = \begin{cases} 2N/p + O(1) & \text{si } p \equiv 1 \ (4) \\ 0 & \text{si } p \equiv 3 \ (4) \\ N/2 + O(1) & \text{si } p = 2 \end{cases}$$

Por tanto la elección natural es $X = N$, $g(2) = 1$, $g(p) = 2$ si $p \equiv 1$ (4) y $g(p) = 0$ si $p \equiv 3$ (4). En general se tiene por el teorema chino de resto $A_d = g(d)N/d + O(2^{\nu(d)})$ para $d$ libre de cuadrados. Así que con $D \ll X/\log^{C+1} X$ se tiene asegurado (2.5). Un punto más delicado es comprobar que (2.9) se cumple porque ahora a la suma sólo contribuyen la "mitad" de los primos (los de la forma $4k+1$) pero como para cada uno de ellos el valor de $g$ es 2 una cosa compensa a la otra (??); la explicación rigurosa de que los primos de esta forma son la mitad requiere un caso particular del teorema de los números primos en progresiones aritméticas.

La condición $s > 2$ que se necesita para tener una cota inferior no trivial lleva a tomar como $z$ aceptable $N^{1/2-\epsilon}$ con $\epsilon$ arbitrariamente pequeño. Es decir, por el teorema se cumple $S(\mathcal{A}, N^{1/2-\epsilon}) > 0$ para $N$ grande. Teniendo en cuenta que el mayor elemento de $\mathcal{A}$ es $N^2 + 1$, esto implica

**Corolario 2.4.3** *Hay infinitos números de la forma $n^2+1$ con a lo más cuatro factores primos.*

Por último un ejemplo doble de Selberg sin trabajar, a modo de comentario, que tiene gran interés teórico. Sean

$$\mathcal{A}^{\text{par}} = \{n \leq 2N \ : \ \lambda(n) = 1\} \qquad \text{y} \qquad \mathcal{A}^{\text{impar}} = \{n \leq 2N \ : \ \lambda(n) = -1\}$$

donde $\lambda$ es la función de Liouville que vale 1 si el número de factores primos (contando multiplicidades) es par y $-1$ si es impar. Se cumple $A_d^{\text{par}} = \frac{1}{2}\sum_{n \leq 2X/d}(1 + \lambda(d)\lambda(n))$ y se conoce por métodos como los empleados en la demostración del teorema de los números primos que $\sum_{n \leq N} \lambda(n) = O(N/\log^C N)$ para cualquier $C > 0$, por tanto para $X = D = N$, y $g(d) = 1$ se cumplen las hipótesis (2.5) y (2.9). Lo mismo se aplica a $\mathcal{A}^{\text{impar}}$.

Por la peculiar estructura de $\mathcal{A}^{\text{par}}$ y $\mathcal{A}^{\text{impar}}$ es posible aplicar el proceso de iteraciones de identidades de Buchstab indicado anteriormente y concluir [He] que para cada $s > 1$

$$S(\mathcal{A}^{\text{par}}, z) = f(s)NV(z) + O(E) \quad \text{y} \quad S(\mathcal{A}^{\text{impar}}, z) = F(s)NV(z) + O(E)$$

con $E = N/(\log N)^2$. Esto es lo mejor que se podría obtener con el teorema (salvo términos de error). Es decir, el teorema de criba es óptimo en el sentido de que los términos principales no se pueden mejorar, ya que para $\mathcal{A}^{\text{par}}$ y $\mathcal{A}^{\text{impar}}$ se alcanzan.

Como aspecto negativo, nótese que los conjuntos $\mathcal{A}^{\text{par}}$ y $\mathcal{A}^{\text{impar}}$ son indistinguibles desde el punto de vista de la criba (ya que $A_d^{\text{par}}$ y $A_d^{\text{impar}}$ son similares), y sin embargo los comportamientos asintóticos de $S(\mathcal{A}^{\text{par}}, z)$ y $S(\mathcal{A}^{\text{impar}}, z)$ son diferentes. Esto limita la posibilidad de obtener ciertas fórmulas asintóticas con métodos de criba si no se introduce información adicional que no está en las hipótesis del teorema de Jurkat-Richert.

# Bibliografía


[Ba-Ha-Pi] R.C. Baker, G. Harman, J. Pintz. The difference between consecutive primes. II. Proc. London Math. Soc. (3) 83 (2001), no. 3, 532–562.

[Bo] E. Bombieri. The asymptotic sieve. Rend. Accad. Naz. XL (5) 1/2 (1975/76), 243–269 (1977).

[Br] V. Brun. Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare. Archiv for Math. og Naturvid B34 (2001), no. 8.

[Ci-Co] J. Cilleruelo y A. Córdoba. La teoría de los números. Mondadori, Madrid, 1992.

[Ha-Ri] H. Halberstam, H.-E. Richert. Sieve methods. London Mathematical Society Monographs, No. 4. Academic Press, London-New York, 1974.

[Ha-Wr] G.H. Hardy, E. Wright. An introduction to the theory of numbers. Fifth edition. The Clarendon Press, Oxford University Press, New York, 1979.

[He] D.R. Heath-Brown. Lectures on sieves. Proceedings of the Session in Analytic Number Theory and Diophantine Equations, Bonner Math. Schriften, 360, Univ. Bonn, Bonn, 2003.

[Sh-Wr] D. Shanks; J.W. Wrench. Brun's constant. Math. Comp. 28 (1974), 293–299; corrigenda, ibid. 28 (1974), 1183.

[Se] A. Selberg. Collected papers. Vol. II. Springer-Verlag, Berlin, 1991.