Complejidad de los números naturales

por

J. Arias de Reyna

1. Introducción

1. Complejidad de un número natural

Voy a hablar de un tema aparentemente menor, que puede comprender un estudiante con catorce años, pero que encierra dificultades muy profundas. Llevo algún tiempo interesado en uno de los problemas que considero más importantes de entre los que tienen planteados los matemáticos: el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$. En este caso, la primera dificultad es exponer el problema de forma asequible a un matemático tradicional, digamos que a un especialista en Análisis Matemático. No es ésta una cuestión banal, pues creo que debe poder plantearse como un problema de acotación, y que entiendan el problema los analistas puede ser el paso principal hacia la solución. La cuestión de la que quiero hablar aquí surgió en un intento de conseguir esta explicación.

Empecemos con la pregunta principal: dado un número natural n, ¿cuántos unos son necesarios para escribir n? Por ejemplo,

$$19 = 1 + (1 + 1)(1 + 1 + 1)(1 + 1 + 1),$$

luego no se necesitan más de 9 unos para escribir el número 19. Decimos entonces que la complejidad de 19 es menor o igual a 9, lo que escribiremos abreviadamente por $\|19\| \leq 9$. Naturalmente la complejidad de 19 será el número de unos en la representación de 19 que menos unos utilice. Sólo se admiten expresiones con las operaciones de suma y de producto.

Los primeros valores de la función complejidad pueden calcularse con no demasiado trabajo:

$$1, 2, 3, 4, 5, 5, 6, 6, 6, 7, 8, 7, 8, 8, 8, 8, 9, 8, 9, 9, \ldots$$

Vemos que no forman una sucesión monótona: $8 = \|11\| > \|12\| = 7$.

Cuando en algún problema matemático surge una sucesión de números naturales, hay algo que debemos hacer: consultar *The Encyclopedia of Integer Sequences* de Sloane y Plouffe [SP]. En ella encontramos la sucesión anterior y nos remite al artículo de Guy [G], donde se la define y analiza.

2. Complejidad de un número natural

Hemos definido la complejidad como una función $n \mapsto \|n\|$ de $\mathbf{N} \to \mathbf{N}$ tal que para todo par de números naturales m y n se tiene

$$\|1\| = 1, \qquad \|m+n\| \leq \|m\| + \|n\|, \qquad \|m \cdot n\| \leq \|m\| + \|n\|.$$

De hecho es la mayor función que cumple estas condiciones. Para probar ésta y otras afirmaciones es útil introducir otro concepto: el de expresión.

2. Definición de las expresiones

Una expresión es una sucesión de símbolos. Los símbolos permitidos son los cuatro siguientes `x, +, (, )`. No toda sucesión de símbolos es una expresión. Ejemplos de expresiones son:

`(x + x)`; `(x+(xx))`; `(x+((x+x)((x+(x+x))(x+(x+x)))))`.

La definición formal es una definición inductiva:

(a) `x` es una expresión.

(b) Si `A` y `B` son expresiones, también lo son `(A+B)` y `(AB)`.

(c) Sólo son expresiones las sucesiones finitas de símbolos que resulten de aplicar reiteradamente las reglas (a) y (b).

Definimos el valor de una expresión `A` como el número $v(\mathtt{A})$ que resulta de sustituir `x` por 1 y efectuar las operaciones indicadas. De nuevo usamos la inducción para definir el valor v: $v(\mathtt{x}) = 1$, y si `A` y `B` son expresiones $v(\mathtt{(A+B)}) = v(\mathtt{A}) + v(\mathtt{B})$ y $v(\mathtt{(AB)}) = v(\mathtt{A})v(\mathtt{B})$.

Dada una expresión, podemos definir su complejidad como el número de letras iguales a `x` que contiene, por ejemplo, $\|\mathtt{(x+(xx))}\| = 3$. Sea $\mathcal{E}$ el conjunto de las expresiones. La definición de la complejidad puede expresarse ahora en la forma

$$\|n\| = \inf\{\|\mathtt{A}\| : \mathtt{A} \in \mathcal{E} \text{ y } \ v(\mathtt{A}) = n\}.$$

Si queremos calcular el valor de $\|n\|$ debemos usar la proposición siguiente:

Proposición 1 *Para todo número natural $n \in \mathbf{N}$,*

$$\|n\| = \min\left\{(\|d\| + \|n/d\|, \|j\| + \|n-j\|) : \begin{array}{ll} 2 & \leq d \leq \sqrt{n}, d/n \\ 1 & \leq j \leq n/2 \end{array}\right\}$$

Prueba. Supongamos que $n > 1$. Sea `E` una expresión óptima de n, es decir una que dé su complejidad, $\|n\| = \|\mathtt{E}\|$. Ahora la expresión será o

bien $\mathtt{E} = \mathtt{(A + B)}$ ó bien $\mathtt{E} = \mathtt{(AB)}$. Pongamos $a = v(\mathtt{A})$, $b = v(\mathtt{B})$. De este modo, ó bien $n = a+b$ y $\|n\| = \|a\|+\|b\|$, ó bien $n = ab$ y $\|n\| = \|a\|+\|b\|$. En el primer caso, si j el menor de los dos, a y b, se tiene $1 \leq j \leq n/2$, y en el segundo, si d es el menor de los dos, d es un divisor de n con $2 \leq d \leq \sqrt{n}$. Naturalmente, para que el razonamiento anterior sea válido, debemos comprobar que, si $\mathtt{E}$ es una expresión óptima de n, entonces $\mathtt{A}$ y $\mathtt{B}$ deben ser expresiones óptimas de a y b. Dejamos dicha comprobación al lector. □

Usando el esquema anterior hemos calculado, con el programa Mathematica, los valores de $\|n\|$ para $1 \leq n \leq 200\,000$.

3. Cotas

Proposición 2 *Sea $P\colon \mathbf{N} \to \mathbf{R}$ una aplicación tal que*

$$P(1) = 1, \quad P(n+m) \leq P(n) + P(m), \quad P(n \cdot m) \leq P(n) + P(m).$$

Entonces, para todo $n \in \mathbf{N}$, se tiene $P(n) \leq \|n\|$.

Prueba. Vemos que, para toda expresión $\mathtt{A}$, se tiene $P\big(v(\mathtt{A})\big) \leq \|\mathtt{A}\|$. Usamos inducción. Es cierto para $\mathtt{A} = \mathtt{x}$, y, si es cierto para $\mathtt{A}$ y $\mathtt{B}$, es también cierto para $\mathtt{(A+B)}$ y $\mathtt{(AB)}$. En efecto, para el producto:

$$P\big(v(\mathtt{(AB)})\big) = P\big(v(\mathtt{A})v(\mathtt{B})\big) \leq P\big(v(\mathtt{A})\big) + P\big(v(\mathtt{B})\big) \leq \|\mathtt{A}\| + \|\mathtt{B}\| = \|\mathtt{(AB)}\|,$$

y un argumento análogo vale para la suma. (Observar que, por la definición de v, se tiene $v(\mathtt{(A+B)}) = v(\mathtt{A}) + v(\mathtt{B})$ y $v(\mathtt{(AB)}) = v(\mathtt{A})v(\mathtt{B})$).

Basta ahora tomar ínfimo en $P\big(v(\mathtt{A})\big) \leq \|\mathtt{A}\|$ para todas las expresiones $\mathtt{A}$ tales que $n = v(\mathtt{A})$. Se obtiene entonces $P(n) \leq \|n\|$. □

Corolario 3 *Para todo número natural n, $\log_2(1+n) \leq \|n\|$.*

Prueba. Basta comprobar las propiedades de $P(n) = \log_2(1+n)$. □

Más adelante en el corolario 9 mejoraremos esta desigualdad.

3. Cotas superiores

A continuación establecemos una cota superior. Con este objeto introducimos una función $L\colon \mathbf{N} \to \mathbf{N}$.

Definición 4 *Definimos la función L inductivamente:*
(a) $L(1) = 1$.

(b) Si p es un número primo, $L(p) = 1 + L(p-1)$.

(c) Si $n = p_1 p_2 \cdots p_k$ es un producto de números primos iguales o diferentes, entonces $L(p_1 p_2 \cdots p_k) = L(p_1) + L(p_2) + \cdots L(p_k)$.

Con esta definición es claro que si $n = ab$, siendo a y b mayores o iguales a 2, entonces se tiene $L(n) = L(a) + L(b)$.

Proposición 5 *Para todo $n \in \mathbf{N}$, se tiene*

$$\|n\| \leq L(n).$$

Prueba. Podemos probarlo por inducción. Para $n = 1$, tenemos $\|1\| = L(1) = 1$. Supongamos que se cumple $\|k\| \leq L(k)$, para todo $k < n$. Pueden darse dos posibilidades: Si $n = p$ es un número primo,

$$\|p\| \leq \|p-1\| + \|1\| = \|p-1\| + 1 \leq L(p-1) + 1 = L(p).$$

Si $n = ab$ con a y $b > 2$,

$$\|n\| \leq \|a\| + \|b\| \leq L(a) + L(b) = L(ab) = L(n).$$

□

Proposición 6 *Para todo $n \geq 2$ se tiene*

$$L(n) \leq \frac{3}{\log 2}(\log n).$$

Prueba. En primer lugar, puesto que $L(2) = 2$, el resultado es cierto para $n = 2$.

Supongamos ahora que $n \geq 3$, y que la desigualdad es válida para números naturales menores que n.

Si $n = p$ es primo, se tiene

$$L(p) = 1 + L(p-1) = 1 + 2 + L\Big(\frac{p-1}{2}\Big) \leq 3 + \frac{3}{\log 2}\log\Big(\frac{p-1}{2}\Big). \quad (1)$$

Queremos que esto sea

$$\leq \frac{3}{\log 2}(\log p).$$

Es decir, basta comprobar que

$$3 \leq \frac{3}{\log 2}\log\Big(\frac{2p}{p-1}\Big), \quad (2)$$

lo cual se cumple para $p \geq 3$.

Si $n = ab$, con a y $b \geq 2$, se tiene

$$L(ab) = L(a) + L(b) \leq \frac{3}{\log 2}(\log a) + \frac{3}{\log 2}(\log b) = \frac{3}{\log 2}(\log ab).$$

□

Nota 1. No sabemos si la constante $3/\log 2$ en el teorema anterior es óptima. Analizando la prueba, sospechamos que el cociente $L(n)/\log n$ es grande cuando $n = p_k$ sea un primo tal que exista una sucesión de primos $(p_j)_{j=1}^{k}$ de forma que $p_j = 2p_{j+1} + 1$. Por ejemplo, los números 89, 179, 359, 719, 1439, 2879 son todos primos, y el máximo valor del cociente $L(n)/\log n$ que conocemos es

$$\frac{L(2879)}{\log 2879} = 3.766384578\cdots < 4.328085123\cdots = \frac{3}{\log 2}.$$

La diferencia principal entre las dos funciones $L(\cdot)$ y $\|\cdot\|$ consiste en que $L(\cdot)$ es multiplicativa y $\|\cdot\|$ no. Para cada pareja de números n y m mayores que 1, la función L verifica $L(nm) = L(n) + L(m)$. En cambio, existen n y m mayores que 1 tales que $\|nm\| < \|n\| + \|m\|$. Diremos que $n \cdot m$ es una mala factorización.

En la figura 1 situamos un punto en (n, m) cada vez que $n \times m$ es mala factorización. La figura cubre todos los factores n ó $m \leq 60$.

Naturalmente $1 \cdot m$ es siempre mala factorización, pero en la figura aparecen otras regularidades sorprendentes. Así, saltan a la vista ciertas alineaciones de puntos, las más prominentes se situan en $n = 23$, 41 y 59, que merecen una explicación.

Estos números, diríamos que malos factores, parecen tener una complejidad grande. Definimos la sucesión de *números con complejidad grande* n_k: son aquellos tales que n_k es la menor solución de $\|n\| = k$. Los primeros valores de esta sucesión son

1, 2, 3, 4, 5, 7, 10, 11, 17, 22, 23, 41, 47, 59, 89, 107, 167, 179, 263, 347, 467 , 683, 719, 1223, 1438, 1439, 2879, 3767, 4283, 6299, 10079, 11807, 15287, 21599, 33599, ...

que aparece en [SP] con alguna errata. Encontramos así la referencia a Rawsthorne [R].

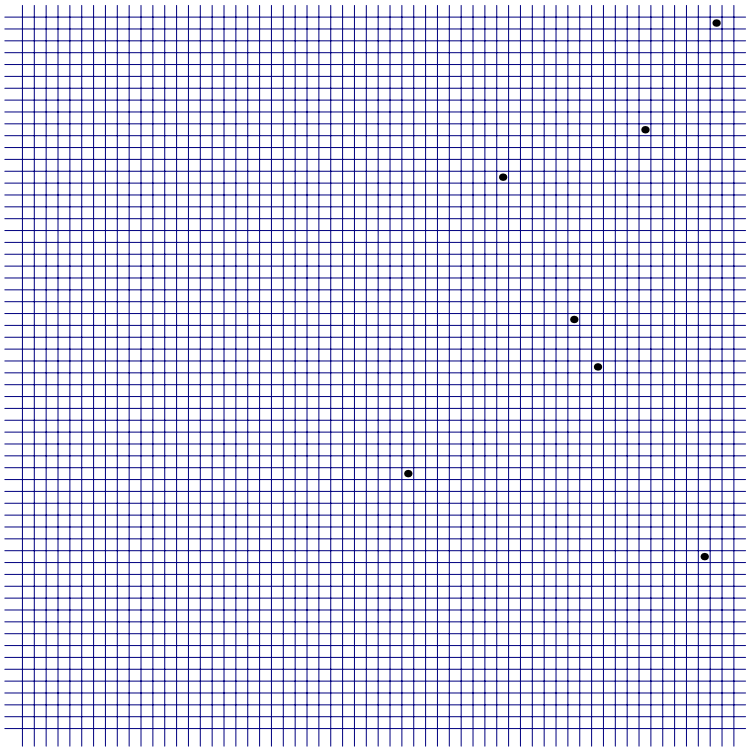

Figura 1. Malos factores

4. Valores medios

Existe otra prueba de que $\|n\| \leq 3 \log n / \log 2$. Consiste en observar que, si escribimos n en binario $n = \sum_{j=0}^{k-1} \varepsilon_j 2^j + 2^k$, tenemos una forma de expresar n:

$$n = \varepsilon_0 + 2(\varepsilon_1 + 2(\varepsilon_2 + \cdots + 2(\varepsilon_{k-2} + 2(\varepsilon_{k-1} + 2)) \cdots)),$$

donde podemos sustituir cada 2 por $1 + 1$ y cada cifra ε_j es 0 ó 1. De este modo obtenemos una expresión de n usando a lo más $2k + k$ unos, donde k cumple $2^k \leq n < 2^{k+1}$. Luego $\|n\| \leq 3 \log n / \log 2$.

El razonamiento anterior prueba que la función $L_2(n) = 2k + \varepsilon_0 + \varepsilon_1 + \cdots \varepsilon_{k-1}$ es otra cota superior de $\|n\|$. La comparación de $L_2(n)$ y de $L(n)$ no es fácil. Entre los primeros 1000 números, generalmente $L(n)$ es menor, pero esto tiene excepciones. La primera es $L_2(161) = 16 < 17 = L(161)$. En este rango la diferencia es pequeña.

La función $L_2(n)$ puede usarse para obtener información sobre la función $\|\cdot\|$. Consideremos el conjunto de los números n que se escriben en binario en la forma $1\varepsilon_{k-1}\dots\varepsilon_0$, es decir, con $k+1$ cifras. Según la expresión anterior, tenemos

$$\|n\| \le 2k + \varepsilon_0 + \cdots + \varepsilon_{k-1}.$$

Podemos pensar que ε_j son variables aleatorias independientes de media $1/2$. La desigualdad de Chernoff (ver [C] y para una exposición sencilla [AS]) nos dice que

$$\mathbb{P}\left(\left|\sum \varepsilon_j - k/2\right| < x\sqrt{k}\right) \ge 1 - 2e^{-2x^2}.$$

Se sigue que $\mathbb{P}(\|n\| \le 2k + k/2 + x\sqrt{k}) \ge 1 - 2e^{-2x^2}$. Finalmente, con $x = \sqrt{\log k}$,

$$\mathbb{P}\left(\|n\| > 5k/2 + \sqrt{k\log k}\right) \le 2k^{-2}.$$

Luego entre los 2^k valores de n con $2^k \le n < 2^{k+1}$ a lo más $(2/k^2)2^k$ verifican $\|n\| > 5k/2 + \sqrt{k\log k}$. Los demás, la mayor parte, cumplen

$$\|n\| \le \frac{5k}{2} + \sqrt{k\log k} = \frac{5}{2}\frac{\log n}{\log 2} + O(\sqrt{\log n \log\log n}).$$

De algún modo podemos decir que para casi todos los valores grandes de n se tiene

$$\|n\| \le \frac{5}{2}\frac{\log n}{\log 2} + O(\sqrt{\log n \log\log n}).$$

La cota superior $L(n)$ es extraordinariamente buena para valores pequeños de n. Por ejemplo, entre los primeros 220 valores de n, $L(n) = \|n\|$, salvo para los indicados en la tabla siguiente:

n	$\|\|n\|\|$	$L(n)$
46	12	13
47	13	14
55	12	13
82	13	14
83	14	15
92	14	15
94	15	16
110	14	15

n	$\|\|n\|\|$	$L(n)$
115	15	16
118	15	16
121	15	16
138	15	16
139	16	17
141	16	17
145	15	16
161	16	17

n	$\|\|n\|\|$	$L(n)$
164	15	16
165	15	16
166	16	17
167	17	18
184	16	17
188	17	18
217	16	17
220	16	17

En estos casos la cota $L_2(n)$ es igual o mayor que $L(n)$, salvo para el valor 161.

Las dos funciones $L(n)$ y $\|n\|$ coinciden en 771 valores de n para $1 \le n \le 1000$, siendo la diferencia igual a 1 en los otros 229 casos, salvo unas pocas excepciones.

4. Valores particulares

5. Números con complejidad pequeña

Una cota inferior para la complejidad $\|n\|$ resulta de resolver la cuestión de qué número N podemos alcanzar con m unos. Esto es, dado m, cuál es el mayor número natural N tal que $\|N\| = m$. La respuesta, *grosso modo*, es que debemos agrupar los m unos disponibles en grupos de 3 y multiplicarlos. Para ver esto definimos el concepto de expresión extremal. Sea $\mathtt{M}_m$ una expresión con $\|\mathtt{M}_m\| = m$, (es decir, $\mathtt{M}_m$ está formada con m $x's$ y las operaciones de suma y producto), y tal que su valor $v(\mathtt{M}_m)$ sea máximo entre las expresiones formadas con m unos, esto es

$$N = v(\mathtt{M}_m) = \sup_{\|\mathtt{A}\|=m} v(\mathtt{A}).$$

Diremos que $\mathtt{M}_m$ es extremal.

Afirmamos entonces que $\|N\| = m$. En efecto, por ser $N = v(\mathtt{M}_m)$ y $\|\mathtt{M}_m\| = m$, se tiene $\|N\| \leq m$. Supongamos, por contra, que fuese $\|N\| < m$. Existiría entonces una expresión $\mathtt{B}$ tal que $v(\mathtt{B}) = N$ y $\|\mathtt{B}\| = \|N\| < m$. Sea d tal que $m = d + \|\mathtt{B}\|$. Podemos construir una expresión $\mathtt{C}$ de la forma $\mathtt{C} = \mathtt{B} + \mathtt{x} + \cdots + \mathtt{x}$, y tal que $\|\mathtt{C}\| = \|\mathtt{B}\| + d = m$ y $v(\mathtt{C}) = v(\mathtt{B}) + d > N$. Esto contradice la definición de $\mathtt{M}_m$.

Es fácil comprobar que las siguientes expresiones son extremales

$$\mathtt{M}_1 = \mathtt{x}, \quad \mathtt{M}_2 = \mathtt{(x + x)}, \quad \mathtt{M}_3 = \mathtt{(x + (x+x))},$$

$$\mathtt{M}_4 = \mathtt{(x+x)(x+x)}, \quad \mathtt{M}_5 = \mathtt{(x+(x+x))(x+x)}, \ldots$$

Como vemos, dado m, la expresión extremal $\mathtt{M}_m$ no es única, por ejemplo $\mathtt{M}_4 = \mathtt{(x+(x+(x+x)))}$ es otra posibilidad.

Usaremos una notación poco precisa, por ejemplo, escribiremos $\mathtt{M}_3^a\mathtt{M}_2$ para denotar cualquier expresión que tenga esa forma sin precisar cómo construimos el producto a partir de los factores. Así $\mathtt{M}_3^4$ puede denotar cualquiera de las expresiones $\mathtt{((M_3M_3)(M_3M_3))}$, $\mathtt{(M_3(M_3(M_3M_3)))}$, o cualquier otra forma de agrupar los factores.

Proposición 7 *Sean* $\mathtt{M}_2 = \mathtt{(x + x)}$, $\mathtt{M}_3 = \mathtt{(x + (x+x))}$ *y* $\mathtt{M}_4 = \mathtt{(x+x)(x+x)}$. *Para* $n > 1$, *las expresiones* $\mathtt{M}_n$ *definidas por*

$$\mathtt{M}_n = \begin{cases} \mathtt{M}_3^k & \text{si } n = 3k, \\ \mathtt{M}_3^{k-1}\mathtt{M}_4 & \text{si } n = 3k+1, \\ \mathtt{M}_3^k\mathtt{M}_2 & \text{si } n = 3k+2, \end{cases} \qquad \textit{son extremales.}$$

Prueba. Directamente podemos comprobar que el resultado es válido para $n = 2$, 3 y 4.

Supongamos que es válida para todo $s < n$ y tratemos de probarlo para $n \geq 5$. Ciertamente existe una expresión extremal $\mathtt{K}$ con $\|\mathtt{K}\| = n$. Existen entonces dos expresiones $\mathtt{A}$ y $\mathtt{B}$ tales que $\mathtt{K} = \mathtt{(A+B)}$ o bien $\mathtt{K} = \mathtt{(AB)}$. Tanto $\mathtt{A}$ como $\mathtt{B}$ son extremales, en otro caso $\mathtt{K}$ no lo sería. Podemos cambiar $\mathtt{A}$ y $\mathtt{B}$ por expresiones extremales de la misma complejidad y la expresión resultante $\mathtt{K}'$ seguirá siendo extremal. Por tanto sin restringir la generalidad podemos suponer, usando la hipótesis de inducción, que $\mathtt{A}$ y $\mathtt{B}$ son de la forma dada en el enunciado, o bien $\mathtt{A} = \mathtt{x}$ y $\mathtt{B}$ como en el enunciado

El caso de ser $\mathtt{K} = \mathtt{(A+B)}$ sólo es posible si $v(\mathtt{A})$ o $v(\mathtt{B}) = 1$, (en otro caso la expresión $\mathtt{(AB)}$ contradice la extremalidad de $\mathtt{K}$). Pero $\mathtt{K} = \mathtt{(x+M}_3^k\mathtt{)}$, $\mathtt{K} = \mathtt{(x+M}_3^{k-1}\mathtt{M}_4\mathtt{)}$, o $\mathtt{K} = \mathtt{(x+M}_3^k\mathtt{M}_2\mathtt{)}$ es imposible con $n \geq 5$. Pues estas expresiones claramente no son extremales. (Compararlas con $\mathtt{M}_3^{k-1}\mathtt{M}_4$, $\mathtt{M}_3^k\mathtt{M}_2$, o $\mathtt{M}_3^{k+1}$ respectivamente).

Llegamos pues a la conclusión de que $\mathtt{K} = \mathtt{(AB)}$, siendo $\mathtt{A}$ y $\mathtt{B}$ de la forma dada en el enunciado. Algunas de las combinaciones no son posibles: por ejemplo, $\mathtt{A} = \mathtt{M}_3^k\mathtt{M}_2$ y $\mathtt{B} = \mathtt{M}_3^{j-1}\mathtt{M}_4$ no es posible pues $\mathtt{M}_3^{k+j-1}\mathtt{M}_4\mathtt{M}_2$ es mejorada por $\mathtt{M}_3^{k+j-1}$ y $\mathtt{K}$ no sería extremal. Un estudio caso por caso, demuestra que $\mathtt{K}$ es de la forma dada en el enunciado. □

De lo anterior se sigue

Corolario 8 *Para $a = 0$, 1 ó 2 y $b \in \mathbf{N}$ se tiene*

$$\|2^a 3^b\| = 2a + 3b, \qquad a = 0, 1, 2.$$

Todo número natural $n > 1$ se escribe de manera única en la forma $n = 2a + 3b$ siendo $a = 0$, 1 ó 2. En ese caso $2^a 3^b$ es el mayor número m tal que $\|m\| = n$. Por tanto $m > 2^a 3^b$ implica $\|m\| > 2a + 3b$.

Definimos $g\colon \mathbf{N} \to \mathbf{N}$ en la forma

$$g(n) = \begin{cases} 3a & \text{si } n \in \left[3^a, 3^a + 3^{a-1}\right), \\ 3a+1 & \text{si } n \in \left[3^a + 3^{a-1}, 2 \cdot 3^a\right), \\ 3a+2 & \text{si } n \in \left[2 \cdot 3^a, 3^{a+1}\right). \end{cases}$$

Se tiene entonces que para todo n, $g(n) \leq \|n\|$.

Corolario 9 *Para todo $n \geq 1$ tenemos*

$$3\frac{\log n}{\log 3} \leq \|n\| \leq L(n) \leq 3\frac{\log n}{\log 2}.$$

Prueba. Sólo debemos probar la primera desigualdad. Si $n = 3^a$, directamente vemos que se cumple la igualdad. Si $x \in \left(3^a, 3^a + 3^{a-1}\right]$, se tiene $\|x\| \geq 3a + 1$. Entonces

$$\|x\| \geq \|3^a\| + 1 = 3a + 1 \geq 3\frac{\log(4 \cdot 3^{a-1})}{\log 3} \geq 3\frac{\log x}{\log 3}.$$

Análogamente para $x \in \left(4 \cdot 3^{a-1}, 2 \cdot 3^a\right]$ tenemos

$$\|x\| \geq \|4 \cdot 3^{a-1}\| + 1 \geq 3a + 2 \geq 3\frac{\log(2 \cdot 3^a)}{\log 3}.$$

Finalmente para $x \in \left(2 \cdot 3^a, 3^{a+1}\right]$, basta comprobar que

$$\|x\| \geq \|2 \cdot 3^a\| + 1 = 3a + 3 \geq 3\frac{\log(3^{a+1})}{\log 3}.$$

□

5. El problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ y la complejidad de los naturales

6. Idea del problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$

Antes de explicar en que consiste este problema debemos explicar lo que se entiende por las clases **P** y **NP**. En primer lugar consideraremos un alfabeto finito A. Designamos por A^* el conjunto deb las *palabras* es decir sucesiones finitas de elementos de A.

Se llama *lenguaje* a un subconjunto $S \subset A^*$. S está en la clase **P** si existen un algoritmo T y un polinomio $p(x)$ tales que con el dato de entrada x, T proporciona la salida $T(x)$, de forma que $T(x) = 1$ si $x \in S$ y $T(x) = 0$ si $x \notin S$. Además T proporciona la salida $T(x)$ en un tiempo acotado por $p(|x|)$ (aquí $|x|$ denota la longitud de la palabra x). Se dice en este caso que T es un algoritmo polinómico. En pocas palabras podemos decir que **P** es la clase de los lenguajes reconocibles en tiempo polinómico. Es importante hacer notar que este concepto es muy estable frente a las posibles definiciones de lo que es un algoritmo, lo que se entiende por el tiempo en que T actúa, o incluso también si consideramos el "mismo" conjunto con distinto alfabeto (como puede ser dar un conjunto de naturales en distintas bases de numeración). Es decir no varía la clase si se dan definiciones sensatas de estos conceptos.

La siguiente clase **NP** es la de los lenguajes reconocibles por algoritmos polinómicos no deterministas. Es decir, $S \subset A^*$ está en **NP** si existen un algoritmo T y un polinomio $p(x)$ de forma que para cada $x \in S$ existe

$y \in A^*$ tal que $|y| \leq p(|x|)$ y con el dato de entrada (x, y) el algoritmo proporciona la salida $T(x, y) = 1$ en tiempo acotado por $p(|x|)$. En cambio si $x \notin S$, $T(x, y) = 0$ para todo y con $|y| \leq p(|x|)$.

Se dice que en este caso T es un algoritmo no determinista porque para deducir que $x \in S$ debemos escoger antes un y. Si ya conocemos este y el proceso es rápido. En cambio si no conocemos y, podríamos en teoría ensayar cada posible y, pero esto lleva a un tiempo $\geq |A|^{p(|x|)}$ que en la práctica suele ser prohibitivo.

De nuevo la clase **NP** es muy estable ante las posibles imprecisiones de nuestra definición. Además una enorme cantidad de problemas prácticos resultan ser de esta clase.

Es fácil comprobar que $\mathbf{P} \subset \mathbf{NP}$. La pregunta es si estas dos clases son o no la misma. Para comprender algo más la dificultad observemos lo siguiente:

Nuestra experiencia como matemáticos nos hace ver que comprender una prueba, mejor dicho comprobar la corrección de una prueba, es una tarea que es de tipo **P**. Es decir el tiempo requerido es proporcional a la longitud de la pueba.

En cambio comprobar si una conjetura x es un teorema requiere antes escribir la prueba y y despues aplicar el procedimiento anterior de comprobar la corrección de (x, y). El conjunto de los teoremas no está en la clase **NP** pues en general sabemos que la longitud de la prueba $|y|$ no está acotada por la longitud del enunciado $|y| \not\leq p(|x|)$. Pero, para cada polinomio $p(t)$, sí está en la clase **NP** el siguiente conjunto

$$\mathcal{T}_p = \left\{ \begin{array}{ll} x & : x \text{ es un teorema que admite una demostración de} \\ & \text{longitud} \leq p(|x|) \end{array} \right\}$$

Puede alguien pensar que estas definiciones son imprecisas, pero la lógica formal permite precisarlas adecuadamente.

Más aún si fuese $\mathbf{P} = \mathbf{NP}$ y la prueba fuera suficientemente constructiva (técnicamente, que se encontrara un algoritmo polinómico para un problema **NP**-completo), existiría un algoritmo polinómico que permitiría no sólo decidir que $x \in \mathcal{T}_p$, sino además escribir la prueba de x en tiempo polinómico. Yo diría que los matemáticos habríamos dejado de ser necesarios.

Cuando uno recuerda los logros de este siglo XX: demostración del teorema de Fermat, clasificación de los grupos simples finitos, convergencia de las series de Fourier de funciones en L^p, hipótesis de Riemann para las variedades algebráicas sobre cuerpos de característica p, independencia de la hipótesis del continuo, y tantos otros, da la impresión de que existe un algoritmo que decide si $x \in \mathcal{T}_p$, buscando directamente la prueba y, no por ensayo y error sino de otro modo. El algoritmo consiste en tomar alumnos

prometedores, darles posibilidad de viajar y hablar con los especialistas en el tema, intentar la solución de problemas análogos, estudiar la solución de problemas relacionados, etc, ...

7. Conexión de la complejidad de los naturales con el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$

Consideremos la afirmación $\|4787\| = 28$. Podemos descomponer esta afirmación en dos. La primera de ellas, $\|4787\| \leq 28$, tiene una prueba muy fácil

$$4787 = 2 + 3(2 + 3^2)(1 + 2^4 3^2) \qquad (*)$$

Sin embargo, la otra parte de la afirmación $\|4787\| \geq 28$, tiene una prueba mucho más laboriosa. Ahora mismo desconozco otro camino que no lleve implícito calcular los valores de $\|n\|$ para todo $n \leq 4787$, tarea que, en mi ordenador personal, requirió varias horas.

Naturalmente que esto no quiere decir que sea fácil encontrar pruebas como la $(*)$.

Consideremos los dos conjuntos

$$A = \{(n, c) \in \mathbf{N}^2 : \|n\| \leq c\}, \qquad B = \{(n, c) \in \mathbf{N}^2 : \|n\| > c\}.$$

El hecho que hemos notado de que si es cierto $(n, c) \in A$, existe una prueba relativamente corta de ello, suele enunciarse diciendo que A está en la clase **NP**.

Con poca precisión, un conjunto A está en **NP** si, para probar que $x \in A$, se requiere una búsqueda exhaustiva que, en principio, es de tamaño exponencial respecto al de x, en cambio, una vez hallada la prueba, es fácilmente reconocible (en tiempo polinómico respecto del tamaño de x). Mayor información puede encontrarse en el libro [GJ]. Estos problemas recuerdan el de buscar una aguja en un pajar. Una vez que la encontramos no hay duda de que la tarea se ha realizado, pero en principio se nos impone como inalcanzable pues la paja es lo suficientemente parecida a la aguja como para que no encontremos más medio que mirar detenidamente y con buen orden.

Lo que plantea el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ es si, en estas situaciones en que existe una prueba corta, hay siempre un camino directo para encontrarla. Decir que $\mathbf{P} = \mathbf{NP}$ equivale a decir que existe un camino que permite encontrar la prueba rápidamente, sin titubeos. A primera vista esto parece descabellado, pero la demostración rigurosa de que $\mathbf{P} \neq \mathbf{NP}$ se sigue resistiendo despues de veintisiete años de estudio.

Recientemente Microsoft ha fundado un instituto de investigación que, entre otros, ha contratado a Michael Friedman, medalla Fields en el año

1986. Friedman tiene intención de atacar el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ y Microsoft piensa invertir 2.600 millones de dólares anualmente para subvencionarlo.

Parece que todo apunta a que la afirmación $\mathbf{P} = \mathbf{NP}$ es falsa, pero no todo es tan simple. A veces, tareas que parecían requerir una búsqueda exhaustiva han resultado más fáciles. Vamos a verlo con un ejemplo:

Sea $\mathcal{C} \subset \mathbf{N}$ el conjunto de los números compuestos. A primera vista parece que el único medio de probar que n es compuesto es dividir n por cada número $m \leq \sqrt{n}$ y comprobar si el resto es cero en algún caso. El tamaño de n es del orden del número de cifras que lo expresan, es decir del orden de $\log n$. El número de comprobaciones es $\sqrt{n} = e^{(\log n)/2}$, que crece exponencialmente con $\log n$. Si realmente n es compuesto, existe una demostración corta: exhibir un divisor d de n. Esto es decir que $\mathcal{C}$ está en la clase $\mathbf{NP}$.

En realidad no es tan difícil decidir si n es compuesto. Si n es primo y b es primo con n se tiene $b^{n-1} \equiv 1 \bmod n$. Una idea algo más elaborada, si n es primo y $n - 1 = 2^s t$, y considero los restos de b^t, b^{2t}, $\ldots$, $b^{2^s t}$ módulo n, el último de ellos diferente de 1 debe ser igual a -1. En otro caso, es seguro que n es compuesto.

El anterior es el famoso test de Miller-Rabin. Se sabe que, si la Hipótesis de Riemann generalizada es cierta, entonces, si n es compuesto, falla la prueba de Miller-Rabin para algún $b < 2(\log n)^2$. Luego, bajo la hipótesis anterior, tenemos un algoritmo rápido (es decir polinómico) para decidir si n es compuesto: Realizar el test de Miller-Rabin para todo $b < 2(\log n)^2$.

Otro motivo para plantear el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ es la existencia de los llamados *problemas* $\mathbf{NP}$ *completos*. Se llaman de este modo conjuntos $B \subset \mathbf{N}$ tales que B es de clase $\mathbf{NP}$ y, para los cuales si $B \in \mathbf{P}$, se seguiría que $\mathbf{P} = \mathbf{NP}$.

Desde los tiempos de Euclides los matemáticos tienen clara la noción de algoritmo. Turing dio un paso más, y mediante un esfuerzo de introspección, consiguió dar una definición precisa. Su imagen mental es la de un matemático, libreta en mano, calculando. Abstrayendo la situación, consigue dar con la idea de un ordenador moderno. A partir de la definición de Turing es posible cuantificar el tiempo que un ordenador tarda en realizar una determinada tarea y definir con precisión las clases $\mathbf{P}$ y $\mathbf{NP}$.

La primera conexión de la complejidad de los naturales con el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ consiste en el hecho de que $\mathbf{P} = \mathbf{NP}$ implica la existencia de un algoritmo rápido para calcular $\|n\|$. Para ciertas constantes C y $k \in \mathbf{N}$, existiría un algoritmo que decidiría el valor de $\|n\|$ en un tiempo $\leq C(\log n)^k$.

8. Complejidad de las funciones booleanas

Existe una segunda conexión, esta vez estructural, entre la complejidad de los naturales y el problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$. Para establecer esta conexión debemos definir otro concepto análogo, el de la complejidad de las funciones booleanas.

El conjunto $\{0,1\}$ tiene una estructura de cuerpo. Basta dotarlo de las operaciones suma y producto módulo 2. Para cada número natural n, consideremos el conjunto $\mathcal{F}_n$ de todas las aplicaciones $f\colon \{0,1\}^n \to \{0,1\}$. Podemos dotarlo de una estructura de anillo si tomamos el producto y la suma respecto de la estructura de cuerpo del conjunto de llegada $\{0,1\}$.

Por ejemplo, tenemos las funciones constantes $\mathbf{1}$, $\mathbf{0}$ y las componentes π_j definidas por ser $\pi_j(\mathbf{x}) = \pi_j(x_1, x_2, \ldots, x_n) = x_j$.

El álgebra $\mathcal{F}_n$ está generada por estas funciones. Es decir, podemos escribir cualquier función $f \in \mathcal{F}_n$ como un polinomio en las funciones anteriores. Para verlo, dado $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{0,1\}^n$, definimos la función $f_\varepsilon = \prod_j (\delta_j + \pi_j)$, donde, para cada j, $\delta_j = 1 + \varepsilon_j$. Entonces $f_\varepsilon(\mathbf{x}) = 0$, salvo para $\mathbf{x} = \varepsilon$. Así pues, para cualquier función g, podemos escribir

$$g = \sum_{\varepsilon \in S} f_\varepsilon,$$

donde S es el conjunto de ε para los que $g(\varepsilon) = 1$.

Igual que en el caso de los naturales, podemos definir la complejidad de los elementos de $\mathcal{F}_n$. Será la mayor función $f \mapsto \|f\|$ tal que

$$\|\mathbf{0}\| = \|\mathbf{1}\| = 0; \quad \|\pi_j\| = 1; \qquad \|f+g\| \le \|f\| + \|g\|; \quad \|fg\| \le \|f\| + \|g\|.$$

Para cada $\theta \in (0,1)$, la mayor parte de los elementos de $\mathcal{F}_n$ tienen complejidad $\ge 2^{\theta n}$. La demostración de este resultado se hace contando cuántos elementos tienen complejidad k, digamos que a_k. No es difícil ver que $a_0 = 2$, $a_1 = 2n$. Si $\|f\| = j$ y $\|g\| = k - j$ se obtienen, a lo sumo, cuatro elementos de complejidad $\le k$ que son $f+g$, fg, $1+f+g$ y $1+fg$. Con estas observaciones se deduce que

$$a_k \le 4(a_1 a_{k-1} + a_2 a_{k-2} + \cdots + a_{k-1} a_1).$$

De lo anterior se sigue entonces que $a_k \le A_k$, donde A_k está definida por

$$A_0 = 2; \quad A_1 = 2n; \quad A_k = 4 \sum_{j=1}^{k-1} A_j A_{k-j}.$$

De esta definición obtenemos

$$\sum_{k=0}^{\infty} A_k x^k = \frac{17 - \sqrt{1 - 32nx}}{8}; \qquad A_k = \frac{1}{2(2k-2)} \binom{2k-2}{k} (8n)^k.$$

Luego

$$a_k \leq A_k \sim \frac{2^{5k}}{8\sqrt{2\pi}k^{3/2}}n^k.$$

Deducimos que para x grande

$$\sum_{k=0}^{x} A_k \leq c\sum_{k=0}^{x}(32n)^k \leq c'(32n)^x \leq Ae^{Bx\log n},$$

luego si $x < 2^{\theta n}$, con $0 < \theta < 1$, se deduce que

$$\sum_{k=0}^{x} A_k \ll \mathrm{card}(\mathcal{F}_n) = 2^{2^n},$$

que prueba la aserción.

Cada construcción de $f(x_1, \ldots, x_n)$ como un polinomio permite deducir una expresión de la forma $\|f\| \leq a$. Pero de la expresión polinómica podemos deducir también algo más práctico: un circuito que permite calcular $f(x_1, \ldots, x_n)$ a partir de los datos de entrada x_j.

Como en el caso de los naturales, lo difícil es probar desigualdades del tipo $\|f\| > a$. De hecho la situación es sorprendente: hemos visto que, entre las funciones de n variables, lo usual es que la complejidad sea mayor que $2^{\theta n}$. Según esto, debiera ser fácil definir sucesiones *concretas* de funciones (f_n), donde f_n depende de n variables, y tales que $\|f_n\| \geq 2^{\theta n}$. Por el contrario, lo más que se ha prodido exhibir es $\|f_n\| > p(n)$, siendo p un polinomio de grado pequeño (Cfr. [Z], [H]). El problema aquí no es probar que *existen* sucesiones con $\|f_n\| > 2^{\theta n}$, que, como vemos, es relativamente simple, sino definir, explícitamente, una sucesión de funciones concreta para la que esto sea así. Cuando hablamos de "definir explicitamente" nos referimos a un concepto técnico que necesita una explicación. Debemos excluir soluciones fáciles como *sea f_n la primera función de n variables con complejidad máxima*. Consideramos que (f_n) está definida explícitamente si existe un algoritmo que determina el valor $f_n(x_1, \ldots, x_n)$ en un tiempo razonable.

El problema $\mathbf{P} \stackrel{?}{=} \mathbf{NP}$ nos lleva a considerar una sucesión especial de funciones booleanas. Sea a un número natural y consideremos $n = \binom{a}{2}$ el número de parejas. Nuestras variables van a ser

$$x_{12}, x_{13}, x_{23}, x_{14}, x_{2,4}, x_{34}, \ldots, x_{1a}, x_{2a}, \cdots, x_{a-1,a}.$$

De este modo, cada conjunto de valores de estas variables $\in \{0,1\}^n$ puede interpretarse como un grafo con a vértices, y donde $x_{jk} = 1$ si y sólo si los vértices j y k están unidos por una arista del grafo. Para cada $b \leq a$

sea entonces $f_b^a(x_{12},\dots,x_{a-1,a})$ la función que vale 1 si y sólo si existen b vértices tales que estén todos conectados en el grafo.

Cabe pensar que $\|f_b^a\| \geq \binom{a}{b}$, ya que para conocer el valor de f_b^a en un grafo necesitamos comprobar cada conjunto de b vértices. Se puede probar que, si esto es así, entonces $\mathbf{P} \neq \mathbf{NP}$. De este modo probar $\|f_b^a\| \geq \binom{a}{b}$ se convierte, a mi parecer, en el camino más prometedor de probar que $\mathbf{P} \neq \mathbf{NP}$.

Volviendo a la complejidad de los naturales, un problema análogo al anterior es el siguiente, planteado por Guy [G]:

Problema *¿Existe una sucesión de naturales (a_n) tales que*

$$\lim_{n\to\infty} \frac{\|a_n\|}{\log a_n} > \frac{3}{\log 3} \qquad ? \tag{1}$$

Un candidato es la sucesión 2^n. Todos los valores calculados hasta ahora cumplen $\|2^n\| = 2n$. Selfridge pregunta (cfr. [G]) si existe n tal que $\|2^n\| < 2n$.

Si para algún valor de n existiera k tal que $2^n = 3^k$, (cosa imposible por otra parte), la segunda expresión daría un valor de $\|2^n\| < 2n$. Naturalmente la ventaja sería mayor mientras mayor fuera n. Aunque lo anterior es imposible, no cabe descartar que se den otro tipo de casualidades que hagan posible $\|2^n\| < 2n$. Por ejemplo, si el desarrollo en base 3 de 2^n tuviera cifras con poco peso. De nuevo esto tiene pocas probabilidades de ocurrir; ahora bien, podría tratarse de otro tipo de expresión de 2^n. La situación aquí es que, si tengo un número que ya puedo expresar en la forma

$$(1+1)(1+1)\cdots(1+1),$$

parece dificil encontrar otra expresión que con menos unos conduzca al mismo resultado. Tenemos un casi-ejemplo trivial $4 = (1+1)(1+1) = 1+1+1+1$. Aquí aparece el mismo número de unos en ambos lados, por esto lo llamo casi-ejemplo. Pero pueden darse casi ejemplos no triviales, como el que sigue:

$$2^{27} = 1 + (1 + 2\cdot 3)(1 + 2^3\cdot 3^2)(1 + 2^9\cdot 3^3(1 + 2\cdot 3^2)).$$

Basta sustituir 2 por $1+1$ y 3 por $1+1+1$ para obtener una expresión alternativa de 2^{27} con 57 unos, y en la que la estructura multiplicativa del número 2^{27} deja de usarse.

La igualdad anterior prueba que $\|2^{27}-1\| \leq 56$. A pesar de una intensa búsqueda no he conseguido encontrar $n > 2$ tal que $\|2^n - 1\| < 2n-1$, sin embargo creo que esto puede ocurrir.

La evidencia parece estar del lado de que existe la sucesión que cumpla (1). Basta observar el gráfico en la figura 2. En él se ha situado un pequeño disco con centro en cada punto $(n, \|n\|)$ con $1 \leq n \leq 2000$ y también se han dibujado las gráficas de las curvas suaves que acotan a $\|n\|$, es decir $3(\log t)/\log 3$ y $3(\log t)/\log 2$, así como de la curva $5\log t/2\log 2$. Los puntos se unen y forman en la figura unas lineas paralelas al eje x. Vemos que la cota superior parece muy mala y que aparentemente $\|n\| \leq 5\log t/2\log 2$, cuando sólo hemos probado que aproximadamente esta desigualdad se cumple para casi todo $n \in \mathbf{N}$.

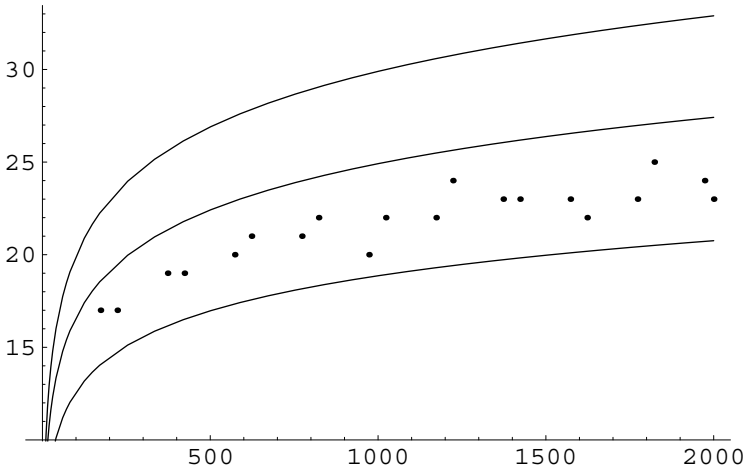

Figura 2

Pero esta figura nada dice sobre la existencia del límite $\lim \|n\|/\log n$, que es de lo que se trata. Sólo vemos que entre los 2000 primeros valores de n esta sucesión oscila entre unos límites próximos a $5/2\log 2$ y $3/\log 3$.

6. Conjeturas

He calculado mediante la proposición 1 la complejidad de los primeros 200.000 números naturales. Observando estos números, saltan a la vista ciertas regularidades. Las llamaré conjeturas sobre el comportamiento de la función $\|\cdot\|$, aunque no tengo mucha confianza de que se mantengan para números mayores.

Los orígenes de estas conjeturas son tablas como la que sigue:

3	6	9	12	15	18	21	24
10	100	1000	10000	100000	1000000	10000000	100000000
	22	220	2200	22000	220000	2200000	22000000
	21	210	2101	21010	210100	2101000	21010000
		202	2100	21000	210000	2100000	21000000
		201	2020	20200	202000	2020000	20200000
		$\overline{122}$	2010	20100	201000	2010000	20100000
			2002	20020	200222	2002220	20022200
			2001	20010	200200	2002000	20020000
			$\overline{1221}$	20002	200100	2001000	20010000
			1220	20001	200020	2000200	20002000
			1212	$\overline{12221}$	200010	2000100	20001000
			1211	12210	200002	2000020	20000200
			1201	12200	200001	2000010	20000100
			1122	12122	$\overline{122210}$	2000002	20000020
			1121	12120	122100	2000001	20000010
			1112	12111	122000	$\overline{1222100}$	20000002
				12110	121220	1221000	20000001
				12102	121200	1220000	$\overline{12221000}$
				12101	121121	1212200	12210000
				12012	121110	1212000	12200000
				12010	121100	1211210	12122000
				12001	121022	1211100	12121201
				11221	121020	1211000	12120000

En ella tenemos escritos en columna los números de complejidad $3n$ ($n = 1$, $2, \ldots, 8$), escritos en base 3 y ordenados de mayor a menor.

La primera observación: $\|3n\| = 3 + \|n\|$ es errónea. $\|107\| = 16$ y $\|321\| = \|1 + 2^6 5\| = 18$. Las que sí parecen ciertas son las siguientes conjeturas:

Conjetura 1 *Para cada número natural n, existe un entero $a \geq 0$ tal que $\|3^j n\| = 3(j-a) + \|3^a n\|$, para todo número natural $j \geq a$.*

Definimos el conjunto $A = \{n \in \mathbf{N} : \|3^j n\| = 3j + \|n\| \text{ para todo } j\}$.

Conjetura 2 *Para todo par de números naturales p y q, existe $a \geq 0$ tal que, para $j \geq a$, se tiene $\|p(q3^j + 1)\| = 3j + 1 + \|p\| + \|q\|$.*

Al observar la tabla anterior, vemos que los mayores números de complejidad $3n$ son los números naturales contenidos en la sucesión $(3^n a_n)$,

donde a_n viene dada por

$$1, \frac{2(3+1)}{3^2}, \frac{2^6}{3^4}, \frac{2\cdot 3+1}{3^2}, \frac{2(3^2+1)}{3^3}, \frac{2\cdot 3^2+1}{3^3}, \frac{2^9}{3^6}, \frac{2(3^3+1)}{3^4}, \frac{2\cdot 3^3+1}{3^4}, \ldots,$$

$$\ldots, \frac{2(3^k+1)}{3^{k+1}}, \frac{2\cdot 3^k+1}{3^{k+1}}, \ldots$$

Conjetura 3 *Existen tres sucesiones transfinitas de números racionales* $(a_\alpha)_{\alpha<\xi}$, $(b_\alpha)_{\alpha<\xi}$, $(c_\alpha)_{\alpha<\xi}$, *tales que los (mayores) números de complejidad* $3n$ *(respectivamente* $3n+1$, $3n+2$*) son los (primeros) números naturales contenidos en la sucesión* $(3^n a_\alpha)$, *(resp.* $(3^n b_\alpha)$, $(3^n c_\alpha)$*).*

ξ *es un ordinal numerable infinito tal que* $\omega\xi = \xi$.

Estas sucesiones comienzan del siguiente modo

(a_α), $1, \frac{8}{9}, \frac{64}{81}, \frac{7}{9}, \frac{20}{27}, \cdots \to \frac{2}{3}$ $\quad \frac{160}{243}, \frac{52}{81}, \cdots \to \frac{16}{27}$ $\quad \frac{1280}{2187}, \frac{140}{243}, \cdots \to \frac{5}{9} \ldots$

(b_α), $\frac{4}{3}, \frac{32}{27}, \frac{10}{9}, \frac{256}{243}, \frac{28}{27}, \cdots \to 1$ $\quad \frac{80}{81}, \frac{26}{27}, \cdots \to \frac{8}{9}$ $\quad \frac{640}{729}, \frac{70}{81}, \cdots \to \frac{64}{81} \ldots$

(c_α), $2, \frac{16}{9}, \frac{5}{3}, \frac{128}{81}, \frac{14}{9}, \cdots \to \frac{4}{3}$ $\quad \frac{320}{243}, \frac{35}{27}, \cdots \to \frac{32}{27}$ $\quad \frac{95}{81}, \frac{2560}{2187}, \cdots \to \frac{10}{9} \ldots$

donde los puntos suspensivos indican sucesiones infinitas, y los límites indicados no pertenecen a las sucesiones.

Conjetura 4 *Las tres sucesiones son decrecientes. Los denominadores de cada término* a_α, b_α *o* c_α *son potencias de* 3.

Conjetura 5 *Los números de la sucesión* (a_α), *son los números del conjunto*

$$\left\{ \frac{n}{3^{\|n\|/3}} : \|n\| \equiv 0 \bmod 3, \quad \text{y} \quad n \in A \right\},$$

ordenados en orden decreciente.

Conjetura 6 *Los números de la sucesión* (b_α), *son los números del conjunto*

$$\left\{ \frac{n}{3^{(\|n\|-1)/3}} : \|n\| \equiv 1 \bmod 3, \quad \text{y} \quad n \in A \right\},$$

ordenados en orden decreciente.

Conjetura 7 *Los números de la sucesión* (c_α), *son los números del conjunto*

$$\left\{ \frac{n}{3^{(\|n\|-2)/3}} : \|n\| \equiv 2 \bmod 3, \quad \text{y} \quad n \in A \right\},$$

ordenados en orden decreciente.

Lo que sigue es más tentativo y sólo está basado en unos pocos casos.

Conjetura 8 *Para todo ordinal $\beta < \xi$, se tiene*

$$\lim_{n\to\infty} a_{\beta\omega+n} = c_\beta/3, \quad \lim_{n\to\infty} b_{\beta\omega+n} = a_\beta, \quad \lim_{n\to\infty} c_{\beta\omega+n} = b_\beta.$$

Esta es la base para la afirmación sobre el valor de ξ, que parece debe ser al menos $\xi = \omega^\omega$, ya que es la menor solución de $\omega\xi = \xi$.

Las siguientes afirmaciones, junto con la conjetura 8, permiten predecir hasta cierto punto los valores de las sucesiones transfinitas.

Conjetura 9 *Los números de la sucesión $b_{\beta\omega+n}$ que converge a $a_\beta = b/3^a$ (con $\|b\| = 3a$) son números de las sucesiones*

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{donde} \quad b = pq,\ y,\ \|p(q3^j+1)\| = 3a+3j+1,$$

y aquellos términos esporádicos de la sucesión $2^{3j+2}/3^{2j+1}$ que estén contenidos entre $\sup_{\gamma<\beta} a_\gamma$ y a_β.

Conjetura 10 *Los números de la sucesión $c_{\beta\omega+n}$ que converge a $b_\beta = b/3^a$ (con $\|b\| = 3a+1$) son números de las sucesiones*

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{donde} \quad b = pq,\ y,\ \|p(q3^j+1)\| = 3a+3j+2,$$

y aquellos términos esporádicos de la sucesión $2^{3j+1}/3^{2j}$ que estén contenidos entre $\sup_{\gamma<\beta} a_\gamma$ y a_β.

Conjetura 11 *Los números de la sucesión $a_{\beta\omega+n}$ que converge a $c_\beta/3 = b/3^a$ (con $\|b\| = 3a-1$) son números de las sucesiones*

$$\frac{p(q3^j+1)}{3^{a+j}}, \quad \text{donde} \quad b = pq,\ y,\ \|p(q3^j+1)\| = 3a+3j,$$

y aquellos términos esporádicos de la sucesión $2^{3j}/3^{2j}$ que estén contenidos entre $\sup_{\gamma<\beta} a_\gamma$ y a_β.

En las conjeturas 9, 10 y 11 debe tenerse en cuenta que algunos términos provienen de sucesiones posteriores. Así, el término $c_\omega = 320/243$ es el término correspondiente a $j = 0$ de la sucesión $2^6(4\cdot 3^j+1)/3^{j+5}$, que converge a $b_3 = 256/243$.

Las anteriores conjeturas permiten predecir, por ejemplo, los 200 mayores números de complejidad 30.

Los números de complejidad 14 divididos por 81, son los números

$$
\begin{array}{llllllllllll}
c_0 = & \frac{162}{81}, & c_1 = & \frac{144}{81}, & c_2 = & \frac{135}{81}, & c_3 = & \frac{128}{81}, \\
c_4 = & \frac{126}{81}, & c_5 = & \frac{120}{81}, & c_6 = & \frac{117}{81}, & c_7 = & \frac{114}{81}, \\
c_9 = & \frac{112}{81}, & c_{10} = & \frac{111}{81}, & c_{11} = & \frac{110}{81}, & c_{13} = & \frac{109}{81}, \\
c_{\omega+1} = & \frac{105}{81}, & c_{\omega+2} = & \frac{104}{81}, & c_{\omega+3} = & \frac{102}{81}, & c_{\omega+6} = & \frac{100}{81}, \\
c_{\omega+8} = & \frac{99}{81}, & c_{\omega+10} = & \frac{98}{81}, & c_{\omega+14} = & \frac{97}{81}, & c_{2\omega} = & \frac{95}{81}, \\
c_{2\omega+3} = & \frac{93}{81}, & c_{2\omega+5} = & \frac{92}{81}, & c_{2\omega+8} = & \frac{91}{81}, & c_{3\omega+4} = & \frac{88}{81}, \\
c_{3\omega+7} = & \frac{87}{81}, & c_{3\omega+15} = & \frac{86}{81}, & c_{4\omega+2} = & \frac{85}{81}, & c_{5\omega+1} = & \frac{83}{81}, \\
c_{\omega^2+\omega+2} = & \frac{79}{81}, & c_{\omega^2+2\omega+3} = & \frac{77}{81}, & \frac{71}{81}, & \frac{69}{81}, & \frac{67}{81}, & \frac{59}{81},
\end{array}
$$

A los cuatro últimos no tengo suficientes datos para asignarles el ordinal correspondiente.

Bibliografía

[AS] ALON, N., SPENCER, J. H.: "The probabilistic method", *John Wiley and Sons, New York*, (1992)

[C] CHERNOFF, H.: "A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations", *Annals of Mathematical Statistics*, **23** (1952), 493—509

[GJ] GAREY, M. R., JOHNSON, D. S.: "Computers and Intractability, a guide to the theory of NP-completeness", *W. H. Freeman and Co.*, (1979)

[G] GUY, R. K.: "What is the least number of ones needed to represent n using only + and $\times$ (and parentheses)?", *American Mathematical Monthly*, **93** (1986), 189—190

[H] HASTAD, J. The Shrinkage exponent of de Morgan formulas is 2, *Siam J. Comput.* **27**, (1998), 48–64

[MP] MAHLER, K., POPKEN, P.: "On a maximum problem in arithmetic (Dutch)", *Nieuw. Arch. Wiskunde*, (3) **1** (1953), 1—15

[R] RAWSTHORNE, D. A.: "How many 1's are needed?", *Fibonacci Quart.*, **27** (1989), 14—17

[SP] SLOANE, N. J. A., PLOUFFE, S.: "The Encyclopedia of Integer Sequences", *Academic Press, London*, (1995)

[Z] ZWICK, U.: "A $4n$ lower bound on the combinatorial complexity of certain symmetric boolean functions over the basis of unate dyadic boolean functions", *Siam J. Comput.*, **20** (1991), 499—505

J. Arias de Reyna. Facultad de Matemáticas, Universidad de Sevilla.
P.O. Box 1160, 41080 Sevilla.
e-mail: arias@cica.es